Numerical Sensitivity Analysis on Anatomical Landmarks with regard to the Human Knee Joint

István Bíró[1], Béla M. Csizmadia[2], Gusztáv Fekete[3,4]

[1] Institute of Technology
Faculty of Engineering, University of Szeged
Mars tér 7, H-6724 Szeged, Hungary
biro-i@mk.u-szeged.hu

[2] Institute of Mechanics and Machinery
Faculty of Mechanical Engineering, Szent István University
Páter Károly utca 1, H-2100 Gödöllő, Hungary
csizmadia.bela@gek.szie.hu

[3] Research Academy of Grand Health
Ningbo University
Fenghua Road 818, 315000 Ningbo, China
feketegusztav@nbu.edu.cn

[4] Savaria Institute of Technology
Faculty of Natural and Technical Sciences, University of West Hungary
Károlyi Gáspár tér 4, H-9700 Szombathely, Hungary
fekete.gusztav@ttk.nyme.hu

Abstract: For determining kinematical landmarks of the human knee joint, anatomical conventions and reference frame conventions are used by most of research teams. Considering the irregular shapes of the femur and the tibia, the anatomical coordinate systems can be positioned with 1-2 mm and 2-4 degree position deflections. However, most of the anatomical landmarks do not appear as a dot, rather as a small surface. For this reason, the optical positioning of anatomical landmarks can only be achieved by additional 1-2 mm position deflection. It has already been proven by other authors that the application of reference frames with different positions and orientations cause significant differences in the obtained kinematics (specific anatomical landmarks and angles) of human knee joint. The goal of this research is to determine the relationship between certain anatomical landmarks and reference frames having different positions and orientations. The investigations were carried out on cadaver knees by means of actual measurements and numerical processing.

Keywords: sensitivity analysis; anatomical landmarks; knee joint; rotation; ad/abduction

1 Introduction

In the current literature several anatomical reference frames are applied by research teams [4, 15, 16, 17]. In spite of the general intention to standardize the position and orientation of the applied reference frames, as results of invasive and noninvasive investigations, the shape of the published kinematical diagrams are quite differing. This is not only a problem in the development of surgical robot systems [27, 28], but also in kinematic-based prosthesis design. It was verified by Pennock and Clark [11] that the different position and orientation of reference frames fastened to femur and tibia yields to significant diversity of kinematical diagrams.

To solve this problem a new reference frame convention was proposed by them. Various flexion axes have been used in the literature to describe knee joint kinematics. Among others, Most et al. [8] studied how two, widely accepted and used, flexion axes (transepicondylar axis (TEA) and the geometric center axis (GCE)) correlate with each other regarding the femoral translation and the tibial rotation. Their results suggested that kinematical calculation is sensitive to the selection of flexion axis.

Patel et al. [10] compared their own results, based on MRI images, with kinematical diagrams published by other research teams. They found similar and considerably different ones. In the study of Zavatsky et al. [14], both tibial-rotation and ad/abduction diagrams are quite different compared to the results published by other researchers. Similarly, significant difference appeared in the kinematical diagram related to the experimental study of Wilson et al. [12], who carried out tests on twelve different cadaver specimens. They assumed that the explanation of the diversity in the kinematical curves is due to the spatial orientation, the test rig design and the load.

In the presented research it will be examined how the different reference frames affect the relationship between the rotation, ad/abduction and the translation of certain anatomical landmarks on of cadaver specimens as a function of flexion angle. The investigated movement is slow knee flexion-extension, since the kinematics can be more precisely observed if the force ratios are lower than, e.g. in case of jump-down [26]. The changes in the kinematics, related to the anatomical angles and landmarks, were determined and plotted by systematic modification of some coordinates of the anatomical landmarks.

2 Methods

2.1 Cadaver Specimens

A uniquely designed and manufactured test rig, which was also equipped with a data acquisition system to track the motion, was used to perform the experiment [7]. After the experiments the sensitivity analysis was carried out. In the presented research, eight fresh frozen human cadaveric knee specimens (5 female knees, 3 male knees; average age 56±6 years; age range 49–63 with an average BMI of 25.21±3.8) were used for the kinematical investigation.

The specimens were stored at a temperature of 0-1 C° while storage time was between 4-6 days. Each of them was manually checked to ensure the active function arc of flexion/extension (0-120°) [18]. The lengths of the knee joint specimens were approximately 40-45 cm at the center of the capsule. After resection of the knee joint specimens, the motion investigations took place within one hour. A suitable quality of capsules and bones was assured by previous radiographic images. During preparation, the skin and soft tissues were removed while the joint capsule, ligaments and muscles were left intact.

2.2 Description of the Experiment

As a first step of the kinematical investigation, coordinates of anatomical landmarks, *fh, me, le, hf, tt, lm, mm* were recorded on the total cadaver body lying on its back (Figure 1), where the tripods of the Polaris optical tracking system [19] have already been attached. The accuracy of the system is 0.5 mm (volumetric: 0.25 mm RMS).

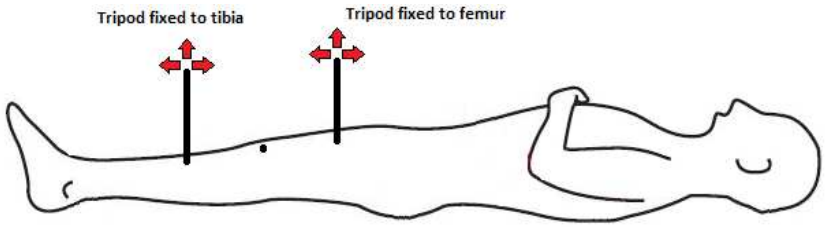


Figure 1
Anatomical landmarks on femur and tibia (right side) [6]

The anatomical landmarks are:

- Coordinates of centre of the femoral head (*fh*),
- Coordinates of medial and lateral epicondyles (*me, le*),

- Coordinates of apex of the head of the fibula (*hf*),
- Coordinates of prominence of the tibial tuberosity (*tt*),
- Coordinates of distal apex of the lateral and medial malleolus (*lm, mm*).

In addition, the following points must be defined:

- The origin (O_t) of the anatomical coordinate system, which is the midpoint of the junction-line between the medial (*me*) and lateral (*le*) epicondyles,
- The y_t axis of the coordinate system, which is the line between the origin and the center of the femoral head (*fh*), pointing upward with positive direction,
- The x_t axis of the coordinate system is perpendicular to the quasi-coronal plane, defined by the three anatomical points (*hf, me, le*). It has positive direction to the anterior plane,
- The z_t axis of the coordinate system is mutually perpendicular to the x_t and the y_t axis with positive direction to the right.

Figure 2 represents these landmarks and points.

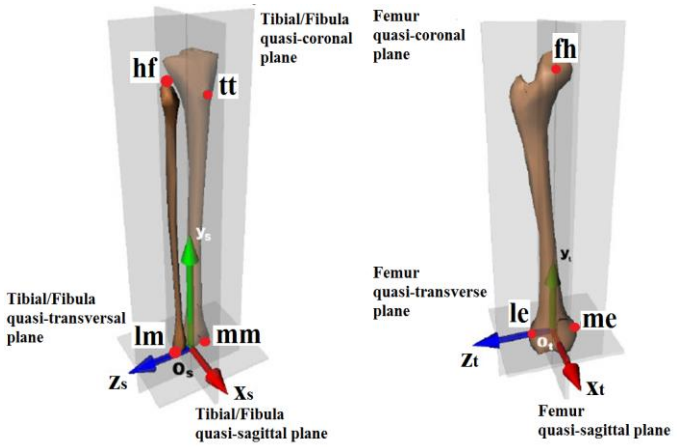


Figure 2
Anatomical landmarks on femur and tibia (right side) [6]

First of all, the coordinates of medial and lateral epicondyles (*me, le*) were pinpointed and recorded in the absolute coordinate system (XYZ) (Figure 3).

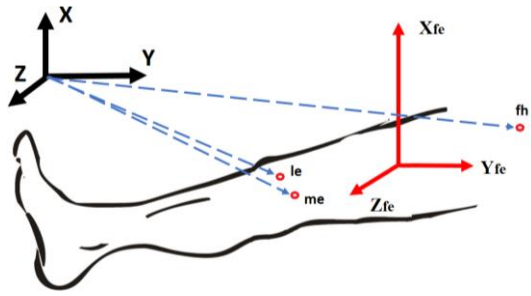


Figure 3
Pinpointing and recording anatomical landmarks on femur (right side)

Then by the use of a simple transformation (eq. (1) and eq. (2)) [25], the registered points (*le*, *me*, *fh*) can be transferred into the coordinate-system attached to the femur (X_{fe}-Y_{fe}-Z_{fe}) (Figure 4):

$$\mathbf{x}_a = \mathbf{T}^{-1} \cdot \mathbf{x}_r \tag{1}$$

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & T_{13} & 0 \\ T_{21} & T_{22} & T_{23} & 0 \\ T_{31} & T_{32} & T_{33} & 0 \\ X_O & Y_O & Z_O & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2}$$

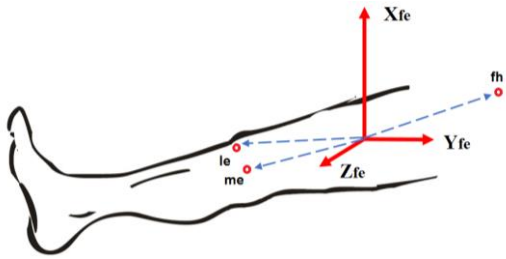


Figure
Pinpointing and recording anatomical landmarks on femur (right side)

This step was followed by circularly moving the thigh in order to determine the center of the femoral head (*fh*) in the absolute coordinate system (XYZ) (Figure 4). After this, the same transformation (eq. (1) and eq. (2)) was carried out to transform the femoral head (*fh*) into the X_{fe}-Y_{fe}-Z_{fe} coordinate system.

Four screws were attached to the femur, which represent the reference points. These reference points are pinpointed and recorded in the absolute coordinate system (XYZ) (Figure 5a). Then, they were similarly transformed into the X_{fe}-Y_{fe}-Z_{fe} coordinate system (Fig. 5b).

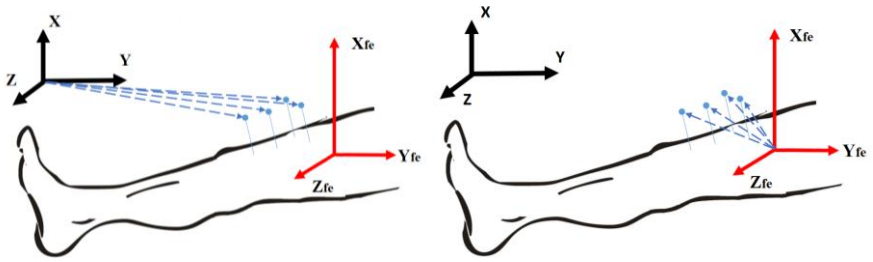


Figure 5a and Figure 5b
Pinpointing and transforming reference points on femur (right side)

Closely the same procedure must be carried out on the tibia as well. The coordinates of the head of the fibula (*hf*), the tibial tuberosity (*tt*) and the lateral and medial malleolus (*mm, lm*) were pinpointed and recorded in the absolute coordinate system (XYZ) (Figure 6).

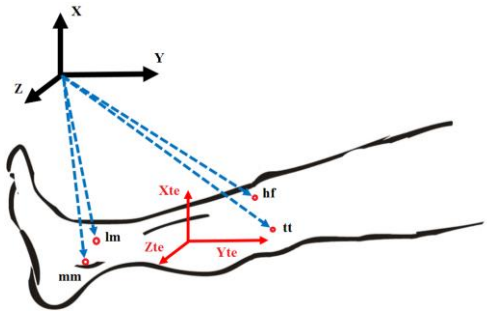


Figure 6
Pinpointing anatomical landmarks on tibia in the absolute (XYZ) coordinate system (right side)

Then again, by the use of a simple transformation (eq. (1) and eq. (2)) [25], the registered points (*hf*, *tt*, *lm*, *mm*) can be transferred into the coordinate-system attached to the tibia (X_{te}-Y_{te}-Z_{te}) (Figure 7):

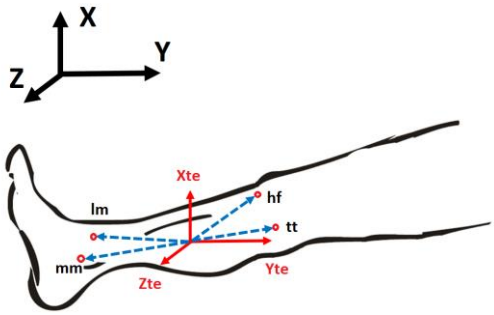


Figure 7
Pinpointing anatomical landmarks on tibia in the relative (*Xte, Yte, Zte*) coordinate system (right side)

The identification of the landmarks on the intact cadaver knee was followed by the resection of the cadaver specimens (Fig. 8). The flexion/extension movement on the cadaver specimens were carried out in the test rig.

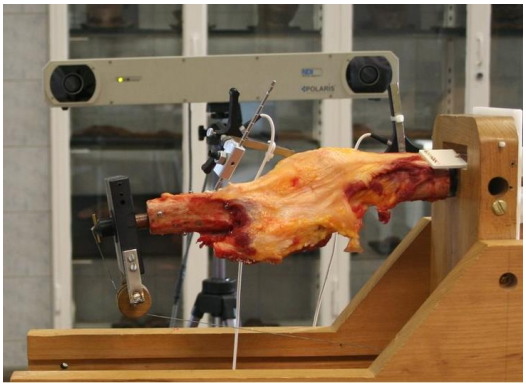


Figure 8
Cadaver specimen with two positioning sensors

During the measurement, the following conditions were kept: a) the knee joint carried out unconstrained motion; b) the test rig is equipped with a Polaris optical tracking system [19], which allows motion data acquisition under flexion/extension movement with a configuration of moving (rotating) tibia and fixed femur; c) motion data can be recorded in any arbitrary flexed position; d) the angular velocity of tibia related to femur is adjustable; e) the acting forces of the knee joint can be measured during flexion/extension; f) the tibia starts its motion by moving downwards from the extended position. The tibia is loaded in the sagittal plane of the knee joint. The loading force, acting on the end of tibia, enables the unconstrained flexion/extension motion while the test rig assures at least 120 degrees in flexion. Each specimen was manually flexed and extended five times before its actual installation into the test rig. Two plastic beams were fixed on both sides of the specimen. A rubber strip, representing the muscle model, was fixed to the quadriceps tendon of the cadaver knee joint. The trend of the quadriceps force function, regarding the magnitude of the force, is approximately linear up to a 70-80 degree of flexion angle [21, 22, 23, 24]. For this reason, a rubber strip was used to model the muscle, its characteristic was tested and showed linear behavior in the applied interval.

Anatomical conventions, reference frame convention and joint system convention of the VAKHUM project [6] were applied in the presented research. These conventions are based on current international standards (e.g. from the International Society of Biomechanics).

The main steps of the measurement are the followings [1, 2, 3, 7]:

Table 1
Steps of the experiment protocol

	Description
1	During the preparation of the cadaver knee joint, 4 + 4 screws are to be fixed into the femur and tibia. Screw heads are reference points in the course of measurement.
2	The position of the screw heads (as reference points) have to be recorded in the reference frames of the sensors.
3	The position sensors have to be attached to the lying cadaver body (femur and tibia).
4	The position of the anatomical landmarks (*le, me, tt, hf, mm, lm*) have to be recorded in the reference frames of the sensors.
5	The position of anatomical landmark *fh* in absolute coordinate-system of Polaris optical tracking system (pelvis is immovable) is to be determined by manually constrained thigh circle.
6	After the removal of the sensors, the resection has to be performed on the knee joint capsule.
7	The capsule has to be fixed into the test rig (Fig. 2).
8	The sensors have to be re-attached to both sides of the capsule (femur and tibia), obviously in a different position compared to the previous ones.
9	Data acquisition regarding the position of the 4 + 4 screw heads has to be repeated.
10	The position data of the anatomical landmarks (*le, me, tt*) in the coordinate-systems of sensors (in a modified position) has to be recorded.
11	The flexion/extension data of the knee joint can be continuously recorded by the sensors attached to the tibia and femur.

2.3 Description of the Parameters and Coordination Frames

Motion components of the knee joint were calculated according to the convention of Pennock and Clark [11]. The calculation is based on a three-cylindrical mechanism using Denavit-Hartenberg [5] parameters. In order to plot the kinematical diagrams, several reference frames were used such as the absolute reference frame of the Polaris optical tracking system, the reference frames of the sensors attached to the femur and the tibia, and the anatomical reference frames attached to the femur and the tibia (Fig. 8 and Fig. 9).

With the object to create the determining matrix-equation, the required anatomical landmarks on the femur and the tibia P_i *(i=1,2,.....,7)* are the following: femoral head (*fh):* P_1 (X_1,Y_1,Z_1), medial epycondylus *(me):* P_2 (X_2,Y_2,Z_2), lateral epycondylus *(le):* P_3 (X_3,Y_3,Z_3), lateral malleolus (*lm*): P_4 (X_4,Y_4,Z_4), medial malleolus (*mm*): P_5 (X_5,Y_5,Z_5), tibial tuberosity (*tt*): P_6 (X_6,Y_6,Z_6), apex of the head of the fibula (*hf*): P_7 (X_7,Y_7,Z_7). These points are appointed on Fig. 8 and Fig. 9.

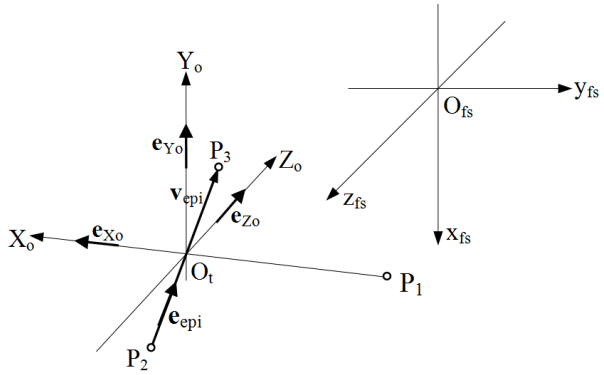


Figure 8
Position of anatomical reference frame $X_oY_oZ_o$ joined to femur in reference frame fs (cadaver lying on his back, investigated right leg, view from medial side)

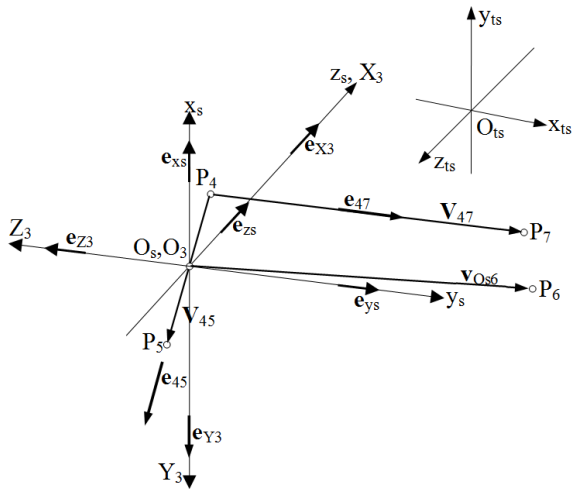


Figure 9
Position of anatomical reference frame $x_sy_sz_s$ and $X_3Y_3Z_3$ joined to tibia in reference frame ts (cadaver lying on his back, investigated right leg, view from medial side)

The mathematical relationship between the applied coordinate-systems can be described by the following matrix-equation:

$$[\mathbf{A}][\mathbf{B}][\mathbf{C}]^{-1}[\mathbf{D}]^{-1} = [\mathbf{E}]. \tag{3}$$

where:

[**A**]: fourth-order transformation matrix between reference frame *ts* (attached to tibia sensor) and anatomical reference frame $X_3Y_3Z_3$ fixed to tibia:

$$\mathrm{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} & 0 \\ A_{21} & A_{22} & A_{22} & 0 \\ A_{31} & A_{32} & A_{32} & 0 \\ A_{41} & A_{42} & A_{42} & 1 \end{bmatrix}. \tag{4}$$

Elements of matrix [**A**] and vector operations-notations to determine them:

$$N_3 = \sqrt{(X_5 - X_4)^2 + (Y_5 - Y_4)^2 + (Z_5 - Z_4)^2}$$

$$N_4 = \sqrt{(X_7 - X_4)^2 + (Y_7 - Y_4)^2 + (Z_7 - Z_4)^2}$$

$$N_5 = \sqrt{(2X_6 - X_4 - X_5)^2 + (2Y_6 - Y_4 - Y_5)^2 + (2Z_6 - Z_4 - Z_5)^2}$$

$$A_{21} = \frac{(Y_7 - Y_4)(Z_5 - Z_4) - (Y_5 - Y_4)(Z_7 - Z_4)}{N_3 N_4} = \frac{\alpha}{N_3 N_4}$$

$$A_{22} = -\frac{(X_7 - X_4)(Z_5 - Z_4) - (X_5 - X_4)(Z_7 - Z_4)}{N_3 N_4} = -\frac{\beta}{N_3 N_4}$$

$$A_{23} = \frac{(X_7 - X_4)(Y_5 - Y_4) - (X_5 - X_4)(Y_7 - Y_4)}{N_3 N_4} = \frac{\chi}{N_3 N_4}$$

$$A_{11} = \frac{(2Y_6 - Y_4 - Y_5)\chi + \beta(2Z_6 - Z_4 - Z_5)}{N_3 N_4 N_5}$$

$$A_{12} = -\frac{(2X_6 - X_4 - X_5)\chi - \alpha(2Z_6 - Z_4 - Z_5)}{N_3 N_4 N_5}$$

$$A_{13} = -\frac{(2X_6 - X_4 - X_5)\beta + \alpha(2Y_6 - Y_4 - Y_5)}{N_3 N_4 N_5}$$

$$A_{31} = -\frac{\begin{array}{c}((2X_6 - X_4 - X_5)\chi - \alpha(2Z_6 - Z_4 - Z_5))\chi \\ + \beta((2X_6 - X_4 - X_5)\beta + \alpha(2Y_6 - Y_4 - Y_5))\end{array}}{N_3^{\,2} N_4^{\,2} N_5}$$

$$A_{32} = -\frac{\begin{array}{c}((2Y_6 - Y_4 - Y_5)\chi + \beta(2Z_6 - Z_4 - Z_5))\chi \\ + \alpha((2X_6 - X_4 - X_5)\beta + \alpha(2Y_6 - Y_4 - Y_5))\end{array}}{N_3^{\,2} N_4^{\,2} N_5}$$

$$A_{33} = \frac{\begin{array}{c}((2Y_6 - Y_4 - Y_5)\chi + \beta(2Z_6 - Z_4 - Z_5))(-\beta) \\ + \alpha((2X_6 - X_4 - X_5)\chi - \alpha(2Z_6 - Z_4 - Z_5))\end{array}}{N_3^{\,2} N_4^{\,2} N_5}$$

$$[A_{41}; A_{42}; A_{43}] = \left[\frac{1}{2}(X_4 + X_5); \frac{1}{2}(Y_4 + Y_5); \frac{1}{2}(Z_4 + Z_5)\right]$$

[**B**]: fourth-order transformation matrix between the absolute reference frame and reference frame *ts* attached to tibia sensor:

$$B = \begin{bmatrix} B_{11} & B_{12} & B_{13} & 0 \\ B_{21} & B_{22} & B_{22} & 0 \\ B_{31} & B_{32} & B_{32} & 0 \\ B_{41} & B_{42} & B_{42} & 1 \end{bmatrix}. \tag{5}$$

Elements of matrix [**B**] in which O_{Xts}, O_{Yts}, O_{Zts}, are the coordinates of the origin of reference frame attached to tibia sensor in the absolute reference frame and Ψ_{ts}, Θ_{ts}, Φ_{ts} are Euler-angles between the same reference frames. These data were measured by Polaris optical tracking system.

$$\begin{bmatrix} B_{11} & B_{12} & B_{13} & 0 \\ B_{21} & B_{22} & B_{23} & 0 \\ B_{31} & B_{32} & B_{33} & 0 \\ B_{41} & B_{42} & B_{43} & 1 \end{bmatrix} = \begin{bmatrix} \cos\Theta_{ts}\cos\Psi_{ts} & \cos\Theta_{ts}\sin\Psi_{ts} & -\sin\Theta_{ts} & 0 \\ -\cos\Phi_{ts}\sin\Psi_{ts} + \sin\Phi_{ts}\sin\Theta_{ts}\cos\Psi_{ts} & \cos\Phi_{ts}\cos\Psi_{ts} + \sin\Phi_{ts}\sin\Theta_{ts}\sin\Psi_{ts} & \sin\Phi_{ts}\cos\Theta_{ts} & 0 \\ \sin\Phi_{ts}\sin\Psi_{ts} + \cos\Phi_{ts}\sin\Theta_{ts}\cos\Psi_{ts} & -\sin\Phi_{ts}\cos\Psi_{ts} + \cos\Phi_{ts}\sin\Theta_{ts}\sin\Psi_{ts} & \cos\Phi_{ts}\cos\Theta_{ts} & 0 \\ O_{Xts} & O_{Yts} & O_{Zts} & 1 \end{bmatrix}$$

[**C**]: fourth-order transformation matrix between the absolute reference frame and reference frame *fs* attached to femur sensor:

$$[C] = \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 \\ C_{21} & C_{22} & C_{22} & 0 \\ C_{31} & C_{32} & C_{32} & 0 \\ C_{41} & C_{42} & C_{42} & 1 \end{bmatrix}. \tag{6}$$

Elements of matrix [**C**] in which O_{Xfs}, O_{Yfs}, O_{Zfs}, are the coordinates of the origin of reference frame attached to femur sensor in the absolute reference frame and Ψ_{fs}, Θ_{fs}, Φ_{fs} are Euler-angles between the same reference frames. These data were measured by Polaris optical tracking system.

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 \\ C_{21} & C_{22} & C_{23} & 0 \\ C_{31} & C_{32} & C_{33} & 0 \\ C_{41} & C_{42} & C_{43} & 1 \end{bmatrix} = \begin{bmatrix} \cos\Theta_{fs}\cos\Psi_{fs} & \cos\Theta_{fs}\sin\Psi_{fs} & -\sin\Theta_{fs} & 0 \\ -\cos\Phi_{fs}\sin\Psi_{fs}+\sin\Phi_{fs}\sin\Theta_{fs}\cos\Psi_{fs} & \cos\Phi_{fs}\cos\Psi_{fs}+\sin\Phi_{fs}\sin\Theta_{fs}\sin\Psi_{fs} & \sin\Phi_{fs}\cos\Theta_{fs} & 0 \\ \sin\Phi_{fs}\sin\Psi_{fs}+\cos\Phi_{fs}\sin\Theta_{fs}\cos\Psi_{fs} & -\sin\Phi_{fs}\cos\Psi_{fs}+\cos\Phi_{fs}\sin\Theta_{fs}\sin\Psi_{fs} & \cos\Phi_{fs}\cos\Theta_{fs} & 0 \\ O_{Xfs} & O_{Yfs} & O_{Zfs} & 1 \end{bmatrix}$$

[**D**]: fourth-order transformation matrix between reference frame *fs* (attached to femur sensor) and anatomical reference frame $X_oY_oZ_o$ fixed to femur:

$$[D] = \begin{bmatrix} D_{11} & D_{12} & D_{13} & 0 \\ D_{21} & D_{22} & D_{22} & 0 \\ D_{31} & D_{32} & D_{32} & 0 \\ D_{41} & D_{42} & D_{42} & 1 \end{bmatrix}. \qquad (7)$$

Elements of matrix [**D**] and vector operations-notations to determine them:

$$N_1 = \sqrt{(X_2 + X_3 - 2X_1)^2 + (Y_2 + Y_3 - 2Y_1)^2 + (Z_2 + Z_3 - 2Z_1)^2}$$

$$[D_{11}; D_{12}; D_{13}] = \left[\frac{X_2 + X_3 - 2X_1}{N_1}; \frac{Y_2 + Y_3 - 2Y_1}{N_1}; \frac{Z_2 + Z_3 - 2Z_1}{N_1} \right]$$

$$N_2 = \sqrt{(X_3 - X_2)^2 + (Y_3 - Y_2)^2 + (Z_3 - Z_2)^2}$$

$$D_{21} = \frac{(Y_3 - Y_2)(Z_2 + Z_3 - 2Z_1) - (Y_2 + Y_3 - 2Y_1)(Z_3 - Z_2)}{N_1 N_2}$$

$$D_{22} = -\frac{(X_3 - X_2)(Z_2 + Z_3 - 2Z_1) - (X_2 + X_3 - 2X_1)(Z_3 - Z_2)}{N_1 N_2}$$

$$D_{23} = \frac{(X_3 - X_2)(Y_2 + Y_3 - 2Y_1) - (X_2 + X_3 - 2X_1)(Y_3 - Y_2)}{N_1 N_2}$$

$$D_{31}=\frac{(Y_2+Y_3-2Y_1)((X_3-X_2)(Y_2+Y_3-2Y_1)-(X_2+X_3-2X_1)(Y_3-Y_2))}{N_1{}^2N_2}+$$
$$+\frac{((X_3-X_2)(Z_2+Z_3-2Z_1)-(X_2+X_3-2X_1)(Z_3-Z_2))(Z_2+Z_3-2Z_1)}{N_1{}^2N_2}$$
$$D_{32}=-\frac{(X_2+X_3-2X_1)((X_3-X_2)(Y_2+Y_3-2Y_1)-(X_2+X_3-2X_1)(Y_3-Y_2))}{N_1{}^2N_2}+$$
$$+\frac{((Y_3-Y_2)(Z_2+Z_3-2Z_1)-(Y_2+Y_3-2Y_1)(Z_3-Z_2))(Z_2+Z_3-2Z_1)}{N_1{}^2N_2}$$
$$D_{33}=\frac{(X_2+X_3-2X_1)(-(X_3-X_2)(Z_2+Z_3-2Z_1)+(X_2+X_3-2X_1)(Z_3-Z_2))}{N_1{}^2N_2}-$$
$$-\frac{((Y_3-Y_2)(Z_2+Z_3-2Z_1)-(Y_2+Y_3-2Y_1)(Z_3-Z_2))(Y_2+Y_3-2Y_1)}{N_1{}^2N_2}$$
$$[D_{41};D_{42};D_{43}]=\left[\frac{1}{2}(X_2+X_3);\frac{1}{2}(Y_2+Y_3);\frac{1}{2}(Z_2+Z_3)\right]$$

[**E**]: fourth-order transformation matrix between reference frame $X_3Y_3Z_3$ and reference frame $X_oY_oZ_o$.

$$E=\begin{bmatrix}E_{11} & E_{12} & E_{13} & 0\\ E_{21} & E_{22} & E_{22} & 0\\ E_{31} & E_{32} & E_{32} & 0\\ E_{41} & E_{42} & E_{42} & 1\end{bmatrix}. \tag{8}$$

Elements of matrix [**E**] (Θ_1, Θ_2, Θ_3, d_1, d_2, d_3, are the obtained kinematical parameters of human knee joint).

$$E_{11}=\cos\Theta_1\cos\Theta_2\cos\Theta_3+\sin\Theta_1\sin\Theta_3;$$

$$E_{12}=\sin\Theta_1\cos\Theta_2\cos\Theta_3-\cos\Theta_1\sin\Theta_3;$$

$$E_{13}=\sin\Theta_2\cos\Theta_3;$$

$$E_{21}=-\cos\Theta_1\cos\Theta_2\sin\Theta_3+\sin\Theta_1\cos\Theta_3;$$

$$E_{22}=-\sin\Theta_1\cos\Theta_2\sin\Theta_3-\cos\Theta_1\cos\Theta_3;$$

$$E_{23}=-\sin\Theta_2\sin\Theta_3;$$

$$E_{31}=\cos\Theta_1\sin\Theta_2;\ E_{32}=\sin\Theta_1\sin\Theta_2;\ E_{33}=-\cos\Theta_2;$$

$$E_{41} = d_3 \cos\Theta_1 \sin\Theta_2 + d_2 \sin\Theta_1;$$
$$E_{42} = d_3 \sin\Theta_1 \sin\Theta_2 - d_2 \cos\Theta_1;$$
$$E_{43} = -d_3 \cos\Theta_2 + d_1.$$

Elements of matrix [**E**] contain the six kinematical parameters of human knee joint complying with the constraint conditions of the three-cylindrical mechanism model, which has been described by the Denavit-Hartenberg (later on HD) parameters – was used (Fig. 10). By the use of the HD parameters, the variables are reduced from six to four (Θ_i, d_i, l_i, α_i). These parameters fit well to the geometrical particularities of the applicable bodies and constraints [[5]].

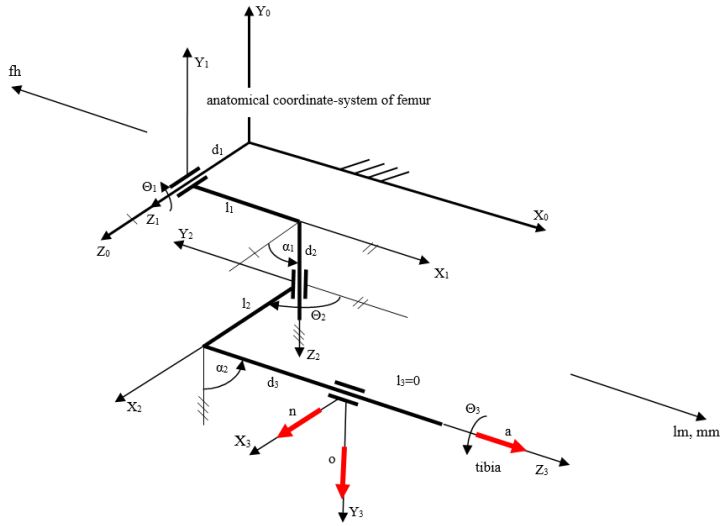


Figure 10
Human knee joint model in extended position [5]

In Fig. 10, the *HD* coordinates can be seen. The parameters of α_i, l_i, $(i=1,2,3)$ can be adjusted optionally according to the special geometry of knee joint. On the basis of the published recommendations [[11]] the following data is supported as a correct setting: $\alpha_1=\alpha_2=90^o$, $\alpha_3=0^o$, $l_1=l_2=l_3=0$. The application of the model enables the calculation of the following quantities (Fig. 10):

Θ_1 – flexion, plotted as *0* degree,

Θ_2 – ad/abduction, plotted as *90* degree,

Θ_3 – rotation of the tibia, plotted as *0* degree,

d_1, d_2, d_3 – moving on accordant axes.

2.4 Description of the Sensitivity Analysis

The aim of the numerical sensitivity analysis is to describe how the rotation, ad/abduction and translation of the (*fh, me, le, lm, mm, tt, hf*) depend on the position of the anatomical points therefore it also depends on the anatomical peculiarity of each subject.

This is also interpreted by the coordinate systems since they are defined by these anatomical points. The anatomical coordinate systems can be defined within a few mm including the deviation or error of the irregular form of the femur and tibia, and their anatomical peculiarities as well.

The detection of the position of the centre of femoral head (P_1), the apex of the head of the fibula (P_7) and prominence of the tibial tuberosity (P_6) cause only small angular deviation, since these anatomical landmarks are located relatively far from the origins of the coordinate systems (Fig. 3 and Fig. 4). The origins of the coordinate-systems are determined between the epicondyles and the apices of the lateral and medial malleolus, therefore the effect of position deflection is quite significant on the position and the orientation of the reference frame.

The numerical sensitivity analysis was carried out by a matrix-equation including 21 position coordinates of seven anatomical landmarks (three on the femur and four on the tibia). The position and orientation of the tibia related to the femur was changing step-by-step as a function of flexion angle.

At each step the position and the orientation of the tibia sensor changes therefore each time step another equation-system is generated.

By solving these equation-systems, three rotational and three translational parameters can be obtained, which determine the position and orientation of the tibia related to the femur.

As a first step of sensitivity analysis, the positions of epicondyles (points P_2 and P_3) were modified step-by-step *(±2 mm*) in the quasi-transverse plane in the opposite direction (cadaver lying on his back, extended position). In the second phase the positions of apices of the lateral and medial malleolus (points P_4 and P_5) were modified step-by-step (*±2 mm*) in quasi-transverse plane in the opposite direction as well. Difference compared to the basic functions (bold curve) are plotted in Fig. 10 and Fig. 11. The basis function contains some error since it was calculated from measured data.

3 Results

As we look at the group of curves created by systematic step-by-step position modification of the epicondyles, it is observable that in the first phase of flexion (until 40 degrees) the rotation-flexion curves are shifted parallel, while over 40 degrees their slopes are significantly different. The ad/abduction-flexion curves start from almost the same point, but the slope and the shape of the functions are quite divergent. The effect of the position modification of the epicondyles is negligible on the displacement components of the knee joint model (Fig. 11abc).

In case of systematic step-by-step position modification of the lateral and medial malleolus, the rotation-flexion curves are shifted parallel moreover the ad/abduction-flexion curves are nearly unchanged. The medio-lateral and antero-posterior translational curves are shifted parallel furthermore the proximal-distal translation along the tibial axis is quite unchanged (Fig. 12abc).

It is also apparent that the shapes of investigated kinematical diagrams depend considerably on the position and orientation of the reference frames fastened to femur and tibia. The shape and position of the kinematical curves were modified due to the different recorded position of the anatomical landmarks.

By the obtained new information, the kinematical results, carried out on different cadaver subjects, becomes comparable. Rotation-flexion and ad/abduction-flexion diagrams (Fig. 11a) are similar to the ones found in the literature.

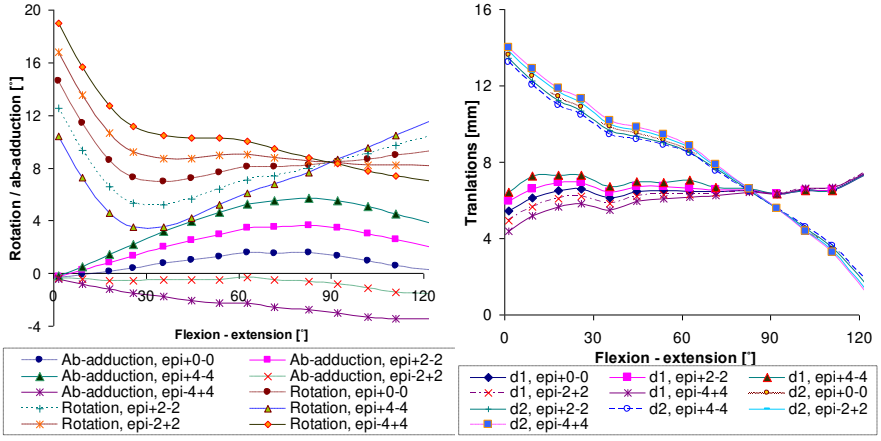


Figure 11a and Figure 11b
Coordinate modification of the epycondyles

Every single curve was plotted after systematic modification of the epicondyles position in the femur quasi-transverse plane.

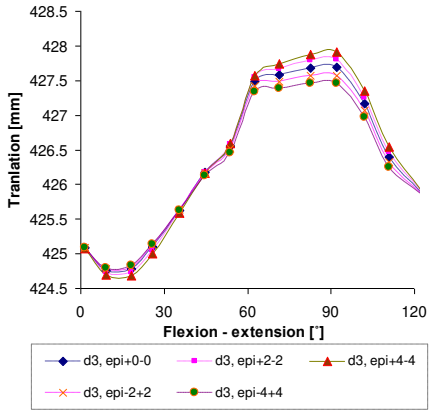


Figure 11c
Coordinate modification of the epycondyles

The curves point out well how the rotation-flexion and the ad/abduction-flexion diagrams depend on the position of the epicondyles. The modification of the condyle peaks did not influence significantly the shape of the translational diagrams (Fig. 11b-c).

As a second step, the effect of the coordinate modification of the malleolus is considered. As it is seen in Fig. 12a, the rotation-flexion curves are shifted parallel, due to the systematic position modification of the lateral and medial malleolus in the tibia quasi-transverse plane.

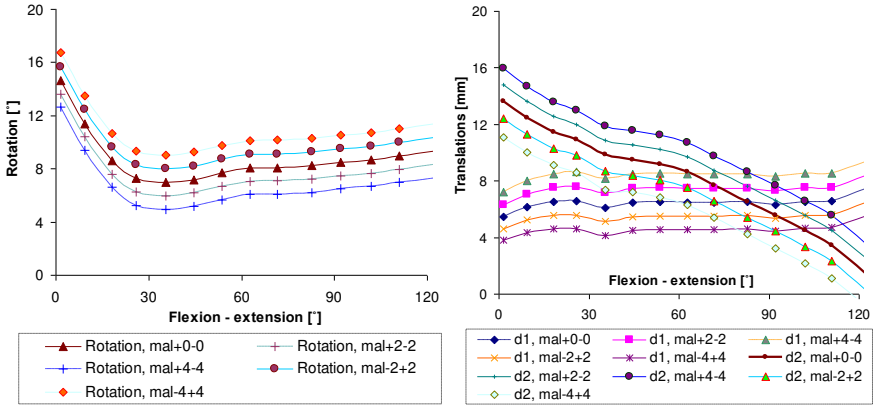


Figure 12a and Figure 12b
Coordinate modification of the malleolus

Translational curves *d1* and *d2* are shifted parallel (Fig. 12b), while the ad/abduction -flexion curves (Fig. 12c) and the curves *d3* are nearly the same (Fig. 11c).

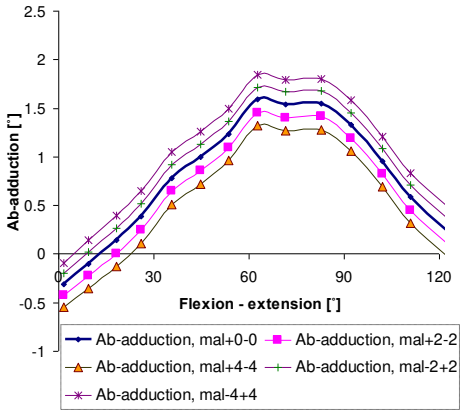


Figure 12c
Coordinate modification of the malleolus

On the basis of the obtained groups of curves, the analysis of other kinematical diagrams becomes possible regarding the anatomical features of the investigated person and the accuracy/inaccuracy of positioning of anatomical landmarks.

Conclusions

By the obtained new information, the kinematical results, carried out on different cadaver subjects, becomes comparable. Rotation-flexion and ad/abduction-flexion diagrams are similar to the ones found in the literature, nevertheless the presented measurement and processing methods, together with the sensitivity analysis, can quantitatively show how kinematical parameters of the knee joint (ad/abduction, rotation) are influenced by the coordinate systems fixed to the femur and tibia under flexion-extension. In this demonstration, every single curve was plotted after systematic modification of the epicondyles position in the femur quasi-transverse plane, thus the curves point out well how the rotation-flexion and the ad/abduction-flexion diagrams depend on the position of the epicondyles.The modification of the condyle peaks did not influence significantly the shape of the translational diagrams. As it is demonstrated, the rotation-flexion curves are shifted parallel, due to the systematic position modification of the lateral and medial malleolus in the tibia quasi-transverse plane. Translational curves *d1* and *d2* are also shifted parallel. The ad/abduction-flexion curves and the curves *d3* are nearly the same. On the basis of the obtained groups of curves, the analysis of other kinematical diagrams becomes possible regarding the anatomical features of the investigated person and the accuracy/inaccuracy of positioning of anatomical landmarks.

Acknowledgements

The presented research was supported by University of Szeged, Szent István University, University of West Hungary and by the Zhejiang Social Science Program – Zhi Jiang youth project (Project number: 16ZJQN021YB).

References

[1] Bíró, I., and Csizmadia, B. M.: Translational Motions in Human Knee Joint Model, *Review of Faculty of Engineering, Analecta Technica Szegedinensia*, **2-3**: 23-29 (2010)

[2] Bíró, I., Csizmadia, B. M., Katona, G.: New Approximation of Kinematical Analysis of Human Knee Joint, *Bulletin of Szent István University*, 330-338 (2008)

[3] Bíró, I., Csizmadia, B. M., Krakovits, G., Véha, A.: Sensitivity Investigation of Three-Cylinder Model of Human Knee Joint, *IUTAM Symposium on Dynamics Modelling and Interaction Control in Virtual and Real Environments*, Book Series **30**: 177-184, Budapest, 7-11 June (2010)

[4] Bull, A. M. J., Amis, A. A.: Knee Joint Motion: Description and Measurement," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, **212**: 357-372 (1998)

[5] Denavit, J., Hartenberg, R. S.: A Kinematic Notation for Lower-Pair Mechanism Based on Matrices, *ASME Journal of Applied Mechanics*, **22**: 215-221 (1955)

[6] http://www.ulb.ac.be/project/vakhum

[7] Katona, G., Csizmadia, B. M., Bíró, I., Andrónyi, K., Krakovits, G.: Motion Analysis of Human Cadaver Knee-Joints using Anatomical Coordinate System, *Biomechanica Hungarica*, **1**: 93-100 (2009)

[8] Most, E., Axe, J., Rubash, H., Li, G.: Sensitivity of the Knee Joint Kinematics Calculation to Selection of Flexion Axes, *Journal of Biomechanics* **37**: 1743-1748 (2004)

[9] Northern Digital Inc.: Polaris Optical Tracking System, Application Programmer's Interface Guide (1999)

[10] Patel, V. V., Hall, K., Ries, M., Lotz, J., Ozhinsky, E., Lindsey, C., Lu, Y., Majumdar, S.: A Three-Dimensional MRI Analysis of Knee Kinematics, *Journal of Orthopaedic Research* **22**: 283-292 (2004)

[11] Pennock, G. R., Clark, K. J.: An Anatomy-based Coordinate System for the Description of the Kinematic Displacements in the Human Knee, *Journal of Biomechanics* **23**: 1209-1218 (1990)

[12] Wilson, D. R., Feikes, J. D., Zavatsky, A. B., O'Connor, J. J.: The Components of Passive Knee Movement are Coupled to Flexion Angle, *Journal of Biomechanics* **33**: 465-473 (2000)

[13] Wu, G., Canavagh, P. R., ISB Recommendation for Standardization in the Reporting of Kinematic Data, *Journal of Biomechanics*, **28**: 1257-1261 (1995)

[14] Zavatsky, B., Oppold, P. T., Price, A. J.: Simultaneous In Vitro Measurement of Patellofemoral Kinematics and Forces, *Journal of Biomechanical Engineering*, **126**: 351-356 (2004)

[15] Stagni, R., Fantozzi, S., Capello A.: Propagation of Anatomical Landmark Displacement to Knee Kinematics: Performance of Single and Double Calibration, *Gait and Posture*, **24**: 137-141 (2006)

[16] Rapoff, A. J., Johnson, W. M., Venkataraman S., Transverse Plane Shear Test Fixture for Total Knee System, *Experimental Techniques*, **27**: 37-39 (2003)

[17] Fekete, G., De Baets, P., Wahab, M. A., Csizmadia M. B., Katona, G., Vanegas-Useche V. L., Solanilla, J. A.: Sliding-Rolling Ratio during Deep Squat with regard to Different Knee Prostheses, *Acta Polytechnica Hungarica*, **9** (5): 5-24 (2012)

[18] Freeman, M. A. R.: How the Knee Moves. *Current Orthopaedics*, **15**: 444-450 (2001)

[19] http://www.ndigital.com/medical/polarisfamily-techspecs.php

[20] Singerman, R., Berilla, J., Archdeacon, M., Peyser, A.: In Vitro Forces in the Normal and Cruciate Deficient Knee during Simulated Squatting Motion. *Journal of Biomechanical Engineering*, **121**: 234-242 (1999)

[21] Cohen, Z. A., Roglic, H., Grelsamer, R. P., Henry, J. H., Levine, W. N., Mow, V. C., Ateshian, G. A.: Patellofemoral Stresses during Open and Closed Kinetic Chain Exercises – An Analysis using Computer Simulation. *The American Journal of Sports Medicine*, **29**: 480-487 (2001)

[22] Mason, J. J., Leszko, F., Johnson, T., Komistek, R. D.: Patellofemoral Joint Forces. *Journal of Biomechanics*, **41**: 2337-2348 (2008)

[23] Fekete, G.: Fundamental Questions on the Patello- and Tibiofemoral Knee Joint: Modelling Methods related to Patello- and Tibiofemoral Kinetics and Sliding-Rolling Ratio under Squat Movement. Scholar's Press, Saarbrücken, Germany, 2013

[24] Fekete, G., Csizmadia, B. M., Wahab, M. A., De Baets, P., Vanegas-Useche, L. V., Bíró, I.: Patellofemoral Model of the Knee Joint under Non-Standard Squatting. *Dyna*, **81** (183): 60-67 (2014)

[25] Bíró, I., Fekete, G.: Approximate Method for Determining the Axis of Finite Rotation of Human Knee Joint, *Acta Polytechnica Hungarica*, **11** (9): 61-74 (2014)

[26] Melińska, A., Czamara, A., Szuba, L., Będziński, R., Klempous, R.: Balance Assessment during the Landing Phase of Jump-Down in Healthy Men and Male Patients after Anterior Cruciate Ligament Reconstruction. *Acta Polytechnica Hungarica*, **12** (6): 77-91 (2015)

[27] Takács, Á., Kovács, L., Rudas, I. J., Precup, R-E., Haidegger, T.: Models for Force Control in Telesurgical Robot Systems. *Acta Polytechnica Hungarica*, **12** (8): 95-114 (2015)

[28] Hošovský, A., Piteľ, J., Židek, K., Tóthová, M., Sárosi, J., Cveticanin, L.: Dynamic Characterization and Simulation of Two-Link Soft Robot Arm with Pneumatic Muscles. *Mechanism and Machine Theory*, **103**: 98-116 (2016)

Utilizing of Phase Shift Transformer for increasing of Total Transfer Capacity

Zsolt Čonka, Michal Kolcun, György Morva

Technical University of Košice, Department of Electric Power Engineering, Mäsiarska 74, 040 01 Košice, Slovakia

Óbuda University, Bécsi út 96/b, 1034 Budapest, Hungary, morva@uni-obuda.hu

Abstract: The Construction of new international power lines is a very expensive and time-consuming task. The use of specialized equipment that can increase the Total Transfer Capacity can replace the need to construct new lines. One such device is a Phase Shift Transformer (PST). By optimal distribution of power flow on transmission lines we could increase TTC on profile. Detailed analysis is presented for Slovakia – Hungary and Slovakia – Czech interconnections in the Central East Europe network. Simulation results with and without PST are used to analyse the impact of PST on the TTC between Slovakia – Hungary and Slovakia – Czech interconnection.

Keywords: phase shift transformer; total transfer capacity; FACTS

1 Introduction

In densely interconnected power systems such as ENTSO-E, the physical power flows differ from the planned power exchanges. This leads to overloaded interconnecting power lines. A possible solution could be the construction of new lines or the use of special equipment that allows us to regulate the power flow thus Net Transfer Capacity could be increased.

The power transmission capacity indicates how much power can be transferred between two power systems without endangering the systems' security. The perfect calculation of the transmission capacity is very important even for daily operation as well as on the electricity market. Both the operators and planners need to consider the maximal transfer capacity, system barriers and the amount of electricity that must be transferred.

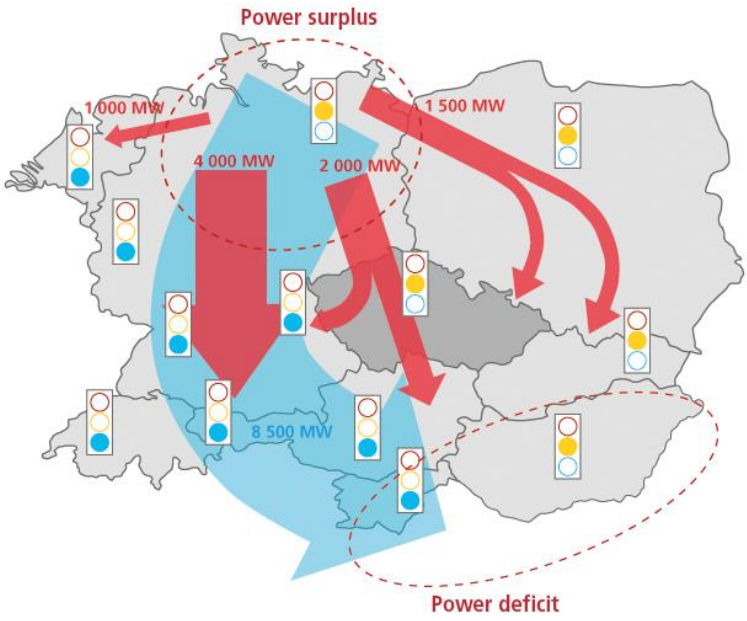


Figure 1
Power surplus in Central Eastern Europe's power system [13]

The calculation of transmission capacity should be repeated several times, in order to prevent inaccuracies in the calculation and thus prevent possible overloading of equipment, loss of stability or even the accidental damage of equipment. Inaccurate estimates of transfer capacity leads to the unnecessary reduction of the amount of transmitted power, and it is inefficient use of networks.

Amount and diversity of the regulatory returns are also increased by power transfer. In fact, the electricity market is enormously competitive. There is a very strong economic impulse to improve the accuracy and efficiency of the calculation of the transport capacity for use by system operators, designers and energy traders [1].

The difference between physical flows and programmed exchanges leads to overloaded interconnecting lines between Slovakia and Hungary and subsequently between Slovakia and the Czech Republic. It also leads to increasing the amount of regulatory returns. Figure 2 shows the power flow across Slovakia's power system. There are several ways to increase transfer capacity between the two power systems. One way is to build new power lines, but this is quite a time-consuming task. Another way is to utilize special devices which are especially designed to improve transfer capacity of existing power lines. One of the most promising devices, is the Phase Shift Transformer (PST).

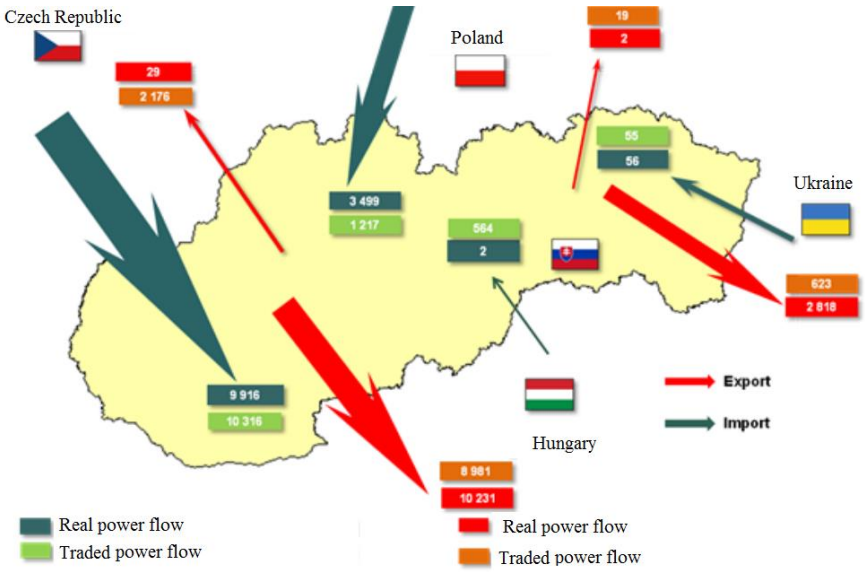


Figure 2
Power flow across Slovak Power system [1]

1.1 Total Transfer Capacity (TTC)

For establishing transmission margins and power transfer capacities between neighbouring power systems the following three terms are used:

Total Transfer Capacity (TTC) is the maximum exchange program between two areas compatible with operational security standards applicable at each system in future network conditions, generation and load patterns were known perfectly in advance.

Transmission Reliability Margin (TRM) is a safety margin that can handle uncertainties to calculate TTC values arising from:

1) Adverse deviations of physical flows during operation due to the physical load management of frequency,
2) Emergency exchanges between power systems to cope with unexpected unbalanced situations in real time,
3) Inaccuracies, faulty data collection.

Net Transfer Capacity (NTC) is defined as:

$$NTC = TTC - TRM \tag{1}$$

Net Transfer Capacity is the maximum amount of electricity that can be transferred which satisfies safety criteria N-1.

Total Transfer Capacity between two power systems determines the transmission system operators on both sides of the interconnection. The calculation of the transfer capacity includes more than just the capacities of the power lines. Due to the fact that the distribution of power flows between different components of TS, N-1 criterion must be fulfilled. N-1 criterion determines that the grid must remain in operation after unexpected loss of an unexpected element of the transmission system. The calculation of transfer capacity between two power systems is realized in both power systems' operators on each side of the interconnection. By using simulation programs, the maximal feasible amount of electricity that can be transmitted may be calculated. If the results of each TSO differ from each other, the lower value is used.

The aim of this process is to put power into the electricity market, as high capacity as is possible, while still observing transmission system limits. The ability of transferring power is calculated for each state of operating conditions. This process also allows not only exchanges between power systems but also transfers within each sub-system.

This process also includes transient and static simulations in order to determine maximal feasible amount of electricity in eight directions through interconnection with maintaining N-1 criterion, voltage collapse, and overcurrent. Figure 3 below shows cross-border power exchange between the power systems [1], [2], [3], [4], [5] [14].

Figure 3
Cross-border import/export capacity [13]

1.2 Phase Shift Transformer

Phase shift transformer (PST) is often used to control active power flow at the interface between two large and solid independent networks. The control of active power flow is achieved by adjusting the phase angle of the voltages at the terminals where the PST is installed. These are very complicated power transformers with several windings and junctions [6].

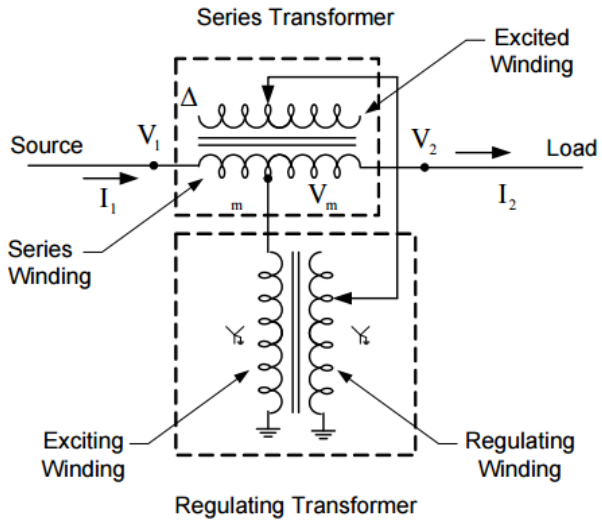


Figure 4
Symmetric Phase Shift Transformer (PST) [6]

Generally, phase shift transformers are built for a transmission system consists of two terminals and three-phase units. The source terminal is the terminal where the power is injected into the transformer and the terminal where the power flows out from transformer I called load terminal. Changes in phase angle between the terminal voltages are carried out by injecting additional voltage into entering voltage. When the series winding injects additional voltage in source terminal into neutral input voltage, the phase angle of voltages between input and output terminals of PST changes. As a result, the direction of power flow also changes. With an active control of injected amount of additional voltage, the power flow across the power line can be continuously controlled as well [6], [8], [9], [10].

1.3 TTC Calculation

The simulation below (Figure 5) contains four variants of interconnection between Slovakia and Czech Republic, three variants between Hungary and Slovakia and two variants between Slovakia and Hungary. In the enlarged power systems

400 kV interconnection lines are installed between Križovany (SK) and Gönyű (HU). All the possible variations were simulated on the 2014 Central Eastern Europe system model. Simulations were created by a simulation program namely: NEPLAN.

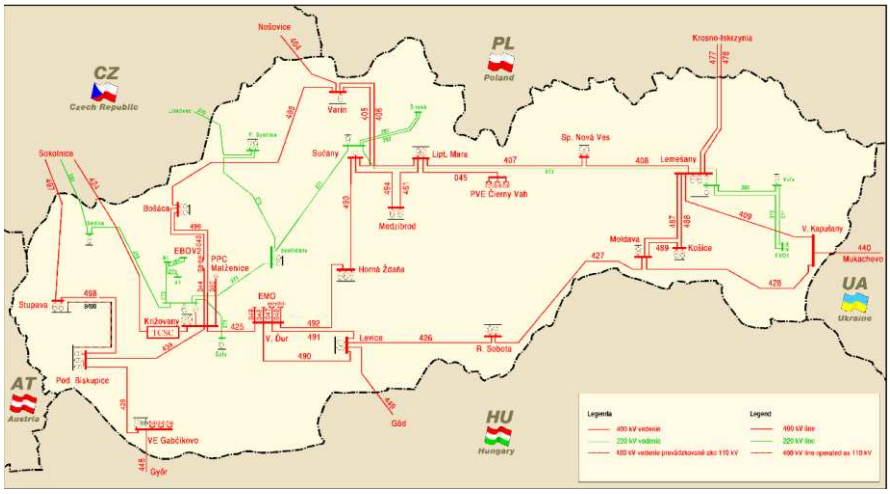


Figure 5
Slovakian power system for 2014

1.1.1 Variant SK-CZ I

Table 1 below shows the values of total transfer capacity on the Slovakian - Czech cross border connection. This model includes all generators, branches, transformers and other transmission system elements. Simulations for this variant were made without phase shift transformer. The model shows the actual values of transfer capacities between Slovakia and the Czech Republic. All simulations have been complied with safety criterion N-1. The limit element for this case was the line V 270, when line V 404 was switched off.

Table 1
TTC values for variant SK-CZ I

Branch	Line loads [MW]	TTC [MW]	Limit element (N-1)
404 (Nošovice – Varín)	-874	2527,4	270 (404)
424 (Sokolnice – Križovany)	-533,2		
497 (Sokolnice – Stupava)	-909,3		
280 (Sokolnice – Senica)	-100,9		
270 (Liskovec – P. Bystrica)	-110		

1.1.2 Variant SK-CZ II

Table 2 below shows the values of total transfer capacity on the Slovakian - Czech cross border connection with installed phase shift transformer on line V 404. In this case the simulations were created with phase shift transformer, and show the potential value of transfer capacities between Slovakia and the Czech Republic with PST installed on line V404. The simulations have been complied with safety criterion N-1. The limit element for this case was the line V 280, when line V 497 was switched off.

Table 2
TTC values for variant SK-CZ II

Branch	Line loads [MW]	TTC [MW]	Limit element (N-1)
404 (Nošovice – Varín)	-1340	2681,4	280 (497)
424 (Sokolnice – Križovany)	-342,9		
497 (Sokolnice – Stupava)	-658,2		
280 (Sokolnice – Senica)	-241		
270 (Liskovec – P. Bystrica)	-99,3		

1.1.3 Variant SK-CZ III

Table 3 below shows the values of total transfer capacity on the Slovakian - Czech cross border connection with installed phase shift transformer on line V 424. The simulations for this variant have been accomplished with phase shift transformer, and show the potential value of transfer capacities between Slovakia and the Czech Republic with PST installed on line V424. The simulations have been complied with safety criterion N-1. The limit element for this case was the line V 270, when line V 407 was switched off.

Table 3
TTC values for variant SK-CZ III

Branch	Line loads [MW]	TTC [MW]	Limit element (N-1)
404 (Nošovice – Varín)	-582,8	2698,4	270 (404)
424 (Sokolnice – Križovany)	-1298,4		
497 (Sokolnice – Stupava)	-494,2		
280 (Sokolnice – Senica)	-171		
270 (Liskovec – P. Bystrica)	-152		

1.1.4 Variant SK-CZ IV

Table 4 below shows the values of total transfer capacity on the Slovak - Czech cross border connection with installed phase shift transformer on line V 497. The simulations for this variant have been accomplished with phase shift transformer, and shows the potential value of transfer capacities between Slovakia and the Czech Republic with PST installed on line V497. Simulations have been complied with safety criterion N-1. The limit element for this case was the line V 270, when line V 404 was turned off.

Table 4
TTC values for variant SK-CZ IV

Branch	Line loads [MW]	TTC [MW]	Limit element (N-1)
404 (Nošovice – Varín)	-689	2727,1	270 (404)
424 (Sokolnice – Križovany)	-354,2		
497 (Sokolnice – Stupava)	-1311,2		
280 (Sokolnice – Senica)	-213,6		
270 (Liskovec – P. Bystrica)	-159,1		

1.1.5 Variant SK-HU I

Table 5 below shows the values of total transfer capacity on Slovakian - Hungarian cross border connection. This model includes all generators, branches, transformers and other transmission system elements. The simulations for this variant have been accomplished without phase shift transformer, and show the actual values of transfer capacities between Slovakia and Hungary. Simulations have been complied with safety criterion N-1. The limit element for this case was the line V 448, when line V 449 was switched off.

Table 5
TTC values for variant SK-HU I

Branch	Line loads [MW]	TTC [MW]	Limit element (N-1)
448 (Gabčíkovo-Györ)	1089,5	1946,7	448 (449)
449 (Levice-Göd)	857,2		

1.1.6 Variant SK-HU II

Table 6 below shows the values of total transfer capacity on Slovakian - Hungarian cross border connection with installed phase shift transformer on line V 449. Simulations for this variant have been accomplished with phase shift transformer, and show the potential value of transfer capacities between Slovakia

and Hungary with PST installed on line V449. The limit element for this case was the line V 490, when line V 491 was switched off.

Table 6
TTC values for variant SK-HU II

Branch	Line loads [MW]	TTC [MW]	Limit element (N-1)
448 (Gabčíkovo-Györ)	1089,5	2394	490 (491)
449 (Levice-Göd)	1304,5		

1.1.7 Variant SK-HU III

Table 7 below shows the values of total transfer capacity on Slovakian - Hungarian cross border connection with installed phase shift transformer on line V 449. The Simulations for this variant have been accomplished with phase shift transformer, and shows the potential value of transfer capacities between Slovakia and Hungary with PST installed on line V448. The limit element in this case was the line V 440, when line V 449 was turned off.

Table 7
TTC values for variant SK-HU III

Branch	Line loads [MW]	TTC [MW]	Limit element (N-1)
448 (Gabčíkovo-Györ)	1315,5	2511	440 (449)
449 (Levice-Göd)	1195,5		

1.1.8 Variant SK-HU IV

Table 8 below shows the values of total transfer capacity on Slovakian – Hungarian cross border connections without any installed phase shift transformer. The power system was enlarged by a new 400 kV line (V 400) from Križovany (Slovakia) and Gönyű (Hungary). Simulations for this variant have been accomplished without any phase shift transformer, and show the potential value of transfer capacities between Slovakia and Hungary with the new 400 kV line. The limit element for this case was the line V 440, when line V 449 was switched off.

Table 8
TTC values for variant SK-HU IV

Branch	Line loads [MW]	TTC [MW]	Limit element (N-1)
448 (Gabčíkovo-Györ)	762,5	2779,4	449 (440)
449 (Levice-Göd)	1226,8		
400 (Križovany-Gonyu)	790,1		

Table 9 below shows the values of total transfer capacity of enlarged Slovakian - Hungarian cross border connections with installed phase shift transformer on line 449. Simulations for this variant shows the potential value of transfer capacities between Slovakia and Hungary with a new 400 kV line and phase shift transformer. The limit element for this case was the line V 440, when line V 448 was switched off.

Table 9
TTC values for variant SK-HU V

Branch	Line loads [MW]	TTC [MW]	Limit element (N-1)
448 (Gabčíkovo-Györ)	873,3	3066,9	448 (440)
449 (Levice-Göd)	1329,7		
400 (Križovany-Gonyu)	863,9		

Conclusions

The simulations' results suggest new ways of increasing transfer capacity between Slovakia and Hungary as well as between Slovakia and Czech Republic. The results also show that the most suitable location of a phase shift transformer would be in the Slovakian power system. The best place for installing a PST transformer on cross-border connection would be between Slovakia and the Czech Republic, on line 497. The value of total transfer capacity with PST installed on line 497 was 2727,1 MW. The total transfer capacity with PST installed on line 424 was 2698,4 MW, and the value of total transfer capacity with phase shift transformer on line 404 was 2681,4 MW.

The most suitable location for installing a phase shift transformer on inter-connection between Slovakia and Hungary was on line V448. The value of total transfer capacity with a phase shift transformer installed on line 448 was 2511 MW. The total transfer capacity with PST installed on line 449 was 2394 MW.

Enlarging the power system by adding an interconnection line between Slovakia and Hungary has caused an increase of TTC from 1946,7 MW to 2779,4 MW. By installing PST in series with line 449 has increased TTC up to 3066,9 MW.

Acknowledgement

This work was supported by the Slovakian Research Agency No. VEGA 1/0388/13 project.

References

[1] I. Dobson, S. Green, R. Rajaraman, Ch. L. DeMarco, F. L. Alvarado, R. Zimmerman.: Electric Power Transfer Capability: Concepts, Applications, Sensitivity, Uncertainty, Power Systems Engineering Research Center November 2001 online [12.05.2015] <http://www.pserc.cornell.edu/tcc/tutorial/TCC_Tutorial.pdf>

[2] Calculation methods online [12.05.2015] <http://www.elia.be/en/products-and-services/cross-border-mechanisms/transmission-capacity-at-borders/calculation-methods>

[3] Ministry of Economy of Slovak Republic: Návrh Energetickej politiky Slovenskej republiky, May 2014 Online: cit.[19. feb. 2015] <http://www.rokovania.sk/File.aspx/ViewDocumentHtml/ Mater-Dokum-165253?prefixFile=m_>

[4] FAQ CEPS, a.s. online [19. Feb. 2015] <http://www.ceps.cz/ENG/Media/Pages/FAQ.aspx>

[5] Definitions of Transfer capacities in liberalised Electricity Markets online [19. Feb. 2015] <https://www.entsoe.eu/fileadmin/user_upload/_library/ntc/entsoe_transfer CapacityDefinitions.pdf>

[6] C. Grande-Moran: Phase-shifting Transformer Modeling in PSS®E online [12.05.2015] <http://ewh.ieee.org/soc/pes/newyork/NewSite/PDFs/Protection%20Related/Phase_Shifting_Transformer_Modeling.pdf>

[7] ENTSO-E, Network Code on Capacity Allocation and Congestion Management, September 2012

[8] ENTSO-E, Network Code on Forward Capacity Allocation, October 2013

[9] K. Noháč, L. Noháčová: Detekce nestandardních stavů v sítích s obnovitelnými zdroji energie“. Proceedings of 7th International scientific symposium Elektroenergetika 2013, September 18-20, 2013, Stará Lesná, Slovakia

[10] M. Mešter, M. Hvizdoš : The fundamentals of wide area monitoring in power system. In: Energetika un elektrotehnika, 2005, Vol. 15, No. 4, pp. 110-115, ISSN 1407-7345

[11] R. Jakubčák, Ľ. Beňa, M. Kmec: Possibilities of Using Facts Devices In Power System. In: Acta Electrotechnica et Informatica. Roč. 13, č. 3 (2013), s. 8-11. - ISSN 1335-8243

[12] Program rozvoja SEPS, a.s., na roky 2011 – 2020. [online]. [cit. 08.08.2011]. Available on internet: http://www.sepsas.sk/seps/Dokumenty/PR2020.pdf

[13] FAQ CEPS, a.s. online [19. Feb. 2015] http://www.ceps.cz/ENG/Media/Pages/FAQ.aspx

[14] Z. Čonka, M. Kolcun: Impact of TCSC on the Transient Stability In: Acta Electrotechnica et Informatica. Roč. 13, č. 2 (2013), s. 50-54. - ISSN 1335-8243 Available on internet :www.versita.com/aei. (2013)

A Survey on Hardware and Software Solutions for Multimodal Wearable Assistive Devices Targeting the Visually Impaired

Ádám Csapó[1], György Wersényi[2], and Myounghoon Jeon[3]

[1]Széchenyi István Univesity, Department of Informatics, Egyetem tér 1, H-9026 Győr, Hungary (e-mail: csapo.adam@sze.hu)

[2]Széchenyi István Univesity, Department of Telecommunications, Egyetem tér 1, H-9026 Győr, Hungary (e-mail: wersenyi@sze.hu)

[3]Michigan Technological University, Department of Cognitive and Learning Sciences and the Department of Computer Science, Houghton, Harold Meese Center, Houghton, MI 49931 USA (e-mail: mjeon@mtu.edu)

Abstract: The market penetration of user-centric assistive devices has rapidly increased in the past decades. Growth in computational power, accessibility, and "cognitive" device capabilities have been accompanied by significant reductions in weight, size, and price, as a result of which mobile and wearable equipment are becoming part of our everyday life. In this context, a key focus of development has been on rehabilitation engineering and on developing assistive technologies targeting people with various disabilities, including hearing loss, visual impairments and others. Applications range from simple health monitoring such as sport activity trackers, through medical applications including sensory (e.g. hearing) aids and real-time monitoring of life functions, to task-oriented tools such as navigational devices for the blind. This paper provides an overview of recent trends in software and hardware-based signal processing relevant to the development of wearable assistive solutions.

Keywords: assistive technology; blind user; haptics; spatial rendering; sonification

1 Introduction

The first (assistive) wearable devices developed in the 1990s tried to address the basic needs of specific target groups. Particular devices were primarily considered as medical devices incorporating the basic functionalities of sensors and actuators (e.g. microphones, amplifiers, vibro-/electrotactile actuators) in order to complement human sensorimotor capabilities.

In the case of visually impaired users, the ability to travel in both familiar and unfamiliar [1, 2], as well as indoor and outdoor [3, 4] environments is crucial. Such a large variety of contexts creates a myriad of challenges, both in terms of functionality and safety. Key issues include avoidance of obstacles, accessibility to salient points of the environment (e.g. finding doors, obtaining key notifications from signs and other information displays), and even co-presence (participating to the fullest possible extent in social interactions that occur within the environment).

However, given that the processing of visual information requires powerful sensory equipment as well as significant resources for processing incoming data, the earliest electronic travel aids (ETAs) were quite bulky and processed information at relatively low resolutions. Significant changes to this state of affairs only became possible in the 21st Century, in parallel with exponential reduction in the size, weight, and processing capacity of both dedicated electronic aids and mobile devices such as smartphones and tablets. Even today, the widespread adoption of SSDs (sensory substitution devices) outside of academia seems to face steep challenges. As suggested by Elli et al., this may be due to the fact that effectively coordinating issues arising from the different perspectives of ergonomics, neuroscience, and social psychology is a tricky business [5].

Even with these challenges, taking into account that visually impaired users have at best very basic visual capabilities through residual vision, the role of both substitutive and augmentative feedback through the senses of audition and taction is widely seen as essential. While these senses can be used both separately and in parallel, the ways in which information is allocated to them, and the ways in which information is presented (i.e. filtered, processed, and displayed) to users cannot be arbitrary. While it is often said (but sometimes questioned) that healthy individuals process up to 90% of information of the environment through vision [6], even the information obtained through this single modality is heterogeneous. Thus, both 2D and 3D information are acquired in foveal and peripheral parts of the retina and at different degrees of attention; not all events and objects are attended to directly - even if they reach consciousness to some extent - and it is important that the structure of such details be reflected in the structure of substitutive and augmentative feedback.

At the same time, it is important to take into consideration how the modalities of audition and taction are fundamentally different from vision - at least in the ways we use them. Users are able to focus through these modalities less acutely, and the spatial resolution of incoming information is also lower than vision. With audition, perception (or rather, estimation) of distance is based partially on sound pressure level, which is very inaccurate [7, 8]. Presenting multiple (concurrent) audio sources result in increased cognitive load, reduced localization, and reduced identification accuracy. As a result, mapping 3D visual information to pure auditory cues has had limited success. The crucial realization is that while the visual modality would be capable of effectively suppressing irrelevant information, in the case of auditory modality this kind of suppression has to be

carried out by design, prior to auditory mapping and rendering through sonification and directional stimulation. With haptics, many of the conclusions are similar to audition. Here, the exclusion of visual feedback also results in a relative strengthening of substance parameters (in this case, hardness and texture) and a weakening of shape salience [9]. This tendency can be countered by appropriate "exploratory procedures", in which subjects move their hands across the given surface in specific ways to obtain relevant information [10, 11]. While such procedures can potentially be richer in the case of haptic as opposed to auditory exploration, enabling users to capitalize on such procedures requires that the structure of the stimulus be assembled in appropriate ways, and whenever the stimuli are temporal, parameters such as burst frequency, number of pulses per burst, pulse repetition frequency, waveform shape and localization also need to be taken into consideration [12]. The general conclusion is that in both cases of audition and haptic/tactile feedback, results are optimal if a selective filtering and feature encoding process precedes the mapping and rendering phases.

The following sections of this paper focus on representations and signal processing techniques associated with the auditory and haptic / tactile modalities, respectively. Finally, an application-oriented section concludes the paper.

2 Audio

Sound is the most important feedback channel for the visually impaired. Aspects of sound such as reverberation (echolocation), pressure, and timbre convey useful information about the environment for localization, navigation, and general expectations about the surrounding space. In fact, in the case of healthy individuals, localization of sound sources is the most important part in identifying hazardous objects, obstacles, or even the “free path” to navigate through [13]. Both ecologically motivated and abstract (artificial) sounds can deliver crucial supplementary information on real life, and increasingly virtual reality situations.

Auditory Localization

For localizing sound sources, the auditory system uses both monaural (using one ear) and interaural (using two ears) cues [14]. Monaural cues are essentially responsible for distance perception and localization in the median plane (elevation perception). For example, distant sounds are usually softer and sounds that gradually grow louder are perceived as approaching. Interaural cues are based on the time, intensity, and phase differences between the two ears in the case of sound sources outside the median plane. Interaural Time Differences (ITD) mean that the arrival times of the sounds to the two ears differ with respect to each other. Interaural Level Differences (ILD) mean that the sound intensities perceived by the two ears differ due to the “head shadow”, i.e., due to the shape of

the listener's head blocking certain high-frequency sound components. Interaural Phase Differences (IPD) are also due to the shape of the head, which alters the phases of the sounds reaching the two ears. These mechanisms are better suited towards the identification of horizontal directions. For all of these phenomena, the hearing system uses the filtering effects of the outer ears, head, and torso. These response characteristics are called Head-Related Transfer Functions (HRTFs) and they can be measured, stored, and used for reverse engineering usually in the form of digital IIR and/or FIR filters [15-18].

Actual localization performance of human subjects depends on various other factors as well, such as:

- real-life environment vs. virtual simulation,
- training and familiarity with the sounds,
- type (bandwidth, length) and number of sound sources,
- spatial resolution and overall accuracy of the applied HRTFs (if any),
- head-tracking,
- other spatial cues (reverberation, early reverb ratio, etc.).

General findings show (1) decreased localization performance in the vertical directions in contrast to horizontal plane sources, (2) preference for individually measured HRTFs during binaural playback, and (3) increased error rates when using headphones [19-24].

Methods for Directional Simulation

Several methods exist for the spatialization of sound sources. As distances are generally mapped to sound level only, the localization task using directional information is usually associated with sources that are at a constant distance. Hence, localization has to be tested both in the horizontal and vertical planes.

Sounds can be played back over loudspeakers or headphones. Loudspeaker systems incorporate at least two channels (stereo panning, panorama stereo), but otherwise can range from traditional 5.1 multi-channel systems to ones with up to hundreds of speakers. Headphones are generally two-channel playback systems using electrodynamic transducers for airborne conduction. Traditional open or closed type headphones (especially, if they are individually free-field equalized [25-28]) are the best solutions. However, if they cover the entire ears and block the outside world, blind users will refrain from using them. Therefore, loudspeaker solutions are not applicable for mobile and wearable applications. Other types of headphones exist such as multi-speakers, multi-channel (5.1) solutions, partly covering the ears, or bone conduction phones. People are likely to think that the latter technology may provide lossy information compared to everyday air conduction phones; however, research has shown that virtual three-

dimensional auditory displays can also be delivered through bone-conduction transducers using digital signal processing, without increased percept variability or decreased lateralization [29, 30].

Signal processing solutions for spatial simulation include the following methods:

- Amplitude panning (panorama). Simple panning between the channels will result in virtual sources at a given direction. Mono is not suitable for delivering directional information. Stereo panning is limited for correct sound source positioning between the two speakers. The classical setup is a 60-degree triangle of listener and loudspeakers. Amplitude panning can be used for two-channel headphone playback, not necessarily limited to ±30 degrees, but it can be up to ±90 degrees.

- HRTF filtering is for binaural playback over headphones. HRTFs are stored in a form of digital filters in a given number of length (taps, filter coefficients) and spatial resolution (number of filters in the horizontal and vertical plane). Real-time filtering or pre-filtered pre-recorded samples are needed, together with some kind of interpolation for missing filters (directions). Although this is a commonly applied method, localization performance is sometimes low and errors, such as front-back-confusions in-the-head localization and others influence the performance [31-33].

- Wave-Field Synthesis (WFS) incorporates a large number of individually driven speakers and it is not designed for wearable applications [34]. The computational load is very high, but the localization does not depend on or change with the listener's position.

- Ambisonics uses a full sphere of loudspeakers, not just in the horizontal plane but also in the vertical plane (above and below the listener). It is not the traditional multi-channel system and the signals are not dedicated speaker signals. They contain speaker-independent representation of a sound field that has to be decoded to the actual speaker setup. Its advantage is that the focus is on source direction instead of loudspeaker positions and it can be applied to various setups. However, using multiple speakers and high signal processing makes it unavailable in wearable devices [35].

Further Aspects of Auditory Representation

Besides directional information, a number of mapping strategies have been applied to wearable assistive devices making it possible to communicate bits of information with semantics that lie outside of the scope of auditory perception.

Traditional auditory cues

Auditory icons were defined by Gaver in the context of 'everyday listening' - meaning that one listens to the information behind the sound as a “caricature" of physical-digital phenomena [36-38]. This was the first generalization of David

Canfield-Smith's original visual icon concept [39] in modalities other than vision. Around the same time, earcons were defined by Blattner, Sumikawa, and Greenberg as "non-verbal audio messages used in the user interfaces to provide users with information about computer objects, operation, or interaction". Today, the term is used exclusively in the context of 'abstract' (rather than 'representational', see [40]) messages, i.e., as a concept that is complementary to the iconic nature of auditory icons.

Sonification

Whenever a data-oriented perspective is preferred, as in transferring data to audio, the term 'sonification' is used, which refers to the "use of non-speech audio to convey information or perceptualize data" [41] (for a more recent definition, the reader is referred to [42]).

Sonification is a widely used term to cover applications of sound as information. However, spatial simulation and directional information is not generally part of sonification: such techniques are generally applied separately. From a visual perspective, sonification focuses on finding an appropriate mapping between visual events and auditory counterparts, that is, how certain visual objects or events should sound like. Several methods have been proposed for finding such mappings based on the conceptual structure of the problem domain (e.g. [43]). For example, Walker used magnitude estimation to identify optimal sonification for diverse tasks with sighted and visually impaired users [44]. Following the specification of an appropriate mapping, the parameters of the individual sound events may also be modified through time, based on real-time changes in the physical attributes of the visual objects or environment that is represented. Whenever the semantic content to be mapped (e.g., an image in the case of auditory substitution of vision) has more parameters than can easily be associated with sound attributes, the following direct kinds of parameter mappings are used most frequently:

- frequency of sound (increasing frequency usually means that an event parameter is increasing, or moving to the right / upward)

- amplitude of sound is often mapped to distance information (increasing loudness means that a parameter is increasing or approaching)

- timing in case of short sound events (decreasing the time interval between sound samples conveys an impression of increasing urgency, or a shortening of a distance)

- timbre, i.e. the "*characteristic quality of sound, independent of pitch and loudness, from which its source or manner of production can be inferred*" [45] can represent iconic features, such as color or texture of the visual counterpart

Besides such "direct" mapping techniques, various analogy-based solutions are also possible. For example, sonification can reflect the spatio-temporal context of

events, sometimes in a simplified form as in 'cartoonification' [46-49]. In Model-Based Sonification [50], the data are generally used to configure a sound-capable virtual object that in turn reacts on excitatory interactions with acoustic responses whereby the user can explore the data interactively. This can be extended to interactive sonification where a higher degree of active involvement occurs when the user actively changes and adjusts parameters of the sonification module, or interacts otherwise with the sonification system [51].

Whenever such mappings are selected intuitively, empirical evaluations should be carried out with the target user group (i.e., whatever the designer considers to be an intuitive sonification does not necessarily correspond to the target group's expectations [52]). In such cases, cognitive load due to a potentially large number of features is just as important as is the aspect of semantic recognizability.

Speech and music

Using speech and music has always been a viable option in electronic aids, and has always been treated somewhat orthogonally to the aspects of sonification described above. Speech is convenient because many steps of sonification can be skipped and all the information can be directly "mapped" to spoken words. This, however, also makes speech-based communication relatively slow, and language-dependent text-to-speech applications have to be implemented. Previously, concatenative synthesis (the recording of human voice) tended to be used for applications requiring small vocabularies of fewer than 200 words, whereas Text-to-Speech (TTS) was used for producing a much larger range of responses [53]. There were two types of TTS synthesis techniques depending on signal processing methods. The technique of diphone synthesis uses diphones (i.e., a pair of phones) extracted from the digitized recordings of a large set of standard utterances. The formant synthesis uses a mathematical model of formant frequencies to produce intelligible speech sounds. Nowadays, these techniques are integrated and used in parallel in electronic products [54].

Recently, several novel solutions have emerged, which try to make use of the advantageous properties of speech (symbolic), and iconic and indexical representations [55]. These speech-like sounds, including spearcons, spemoticons, spindexes, lyricons, etc. are a type of tweaked speech sound, which uses part of the speech or the combinations of speech and other sounds [56-61]. Spindexes are a predesigned prefix set and can be automatically added to speech items. Lyricons are a combination of melodic speech and earcons and thus, require some manual sound design. However, spearcons and spemoticons can be algorithmically made on the fly. Spearcons' time compression is accomplished by running text-to-speech files through a SOLA (Synchronized Overlap Add Method) algorithm [62, 63], which produces the best-quality speech for a computationally efficient time domain technique. Spemoticons can be made in the interactive development environment by manipulating the intensity, duration and pitch structure of the generated speech [64].

Some of these auditory cues can represent a specific meaning of the item, but others can represent an overall structure of the system. For example, each spearcon can depict a specific item like auditory icons with a focus on "what" an item is and provide a one-on-one mapping between sound and meaning. In contrast, spindex cues can provide contextual information, such as the structure and size of the auditory menus, and the user's location or status like earcons with a focus on "where" the user is in the system structure. Interestingly, lyricons or the combinations of melodic speech and earcons can represent both the semantics of the item (speech) and the contextual information of the system (earcons).

Music or musical concept can also be used for electronic devices. Music is pleasant for long term, can be learned relatively fast, and can be both suitable for iconic and continuous representation. Chords or chord progression can deliver different events or emotions. Different instruments or timbre can represent unique characteristics of objects or events. By mapping a musical scale to menu items, auditory scrollbars enhanced users' estimation of menu size and their relative location in a one-dimensional menu [65]. Using a short portion of existing music, musicons have been shown successful application as a reminder for home tasks (e.g., reminder for taking pills) [66]. On the other hand, intuitive mappings between music and meaning are inherently difficult given the subjective characteristics of music. First, individual musical capabilities differ from person to person. Second, the context of the application has to be considered to alleviate the possibility of both misinterpretations (e.g., when other sound sources are present in the environment at the same time) and of "phantom" perceptions, in which the user thinks that she has perceived a signal despite the fact that there is no signal (understanding such effects is as important as guaranteeing that signals that are perceived are understood correctly).

Hybrid solutions As a synthesis of all of the above techniques, increasingly ingenious approaches have appeared which combine high-level (both iconic and abstract, using both music and speech) sounds with sonification-based techniques. Walker and his colleagues tried to come up with hybrids integrating auditory icons and earcons [67]. Speech, spearcons, and spindex cues have also been used together in a serial manner [68]. Jeon and Lee compared subsequent vs. parallel combinations of different auditory cues on smartphones to represent a couple of submenu components (e.g., combining camera shutter and game sound to represent a multimedia menu) [69]. In the same research, they also tried to integrate background music (representing menu depths) and different auditory cues (representing elements in each depth) in a single user interface. Recently Csapo, Baranyi, and their colleagues developed hybrid solutions based on earcons, auditory icons and sonification to convey tactile information as well as feedback on virtual sketching operations using sound [70, 71]. Such combinations make possible the inclusion of temporal patterns into audio feedback as well, e.g. in terms of the ordering and timing between component sounds.

3 Haptics

Haptic perception occurs when objects are explored and recognized through touching, grasping, or pushing / pulling movements. “Haptic” comes from the Greek, “haptikos”, which means “to be able to touch or grasp” [72]. Depending on the mechanoreceptor, haptic perception includes pressure, flutter, stretching, and vibration and involves the sensory and motor systems as well as high-level cognitive capabilities. The terms "haptic perception" and "tactile perception" are often used interchangeably, with the exception that tactile often refers to sensations obtained through the skin, while haptic often refers to sensations obtained through the muscles, tendons, and joints. We will use this same convention in this section.

Haptic/tactile resolution and accuracy

In a way similar to vision and hearing, haptic / tactile perception can also be characterized by measures of accuracy and spatial resolution. However, the values of such measures are different depending on various body parts and depending on the way in which stimuli are generated. From a technological standpoint, sensitivity to vibration and various grating patterns on different areas of the body influences (though it does not entirely determine) how feedback devices can be applied. Relevant studies have been carried out in a variety of contexts (see e.g., [73]). The use of vibration is often restricted to on/off and simple patterns using frequency and amplitude changes. As a result, it is generally seen as a way to add additional information along auditory feedback about e.g. warnings, importance, or displacement.

With respect to the resolution of haptic perception, just-noticeable-differences (JNDs) ranging from 5 to 10% have been reported [74]. The exact value depends on similar factors as in the case of tactile discrimination, but added factors such as the involvement of the kinesthetic sense in perception, and even the temperature of the contact object have been shown to play a significant role [75].

The overall conclusion that can be drawn from these results is that the amount of information that can be provided using tactile and haptic feedback is less than it is through the visual and auditory senses. As we will see later, information obtained through taction and force is also less conceptual in the sense that it seems to be less suitable for creating analogy-based information mapping.

Haptic/tactile representations

As in the case of audio, several basic and more complex types of tactile (and more generally, haptic) representations have been proposed as detailed in this section. One general tendency in these representations that can be contrasted with audio representation types is that the distinction between iconic and abstract representations is less clear-cut (this may be evidence for the fact that people are

generally less aware of the conceptual interpretations of haptic/tactile feedback than in the case of auditory feedback).

Hapticons (haptic icons) Haptic icons were defined by MacLean and Enriquez as "brief computer-generated signals, displayed to a user through force or tactile feedback to convey information such as event notification, identity, content, or state" [76]. In a different paper, the same authors write that "Haptic icons, or hapticons, [are] brief programmed forces applied to a user through a haptic interface, with the role of communicating a simple idea in a manner similar to visual or auditory icons" [77].

These two definitions imply that the term haptic icon and hapticon can be used interchangeably. Further, the discussions of MacLean and Enriquez refer both to 'representational' and 'abstract' phenomena (while the definitions themselves reflect a representational point of view, the authors also state that "*our approach shares more philosophically with [earcons], but we also have a long-term aim of adding the intuitive benefits of Gaver's approach…"[*76]). All of this suggests that the dimensions of representation and meaning are seen as less independent in the case of the haptic modality than in vision or audio. Stated differently, whenever an interface designer decides to employ hapticons/haptic icons, it becomes clear that the feedback signals will provide information primarily 'about' either haptic perception itself, or indirectly about the occurrence of an event that has been linked - through training - to the 'iconic' occurrence of those signals. However, the lack of distinction between haptic icons and hapticons also entails that designers of haptic interfaces have not discussed the need (or do not see a possibility, due perhaps to the limitations of the haptic modality) to create higher-level analogies between the form of a haptic signal and a concept from a different domain.

Tactons (tactile icons) Brewster and Brown define tactons and tactile icons as interchangeable terms, stating that both are "structured, abstract messages that can be used to communicate messages non-visually" [78]. This definition, together with the conceptual link between the 'representational' and 'abstract' creates a strong analogy between tactons and hapticons. In fact, tactons may be seen as special kinds of hapticons which make use of Blattner, Sumikawa and Greenberg's original distinction between signals that carry a certain meaning within their representation, and those which do not (which can be generalized to other modalities) is nowhere visible in these definitions. Once again, this may be due to limitations in the 'conceptual expressiveness' of the haptic/tactile modality. Subsequent work by Brewster, Brown and others has suggested that such limitations may be overcome. For example, it was suggested that by designing patterns of abstract vibrations, personalized cellphone vibrations can be used to provide information on the identity of a caller [79]. If this is the case, one might wonder whether a separation of the terms haptic icon and hapticon, as well as those of tactile icon and tacton would be useful.

4 Applications

When it comes to developing portable and wearable assistive solutions for the visually impaired, power consumption is generally a crucial issue. If several computationally demanding applications are required to run at the same time, the battery life of even the highest quality mobile devices can be reduced to a few hours. Possible solutions include reducing the workload of the application (in terms of amount of information processed, or precision of processing); or using dedicated hardware alongside multi-purpose mobile devices, such as smartphones or tablets. In this section, we provide a broad overview of the past, present, and (potential) future of devices, technologies, and algorithmic solutions used to tackle such challenges.

Devices

State-of-the-art mobile devices offer built-in sensors and enormous computational capacity for application development on the Android and iOS platform [80, 81]. Without dedicated hardware, software-only applications provide information during navigation using the GPS, compass, on-line services and others [82-84]. However, microcontrollers and system-on-a-chip (SoC) solutions – e.g. Arduino or Raspberry Pi - are also gaining popularity, as the costs associated with such systems decrease and as users are able to access increasingly sophisticated services through them (e.g., in terms of accessing both versatile computing platforms, such as Wolfram Mathematica, and dedicated platforms for multimodal solutions, such as the Supercollider, Max/MSP [85] and PureData [86]).

Most of the above devices afford generic methods for the use of complementary peripheral devices for more direct contact with end users. In the blind community, both audio and haptic/tactile peripherals can be useful. In the auditory domain, devices such as microphones, filters / amplifiers and headphones with active noise cancelling (ANC) for the combination and enhancement of sound are of particular interest. In the haptic and tactile domains, devices such as white canes, "haptic" bracelets and other wearables with built-in vibrotactile or eletrotactile displays are often considered. While the inclusion of such peripherals in solutions increases associated costs (in terms of both development time and sales price), it also enables the implementation of improved functionality in terms of number of channels, "stronger" transducers, and the ability to combine sensory channels both from the real world and virtual environments into a single embedded reality.

Assistive Technologies for the Blind

In this section, we provide a brief summary of systems which have been and are still often used for assistive feedback through the auditory and haptic senses.

Assistive technologies using audio

Auditory feedback has increasingly been used in assistive technologies oriented towards the visually impaired. It has been remarked that both the temporal and frequency-based resolution of the auditory sensory system is higher than the resolution of somatosensory receptors along the skin. For several decades, however, this potential advantage of audition over touch was difficult to be taken due to the limitations in processing power [80].

Systems that have been particularly successful include:

- SonicGuide, which uses a wearable ultrasonic echolocation system to provide users with cues on the azimuth and distance of obstacles [87, 88].
- LaserCane, which involves the use of a walking cane and infrared instead of ultrasound signals [87, 89].
- The Nottingham Obstacle Detector, which is a handheld device that provides 8 gradations of distance through a musical scale based on ultrasonic echolocation [90].
- The Real-Time Assistance Prototype (RTAP), which is a camera-based system, equipped with headphones and a portable computer for improved processing power that even performs object categorization and importance-based filtering [91].
- The vOICe [92] and PSVA [87] systems, which provide direct, retinotopic temporal-spectral mappings between reduced-resolution camera-based images and audio signals.
- System for Wearable Audio Navigation (SWAN), which is developed for safe pedestrian navigation, and uses a combination of continuous (abstract) and event-based (conceptual) sounds to provide feedback on geometric features of the street, obstacles, and landmarks [93, 94].

Text-to-speech applications, speech-based command interfaces, and navigational helps are the most popular applications on mobile platforms. Talking Location, Guard my Angel, Intersection Explorer, Straight-line Walking apps, The vOICe, Ariadne GPS, GPS Lookaround, BlindSquare etc. offer several solutions with or without GPS for save guidance. See [80] for a detailed comparison and evaluation of such applications.

Assistive technologies using haptics

Historically speaking, solutions supporting vision using the tactile modality appeared earlier than audio-based solutions. These solutions generally translate camera images into electrical and/or vibrotactile stimuli.

Systems that have been particularly successful include:

• The Optacon device, which transcodes printed letters onto an array of vibrotactile actuators in a 24x6 arrangement [95-97]. While the Optacon was relatively expensive at a price of about 1500 GBP in the 1970s, it allowed for reading speeds of 15-40 words per minute [98] (others have reported an average of about 28 wpm [99], whereas the variability of user success is illustrated by the fact that two users were observed with Optacon reading speeds of over 80 wpm [80]).

• The Mowat sensor (from Wormald International Sensory Aids), which is a hand-held device that uses ultra-sonic detection of obstacles and provides feedback in the form of tactile vibrations inversely proportional to distance.

• Videotact, created by ForeThought Development LLC, which provides navigation cues through 768 titanium electrodes placed on the abdomen [100].

• A recent example of a solution which aims to make use of developments in mobile processing power is a product of a company, "Artificial Vision For the Blind", which incorporates a pair of glasses from which haptic feedback is transmitted to the palm [101, 102].

Today, assistive solutions making use of generic mobile technologies are increasingly prevalent. Further details on this subject can be found in [80].

Hybrid solutions

Solutions combining the auditory and haptic/tactile modalities are still relatively rare. However, several recent developments are summarized in [80]. Examples include the HiFiVE [103, 104] and SeeColOR [104] systems, which represent a wide range of low to high-level visual features through both audio and tactile representations. With respect to these examples, two observations are made: first, audio and taction are generally treated as a separate primary (i.e., more prominent with a holistic spatio-temporal scope) and secondary (i.e., less prominent in its spatio-temporal scope) modality, respectively; and second, various conceptual levels of are reflected in signals presented to these modalities at the same time.

Algorithmic challenges and solutions

In general, designing multimodal applications shows tradeoffs between storing stimuli beforehand and generating them on the fly. While the latter solution is more flexible in terms of real-time parametrization, it requires more processing.

Signal generation

An important question on any platform is how to generate the desired stimuli. In this section, we summarize key approaches in auditory and haptic/tactile domains.

Auditory signals With the advance of technology, multiple levels of auditory signals can be generated from electronic devices, including assistive technologies. However, auditory signals can be largely classified into two types. First, we can generate auditory signals using a buzzer. The buzzer is a self-sufficient sound-

generation device. It is cheap and small. It does not require any additional sound equipment or hardware. The different patterns of auditory signals can be programmed using even lower level programming languages (e.g., C or Assembly), varying the basic low-level signal processing parameters of *attack*, decay, *sustain* and *release*. With these, the parameters of *sound frequency*, *melodic pattern (including polarity)*, *number of sounds*, *total sound length*, *tempo of the sound*, and *repetitions within patterns* can be adjusted [105, 106]:

Nowadays, high quality auditory signals can also be generated using most mobile and wearable devices, including compressed formats like MP3 or MPEG-4 AAC. In this case, additional hardware is required, including an amplifier and a speaker system. Of course, the above variables can be manipulated. Moreover, timbre, which is a critical factor in mapping data to sounds [107] or musical instruments can represent particular functions or events on the device. More musical variables can also be controlled, such as chord, chord progression, key change, etc. in this high level configuration. Auditory signals can be generated as a pre-recorded sound file or in real-time through the software oscillator or MIDI (musical instrument digital interface). Currently, all these sound formats are supported by graphic programming environments (e.g., Max/MSP or PureData) or traditional programming languages via a sound specific library (e.g. JFugue [108] in Java).

Tactile signals Mobile phones and tablets have integrated vibrotactile motors for haptic feedback. Usually, there is only one small vibrator installed in the device that can be accessed by the applications. The parameters of *vibration length*, *durations of patterns* and *repetitions within patterns* can be set through high-level object-oriented abstractions. Many vibration motor configurations are by design not suited to the real-time modification of vibration amplitude or frequency. Therefore, while such solutions offer easy accessibility and programmability, they do so with important restrictions and only for a limited number of channels (usually one channel). As a result, very few assistive applications for blind users make use of vibration for purposes other than explorative functions (such as 'zooming in' or 'panning') or alerting users, and even less use it to convey augmentative information feedback.

From the signal processing point of view, smartphones and tablets run pre-emptive operating systems, meaning that more than a single application can run "simultaneously", and even processes that would be critical for the assistive application can be interrupted. Although accessing a built-in vibration motor is not a critical application in and of itself, a comprehensive application incorporating other features (such as audio and/or visual rendering) together with haptics can be problematic. For example, spatial signals can be easily designed in a congruent way between audio and haptic (e.g., left auditory/haptic feedback vs. right auditory/haptic feedback). When it comes to frequency, the frequency range does not match with each other in one-on-one mapping. This is why we need to empirically assess the combination of multimodal feedback in a single device.

Latency and memory usage

Latency is a relatively short delay, usually measured in milliseconds, between the time when an audio signal enters and when it outputs from a system. This includes hardware processing times and any additional signal processing tasks (filtering, calculations etc.). The most important contributors to latency are DSP, ADC/DAC, buffering, and in some cases, travelling time of sound in the air. For audio circuits and processing pipelines, a latency of 10 milliseconds or less is sufficient for real-time experience [109], based on the following guidelines [110]:

Less than 10 ms - allows real-time monitoring of incoming tracks including effects.

At 10 ms - latency can be detected but can still sound natural and is usable for monitoring.

11-20 ms - monitoring starts to become unusable, smearing of the actual sound source, and the monitored output is apparent.

20-30 ms - delayed sound starts to sound like an actual delay rather than a component of the original signal.

In a virtual acoustic environment, the total system latency (TSL) refers to the time elapsed from the transduction of an event or action, such as movement of the head, until the consequences of that action cause the equivalent change in the virtual sound source location [111]. Problems become more significant if signal processing includes directional sound encoding and rendering, synthesizing and dynamic controlling of reverberation, room and distance effects. Several software applications have been recently developed to address this problem for various platforms such as Spat [112], Sound Lab (SLAB) [113], DirAC [114], etc.

It has been noted that real-time audio is a challenging task in VM-based garbage collected programming languages such as Java [115]. It was demonstrated that a low latency of a few milliseconds can nevertheless be obtained, even if such performance is highly dependent on the hardware-software context (i.e., drivers and OS) and cannot always be guaranteed [115]. Since version 4.1 ("Jelly Bean"), Android has included support for audio devices with low-latency playback through a new software mixer and other API improvements. Thus, improved functionality in terms of latency targets below 10 ms, as well as others such as multichannel audio via HDMI is gaining prevalence. The use of specialized audio synthesis software such as Csound [116], Supercollider, Max/MSP, and PureData can also help achieve low latency.

Training challenges

An important challenge when deploying assistive applications lies in how to train prospective users in applying them to their own real-life settings. In such cases, serious gaming is one of the options used for training and maintaining the interest of users. The term refers to a challenging and motivating environment for training

in which users adapt to the system in an entertaining, and as a result, almost effortless way. Examples might include giving users the task of finding collectibles, earning rankings, or challenging other competitors by improving in the task. Especially blind users welcome audio-only games as a source of entertainment [117, 118]. As far as mobile applications are concerned, it can be stated as a general rule that regardless of subject matter, serious games usually do not challenge memory and processing units further, as they use limited visual information (if at all), and also often rely on reduced resolution in audio rendering.

5 Summary

This paper provided an overview of the state-of-the-art that is relevant to the key design phases behind portable and wearable assistive technologies for the visually impaired. Specifically, we focused on aspects of stimulus synthesis, semantically informed mapping between data / information and auditory / tactile feedback, as well as signal processing techniques that are useful either for curbing computational demands, or for manipulating the information properties of the feedback signals. Through a broad survey of existing applications, the paper demonstrated that audio-tactile feedback is increasingly relevant to transforming the daily lives of users with visual impairments.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

References

[1] F. Hosseini, S. M. Riener, A. Bose and M. Jeon, "Listen2dRoom": Helping Visually Impaired People Navigate Indoor Environments using an Ultrasonic Sensor-based Orientation Aid, in *Proc. of the 20th International Conference on Auditory Display (ICAD2014)*, NY, 2014, 6 pages

[2] M. Jeon, A. Nazneen, O. Akanser, A. Ayala-Acevedo and B. N. Walker, "Listen2dRoom": Helping Blind Individuals Understand Room Layouts. In *Extended Abstracts Proc. of the SIGCHI Conference on Human Factors in Computing Syst. (CHI'12)*, Austin, TX pp. 1577-1582, 2012

[3] A. M. Ali and M. J. Nordin, "Indoor Navigation to Support the Blind Person using Weighted Topological Map," *in Proc. of the Int'l Conf. on Elect. Eng. and Inf.*, Malaysia, pp. 68-72, 2009

[4] B. N. Walker and J. Lindsay, "Development and Evaluation of a System for Wearable Audio Navigation," *in Proc. of Hum. Fact. and Erg. Society*

Annual Meeting Proceedings 09/2005, Vol. 49, No. 17, Orlando, USA, pp. 1607-1610, 2005

[5] G. V. Elli, S. Benetti and O. Collignon, "Is There a Future for Sensory Substitution Outside Academic Laboratories?," *Multisensory Research*, Vol. 27, No. 5/6, pp. 271-291, 2014

[6] M. Sivak, "The Information that Drivers Use: Is It Indeed 90% Visual?," *Perception*, Vol. 25, pp. 1081-1089, 1996

[7] S. O. Nielsen, "Auditory Distance Perception in Different Rooms," *Auditory Engineering Society 92nd Convention*, Vienna, Austria, paper nr. 3307, 1992

[8] P. Zahorik and F. L. Wightman, "Loudness Constancy with Varying Sound Source Distance," *Nature Neuroscience*, Vol. 4, pp. 78-83, 2001

[9] R. Klatzky, S. Lederman and C. Reed C, "There's More to Touch than Meets the Eye: The Salience of Object Attributes for Haptics With and Without Vision," *J. of Exp. Psychology*, Vol. 116, No. 4, pp. 356-369, 1987

[10] R. Klatzky and S. Lederman, "Intelligent Exploration by the Human Hand," *Dextrous Robot Hands*, pp. 66-81, 1990

[11] S. Lederman and R. Klatzky, "Hand Movements: A Window into Haptic Object Recognition," *Cognitive Psychology*, Vol. 19, pp. 342-368, 1987

[12] K. A. Kaczmarek and S. J. Haase, "Pattern Identification and Perceived Stimulus Quality as a Function of Stimulation Current on a Fingertip-scanned Electrotactile Display," *IEEE Trans Neural Syst Rehabil Eng*, Vol. 11, pp. 9-16, 2003

[13] D. Dakopoulos and N. G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," *IEEE Trans. on Syst. Man and Cybernetics Part C*, 40(1), pp. 25-35, 2010

[14] J. Blauert, *Spatial Hearing*, The MIT Press, MA, 1983

[15] C. I. Cheng and G. H. Wakefield, "Introduction to HRTFs: Representations of HRTFs in Time, Frequency, and Space," *J. Audio Eng. Soc.*, Vol. 49, pp. 231-249, 2001

[16] H. Möller, M. F. Sorensen, D. Hammershoi and C. B. Jensen, "Head-Related Transfer Functions of Human Subjects," *J. Audio Eng. Soc*., Vol. 43, pp. 300-321, 1995

[17] D. Hammershoi and H. Möller, "Binaural Technique – Basic Methods for Recording, Synthesis, and Reproduction," in J. Blauert (Editor) *Comm. Acoustics*, pp. 223-254, Springer, 2005

[18] F. Wightman and D. Kistler, "Measurement and Validation of Human HRTFs for Use in Hearing Research," *Acta acustica united with Acustica*, Vol. 91, pp. 429-439, 2005

[19] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modeling, Simulation, Algorithms, and Acoustic Virtual Reality*, Springer, Berlin, 2008

[20] A. Dobrucki, P. Plaskota, P. Pruchnicki, M. Pec, M. Bujacz and P. Strumiłło, "Measurement System for Personalized Head-Related Transfer Functions and Its Verification by Virtual Source Localization Trials with Visually Impaired and Sighted Individuals," *J. Audio Eng. Soc*., Vol. 58, pp. 724-738, 2010

[21] Gy. Wersényi, "Virtual localization by blind persons," *Journal of the Audio Engineering Society*, Vol. 60, No. 7/8, pp. 579-586, 2012

[22] Gy. Wersényi, "Localization in a HRTF-based Virtual Audio Synthesis using additional High-pass and Low-pass Filtering of Sound Sources," *Journal of the Acoust. Science and Technology Japan*, Vol. 28, pp. 244-250, 2007

[23] P. Minnaar, S. K. Olesen, F. Christensen and H. Möller, "Localization with Binaural Recordings from Artificial and Human Heads," *J. Audio Eng. Soc.*, Vol. 49, pp. 323-336, 2001

[24] D. R. Begault, E. Wenzel and M. Anderson, "Direct Comparison of the Impact of Head Tracking Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source," *J. Audio Eng. Soc*., Vol. 49, pp. 904-917, 2001

[25] H. Möller, "Fundamentals of Binaural Technology," *Appl. Acoustics*, Vol. 36, pp. 171-218, 1992

[26] P. A. Hill, P. A. Nelson and O. Kirkeby, "Resolution of Front-Back Confusion in Virtual Acoustic Imaging Systems," *J. Acoust. Soc. Am*., vol. 108, pp. 2901-2910, 2000

[27] H. Möller, D. Hammershoi, C. B. Jensen and M. F. Sorensen, "Transfer Characteristics of Headphones Measured on Human Ears," *J. Audio Eng. Soc*., Vol. 43, pp. 203-216, 1995

[28] R. Algazi and R. O. Duda, "Headphone-based Spatial Sound," *IEEE Signal Processing Magazine*, Vol. 28, No. 1, pp. 33-42, 2011

[29] R. M. Stanley, "Toward Adapting Spatial Audio Displays for Use with Bone Conduction". M.S. Thesis. Georgia Institute of Technology. Atlanta, GA. 2006

[30] R. M. Stanley, "Measurement and Validation of Bone-Conduction Adjustment Functions in Virtual 3D Audio Displays," Ph.D. Dissertation. Georgia Institute of Technology. Atlanta, GA. 2009

[31] O. Balan, A. Moldoveanu, F. Moldoveanu, "A Systematic Review of the Methods and Experiments Aimed to Reduce Front-Back Confusions in the

Free-Field and Virtual Auditory Environments," submitted to *International Journal of Acoustics and Vibration*

[32] E. M. Wenzel, "Localization in Virtual Acoustic Displays," *Presence*, Vol. 1, pp. 80-107, 1992

[33] P. A. Hill, P. A. Nelson and O. Kirkeby, "Resolution of Front-Back Confusion in Virtual Acoustic Imaging Systems," *J. Acoust. Soc. Am.*, Vol. 108, pp. 2901-2910, 2000

[34] M. M. Boone, E. N. G. Verheijen and P. F. van Tol, "Spatial Sound-Field Reproduction by Wave-Field Synthesis," *J. Audio Eng. Soc*., Vol. 43, No. 12, pp. 1003-1012, 1995

[35] J. Daniel, J. B. Rault and J. D. Polack, "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions," *in Proc. of 105th AES Convention*, paper 4795, 1998

[36] W. Gaver, "Auditory Icons: Using Sound in Computer Interfaces," *Human Comput Interact,* Vol. 2, No. 2, pp. 167-177, 1986

[37] W. Gaver, "Everyday Listening and Auditory Icons," Ph.D. dissertation, University of California, San Diego, 1998

[38] W. Gaver, "The SonicFinder: an Interface that Uses Auditory Icons," *Human Comput Interact,* Vol. 4, No. 1, pp. 67-94, 1989

[39] D. Smith, "Pygmalion: a Computer Program to Model and Stimulate Creative Thought," Ph.D. dissertation, Stanford University, Dept. of Computer Science, 1975

[40] M. Blattner, D.Sumikawa and R. Greenberg, "Earcons and Icons: Their Structure and Common Design Principles," *Human Comput Interact,* Vol. 4, No. 1, pp. 11-44, 1989

[41] G. Kramer (ed), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Santa Fe Institute Studies in the Sciences of Complexity. Proceedings volume XVIII. Addison-Wesley, 1994

[42] T. Hermann, A. Hunt and J. G. Neuhoff, *The Sonification Handbook,* Logos, Berlin, 2011

[43] Á. Csapó and Gy. Wersényi, "Overview of Auditory Representations in Human-Machine Interfaces," *Journal ACM Computing Surveys (CSUR)*, Vol. 46, No. 2, Art.nr. 19., 19 pages, 2013

[44] B. N. Walker, "Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification," *Journal of Experimental Psychology*: *Applied*, Vol. 8, pp. 211-221, 2002

[45] See 72

[46] P. Lennox, J. M. Vaughan and T. Myatt, "3D Audio as an Information-Environment: Manipulating Perceptual Significance for Differentiation and Pre-Selection," *in Proc. of the 7th International Conference on Auditory Display (ICAD2001),* Espoo, Finland, 2001

[47] P. Lennox, and T. Myatt, "Perceptual Cartoonification in Multi-Spatial Sound Systems," *in Proc. of the 17th Int. Conference on Auditory Display (ICAD2011*), Budapest, Hungary, 2011

[48] See 71

[49] See 70

[50] T. Hermann, Sonification for Exploratory Data Analysis. Ph.D. dissertation. Bielefeld University. Bielefeld, Germany. 2002

[51] T. Hermann, "Taxonomy and Definitions for Sonification and Auditory Display," *Proc of the 14th International Conference on Auditory Display*, Paris, France, 2008

[52] G. Kramer, B. N. Walker, T. Bonebright, P. Cook, J. H. Flowers, N. Miner and J. Neuhoff, "Sonification Report: Status of the Field and Research Agenda," Faculty Publications, Department of Psychology. Paper 444. http://digitalcommons.unl.edu/psychfacpub/444, 2010

[53] M. Jeon, S. Gupta and B. N. Walker, "Advanced Auditory Menus II: Speech Application for Auditory Interface," Georgia Inst. of Technology Sonification Lab Technical Report. February, 2009

[54] M. Jeon, "Two or Three Things You Need to Know about AUI Design or Designers," *in Proc. of the 16th International Conference on Auditory Display (ICAD2010),* Washington, D.C., USA, 2010

[55] M. Jeon, "Exploration of Semiotics of New Auditory Displays: A Comparative Analysis with Visual Displays," *in Proc. of the 21st Int. Conference on Auditory Display (ICAD2015*), Graz, 2015

[56] Gy. Wersényi, "Auditory Representations of a Graphical User Interface for a Better Human-Computer Interaction," in S. Ystad et al. (Eds.): Auditory Display. CMMR/ICAD 2009 post proceedings edition, LNCS 5954, Springer, Berlin, pp. 80-102, 2010

[57] B. Gygi and V. Shafiro, "From Signal to Substance and Back: Insights from Environmental Sound Research to Auditory Display Design," *in Proc. of the International Conference on Auditory Display ICAD 09*, pp. 240-251, 2009

[58] B. N. Walker and A. Kogan, "Spearcon Performance and Preference for Auditory Menus on a Mobile Phone," *Universal Access in Human-Computer Interaction - Intelligent and Ubiquitous Interaction Environments*, Lecture Notes in Computer Science Volume 5615, pp. 445-454, 2009

[59] See 64

[60] M. Jeon and B. N. Walker, "Spindex (Speech Index) Improves Auditory Menu Acceptance and Navigation Performance," *Journal ACM Transactions on Accessible Computing (TACCESS)*, Vol. 3, No. 3, article no. 10, 2011

[61] M. Jeon, and Y. Sun, "Design and Evaluation of Lyricons (Lyrics + Earcons) for Semantic and Aesthetic Improvements of Auditory Cues," *in Proc. of the 20th International Conference on Auditory Display (ICAD2014),* 2014

[62] D. J. Hejna, "Real-Time Time-Scale Modification of Speech via the Synchronized Overlap-Add Algorithm," M.S. Thesis, Dept. of Electrical Engineering and Computer Science, MIT, 1990

[63] S. Roucos, and A. M. Wilgus, "High Quality Time-Scale Modification for Speech," *in Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, pp. 493-496, New York, NY, USA. 1985

[64] G. Németh, G. Olaszy and T. G. Csapó, "Spemoticons: Text to Speech Based Emotional Auditory Cues," *in Proc. of ICAD2011*, Budapest, Hungary, 2011

[65] P. Yalla, and B. N. Walker, "Advanced Auditory Menus: Design and Evaluation of Auditory Scrollbars," *in Proc. of ACM Conference on Assistive Technologies (ASSETS'08*), pp. 105-112, 2008

[66] M. McGee-Lennon, M. K. Wolters, R. McLachlan, S. Brewster and C. Hall, "Name that Tune: Musicons as Reminders in the Home," *in Proc. of the SIGCHI Conference on Human Factors in Computing System*, Vancouver, BC, Canada, pp. 2803-2806, 2011

[67] B. N. Walker, J. Lindsay, A. Nance, Y. Nakano, D. K. Palladino, T. Dingler and M. Jeon, "Spearcons (Speech-based earcons) improve navigation performance in advanced auditory menus," *Human Factors*, Vol. 55, No. 1, pp. 157-182. 2013

[68] M. Jeon, T. M. Gable, B. K. Davison, M. Nees, J. Wilson and B. N. Walker, "Menu Navigation with In-Vehicle Technologies: Auditory Menu Cues Improve Dual Task Performance, Preference, and Workload," *Int. J. of Human-Computer Interaction*, Vol. 31, No. 1, pp. 1-16. 2015

[69] M. Jeon and J.-H. Lee, "The Ecological AUI (Auditory User Interface) Design and Evaluation of User Acceptance for Various Tasks on Smartphones," in M. Kurosu (Ed.), Human-Computer Interaction: Interaction Modalities and Techniques (Part IV), HCII2013, LNCS, Vol. 8007, pp. 49-58, Heidelberg, Springer. 2013

[70] Á. Csapó and P. Baranyi, "CogInfoCom Channels and Related Definitions Revisited," *in Proc. Intelligent Systems and Informatics (SISY), 2012 IEEE 10th Jubilee International Symposium*, Subotica, Serbia, pp. 73-78, 2012

[71] Á. Csapó, J. H. Israel and O. Belaifa, "Oversketching and Associated Audio-based Feedback Channels for a Virtual Sketching Application," *in Proc. 4th IEEE International Conference on Cognitive Infocommunications*, Budapest, Hungary, pp. 509-513, 2013

[72] S. Yantis, *Sensation and Perception*, NY: Worth Publishers, 2014

[73] R. van Bowen, K. O. Johnson, "The Limit of Tactile Spatial Resolution in Humans," *Neurology*, Vol. 44, No. 12, pp. 2361-2366, 1994

[74] S. Allin, Y. Matsuoka and R. Klatzky, "Measuring Just Noticeable Differences for Haptic Force Feedback: Implications for Rehabilitation," in *Proc. 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems,* Orlando, USA, pp. 299-302, 2002

[75] M. A. Srinivasan and J. S. Chen, "Human Performance in Controlling Normal Forces of Contact with Rigid Objects," *Adv. in Robotics, Mechat. and Haptic Interfaces*, pp. 119-125, 1993

[76] K. Maclean and M. Enriquez, "Perceptual Design of Haptic Icons," in *Proc. of Eurohaptics*, pp. 351-363, 2003

[77] M. Enriquez and K. MacLean, "The Hapticon Editor: a Tool in Support of Haptic Communication Research," *in Proc. of the 11th symposium on haptic interfaces for virtual environment and teleoperator systems (HAPTICS'03),* Los Angeles, USA, pp. 356-362, 2003

[78] S. Brewster and L. Brown, "Tactons: Structured Tactile Messages for Non-Visual Information Display," *in Proc. of the 5th Conf. on Australasian User Int. (AUIC'04*), Vol. 28, pp. 15-23, 2004

[79] L. M. Brown and T. Kaaresoja, "Feel Who's Talking: Using Tactons for Mobile Phone Alerts," pp. 1-6, CHI 2006

[80] Á. Csapó, Gy. Wersényi, H. Nagy and T. Stockman, "A Survey of Assistive Technologies and Applications for Blind Users on Mobile Platforms: a Review and Foundation for Research," *J Multimodal User Interfaces*, Vol. 9, No. 3, 11 pages, 2015

[81] Á. Csapó, Gy. Wersényi and H. Nagy, "Evaluation of Reaction Times to Sound Stimuli on Mobile Devices," in *Proc. of ICAD15*, Graz, Austria, pp. 268-272, 2015

[82] S. A. Panëels, D. Varenne, J. R. Blum and J. R. Cooperstock, "The Walking Straight Mobile Application: Helping the Visually Impaired avid Veering," i*n Proc. of ICAD13*, Lódz, Poland, pp. 25-32, 2013

[83] Gy. Wersényi, "Evaluation of a Navigational Application Using Auditory Feedback to Avoid Veering for Blind Users on Android Platform," *Acoust. Soc. Am.*, Vol. 137, pp. 2206, 2015

[84] L. Dunai, G. P. Fajarnes, V. S. Praderas, B. D. Garcia and I. L. Lengua, "Real-Time Assistance Prototype—A New Navigation Aid for Blind People," *in Proc. IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society*, pp. 1173-1178, 2010

[85] https://cycling74.com/

[86] https://puredata.info/

[87] C. Capelle, C. Trullemans, P. Arno and C. Veraart, "A Real-Time Experimental Prototype for Enhancement of Vision Rehabilitation Using Auditory Substitution," *IEEE Transactions on Biomedical Engineering*, Vol. 45, No. 10, pp. 1279-1293, 1998

[88] L. Kay, "A Sonar Aid to Enhance Spatial Perception of the Blind: Engineering Design and Evaluation," *IEEE Radio and Electronic Engineer*, Vol. 44, No. 11, pp. 605-627, 1974

[89] E. F. Murphy, "The VA – Bionic Laser Can for the Blind. In The National Research Council,", Evaluation of Sensory Aids for the Visually Handicapped, NAS, pp. 73-82, 1971

[90] D. Bissitt and A. D. Heyes, "An Application of Bio-Feedback in the Rehabilitation of the Blind," *Applied Ergonomics*, Vol. 11, No. 1, pp. 31-33 1980

[91] L. Dunai, G. P. Fajarnes, V. S. Praderas, B. D. Garcia and I. L. Lengua, "Real-Time Assistance Prototype—A New Navigation Aid for Blind People," *in Proc. IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society*, pp. 1173-1178, 2010

[92] P. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Transactions on Biomedical Engineering,* Vol. 39, No. 2, pp. 112-121, 1992

[93] B. N. Walker and J. Lindsay, "Navigation Performance with a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice," *Hum Factors*, Vol. 48, No. 2, pp. 265-278, 2006

[94] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias and F. Dellaert, "SWAN: System for Wearable Audio Navigation," *in Proc. of the 11th International Symposium on Wearable Computers (ISWC 2007*), USA, 8 pages, 2007

[95] V. Levesque, J. Pasquero and V. Hayward, "Braille Display by Lateral Skin Deformation with the STReSS Tactile Transducer," *in Proc. of the 2nd joint eurohaptics conference and symposium on haptic interfaces for virtual environment and teleoperator systems*, Tsukuba, pp. 115-120, 2007

[96] P. Bach-y Rita, M. E. Tyler and K. A. Kaczmarek, "Seeing with the Brain," *Int J Hum Comput Interact*, Vol. 15, No. 2, pp. 285-295, 2003

[97] J. C. Bliss, M. H. Katcher, C. H. Rogers, R. P. Shepard," Optical-to-Tactile Image Conversion for the Blind," *IEEE Trans Man Mach Syst* Vol. 11, pp. 58-65, 1970

[98] J. C. Craig, "Vibrotactile Pattern Perception: Extraordinary Observers," *Science*, Vol. 196, No. 4288, pp. 450-452, 1977

[99] D. Hislop, B. L. Zuber and J. L. Trimble, "Characteristics of Reading Rate and Manual Scanning Patterns of Blind Optacon Readers," *Hum Factors* Vol. 25, No. 3, pp. 379-389, 1983

[100] http://www.4thtdev.com. Accessed Mar 2015

[101] http://unreasonableatsea.com/artificial-vision-for-the-blind/. Accessed Mar 2015

[102] http://www.dgcs.unam.mx/ProyectoUNAM/imagenes/080214.pdf. Accessed Mar 2015

[103] D. Dewhurst, "Accessing Audiotactile Images with HFVE Silooet," *in Proc. Fourth Int. Workshop on Haptic and Audio Interaction Design*, Springer-Verlag, pp. 61-70, 2009

[104] D. Dewhurst, "Creating and Accessing Audio-Tactile Images with "HFVE" Vision Substitution Software," *in Proc. of 3rd Interactive Sonification Worksh*, Stockholm, pp. 101-104, 2010

[105] M. Jeon, "Two or Three Things You Need to Know about AUI Design or Designers," *in Proc. of the 16th International Conference on Auditory Display (ICAD2010*), Washington, D.C., USA, 2010

[106] M. Jeon, "Auditory User Interface Design: Practical Evaluation Methods and Design Process Case Studies," *The International Journal of Design in Society*, Vol. 8, No. 2, pp. 1-16, 2015

[107] B. N. Walker and G. Kramer, "Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping, and Meaning Making," Ed. by J. Neuhoff, Ecol. Psychoacoustics. NY, Academic Press. 2004

[108] http://www.jfugue.org/

[109] S. M. Kuo, B. H. Lee and W. Tian, *Real-Time Digital Signal Processing,* Wiley, 2013

[110] Adobe Audition User's Manual

[111] E. M. Wenzel, "Effect of Increasing System Latency on Localization of Virtual Sounds," *in Proc. of 16th International Conference: Spatial Sound Reproduction*, 1999

[112] J.-M. Jot, "Real-Time Spatial Processing of Sounds for Music, Multimedia and Interactive Human-Computer Interfaces," *Multimedia Systems*, Vol. 7, No. 1, pp. 55-69, 1999

[113] E. M. Wenzel, J. D. Miller and J. S. Abel, "Sound Lab: A Real-Time, Software-Based System for the Study of Spatial Hearing," *in Proc. of the 108 AES Convention*, Paris, France, 2000

[114] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *Journal Audio Eng. Soc*., Vol. 55, No. 6, pp. 503-516, 2007

[115] N. Juillerat, S. Muller Arisona, S. Schubiger-Banz, "Real-Time, Low Latency Audio Processing in JAVA," *in Proc. of Int. Comp. Music Conference ICMC2007*, Vol. 2, pp. 99-102, 2007

[116] http://www.csounds.com/manual/html/UsingRealTime.html

[117] http://www.audiogames.net/

[118] O. Balan, A. Moldoveanu, F. Moldoveanu, M-I. Dascalu," Audio Games- a Novel Approach towards Effective Learning in the Case of Visually-impaired People," *In Proc. 7th Int. Conference of Education, Research and Innovation*, Seville, Spain, 7 pages, 2014

Biometric Verification of Maternity and Identity Switch Prevention in Maternity Wards

Komlen Lalović[1], Nemanja Maček[2], Milan Milosavljević[3], Mladen Veinović[4], Igor Franc[5], Jelena Lalović[6], Ivan Tot[7]

[1,3,4] Faculty of Informatics and Computing, Singidunum University, 32 Danijelova Street, 11000 Beograd, Serbia; e-mails: komlen.lalovic.13@singimail.com, mmilosavljevic@singidunum.ac.rs, mveinovic@singidunum.ac.rs

[2,5] SECIT Security Consulting, 21/3 Aksentija Maksimovića Street, Pančevo, Serbia; e-mail: nmacek@secitsecurity.com, ifranc@secitsecurity.com

[6] Golden Mind Ltd, 11 Pohorska Street, 11070 Novi Beograd, Serbia, e-mail: office@GoldenMind.rs

[7] Military Academy, 33 Pavla Jurišića Šturma Street, 11000 Beograd, Serbia, e-mail: ivan.tot@va.mod.gov.rs

Abstract: This paper presents a novel approach to noninvasive biometric maternity verification and baby-switching prevention in maternity wards based on dual biometric fingerprint scanner. The proposed system is expected to solve issues that have occurred in many countries regarding stealing or mixing the identities of newborn babies. The solution is based on a device that acquires biometric samples from the mother and her newborn baby right after birth and generates the unique reference of the pair. The same device is used to verify the bound of biometric samples before the mother and the baby are allowed to leave the maternity ward. The privacy of stored templates is provided by cancelable biometrics while the auxiliary data are secured with strong cryptographic protection. The main contribution of the proposed approach is a very high level of proof of maternity for each newborn baby, as well as, the prevention of replacing identities of newborn babies in maternity wards, with a system that can be realized via low cost hardware.

Keywords: biometrics; fingerprint; minutiae; maternity; identity

1 Introduction

Babies switched in maternity hospitals are babies who are interchanged with each other at birth and raised by non-biological parents. According to Brandon Gaille, about 28,000 babies out of 4 million total births are switched every year [1]. Although Gaille stated that many of these cases are solved at some point before

families leave the hospital, the objective risks of switching babies still exist and these issues need to be resolved. The major cause behind baby-switching is human error. As an example, officials of Smith N. Mercy hospital stated definitively, that human error was the reason for such a mistake [2]. Once an incident is discovered, lawsuits against hospitals may occur. Two mothers whose infant sons were switched by medical personnel at Heartland Regional Medical Center brought a lawsuit against the hospital [3]. The switching of two Russian newborn girls was discovered after the ex-husband of one of the mothers had refused to pay alimony on the basis that she looked nothing like him and requested a DNA test [4]. Both families raised a lawsuit against the hospital. Crane provided a more detailed analysis on hospital liability and resulting custody issues [5]. Another aspect that should also be seriously considered is the maternity patient's levels of anxiety. Davis et al. stated that media reports in the USA of baby-switching caused anxiety of a number of patients, and, according to an experimental study, 10% of the mothers reported anxiety about baby-switching [6]. According to Rusting, baby-switching is a problem that can have the same impact on a healthcare facility as an infant kidnapping [7]. The fears of maternity health providers, including. but not limited to baby-switching, and the impact on the women they care for are discussed by Dahlen and Caplice [8]. Although some hospitals implement certain precautionary procedures, no 100% proof of maternity verification technique has been reported in the literature.

This paper presents a maternity verification approach based on fingerprint biometric trait matching. Biometric verification is the process of validating the identity of individuals according to their physiological or behavioral qualities [9], [10]. A fingerprint is the only physiological quality that is completely formed during the prenatal period [11]. By the end of the 7^{th} month of pregnancy, minutae point structure on each finger of the fetus is formed and the ridge shape remains constant during their entire life. Prenatally formed shape of fingerprint ridges and valleys allows biometric template acquisition of a newborn baby that can later be used for maternity verification, even if the baby is born premature, e.g. in the 8^{th} month of pregnancy [12]. Unlike fingerprint, the other physiological traits are unstable at birth. For example, the pigmentation of child's eye is changes until the age of four [13], resulting in an unstable iris acquisition and verification process for newborn babies. That is why the proposed approach employs the fingerprint biometrics. Biometric traits of a baby and the mother are acquired right after the birth on a dual fingerprint scanner; the unique reference is generated and the data is stored in a secure manner on the device. At any time, the identity of the mother and the baby can be verified on the same device, e.g., before breast-feeding. Before leaving the maternity hospital a final verification is performed, and the data are securely wiped from the device. The proposed approach employs cancelable biometrics to provide the privacy of stored templates and cryptography to provide security of stored auxiliary data.

2 Dual Fingerprint Scanner

Device for biometric identification of maternity [14] is dual fingerprint scanner, registered patent number 1412 U1 at the Institute of Intelectual Property, Republic of Serbia. It won the Belgrade City Award for the best Patent Innovation in the field of natural sciences in 2014. According to the International classification of patents, this patent is classified with a symbol G06F21/00, as it belongs to the biometric systems – fingerprint scanning devices. Fingerprint scanners use different algorithms and methods to authenticate or verify the identity of individuals. However, no device has been found in the National Database of Patents [15] that operates as dual fingerprint scanner or a fingerprint scanner capable of providing any kind of match between two persons. An example of single fingerprint scanner can be found in the patent confirmation 13848069.4, issued on April 2, 2013, with remark WO2014059761 and classification G06F21/00 ("Fingerprint identification device"). The device used in this research employs two optical sensors that simultaneously acquire biometric samples of two different persons. Once the acquisition of samples is complete in the enrollment phase, the device extracts features, generates cancelable biometric templates and unique ID reference that bonds those templates. Generated reference and cancelable templates are stored on the device in a secured manner. During the verification phase, the stored reference is used as an undeniable proof of identity match of two persons that provided their samples in enrollment phase.

2.1 Detailed Device Description

The device for biometric identification of maternity, contains the following components:

- Ignition switch that can be connected with delay timer (I),
- Display that can provide information on all current activities and results of generating unique ID reference,
- Start button (S1) that starts the fingerprint scanning,
- Store button (S2) that secures and stores the data after scanning,
- Reset button (R) that resets acquired and processed data, after storing it,
- Numerical keyboard that is used to acquire PIN,
- Two fingerprint scanning sensors (S1 and S2) for a mother and the baby. Larger sensor is used to scan mother's finger with 500 dpi resolution. Smaller sensor scans the baby's finger with at least 1000 dpi resolution as the ridge structure image is harder to acquire due to the size of the finger [16].

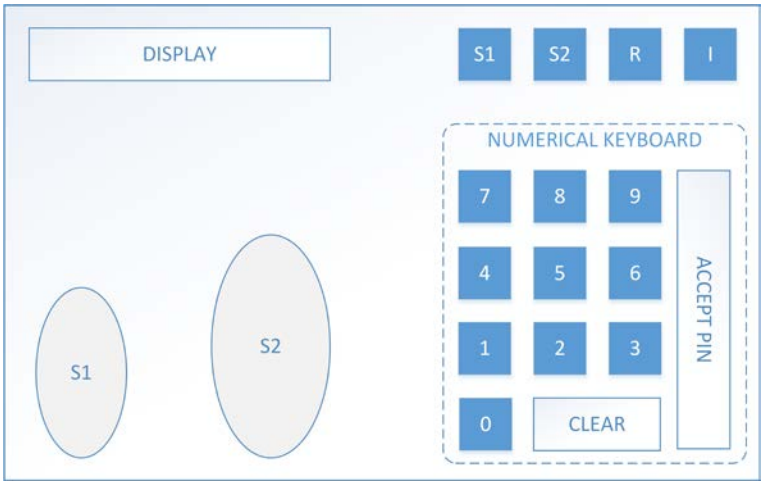


Figure 1
Device for biometric identification of maternity

The device operates as follows: the device started via ignition switch performs a self-check and provides the user with the information if it functions properly or not. After the start button is pressed, the device requires mother's and baby's fingers to be placed on sensors and starts scanning their fingerprints simultaneously. After the scanning, the unique ID reference is generated. Mother enters a PIN code on the numerical keyboard and accepts it. Once the store button is pressed, the data is secured and stored in the memory of the device, as described in chapters 3 and 4 of this paper.

The device employs memory cards for storing the acquired data. It should be stated that the device is still in the development phase and that additional functionalities, such as storing the data on a remote server via encrypted wireless communication channel, will be added subsequentially.

3 Proposed Maternity Verification Algorithm

The main idea behind our approach is to employ simultaneous dual fingerpint scanning, cancelable biometrics and cryptographic protection of the auxiliary data. The proposed approach that employs noninvasive biometrics is expected to provide 100% proof of maternity verification, template privacy and stored data security.

During the enrollment phase, the system operates as it is shown in Figure 2.

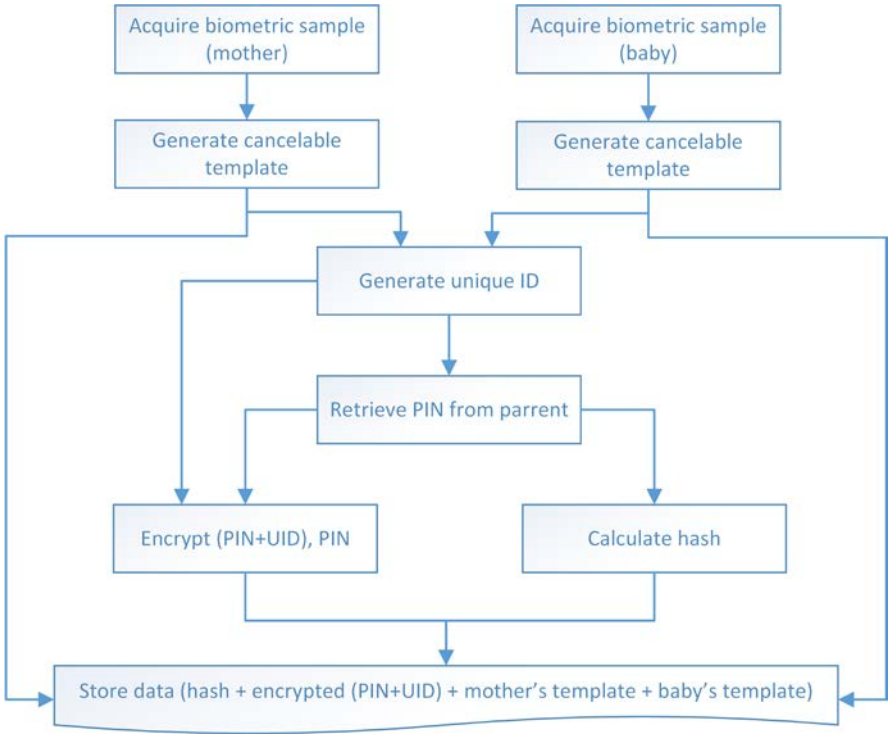


Figure 2
Enrollment phase

1) Biometric data from a mother and the baby are acquired by the sensor simultaneously. If the scan is unsuccessful, the scanning is repeated.
2) Data is preprocessed and both cancelable biometric templates are generated with non-invertible transforms.
3) Unique ID reference (UID) is generated.
4) Mother enters a 6-digit long PIN code on the the numerical keyboard and accepts it.
5) Unique ID is concatenated to and encrypted with provided PIN using symmetric encryption.
6) Hash of the PIN code is calculated.
7) A record is added to database containing the calculated hash of the PIN, the encrypted concatenation of UID and PIN, and both cancelable templates. The entire record may additionally be encrypted with a system-wide key (this functionality will be added subsequentially).

During the verification phase, the system operates as it is shown in Figure 3.

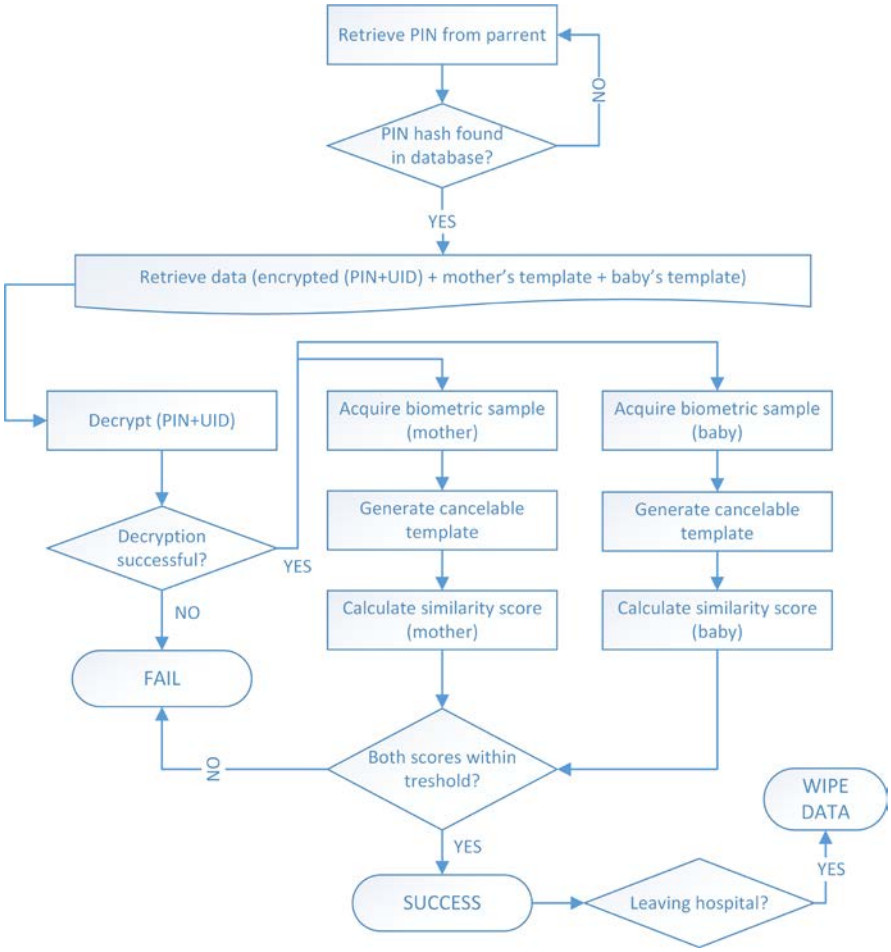


Figure 3
Verification phase

1) The mother is asked to provide a PIN code.
2) Hash of provided PIN is calculated. If no corresponding hash is found in the database, the mother is asked to retry.
3) If the corresponding hash is found, PIN is used to decrypt the stored cyphertext. If the PIN retrieved from the decrypted cyphertext matches the provided one, decryption is considered to be successful and the device continues to operate. If the decryption fails, device stops.
4) Biometric data from the mother and the baby are acquired by the sensor simultaneously. If the scan is unsuccessful, the scanning is repeated.

5) Data are preprocessed and both cancelable biometric templates are generated with non-invertible transforms.
6) Generated templates are compared with the ones stored in the database using cancelable biometrics. The verification is considered successful only if the similarity scores calculated between generated cancelable biometric templates and the one stored in the database are above predefined threshold, both for the mother and the baby.
7) If the verification is final (mother is leaving the maternity hospital with her baby) the data are securely wiped from the database.

4 Biometric Template Generation and Verification

A fingerprint is the pattern of ridges and valleys on the surface of a fingertip [17], the formation of which is determined during the first seven months of fetal development [10]. Fingerprint ridges are noncontinuous lines. Ridges end and form ridge termination points, or split and form bifurcation points [17]. These points are called minutiae points and they are used as features in fingerprint recognition systems. According to Jain et al., the accuracy of the currently available fingerprint recognition systems is adequate for verification systems and medium-scale identification systems [10].

This research adopts the minutiae extraction method presented by Jagadeesan and Duraiswamy [18]. Details on the minutiae extraction can also be found in [19]. The process involves image preprocessing, segmentation, orientation field estimation, image enhancement and minutiae extraction. The first operation applied to the image in the preprocessing phase is histogram equalization that increases the local contrast of the image, resulting in overall contrast improvement. Wiener filter [20] removes blur and additive noise from the picture without altering ridge structures. Filtered image is further segmented and regions of interest are separated from the rest of the image. Let N denote the size of the block and $\mu(I)$ the mean pixel value of the block. Block I is considered to be foreground block if its variance is greater than the threshold τ_s:

$$\sigma^2(I) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(I(i,j) - \mu(I)\right)^2 > \tau_s . \tag{1}$$

The local orientation of ridge valley structures is further estimated. This operation is also a block-wise operation and it employs gradient vectors that indicate the highest deviation of gray intensity normal to the edge of ridge lines [21]. The resulting image is enhanced by Gaussian low-pass filter followed by the 2-D Gabor filter. Let f_0 denote ridge frequency, θ the orientation of the filter, σ_x and σ_y standard deviations of the Gaussian envelope along x and y axes, and $[x_\theta, y_\theta]$

coordinates of [x, y] after clockwise rotation of the Cartesian axes by $0.5\pi–\theta$. The 2-D Gabor filter is given by:

$$G\left(x, y, \theta, f_0\right) = \exp-\frac{1}{2}\left(\frac{x_\theta^{\ 2}}{\sigma_x^{\ 2}} + \frac{y_\theta^{\ 2}}{\sigma_y^{\ 2}}\right)\cos(2\pi f_0 x_\theta), \tag{2}$$

$$x_\theta = x\sin\theta + y\cos\theta, \tag{3}$$

$$y_\theta = -x\cos\theta + y\sin\theta. \tag{4}$$

The output if 2-D Gabor filter is binarized, resulting in the image with two levels of interest: ridges (black) and valleys (white pixels). Morphological operators are applied in order to eliminate the noise resulting from spurs and line breaks, followed by thinning algorithm presented by Lam et al. [22] that reduces the width of ridge lines. The thinning algorithm segments the image into two subfields as in the checkboard pattern, employs Hilditch crossing number and defines four pixel removal conditions that are used in the iterations of the algorithm. The result of the algorithm is the image composed of one pixel wide ridges, with clearly visible minutiae, as shown in Figure 4.

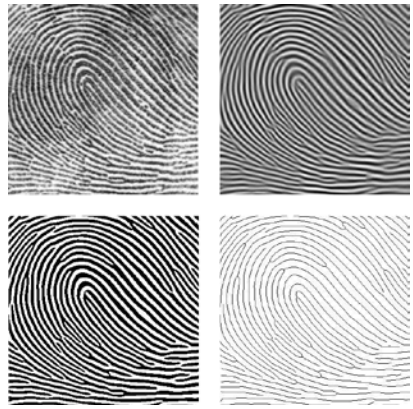


Figure 4
Original image (top left), enhanced image (top right), binarized image (bottom left) and thinned image (bottom right)

Let p_1, p_2, ... $p_8 \in [0, 1]$ denote neighbor pixels of pixel p, starting from the one on the right and counting counter clockwise. Crossing number $X_R(p)$ is calculated for each pixel in the resulting image, as the number of transitions from white to black and vice versa when points are traversed in order:

$$X_R\left(p\right) = \sum_{i=1}^{8}\left|p_{i+1} - p_i\right|. \tag{5}$$

Pixel p is identified as ridge termination point if $X_R(p) = 2$, and valley termination point (bifurcation) if $X_R(p) = 6$.

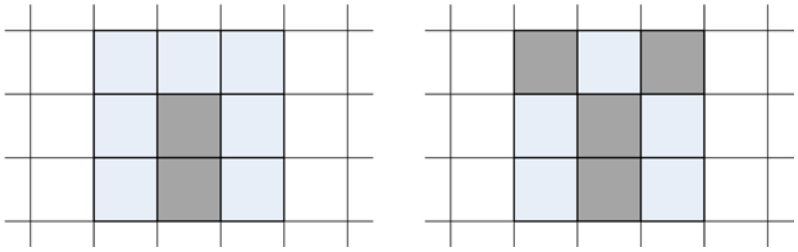


Figure 5
Termination point (centered pixel on the left) and bifurcation point (centered pixel on the right)

Fingerprint template is given by two-dimensional vector that contains cartesian coordinates of n extracted minutiae points:

$$F = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}. \quad (6)$$

Non-invertible transforms are applied to extracted feature vector to preserve privacy of stored biometric data. According to Ratha et al., the transformation is invertible if the minutiae positions after transformation are highly correlated to minutiae positions before transformation [23]. Having that said, the goal of the transform is to eliminate minutiae correlation to the maximum possible extent and provide tolerance to brute force attacks.

This research employs minutia shuffling transform that depends on the PIN. The transform proposed by authors operates as follows. The coordinate system is divided into 4 x 4 cells. Each minutiae is clockwise relocated from one cell to another depending on the one digit of PIN, while the relative position from top-left corners of the originating and destination cells remains unchanged (as shown in Figure 6). The transformation starts with the first PIN digit and top left minutiae and ends up with the one at the bottom right. Each seventh minutiae is transformed using the first PIN digit. The transform is non-invertible as is impossible to determine the original cell of the minutiae. More details on security of cancelable fingerprint templates can be found in [24]. The impact of the transform on matching accuracy is discussed in Section 5 of this paper.

Generated cancelable template consists of shuffled minutia point coordinates. Templates are generated both for the mother's and the baby's fingerprints and stored on the device alongside encrypted UID and hash of the PIN.

During the verification phase, the same procedure is performed and another vector is generated. Missing points are discarded, the sum of the squared differences between two feature vectors is calculated, normalized by the number of remaining non-discarded values, and the matching score is compared with a threshold.

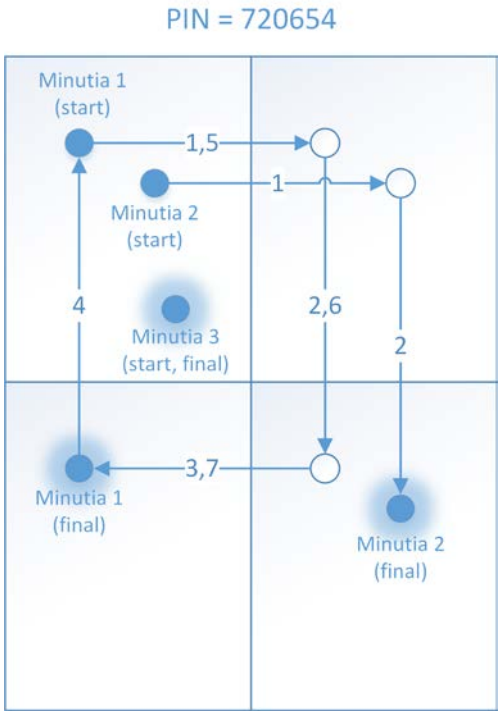


Figure 6
PIN-dependent non-invertible transform

5 Experimental Evaluation

Two sets of experiments were conducted to evaluate the proposed solution. First, experiments that provide a proof that scanning baby's fingerprints is possible were conducted. The type of sensor that provides lowest Failure to Enroll rates and high verification performance were also identified. These experiments were run with prior written permission from parents. Due to legal and ethical aspects of children fingerprint acquisition for the purpose of experimentation, images from CASIA-FingerprintV5 [25], collected by the Chinese Academy of Sciences' Institute of Automation, were used as inputs in the second set of experiments. The second set of experiments evaluates the performance of the proposed system.

5.1 Baby's Fingerprints and Commercial Scanners

In the first set of experiments, the fingerprints of a 3 month old child were acquired with three different types of commercial scanners: optical, thermal and capacitive. Optical scanners take an image of the fingerprint using a camera. The device used in this experiment is Futronic FS80H USB 2.0 scanner. Capacitive scanners use capacitors to form an image of the fingerprint from electrical current. The device used in this experiment is Eikon II Swipe. Thermal scanners sense the temperature differences on the contact surface, between fingerprint ridges and valleys. The device used in this experiment is id3 Semiconductors Certis Image that employs Atmel FingerChip thermal sensor. More information on these devices is available on Fulcrum Biometrics [26]. In this experiment, ten attempts were made to acquire a biometric template of each finger of the right hand. Acquisition results and Failure to Enroll rates (FTE) are given in Table 1.

Table 1
Baby's fingerprint FTE on optical, capacitive and thermal scanner

	Sensor type		
Finger	**Opical**	**Capacitive**	**Thermal**
Thumb	10	7	2
Index finger	10	7	2
Middle finger	10	6	1
Ring finger	10	5	0
Pinky	9	4	0
% successful	98%	56%	10%
Failure to Enroll (%)	2%	44%	90%

According to the results given in Table 1, it can be concluded that the optical scanner outperforms both capacitive and thermal ones. The optical scanner failed to enroll only the pinky finger once, due to smaller size of ridge structures. Due to acceptable 2% Failuire to Enroll rate, the optical scanner is further used to evaluate the success of identity verification. The templates of each finger are created and stored, and each finger is further verified against 10 reacquired biometric samples. The optical scanner failed to verify both ring finger and pinky twice, which results in 4% False Rejection Rate (FRR).

5.2 Experimental Evaluation of the Proposed System

The Performance of the proposed solution is experimentally evaluated using MATLAB R2015a (feature extraction) and Python 2.7 (matching, cancelable template generation and scripting). We have conducted our experiments as follows. CASIA database contains five images of each finger for 500 subjects. A set containing 10 pairs of index finger images was created and used to enroll

genuine users. One image in each pair represents the mother and the other represents the child. Each pair was associated with UUID and randomly generated 6-digit PIN during the enrollment phase. The accuracy of the system was measured by verifying each pair of fingerprints with: 10 pairs that contain remaining images belonging to those people (images not used to enroll users), 5 pairs that contain genuine image of one person and an imposter image for the other person and 5 pairs that contain both images belonging to imposters. Matching scores have been scaled to range [0, 1] and compared with several threshold values. The verification is considered to be successful if matching scores for both fingers in the pair are above the threshold. The definition of False Acceptance Rate (FAR) and False Rejection Rate (FRR) is slightly altered as the system verifies two fingerprints at the same time. False acceptance in this scenario occurs if the system identifies at least one imposter fingerprint in the pair as genuine user during the verification. False rejection in this scenario occurs if the system identifies at least one genuine fingerprint in the pair as imposter. Experiments were conducted with the aim to evaluate the overall accuracy of the fingerprint matcher, the impact of cancelable biometric template protection and overall security of the system.

First the fingerprint matcher that does not involve cancelable biometric templates and PIN protection was tested. The results of the experiments are given in Table 2 and graphically presented on Fig. 7.

Table 2

Verification summary of the dual fingerprint matcher for different treshold values

Verification matrix			**Treshold**					
Input pair	#	**Identified as**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**
Both genuine	100	Genuine	96	97	99	99	100	100
		Imposter	4	3	1	1	0	0
Genuine and imposter	50	Genuine	1	1	2	2	5	7
		Imposter	49	49	48	48	45	43
Both imposter	50	Genuine	0	1	1	1	4	5
		Imposter	50	49	49	49	46	45
False rejection rate (%)			4%	3%	1%	1%	0%	0%
False acceptance rate (%)			1%	2%	3%	3%	9%	12%

Experimentally obtained results given in Table 2 were not surprising as false rejection rate decreases and false acceptance rate increases with the threshold being loosen up (see Fig. 7).

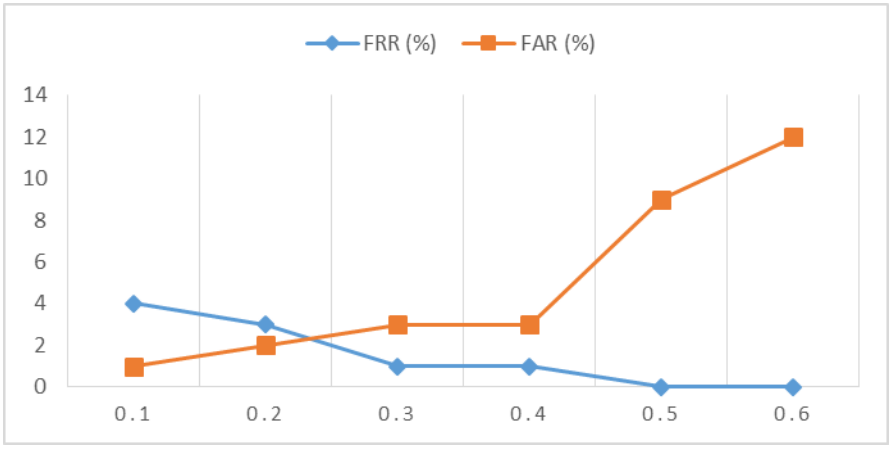


Figure 7
FAR and FRR of the dual fingerprint matcher

The second set of experiments introduces cancelable biometrics to the system. The results of the fingerprint matcher that employs cancelable biometric template generation algorithm proposed by authors are given in Table 3.

Table 3
Verification summary of the dual fingerprint matcher that employs cancelable biometric templates for different threshold values

Verification matrix			**Treshold**					
Input pair	**#**	**Identified as**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**
Both genuine	100	Genuine	93	94	96	98	99	100
		Imposter	7	6	4	2	1	0
Genuine and imposter	50	Genuine	1	2	2	3	7	9
		Imposter	49	48	48	47	43	41
Both imposter	50	Genuine	1	1	1	2	6	6
		Imposter	49	49	49	48	44	44
False rejection rate (%)			7%	6%	4%	2%	1%	0%
False acceptance rate (%)			2%	3%	3%	5%	13%	15%

As we have expected, the non-invertible transform has introduced additional errors to the matcher. FRR is increased for lower threshold values, while FAR increases to the unacceptable level for loosen thresholds. Optimal threshold for this system is found via equal error rate (EER), as presented on Fig. 8.

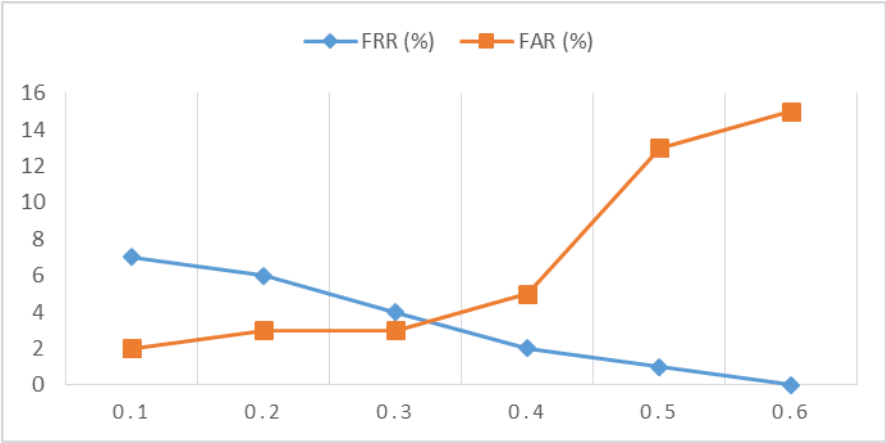


Figure 8
FAR and FRR of the dual fingerprint matcher that employs cancelable biometric templates

Threshold value $t = 0.325$ results with an 3.5% EER. This error is slightly higher when compared to a system that does not employ cancelable templates: 2.3% EER occurs on threshold value $t = 0.23$. Having in mind that cancelable biometrics protects the privacy of stored templates and prevents possible malicious misuses, this small increasing of EER is further treated as negligible.

Once optimal EER is determined, the PIN and cryptographic protection are further introduced to the system. Minutiae points were extracted from the set of CASIA fingerprint images and Python scripts were run to simulate the enrollment and verification phases of the entire system, as presented on Fig. 2 and Fig. 3. PINs were generated randomly during the enrollment and stored in a sepatate file, as they are required for the verification. With the exception of the PIN being stolen or compromised in any other way, it can easily be concluded that the cryptographic protection eliminates any chance of false accepts occuring. 3.25% False Rejection Rate results with one reject in approximately 30 attempts, which is an acceptable result for intended application as users are allowed to repeat the verification step if it fails.

5.3 Discussion

As stated in the section 3 of the paper, the main goal was to reach a very high level of proof of maternity using noninvasive biometrics and provide the sufficient level of security with cancelable biometrics (non-invertible transforms). Although no dual fingerprint scanners are found in National Database of Patents, the Republic of Serbia [15], some commercial solutions do exist on the market, such as [27]. By conducting a detailed background research, these are used to enroll or verify the identity of one person using two different fingers, thus reassembling

single trait – multiple units scenario in multimodal biometric system classification of Ross and Jain [28]. Our approach to dual fingerprint scanning is different and unique: it generates cancelable templates from two fingers belonging to different people and makes a bond between them, thus finding an application in maternity verification and identity theft prevention in maternity wards.

The performance and security of the proposed solution, as well as some ethical aspects behind the application of the device are further discussed. We have identified the performance of the system as satisfactory: according to the experimental evaluation, system rejects 3.5% genuine users. However, users are allowed to retry and therefore this does not set an obstacle to keeping the high level of security by employing tight threshold and non-invertible transform. The cryptographic protection of auxiliary data eliminates the chance of false acceptance as only the users who know the PIN can proceed to the verification. The security of the system depends on several factors, and an adversary who wants to steal the baby, for example, would have to override the auxiliary data encryption and non-invertible tranform. Some may set up ethical obstacles to the application of this device by asking the following question: "can the biometric templates be misused?" The answer is no. Even if the PIN is known, it is impossible to regenerate original templates from cancelable ones using anything other than brute force. A brute force attack on 4x4 cell shuffler would require in $(4x4)^{(4x4)}$ attempts, which is approximately 18.5×10^{18} attempts. And all the data are securely deleted from the storage once the mother and the child leave the hospital.

By conducting informal inquiries, one critical issue in the whole approach was identified: the problem of forgetting PIN codes. Mother is emotionally driven at the time of birth and a reasonable chance exists for her to forget the PIN she has entered right after birth. Authors are discussing several methods that will provide recovery of the data, such as some key escrow techniques.

Conclusions

Although various anti baby-switching procedures are implemented in hospitals world wide, no 100% proof maternity verification method is reported in the literature or, to the best of our knowledge, used in any maternity wards. The major contribution of the approach presented in this paper based on a patented dual fingerprint scanner is the elimination of anxiety caused by baby-switching and the provision of a proof to each mother that she will leave the maternity hospital with her own child. Additionally, each mother may require biometric verification of her baby during breast-feeding, for example. The proposed approach also removes the possibility of subsequent lawsuits and traumas resulting from baby-switching. As the device stores the templates processed with PIN-dependent non-invertible transforms and securely wipes them when the family leaves the maternity hospital, the concern about biometric template privacy or identity theft is also eliminated. As an average a high-quality fingerpint scanner costs no more that $100-$150 and

the algorithms presented in this paper can be easily implemented on a low-cost hardware, it is expected that this device will be affordable for each hospital.

References

[1] B. Gaille: 20 Rare Babies Switched at Birth Statistics. Tips from the Blog Millionaire. July 29, 2014

[2] Williston Herald. November 11, 2009

[3] John D. Homan, TheSouthern.com, April 11, 2008

[4] Steve Rosenberg, BBC News, Moscow, October 10, 2011

[5] T. R. Crane: Mistaken Baby Switches: an Analysis of Hospital Liability and Resulting Custody Issues. Journal of Legal Medicine, 21(1), pp. 109-124, 2000

[6] J. D. Davis, M. K. Moran, E. O. Horger III, A. N. Dajani: Pregnancy Anxieties and Natural Recognition in Baby-Switching. British Journal of Nursing, 10(11), pp. 718-726, 2001

[7] R. R. Rusting: Baby Switching: an Underreported Problem that Needs to be Recognized. Journal of Healthcare Protection Management: Publication of the International Association for Hospital Security, 17(1), pp. 89-100, 1999

[8] H. G. Dahlen, S. Caplice: What do Midwives Fear?. Women and Birth, 27(4), pp. 266-270, 2014

[9] A. K. Jain, A. Ross: Introduction to Biometrics. In "Handbook of Biometrics", A. Jain et al. (Eds), Springer, 2008

[10] A. K. Jain, A. Ross, S. Prabhakar: An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, pp. 4-20, 2004

[11] K. Moore, T. V. N. Peraud, M. Torchia: Before We Are Born, 9th Edition. Elsevier UK, 2014

[12] G. Schoenwolf, S. Bleyl, P. Brauer, P. Francis-West: Larsen's Human Embryology, 5th Edition. Elsevier Health Sciences, Churchill Livingstone, 2014

[13] A. R. Kavsaoğlu, P. Kemal, M. R. Bozkurt: A Novel Feature Ranking Algorithm for Biometric Recognition with PPG Signals. Computers in Biology and Medicine 49, pp. 1-14, 2014

[14] K. Lalović, M. Milosavljević, I. Tot, N. Maček: Device for Biometric Verification of Maternity. Serbian Journal of Electrical Engineering, Vol. 12, No. 3, pp. 293-302, 2015

[15] http://www.zis.gov.rs/pocetna.1.html

[16] C. Lee, H. S. Shin, J. Park, M. Lee: The Optimal Attachment Position for a Fingertip Photoplethysmographic Sensor with Low DC. Sensors Journal, IEEE, 12(5), pp. 1253-1254, 2012

[17] T. Van der Putte, J. Keuning: Biometrical Fingerprint Recognition: don't Get Your Fingers Burned. In Smart Card Research and Advanced Applications (pp. 289-303) Springer US, 2000

[18] A. Jagadeesan, K. Duraiswamy: Secured Cryptographic Key Generation From Multimodal Biometrics: Feature Level Fusion of Fingerprint and Iris. International Journal of Computer Science and Information Security, Vol. 7, No. 2, pp. 28-37, 2010

[19] N. Maček, B. Đorđević, J. Gavrilović, K. Lalović: An Approach to Robust Biometric Key Generation System Design. Acta Polytechnica Hungarica, Vol. 12, No. 8, pp. 43-60, 2015

[20] H. Furuya, S. Eda, T. Shimamura: Image Restoration via Wiener Filtering in the Frequency Domain. WSEAS Transactions on Signal Processing, 5(2), pp. 63-73, 2009

[21] Y. Wang, J. Hu, F. Han: Enhanced Gradient-based Algorithm for the Estimation of Fingerprint Orientation Fields. Applied Mathematics and Computation, Special Issue on Intelligent Computing Theory and Methodology, Vol. 185, No. 2, pp. 823-833, 2007

[22] L. Lam, S. W. Lee, C. Y. Suen: Thinning Methodologies - a Comprehensive Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(9), pp. 869-885, 1992

[23] N. K. Ratha, S. Chikkerur, J. H. Connell, R. M. Bolle: Generating Cancelable Fingerprint Templates. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(4), pp. 561-572, 2007

[24] C. Li, J. Hu: Attacks via Record Multiplicity on Cancelable Biometrics Templates. Concurrency and Computation: Practice and Experience, 26(8), pp. 1593-1605, 2014

[25] Biometrics Ideal Test, http://biometrics.idealtest.org

[26] http://www.fulcrumbiometrics.com/

[27] http://www.m2sys.com/dual-fingerprint-biometrics/

[28] A. Ross, A. K. Jain: Multimodal Biometrics: An Overview. In Signal Processing Conference, 2004 12th European, pp. 1221-1224, 2004, IEEE

Hydrodynamic Modelling and Analysis of a New-developed Mobile Refrigerated Container

László Kátai*, Tibor Várszegi*, István Oldal*, László Zsidai**

*Szent István University, Faculty of Mechanical Engineering, Institute of Mechanics and Machinery, Páter K. u. 1, 2100 Gödöllő, Hungary, e-mail: katai.laszlo@gek.szie.hu, varszegi.tibor@gek.szie.hu, oldal.istvan@gek.szie.hu

**Szent István University, Faculty of Mechanical Engineering, Institute of Industrial Technologies, Páter K. u. 1, 2100 Gödöllő, Hungary, e-mail: zsidai.laszlo@gek.szie.hu

Abstract: In this paper the hydrodynamic air flow relations in two adjustable compartment units of the mobile refrigerated container were examined. Analytical solution of the hydrodynamic equations describing the cooling air flow motion is not possible. We applied two possible methods for the solution: (1) – small scale test, (2) – approximate solution of the ordinary differential equations with finite elements simulation. The objective of the performed tests is to define the cooling air flow kinematical conditions, such as air velocity distribution in empty and fully loaded state, testing the shape of the air inlet profile and furthermore to determine the necessity of auxiliary ventilators adaptation. According to the performed test results a proposal to the final development of the cooling and ventilation elements has been made.

Keywords: cfd analysis; refrigerated container; small scale test; 3D model; FEM

1 Introduction

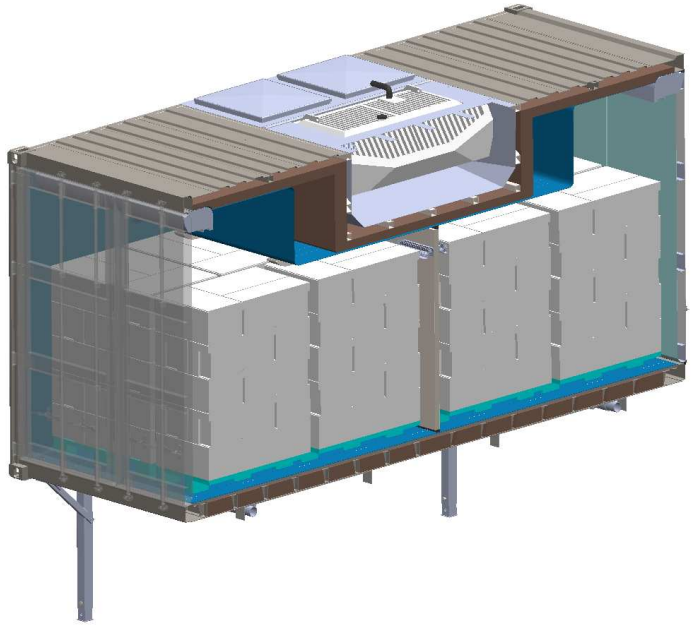
Due to HACCP requirements the storage and transportation of food products and food raw materials at a suitable temperature is an extremely important issue in logistics. Currently, the availability of technical solutions regarding the transportation part of the cooling chain is limited; the use of transportation vehicles with refrigerated cargo holds is the most common. However, even if for a short time, the cooling chain can be compromised due to technical deficiencies and human error. The solution to the technical deficiencies is the design of intelligent refrigerated compartments to be implemented in the project. One of the conditions of efficient and reliable refrigeration for the whole load is evenly distributed flow in the whole compartment.

In this article we have examined air flow distribution in the newly-designed mobile refrigerated container compartments, for which work the background is provided by the research and development GOP project consortium.

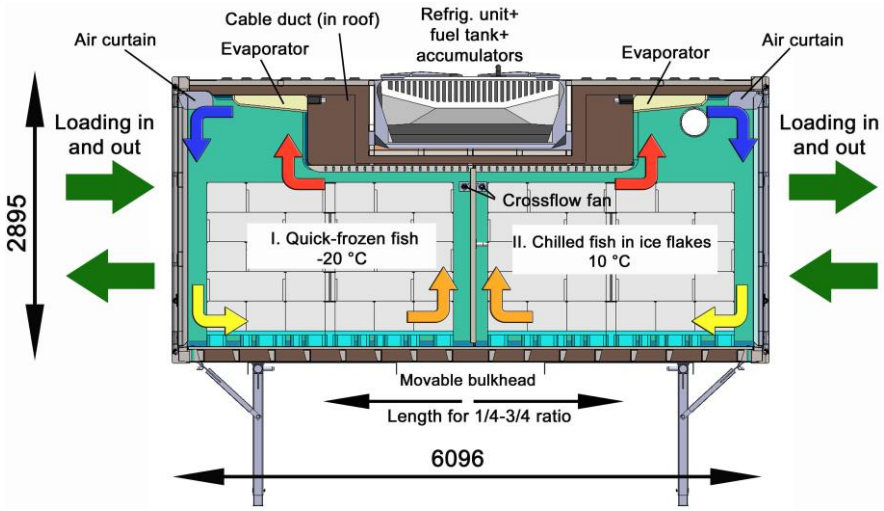
1.1 The Structure and Operation of the Designed Cooling Container

Extensive international literature deals with container transportation and the structure of containers. [1…3, 7, 8] The selected container type determines the whole transportation system since emission level functions of different pollutants are also applicable as route costs of transportation assignment problems, thus taking fuel consumption and environmental impact into account during process design. [6, 9]

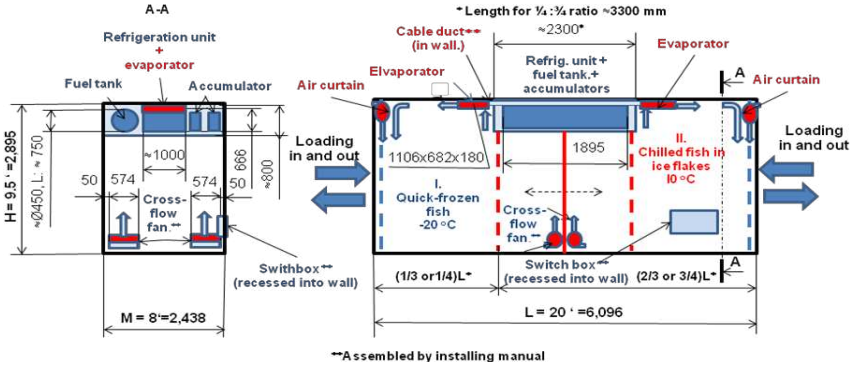
The designed container shown in Figure 1 can be loaded via both ends, it is suitable for road and rail transportation, it is 20 feet long, with 2 temperature settings, adjustable compartments, and assembled with a diesel refrigeration unit.



a)



b)



c)

Figure 1

a) A Perspective view of the container b) The cooling characteristics, flow directions c) The diagram of the cooling system and air circulation

Its main characteristics: (1) – own diesel engine for refrigeration unit, (2) – 2 refrigeration compressors (one diesel unit + a second electrical one for stand-by state), (3) – cross flow air cooling of condenser of the refrigeration unit relative to the direction of travel of the vehicle, (4) – refrigeration unit is located in the gravity axis of the container, (5) – easy accessibility of the refrigeration units to assemble and maintain them, (6) – installation of DC cross flow fans, (7) – double air curtains, (8) – installed with GPS positioning and temperature transmitter. The applied 3 cylinder 4-cycle liquid cooled 0.854 L diesel engine has a 16 kW maximum output power (at 3600 rpm). Due to its overall dimensions (LxWxH 518 x 454 x 576 mm) it can easily fit into the space and has been designed in the

middle section of the container roof. The diesel drive of the refrigeration unit makes the cooling of compartments independent of the speed and direction of travel of the vehicle and can be even used in the stand-by position. The independent diesel engine consumes less fuel and reduces the impact on the environment. The installation of two separate refrigeration compressors increases the safety of refrigeration transportation. The dual energy supply (diesel and electrical) makes the refrigeration system more flexible and makes it possible (in stand-by state) to specify and modify the type of driving power depending on the cost of energy sources. Owing to the cross direction position of the condensing unit relative to the direction of travel of the vehicle the operation of refrigeration unit becomes independent of the direction of travel of the vehicle [4, 5]. Due to the mounting of the refrigeration unit and its accessories in the gravity axis of the container the route traceability of the vehicle is safer. The refrigeration unit and its accessories can be accessed by a built-in telescopic ladder so they can be assembled and maintained on the spot and can also be removed. The cross flow fans making the airflow more uniform within the container spaces are located and fixed on the movable bulkheads, they can be moved along with the bulkheads, and can be powered by DC accumulators both in road and stand-by state. The temperature of compartments is decreased by about 10 degrees centigrade by the application of the double air curtains when the container is opened for loading in and out. The GPS system makes it possible to track the route, the mobile and stand-by position of the vehicle, its fuel consumption and the temperature of the compartments, furthermore if the refrigerated mobile container breaks down; it helps organize a rescue action.

Due to the "unique" design of the container it had become obvious that the establishment of the flow conditions of the (cooling) air was not possible through analytical mathematical methods. Therefore, in order to justify the "concept" "plank" (experimental model) measurements and finite-element simulation were performed. Numeric simulation is used by Son H. Ho et al. (2010) to examine the flow conditions of cooling air for cooling-houses with good results [10]. Several studies deal with CFD temperature distribution analysis of refrigerated containers [11, 12].

2 Material and Method

2.1 Tools of Affinity Experiments

The measurements were taken in the controlled temperature range - Tc = (0÷2) ^{o}C – with cooled air, while in the laboratory the temperature of the ambient air surrounding the experimental model was: Ta = 23 ^{o}C.

The experimental model simulates one of the compartments of the container to be developed in which air of T_{air} = (0 ÷ 2) °C temperature circulates. For this, i.e. to create the experimental model the existing FRIGOR – BOX type cooling chamber, recommended for T = (0 ÷ 2) °C temperature was used.

The ratio of the plank model (experimental model):

$$i = 2226/1140 = 1{,}952 \sim 2$$

This resulted from the fact that the width of the experimental model chamber was the same as the original width of the FRIGOR – BOX cooling chamber, which was: sk = 1140 mm.

Figure 2 shows the structure and the main components of the experimental model.

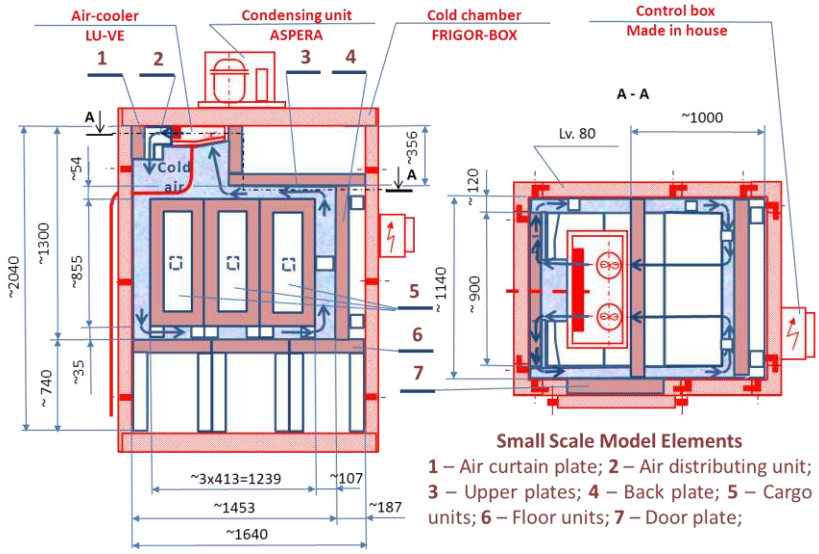


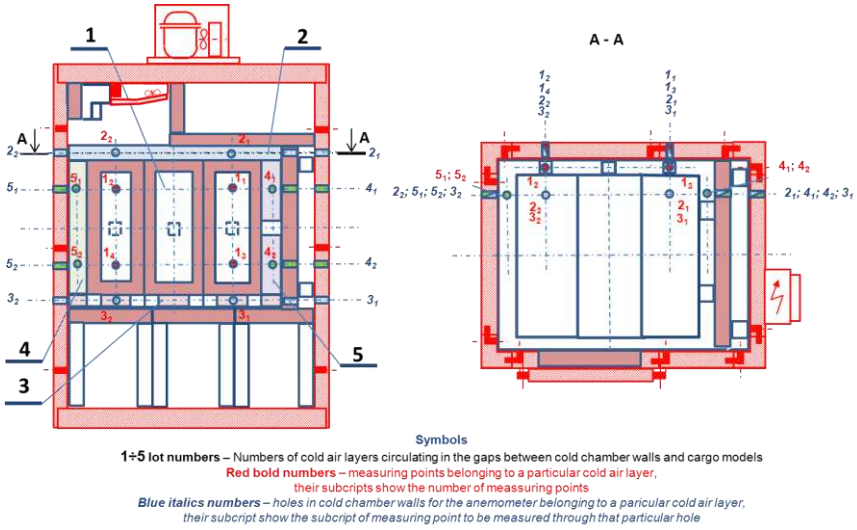
Figure 2
The sketch, dimensions and components of the experimental model

The construction of the plank model (experimental model) (see Figure 2): was made by converting the FRIGOR – BOX cooling chamber, by inserting walls from Styrofoam components. The width of the cargo hold of the experimental model was identical to the original width of the cooling chamber, sk = sm. However, the shape, the height (hk) and the length (lk) of the original cooling chamber was different from the shape and the dimensions calculated above of the prototype in the original development concept and the corresponding plank (experimental model): [hk> (hm + hmb)], and (lk> lm). Therefore, an upper plate (3), back plate (4), floor unit (6) and door plate (7) - to eliminate the door recess - were built into the original chamber. With this, a cargo hold corresponding to the shape of the prototype cooling chamber and dimensions corresponding to the size reduction were generated on the plank model (experimental model).

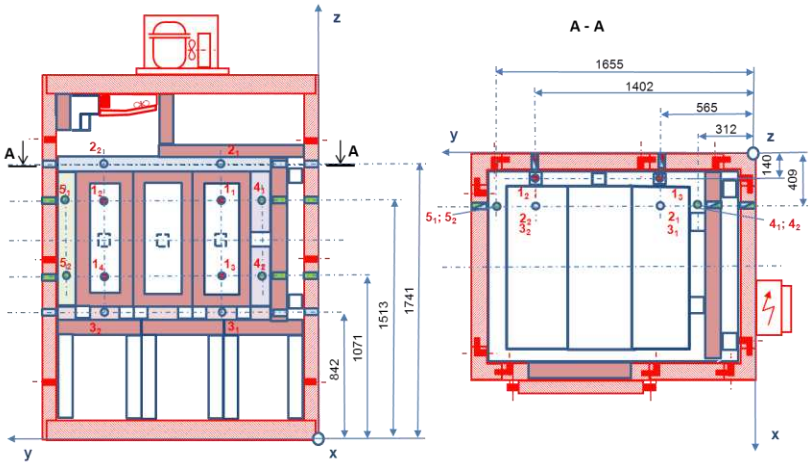
In the plank model the velocity of the air flowing in the gaps between the plank model and the cargo model was measured by a TESTO 435 anemometer. The test kit contained both thermal (hot wire) and wheel sensors which could be connected to a common measuring, display and data recording unit.

The thermal (hot wire) sensor measured air temperature in addition to air velocity.

The measurement points were determined in the middle of the gaps of the experimental model, in the midline of the width of the two outer cargo models, at the same distance from the corners, on one side of the longitudinal symmetry line of the experimental model because the refrigerated container is symmetric to its longitudinal axis. It was assumed that in the other (mirror image) part the flow pattern and the velocities were identical. The marking of the measuring points is shown in Figure 3/a, their location, coordinates are shown in Figure 3/b. In the case of the marking of the measurement location: the number marks the serial number of the gap, the corresponding index marks the serial number of the bore. The number of measurement points: 12, the number of bores: 16, because in the horizontal gaps the velocities in the direction of axes "x" and "y" must be measured through two separate bores. The bores were plugged in with plastic pipe caps when not being measured.



a)



b)

Figure 3
a) Marking of the measurement points and the inlet of the sensors, b) Coordinates of the measuring points and the bores serving as inlet and fixing for the sensors

2.2 Method of the Experimental Model Experiment

The velocity of the air distributed and circulated by the fan of the cooling equipment was measured in the gaps between the experimental model (plank model) and the cargo model - in the direction of space coordinates "x", "y" and "z".

2.2.1 The Examination of Similarity

The cooling air flows in the gap between the inner walls of the refrigerated container and the cargo in all the – i.e. x - y - z – directions of space. Physical similarity was expressed with the identical Reynolds (Rem) number – Rep = Rem of the prototype (Rep) and the gaps of the experimental model [13, 14, 15, 16]. This can be defined with the equation of the Reynolds number as:

$$Re_p = (d_p \times v_p)/\nu = Re_m = (d_m \times v_m)/\nu$$

where: ν is the kinematic viscosity of air, which is the same on the prototype and in the case of the experimental model, so it drops out of the equation,

d_p and d_m – the identical diameter of the gaps on the prototype (p – index), and on the experimental model (m – index), m,

v_p and v_m – the flow velocities in the gaps on the prototype (p – index) and on the experimental model (m – index), m/s.

Mentioned before, the cooling media (cold air) used had the same temperature both in the experimental model and the prototype. Thus, the "Pr" number (Prandtl number) of the cooling media is, also, the same in both cases:

$$Pr_p=(c_p \times \mu)/k = Pr_m$$

Where: P_{rp} and P_{rm} – Prandtl number for cooling air in prototype (p – index) and in experimental model (m – index)

c_p – specific heat at constant pressure, kJ/kg K

μ - Dinamic viscosity, Pa s

k – thermal conductivity, W/m K

In order for the results of the hydrodynamic experiments conducted on the experimental model to be transferred to the prototype, the "geometric" and "physical (flow)" similarity between the experimental model and the prototype equipment is fundamental. The experimental model was created and the measurements taken were performed in light of these principles [17, 18].

2.3 The Method of the Hydrodynamic Simulation

The steady state of hydrodynamic modelling is examined. This means the cargo is in the compartment, the door is closed, the cooling unit is working continuously and at a constant output. The freezer and cooling compartments are examined together, from a hydrodynamic point of view with almost identical geometries identical, hydrodynamic state is assumed. The geometries must be examined one by one in the case of each partition wall position. In the longest case, which is the most unfavourable from a hydrodynamic point of view, it is examined for several aspects [19].

Geometric modelling

During geometric modelling the inside air is modelled, all connected bodies are considered as outer contour. The body and the peripheral conditions are also symmetric, so it is sufficient just to model half of the body and the other half is modelled with symmetry peripheral condition (green in the figure). (Except in the case of asymmetric cargo because then the model of the entire body is needed.)

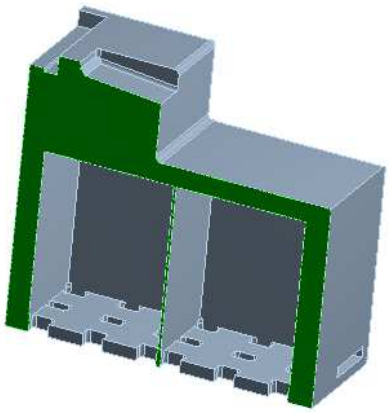
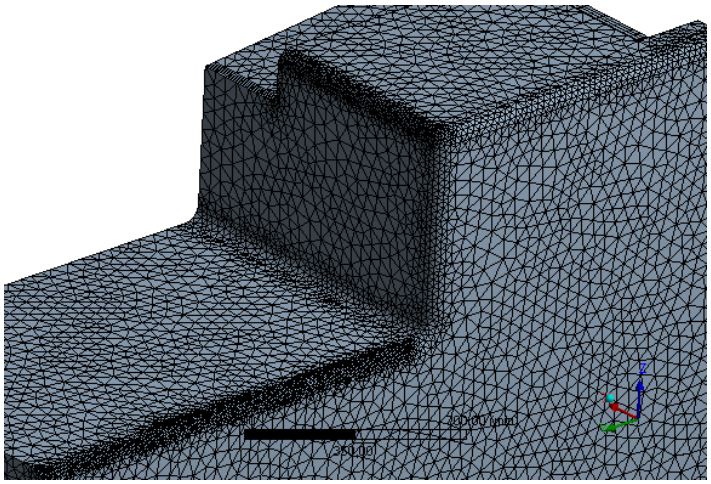


Figure 4
Geometric model of the examined space

Finite-element modelling

In the case of hydrodynamic simulation another important part of modelling is breaking down to elements. With modelling in addition to the mesh of the necessary size the special background layer mesh is also necessary along the walls.



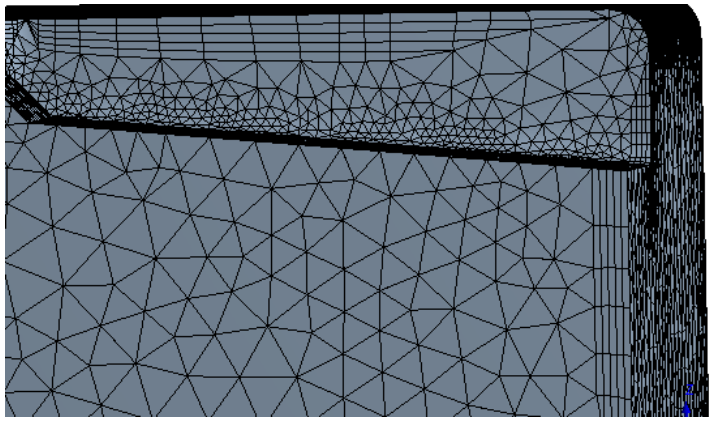


Figure 5
Boundary layer mesh

When modelling the environment, the air flowing in and out of the cooling unit as well as the air flowing in and out of the auxiliary fan, furthermore, the symmetry and the wall impact must be also considered. The modelling of the wall impact is shown in Figure 6.

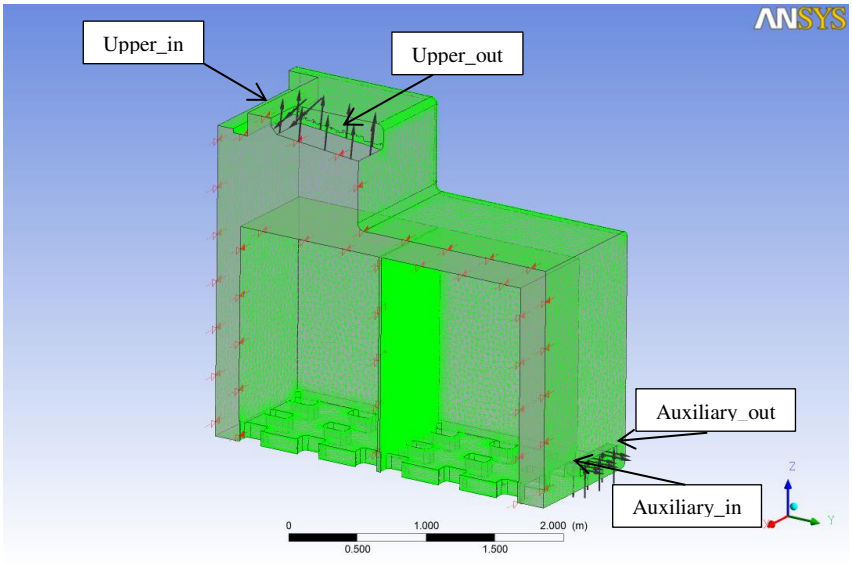


Figure 6
Modelling the wall impact

The material and the settings of the simulation (e.g. the turbulence model) are given in the following table:

Table 1
Simulation settings

Materials	
Air at 25 C	
Fluid Definition	Material Library
Morphology	Continuous Fluid

Settings	
Buoyancy Model	Non Buoyant
Domain Motion	Stationary
Reference Pressure	1.0000e+00 [atm]
Heat Transfer Model	Isothermal
Fluid Temperature	25 [C]
Turbulence Model	k epsilon
Turbulent Wall Functions	Scalable

The boundary conditions was set according to the following table:

Table 2

Boundary conditions

Boundaries	
Boundary –Upper_in	
Type	INLET
Location	**Upper_in**
Settings	
Flow Regime	Subsonic
Mass And Momentum	Normal Speed
Normal Speed	9.5000e+00 [m s^-1]
Turbulence	Medium Intensity and Eddy Viscosity Ratio
Boundary – Auxiliary_in	
Type	INLET
Location	Auxiliary_in
Settings	
Flow Direction	Normal to Boundary Condition
Flow Regime	Subsonic
Mass And Momentum	Mass Flow Rate
Mass Flow Rate	5.2900e-02 [kg s^-1]
Turbulence	Medium Intensity and Eddy Viscosity Ratio

Boundary –Upper_out	
Type	OUTLET
Location	Upper_out
Settings	
Flow Regime	Subsonic
Mass And Momentum	Mass Flow Rate
Mass Flow Rate	7.2900e-02 [kg s^-1]
Boundary - Auxiliary_out	
Type	OUTLET
Location	Auxiliary_out
Settings	
Flow Regime	Subsonic
Mass And Momentum	Mass Flow Rate
Mass Flow Rate	5.2900e-02 [kg s^-1]

Since the container has a movable partition wall, in the case of the studied space part simulation was performed for several geometric settings (position of the bulkhead). The layout of the studied cases is illustrated in Figure 7.

The aim was to determine the flow velocities in the whole geometry for the listed cases. The boundary conditions of the simulation were set up in the following way:

- The evaporator fans perform the blowing in; the particulars of the evaporator fan:

- opening: about 88 x 600 mm
- airflow: about. 1680 m3/h
- blow-in velocity: 9.5 m/s
- throw distance: about 7.5 m
- type: 2 axial fans, pressure: about 50 mm H_2O

- Auxiliary fans:

type: 2 ZIEHL-ABEGG QR 06A-50BP 90o, cross-flow fan airflow: 610 m3/h.

Accordingly, the boundary conditions were given in the following way:

- evaporator blow-in velocity: 9.5 m/s,
- suction side of evaporator 1680 m3/h air flow = 0.5883 kg/s, only half of this is modelled due to the use of the symmetry, thus the airflow: 0.29167 kg/s,
- auxiliary fans: 610 m3/h = 0.2118 kg/s on both the outlet and inlet side.

There are three simulations for each geometric model:

- the auxiliary fans are at the bottom of the partition wall,
- the auxiliary fans are at the top of the auxiliary fans,

the auxiliary fans are not switched on.

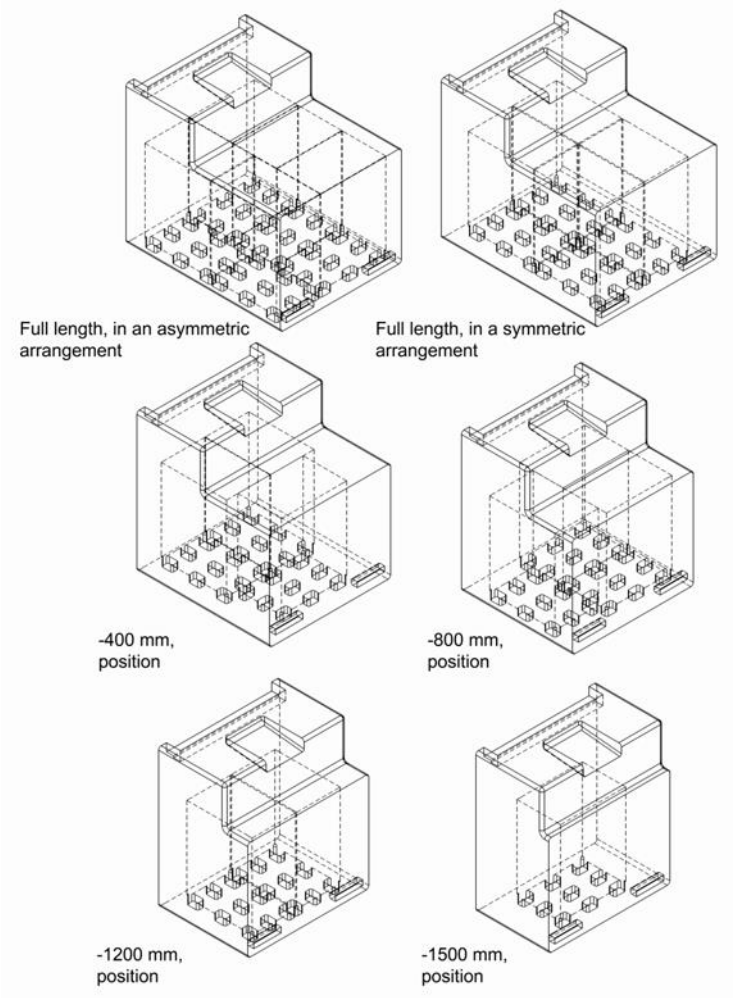


Figure 7
Layouts of the studied inner space

3 Results and Discussion

3.1 Results, Findings of the Experimental Model Experiment

During the measurement the following were modified:

- the "loadedness" of the experimental model, which means that measurements were taken in two states – loaded (with cargo model) and unloaded (without cargo model),
- thedelivery of the fan (V) and the velocity of the blow-in (v_o).

The measurement parameters are summarized in Table 1.

Table 1

Measurement settings

Serial number of the measurement setting	Delivery of the fan (V) m^3/h	Blow-in velocity of the air (v_o) m/s	Loadedness of the experimental model
1	850	4.75	empty (without cargo)
2	850	4.75	loaded
3	651	3.64	loaded

The circulation of the air in the gaps in the symmetry planes

The profile of the air flow velocity is determined from the test results of each measurement setting around the experimental model cargo in the gaps in the symmetry planes "xy", "xz" and "yz" perpendicular to each other – going through the centre of gravity of the cargo model. The illustration of the symmetry planes is shown in Figure 8. It is determined how the air velocity changes while it goes around the gaps in these symmetry planes, from being blown out of the fan until being suctioned in the fan, during the circulation occurring in the given plane. For a better visual representation, the speed profiles derived from the measurement data are illustrated graphically, in figures.

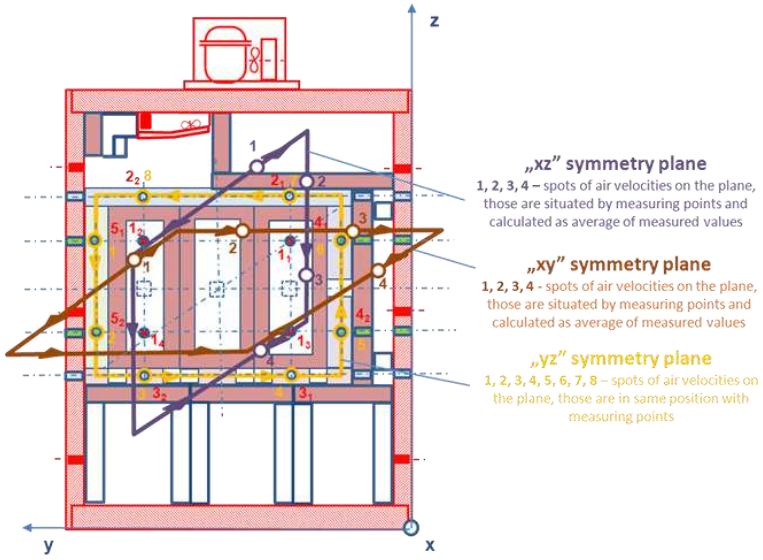


Figure 8

Interpretation of the symmetry planes, the gaps and the flow (circulation) directions.

As an example, Figure 9 shows the measured speed profiles in the case of measurement setting 1.

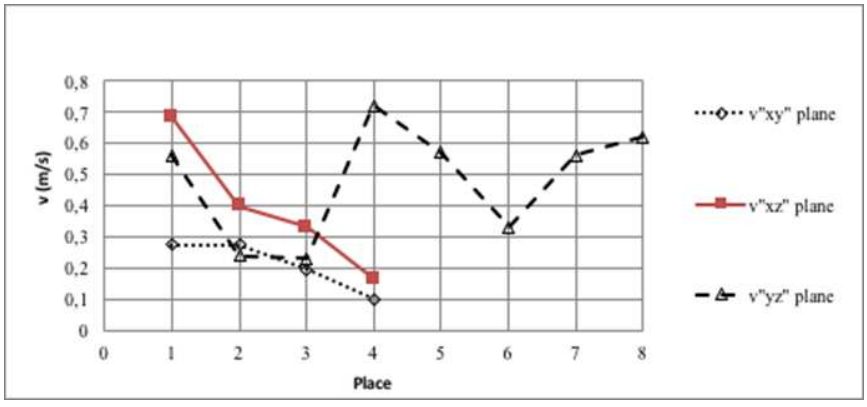


Figure 9

Profile of the air flow velocity: in symmetry plane "xy",in symmetry plane "xz", in symmetry plane "yz",with empty experimental model

Figure 10 shows the results obtained from measurement setting 2, and Figure 11shows the results obtained from measurement setting 3.

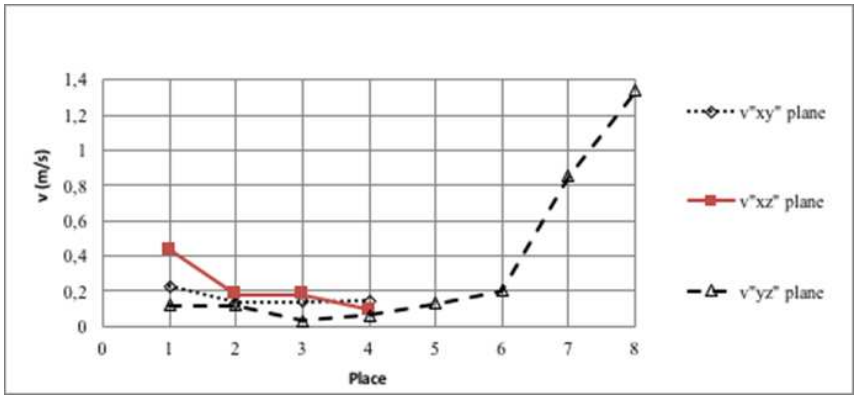


Figure 10

Profile of the air flow velocity:in the gap in symmetry plane "xy",in the gap in symmetry plane "xz",in the gap in symmetry plane "yz", in loaded state with delivery rate of V_2 = 850 m^3/h

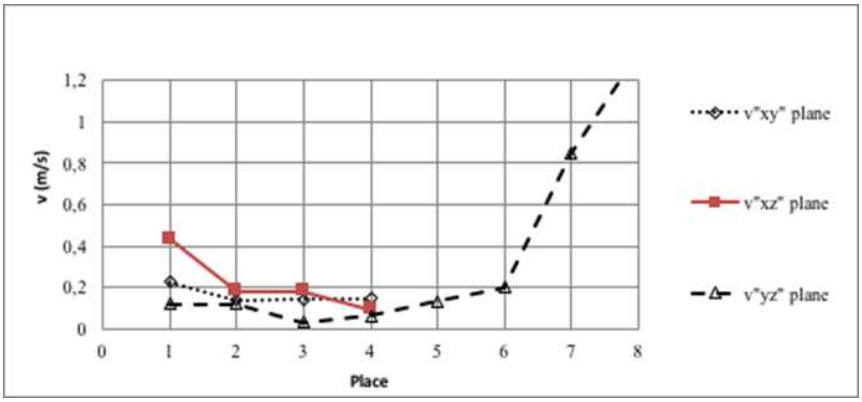


Figure 11

Profile of the air flow velocity:in the gap in symmetry plane "xy",in the gap in symmetry plane "xz",in the gap in symmetry plane "yz", in loaded state with delivery rate of V_3 = 651 m^3/h

General findings:

The thermal anemometer used for measuring the velocities and temperatures showed significantly varying air velocities many times in the same points, furthermore, it showed slightly different temperatures in the same points, but turned into different directions. Therefore, in the case of both parameters the average values are included in the figures showing the evaluation.

The alternation, change of the velocities and the temperatures can be explained by the turbulent flow of the air. Since the measurement points in the experimental model are situated "relatively" near the corners of the cargo hold, turbulent flow was "expected".

Based on the temperatures measured during the tests which fluctuated between 0 and 10 ^{o}C it can be concluded that the air only cooled down to the value $T_{air,b}$ = (0 ÷ 2) ^{o}C set on the thermostat near sensor "A" of the evaporator. Elsewhere, it was much warmer than this, which can be explained by the fact that the walls making up the cargo hold of the experimental model and the cargo model did not cool down to the above temperature, so they projected heat energy into the air. This also influenced the air flow, but the deviances caused by the temperature differences of cooling air were negligible for the calculation of the results, see physical parameters of air in reference of Kiss R. (1980) [20].

At the same time, the results measured for temperature have indicated that so-called "heat recesses (temperature peaks)" can come into effect in the compartments of the container, furthermore underlined the importance of proper pre-cooling of both the compartments and the cargo prior to loading in, separately.

For solving the problems caused by fluctuating and not proper air temperature, on one hand a new telemetric system has been developed in the project and that among others applied several temperature sensors locating in the area of temperature peaks to control the refrigeration system. Those devices are not discussed and prescribed in the current paper.

On the other hand, for improving air circulation affectivity and blowing cold air into the "heat recess areas", new additional duct + nozzle system and two auxiliary fans were built in each compartment of the prototype container. Principles of the newly developed elements will be illustrated and described in the section titled "Conclusions".

3.2 Results of the Hydrodynamic Simulation

Velocity distributions were examined in horizontal plane. One specific test result is shown in Figure 12.

A too low flow velocity is unfavourable in terms of heat distribution, harmful to the quality of cooling. A too high flow velocity increases the cooling power demand. Since in our case the quality of cooling is important, the primary aim is to avoid too low velocities. With the actual fans the maximum velocity does not change in a wide range anyway, so its impact is less significant. During the tests velocity values lower than 0.5 m/s are examined, velocities lower than this can only occur on small, local surfaces, which can assure refrigerated transportation of proper quality.

The results are summarized in Table 2.

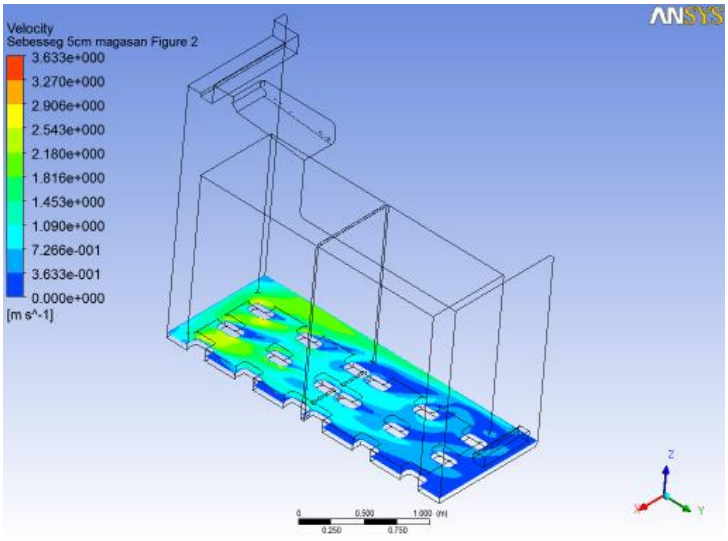


Figure 12

Full length, symmetrical arrangement, without auxiliary fan, in the case of horizontal blow-in the velocity distribution at the feet of the pallets

Table 2

Measurement results

	v max	% of areas of velocity lower than 0.5 m/s
Asymm, full, no auxiliary fan	2.025	50
Symm, full, no auxiliary fan	5.576	15
Symm, -400, no auxiliary fan	6.6	5
Symm, -800, no auxiliary fan	3.651	15
Symm, -1200, no auxiliary fan	5.38	3
Asymm, full, upper auxiliary fan	2.12	70
Symm, full, upper auxiliary fan	7.393	10
Symm, -400, upper auxiliary fan	6.817	10
Symm, -800, upper auxiliary fan	4.92	10
Symm, -1200, upper auxiliary fan	6.435	5
Asymm, full, lower auxiliary fan	2.247	20
Symm, full, lower auxiliary fan	5.411	5
Symm, -400, lower auxiliary fan	6.792	5
Symm, -800, lower auxiliary fan	4.879	5
Symm, -1200, lower auxiliary fan	6.722	3
Symm, full, with horizontal blow-in	3.633	40
Symm, full, with horizontal blow-in, lower auxiliary fan	4.097	10

Based on the measurements it can be concluded that the most unfavourable layout is the longest cooled space with symmetrical cargo arrangement. The results of this case are shown in Figure 13.

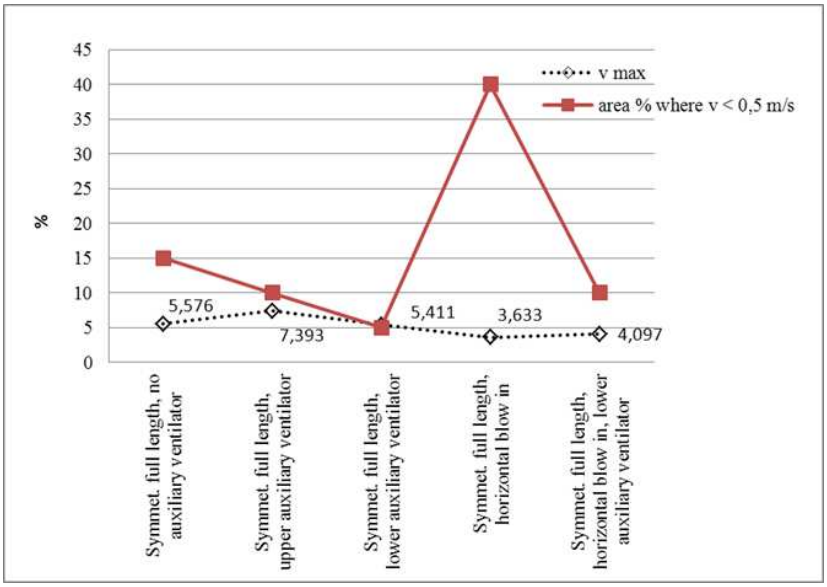


Figure 13

Longest cooled space, the maximum air velocities with symmetrical cargo arrangement, and the ratio of areas of velocity lower than 0.5 m/s

The ratio of areas of velocity lower than 0.5 m/s was examined. At lower air velocities than this local thermal bridges may be generated, which adversely affect the quality of cooling, so this value must be reduced to the lowest possible state. In reality the situation is somewhat more favourable than the calculation because simulation assumes a steady state, which does not exist due to the well visible turbulences, so air movement is more intensive in areas where low velocity and high velocity areas are next to each other. It can be concluded that horizontal blow-in is unsuitable to generate proper air movement. 40% of the area air velocity is too low, which is detrimental in itself, as opposed to the next case there is another problem here that the area is coherent, i.e. there is an almost stationary zone in the middle of the container, so there is no cooling. This can be modified with an auxiliary fan to an acceptable level, as the last pair of results show.

As a comparison Figure 14 shows the results with the shortest cooling compartment setting. It can be concluded that in the case of this small space already the air movement generated by the properly directed condensing unit is sufficient to achieve the proper cooling quality. Spaces shorter than this were not tested, it is apparent the flow conditions are better and better in smaller spaces, as it had been expected.

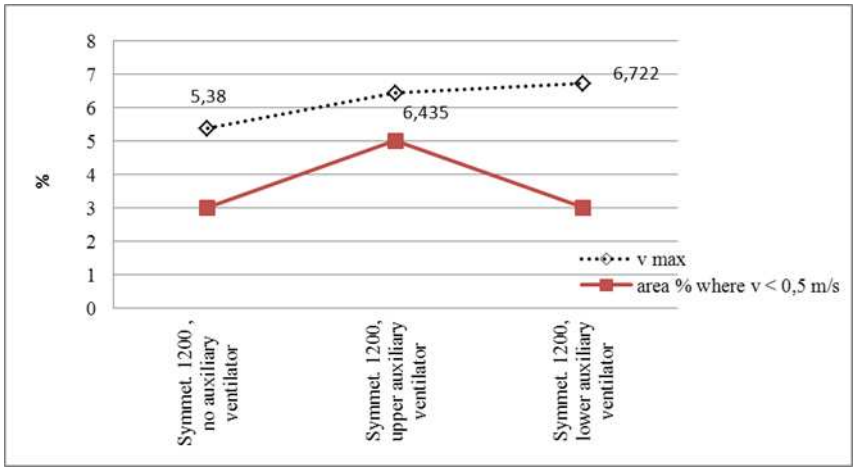


Figure 14

Shortest cooled space,the maximum air velocities with symmetrical cargo arrangement, and the ratio of areas of velocity lower than 0.5 m/s

Conclusions

With the original design of the condensing unit the container is not suitable for proper cooling with all the cooling space arrangements. Due to the horizontal blow-in the suitable air circulation is not generated. This is supported by the experimental model experiments in addition to the simulations.

The application of the multijet technology with several auxiliary fans and directed evaporator blow-out is required to assure optimum quality of cooling.

The auxiliary fans at the top position improve the flow with little efficiency and increase the maximum speed, so their application does not provide any obvious advantages.

The optimum position of the auxiliary fans is the bottom of the partition wall. The flow assistance effect is the most efficient; the uniformity of the flow is, also the best in this way.

The increase in the throw of the air could be supported by the previously proposed auxiliary fans and the use and installation of the air distribution unit containing the jets. Figure 15 shows a proposal for the construction of this.

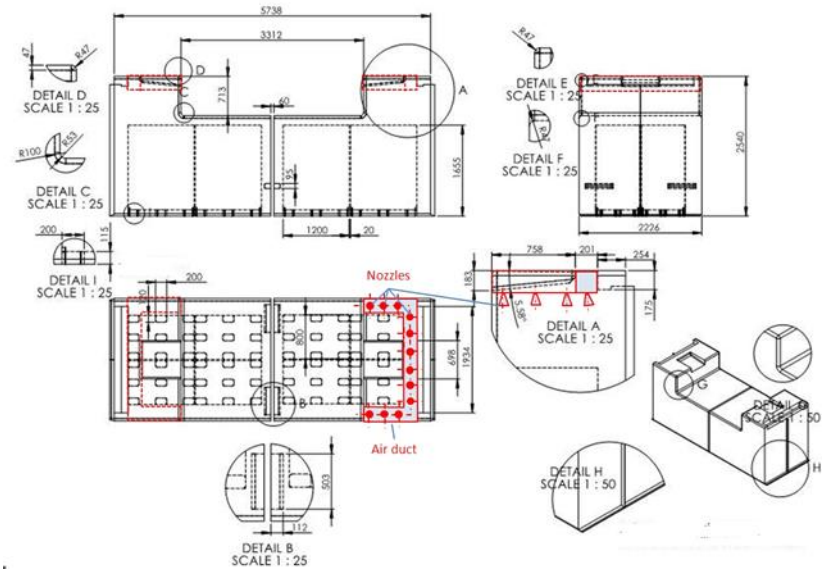


Figure 15

Sketch of the installation of the air distribution unit featuring the blow-in jet nozzles (throw jet nozzles)

Acknowledgement

This work was supported by the **GOP-1.1.1.-11-2012-0322** identification number project titled: **"Research and development of new type intelligent mobile refrigerated compartments with multi-dimension and injection airflow technique and GPS registration."**

Among the authors, László Zsidai would like to thank MTA (Hungarian Academy of Sciences) for supporting this work in the frame of the research fellowship BOLYAI (BO/00127/13/6).

References

[1] Prezenszki J. (et.al.): Logistics II. Logistics Development Centre, 1999

[2] Scharnow, U./Scharnow, R.: Container, Gutart und Qualitätserhaltungwährend des Seetransports [Container, type of cargo and retention of quality during maritime transport] (Part 1) DDR-Verkehr, Berlin 19 (1986) 1, pp. 20-23; Continuation and conclusion 19 (1986) 2, pp. 50-55

[3] Grote-Antonson: Handbook of Mechanical Engineering: ISBN 978 3 540 49131 6; Springer, New York, 2008

[4] Beke Gy.: Freezing Industrial Handbook Vol. I, Mezőgazda Kiadó, Budapest 2001

[5] Várszegi T.: Food Industrial Technologies and Machines. Szent István University, Gödöllő, 2009

[6] Edit Schmidt: Elementary Investigation of Transportation Problems. Acta Polytechnica Hungarica Vol. 6, No. 2, 2009, ISSN 1785-8860

[7] Kasmire, R. F. and Hinsch, R. T. (1987) Maintaining Optimum Transit Temperatures in Refrigerated Truck Shipments of Perishables. University of California Perishables Handling Transportation Supplement No. 2

[8] Y. Wild: Containerhandbuch. GDV 2014.1500 p.

[9] A. Csikós et al.: Real-Time Estimation of Emissions Emerging from Motorways Based on Macroscopic Traffic Data. Acta Polytechnica Hungarica Vol. 8, No. 6, 2011, ISSN 1785-8860

[10] Son H. Ho, Luis Rosaria, Muhammad M. Rahman: Numerical Simulation of Temperature and Velocity in a Refrigerated Warehouse. International Journal of Refrigeration Vol. 32, pp. 1015-1025 (2010) ISSN 0140-7007

[11] Yuta Umeno et al: The Use of CFD to Simulate Temperature Distribution in Refrigerated Containers. Engineering in Agriculture, Environment and Food, Vol. 8, Issue 4, October 2015, pp. 257-263, ISSN: 1881-8366

[12] Thijs Defraeye et al: Integral Performance Evaluation of the Fresh-Produce Cold Chain: A Case Study for Ambient Loading of Citrus in Refrigerated Containers. Postharvest Biology and Technology Vol. 112, 2016, pp. 1-13, ISSN: 0925-5214

[13] Holman, J. P. (1992) Heat Transfer. McGraw-Hill Book Company. Singapore

[14] Al-Shemmeri, T. T. (2012) Engineering Fluid Mechanics. www.bookboon.com. E-book

[15] Fejes, G. és Fábry, Gy. (1975) Vegyipari Gépek és Műveletek II Tankönyvkiadó, Budapest

[16] Fejes, G. és Tarján, G. (1973) Vegyipari Gépek és Műveletek. Tankönyvkiadó, Budapest

[17] Tijskens, L. M. M. – Hertog, M. L. A. T. M. – Nicolaï (2001) Food Process Modelling. Woodhead Publishing Ltd. Cambridge, England

[18] Liu, C. (2012) Advances in Modeling of Fluid Dynamics. www.e-booksdirectory.com. E-book

[19] Nakasone, Y. – Stolarski, T. A. – Yoshimoto, S. (2006) Engineering Analysis with ANSYS Software. ELSEVIER. E-book

[20] Kiss, R. (1980) Légtechnikai Adatok. Műszaki Könyvkiadó, Nudapest

Control of LED Lighting Equipment with Robustness Elements

Alexander Baklanov, Svetlana Grigoryeva

D. Serikbayev East Kazakhstan State Technical University
Faculty of Information Technology and Power Engineering
A. K. Protazanov Str. 69, 070004, Ust-Kamenogorsk, Kazakhstan
e-mail: {ABaklanov, SGrigorieva}@ektu.kz

György Györök

Óbuda University, Alba Regia Technical Faculty
Budai út 45, H-8000 Székesfehérvár, Hungary
gyorok.gyorgy@amk.uni-obuda.hu

Abstract: This paper presents the results of the study into the control of light-emitting-diode lighting equipment. It also deals with the issues of robustness in the organization of controlling key parameters of a lighting system. An experimental facility for research of using high-power LEDs and controlling them with compensation of uncertainties was developed. The results of that development allows for showing the effectiveness of the use of additional control loops with elements of robust control. Additionaly, there are suggestions to improve systems both of indoor and outdoor led lighting devices of increased power.

Keywords: robustness; high power LED; control system; controller; Peltier element; SCADA system

1 Introduction

Modern control systems are characterized by a high degree of complexity of a control object resulting from the incomplete priori information about, object parameters, properties of external acting disturbances and the presence of non-stationary and nonlinear dynamic characteristics of the object.

Classical methods of synthesis of control systems assume that the object and controller models are known, and have constant parameters. However, the model of a real physical system will always be inaccurate for the following reasons:

change of parameters due to certain circumstances; dynamic properties not included in the model; time lag not accounted for; change of the operating point location; sensor background noise and unpredictable external disturbances.

The development of electronic engineering requires development of reliable electronic systems that are implemented by way of expanding adaptivity to external and internal factors. The development of systems with a high degree of adaptation can be implemented with the use of robust control, i.e. control with a certain margin of stability.

The problem of adaptive and robust control under restrictions of external and internal uncertainties is a major issue in modern control theory. The most common methods of robust control are based on the use of regulators that ensure a certain stability margin of the closed-loop system [1, 2], various types of observers [3, 4], etc. One of the effective ways of controlling uncertain objects is through compensation of disturbances. The problem of compensating for uncontrolled disturbances is discussed in a series of papers [5, 6, 7, 8], in which various schemes of control systems architecture are considered.

2 Application of the Theory of robustness in Electronic Control Systems

In a general case a synthesized (according to the criteria of the sustainability) control system may have either a low sensitivity to changes of parameters of a control object or, conversely, a greater one. In the first case we speak of the system structural or its robustness, while in the second case, the system is almost inoperative as the slightest parameter drift leads to the loss of stability. The implementation of such systems is based on designing an electronic circuit which allows for maintaining the efficiency of the whole control system provided there are uncertainties in its description [9, 10].

The general diagram of the electronic system of controlling with uncertainties associated with the internal and external factors is shown in Figure 1. The internal factors include errors connected with the operation of the circuit, changing parameters of circuit components, engineering errors, errors in the firmware of the scheme, the increase in internal noise, etc. External factors include environmental influences and human errors related to improper operation of the electronic system, high ambient temperature and exposure to different radio waves and other types of electromagnetic radiation. The interaction of control systems with parameter change uncertainties is carried out by means of tracking system performance errors and controlling the correct operation of each element of the system.

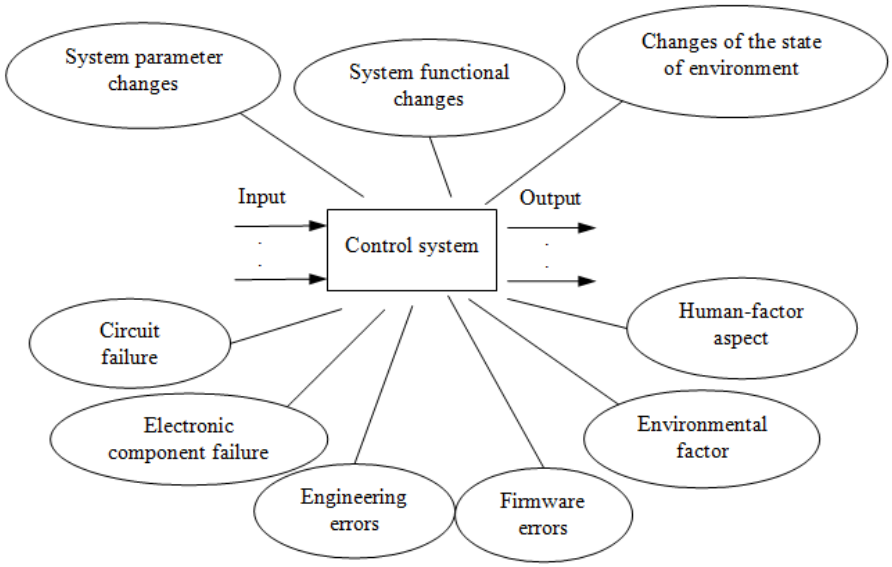


Figure 1
An overall view of the electronic control system with account for external and internal uncertainties

The system shown in Figure 1, may be presented by a general scheme describing robust control systems in Figure 2 [11], where P is the nominal control system; C is the synthesized controller; Δ is the set of all possible uncertainties in the description of the system; w is the vector of a signal including noises, disturbances and reference signals (external input of the system); z is the vector of the signal including controlled signals and tracking errors (the system output); u is the control signal, and y is the monitor signal; the signals η and v are uncertainty input and output respectively.

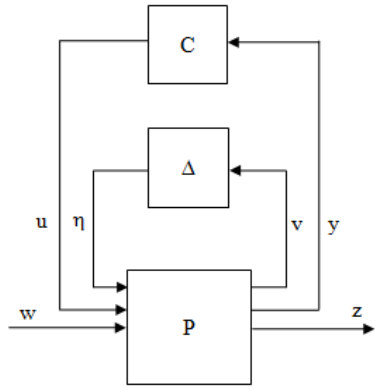


Figure 2
General scheme of a robust control system

The papers [12, 13, 14] investigate the robust control system using different architectures of control circuits. Let us consider a system with a parallel architecture of robust control.

Figure 3 presents the diagram of a system with robust control based on the parallel architecture of monitoring and control. System input w(t) includes the reference signal and all the sum signals of noise and interference of the whole system. The output signal z(t) includes the monitored signal and the tracking error signal. The uncertainty system Δ_n is supplied by the contours of processing and controlling input signals $v_n(t)$ and output $\eta_n(t)$ control signals. This system allows implementing robust control on account of compensating for uncertainties and thereby increases the stability of the system in the optimal mode. In other words, it extends the control range of the overall system. This scheme has a parallel architecture and allows for controlling the compensation of uncertainties independently.

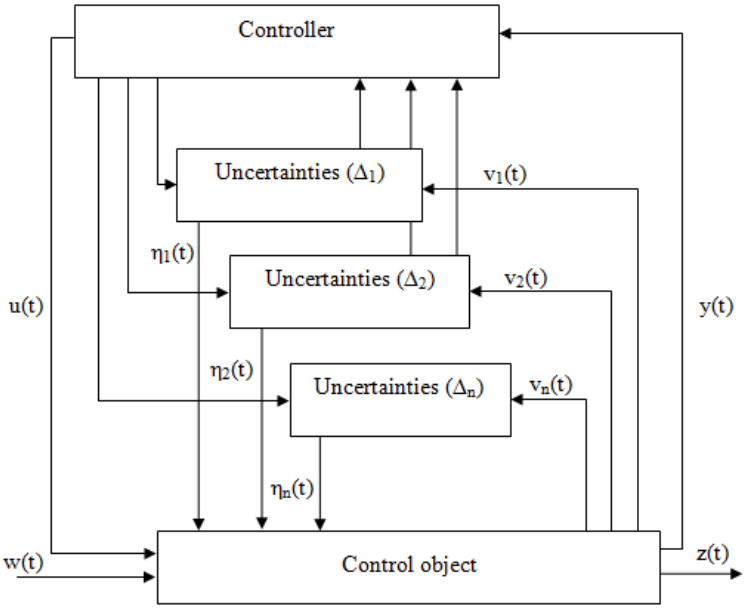


Figure 3
The control system based on the parallel architecture

In the implementation of schemes of LED lighting automatic control there are both external and internal perturbing factors (uncertainties) affecting the quality of lighting and the operation time of a lighting device. For example, the current oscillations lead to scintillation and degradation of the crystal associated with the current instability within the crystal [15]. The increase of the temperature of the crystal also leads to the loss of optical power of the LED and at the end, to its complete degradation [16].

In real LED lighting systems these factors do not allow for ensuring, specified by the manufacturer, resources of 50 thousand hours of continuous operation with retention of the original optical characteristics.

In real conditions of premise lighting there is an additional light bias associated with natural light. At that this illumination may change over time within a significant range, and also have unpredictable, fast changing light conditions changes. Taking into account that the external natural light can improve the stability of the automated system and provide the required light level.

3 Description of the LED Lighting System Regulation Scheme

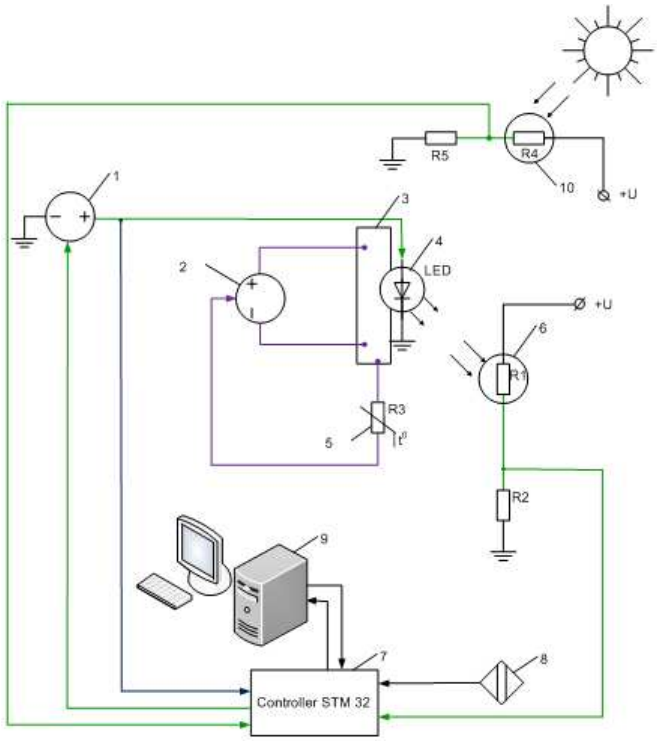
In order to implement an automation system which takes into account the above listed uncertainties there was developed the scheme [17] shown in Figure 4. Due to the first control loop this scheme maintains the desired temperature of the LED crystal. The second control loop provides the necessary light intensity due to changes of the driver operation modes.

Let us consider the operational procedure for this scheme. The first control loop maintains the desired temperature in the range which provides stable performance of the LED 4, and ensures its non-degradable operation within the asserted according to the manufacturing company period of time. The adjustment is carried out automatically by modifying the operating voltage of Peltier element 3 with the help of the designed power source 2. A thermally sensitive resistor 5 serves as the temperature sensor.

The second regulation loop provides a constant level of light intensity by controlling the LED 4 power by using the PWM regulation of driver 1. The operation of the driver is controlled by a STM32 microcontroller. A monitor signal for control is transmitted from the photoresistor 6 to the controller 7 which executes the program with the necessary LED operational scenario. The control of sunlight is carried out by using the photoresistor10. The signal from the photoresistor enters the analog input of the microcontroller and allows for maintaining the total light intensity of the premises constantly. This loop provides compensation of the natural light, as well as taking into account the combined illumination of the working area.

The disturbances associated with LED current changes can be controlled using the analog input of the controller and are displayed on the personal computer monitor screen. The signal of the current related uncertainty compensation is obtained by inverting the input signal and is added to the LED driver 1 control signal.

To ensure the operation of the lighting system only in the presence of people the motion sensor 8 is used. The signal from the sensor is communicated to the controller and provides enabling or disabling the system depending on the presence of people. The recording of the intelligent system algorithm of the microcontroller operation is carried out by a personal computer 9 with the necessary software.



1 – LED control driver; 2– the power source of the Peltier element temperature control; 3 – Peltier element; 4 – LED; 5 – thermally sensitive resistor; 6, 10 – photoresistor; 7 – STM32 microcontroller; 8 – motion sensor; 9 – PC with the intelligent control software

Figure4
The LED lighting system regulation scheme

The study of this system operation (Figure 4) may be performed using a structural diagram shown in Figure 5. This diagram shows the three control loops for optimum operation of a lighting device (an object of control P(s)), while the main control loop provides regulation of the light intensity level due to the controller $C_1(s)$ (LED driver) control. The instabilities of temperature occurring in the system are monitored by the temperature sensor $H_T(s)$ and controlled by the controller $C_2(s)$ based on the use of the Peltier plate. The current fluctuations present in the system are compensated using the controller $C_1(s)$.

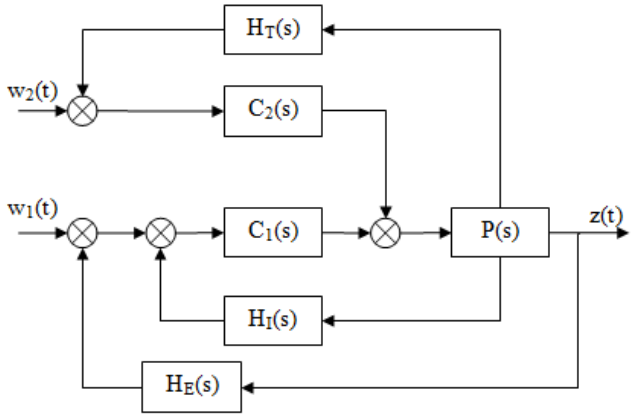


Figure 5
Block diagram of the automated control system for the LED lighting system

To ensure that robust control is implemented using the principles of parallel architecture (Figure 3) the discussed above block diagram (Figure 5) can be transformed into a system of control with elements of robustness (Figure 6). The changes in temperature, current and light intensity can be represented in the form of uncertainties, i.e. not predicted changes.

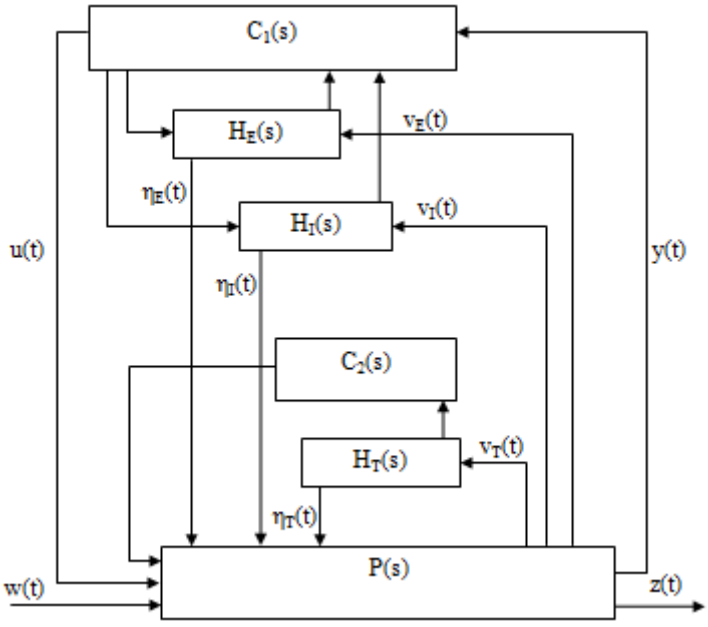


Figure 6
The robust control of a lighting system using LEDs

The following block diagram shows the system with the compensation for uncertainties $v_T(t)$, $v_I(t)$, $v_E(t)$, the uncertainties of the temperature, current and light intensity respectively. With a robust control the compensation for uncertainties irrespective of the main control loop is controlled by controller $C_1(s)$.

The output signals of the uncertainties $\eta_I(t)$, $\eta_E(t)$, $\eta_T(t)$ are necessary for controlling compensation. The information about the compensation quality is obtained by comparing the input signal $v_i(t)$ to the output signal $\eta_i(t)$.

The compensation for the temperature uncertainty controlled by sensor $H_T(s)$, is performed on account of maintaining a constant temperature of the LED crystal with the help of the controller $C_2(s)$.

The disturbances associated with current and light intensity are eliminated by using controller $C_1(s)$ where the disturbance signals coming from sensors $H_I(s)$ and $H_E(s)$ are summed. Thanks to this, the LED lighting control stabilization system becomes more stable. This allows maintaining an operating life of the LED crystal and, as a consequence, increasing the operation time of both the LED and the system as a whole.

It is shown in Figure 7 the model made in program Simulink of bypass of robust control LED element, compensating thermal uncertainty.

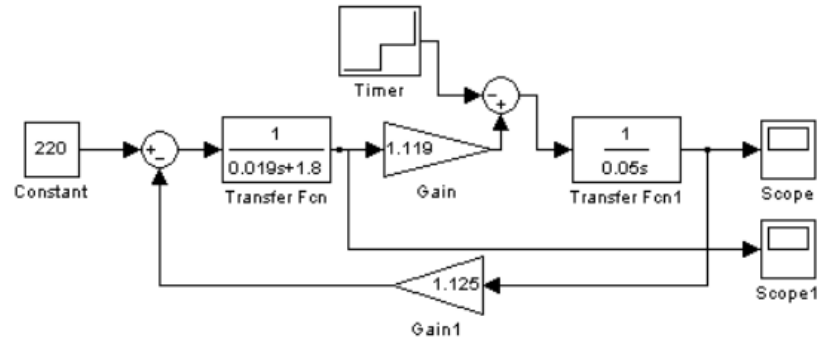


Figure 7
The model of robust control LED element

The voltage control loop is selected on the Peltier element to calculate the local automatic control system. Where there is temperature sensing of the LED. The temperature should be maintained at a level of 40^0C. The resulting transition process is shown in Figure 8.

In the absence of voltage the Peltier element LED is heated to 85^0C for 40 seconds. This process exceeds the optimum temperature of the LED operation. After applying voltage to the Peltier element the LED returns to normal operating mode with the corresponding nominal temperature of 40^0C for 60 seconds, at the same time the voltage at the Peltier element corresponds to 4.5 V (Figure 12).

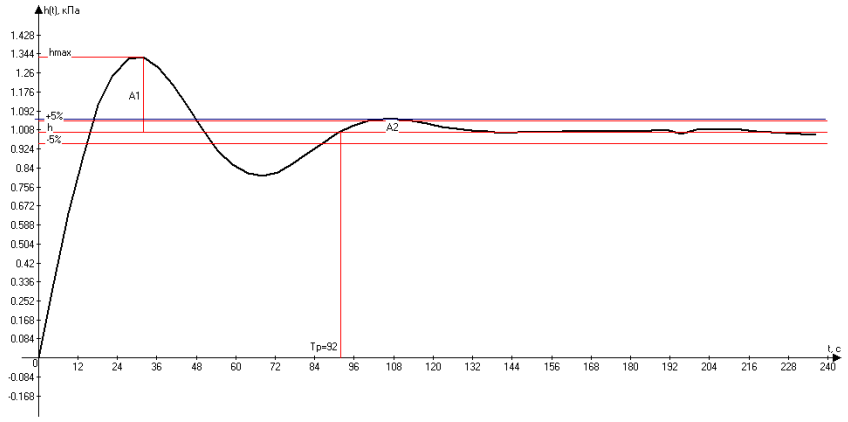


Figure 8
Graph of the transition process in the system of automatic regulation of the LED temperature

4 Description of the Experimental Unit

For the study of automated control system for a LED lighting device we developed an experimental unit with elements of robustness (Figure 9).

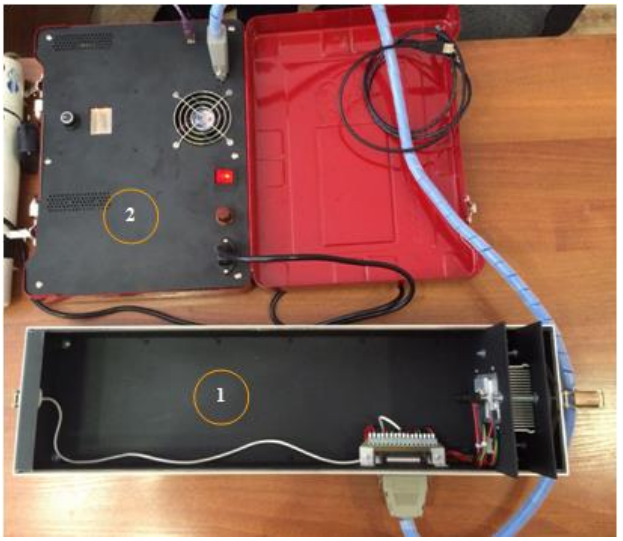


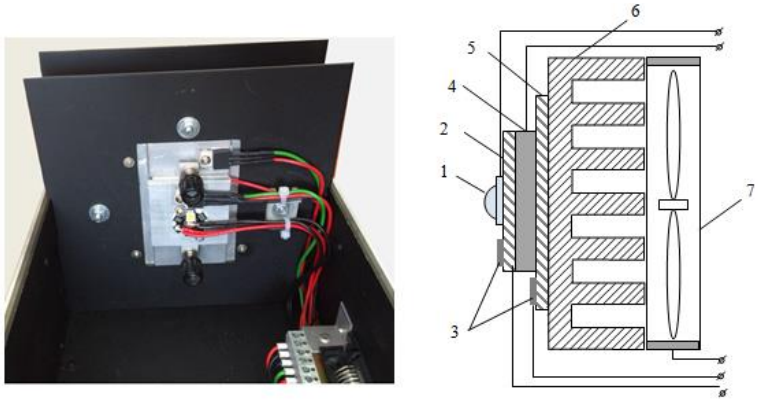
Figure 9
General view of the experimental unit for the LED automated control

It consists of two blocks: Block 1 – a lighting module, which is a cased LED with a light intensity sensor. Position 1 in Figure 9 shows the lower part of the housing; Block 2 – the LED controller and power elements and the Peltier plates (Figure 9, Position 2).

Communication between the blocks is carried out by means of the RS32 bus, Block 2 also has a connection to the PC through the USB port. There is SCADA system TRACE MODE 6 installed on the host PC; it is supplied with a specially developed application for monitoring the controller, receiving data from it and adding them to the archive using the ACCESS database management system.

The lighting module consists of a light-tight box where the following are installed: the Nichia NCSL219B LED [18], the system of supporting the LED crystal necessary temperature, the light intensity control system, and a source of additional lighting.

The LED is mounted on an aluminum plate along with the temperature sensor for the crystal temperature control. The plate is mounted to one side of the Peltier element through an organic-silicon heat-conducting paste with the aid of spring clips. To the other side of the element, with the help of organic-silicon heat-conducting paste, there is mounted a massive heat sink with a fan for removing or applying heat, depending on the function of heating or cooling, respectively. The temperature sensor placed on the radiator is necessary to prevent overheating of the Peltier element. This design is shown in Figure 10.



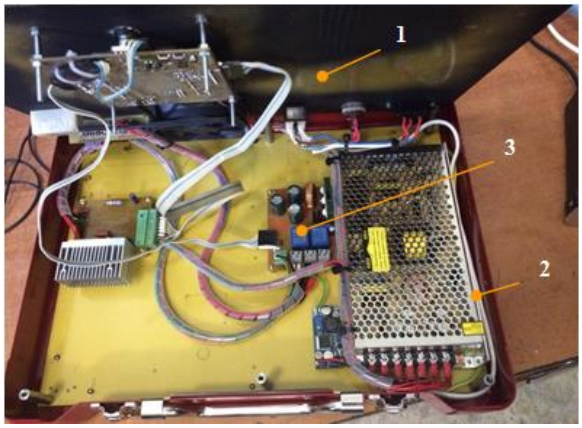
1 – LED; 2, 5 – aluminum plate; 3 – temperature sensors; 4 – Peltierplate; 6 – copper radiator; 7 – fan

Figure 10
The LED crystal temperature demonstrating construction

Across the LED, on its axis, there is a light intensity sensor – a calibrated light meter that measures the total luminous density from an artificial (tested) and a natural lighting source. To simulate natural lighting we used an additional LED mounted in the casing and powered from the IT-300 power source.

The controller module is functionally composed of the main board on a 32-bit microprocessor ARM Cortex M4 STM32F401 [19] connected to the LED control block and the heater control block (Figure 11).

This microcontroller performs measurement processes of the feedback signal and the calculation of the manipulated variable according to the algorithm of PID-regulator operating. The feedback signal from the temperature sensor is fed to an analog-digital converter of the microcontroller. Selection of PID coefficients was carried out by successive changes in these factors and measuring the response of the system with their values. Thus, we have the system response to external factors, shown in Figure 14.



1 –controller block; 2 – heater control block; 3 – LED control block

Figure 11

The exterior appearance of the controller board and the blocks of controlling the LED

The LED control block regulates the LED current depending on the task, monitors the temperature of the crystal, measures light density and sends data to the controller main board. The heater control block controls the LED crystal temperature using the Peltier element, controls the radiator temperature, and controls the air flow fan. The heater control block is connected to the main board in the same way as the LED control block. The main controller board receives data from the LED control block and the heater control block. On the basis of these parameters it controls the LED current and temperature regulators, sends the data to the SCADA system via USB for archiving and further processing.

To control the operation of the whole system we developed the software the interface of which is shown in Figure 12.

In real-time the interface switches to the working mode which shows the ongoing processes (Figure 13). This interface shows the behavior of temperature and light intensity (temperature – Line 1, light intensity –Line 2, power – Line3).

Figure 12

The main window of the custom application

Figure 13

Dialog box of the analysis of LED characteristics through time

5 Results of the Experiment

As a result of the study of the LEDs operational modes it was found that such a control system with compensation for temperature and current uncertainties (using robustness) makes it possible to obtain a stable operation mode (in a wide range of changes in the environment parameters) of a LED at the temperature of 40^0C. This mode allows for eliminating the degradation of the LED matrix, and providing the highest light intensity values.

Figure 14 shows the transient process of setting a certain temperature with the help of the developed automation system, i.e. it actually shows the additional control loop that ensures the robustness of the system. The parameters were taken at LED I = 350 mA, LED T = 40^0C.

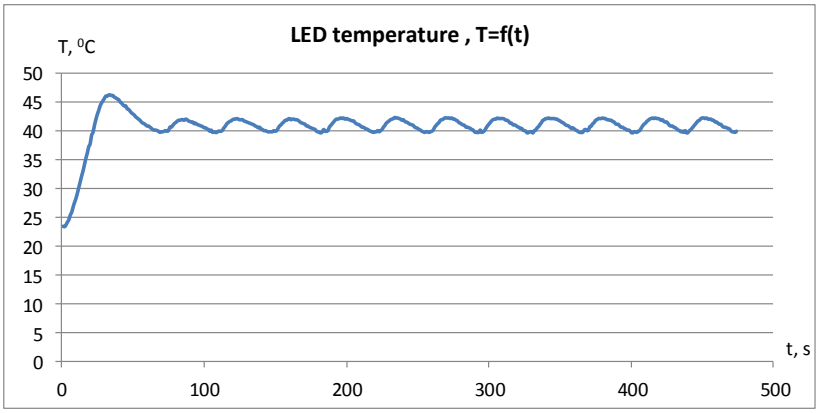


Figure 14
Transient process of temperature setting with the help of the designed system of regulation

When the temperature stabilizes (the uncertainty of temperature fluctuation is eliminated) the output power of LEDs has a more stable character. The power does not change through time, either upward or downward. The only change in the power is associated with the white noise at the output of a stabilized power supply. This dependence of the power on time is shown in Figure 15. The parameters were taken at LED I = 350 mA , LED T = 40^0C.

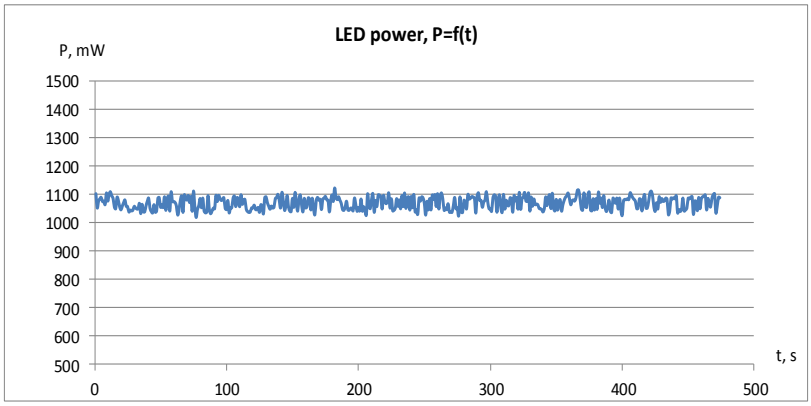


Figure 15
Dependence of the LED power on time at the stabilized temperature

Despite the fact that there is a main power control loop at significant temperature changes that occurs under natural conditions of lighting device exploitation, the control range is not enough to support the target power. As stated above in our case, using a robust approach (maintaining constant LED crystal temperature), we obtain stabilized power during the LED long-term exploitation.

Conclusion

The results of the experiment showed that the use of current approaches to control complex systems allows ensuring the stability of system operational modes. In this case, using the theory of robustness and on this basis creating additional control loops led to eliminating the factors that negatively affect the operation of the system. On this account the system, as seen in Figures 14, 15, becomes more stable and provides long-term stable operation of high power lighting LEDs. The application of proposed method for high power lighting equipment will significantly increase its service life, and also help to avoid degradation of LED crystals. The lighting equipment can be used for multiple purposes: in offices, industrial premises and for outdoor lighting.

This technique eliminates both external and internal uncertainties, adding the control loops in accordance with the robust approach (Figure 3). The further development of the methodology is related to the study of robust stability and its influence on the regulation of light intensity in aggressive conditions of high temperature and high humidity.

References

[1] K. Zhou, J. C. Doyle. Essentials of Robust Control. Upper Saddle River, NJ: Prentice Hall, 1998, p. 411

[2] Vincent J. Ginter, Jeff K. Pieper. *Robust Gain Scheduled Control of a Hydrokinetic Turbine*. IEEE Transactions on Control Systems Technology, 2011, Vol. 19, No. 4, pp. 805-817

[3] B. T. Polyak, R. Tempo. *Probabilistic Robust Design with Linear Quadratic Regulators*. Systems & Control Letters, 2001, Vol. 43, pp. 343-353

[4] I. Furtat, A. Tsykunov. *Robust Control of Time-Varying Plants with Unknown Variable Relative Degree*. Large-scale Systems Control ICS RAS, Vol. 33, 2011, pp. 91-112

[5] A. M. Tsykunov. *Robust Tracking System with Compensation of Perturbations and Noises*. Vestnik AGTU, 2014, Vol. 1, pp. 54-61

[6] A. A. Bobtsov. *An Algorithm of Robust Output-based Control of Linear Object with Compensation of an Unknown Deterministic Disturbance*. Journal of Computer and Systems Sciences International, Vol. 42 (2), 2003, pp. 251-256

[7] K. V. Voronov, V. O. Nikiforov. *A Dynamic Output Controller with Constant Perturbation Compensation*. Automation and Remote Control, 2003, Vol. 64:2, pp. 181-190

[8] P. Kucsera, Z. Béres. *Hot rolling mill Hydraulic Gap Control (HGC) thickness control improvement*. Acta Polytechnica Hungarica, 2015, Vol. 12, No. 6, pp. 93-106

[9] Gy. Györök, J. Lakner, and M. Makó, *Robust Electronic Application by PSoC*. 10th IEEE Jubilee International Symposium on Applied Machine Intelligence and Informatics-SAMI 2012, Slovakia, Herlany, pp. 405-409, 2012

[10] Gy. Györök. *A Special Case of Electronic Power Controll of Induction Heating Equipment.* Acta Polytechnica Hungarica, 2014, 11:(5) pp. 235-246

[11] K. Zhou and Z. Ren. *A New Controller Architecture for High Performance Robust, and Fault-Tolerant Control.* IEEE Transactions on Automatic Control, 2001, Vol. 46(10), pp. 1613-1618

[12] Gy. Gyorok. *Reconfigurable Control in Robust Systems by FPAA*. Intelligent Systems and Informatics, 2008. SISY 2008. 6th International Symposium on. IEEE, 2008

[13] Gy. Gyorok. *The FPAA Realization of Analog Robust Electronic Circuit*. IEEE International Conference on Computational Cybernetics: ICCC 2009, pp. 179-183

[14] Gy. Gyorok. *Embedded Hybrid Controller with Programmable Analog Circuit.* Intelligent Engineering Systems (INES), 2010 14th International Conference on. IEEE, 2010

[15] M. Meneghini, G. Meneghesso, E. Zanoni, M. D. Lago and N. Trivellin. *Development ofrobust LED lighting sources.* Solid-state lighting, 2013, Vol. 5, pp. 61-64

[16] P. Swietoniowski, T. Binkowski. *The Influence of Temperature on Optical and Electrical Parameters of Medium and High Power LEDs.* Przegląd Elektrotechniczny, 2014, Vol. 5, pp. 200-203

[17] A. Baklanov, S. Grigoryeva, Gy. Györök. *Intelligent control of LED luminaries*. 9th International Symposium on Applied Informatics and Related Areas - AIS 2014, Székesfehérvár, 2014. pp. 87-91 (ISBN:978-615-5460-21-0)

[18] Nichia. LED NCSL219B. https://www.nichia.co.jp

[19] STMicroelectronics. STM32F4 Series microcontrollers https://www.st.com

A Novel Combinational Relevance Feedback Based Method for Content-based Image Retrieval

Mohammadreza Azodinia and András Hajdu

Department of Computer Graphics and Image Processing, University of Debrecen, Egyetem tér 1, 4032 Debrecen, Hungary;
Email: m.r.azodinia@inf.unideb.hu, hajdu.andras@inf.unideb.hu

Abstract: Due to the extensive use of images in various fields, using effective approaches to retrieve the most related images given, a query image is of great importance. Content-based image retrieval is the approach commonly used to address this issue. The content-based image retrieval systems use many techniques to provide more accurate and comprehensive answers, among which, is relevance feedback. Relevance feedback is used by the system to help it retrieve more relevant images in response to a query. In this paper, we have proposed a novel relevance feedback method that is able to improve the precision of the content-based retrieval systems. The proposed method is based on multi-query relevance feedback, and similarity function refinement.

Keywords: CBIR image retrieval; CBVIR information retrieval; image database; recommender system

1 Introduction

Nowadays, digital images are commonly utilized in many areas from biology and medicine, to face or finger print recognition. Therefore, effective methods for searching and retrieval of images have received considerable interest. In traditional image retrieval systems keywords and captions representing the images were used to retrieve the images [1, 2]. These captions were generated manually, which was quite inefficient and expensive. Furthermore, the manual generation of the keywords is unable to capture every noticeable keyword that describes an image. In order to overcome the major deficiencies of the conventional approaches, content-based image retrieval (CBIR), also known as query by image content (QBIC), and content-based visual information retrieval (CBVIR), was introduced. The term "content-based" alludes that the contents of the image are used in the retrieval process, rather than the metadata such as keywords, or captions associated with the image which are traditionally used.

Most CBIR systems rely on low-level image features, i.e., color, texture and shape which are extracted from each image. These features are arranged so that they form an appropriate feature vector (FV). When the system receives a query image, the feature vector is extracted from it, and then this image can be compared to the images stored in the system. The images with the minimum distance to the query image are displayed as the result. In order to increase the overall efficiency of the system, the low-level features are usually stored in a database.

The main advantage of the CBIR systems is that the retrieval process is automatic, and it is not necessary to manually provide any metadata for images. However, one of the problems the CBIR systems have to face is that images are complex to manage. Also, image indexing is not a trivial task. Furthermore, CBIR systems require a mapping from the high-level user perceptions into low-level image features, which is known as the "semantic gap" problem. A survey on this concept is presented in [3].

In order to reduce the semantic gap, many different approaches such as image semantic classification, short-term learning (STL), and long-term learning (LTL) have been proposed, which are based on the relevance feedback technique. The relevance feedback technique was used in the CBIR for the first time by Rui in 1998 [4].

In the relevance feedback technique the user interacts with the system until the desired results are achieved. This interaction helps the system to establish a meaningful relation between the low-level features and the high-level perceptions, which reduces the semantic gap. Indeed, learning by means of relevance feedback is done in two ways, namely: short-term learning, and long-term learning. Short-term learning is done according to a query from the user in order to lead the system towards the user's desired results. In every interaction with the user, the system tries to find out how the user thinks. Short-term learning can be categorized into four different approaches:

- Classification-based methods;
- Query refinement methods;
- Similarity refinement methods;
- Multi-query methods.

2 Short-Term Learning Methods

Considering short-term learning and long-term learning we have put our focus on the former in this paper; therefore, in this section, the four sub-categories of short-term learning are briefly explained along with some researches conducted based on each method.

2.1 Classification Methods

The methods belonging to this category train a classifier using the training set, which is provided by the user and contains some relevant and irrelevant images regarding the query images. This classifier is then used to distinguish between the relevant and irrelevant images regarding the unseen query images. One of the most widely used classifiers to classify the images as described, is the Support Vector Machine (SVM). For instance, in [5] fuzzy SVM has been used in their proposed method in order to separate the relevant and irrelevant images. Indeed, in this method that combines the short-term and long-term learnings, the short term learning technique splits an image into several regions, then the fuzzy support vector machine learning is applied to these regions. Subsequently, the long-term learning approach tries to adaptively learn the semantic concepts represented in each image by means of relevance feedback and semantic clustering.

Djordjevic et al. [6] proposed a method whereby the images are partitioned into small blocks and the low-level features are extracted from each block. Then, the blocks are clustered, and some representatives are achieved for each image. After receiving the relevance feedbacks from the users, the relevant and irrelevant images are used by a SVM classifier.

2.2 Similarity Refinement Methods

The similarity refinement process adjusts the similarity function during the feedback iterations in order to refine the query response. Each CBIR system uses a similarity function to retrieve the most similar images to the query image. This similarity function is typically based on various low-level features, each of which may have different weights. The weight of these features are adjusted in the similarity refinement process.

The similarity of two images Q and T can be computed using the following formula:

$$D(Q,T \mid \mathrm{W}) = \sum_{i=1}^{n} w_i d(\mathrm{q}_i, \mathrm{t}_i) \tag{1}$$

where $W = [w_1, w_2, ..., w_n]$ is the weight vector which determines the weight of each low-level feature used in the similarity measure; q_i and t_i are the values of the ith feature from the feature vectors representing the images Q and T respectively, n is the length of the feature vectors, and $d(.)$ is a distance function.

An approach to refine the similarity function is to use the variance of features of the relevant and irrelevant images. In [4] after each feedback iteration, where the relevant and irrelevant images are specified by the user, the weight of each feature

of the feature vector is adjusted by the inverse of the standard deviation of that feature over the relevant features.

Cheng et al. [7] proposed a method wherein the users provide a ranking for the images in the feedback rounds. Subsequently, the system computes the ranking of exactly the same images based on the current weight vector for the features. Based on the similarity between these two kinds of rankings, the weights of the similarity function are adjusted.

In [8] an optimization function is defined and the weights of the features have been adjusted using the Laplace method. In this method some pre-stored images are used for the optimization task.

Another approach to similarity function refinement is to define a cost function and use the Gradient-descent method to adjust the weights of the features [9]. The main aim of the methods that take this approach is to find a better cost-function. The idea behind this approach is borrowed from the supervised learning [10].

2.3 Multi-Query Methods

In these methods the retrieval is performed using several queries. There are several ways to create these query vectors based on a given query. One of the most widely used approaches is clustering the images that are recognized to be relevant according to the user's feedbacks and to use the centers of these clusters as new queries. Salvador et al. proposed a method to estimate the proper number of clusters [11].

Kim et al. [12] have used the Q-clustering method in order to create several queries from the primary query. In their method, first the query image is given to the system, then the system retrieves M images from the database which have the least distance to the query image. These images are categorized by the user as the relevant or irrelevant images. At the first iteration, the relevant images are classified and during the rest of iterations the images are categorized into these pre-defined categories. The Bayesian classifier is used for the classification task. Also, the authors proposed another method which uses hierarchical clustering instead of the Q-clustering [13].

Herraez et al. [14] used genetic algorithms to create new query vectors by applying the genetic operators, i.e. mutation and crossover, on the relevant images. The new query vectors are created by mapping each child into one of the images from the image database.

2.4 Query Refinement Methods

This method was first proposed in 1997 in a system called MARS [15], where the average of all the images that were considered to be relevant to the query was computed and the result was used as the new query.

In [15] both the relevant and irrelevant images have been used to improve the query vector. Indeed, this method, which is known as Ricchio's method, tries to move the query vector closer to the relevant images and move it away from the irrelevant ones during each step.

3 Architecture of CBIR Systems

As mentioned before, a CBIR system receives a query image and returns the most similar images to the query image among the images stored in the images database. The queries are processed based on the low-level features of the images. The query-processing module is responsible for extracting the features of the query image and comparing the feature vector to the stored images. Once the similarities between the query image and the images stored in the database are computed, the stored images are ranked based on their distance to the query image, and the most similar ones are retrieved.

In order to increase the efficiency of the system, the features vectors representing each image are extracted and stored in the database. This process, which is done by data insertion subsystem, is usually performed off-line. It is worth mentioning that like other forms of databases, the image databases are typically indexed too. The image indexing is performed based on the feature vectors extracted from each image, and utilizes structures such as M-tree [16] or Slim-tree [17] to accelerate the similarity computation and the retrieval processes. Last by not least, it should be noted that various CBIR systems may vary this architecture or the actions carried out in each phase according to their needs. For instance, the method proposed by Santhosh et al. [18] includes the conversion of images to gray scale and image segmentation through a clustering method.

4 Proposed Method

Firstly, we should mention that in the proposed system, some general low-level features are extracted from the images contained in the database of images. The database is represented by $X = \{X_1, X_2, ..., X_N\}$ where, X_i refers to the ith image. For every image X_i there is a feature vector F_i. Therefore, as shown in Figure 1, the image database contains a database of feature vectors

$F = \{F_1, F_2, ..., F_N\}$ corresponding to X . We extracted four types of low-level visual features, three of which are color features and one of them is a texture feature.

4.1 Features

Color Moments

The mean, variance and skewness of an image which are known as the color moments (respectively, known as the first, second, and third color moments) characterize the color distribution in an image.

The first moment can be calculated by using the following formula:

$$\mu_i = \sum_{j=1}^{N} \frac{1}{N} p_{ij} \tag{2}$$

where *N* is the number of pixels in the image and p_{ij} denotes the value of the j^{th} pixel of the image at the i^{th} color channel.

The second color moment can be calculated by using the following formula:

$$\sigma_i = \sqrt{(\frac{1}{N} \sum_{j=1}^{N} (p_{ij} - \mu_i)^2)} \tag{3}$$

The third color moment can be calculated by using the following formula:

$$\gamma_i = \sqrt[3]{(\frac{1}{N} \sum_{j=1}^{N} (p_{ij} - \mu_i)^3)} \tag{4}$$

Color histogram

A color histogram denotes a representation of the statistical distribution of the colors in an image, and disregards the spatial location of the colors. Each bin of the histogram corresponds to a color in the quantized color space. The number of the bins is specified by the number of colors in the image.

The color histogram can be measured for any kind of color space. The number of elements in a histogram is specified by the number of bits in each pixel of the image. Indeed, an n-bit image has 2^n values (or intensities) lying in the interval $[0, 2^n - 1]$.

Although the color histogram is relatively invariant under the translation and rotation about the viewing axis and varies just slightly with the angle of the view,

it is significantly affected by the changes in the illumination, shading, highlights, and the inter reflections. To overcome these drawbacks, Gevers et al. [19] proposed a robust histogram for object recognition that is invariant to the aforementioned changes. Domke and Aloimonos [20] proposed a method to create a color histogram which makes the histogram invariant under any mapping of the surface that is locally affine, which includes a very wide range of the viewpoint changes or deformations. The main idea is that the pixels in the two images can be weighted using the gradients in different color channels. Therefore, a deformation of a region of the image will change both the image and the weights equally. We have used the method proposed in [20] in this study.

Edge histogram

Edges are among the important features of images. In the MPEG-7 standard the edge histogram is used in order to capture the edge distribution of the image. However, its drawback is that it contains only the local edge distribution with 80 bins. In this paper, we have used the method proposed in [21], which generates two other types of the edge histograms: the global, and semi-global histogram bins.

Gabor features

The texture features of each image are captured using the Gabor filter. The Gabor filter provides optimal joint resolution in both frequency and spatial domains [22]. We have used 30 filters with five different scales and six different orientations [23]. We first create a grayscale equivalent for each image and normalize it so as to have dimensions of 256×256 to speed up the computation using the FFT method. After filtering the images, their means and standard deviations are computed which provide 60 features for each image.

5.2 Weight Adjustment

In our proposed method, the query vector is broken into several multi-query vectors in regard to every feature space to perform the short-term learning. Initially, the user feeds the system with a query image Q, then, the features of the query image are extracted. These features are illustrated in the following section. The extracted feature vector is then used by the similarity function to retrieve *P* relevant images from the stored images. The images are then partitioned into two categories by the user: relevant images, and irrelevant images. At the next stage, the multi-query approach is used. This is done by clustering the images that are categorized as relevant, and using each cluster center as a new query. Subsequently, the minimum distance of each image from the cluster centers is considered as its similarity to the query. We have used the WPGMC[1] algorithm to perform the clustering.

[1] Weighted Pair Group Method Centroid

In order to combine the results of the retrieval based on each feature *i*, the ranking of each image in the database is computed when images are retrieved based solely on that feature. This ranking is then used to adjust the weight of each feature. Suppose we have *N* images in the database, and we offer the user only the *M* first retrieved images. First, the rankings of each of these *M* images are added together. Then according to the difference between the summation achieve at stage *t* and stage *t*+1 the new weights are computed. It should be noted that initially the weight of all the features are equal. In order to shed more light on this process, we bring an example. Suppose the database contains 8 images, i.e., $\{A, B, C, D, E, F, G, H\}$, and the system shows the user the first 4 images, i.e., $\{I_1, I_2, I_3, I_4\}$. Furthermore, suppose we use three features f_1, f_2, f_3. Also, the first four features retrieved at stage *t* are A,B,C, and D, and the first four features retrieved at stage *t*+1 are E,F,G, and H. The orders of the images retrieved using each feature separately at stages *t* and *t*+1 are represented in Table1 and Table2 respectively.

Table 1
The orders of the images retrieved using each feature separately at stage *t*

Retrieval based on each feature		
f_1	f_2	f_3
C	A	F
E	F	D
F	B	E
H	G	G
B	C	C
D	D	B
A	E	H
G	H	A

Table 2
The orders of the images retrieved using each feature separately at stage *t+1*

Retrieval based on each feature		
f_1	f_2	f_3
A	C	G
B	D	H
H	A	E
C	F	B
G	B	A
F	G	F
E	E	C
D	H	D

Considering Table 1, Table 2, and the aforementioned assumptions, we can deduce the results shown in Table 3.

It is worth mentioning that we call the sum of feature-wise ranking of the selected images the *feature-wise sum of the rankings.* The *normalized feature-wise sum of the rankings* for each feature is computed by dividing the feature-wise sum of the rankings for that feature by the summation of the feature-wise sum of the rankings.

Table 3
The rankings of the first 4 images according to each feature

Stage		Image	The ranking based on each feature		
			f_1	f_2	f_3
t	I_1	A	7	1	8
	I_2	B	5	3	6
	I_3	C	1	5	5
	I_4	D	6	6	2
	The feature-wise sum of the rankings		19	15	21
	The normalized feature-wise sum of the rankings		19/55	15/55	21/55
$t+1$	I_1	E	7	7	3
	I_2	F	6	4	6
	I_3	G	5	6	1
	I_4	H	3	8	2
	The feature-wise sum of the rankings		21	25	12
	The normalized feature-wise sum of the rankings		21/58	25/58	12/58

As implied from Table 3, the normalized total ranking of the first 4 images considering only feature f_1 has decayed from 0.34 to 0.36 so its corresponding weight should decrease accordingly. On the other hand, the normalized total ranking of the first 4 images considering only feature f_3 has improved from 0.38 to 0.2, therefore, its corresponding weight should increase accordingly. It is clear from this example that the values of the feature-based normalized rankings always sum up to 1. This is exactly the case for the weights of each feature in the similarity function. Therefore, considering the sign of the changes in these values, the total changes will always be equal to 0. Consequently, the difference between the normalized feature-wise sum of rankings of that feature at stages *t* and *t*+1 is added the weight of each feature. That is, the weight of the feature f_i at stage *t*+1 is computed using the following formula:

$$w_i^{t+1} = w_i^t + \left(\sum_{j=1}^{M} r_{ij}^t - \sum_{j=1}^{M} r_{ij}^{t+1}\right) \tag{5}$$

where r_{ij}^{t} is the ranking of the j^{th} relevant image when it is retrieved using only the i^{th} feature, and M is the number of the images that are given to the user. It should be reminded that the relevance of the images is determined by the user.

5.3 Similarity Distance

Thanks to its simplicity, the Euclidean distance is the most commonly used distance metric among all the image distance metrics.

Given two M by N images, $I = (i^1, i^2, ..., i^{MN})$ and $J = (j^1, j^2, ..., j^{MN})$, the Euclidean distance is computed by the following equation:

$$d_E{}^2(I, J) = \sum_{k=1}^{MN} (i^k - j^k)^2 \tag{6}$$

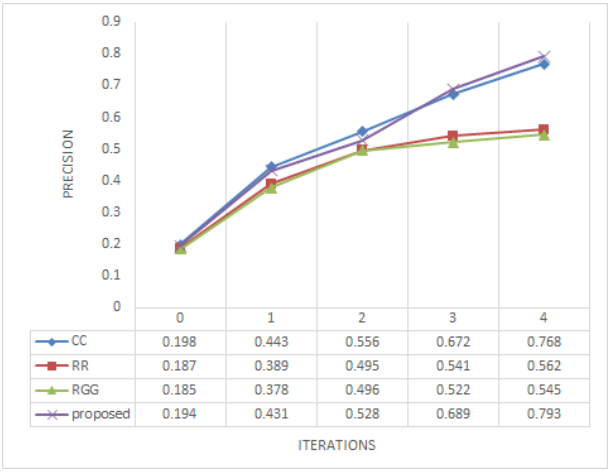
6 Experiments and Results

In order to evaluate the proposed method, Matlab running on a PC with an Intel CORE 2 Due 2.20 GHz processor and 4 GBs of RAM was implemented. A collection composed of 20000 JPEG formatted images borrowed from the Hamshahri2 dataset was used in our experiments [24]. Furthermore, 50 query images chosen from 50 different semantic groups were examined in the experiments. Each of these 50 images was used in a scenario wherein the image was given to the user as the query image, then the system offered the user a predefined number of images as the relevant images. Thereafter, the user was asked to provide a feedback and to decide which images were relevant and which ones were not. The relevant images were used for updating the similarity function according to the proposed method. The updated similarity function was used to retrieve images again. This loop was repeated 4 times for each image. We have compared our proposed method with the following methods:

- CC: The Classifier Combination, which is based on the multi-query approach.
- RR: This method uses the Rocchio's method [16] for the query vector refinement and uses the Rui's method [4] for the distance function refinement;
- RGG: This method uses the Rocchio's method [16] for the query vector refinement and uses the method proposed by Guldogan & Gabouj [25] for the distance function refinement.

The precision of the proposed method is compared to the precisions of the other methods and the results are depicted in Figure 2. It is worth mentioning that the results achieved at each iteration are averaged over all the categories. The equation 7 is used to compute the precision of each method.

$$Precision = \frac{The\ number\ of\ the\ relevant\ images\ in\ the\ first\ j\ positions\ of\ the\ retrieved\ images}{j} \quad (7)$$



	0	1	2	3	4
CC	0.198	0.443	0.556	0.672	0.768
RR	0.187	0.389	0.495	0.541	0.562
RGG	0.185	0.378	0.496	0.522	0.545
proposed	0.194	0.431	0.528	0.689	0.793

Figure 1

The average precision graph depicted for the top 20 retrieved images

The average time for the retrieval of images by means of each method is shown in Figure 3.

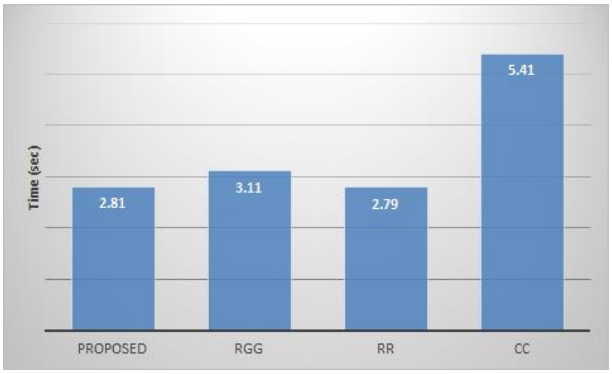


Figure 2

The average time for the retrieval of images by different methods

It can be alluded from the above results that the proposed method not only produces more accurate results, but also achieves the results in a reasonable time.

Furthermore, the precisions of each method for some randomly selected semantic groups are shown in Table 4. The best result is shown by bold face for each case.

Table 4
Comparison of the precision of the proposed method with other methods for the top 20 retrieved images

Semantic group	CC	RR	RGG	proposed
Sports	0.8185	0.5643	0.5912	**0.9261**
Buildings	0.7203	0.6485	0.7943	**0.8527**
Animals	**0.9918**	0.9561	0.8924	0.9840
Weather Forecast	0.5470	0.7511	**0.7519**	0.7455
Accidents	0.4149	0.2206	0.2418	**0.5613**
Cars	0.8651	0.8769	0.8133	**0.8774**

According to Table 4 the proposed method outperforms the other well-known methods in most cases.

Conclusion and Future Work

In this paper we developed a relevance feedback based approach to content-based image retrieval system. The proposed method combines the color, spatial, and texture features and adjusts the similarity function in order to retrieve the most relevant images from an image database in response to a query image. Our approach takes advantage of the multi-query methods as well, which results in better precision while the method is still fast. For the future studies we will try to combine our approach with some supervised learning methods. Also, we will try to use and evaluate other distance measures to compute the distance between the feature vectors.

Acknowledgement

This work was supported in part by the project VKSZ_14-1-2015-0072, SCOPIA: Development of diagnostic tools based on endoscope technology supported by the European Union, co-financed by the European Social Fund.

References

[1] V. E. Ogle and M. Stonebraker. Chabot: Retrieval from Relational Database of Images. IEEE Computer, 28 (9): 40-48, Sept. 1995

[2] H. Lieberman, E. Rosenzweig, and P. Singh. Aria: An Agent for Annotating and Retrieving Images. IEEE Computer, 34 (7): 57-62, 2001

[3] Ying Liu, Dengsheng Zhang, Guojun Lu. Regionbased Image Retrieval with High-Level Semantics using Decision Tree Learning. Pattern Recognition 41, pp. 2554-2570, 2008

[4] Rui, Y., Huang, S., Ortega, M., Mehrotra, S., Relevance Feedback: a Power Tool for Interactive Content-based Image Retrieval, IEEE Transactions on Circuits and Video Technology, Vol. 8, No. 5, pp. 25-36, 1998

[5] Barrett, S., Content-based Image Retrieval: a Short Term and Long-Term Learning Approach, 2007

[6] Djordjevic, D., Izquierdo, E., An Object- and User-driven System for Semantic-based Image Annotation and Retrieval, IEEE transactions on circuits and systems for technology, Vol. 17, No. 3, pp. 313-323, 2007

[7] Cheng, P. C., Chien, B. C., Ke, H. R., Yang, W. P., A Two-Level Relevance Feedback Mechanism for Image Retrieval, Expert Systems with Applications, Vol. 34, pp. 2193-2200, 2008

[8] Hoi, S. C. H., Liu, W., Chang, S. F., SemiSupervised Distance Metric Learning for Collaborative Image Retrieval, ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMCCAP), Vol. 6, No. 3, pp. 1-26, 2010

[9] Deselaers, T., Paredes, R., Vidal, E., Ney, H., Learning Weighted Distances for Relevance Feedback in Image Retrieval, 19th International Conference on Pattern Recognition (ICPR), 2008

[10] Paredes, R., Vidal, E., Learning Weighted Metrics to Minimize Nearest Neighbor Classification Error, PAMI, Vol. 28 (7), pp. 1100-1110, 2006

[11] Salvador, S., Chan, P., Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms, Technical Report CS-2003-18, Florida Institute of Technology, 2003

[12] Kim, D. H., Chung, C. W., Qcluster: Relevance Feedback Using Adaptive Clustering for Content-Based Image Retrieval", Proc. ACM SIGMOD, pp. 599-610, 2003

[13] Kim, D. H., Chung, C. W., Barnard, K., Relevance Feedback using Adaptive Clustering for Image Similarity Retrieval, The Journal of Systems and Software, Vol. 78, pp. 9-23, 2005

[14] Herraez, M. A., Ferri, F. J., Picot, S. M., Distancebased Relevance Feedback using a Hybrid Interactive Genetic Algorithm for Image Retrieval", Applied Soft Computing, Vol. 11, pp. 1782-1791, 2011

[15] J. Rocchio. Relevance Feedback in Information Retrieval. In The SMART Retrieval System: Experiments in Automatic Document Processing, pp. 313-323. PrenticeHall, Englewood Cliffs, NJ, USA, 1971

[16] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. InProceedings of 23rd International Conference on Very Large Data Bases, pp. 426-435, Athens, Greece, 1997

[17] C. Traina, B. Seeger, C. Faloutsos, and A. Traina. Fast Indexing and Visualization of Metric Datasets Using Slim-Trees.IEEE Transactions on Knowledge and Data Engineering, 14 (2): 244-260, March/April 2002

[18] Mathew S., Balas V., Zachariah P., and Samue P., "A Content-based Image Retrieval System Based on Polar Raster Edge Sampling Signature," Acta Polytechnica Hungarica, Vol. 11, No. 3, pp. 25-36, 2014

[19] Theo Gevers and Harro Stokman. Robust Histogram Construction from Color Invariants for Object Recognition. IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 26, No. 1, pp. 113-118, 2004

[20] J. Domke, Y. Aloimonos, "Deformation and Viewpoint Invariant Color Histograms", Proc. BMVC (British Machine Vision Conference), September 2006, Edinburgh, UK

[21] Park DK, Jeon YS, .Won CS, Efficient Use of Local Edge Histogram Descriptor. Proceedings of the 2000 ACM workshops on Multimedia: 51-54, 2000

[22] B. S. Manjunath, et al., Introduction to MPEG-7, Wiley, New York, 2002

[23] Manjunath BS, Ma WY, Texture Feature for Browsing and Retrieval of Image Data. IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (8): 837-842, 1996

[24] Abolfazl AleAhmad, Hadi Amiri, Ehsan Darrudi, Masoud Rahgozar, & Farhad Oroumchian. (2009) Hamshahri: A Standard Persian Text Collection. Knowledge-based Systems, 22 (5), 382-387

[25] Guldogan, E. & Gabbouj, M., Dynamic Feature Weights with Relevance Feedback in Content-based Image Retrieval. In 24th International Symposium on Computer and Information Science. IEEE, pp. 56-59, 2009

COP and Economic Analysis of the Heat Recovery from Waste Water using Heat Pumps

Jozsef Nyers

Obuda University Budapest, Hungary
Szent Istvan University Gödöllő-Budapest, Hungary
Subotica Tech, Marko Oreskovic 16, 24000 Subotica, Serbia
e-mail: nyers@uni-obuda.hu

Abstract: This article presents a newly designed and realized system for heat recovery from waste water. The waste water is the heat source, the temperature of which is about 44-49 ^oC. The heat recovery process is performed using a plate heat exchanger and two heat pumps. The mentioned components are serially linked. The heat sink is clean water, which is pumped from a 40 meter deep well and is heated up from 13-14 ^oC to 50-55 ^oC. Waste water is cooled down from 44-49 ^oC to 13-14 ^oC. The mass flow rate of the clean and waste water is approximately the same. This heat recovery process reaches a significant coefficient of performance (COP) with 6.03-6.5 meaning that 83-85% of the energy demand for heating clean water is provided by the recovery process and electric energy is consumed only for 17-15%. From the aspect of energy efficiency: the plate heat exchanger participates by about 49%, the low temperature heat pump by 30%, the high temperature heat pump by 21% in the heat recovery.

Keywords: heat recovery; heat pump; waste water; COP; payback period

NOMENCLATURE

$\dot{q}$: heat flux, W
Q *:* total heat, J
$\dot{m}$: mass flow rate, kg/s
C_p : specific heat, p=const., J/kgK
P = performance or power, W
T*:* temperature, K, ^oC
ΔT : temperature difference, K, ^oC
COP : coefficient of performance
PHE: plate heat exchanger
HP1: low-temperature heat pump
HP2: high-temperature heat pump
HRS: heat recovery system

Greek symbols

Δ : different,
τ : time duration,

Subscripts and Superscripts

w : water,
i : input,
o : output,
e : evaporator,
c : condenser,
hp : heat pump,
phe or ex: plate heat exchanger,

rs : recovery system,

1 : first heat pump, filled with R 407C,

2 : second heat pump, filled with R 134a,

p : pressure,

pump : centrifugal pump,

comp : compressor

te : total electric power demand,

dis : dissipation,

save : save,

max : maximum,

min : minimum.

1 Introduction

In the present, world-leading scientific journals only a few scientific-research articles can be found related to heat recovery systems applying some kind of a heat pump. Fangtian Sun et al. [1] presented a new waste heat district heating system based on ejector heat exchangers -EHE and absorption heat pumps -AHP. The goal was to decrease heating energy consumption of an existing combined heat-power system by recovering waste heat of exhausted steam from a steam turbine. A new ejector heat exchanger based on ejector refrigeration cycle (absorption heat pump) was developed and used. The new EHE+AHP system can decrease consumption of steam extracted from a steam turbine by 41.4% and increase heat capacity of the existing heating network by 66.7% without changing the mass flow rate of circulating water. The heating cost of the new EHE+AHP system is 8.62 U/GJ less than that of traditional systems. The payback period of the investment was about 2 years. Jian Sun et al. [2] the same group of authors [1] published this article. The problem was the same, but in this study they focused on the mathematical description and measurement of the process in the heat recovery unit (HRU). HRU is composed of a condenser and absorption heat pump. The heat source is exhausted steam from the steam turbine and the heat sink is a combined heat-power system, CHP. Mathematical simulation is validated through measurements. It was found that the COP of the AHP 1.74 and 25 MW heat from exhausted steam is recovered by the HRU. Marija S. Todorovic and Jeong Tai Kim [3] analyzeda cost-effective technical solution for replacing fossil fuel and electricity with solar energy for water heating for different purposes (for pools, sanitary water, washing) in a spa. Bicui Ye et al. [4] proposed a new open absorption heat pump system in order to improve the thermal efficiency of the drying process. The aim was to recover the latent heat from exhausted moist gas. A mathematical model was established in order to analyze the performance of the new system. Simulation results indicated an improved COP of the new system compared to that of the double-stage close absorption heat pump system. The COP of the new system varied from 1.52 to 1.97 and the efficiency of heat recovery varied from 15.1% to 54.8% when the temperature of heat source varied from

135 °C to 175 °C and saturated steam of 100 °C was produced. Ming Qu et al. [5] dealt with heat recovery from flue gases exiting from the gas boiler. These gases contain large amounts of water vapor and a large amount of heat at high temperatures, 150-200 °C. For that reason the efficiency varied,only from 70% to 80%. Applying an absorption heat pump, part of the latent heat from the flue gas can be recovered. A mathematical model was created in order to test the COP of the HRS. Results of the model were verified by measurements on the performed system. The results are presented in the diagrams and indicate significant improvements of the COP. Zhao Yang et al. [6] presented a novel "travelling-wave thermo-acoustic heat pump" for the production of high temperature steam from 80 to 200 ^{o}C.This system comprises three linear pressure wave generators which are coupled with three heat pumps into one single closed loop. Theoretical simulations were performed at varied waste-heat temperatures (40−70 °C) and different hot-end temperatures (120−150 °C). The computing results show that this new heat pump system has a high efficiency of about 50%–60% related to Carnot.

Jozsef Nyers et al. [7], [8], [9] in earlier works has dealt with heat pumps and using the heat pumps for heating-cooling and heat recovery.

Analyzing the literature above the following content can be summarized: in the articles [1], [2], [6] the heat is recovered from exhausted steam of a steam turbine by applying the absorption heat pumps or in [6] a travelling-wave thermo-acoustic heat pump. Authors presented cases [4], [5], when waste heat is recovered from moist gases using also absorption heat pumps. In the present article the goal is the same as mentioned before, but the solution and realization are different. Heat is recovered from waste water using a plate heat exchanger and vapor-compression heat pumps.

It can be seen that in the contemporary energy sector that heat recovery occupies a more and more significant place. By applying the heat recovery system enormous energy savings are achieved, usually over 80%. The payback period of the investment is very short. In Subotica, Serbia an opportunity appeared for designing and realizing a heat recovery system for a jeans factory.

During the production process washing of jeans is necessary. For washing, a large quantity of hot water, 5220 kg/h at atemperature from 50^{o}C to 55 ^{o}C is needed. Clean water for the process is taken from two more than 40 meter deep wells. For the water pumping centrifugal pumps are applied. The temperature of the clean well water is about 13.8^{o}C. For heating up this quantity of clean well water from 13.8 ^{o}C to 50 ^{o}C or 55 ^{o}C, approximately 220-248 kW power is needed. Applying the heat recovery system 85% of 248 kW power demand is recovered from the waste water. The achieved energy efficiency is very high, coefficient of performance reaches a maximum value of COP_{max} = 6.5 or a minimum COP_{min} = 6.03.

The system components are serially linked and therefore during the operation, the quantity of clean water entering the washing machine is approximately equal to the waste water leaving the washing machine simultaneously. Temperature of the waste water which leaves the system ranges from 44°C to 49°C. While applying the heat recovery system the heat from the waste water is mostly transferred to the clean water. The temperature and mass flow rate of the incoming clean water and the outgoing waste water is nearly the same (about 13–14°C).

System components that participate in the heat recovery system are: plate heat exchanger, heat pump for low (R 407C) and heat pump for high temperatures (R 134a). Study of system's energy efficiency based on the total coefficient of performance and the partial COP of individual components was determined, as well. The aim of partial analysis of COP was to find out the participation of individual components in the total COP. Analysis and definition of partial COP is the first step to improve the system's energy and economic efficiency.

Indeed, the dimensioning of the thermal components was conducted classically, without optimization. The task of further research is to find the optimal structure and dimensions of the system's thermal and energy components in order to improve the energy and economic efficiency of the new heat recovery system.

2 Physical Model of the Recovery System

2.1 Structure of the Recovery System

The heat recovery system is composed of the following components:

- (1) Two 40 meter deep wells as the source of clean water,
- (2) Two centrifugal pumps for forwardingand pumping the clean water from the well,
- (3) Plate heat exchanger,
- (4) Heat pump for low temperatures,
- (5) Heat pump for high temperatures,
- (6) Tank for hot clean water,
- (7) Washing machines,
- (8) Tank for hot waste water,
- (9) Two centrifugal pumps for pumping and transporting the waste water,
- (10) Sewer system for waste water draining,
- System for automatic control,
- System of measuring instruments.

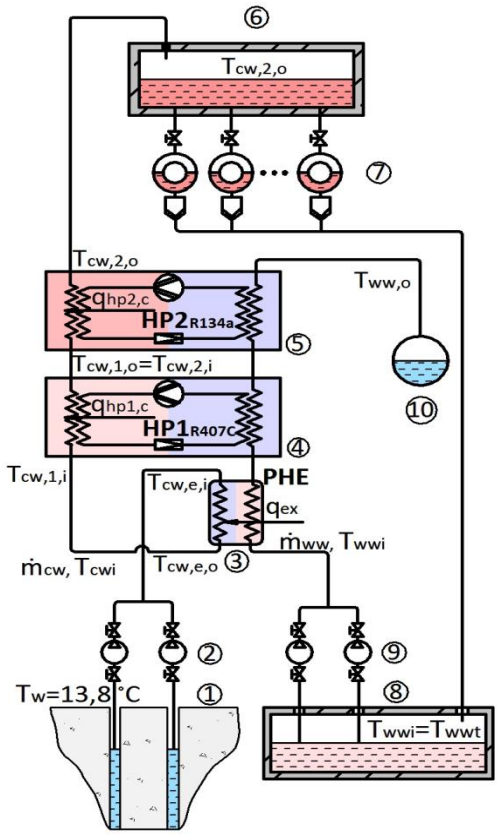


Figure 1
The functionalscheme of the heat recovery system with system's parameters

3 System Operation

3.1 General Description

The main energy components of the heat recovery system are the plate heat exchanger and the heat pump for higher and lower temperature range.

The first stage of heat recovery is realized in the counter flow plate heat exchanger. On one side of the exchanger clean cold water flows from the well, while on the other side, hot waste water flows from the washing machine. For both fluids, the flow energy is provided by centrifugal pumps. As a result, heat is recovered from the hot waste water to cold clean water.

The second stage of the heat recovery is realized in the first heat pump. The clean water flows through a shell tube condenser, while the waste water flows through the evaporator of the heat pump. The refrigeration circuit by applying mechanical work transports the heat from the waste water to the clean water.

The third stage of heat recovery occurs in the second heat pump. The clean water from the first pump's condenser flows through the condenser of the second heat pump. The waste water from the first pump's evaporator also flows through the evaporator of the second heat pump. The refrigeration circuit transports the heat between waste and clean water.

The clean heated water is gathered in the thermal insulated tank, while the cooled waste water is released into the sewer after filtration.

The clean hot water tank is elevated 5 m and therefore water by free fall can flow into the washing machine. From the washing machine the water is then released into the pipelines and by free fall it flows into the tank. The tank is placed on the ground. Pumpingthe clean water from the well is realized by the use of centrifugal pumps.

3.2 Supervision of the System

For the time being the supervision of the entire system is done visually with the exception of the heat pumps, supervision of which is automated.

Supervision is realized by measuring instruments: thermometers and manometers.

The manometers are built in the inlet pipeline of the pumps, in front of the plate surface exchanger.

The thermometers measure the water temperature in both tanks, in front of and after the plate exchanger, as well as the temperature of the incoming and outgoing water in both heat pumps.

4 Mathematical Model

4.1 Introduction

When analyzing the heat recovery system, the basic question is the energy and economic efficiency of the process. The answer can be found by determining the total and partial COP of the system. This means the quantity of total or partial recovered heat in all heat exchangers and all the invested mechanical work necessary for driving the electric motors should be found.

For the analysis adequate mathematical model is required. The mathematical description is conducted for the stationary operation mode of the heat recovery process.

4.2 Approximations and Assumptions

During the steady operation mode of the system all parameters are in a greater or lesser extent change. Some of them can be considered without greater error constant. For example: Mass flow rate of the clean and waste water can be considered permanent because the geometry of the pipe network and the power of centrifugal pumps in the operation manner do not change.

Hydraulic resistance of the pipeline due to unchanging geometry is constant.

Power of the centrifugal pumps is unchangeable.

Temperature of well wateris non-variable. Therefore, the input temperature of well water in the plate heat exchanger is constant.

The time variable system parameters are:

1) The clean hot water outlet temperature from the tank, which is variable. The reason, the clean hot water consumption from the tank is stochastic, depends on the simultaneity of filling of the four washing machines. When the consumption rises, the quantity of the cold water will rise as well; this in turn lowers the temperature of the clean hot water.
2) Due to the temperature variation of hot clean water the temperature of the waste water that enters and leaves the washing machines also varies.
3) Varies the input as well as output temperature of the waste water on the input and output of the heat exchangers, as well.

4.3 Heat Transfer in the Recovery System

The heat transfer in the recovery system is realized in three stages, in the plate exchanger, as well as, in the first and the second heat pump's heat exchangers (evaporators and condensers).

The mentioned components of the system are serially linked; therefore the total heat flux of the recovery system is sum of three partial fluxes:

$$\dot{q}_{rs} = \frac{\sum_{i=1}^{3} Q_i}{\tau} \tag{1}$$

After substitution

$$\dot{q}_{rs} = \dot{q}_{ex} + \dot{q}_{1hp,c} + \dot{q}_{2hp,c} \tag{2}$$

Where:

$Q_i\,[J]$ - Total transferred heat in heat exchangers for τ time

$\tau\ [sec\ ]$- Time duration,

$\dot{q}_{rs}\ [W]$- Total heat flux in the heat exchangers,

$\dot{q}_{ex}\ [W]$- Heat flux in the plate heat exchanger,

$\dot{q}_{1hp,c}\ [W]$-Heat flux in the condenser of the first heat pump with R 407C,

$\dot{q}_{2hp,c}\ [W]$- Heat flux in the condenser of the second heat pump with R 134a.

4.4 Heat Transferred from the Waste Water

Waste water is the heat source in the recovery process. Due to the heat transfer in the plate heat exchanger and in the evaporators of heat pumps the waste water's temperature decreases.

Recovered heat flux from the waste water in all heat exchangers is the following.

$$\dot{q}_{ww} = C_{pw} \cdot \dot{m}_{ww} \cdot \Delta T_{ww}\ [W] \tag{3}$$

The temperature difference between input and output temperatures of the waste water in all serial linked heat exchangers is:

$$\Delta T_{ww} = T_{ww,i} - T_{ww,o} \tag{4}$$

Where:

$C_{pw}\left[\frac{\mathrm{J}}{\mathrm{kgK}}\right]$ - Isobaric specific heat of the waste water,

$\dot{m}_{ww}\left[\frac{\mathrm{kg}}{\mathrm{s}}\right]$ - Mass flow rate of the waste water throughout the system,

$T_{ww,i}\ [\mathrm{K}]$- Temperature of the waste water on the input of the plate heat exchanger,

$T_{ww,o}\ [\mathrm{K}]$- Temperature of the waste water on the output of the second heat pump evaporator.

4.5 Heat Transferred to the Clean Water

During the recovery process by applying the plate heat exchanger and the heat pumps the heat-energy from the waste water is recovered which heats up the clean water from 13-14 ^{o}C to 50-55 ^{o}C.

Heat flux equals the increment of the clean water enthalpy

$$\dot{q}_{cw} = C_{pw} \cdot \dot{m}_{cw} \cdot \Delta T_{cw} \tag{5}$$

Increase ofthe clean water temperature

$$\Delta T_{cw} = T_{cw,i} - T_{cw,o} \tag{6}$$

Where:

$C_{pw}\left[\frac{\mathrm{J}}{\mathrm{kgK}}\right]$ - Isobaric specific heat of the clean water,

$\dot{m}_{cw}\left[\frac{\mathrm{kg}}{\mathrm{s}}\right]$ - Mass flow rate of the clean water,

$T_{cw,i}$ $[\mathrm{K}]$- Input temperature of the clean water,

$T_{cw,o}$ $[\mathrm{K}]$- Output temperature of the clean water.

4.6 Electric Power Demand

Electric power demand of system equals the electricity demand of the electric motors for driving the compressors and centrifugal pumps is as follows.

$$P_{te} = \sum_{1}^{2} P_{comp} + \sum_{1}^{2} P_{w\,pump} + \sum_{1}^{2} P_{c\,pump}\,. \tag{7}$$

After substitution the partial electricity demand:

$$P_{te} = P_{comp1} + P_{comp2} + P_{w\,pump1} + P_{w\,pump2} + P_{c\,pump3} + P_{c\,pump4}\,. \tag{8}$$

Where:

P_{comp} [W]- Power demand of the electric motors for driving the compressors.

$P_{w,c,pump}$ [W]- Power demand of the electric motors for driving the well and the centrifugal pumps.

4.7 Coefficient of Performance-COP

The coefficient of performance, COP is a ratio between the heat fluxes transferred tothe clean hot water and the electric power demand of all electric motors for driving the compressors and centrifugal pumps.

$$COP = \frac{q_{cw}}{P_{te}} \tag{9}$$

After substitution all the transferred heat fluxes and electric power demand of all electric motors

$$COP = \frac{C_{pv} \cdot \dot{m}_{cw} \cdot \Delta T_{cw}}{\sum_{1}^{2} P_{comp} + \sum_{1}^{2} P_{w\,pump} + \sum_{1}^{2} P_{c\,pump}} \tag{10}$$

Determination of the COP is possible through the evaporators or in other words through the waste water in the evaporators, as well. In this case, the total heat flux is calculated so that to the heat flux (q_{ww}) of the waste water in the evaporators is added to the consumed electric power of both the compressors (P_{comp}) and two centrifugal pumps ($P_{c\,pump}$) in hot water circuit (11).

$$COP = \frac{q_{ww} + \sum_{1}^{2} P_{comp} + \sum_{1}^{2} P_{c\,pump}}{P_{te}} \tag{11}$$

After substitution

$$COP = \frac{C_{pv} \cdot \dot{m}_{ww} \cdot \Delta T_{ww} + \sum_{1}^{2} P_{comp} + \sum_{1}^{2} P_{c\,pump}}{\sum_{1}^{2} P_{comp} + \sum_{1}^{2} P_{w\,pump} + \sum_{1}^{2} P_{c\,pump}} \tag{12}$$

4.8 Partial COPof the Individual Components

In this chapter the partial participation of individual components in the total COP was determined. Particularly, the COP of the heat exchanger, the first and second heat pump is determined in order to see exactly the percentage contribution and the energy efficiency of the components.

The total heat flux which clean water receives is the sum of the partial fluxes realized in the heat exchanger, and the condensers of the first and second heat pump.

$$\dot{q}_{rs} = \dot{q}_{ex} + \dot{q}_{1hp,c} + \dot{q}_{2hp,c} \tag{13}$$

Heat flux in the plate heat exchanger

$$\dot{q}_{ex} = C_{pv} \cdot \dot{m}_{cw} \cdot (T_{cw,e,o} - T_{cw,e,i}) \tag{14}$$

Heat flux in the 1. heat pump's heat exchanger i.e. in the condenser

$$\dot{q}_{1hp,c} = C_{pv} \cdot \dot{m}_{cw} \cdot (T_{cw,1,o} - T_{cw,1,i}) \tag{15}$$

Heat flux in the 2. heat pump's heat exchanger i.e. in the condenser

$$\dot{q}_{2hp,c} = C_{pv} \cdot \dot{m}_{cw} \cdot (T_{cw,2,o} - T_{cw,2,i}) \tag{16}$$

Total coefficient of performance of the system

$$COP = \frac{\dot{q}_{rs}}{P_{te}} \tag{17}$$

$$COP = \frac{\dot{q}_{ex} + \dot{q}_{1hp,c} + \dot{q}_{2hp,c}}{P_{te}} \tag{19}$$

$$COP = \frac{\dot{q}_{ex}}{P_{te}} + \frac{\dot{q}_{1hp,c}}{P_{te}} + \frac{\dot{q}_{2hp,c}}{P_{te}} \tag{20}$$

Partial coefficients of performance of the components

$$COP_{ex} = \frac{\dot{q}_{ex}}{P_{te}} [-] \tag{21}$$

of the heat exchanger

$$COP_{1hp,c} = \frac{\dot{q}_{1hp,c}}{P_{te}} [-] \tag{22}$$

of the 1 heat pump

$$COP_{2hp,c} = \frac{\dot{q}_{2hp,c}}{P_{te}} [-] \tag{23}$$

of the 2 heat pump

Percentage share of partial COPs, in the total COP.

$$COP_{ex,\%} = \frac{COP_{ex}}{COP} \cdot 100 \ [\%] \tag{24}$$

of the heat exchanger

$$COP_{1hp,c,\%} = \frac{COP_{1hp,c}}{COP} \cdot 100 \ [\%] \tag{25}$$

of the 1 heat pump

$$COP_{2hp,c,\%} = \frac{COP_{2hp,c}}{COP} \cdot 100 \ [\%] \tag{26}$$

of the 2 heat pump

Total coefficient of performance equalsthe sum of partial coefficients of performance

$$COP = COP_{ex} + COP_{1hp,c} + COP_{2hp,c} \quad [-] \tag{27}$$

4.9 Heat Dissipation during the Process

During the operation of the system through each of the components' surface heat dissipation occurs into the colder air of the environment. The quantity of the dissipated heat flux is approximately equal to the difference between the clean hot water fluxthat leaves the condenser of 2. heat pump and the waste water flux which enters the heat plate exchanger.

Dissipate heat flux in the whole heat recovery system can be calculated as follows

$$\dot{q}_{dis} = \dot{q}_{cw,o} - \dot{q}_{ww,i} \tag{28}$$

$$\dot{q}_{dis} = C_{pw} \cdot \dot{m}_{cw} \cdot T_{cw,o} - C_{pw} \cdot \dot{m}_{ww} \cdot T_{ww,i} \tag{29}$$

In fact, in steady operation mode of system, mass flow rate is unchangeable. Therefore, the mass flow rate of the clean and the waste water is approximately equal.

$$\dot{m}_w \cong \dot{m}_{ww} \cong \dot{m}_{cw} \tag{30}$$

After substitution the dissipate heat

$$\dot{q}_{dis} = C_{pw} \cdot \dot{m}_w \cdot (T_{cw,o} - T_{ww,i}) \tag{31}$$

4.1 Energy Savings as a Percentage

Saved energy flux is equal to the difference between obtained and invested energies

$$\dot{q}_{save} = \dot{q}_{wc} - P_{te} \tag{32}$$

Energy savings as a percentage is obtained if the saved energy flux is divided by the energy flux demand.

$$\frac{\dot{q}_{save}}{\dot{q}_{wc}} \cdot 100 = \frac{\dot{q}_{wc} - P_{te}}{\dot{q}_{wc}} \cdot 100 = \left(1 - \frac{P_{te}}{\dot{q}_{wc}}\right) \cdot 100 \tag{33}$$

After substituting the COP

$$\left(1 - \frac{1}{COP}\right) \cdot 100 \quad [\%] \tag{34}$$

5. Numerical Energy Analysis

5.1 Introduction

Measurement of the system parameters has been carried out during steady-state operation mode of the heat recovery system. All vital parameters are measured in order to determine the energy efficiency through the coefficient of performance (COP) and the payback period of the investment.

The measuring is conducted by applying:

- Glass thermometers filled with alcohol, with a resolution of 0.1 °C and measurement range from-10 to +110 °C. Figure 2
- Turbine flow meter: dimension 5/4", accuracy 2% at nominal volumetric flow rate of 5 m^3/h. Figure 3.
- Manual electrical Watt meter.

5.1.1 Unchangeable Measured Parameters of the System

- Mass flow rate of the clean and waste water
  $\dot{m}_{cw} \cong \dot{m}_{ww} \cong 87\ [kg/min] = 5220[kg/h]$
- Temperature of the clean well water $T_{cwi} = 13.8\ \left[^{o}C\right]$
- Electric power demand of the well pumps $P_{w\,pump} = 1\,359\ [W]$
- Electric power demand of the circulation pumps $P_{c\,pump} = 906\ [W]$

6 Numerical Results

6.1 COP Maximum

- Temperature of the clean water, flows out of the condenser $T_{cw,o}$=54.7 [^{o}C]
- Temperature of the waste water at the input of plate heat exchanger $T_{ww,i}$=48.8 [^{o}C]
- Electric power demand of the heat pump1. P_{hp1}= 19 338 [W]
- Electric power demand of the heat pump2. P_{hp2}=15 337 [W]

Recovered heat to the clean water

$$\dot{q}_{cw} = C_{pv} \cdot \dot{m}_{cw} \cdot \Delta T = 1.163 \frac{Wh}{kg\,^{o}C}\ 5\,220 \frac{kg}{h}\ (54.7 - 13.8)^{o}C$$

$\dot{q}_{cw} = 248\,298\,[W]$

Total electrical power demand

$$P_{te} = \sum_{1}^{2} P_{comp} + \sum_{1}^{2} P_{w\,pump} + \sum_{1}^{2} P_{c\,pump}$$

$$P_{te} = 19\,338 + 15\,337 + 2 \cdot 1359 + 2 \cdot 906 = 38\,205\,[W]$$

COP maximum of recovery system based on measured values

$$COP_{max} = \frac{q_{cw}}{P_{te}} = \frac{248\,298\,[W]}{38\;205\,[W[} = 6{,}499 \approx 6.5$$

$COP_{max} = 6.499 \approx \mathbf{6.5}$

6.2 COP Minimum

- Temperature of the clean water which flows out of the condenser $T_{cw,o}$=51.6 [^oC]
- Temperature of the waste water on the input of the plate heat exchanger $T_{ww,i}$=44.5 [^oC]
- Electric power demand of the heat pump1. P_{hp1}= 17 847 [W]
- Electric power demand of the heat pump2. P_{hp2}=14 168 [W]

Recovered heat to the clean water

$$Q_{CW} = C_{pv} \cdot \dot{m}_{CW} \cdot \Delta T = 1.163 \frac{Wh}{kg\,^oC}\; 5\,220 \frac{kg}{h}\; (50.1 - 13.8)^o C$$

Q_{cw} = 220 372 [W]

The value of the total invested electrical power:

$$P_{te} = \sum_{1}^{2} P_{comp} + \sum_{1}^{2} P_{w\,pump} + \sum_{1}^{2} P_{c\,pump}$$

P_{te} = 17 847 + 14 168 + 2 · 1359 + 2 · 906 [W]

P_{te} = 36 545 [W]

COP of the recovery system based on the measured values.

The value ofenergy efficiency ratio of recovery:

$$COP_{min} = \frac{q_{cw}}{P_{te}} = \frac{208\,928\,[W]}{34\;675\,[W[} = 6{,}03$$

$COP_{min} = 6.03 \approx \mathbf{6.0}$

6.3 Partial COPs

6.3.1 Partial COP of Heat Plate Exchanger

- Input temperature of clean water at the entrance of the plate heat exchanger $T_{ex,i}$=13.8°C
- Output temperature of clean water at the exit of the plate heat exchanger $T_{ex,i}$=33.9°C

$$COP_{ex} = \frac{\dot{q}_{ex}}{P_{te}} = \frac{C_{pw} \cdot \dot{m}_{cw} \cdot \Delta T_{cw}}{P_{te}} = \frac{1.163 \frac{Wh}{kg\,^oC}\ 5\,220 \frac{kg}{h}\ (33.9 - 13.8)\,^oC}{38\,205\,[W]} = 3.1939\,[-]$$

$$COP_{ex,\%} = \frac{COP_{ex}}{COP} \cdot 100 = \frac{3.1939}{6.499} \cdot 100\% = 49.15\,[\%]$$

6.3.2 Partial COP of Heat Pump1

- Input temperature of clean water at the entrance of the heat pump1 condenser $T_{1ht,c}$=33.9 °C
- Output temperature of clean water at the exit of the heat pump1 condenser $T_{1ht,c}$=46.2 °C

$$COP_{1hp,c} = \frac{\dot{q}_{1hp,c}}{P_{te}} = \frac{C_{pw} \cdot \dot{m}_{cw} \cdot \Delta T_{cw}}{P_{te}} = \frac{1.163 \frac{Wh}{kg\,^oC}\ 5\,220 \frac{kg}{h}\ (46.2 - 33.9)\,^oC}{38\,205\,[W]}\,[-]$$

$$COP_{1hp,c} = 1.9545[-]$$

$$COP_{1hp,c,\%} = \frac{COP_{1hp,c}}{COP} \cdot 100 = \frac{1.9545}{6.499} \cdot 100\% = 30.03\,[\%]$$

6.3.3 Partial COP of Heat Pump2

- Input temperature of clean water at the entrance of the heat pump2 condenser $T_{2ht,c}$ =46.2 °C
- Output temperature of clean water at the exit of the heat pump2 condenser $T_{2ht,c}$ =54.7 °C

$$COP_{2hp,c} = \frac{\dot{q}_{2hp,c}}{P_{te}} = \frac{C_{pw} \cdot \dot{m}_{cw} \cdot \Delta T_{cw}}{P_{te}} = \frac{1.163 \frac{Wh}{kg\,^oC}\ 5\,220 \frac{kg}{h}\ (54.7 - 46.2)\,^oC}{38\,205\,[W]}\,[-]$$

$$COP_{2hp,c} = 1.3511[-]$$

$$COP_{2hp,c,\%} = \frac{COP_{2hp,c}}{COP} \cdot 100 = \frac{1.3511}{6.499} \cdot 100\% = 20.79\,[\%]$$

6.3.4 Validity of COP

Total COP of heat recovery system can be calculated directly or through partial COPs the final result is the same. The proof is next.

$$COP = COP_{ex} + COP_{1hp,c} + COP_{2hp,c}$$

$$COP = 3.1939 + 1.9545 + 1.3511 = 6.499 \quad [-]$$

$$COP_{\%} = 49.15 + 30.03 + 20.79 = 99.97 \quad [\%]$$

6.3.5 Saved Energy in Percentage

Maximum of saving

$$\left(1 - COP^{-1}\right) \cdot 100 = \left(1 - 6.5^{-1}\right) \cdot 100\,[\%] = 85\%$$

Minimum of saving

$$\left(1 - COP^{-1}\right) \cdot 100 = \left(1 - 6.0^{-1}\right) \cdot 100\,[\%] = 83.3\%$$

7 Numerical Economic Analysis

7.1 Inputdata

Current average price of industrial electrical energy in Serbia, 2015 is 0.0463 [Euro/ kWh].

Predicted average consumption of clean water 5220 [kg/h]

Power for heating 5220 [kg/h] clean water from 13.8 [°C] to 55 [°C] is 248 [kW].

Nominal working hours for two shifts per day is 16 hours.

Coefficient utilization of nominal working hours is about 0.8.

7.2 Calculation

Consumption of hot clean water is a stochastic quantity because filling the four washing machines is periodical. Determination of the necessary clean hot water quantity was carried out, together with the amount estimation and the effective time of consumption.

Nominal consumption of heat energy 248 kW · 0.8 · 16 h = 3174 [kWh/day]

Predicted number of workdays a year

(52- 4) [week/ year] · 5 [day/week] = 240 [day/year]

Annual consumption of heat energy

3174 [kWh] · 240 [day/ year] = 761 760 [kWh/year]

Maximum consumed electrical energy annually of all components of the recovery system

38,2 [kW] · 0.8 · 16 [h] · 240 [day/ year]= 117 350 [kWh/ year]

Annual electrical energy savings of the recovery system

761 760 [kWh /year] -117 350 [kWh/year] = 644 410 [kWh/year]

Value of the saved electrical energy per year

644 410 [kWh / year] 0.0463 [euro / kWh] = 29 836 [Euro / year]

Approximate total necessary investment for realization of the recovery system:

60 000[Euro]

Expected payback period of the invested money when an electrical heater recovery system is used for clean water heating:

60 000 [Euro] / 29 836 [euro/year] = 2.01 [year]

8 Results in the Graphics

Numerically obtained results in graphical presentation.

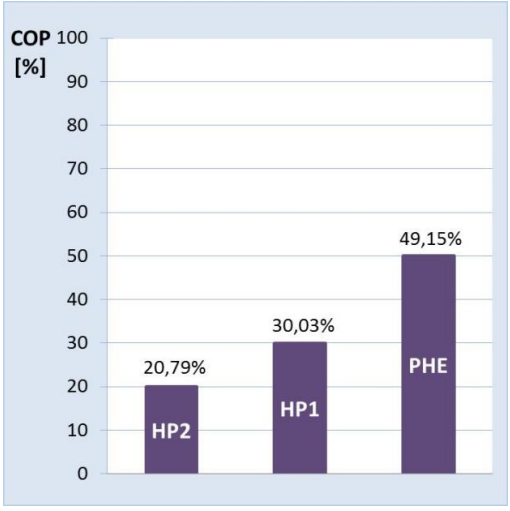


Figure 2

The percentage share of individual components in the heat recovery process: PHE-plate heat exchanger 49.15%, HP1- heat pump filled with R 407C 30.03%, HP2- heat pump filled with R 134a 20.79%

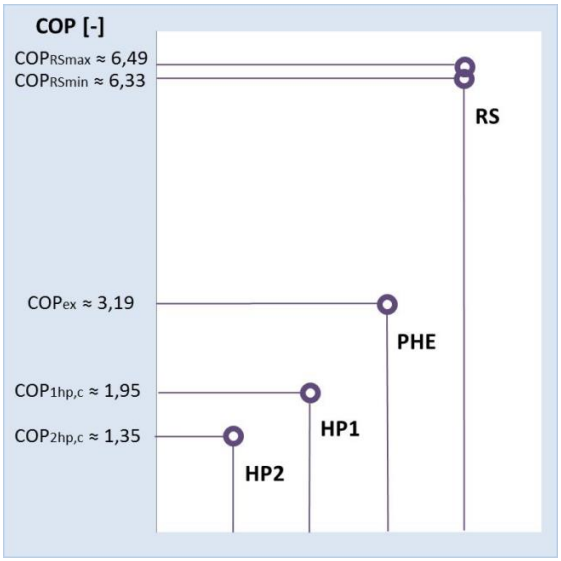


Figure 3
COP of the components and of complete heat recovery system

Conclusion

The recovery system described in this article is quite efficient both in energetic and economic terms.

- The system's total coefficient of performance is very high: COP is from 6 to 6.5
- The component's partial COPs are the following: 3.19 [-]for the heat plate exchanger, 1.95 for the heat pump1, 1.35 for the heat pump2
- The total energy savings expressed in percentages:

  Minimally (1-1/6) 100% = 83.3% or

  Maximally (1-1/6,5) 100% = 85%
- The energy savings of the components in percentages: Plate heat exchanger 49%, low temperature heat pump 30%, high temperature heat pump 21%
- The payback period of the investment is relatively short, about 2 years.

In addition to energy and economic advantages this solution is also very favorable in terms of environment protection. The recovery system protects the environment on direct and indirect way.

Directly:

1) The waste water does not enter in the sewer while it is still hot.

2) The waste water is cleaned by filtration before getting out of the recovery system.

Indirectly:

1) By saving a great quantity of heat energy the electrical energy use is significantly reduced.

2) Less use of electricity, less working power plant the result is a reduced CO_2 emission.

The presented results are valid for the economic situation in Serbia, 2015.

The task of further research is to find the optimal structure and dimensions of the system's thermal and energy components in order to improve the energy and economic efficiency of the new heat recovery system.

References

[1] Fangtian Sun, Lin Fu, Jian Sun and Shigang Zhang:"A New Waste Heat District Heating System with Combined Heat and Power (CHP) Based on Ejector Heat Exchangers and Absorption Heat". IJ. Energy Vol. 69 (2014) pp. 516-524

[2] Jian Sun, Lin Fu, Fangtian Sun and Shigang Zhang: "Study on a Heat Recovery System for the Thermal Power Plant Utilizing Air Cooling Island". IJ. Energy Vol. 74 (2014) pp. 836-844

[3] Marija S. Todorovic, Jeong Tai Kim:" In Search for Sustainable Globally Cost-Effective Energy Efficient Building Solar Syste1.m – Heat Recovery Assisted Building Integrated PV Powered Heat Pump for Air-Conditioning, Water Heating and Water Saving". Energy and Buildings, Volume 85, December 2014, pp. 346-355

[4] Bicui Ye, Jun Liu, Xiangguo Xu, Guangming Chenand Jiao Zheng:"A New Open Absorption Heat Pump for Latent Heat Recovery from Moist Gas". Energy Conversion and Management, Volume 94, April 2015, pp. 438-446

[5] Ming Qu, Omar Abdelaziz and Hongxi Yin:"New Configurations of a Heat Recovery Absorption Heat Pump Integrated with a Natural Gas Boiler for Boiler Efficiency Improvement". Energy Conversion and Management, Volume 87, November 2014, pp. 175-184

[6] Zhao Yang, Yang Zhuo, Luo Ercang and Zhou Yuan: "Travelling-Wave Thermo Acoustic High-Temperature Heat Pump for Industrial Waste Heat Recovery". Energy, Volume 77, 1 December 2014, pp. 397-402

[7] J. Nyers, L. Garbai, A. Nyers: "Analysis of Heat Pump's Condenser Performance by means of Mathematical Model". International J. Acta Polytechnica Hungarica, Vol. 11, No. 3, pp. 139-152, 2014, DOI: 10.12700/APH.11.03.2014.03.9

[8] György Györök: A Special Case of Electronic Power Control of Induction Heating Equipment. Acta Polytechnica Hungarica Vol. 11, No. 5, pp 235-246, 2014. DOI: 10.12700/APH.11.05.2014.05.14

[9] László Kajtár, Miklós Kassai, László Bánhidi: Computerised Simulation of the Energy Consumption of Air Handling Units, 2011, Energy and Buildings, ISSN: 0378-7788, (45) pp. 54-59

[10] Miklos Kassai: Effectiveness and Humidification Capacity Investigation of Liquid-to-Air Membrane Energy Exchanger under Low Heat Capacity Ratios at Winter Air Conditions. Journal of Thermal Science. Vol. 24, No. 4, pp. 391-397, DOI: 10.1007/s11630-015-0800-4 (2015)

[11] S. Pavlović, D. Vasiljević, V. Stefanović, Z. Stamenković, S. Ayed: "Optical Model and Numerical Simulation of the New Offset Type Parabolic Concentrator with Two Types of Solar Receivers". Facta Universitatis Series: Mechanical Engineering Vol. 13, No 2, pp. 169-180, 2015

[12] B. Bokor, L. Kajtár: Luftkollektoren: Nutzung der Sonnenenergie in der Gebäudelüftung e-nova Internationaler Kongress 2014, Nachhaltige Gebaude, 13-14. November 2014. Pinkenfeld, pp. 305-310, ISBN: 978-3-7011-0316-4

[13] M. Todorović, D. Živković, M. Mančić, D. Milčić: Dynamic Behaviour of Hot Water Boilers during Start UP. Facta Universitatis Series: Mechanical Engineering Vol. 12, No. 1, pp. 85-94 (2014)

[14] F. Kalmár, GY. Csomós, Á. Lakatos: "Focus on Renewable Energy (energy policy)—Editorial"International Review of Applied Sciences and Engineering 4 (1), 95-95 (2013) DOI: http://dx.doi.org/10.1556/IRASE.4.2013.1.13

[15] Magyar Zoltán, Révai Tamás: Thermal Insulation of the Clothing 2nd Royal Hungarian Army in Winter Campaign in the Light of Thermal Manikin Measurements. Acta Polytechnica Hungarica 11:(7) pp. 197-207 (2014)

[16] Á. Lakatos, S. Szigeti, F. Kalmár: "Measurement of the Water Uptaking Capability of a Thermal Insulating Paint". International Review of Applied Sciences and Engineering 4 (2), 157-161 (2013) DOI: http://dx.doi.org/10.1556/IRASE.4.2013.2.10

The Use of Linear Programming for Determining Number of Fire-Fighters on Shifts in Case of Special Events

Obrad Stevanović[1], Dalibor Kekić[1], Valentin Kónya[2], Miloš Milenković[3]

[1] Academy of Criminalistic and Police Studies, Department of Police Sciences, Cara Dušana 196, 11080 Zemun, Belgrade, Serbia

[2] Faculty of Economics, University of Novi Sad, Segedinski put 9-11, 24000 Subotica, Serbia

[3] University of Belgrade, Faculty of Organizational Sciences, Jove Ilića 154, 11000 Belgrade, Serbia

obrad.stevanovic@kpa.edu.rs, dalibor.kekic@kpa.edu.rs, valentink@uns.ac.rs, mijatov51804@fon.bg.ac.rs

Abstract: This paper focuses on finding the most suitable number of fire-fighters who will be engaged on special events, such as concerts, sport tournaments, music events or similar events, while at the same time ensuring the number of fire-fighters in a particular fire station for each shift will be enough to ensure the overall social welfare of a community and remain at the lowest possible costs. Due to the risk of fire and other public welfare risks, it is often common fire-fighters are required to be engaged in special events with their duty primarily at those special event locations. Their main task certainly is to protect all people in a community and its assets by quick response in case of an emergency. Consequently, it remains necessary to provide a sufficient number of fire-fighters who will remain at the fire station so that they meet the standard requirements of them during each shift. In cases of temporary absences and permanent attrition in the fire-fighter workforce, sometimes commanders need to use overtime and hiring of additional fire-fighters to satisfy the needs of organizers of special events. In those situations, they also have to think about minimizing total staffing costs and budget restraints. This paper uses linear programming to develop a model for determining the optimal number of fire-fighters to have on staff during additional demands of special events still covering the standard emergency coverage within a community. Also, the aim of this paper is to present a model which will facilitate the decisions of commanders of fire stations when determining the number of fire-fighters required for ordinary and extraordinary activities.

Keywords: fire-fighter; model; constraints; optimization

1 Introduction

Disasters are appearing more and more often in our world than ever before. They cause victims and also a huge amount damages in many areas of life, e.g. energy distribution, flow of transportation and logistics, health care demands, availability of education, financial costs, housing and many others. Despite overwhelming disasters, every day events happen such as fires, explosions and other similar day to day tragedies. In all cases the main emergency forces protecting and rescuing the public during these occurrences are our fire-fighters. Statistics shown in the large number of calls, which they receive for assistance on a daily basis, around the world, speak enough about their great contribution to society and the phenomenal demands placed on them.

Looking at the history and development of the fire service, in its early stages individuals were primarily part of the service as volunteers. However, as the requirements of the profession became more sophisticated and the volume of calls increased, it became necessary to employ qualified, full-time, well-trained fire-fighters [1, 16]. Production development, new technologies, as well as climate change, created a growing number of risks that endangered people, material and cultural assets. In connection with this fact, those influences on society created increases in the need for and engagement of professional fire-fighters in our communities. However, as most often happens the number of fire-fighters has not kept up with the growth of those influences.

In addition to the regular requirements of community protection, care and welfare of citizens and their property in the event of social and special events fire-fighters are needed to be present at them in official capacity. Sport games and competitions, music events, visits of foreign statesmen and similar, are examples of the previously mentioned events. So, the Fire Station Commanders and Emergency Coordinators have to make decisions of the number of fire-fighters that will be engaged in those special events and at same time what manpower will remain in the fire station covering routine emergency standby requirements. These management personnel in many cases have difficult decisions to make when deciding on the appropriate number of engaged fire-fighters, most especially when comparing the workload to the number of available personnel.

One of the limiting conditions is that those workers may not work constantly because of “Occupational Health Risk Limitations” which are required due to the exposures and hazards found within the work environment. Usually, fire-fighters work during two shifts as is the case with fire service in the Republic of Serbia. After a daily shift of 12 hours, which starts at 7 am and finishes at 7 pm they have a break of 12 hours. The opposite shift, which starts at 7 pm and finishes at 7 am has a break of 24 hours. So, in case of a special event and with a limited number of staffed fire-fighters it is not so easy to make the right decision which number of them will be can be assigned considering the restrictions that apply within the work-rest cycle. Another consideration is connected to the costs of additional engagement of fire-fighters. Fire-fighters are, generally, officers of some governmental institution, for example Ministry of Interior or Directorate for

Emergency Situations. In times of economic crisis, the budget of the state institutions is quite limited. Most often Commander of Fire Stations can deploy only that additional number of fire-fighters, in case like special events, which is consistent with the assigned budget. By linear programming, we can define a model with the previously mentioned constraints. This can help in determining the optimal number of engaged fire-fighters for any given shift, of whom part may be allocated to the special event.

2 Theoretical Background

During the last two decades, a great number of models that solve problems connected with determining number of security forces have been developed. However, the use of linear programming in this sense mostly is directed towards the police force. Taylor and Huxley [2] developed a police patrol scheduling system (PPSS). Currently, this system is used by all San Francisco (SF) police precincts which use PPSS to schedule all of their officers. It is estimated that PPSS saves the SF police more than $5 million annually. This system, especially, proved to be beneficial when it was used for planning the required number of personnel and additional requirements for hiring for concerts, sporting events or rallies.

In the last decade a large number of authors have dealt with the theme of optimizing the number of operational forces that are required to react in case of an emergency events or situation. Also, some of them were focused on developing the most appropriate model for determining the number of fire-fighters which are required for just one shift to satisfy all the demands of a local society. So, some of them use stochastic models based on the queuing theory. Although the operation of the fire stations, as an example of one unit intended for reacting in cases of emergencies, is very complex, many researchers try to focus on the most important parts of the staff process and build simpler stochastic queuing models to study the problem of staffing. These models strive to seek a balance between tractability and validity and obtain some very valuable results [3, 4–6].

When defining the model, significant attention of some research is paid to financial constraints in hiring of additional numbers of fire-fighters, as is the case of during special events. Fry, Magazine and Rao [7] examine the problem of determining the annual staffing level that minimizes total expected costs for a fire department, subject to minimum service-level-based staffing requirements. They develop a quantitative model that allows for stochastic temporary absences, permanent wastage, and limited hiring opportunities, which takes into account the unique fire-fighter work schedule.

Thesis about optimizing fire department operations through work schedule analysis, alternative staffing, and non-productive time reduction, done by the Naval postgraduate school [8], conducts a policy analysis exploring how current fire department policies can be modified to optimize employee availability to lead

to higher staffing levels and lower sick leave and injury leave usage. Work schedule modification and alternative staffing models are the two options examined in this thesis. Also, many authors try to evaluate different shift work systems in a rigorous, controlled environment, using measures relevant to the cognitive demands of incident management and firefighting [9].

3 Methodology

In this paper, the model for finding the optimal number of fire-fighters required on shifts during special events is suggested. Many constraints affect defining the optimal number of fire-fighters in those situations. The most influential constraint is the minimal requested number of fire-fighters which are needed by organisers of special events, such as: musical concerts, sport matches and the similar. On the other hand, based on risk assessments or statistical data of the fire station what each shift commander requires to satisfy the minimal number of professionals who are engaged. Also, a limiting factor should be the budget of fire service. In a times where all state institutions have limited available funds, all issues and costs of support must be taken into account when adding additional manpower to the workforce. Another, interesting question and at same time limiting factor for leaders and managers when making decisions about the number of engaged fire-fighters, is what to do in case of vacations or employee absence due to sick leave. Linear programming is used for minimization of engaged fire-fighters per a single shift from which part of the manpower will be sent to a special event. All mentioned constraints are taken into account during model definition.

3.1 Model Definition

In order that the commander of the fire station make the right decision concerning the number of fire-fighters in one shift during special events, with minimal costs, the model has to be correctly defined with all real constraints [15]. Firstly, this problem should only occur when organiser of special event has a legal obligation to have a fire-fighter on site or they recognize a need for them because of safety concerns. Usually, after a meeting between organizers and representatives of the security services, decisions will be made about necessary forces to be engaged. So, commanders of territorially competent fire station will receive the request for engagement of fire-fighters for that special event, in the form of an official letter sent by the organizers. Practice shows that on the basis of specific criteria, such as the expected number of spectators, time of day and experiences from previous similar events will be given and the minimum requirements for the number of fire-fighters present. An additional problem may be that there are unequal required numbers of fire-fighters during different periods of time during the event. The main part of the letter sent to the fire station should provide this information in table form.

Table 1
Example of table form of the required number of fire-fighters

Time period of special event	The minimal number of fire-fighters
Time period 1	n_1
Time period 2	n_2
Time period n	n_i

In addition to the new taking the commander of the fire station has an obligation to provide a sufficient number of fire-fighters who will be at the fire station ready for action during the ordinary and extraordinary circumstances. Most fire departments around the world are organized to operate in two shifts that last twelve hours. For the purposes of this paper such a mode will be taken into account.

Table 2
Review of shift duration and minimal number of fire-fighters per one shift

Shift	Shift start	Shift end	Minimal number of fire-fighters per shift
1	Time1	Time2	N_1
2	Time2	Time1	N_2

We want to minimize the total number of fire-fighters scheduled, for a special event and also for the fire station, so the model will be defined as follows:

$$\min f(x) = x_1 + x_2 \tag{1}$$

subject to:

$$x_1 \geq n_i \tag{2}$$

$$x_2 \geq n_i \tag{3}$$

$$x_1 + x_2 \geq n_i \tag{4}$$

$$x_1 \geq N_1 \tag{5}$$

$$x_2 \geq N_2 \tag{6}$$

$$x_1, x_2 \geq 0 \tag{7}$$

where x_1 is the number of fire-fighters who will work in the first shift, and x_2 is the number of fire-fighters who will work in the second shift.

The objective function (1) is to minimize the total number of engaged fire-fighters on both shifts. Constraint (2) shows the total number of fire-fighters on the first shift have to be more or equal to the required number of them in the time period i (i=1,2…, the last time period) of the special event. It is important to mention that organizers usually have different demands for fire-fighters during different time periods of a special event. For example, organizers of the New Year celebration divide the whole celebration into different periods and based on mentioned criteria

define different numbers of fire-fighters needed for a certain periods of time. It is clear that the number of fire-fighters on the first shift will be limited to the required minimum for the period of special events that fall under the first shift. Constraint (3) shows the total number of fire-fighters in the second shift have to be more or equal to the required number of them during the time period i (i=1,2…, the last time period) of the special event. Also, the number of fire-fighters on the second shift will be limited to the required minimum for the period of the special events that fall under the second shift. Constraint (4) shows that in the case that a certain period of time of the special event refers to both shifts the total number of fire-fighters in shifts must meet the required minimum number of fire-fighters during this period. Constraint (5) points that the number of fire-fighters in the first shift have to be more or equal to the minimal number of fire-fighters per that shift. The same applies to constraint (6), only it relates to the second shift. Practice shows that on the basis of certain standards each shift requires a certain number of fire-fighters. Also, risk assessment is a very useful tool for determining those standards. Additionally, the number of fire-fighters that are needed for each shift may vary according to time of year, from natural and environmental and other disasters issues. The warnings and regulations are established and given by competent authorities concerning the possible dangers. In many countries around the world specific policies and guidelines precisely define the numbers of fire-fighters for each shift. An example of this is in the United States of America where the National Fire Protection Association (NFPA) publishes standards for fire departments to use a guide. Although they are not required to follow, the publications often become the way of conducting business in the fire service. The NFPA Index was examined for standard(s) of recommended numbers of fire-fighters when performing fire ground operations or as required number per shift [10]. For adequate determination of numbers N_1 and N_2, which will be used for constraints, can serve as standards for the required number of fire-fighters by population protected. European Union norms envisaged that per every 1,000 inhabitants a country must have at least one fire-fighter [11]. As usual, the nonnegative constraint (7) completes the formulation.

In some cases, the objective function of the mathematical model and the standard of scheduling will result in the least expenditures of the daily budget and not reduce the normal levels of employees. This may lead to a change of the objective function and that may lead to a change to the optimal solution. The commander of a fire station that is responsible for determining the number of fire-fighters is well acquainted with the fact that daily allowances which would be utilized to engage fire-fighters on the second shift are larger than daily allowances of fire-fighters who work during the first shift. In accordance with this and in a situation of where you have limited funds it is sometimes necessary to include this criterion in defining the mathematical model. What Fred McChesney wrote some 10 years ago is even more true today that taxpayers are unlikely to support budget increases for fire departments if they see inactive firemen lolling about the fire station [12]. So, cities have created new, highly visible jobs for their firemen. The Wall Street Journal reported recently, “In Los Angeles, Chicago and Miami, for example, 90% of the emergency calls to firehouses are to accompany ambulances to the scene of

auto accidents and other medical emergencies. Elsewhere, to keep their employees occupied, fire departments have expanded into neighbourhood beautification programs, gang intervention, substitute-teaching and other non-firefighting pursuits." [13]. So, the engagement of fire-fighters with special events can certainly be adequate to justify their manpower numbers at fire stations, but in determining the manpower levels financial costs must be taken into account. When we know the amount of daily allowance for hiring fire-fighters, to the first, and the second shift model, we can have the following form:

$$\min f(x) = C1 \cdot x_1 + C2 \cdot x_2 \tag{8}$$

where x_1 still is the number of fire-fighters who will work in the first shift, and x_2 is the number of fire-fighters who will work in the second shift. The objective function (8) is to minimize the total cost of engaged fire-fighters in both shifts. Constraints (2-7) in this case the same form remains. C_1 is the value of cost for the daily allowance for one fire-fighter on the first shift, while C_2 is the value of cost for the daily allowance for one fire-fighter on the second shift. Those values are not same because value C_2 has to be bigger than the first one as it refers to night shift which according to existing regulations requires a higher fee.

The duties of a fire-fighter carries with them certain risks, which often have a negative impact on their health. Inherent in working for the fire service or emergency services is the requirement to work a shift schedule that includes nights, a determined health risk. The challenge for administrators is to provide the necessary level of service to the community while recognizing that night work has consistently been recognized as a serious risk factor for workers' health as it interferes with our basic biological functions, social relations, and psychological/mental health [1]. All those indicate the possible absence of fire-fighters due to health problems. Also, the interventions of fire-fighters are often performed in difficult conditions and injuries at work are common. Also, fire-fighters may be absent from work due to vacation days that they normally earn and acquire. In connection with the previously mentioned appears an additional constraint which is:

$$x_1 + x_2 \leq N \tag{9}$$

where, in addition to the above described variables x_1 and x_2, N is total number of available fire-fighters on which the commander of a fire station can rely, including all absent.

4 Results

This research is based on a fire station whose number of employed fire-fighters, according to the standards of the European Union, serve a populated area of about 50,000 inhabitants and which received from organizers of a celebration of a New Year special event a request for engaging fire-fighters at various different

locations. In this paper, taking into consideration that mentioned constrains are relevant for making decision about the optimal number of engaged fire-fighters, we summarized criteria from previous studies in this area, as the outcomes of the literature review, in order to more precisely understand the influence of demands for additional commitment arising out of special events to make right decisions of manpower. Finally, tables shown in the following part of the paper represent the application of dual model as a method of linear programming.

At the beginning of the case study it should be noted that the organizers of the New Year celebrations had already determined, based on the experience of previous years, how many fire-fighters were necessary to protect different areas in the period of 7.00 am on 31st December to 7.00 am on 1st January. They found, that the number of needed fire-fighters depends on the time of day as shown in Table 3.

Table 3
Required number of fire-fighters

New Year celebrations time period	Minimal number of fire-fighters
7 am 31st December – 4 pm 31st December	5
4 pm 31st December – 24:00 31st December	8
0:00 1st January – 7 am 1st January	12

Fire-fighters from this fire station work in shifts as is shown in the table below. It should be pointed out that third shift is the inter-shift or spare shift that in some situations is inserted into the manpower scheme as the need arises, and the first two shifts are standard shifts.

Table 4
Shift duration

Shift	Shift start	Shift end	Number of fire-fighters per shift as variable of model
1	7 am	7 pm	x_1
2	7 pm	7 am	x_2
3	4 pm	4 am	x_3

Initially, we should formulate a mathematical model of the given problem. Target function will be the minimization of the total number of engaged fire-fighters which will satisfied the order from organizers. The model would seem:

$$\min f(x) = x_1 + x_2 + x_3 \tag{10}$$

subject to:

$$x_1 \geq 5 \quad \text{required number of fire-fighters between 7 am and 4 pm} \tag{11}$$

$$x_1 + x_3 \geq 8 \quad \text{required number of fire-fighters between 4 pm and 7 pm} \tag{12}$$

$$x_2 + x_3 \geq 8 \quad \text{required number of fire-fighters between 7 pm and 24:00} \tag{13}$$

$x_2 + x_3 \geq 12$ required number of fire-fighters between 0:00 and 4 am (14)

$x_2 \geq 3$ required number of fire-fighters between 4 am and 7 am (15)

$xj \geq 0\, j = 1,2,3$ (16)

Constraints in this model are in accordance with the required number of fire-fighters at the indicated time periods of the New Year celebrations and duration of the shifts. For determining the optimal solution, it is recommended solving through the dual model as a method of linear programming, which will be shown below.

$$\max \varphi(y) = 5y_1 + 8y_2 + 8y_3 + 12y_4 + 3y_5 \quad (17)$$

subject to:

$$y_1 + y_3 \leq 1 \quad (18)$$

$$y_3 + y_4 + y_5 \leq 1 \quad (19)$$

$$y_2 + y_3 + y_4 \leq 1 \quad (20)$$

$$yi \geq 0 i = 1,2,3,4,5 \quad (21)$$

$$\max \varphi(y) = 5y_1 + 8y_2 + 8y_3 + 12y_4 + 3y_5 + 0 \cdot (j_1 + j_2 + j_3) \quad (22)$$

subject to:

$$y_1 + y_3 + j_1 = 1 \quad (23)$$

$$y_3 + y_4 + y_5 + j_2 = 1 \quad (24)$$

$$y_2 + y_3 + y_4 + j_3 = 1 \quad (25)$$

$$yi \geq 0 i = 1,2,3,4,5 \quad (26)$$

Table 5

T_0 of dual model

			5	8	8	12	3	0	0	0	
Cb	Base	CFA	y_1	y_2	y_3	y_4	y_5	j_1	j_2	j_3	CFA/ y_4
0	j_1	1	1	1	0	0	0	1	0	0	-
← 0	j_2	1	0	0	1	**1**	1	0	1	0	1
0	j_3	1	0	1	1	1	0	0	0	1	1
-f		0	5	8	8	12	3	0	0	0	

↑

Table 6

T_1 of dual model

			5	8	8	12	3	0	0	0	CFA/
Cb	Base	CFA	y_1	y_2	y_3	y_4	y_5	j_1	j_2	j_3	y_2
0	j_1	1	1	1	0	0	0	1	0	0	1
12	y_4	1	0	0	1	1	1	0	1	0	-
← 0	j_3	0	0	**1**	0	0	-1	0	-1	1	0
-f		-12	5	8	-4	0	-9	0	-12	0	

↑

Table 7

T_2 of dual model

			5	8	8	12	3	0	0	0	CFA/
Cb	Base	CFA	y_1	y_2	y_3	y_4	y_5	j_1	j_2	j_3	y_1
← 0	j_1	1	**1**	0	0	0	1	1	1	-1	1
12	y_4	1	0	0	1	1	1	0	1	0	-
8	y_2	0	0	1	0	0	-1	0	-1	1	0
-f		-12	5	0	-4	0	-1	0	-4	-8	

↑

Table 8

T_3 of dual model

			5	8	8	12	3	0	0	0
Cb	Base	CFA	y_1	y_2	y_3	y_4	y_5	j_1	j_2	j_3
5	y_1	1	1	0	0	0	1	1	1	-1
12	y_4	1	0	0	1	1	1	0	1	0
8	y_2	0	0	1	0	0	-1	0	-1	1
-f		-17	0	0	-4	0	-6	-5	-9	-3
			↑ s_1	↑ s_2	↑ s_3	↑ s_4	↑ s_5	↑ x_1	↑ x_2	↑ x_3

From Table 8 we can read the result of solving our dual model. The results shown in the table is - 17 (-f). That number of fire-fighters is the minimum number required to satisfy the needs of the organizers of the New York celebration to cover all areas of the event. From them 5 would be on the first shift (x_1), 9 on the second(x_2) and 3 fire-fighters on the third shift (x_3). At the same time, in the period between 7 pm and 24:00 would be 4 more fire-fighters than is minimally necessary (s_3). Also, in the period between 4 am and 7 am would be 6 additional fire-fighters than would be minimally necessary (s_5).

With those constraints in the primary model we should add additional constraints, such as those relating to limited financial resources when hiring fire-fighters. The commander of the fire station, who is responsible for determining the number of

engaged fire-fighters analyses daily allowances which would be paid in salaries to both fire-fighters concludes that the manpower on the second and third shift have to be paid double the daily allowances than required on the first shift. Therefore, it is concluded that the criteria in the engagement of the fire-fighters should not be hiring fewer numbers, but to minimize the paid daily allowances. In this case there would be a change of the objective function. The new objective function is:

$$\min f(x) = x_1 + 2x_2 + 2x_3 \tag{27}$$

However, in such a defined difference in wages there would not have been a change in the structure of the optimal solution. Change of the objective function of the primary model causes a change in the free member's limitations of the dual model. Solving the dual again will determine that all values in optimal base are nonnegative and that there is no change in the structure of the optimal solution.

Also, constraints relating to the minimum number of fire-fighters per shift also could influence the final results. Different standards or risk assessments are useful documents for defining those constraints which may be required to be modified or satisfied at any moment. So, for purpose of this part the paper emphasis is placed on constraints given by the organizers of special events. The same applies to the number of available fire-fighters in relation to the number of absent fire-fighters for various reasons. The solving procedure would be the same, only there would have been a larger number of constraints.

5 Discussion

The final results obtained in table 8 shows the aim of this research which was to obtain the required number of engaged fire-fighters per defined shift. These results are obtained on the basis of the defined mathematical model where the objective function minimization of total number of fire-fighters including constraints provided by the organizers of special event be defined in the case study.

According to this function and the constraints, which are used as part of one case study, the optimal solution of primal model is multiple because the optimal solution of the dual model is degenerated, respectively that in two periods of time there would be more fire-fighters than is necessary. By involving the constraints relating to the payment of daily allowances it can be seen that there are no changes in the optimal solution.

We can conclude that, through solving a mathematical model, decisions about the number of fire-fighters who will satisfy the requirements of the organizers of special events can be done in a more accurate way. Obtained results presented how different orders for engagement of fire-fighters with the goals of desired safety of all participants can be realized in very exact way. By implementation of the organizer's requirements into a mathematical model built up on the basis of linear programming, it is possible to make decisions more precisely. Final results

achieved through solving a dual model are more accurate than some made only on the basis of previous experience.

What is frightening about these kinds of problems is that no matter how successful you have been in applying linear programming to solve problems, the next problem you confront may defy your best efforts to get a handle on it [14]. Probably, the "force-device" calculation model for determining the optimal number of fire-fighters on shifts in case of special events is not the simplest version. Commanders of fire stations in real practice have a lot of tasks to solve every day. Manual use of linear programming, without any software, for them will be very hard. Therefore, the desired solution would be to utilize an appropriate venue - "force-device'' calculation model.

Conclusion

The goal of this paper is to recommend the use of linear programming when it is very difficult to make decisions about the right number of fire-fighters who will be present during a special event and at same time their maintaining their standard social requirements, working with limited financial resources and ensuring the safety of all citizens will not be jeopardized. A special part of this paper belongs to the practical use of linear programming which is shown through case study, and its implementation while determining the number of fire-fighters on shifts during particular special events. In making such decisions it is recommended that decision makers use this mathematical model to find the suitable solution. Some considerations are how difficult is it to train commanders of fire stations, which will be making these decisions, to use reality linear programming and the implementation that would follow. It is possible, but it is essential that this method be mastered by seasoned administrator of fire-fighters within the central part of state institutions before implementation. So, the requirements of organizers of special events will be submitted to the commander of a fire station who in turn will to involve the administrative evaluator (software user) who will utilize the model to analyze the requirements of manpower for any given event.

Also, a special computer program can be created for this purpose in connection with linear programming and our defined model. This software can be created to further obtain significant data in the over-all management of special events resulting in a very useful tool for the managers. This software, based on predefined constraints, would automatically solve the model and provide results with number of fire-fighters per each shift in case of special events and may also, provide statistical data and reports for use of the manager. LINDO and LINGO are examples of software products which can be useful base for making appropriate applications or probably sufficient for determining number of fire-fighters on shifts in case of special events.

References

[1] Cohen, I., and Plecas, D.: A Review of the Research Literature on 24-Hour Shifts for Fire-fighters, Centre for Public Safety and Criminal Justice Research, University of the Fraser Valley, 2015

[2] Taylor, P., and Huxley, S.: A Break from Tradition for the San Francisco Police: Patrol Officer Scheduling Using an Optimization-Based Decision Support System, San Francisco, 1989

[3] Zeltyn, S., Marmor, Y. N., Mandelbaum, A., et.al.: Simulation-based Models of Emergency Departments: Operational, Tactical, and Strategic Staffing, ACM Transactions on Modeling and Computer Simulation, Vol. 21, No. 4, (2011)

[4] Green, L. V., Kolesar, P. J., Soares, J.: Improving the Sipp Approach for Staffing Service Systems that have Cyclic Demands, Operations Research, Vol. 49, No. 4, (2001) pp. 549-564

[5] Green, L. V., Soares, J., Giglio, J. F., Green, R. A.: Using Queueing Theory to Increase the Effectiveness of Emergency Department Provider Staffing, Academic Emergency Medicine, Vol. 13, No. 1, (2006) pp. 61-68

[6] Yom-Tov, G. B., and Mandelbaum, A.: Erlang-R: a Time-Varying Queue with ReEntrant Customers, in Support of Healthcare Staffng, Working Paper, Technion-Israel Institute of Technology, 2011

[7] Fry, M., Magazine, M., Rao,U.: Fire-fighter Staffing including Temporary Absences and Wastage, Operations Research, Vol. 54, No. 2 (2006) pp. 353-365

[8] Payne, J.: Optimizing Fire Department Operations through Work Schedule Analysis, Alternative Staffing, and Nonproductive Time Reduction, Naval postgraduate school, Monterey, 2014

[9] Short, M., et al.: Shift length: When less is more, Fire note, No. 120, 2014

[10] Monday, C.: Optimal Staffing Levels for Fire-Fighter Effectiveness and Mitigation of Fire-Fighter Injuries, Martinsville Virginia, 2000

[11] Government of the Republic of Serbia: Strategy of fire protection for the period 2012-2017, The Official Gazette of the Republic of Serbia, 2012

[12] Mc Chesney, F.: Smoke and errors, Library of economics and liberty, 2002

[13] Johnson, R.: As Blazes Get Fewer, Fire-Fighters Take on some Eclectic New Roles, The Wall Street Journal, 2001

[14] Kane, J.: The Art of Linear Programming, Professional Statistical Services, 2012

[15] Kónya, V., Grubić-Nešić, L., Matić, D.: The Influence of Leader-Member Communication on Organizational Commitment in a Central European Hospital, Acta Polytechnica Hungarica, Vol. 12, No. 3 (2015) pp. 109-128

[16] Kónya, V., Matić, D., Pavlović, J.: The Influence of Demographics, Job Characteristics and Characteristics of Organizations on Employee Commitment, Acta Polytechnica Hungarica, Vol. 13, No. 3 (2016) pp. 119-138

An Analysis of Employer Requirements of University Graduates

Tünde Tóth-Téglás[1], Erika Hlédik[2], Laura Fónadová[3]

[1] Óbuda University, Faculty of Business and Management, Tavaszmező utca 17, H-1084 Budapest, Hungary; tothne.teglas.tunde@kgk.uni-obuda.hu

[2] Budapest University of Technology and Economics, Faculty of Economic and Social Sciences, Műegyetem rkp. 3, H-1111 Budapest, Hungary; hledike@erg.bme.hu

[3] Masaryk University Brno, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, Czech Republic; laura@econ.muni.cz

Abstract: It is of equally outstanding importance for university graduates and universities alike to learn what requirements workplaces have of graduates and what knowledge and skills they are expected to have. Previous research in Hungary conducted by GVI MKIK and DPR examined employer and higher education graduate views based on the importance of expected competencies. In our research we focused on understanding employer requirements related to the recruitment of graduates. Therefore, we examined the opinions of companies that assumingly had a conscious strategy for employing graduates, and we carried out comprehensive analyses of the findings of comprehensive research. Our research findings reveal that the large companies surveyed assigned primary importance to competencies ensuring present performance; at the same time, the aspect of competencies enabling future performance also emerged as relevant. It was also analysed in our study to what extent the companies surveyed had different requirements of economic and technical graduates. The research furthermore explored with the help of factor analysis in what ways the respective competency requirements were interrelated in the employers' approach. Although our research was conducted among Hungarian companies exclusively, we believe that the findings related to the role of corporate size and field of specialisation may generate conclusions of interest for research carried out in other countries as well.

Keywords: workplace requirements; competencies; skill requirements; graduate recruitment process

1 An Analysis of Employer Requirements of Graduates

Understanding what requirements workplaces have of graduates, what knowledge and skills they are expected to have is extremely important for university graduates and universities. Previous research in Hungary conducted by the Institute for Economic and Enterprise Research of the Hungarian Chamber of Commerce and Industry (GVI MKIK) examined employer views based on the importance of expected competencies. One of the main findings of this research was that the demand for higher education graduates in the technical field is more significant than that of graduates in the field of economy. The research showed that the most important competencies appeared to be precision, self-reliance, high work endurance, ability to work in teams and ability to achieve goals. Our research focuses on understanding employer requirements related to the recruitment of graduates. We examined large companies, who have a conscious strategy for employing graduates. As previous research shows there is a difference on the demand side, therefore our aim was to investigate whether there exists a difference also in the competency requirements toward technical and economic higher education graduates.

1.1 The Higher Education and the World of Work

Considering the significant changes apparent at the labour market as well as the growing efforts aimed to improve the quality of 'graduate output', understanding labour market requirements has become an important field of study recently.

Szabó and Hámori point out that "at the information age, companies undergo constant transformation and reorganization, their strategy and organization are both flexible." [26] The trends experienced in economic transformation, globalisation, networking, the emergence of new technologies and the changes in consumer habits generate rearrangements in the form and content framework of the world of work as well. Researchers conducting career tracking surveys on graduates in the European region formulate these changes the following way: "While traditional organisations regarded employee care as a basic value, today's workplaces rather expect their employees to have a kind of responsible and self-caring attitude and make efforts in order that their employees identify with the organisation's values and goals."[9] In the 'looser' relations of work performance, the emphasis is on task performance, which requires of employees greater independence, fast response, pro-activity and flexible adaptation and significantly "influences the value and contents of the competencies necessary". [13]

In the system of changing environmental factors, companies re-interpret their recruitment and training strategies as well. There is no clear formulation of employer requirements of employees, either, making up for which shortcoming can be regarded as a task which is a key to future success. In the technical

literature on management, McClelland examines, based on his definition of competency, the relationship between workforce proficiency and employer requirements by researching into the criteria ensuring successful work performance. [18] In a labour market undergoing a transformation as an effect of economic and social changes, the interpretation of success criteria can be interpreted as a relationship dynamically changing both on employers' and employees' sides. [16] Employers constantly look for the personal abilities and knowledge that ensure the implementation of their strategic plans both in the short and long run. [4] On the other hand, the 'competency set' of an organisation, which determines the company's performance, is also influenced by the employees' proficiency, motivations and competency mismatches. [1] Accordingly, the practice of competency management is based on the broad-scale and regular communication of the company's management, HR specialists and employees, in the course of which the specification of recruitment criteria and training goals and the elaboration of employee assessment factors are based on the formulation of current and future employer competency requirements that support serving the strategic goals.

Some employees find it easier, others more difficult to adapt to the requirements that are new or under transition. *"Technological change generates a conflict between the demand for and the supply of competencies.*" [9] The salary differences emerging between high and low skilled employees draw attention to the fact that "the demand for high skilled workforce has emerged in the economy. Inequalities in remuneration are rooted in the fact that the supply side is unable to keep pace with the needs of the demand side." [9] Cedefop's European skills and jobs (ESJ) survey pointed out that "skill mismatch" might be one of the reasons why employers are unable to fill vacancies, despite the high level of unemployment in certain areas. [3] Skill mismatch is "the gap between the skills required on the job and those possessed by individuals". [3] This gap may be caused partly because of the increase in company expectations or changes in job requirements observed in the labour market. [10] On the other hand, the gap between required and possessed competencies may have its roots in imperfect information about job opportunities, but also in individual educational and skill shortage or underlying problems of higher education. [17] The role of higher educational institutions is essential, as it is shown in research conducted by Hungarian institutions, there is a gap between theoretical oriented practice of universities and the practice oriented requirements of the labour market. [8]

This process generates new tasks for institutions preparing graduates. Based on the educational implications of EU 2020 strategy [6], the initiative "*Agenda for new skills and jobs"* aims to increase the match between the labour market supply and demand side by supporting life-long learning and skills development. [19] Interpreting education as a value creating process, its task is to 'deliver' the competencies required by society and the labour market. "It is increasingly through this that the content requirements and achievements of studies are

monitored and developed in higher education." [15] The concept of competency was used in pedagogy in a different interpretation before (discussed in Tóthné Téglás [31]), but there is a shift from this interpretation of competence towards interpretation used management practice in order to understand better the expectations of labour market. In this context the process of learning is approached from the direction of the achievements attained [11] [29] [30] [12], and various learning and carrier building opportunities are available for students. [14] [21] In such an *output approach, institutions make efforts to learn about labour market requirements following a pull strategy* and aim to train graduates employable in the labour market. [2]

Correspondence between the level and field of studies and the requirements of a position can be examined by identifying under and overqualified areas, assessing learning skills and measuring to what extent students are prepared for future tasks as well as the performance of indeterminable, innovative tasks. [28] From the beginning of the 2000's the international graduate career tracking system research (CHEERS, REFLEX, HEGESCO) focused on the identification of general competencies essential in the rapidly changing environment and the contribution of higher education. In those research projects the role of competencies, too, was examined among factors determining professional success. Even though no factors were found to have outstanding effects, it could be concluded as regards to competencies, that they primarily influenced the subjective elements of work. [22] In the international graduate career tracking system research, the examination of workplace requirements measures the necessary and the acquired levels of competency based on fresh graduate's opinions several years after their entry to work. As the further development of previous research, the Tuning project carried out by cooperating European universities "initiated a series of consultations at the European level with employers, higher education graduates and university lecturers participating in order to determine the competencies considered to be of key importance for a higher education degree." [16] The main finding of this research was that there were distinctive differences between graduate's and employer's responses although the correlation between the competencies considered most important was high ($r = 0.9$).

The role of competency assessment in higher education research is to implement education better adjusted to labour market needs. Differences in the interpretation of competency in pedagogy and management made the compilation of competency lists applied in competency research difficult. As an effect of employer requirements, both cognitive and non-cognitive elements determining success at work have been included in the range of competencies assessed in the research. [16] A uniform competency model generally accepted in the respective fields is not available either in higher education research or in management; several approaches can be applied. [31] [23] The criteria for outstanding performance in a given position can be investigated by examining the characteristic features of the people filling the positions or by considering the

skills and knowledge required to fulfil the tasks. These correspond to what are referred to as the income and outcome approaches to competencies. [32] [27] The competency models can differ in what components they consider important. Applying the income approach, American models emphasise personal and social skills, while in the British approach using the outcome approach of competencies, the role of functional skills is underlined. The models worked out in the course of developing German technical training complemented personal and social competencies by professional and methodological/learning competences. [32] Accordingly, various competency lists can be found in the research that makes it more difficult to compare results.

1.2 Hungarian Research Findings

Labour market requirements of graduates can be examined on the basis of employer views or the experience of fresh graduate entrants. In what follows, we summarize the findings of research conducted in Hungary. The Institute for Economic and Enterprise Research of the Hungarian Chamber of Commerce and Industry (GVI MKIK) set out to survey the current situation and changes of the demand side, i.e. the labour market, within the framework of which they have made several surveys of the characteristics of the current and future employment of graduates among Hungarian enterprises since 2005. *Research conducted by GVI MKIK in 2008 and 2011 (herein after GVI 2008 and GVI 2011) examined employer views also on the basis of experience from recruitment process and the importance of expected competencies*. [33]

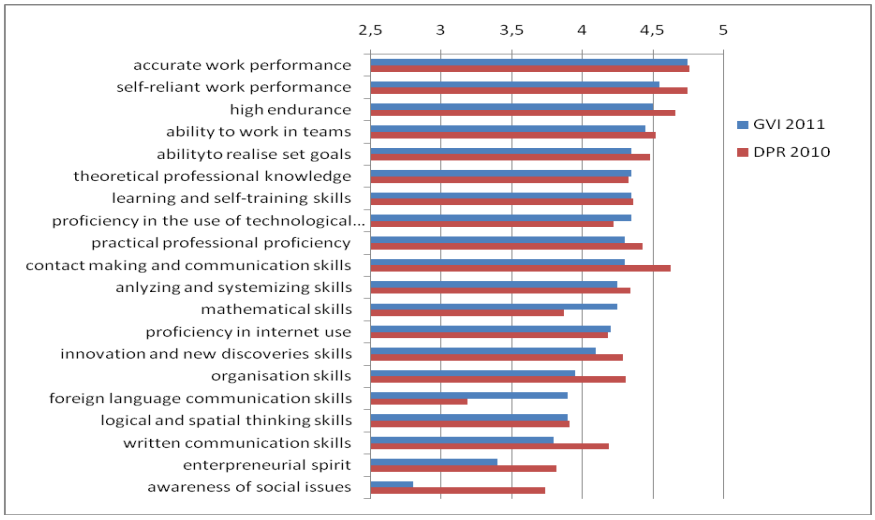


Figure 1

Competency requirements as perceived by graduates with a higher education degree and employers' actual competency requirements based on research by GVI 2011 and DPR2010 [16] [33]

Considering research focusing on issues related to the quality of fresh graduate output, national graduate career tracking system research (hereinafter DPR) conducted by Educatio Social Services Nonprofit Ltd. analyses the opinion and the workplace experience of persons who graduated a few years ago and have already entered the labour market. In their research programme, they primarily examine the roles of socio-demographic, motivational, institutional and field-specific features in the process of transition from higher education to the world of labour. Building on experience from the international Tuning-project, the comprehensive graduate career tracking system research conducted in Hungary in 2010 (hereinafter DPR 2010) complemented the study of workplace adjustment, in addition to examining qualifications, by examining the competencies required for the scopes of work, with the aim to measure adjustment in a more complex way. [16] *Thus, the DPR 2010 research surveyed the requirements perceived by employees on the basis of the workplace experience of graduates of various fields, ranking the importance of the respective competencies as experienced in their work.* In the next section we summarize the finding of this research (DPR 2010, GVI 2011) because since then this research has not measured competencies.

The DPR 2010 domestic research conducted with a high number of subjects (N=4511), surveying the opinions of graduates who had graduated three years before and had already entered the labour market, and the employer surveys performed by GVI MKIK in 2008 (N=3200) and repeated in 2011 (N=1313) give an opportunity for the comparison of the views of the two sides. [16] [33]

The research findings recorded by GVI in 2011 revealed a radically different order of preferences compared to three years before. Examining the values on the same scale as the one used for the categories in DPR 2010, research participants evaluated the importance of competencies along a much wider span than before (the lowest mean value being 2.8 and the highest 4.75), which may indicate that they perceived preferences more precisely. *Requirements having become clearer is also reflected* by the fact that the corporate preferences shown in 2011 reflected a much similar picture as the views of graduates interviewed in 2010 (DPR 2010). When arranging data into quartiles, *the most important competencies appear to be precision, self-reliance, high work endurance, ability to work in teams and ability to achieve goals* (Figure 1). The second quartile includes competencies related to professional proficiency, practical experience and learning. Interestingly, *innovation skills, entrepreneurial spirit and foreign language skills, which could be key factors amidst the changing economic requirements to future success, were ranked among less important competencies*. Considering this it can be assumed that the emergence of clear requirements can in fact be attributed to a temporary, but strong appearance of self-sustaining functions related to the economic crisis. It is with language skills that the opinions of employers and graduates differ to the greatest extent; graduates regard it even less decisive for their work than employers. At the same time, awareness of social issues, entrepreneurial spirit and written communication skills were competencies regarded less important by

employers, while graduates ranked these requirements higher than the actual requirements placed upon them.

In the DPR 2010 research, the competency categories of the Tuning project were applied for the sake of comparability. On the basis of a person-oriented income approach, the thus applied categories formulate basic skills deducible from the original arrangement of the competencies. In the course of the analysis examining the co-occurrences of young graduates' responses revealed by factor analysis, the 20 competencies used in the research, displaying a relatively clear structure, were arranged in four dimensions accounting for 49% of the variance on the whole. [16] Abilities and skills related to self-reliant work performance and contact maintenance were assigned in the *Cooperation* factor. The second, *Instrumental-cognitive* factor includes the use of technical devices and basic cognitive competences (writing, reading, computation) and, even though assigned medium weight, the knowledge of foreign languages was also put into this category according to respondents' opinions. Entrepreneurial spirit and openness towards innovation were categorised under the third factor, *Social initiatives*. The fourth factor of *Professionalism* includes theoretical and practical professional proficiency and the ability to develop these.

As a way of summary, the domestic DPR 2010 research outlined above explored the competency requirements perceived by graduates, while in the analyses performed by GVI, the ranking of importance of requirements placed upon fresh graduates was made on the basis of corporate opinions. In our own research, we also examined corporate requirements of graduates in the case of technical and economic employees. At the time the data were recorded for our research, the research findings of GVI 2011 were not yet known so we were unable to use the experience; therefore, we used the research findings retrospectively for our comparative analysis.

2 Corporate Survey

2.1 Objective and Methodology of the Survey

In our survey we examined the opinion of the demand side through a relatively narrow corporate group, also narrowing the field included in the survey. We examined the workplace requirements of primarily large companies actively recruiting fresh graduates from technical and economic fields. Our research findings created opportunity for examining new aspects: the differences according to corporate size and fields of specialisation. We focused on the HR representatives who participate in job fairs, because we believe that they have clear ideas about requirements for fresh graduates to meet, and have the appropriate insight into corporate requirements of graduates. It can be assumed

that companies exhibiting at a job fair have a conscious concept on the recruitment of graduates, have requirements in place and await fresh graduates with adequate recruitment and training programmes. For this reason, we believe that *we managed to reach an important target group from the point of view of our research since we were able to address companies actively looking for graduates and explore their views*.

We narrowed down our focus to two segments of graduates namely technical and economic fields, because according to KSH in 2012 most of the tertiary graduates belong in these two segments (38%). They reasonably face and in future will have to face more competition while seeking for job. Our analysis of the arrangement of competency requirements delivered findings that prove interesting especially in comparison with the findings of the DPR 2010 research.

Our primary research was conducted at the Faculty of Business and Management of Óbuda University (hereinafter OE). The aim of our research was *to explore workplace requirements of economic and technical graduates*. In the OE survey, we interviewed the HR representatives of companies at 4 job fairs held in the Central Hungarian region (in Kecskemét, Gödöllő and 2 in Budapest: organised at Óbuda University and by HVG Publisher respectively) with the help of interviewers[1]. Considering that there was convenient sampling, our findings cannot be regarded as representative considering company size or ownership structure in Hungary. On the other hand, we believe we managed to reach an important segment in term of our research (HR representatives participating in job fairs, who have clear requirements for fresh graduates).

In our research we sought responses to the following questions:

1) What requirements do companies actively looking for graduates have of graduates? What are competencies of outstanding and of lesser importance?
2) How are competency requirements arranged in the respondents' ways of thinking?
3) Do corporate requirements of technical and economic graduates differ?

In the first part of the questionnaire, we asked about the characteristics of the respondents' companies and the recruitment procedures applied by them. Based on the findings of the GVI 2011 research, companies attribute several advantages to employing graduates. According to two thirds of the companies participating in their research, young people bring new spirit and new ideas to the company, they are easy to shape, they become most valuable work force in the medium run already and take on more workload. [32] We therefore regarded it as important to first of all examine for what positions the companies exhibiting at job fairs look

[1] Students Enikő Kardos and Lajos Barcza performed the interviews within the framework of their activities in the Scientific Student Association.

for graduates. We also wanted to know *to what extent companies are conscious in shaping their requirements in the recruitment process of graduates* and whether they used competency-based recruitment tools. According to research conducted in Northern Hungary, the practice of large multinational companies is characterised by conscious and planned training activity. [20] Considering that corporate training and recruitment are systems closely built upon one another, it could be assumed that the organisations of respondents, most of whom belonged to corporate circles, used consciously designed and shaped procedures for determining recruitment criteria and implementing the recruitment procedure.

Based on the list of competencies presented in the second part of the questionnaire, we asked respondents to mark the importance of competencies expected with reference to the positions concerned on a 7 point Likert scale. The competencies used in the research were formulated according to the categories generally used in education. We designed our competency list based on Sonntag & Schaffer-Rauser's model [7]. This competency model includes professional, methodological, social and personal skills as well, so, similar to DPR research, *it examines cognitive and non-cognitive competencies alike*. Based on the competency categories of this model, we finalised the list by consulting researcher and experts of this topic at the university. We add new competencies compared to the competency lists used in earlier research, which we judged could be important in scopes of work performed by higher education graduates. These included among others ethical behaviour and demand for multidisciplinary knowledge.

In the questionnaire we asked respondents to evaluate separately requirements towards graduates having technical and economic qualifications in order to understand the differences in the competencies required. Our original sample of respondents was 98 companies, but owing to deficiencies in filling in some of the questionnaires, the final number of questionnaires that could be processed was 88. Considering that respondents could make the competency evaluation with reference to the positions to fill, we were able to use 111 work scope evaluations when analysing the competency evaluations, of which 73 referred to positions requiring technical and 38 requiring economic qualifications. The data were analysed with the help of the SPSS 19 programme package.

2.2 Characteristics of the Companies Participating in the Survey

We aimed to address companies actively looking for graduates. 78% of the companies participating in the research *liked to employ graduates*, out of which 48% said that they very much liked to recruit graduates. Companies do not only look for trainees at job fairs; *92%* of the responding corporations *employed* graduates *full-time*. At the same time, *internship and trainee programmes* are popular as well: 53% of the companies interviewed advertised such positions. These positions offered excellent opportunities to graduates and *assumed*, on the

companies' part, a conscious and *planned workforce asset training programme* by which the company prepared for its future workforce demand by recruiting talented young people. 18% of the respondents looked for graduates for temporary employment, which could be either genuine temporary demand or often served as a trial period in the course of which the workplace management were able to survey the young people's abilities and their attitude to work in the practice.

As regards their fields of activity, the companies in the sample having a seat in Hungary operate in a wide range of fields. As regards size, *it was primarily large companies that were included in the sample*, which can partly be explained by their presence at the job fair. In *83%* of the companies interviewed, the number of employees was *above 250 people*; in the case of 14%, it was between 50-250 people and in the case of 3% it was under 50. Considering ownership relations, *75% of the sample were foreign-owned, 9 % were of mixed ownership* and 15% Hungarian-owned.

The sample included companies of larger sizes the majority of which were foreign owned or had mixed ownership, which assumingly had a relatively conscious recruitment strategy. Findings only partly confirmed the latter assumption. The *recruitment practice of the companies interviewed is based on a competency-based concept only in the case of less than half of the companies in the sample*. In the practice of only 38% of the companies interviewed do they use uniform competency profiles for scopes of work or the company as a whole; 55% of them conduct competency-based interviews and only 31% recruit graduate employees by using an Assessment Centre. In the recruitment of graduates, the most decisive recruitment tool applied is a job interview with the manager (84%); 58% of the companies have applicants also fill in an aptitude test, while only 8% of the companies require that written technical papers be submitted.

3 The Importance of Required Competencies

The HR representatives participating in the survey evaluated the importance of the competencies listed with reference to the advertised positions on a 7 point Likert scale. Table 1 shows the competencies in order of their importance perceived by companies actively looking for graduates. Mean scores range between values of 6.39 and 3.47. (Table 1) In our analysis, we first present the results calculated for the total sample, followed by the presentation of the differences between the two fields of study.

The most important competency requirement among participating HR representatives was problem-solving skill (M=6.39, SD=0.755). This is followed by self-reliant and precise work performance (M=6.38, SD=0.825), cooperation with others (M=6.23, SD=0.835), information acquisition and processing skills (M=6.06, SD=0.920) and foreign language proficiency (M=6.01, SD=1.404), and

finally, ethical behaviour (M=0.6, SD=1.206). Competencies ranked in the first quartile according to importance serve *current performance* primarily. Companies actively looking for graduates expressed that most important competency they expected from graduates were self-reliant and efficient work performance 'from the very first moment'. These competencies closely related to work performance that determines the success of their recruitment.

The further competencies in the order of importance judged as important by companies, i.e. commitment to further continuous learning, openness to changes, flexibility, self-confidence and inner motivation, workload tolerance and endurance, and oral communication skills can still be categorised as competencies fostering *independent work performance*. However, this second quartile already includes competencies that may serve future work performance. These are: creativity, openness and flexibility, and information acquisition and processing skills and the ability to manage changes. These competencies were less highly ranked in previous research and were often in the less important part of the list. This may suggest that in the sample including large companies primarily, companies are further-sighted and their preferences include, in addition to the criteria for present success, success factors for the future as well. Therefore, they formulate much more complex requirements of higher education graduates.

According to the findings of the survey, companies ranked theoretical and practical professional proficiency in the third quartile. This result could suggest that there is less emphasis on professional proficiencies, but in fact these requirements, too, had very high scores. Theoretical knowledge was ranked 5.65 and applied practical proficiency 5.28 on the 7 point scale. Thus, in fact results do not reflect a 'depreciation' of professional proficiency but a growing role of soft skills related to other work performance. In complex scopes of work requiring a higher education degree, successful work performance is not guaranteed by high professional proficiency in itself. Without personal and social competencies, professional tasks cannot be successfully accomplished.

Among less important competencies, but still in the third quartile we also find IT skills, tolerance, strategic thinking as well as decision making and risk taking skills. The companies interviewed regard management and leadership skills and multidisciplinary knowledge as the least important among the requirements placed on graduates. The *competencies ranked less important are primarily used in the performance of duties related to managerial duties,* which is proven by the other competencies at the bottom of the ranking whose ranking is lower even in absolute terms, like negotiation techniques, high-level writing skills and influencing and supervisory skills. On the basis of the evaluation monotony tolerance does not seem to be relevant for the recruitment of graduates, either.

Table 1

The importance of respective competencies based on the OE research

(Scale of 7; 1- not important at all, 7 – extremely important; mean, standard deviation, N=111)

Source: authors' calculations

	economic qualification		technical qualification		Total	
	M	SD	M	SD	M	SD
Problem solving skills	6,38	,721	6,40	,777	6,39	,755
Self-reliant, precise work performance; initiative approach	6,65	,484	6,24	,927	6,38	,825
Cooperation with others, working in group	6,50	,655	6,10	,885	6,23	,835
Information acquisition and processing skills	6,36	,798	5,92	,946	6,06	,920
High proficiency in at least one foreign language	6,14	1,475	5,95	1,373	6,01	1,404
Commitment to ethical action	6,22	1,134	5,89	1,234	6,00	1,206
Commitment to further continuous learning	5,81	1,175	6,08	1,057	5,99	1,100
Openness to changes, flexibility	6,16	,764	5,89	1,061	5,98	,977
Self-confidence and inner motivation	6,14	1,110	5,88	1,079	5,96	1,091
Responsibility taking skills	6,08	1,064	5,88	1,260	5,95	1,197
Creative thinking skills	5,68	1,292	6,03	1,093	5,91	1,170
Workload tolerance and endurance	6,05	,998	5,81	1,002	5,89	1,003
Oral communication skills	6,35	,824	5,50	1,252	5,78	1,194
Ability of manage conflict	5,97	,910	5,68	1,035	5,77	1,001
Ability of manage changes	5,78	,976	5,63	1,173	5,68	1,108
High professional proficiency, up-to-date theoretical knowledge	5,14	1,291	5,90	1,157	5,65	1,250
Proficiency in the use of modern IT	5,20	1,389	5,48	1,237	5,39	1,289
Awareness of the practical methods applied	4,89	1,326	5,47	1,311	5,28	1,338
Tolerance of other (various) views	5,44	1,340	5,08	1,372	5,20	1,367
Decision-making skills	5,32	1,292	4,99	1,494	5,10	1,433
Strategic thinking skills	4,70	1,288	4,89	1,309	4,83	1,299
Risk-taking skills	4,78	1,377	4,68	1,433	4,72	1,409
Presentation skills	5,00	1,309	4,54	1,426	4,69	1,400
High-level writing skills	5,11	1,308	4,41	1,489	4,65	1,463
Application of learning techniques and strategies	4,64	1,376	4,65	1,385	4,64	1,375
Tolerance of monotony	4,86	1,549	4,46	1,357	4,59	1,429
Application of negotiation techniques	4,69	1,618	4,11	1,505	4,30	1,561

Ability to influence and supervise others	4,22	1,530	3,75	1,489	3,91	1,512
Management and leadership skills	4,03	1,647	3,44	1,500	3,64	1,568
Proficiency in several fields of science	3,33	1,454	3,54	1,452	3,47	1,449

Considering the findings, *there are no clearly formulated competency requirements related to learning*. The application of learning techniques and strategies is ranked 25th in the list, which is even more surprising given the fact that commitment to continuous learning was ranked 7th in the order of preferences. As regards competencies newly included in our competency list, *ethical behaviour* had a high ranking, while the survey did not confirm any demand for multidisciplinary knowledge.

On the basis of the research findings, there seem to emerge some remarkable differences between competency requirements for technical and economic positions. As the highlighted figures of Table 1 show *there are higher requirements as regard to oral and written communication as well as management skills for economic positions as compared to technical jobs. At the same time, technical fields pose higher requirements in professional and practical proficiency and creative thinking*. Self-reliance, endurance, cooperation skills, information management, decision making and negotiation skills were ranked somewhat higher for economic positions. Strategic thinking and IT skills, on the other hand, proved somewhat more important for technical fields. Learning competences have about the same place in the ranking of the two fields, while the mean values are a little higher for technical graduates.

4 Structure of Competencies

The 30 items of competencies were subjected to principal component analysis (PCA) using SPSS. Prior to performing PCA the suitability of data for factor analysis was assessed. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy value was 0.73, and the Barlett's test of Sphericity reached statistical significance, supporting the factorability of the correlation matrix. Using Kraiser's criterion principal component analysis revealed the presence of seven components with eigenvalues exceeding 1, explaining 67% of total variance. To aid in the interpretation of these components, Varimax rotation was performed. The rotated solution is presented in Table 2.

Table 2
Competency factors based on the OE research
Source: drafted by the authors

Competency components	**Competencies**	**Factor weight**
Adaptation competencies	Workload tolerance, endurance	.781
	Commitment to ethical action	.765
	Responsibility taking skills	.749
	Ability of manage conflict	.709
	Risk taking skills	.678
	Self-confidence, inner motivation	.663
	Ability of manage changes	.573
Communication skills	Application of negotiation techniques	.811
	Presentation skills	.775
	Management and leadership skills	.760
	Ability to influence and supervise others	.605
	Oral communication skills	.538
	Proficiency in several fields of science	.534
	High-level writing skills	.518
Openness	Openness to changes, flexibility	.690
	Creative thinking skills	.687
	Commitment to further continuous learning	.670
	Tolerance of other (various) views	.636
	Application of learning techniques and strategies	.557
Problem and task solving	Use of modern IT	.704
	Problem solving skills	.689
	Information acquisition and processing skills	.656
	Self-reliant, precise work performance; initiative approach	.493
Professional proficiency	Awareness of the practical methods applied	.825
	High professional proficiency, up-to-date technical knowledge	.806
Cooperation	High proficiency in at least one foreign language	.849
	Tolerance of monotony	.661
	Cooperation with others, working in group	.520
Decision-making	Decision-making skills	.716
	Strategic thinking skills	.475

The first component, which can be regarded as the competency group of A*daptation competencies* includes workload tolerance, ethical behaviour, responsibility taking, conflict management, risk taking, self-confidence and ability

to manage changes. This component could be interpreted as a requirement of ability to adapt to a rapidly changing environment, and to integrate into the organization. The second component consists of competencies that represent various *Communication skills,* since this component includes items as negotiation techniques, presentation skills, management and leadership skills, ability to influence and supervise others, oral and written communication skills and proficiency in several fields of science. Openness to change, flexibility, tolerance, creative thinking and learning competences are placed in the third component, which we termed the component of *Openness*. The fourth component includes competencies determining *Problem-solving ability*, like problem-solving skills, information management skills, use of IT, self-reliant and precise work performance and initiative approach. The fifth component is made up by just two competencies, high-level professional proficiency and awareness of the practical methods applied. We called this component *Professional proficiency*. Unlike in different research findings, cooperation with others was ranked only to the sixth component. Although proficiency in foreign languages and tolerance of monotony "fit" this category to a lesser extent, considering the importance of social competency we termed this factor *Cooperation*. The seventh component can be described as a requirement for *Decision-making competencies*, and this is associated with the competencies of decision-making skills and strategic thinking.

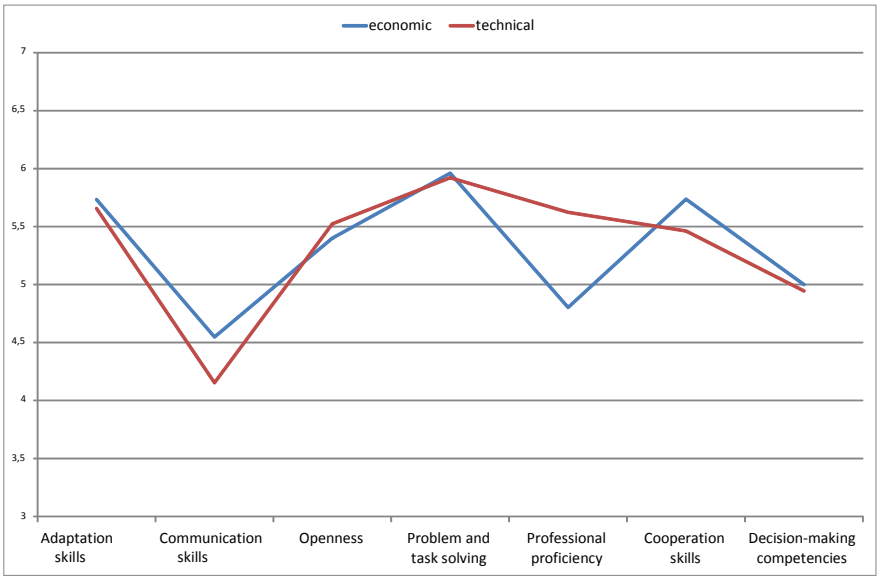


Figure 2

Means of indices calculated on the basis of the competency factors in the case of economic and technical graduates, based on the OE research

Source: drafted by the authors

In the interpretation of cooperation and learning competences we attained interesting new findings compared to previous research. Cooperation and competencies serving self-reliant work performance are usually interpreted as competences built upon one another in groupings, while in our research the HR representatives of the companies interviewed *interpreted cooperation separately from competencies determining individual performance*. In our sample, learning competences were clearly arranged in the factor of openness and creativity. This may imply that *the companies interviewed think in broader terms about learning competences, which may also refer to the method and content of learning*. Learning is thus evaluated as a tool to future adaptation especially by the HR representatives of large companies, in contrast to previous research findings where respondents clearly relate learning to professional competencies.

In order to analyse the differences related to positions requiring technical or economic qualifications, we created new variables (competency factor index) by adding scores of items belonging to a given factor and then we calculated an average. The averages of the competency factor indexes are shown in Figure 2. The ANOVA analysis and the nonparametric test (Mann-Whitney U test) performed revealed that in the case of five factors out of seven the expectations towards graduates were very similar; however, there were clear differences as regards leadership skills ($F=5.610$; $p=0.02$) and professional proficiency ($F=6.542$; $p=0.012$). *In the companies' evaluations, the competencies of communication skills proved more important for economic graduates, while professional proficiency was judged more important in the technical field*[2].

Although general requirements and those related to specific scopes of work are difficult to separate, *the differences* apparent from the findings *may also suggest that thinking in terms of general competencies is reflected in different ways in the case of different professions*. For an employee starting an engineer's career, professional proficiency and awareness of practical solutions are already essential at the time of entry so that he can join the work of an engineers' team; this basic proficiency to be acquired at school is even necessary for the self-reliant performance of subtasks. In the case of scopes of work requiring economic qualifications it holds true certainly in the beginning and probably later as well that the role of soft skills is more decisive for the successful performance of professional tasks. Economic graduates are more likely to perform tasks requiring complex communication skills (negotiation, presentation, coordination, organization, supervision) and operational management tasks later on than their colleagues working in technical fields. We would have found it interesting to explore the differences in learning competences as well, but in our sample there proved to be no difference between the two fields as regards this competence.

[2] The variance analysis and nonparametric test performed on the original factor variables confirmed the same correlation.

Conclusions

In our analysis we examined the requirements placed on graduates in the world of work. Building on the data of secondary research, we examined workplace adjustment in the course of the transition from school to work by presenting our own research findings. We first compared the findings of research conducted in the two most important target groups affected: of GVI 2011 examining the requirements of the employer side and of DPR 2010 exploring the competency requirements perceived by graduates. In our research, we primarily examined corporate requirements placed on technical and economic graduates. In our sample we addressed companies actively looking for graduates, assuming that these companies had conscious requirements of entrants to the labour market.

Our results show that competencies serving *present performance* proved the most important for companies, which *implies that self-reliant and efficient work performance is expected of graduates 'from the first moment'*. In the second quartile of competencies we found different results compared to the GVI research. In the GVI corporate sample of mixed composition, professional competencies were listed in this quartile, while in our research based on a sample of primarily large companies, competencies relevant for the future. However, in line with previous research our survey confirmed that successful work performance is not guaranteed by high professional proficiency in itself; without personal or social competencies, professional tasks cannot be successfully performed. Examining the co-occurrence of competencies based on employer responses in our research it can be concluded that in the way of thinking of corporate representatives, requirements are arranged very differently from the way the opinions of graduates participating in the DPR2010 research were arranged. In the examination of differences between the requirements of technical and economic graduates, it was the comparison of group means per factor that generated interesting findings.

Similar results bring knowledge within a Czech context.[3] Research by the National Institute of Education of the CR focused on the requirements and expectations of employers. In the responses of employers operating in the tertiary sectors lightly prevailed answers preferring rather broad occupational skills than narrow and deep professional skills of tertiary education graduates. In all monitored educational categories employers in the tertiary sector attach the greatest importance to ability to take responsibility, willingness to learn, read and understand the job instructions and solve the problem. In view of the employers in the CR, proficiency in foreign languages represents an attribute of the smallest importance. [5] On the other hand, other research reports about that, "in comparison to other European countries the employers across the Czech business sector were the most likely highlight the importance of sector specific skills". [24] [25]

[3] Most of the studies carried out in the Czech Republic are mainly concerned with the chances and position of the graduates in the labour market or with the economic returns of education or with the assessment of adult competencies (e.g. research project PIAAC) [24]

As a continuation of our research, we believe it would be interesting to examine the effects of corporate size and fields of specialisation. Considering the difference apparent between the GVI research and our sample of large companies the question arises whether corporate size determines a different pattern of competency requirements. When surveying the competency needs of the corporate sector, requirements arising from the needs of the organisation or from the scope of duties are difficult to separate in the responses. Analyses per field of specialisation would make it possible to separate needs of various 'origin' and to map other effects. The further examination of different preferences in professional and leadership skills could reflect upon the different 'competency management' decisions in the respective fields or indicate different interpretations of the new requirements of the changing economy. The examination of recruitment criteria used in relation to graduates is not sufficient for providing information on competencies to become useful in the future, either. It would be interesting to examine how requirements placed on colleagues change later on, what learning periods they are provided and what new skills they are expected to have several years after their starting their jobs.

The differences between the factors explored in our research and those revealed in the DPR research provide an insight into the respondents' different thinking schemes, indicating that the ways of thinking of fresh graduates and corporate representatives in terms of the respective competency categories may be very different. The differences revealed in our research on requirements towards graduates of technical and economic specialization, may also imply that in the case of different professions, the ways of thinking of participants in terms of general competencies required for these different positions are reflected differently. Competency-based thinking aims at formulating common interpretations, thereby launching a continuous dialogue between players on issues of requirements, learning and compliance with requirements. For researchers at the same time, examining how the respective competencies are arranged in the players' views may be an important starting point for further research, since the differences revealed in the findings of research mapping workplace requirements may too root in different interpretations.

Revealing differences between various countries (e.g. post transformation) may be another interesting way to continue our research and could serve exploring the differences perceived between the requirements or may generate new results with respect to the potentially decisive roles of corporate size or the fields of specialisation.

References

[1] Amador, L. B., Nicolás Á. L., Vila L. E., Education and Competency Mismatches: Job Satisfaction Consequences for Workers, XVI Jornadas ASEPUMA – IV Encuentro International Rect@VolActas_16Issue1:105

[2] Ádler, J., Stocker, M., Competency-based, Output-oriented Education and

Ideal Employment, BCE Business Economy Institute for Competitiveness Researches, Budapest, 2012

[3] CEDEFOP, Matching Skills and Jobs in Europe, 2015, http://www.cedefop.europa.eu/en/publications-and-resources/publications/8088

[4] Díaz-Fernandez M., López-Cabrales A., Valle-Cabrera R., A Contingent Approach to the Role of Human Capital and Competencies in Firm Strategy. Business Research Quartely, 17, pp. 205-222, 2014

[5] Doležalová, G., The Needs of Employers and the Readiness of Graduates - Research in the Tertiary Sector, Praha, 2014, http://www.nuv.cz/uploads/Vzdelavani_a_TP/TER_Potreby_zamestnavatelu_pro_www.pdf

[6] EU 2020 Strategy, http://ec.europa.eu/eu2020

[7] Farkas, A., Comparison of Employers' and Student's Ratings of Importance of Expected Technical Skills, MEB 2008 – 6th International Conference on Management, Enterprise and Benchmarking, May 30-31, 2008, Budapest

[8] Forgács, K., Kaucsek, Gy., Simon, P., Assessment of the Competent Labour Force with Psychological Test and Script Analysis. Munkaügyi Szemle, 46/9, pp. 12-18, 2002

[9] García-Aracil, A., Rolf van der Velden, Competencies for Young European Higher Education Graduates: Labour Market Mismatches and Their Payoffs. High Educ 55: 219239, 2008

[10] Hajdú, M., Nyírő, Zs., Türei, G., Tóth I. J., Varga A., Data on Higher Education and Employment of Graduates. Facts and Relationships. MKIK GVI, 2015

[11] Halász, G., Előszó. In: Demeter, K. (eds.), Competency: Challenges and Interpretations, National Institute of Education, Budapest, 2006

[12] Kálmán O., Czető K., The Development of the Learning Outcomes Approach in Hungarian Higher Education, XIV. National Education Conference Debrecen, In: Buda A. (ed.), Training and Education - Practice and Science. 8, 2014

[13] Kiss, I., Répáczky, R., New Skills and Work Places: Competence-based Services in Order to Higher Employment, In: Csernovitz, A., Szegedi, E. (eds.), Labour market oriented higher education, TEMPUS, Budapest, 2012

[14] Kiss I., Higher Education Students Life Management, Career-Building Issues, Career Counseling III. 7-8. pp. 6-16, 2011

[15] Kiss, P., Introduction. In: Kiss, P. (eds.), Competency-based Assessment in Higher Education, Educatio, Budapest, 2010

[16] Kiss, P., Demand for Competencies of Graduates and Work Satisfaction, In: Garai, O., Horváth, T, Kiss, L., Szép, L., Veroszta, Zs. (eds.), Graduate

Survey IV, 2011, Educatio, Budapest, 2011

[17] Lazányi, K., Backgrounds of Skills Mismatches in the Labour Market, Status of Economic Students,. Munkaügyi Szemle 57:(3) pp. 50-62, 2013

[18] McClelland D. C., Testing for Competency Rather than for „Intelligence". American Psychologist, 28, 1-14, 1973

[19] New Skills for New Jobs, http://ec.europa.eu/social

[20] Polónyi, I., Interest of Employers in Training, In: Polónyi, I., Barizsné, H. E. (eds.), Adult's Training, Corporate Training, Debrecen University, Economic Faculty, 2004

[21] Sági M., Employment Strategies of Career Starter Graduates Pályakezdő diplomások munkába állási stratégiái. In: Garai O., Veroszta Zs., Frissdiplomások 2011, Educatio, 2012

[22] Schomburg, H., The Professional Success of Higher Education Graduates. European Journal of Education. 42 (1), 35-57

[23] Simon, P., Competencies Learnt in Formal and Informal Way, Prepared: „Personal factors of Hungarian economy" Researches, Budapest, 2006

[24] Straková, J., Vesely, A., V. (eds.), Conditions of Success at the Work and in Life, International Survey of Adult Skills OECD PIAAC, Praha 2013

[25] Employers' Perception of Graduate Employability. Analytical Report. Flash Euro barometer 304 – The Gallup Organization, 2010

[26] Szabó, K., Hámori, B, Information Economy. Akadémia, Budapest, 2006

[27] Szelestey, J., Theoretical Background of Competency Modelling, 2008 http://www.erg.bme.hu/szakkepzes/4felev/SelesteyKompetencia.pdf

[28] Teichler, U., „Employability" or „Professional Relevance": Changes in Relationships between Higher Education and the World of Work on the Way towards the European Higher Education Area. In: Teicler, U.: *Higher Education and the Word of Work*. Sense Publishers, Rotterdam, 2009

[29] Tóth P., The Role of Individual Differences in Learning. Acta Polytechnica Hungarica 11:4 pp. 183-197, 2014

[30] Tóth P., Maior E., Pogátsnik M. (ed.), Teacher and Student Competencies in Empirical Researches. Budapest: Ágoston Trefort Centre for Engineering Education, 2015

[31] Tóthné Téglás T., Competence from the View of Employers, Employee and High Education. Business Development in the 21st Century, Budapest, 2016

[32] Winterton, Delamara, What is Competency? Human Resource Development International, Vol. 8, March, No. I, 27-46, 2005

[33] Várhalmi, Z., Graduates in Competitive Sector – 2011, MKIK GVI, Budapest, 2012

Fingertip-based Real Time Tracking and Gesture Recognition for Natural User Interfaces

Georgiana Simion, Ciprian David, Vasile Gui, Cătălin-Daniel Căleanu

Politehnica University of Timişoara, Faculty of Electronics and Telecommunications, 2 Pârvan Av., 300223, Timişoara, Romania, {georgiana.simion, ciprian.david, vasile.gui, catalin.caleanu}@upt.ro

Abstract: The widespread deployment of Natural User Interface (NUI) systems in smart terminals such as phones, tablets and TV sets has heightened the need for robust multi-touch, speech or facial recognition solutions. Air gestures recognition represents one of the most appealing technologies in the field. This work proposes a fingertip-based approach for hand gesture recognition. The novelty in the proposed system is the tracking principle, where an improved version of the multi-scale mode filtering (MSMF) algorithm is used, and in the classification step, where the proposed set of geometric features provides high discriminative capabilities. Empirically, we conducted an experimental study involving hand gesture recognition – on gestures performed by multiple persons against a variety of backgrounds – in which our approach achieved a global recognition rate of 95.66%.

Keywords: fingertip detection; hand gesture recognition; image classification; image edge detection; object recognition

1 Introduction

The Human-Computer Interface (HCI) domain represents one of the most dynamic research fields. It started as CI (cards, tapes, character CRT) only to become more interactive (keyboards, joysticks, realistic rendering). Nowadays, a second revolutionary change in HCI awaits- "Natural User Interface (NUI)". Among the key factors that contribute to the NUI revolution one must mention the availability of appropriate hardware platforms (CPU, GPU), existing software frameworks and, perhaps the most important thing, the user's demand and acceptance. According to the NUI principles, the machines should be able to interpret and communicate close to the way people do, using multi-touch, gestures, speech and various other sensors [1].

The use of the hand as a NUI interfacing method has recently grown to be more and more popular. Currently, we use the hand as a direct input device, we have

touch screens in our pockets embedded in smart phones with no keypads, touristic information points based on touch screens. Now the ultimate frontier is to use hand gestures without a direct contact to interact with the surrounding devices. Computer vision-based techniques can offer such non-contact solutions.

Some of the cutting-edge technologies employed by these solutions are briefly presented in this research, of those, the most promising technologies are the 3D acquisition solutions based on either structured light approach [2-3] or the Time-of-Flight (TOF) principle [4-6]. One of the most important advantage of these approaches is simplifying the hand segmentation. The electrical-field-based 3D gesture controller recently patented by Microchip Technology Inc. offers a low power, precise, fast and robust hand position tracking with free-space gesture recognition. These solutions have applications in various fields from virtual reality to healthcare, and from vehicle interfaces to desktops and tablets.

Although promising, these technologies still have some drawbacks, e.g. for the 3D cameras, the cost is still prohibitive and for the e-filed solutions the functional range is relatively small.

In our proposed framework, the images are acquired using a 640 x 480 VGA notebook integrated webcam. The target application is a non-direct contact interface which can be used in an info-point. The hand is extracted using the background subtraction technique [7], which is based on a codebook representation. Fingertips are tracked using our previously developed algorithm based on MSMF. For the classification step a set of six gestures, easily performed by the user, is chosen. The suggested set of geometric features provides high discriminative capabilities. Empirically, we have conducted an experimental study involving hand gesture recognition – on gestures performed by multiple persons against various backgrounds – in which our approach has achieved a global recognition rate of 95.66%.

The remainder of this paper is organized as follows. Section II describes recent solutions in fingers and fingertips detection. Details regarding the fingertips tracker used within the context of our approach are presented in Section III. Section IV gives the gesture classification method. Section V shows the experimental results and Section VI concludes the paper.

2 Related Work

There are different ways to use hand gestures in order to build a NUI. In literature, one can find several articulated hand models like those employed by [3], [8-10], but the proposed model uses fingertips and the palm region. Fingers and fingertips have more simple and stable shapes and therefore can be easily detected and tracked more reliably than hands [11-16]. Thus, they are more suitable in

applications where hands are used for pointing [17], since occlusions are less frequent in such cases. A commonly used method to detect fingers and fingertips is the mathematical morphology and related solutions [18-20]. Some approaches use Hu invariant moments [21-22], others k-curvature algorithm [23], combined with template matching [11], [19]. Noise, scale, and hand direction have a significant impact on finding fingers and fingertips using skeleton [24] or dominant points [12] principles.

Color information [11-14] is also an important feature in finger tracking because it is used to detect the hand. Remarkably, skin hue has relatively low variation between different people and different races. In addition, it is invariant to moderate illumination changes. For objects that are achromatic or have low saturations, the hue information is unreliable and influenced by the illumination level. The methods based on skin color may fail under complicated or dynamic lighting although motion information, gradient or 3D based solutions [15-16], [25-27] alleviate some of the drawbacks.

Another key element in finger-based approaches is edge detection [28]. If edge detection was not so much affected by complex backgrounds or hand texture, the hand shape could be well defined. Yet, some improvement in the process of edges selection is expected because of the post-processing operation.

Using the detected edges or segmentation, the hand contour can be obtained. It is the case of the following approaches [25], [29-30]. Unfortunately, most of them were tested against uniform backgrounds and proved highly sensitive to the hand posture. For better results, the segmentation operation could be augmented by hand tracking and recognition. In [31] the hand contour is extracted using color information and motion information whereas the skin regions, which are in movement, are considered part of the hand region. The fingertips are extracted using high values of curvature.

The conclusion drawn in [19] is that gestures with “one” or “two” fingers are the most reliably recognized. Using two fingers or just one finger, a large set of gestures can be defined and recognized. For example, in the work [32] a robust method for defining dynamic single finger-based gesture is proposed. This shows that the set of meaningful gestures that can be defined is large enough to create a comprehensive gesture dictionary. However, the proposed dynamic gesture recognition algorithm does not exploit the full potential of the tracking system used. In [32] only one active finger gesture is used for recognition. In contrast, the present work proposes extending the usage of the tracking system to perform two or more finger-based gesture recognition. We also avoid the complex problems of model definition and elaborated classification scheme typically employed in gesture recognition [33] by designing a geometrical approach, which brings important improvements in the frame processing speed.

3 The Tracker Principle

This section contains a brief overview of the used finger tracking method [34]. The proposed approach offers significant improvements with respect to real-time capabilities in processing the foreground and environment variations (indoors, outdoors, dynamic background, illumination changes).

Based on the background subtraction technique from [7] the foreground objects are extracted, resulting in a considerable reduction of the processing data. A codebook representation of the background model is used to quantize the background model at each pixel location. The model is adaptable to illumination changes. The background model based on codebook representation allows for the fast update of the model, being easy to adapt when dealing with dynamic backgrounds.

In the next stage, the finger features are extracted. These finger features are line strips, obtained by the horizontal and vertical scanning of the segmented foreground. We use both horizontal and vertical scans in order to achieve rotation invariance of the hand tracking system.

Looking for similarities in length and thickness, and considering the centers of these segments, the algorithm identifies the fingers' positions. The slope of the line containing the identified strip centers is then computed. First, the strips are filtered using neighboring information, and afterwards are clustered in order to detect the fingers. The clustering method used in our work is the multi-scale mode filtering (MSMF) [35] algorithm, which in fact is our multi-scale version of the Mean Shift (MS) algorithm. In MSMF algorithm, the scale parameter varies during iterations, allowing a course to fine tuning approach. This strategy proved to be beneficial in avoiding false maxima. The final point reached is the maximum point of probability density and its location is our solution.

As an important improvement, the tracking algorithm presented in [34] was extended to detect also the palm of the hand. Basically, palm detection and tracking work mainly the same as the finger tracking algorithm with a difference in the scale of the extracted line. As noticeable next, palm detection is very important in our approach, as it allows defining a supplementary point that is used to design the final set of natural gestures to be recognized.

This approach [34] is based on sparse features and it is part of the recent direction in tracking, known as tracking by detection [36-37]. Using sparse features provides computational efficiency, robustness and has high accuracy. The clustering algorithm, the careful selection of the finger features and the shape information regarding the fingers enhance the robustness of this tracking method.

We demonstrated the robustness and the dynamical accuracy of the tracker against occlusion in [34].

4 Gesture Classification

The envisaged NUI uses six simple gestures, easily performed by any user. These gestures were carefully chosen, taking into consideration the anatomic features of the hand and also the envisaged touchless info-point application. Even if the number of the defined gestures is not that large, still they can be versatilely used for multiple actions. Using these gestures, a virtual mouse could be easily implemented. All these gestures imply the use of two fingers only.

The six gestures definitions are in the next section. The two fingertips based gestures newly introduced offer a convenient and natural way to manipulate the content of the aimed application. Contrasting, the three fingertips based gestures are more difficult to be performed and the number of gestures that can be defined is not superior due to the lack of accuracy in defining the geometric distances and angles.

The algorithm initially detects all fingertips, except the thumbs' fingertip, and the palm of the hand as shown in Figure 1. The thumb is not of interest in the proposed context of dynamic hand gesture recognition because it is not used to gesticulate. Besides, its relative position to the palm area is difficult to estimate.

Figure 1
The fingertip detection stage

Context awareness is also used in our approach: once two fingertips are detected, the user can start gesticulating. The user knows when to start performing a gesture (a start message is printed on the screen, as shown in Figure 2. The two fingertips must be detected in at least three consecutive frames in order to start the routine that classifies the gesture. The algorithm computes several distances and angles as detailed below.

Figure 2
The system is ready for interpreting gestures

The first distance, denoted by *dist1*, is the distance between the index fingertip and the center of the palm (1); the second distance, denoted by *dist2*, represents the distance between the middle fingertip and the center of the palm (2); and finally *dist3* represents the distance between the two fingertips (3).

$$dist1 = \sqrt{\left(x_{palm} - x_{index}\right)^2 + \left(y_{palm} - y_{index}\right)^2} \tag{1}$$

$$dist2 = \sqrt{\left(x_{palm} - x_{middle}\right)^2 + \left(y_{palm} - y_{middle}\right)^2} \tag{2}$$

$$dist3 = \sqrt{\left(x_{index} - x_{middle}\right)^2 + \left(y_{index} - y_{middle}\right)^2} \tag{3}$$

where (x_{palm}, y_{palm}) are the coordinates of the central point of the palm region and (x_{index}, y_{index}) are the coordinates of the central point of the index fingertip, while (x_{middle}, y_{middle}) are the coordinates of the central point of the middle fingertip.

The computed angles are: *alpha*, which is the angle between the two fingers (4); *beta*, the index angle with Ox axis (5) and *gamma*, the middle finger angle with Ox axis (6). See Figure 3 for details.

$$alpha = beta - gamma \tag{4}$$

$$beta = \arctan\frac{y_{palm} - y_{index}}{x_{palm} - x_{index}} \tag{5}$$

$$gamma = \arctan\frac{y_{palm} - y_{middle}}{x_{palm} - x_{middle}} \tag{6}$$

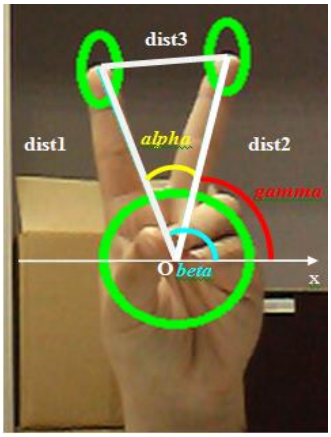


Figure 3
The computed distances and angles

Using these distances and angles the following gestures are defined: Select, clickLeft, clickRight, MoveLeft, MoveRight and DoubleClick. Each gesture is characterized by a set of parameters. These parameters are presented in Table I. It is important to observe that the distances *dist1, dist2*, *dist3* are normalized with the finger thickness, so they are not influenced by the distance between the user and the camera.

Table 1
Parameters of the defined gestures

Gesture	**The parameters describing the gesture**	**The values for the gestures' parameters**
Select	*alpha, dist3*	$alpha < a, dist3 <= \sigma$
clickLeft	*alpha, dist1,dist2*	$(dist2 < meanD2) \&\& ((meanAlpha-\tau) < alpha < (meanAlpha+\tau)) \&\& (dist1 >= meanD1)$
clickRight	*alpha, dist1,dist2*	$(dist1 < meanD1) \&\& ((meanAlpha-\tau) < alpha < (meanAlpha+\tau)) \&\& (dist2 >= meanD2)$
MoveLeft	*alpha, beta, gamma*	$(beta < (meanBeta-\tau)) \&\& (gamma < (meanGamma-\tau)) \&\& ((meanAlpha-\tau) < alpha < (meanAlpha+\tau))$
MoveRight	*alpha, beta, gamma*	$(beta > (meanBeta+\tau)) \&\& (gamma > (meanGamma+\tau)) \&\& ((meanAlpha-\tau) < alpha < (meanAlpha+\tau)$
DoubleClick	*dist1,dist2*	$(dist1 < meanD1) \&\& (dist2 < meanD2)$

In order to evaluate the temporal evolution of these distances, the mean distances *meanDi*, $i = 1..3$, are computed and later compared. This is done using only the frames associated with the performed gesture. In each new frame we compute the

meanDi as the sum between *disti* , i = 1..3, computed in all previous frames and the current frame, divided by the number of the current frame. The temporal evolution of angles *alpha*, *beta*, *gamma* is evaluated computing the mean values: *meanAlpha*, *meanBeta*, *meanGamma* similar to *meanDi*. The *end of gesture* occurs when the fingertips are no longer tracked. The value of angle *alpha* has a maximum range of 60º. The value for parameters a, σ and τ from Table 1 are experimentally determined.

The six gestures can be seen in Figure 4 a to f.

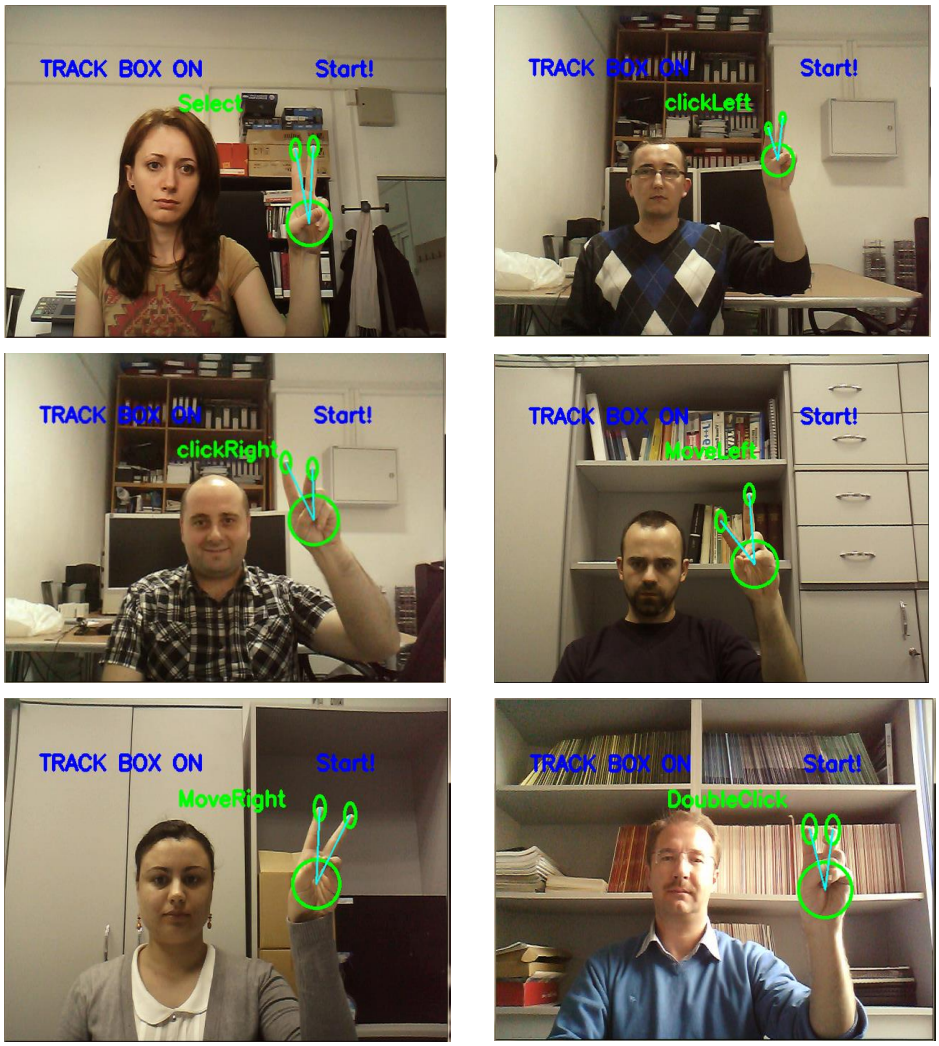


Figure 4
The defined gestures

5 Experimental Results

In order to test our method, ten subjects, both male and female, were asked to perform the six gestures. The subjects were quickly trained to perform the gestures. They had no difficulties in learning how to gesticulate correctly.

To test the robustness of our framework various environments were chosen. More exactly, the experiments were conducted in rooms with different illumination settings and complex backgrounds. In order to increase the difficulty of the recognition task, we deliberately chose to have skin-like objects in the background scene. Also the hand overlaps those objects. Considering the typical operating distance for an info-point application, the subjects were placed at a distance of 40 cm up to 120 cm from the camera. In all our experiments the value of parameter a was 25, the value of parameter σ was 2 and the value of parameter τ was equal to 10 degrees.

The proposed algorithm was implemented using OpenCV 2.4 library and compiled using Visual Studio 2010. The proposed solution ran smoothly using Windows 7 and a Dell Vostro 3500 notebook hardware platform based on Intel® Core™ i3 CPU, M370 @ 2.40 Ghz, 3.00 GB RAM. The images were acquired with the 640 x 480 VGA integrated webcam of the notebook.

The recognition rate for each gesture was experimentally determined (see Table 2) taking into consideration a total number of 50 image sequences (10 Individuals x 5 backgrounds) for each of the 6 gestures. The total recognition rate was about 95.66%, comparable with other previously reported results [12], [31].

Table 2
Average rate of our hand posture detection

Gesture	Recognition rate
Select	100%
clickLeft	100%
clickRight	100%
MoveLeft	92%
MoveRight	96%
DoubleClick	86%

The confusion matrix is shown in Table 3.

Table 3
The confusion matrix for the six classes

	Select	clickLeft	clickRight	MoveLeft	MoveRight	Double Click
Select	100	0	0	0	0	0
clickLeft	0	100	0	0	0	0
clickRight	0	0	100	0	0	0
MoveLeft	0	8	0	92	0	0
MoveRight	0	0	4	96	0	0
DoubleClik	8	4	2	0	0	86

The tests results performed in challenging illumination conditions are shown in Figure 5.

Another set of tests were conducted in order to check the robustness to occlusions and pose/perspective/distance to the camera variations. The results can be seen in Figures 6 and 7.

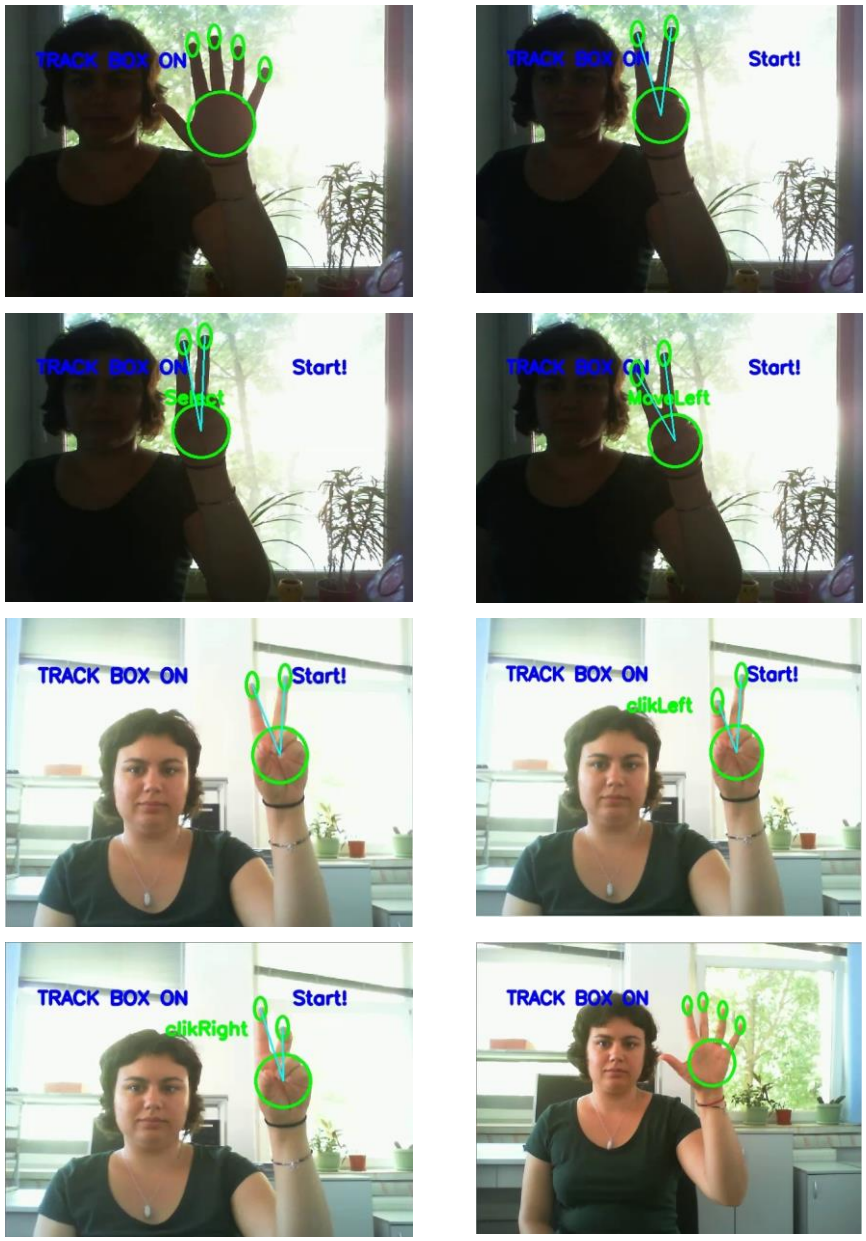


Figure 5
Experiments with different illumination conditions

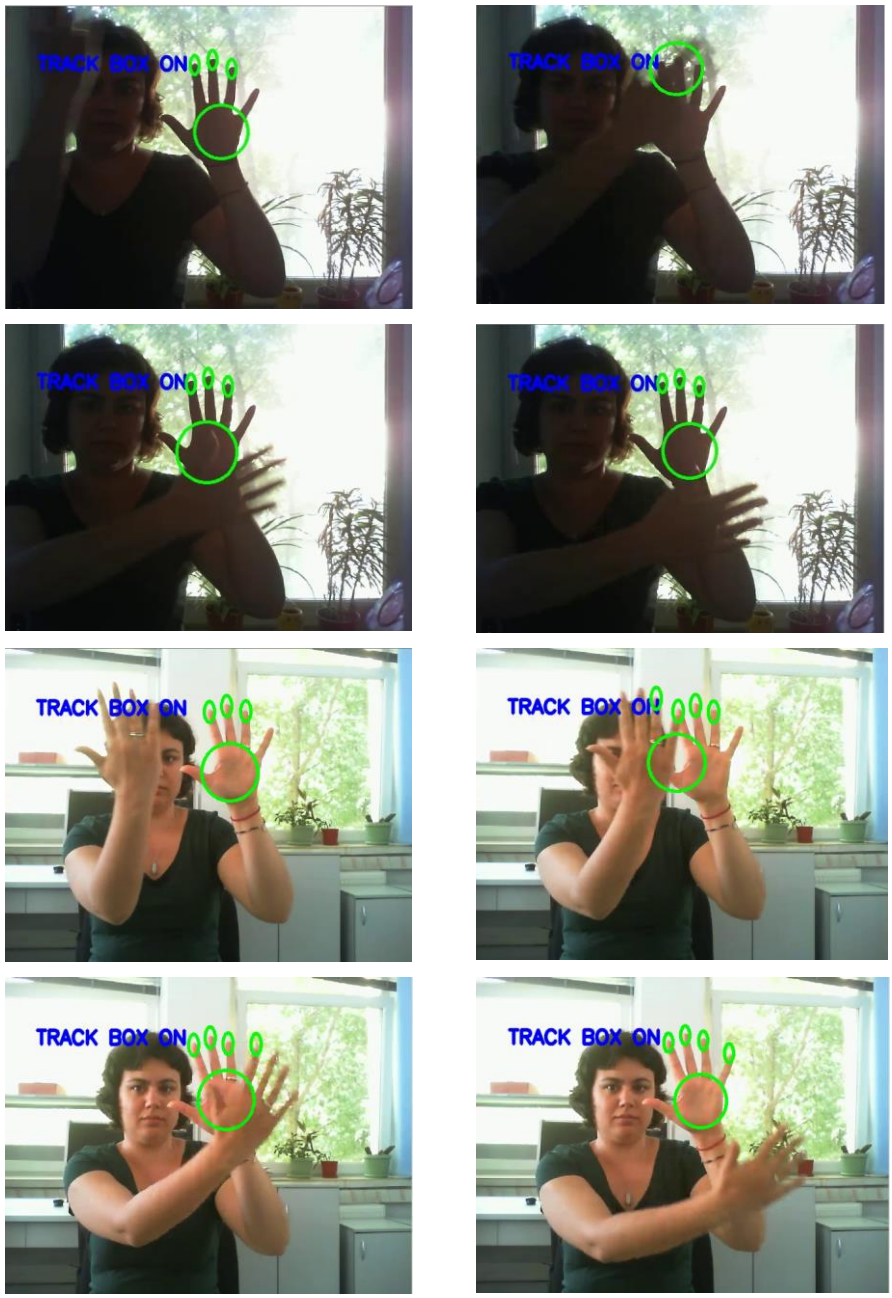


Figure 6
Experiments with occlusions in different illumination conditions

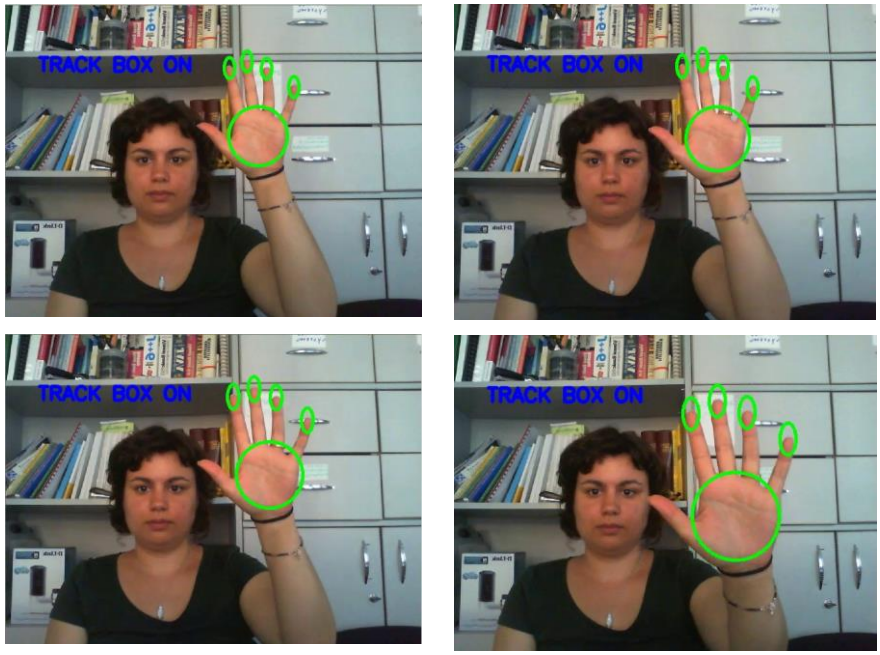


Figure 7
Experiments related to scale adaptation

Conclusion

This paper describes a novel vision-based real-time approach for recognizing basic two fingers gestures. Our method does not imply the acquisition of additional (usually more expensive) hardware, e.g., Leap Motion [38] or Kinect controller [39] nor wearable devices like in the case of Myo Armband [40] but employs a common 2D video camera integrated in most laptops or an inexpensive USB connected external web cam. Finger features are first extracted by looking for similarities in length and thickness. For this step, the MS algorithm with a multi-scale filtering capability is used. Then, the fingertips are also tracked by using the palm region of the hand. This enables us to define a supplementary point that is used to design the final set of natural gestures to be recognized. The main characteristics of the proposed framework for fingertips detection and tracking are robustness and high accuracy with reduced computational costs. For the classification step, the scale-invariant distances and angles are calculated. By analyzing the position of just two fingers, various interface actions, such as clicking and moving, are recognized. The system performances were analyzed further with respect to the real-time capability and the recognition rate. Experimentally, we have proven that the proposed approach provides a good classification rate (95.66% for a total of 300 test images) along with tight timing constraints (an average response time of 30 ms has been determined), which

enables humans to perceive interactivity. Thus the proposed architecture is suitable to support different types of applications, e.g. touchless info-point machine or smart TVs. As a future development, we estimate that the current framework could be easily extended to work within a 3D vision system.

References

[1] G. Simion, C. Caleanu, "A ToF 3D Database for Hand Gesture Recognition", 10th International Symposium on Electronics and Telecommunications (ISETC), Timisoara, Romania, 2012, pp. 363-366

[2] F. Malassiotis, N. Tsalakanidou,V. Mavridis, N. Giagourta, N. Grammalidis, M.G Strintzis, " A Face and Gesture Recognition System Based on an Active Stereo Sensor," In: Proceedings 2001 ICIP, Thessaloniki, Greece, 7-10 Oct. 2001, Vol. 3, pp. 955-958

[3] M. Bray, Koller-Meier, L. V Gool, "Smart Particle Filtering for 3D Hand Tracking", in Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Los Alamitos, CA, USA, 2004, pp. 675-680

[4] P. Breuer, C. Eckes, S Muller, "Hand Gesture Recognition with a Novel IR Time-of-Flight Range Camera - A Pilot Study", Proceedings of Mirage 2007, Computer Vision / Computer Graphics Collaboration Techniques and Applications, 2007, pp. 247-260

[5] E. Kollorz, J. Penne, J. Hornegger, A. Barke, "Gesture Recognition with a Time-of-Flight Camera,"Int. J. Intell. Syst. Technol. Appl., 5(3/4), pp. 334-343, 2008

[6] J. Molina, M. Escudero-Viñolo, A. Signoriello, M. Pardás, C. Ferrán, J. Bescós, F. Marqués, J. Martínez, "Real-Time User Independent Hand Gesture Recognition from Time-of-Flight Camera Video Using Static and Dynamic Models," Machine Vision and Applications, 2011, pp. 1-18

[7] K. Kim, T. H. Chalidabhongse, D. Harwood, L. Davis, "Real-Time Foreground–Background Segmentation Using Codebook Model," in Real-time Imaging, Vol. 11, No. 3, pp. 167-256, 2005

[8] M. de La Gorce, D. J. Fleet, N. Paragios, " Model-based 3D Hand Pose Estimation from Monocular Video" IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 33, No. 9, pp 1793-1805, 2011

[9] Y. Wu, Y. L. Lin, T. S. Huang, "Capturing Natural Hand Articulation," in: Eighth IEEE International Conference on Computer Vision, 2001, pp. 426-432

[10] B. Stenger, P. R. S. Mendonca, R. Cipolla, "Model-based 3D Tracking of an Articulated Hand" IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 310-315, 2001

[11] J. M. Kim, W. K. Lee, "Hand Shape Recognition Using Fingertips," 5th Int. Conf. Fuzzy Syst. Knowledge Discovery, 2008, pp. 44-48

[12] D. Lee, Y. Park, "Vision-based Remote Control System by Motion Detection and Open Finger Counting," IEEE Trans. Consumer Electron., Vol. 55, No. 4, pp. 2308-2313, 2009

[13] L. Bretzner, I. Laptev, T. Lindeberg, "Hand Gesture Recognition Using Multiscale Color Features, Hieracrchichal Models and Particle Filtering," in Proceedings of Int. Conf. on Automatic face and Gesture recognition, Washington D.C., 2002, pp. 423-428

[14] M. Donoser, H. Bischof, "Real Time Appearance-based Hand Tracking," Pattern Recognition, 2008, ICPR 2008, 19th International Conference on, Vol., No., pp. 1-4, 2008

[15] R. Belaroussi, M. Milgram, "A Real Time Fingers Detection by Symmetry Transform Using a Two Cameras System," LNCS, Vol. 5359, 2008, pp. 703-712

[16] O. Gallo, S. M. Arteaga, J. E. Davis, "A Camera-based Pointing Interface for Mobile Devices," IEEE Int. Conf. Image Processing, 2008, pp. 1420-1423

[17] L. Song, M. Takatsuka, "Real-Time 3D Finger Pointing for an Augmented Desk," In Australasian conference on User interface, Vol. 40, Newcastle, Australia, 2005, pp. 99-108

[18] C. von Hardenberg, F. Bérard, "Bare-Hand Human-Computer Interaction", Proceedings of the ACM workshop on Perceptive User Interfaces, Orlando, Florida, USA, 2001, pp. 1-8

[19] K. Oka, Y. Sato, H. Koike, "Real-Time Fingertip Tracking and Gesture Recognition", IEEE Computer Graphics and Applications, pp. 64-71, 2002

[20] S. Malik, C. McDonald, G. Roth, "Finger Detection via Blob Finding Flood fill: Hand Tracking for Interactive Pattern-based Augmented Reality" Proceedings of the In.ternational Symposium on Mixed and Augmented Reality (ISMAR'02), 2002

[21] P. Premaratne, Q. Nguyen, "Consumer Electronics Control System Based on Hand Gesture Moment Invariants," Computer Vision, IET, Vol. 1, No. 1, pp. 35-41, 2007

[22] L. Yun Liu, G. Zhijie, Yu Sun, "Static Hand Gesture Recognition and its Application based on Support Vector Machines," Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, SNPD '08. Ninth ACIS International Conference on, 2008, pp. 517-521

[23] L. Taehee, T. Hollerer, "Handy AR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking," Wearable Computers, 2007 11th IEEE International Symposium on, 2007, pp. 83-90

[24] J. MacLean, R. Herpers, C. Pantofaru, L. Wood, K. Derpanis, D. Topalovic, J. Tsotsos, "Fast Hand Gesture Recognition for Real-Time Teleconferencing Applications," Int. Workshop Recognition, Analysis Tracking of Faces Gestures Real Time Syst., 2001, pp. 133-140

[25] B. Stenger, A. Thayananthan, P. Torr, R. Cipolla, "Modelbased Hand Tracking Using a Hierarchical Bayesian Filter," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 28, No. 9, pp. 1372-1384, 2006

[26] J. Alon, V. Athitsos, Y. Quan, S. Sclaroff, "A Unified Framework for Gesture Recognition and Spatiotemporal Gesture Segmentation," IEEE Trans. Patt. Anal.Mach. Intell., Vol. 31, No. 9, pp. 1685-1699, 2009

[27] H. Ying, J. Song, X. Ren, W. Wang, "Fingertip Detection and Tracking Using 2D and 3D Information," Proc. 7th World Congress Intell. Control Autom., 2008, pp. 1149-1152

[28] S. C. Crampton, M. Betke, "Counting Fingers in Real Time: A Webcam-Based Human-Computer Interface with Game Applications," Proc. Conf. Universal Access Human-Comput. Interaction, 2003, pp. 1357-1361

[29] A. M. Burns, B. Mazzarino, "Finger Tracking Methods Using EyesWeb," Proceedings of the 6th international conference on Gesture in Human-Computer Interaction and Simulation, Vol. 3881, 2006, pp. 156-167

[30] L. Gui, J. P. Thiran, N. Paragios, "Joint Object Segmentation and Behavior Classification in Image Sequences", Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Minneapolis 2007, pp. 1-8

[31] D. Lee, S. Lee, "Vision-based Finger Action Recognition by Angle Detection and Contour Analysis" ETRI Journal, Vol. 33, No. 3, pp. 415-422, 2011

[32] C. David, V. Gui, P. Nisula,V. Korhonen, "Dynamic Hand Gesture Recognition for Human-Computer Interactions," Proc. of 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI), 2011, pp. 165-170

[33] J. Davis, M. Shah, "Recognizing Hand Gestures", European Conference on Computer Vision, 1994, pp. 331-340

[34] V. Gui, G. Popa, P. Nisula, V. Korhonen, "Finger Detection in Video Sequences Using a New Sparse Representation," Acta Technica Napocensis Electronics and Telecommunications, Vol. 52, No. 1, 2011

[35] V. Gui, "Edge Preserving Smoothing by Multiscale Mode Filtering". European Conference on Signal Processing, EUSIPCO'08, Lausanne, pp. 25-29, 2008

[36] S. Gu, Y. Zheng, C. Tomasi, "Efficient Visual Object Tracking with Online Nearest Neighbor Classifier". The 10th Asian Conference on Computer Vision, ACCV Queenstown, New Zealand, 2010

[37] J. Santner, C. Leistner, A. Sa_ari,T. Pock, H. Bischof, "PROST Parallel Robust Online Simple Tracking", IEEE CVPR 2010, pp. 723-730

[38] https://www.leapmotion.com/

[39] www.xbox.com

[40] https://www.myo.com/

Support Vector Machine and Fuzzy Logic

József Menyhárt[1], Róbert Szabolcsi[2]

[1*] Corresponding author, Doctoral School of Safety and Security Sciences, Donát Bánki Faculty of Mechanical and Safety Engineering, Óbuda University, Népszínház u. 8, H-1081 Budapest, Hungary,
E-mail: jozsef.menyhart@eng.unideb.hu

[2] Department of Mechatronics, Donát Bánki Faculty of Mechanical and Safety Engineering, Óbuda University, Népszínház u. 8, H-1081 Budapest, Hungary
E-mail: szabolcsi.robert@bgk.uni-obuda.hu

Abstract: The article gives a short description of the history of the Support Vector Machine (SVM) method and fuzzy logic and their main parameters. It describes how SVM can be used for classification and regression calculation. In addition, it includes the description of the fuzzy logic and the three main defuzzification methods. Examples will be given to show simultaneous use of the two main Support Vector Machine-methods. It ends with the authors' description of the new possibilites for the use of the SVM and fuzzy logic in evaluation of the parameter deviation in the batteries of electric vehicles as well as those of robots.

Keywords: SVM; classification; regression; fuzzy

1 Introduction

Nowadays more and more intelligent machines and tools are used in everyday life, which makes life easier for their users. The engineers working on these kinds of machines make efforts to copy human intelligence. Their goal is to create machines with higher efficiency. Machine intelligence is a combination of different types of methodologies, for instance, fuzzy logic, neural networks, neural computing, anytime techniques, probability theory, etc. [1]

The methods mentioned above work well when complementing each other. However, they cannot provide a useful solution if one of them is used by itself. These methods share a feature: they take advantage of the inaccuracy and uncertainty associated with the tolerances and the boundaries to make a better representation model of the real world. [1]

In a previous work [32] lithium polymer battery parameters were analyzed by Fuzzy logic. The result of this paper presents a brand new operational strategy for

the batteries, which are very sensitive to their operational parameters. The authors would like to find an ideal method for Fuzzy logic which enables us to find ideal limits in the batteries operation parameters. Artificial intelligence equipment systems play a key role in the development of modern industrial applications and vehicles. Thanks to these techniques both the efficiency and the availability can be increased. This paper would like to describe the Support Vector Machine (SVM) and the Fuzzy logic methodologies and some applications where they are used simultaneously.

The study gives a short summary and the syntax of the two methods. The last chapter provides the reader with examples in different practical areas.

The authors describe the usage of SVM during energy supply. The empirical results show that batteries can be used with operation parameters different from those prescribed by the producers. But these operation conditions raise a safety technique question as it is essential not to exceed the limit where the batteries still work with reasonable safety. The authors seek the answer to this question.

This question is of crucial importance from a practical perspective. The performance of the machines is strongly influenced by daily unexpected events.

2 Related Works

The importance of battery parameter check is mentioned in the works of G.F. Guo and his co-authors pointing out the problem of SOC for the battery management system. During their work they checked voltage, amperage and temperature parameters and used the PSO-SVM method for optimization. [23]

In [24], which is a study that considers the batteries as complex electro-chemical cells, we can read about their importance. The Relevance Vector Machine and Particle Filters (PF) were used to check the uncertainty properties of the batteries during use.

Another field where the SVM method was used is the energy management of solar systems. Yen Yee Chia and his co-authors checked the energy flow and its optimization in their studies and conducted research work about supercapacitator-battery hybrid energy storage systems which are a part of solar systems. [25]

The studies mentioned above all deal with the optimization of the original energy storage systems. The increase of the availability of the present systems is also of primary importance. The research results prove that the batteries can work with parameter deviation and they can be used for a longer period of time. The SVM method enables us to determine new operation limit(s) for the batteries if we choose the right hyperplane. Since, this method has great importance in practice, maybe it can be seen from a new perspective.

3 Support Vector Machine

3.1 Theoretical Background of SVM Method

It is undoubtedly true that there is an ever-increasing demand for both mathematical and computer modelling, which runs parallel to the progress in informatics. However, these techniques have to check repetitive and similar schemes, and it is necessary to classify these schemes during their operation. The fastest growing area of artificial intelligence is statistical learning algorithms. In other words, we can call it the theory of machine learning, which enables us to draw conclusions and make generalizations from the learning sets and the sample sets (which means real time observations). The Support Vector (SV) method, which is the most important method, leans on a statistical base. The Support Vector group is made up from different variants of Support Vector Machine (classification) and Support Vector Regression (regression). Their names indicate that not all the elements of the available training set are used for the statistical model. The SVM is a special neural network, or, is a statistical learning theory in some cases. During the last 30 to40 years we have witnessed formation and development of SVM. V.N. Vapnik has done significant work in this topic. [2] [3] [4] [5] [6] [7] [14] [36]

The main task of the SVM method is to find the optimal solution but at the current level of development it can still be used for approximating functions and classifications as well. The SVM is applied in different areas like character recognition, image processing, bioinformatics, data mining etc. The SVM's decision-making process is quick and for this reason it can be used in real time applications. The SVM is applicable for real time applications thanks to its ability to make and implement decisions very quickly. Chapter 3 and the subsections are based on [6]. [2] [3] [4] [5] [6] [7] [14] [36]

3.2 SVM used for classification

Suppose the following [6] [14] [36]:

$$(x_1, y_1) \ldots (x_n, y_n);\ x_i \in R^d;\ y_i \in \{-1,1\} \tag{1}$$

The elements of the samples are derived from two classes. The elements of $x_i \epsilon R^d$ from class A_1 then use $y_i = 1$, if from class A_2 then use $y_i = -1$. The training set is linearly separable if we know a hyperplane one side of which contains only the elements of class A_1, while the other side contains the elements of class A_2. [6] [14] [36]

So

$$\langle x, \varphi \rangle = c \tag{2}$$

is linearly separable with a hyperplane, if

$$\langle x_i, \varphi \rangle > c, \text{ if } y_i = 1, \tag{3}$$

$$\langle x_i, \varphi \rangle < c, \text{ if } y_i = -1, \tag{4}$$

where $\varphi \epsilon R^d$ unit vector is an inner product between *a* and *b* where $c \in R$ $\langle a, b \rangle$. The SVM always looks for the most optimal hyperplane. Figure 1 shows two non-optimal separator lines and another one, which was made by the SVM method. [6] [14] [36]

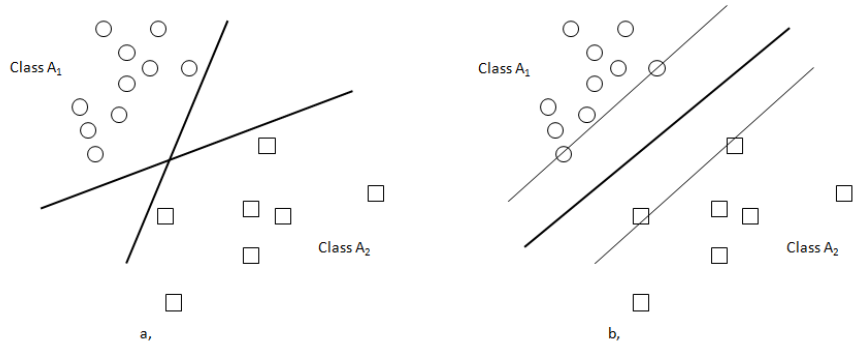


Figure 1
Separation with a hyperplane, two not optimal separations in case a, while case b contains a separation with maximum edge [5] [6] [36]

The thin lines in Figure 1 in Case *b* are edges or margins. The thick line in the middle is called the band limit. The points which fit on the margins are the support vectors. The separating hyperplane is as far away as possible from the sample points. The support vectors are the closest to the optimum hyperplane. [6] [7] [14] [36]

3.3 SVM Used for Regression

The Support vector based statistical classification and regression methods are parts of the statistical learning algorithms group. The Support Vector Machine (SVM) algorithm was the first method for classification, which appeared in the first form of Support Vector Regression in 1997. The SVM can rank an optional number of vectors into classes in the case of binary classification where the model is based on learning vectors. Figure 2 shows a Support Vector Machine architecture. [5] [6] [14] [15] [36]

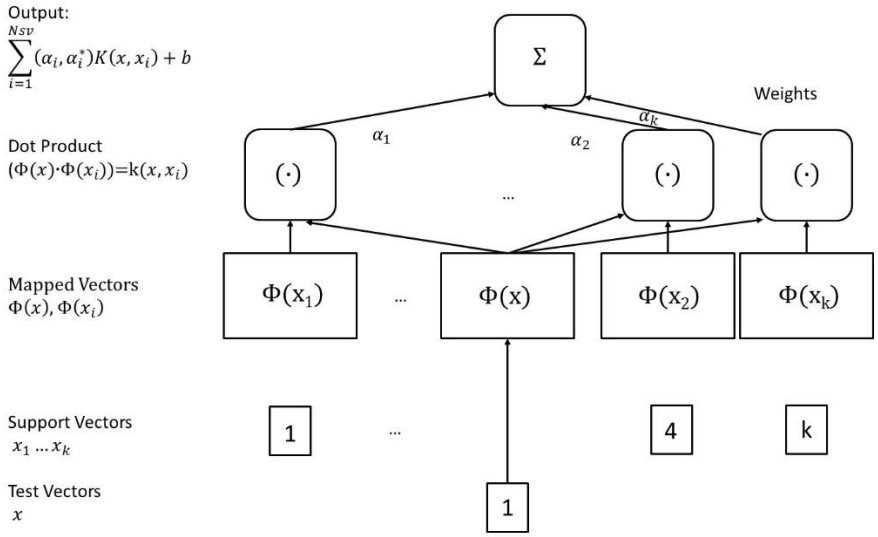


Figure 2
Support Vector Machine architecture [15]

The square of the distance is widely used in the classical areas of mathematics. We can find it in the case of L2 and I2 spaces and in the smallest square methodology of numerical mathematics. The variance can be found in probability theory, the classical regression analysis also use this. [5] [6] [14] [15] [36]

Form:

$$\big(y - f(x, \alpha)\big)^2, \tag{5}$$

Where x is the input and y is the output. We use loss functions (like those used in robust mathematics) in the case of SVM, such as the ε-insensitive loss function. [5] [6] [14] [36]

The ε-insensitive loss function means that the system is not sensitive to deviations which are smaller than ε. Higher deviations than ε are not used quadratically, we use them linearly. [5] [6] [14] [36]

That is

$$L_\varepsilon\big(y - f(x, \alpha)\big) = |y - f(x, \alpha)|_\varepsilon \tag{6}$$

a linear ε-insensitive loss function, where

$$|y - f(x, \alpha)|_\varepsilon = \begin{cases} 0, & \text{if } |f(x, \alpha)| \le \varepsilon, \\ |f(x, \alpha) - y| - \varepsilon, & \text{if } |f(x, \alpha)| > \varepsilon, \end{cases} \tag{7}$$

and it can be seen in Figure 3. [6] [14] [36]

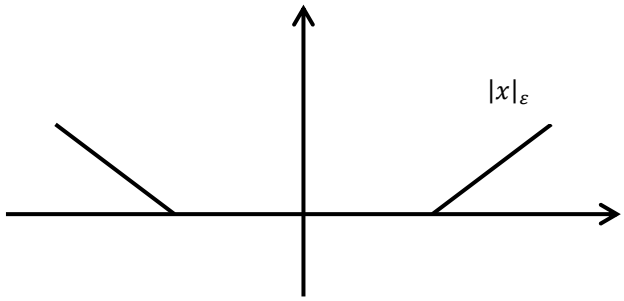


Figure 3
The ε – insensitive loss function [6] [14] [36]

3.4 New Approach of SVM Used for Operating Conditions

In the case of electric and autonomus vehicles the range gets a higher priority. The empirical data show that the batteries can work with parameter deviations.

During the research work the authors describe two sets (or classes):

- prescribed data (during normal battery use)
- empirical data (during aperiodic (random) battery use, parameter deviation)

Table 1 and Table 2 contain the data about the two sets.

Table 1
Prescribed data

Prescribed										
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Max Voltage	4,201	4,204	4,209	4,200	4,205	4,208	4,206	4,200	4,201	4,200

Table 2
Empirical data

Empirical										
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Max Voltage	4,392	4,400	4,390	4,410	4,438	4,410	4,378	4,420	4,400	4,397

The prescribed data (normal use) show a linear result, the inaccuracy during the charge is negligible. The empirical results (random use) show a scatter. An optimal hyperplane can be calculated with the SVM method. This new hyperplane can help the users to use the batteries with a satisfactorysafety level. The difference between the prescribed and emperical results is shown in Figure 4.

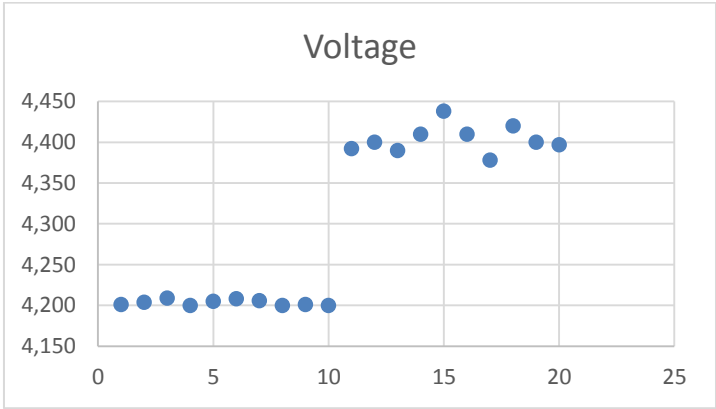


Figure 4
Difference between the measured results

The optimal hyperplane enables the user to choose the right voltage results for a better operational condition. If they choose the new parameters, the range of the vehicles can grow.

The authors address the following questions relating to the topic:

- Where is the ideal place of the optimal hyperplane, between the prescribed and empirical data where it can operatewithin the proper safety range?

  or

- Where in the empirical data is it demonstrated thatare the voltage results higher but the safety level of the vehicle is worse, than in the first case?

The questions are of great importance, in both cases we can find pro and contra arguments. The fuzzy logic may help us to find the answers to these questions.

4 The Theoretical Background of Fuzzy Logic

The first references to fuzzy logic can be found in the works of Lofti Zadeh who achieved outstanding results in this topic in 1965. He examined the blurred boundaries of the truth values of different colloquial concepts from a mathematical approach (calculation with words). He assigned a value between [0;1] (closed interval) to every logical statement during the modelling process. [8] [9] Figure 5 presents the difference between non-fuzzy and fuzzy logic.

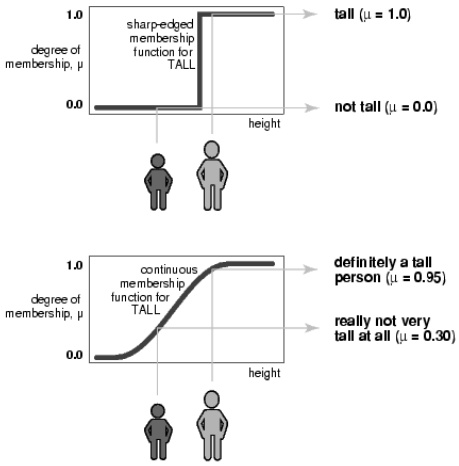


Figure 5
Non-Fuzzy and Fuzzy logic [16]

We talk about nice weather, high speed, and good restaurants with a blurred meaning in our everyday life. If we were surrounded by exact definitions, verbal communication would become impossible so the lack of precision in the vernacular is necessary in our life. If mathematics is used to describe the complexity of the real world, we have to do it in a numerical form. The Fuzzy is the only way to describe human knowledge, expertise and experience in a mathematical form. [8] [9] [10] [19] [21]

If we happen to read the measuring results incorrectly from the gauges in practice, this mistake can lead to inaccuracy during our work. This is the reason why we can use Fuzzy logic when such problems occur. This inaccuracy can be represented by so-called membership functions. The process in fuzzy systems is illustrated in Figure 5. [8] [9] [10] [21] [22]

The first step is fuzzification, which is in fact (about) giving production values to the system. We have to define the categories and the membership functions for the model. The main factors need to be examined. It is important to select an appropriate number of categories. If the number of the categories is increased, we can get an exact picture of the system but it can lead to unnecessary calculations. Having too much information can be as dangerous as having too little. [8] [10] [21] [22] [33] [34]

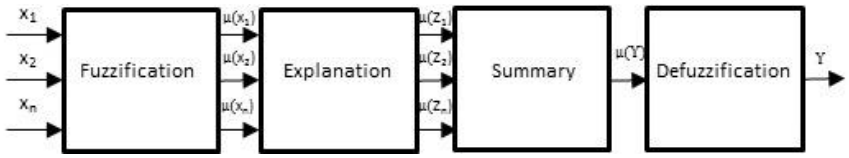


Figure 6
The fuzzy system and process [8] [10] [33] [34]

Every category has to use a membership function, we know several ways to define the categories. The μ(x;a) membership function gives the rate of feature x in set A. Another important step is scale definition. It is practical to use it in the following scales: 0-10, 1-10, 0-100, 1-100. The target of the scale is to find the simplest possible way to compare and evaluate the examined values. [8] [9] [10]

It is essential that the system uses the right rule in the explanation stage. The rules come from the defined categories. If the rules work well, the system can interpret the values. This is how the rules of a fuzzy model are created. [8] [9] [10]

In the final step, we link the values - which we got from the explanation - which differ form 0 in the light of the features of the controlled process, the fuzzy process. The result of this is a fuzzy set. As this is a preconlcusion, this result is not relevant in practical usage. It can be interpreted in the next step called defuzzification. [8] [9] [10]

The defuzzification is the last step of the process. In this step, we have to choose the exact value based on the fuzzy conclusion and this value - depending on the application and the model - is the most representative of the set. The meaning of the defuzzification can be different depending on the usage. We can use different types of defuzzification methods to get the right results. [10] [33] [34]

The most important ones among them are:

- Center of Gravity (COG);
- Center of Area (COA);
- Weighted Mean of Maximum

The Center of Gravity method is one of the most commonly used defuzzification methods. The main advantages of COG are that it is easy to use with triangular and trapezoidal rules and it can show continuous behavior during direct navigating. The COG in general form [10] [33] [34]:

$$Y_{COG} = \frac{\sum_{i=1}^{n} \int_{-\infty}^{+\infty} \mu_i(z) z dz}{\sum_{i=1}^{n} \int_{-\infty}^{+\infty} \mu_i(z) dz} \tag{8}$$

The Center of Area method is very similar to the Center of Gravity. The difference between the two methods is in the calculation. The COG uses areas which are covered by part conclusion while the COA uses only the cumulative consequences. The form of COA [10] [33] [34]:

$$Y_{COA} = \frac{\int_{-\infty}^{+\infty} \mu_{\sum}(z) z dz}{\int_{-\infty}^{+\infty} \mu_{\sum}(z) dz} \quad (9)$$

The Weighted Mean of Maximum is the most frequently used method for defuzzification. The method's result shows the biggest membership value. If it reaches the highest result in one interval, then we have to calculate the mean value of this stage. Its form is [10] [33] [34]:

$$Y_{WMM} = \frac{\sum_{i=1}^{n} \mu_i z_i}{\sum_{i=1}^{n} \mu_i} \quad (10)$$

Menyhárt and Pokorádi have used Fuzzy logic to examine battery parameter deviations and operation condition monitoring in their previous research. [8]

5 Fuzzy Logic and SVM Applications

The SVM is a machine learning technique based on statistical methods. The SVM can be used for classification or regression analysis and its aim is to find the most optimal hyperplane. Thanks to this feature, it is different from other neural networks. [6]

The variables of Fuzzy systems are based on fuzzy sets. These features make it possible to give a numerical description of empirical and linguistic skills. [10]

They cannot support independent machine learning and adaptation. At first glance (it seems as if) the two machine learning methods cannot be used and combined at the same time. This paper presents how the two methods can work together in different applications in everyday life.

[7] shows a new type of use of Support Vector Machine, which forms a kernel function based on a fuzzy rule base. The proposed network combines the characteristics of SVM and Fuzzy systems. It is high general performance, even if the dimension of the input space is very high, structured and it gives a numerical representation of knowledge and ability using linguistic fuzzy rules, in order to bridge the semantic gap between the low-level descriptors and the high-level semantics of an image. The authors checked different types of images with this special Fuzzy-SVM network putting the main focus on urban and beach pictures. [7]

In order to avoid the scale effect, every feature should contain more or less the same numerical values. In this case the MPEG-7 descriptors are already scaled into integer values of equivalent magnitude. The authors used the Personal Content Services database for their work. It contains 767 high quality color images which are divided into two classes (beach and urban) (Figure 7). 40 images were used from the beach category and 20 from the urban during the system training. The remaining 707 were used for the evaluation. [7]

Figure 7
Beach and urban pictures [17] [18]

The Fuzzy SVM network made it possible to extract linguistic fuzzy rules during the classification. An extracted fuzzy rule for the case of the Edge Histogram descriptor [7]:

"IF the number of 0 ± edges on the upper part of the image is low AND the number of 45 ± edges on the upper part of the image is medium AND . . . AND the number of non-directional edges on the lower part of the image is high, THEN the image belongs to class Beach". [7]

The proposed network was successfully applied to the problem of image classification. The fusion of the two methods was very useful as it can provide a lingusitic description of the underlying classification mechanism. The authors refer to their future work: they would like use more MPEG-7 descriptors and more classes. [7]

At the beginning of their work, Yixin Chen and his co-author James Z. Wang give a description of Fuzzy logic: this research topic has been very popular for a long time. The Support Vector Machine method is mentioned, it has a good generalization ability but what is even more important: SVM works well in multidimensional spaces. They describe the connection between fuzzy classification and kernel machines and they look for a link between fuzzy rules and kernel functions and present an algorithm for classification. [12]

Takuya Inoue and his colleague Shigeo Abe at Kobe University in Japan also did a similar research, which was about classification problems. [11]

Boumediene Allaoua and Abdellad Laoufi - unlike the previous examples – used the methods mentioned above in the vehicle industry. They present a new sliding mode fuzzy control scheme for torque control of induction motors. The control principle which they developed is based on sliding mode fuzzy control combined with SVM. The sliding mode fuzzy control contributes to the robustness of induction motor wheel drives and the space vector modulation improves many other properties of the electric system. They used the Lyapunov direct method with fuzzy logic. Figure 8 shows their system. [13]

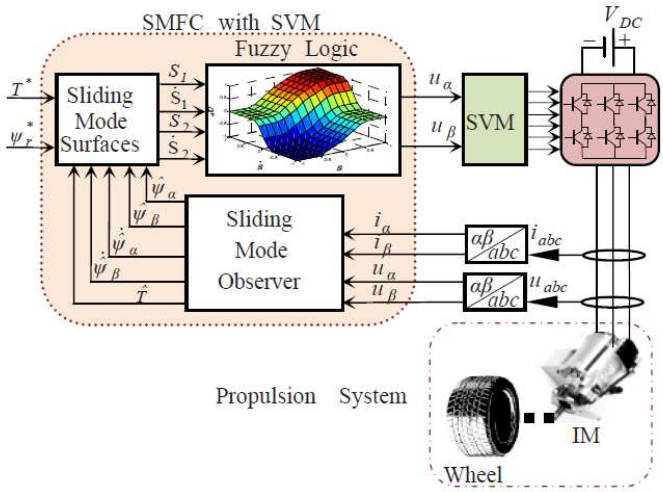


Figure 8
The schematic figure of the system [13]

Compared with the classical PI-control method, the conventional SMC method and SMFC with SVM technique, this new scheme has low torque ripple, low current distortion, and high performance dynamic characteristics of the developed system. This new control scheme can achieve high accuracy in torque tracking to various reference torque signals and shows robustness to external load disturbances. In summary, their system is simple, accurate and it has high reliability. [13]

Conclusion

These examples perfectly represent the combination of Fuzzy logic and SVM methods in different areas. There is an infinite number of problems where we can use these methods but the most important thing is in [13]. In vehicle industry and other industrial applications we would like to use machines with high reliability and outstanding operating performance. Current knowledge shows that Fuzzy logic and SVM can be used to create more than sufficient regulatory systems.

Intelligent power supply systems and power stations are very important during the development of alternative fuel vehicles and autonomous vehicles. [8] presents a Fuzzy rule base for the monitoring of an electric car's battery condition. These points are the base for future work. This work would like to present an energy regulation system with Fuzzy logic and SVM.

This study contains a short description of the history of SVM. The reader can get information about the SVM method for regression analysis and classification. The main focus of this method is to find the most optimal hyperplane.

The article contains another machine learning technique, which is the Fuzzy logic. It provides the reader with details of the story of Fuzzy logic and the three most widely used defuzzification methods.

The study presents the advantages of the simultaneous combination of Fuzzy and SVM. It is possible to find solutions to different types of classification problems with these methods. But we can find the most substantial opportunity in the vehicle industry. The articles contains two examples, one of classification and one of vehicle industry.

The focus of our future work will be the determination of the optimal hyperplane. We also have to distinguish a milk run robot from a military robot. In the case of military different missions and events. The next research work focuses on the simultaneous usage of the Support Vector Machine and Fuzzy logic.

The system has to check the telemetry data in real time and it is supposed to be able to define which hyperplane is the best in practice, in different real situations.

References

[1] A. R. Várkonyiné-Kóczy: Intelligent Modelling Systems: When is It a Solution for the Unsolvable?, Lecture notes, Óbuda University, Institute of Mechatronics and Vehicle Engineering, 2014

[2] S. Haykin: Neural Networks and Learning Machines, Third Edition, Mc Master University Hamilton, Ontario Canada, Prentice Hall (2009) ISBN-13: 978-0-13-147139-9; ISBN-10: 0-13-147139-2

[3] A. Smola, B. Schölkopf: A Tutorial on Support Vector Regression, Kluwer Academic Publishers, Manufactured in The Netherlands, Statistics and Computing 14 (2003) ISSN: 0960-3174. pp. 199-222

[4] S. Gunn: Support Vector Machines for Classification and Regression, Technical Report, University of Southampton, Faculty of Engineering Science and Mathematics, School of Electronics and Computer Science (1998)

[5] G. Horváth: Neural Networks and Their Technical Applications, Publisher of Műegyetem, ISBN 96334205771 (1998)

[6] I. Fazekas: Neural Networks, University of Debrecen, Faculty of Informatics, TÁMOP-4.1.2.A/1-11/1-2011-0103 (2013)

[7] E. Spyrou, G. Stamou, Y. Avrithis, S. Kollias: Fuzzy Support Vector Machines for Image Classification fusion MPEG-7 visual descriptors url: http://doras.dcu.ie/353/1/lncs_3697.pdf Download: 2015.01.17.

[8] J. Menyhárt, L. Pokorádi: The Fuzzy Rule-based Estimation of the Battery Condition, Hadmérnök Katonai Műszaki Tudományok Online, ISSN 1788-1919, 2014/2. pp. 48-55

[9] T. Portik., L. Pokorádi: Possibility of Use of Fuzzy Logic in Management, 16th Building Services, Mechanical and Building Industry Days, 147 p., Conference: Debrecen, Hungary, 2010.10.14-2010.10.15, University of Debrecen, ISBN 978-963-473-422-2, 1, GEOREN, (2010) pp. 353-360

[10] L. Pokorádi: Modelling of Systems and Processes, Campus Publisher, Debrecen, Hungary (2008)

[11] T. Inoue, S. Abe: Fuzzy Support Vector Machines for Pattern Classification, Kobe University Repository: Kernel (2015)

[12] Chen Y., Wang J. Z.: Support Vector Learning for Fuzzy Rule-based Classification Systems, IEEE Transactions on fuzzy systems, Vol. 11 (2003)

[13] Allaoua B., Laoufi A.: A Novel Sliding Mode Fuzzy Control Based on SVM for Electric Vehicles Propulsion System, ScienceDirect, Energy Procedia, (2013), pp. 120-129

[14] T. Borsavölgyi, T. Mankovits, I. Kocsis: The Application of SVM for the Optimization of Technical Tasks, Technical Science in the Eastern Hungarian Region (2011) pp. 499-507

[15] Aceves-Fernandez Marco Antonio, Sotomayor-Olmedo Artemio, Gorrostieta-Hurtado Efren, Pedraza-Ortega Jesus Carlos, Ramos-Arreguín Juan Manuel, Canchola-Magdaleno Sandra and Vargas-Soto Emilio: Advances in Airborne Pollution Forecasting Using Soft Computing Techniques, Intech, Open Science – Open Minds, Url: http://www.intechopen.com/books/air-quality-models-and-applications/advances-in-airborne-pollution-forecasting-using-soft-computing-techniques , Download: 2015.05.10. 15:32

[16] Faculty of Electrical Engineering, Czeh Technical University in Prague, Department of Radio Engineering K 13137, CTU FEE Prague, Fuzzy Logic Toolbox, url: http://radio.feld.cvut.cz/matlab/toolbox/fuzzy/fuzzytu3.html Download: 2015.09.19. 14:19

[17] SantaBanta.com: Beaches, url:http:// www.santabanta.com/photos/beaches/2122066.htm, Download: 2015.09.19. 14:10

[18] The Telegraph: China's urban population exceeds rural for first time, url: http://www.telegraph.co.uk/news/worldnews/asia/china/9020486/Chinas-urban-population-exceeds-rural-for-first-time-ever.html, Download: 2015.09.19. 14:15

[19] Zs. Cs. Johanyák: Fuzzy Modeling of Thermoplastic Composites' Melt Volume Rate, Computing and Informatics, Vol. 32 (2013) No 4, ISSN 1335-9150, pp. 845-857

[20] Zs. Cs. Johanyák and J. Kovács: Fuzzy Model-based Prediction of Ground-Levele Ozone Concentration, Acta Technica Jaurinensis, Series Intelligentia Computatorica, ISSN 1789-6932, Vol. 4, No. 1 (2011) pp. 113-126

[21] L. Pokorádi: Risk Assessment Based upon Fuzzy Set Theory, 15th Building Services, Mechanical and Building Industry Days. 540 p., Conference: Debrecen, Hungary, 2009.10.15-2009.10.16, University of Debrecen, ISBN 978-963-473-315-7 (2009) pp. 311-318

[22] L. Pokorádi: Application of Fuzzy Set Theory for Risk Assessment: Zastosowanie Teorii Zbiorów/ukladów Rozmytych do Oceny Ryzyka, ISSN 1895-8281, No 2,3 (14, 15) Journal of Konbin 14-15, Warsaw, Poland (2010) pp. 187-196

[23] Gui Fang Guo, Lin Shui, Xiao Lan Wu, Bing Gang Cao: SOC Estimation for Li-Ion Battery Using SVM Basen on Particle Swarm Optimization, Advanced Materials Research, Vol. 1051 (2015) ISSN: 1662-8985, pp. 1004-1008

[24] B. Saha, S. Poll, K. Goebel, J. Christophersen: An Integrated Approach to Battery Health Monitoring Using Bayesian Regression and State Estimation, National Aeronautics and Space Administration, Intelligent Systems Division url: http://ti.arc.nasa.gov/m/pub-archive/1367h/1367%20%28Goebel%29.pdf , Download: 2015.10.28

[25] Yen Yee Chia, Lam Hong Lee, Niusha Shafiabady, Dino Isa: A Load Predictive Energy Management System for Supercapacitor-Battery Hybrid Energy Storage System in Solar Application Using the Support Vector Machine, Elsevier, Applied Energy, Vol. 137 (2015) ISSN: 0306-2619, pp. 588-602

[26] Chih-Min Lin, Yi-Jen Mon, Ching-Hung Lee, Jih-Gau Juang, Imre J. Rudas: ANFIS-based Indoor Location Awareness System for the Position Monitoring of Patients, Acta Polytechnica Hungarica, Vol. 11, No. 1 (2014) ISSN 1785-8860, pp. 37-48

[27] Chiu-Hsiung Chen, Chang-Chih Chung, Fei Chao, Chih-Min Lin, Imre J. Rudas: Intelligent Robust Control for Uncertain Nonlinear Multivariable Systems using Recurrent Cerebellar Model Neural Networks, Acta Polytechnica Hungarica, Vol. 12, No. 5 (2015) ISSN 1785-8860, pp. 7-33

[28] A. R. Varkonyi-Koczy, A. Rovid: Soft Computing Based Point Correspondence Matching for Autoatic 3D Reconstruction, Acta Polytechnica Hungarica, Vol. 2, No. 1 (2005) ISSN 1785-8860, pp. 33-44

[29] S. Preitl, R-E. Precup, J. Forod, B. Bede: Iterative Feedback Tuning in Fuzzy Control Systems. Theory and Applications, Acta Polytechnica Hungarica, Vol. 3, No. 3 (2006) ISSN 1785-8860, pp. 81-96

[30] A. R. Varkonyi-Koczy, A. Rovid, Sz. Balogh, T. Hashimoto, Yoshifumi Shimodaira: High Dynamic Range Image Based on Multiple Exposure Time Synthetization, Acta Polytechnica Hungarica, Vol. 4, No. 1 (2007) ISSN 1785-8860, pp. 5-15

[31] Gy. Bardossy, J. Fodor: Assessment of the Completeness of Mineral Exploration by the Application of Fuzzy Arithmetic and Prior Information, Acta Polytechnica Hungarica, Vol. 2, No. 1 (2005) ISSN 1785-8860, pp. 15-31

[32] L. Pokorádi, J. Menyhárt: Electric Vehicles' Battery Parameter Tolerances Analysis by Fuzzy Logic, Proceedings of the 11th IEEE International Symposium on Applied Computational Intelligence and Informatics SACI 2016, Budapest: IEEE, 2016. pp. 361-364, ISBN 978-1-5090-2379-0 Timisoara, Romania, 2016.05.12-2016.05.14

[33] A. Rikalovic, I. Cosic: A Fuzzy Expert for Industrial Location Factor Analysis, Acta Polytechnica Hungarica, Vol. 12, No. 2, (2015) ISSN 1785-8860, pp. 33-51

[34] L. Bessissa, L. Boukezzi, D. Mahi: A Fuzzy Logic Approach to Model and Predict HV Cable Insulation Behaviour under Thermal Aging, Acta Polytechnica Hungarica, Vol. 11, No. 3 (2014) ISSN 1785-8860, pp. 107-123

[35] C. B. Regaya, A. Zaafouri, A. Chaari: A New Sliding Mode Speed Observer of Electric Motor Drive Based on Fuzzy-Logic, Acta Polytechnica Hungarica, Vol 11, No. 3 (2014) ISSN 1785-8860, pp. 219-232

[36] V. N. Vapnik: Statistical Learning Theory, AT&T Research Laboratories, A Wiles-Intercience Publication, John Wiles & Sons In., 1998, ISBN 0-471-03003-1

Risk Management Impact Assessment on the Success of Strategic Investment Projects: Benchmarking Among Different Sector Companies

Filip Jovanović[1], Nenad Milijić[2], Makedonka Dimitrova[3], Ivan Mihajlović[2]

[1] Faculty of Project and Innovation Management, Boze Jankovica 14, 11000 Belgrade, Serbia

[2] University of Belgrade, Technical Faculty in Bor, Management Department, Vojske Jugoslavije 12, 19210 Bor, Serbia

[3] University American College Skopje, Faculty of Business Economics and Management, Blvd. Treta makedonska brigada no.60, Skopje 1000, Macedonia

filip.jovanovic@mozzartbet.com, nmilijic@tf.bor.ac.rs, dimitrova@uacs.edu.mk, imihajlovic@tf.bor.ac.rs

Abstract: The paper explores the impact of the main elements of the strategic investment projects management process in South East Europe (SEE), on achieving project objectives. The impact of different projects elements on the project's risk management process was also addressed. The initial hypotheses are based on literature review in the field. The proposed hypotheses were tested by the SEM (Structural Equation Modeling) methodology on a sample of 311 strategic investment projects in Serbia, Bosnia and Herzegovina as well as Macedonia. The strategic projects were analyzed in companies across various business sectors thus the obtained results represent a benchmark of risk factors' significance according to the project team members and managers. The benchmarking of risk factors significance was conducted by using the PROMETHEE – GAIA methodology. The results verify the adequacy of the hypothetical framework.

Keywords: risk management; strategic investment projects; project objectives

1 Introduction

Strategic projects (SP) are of great significance for organizations, as they enable environment for realization of strategic objectives. The SP may vary from major investment projects, such as, construction of a new factory or plant, introduction

of modern technology, improvement of energy efficiency etc., to somewhat smaller projects such as information system introduction, development and implementation of a new organizational structure, quality system introduction or similar.

Asrilhant et al. suggested that the strategic projects are necessary when an organization strives to achieve its long-term objectives and development [1]. In this respect, Schoemaker defines the strategic projects as a manner of implementation and realization of a sound organizational vision [2]. Moreover, SPs are connected to the main investments of the companies and often bear considerable risk and uncertainty, intangible benefits and enable attractive long-term financial benefit [3, 4].

The strategic project management process helps achieve the projects' successful implementation, including the financial and the non-financial outcomes, as well as, the benefits. Strategic project management includes two primary phases: the evaluation phase and the monitoring phase [5]. The evaluation phase includes development, planning and the evaluation of the strategic projects as well as the project approval. The strategic projects' monitoring phase envisages the project management, the control and possible adaptation processes in case of necessity [1]. Efficient strategic project management may be achieved if the responsible project manager and the project team perform in accordance with the evaluation and monitoring phase requirements; all based on the predefined influence elements [6]. The process involves application of contemporary methods and techniques necessary to achieve efficient management of the strategic projects [7]. In practice, strategic investment projects most frequently apply the following techniques for their evaluation and monitoring: the return on investment, the net present value, the internal rate of return, the repayment period, the cost benefit analysis, the sensitivity analysis, the decision tree, the risk analysis, the forecasting methods, the game theory, the simulation etc. [1]. Selection process of the methods and techniques to be used in the strategic project management process involves prior analysis and assessment to which extend certain methods and techniques would be applied, if applicable at all [8].

The evaluation and the forecasting of future undertakings, processes and activities are accompanied by uncertainty and risk. Some future project settings might bear risk that could lead to negative and in rare cases positive impact on the project implementation [9]. This is the reason that the project management process entails appropriate project risk management. According to Kerzner, the risk management is an action or an exercise for dealing with risk [10]. The project risk management includes several related actions, such as the risk planning, the identification and analysis of risk events, the development of strategies for risk handling and monitoring, all oriented towards the project success [11]. Successful risk control and management in the strategic investment projects may lead towards the successful projects' implementation. The level of success may be estimated by application of an appropriate criterion [12].

This research paper focuses on the strategic investment projects' (SIP) management processes in the companies from the industrial and the non-industrial sectors in South-Eastern Europe (SEE), in particular from Serbia, Macedonia and Bosnia and Herzegovina. Research results from all three countries were analyzed simultaneously. We can explain this decision based on the similarity of all three countries. Namely, not long ago they were Member States of one country, Yugoslavia. Even today, political, economic, social and technological factors do not differ much in all three environments and therefore, we assumed that there should not be much differences in the responses. This research endeavor targets the risk management process, primarily in terms of understanding the effective management process at SIP. Therefore, the objective of this paper is to analyze the impact of the main elements of the risk management processes at SIP in SEE on the achievement of the project goals as well as the impact of the project elements on the risk management process. The importance of the risk factors, were determined, by benchmarking, in accordance with the opinion of the project managers and team members operating across different sectors.

2 Theoretical Framework and Research Hypotheses

The research framework of this work was preceded by extensive analysis of the available literature in the field of SIP risk management. Numerous recently published research papers pointed out to the relationship between the appropriate methods for the risks analysis and assessment based on the factors of significance for the decision makers, for evaluation of the overall success of SIP [13, 14, and 15]. Therefore, we have defined the first research hypothesis of this paper as follows:

H1: The application of the methods for risk analysis and assessment is directly related to the risk factors significant for the success of the company's projects

Given the significant relationship between the methods used for the analysis and assessment of risks and the risk factors that are primarily manageable [16, 17, 18], we have defined the second research hypothesis as follows:

H2: The application of methods for risk analysis and assessment is directly related to risk factors of significance for project risk management

The contemporary literature dealing with project risk management identifies that the analysis of risk factors is significant for the project's success and their relationship with the risk management success [19, 20, 21]. The following hypothesis is based on the previous findings:

H3: The risk factors of significance for the success of a company's projects influence the risk management results

Analysis of the factors significant for risk management and their relationship with the success of risk management is featured in a many research publications as well [22, 23, 24]. Based on results of the authors of the aforementioned papers, the following hypothesis is proposed as:

H4: The factors of significance for project risk management affect the risk management results

On the other hand, some of the latest papers dealing with the correlation of elements of significance to SIP control, and the applied methods for project management [25, 26, 27], have led to the proposal of the following hypothesis:

H5: The elements of significance for the control of the strategic investment project are directly correlated to the methods for strategic investment project management

Apart from that, numerous authors emphasize that the applied methods for SIP management are directly related to a significant criteria for measuring the project success [28, 29, 30]. Therefore, the following hypothesis is further defined as:

H6: The methods for strategic investment project management are directly correlated to the main criteria for measuring project success

In addition to the six research hypotheses, the following statements would be considered axioms, since they are widely accepted in theory and practice [9, 31, 32, 33].

A1: The results of the project risk management affect the results of the strategic investment project management

A2: The main criteria for measuring the project success affect the results of the strategic investment project management

3 Sample and the Data Collection

The research objective of this paper is to understand the extent to which the risk management process of SIP may be characterized by a set of elements, set forth by the project managers as the most significant ones and which are in accordance with the wide range of data available in the literature related to the topic. The identified elements of significance were then grouped in several groups of research questions. The data collection in this particular research used the questionnaire based methodology [1, 31].

The questionnaire consisted of two parts. The first part consisted of 10 (ten) control questions of a demographic character relevant to the surveyed sample (project-oriented companies, respondents and projects). The second part of the

questionnaire featured 42 questions in the field of risk management and significant factors for SIP management, all divided into appropriate groups defined in accordance to proposed hypothesis, as presented in the Appendix 1. Based on the questionnaire, the opinions of the project managers on the importance of the methods for analysis and assessment of project risk were reviewed (1st group of questions – G1), along with risk factors having the greatest impact on the projects' success (2nd group of questions – G2), then the most significant factors for the project risk management (3rd group of questions – G3), the elements significant in the process of SIP control (4th group of questions – G4), the methods for the evaluation and control of SIPs (5th group of questions – G5), as well as the main criteria for measuring the success of SIPs (6th group of questions – G6). The impact of each of these groups of factors on the achieved results of project risk management (Key Question 1 - Q1), as well as the success of management of SIPs of the investigated companies (Key Question 2 – Q2) were then analyzed.

The Likert scale was used to measure the different levels of significance, where 1 represents the lowest significance (I absolutely disagree) and 5 represent the highest significance (I absolutely agree). Also, certain questions were of a dichotomous character (yes/no). In order to collect relevant data, project managers and project team members were sampled, in particular those in charge for SIP management or those involved with SIP on frequent basis. A total of 400 questionnaires were sent to potential respondents, 311 of which, were fully completed within the stipulated deadline from Serbia, Bosnia and Herzegovina and Macedonia. A relatively high response rate was achieved, owing to the persistence and the direct contact between the researchers of this paper and the surveyed managers and employees. Employees–non-managers were also included in the research, with the condition that they were engaged in the project teams of the company. Detailed demographic indicators of the companies are presented in Table 1.

Table 1
Profiles of companies, respondents and projects

Characteristics			*N*	%
Company (*N*=311)	Field of the company	IT	50	16.08
		Finance	13	4.18
		Energy sector	29	9.32
		Public administration	31	9.97
		Traffic	13	4.18
		Education	49	15.76
		Scientific-research	71	22.83
		Other	55	17.68
	Number of employees	<10	57	18.33
		11-50	102	32.80
		51-250	121	38.91

		251-1000	31	9.96
		>1000	0	0
Respondent (*N*=311)	Age	<29	58	18.65
		30-44	81	26.04
		45-54	65	20.90
		>55	107	34.41
	Years spent in the company	<5	172	55.31
		6-15	89	28.62
		16-25	16	5.14
		>26	34	10.93
	Years of service	<5	75	24.12
		6-15	82	26.37
		16-25	29	9.32
		>26	125	40.19
	Level of education	Secondary vocation. educ.	31	9.97
		High education	41	13.18
		Higher education	17	5.47
		HE - Master	82	26.37
		MA	21	6.75
		PhD	119	38.26
	Field of education	Technical-technological	151	48.55
		Legal-economics	49	15.76
		Social-humanistic	81	26.04
		Natural- mathematics	12	3.86
		Other	18	5.79
	Position in the company	Top manager / director	123	39.55
		Middle management	42	13.50
		Operational level of management	77	24.76
		Employees	69	22.19
Project (*N*=311)	Type of projects	ICT	41	13.18
		R&D	34	10.93
		Construction	17	5.47
		Public administration	29	9.32
		Scientific-research	83	26.69
		Other	107	34.41
	Duration of project implementation	<6 months	69	22.19
		6 month – 2 years	213	68.49
		2-5 years	29	9.32
		>5 years	0	0

4 Results and Discussion

In the following text, the results of the data analysis are presented, in order to identify the relationship among elements of SIPs, as outlined in the initial hypothetical framework. Subsequent to the data entry into a database, we have proceeded to data analysis using the corresponding statistical analysis tools. The statistical analysis included the measurement of adequacy of the whole sample and the validation of the data structure. Then the analysis of the reliability of the risk management indicators of SIPs, placed within the appropriate factor groups, was performed along with testing the initial hypothetical frameworks through the application of structural equations modeling. The statistical analysis of the collected data was performed using the software packages SPSS 18.0 and LISREL 8.80.

4.1 Measures of the Sampling Adequacy and Structure Validation

The measure of sampling adequacy analysis was performed by using the Kaiser–Meyer–Olkin (KMO) and Bartlett tests. Based on the recommendations of other authors, the minimum acceptable value of the Kaiser–Meyer–Olkin indicator is 0.6 [34]. The analysis showed that the coefficient of the Kaiser–Meyer–Olkin (KMO) sampling adequacy test, for the results of the questionnaire conducted in this research, is 0.738, indicating that the collected data is suitable for the application of factor analysis. Aside from that, the Bartlett test of sphericity shows the importance ($\chi^2 = 592.16$, $p < 0.000$), indicating that there are acceptable correlations among items within the measuring instrument, and that the correlation matrix is not a unit matrix [35].

4.1.1 Correlation Matrix

The correlations between the six factor groups and the two key questions on risk management of SIPs within the proposed model are presented in Table 2. Based on the results in Table 2, it can be concluded that the majority of coefficients are near or above the value of 0.5 pointing to a significant internal correlation between the listed SIP management factors, and thus the use of factor analysis in further research is justified [36]. Moreover, most of the correlation coefficients have statistical significance at the level of 0.01.

Table 2
Internal correlations between 6 factor groups and two key questions of significance for strategic investment project risk management

Coefficient	G1	G2	G3	G4	G5	G6	Q1	Q2
G1	1.000							
G2	0.403**	1.000						
G3	0.651**	0.618**	1.000					
G4	0.659**	0.467**	0.561**	1.000				
G5	0.781**	0.173	0.446**	0.619**	1.000			
G6	0.417**	0.806**	0.721**	0.550**	0.218*	1.000		
Q1	0.614**	0.491**	0.390**	0.555**	0.477**	0.474**	1.000	
Q2	0.637**	0.422**	0.497**	0.583**	0.663**	0.474**	0.819**	1.000

Note: The level of statistical significance * $p < 0.05$; ** $p < 0.01$

4.1.2 Factor Analysis

The factor analysis was performed in order to extract questions that should be kept within the main factors of control and evaluation of SIPs, as well as the analysis and the management of project risks, placed in the groups G1-G6. The relationships between the measured variables are such that their regrouping into smaller sets of variables may be performed, representing a more concise and understandable structure of the examined field [35].

Based on the results of the factor analyses, it was concluded that, in almost all question groups, certain questions should be eliminated from the final model or further divided in the subgroups. Accordingly, for the group G1 – the analyses of the methods for the assessment and analysis of project risks (Appendix 1), the factor analysis revealed that questions 1.1-1.5 should remain in this factor group, while the questions 1.6. - 1.8 should be removed due to the low factor loadings. Similarly, only the questions 2.1, 2.2 and 2.3, from the group G2 – the results of the risk factors having impact on the project success, should remain in this factor group, while the questions 2.4, and 2.5, should be avoided from the following analysis. On the other hand, considering the large number of questions in the group G3, it should be divided into two factors (the subgroups G3.1 and G3.2), with the group G3.1 including the questions 3.1 - 3.3 and 3.5, and the group G3.2 including all other questions. Almost all the questions in the group G4 – elements of significance for SIP evaluation and control, are in the same factor, with the exception of the question 4.5. The questions in the group G5 – the methods of SIP evaluation and control are one-dimensional, as all of them should remain in just one factor, according to their high factor loadings. The questions in the group G6 – the criteria for measuring SIP success should all remain within one factor, except for the question 6.6, which should be removed from the further analysis. The accuracy of such segmentation of the questions, resulting after the factor

analysis, was verified using the analysis of reliability of the listed grouped indicators of SIP risk management, presented in the following text.

4.2 Analysis of the Reliability of SIP Risk Management Indicator Factors

The evaluation of the internal consistence of the concentration of questions into factors of the indicators of SIP risk management, resulting from the factor analysis was performed by using the Cronbach alpha, the Spearman–Brown and the Ω tests [37]. According to these tests, presented in Table 3, values of the Cronbach α, the Spearman–Brown and the Ω coefficient higher than 0.70, represent a good enough internal consistency to enable accurate modeling of the questionnaire results within the surveyed population [38].

Table 3
Coefficients of internal consistency of the final groups of questions (after the factor analysis)

Question groups	Number of questions within a group	Cronbach α coefficient	Spearman–Brown coefficient	Ω coefficient
G1	5	0.851	0.835	0.858
G2	3	0.741	0.852	0.800
G3	10	0.822	0.369	0.889
G4	4	0.896	0.813	0.898
G5	6	0.924	0.901	0.928
G6	5	0.861	0.732	0.931

4.3 Structural Model

The Structural Equation Modeling (SEM) represents a multivariate statistical technique used to analyze the structural relations between the dependent and latent variables. Its greatest advantage is the possibility for multivariate analysis at one time.

Two types of variables are identified in the analysis i.e. endogenous – variables determined by the system or variables deriving from the models (representing an equivalent to a dependent variable), and exogenous – variables outside the models, meaning that their values are accepted as given (representing the equivalent to an independent variable). SEM tests and evaluates the relationships in a model using a combination of statistical data and qualitative causal assumptions. It can be used for the verification or development of hypotheses, or for the confirmative or research purposes. One of the great strengths of this type of modeling is the ability to create latent variables, variables that may not be measured directly but are assessed in the model based on other measured variables. These variables are

actually grouping factors of certain measured variables. In order to develop a model of structural equations, the grouping of the initial measured variables by application of, for example, factor analysis, is performed beforehand.

Based on the conclusions drawn from the results of the statistical analysis, especially the factor analysis described in the previous text, and in order to verify the initial research hypotheses, the structural models for the analysis of risk in the examined SIPs were set. The models are presented in Figure 1.

The following correlations have been verified through the structural equation model (Figure 1a):

- The applied methods of the project risk assessment and analysis (G1) are positively related to the risk factors that affect SIP success (G2). The correlation level is 0.41. It may therefore be concluded that the research hypothesis H1 has been verified.
- The applied methods of the project risk assessment and analysis (G1) are also positively related to the risk factors that affect SIP success (G3). The correlation level is 0.60. Verification of the hypothesis H2.
- The risk factors affecting the overall project success (G2) are positively related to the achieved results of the project risk management (Q1). The correlation level is 0.32. Verification of the hypothesis H3.
- The factors affecting the project risk management (G3) are also positively related to the achieved results of the project risk management (Q1). The correlation level is 0.37. Verification of the hypothesis H4.
- The achieved results of the project risk management (Q1) are positively related to the success of SIP risk management (Q2). The correlation level is 0.74. Axiom A1.

The following correlations have also been verified (Figure 1b):

- The elements of the significance in the process of SIP control (G4) are positively related to the methods for evaluation and control of SIPs (G5), with the correlation level of 0.41. Verification of the hypothesis H5.
- The methods for evaluation and control of SIPs (G5) are positively related to the criteria for the measuring of SIP success (G6), with 0.40 correlation. Verification of the hypothesis H6.
- Finally, the criterion for measuring the success of SIPs (G6) is directly related to the success of SIP management (Q2), with correlation level of 0.59. Axiom A2.

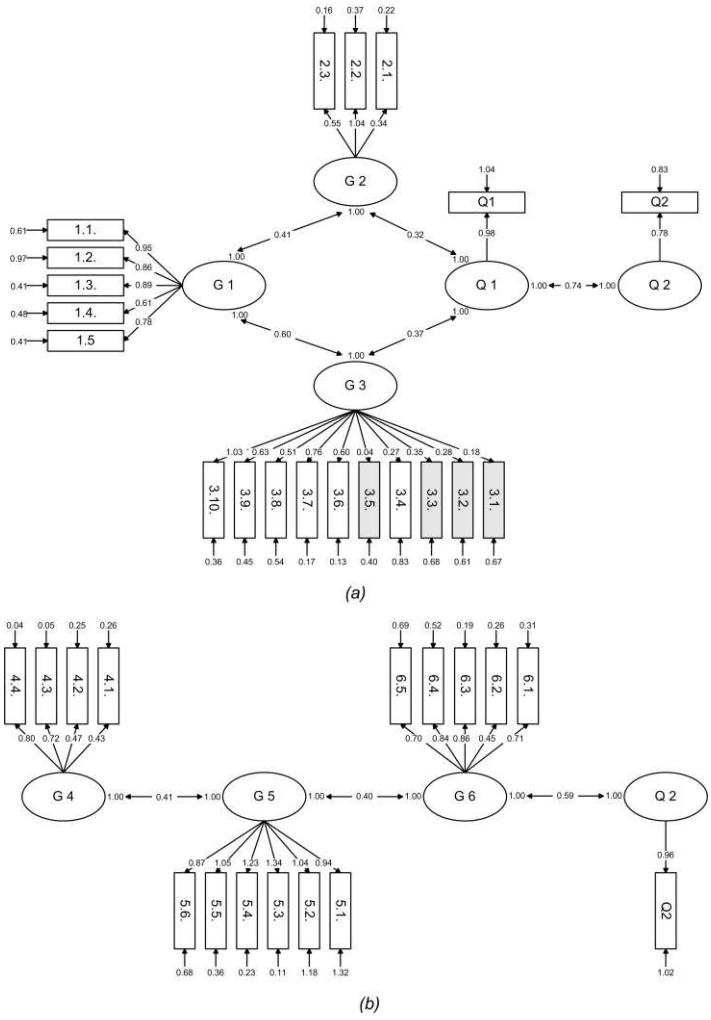


Figure 1

Structural models for the verification of the initial research hypotheses

Considering the existence of a strong positive direct correlation between the listed question groups, also proven in the form of combination of the direct and indirect correlations using SEM, it may be concluded that all of the hypotheses proposed in this research have been verified. The relationship between the two hypothetical models (Figure 1a and Figure 1b), may be established through a common key question, as present in both cases – the key question Q2.

4.4 Ranking the Importance of the Project Risks

To assess the influence of the project types, based on the field of the company, and the position of employees in organizations related to the success on risk management of the SIPs, in addition to the success of the projects achieving predefined goals, the Multi-Criteria Decision Analysis (MCDA) was used [39]. The MCDA methodology was based on the PROMETHEE II technique sustained with the GAIA plane representation. Obtained MCDA results represent a comparison of the identified project risks in analyzed companies across different sectors. This way, it could be identified which of the risk factors are of largest influence on success of quality, time and cost management at SIPs [40, 41]. During the process of application of the PROMETHEE II methodology, the alternatives were ranked based on the values of their net flows, which represents a complex ranking procedure. Moreover, this method could use the DECISION LAB software, enabling graphical visualization of the obtained results in the form of the GAIA plane (Geometrical Analysis for Interactive Aid) [42].

The influence of the two demographic parameters (the project type and the position of the respondent in the organization) on the SIP risk management was applied to the methodology as well. The initial data used in the PROMETHEE procedure is presented in the Table 4. The values in the Table 4 are presenting the average ranking of the individual groups of questions assessed by the employees which are involved in different SIPs in the SEE region, at different work place locations. For each work place, which is actually representing the position of the employee in the project team (top manager P1, middle management level P2, operational management level P3 and employee P4), given the researched project types (ICT (T1), I&R (T2), civil construction works (T3), public sector (T4), scientific research work (T5) and else (T6)), the average ranking of the employees was calculated and presented in the Table 4.

Given that the PROMETHEE methodology includes the weight coefficient for each selected criterion, and as such is associated with certain decision functions, the results are presented in the Table 5. The weight coefficients can be described as raw indicators of relative importance of each selected criterion in the analysis. The entropic approach was used for the weights determination [43]. The Min/Max values were based on the type of the questions and their potential impact on the investigated factors.

The data presented in the Table 4 has been analyzed by using the software Decision Lab 2000. The results of the full ranking based on the PROMETHEE II method are presented in the Table 6 and the Figure 2. The visual presentation of obtained ranking is presented in the Figure 3.

Table 4
Initial data used in the PROMETHEE multi criteria decision making analyses for the strategic investment projects (SIP)

Criteria / Alternatives	Applied methods of project risk assessment and analysis (G1)	Risk factors that affect SIP success (G2)	Risk factors that affect SIP success (G3)	Elements of significance in the process of SIP control (G4)	Methods for evaluation and control of SIP (G5)	Main criteria for the measurement of SIP success (G6)	Did applied management method give good results? (Q1)	Was the SIP management successful? (Q2)
ICT, Top manager / director (T1P1)	3.40	3.90	4.00	4.10	2.60	4.30	2.00	2.00
ICT, Operational level of management (T1P3)	1.50	2.80	2.60	2.70	2.10	2.30	1.00	1.00
R&D, Top manager / director (T2P1)	3.10	3.70	3.60	4.40	3.60	3.80	2.00	2.00
R&D, Employees (T2P4)	2.60	4.10	3.80	4.10	2.90	4.80	2.00	2.00
Construction, Middle management (T3P2)	2.70	4.20	4.20	4.00	3.80	4.50	2.00	2.00
Public administration, Middle management (T4P2)	2.10	3.00	3.20	2.50	2.70	3.70	1.00	1.00
Public administration, Employees (T4P4)	2.90	2.60	4.00	4.10	3.60	3.00	1.00	1.00
Scientific-research, Middle management (T5P2)	2.20	4.00	4.20	3.90	1.60	4.20	1.00	1.00
Scientific-research, Operational level of management (T5P3)	2.50	3.70	4.10	2.80	2.40	3.70	1.10	1.10
Scientific-research, Employees (T5P4)	3.10	3.80	3.70	3.70	2.80	3.90	2.00	1.40
Other, Top manager / director (T6P1)	2.40	3.50	3.80	3.30	2.30	3.90	1.30	1.30
Other, Operational level of management (T6P3)	2.20	3.50	3.80	3.70	2.60	4.30	1.10	2.00

Table 5
Preference functions and weight coefficient of criterions

Criterion	G1	G2	G3	G4	G5	G6	Q1	Q2
Weight coefficient	0.1246	0.1235	0.1232	0.1240	0.1253	0.1241	0.1279	0.1275
Preference function	Level	Level	Level	Level	Level	Level	Level	Level
Min/Max	MAX	MAX	MAX	MAX	MAX	MAX	MAX	MAX

Based on the results of the MCDA, presented in the Table 6 and the Figures 2 and 3, it can be concluded that employees at both levels of research & development projects (T2P1 and T2P4) are completely satisfied with the success of the project risks management as well as the success of the projects. This was also the opinion for the middle level managers of the civil works construction projects (T3P2).

Conversely, the employees which strongly emphasized the negative outcomes from the project risk management they were involved in as well as the low success of the projects, belong to the operational level of managers at scientific/research projects (T5P3) and middle level managers in public institutions (T4P2). The lowest ranking of the investigated factors was obtained among the ICT projects upon opinion obtained from the operational level managers (T1P3).

The position of the considered alternatives (triangles) identifies the strengths and the weaknesses of the actions determining the criterions (rectangles), as presented in the Figure 3. The closest the alternative lies to the direction of the criterion vector, the alternative is ranked better from the point of this criterion [42]. This way, as presented in the GAIA plane (Figure 3) the criterions Q1, Q2 and G4 have largest influence on the ranking of the alternatives, based on their position towards the *pi* vector – the decision stick (the red vector in the plane). In contrast, the criterions G2 and G6 have the lowest influence on their decisions, based on the position relative to the *pi* vector.

Table 6

Results of complete ranking of the success in the risk management of the strategic investment projects based on the opinions of employees at different project team positions

Rank	Alternatives	Φ^+	Φ^-	Φ
1	T3P2	0.2789	0.0057	0.2733
2	T2P1	0.2679	0.0225	0.2454
3	T1P1	0.2223	0.0171	0.2053
4	T2P4	0.2278	0.0227	0.2051
5	T5P4	0.1480	0.0630	0.0850
6	T6P3	0.1195	0.0857	0.0338
7	T4P4	0.1246	0.1762	-0.0516
8	T5P2	0.0844	0.1660	-0.0816
9	T6P1	0.0620	0.1541	-0.0922
10	T5P3	0.0507	0.1654	-0.1147
11	T4P2	0.0339	0.2834	-0.2494
12	T1P3	0.0000	0.4584	-0.4584

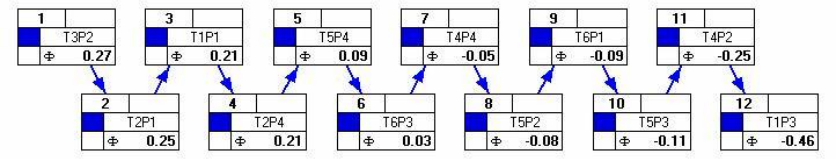


Figure 2

Results of the complete ranking of the success of the strategic investment project risk management, based on the PROMETHEE II methodology

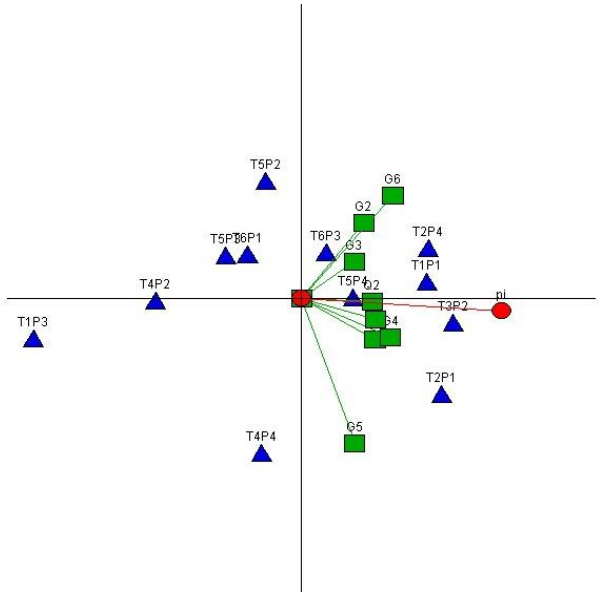


Figure 3
Graphic presentation at GAIA plane for ranking of the success of the risk management of the strategic investment projects

Conclusion

Strategic investment projects represent the very foundation of economic development and the development of Society, in general. Detailed analysis and approaches to the management of every single element of these types of projects is required given their level of significance. The potential risks and their adverse effects, represent a particularly weighty part of that process, especially the possibility to eliminate or minimize them or respond in a suitable way. The statistical analysis of the data collected through a survey of employees in 311 companies and organizations within the area of SEE, dealing with SIP management within their operations, leads to the following conclusions. The initial questions in the questionnaire, as well as the initial hypothetical framework of the research, have been defined based on the analysis of the available literature in the field. Accordingly, six research hypotheses were proposed based on the relation between application of the methods for risk analysis and identified risk factors influencing the SIPs success in SEE.

Conducted factor analysis, according to the question groups in the initial questionnaire, indicated that subgroups should be formed within certain groups, whereas certain questions should be eliminated from the further analysis. The final groups of questions, classified by factors, were used to develop a structural model, used for verification of the previously set hypothetical framework.

Based on the formed structural model for risk analysis of SIPs, it can be concluded that the hypotheses have been verified. The obtained results could be of use for the risk analysis and the strategic project management, connecting the identified project risks with applicable methods for risk analysis, in the real life project management practice. Finally, the obtained results, presented in this paper, represent a solid foundation for the continuation of the research and the development of a final measurement instrument and model of SIP risk management.

We suggest that future research should apply the GLM multivariate procedure and focus on the influence of the most significant demographic characteristics of the projects (project type, project duration, project budget and type of project management), as well as, the different methodologies for risk management of the projects, on the overall SIP success. Also, additional research endeavors should target new and larger samples of respondents.

References

[1] Asrilhant, B., Meadows, M., Dyson, R., Techniques to Support Successful Strategic Project Management in the UK Upstream Oil and Gas Sector, European Management Journal, 24(2-3) (2006) pp. 214-225

[2] Schoemaker, P. J. H., How to Link Strategic Vision to Core Capabilities, Sloan Management Review, 34(1) (1992) pp. 67-72

[3] Buckley, A., International Investment: Value Creation and Appraisal, Copenhagen Business School, Denmark, 1998

[4] Foss, N. J., Resources and Strategy: Problems, Open Issues, and Ways Ahead, In Resources, Firms and Strategies: A Reader in the Resource-based Perspective, ed. N. J. Foss, pp. 345-365. Oxford University Press, Oxford, 1997

[5] Turner, R., Projects and Project Management, In Gower Handbook of Project Management, (eds) J. R. Turner and S. J. Simister (3rd ed.) pp. 65-76, Spring, UK, 2000

[6] Brook, J. W., Pagnanelli, F., Integrating Sustainability into Innovation Project Portfolio Management – A Strategic Perspective, Journal of Engineering and Technology Management, 34 (2014) pp. 46-62

[7] Huang, Y. C., Ma, R., Lee, K. W., Exploitative Learning in Project Teams: Do Cognitive Capability and Strategic Orientations Act as Moderator Variables? International Journal of Project Management, 33 (2015) pp. 760-771

[8] Mooraj, S., Oyon, D., Hostettler, D., The Balanced Scorecard: a Necessary Good or an Unnecessary Evil? European Management Journal, 17(5) (2009) pp. 481-491

[9] Project Management Institute (PMI), A Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Fifth edition, Newtown Square, PA: Author, 2013

[10] Kerzner, H., Project Management: A Systems Approach to Planning, Scheduling and Control, John Wiley & Sons, Inc., Hoboken, New Jersey, 2009

[11] Kumar, M., Gregory, M., An Exploration of Risk Management in Global Industrial Investment, Risk Management, 15(4) (2013) pp. 272-300

[12] Marmier, F., Deniaud, I. F., Gourc, D., Strategic Decision-Making in NPD Projects according to Risk: Application to Satellites Design Projects, Computers in Industry, 65 (2014) pp. 1107-1114

[13] Hussein, B. A., Klakegg, O. J., Measuring the Impact of Risk Factors Associated with Project Success Criteria in Early Phase, Procedia - Social and Behavioral Sciences, 119 (2014) pp. 711-718

[14] Peixoto, J., Tereso, A., Fernandes, G., Almeida, R., Project Risk Management Methodology: A Case Study of an Electric Energy Organization, Procedia Technology, 16 (2014) pp. 1096-1105

[15] Carvalho, M. M., Patah, L. A., Bido, D. S., Project Management and its Effects on Project Success: Cross-Country and Cross-Industry Comparisons, International Journal of Project Management, 33 (2015) pp. 1509-1522

[16] Enescu, M., Enescu, M., Tudorescu, N., Points of View Regarding the Evaluation of Decisional Factors in Risk Management Associated to Large Combustion Plants, Procedia Economics and Finance, 16 (2014) pp. 110-114

[17] Marcelino-Sádaba, S., Pérez-Ezcurdia, A., Lazcano, A. M. E., Villanueva, P., Project Risk Management Methodology for Small Firms, International Journal of Project Management, 32 (2014) pp. 327-340

[18] Ameyaw, E. E., Chan, A. P. C., Evaluation and Ranking of Risk Factors in Public–Private Partnership Water Supply Projects in Developing Countries using Fuzzy Synthetic Evaluation Approach, Expert Systems with Applications, 42 (2015) pp. 5102-5116

[19] Kutsch, E., Hall, M., The Rational Choice of not Applying Project Risk Management in Information Technology Projects, Project Management Journal, 40(3) (2009) pp. 72-81

[20] Espinoza, R. D., Separating Project Risk from the Time Value of Money: A Step toward Integration of Risk Management and Valuation of Infrastructure Investments, International Journal of Project Management, 32 (2014) pp. 1056-1072

[21] Rodrigues-da-Silva, L. H., Crispim, J. A., The Project Risk Management Process, a Preliminary Study, Procedia Technology, 16 (2014) pp. 943-949

[22] Hwang, B.-G., Zhao, X., Gay, M. J. S., Public Private Partnership Projects in Singapore: Factors, Critical Risks and Preferred Risk Allocation from the Perspective of Contractors, International Journal of Project Management, 31 (2013) pp. 424-433

[23] Ihuah, P. W., Kakulu, I. I., Eaton, D., A Review of Critical Project Management Success Factors (CPMSF) for Sustainable Social Housing in Nigeria, International Journal of Sustainable Built Environment, 3 (2014) pp. 62-71

[24] Shrivastava, S. V., Rathod, U., Categorization of Risk Factors for Distributed Agile Projects, Information and Software Technology, 58 (2015) pp. 373-387

[25] Lončar, D., Applicative Model for Appraisal of Investment Projects Based on Real Options Methodology, Serbian Journal of Management, 6(2) (2011) pp. 269-282

[26] Pandremmenou, H., Sirakoulis, K., Blanas, N., Success Factors in the Management of Investment Projects: A Case Study in the Region of Thessaly, Procedia - Social and Behavioral Sciences, 74 (2013) pp. 438-447

[27] Kiliç, M., Kaya, I., Investment Project Evaluation by a Decision Making Methodology Based on Type-2 Fuzzy Sets., Applied Soft Computing, 27 (2015) pp. 399-410

[28] Drury-Grogan, M. L., Performance on Agile Teams: Relating Iteration Objectives and Critical Decisions to Project Management Success Factors, Information and Software Technology, 56 (2014) pp. 506-515

[29] Lappe, M., Spang, K., Investments in Project Management are Profitable: A Case Study-based Analysis of the Relationship between the Costs and Benefits of Project Management, International Journal of Project Management, 32 (2014) pp. 603-612

[30] Todorović, M. Lj., Petrović, D. Č., Mihić, M. M., Obradović, V. Lj., Bushuyev, S. D., Project Success Analysis Framework: A Knowledge-based Approach in Project Management, International Journal of Project Management, 33 (2015) pp. 772-783

[31] Liu, S., Wang, L., Understanding the Impact of Risks on Performance in Internal and Outsourced Information Technology Projects: The Role of Strategic Importance, International Journal of Project Management, 32 (2014) pp. 1494-1510

[32] Junkes, M. B., Tereso, A. P., Afonso, P. S. L. P., The Importance of Risk Assessment in the Context of Investment Project Management: a Case Study, Procedia Computer Science, 64 (2015) pp. 902-910

[33] Davis, K., A Method to Measure Success Dimensions Relating to Individual Stakeholder Groups, International Journal of Project Management, 34 (2016) pp. 480-493

[34] Kaiser, H. F., An Index of Factorial Simplicity, Psychometrika, 39 (1974) pp. 31-36

[35] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., Multivariate Data Analysis, 6th Edition, Pearson Prentice Hall, Upper Saddle River, NJ, 2006

[36] Gorsuch, R. L., Factor Analysis, Hillsdale, NJ: Lawrence Erlbaum, 1983

[37] Cronbach, L. J., Coefficient Alpha and the Internal Structure of Tests, Psychometrika, 16 (1951) pp. 297-334

[38] Nunnally, JM., Psychometric Theory, third ed. McGraw-Hill, New York, 1994

[39] Roy, B., Vincke, P., Multicriteria Analyses: Survey and New Directions, European journal of operational research, 8 (3) (1981) pp. 207-218

[40] Csala, V., Szalay, T., Farkas, B., Markos, S., Application Benchmark of Three Micro Hole Machining Processes for Manufacturing the Nozzle of a Medical Water Jet Machine, Acta Polytechnica Hungarica, 12(2) (2015) pp. 53-69

[41] Živković, Ž., Nikolić, Dj., Djordjević, P., Mihajlović, I., Savić, M., Analytical Network Process in the Framework of SWOT Analysis for Strategic Decision Making (Case Study: Technical Faculty in Bor, University of Belgrade, Serbia), Acta Polytechnica Hungarica, 12(7) (2015) pp. 199-216

[42] Brans, J. P., Mareschal, B., The PROMCALC & GAIA Decision Support System for Multicriteria Decision Aid, Decision Support Systems, 12 (1994) pp. 297-310

[43] Zhi-hong, Z., Yi, Y., Jing-nan, S., Entropy Method for Determination of Weight of Evaluating Indicators in Fuzzy Synthetic Evaluation for Water Quality Assessment, Journal of Environmental Sciences, 18 (2006) pp. 1020-1023

Appendix 1
Results of the descriptive statistics of answers to the questionnaire questions

	N	Range	Mean		Std. Deviation	Var.
	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
1.1. Questionnaire is used as a method for risk analysis and assessment.	311	4	2.97	.143	1.227	1.506
1.2. Sensitivity analysis is used as a method for risk analysis and assessment.	311	3	2.66	.152	1.306	1.706
1.3. Probability analysis is used as a method for risk analysis and assessment.	311	3	2.55	.128	1.100	1.209
1.4. Decision tree is used as a method for risk analysis and assessment.	311	3	2.43	.107	.923	.852
1.5. The simulation method is used as a method for risk analysis and assessment.	311	4	2.28	.118	1.014	1.028
1.6. Brainstorming is used as a method for risk analysis and assessment.	311	4	2.49	.152	1.306	1.705
1.7. The Delphi method is used as a method for risk analysis and assessment.	311	4	2.84	.143	1.228	1.508
1.8. WBS is used as a method for risk analysis and assessment.	311	3	2.08	.129	1.107	1.226
2.1. Potential loss as a risk factor affects the project success.	311	2	3.54	.067	.578	.334
2.2. Market risk as a risk factor affects the project success.	311	4	3.27	.140	1.208	1.460
2.3. Foreign exchange risk as a risk factor affects the project success.	311	2	3.32	.080	.685	.469
2.4. Cost assessment risk as a risk factor affects the project success.	311	3	3.51	.082	.707	.500
2.5. Human resource as a risk factor affects the project success.	311	2	3.89	.083	.713	.509
3.1. The probability of the occurrence of a risk event is a significant factor for risk management.	311	4	3.76	.098	.841	.707
3.2. The frequency of risk events is a significant factor for risk management.	311	4	3.68	.096	.829	.688
3.3. The connection between different risk events is a significant factor for risk management.	311	4	3.55	.104	.894	.798
3.4. Potential loss is a significant factor for risk management.	311	3	3.97	.110	.950	.903
3.5. Timeline risk is a significant factor for risk management.	311	2	3.70	.074	.635	.404
3.6. Cost assessment risk is a significant factor for risk management.	311	3	3.84	.082	.703	.494
3.7. Financing risk is a significant factor for risk management.	311	4	3.73	.101	.865	.748
3.8. Market risk is a significant factor for risk management.	311	4	3.53	.104	.895	.801
3.9. Foreign exchange risk is a significant factor for risk management.	311	4	3.20	.107	.921	.849

3.10. The human factor is a significant factor for risk management.	311	4	3.72	.138	1.188	1.412
4.1. Planned budget is a significant element for project control and evaluation.	311	2	4.27	.078	.668	.447
4.2. Available resources are a significant element for project control and evaluation.	311	2	4.31	.079	.681	.464
4.3. Project team capacity is a significant element for project control and evaluation.	311	2	4.28	.087	.750	.562
4.4. Project manager knowledge and skills are a significant element for project control and evaluation.	311	3	4.19	.096	.822	.676
4.5. Planned time of project implementation is a significant element for project control and evaluation.	311	2	4.16	.072	.620	.384
5.1. WBS is used as a method for SIP* management.	311	4	2.89	.173	1.486	2.207
5.2. Milestone is used as a method for SIP management.	311	4	2.85	.175	1.505	2.265
5.3. Organization methods are used for SIP management.	311	4	3.32	.160	1.376	1.893
5.4. Cost assessment methods are used for SIP management.	311	4	3.68	.153	1.315	1.729
5.5. The Gantt chart is used as a method for SIP management.	311	4	3.58	.140	1.205	1.452
5.6. Network planning techniques are used as methods for SIP management.	311	4	2.85	.140	1.201	1.443
6.1. Client satisfaction is the main criterion for project success measurement.	311	2	4.15	.105	.902	.813
6.2. Planned time of implementation is the main criterion for project success measurement.	311	2	3.81	.079	.676	.457
6.3. Planned budget is the main criterion for project success measurement.	311	3	3.92	.112	.962	.925
6.4. Project quality is the main criterion for project success measurement.	311	4	3.92	.129	1.107	1.226
6.5. The achievement of other objectives is the main criterion for project success measurement.	311	4	3.55	.126	1.087	1.182
6.6. Top management satisfaction is the main criterion for project success measurement.	311	2	3.55	.077	.665	.442
*Q1 Has the methodology used for project risk management provided good results?	311	2	1.32	.110	.952	.907
*Q2 Has SIP management in your organization been successful?	311	2	1.32	.110	.952	.907

* SIP – Strategic Investment Project; ** Questions Q1 and Q2 are questions with a dichotomous character (yes/no questions). The answer YES was ranked with a rating of 2 and NO was ranked with a 0 rating when the results were processed.