FUSION OF 2D AND 3D DATA IN THREE-DIMENSIONAL FACE RECOGNITION

Alexander M. Bronstein

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

Michael M. Bronstein, Eyal Gordon, Ron Kimmel

Department of Computer Science
Technion - Israel Institute of Technology
Haifa 32000, Israel

ABSTRACT

We discuss the synthesis between the 3D and the 2D data in three-dimensional face recognition. We show how to compensate for the illumination and facial expressions using the 3D facial geometry and present the approach of canonical images, which allows to incorporate geometric information into standard face recognition approaches.

1. INTRODUCTION

In face recognition, it is desired to be able to identify different instances of the same face, independent of external factors (illumination conditions, head pose relative to the camera, use of cosmetics, etc.) and internal factors (facial expressions). The ultimate goal of face recognition is to find some invariant representation of the face, which would be insensitive to all these changes. Unfortunately, a 2D image of the face can be significantly altered as the result of these factors. Generative approaches achieved certain success in coping with variations in illumination and changes in the head pose [1], but are sensitive to facial expressions.

A relatively new trend in face recognition is an attempt to use 3D imaging [2, 3, 4]. Besides a conventional face picture (reflectance image), three-dimensional data carry all the information about the geometry of the face. In [5, 6], a new approach able to cope with problems resulting from the non-rigid nature of the human face was introduced. Assuming that many of human facial expressions are near-isometric transformations of the facial surface, using embedding into low-dimensional Euclidean space the facial surface can be converted into a representation, which is invariant under such transformations. Such a representation facilitates the recognition in the presence of facial expressions.

The focus of this paper is on the synthesis between the 3D and the 2D data in the framework of [6]. Besides making the facial surface geometry available, 3D imaging also benefits from the ability to compensate for the illumination in the reflectance images of the face, i.e. estimate the albedo of the face. When the facial geometry together with the albedo is embedded into a plane, a 2D illumination- expression- and pose-invariant representation of the face is obtained. Standard techniques can be then employed to carry out the recognition.

This research was supported by Dvorah Fund of the Technion, Bar Nir Bergreen Software Technology Center of Excellence and the Technion V.P.R. Fund - E. and J. Bishop Research Fund. The authors are listed in alphabetical order.

2. IMAGE FORMATION

Let us assume for simplicity that the facial surface is represented by the graph of a function $z(x, y)$, has a Lambertian reflectance with albedo $\rho(x, y)$ and is viewed orthographically (see Figure 1). We will denote by $\mathbf{n}(x, y)$ the inward pointing unit normal vector, which can be expressed as

$$\mathbf{n}(x,y) = \frac{(z_x(x,y), z_y(x,y), -1)}{\sqrt{1 + \|\nabla z(x,y)\|_2^2}} \tag{1}$$

in terms of the $x-$ and $y-$ partial derivatives of z. Also, we will use the notation $\mathbf{b}(x, y) = \rho(x, y)\mathbf{n}(x, y)$. We will further assume that $\mathbf{b}(x, y)$ is discretized on an even $N \times M$ grid and represented by a matrix $B \in \mathbf{R}^{MN\times 3}$ containing $\mathbf{b}(x, y)$ as rows.

When the surface is illuminated by a distant point light source described by intensity α and unit direction vector $\mathbf{l}$, a reflectance image $\mathbf{r}$ (with the same ordering of pixels as B) is formed according to

$$\mathbf{r} = \max(B\mathbf{s}, \mathbf{0}), \tag{2}$$

where $\mathbf{s} = \alpha\mathbf{l}$. Zero values correspond to attached shadows. Cast shadows that may appear since the object is non-convex are ignored [1]. If the illumination is structured, the source intensity is a function of x, y, i.e. $\mathbf{s}(x, y) = \alpha(x, y)\mathbf{l}$. Under the assumption that the camera has a linear response, the superposition principle holds: when the surface is illuminated by sources $\mathbf{s}_1, \mathbf{s}_2$, the resulting reflectance image is $\mathbf{r} = \max(B(\mathbf{s}_1 + \mathbf{s}_2), \mathbf{0})$.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

3. RECONSTRUCTION

At the first step a combined 2D-3D face recognition system has to acquire the geometry of the face and its reflectance characteristics. The presented algorithm will not require the explicitly surface, but it will be rather sufficient to reconstruct the normal field $\mathbf{n}(x, y)$. The latter will also suffice to estimate the albedo $\rho(x, y)$ from the reflectance image. We highlight here several range acquisition techniques and discuss the issues of the normal field and albedo estimation.

3.1. Photometric stereo

When $K \geq 3$ reflectance images $\mathbf{r}_1, ..., \mathbf{r}_K$ resulting from light sources $\mathbf{s}_1, ..., \mathbf{s}_K$ are available, it is possible to reconstruct both the surface normals and the albedo, if $\text{rank}(S) = 3$ (where $S = (\mathbf{s}_1, ..., \mathbf{s}_K) \in \mathbf{R}^{3\times K}$). We assume that there is no ambient light; in practice, ambient light can be compensated for by subtraction of the "darkness" image (reflectance image formed by ambient illumination only). The methods of shape reconstruction from illuminations at different directions are usually termed as *photometric stereo*.

Let $R = (\mathbf{r}_1, ..., \mathbf{r}_K) \in \mathbf{R}^{NM\times K}$ and assume for a moment that there are no attached shadows, i.e. the reflectance images are given by $X = BS$. In this case, the matrix B can be recovered by pseudoinversion of S, i.e. $B = RS^{\dagger}$, resulting from the solution of the problem $\min_B \|R - BS\|_F^2$. The unit normal vectors are obtained by normalizing the rows of B. The albedo at each pixel, accordingly, is the norm of the respective row of B.

Attached shadows are problematic in such a solution, since at shadowed pixels the linear relationship between B and R is no more true. Hence, shadowed pixels must be excluded from computation. Assume that in all the reflectance images we are able to determine whether pixel (x, y) is shadowed (e.g. by thresholding, see [1]), and find that in $L \geq 3$ images $\mathbf{r}_{m_1}, ..., \mathbf{r}_{m_L}$ the pixel is not shadowed. Only these images will be used for reconstruction at pixel (x, y), i.e. $\mathbf{b}(x, y) = (r_{m_1}(x, y), ..., r_{m_L}(x, y))\, \tilde{S}^{\dagger}$, where $\tilde{S} = (\mathbf{s}_{m_1}, ..., \mathbf{s}_{m_L})$.

3.2. Spatially-multiplexed photometric stereo

Instead of acquiring a set of K images, each illuminated by a single distinct source $\mathbf{s}_k$, one can use *multiplexed illumination*, wherein each image is acquired under a superposition of illuminations emerging from L light sources with different weights:

$$\mathbf{r}_k = \max(B(\alpha_{1k}\mathbf{l}_1 + ... + \alpha_{Lk}\mathbf{l}_L), \mathbf{0}), \qquad (3)$$

where α_{ij} is the known relative contribution of the i-th light source to the j-th image. If the weights are selected in such a way that the matrix (α_{ij}) is invertible, it is possible to compute (demultiplex) reflectance images resulting from each of the single distant sources $\mathbf{l}_1, ..., \mathbf{l}_K$. In [7] it was shown that e.g. Hadamard matrices can be used for binary multiplexing. After demultiplexing, the normal field and the albedo are reconstructed similarly to the standard photometric stereo. However, multiplexed illumination is advantageous, since it allows to achieve higher SNR [7].

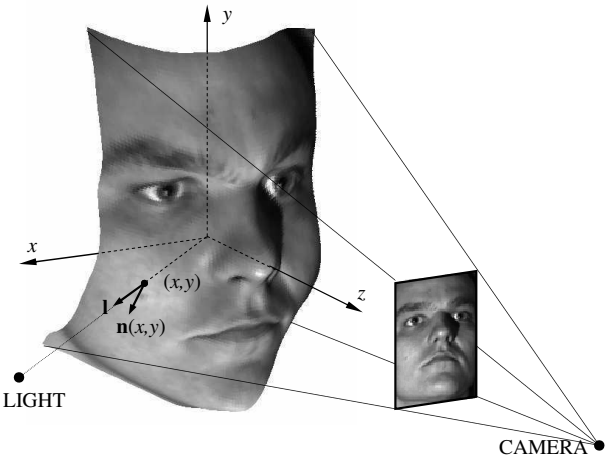

Fig. 1. Image acquisition in photometric stereo.

3.3. Structured light

The facial surface can also be acquired using the *structured* or *coded light* techniques [8, 9], wherein the range sensor, consisting of a camera and projector, projects a pattern (or a set of patterns) onto the face and reconstructs the depth of each pixel by means of triangulation. In this case, the normal field $\mathbf{n}(x, y)$ can be estimated locally from the depth data, whereas the illumination direction and intensity $\mathbf{s}(x, y)$ are known at each pixel. Excepting the occluded regions, the albedo can be reconstructed according to

$$\rho(x, y) = \frac{r(x, y)}{\langle \mathbf{n}(x, y), \mathbf{s}(x, y) \rangle}. \qquad (4)$$

In occluded pixel, albedo can be interpolated from neighbour non-occluded pixels.

4. BENDING-INVARIANT REPRESENTATION

Due to the non-rigid nature of the human face, it undergoes deformations as the result of facial expressions. In [6], an isometric model for facial expressions was used (see [15] for an experimental proof of this model). Facial expressions can be approximated by isometric transformation, i.e. transformations do not stretch or tear the surface (or more rigorously, preserve the surface metric). Hence, faces can be thought of as an equivalence classes of surfaces obtained by isometric transformations. A representation of these equivalence classes, called the *bending invariant canonical form*,

was proposed in [5] in the context of deformable surface matching method.

The key idea of bending invariant representation is the notion of *embedding*, i.e. mapping of the facial surface from a space with a non-Euclidean metric to a low-dimensional Euclidean space. This is accomplished by first computing the distances between the points of the surface, and then finding a set of corresponding points in the embedding space, such that the mutual distances between the new set of points are as close as possible to the original ones. The following briefly describes the algorithm.

4.1. Geodesic distance measurement

Our model of the face as a graph of $z(x,y)$, is in fact a particular case of a parametric manifold, represented by a mapping $X : \mathbf{R}^2 \rightarrow \mathbf{R}^3$ from the parameterization plane $U = (u^1, u^2) = (x,y)$ to $X(U) = (x,y,z(x,y))$. The derivatives of X with respect to u^i are defined as $X_i = \frac{\partial}{\partial u^i} X$ and constitute a non-orthogonal coordinate system on the manifold:

$$X_1 = (1, 0, z_x(x,y)), \;\; X_2 = (0, 1, z_y(x,y)). \qquad (5)$$

The *metric tensor*

$$(g_{ij}) = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} X_1 \cdot X_1 & X_1 \cdot X_2 \\ X_2 \cdot X_1 & X_2 \cdot X_2 \end{bmatrix} \qquad (6)$$

allows to express a distance element on the manifold as $ds = \sqrt{g_{ij} u^i u^j}$ (here we use Einstein summation convention), and consequently, to measure the *geodesic distance*[1] $\delta(x,y)$ between any pair of points x, y on the manifold.

When the parametric manifold is given in a discrete form, the parametric version of the Fast Marching method (FMM) [10] can be used in order to compute the geodesic distances numerically. A remarkable property of parametric FMM is that it does not use the manifold explicitly and requires only the knowledge of the metric tensor g_{ij}. According to (6), the metric tensor is completely defined by the surface gradients $z_x(x,y), z_y(x,y)$ or alternatively, by the surface normals $\mathbf{n}(x,y)$. This allows to bypass the surface reconstruction [11] on the geodesic distance computation stage.

4.2. Embedding

The output of FMM is a $MN \times MN$ matrix Δ of squared mutual geodesic distances $\delta^2(x,y)$ between every pair of points on the surface[2]. The matrix Δ is invariant under isometric surface deformations, but is not a unique representation of isometric surfaces, since it depends on arbitrary ordering and the selection of the surface points. Treating the squared mutual distances as a particular case of dissimilarities, we apply a multidimensional scaling (MDS) technique in order to embed the surface points with their geodesic distances in a low-dimensional Euclidean space $\mathbf{R}^m$ [12, 13, 5].

In [6] a particular MDS algorithm, the *classical scaling*, was used. The embedding into $\mathbf{R}^m$ is performed in two stages. Firstly, the matrix Δ undergoes double-centering:

$$B = -\frac{1}{2} J \Delta J \qquad (7)$$

(here $J = I - \frac{1}{n} U$; I is a $n \times n$ identity matrix, and U is a matrix of ones). Secondly, eigenvectors e_i, corresponding to m largest eigenvalues of B, are used as the embedding coordinates

$$x_i^j = e_i^j; \qquad i = 1, ..., MN; \quad j = 1, ..., m \; . \qquad (8)$$

where x_i^j denotes the j-th coordinate of the vector x_i. As alternatives to the classical scaling, iterative algorithms, such as LS MDS can be used [12].

The set of points x_i obtained by the MDS is referred to as the bending-invariant canonical form of the surface; when $m = 3$, it can be plotted as a surface. Standard alignment and 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 [5, 6].

4.3. Canonical images

Of special interest in the context of this paper is the case of $m = 2$, i.e. when the embedding is performed into a plane. In this case, embedding can be thought of as warping of the manifold parameterization plane U; a transformation mapping the coordinate system (x,y) to (x',y'). Such warping compensates for isometric transformations of the surface on which the albedo is drawn. Plotting the albedo α in the new coordinates, we obtain the *canonical image* (Figure 2).

The canonical image is invariant under isometric facial expressions, illumination, and head orientations (ignoring occlusions) and to some extent serves as a registration (alignment) method between two faces. As the next stage, standard linear methods (e.g. eigenfaces) can be applied in the canonical images space in order to perform recognition.

5. RESULTS

In this section we exemplify the canonical image computation. Eight photometric images from Yale Database B [14] with illumination angle of up to 20° were used for surface and albedo reconstruction. The images were cropped and subsampled to 50% of the original size. Embedding was

[1] The length of the shortest path on the manifold connecting the points.

[2] In practical implementation, the surface is cropped and subsampled prior to the FMM application to reduce the computational effort. For clarity of presentation, the preprocessing stage is ignored here.

carried out using classical scaling. The result is presented in Figures 2–3. Due to space limitations we do not present here results of face recognition; extensive testing of our approach appears in [15].

Fig. 2. Examples of images used for albedo reconstruction.

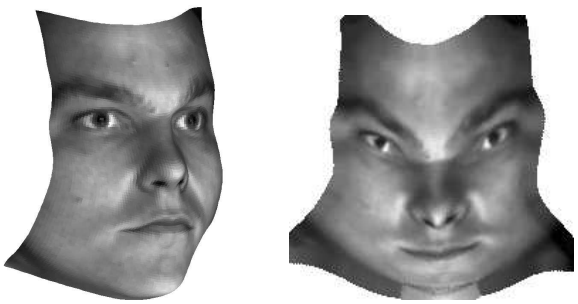

Fig. 3. Estimated albedo rendered on the reconstructed facial surface (left) and the illumination-compensated canonical image (right) obtained by embedding into the plane.

6. CONCLUSIONS

In this paper, we considered the problem of fusing the 2D and the 3D data in three-dimensional face recognition. As the working framework, we focused on the geometric face recognition approach proposed in [6]. The availability of the facial geometry, combined with known illumination direction allows to extract the albedo of the face, which is invariant to illumination. We highlighted how to estimate the albedo from photometric stereo and structured light. The albedo image can be then "flattened" using the MDS procedure applied to the matrix of geodesic distances on the face measured using the Fast Marching method. The resulting canonical image incorporates the geometric invariants of the face (the geodesic distances), which appear to be nearly-invariant to facial expression. Hence, apart from being illumination-invariant, the canonical image is also nearly-invariant to facial expressions. For recognition purposes, these invariant representations can be compared using the classical techniques such as eigendecomposition.

7. REFERENCES

[1] A. S. Gheorghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. PAMI*, vol. 23, no. 6, 2001.

[2] C. Beumier and M. P. Acheroy, "Automatic face authentication from 3D surface," in *Proc. British Machine Vision Conf.*, 1988, pp. 449–458.

[3] J. Huang, V. Blanz, and V. Heisele, "Face recognition using component-based SVM classification and morphable models," *SVM*, pp. 334–341, 2002.

[4] N. Mavridis, F. Tsalakanidou, D. Pantazis, S. Malassiotis, and M. G. Strintzis, "The HISCORE face recognition application: Affordable desktop face recognition based on a novel 3d camera," in *Proc. Int'l Conf. Augmented Virtual Environments and 3D Imaging*.

[5] A. Elad and R. Kimmel, "Bending invariant representations for surfaces," in *Proc. CVPR*, 2001, pp. 168–174.

[6] A. Bronstein, M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition," in *Proc. Audio and Video-based Biometric Person Authentication*, 2003, pp. 62–69.

[7] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, "A theory of multiplexed illumination," in *Proc. ICCV*, 2003, pp. 808–815.

[8] M. Proesmans, L. Van Gool, and A. Oosterlinck, "One-shot active shape acquisition," in *Proc. Internat. Conf. Pattern Recognition*, 1996, vol. C, pp. 336–340.

[9] A. Bronstein, M. Bronstein, E. Gordon, and R. Kimmel, "High-resolution structured light range scanner with automatic calibration," Tech. Rep. CIS-2003-06, Dept. of Computer Science, Technion, Israel, 2003.

[10] R. Kimmel and J. A. Sethian, "Computing geodesic on manifolds," in *Proc. US National Academy of Science*, 1998, vol. 95, pp. 8431–8435.

[11] A. Bronstein, M. Bronstein, R. Kimmel, and A. Spira, "Face recognition from facial surface metric," in *Proc. ECCV*, 2004, to appear.

[12] I. Borg and P. Groenen, *Modern multidimensional scaling - theory and applications*, Springer-Verlag, Berlin Heidelberg New York, 1997.

[13] E. L. Schwartz, A. Shaw, and E. Wolfson, "A numerical solution to the generalized mapmaker's problem: flattening nonconvex polyhedral surfaces," *IEEE Trans. PAMI*, vol. 11, pp. 1005–1008, 1989.

[14] "Yale face database B," Online: http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html.

[15] Michael M. Bronstein, Alexander M. Bronstein, and Ron Kimmel, "Three-dimensional face recognition," Tech. Rep. CIS-2004-04, Dept. of Computer Science, Technion, Israel, 2004.