ELEKTROTEHNIŠKI VESTNIK 79(1-2): 1–6, 2012
ENGLISH EDITION

Robust 3D Face Recognition

Janez Križaj, Vitomir Štruc, Simon Dobrišek

University of Ljubljana, Faculty of Electrical Engineering, Tržaška 25, 1000 Ljubljana, Slovenia
E-mail: janez.krizaj@fe.uni-lj.si

Abstract. Face recognition in uncontrolled environments is hindered by variations in illumination, pose, expression and occlusions of faces. Many practical face-recognition systems are affected by these variations. One way to increase the robustness to illumination and pose variations is to use 3D facial images. In this paper 3D face-recognition systems are presented. Their structure and operation are described. The robustness of such systems to variations in uncontrolled environments is emphasized. We present some preliminary results of a system developed in our laboratory.

Keywords: biometric systems, face recognition, 3D images, features

1 INTRODUCTION

Systems for the biometric recognition of individuals assess the identity of these individuals on the basis of their physiological or behavioral characteristics, like fingerprint, face, speech, gait and iris patterns. The scope of these systems includes various kinds of access control (border crossing, access to personal information), forensics, law enforcement and others. Face-recognition systems are some of the most popular among all biometric systems. This is mostly due to their non-intrusive nature, as the data acquisition can be performed from a distance, even without the subject's cooperation. Special attention in such biometric systems is being paid to developing the so-called smart surveillance technologies, especially to developing portals for the automatic control of border crossings [1].

Although people recognize faces without any special effort, automatic face recognition with a computer presents a considerable difficulty and challenge, especially if the images are acquired in an uncontrolled environment. The main factors that affect the accuracy of face-recognition systems are the variability in the illumination and the pose of the faces, expressions, occlusions (scarf, beard, glasses), time delay (signs of aging), makeup and similar occurrences during the image-acquisition task. The presence of these factors in the facial images can lead to a diminished recognition reliability.

In order to improve the robustness and reliability of face-recognition systems, various data-acquisition techniques were employed, including video, infra-red images, multiple consecutive shots from different angles and 3D images. The benefits of using 3D images for face recognition encompass the invariance of 3D data to illumination conditions and the ability to rotate the 3D facial data to a normal pose [2]. Despite this, most of the 3D face-recognition systems are affected by facial expression, occlusion and time delay.

This paper first discusses the basic structure and operation of 3D face-recognition systems. Some examples of popular methods used in these systems are also presented - from the 3D image-acquisition techniques to the classification methods. The preliminary results of our own 3D face-recognition system developed in our laboratory are also introduced.

2 STRUCTURE OF 3D FACE-RECOGNITION SYSTEMS

A typical 3D face-recognition system is built from the following units (Fig. 1): an image-acquisition and pre-processing unit, a feature-extraction unit and a similarity-measure and classification unit. In the subsequent sections, the units are presented in detail with examples of implementations that have emerged in the literature.

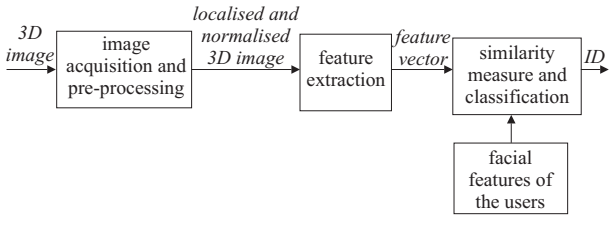


Figure 1. Block diagram of a 3D face-recognition system.

Received Januar 12, 2012
Accepted Februar 2, 2012

3 IMAGE ACQUISITION AND PRE-PROCESSING

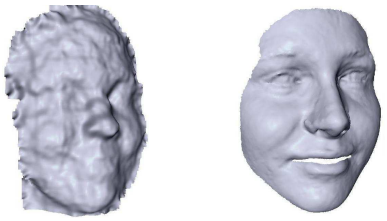
It is assumed that the representation of faces with 3D images has several advantages over 2D images. However, advantages like invariance to illumination and robustness to the rotation of faces in 3D space do not hold completely in reality [3]. Rotation and scale normalization can be computationally quite expensive and the existing methods are not always convergent. Similarly, the fact that the 3D data is illumination invariant is not always true - strong light sources and reflective surfaces can significantly affect the 3D sensor reading. Therefore, the raw 3D images usually contain some degree of noise, which can be recognized as spikes (reflective regions - oily skin) and holes (missing data in the transparent regions - eyes, eyebrows, hair, beard) in the scanned image. Examples of the above distortions are given in Fig. 2.

(a) (b) (c)

Figure 2. Examples of scans with distortions: (a) nose absence, (b) noise around the eye region, (c) distortion due to face motion during image acquisition.

3.1 3D sensors

The current 3D sensors fall into two categories, i.e., active sensors and passive sensors, of which active sensors are more suitable for the 3D face-recognition task. The passive 3D sensors reconstruct the 3D images indirectly from the 2D images or video. The active 3D sensors or scanners use laser light or structured light patterns to capture the 3D data. Among the active 3D sensors used to capture face scans structured-light, the 3D scanners are the most popular. These scanners use projected light patterns and a camera to measure the 3D shape of an object. The projector emits multiple light stripes (usually infra-red light or laser light) onto a 3D-shaped surface, while the camera acquires the light patterns that are distorted from other perspectives than that of the projector. The 3D shape of an object can be reconstructed from the differences between the projected and acquired patterns. This technique is used in a large number of commercially available 3D sensor systems: Konica Minolta Range 5 / Vivid9i / Vivid910 (examples of scans acquired with Konica Minolta Vivid910 can be seen in Fig. 2), Cyberware PX, 3DFaceCam, FaceSCAN, FastSCAN, IVP Ranger M50. Recently, some low-cost alternatives for the 3D data acquisition have emerged in the market, such as the Microsoft Kinect sensor and Asus Xtion PRO. Although these sensors have numerous limitations, such as a low depth resolution and depth of field, it is possible to obtain an adequate 3D model of an object with a suitable pre-processing step. A representative example is the method in [4], where a reconstructed 3D face model (Fig. 3b) can be obtained with an iterative adaptation of the average face model to the 3D scan acquired by the Kinect sensor (Fig. 3a).



(a) (b)

Figure 3. 3D face image acquired by the Kinect sensor: (a) raw image, (b) adapted 3D face model.

3.2 3D image pre-processing

The output of a common 3D sensor is a set of 3D points of a scanned surface, with the values of the x, y and z components at each point. The 3D data is usually presented as a point cloud (Fig. 4(a)) or a range image (Fig. 4b). The point cloud is a set of (x, y, z) coordinates of scanned points from the object surface. A range image (or a depth image) can be obtained by the projection of scanned 3D points onto the (x, y) plane. Therefore, the range image is formatted in a similar way to a 2D intensity image, but with the difference that in the range image the pixel intensities are proportional to the depth components of a scanned object (z coordinates).

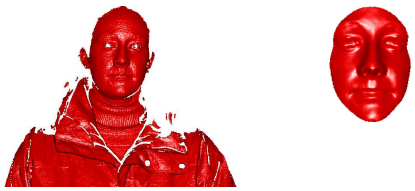
(a) (b) (c)

Figure 4. 3D data representation: (a) point cloud, (b) depth image, (c) shaded depth image.

In the acquired 3D image, the face detection and localization are usually performed first. Detection denotes a process where the presence and the number of faces in the image are determined. Assuming that the image contains only one face, the localization task is to find the location and size (and sometimes also the orientation) of a facial region. Most methods for face localization in 3D images are based on an analysis of the local curvedness

of the facial surface [5–7]. This gives us a set of possible points for the locations of the characteristic facial parts, such as the location of the nose, eyes and mouth, through which the exact location, size and orientation of the facial area can be determined. Based on the locations of these points, the face area can be cut from the rest of the image and eventually re-scaled and rotated to the normal pose (Fig. 5).

Since the raw images generally contain some degree of noise, the images are usually filtered before subsequent processing. Usually, low-pass filters are used to filter out high-frequency noise (spikes), while the missing data are substituted by the interpolation of adjacent points on the facial surface [5, 8].

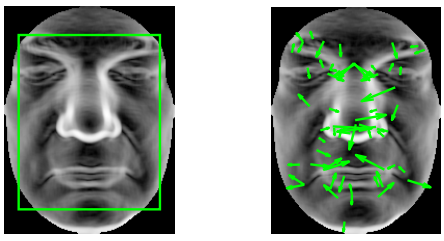


(a) (b)

Figure 5. 3D image pre-processing: (a) raw image, (b) localized and filtered facial region.

4 Feature Extraction

The purpose of feature extraction is to extract the compact information from the images that is relevant for distinguishing between the face images of different people and stable in terms of the photometric and geometric variations in the images. One or more feature vectors are extracted from the facial region. Depending on how the facial region is treated, the existing feature-extraction methods can be divided into the groups described below.



(a) (b)

Figure 6. Conceptual presentation of feature types: (a) global features, (b) local features.

4.1 Global-feature extraction methods

These methods extract the feature vector from the whole face region (Fig. 6a). The majority of the global 3D facial-feature-extraction methods have been derived from methods originally used on 2D facial images, where 2D gray-scale images are replaced by range images. *Principal component analysis* (PCA) is the most widespread method for global-feature extraction. PCA was first used for feature extraction from 2D face images and later also for feature extraction from range images [9]. Other popular global-feature extraction methods, such as *linear discriminant analysis* (LDA) [10] and *independent component analysis* (ICA) [11], were also used on range images.

The advantages of global features are: a considerable reduction of the data dimensionality, and the spatial relationship among the different parts of the face is retained (in the case of local features this information is generally lost). The main disadvantage of global-feature methods is that these methods require the precise localization and normalization of the orientation, scale and illumination. Changes in these factors can affect the global facial features, resulting in a decreased recognition rate. In global-feature-based recognition systems, localization and normalization are often performed by the manual labeling of characteristic points on the face (which makes the whole process semi-automatic). Automatic localization and normalization is generally achieved using the *iterative closest-point* algorithm (ICP) [12]. However, the ICP algorithm is computationally expensive and does not always converge to a global maximum. The global-feature-based approaches are normally also sensitive to facial expression and occlusion.

4.2 Local-feature extraction methods

Local-feature extraction methods extract a set of feature vectors from a face, where each vector holds the characteristics of a particular facial region. The use of global features is prevalent in face-recognition systems based on images acquired in a controlled environment. The local features are at an advantage over the global features in uncontrolled environments, where the variations in facial illumination, rotation, expressions and scale are present, since a local analysis of facial parts provides a better basis to deal with such variations.

The process of local-feature extraction can be divided into two parts. In the first part, the interest points on the face region are detected. In the second part, the interest points are used as locations at which the local feature vectors are calculated.

There are several methods for interest-points detection. The interest points can be detected as extrema in the scale-space, resulting in the invariance of features to the scale. This approach of interest-points detection is used in the *scale-invariant feature transform* (SIFT) [13, 14] and the *speeded-up robust features* (SURF) [15]. The interest points can also be detected as follows: on the basis of the local curvedness analysis [16]; by the alignment of faces with a face model in which interest-point locations are marked *a priori* [17]; by the *elastic bunch graph method* (EBGM) [18] as nodes of the elastic graph; and as nodes of a rectangular grid covering

the facial region [8, 19, 20]. The latter approach is equivalent to detecting the local features on a block basis, where the feature vectors are extracted by the sliding-block technique.

Interest-point detection is followed by the extraction of local feature vectors at the locations of the interest points. In the earlier approaches, local features were generally defined from the geometric relations among the interest points (location of the points, distances and angles between the points, distance ratios, geodesic distances). For the description of the local surface around the interest points, the latter approaches normally use: differential geometry descriptors (mean curvature, Gaussian curvature, shape index) [8, 21], point signatures [22], Gabor filters [18], *coefficients of discrete cosine transform* (DCT) [19, 20] and orientation histograms [13].

Local features have several advantages over global features. Due to the nature of the local-feature-based approaches, the recognition performance is less affected by the imprecise face localization and normalization than in the case of global features. Therefore, some local-feature-based approaches do not require the normalization of illumination, rotation and scale variations. The local approaches are also less sensitive to expression variations.

4.3 Hybrid-feature extraction methods

Hybrid-feature-based approaches use both of the above-mentioned feature types. Data from the global and local approaches can be fused at the feature level [23], similarity level [24] or decision level [25].

5 SIMILARITY MEASURE AND CLASSIFICATION

The last step of the 3D face-recognition process presents a similarity measure between the test face and the faces from the system's database. The decision that follows depends on the purpose of the recognition system. Recognition systems can be divided according to their purpose into verification systems and identification systems. Verification systems validate the identity claim of the test person by comparing their features to the template associated with the claimed identity. If the similarity with the template is high enough, the system recognizes the person as a client, or if the similarity is too low, the person is recognized as an impostor. The identification system conducts a one-to-many comparison by searching for the maximum similarity between the test-person features and all the templates stored in the database. The output of the identification system is the person's identity or the answer that the person does not fit to any model from the database.

In the case of global-feature-based approaches, where each face is represented by one feature vector, the similarity between two faces is defined on the basis of a certain distance measure (L_1 norm, L_2 norm, cosine distance, Mahalanobis distance) between the feature representations of these two faces. Local-feature approaches may require a different similarity-measure procedure, since each face is generally represented by a *variable* number of local-feature vectors, and therefore we do not have a unified representation of faces. When comparing two faces represented by the local features, we usually also do not know which local-feature vectors belong to the same face regions of the two faces. For the reasons outlined above, the unified encoding of faces represented by local features is often utilized with the parameters of the *Gaussian mixture model* (GMM) [26] or the *hidden Markov model* (HMM) [19, 20]. The comparison between two faces represented by local features can also be performed directly by comparing each local-feature vector of one face to each local-feature vector of the other face. In this case, the similarity measure is based on the number of the matched feature vectors or on the sum of the distance measures among the matching feature vectors.

The classification process normally utilizes the result of the similarity measure. The most common and the simplest classification technique is the *nearest-neighbor* classifier (1-NN), while other popular classification techniques include *support vector machine* (SVM) [27, 28] and *likelihood ratio* [20, 26].

6 SYSTEM IMPLEMENTATION AND EXPERIMENTS

We implemented a 3D face-verification system based on the local features and GMM models (see Fig. 7). This system represents a popular procedure used for 2D face recognition as well as for 3D face recognition. The system will serve as a starting point for further research in which we will try to justify the applicability of local features in robust 3D face recognition. The system performance was compared to the global PCA method.

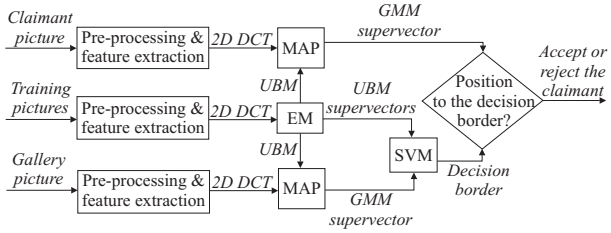


Figure 7. Block diagram of the implemented system.

In the pre-processing step, a low-pass filter was used to remove the high-frequency noise, while the missing data were compensated by the interpolation of adjacent points on the facial surface. Only a rough automatic face

localization was performed, based on the work in [16]. The local-feature vectors were extracted on a block-by-block basis. From each block, 2D DCT coefficients were calculated and the first ten low-frequency coefficients were used to build the feature vector. The distribution of feature vectors from each face image was described by the GMM model. A GMM model, consisting of K Gaussian components, is defined by the following parameters: weights $\{\pi_k\}_{k=1}^K$, mean vectors $\{\boldsymbol{\mu}_k\}_{k=1}^K$ and covariance matrices $\{\boldsymbol{\Sigma}_k\}_{k=1}^K$. The *expectation maximization* algorithm (EM) [29] was utilized to set these parameters. Due to the small number of feature vectors for each face, the parameters of the so-called *universal background model* (UBM) were determined first. The UBM is a GMM trained on all the feature vectors from the training set. The GMM of each person was adapted from the UBM by the *maximum a posteriori* estimation (MAP) [30], where only the mean vectors were adapted. GMM-based verification systems normally use the *likelihood ratio* test for the classification task, while in our system the SVM-based classifier is employed. An unified representation of the images has to be made to utilize the SVM classification. For this purpose, the mean vectors from each face were stacked one over the other to form the so-called *supervector* for each face. In the enrollment phase, when the person is introduced to the system, the SVM constructs the decision border between the supervector of the enrolled person and the supervectors from all the training images. In the test phase, the claimant is accepted or rejected with respect to the position of the claimant's supervector to the decision border.

The experiments were performed on the *face recognition grand challenge* (FRGC) data set [9]. Fig. 8 shows the verification performance in the form of a *receiver operating characteristic* curve or ROC. This curve plots the *false-acceptance rate* against the *true-acceptance rate* for all possible operating points, i.e., thresholds.

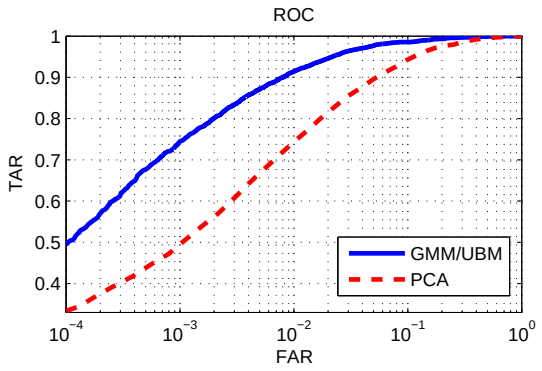


Figure 8. ROC curve of the proposed system compared to the global PCA approach.

The presented method expectedly outperforms the PCA approach. This stems mainly from the fact that in the PCA approach, global-feature vectors are used, while the GMM models utilize the local-feature vectors, resulting in an improved robustness to imprecise localization, expression variations and occlusions. These factors are comprised in most of the images from the FRGC data set employed in our experiments.

7 Conclusion

In this paper we present some of the 3D face-recognition systems. The whole recognition process is described from the image-acquisition stage to the classification task. Operation of such systems in an uncontrolled environment is highlighted, where the recognition performance can be affected by numerous variations during the image acquisition. The recognition systems based on local features are generally more robust to these variations, as can also be seen from the results of our experiments.

Acknowledgement

The research leading to the above results has received funding from the European Union's Seventh Framework Programme (FP7-SEC-2010-1) under grant agreement number 261727 and the bilateral programme named Fast and reliable 3D Face Recognition (BI-BG/11-12-007).

References

[1] C. Busch and A. Nouak, "3d face recognition for unattended border control." in *Security and Management'08*, 2008, pp. 350–356.

[2] G. Medioni and R. Waupotitsch, "Face modeling and recognition in 3-d," in *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, ser. AMFG '03, 2003, pp. 232–240.

[3] S. Huq, B. Abidi, S. G. Kong and M. Abidi, "A survey on 3d modeling of human faces for face recognition," in *3D Imaging for Safety and Security*, ser. Computational Imaging and Vision, 2007, vol. 35, pp. 25–67.

[4] M. Zollhöfer *et al.*, "Automatic reconstruction of personalized avatars from 3d face scans," *Computer Animation and Virtual Worlds (Proceedings of CASA 2011)*, vol. 22, pp. 195–202, 2011.

[5] A. Mian, M. Bennamoun and R. Owens, "Face recognition using 2d and 3d multimodal local features," *ISVC '06*, vol. 860, pp. 860–870, 2006.

[6] A. Colombo, C. Cusano and R. Schettini, "3d face detection using curvature analysis," *Pattern Recogn.*, vol. 39, pp. 444–455, 2006.

[7] S. Mehryar, K. Martin, K. Plataniotis and S. Stergiopoulos, "Automatic landmark detection for 3d face image processing," in *CEC '10*, 2010, pp. 1–7.

[8] T. Inan and U. Halici, "3-d face recognition with local shape descriptors," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 577–587, 2012.

[9] P. J. Phillips *et al.*, "Overview of the face recognition grand challenge," in *CVPR '05*, 2005, pp. 947–954.

[10] T. Heseltine, N. Pears and J. Austin, "Three-dimensional face recognition: A fishersurface approach," in *Image*

Analysis and Recognition, ser. LNCS, 2004, vol. 3212, pp. 684–691.

[11] C. Hesher, A. Srivastava and G. Erlebacher, "A novel technique for face recognition using range imaging," in *Proc. Seventh Int Signal Processing and Its Applications Symp*, vol. 2, 2003, pp. 201–204.

[12] B. Amor, M. Ardabilian and L. Chen, "New experiments on icp-based 3d face recognition and authentication," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3, 2006, pp. 1195–1199.

[13] T. R. Lo and J. P. Siebert, "Local feature extraction and matching on range images: 2.5d sift," *Comput. Vis. Image Underst.*, vol. 113, pp. 1235–1250, 2009.

[14] C. Maes *et al.*, "Feature detection on 3d face surfaces for pose normalisation and recognition," in *BTAS '10*, 2010, pp. 1–6.

[15] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision and Image Understanding (CVIU)*, 2008, pp. 346–359.

[16] P. Szeptycki, M. Ardabilian and L. Chen, "A coarse-to-fine curvature analysis-based rotation invariant 3d face landmarking," in *BTAS '09*, 2009, pp. 32–37.

[17] M. O. Irfanoglu, B. Gokberk and L. Akarun, "3d shape-based face recognition using automatically registered facial surfaces," in *ICPR'04*, 2004, pp. 183–186.

[18] M. Husken, M. Brauckmann, S. Gehlen and C. Von der Malsburg, "Strategies and benefits of fusion of 2d and 3d face recognition," in *CVPR'05*, 2005, pp. 174–182.

[19] F. Cardinaux, C. Sanderson and S. Bengio, "User authentication via adapted statistical models of face images," *Signal Processing, IEEE Transactions on*, vol. 54, pp. 361–373, 2006.

[20] C. McCool, J. Sanchez-Riera and S. Marcel, "Feature distribution modelling techniques for 3d face verification," *Pattern Recogn. Lett.*, vol. 31, pp. 1324–1330, 2010.

[21] N. Alyüz, B. Gökberk and L. Akarun, "Regional registration and curvature descriptors for expression resistant 3d face recognition," in *Signal Processing and Communications Applications Conference, 2009. SIU 2009. IEEE 17th*, 2009, pp. 544–547.

[22] Y. Wang, C. S. Chua and Y. K. Ho, "Facial feature detection and face recognition from 2d and 3d images," *Pattern Recognition Letters*, vol. 23, no. 10, pp. 1191–1202, 2002.

[23] F. Al-Osaimi, M. Bennamoun and A. Mian, "Integration of local and global geometrical cues for 3d face recognition," *Pattern Recognition*, vol. 41, no. 3, pp. 1030–1040, 2008.

[24] J.-P. Vandeborre, V. Couillet and M. Daoudi, "A practical approach for 3d model indexing by combining local and global invariants," in *3D Data Processing Visualization and Transmission. Proceedings*, 2002, pp. 644–647.

[25] B. Gökberk, A. Salah and L. Akarun, "Rank-based decision fusion for 3d shape-based face recognition," in *Audio- and Video-Based Biometric Person Authentication*, ser. LNCS, 2005, vol. 3546, pp. 1019–1028.

[26] C. McCool, V. Chandran, S. Sridharan and C. Fookes, "3d face verification using a free-parts approach," *Pattern Recogn. Lett.*, vol. 29, no. 9, pp. 1190–1196, 2008.

[27] H. Bredin, N. Dehak and G. Chollet, "Gmm-based svm for face recognition," in *ICPR '06*, vol. 3, 2006, pp. 1111 –1114.

[28] A. Moreno, A. Sanchez, J. Velez and J. Diaz, "Face recognition using 3d local geometrical features: Pca vs. svm," in *ISPA '05*, 2005, pp. 185–190.

[29] T. Moon, "The expectation-maximization algorithm," *Signal Processing Magazine, IEEE*, vol. 13, pp. 47–60, 1996.

[30] D. A. Reynolds, T. F. Quatieri and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," in *Digital Signal Processing*, 2000, pp. 19–41.

Janez Križaj received his B.Sc. degree in 2008 from the Faculty of Electrical Engineering of the University of Ljubljana. Currently he is a Ph.D. student and junior researcher at the Laboratory of Artificial Perception, Systems and Cybernetics at the same faculty. His research interests include computer vision, image processing and pattern recognition.

Vitomir Štruc received his B.Sc. and Ph.D. degrees in electrical engineering from the University of Ljubljana in 2005 and 2010, respectively. He is currently working as a PostDoc at the Laboratory of Artificial Perception, Systems and Cybernetics at the same faculty. His research interests include pattern recognition, machine learning and biometrics.

Simon Dobrišek received his B.Sc., M.Sc. and Ph.D. degrees in electrical engineering in 1990, 1994 and 2001, respectively, from the Faculty of Electrical Engineering of the University of Ljubljana. In 1990 he became a Research Staff Member in the Laboratory of Artificial Perception, Systems and Cybernetics at the same faculty, where he is presently a teaching assistant and research fellow. His research interests include pattern recognition and artificial intelligence. He participates in research projects on spoken language technology and biometric security systems.