Journal of ELECTRICAL ENGINEERING, VOL. 58, NO. 3, 2007, 134–139

BASIS OF EIGENFACES FOR TRACKING OF HUMAN HEAD

Ján Mihalík — Miroslav Kasár *

The paper deals with the construction of the basis of eigenfaces from a training set of face images using the wireframe 3D model Candide of the human head. The wireframe 3D model is adapted to the face images from the training set, because in general they are obtained from different sources. Afterwards the face images with the adapted model are geometrically normalized to remove texture variations caused by their global and local motion and geometrical differences between individuals. As a result there are more similar geometrically normalized textures of the human faces than before the normalization process. To get completely pre-processed textures of the human faces after their geometrical normalization they are centred and energetically normalized to minimize the effect of global lighting variation. By applying the principal component analysis to these completely pre-processed textures the basis of eigenfaces is obtained. Finally the basis of eigenfaces combined with an algorithm of adapting of 3D model is applied on tracking of the human head in head-and-shoulder videosequence.

K e y w o r d s: eigenfaces, 3D model, Candide, texture, shape, principal component analysis, tracking system

1 INTRODUCTION

Most of the methods of image coding achieve data compression by means of employing the intraframe or interframe correlation [1]. For coding of videosequences at very low bit rates, the model-based coding [2] gives a higher efficiency. In the model-based coder, the human head in the videosequence is described by a 3D head model that has to track its moving. There are two basic classes of tracking systems: attribute-based (or feature-based) and template-based (or example-based) systems.

Using the attribute-based system [3, 4], specific attributes like elliptic shape, certain colours or facial features are searched for in the face image. An example of this is the tracking system which uses the Smallest Univalue Segment Assimilating Nucleus (SUSAN) algorithm for extracting the important facial features. The first step consists of the establishment of rectangular search regions for single facial features. Then the facial features are extracted by applying the SUSAN algorithm [5] inside the rectangular search regions. In the final step the head model is adapted using the extracted facial features. Such systems are usually quite fast, but less robust.

Using the template-based system [6, 7], objects similar to the example (face) are searched for in the face image. An example of this is the tracking system using the basis of eigenfaces (BEF) to track the human head that will be described in this paper. Template-based systems tend to be more robust but also more computationally demanding than attribute-based systems.

For construction of BEF, face images with an adapted wireframe 3D model are required. The face images with adapted models have been normalized (geometrically, energetically) and then the principal component analysis (PCA) was applied to the completely pre-processed textures. The combination of BEF with an algorithm for adapting of 3D model to face image creates the tracking system of the human head.

In section 2 the wireframe 3D model Candide and its parameterization are described. Geometrical normalization of the human face is described in section 3 and in section 4 the basis of eigenfaces and their construction is presented. In section 5 application of BEF on tracking of the human head is presented.

2 CANDIDE AND ITS PARAMETERIZATION

Candide (Fig. 1) is a wireframe 3D model [8] which contains 113 vertices and 184 triangles (polygons). The geometry of 3D model Candide is parameterized according to

$$V = \overline{V} + zS + wA \tag{1}$$

where the resulting vector V contains (h, v, r) coordinates of vertices of the model. Vector $\overline{V}$ is the standard shape of the model, and the columns of S and A matrices are the shape and animation units respectively. Finally z and w are the shape and animation parameters.

To perform the global motion of the model six more parameters for rotation, scaling and translation are needed. Thus (1) can be replaced by

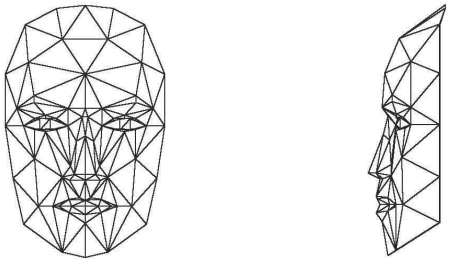
$$V = sR(\overline{V} + zS + wA) + T \tag{2}$$

where $R = R(\Theta_h, \Theta_v, \Theta_r)$ is the rotation matrix, s is a scale, and $T = T(t_h, t_v)$ is 2D translation vector. The geometry and motion of the model are thus parameterized by the parameter vector

$$p = [\Theta_h, \Theta_v, \Theta_r, s, t_h, t_v z, w]^\top . \tag{3}$$

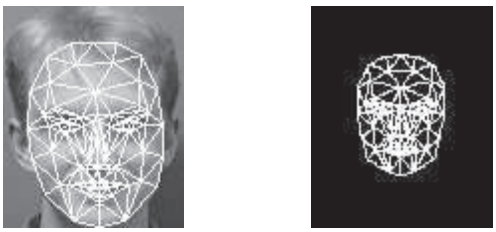
* Laboratory of Digital Image Processing and Videocommunication, Department of Electronics and Multimedia Communication, FEI TU Košice Park Komenského 13, 041 20 Košice, Slovakia E-mail: Jan.Mihalik@tuke.sk, Miroslav.Kasar@tuke.sk

ISSN 1335-3632 © 2007 FEI STU



a) b)

Fig. 1. 3D model Candide a) front view, b) profile.



a) b)

Fig. 2. (from left) Adapted model, standard shape.

a) b) c)

Fig. 3. (from left) Adapted model, standard shape with texture, geometrically normalized texture.

a)

b)

c)

Fig. 4. (top) Original images of human faces, (middle) geometrically normalized textures, (bottom) relevant parts of normalized textures.

The texture x mapped on the surface of the wireframe 3D model [9] is a standard-shaped image, being a linear combination of the basis of eigenfaces. This is formulated as

$$x = \overline{x} + U_x b_x \tag{4}$$

where $\overline{x}$ is the mean texture, the columns of matrix U_x are the eigenfaces and b_x is the vector of texture coefficients.

3 GEOMETRICAL NORMALIZATION OF HUMAN FACE

Geometrical normalization of the human face used to obtain its normalized texture removes texture variations caused by its global and local motion and geometrical differences between individuals. Geometrical normalization is a non-linear transformation [10] which warps the source image of the human face with the adapted model Candide to the standard shape of the model Candide with a given scale (Fig. 2).

Texture mapping on the standard shape is performed by the following algorithm. First, some terms are defined:

- Polygonal mesh (M) of the standard shape (destination mesh) contains triangles $M_0 \dots M_N$. Each triangle is a triplet of 2D vertex coordinates

$$M = \begin{bmatrix} h_1 & v_1 \\ h_2 & v_2 \\ h_3 & v_3 \end{bmatrix} . \tag{5}$$

- Polygonal mesh (M') of the adapted model (source mesh).
- Destination textures $f(h, v)$ corresponding to the standard shape.
- Source textures $f'(h', v')$ corresponding to the adapted model.

Then the algorithm of texture mapping is:

1. In triangle M, all points are stepped using the scan-line algorithm [11]. Barycentric coordinates [12] are calculated for each point coordinates (h, v) of the triangle from the destination mesh of the standard shape. That means, a, b, c need to be found

$$(h, v) = (a, b, c)M . \tag{6}$$

They are valid if (and only if)

$$\begin{aligned} 0 \leq a \leq 1 , \\ 0 \leq b \leq 1 , \\ 0 \leq c \leq 1 . \end{aligned} \tag{7}$$

2. Using barycentric coordinates the source coordinates (h', v') are computed for the corresponding triangle in the source mesh

$$(h', v') = (a, b, c)M' . \tag{8}$$

3. Interpolation of the destination texture $f(h, v)$ according to the relations (6) and (8) is calculated

$$f(h, v) = f'(h', v') . \tag{9}$$

This approach is repeated for all triangles of the standard shape. Results of texture mapping are illustrated in Fig. 3.

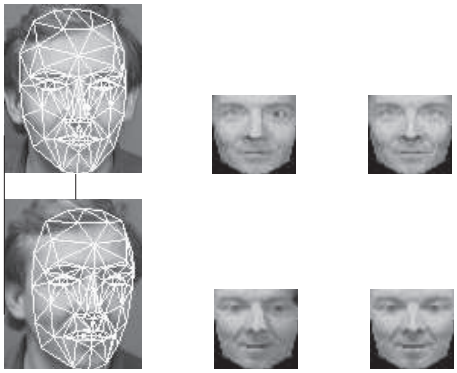


Fig. 5. (from left) Input images with adapted model, geometrically normalized textures,and geometrically normalized textures using texture symmetry.

Figure 4 displays 3 different original images of human faces (top) and 3 geometrically normalized textures appertaining to them (middle). Note that the faces are more similar to each other after the normalization process than before.

After obtaining the normalized texture it is important to select its relevant part which contains the most important features of the human face (eyes, lip, nose ...). Figure 4 displays in the bottom the relevant parts of normalized textures of the human faces.

One reason of obtaining an error normalized texture using geometrical normalization of the human face is its big turn in the source image. In order to eliminate this error, texture symmetry is exploited. This is applied to the normalized textures in which the turn is bigger than a given angle. Accurately adapted one half of the normalized texture is then expanded to the other half specularly. After this expansion the acquired geometrical normalized texture is more accurate because the error caused by the big turn of the human face is removed.

Figure 5 displays two geometrically normalized textures (relevant parts). The first texture is obtained from the frontal view (top), the second texture is obtained from the non-frontal view (bottom) of the human face.

From Fig. 5 it is possible to see that using the texture symmetry for the frontal view is not visible, therefore it is not exploited in this case. The texture symmetry for the non-frontal view gives a more accurate texture, the black place (right top on the geometrically normalized texture) is removed.

4 BASIS OF EIGENFACES

Before applying PCA to the relevant parts of geometrically normalized textures it is necessary to minimize the effect of global lighting variation. To minimize the effect, geometrically normalized texture y is centered (the sum of texture elements y_i is zero) and energetically normalized (the variance of texture elements y_i is unity) according to the relationship

$$x = \frac{y - \alpha}{\sigma} \tag{10}$$

where the vector x is the completely pre-processed texture independent on the geometry of the human face and lighting. Then the values of α and σ are given by

$$\alpha = \frac{1}{n} \sum_{i=1}^{n} y_i \,, \tag{11}$$

$$\sigma = \sqrt{\sum_{i=1}^{n} (y_i - \alpha)^2} \,. \tag{12}$$

By applying PCA [13] to the training set of the completely pre-processed textures x the basis of eigenfaces is obtained.

In the first step of PCA, the covariance matrix C of the input matrix $F = [x_1\, x_2\, \ldots\, x_M]$, $x_j = [x_{ij}]$, $I = 1, \ldots, N$, $j = 1, \ldots, M$, must be found. The input matrix F represents the training set and consists of M N-dimensional column vectors of the completely pre-processed textures. The textures consisting of R rows and K columns produce the column vectors x_j which consist of $N = R \times K$ elements. Then their covariance matrix can be obtained from the following relation

$$C = BB^{\top} \tag{13}$$

where the columns of matrix B are the vectors b_j that differ from the vectors x_j by the expected value m_F of the columns of matrix F. Then

$$B = [x_1 - m_F\, x_2 - m_F\, \ldots\, x_M - m_F] \,, \tag{14}$$

$$m_F = \frac{1}{M} \sum_{i=1}^{M} x_i \,. \tag{15}$$

If (13) is developed using (14) and (15), the symmetric covariance matrix is obtained

$$C = \begin{bmatrix} c_1^2 & c_{12} & \ldots & c_{1N} \\ c_{12} & c_2^2 & \ldots & c_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1N} & c_{2N} & \ldots & c_{NN} \end{bmatrix} \tag{16}$$

where c_i^2 is the variance of i^{th} texture element (tel) and c_{ij}, $i \neq j$, is the covariance between i^{th} and j^{th} tels. The eigenvectors of the covariance matrix C can be calculated from the following equation

$$Cu_i = \beta_i u_i \tag{17}$$

where u_i are the eigenvectors (eigenfaces) and β_i are the eigenvalues of matrix C, $i = 1, \ldots, N$. The eigenvectors u_i create the columns of matrix $U = [u_1\, u_2\, \ldots\, u_N]$.

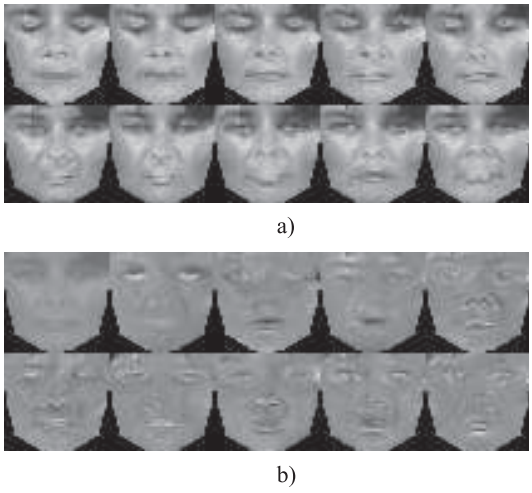


Fig. 6. (top) Set of 10 completely pre-processed textures, (bottom) basis of 10 eigenfaces.

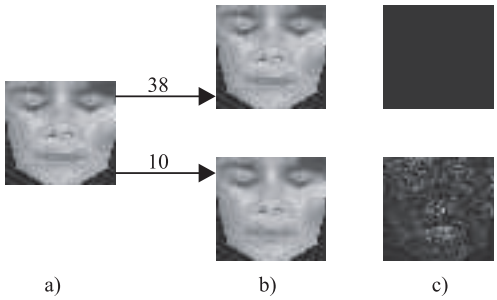


Fig. 7. Completely pre-processed texture, synthesized textures, residual images (5 times enhanced).

Fig. 8. (from left) Input image with well (top) and wrong (bottom) adapted model, the image mapped onto the model, the texture from input frame $j(p)$, synthesized texture $x(p)$, residual image $r(p)$.

From (17) it is visible that the calculation of the eigenvectors involves operations on the covariance matrix C. Even if small completely pre-processed textures are used for training, the size of covariance matrix can be too large to handle it by common computing equipment. If the number M of the textures in the matrix F is considerably smaller than the dimension N of their vectors, it is possible to reduce the computational effort by application of the singular value decomposition (SVD) [14].

SVD allows to express the eigenvectors (eigenfaces) of the matrix $C = BB^\top$ as the linear combination of the eigenvectors of the matrix $H = B^\top B$. The eigenvectors of matrix H can be calculated as follows

$$Hg_i = \lambda_i g_I \tag{18}$$

where g_i are the eigenvectors and λ_i are the eigenvalues of the matrix H, $i = 1, \ldots, M$. The eigenvectors g_i create the columns of matrix $G = [g_1\ g_2\ \ldots\ g_M]$. Then the eigenvectors u_i can be obtained as

$$u_i = \frac{1}{\sqrt{\lambda_i}} B g_i \tag{19}$$

where $i = 1, \ldots, M$. Matrix U has N eigenvectors and values. However since it is formed from only M input vectors, the size of U is less than or equal to M. Then only M of its eigenvectors have nonzero eigenvalues. Computation of the eigenvectors from the $N \times N$ matrix C reduces to computation of the eigenvectors from the $M \times M$ matrix H.

Figure 6 displays 10 completely pre-processed textures from the training set of 38 human faces. The training set was obtained by decimation with the factor 4 of the videosequence "Miss America". In the same figure the basis of 10 eigenfaces corresponding to the largest eigenvalues and calculated by the above procedure is shown.

Figure 7 shows the completely pre-processed texture for which the forward and inverse transform by using the designed BEF is applied. The synthesized textures are obtained, when the number of eigenvectors (eigenfaces) in the transform matrix U is 38 (top) or 10 (bottom). While for 38 eigenfaces the residual image is zero for the other numbers it is not valid. However already for 10 eigenfaces very high accuracy of the synthesized texture is obtained with the summed squared error 56.

5 TRACKING OF HUMAN HEAD

The tracking system of the human head is based on BEF combined with an algorithm for adapting the model to the face image (frame). Adaptation to the image assumes that reasonable starting approximation is known. Then the optimal adaptation of the model to the frame of the videosequence with the human head means to find the parameter vector p that minimizes the distance between the model and the frame. As the initial vector for the actual frame the parameter vector p is used that adapts the model in the previous frame of the videosequence, assuming that the motion from one frame to another one is small enough. The summed squared error (SSE) between the synthesized texture and the texture from the input frame is chosen as distance measure. The texture from the input frame is acquired by the same way as at the construction of BEF. It will be denoted as $j(p)$ because it responds to the parameter vector p. The synthesized texture will be denoted as $x(p)$ and it is acquired from the texture $j(p)$ by BEF. Residual image is given as

$$r(p) = j(p) - x(p)\,, \tag{20}$$

Table 1. SSE in dependence on translation of the model in the input frame.

$t_v \downarrow$ / $\rightarrow t_h$	−5	−4	−3	−2	−1	0	1	2	3	4	5
−5	668					469					726
−4						367					
−3						287					
−2				271		199					
−1						95					
0	481	384	250	146	79	0	80	183	308	429	519
1						133					
2						303					
3						477					
4						595					
5	839					685					865

Table 2. SSE for variation of rotation angle and translation vector T of the model in the input frame.

$T = T(t_h m t_v)$		(0, 0)	(−2, −2)	(5, 5)
Θ_r (degrees)	−20	873	941	881
	−15	723	783	854
	−10	543	589	816
	−5	273	370	802
	0	0	271	865
	5	246	332	954
	10	592	546	1019
	15	912	806	1162
	20	1117	1000	1218

and the SSE

$$e(p) = \|r(p)\|^2. \tag{21}$$

Figure 8 displays single textures $j(p)$, $x(p)$ and residual images $r(p)$ for input images with well and wrong adapted model. There is visible that textures $j(p)$ and $x(p)$ are more different for wrong adapted model, therefore SSE of the residual image $r(p)$ is bigger in this case.

The goal is to find the parameter vector p that minimizes $r(p)$ and $e(p)$. After calculation $r(p)$ and $e(p)$ for the given p, the parameter vector p is updated by Δp and new SSE is calculated as follows

$$e_k = e(p + k\Delta p)\,. \tag{22}$$

If $e_k < e(p)$, $k = 1$, otherwise $k = 0.5$ or 0.25. Vector Δp is a vector of little changes, which gives a probable direction in the search space. Finally, p is updated accordingly

$$p + k\Delta p \rightarrow p \tag{23}$$

and an algorithm for adapting the model is based on iteration procedure of eq. (22) and (23) until convergence [15].

Table 1 shows SSE for a frame from the training set in dependence on variation of the translation vector $T = T(t_k, t_v)$ of the model.

Table 2 shows SSE for the same frame with various positions of the model in dependence on variation of the rotation angle Θ_r and the translation vector T. The forward and inverse transforms, used for both cases, have been calculated by using the designed BEF with all 38 eigenfaces of the size 40×42.

From Tabs. 1 and 2 follow out that SSE=0 for the full BEF and optimal position ($T = (0, 0)$ and $\Theta_r = 0$) of the model in the input frame. As it is seen in Tab. 1 translation in vertical direction of the model from the optimal position gives larger SSE compared to that one for translation in horizontal direction. Sensitivity of SSE on rotation by Θ_r at zero translation is seen from the first column in Tab. 2. In next columns of Tab. 2 the affect both translation and rotation of the model on the values of SSE can be seen. In this case of moving of the model by translation together with its rotation the values of SSE increase. Afterwards tracking of the human head on the basis of eq. (22) and (23) is very effective using BEF when SSE changes enough in dependence on motion parameters.

For the frames outside the training set and the full BEF the measured SSE is not zero at the optimal position of the model but changes from 78 to 124. It does not matter from the point of view of tracking of the human head in the frames because the sensitivity of SSE on replacing of the model from the optimal position is kept.

When BEF included only 10 eigenfaces the synthesized textures are still sufficiently accurate as it is seen from Fig. 7. In the case of using of the non full BEF for tracking of the human head in the frames inside or outside of the training set the measured SSE next increases about 56 for all positions of the model. The sensitivity of SSE is always kept in dependence on moving of the model by its translation and rotation.

6 CONCLUSION

Presented BEF is very important part of the tracking system, which is used for tracking of the motion of the human head in the head-and-shoulder videosequences.

From Tabs. 1 and 2 it is possible to see that by choosing a more accurate texture, smaller SSE is accrued. This proves that BEF in the tracking system of the human head enables adapting the model to the face images of the input videosequence.

Improvement of the tracking system can be realized by means of an automatic search algorithm. Recently we are developing the tracking system by the algorithm, for which reasonable starting approximation is known. Procedure of optimization of the parameter vector p is the same, only the update vector is found by multiplying the residual image with an update matrix. The matrix is created in advance by training set of the face images with the model correctly adapted. The tracking system with the

automatic search algorithm on the basis BEF can adapt the model to the new frame in real time. Disadvantage of this system is more complex training process.

The proposed BEF can be used by other analysis/synthesis tasks like facial detection and recognition.

Acknowledgement

The work was supported by the Grant Agency of Ministry of Education and Slovak Academy of Science under Grant No. 1/3133/06.

References

[1] MIHALIK, J.: Image Coding in Videocommunication, Mercury-Smekal, Košice, 2001. (In Slovak)

[2] PANDZIC, I.—FORCHHEIMER, R.: MPEG-4 Facial Animation: The Standard, Implementation and Aplications, John Wiley and Sons, 2002.

[3] HESS, M.—MARTINEZ, G.: Facial Feature Extraction Based on the Smallest Univalue Segment Assimilating Nucleus (SUSAN) Algorithm, Picture Coding Symposium, 2004.

[4] JACQUIN, A.—ELEFTHERIADIS, A.: Automatic Location Tracking of Faces and Facial Features in Videosequences, International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 142–147.

[5] SMITH, S.—BRADY, J.: Susan — New Approach to Low Level Image Processing, International Journal of Computer Vision **23** No. 1 (1997), 45–78.

[6] AHLBERG, J.: Using the Active Appearance Algorithm for Face and Facial Feature Tracking, 2nd International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems, 2001, pp. 68–72.

[7] ANTOSZCZYSZYN, P.—HANAH, J.—GRANT, P.: A New Approach to Wire-Frame Tracking for Semantic Model-Based Moving Image Coding, Signal processing: Image Communication **15** (2000), 567–580.

[8] AHLBERG, J. Candide-3 – Updated Parameterized Face: Report No. LiTH-ISY-R-2326, Dept. of EE, Linköping University, 2001.

[9] MIHALÍK, J.—MICHALČIN, V.: 3D Motion Estimation and Texturing of Human Head Model, Radioengineering **13** No. 1 (2004), 26–31.

[10] GALLIER, J.: Curves and Surfaces in Geometrical Modeling – Theory and Algorithms, Morgan Kaufmann Publishers, 2000.

[11] FOLEY, J. D.—van DAM, A.—FEINER, S. K.—HUGHES, J. F.: Computer Graphics, Principles and Practices, 2nd edition, Addison-Wesley, 1990.

[12] MIHALÍK, J.—MICHALČIN, V.: Texturing of Surface of 3D Human Head Model, Radioengineering **13 No.4,** (2004), 44–47.

[13] COOTES, T. F.—TAYLOR, C. J.: Statistical Models of Appearance for Computer Vision, Technical report, University of Manchester, 2004.

[14] MURAKAMI, H.—KUMAR, V.: Efficient Calculation of Primary Images from a Set of Images, IEEE Trans. Pattern Anal. and Mach. Intell. **4** No. 5 (1982), 511–515.

[15] DORNAIKA, F.—AHLBERG, J.: Face Model Adaptation Using Robust Matching and the Active Appearance Algorithm, IEEE Workshop on Applications of Computer Vision, 2002, 3-7.

Received 7 March 2006

Ján Mihalík graduated from the Technical University in Bratislava in 1976. In 1979 he joined the Faculty of Electrical Engineering and Informatics of the Technical University of Košice, where received his PhD degree in Radioelectronics in 1985. Currently, he is Full Professor of Electronics and Telecommunications and head of the Laboratory of Digital Image Processing and Videocommunications at the Department of Electronics and Multimedia Telecommunications. His research interests include information theory, image and video coding, digital image and video processing and multimedia videocommunications.

Miroslav Kasár was born in 1980 in Hnúšťa, Slovakia. He received the Ing (MSc) degree from the Technical University of Košice in 2003. At present he is a PhD student at the Department of Electronics and Multimedia Telecomunications of the Technical University, Košice. His research interest includes video coding with a very low bit rate.