Three-Dimensional Face Recognition *

Michael M. Bronstein
Department of Computer Science
Technion – Israel Institute of Technology, Haifa 32000, Israel

Alexander M. Bronstein
Department of Electrical Engineering
Technion – Israel Institute of Technology, Haifa 32000, Israel

Ron Kimmel
Department of Computer Science
Technion – Israel Institute of Technology, Haifa 32000, Israel

May 18, 2004

Abstract.

An expression-invariant 3D face recognition approach is presented. Our basic assumption is that facial expressions can be modelled as near-isometric transformations of the facial surface. We first prove our assumption by a simple experiment. Next, using the Elad-Kimmel bending-invariant canonical forms we construct a representation of the facial surface, which is invariant to its isometric transformations. The result is an efficient and accurate face recognition algorithm, robust to facial expressions that can distinguish between identical twins (the first two authors). We demonstrate the performances of the proposed algorithm and compare it to classical face recognition methods.

The numerical methods employed by our approach do not require the facial surface explicitly. The surface gradients field, or the surface metric, are sufficient for constructing the expression-invariant signature of any given face. It allows us to perform the 3D face recognition task while avoiding the surface reconstruction stage.

Keywords: three-dimensional face recognition, bending invariant forms, facial expressions, differential geometry, multidimensional scaling

1. Introduction

Automatic face recognition has been traditionally associated with the fields of computer vision and pattern recognition. Face recognition is considered natural, less intimidating, and a widely accepted biometric identification method (Ashbourn, 2002; Ortega-Garcia et al., 2004). As such, it has the potential of becoming the leading biometric technology. Unfortunately, it is also one of the most difficult recognition

* This research was partially supported by Dvorah Fund of the Technion, the Bar Nir Bergreen Software Technology Center of Excellence and the Technion V.P.R. Fund - E. and J. Bishop Research Fund.

tasks. So far, all existing solutions provide only partial, and usually unsatisfactory, answers to the market needs.

In the context of face recognition, it is common to distinguish between the problem of *verification* and that of *identification*. In the first case, the enrolled individual claims identity of a person whose template is stored in the database. We refer to the data used for a specific recognition task as a template. A face recognition algorithm needs to compare a given face with a given template and verify their equivalence. Such a setup (one-to-one matching) can occur when biometric technology is used to secure financial transactions. In this case, the user can be usually assumed to *collaborate*, willing to assist the biometric system.

The second case is more difficult. Identification implies that the enrolled individual should be compared with all templates stored in the database. The face recognition algorithm should then match a given face with one of the individuals in the database. Finding a terrorist in a crowd (one-to-many matching) is one such application. Needless to say, no collaboration can be assumed in this case.

Despite its straightforward definition, face recognition is a non-trivial task. The facial reflectance image, acquired using a traditional 2D camera, varies due to external and internal factors. External factors include different lighting conditions, head pose relative to the camera, the use of make-up, facial hair, etc. Internal factors include movements of the mimic muscles, which are responsible for facial expressions, and the process of aging. In some cases, the variability in a facial image due to these factors can be even greater than a change in the person's identity (see Figure 1). Theoretically, it is possible to recognize an individual's face reliably in different conditions (illumination, pose, etc.) provided that the same person *has been previously observed in similar conditions*. However, in practice only few observations of the face are usually available. Actually, in some cases only a single observation is given.

Figure 1. Face recognition with varying lighting, head pose, and facial expression is a non-trivial task.

The difficulty to adapt to new viewing conditions is one of the inherent drawbacks of *appearance-based* methods, which include the classical

eigenfaces method (Turk and Pentland, 1991) and similar approaches (Sirovich and Kirby, 1987; Hallinan, 1994; Pentland et al., 1994). These methods use low-dimensional representations of the facial image for the recognition task. They perform well only when the enrolled image is acquired in conditions similar to those in which the template image was acquired (Gheorghiades et al., 2001).

Invariant face recognition refers to methods which are insensitive to one or all of the above mentioned factors. Such methods can be divided into two main classes. The first is the *invariant representation* approach. It attempts to find a set of features which would be invariant, or at least insensitive, to possible changes in the facial image. In fact, the first automatic face recognition algorithms (Bledsoe, 1966; Kanade, 1973; Goldstein et al., 1971) and follow-up works (see surveys (Samil and Iyengar, 1992; Li and Lu, 1999) and references cited therein) were based on the so-called *fiducial points* (eyes, nose, mouth, etc.) and their geometric relations (angles, lengths, and ratios). Unfortunately, very few fiducial points that can be reliably extracted from a 2D facial image are insensitive to illumination and pose (Cox et al., 1996); almost none of these kind of features are invariant under facial expressions. Moreover, such measures are usually insufficient for accurate recognition (Brunelli and Poggio, 1993). Elastic graph matching was used in (Wiskott, 1995; Wiskott et al., 1997) to account for fiducial points displacement due to flexibility of the facial surface. This method yield limited success as the attributed graph is merely a flat representation of a curved object.

The second class of methods is the *generative* approach. Assuming some model of facial image formation, generative methods use observations of the face to find the model parameters, in order to synthetically generate facial images in new poses and illumination conditions. These approaches produce sufficiently good results when coping with variable illumination, especially if a Lambertian reflectance model is assumed. In their recent papers, Gheorghiades et al. (1998, 2001) showed a solution based on the illumination cones model, which is insensitive to illumination and moderate head orientations. Facial expressions appear to be more problematic to synthesize. Approaches modelling facial expressions as warping of the facial image do not capture the true geometric changes of the facial surface, and are therefore useful mainly for graphics applications. That is, the results may look natural yet fail to represent the true nature of the expression.

1.1. Two-dimensions vs. three-dimensions

A recent trend is the approach of *three-dimensional face recognition*. These methods try to use the geometry of the human face in addition to the conventional flat image. In 3D face recognition, the enrollment requires a sensor which acquires depth information – usually referred to as *depth* or *range camera*. With slight abuse of terminology, we refer to 3D face recognition as those methods that use the facial surface information; while the term 2D face recognition is reserved for those using only the 2D reflectance image of the face[1].

In some sense, this relatively new research direction breaks the long-term tradition of 2D face recognition methods, that is of mimicking the functions of the human eye. It appears that the face recognition in the human visual system is indeed based on flat, two-dimensional intensity information rather than the curved facial surface in three dimensions. This can be demonstrated by a simple visual experiment, which we named "face swap" (or "face/off"). In Figure 2, the facial albedo of one individual (Alice) is mapped onto the facial surface of another individual (Bob). The albedo (that represents the intrinsic color and texture) leads us to recognize the subject at the bottom as Alice, although the true facial surfaces of the two individuals are significantly different.

Though practically "overlooked" by the human eye, the geometry of a face potentially conceals useful invariant features. The facial surface is insensitive to illumination, head pose, as well as the use of makeup. Moreover, 3D data can be used to produce invariant measures out of the 2D data. For example, the albedo, which is insensitive to illumination, can be estimated from the 2D reflectance, if the facial surface geometry is known. Prior results prove the utility of 3D geometry for face recognition. Gordon (1997) showed that combining frontal and profile views can improve recognition accuracy. Beumier and Acheroy (1988) proposed augmenting face recognition systems with a range camera that captures the facial surface geometry, in addition to the facial image. Improvement over traditional 2D face recognition algorithms was also shown in (Mavridis et al., ; Huang et al., 2002).

However, facial expressions do change significantly not only the 2D image, but also the geometric structure of the face. The basic weakness of previous 3D face recognition approaches is the treatment of the facial

[1] Gheorghiades et al. (2001), for example, use implicitly the facial surface recovered from photometric stereo to generate synthetic facial images with new poses and illumination. While Chen and Medioni (2001) used stereo vision for 3D face recognition.

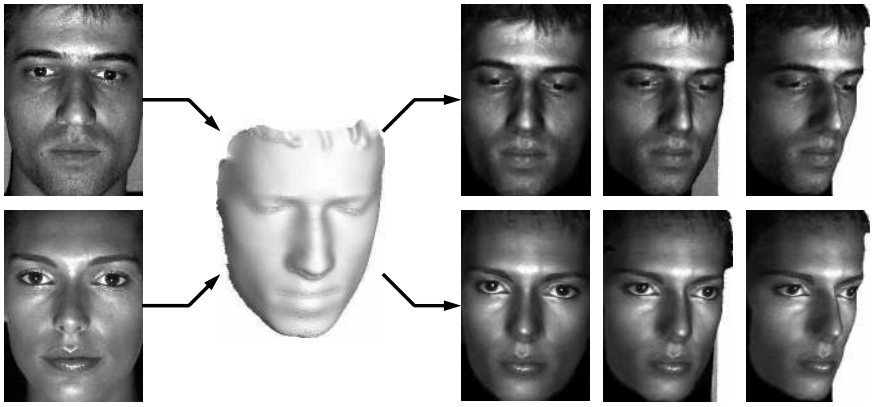

Figure 2. Does depth play a dominant role in face recognition in the human visual system? Left: A frontal picture of subjects Bob (top) and Alice (bottom). Center: Facial surface of Bob. Right: Synthetic views of Bob's face produced using Bob's facial surface (top); and synthetic views of Alice's face textured mapped onto Bob's facial surface (bottom). Though the facial features appear slightly distorted, most human observers preferred to recognize the bottom pictures as Alice, rather than Bob.

surface as a rigid object - an assumption which is inaccurate for human faces. Consequently, these methods can not handle facial expressions.

Our approach is based on treating the facial surface as a deformable object. Empirical observations show that the deformations of the face resulting from mimics can be modelled as near-isometric transformations (Bronstein et al., 2003b). Consequently, the intrinsic geometric properties such as the geodesic distances between points on the facial surface are preserved. Thus, a representation of the face invariant to isometric transformations would also be invariant to facial expressions.

The geodesic structure of an isometric surface is captured by the inter geodesic distance between points on the surface. The question is how should we efficiently compare between the geometric structure of such large arrays of mutual geodesic distances. One computationally efficient invariant representation can be constructed by embedding the surface in a low-dimensional Euclidean space – a procedure known as *flat embedding* or *generalized map-making*. Embedding into a plane was shown to be useful in the analysis of cortical surfaces (Schwartz et al., 1989), and in texture mapping (Zigelman et al., 2002; Grossman et al., 2002). Embedding into higher dimensional Euclidean spaces was shown to be an efficient way to perform matching of deformable objects (Elad and Kimmel, 2001). This method, producing geometric signatures called *bending-invariant canonical forms*, is the core of our 3D face recognition system. It consists of measuring the geodesic distances between points on the facial surface and then using multidimensional

scaling to carry out the flat embedding. This way, the task of comparing deformable objects like faces is transformed into a much simpler problem of rigid surface matching, at the expense of loosing some accuracy, which appears to be insignificant in this case.

An important property of the numerical algorithms implemented in our system, is that we actually do not need the facial surface to be given explicitly. All stages of our recognition system, including pre-processing and computation of geodesic distances can be carried out given only the metric tensor of the surface, or equivalently, the surface gradients. It allows us to use simple and cost-efficient 3D acquisition techniques like photometric stereo. Avoiding the explicit surface reconstruction also saves computational time and reduces the numerical errors (Bronstein et al., 2004b).

The organization of this paper goes as follows: The second section starts with modelling faces as Riemannian manifolds[2]. Such a formulation serves as a unifying framework for different procedures described later on. Facial expressions are modelled as isometric deformations of the facial surface. A simple experiment justifies this assumption. The presentation of the rest of the sections tracks the data flow of our algorithm. In Section 3, we describe depth acquisition techniques suitable for 3D face recognition. Section 4 is dedicated to post-processing of the acquired 3D data. Section 5 deals with measuring geodesic distances on discrete facial surfaces. Section 6 introduces the concept of multidimensional scaling and its application to flat embedding. Section 7 unites all the building blocks in the previous sections into a 3D face recognition system and addresses some implementation considerations. In Section 8 we show experimental results assessing the performance of our method. Section 9 concludes the paper.

2. The geometry of human faces

We model a human facial surface as a 2-dimensional smooth connected parametric manifold $\mathcal{M}$, represented by a coordinate chart from a compact subset $U \subset \mathbb{R}^2$ to $\mathbb{R}^3$, given by

$$\mathbf{x}(U) = (x^1(u^1, u^2), x^2(u^1, u^2), x^3(u^1, u^2)), \tag{1}$$

such that $\mathcal{M} = \mathbf{x}(U)$. We assume that the functions $x^1, ..., x^3$ are smooth ($\mathcal{C}^\infty$), and that the vectors $\partial_i \mathbf{x} = \frac{\partial}{\partial u^i}\mathbf{x}$ $(i = 1, 2)$ are linearly independent at every point. We will further assume, w.l.o.g., that the manifold can be written as a graph of a function, e.g. $x^3 = x^3(x^1, x^2)$,

[2] Hereinafter, the term "manifold" is used synonymously to "surface."

such that x^3 can be referred to as the *depth* coordinate. Also, for convenience, in the following the parameterization coordinates $\mathbf{u} = (u^1, u^2)$ will be identified with the coordinates in the image acquired by the camera (see Section 3). Similarly, we define the facial *albedo* $\rho(u^1, u^2)$.

The derivatives $\partial_i \mathbf{x}$ constitute a local non-orthogonal coordinate system on $\mathcal{M}$, and span an affine subspace of $\mathbb{R}^3$ called the *tangent space* and denoted by $T_{\mathbf{x}}\mathcal{M}$ for every $\mathbf{x} \in \mathcal{M}$. In order to consider the non-Euclidean geometry of the manifold, we introduce a bilinear symmetric non-degenerate form (tensor) g called the *Riemannian metric* or the *first fundamental form*. It can be identified with an inner product on $T_{\mathbf{x}}\mathcal{M}$. The Riemannian metric is an intrinsic characteristic of the manifold and allows us to measure local distances on $\mathcal{M}$ independently of the coordinates (Kreyszig, 1991). The pair $(\mathcal{M}, g)$ is called a *Riemannian manifold*.

In coordinate notation, a distance element on the manifold can be expressed via the *metric tensor* $g_{ij}(\mathbf{x})$ as

$$ds^2 = g_{ij} du^i du^j, \quad i, j = 1, 2; \tag{2}$$

where repeating super- and subscripts are summed over according to Einstein's summation convention. The metric tensor g_{ij} of the manifold is given explicitly by

$$g_{ij} = \partial_i \mathbf{x} \cdot \partial_j \mathbf{x}, \quad i, j = 1, 2. \tag{3}$$

The unit *normal* to $\mathcal{M}$ at $\mathbf{x}$ is a vector orthogonal to the tangent plane $T_{\mathbf{x}}\mathcal{M}$ and can be written as a cross-product

$$\mathbf{n}(\mathbf{x}) = \frac{\partial_1 \mathbf{x} \times \partial_2 \mathbf{x}}{\|\partial_1 \mathbf{x} \times \partial_2 \mathbf{x}\|_2}. \tag{4}$$

Differentiating the obvious relation $\partial_i \mathbf{x} \cdot \mathbf{n} = 0$ w.r.t. u^j yields another intrinsic characteristic of the manifold called the *second fundamental form*, which is given in coordinate notation as

$$b_{ij} = -\partial_i \mathbf{x} \cdot \partial_j \mathbf{n} = \partial_j (\partial_i \mathbf{x}) \cdot \mathbf{n}. \tag{5}$$

The importance of the second fundamental form is that the maximum and the minimum eigenvalues $\lambda_{\max}, \lambda_{\min}$ of the tensor $b_i^j = b_{ik} g^{kj}$ are the *principal curvature* values of the manifold at point $\mathbf{x}$. The corresponding eigenvectors of (b_{ij}) are the principal curvature directions. The value $b_i^i = (\lambda_{\max} + \lambda_{\min})$ is called the *mean curvature* and is denoted by H; the determinant $\det(b_i^j) = \lambda_{\max}\lambda_{\min}$ is called the *Gaussian curvature* and is denoted by K.

Since our manifold is connected and compact, the Riemannian metric induces a distance metric. In order to define it, let $\mathbf{x}, \mathbf{y} \in \mathbf{x}(U)$ be

two surface points and let $\mathbf{c} : [0,1] \to \mathcal{M}$ be a smooth parametric curve from $\mathbf{x}$ to $\mathbf{y}$ on the surface $\mathcal{M}$. The length of $\mathbf{c}(t)$ is defined by

$$\ell(\mathbf{c}) = \int_0^1 \left\| \frac{d\mathbf{c}(t)}{dt} \right\| dt. \tag{6}$$

Then, the distance from $\mathbf{x}$ to $\mathbf{y}$ on $\mathcal{M}$ is given by

$$d(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{c}(t)} \ell(\mathbf{c}). \tag{7}$$

The paths of minimum length, resulting from the extrema of the functional $\ell(\mathbf{c})$ are called *minimal geodesics*, and $d(\mathbf{x}, \mathbf{y})$ is called the *geodesic distance*.

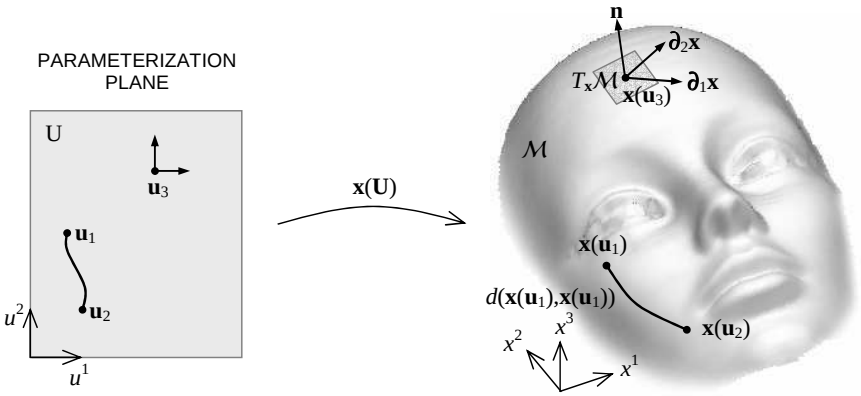

Figure 3. Definition of the manifold $\mathcal{M}$ by the mapping $\mathbf{x} : U \to \mathcal{M}$ and examples of the differential geometric notions mentioned in the text.

2.1. Facial expressions as isometric transformations

Let $f : (\mathcal{M}, g) \to (\mathcal{M}', g')$ be a diffeomorphism on the facial manifold $\mathcal{M}$. The resulting surface $\mathcal{M}'$ can be parameterized by $\mathbf{x}'(U) = f(\mathbf{x}(U))$. We will also assume that the vectors $\partial_i \mathbf{x}'$ are linearly independent. That is, f is an *allowable* mapping (Floater and Hormann, 2004). The transformation f is called *isometric*, or an *isometry*, if it preserves the metric tensor, that is $g(\mathbf{x}) = g'(\mathbf{x}')$. A classical result, known as Theorema Egregium (Gauss, 1827), claims that the Gaussian curvature can be expressed entirely in terms of the metric. Therefore, $K = K'$ is a necessary condition for an isometry.

From the point of view of Riemannian geometry, *isometric manifolds* which are two manifolds differing by an isometry, are indistinguishable. In the same way as g induces a distance metric on $\mathcal{M}$, the tensor g'

induces a distance metric on $\mathcal{M}'$. Consequently, f preserves the distance between every pair of points on $\mathcal{M}$. That is, for every $\mathbf{u}_1, \mathbf{u}_2 \in U$

$$d(\mathbf{x}(\mathbf{u}_1), \mathbf{x}(\mathbf{u}_2)) = d(\mathbf{x}'(\mathbf{u}_1), \mathbf{x}'(\mathbf{u}_2)). \tag{8}$$

Next, we apply this model to faces. Facial expressions result from the movement of mimic muscles (Ekman, 1973). We assume that natural deformations of the facial surface can be modelled as isometric transformations. In other words, facial expressions give rise to nearly isometric manifolds. This allows us to construct an expression-invariant representation of the face, based on its intrinsic geometric structure that can be captured by geodesic distances.

Remark 1. *Some extreme facial expressions, specifically, an open mouth, change the* topology *of the facial manifold. Such transformation can be dealt with by enforcing a fixed topology. For example, assuming that the mouth is shut and thereby "gluing" the lips when the mouth is open. Alternatively, we can assume the mouth to be always open, and then "disconnecting" the lips by introducing a cut in the surface when the mouth is closed. For simplicity, we assume hereinafter that expressions preserve the facial manifold topology.*

Verifying quantitatively that facial expressions are indeed isometric is possible by tracking a set of feature points on the facial manifold and measuring how the distances between them change due to facial expressions. Unfortunately, as noted in Section 1, there are very few points that can be located accurately on a human face. In order to overcome this difficulty, we placed 82 white round markers (approximately 2 mm in diameter) as invariant fiducial points, onto our subject's face. Such markers can be easily detected under a variety of facial expressions (Figure 4, left). In this experiment, we used 17 faces with weak, medium and strong facial expressions (see Figure 4, right)[3]. The reference points were manually labelled; then the Fast Marching algorithm (see Section 5) was used to compute the geodesic distances between them.

Figure 5 depicts a plot of the distances between the points in faces with different expressions as function of the corresponding distances in the reference "neutral" face (we used an average of the distances on 4 faces with neutral expression). Each point represents a distance; deviation from the diagonal represents the absolute error w.r.t. the reference distance. Ideally, all the points should lie along the diagonal line. We present two separate plots for weak and strong facial expressions.

[3] For the sake of consistency with our assumption of fixed topology of the facial manifold, we present here results on facial expressions with closed mouth only.

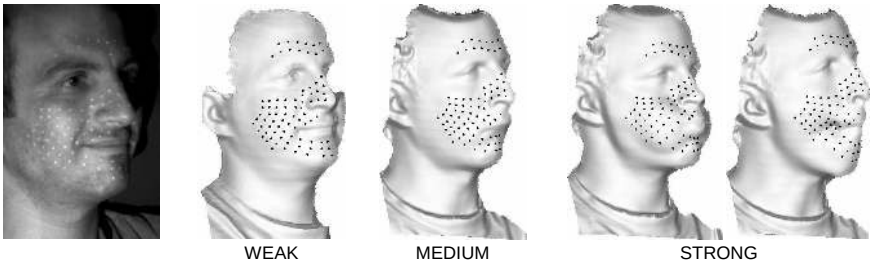

Figure 4. Facial expressions experiment. Left: facial image with the markers. Right: example of four facial expressions with marked reference points.

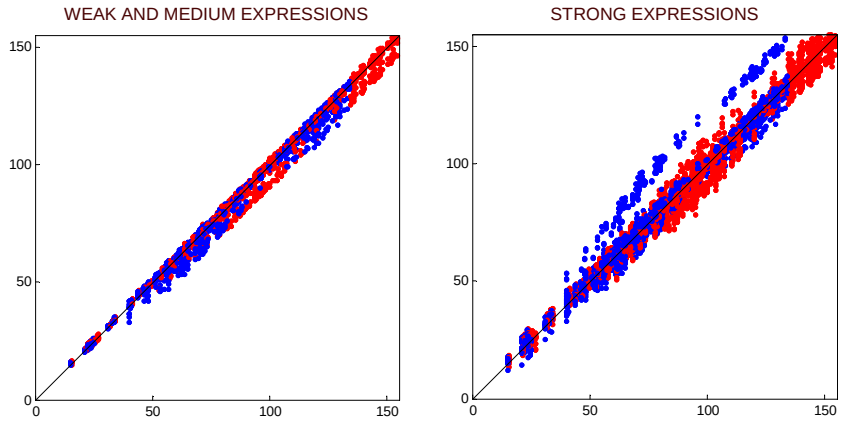

Figure 5. Geodesic (red) and Euclidean (blue) distances between the points in faces with medium (left) and extreme (right) expressions as function of the corresponding distances in the reference (neutral) face. Scale is given in mm.

In order to quantify the changes that the distances undergo, we use two measures: the absolute error w.r.t the reference distances: $\epsilon_i^{abs} = d_i - d_i^{ref}$, and the relative error $\epsilon_i^{rel} = (d_i - d_i^{ref})/d_i^{ref}$ (here d_i denotes the i-th distance and d_i^{ref} is the corresponding reference distance). The relative error is a more meaningful criterion, since on non-flat surfaces geodesic distances are larger than Euclidean ones (the mean of the geodesic distances was 61.23mm; while the mean of the Euclidean distances was 57.71mm, i.e. 6% smaller).

Table I presents the standard deviation and the mean absolute value of ϵ_i^{abs} and ϵ_i^{rel}. Figure 6 presents the standard deviation of ϵ_i^{abs} and ϵ_i^{rel} for distances divided into short (up to 50mm), medium ($50 - 100$mm) and long (greater than 100mm) ones. Comparing the behavior of Euclidean and geodesic distances (in sense of the standard deviation of ϵ^{rel}), we observe the following trends: First, geodesic distances outperform the Euclidean ones (i.e. result in smaller error) by up to 54.14%. Secondly, the factor by which geodesic distances outperform

Table I. Standard deviation and the mean absolute value of the absolute error ϵ_i^{abs} and the relative error ϵ_i^{rel}. Shown separately for weak and strong facial expressions.

	Weak & medium		Strong	
	Geodesic	Euclid.	Geodesic	Euclid.
Absolute error std. dev. (mm)	2.95	2.76	4.07	5.35
Absolute error abs. mean (mm)	1.97	2.00	2.89	3.10
Relative error std. dev. (%)	3.13	3.59	4.07	7.15
Relative error abs. mean (%)	2.14	2.62	3.26	4.26

the Euclidean ones grows (from 14.70% to 54.14%) when allowing for strong facial expressions. Finally, examining Figure 6 we conclude that the inaccuracy of the Euclidean distances as opposed to the geodesic ones appear more acutely in medium and large distances rather than short ones; this result is explainable since locally every curved surface is approximately flat.

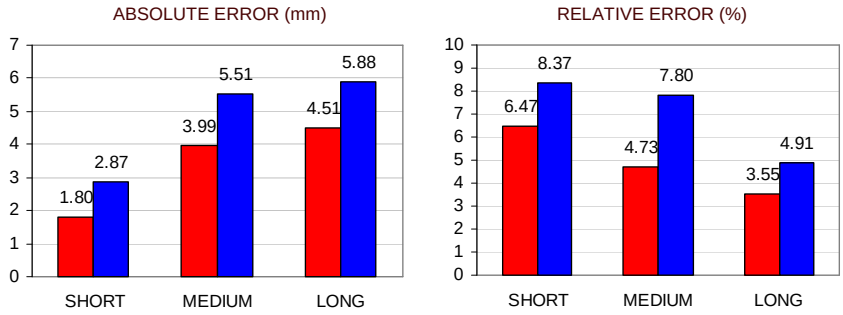

Figure 6. Standard deviation of the absolute error (ϵ^{abs}, in mm) and the relative error (ϵ^{rel}, in %) on short, medium and long distances. Red and blue represent geodesic and Euclidean distances, respectively.

The conclusion of this experiment is two-fold. First, the changes that geodesic distances undergo due to facial expressions are insignificant (up to 3.26% relative error on the average), which justifies our isometric model. Secondly, Euclidean distances are much more susceptible to changes due to facial expressions compared to the geodesic ones. This observation will be reinforced by the results in Section 8, where we compare our method based on the isometric model assumption to a method that treats facial surfaces as a rigid objects.

3. Three-dimensional data acquisition

We assume a Lambertian reflectance model, according to which the reflectance of an infinitesimal surface patch at $\mathbf{x}(\mathbf{u})$ with outward pointing unit normal vector $\mathbf{n}(\mathbf{u})$ and albedo $\rho(\mathbf{u})$ is given by

$$r(\mathbf{u}) = \max\left(0, \rho(\mathbf{u})\mathbf{n}(\mathbf{u}) \cdot \mathbf{l}(\mathbf{u})\right), \tag{9}$$

where $\mathbf{l}(\mathbf{u})$ is the light direction (in the general case, if the light source is not distant, i.e. the rays are not parallel, the light direction in every point $\mathbf{u}$ is different). Pixels in which $r(\mathbf{u}) = 0$ are shadowed.

3.1. Surface reconstruction

A camera, described by a perspective projection, maps the point $\mathbf{x}$ in the three-dimensional world coordinate system into the point $\mathbf{x}_c$ in the two-dimensional image plane coordinate system. All the points lying on the ray $\mathbf{o}_c\mathbf{x}_c$ (bold line in Figure 7) are projected to $\mathbf{x}_c$. In passive stereo, a second camera is used, adding another non-collinear view $\mathbf{o}_{c'}\mathbf{x}_{c'}$, that can be used to resolve this ambiguity by triangulation. The location of $\mathbf{x}$ is uniquely determined by the intersection point of the two rays $\mathbf{o}_c - \mathbf{x}_c$ and $\mathbf{o}_{c'} - \mathbf{x}_{c'}$). This procedure requires to determine the corresponding $\mathbf{x}_{c'}$ for every $\mathbf{x}_c$ (referred to as the *correspondence problem*).

In active stereo, the second view effect is obtained by an "active" projector rather than a second "passive" camera. A typical active stereo setup is shown in Figure 7. The projector is also described by a perspective projection, which maps $\mathbf{x}$ into $\mathbf{x}_p$ in a one-dimensional projector coordinate system. The projector casts a light code onto the object, which allows us to determine $\mathbf{x}_p$ corresponding to each point $\mathbf{x}_c$ in the camera image plane. World coordinates $\mathbf{x}$ are obtained unambiguously from the intersection of the ray $\mathbf{o}_c - \mathbf{x}_c$ and the plane $\mathbf{o}_p\mathbf{x}_p$, assuming that the latter are known (and in practice, obtained by a calibration procedure).

Active stereo techniques differ mainly by the illumination patterns used to encode the projection plane $\mathbf{x}_p$. Time- (Posdamer and Altschuler, 1982; Horn and Kiryati, 1999), gray level intensity- (Carrihill and Hummel, 1985), color- (Tajima and Iwakawa, 1990), and space-multiplexed (Hugli and Maitre, 1989; Vuylsteke and Oosterlinck, 1990) codes are commonly used. A Grey time-multiplexed coding scheme was adopted for facial surface acquisition in this work (Bronstein et al., 2003a). Positioning the camera and the projector in front of the subject, such that occlusions are avoided, allows us to capture all the facial surface.

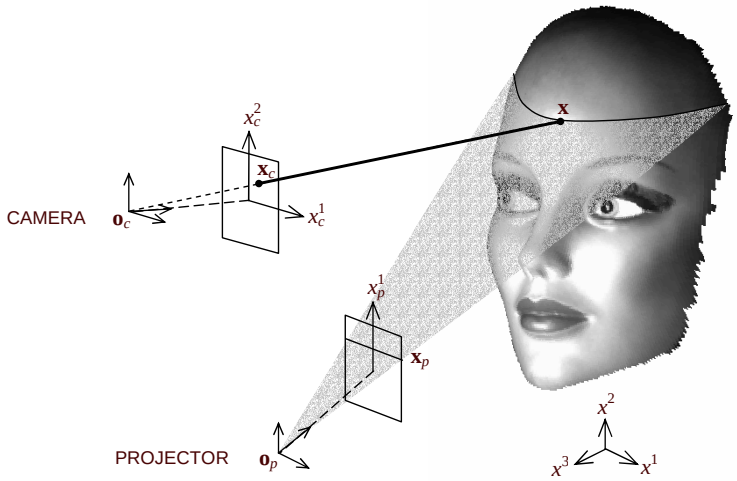

Figure 7. Active stereo: The location of the point $\mathbf{x}_w$ in the world coordinates cannot be determined only by its 2D coordinated $\mathbf{x}_c$ in the camera coordinate system. All points with the same $\mathbf{x}_c$ lie along a line (bold). If in addition, a 1D coordinate $\mathbf{x}_p$ of the point in the projector system is known, $\mathbf{x}$ can be computed by triangulation.

Remark 2. *Extreme rotations of the head may result in occlusions. In our current implementation, we use only a single camera, and thus limit the head poses to bounded deviations from the frontal position. This limitation is merely technical - if insensitivity to larger angles is required, a multiple-view acquisition* [see e.g. (Hung et al., 1999)] *can be used.*

Under the assumption of a frontal view, we identify the image coordinates with the parameterizations coordinates (u^1, u^2), which are now partitioned into a uniform grid of N_0 pixels. Then, for every pixel $\mathbf{u}_i$ $(i = 1, ..., N_0)$ we have an estimate of $\mathbf{x}(\mathbf{u}_i)$. In other words, the scanner produces a sampled version $\{\mathbf{x}_i = \mathbf{x}(\mathbf{u}_i)\}_{i=1}^{N_0}$ of the facial manifold. Note that though the parameterizations plane is sampled on a regular grid of pixels, the samples along x^1 and x^2 are neither necessarily regular nor uniform. Specifically, in our implementation, the range data is stored in three double-precision matrices, each of size 320×240, corresponding to the values of x^1, x^2 and x^3 in each pixel. Thereby, the scanner output is a cloud of $N_0 = 76.8 \times 10^3$ points in 3D.

The active stereo setup allows albedo estimation from the reflectance image $r(\mathbf{u}_i)$ observed by the camera, under the assumptions of Lambertian reflection and linear camera response. First, the world coordinates

of a point $\mathbf{x}(\mathbf{u}_i)$ and the normal to the surface $\hat{\mathbf{n}}(\mathbf{u}_i)$ at that point are estimated. Next, illumination direction and relative intensity $\mathbf{l}(\mathbf{u}_i)$ are extracted. The illumination direction is given by $\mathbf{x}(\mathbf{u}_i) - \mathbf{o}_p$, and the intensity is dictated by the structure of the projected light code. The albedo in pixel i is given by

$$\rho_i = \frac{r(\mathbf{u}_i)}{\mathbf{l}(\mathbf{u}_i) \cdot \hat{\mathbf{n}}(\mathbf{u}_i)}, \tag{10}$$

where r_i is the reflectance observed in pixel $\mathbf{u}_i$.

3.2. Metric reconstruction

Active stereo methods estimate the absolute location of each and every point on the facial surface. As an alternative, it is possible to estimate the gradient $\nabla x^3(\mathbf{u})$, or the normal field $\mathbf{n}(\mathbf{u})$ of the surface. In this case, the natural parameterizations of the manifold is $(u^1, u^2, x^3(u^1, u^2))$.

One of the classical method for recovering the field of normals under the assumption of Lambertian model is the *photometric stereo* approach. Given K reflectance images $r^{(1)}(\mathbf{u}_i), ..., r^{(K)}(\mathbf{u}_i)$ acquired under illumination by a distant point light source at different directions $\mathbf{l}^{(1)}, ..., \mathbf{l}^{(K)}$, it is possible to estimate the unit normal vector in every pixel (Gheorghiades et al., 2001). Proesmans et al. (1996) and Winkelbach and Wahl (2001) proposed a method to recover the surface gradient field by projecting two stripe patterns and studying their deformations.

Given the surface gradients $\partial_1 x^3, \partial_2 x^3$, the surface $x^3(\mathbf{u})$ can be reconstructed by solving the Poisson equation (Gheorghiades et al., 2001; Kimmel and Yavneh, 2003). However, as we will see, our algorithm does not require the facial surface in an explicit form, and thus the gradient field integration stage can be avoided.

4. Preprocessing

The points cloud obtained from the scanner undergoes processing by cropping and smoothing. The first stage is required to extract the approximate facial contour and remove problematic points in which the reconstruction is inaccurate. Those usually appear as spikes (see Figure 8, second column). The facial contour is defined by a binary *mask* image $\mu_i = \mu(\mathbf{u}_i)$. The mask computation involves several thresholds, defined according to the fidelity level and the value of the discrete gradient norm $\|\nabla x^3\|_2^2 = \partial_i x^3 \partial^i x^3$ (both are used to remove the spikes),

and the depth histogram, that allows us to separate the face from the background. Finally, morphological operations are applied to the mask in order to remove non-connected regions and isolate the facial contour as a single object. Holes inside the facial contour are closed by interpolation. We use $\Omega_1 = \{\mathbf{u}_i : \mu_i = 1\}$ to denote the pixels in the facial contour, and denote the number of pixels in the mask by $|\Omega_1| = N_1$.

4.1. Beltrami flow

The second stage is required to remove the quantization noise of the scanner. We use a variant of the Beltrami flow for edge-preserving selective smoothing, which was proven to be a powerful method for color image processing (Sochen et al., 1998; Kimmel, 2003). The method is efficient and does not require the surface to be given explicitly.

Our smoothing procedure resembles the anisotropic diffusion in (Tasdizen et al., 2002; Fleishman et al., 2003). The key idea of such smoothing is iterative minimization of the functional

$$S(\mathbf{x}, g) = \int_U d^2\mathbf{u}\sqrt{g}g^{ij}\partial_i x^i \partial_j x^j \delta_{ij}, \tag{11}$$

which can be considered as a generalization of the L_2 norm to curved spaces. The minimum of functional S stems from the Euler-Lagrange conditions and results in the following gradient descent evolution equation

$$\partial_t x^i = \frac{1}{\sqrt{g}}\frac{\delta S}{\delta x^i}; \;\; i = 1, ..., 3; \tag{12}$$

where t is a time parameter, and the manifold evolves in time, according to its "mean curvature flow" towards a minimal surface. Our manifold is given as a graph of a function $x^3(x^1, x^2)$, filtering just the depth coordinate x^3 according to this result gives rise to the partial differential equation

$$\partial_t x^3 = \frac{1}{\sqrt{g}}\partial_i(\sqrt{g}g^{ij}\partial_j x^3); \;\; i, j = 1, 2. \tag{13}$$

where g^{ij} is the inverse metric tensor. This PDE is a non-Euclidean version of the diffusion equation (Spira et al., 2004). Smoothing is performed by numerical solution of (13) for a time proportional to the noise variance, a process that can be implmented very efficiently. The processing is limited to the facial contour Ω_1; Neumann boundary conditions $\nabla x^3|_{\partial\Omega_1} = \mathbf{0}$ are used.

Figure 8. Pre-processing of the 3D scanner data, left to right: points cloud (sub-sampled 1:5); facial surface before processing; extracted facial contour; Beltrami smoothing result. Surfaces are rendered using Phong shading.

When the facial manifold is not given explicitly, and all we have is its gradients on a uniform grid $(x^1, x^2, x^3) = (u^1, u^2, x^3(u^1, u^2))$, it is possible to evolve the gradient field instead of the surface.

5. Measuring geodesic distances on the facial manifold

The next step in the processing requires measuring geodesic distances on the facial surface. We use the Fast Marching Method (FMM), which measures the distances from one point to the rest of the N_1 surface points in $\mathcal{O}(N_1)$ operations.

The FMM algorithm, introduced by Sethian (1996), is based on upwind finite difference approximations for solving the *eikonal equation*

$$\|\nabla\varphi(\mathbf{x})\| = 1, \tag{14}$$

where φ is the distance map from the sources $\mathbf{s}_1$, such that $\varphi(\mathbf{s}_1) = 0$. The FMM was extended to triangulated manifolds by Kimmel and Sethian (1998).

The classical FMM uses an orthogonal coordinate system (regular grid). The numerical stencil for an update of a grid point consists of vertices of a right triangle. In the case of triangulated manifolds, the stencils used by the Fast Marching algorithm are not necessarily right triangles. If a grid point is updated through an obtuse angle, a consistency problem may arise. To cope with this problem, Kimmel and Sethian proposed to split obtuse angles by unfolding triangles as a preprocessing stage.

A variation of the FMM for parametric manifolds was presented by Spira and Kimmel (2003, 2004). The main advantage of this method is that the computations are done on the uniform Cartesian grid in the parametrization plane, and not on the manifold like in the original version of Kimmel and Sethian. The numerical stencil is calculated

directly from the local metric value, and therefore, no unfolding is required (see details in (Spira and Kimmel, 2004)). In our application, the advantage of using the parametric FMM is that the surface is not required explicitly in order to measure distances; and the metric given on the parametric plane is sufficient (Bronstein et al., 2004b).

5.1. Geodesic mask

The geodesic mask is defined as $\Omega_2 = \{\mathbf{u}_i \in \Omega_1 : \min_l d(\mathbf{x}_i, \mathbf{s}_l) \leq R\}$, where $\mathbf{s}_1, ..., \mathbf{s}_L$ are some source points. Roughly speaking, it is a the interior of a "geodesic circle" with radius R. We denote $|\Omega_2| = N$ - the number of points remaining after all the processing stages.

The source points are chosen from some invariant features, such as the tip of the nose or the eye sockets. Such points can be identified robustly from the geometric information. Specifically, we used two sources: the nose tip and the bridge of the nose. The geodesic mask is computed by measuring geodesic distances from $\mathbf{s}_1, ..., \mathbf{s}_L$ to all the points $\mathbf{x}_i$ $(i = 1, ..., N_1)$ in the mask Ω_1, computed at the preprocessing stage, using the Fast Marching. Then, we crop the points for which the distance exceeds some radius R. A typical value of R for a human face is $90 - 110$ mm.

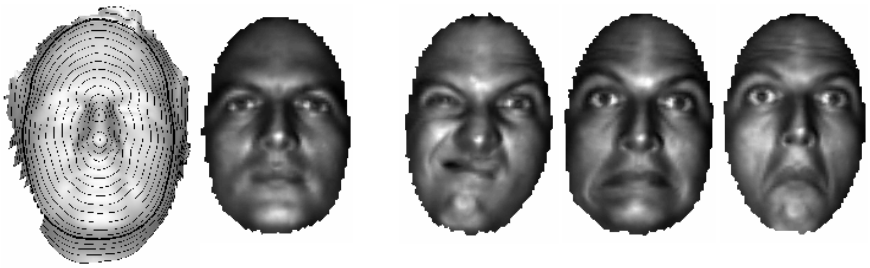

Figure 9. Computation of the geodesic mask. First column: Equidistant contours from two sources located at the tip of the nose and the center of the bridge of the nose. Second column: The resulting geodesic mask. Third through fifth columns: Examples of the geodesic mask insensitivity to facial expressions.

6. Bending-invariant representation

A curved non-Euclidean facial manifold is not a convenient object to process. We would like to find a better representation of the manifold which would be (i) convenient to analyze (Euclidean space is one suitable candidate); and (ii) identical for all isometric manifolds.

A similar problem arouse hundreds of years ago in the context of map making - namely, how to map the spherical surface of the Earth

onto a plane without distorting the distances between geographical objects? (see Figure 10). Such a mapping is also known as *flat embedding*. From the Theorema Egregium it follows that a sphere and a plane are not isometric manifolds, since the plane has zero Gaussian curvature, while the sphere has a positive one. Hence, flat embedding is possible only with some errors. Similarly, when trying to embed an arbitrary curved manifold into any flat space, there will always exist some metric distortion. The question then is how to minimize these distortions.

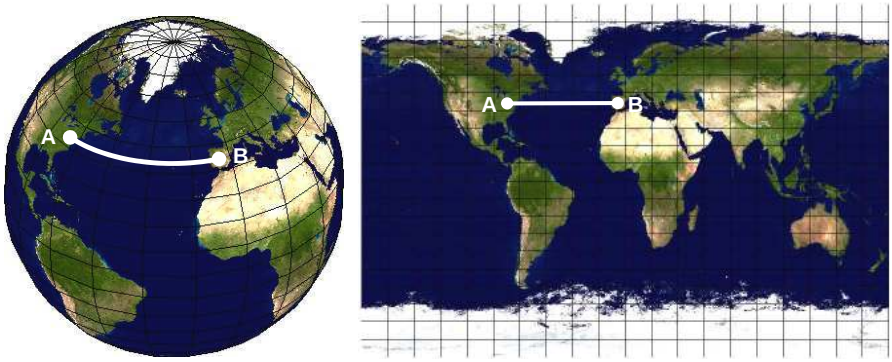

Figure 10. The classical flat embedding problem in map-making: The spherical surface of the Earth (left) is converted into a planar map (right), trying to preserve the distances between every pair of points (here denoted by A and B). Geodesic distances are approximated by Euclidean ones.

We would like to find the "most isometric" embedding, the one that deforms the manifold distances the least. In practice, we have a finite discrete set of N manifold samples $\{\mathbf{x}_i\}_{i=1}^N$ (represented as a $3 \times N$ matrix $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_N)$) and a set of N^2 mutual geodesic distances between these samples. The geodesic distances are represented as a matrix $\mathbf{D}$ with $d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$. We consider a mapping of the form $\varphi : (\mathcal{M}, d) \to (\mathbb{R}^m, d')$, which maps the manifold samples $\mathbf{x}_i$ into points $\mathbf{x}'_i$ in an m-dimensional Euclidean space, such that the geodesic distances d_{ij} are replaced by Euclidean ones $d'_{ij} = \|\mathbf{x}'_i - \mathbf{x}'_j\|_2$.

Obviously, as isometric transformations preserve distances, d_{ij} remain unchanged. The problem is that the discrete set of points $\{\mathbf{x}_i\}_{i=1}^N$ can be ordered arbitrarily, and thus the matrix $\mathbf{D}$ is invariant up to a permutation of rows and columns. Moreover, when sampling a surface, there is no guarantee to sample the surface at similar points. This makes the computation of such an invariant impractical, though we refer to the most recent attempt to directly compare the geodesic matrices (Mémoli and Sapiro, 2004). Flat embedding (though inevitably being an *approximate* representation of the original surface), on the other hand, is invariant up to arbitrary rotation, translation and reflection,

which is much easier to cope with. We will refer to the embedding result $\mathbf{X}'$ as the *bending invariant canonical form* of the manifold.

The embedding error can be measured as a discrepancy between the geodesic and the resulting Euclidean distances using some norm,

$$s(\mathbf{X}';\mathbf{D}) = \|\mathbf{D} - \mathbf{D}'(\mathbf{X}')\|, \tag{15}$$

where $\mathbf{X}' = (\mathbf{x}'_1, ..., \mathbf{x}'_N)$ is an $m \times N$ matrix representing the points in the embedding space, and $\mathbf{D}'$ is the matrix of mutual Euclidean distances depending on the points configuration $\mathbf{X}'$. The function (15) is sometimes referred to as *stress*.

Finding the best approximate flat embedding is possible by minimization of $s(\mathbf{X}';\mathbf{D})$ w.r.t. $\mathbf{X}'$. A family of algorithms used to carry out such approximate flat embedding is usually referred to as *multidimensional scaling* (MDS). These algorithms differ in the choice of the embedding error criterion and the method used for its minimization.

6.1. Multidimensional scaling

One of the first MDS algorithms proposed by Torgerson (1952) and Gower (1966), based on earlier theoretical results of Eckart and Young (1936) and Young and Householder (1938), is now known as the *classical scaling*. This method, being merely algebraic, is very efficient. Another advantage is that the embedding coordinates in classical scaling can be expressed analytically.

The method works with squared geodesic distances, which can be expressed as Hadamard product $\mathbf{\Delta} = \mathbf{D} \circ \mathbf{D}$ (coordinate-wise product of matrices). The matrix $\mathbf{\Delta}$ is first double-centered

$$\mathbf{B} = -\frac{1}{2}\mathbf{J}\mathbf{\Delta}\mathbf{J} \tag{16}$$

(here $\mathbf{J} = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathbf{T}}$ and $\mathbf{I}$ is an $N \times N$ identity matrix). Then, the eigendecomposition $\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ is computed, where $\mathbf{V} = (\mathbf{v}_1, ..., \mathbf{v}_N)$ is the matrix of eigenvectors of $\mathbf{B}$ corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$. Denoting by $\mathbf{\Lambda}_+$ the matrix of first m positive eigenvalues and by $\mathbf{V}_+$ the matrix of the corresponding eigenvectors, the coordinate matrix in the embedding space is given by

$$\mathbf{X}' = \mathbf{V}_+\mathbf{\Lambda}_+. \tag{17}$$

Explicitly,

$$\mathbf{x}'_i = \left(\sqrt{\lambda_1}v_1^i, ..., \sqrt{\lambda_m}v_m^i\right), \quad i = 1, ..., N, \tag{18}$$

with v_j^i denoting the i-th coordinate of the vector $\mathbf{v}_j$ where $j = 1, ..., m$. The set of points $\mathbf{x}'_i$ obtained by classical scaling is referred to as the

classical canonical form of the discrete manifold, in order to distinguish it from canonical forms obtained by other MDS methods. In practice, since we are usually interested in embedding into $\mathbb{R}^3$ or $\mathbb{R}^2$, no full eigendecomposition of $\mathbf{B}$ is needed - it is enough to find only the first three or even two eigenvectors. The method of Arnoldi iterations (Arnoldi, 1951), for example, performs this task efficiently.

It can be easily shown [see (Borg and Groenen, 1997)] that classical scaling minimizes the Frobenius norm

$$s(\mathbf{X}) = \left\| \mathbf{V} \left(\mathbf{\Lambda} - \begin{bmatrix} \mathbf{\Lambda}_+ & \\ & \mathbf{0} \end{bmatrix} \right) \mathbf{V}^T \right\|_F, \tag{19}$$

which gives the possibility to define the embedding error as a squared sum of the eigenvalues $\lambda_{m+1}, ..., \lambda_N$ ignored in the embedding procedure.

However, such definition of the metric distortion is not intuitive. One straightforward possibility is to choose the norm in (15) to be Euclidean, and thus have the metric distortion defined as a sum of squared differences

$$s(\mathbf{X}'; \mathbf{D}) = \sum_{i>j} (d_{ij} - d'_{ij})^2, \tag{20}$$

and the MDS is posed as a least-squares (LS) problem. Such a function is non-convex and if minimized by conventional convex optimization algorithms, convergence to a global minimum cannot be guaranteed. An algorithm for stress minimization based on *iterative majorization* was proposed by De-Leeuw (1977) and is known as the SMACOF algorithm. The key idea of majorization is replacing the non-convex stress $s(\mathbf{X}, \mathbf{D})$ by some convex (e.g. quadratic) *majorizing function* and minimizing it w.r.t. $\mathbf{X}$. It can be shown (De-Leeuw and Stoop, 1984) that such optimization converges linearly to the global minimum $\mathbf{X}'$, which we call the *LS canonical form*.

When the embedding is performed into a space with dimensionality $m \leq 3$, the canonical form can be plotted as a surface (Figure 11, third column). Figure 12 depicts canonical forms of one subjects with different facial expressions. It demonstrates that while the facial surface changes are dramatic, the changes between the corresponding canonical forms are small. Extreme facial expressions deviating from our isometric model make the corresponding canonical forms deviate from the neutral one. Yet, the changes in the canonical forms are significantly smaller than the changes between the original facial surface - this point will be exemplified experimentally in Section 8.

Embedding into $\mathbb{R}^2$ is a special case - it can be thought of as "warping" of the facial texture (Figure 11, fourth column). This serves as a way of performing geometry-based registration of 2D facial images.

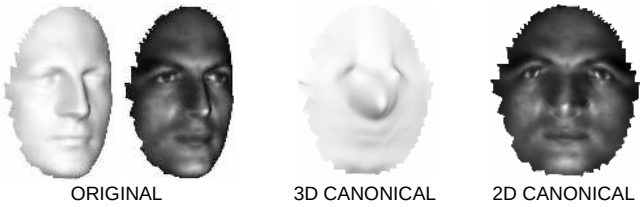

Figure 11. Examples of LS canonical forms of a face. When the embedding is performed in $\mathbb{R}^2$, the MDS procedure can be thought of as warping of the facial texture (fourth column).

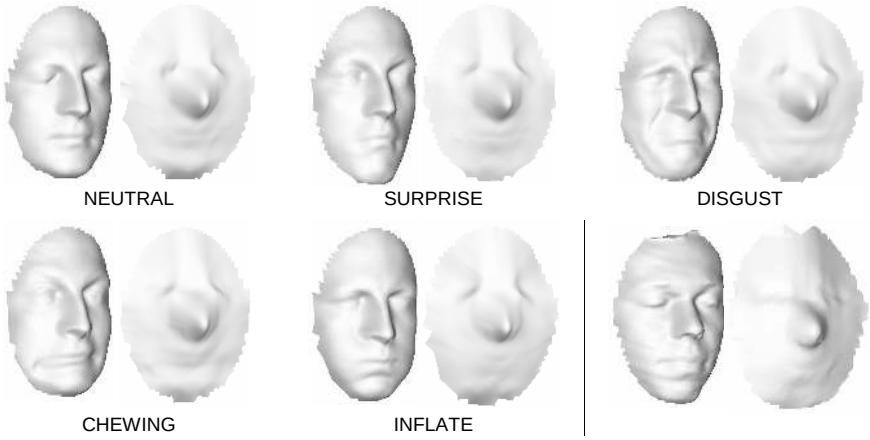

Figure 12. Insensitivity of canonical forms to facial expressions. The canonical forms appear less susceptible to changes compared to the original facial surfaces. For comparison, the face of a different subject and its canonical form are shown.

6.2. Stability analysis

An important property that makes the canonical forms practically useful is that the embedding result $\mathbf{X}'$ changes continuously with the change of $\mathbf{X}$. This guarantees that a small perturbation in the points of the manifold does not change significantly the canonical form.

We show here stability of both the classical and the LS canonical forms; while the second is straightforward (Theorem 2), the first requires more delicate analysis involving matrix perturbation theory. The proofs are given in Appendices .1–.2.

Theorem 1. *Let* $\{\mathbf{x}_i\}_{i=1}^N$ *be samples of the manifold with geodesic distances* d_{ij}*, such that the double-centered matrix* $\mathbf{B}$ *has non-degenerate*

eigenvalues $\lambda_1 < \lambda_2 < ... < \lambda_N$. *Then, a small perturbation of a point* $\mathbf{x}_i$ *results in a small perturbation of the classical canonical form* $\{\mathbf{x}'_i\}_{i=1}^N$.

Remark 3. *The assumption of non-degeneracy of the spectrum of* $\mathbf{B}$ *is valid for curved surfaces such as the facial surface.*

Theorem 2. *Let* $\{\mathbf{x}_i\}_{i=1}^N$ *be samples of the manifold with geodesic distances* d_{ij}. *Then, a small perturbation of a point* $\mathbf{x}_i$ *results in a small perturbation of the least squares canonical form* $\{\mathbf{x}'_i\}_{i=1}^N$.

7. The 3D face recognition system

Bringing all the building blocks described above, we obtain a 3D face recognition system with data flow as depicted in Figures 13 – 14. Figure 13 shows the basic preprocessing stages. The subject's facial surface is first scanned, producing the surface and the reflectance image; due to the acquisition method used, there is a one-to-one correspondence between each point of the surface and each pixel of the image. At the next stage, both the surface and the image are cropped in order to remove the background and thus reduce the number of points and remove the spikes. On the third stage, holes in the surface, that are created by the spike removal procedure, are closed and the surface is selectively smoothed to eliminate the quantization artifacts of the scanner. Using the smoothed surface, the normal field is estimated and used to recover the facial albedo from the reflectance image. At the fourth stage, the surface and the albedo are sub-sampled. We used a factor of 0.35 in both axes. Finally, the surface is triangulated, the nose is located by robust correlation with a 3D template; and a second cropping based on the geodesic mask around the nose is performed. As an alternative, the above stages can be done without the surface reconstruction. Figure 14 depicts an example of such processing when photometric stereo is used for 3D acquisition.

The surface obtained at the preprocessing stage undergoes flattening by MDS to yield a canonical form. The corresponding albedo image is flattened as well. However, since embedding is defined up to an Euclidean and reflection transformation, the canonical surface must be aligned. We performed the alignment by first setting to zero the first-order moments (the center of gravity) $m_{100}, m_{010}, m_{001}$ of the canonical surface to resolve the translation ambiguity (here

$$m_{pqr} = \sum_{i=1}^{N} (x_i^1)^p (x_i^2)^q (x_i^3)^r \tag{21}$$

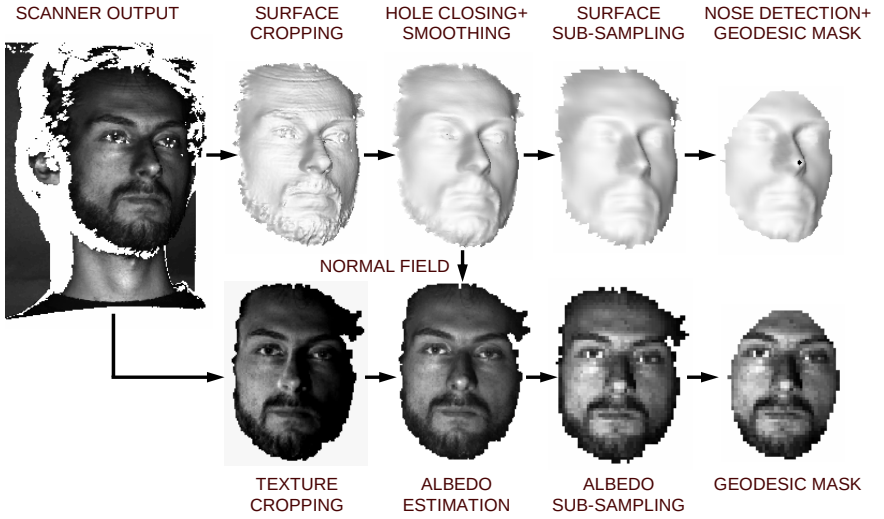

Figure 13. Preprocessing stages in 3D face recognition. Geometry processing (first row, left to right): Raw data acquisition by a 3D scanner; surface cropping; filling in the holes and selective smoothing; sub-sampling; nose detection and geodesic mask computation. Texture processing (second row, left to right): Cropping; illumination compensation; texture sub-sampling; cropping by geodesic mask.

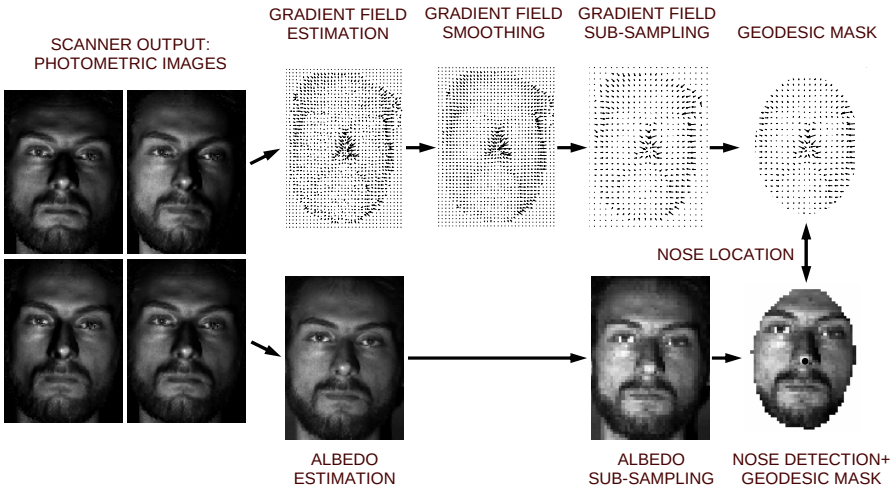

Figure 14. Preprocessing stages in "surface-less" 3D face recognition with photometric stereo scanner. Geometry processing (first row, left to right): Gradient field reconstruction from photometric images; gradient field smoothing; sub-sampling; computation of the geodesic mask by fast marching on parametric manifolds. Texture processing (second row, left to right): Albedo estimation from photometric images; sub-sampling; nose detection and cropping by geodesic mask.

denotes the pqr-th moment); then, the mixed second-order moments $m_{110}, m_{011}, m_{101}$ were set to zero to resolve the rotation ambiguity.

Finally, using the coordinate relations of three fiducial points[4] on the face (e.g. that the nose points to the positive direction of the z-axis, and the x-coordinate of the left eye is smaller than that of the right one) the reflection ambiguity is resolved.

7.1. Surface matching

The final stage of the face recognition algorithm is surface matching. Since the flattening compensates for the isometric surface transformations, standard rigid matching (see e.g. (Gruen and Akca, 2004)) can be used for comparing the canonical surfaces. One of the simplest and most efficient surface matching method proposed by Tal et al. (2001) is based on high-order moments of the surface. The main idea is to represent the surface by its moments m_{pqr} up to some degree P, and compare the moments as vectors in an Euclidean space. Given two canonical surfaces $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N_x}$ and $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{N_y}$, this allows to define the distance between two faces as

$$d_{mom}(\mathbf{X}, \mathbf{Y}) = \sum_{p+q+r \leq P} (m^x_{pqr} - m^y_{pqr})^2. \tag{22}$$

This distance can be improved by allowing for some alignment transformation between the surfaces, e.g.

$$d_{alig}(\mathbf{X}, \mathbf{Y}) = \min_{\mathbf{R},\mathbf{C}} d_{mom}(\mathbf{X}, \mathbf{RY} + \mathbf{C}), \tag{23}$$

where $\mathbf{R}$ is a rotation matrix and $\mathbf{C}$ is a translation vector. However, such metric is inefficient in practice, since an optimization algorithm is required to find the optimal alignment.

In (Bronstein et al., 2003b; Bronstein et al., 2004a) we proposed to treat canonical forms as images. After alignment, both the canonical surface and the flattened albedo are interpolated on a Cartesian grid, producing two images. These images can be compared using standard techniques, e.g. applying eigendecomposition like in eigenfaces or eigenpictures. The obtained representation was called in (Bronstein et al., 2003b) *eigenforms*. The use of eigenforms has several advantages: First, image comparison is simpler than surface comparison, and second, the 2D texture information can be incorporated as a second image in a natural way. In this paper, however, we focus on the 3D geometry, and in the following experiments use only the surface geometry ignoring the texture in 3D face recognition.

[4] The fiducial points are not necessarily facial features; any three points with known coordinate relations can be used.

7.2. Implementation issues

In our implementation, a time multiplexed coded-light method with 10 bit binary Gray code was built and used. Point Grey CCD grayscale camera (640×480 at 60 fps) was used as an acquisition device; Compaq MP2800 DLP projector was used to project the patterns. One 3D scan takes about 150 msec. The resulting raw data was down-sampled to a resolution of 320×240, and then all the above processing filters were applied.

The speed of the scan and the low intensity of the projected light does not cause any unpleasant sensation. A preferable configuration, however, should be invisible (e.g. infra- or near-infrared) light.

8. Results

The data set in our experiments contained 220 faces of 30 subjects - 3 artificial (mannequins) and 27 human. Most of the faces appeared in a large number of instances with different facial expressions. Facial expressions were classified into 10 groups (smile, anger, etc.) and into 4 strengths (neutral, weak, medium, strong). Neutral expressions are the natural postures of the face, while strong expressions are extreme postures. Small head rotations (up to about 10 degrees) were allowed. Since the data was acquired in a course of several months, variations in illumination conditions, facial hair, etc. present. Subjects `Alex` (blue) and `Michael` (red) are identical twins; subjects `Eyal` (dark green) and `Noam` (yellow) are brothers, having visually great similarity (see Figure 15).

8.1. Experiment I - Facial expressions

The goal of the first experiment was to demonstrate the difference between using original facial surfaces and their canonical forms for face recognition under strong facial expressions. Surface matching based on moments of degree up to $P = 5$ (i.e. vectors of dimensionality 216), according to (21), was used. In Experiment I, we used a subset containing 10 human and 3 artificial subjects (Figure 15). Each face appeared in a number of instances (a total of 204 instances), including neutral, weak, medium and strong facial expressions (Figure 16).

Figure 17 visualizes the dissimilarities between faces using LS MDS. Each face on this plot is represented by a point in $\mathbb{R}^2$. Note that it is merely a 2D representation of data originally lying in $\mathbb{R}^{216}$ (it captures about 88% of the high-dimensional information). The first row depicts the dissimilarities between faces with neutral expressions only. The

faces of different subjects (marked with different colors) form tight clusters and are easily distinguishable. Canonical surface matching (left) and facial surface matching (right) methods produce approximately the same results.

This idealistic picture breaks down when we allow for facial expressions (Figure 17, second row). The clusters corresponding to canonical surface matching are much tighter; moreover, we observe that using original surface matching some clusters (red and blue, dark and light magenta, light blue, yellow and green) overlap, which means that a face recognition algorithm based on original surface matching would confuse between these subjects.

Table II shows the values of the ratio of the maximum inter-cluster to minimum intra-cluster dissimilarity

$$\varsigma_k = \frac{\max_{i,j\in\mathcal{C}_k} \eta_{ij}}{\min_{i\notin\mathcal{C}_k, j\in\mathcal{C}_k} \eta_{ij}}, \tag{24}$$

and the ratio of root mean squared (RMS) inter-cluster and intra-cluster dissimilarities

$$\sigma_k = \sqrt{\frac{\frac{2}{|\mathcal{C}_k|^2-|\mathcal{C}_k|}\sum_{i,j\in\mathcal{C}_k,\ i>j} \eta_{ij}^2}{\frac{1}{|\mathcal{C}_k|(|\mathcal{C}|-|\mathcal{C}_k|)}\sum_{i\notin\mathcal{C}_k, j\in\mathcal{C}_k} \eta_{ij}^2}}, \tag{25}$$

($\mathcal{C}_k$ denotes indexes of k-th subject's faces, $\mathcal{C} = \bigcup_k \mathcal{C}_k$ and η_{ij} denotes dissimilarities between faces i and j) for facial and canonical surface matching. These criteria are convenient being scale-invariant; they measure how each cluster is tight and far from other clusters. Ideally, σ_k and ς_k should tend to zero.

When only neutral expressions are used, straightforward facial surface matching outperforms canonical form matching by 4.3% in sense of σ_k and up to 4.5% in sense of ς_k on the artificial subject `Eve`; on other subjects, our algorithm shows slightly better performance than facial surface matching. However, when allowing for facial expressions, our approach outperforms facial surface matching by up to 94.2% in sense of σ_k and up to 181.1% in sense of ς_k.

8.2. Experiment II - comparison of algorithms

The goal of the second experiment was to compare our method to other face recognition algorithms. For this purpose, we simulated a real-life situation of a biometric identification, in which a the face of an enrolled subject is compared to a set of templates stored in the database. The number of templates was 65. Only neutral expressions were used as templates.

Table II. Description of the facial expressions in data set used in experiment I (N - neutral, W - weak, M - medium, S - strong) and the inter-cluster to intra-cluster dissimilarity ratios σ_k and ς_k using original and canonical surface matching. Asterisk denotes brothers. Double asterisk denotes identical twins. Triple asterisk denotes artificial subjects.

Subject	Color	N	W	M	S	σ_k^{can}	σ_k^{orig}	ς_k^{can}	ς_k^{orig}
Michael**	red	6	5	6	-	0.2376	0.2730	171.22	284.97
Alex**	blue	3	1	3	1	0.1682	0.2433	88.55	171.88
Eyal*	green	4	1	7	9	0.1740	0.3113	106.14	388.72
Noam*	yellow	3	-	-	7	0.1983	0.2136	111.27	182.90
Moran	magenta	4	-	4	10	0.1405	0.2732	99.74	272.55
Ian	orange	5	-	16	7	0.1237	0.2584	80.34	220.26
Ori	cyan	8	-	11	10	0.2084	0.3406	158.11	352.74
Eric	d. green	5	3	-	3	0.1879	0.3876	117.14	432.21
Susy	d. magenta	6	-	9	8	0.1394	0.1933	109.38	224.62
David	l. blue	5	2	6	5	0.2040	0.7697	142.92	1129.50
Eve***	black	6	-	-	-	0.0100	0.0115	1.51	1.39
Benito***	grey	7	-	-	-	0.0789	0.0923	45.33	87.35
Liu***	l. grey	8	-	-	-	0.0935	0.1077	49.20	8.75

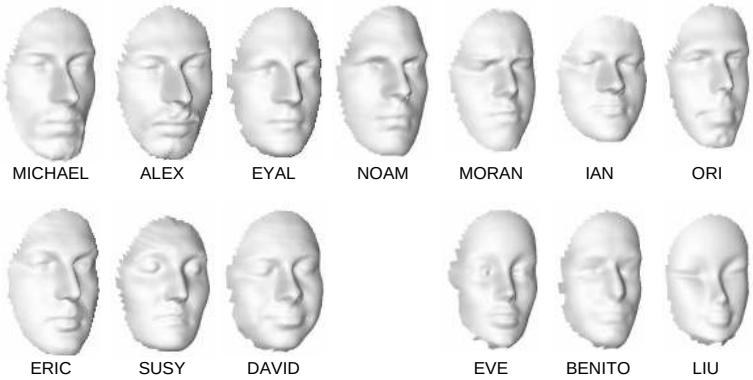

Figure 15. The subjects used in experiment I (shown with neutral expressions). Second row right: three artificial subjects.

The experiment was divided into two parts: First, faces with neutral expressions (total of 87) were used for enrollment; next, faces with all expressions, including extreme ones (total of 220 faces) were enrolled. Three algorithms were tested: canonical form matching, facial surface matching and 2D image-based eigenfaces. Eigenfaces were trained by 35 facial images that do not appear as templates. 23 eigenfaces were used

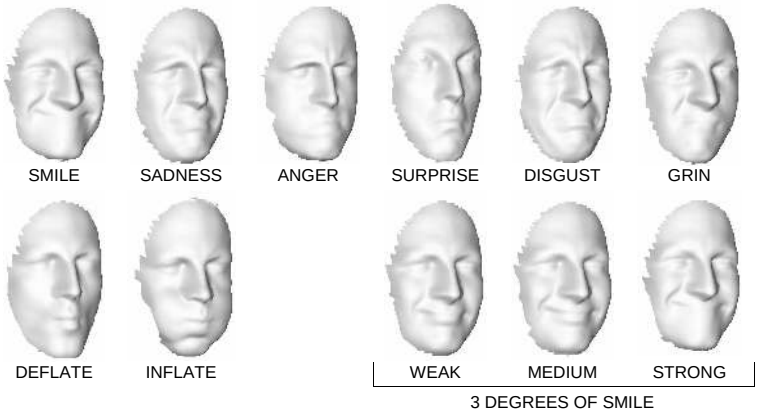

Figure 16. Eight representative facial expressions of subject Eyal. Second row right: three degrees of the smile expression.

for the recognition (the first two eigenfaces were excluded in order to decrease the influence of illumination variability (Gheorghiades et al., 2001)).

Figure 18 shows the receiver operation characteristic (ROC) curves of three algorithms participating in the comparison. The plot represents the tradeoff between the false rejection (FRR) and the false acceptance (FAR) rate, as function of a threshold determining whether the enrolled subject is accepted or rejected. Our algorithm significantly outperforms both the straightforward 3D face recognition (rigid facial surface matching) and the classical 2D algorithm (eigenfaces). Moreover, it shows very low sensitivity to facial expressions - the equal error rate (EER) increases from 1.71% to 2.1% when facial expressions are introduced. This is in contrast to the eigenfaces algorithm, for example, whose performance deteriorates dramatically (by 8.7%) when the database contains facial images with variability due to facial expressions.

For comparison, we should note that a typical performance of a commercial 2D face recognition system tested on about 200 faces shows EER of at least 6% (Mansfield et al., 2001). The performance of our algorithm (even in the very challenging situation when extreme facial expressions are allowed for) approaches the EER typical for fingerprint recognition ($\sim 2\%$), a biometric technology which is traditionally believed to be more accurate than face recognition (Ortega-Garcia et al., 2004).

Figure 19 shows the rank 1, 2 and 3 recognition errors on the full database with facial expressions using the above algorithms. Our approach results in *zero* recognition error. Figure 20 (first row) shows an example of recognition in the second setup (220 faces). The first column

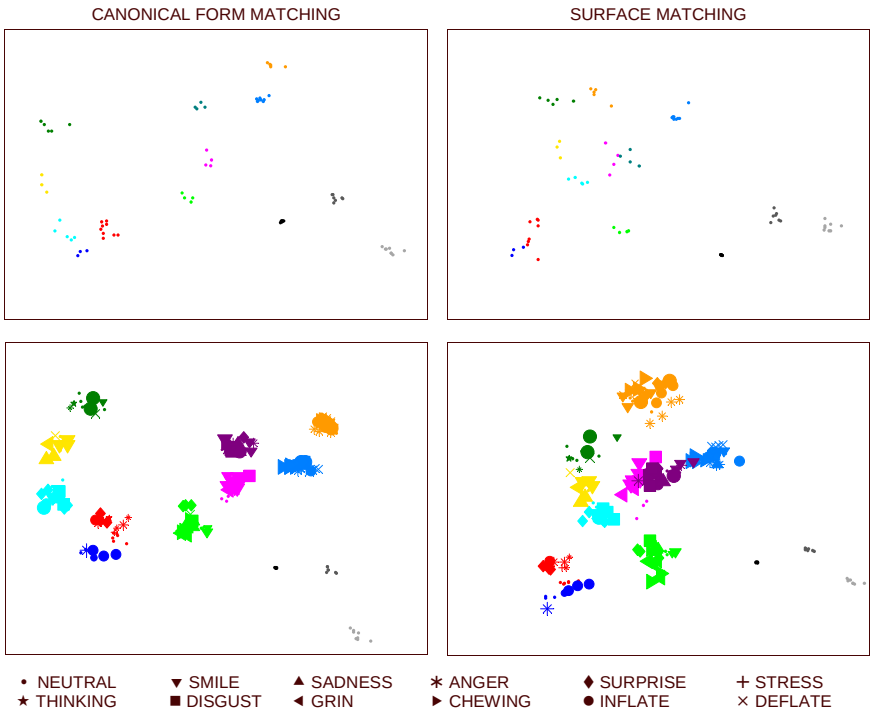

Figure 17. Low-dimensional visualization of dissimilarities between faces using canonical form (left) and original surface matching (right). First row: neutral expressions only. Second row: all expressions. Colors represent different subject. Symbols represent different facial expression. Symbol size represents the strength of the facial expression.

depicts a subject with extreme facial expression; columns two through four depict the matches among the 65 templates using canonical form matching, facial surface matching and eigenfaces. The results are typical for the descried algorithms. Eigenfaces, being image-based, finds the subject `Ori 188` more similar to the reference subject `Moran 129` since they have the same facial expression (strong smile), though these are different subjects. Facial surface matching is confused by 3D features (outstanding inflated cheeks) that appear on the face of subject `Moran 129` due to the facial expression. These features are similar to the natural facial features (fat cheeks) of subject `Susy 276` who has fat cheeks. Finally, canonical surface matching finds a correct match (`Moran 114`), since flattening compensates for the distortion of the face of subject `Moran 129` due to smile.

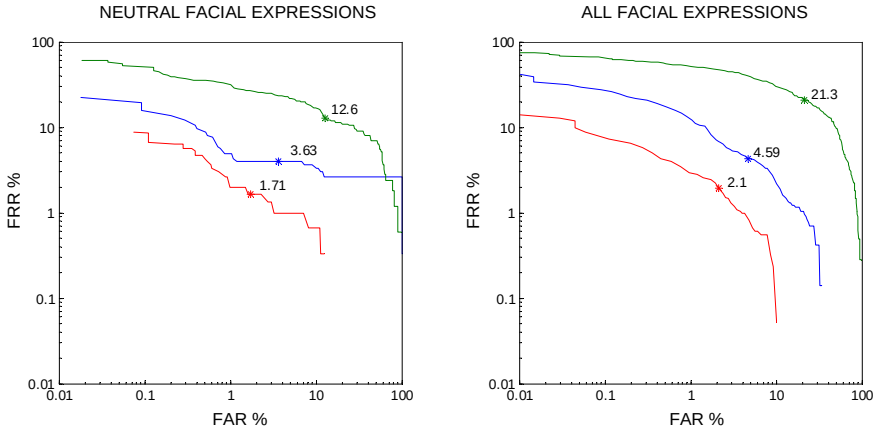

Figure 18. ROC curve of face recognition based on surface matching (blue), canonical surface matching (red) and eigenfaces (green). Left: test on database with neutral facial expressions, right: test on database with all expressions. Star denotes equal error rate.

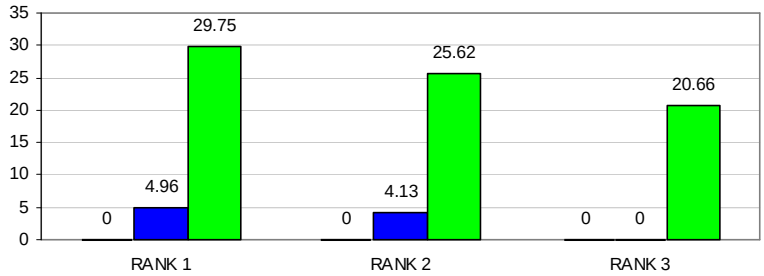

Figure 19. Resulting rank 1, 2 and 3 (left, middle and right, respectively) recognition error in full database with all facial expressions using canonical surface matching (black), facial surface matching (dark gray) and eigenfaces (light gray).

8.3. Experiment III - twins test

In the final experiment, we performed one of the most challenging tests for a face recognition algorithm - an attempt to tell identical twins apart. The full database containing all facial expressions (total of 220 faces; 17 faces of `Michael` and 8 of `Alex`) was used. Recognition was performed between two identical twins, with the other subjects acting as impostors. The matching was performed to templates with neutral expressions (6 for `Michael`, 3 for `Alex`).

The eigenfaces algorithm resulted in 29.41% incorrect matches when enrolling `Michael` and 25% when enrolling `Alex`. Facial surface matching resulted in 17.64% and 0% wrong matches, respectively. Canonical form matching resulted in 0% recognition error for both twins (see

example in Figure 20, second row). A visualization of the dissimilarities between the faces of twins and other subjects is shown in Figure 21.

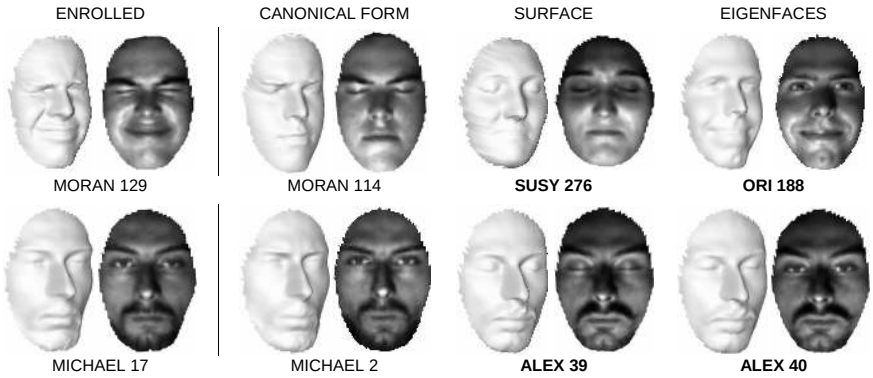

Figure 20. Example of recognition using different algorithms. First column shows the reference subject; second through fourth columns depict the closest (rank 1) matches found by the canonical form matching, facial surface matching and eigenfaces, respectively. Note that only the match using canonical form matching is correct. Numbers represent the subject's index in the database. First row exemplifies Experiment II. Second row exemplifies recognition of identical twins in Experiment III. Wrong matches are emphasized.

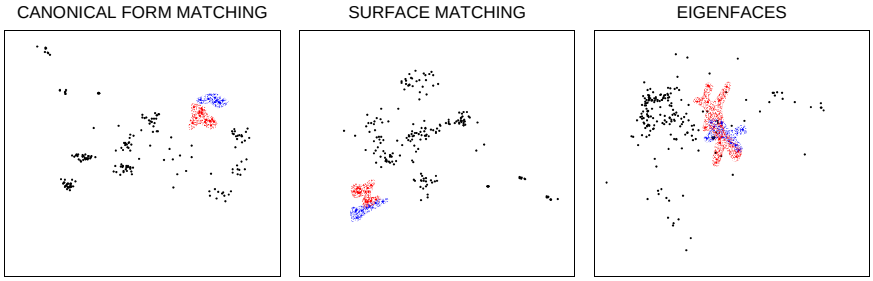

Figure 21. Twins test. A low-dimensional visualization of the dissimilarities between 220 faces used in the twins test. Each face is represented as a point. Red and blue denotes the faces of Michael and Alex, respectively.

9. Conclusions

We presented a geometric 3D face recognition framework which efficiently handles illumination changes due to pose as well as facial expressions. We experimentally proved the obvious fact that the Riemannian (intrinsic) geometry of the facial surface is less susceptible to changes due to facial expressions compared to the Euclidean (extrinsic) geometry. This property allowed us to construct a bending-invariant

representation which is insensitive to expressions. In other words, we introduced an expression invariant signature.

Experimental results showed that our canonical forms approach outperforms classical 2D and prior (rigid) 3D face recognition algorithms. The recognition accuracy of the proposed method is very high, which indicates that for the purpose of biometric identification, treating faces as isometric objects is a promising direction.

Acknowledgements

We are grateful to David Donoho for pointing us to Ekman's publications, to Alon Spira for providing his fast marching on parametric manifolds code, to Moran Feldman and Asi Elad for their fast implementation in C; to Michael Elad for helpful discussions; to Eyal Gordon for providing data for the facial expressions tests and to all the people who agreed to contribute their faces to our 3D face database.

Appendix

.1. Proof of Theorem 1

Without loss of generality, we assume that the perturbed point is $\mathbf{x}_1$, so that $d(\mathbf{x}_1, \tilde{\mathbf{x}}_1) < \epsilon$. Let us denote the perturbed geodesic distances by $\tilde{d}_{ij}$. By the triangle inequality,

$$\tilde{d}_{1j} \;\leq\; d(\mathbf{x}_1, \tilde{\mathbf{x}}_1) + d(\mathbf{x}_1, \mathbf{x}_j) \leq d_{ij} + \epsilon,$$

whereas, $\tilde{d}_{ij}$ for $i > 1$ remains unchanged.

The perturbed geodesic distances matrix $\tilde{\mathbf{D}}$ can be written as $\tilde{\mathbf{D}} = \mathbf{D} + \delta\mathbf{D}$, where

$$\delta\mathbf{D} \;=\; \begin{pmatrix} 0 & \epsilon_2 & \dots & \epsilon_n \\ \epsilon_2 & & & \\ \vdots & & & \\ \epsilon_N & & & \end{pmatrix},$$

and $\epsilon_i \leq \epsilon$. The perturbed matrix of squared geodesic distances $\tilde{\mathbf{\Delta}}$ is given by

$$\tilde{\mathbf{\Delta}} \;=\; \mathbf{\Delta} + \delta\mathbf{\Delta} = (\mathbf{D} + \delta\mathbf{D}) \circ (\mathbf{D} + \delta\mathbf{D}) = \mathbf{\Delta} + 2\mathbf{D} \circ \delta\mathbf{D} + \delta\mathbf{D} \circ \delta\mathbf{D}.$$

Neglecting the second-order term $\delta\mathbf{D} \circ \delta\mathbf{D}$, we obtain $\delta\mathbf{\Delta} = 2\mathbf{D} \circ \delta\mathbf{D}$. The spectral norm of the perturbation of $\mathbf{\Delta}$ is

$$\begin{aligned}\left\| \tilde{\mathbf{\Delta}} \right\|_2 &= \| 2\mathbf{D} \circ \delta\mathbf{D} \|_2 \leq 2 \max d_{ij} \ \| \delta\mathbf{D} \|_2 \\ &= 2 \max d_{ij} \ \max \sqrt{\lambda_i^{\delta\mathbf{D}^T \delta\mathbf{D}}} = 2 \max d_{ij} \ \sqrt{\sum_{i=2}^{N} \epsilon_i^2} < 2\sqrt{N}\epsilon \max d_{ij}.\end{aligned}$$

The perturbed double-centered matrix $\tilde{\mathbf{B}}$ is given by

$$\tilde{\mathbf{B}} = \mathbf{B} + \delta\mathbf{B} = -\frac{1}{2}\mathbf{J}(\mathbf{\Delta} + \delta\mathbf{\Delta})\mathbf{J} = \mathbf{B} - \frac{1}{2}\mathbf{J}\delta\mathbf{\Delta}\mathbf{J}.$$

Since $\| \mathbf{J} \|_2 = 1$, it follows that

$$\| \delta\mathbf{B} \|_2 \leq \frac{1}{2} \| \delta\mathbf{\Delta} \|_2 < \sqrt{N}\epsilon \max d_{ij}.$$

Eigendecomposition of the perturbed double-centered matrix yields $\tilde{\mathbf{B}} = \tilde{\mathbf{V}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{V}}^T$ with the eigenvalues $\tilde{\lambda}_i$ and the corresponding eigenvectors $\tilde{\mathbf{v}}_i$. The perturbed canonical form of the set of points $\{\mathbf{x}_i\}_{i=1}^N$ is therefore given by $\mathbf{x}_i' = \left(\sqrt{\tilde{\lambda}_1}\tilde{v}_1^i, \ \ldots \ , \sqrt{\tilde{\lambda}_m}\tilde{v}_m^i \right)^T$.

A known result from non-degenerate perturbation theory (Stewart and Sun, 1990) states that

$$\begin{aligned}\left| \lambda_i - \tilde{\lambda}_i \right| &\leq \| \delta\mathbf{B} \|_2 < \sqrt{N}\epsilon \max d_{ij}, \\ \frac{1}{2} \sin 2\theta \left(\mathbf{v}_i, \tilde{\mathbf{v}}_i \right) &\leq \frac{\| \delta\mathbf{B} \|_2}{\mathrm{gap}(\mathbf{B})} < \frac{\max d_{ij}}{\mathrm{gap}(\mathbf{B})}\sqrt{N}\epsilon,\end{aligned}$$

where $\theta\left(\mathbf{v}_i, \tilde{\mathbf{v}}_i \right)$ is the acute angle between the vectors $\mathbf{v}_i$ and $\tilde{\mathbf{v}}_i$, and $\mathrm{gap}(\mathbf{B}) = \min_{i \neq j} |\lambda_i - \lambda_j|$ is the spectral gap of the matrix $\mathbf{B}$. $\mathrm{gap}(\mathbf{B})$ is non-zero, since we assume that $\mathbf{B}$ has non-degenerate eigenvalues. Under a small perturbation, the order of the eigenvalues is preserved, i.e. $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \ldots \leq \tilde{\lambda}_N$; from Taylor expansion

$$\sqrt{\tilde{\lambda}_i} - \sqrt{\lambda_i} \approx \frac{(\tilde{\lambda}_i - \lambda_i)}{2\lambda_i},$$

and $\frac{1}{2} \sin 2\theta \left(\mathbf{v}_i, \tilde{\mathbf{v}}_i \right) \approx \theta \left(\mathbf{v}_i, \tilde{\mathbf{v}}_i \right)$. Since $\mathbf{v}_i$ and $\tilde{\mathbf{v}}_i$ have unit length, it follows that

$$\| \mathbf{v}_i - \tilde{\mathbf{v}}_i \|_2 \approx \sin\theta \left(\mathbf{v}_i, \tilde{\mathbf{v}}_i \right) \approx \theta \left(\mathbf{v}_i, \tilde{\mathbf{v}}_i \right) < \frac{\max d_{ij}}{\mathrm{gap}(\mathbf{B})}\sqrt{N}\epsilon$$

Using the triangle inequality and the above relations, the perturbation of the canonical form can be bounded by

$$\left\|\mathbf{x}'_i - \tilde{\mathbf{x}}'_i\right\|_2 = \left\|\sqrt{\lambda_i}\mathbf{v}_i - \sqrt{\tilde{\lambda}_i}\tilde{\mathbf{v}}_i\right\|_2 \leq \sqrt{\lambda_i}\left\|\mathbf{v}_i - \tilde{\mathbf{v}}_i\right\|_2 + \left|\sqrt{\lambda_i} - \sqrt{\tilde{\lambda}_i}\right|\left\|\tilde{\mathbf{v}}_i\right\|_2$$
$$\leq \sqrt{\lambda_i}\left\|\mathbf{v}_i - \tilde{\mathbf{v}}_i\right\|_2 + \frac{\left\|\left\|\delta\mathbf{B}\right\|\right\|_2}{2\sqrt{\lambda_i}} < \left(\frac{\sqrt{\lambda_i}}{\mathrm{gap}(\mathbf{B})} + \frac{1}{2\sqrt{\lambda_i}}\right)\sqrt{N}\epsilon\max d_{ij}.$$

.2. Proof of Theorem 2

For convenience of the proof, we replace the matrix arguments $\mathbf{X}, \mathbf{D}$ in the stress function $s(\mathbf{X}', \mathbf{D})$ with vector ones, $s(\mathbf{x}', \mathbf{d})$, where $\mathbf{x} = \mathrm{vec}(\mathbf{X})$ is a $3N \times 1$ column vector obtained by parsing the matrix $\mathbf{X}$ in column-stack order, and $\mathbf{d} = \mathrm{vec}(\mathbf{D})$ is a $N^2 \times 1$ column vector obtained by parsing the matrix $\mathbf{D}$. Similarly, we denote $\mathbf{x}' = \mathrm{vec}(\mathbf{X}')$.

Since $\mathbf{x}'$ is a local minimizer of the stress function $s(\mathbf{x}, \mathbf{d})$, from the first-order optimality conditions, it follows that

$$\nabla_{\mathbf{x}} s(\mathbf{x}', \mathbf{d}) = \mathbf{0}.$$

Now assume that the points are perturbed such that $\tilde{\mathbf{d}} = \mathbf{d} + \delta\mathbf{d}$. ¿From Taylor expansion (neglecting the higher-order terms) we have

$$\begin{aligned}\nabla_{\mathbf{x}} s(\mathbf{x}' + \delta\mathbf{x}', \mathbf{d} + \delta\mathbf{d}) = \\ \nabla_{\mathbf{x}} s(\mathbf{x}', \mathbf{d}) + \nabla^2_{\mathbf{x}} s(\mathbf{x}', \mathbf{d})\delta\mathbf{x}' + \nabla^2_{\mathbf{d}} s(\mathbf{x}', \mathbf{d})\delta\mathbf{d} = \\ \nabla^2_{\mathbf{x}} s(\mathbf{x}', \mathbf{d})\delta\mathbf{x}' + \nabla^2_{\mathbf{d}} s(\mathbf{x}', \mathbf{d})\delta\mathbf{d}.\end{aligned}$$

From the second-order optimality condition, it follows that the Hessian $\nabla^2_{\mathbf{x}} s(\mathbf{x}', \mathbf{d})$ is positive-definite, i.e. $\mathbf{x}^T \nabla^2_{\mathbf{x}} s(\mathbf{x}', \mathbf{d})\mathbf{x} > 0$ for every $\mathbf{x} \neq \mathbf{0}$.

By requiring that $\mathbf{x}' + \delta\mathbf{x}'$ is the local minimizer of $s(\mathbf{x}, \mathbf{d} + \delta\mathbf{d})$ w.r.t. the first argument, we have

$$\nabla^2_{\mathbf{x}} s(\mathbf{x}', \mathbf{d})\delta\mathbf{x}' + \nabla^2_{\mathbf{d}} s(\mathbf{x}', \mathbf{d})\delta\mathbf{d} = 0.$$

Consequently,

$$\left\|\delta\mathbf{x}'\right\|_2 \leq \left|\frac{\lambda_{\max}^{\nabla^2_{\mathbf{d}} s(\mathbf{x}', \mathbf{d})}}{\lambda_{\min}^{\nabla^2_{\mathbf{x}} s(\mathbf{x}', \mathbf{d})}}\right| \cdot \left\|\delta\mathbf{d}\right\|_2.$$

The bound is finite due to positive definiteness of $\nabla^2_{\mathbf{x}} s(\mathbf{x}', \mathbf{d})$ and since the eigenvalues of $\nabla^2_{\mathbf{d}} s(\mathbf{x}', \mathbf{d})$ are finite.

References

Arnoldi, W.: 1951, 'The Principle of Minimized Iterations in the Solution of the Matrix Eigenvalue Problem'. *Quart. Appl. Math.* **9**, 17–29.

Ashbourn, J.: 2002, *Biometrics: advanced identity verification*. Springer-Verlag, Berlin Heidelberg New York.

Beumier, C. and M. P. Acheroy: 1988, 'Automatic face authentication from 3D surface'. In: *Proc. British Machine Vision Conf.* pp. 449–458.

Bledsoe, W. W.: 1966, 'The model method in facial recognition'. Technical Report PRI 15, Panoramic Research Inc., Palo Alto (CA) USA.

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

Bronstein, A., M. Bronstein, E. Gordon, and R. Kimmel: 2003a, 'High-resolution structured light range scanner with automatic calibration'. Technical Report CIS-2003-06, Dept. of Computer Science, Technion, Israel.

Bronstein, A., M. Bronstein, E. Gordon, and R. Kimmel: 2004a, 'Fusion of 3D and 2D information in face recognition'. In: *Proc. ICIP*. to appear.

Bronstein, A., M. Bronstein, and R. Kimmel: 2003b, 'Expression-invariant 3D face recognition'. In: *Proc. Audio and Video-based Biometric Person Authentication*. pp. 62–69.

Bronstein, A., M. Bronstein, R. Kimmel, and A. Spira: 2004b, 'Face Recognition from Facial Surface Metric'. In: *Proc. ECCV*.

Brunelli, R. and T. Poggio: 1993, 'Face recognition: features vs. templates'. *IEEE Trans. PAMI* **15**(10), 1042–1053.

Carrihill, B. and R. Hummel: 1985, 'Experiments with the intensity ratio depth sensor'. *Computer Vision, Graphics and Image Processing* **32**, 337–358.

Chen, G. and G. Medioni: 2001, 'Building Human Face Models from Two Images'. *Journal of VLSI Signal Processing* **27**(1/2), 127–140.

Cox, I., J. Ghosn, and P. Yianilos: 1996, 'Feature-based face recognition using mixture distance'. In: *Proc. CVPR*. pp. 209–216.

De-Leeuw, J.: 1977, *Recent developments in statistics*, Chapt. Applications of convex analysis to multidimensional scaling, pp. 133–145. Amsterdam: North-Holland.

De-Leeuw, J. and I. Stoop: 1984, 'Upper bounds on Kruskal's stress'. *Psychometrika* **49**, 391–402.

Eckart, C. and G. Young: 1936, 'Approximation of one matrix by another of lower rank'. *Psychometrika* **1**, 211–218.

Ekman, P.: 1973, *Darwin and facial expression; a century of research in review*. New York: Academic Press.

Elad, A. and R. Kimmel: 2001, 'Bending invariant representations for surfaces'. In: *Proc. CVPR*. pp. 168–174.

Fleishman, S., I. Drori, and D. Cohen-Or: 2003, 'Bilateral Mesh Denoising'. In: *Proc. SIGGRAPH*. pp. 950–953.

Floater, M. S. and K. Hormann: 2004, *Advances on Multiresolution in Geometric Modelling*, Chapt. Surface Parameterization: a Tutorial and Survey. Springer-Verlag, Heidelberg. To appear.

Gauss, C. F.: 1827, 'Disquisitiones generales circa superficies curva'. *Commentationes Societatis RegiæScientiarum Gottingensis Recentiores* **6**, 99 146.

Georghiades, A. S., P. N. Belhumeur, and D. Kriegman: 1998, 'Illumination cones for recognition under variable lighting: faces'. In: *Proc. CVPR*.

Gheorghiades, A. S., P. N. Belhumeur, and D. J. Kriegman: 2001, 'From few to many: illumination cone models for face recognition under variable lighting and pose'. *IEEE Trans. PAMI* **23**(6).

Goldstein, A., L. Harmon, and A. Lesk: 1971, 'Identification of human faces'. *Proc. IEEE* **59**(5), 748–760.

Gordon, G.: 1997, 'Face recognition from frontal and profile views'. In: *Proc. Int'l Workshop on Face and Gesture Recognition*. pp. 74–52.

Gower, J. C.: 1966, 'Some distance properties of latent root and vector methods used in multivariate analysis'. *Biometrika* **53**, 325–338.

Grossman, R., N. Kiryati, and R. Kimmel: 2002, 'Computational surface flattening: a voxel-based approach'. *IEEE Trans. PAMI* **24**(4), 433–441.

Gruen, A. and D. Akca: 2004, 'Least Squares 3D Surface Matching'. In: *Proc. ISPRS Working Group V/1 Panoramic Photogrammetry Workshop*. pp. 19–22.

Hallinan, P.: 1994, 'A low-dimensional representation of human faces for arbitrary lighting conditions'. In: *Proc. CVPR*. pp. 995–999.

Horn, E. and N. Kiryati: 1999, 'Toward optimal structured light patterns'. *Image and Vision Computing* **17**(2), 87–97.

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

Hugli, H. and G. Maitre: 1989, 'Generation and use of color pseudo random sequences for coding structured light in active ranging'. In: *Proc. Industrial Inspection*, Vol. 1010. pp. 75–82.

Hung, Y.-P., C.-S. Chen, I.-B. Hsieh, and C.-S. Fuh: 1999, 'Reconstruction of complete 3D object model from multiview range images'. In: *Proc. SPIE Int'l Symposium on Electronic Imaging*. pp. 138–145.

Kanade, T.: 1973, 'Picture processing by computer complex and recognition of human faces'. Technical report, Kyoto University, Dept. of Information Science.

Kimmel, R.: 2003, *Numerical geometry of images*. Berlin Heidelberg New York: Springer-Verlag.

Kimmel, R. and J. A. Sethian: 1998, 'Computing geodesic on manifolds'. In: *Proc. US National Academy of Science*, Vol. 95. pp. 8431–8435.

Kimmel, R. and I. Yavneh: 2003, 'An algebraic multigrid approach for image analysis'. *SIAM Journal on Scientific Computing* **24**(4), 1218–1231.

Kreyszig, E.: 1991, *Differential geometry*. New York: Dover Publications Inc.

Li, S. and J. Lu: 1999, 'Face recognition using nearest feature line'. *IEEE Trans. Neural Networks* **10**(2), 439–443.

Mansfield, T., G. Kelly, D. Chandler, and J. Kane: 2001, 'Biometric product testing final report'. Technical report, Centre for Mathematics and Scientific Computing, National Physical Laboratory, UK.

Mavridis, N., 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*.

Mémoli, F. and G. Sapiro: 2004, 'Comparing point clouds'. IMA preprint series 1978, University of Minnesota, Minneapolis, MN 55455, USA.

Ortega-Garcia, J., J. Bigun, D. Reynolds, and J. Gonzalez-Rodriguez: 2004, 'Authentication gets personal with biometrics'. *IEEE Signal Processing magazine* **21**(2), 50–62.

Pentland, A., B. Moghaddam, and T. Starner: 1994, 'View-based and modular eigenspaces for face recognition'. In: *Proc. CVPR*. pp. 84–91.

Posdamer, J. L. and M. D. Altschuler: 1982, 'Surface measurement by spaceencoded projected beam systems'. *Computer Graphics and Image Processing* **18**(1), 1–17.

Proesmans, M., L. V. Gool, and A. Oosterlinck: 1996, 'One-shot active shape acquisition'. In: *Proc. Internat. Conf. Pattern Recognition*, Vol. C. pp. 336–340.

Samil, A. and P. Iyengar: 1992, 'Automatic recognition of human faces and facial expressions: a survey'. *Pattern recogntion* **25**, 65–77.

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

Sethian, J. A.: 1996, 'A review of the theory, algorithms, and applications of level set method for propagating surfaces'. *Acta numerica* pp. 309–395.

Sirovich, L. and M. Kirby: 1987, 'Low-dimensional procedure for the characterization of human faces'. *J. Optical Soc. Am. A* **2**, 519–524.

Sochen, N., R. Kimmel, and R. Malladi: 1998, 'A general framework for low level vision'. *IEEE Trans. Image Proc.* **7**(3), 310–318.

Spira, A. and R. Kimmel: 2003, 'An efficient solution to the eikonal equation on parametric manifolds'. In: *INTERPHASE 2003 meeting*.

Spira, A. and R. Kimmel: 2004, 'An Efficient Solution to the Eikonal Equation on Parametric Manifolds'. *Interfaces and Free Boundaries*. To appear.

Spira, A., N. Sochen, and R. Kimmel: 2004, *Handbook of Computational Geometry for Pattern Recognition, Computer Vision, Neurocomputing and Robotics*, Chapt. Geometric filters, diffusion flows, and kernels in image processing. Berlin Heidelberg New York: Springer-Verlag. to be published.

Stewart, G. W. and J.-G. Sun: 1990, *Matrix perturbation theory*. Academic Press.

Tajima, J. and M. Iwakawa: 1990, '3D data acquisition by rainbow range finder'. In: *Proc. Int'l Conf. Pattern Recognition*. pp. 309–313.

Tal, A., M. Elad, and S. Ar: 2001, 'Content based retrieval of VRML objects - an iterative and interactive approach'.

Tasdizen, T., R. Whitaker, P. Burchard, and S. Osher: 2002, 'Geometric surface smoothing via anisotropic diffusion of normals'. In: *Proc. IEEE Conf. Visualization*. pp. 125–132.

Torgerson, W. S.: 1952, 'Multidimensional scaling I - Theory and methods'. *Psychometrika* **17**, 401–419.

Turk, M. and A. Pentland: 1991, 'Face recognition using eigenfaces'. In: *Proc. CVPR*. pp. 586–591.

Vuylsteke, P. and A. Oosterlinck: 1990, 'Range image acquisition with a single binaryencoded light pattern'. *IEEE Trans. PAMI* **12**(2), 148–163.

Winkelbach, S. and F. M. Wahl: 2001, *Lecture Notes in Computer Science*, Vol. 2191, Chapt. Shape from 2D edge gradients, pp. 377–384. Springer-Verlag, Berlin Heidelberg New York.

Wiskott, L.: 1995, 'Labeled graphs and dynamic link matching for face recognition and scene analysis'. *Reihe Physik* **53**.

Wiskott, L., J. Fellous, N. Kruger, and C. von der Malsburg: 1997, 'Face recognition by elastic bunch graph matching'. *IEEE Trans. PAMI* **19**(7), 733–742.

Young, G. and A. S. Householder: 1938, 'Discussion of a set of point in terms of their mutual distances'. *Psychometrika* **3**, 19–22.

Zigelman, G., R. Kimmel, and N. Kiryati: 2002, 'Texture mapping using surface flattening via multi-dimensional scaling'. *IEEE Trans. Visualization and computer graphics* **9**(2), 198–207.