Expression-Invariant 3D Face Recognition

Alexander M. Bronstein[1], Michael M. Bronstein[1], and Ron Kimmel[2]

[1] Technion – Israel Institute of Technology
Department of Electrical Engineering, Haifa 32000, Israel
{alexbron,bronstein}@ieee.org
[2] Technion – Israel Institute of Technology
Department of Computer Science, Haifa 32000, Israel
ron@cs.technion.ac.il

Abstract. We present a novel 3D face recognition approach based on geometric invariants introduced by Elad and Kimmel. The key idea of the proposed algorithm is a representation of the facial surface, invariant to isometric deformations, such as those resulting from different expressions and postures of the face. The obtained geometric invariants allow mapping 2D facial texture images into special images that incorporate the 3D geometry of the face. These signature images are then decomposed into their principal components. The result is an efficient and accurate face recognition algorithm that is robust to facial expressions. We demonstrate the results of our method and compare it to existing 2D and 3D face recognition algorithms.

1 Introduction

Face recognition is a biometric method that unlike other biometrics, is non-intrusive and can be used even without the subject's knowledge. State-of-the-art face recognition systems are based on a 40-year heritage of 2D algorithms, dating back to the early 1960s [1]. The first face recognition methods used the geometry of key points (like the eyes, nose and mouth) and their geometric relationships (angles, length, ratios, etc.). In 1991, Turk and Pentland introduced the revolutionary idea of applying principal component analysis (PCA) to face imaging [2]. This has become known as the *eigenface* algorithm and is now a golden standard in face recognition. Later, algorithms inspired by eigenfaces that use similar ideas were proposed (see [3], [4], [5]).

However, all the 2D (image-based) face recognition methods appear to be sensitive to illuminations conditions, head orientations, facial expressions and makeup. These limitations of 2D methods stem directly from the limited information about the face contained in a 2D image. Recently, it became evident that the use of 3D data of the face can be of great help as 3D information is viewpoint- and lighting-condition independent, i.e. lacks the "intrinsic" weaknesses of 2D approaches.

Gordon showed that combining frontal and profile views can improve recognition accuracy [6]. This idea was extended by Beumier and Acheroy, who compared cen-

J. Kittler and M.S. Nixon (Eds.): AVBPA 2003, LNCS 2688, pp. 62-70, 2003.

tral and lateral profiles from the 3D facial surface, acquired by a structured light range camera [7]. This approach demonstrated better robustness to head orientations. Another attempt to cope with the problem of head pose using 3D morphable head models is presented in [8], [9]. Mavridis *et al.* incorporated a range map of the face into the classical face recognition algorithms based on PCA and hidden Markov models [10]. Particularly, this approach showed robustness to large variations in color and illumination and use of cosmetics, and also allowed separating the face from cluttered background.

However, none of the approaches proposed heretofore was able to overcome the problems resulting from the non-rigid nature of the human face. For example, Beumier and Acheroy failed to perform accurate global surface matching, and observed that the recognition accuracy decreased when too many profiles were used [7]. The difficulty in performing accurate surface matching of facial surfaces was one of the primary limiting factors of other 3D face recognition algorithms as well.

In this work, we present a geometric framework for efficient and accurate face recognition using 3D data (patented, [11]). Our method is based on geometric invariants of the human face and performs a non-rigid surface comparison, allowing deformations, typical to the human face due to facial expressions.

2 Non-rigid Surface Matching

Classical surface matching methods, based on finding a Euclidean transformation of two surfaces which maximizes some shape similarity criterion (see, for example, [12], [13], [14]), are suitable mainly for *rigid objects*. Human face can not be considered a rigid object since it undergoes deformations resulting from facial expressions. On the other hand, the class of transformations that a facial surface can undergo is not arbitrary, and empirical observations show that facial expressions can be modeled as *isometric* (or *length-preserving*) transformations. Such transformations do not stretch and do not tear the surface, or more rigorously, preserve the surface *metric*. The family of surfaces resulting from such transformations is called *isometric surfaces*. The requirement of a deformable surface matching algorithm is to find a representation, which is the same for all isometric surfaces.

Schwartz *et al.* were the first to use multidimensional scaling (MDS) as a tool for studying curved surfaces by planar models. In their pioneering work, they applied an MDS technique to flatten convoluted cortical surfaces of the brain, onto a plane, in order to study their functional architecture [15]. Zigelman *et al.* [16] and Grossman *et al.* [17] extended some of these ideas to the problem of texture mapping and voxel-based cortex flattening. A generalization of this approach was introduced in the recent work of Elad and Kimmel [18], as a framework for object recognition. They introduced an efficient algorithm to construct a signature for isometric surfaces. This method, referred to as *bending-invariant canonical forms*, is the core of our 3D face recognition framework.

2.1 Bending-Invariant Canonical Forms

Consider a polyhedral approximation of the facial surface, S. One can think of such an approximation as if obtained by sampling the underlying continuous surface on a finite set of points p_i ($i = 1,\ldots,n$), and discretizing the metric δ associated with the surface

$$\delta(p_i, p_j) = \delta_{ij} . \tag{1}$$

Writing the values of δ_{ij} in matrix form, we obtain the matrix of mutual distances between the surface points. For convenience, we define *squared* mutual distances,

$$(\Delta)_{ij} = \delta_{ij}^2 . \tag{2}$$

The matrix Δ is invariant under isometric surface deformations, but it is not a unique representation of isometric surfaces, since it depends on arbitrary ordering of the points. We would like to obtain a geometric invariant, which is unique for isometric surfaces on one hand, and allows using simple rigid surface matching algorithms to compare such invariants on the other. Treating the squared mutual distances as a particular case of *dissimilarities*, one can apply a dimensionality-reduction technique called *multidimensional scaling* (MDS) in order to embed the surface into a low-dimensional Euclidean space $\mathbf{R}^m$. This is equivalent to finding a mapping between two metric spaces,

$$\varphi : (S, \delta) \to (R^m, d) \quad ; \quad \varphi(p_i) = x_i , \tag{3}$$

which minimizes the embedding error

$$\varepsilon = f\left(\left|\delta_{ij} - d_{ij}\right|\right) \quad ; \quad d_{ij} = \left\|x_i - x_j\right\|_2 . \tag{4}$$

The obtained m-dimensional representation is a set of points $x_i \in \mathbf{R}^m$ ($i = 1,\ldots,n$), corresponding to the surface points p_i. Different MDS methods can be derived using different embedding error criteria [19].

A particular case is the *classical scaling*, introduced by Young and Householder [20]. The embedding in $\mathbf{R}^m$ is performed by double-centering the matrix Δ

$$B = -\tfrac{1}{2} J \Delta J . \tag{5}$$

(here J = I - ½U; I is a $n \times n$ identity matrix, and U is a matrix consisting entirely of ones). The first m eigenvectors e_i, corresponding to the m largest eigenvalues of B, are used as the embedding coordinates

$$x_i^j = e_i^j \quad ; \quad i=1,\ldots,n;\ j=1,\ldots,m , \tag{6}$$

where x_i^j denotes the j-th coordinate of the vector x_i. Eigenvectors are computed using a standard eigendecomposition method. Since only m eigenvectors are required (usually, m=3), the computation can be done efficiently (e.g. by power methods).

We will refer to the set of points x_i obtained by MDS as the *bending-invariant canonical form* of the surface; when m=3, it can be plotted as a surface. Standard rigid surface matching methods can be used in order to compare between two deformable surfaces, using their bending-invariant representations instead of the surfaces themselves. Since the canonical form is computed up to a translation, rotation, and reflection transformation, to allow comparison between canonical forms, they must be *aligned*. This is possible, for example, by setting the first-order moments (center of mass) and the mixed second-order moments to zero (see [21]).

2.2 Measuring Geodesic Distances on Triangulated Manifolds

One of the crucial steps in the construction of the canonical form of a given surface, is an efficient algorithm for the computation of the geodesic distances on surfaces, that is, δ_{ij}. A computationally inefficient distance computation algorithm was one of the disadvantages of the work of Schwartz *et al.* and actually limited practical applications of their method.

A numerically consistent algorithm for distance computation on triangulated domains, henceforth referred to as *fast marching on triangulated domains* (FMTD), was used by Elad and Kimmel [18]. FMTD was proposed by Kimmel and Sethian [22] as a generalization of the *fast marching method* [23]. Using FMTD, the geodesic distances between a surface vertex and the rest of the n surface vertices can be computed in $O(n)$ operations. We use this method for the bending invariant canonical form computation.

3 Range Image Acquisition

Accurate acquisition of the facial surface is crucial for 3D face recognition. Many of commercial range cameras that are available in the market today are suitable for face recognition application. Roughly, we distinguish between *active* and *passive* range sensors. The majority of passive range cameras exploit stereo vision, that is, the 3D information is established from correspondence between pixels in images viewed from different points. Due to the computational complexity of the correspondence problem, passive stereo is usually unable to produce range images in real time.

Active range image acquisition techniques usually use controlled illumination conditions for object reconstruction. One of the most popular approaches known as *structured light*, is based on projecting a pattern on the object surface and extracting the object geometry from the deformations of the pattern [24]. A more robust and accurate version of this approach uses a series of black and white stripes projected sequentially and is known as *coded light*. The patterns form a binary code, that allows the reconstruction of the angle of each point on the surface with respect to the optical axis of the camera. Then one can compute the depth using triangulation.

In this paper, we use the coded light technique for 3D surface acquisition. Using 8 binary patterns, we obtained 256 depth levels which yielded depth resolution of about 1 mm. In our setup, we used an LCD projector with refresh rate of 70Hz controlled via the DV interface. Images were acquired at the rate of 30 frames per second by a black-and-white FireWire CCD camera with a resolution of 640×480 pixels, 8 bit.

4 3D Face Recognition Using Eigenforms

As a first step, using the range camera we acquire a 3D image, which includes the range image (geometry) and the 2D image (texture) of the face. The range image is converted into a triangulated surface and smoothed using spline. Regions outside the facial contour are cropped, and the surface is decimated to a size of approximately 2000-2500 vertices. Next, the bending-invariant canonical form of the face is computed and aligned using the procedure described in Section 2.1.

Since there is a full correspondence between the texture image pixels a_n and the canonical surface vertices $\left(x_n^1, x_n^2, x_n^3\right)$, the face texture image can be mapped onto the aligned canonical surface in the canonical form space. By interpolating a_n and x_n^3 onto a Cartesian grid in the X^1X^2 plane, we obtain the *flattened texture* $\tilde{a}$ and the *canonical image* $\tilde{x}$, respectively. Both $\tilde{a}$ and $\tilde{x}$ preserve the invariance of the canonical form to isometric transformations, and can be represented as images (Fig. 1).

Application of eigendecomposition is straightforward in this representation. Like in eigenfaces, we have a training set, which is a set of duplets of the form $\left\{\tilde{x}_n, \tilde{a}_n\right\}_{i=1}^{N}$. Applying eigendecomposition separately on the set of $\tilde{a}$ and $\tilde{x}$, we produce two sets of eigenspaces corresponding to the flattened textures and the canonical images. We term the respective sets of eigenvectors e_n^a and e_n^x as *eigenforms*.

For a new subject represented by $\left(\tilde{x}', \tilde{a}'\right)$, the decomposition coefficients are computed according to

$$\begin{aligned} \alpha &= \left[e_1^a, \ldots, e_N^a\right]\left(\tilde{a}' - \bar{a}\right), \\ \beta &= \left[e_1^x, \ldots, e_N^x\right]\left(\tilde{x}' - \bar{x}\right), \end{aligned} \tag{10}$$

where $\bar{a}$ and $\bar{x}$ denote the average of $\tilde{a}_n$ and $\tilde{x}_n$ in the training set, respectively. The distance between two subjects represented by $\left(\tilde{x}_1, \tilde{a}_1\right)$ and $\left(\tilde{x}_2, \tilde{a}_2\right)$ are computed as a weighted Euclidean distance between the corresponding decomposition coefficients, (α_1, β_1) and (α_2, β_2).

5 Results

The experiments were performed on a 3D face database consisting of 64 children and 93 adults (115 males and 42 females). The texture the range images (acquired at a resolution of 640×480) were decimated to a scale of 1:8 and cropped outside of the facial contour. The database contained several instances of identical twins (Alex and

Mike). Four approaches were compared: (i) eigendecomposition of range images; (ii) combination of texture and range images in the eigenfaces scheme, as proposed by Mavridis *et al.*; (iii) eigendecomposition of canonical images; and (iv) our eigenforms algorithm.

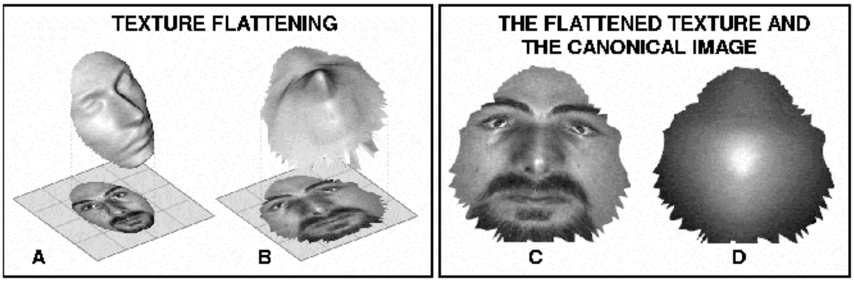

Fig. 1. Texture flattening by interpolation onto the X^1X^2 plane: texture mapping on the facial surface (A) and on the canonical form (B); the resulting flattened texture (C) and the canonical image (D)

Using each of these four algorithms, we found the closest matches between a reference subject and the rest of the database. Different instances of identical twins (Alex and Mike) were chosen as the reference subjects.

Fig. 2 shows significant improvement of the recognition accuracy if canonical images are used instead of range images. Even without using the texture information, we obtain an accurate recognition of twins. Fig. 3 compares between the method of Mavridis *et al*. (eigendecomposition of texture and range images), and our eigenforms method (eigendecomposition of flattened textures and canonical images). Our method made no mistakes in distinguishing between Alex and Mike. One can also observe that a conventional approach is unable to cope with significant deformations of the face (e.g. inflated cheeks), and finds a subject with fat cheeks (Robert 090) as the closest match. This is a result typical for eigenfaces, as well as for straightforward range image eigendecomposition.

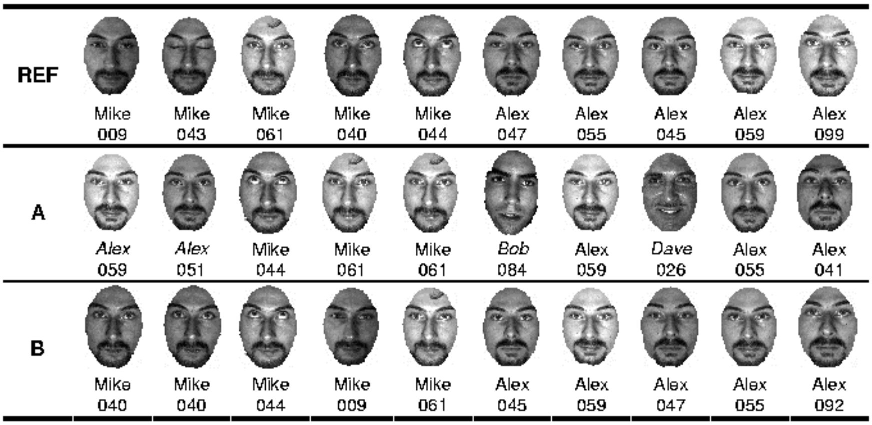

Fig. 2. The closest matches, obtained by eigendecomposition of range images (A) and canonical images (B). Wrong matches are italicized

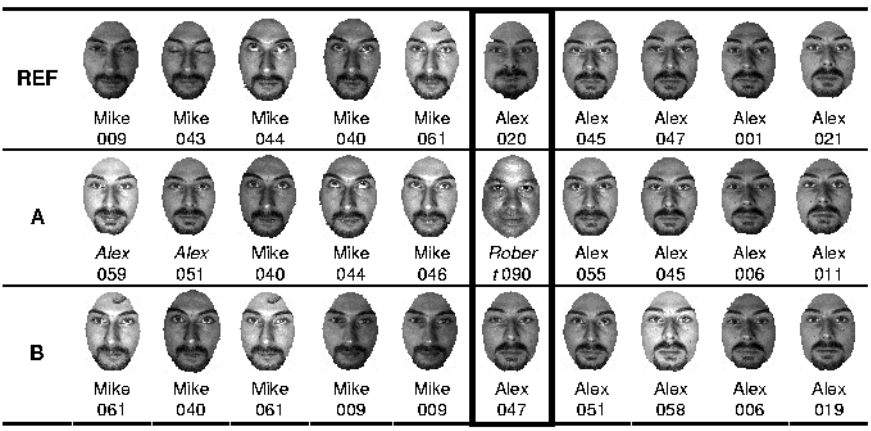

Fig. 3. The closest matches, obtained by the method of Mavridis *et al*. (A) and eigenforms (B). Wrong matches are italicized. Note the inability of the conventional method to cope with subjects with exaggerated facial expression (Alex 20, sixth column)

6 Conclusions

We proposed an algorithm capable of extracting the intrinsic geometric features of facial surfaces using geometric invariants, and applying eigendecomposition to the resulting representation. We obtained very accurate face recognition results. Unlike previously proposed solutions, the use of bending-invariant canonical representation makes our approach robust to facial expressions and transformations typical of non-rigid objects.

Experimental results showed that the proposed algorithm outperforms the 2D eigenfaces approach, and the straightforward incorporation of range images into the eigenfaces framework, proposed by Mavridis *et al*. Particularly, we observed that even very significant deformations of the face do not confuse our algorithm, unlike conventional approaches.

References

[1] Bledsoe, W. W. The model method in facial recognition. Technical report PRI 15, Panoramic Research Inc., Palo Alto (1966).

[2] Turk, M., Pentland, A. Face recognition using eigenfaces, Proc. CVPR, pp. 586-591 (1991).

[3] Belhumeur, V. I., Hespanha, J. P., Kriegman D. J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. PAMI 19(7), pp. 711-720 (1997).

[4] Frey, B. J., Colmenarez, A., Huang, T.S. Mixtures of local linear subspaces for face recognition, Proc. CVPR, pp. 32-37 (1998).

[5] Moghaddam, B., Jebara, T., Pentland, A. Bayesian face recognition. Technical Report TR2000-42, Mitsubishi Electric Research Laboratories (2000).

[6] Gordon, G. Face recognition from frontal and profile views, Proc. Int'l Workshop on Face and Gesture Recognition, pp. 47-52 (1996).

[7] Beumier C., Acheroy, M. P. Automatic face authentication from 3D surface, Proc. British Machine Vision Conf. (BMVC), pp 449-458 (1998).

[8] Huang, J., Blanz, V., Heisele, B. Face recognition using component-based SVM classification and morphable models, SVM 2002, pp. 334-341 (2002).

[9] Blanz, V., Vetter, T. A morphable model for the synthesis of 3D faces, Proc. SIGGRAPH, pp. 187-194 (1999).

[10] Mavridis, N., Tsalakanidou, F., Pantazis, D., Malassiotis, S., Strintzis, M. G. The HISCORE face recognition application: Affordable desktop face recognition based on a novel 3D camera, Proc. Int'l Conf. on Augmented Virtual Environments and 3D Imaging (ICAV3D), Mykonos, Greece (2001).

[11] Bronstein, A. M., Bronstein, M. M., Kimmel, R. 3-Dimensional face recognition, US Provisional patent No. 60/416,243 (2002).

[12] Faugeras, O. D., Hebert, M. A 3D recognition and positioning algorithm using geometrical matching between primitive surfaces, Proc. 7th Int'l Joint Conf. on Artificial Intelligence, pp. 996–1002 (1983).

[13] Besl, P. J.. The free form matching problem. Machine vision for three-dimensional scene, In: Freeman, H. (ed.) New York Academic (1990).

[14] Barequet, G., Sharir, M. Recovering the position and orientation of free-form objects from image contours using 3D distance map, IEEE Trans. PAMI, 19(9), pp. 929–948 (1997).

[15] Schwartz, E. L., Shaw, A., Wolfson, E. A numerical solution to the generalized mapmaker's problem: flattening nonconvex polyhedral surfaces, IEEE Trans. PAMI, 11, pp. 1005-1008 (1989).

[16] Zigelman, G., Kimmel, R., Kiryati, N. Texture mapping using surface flattening via multi-dimensional scaling, IEEE Trans. Visualization and Comp. Graphics, 8, pp. 198-207 (2002).
[17] Grossman, R., Kiryati, N., Kimmel, R. Computational surface flattening: a voxel-based approach, IEEE Trans. PAMI, 24, pp. 433-441 (2002).
[18] Elad, A., Kimmel, R. Bending invariant representations for surfaces, Proc. CVPR (2001).
[19] Borg, I., Groenen, P. Modern multidimensional scaling - theory and applications, Springer (1997).
[20] Young, G., Householder, G.S. Discussion of a set of points in terms of their mutual distances, Psychometrika 3 (1938).
[21] Tal, A., Elad, M., Ar, S. Content based retrieval of VRML objects – an iterative and interactive approach", EG Multimedia, 97 (2001).
[22] Kimmel, R., Sethian, J. A. Computing geodesic on manifolds. Proc. US National Academy of Science 95, pp. 8431–8435 (1998).
[23] Sethian, J. A. A review of the theory, algorithms, and applications of level set method for propagating surfaces. Acta numerica (1996).
[24] Horn, E., Kiryati, N. Towards optimal structured light patterns, Image and Vision Computing, 17, pp. 87-97 (1999).