International Journal of Computer Vision 64(1), 5–30, 2005

Three-Dimensional Face Recognition*

ALEXANDER M. BRONSTEIN, MICHAEL M. BRONSTEIN AND RON KIMMEL
Department of Computer Science, Technion—Israel Institute of Technology, Haifa 32000, Israel

Received June 3, 2004; Revised December 15, 2004; Accepted January 10, 2005

First online version published in April 2005

Abstract. An expression-invariant 3D face recognition approach is presented. Our basic assumption is that facial expressions can be modelled as isometries of the facial surface. This allows to construct expression-invariant representations of faces using the bending-invariant canonical forms approach. 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 a prototype system based on the proposed algorithm and compare its performance 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 representation of any given face. It allows us to perform the 3D face recognition task while avoiding the surface reconstruction stage.

Keywords: expression-invariant 3D face recognition, isometry invariant, facial expressions, multidimensional scaling

1. Introduction

Automatic face recognition has been traditionally associated with the fields of computer vision and pattern recognition. Face recognition is considered a natural, non-intimidating, and 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 pattern recognition problems. 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 *recognition* (or *identification*). In the first case, the enrolled individual (*probe*) claims identity of a person whose template is stored in the database (*gallery*). We refer to the data used for a specific recognition task as a *template*. The 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, for example, in an automatic teller machine (ATM). In this case, the user is usually assumed to be collaborative.

The second case is more difficult. Recognition implies that the probe subject should be compared with all the templates stored in the gallery 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. At current technological level, one-to-many face recognition with non-collaborative users is practically unsolvable. That is, if one intentionally wishes not to be recognized, he can always deceive

*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.

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

any face recognition technology. In the following, we will assume collaborative users.

Even collaborative users in a natural environment present high variability of their faces—due to natural factors beyond our control. The greatest difficulty of face recognition, compared to other biometrics, stems from the immense variability of the human face. The facial appearance depends heavily on environmental factors, for example, the lighting conditions, background scene and head pose. It also depends on facial hair, the use of cosmetics, jewelry and piercing. Last but not least, plastic surgery or long-term processes like aging and weight gain can have a significant influence on facial appearance.

Yet, much of the facial appearance variability is inherent to the face itself. Even if we hypothetically assume that external factors do not exist, for example, that the facial image is always acquired under the same illumination, pose, and with the same haircut and make up, still, the variability in a facial image due to facial expressions may be even greater than a change in the person's identity (see Fig. 1).

1.1. Two-Dimensional Face Recognition: Invariant Versus Generative Approaches

Trying to make face recognition algorithms insensitive to illumination, head pose, and other factors mentioned above is one of the main efforts of current research in the field. Broadly speaking, there are two alternatives in approaching this problem. One is to find features that are not affected by the viewing conditions; we call this the *invariant* approach. Early face recognition algorithms advocated the invariant approach by finding a set of *fiducial points* such as eyes, nose, mouth, etc. and comparing their geometric relations (*feature-based* recognition) (Bledsoe, 1966; Kanade, 1973; Goldstein et al., 1971) or comparing the face to a whole facial template (*template-based* recognition) (Brunelli and Poggio, 1993).

It appears, however, that very few reliable fiducial points can be extracted from a 2D facial image in the presence of pose, illumination, and facial expression variability. As the result, feature-based algorithms are forced to use a limited set of points, which provide low discrimination ability between faces (Cox et al., 1996). Likewise, templates used in template matching approaches change due to variation of pose or facial expression (Brunelli and Poggio, 1993). Using elastic graph matching (Wiskott, 1995; Wiskott et al., 1997) as an attempt to account for the deformation of templates due to flexibility of the facial surface has yielded limited success since the attributed graph is merely a flat representation of a curved 3D object (Ortega-Garcia et al., 2004).

Appearance-based methods that treat facial images as vectors of a multidimensional Euclidean space and use standard dimensionality reduction techniques to construct a representation of the face (*eigenfaces* (Turk and Pentland, 1991) and similar approaches (Sirovich and Kirby, 1987; Hallinan, 1994; Pentland et al., 1994)), require accurate *registration* between facial images. The registration problem brings us back to identifying reliably fiducial points on the facial image independently of the viewing conditions and the internal variability due to facial expressions. As a consequence, appearance-based methods perform well only when the probe image is acquired in conditions similar to those of the gallery image (Gheorghiades et al., 2001).

The second alternative is to *generate* synthetic images of the face in new, unseen conditions. Generating facial images with new pose and illumination requires some 3D facial surface as an intermediate stage. It is possible to use a generic 3D head model (Huang et al., 2002), or estimate a rough shape of the facial surface from a set of observations (e.g. using photometric stereo (Georghiades et al., 1998, 2001)) in order to synthesize new facial images and then apply standard face recognition methods like eigenfaces (Sirovich and Kirby, 1987; Turk and Pentland, 1991) to the synthetic images. Yet, facial

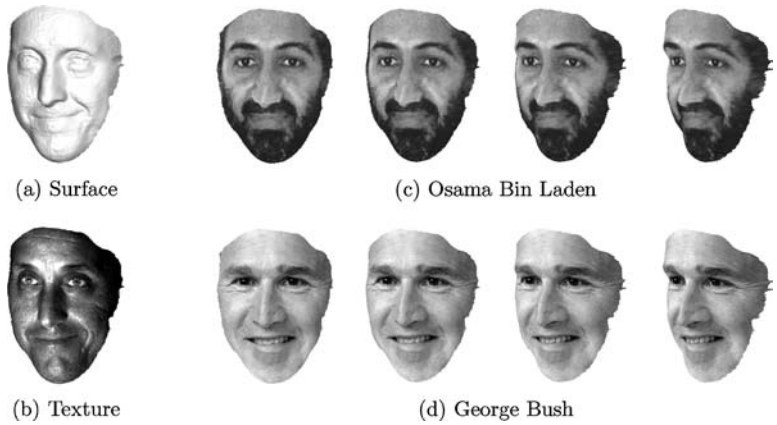


Figure 2. Simple texture mapping on the same facial surface can completely change the appearance of the 2D facial image and make the same face look like George Bush or Osama Bin Laden.

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 computer graphics applications. That is, the results may look natural, but fail to represent the true nature of the expression.

Figure 2 shows a simple visual experiment that demonstrates the generative approach. We created synthetic faces of Osama Bin Laden (Fig. 2(c)) and George Bush (Fig. 2(d)) in different poses by mapping the respective textures onto the facial surface of another subject (Fig. 2(a) and (b)). The resulting images are easily recognized as the world number one terrorist and the forty third president of the United States, though in both cases, the facial geometry belongs to a completely different individual. This is explained by the property of the human visual system, which uses mainly the 2D information of the face to perform recognition.

Simple texture mapping in our experiment allowed to create naturally-looking faces, yet, the individuality of the subject concealed in the 3D geometry of his face was completely lost. This reveals the intrinsic weakness of all the 2D face recognition approaches: the face is a 3D object, and using only its 2D projection can be misleading. Exaggerating this example, if one had the ability to draw any face on his facial surface, he could make himself look essentially like any person and deceive any 2D face recognition method. Practically, even with very modest instruments, makeup specialists in the theater and movie industry can change completely the facial appearance of actors.

1.2. Three-Dimensional Face Recognition

Three-dimensional face recognition is a relatively recent trend that in some sense breaks the long-term tradition of mimicking the human visual recognition system, like the 2D methods attempt to do. As evaluations such as the Face Recognition Vendor Test (FRVT) demonstrate in an unarguable manner that current state of the art in 2D face recognition is insufficient for high-demanding biometric applications (Phillips et al., 2003), trying to use 3D information has become an emerging research direction in hope to make face recognition more accurate and robust.

Three-dimensional facial geometry represents the internal anatomical structure of the face rather than its external appearance influenced by environmental factors. As the result, unlike the 2D facial image, 3D facial surface is insensitive to illumination, head pose (Bowyer et al., 2004), and cosmetics (Mavridis et al., 2001). Moreover, 3D data can be used to produce invariant measures out of the 2D data (for example, given the facial surface, the *albedo* can be estimated from the 2D reflectance under assumptions of Lambertian reflection).

However, while in 2D face recognition a conventional camera is used, 3D face recognition requires a more sophisticated sensor, capable of acquiring depth information—usually referred to as *depth* or *range camera* or *3D scanner*. The 3D shape of the face is usually acquired together with a 2D intensity image. This is one of the main disadvantages of 3D methods compared to 2D ones. Particularly, it prohibits the use

of legacy photo databases, like those maintained by police and special agencies.

Early papers on 3D face recognition revealed the potential hidden in the 3D information rather than presented working algorithms or extensive tests. In one of the first papers on 3D face recognition, Cartoux et al. (1989) approached the problem by finding the plane of bilateral symmetry through the facial range image, and either matching the extracted profile of the face, or using the symmetry plane to compensate for the pose and then matching the whole surface. Similar approaches based on profiles extracted from 3D face data were also described in the follow-up papers by Nagamine et al. (1992), Beumier and Acheroy (1998) and Gordon (1997).

Achermann et al. (1997), Hesher et al. (2003), Mavridis et al. (2001), Chang et al. (2003), and Tsalakanidou et al. (2003) explored the extension of conventional dimensionality reduction techniques, like Principal Component Analysis (PCA), to range images or combination of intensity and range images. Gordon (1992) proposed representing a facial surface by a feature vector created from local information such as curvature and metric. The author noted that the feature vector is similar for different instances of the same face acquired in different conditions, except "variation due to expression" (Gordon, 1992).

Lee and Milios (1990) and Tanaka et al. (1998) proposed performing curvature-based segmentation of the range image into convex regions and compute the Extended Gaussian Image (EGI) locally for each region. A different local approach based on Gabor filters in 2D and point signatures in 3D was presented by Wang et al. (2002).

Finally, many theoretical works including Medioni and Waupotitsch (2003), Achermann and Bunke (2000), as well as some commercial systems, use rigid surface matching in order to perform 3D face recognition. However, facial expressions do change significantly the 3D facial surface not less than they change the 2D intensity image, hence modelling faces as rigid objects is invalid when considering facial expressions.

We note that the topic of facial expressions in 3D face recognition is very scarcely addressed in the literature, which makes difficult to draw any conclusions about the robustness of available algorithms. Many of the cited authors mention the problem of facial expressions, yet, none of them has addressed it explicitly, nor any of the algorithms except in Wang et al. (2002) was tested on a database with sufficiently large (if any) variability of facial expressions.

1.3. The 3DFACE Approach

Our approach, hereinafter referred to as 3DFACE, comes to address explicitly the problem of facial expressions. It treats the facial surface as a deformable object in the context of Riemannian geometry. Our observations show that the deformations of the face resulting from facial expressions can be modelled as isometries (Bronstein et al., 2003b), such that the intrinsic geometric properties of the facial surface are expression-invariant. Thus, finding an expression-invariant representation of the face is essentially equivalent to finding an isometry-invariant representation of the facial surface.

A computationally-efficient invariant representation of isometric surfaces can be constructed by isometrically embedding the surface into a low-dimensional space with convenient geometry. Planar embedding appeared 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, as first presented by Elad and Kimmel (2001), was shown to be an efficient way to perform matching of deformable objects.

Isometric embedding 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 perform the 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 the 3DFACE system is that we actually do not need the facial surface to be given explicitly. All the stages of our recognition system, including preprocessing and computation of geodesic distances can be carried out given only the metric tensor of the surface. It allows us to use simple and cost-efficient 3D acquisition techniques like photometric stereo. Avoiding explicit surface reconstruction also saves computational time and reduces the numerical inaccuracies (Bronstein et al., 2004b).

The organization of this paper is the following: Next section starts with modelling faces as Riemannian

surfaces. Such a formulation serves as a unifying framework for different procedures described later on. Facial expressions are modelled as isometries of the facial surface. A simple experiment justifies this assumption. Section 3 introduces the concept of multidimensional scaling and its application to isometric embedding. Section 4 describes a prototype 3DFACE system and addresses some implementation considerations. In Section 5 we show experimental results assessing the performance of our method and comparing it to other 2D and 3D face recognition algorithms. Section 6 concludes the paper.

2. The Geometry of Human Faces

2.1. A Brief Introduction into Riemannian Geometry

We model a human face as a two-dimensional smooth connected parametric manifold (surface), denoted by $\mathcal{S}$ and represented by a coordinate chart from a compact subset $\Omega \subset \mathbb{R}^2$ to $\mathbb{R}^3$

$$\mathbf{x}(\Omega) = (x^1(\xi^1, \xi^2), x^2(\xi^1, \xi^2), x^3(\xi^1, \xi^2)). \quad (1)$$

We assume that the functions x^1, x^2 and x^3 are $\mathcal{C}^r$ (with r sufficiently large), and that the vectors $\partial_i \mathbf{x} \equiv \frac{\partial}{\partial \xi^i}\mathbf{x}$ $(i = 1, 2)$ are linearly independent. We will further assume that the surface can be represented as a graph of a function, e.g. $x^3 = x^3(x^1, x^2)$, such that x^3 can be referred to as the *depth* coordinate. Also, for convenience, in the following the parameterization coordinates $\boldsymbol{\xi} = (\xi^1, \xi^2)$ will be identified with the coordinates in the image acquired by the camera (see Section 4.1). Similarly, we define the facial *albedo* (or *reflection coefficient*) as a scalar field $\rho : \Omega \to \mathbb{R}_+$.

The derivatives $\partial_i \mathbf{x}$ constitute a local non-orthogonal coordinate system on $\mathcal{S}$, and span an affine subspace of $\mathbb{R}^3$ called the *tangent space* and denoted by $T_{\mathbf{x}}\mathcal{S}$ for every $\mathbf{x} \in \mathcal{S}$. 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 the tangent space. The Riemannian metric is an intrinsic characteristic of the manifold and allows us to measure local distances on $\mathcal{S}$ independently of the coordinates (Kreyszig, 1991). The pair $(\mathcal{S}, g)$ 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

$$ds^2 = g_{ij} d\xi^i d\xi^j, \quad i, j = 1, 2; \quad (2)$$

where repeating super- and subscripts are summed over according to Einstein's summation convention. The metric tensor g_{ij} at every point of a parametric surface can be given explicitly by

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

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

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

Another characteristic of the surface called the *second fundamental form*, is given in coordinate notation as

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

Unlike the metric, the second fundamental form is prescribed by the normal field $\mathbf{N}$, that is, it has to do with the way the manifold is immersed into the ambient space. This distinction plays a crucial role. The metric is responsible for all properties of the manifold which are called *intrinsic*. Propertied expressible in terms of the second fundamental form are called *extrinsic*.

The maximum and the minimum eigenvalues $\lambda_{\max}, \lambda_{\min}$ of the tensor $b_i^j = b_{ik} g^{kj}$ are called the *principal curvatures*. The corresponding eigenvectors of (b_i^j) are the principal curvature directions. The values

$$H = \frac{1}{2}\mathrm{trace}\left(b_i^j\right) = \frac{1}{2}(\lambda_{\max} + \lambda_{\min}); \quad (6)$$

$$K = \det\left(b_i^j\right) = \lambda_{\max}\lambda_{\min}, \quad (7)$$

are called the *mean curvature* and the *Gaussian curvature*, respectively. A classical result, known as *Theorema Egregium* (Gauss, 1827) claims that K is an intrinsic property of the manifold.

Since our surface is connected and compact, the Riemannian metric induces a distance metric. In order to define it, let $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ be two points on the surface and let $\mathbf{c} : [0, 1] \to \mathcal{S}$ be a smooth curve connecting $\mathbf{x}$ and $\mathbf{y}$ on $\mathcal{S}$. The length of $\mathbf{c}$ is defined

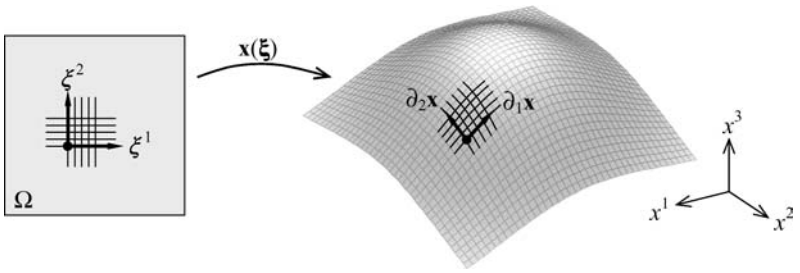


Figure 3. Definition of a parametric manifold and the local coordinate system.

by

$$\ell[\mathbf{c}] = \int_0^1 \left\| \frac{d\mathbf{c}(t)}{dt} \right\| dt. \tag{8}$$

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

$$d(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{c}} \ell[\mathbf{c}]. \tag{9}$$

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*.

2.2. *Facial Expressions as Isometric Transformations*

Let $(\mathcal{S}, g)$ and $(\mathcal{Q}, h)$ be Riemannian manifolds, and let $f : (\mathcal{S}, g) \to (\mathcal{Q}, h)$ be a diffeomorphism. f is called an *isometry* if

$$f^* h = g, \tag{10}$$

where * denotes *pullback* (Gudmundsson, 2004). Riemannian manifolds $(\mathcal{S}, g)$, $(\mathcal{Q}, h)$ related by an isometry are said to be *isometric*.

From the point of view of intrinsic geometry, isometric manifolds are indistinguishable. Particularly, in the same way as g induces a distance metric $d_{\mathcal{S}}$ on $\mathcal{S}$, the tensor h induces a distance metric $d_{\mathcal{Q}}$ on $\mathcal{Q}$. This essentially implies that f preserves the geodesic distance between every pair of points, that is,

$$d_{\mathcal{S}}(\mathbf{x}, \mathbf{y}) = d_{\mathcal{Q}}(f(\mathbf{x}), f(\mathbf{y})), \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{S}. \tag{11}$$

In fact, in our case Eq. (11) is an alternative definition of an isometry. Furthermore, from Theorema Egregium it stems that isometric surfaces have equal Gaussian curvature $K_{\mathcal{S}}(\mathbf{x}) = K_{\mathcal{Q}}(f(\mathbf{x}))$.

We apply this isometric 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 isometries. In other words, facial expressions give rise to nearly isometric surfaces. This will allow us to construct an expression-invariant representation of the face, based on its intrinsic geometric structure.

It must be understood, of course, that the isometric model is only an approximation and models natural expressions excepting pathological cases. One of such pathologies is the open mouth. The isometric model tacitly assumes that the *topology* of the facial surface is preserved, hence, facial expressions are not allowed to introduce arbitrary "holes" in the facial surface. This assumption is valid for most regions of the face except the *mouth*. Opening the mouth changes the topology of the facial surface by virtually creating a "hole". In Bronstein et al. (2005) we extend the isometric model to topology-changing expressions. Here, we assume that the mouth is closed in all facial expressions.

2.3. *Discrete Manifolds*

In practice, we work with discrete representation of surfaces, obtained by taking a finite number of samples on $\mathcal{S}$ and measuring the geodesic distances between them. After sampling, the Riemannian surface becomes a finite metric space $(X, \mathbf{D})$ with mutual distances described by the matrix $\mathbf{D}$ with elements $d_{ij} = d(x_i, x_j)$.

In Sethian (1996) an efficient numerical procedure called the *Fast Marching Method* (FMM), capable of measuring the distances from one source to N points on a plane in $\mathcal{O}(N)$ operations, was introduced (see (Tsitsiklis, 1995) for a related result). The FMM is based on upwind finite difference approximations for solving the *eikonal equation*, which is a differential

formulation of a wave propagation equation

$$\|\nabla v(\mathbf{x})\| = 1. \tag{12}$$

Here $v : \mathbb{R}^2 \to \mathbb{R}_+$ is the distance map from the sources $\mathbf{s}_1, \ldots, \mathbf{s}_K$, such that $v(\mathbf{s}_i) = 0$ serve as boundary conditions. The distance map is constructed by starting from the source point and propagating outwards. The FMM was extended to triangulated manifolds (FMTD) 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 FMTD 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 surfaces, presented by Spira and Kimmel (2003), was adopted here. The main advantage of this method is that the computations are performed on a uniform Cartesian grid in the parametrization plane, and not on the manifold like in the original version of Kimmel and Sethian (1998). 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 PFMM is that the surface is not needed explicitly in order to measure distances; and the metric given on the parametrization plane is sufficient (Bronstein et al., 2004b).

2.4. *Isometric Model Validation*

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. In Bronstein et al. (2005) we presented an experimental validation of the isometric model. We placed 133 markers on a face and tracked how the distances between these points change due to facial expressions. Figure 4 shows the distribution of the absolute change of the geodesic and the Euclidean distances. For more details, see Bronstein et al. (2005).

We conclude from this experiment that the change of the geodesic distances due to facial expressions is insignificant (which justifies our model), and that it is more than twice smaller than the respective change of the Euclidean distances. This implies that the isometric model describes better the nature of facial expressions compared to the rigid one. This observation will be reinforced in Section 5, where we compare our approach to a method that treats facial surfaces as rigid objects.

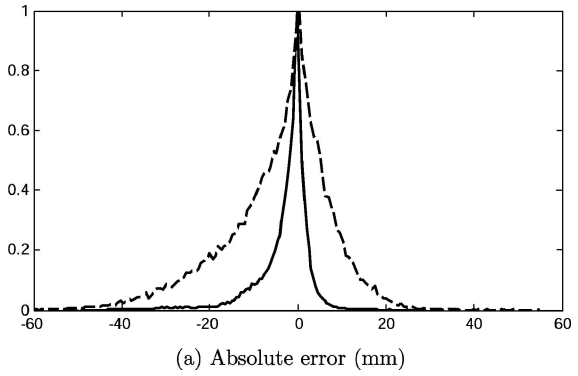


Figure 4. Normalized histogram of the absolute error of the geodesic (solid) and the Euclidean (dashed) distances.

3. Expression-Invariant Representation

A cornerstone problem in three-dimensional face recognition is the ability to identify facial expressions of a subject and distinguish them from facial expressions of other subjects. Under our isometric model assumption, this problem is reduced to finding similarity between isometric surfaces.

Let $(\mathcal{S}, g)$ and $(\mathcal{Q}, h)$ be isometric Riemannian surfaces, representing different expressions of the same face and related by an isometry f. Since f is an isometry, it preserves the geodesic distances, i.e. $d_{\mathcal{S}}(\mathbf{x}_1, \mathbf{x}_2) = d_{\mathcal{Q}}(\mathbf{y}_1, \mathbf{y}_2)$ for every $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{S}$ and $\mathbf{y}_1 = f(\mathbf{x}_1)$, $\mathbf{y}_2 = f(\mathbf{x}_2) \in \mathcal{Q}$. Therefore, the geodesic distances are suitable candidates for an expression-invariant representation of the face.

However, the surfaces $\mathcal{S}$ and $\mathcal{Q}$ are sampled and in practice we have *finite metric spaces* $(\{\mathbf{x}_1, \ldots, \mathbf{x}_{N_S}\}, \mathbf{D}_{\mathcal{S}})$ and $(\{\mathbf{y}_1, \ldots, \mathbf{y}_{N_Q}\}, \mathbf{D}_{\mathcal{Q}})$, respectively. There is neither guarantee that $\mathcal{S}$ and $\mathcal{Q}$ are sampled at the same points, nor even that the number of samples of the two surfaces is the same. Moreover, even if the samples are the same, they can be ordered arbitrarily, up to a permutation of rows and columns. This ambiguity makes impractical the use of $\mathbf{D}$ as an invariant representation, though we refer to the most recent attempt to directly compare the distance matrices (Mémoli and Sapiro, 2004).

3.1. Canonical Forms

An alternative proposed by Elad and Kimmel (2001) is to avoid dealing explicitly with the matrix of geodesic distances and represent the Riemannian surface as a subset of some convenient m-dimensional manifold $\mathcal{M}^m$, such that the original intrinsic geometry is preserved. We call such a procedure *isometric embedding*, and such $\mathcal{M}^m$ the *embedding space*.

The embedding allows to get rid of the extrinsic geometry, which no more exists in the new space. As a consequence, the resulting representation will be identical for all isometries of the surface. Another advantage is related to the fact that a general Riemannian metric is usually inconvenient to work with. The embedding space, on the other hand, can be chosen completely to our discretion. Simply saying, the embedding replaces a complicated geometric structure by a convenient one.

In our discrete setting, isometric embedding is a mapping

$$\varphi: (\{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathcal{S}, \mathbf{D}) \rightarrow (\{\mathbf{x}'_1, \ldots, \mathbf{x}'_N\} \subset \mathcal{M}^m, \mathbf{D}'), \qquad (13)$$

between two finite metric spaces, such that $d'_{ij} = d_{ij}$ for all $i, j = 1, \ldots, N$. The symmetric matrices $\mathbf{D} = (d_{ij}) \equiv (d(\mathbf{x}_i, \mathbf{x}_j))$ and $\mathbf{D}' = (d'_{ij}) \equiv (d'(\mathbf{x}'_i, \mathbf{x}'_j))$ denote the mutual geodesic distances between the points in the original and the embedding space, respectively. Following Elad and Kimmel (2001, 2003), the image of $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ under φ is called the *canonical form* of $(\mathcal{S}, g)$.

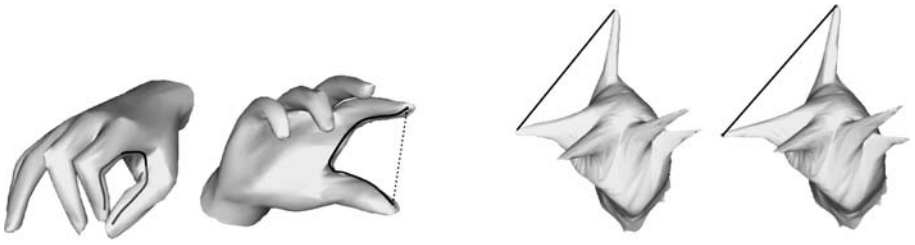
Figure 5 shows isometries of a deformable surface (human hand) and the corresponding canonical forms. In this example, once again, we should emphasize that the original surface (Fig. 5(a)) is a Riemannian manifold immersed into $\mathbb{R}^3$, on which the geodesic distances are induced by the Riemannian metric, while the canonical form (Fig. 5(b)) is a subset of $\mathbb{R}^3$, on which the geodesic distances are replaced by Euclidean ones. Thus, there is *no distinction between intrinsic and extrinsic geometry* in the embedding space.

We should keep in mind that a canonical form is defined up to an isometry group in $\mathcal{M}^m$. In an Euclidean space, for example, the ambiguity is up to a rotation, translation and reflection transformation. Fortunately, these are the only degrees of freedom that canonical forms possess (compared to the vast number of degrees of freedom of the matrix $\mathbf{D}$ itself), and thus can be easily dealt with.

3.2. Embedding Error

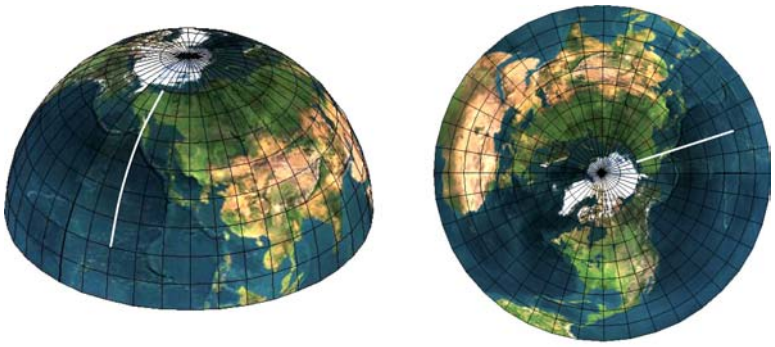
A question of cardinal importance is whether an isometric embedding of a given surface into a given space exists. A particular setting of this question first arose in the context of cartography: is it possible to map the spherical surface of the Earth onto a plane without distorting the distances between geographical objects? Theorema Egregium gives a negative answer to this question: the plane has zero Gaussian curvature, while the sphere has a positive one, and therefore, these surfaces are not isometric (Fig. 6). Moreover, it can be shown that such a simple surface like the sphere is not isometrically embeddable into Euclidean space of any finite dimension. For general, more complicated surfaces, the existence of an isometric embedding is unfortunately a very rare case (Linial et al., 1995).

Hence, we must keep in mind that the embeddings we discuss are *near-isometric*, and consequently, canonical forms only approximate the intrinsic geometry of the original surface. It is therefore necessary to define the measure of how the geodesic distances are distorted,



(a) Original surfaces (b) Canonical forms

Figure 5. Illustration of the embedding problem and the canonical forms. (a) Riemannian surface (hand) undergoing isometric transformations. Solid line shows the geodesic distance between two points on the surface, dotted line is the corresponding Euclidean distance in the space where the surface "lives". (b) After embedding, the hand surfaces become submanifolds of a three-dimensional Euclidean space, and the geodesic distances become Euclidean ones.



(a) Upper hemisphere of the globe (b) Planar map

Figure 6. The embedding problem arising in cartography: The upper hemisphere of the Earth (a) and its planar map (b). The geodesic distances (an example is shown as a white curve) are distorted by such a mapping.

that is, the discrepancy between $\mathbf{D}$ and $\mathbf{D}'$, called the *embedding error*. Here we limit our attention mainly to the embedding error criterion

$$\epsilon(\mathbf{X}'; \mathbf{D}, \mathbf{W}) \equiv \sum_{i<j} w_{ij}(d'_{ij}(\mathbf{X}') - d_{ij})^2, \quad (14)$$

referred to as the *stress*. Here $\mathbf{X}' = (\mathbf{x}'_1, \ldots, \mathbf{x}'_N)$ is a $N \times m$ matrix representing the parametric coordinates of the canonical form points in $\mathcal{M}^m$, and the symmetric $N \times N$ matrix of weights $\mathbf{W} = (w_{ij})$ determines the relative contribution of distances to the error criterion. If some of the distances are missing, the respective weights w_{ij} can be set to zero.

A generic name for algorithms that compute the canonical form by minimization of the stress with respect to $\mathbf{X}'$ is *multidimensional scaling* (MDS). These algorithms differ in the choice of the the embedding error criterion and the optimization method used for its minimization (Borg and Groenen, 1997).

3.3. The Choice of the Embedding Space

Another important question is how to choose the embedding space. First, the geometry of the embedding space is important. Popular choices include spaces with flat (Schwartz et al., 1989; Elad and Kimmel, 2001; Grossman et al., 2002; Zigelman et al., 2002), spherical (Elad and Kimmel, 2002; Bronstein et al., 2005b) or hyperbolic (Walter and Ritter, 2002) geometry. This choice should be dictated by the convenience of a specific space (namely, we want the geodesic distances in $\mathcal{S}'^m$ to have a simple analytic form) and the resulting embedding error, which, in turn, depends on the embedding error criterion.

Secondly, the dimension m of the embedding space must be chosen in such a way that the codimension of $\varphi(\mathcal{S})$ in $\mathcal{M}^m$ is at least 1 (as opposed to dimensionality reduction applications where usually $m \ll \dim(\mathcal{S})$). The reason becomes clear if we think of our manifolds as of function graphs, in our case—functions of two variables $z(x, y)$. Sampling $z(x, y)$ produces a set of points, which when embedded into a two-dimensional manifold, reflect the sampling pattern rather than the intrinsic geometry of the surface. On the other hand, when the surface is embedded into $\mathbb{R}^3$ (or a manifold of higher dimension), the samples will lie on some two-dimensional submanifold of $\mathbb{R}^3$ (the continuous limit of the canonical form), and increasing the sampling density would result in a better approximation of the continuous canonical form.

Embedding with codimension zero (e.g. embedding of a surface into a plane or a two-dimensional sphere $\mathbb{S}^2$) is useful when the manifold is endowed with some additional property, for example, texture. Such embedding can be thought of as an intrinsic parametrization of the manifold and has been explored in the context of medical visualization (Schwartz et al., 1989), texture mapping (Zigelman et al., 2002) and registration of facial images (Bronstein et al., 2004a, 2005a).

3.4. Least-Squares MDS

Minimization of $\epsilon(\mathbf{X}')$ can be performed by first-order, gradient descent-type methods, in which the direction at each step is

$$\mathbf{X}'^{(k+1)} = -\nabla\epsilon\left(\mathbf{X}'^{(k)}\right). \quad (15)$$

The gradient of $\epsilon(\Xi')$ with respect to Ξ' is given by

$$\frac{\partial}{\partial x'^{l}_{k}}\epsilon(\mathbf{X}';\mathbf{D},\mathbf{W}) = 2\sum_{j\neq k} w_{kj}\frac{(d'_{kj}-d_{kj})}{d_{kj}}\left(x'^{l}_{k}-x'^{l}_{j}\right), \tag{16}$$

and can be written as

$$\nabla\epsilon(\mathbf{X}';\mathbf{D},\mathbf{W}) = 2\mathbf{U}\mathbf{X}' - 2\mathbf{B}(\mathbf{X}';\mathbf{D},\mathbf{W})\mathbf{X}', \tag{17}$$

where $\mathbf{U}$ is a matrix with elements

$$u_{ij} = \begin{cases} -w_{ij} & \text{if } i\neq j \\ \sum_{j=1}^{N} w_{ij} & \text{if } i = j \end{cases}, \tag{18}$$

and $\mathbf{B}$ is a matrix with elements

$$b_{ij} = \begin{cases} -w_{ij}d_{ij}d'^{-1}_{ij}(\mathbf{X}') & \text{if } i\neq j \text{ and } d'_{ij}(\mathbf{X}')\neq 0 \\ 0 & \text{if } i\neq j \text{ and } d'_{ij}(\mathbf{X}')=0 \\ -\sum_{j\neq i} b_{ij} & \text{if } i = j \end{cases}. \tag{19}$$

It was observed by Guttman (1968) that the first-order optimality condition, $\nabla\epsilon(\mathbf{X}')=\mathbf{0}$, can be written as $\mathbf{X}'=\mathbf{U}^{\dagger}\mathbf{B}(\mathbf{X}')\mathbf{X}'$, and that the sequence

$$\mathbf{X}'^{(k+1)} = \mathbf{U}^{\dagger}\mathbf{B}\left(\mathbf{X}'^{(k)}\right)\mathbf{X}'^{(k)}, \tag{20}$$

converges to the local minimum of $\epsilon(\mathbf{X}')$ (here $\dagger$ denotes matrix pseudoinverse). The algorithm using this multiplicative update is called SMACOF (De Leeuw, 1977; De Leeuw and Stoop, 1984; Borg and Groenen, 1997). It can be shown to be equivalent to weighted gradient descent with constant step size $\mathbf{X}'^{(k+1)} = -\frac{1}{2}\mathbf{U}^{\dagger}\nabla\epsilon(\mathbf{X}'^{(k)})$, and if a non-weighted stress is used, it is essentially a gradient descent with constant step size $\mathbf{X}'^{(k+1)} = -\frac{1}{2N}\nabla\epsilon(\mathbf{X}'^{(k)})$ (Bronstein, 2004).

SMACOF is widely used for large-scale MDS problems. Its disadvantage is slow convergence in the proximity of the minimum, which is inherent to all first-order methods. Second order (Newton-type) methods (Kearsley et al., 1998) are usually disadvantageous in large-scale MDS problems. Recently, we showed that the multigrid framework can significantly improve the convergence time of the SMACOF algorithm (Bronstein et al., 2005c).

3.5. Classical Scaling

As an alternative to raw stress minimization in Euclidean embedding, it is worthwhile to mention an algebraic algorithm known as *classical scaling* (Torgerson, 1952; Gower, 1966), based on theoretical results of Eckart and Young (1936) and Young and Householder (1938). Classical scaling works with squared geodesic distances, which can be expressed as Hadamard (coordinate-wise) product $\boldsymbol{\Delta} = \mathbf{D}\circ\mathbf{D}$, where $\boldsymbol{\Delta} = (d^2_{ij})$. The matrix $\boldsymbol{\Delta}$ first undergoes double-centering

$$\mathbf{B}_{\boldsymbol{\Delta}} = -\frac{1}{2}\mathbf{J}\boldsymbol{\Delta}\mathbf{J} \tag{21}$$

(here $\mathbf{J} = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}$ and $\mathbf{I}$ is an $N\times N$ identity matrix). Then, the eigendecomposition $\mathbf{B}_{\boldsymbol{\Delta}} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{T}$ is computed, where $\mathbf{V} = (\mathbf{v}_1,\ldots,\mathbf{v}_N)$ is the matrix of eigenvectors of $\mathbf{B}$ corresponding to the eigenvalues $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_N$.

Let us denote by $\boldsymbol{\Lambda}^m_+$ the $m\times m$ matrix of first m positive eigenvalues, by $\mathbf{V}^m_+$ the $N\times m$ matrix of the corresponding eigenvectors, and by $\boldsymbol{\Lambda}^m_-$ the matrix of the remaining $N-m$ eigenvalues. The canonical form coordinates in the embedding space are given by

$$\mathbf{X}' = \mathbf{V}^m_+(\boldsymbol{\Lambda}^m_+)^{1/2}. \tag{22}$$

Classical scaling approximates the matrix $\mathbf{B}_{\boldsymbol{\Delta}}$ by a matrix $\mathbf{B}_{\boldsymbol{\Delta}'} = \mathbf{X}'\mathbf{X}'^{\mathrm{T}}$ of rank m. It is one of the most efficient MDS algorithms. In practical application, since we are usually interested in embedding into $\mathbb{R}^3$ or $\mathbb{R}^2$, no full eigendecomposition of $\mathbf{B}_{\boldsymbol{\Delta}}$ is needed—it is enough to find only the first three or two eigenvectors. In addition, the matrix $\mathbf{B}_{\boldsymbol{\Delta}}$ is symmetric. This allows us to use efficient algorithms such as Arnoldi (Arnoldi, 1951), Lanzcos or block-Lanzcos (Golub and van Loan, 1996; Bai et al., 2000) iterations, which find a few largest eigenvectors of a symmetric matrix (Golub and Saunders, 2004).

The major drawback of classical scaling is that the embedding error criterion (referred to as *strain*) that such a procedure minimizes is much less meaningful in our setting compared to the stress used in LS MDS (Borg and Groenen, 1997).

3.6. Canonical form Alignment

In order to resolve the Euclidean isometry ambiguity, the canonical form has to be aligned. We perform the alignment by first setting to zero the first-order moments (the center of gravity) $\mu_{100}, \mu_{010}, \mu_{001}$ of the

canonical form (here $\mu_{pqr} = \sum_{i=1}^{N} (x_i'^1)^p (x_i'^2)^q (x_i'^3)^r$ denotes the pqr-moment). This resolves the translation ambiguity. Next, the mixed second-order moments $\mu_{110}, \mu_{011}, \mu_{101}$ are set to zero and the axes are reordered according to the second order moments magnitude, where the projection onto the first axis x'^1 realizes the largest variance, and onto the third axis x'^3 the smallest variance. This resolves the rotation ambiguity. Finally, reflections are applied to each axis x^k such that

$$\sum_{i=1}^{N} \text{sign}\left(x_i'^k\right) \geq 0, \tag{23}$$

in order to resolve the reflection ambiguity.

As an alternative, it is possible to use the relations between coordinates of some fiducial points (e.g. nose tip and eyes) in order to perform the canonical form alignment.

3.7. Stability

An important property that makes the canonical forms practically useful is that the embedding result $\mathbf{X}'$ changes continuously with the change of $\mathbf{D}$. This guarantees that a small perturbation in the points of the surface does not change significantly the canonical form. We show here stability of canonical forms obtained by LS and classical MDS; while the first (Theorem 1a) is quite straightforward, the second requires more delicate analysis involving matrix perturbation theory.

Theorem 1 (*Canonical form stability*)**.**

(a) *A small perturbation of the geodesic distances* $\mathbf{D}$ *results in a small perturbation of the LS canonical form satisfying the stationary condition* $\mathbf{X}' = \mathbf{U}^\dagger \mathbf{B}(\mathbf{X}')\mathbf{X}'$.

(b) *Assume that the matrix* $\mathbf{B}_\Delta$ *has non-degenerate eigenvalues* $\lambda_1 < \lambda_2 < \cdots < \lambda_N$. *Then, a small perturbation of the geodesic distances* $\mathbf{D}$ *results in a small perturbation of the classical canonical form.*

Remark 1. The assumption of non-degeneracy of the spectrum of $\mathbf{B}_\Delta$ is usually valid for curved surfaces such as the facial surface. □

Proof: See Appendix.

3.8. Canonical Forms of Facial Surfaces

By performing canonization of a facial surface using MDS, we create an expression-invariant representation of the face, which can be treated as a rigid object. We should keep in mind that the embedding error criterion is *global*, that is, depends on *all* the geodesic distances. Hence, the canonical form is influenced by all the samples of the surface. As an implication, it is very important to always crop the same region of the face (though the samples themselves should not necessarily coincide). This is one of the key roles that an appropriate preprocessing plays.

When embedding is performed into $\mathbb{R}^3$, the canonical form can be plotted as a surface. Examples are shown in Figs. 7(b) and 8. Figure 8 depicts canonical forms of a subject's face with different facial expressions. It demonstrates that although the facial surface deformations are substantial, the deformations of the corresponding canonical forms are much smaller.

Embedding of a facial surface into $\mathbb{R}^2$ (Fig. 7(c)) or $\mathbb{S}^2$ is a zero codimension embedding—it produces an intrinsic parametrization of the surface. Such embedding can be thought of as "warping" of the facial texture, yielding a *canonical texture* (Bronstein et al., 2004a). Canonization in some sense performs a geometry-based registration of 2D facial images. That is, in two canonical textures we are able to identify two pixels as belonging to the same point on the 3D facial surface, no matter how the latter is deformed. As the result, 2D appearance-based methods such as

Figure 7. Example of a facial surface embedding into $\mathbb{R}^3$ (b) and into $\mathbb{R}^2$(c).

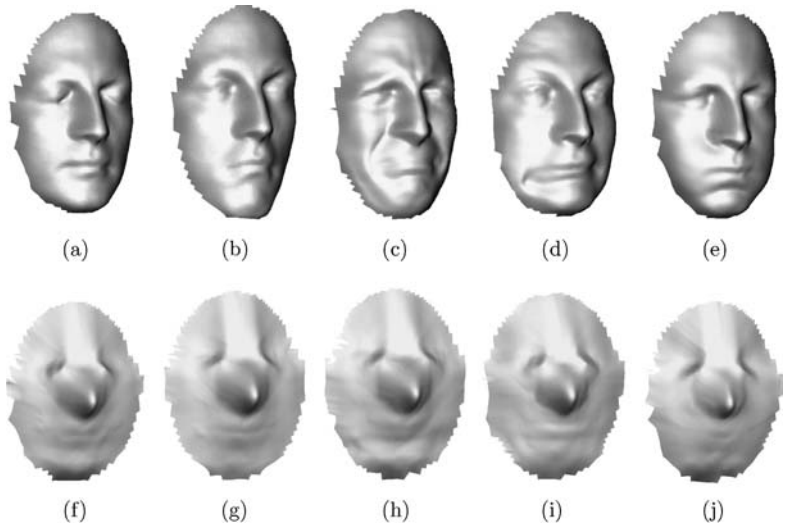


Figure 8. Examples of canonical forms (f)–(j) of faces with strong facial expressions (a)–(e).

eigenfaces can be used to compare canonical textures (Bronstein et al., 2004a).

4. The 3DFACE System

In the Geometric Image Processing laboratory (Department of Computer Science, Technion) we designed a fully automatic prototype 3D face recognition system based on the expression-invariant representation of facial surfaces. Current 3DFACE system prototype is shown in Fig. 9. It operates both in one-to-one and one-to-many modes. The 3DFACE system runs on a dual AMD Opteron64 workstation under Microsoft Windows XP. One of the CPUs is dedicated merely to processing and computation of the canonical forms; another one handles the graphical user interface (GUI) and the visualization.

Data processing in the 3DFACE system can be divided into several stages (Fig. 11). First, the subject's face undergoes a 3D scan, producing a cloud of points

Figure 9. The 3DFACE prototype system and its main components: DLP projector (a), digital camera (b), monitor (c), magnetic card reader (d), mounting (e) .

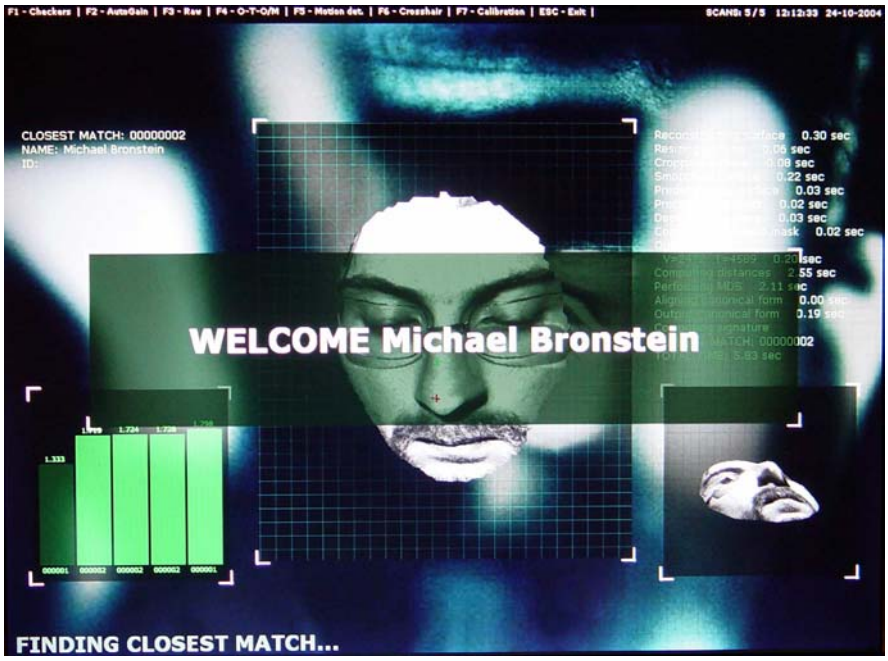


Figure 10. Screenshot from the graphical user interface of the 3DFACE system (one-to-many mode), showing successful recognition of one of the authors.

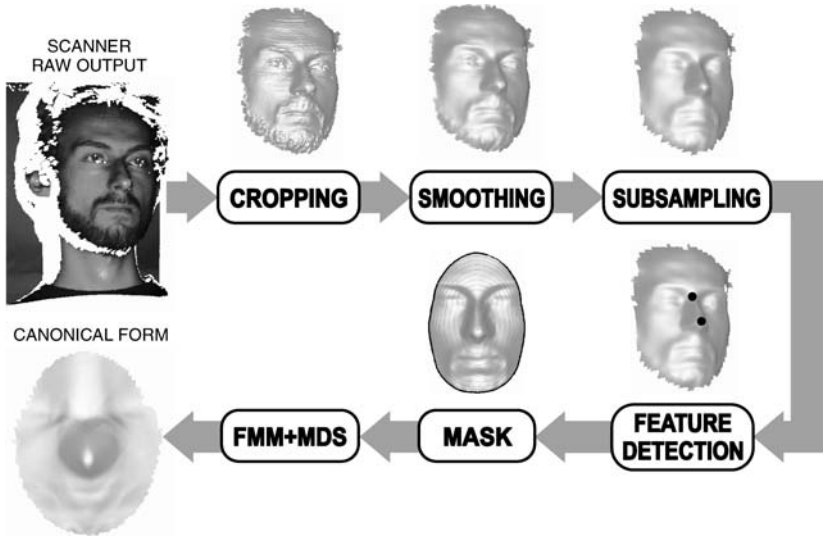


Figure 11. Scheme of the 3DFACE system pipeline.

representing the facial surface. The surface is then cropped, smoothed and subsampled (subsampling is performed in order to reduce the computational complexity of subsequent stages). Then, a feature detector is applied in order to find a few fiducial points (eyes, nose tip and nose apex). Next, a geodesic mask is computed around the nose apex and nose tip. Finally, the facial surface undergoes canonization using an MDS procedure.

In the following, we briefly overview all the processing stages and the basic components of the 3DFACE system pipeline, omitting some insignificant technological and implementation details.

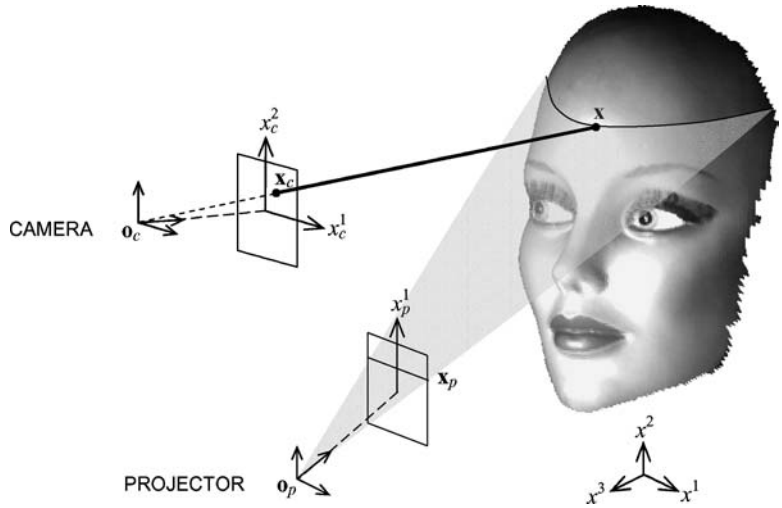


Figure 12. Active stereo principle.

4.1. 3D Acquisition

A single camera, which can be 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 Fig. 12) 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 Fig. 12. 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 our system (Bronstein et al., 2003a). Positioning the camera and the projector in front of the subject, such that occlusions are avoided, allows us to capture the whole facial surface.

Remark 2. Extreme rotations of the head may result in occlusions. In current implementation, we use a single camera, that limits 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 (ξ^1, ξ^2), which are now partitioned into a uniform grid of N_0 pixels. Then, for every pixel $\boldsymbol{\xi}_i$ $(i = 1, \ldots, N_0)$ we have an estimate of $\mathbf{x}(\boldsymbol{\xi}_i)$. In other words, the scanner produces a sampled version $\{\mathbf{x}_i = \mathbf{x}(\boldsymbol{\xi}_i)\}_{i=1}^{N_0}$ of the facial surface. 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 floating point matrices, each of size 320×240, corresponding to the values of x^1, x^2 and x^3 at each pixel. Thereby, the scanner output is a cloud of $N_0 = 76.8 \times 10^3$ points in 3D.

Reflectance image is also acquired. Assuming Lambertian reflectance, it is possible to estimate the albedo in every pixel (Gheorghiades et al., 2001).

4.2. Cropping

After 3D acquisition, the raw scanner data undergoes several preprocessing stages. First, a preliminary cropping is performed, separating the background from the facial region and removing problematic points in which the reconstruction is inaccurate (the latter usually appear as spikes). A histogram of the depth coordinates is used to roughly separate the face from the background. The facial region is defined by a binary 320×240 mask image, whose computation involves several thresholds: for example, pixels with a high value of the discrete gradient norm $\|\nabla x^3\| = \sqrt{(\partial_1 x^3)^2 + (\partial_2 x^3)^2}$ are removed as potential spikes. Morphological operations are then applied to the mask in order to remove non-connected regions and isolate the facial region (which will be denoted by Ω_c) as a single object. Holes inside the facial contour are closed by interpolation.

4.3. Smoothing

The next stage is facial surface smoothing, for which a variant of the *Beltrami flow* is employed. This type of nonlinear filter was proven to be a powerful method in color image processing (Sochen et al., 1998; Kimmel, 2003). Our smoothing procedure resembles the anisotropic diffusion in Tasdizen et al. (2002) and Fleishman et al. (2003); it does not require the surface to be given explicitly.[1]

The key idea of such smoothing is the minimization of the surface area, which is expressed by the so-called *Polyakov functional*

$$F[(\mathcal{S}, g)] = \int_{\Omega_c} d^2\boldsymbol{\xi} \sqrt{g} g^{kl} \partial_k x^i \partial_l x^j \delta_{ij}. \qquad (24)$$

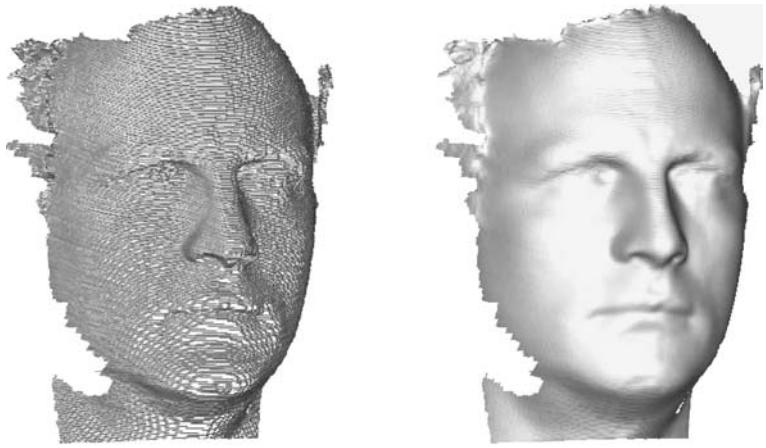
The Polyakov functional can be considered as a generalization of the L_2 norm to curved spaces. Using the Euler-Lagrange lemma, the minimum of F can be achieved by gradient descent of the form

$$\partial_t x^i = \frac{1}{\sqrt{g}} \frac{\delta F}{\delta x^i}; \quad i = 1, \ldots, 3, \qquad (25)$$

which can be written as $\partial_t \mathbf{x} = \Delta_g \mathbf{x}$ using the Laplace-Beltrami operator. t is a time parameter, and the flow can be thought of as evolving the surface in time. Since we assume the surface to be given as a graph of the function $x^3(x^1, x^2)$, filtering just the depth coordinate x^3 according to (25) gives rise to the following PDE

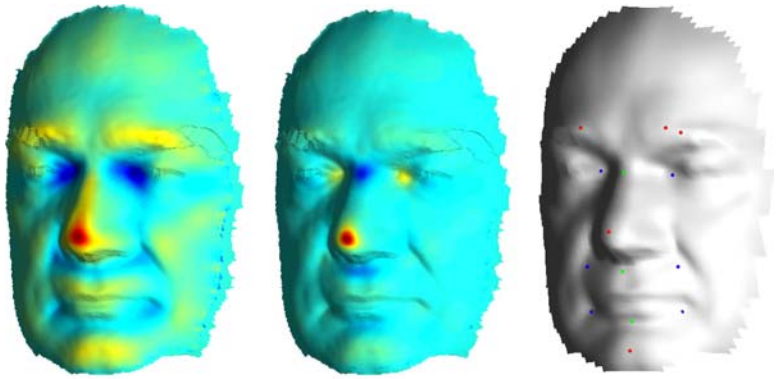
$$\partial_t x^3 = \frac{1}{\sqrt{g}} \partial_i (\sqrt{g} g^{ij} \partial_j x^3). \qquad (26)$$

Smoothing is performed by a numerical solution of (26) for a time proportional to the noise variance. The processing is limited to the facial contour Ω_c; Neumann boundary conditions $\nabla x^3|_{\partial\Omega_c} = \mathbf{0}$ are used.



(a) Before smoothing (b) After smoothing

Figure 13. Pre-processing of the 3D scanner data: (a) raw scanner data after cropping and (b) Beltrami smoothing result. Surfaces are rendered using Phong shading.



(a) Mean curvature (b) Gaussian curvature (c) Candidate points

Figure 14. Geometric feature detection example: mean curvature (a) and Gaussian curvature (b) images computed on the surface smoothed by mean-curvature flow and shown mapped onto the original surface; candidate nose (red), apex (green) and eyes (blue) points (this is a color figure).

4.4. Feature Detection

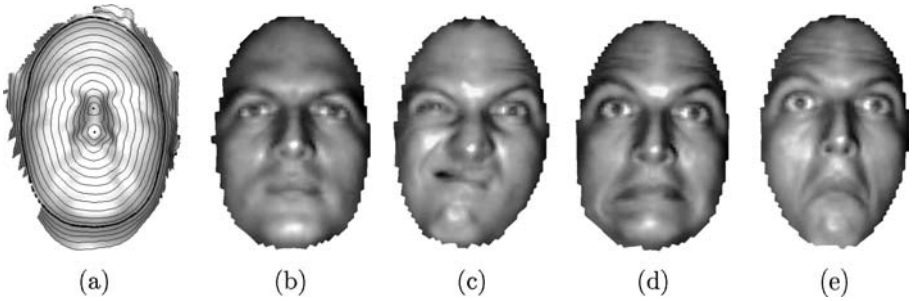
Next, the feature detection stage is performed. We are interested in invariant features that do not change under facial expressions; specifically, we detect the nose tip, the nose apex (the top part of the nose bridge bone) and the eye corners. In order to ensure invariance to head rotations, we use a geometric curvature-based feature detector.

Figure 14(a) and (b) depicts the mean and the Gaussian curvatures of a facial surface. The fiducial points can be very clearly classified according to these two curvatures. Our feature detector first locates the points corresponding to local extrema of K and H. Then, candidate points are specified; e.g. the candidate nose locations are points for which both K and H obtain a local maximum (Fig. 14(c)). Next, we use a set of geometric relations (e.g. that the nose apex is located between the two eyes, above the nose tip, and within certain distance intervals) to choose the best set of candidate points. The failure rate of such a feature detector is below 1%.

4.5. Geodesic Mask

At the last preprocessing stage, the facial contour is extracted by using the *geodesic mask*. The key idea is locating invariant "source" points on the face and measuring an equidistant (in sense of the geodesic distances) contour around it. The geodesic mask is defined as the interior of this contour; all points outside the contour are removed. This allows us to crop the facial surface in a geometrically-consistent manner, insensitively to facial expressions. Geodesic mask plays a crucial role for a good performance of canonical form-based face recognition.

The geodesic mask is computed as a contour around two source points: the nose tip and the nose apex (Fig. 15). The radius of the geodesic mask is set empirically; typical values vary between 80 and 100 mm. Fast



(a) (b) (c) (d) (e)

Figure 15. Geodesic mask computation around the nose tip and the nose apex points (a) and examples of the geodesic mask insensitivity to facial expressions (b)–(e).

Marching is used to compute the geodesic contours. After cropping, the resulting surface contains about 2500–3000 points; this corresponds to approximately 35% of the initial number of surface points.

4.6. Canonization and Surface Matching

After going through all the preprocessing stages, the smoothed, resized, and cropped facial surface undergoes canonization. In current 3DFACE system prototype, we use embedding into $\mathbb{R}^3$ by means of a few (about 40) iterations of the SMACOF algorithm minimizing a non-weighted version of the raw stress. As the initialization, we use the original points coordinates. The resulting canonical form then undergoes alignment according to Section 3.6.

The final stage of the 3DFACE algorithm is canonical form matching. Since the isometric embedding compensates for the isometric surface transformations, standard rigid matching (see e.g., Gruen and Akca, 2004) can be used for comparing the canonical forms. The standard choice in surface matching is the iterative closest point (ICP) method and its variants (Besl and McKay, 1992; Zhang, 1994), yet, it is disadvantageous from the point of view of computational complexity.

We use a simple and efficient surface matching method based on high-order moments (Tal et al., 2001). The main idea is to represent the surface by its moments μ_{pqr} up to some degree $P \leq p + q + r$, and compare the moments as vectors in an Euclidean space. Given two facial surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$ with their corresponding canonical forms $\mathbf{X}_1'$ and $\mathbf{X}_2'$, we can define the moments-based distance between them

$$d^{\text{mom}}(\mathcal{S}_1, \mathcal{S}_2) = \sum_{p+q+r \leq P} (\mu_{pqr}(\mathbf{X}_1') - \mu_{pqr}(\mathbf{X}_2'))^2. \tag{27}$$

In practice, the vectors of moments are those stored in the gallery database and compared to those computed at each enrollment.

4.7. Fusion of 3D and 2D Information

In Bronstein et al. (2003b, 2004a) we proposed to treat canonical forms as images, by performing zero-codimension embedding into a plane.[2] After alignment, both the canonical form and the flattened albedo are interpolated onto a Cartesian grid, producing two images. These images can be compared using standard techniques, such as applying eigendecomposition like in the eigenfaces algorithm. We called the obtained representation *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 in a natural way. Here, however, we focus on the 3D geometry, and in the following experiments present recognition results based only on the surface geometry, ignoring the texture.

5. Results

Our experiments were performed on a data set containing 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 were acquired in a course of several months, variations in illumination conditions, facial hair, etc. present. Subjects `Alex` (blue) and `Michael` (red) are identical twins, having visually great similarity (see Fig. 16).

5.1. Experiment I—Sensitivity to Facial Expressions

The goal of the first experiment was to demonstrate the difference between using original (rigid) 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 52), according to (27), was used. In Experiment I, we used a subset containing 10 human and 3 artificial subjects. Each face appeared in a number of instances (a total of 204 instances), including neutral, weak, medium and strong facial expressions (Fig. 16).

Figure 17 visualizes the dissimilarities between faces using classical scaling. 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 $\mathbb{R}^{52}$ (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 rigid surface matching (right) methods produce approximately the same results.

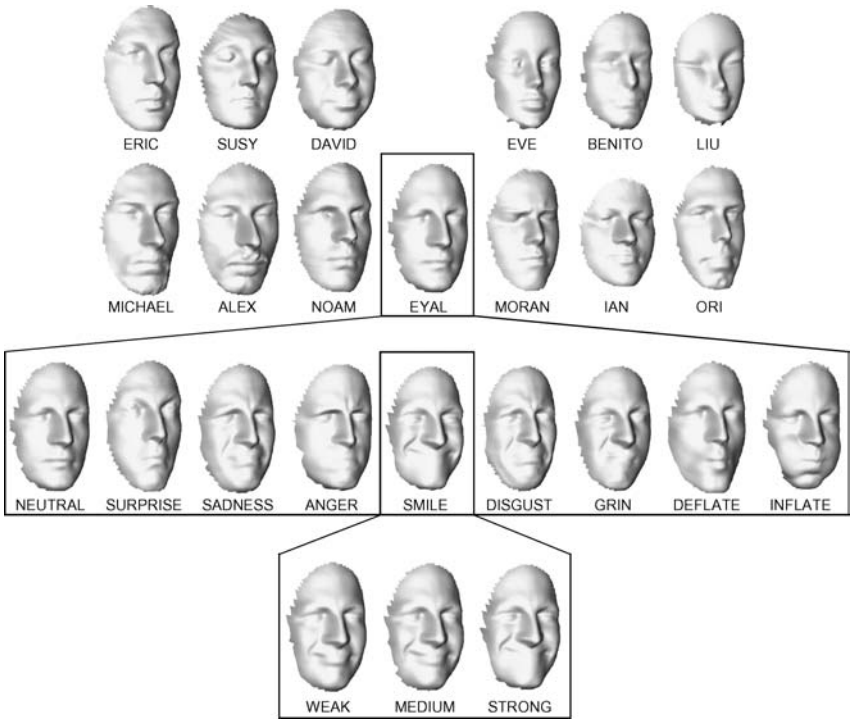


Figure 16. The face database used in Experiment I, containing 10 human and 3 artificial subjects. The facial expression variability is represented by 8 different expressions. Each expression appears in 3 strengths.

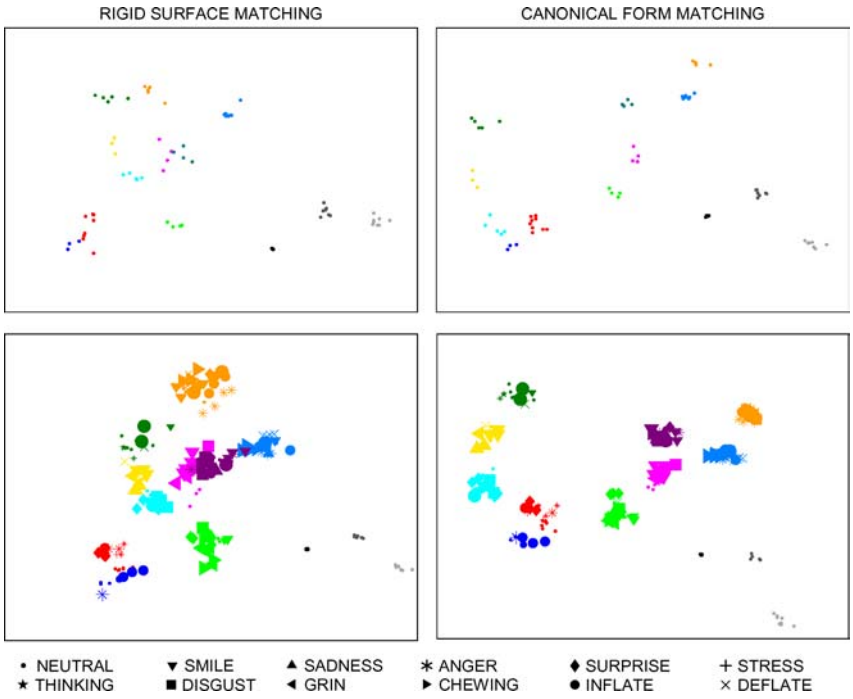


Figure 17. Low-dimensional visualization of dissimilarities between faces using rigid surface matching (left) and canonical form matching (right). First row: neutral expressions only. Second row: all expressions. Colors represent different subject. Symbols represent different facial expression. Symbol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

Table 1. Description of the facial expressions in data set used in experiment I and the inter-cluster to intra-cluster dissimilarity ratios σ_k and ς_k using rigid and canonical surface matching.

						Neutral expressions				All expressions			
Subject	Color	Neut.	Weak	Med.	Str.	σ_k^{can}	σ_k^{rig}	ς_k^{can}	ς_k^{rig}	σ_k^{can}	σ_k^{rig}	ς_k^{can}	ς_k^{rig}
Michael**	red	6	5	6	–	0.48	**0.51**	2.60	**3.71**	0.49	**0.52**	2.60	**3.71**
Alex**	blue	3	1	3	1	0.18	**0.20**	0.53	**0.88**	0.47	**0.64**	1.34	**2.24**
Eyal	green	4	1	7	9	0.20	**0.22**	**0.28**	0.22	0.46	**0.72**	1.34	**2.45**
Noam	yellow	3	–	–	7	0.29	**0.33**	**0.92**	0.52	0.47	**0.70**	1.52	**1.67**
Moran	magenta	4	–	4	10	0.32	**0.52**	0.51	**1.25**	0.42	**0.80**	1.42	**2.97**
Ian	orange	5	–	16	7	0.16	**0.21**	0.31	**0.38**	0.19	**0.42**	0.72	**1.19**
Ori	cyan	8	–	11	10	0.11	**0.18**	0.32	**0.54**	0.38	**0.60**	1.84	**2.53**
Eric	d. green	5	3	–	3	0.22	**0.28**	0.60	**0.77**	0.44	**1.08**	1.09	**3.94**
Susy	d. magenta	6	–	9	8	0.29	**0.32**	0.58	**0.91**	0.35	**0.49**	1.56	**2.45**
David	l. blue	5	2	6	5	0.26	0.26	**1.11**	0.72	0.34	**1.55**	1.95	**7.87**
Eve*	black	6	–	–	–	0.02	0.02	**0.05**	0.03	0.04	0.04	**0.05**	0.03
Benito*	grey	7	–	–	–	0.14	0.14	0.34	**0.44**	0.24	0.24	0.34	**0.44**
Liu*	l. grey	8	–	–	–	0.11	**0.12**	0.36	**0.39**	0.20	0.20	0.36	**0.39**

Shown separately for neutral and all expressions. Worse results are emphasized in boldface. Asterisk denotes artificial subjects. Double asterisk denotes identical twins.

This idealistic picture breaks down when we allow for facial expressions (Fig. 17, second row). The clusters corresponding to canonical surface matching are much tighter; moreover, we observe that using rigid 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 rigid surface matching would confuse between these subjects.

Table 1 shows the values of the ratio of the maximum inter-cluster to minimum intra-cluster dissimilarity

$$\varsigma_k = \frac{\max_{i,j\in\mathcal{C}_k} d_{ij}^{\text{mom}}}{\min_{i\notin\mathcal{C}_k, j\in\mathcal{C}_k} d_{ij}^{\text{mom}}}, \tag{28}$$

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

$$\sigma_k = \sqrt{\frac{\frac{2}{|\mathcal{C}_k|^2-|\mathcal{C}_k|}\sum_{i,j\in\mathcal{C}_k,\ i>j}\left(d_{ij}^{\text{mom}}\right)^2}{\frac{1}{|\mathcal{C}_k|(|\mathcal{C}|-|\mathcal{C}_k|)}\sum_{i\notin\mathcal{C}_k, j\in\mathcal{C}_k}\left(d_{ij}^{\text{mom}}\right)^2}}, \tag{29}$$

($\mathcal{C}_k$ denotes indexes of k-th subject's faces, $\mathcal{C} = \bigcup_k \mathcal{C}_k$ and d_{ij}^{mom} denotes the moment-based distance between faces i and j) for rigid and canonical surface matching. These criteria are convenient being scale-invariant; they measure the tightness of each cluster and its relative distance from other clusters. Ideally, σ_k and ς_k should tend to zero.

Table 1 shows the separation quality criterion ς_k for rigid and canonical surface matching. When only neutral expressions are used, canonical form matching slightly outperforms rigid surface matching on most subjects. The explanation to the fact that canonical forms are better even in case when no large expression variability is present, is that "neutral expression" as a fixed, definite expression, does not exist, and even when the face of the subject seems expressionless, its possible deformations are still sufficiently significant. When allowing for facial expressions, our approach outperforms facial surface matching by up to 358% in sense of σ_k and up to 305% in sense of ς_k.

5.2. *Experiment II—Comparison of Algorithms*

The goal of Experiment II was to perform a benchmark of the 3DFACE method and compare it to other face recognition algorithms. For this purpose, we simulated a biometric identification setup, in which the face of an enrolled subject (probe) was compared to a set of templates stored in the gallery. The probe database contained 30 subjects with different facial expressions. The number of templates in the gallery was 65 (several templates for each subject were used). Only neutral expressions were used as templates. Three algorithms were tested: canonical form matching, facial surface matching and 2D image-based eigenfaces. Eigenfaces

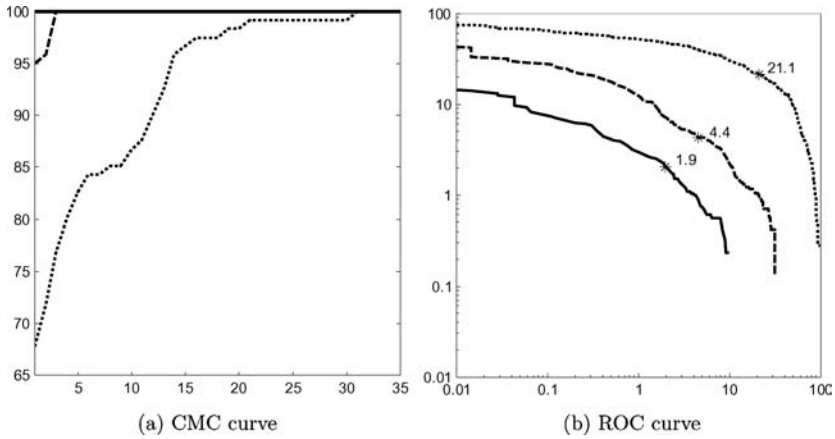


Figure 18. CMC (a) and ROC (b) curves of face recognition based on surface matching (dashed), canonical surface matching (solid) and eigenfaces (dotted). Star denotes equal error rate.

were trained by 35 facial images that did not appear as templates; 23 eigenfaces were used for the recognition (the first two eigenfaces were excluded in order to decrease the influence of illumination (Gheorghiades et al., 2001)).

Figure 18(b) shows the Receiver Operation Characteristic (ROC) curves of three algorithms compared in this experiment. The curve represents the tradeoff between the false acceptance and false rejection rate (FAR and FRR, respectively), as function of a global 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).

Figure 18(a) shows the Cumulative Match Characteristic (CMC) on the full database with facial expressions using the above algorithms. This curve represents the recognition error as a function of the recognition rank (the number of closest matches, at least one of which should be correct). Our approach results in *zero* rank-1 recognition error.

Figure 19 (first row) shows an example of rank-1 recognition on the full database (220 instances with facial expressions). The first column depicts a probe subject with extreme facial expression; columns two through four depict the rank-1 matches among the 65 gallery templates using eigenfaces, rigid surface matching, and canonical form matching. These results are typical for the described algorithms. The eigenfaces algorithm, 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.

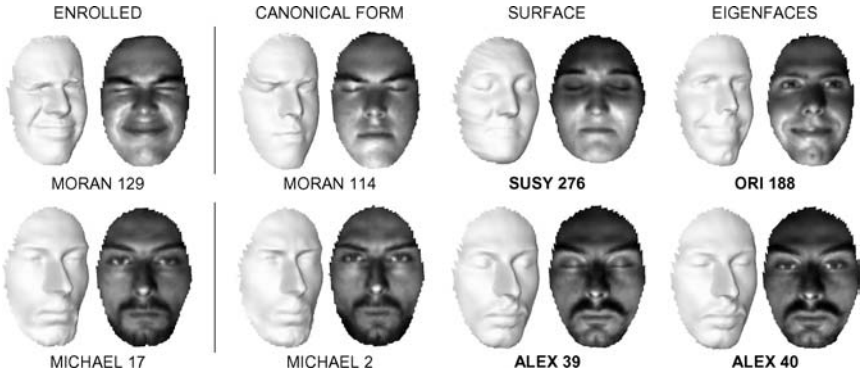


Figure 19. Example of recognition using different algorithms. First column shows the reference subject; second through fourth columns depict the closest (rank 1) matches found by the canonical form matching, facial surface matching and eigenfaces, respectively. Note that only the match using canonical 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.

Facial surface matching is confused by 3D features (outstanding inflated cheeks) that appear on the face of the probe 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 isometric embedding compensates for the distortion of the face of subject `Moran 129` caused by his smile.

5.3. Experiment III—Twins Test

In Experiment III, 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 all 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`. Rigid 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 Fig. 19, second row).

6. Conclusions

We presented a geometric framework for three-dimensional face recognition, that naturally incorporates the non-rigidness of the facial surfaces and allows to cope with facial expressions. Our isometric model reduces the problem of comparing faces in the presence of facial expressions to the problem of isometric surface matching. We use the isometric embedding as a method to construct isometry-invariant representation of facial surfaces, which gives an expression-invariant representation of the face. An implementation of our 3DFACE algorithm shows high recognition accuracy, significantly outperforming standard 2D and 3D approaches, and working favorably even in the presence of very strong facial expressions.

The main advantage of the 3DFACE algorithm is obvious when a large variability of facial expressions is present. The significance of it is first of all in commercial applications, which can never assume collaborative users, and especially in cases when the face recognition is assumed to work in a natural environment. Ideally, the user will be unaware of being scanned and recognized, which implies that the variability of his facial expressions can be significant.

Yet, the results presented in Section 5 suggest that canonical forms are useful not only when extreme facial expressions are present, but also in cases where the faces seem to be completely expressionless. The explanation, as we see it, is the fact that an absolute "neutral expression" does not exist, such that even when apparently without expressions, the use of canonical forms can still be beneficial.

Besides being insensitive to expressions, canonical forms conceal some other favorable properties. First, the obtained representation is practically irreversible. Therefore, given the canonical form it is impossible (or at least very hard) to find the underlying original facial surface. Thus, the canonical form in some sense "hides" the actual identity of the subject stored in the gallery. This is significant in commercial systems where the security of the biometric data is an important issue. Secondly, canonical forms provide an intrinsic parametrization of the facial surface, which leads to an easy registration of the facial images, and consequently, to an easy fusion of 2D and 3D information. Thirdly, embedding has a "regularization" effect on the facial surface: small local artifacts lead to fluctuation of all geodesic distances. The fluctuations in the canonical forms are no more local, but rather "spread" among all points of the canonical form. As a consequence, the canonical forms are less sensitive for example to acquisition and processing artifacts than the original surfaces.

In this paper, we limited our discussion to canonical representation in Euclidean spaces only. However, virtually every space, which is sufficiently simple to possess a closed-form expression for the geodesic distances, is suitable for "housing" the canonical form. For example, in Bronstein et al. (2005a, 2005b), we showed that embedding of the facial geometry into $\mathbb{S}^3$ or the facial albedo into $\mathbb{S}^2$ allows to achieve more than twice lower embedding error (in the sense of stress), without compromising the computational efficiency of the embedding. The pursuit for spaces other than Euclidean is motivated by a simple intuitive conjecture that better representation accuracy leads to better recognition. An empirical evidence that this claim does not hold in reality is presented in Bronstein et al. (2005a).

Appendix

Proof of Theorem 1 (Canonical form stability)**:** (a) For simplicity of the proof, let us assume $w_{ij} = 1$, such that $\mathbf{U}^\dagger = \frac{1}{N}\mathbf{J}$. Consider the stationary condition

$$\mathbf{X}' = \mathbf{U}^\dagger \mathbf{B}(\mathbf{X}';\mathbf{D})\mathbf{X}' = \frac{1}{N}\mathbf{B}(\mathbf{X}';\mathbf{D})\mathbf{X}'. \tag{30}$$

We will write (30) as a nonlinear homogeneous equation in $\mathbf{X}'$ and $\mathbf{D}$:

$$\mathcal{A}(\mathbf{X}';\mathbf{D}) = \mathbf{0}. \tag{31}$$

Now assume that $\mathbf{D}$ undergoes a perturbation, such that $\tilde{\mathbf{D}} = \mathbf{D} + \delta\mathbf{D}$, and $\max|\delta d_{ij}| < \epsilon$. The new canonical form will be denoted by $\tilde{\mathbf{X}}' = \mathbf{X}' + \delta\mathbf{X}'$. The stationary condition with $\tilde{\mathbf{X}}'$ and $\tilde{\mathbf{D}}$ becomes

$$\mathcal{A}(\mathbf{X}' + \delta\mathbf{X}';\mathbf{D} + \delta\mathbf{D}) = \mathbf{0}, \tag{32}$$

and by the Taylor expansion, neglecting $\mathcal{O}(\epsilon^2)$ terms and plugging in Eq. (31), we have

$$\begin{aligned}
&\mathcal{A}(\mathbf{X}' + \delta\mathbf{X}';\mathbf{D} + \delta\mathbf{D}) \\
&\quad = \mathcal{A}(\mathbf{X}';\mathbf{D}) + \nabla_{\mathbf{X}'}\mathcal{A}(\mathbf{X}';\mathbf{D})\delta\mathbf{X}' + \nabla_{\mathbf{D}}\mathcal{A}(\mathbf{X}';\mathbf{D})\delta\mathbf{D} \\
&\quad = \nabla_{\mathbf{X}'}\mathcal{A}(\mathbf{X}';\mathbf{D})\delta\mathbf{X}' + \nabla_{\mathbf{D}}\mathcal{A}(\mathbf{X}';\mathbf{D})\delta\mathbf{D} = \mathbf{0},
\end{aligned} \tag{33}$$

where $\nabla_{\mathbf{X}'}\mathcal{A}(\mathbf{X}';\mathbf{D})$ and $\nabla_{\mathbf{D}}\mathcal{A}(\mathbf{X}';\mathbf{D})$ are four-dimensional tensors denoting the derivatives of $\mathcal{A}$ with respect to $\mathbf{X}'$ and $\mathbf{D}$, respectively.

Let us now write $\mathcal{A}(\mathbf{X}';\mathbf{D})$ in coordinate notation:

$$a_i^j = (\mathcal{A}(\mathbf{X}';\mathbf{D}))_i^j = \frac{1}{N}\sum_k b_{ik}x_k'^j - x_i'^j. \tag{34}$$

Plugging in $b_{ii} = -\sum_{j\neq i} b_{ij}$ according to the definition of $\mathbf{B}$, we obtain

$$a_i^j = \frac{1}{N}\sum_k b_{ik}\left(x_k'^j - x_i'^j\right) - x_i'^j. \tag{35}$$

Differentiating with respect to $\mathbf{D}$ and using (19) yields

$$\begin{aligned}
\frac{\partial a_i^j}{\partial d_{i'j'}} &= \frac{1}{N}\sum_k \left(x_k'^j - x_i'^j\right)\frac{\partial b_{ik}}{\partial d_{i'j'}} \\
&= -\frac{1}{N}\sum_{k\neq i}\left(x_k'^j - x_i'^j\right)\frac{\delta_{ii'}\delta_{kj'}}{\|\mathbf{x}_i' - \mathbf{x}_k'\|}
\end{aligned} \tag{36}$$

$$= \begin{cases} -\frac{1}{N}\left(x_{j'}'^j - x_i'^j\right)\frac{\delta_{ii'}}{\|\mathbf{x}_i' - \mathbf{x}_{j'}'\|} & \text{if } k \neq j' \\ 0 & \text{otherwise} \end{cases}.$$

The derivatives can be bounded by

$$\begin{aligned}
\left|\frac{\partial a_i^j}{\partial d_{i'j'}}\right| &\leq \frac{1}{N}\max_{k\neq l,q}\left(\frac{|x_k'^q - x_l'^q|}{\|\mathbf{x}_k' - \mathbf{x}_l'\|_2}\right) \\
&= \frac{1}{N}\max_{k\neq l}\left(\frac{\|\mathbf{x}_k' - \mathbf{x}_l'\|_\infty}{\|\mathbf{x}_k' - \mathbf{x}_l'\|_2}\right) \equiv M_1,
\end{aligned} \tag{37}$$

and from the regular sampling assumption, it follows that $0 < M_1 < \infty$.

Differentiating with respect to $\mathbf{X}'$ yields

$$\begin{aligned}
\frac{\partial a_i^j}{\partial x_{i'}'^{j'}} &= \frac{1}{N}\sum_k\left(\left(x_k'^j - x_i'^j\right)\frac{\partial b_{ik}}{\partial x_{i'}'^{j'}}\right. \\
&\qquad \left. + b_{ik}\frac{\partial}{\partial x_{i'}'^{j'}}\left(x_k'^j - x_i'^j\right)\right) - \frac{\partial x_i'^j}{\partial x_{i'}'^{j'}} \\
&= -\frac{1}{N}\sum_{k\neq i}\left(x_k'^j - x_i'^j\right)\delta_{ii'}d_{ik}\frac{\left(x_k'^{j'} - x_i'^{j'}\right)}{\|\mathbf{x}_k' - \mathbf{x}_i'\|^3} \\
&\qquad - \frac{1}{N}\sum_k b_{ik}\delta_{ii'}\delta_{jj'} + \frac{1}{N}b_{ii'}\delta_{jj'} - \delta_{jj'}\delta_{ii'}.
\end{aligned} \tag{38}$$

The derivatives can be bounded by

$$\left|\frac{\partial a_i^j}{\partial x_{i'}'^{j'}}\right| \leq \max_{k\neq l}\|\mathbf{x}_k' - \mathbf{x}_l'\|^{-1}\max_{k\neq l}d_{kl} + 2\max_{kl}|b_{kl}| + 1 \tag{39}$$

$$\leq \max_{k\neq l}\frac{d_{kl}'}{d_{kl}} + 2\max_{kl}|b_{kl}| + 1 \equiv M_2, \tag{40}$$

and since all the quantities are bounded, we have $M_2 < \infty$.

Rewriting (33) in coordinate notation and plugging in the bounds we obtain

$$\begin{aligned}
0 = a_i^j &= \sum_{i'=1}^N\sum_{j'=1}^m \frac{\partial a_i^j}{\partial x_{i'}'^{j'}}\delta x_{i'}'^{j'} + \sum_{i'=1}^N\sum_{j'=1}^N \frac{\partial a_i^j}{\partial d_{i'j'}}\delta d_{i'j'} \\
&\leq \sum_{i'=1}^N\sum_{j'=1}^m M_1\delta x_{i'}'^{j'} + \sum_{i'=1}^N\sum_{j'=1}^N M_2\delta d_{i'j'},
\end{aligned} \tag{41}$$

which leads to a bound on the canonical form perturbation

$$\begin{aligned}
\max_{i'j'}\left|\delta x_{i'}'^{j'}\right| &\leq \sum_{i'=1}^N\sum_{j'=1}^m\left|\delta x_{i'}'^{j'}\right| \\
&\leq \frac{M_2}{M_1}\sum_{i'=1}^N\sum_{j'=1}^N|\delta d_{i'j'}| < \frac{M_2}{M_1}N^2\epsilon
\end{aligned} \tag{42}$$

□

(b) 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

$$\begin{aligned} \tilde{\mathbf{\Delta}} &= \mathbf{\Delta} + \delta\mathbf{\Delta} = (\mathbf{D} + \delta\mathbf{D}) \circ (\mathbf{D} + \delta\mathbf{D}) \\ &= \mathbf{\Delta} + 2\mathbf{D} \circ \delta\mathbf{D} + \delta\mathbf{D} \circ \delta\mathbf{D}. \end{aligned}$$

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

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

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

$$\begin{aligned} \tilde{\mathbf{B}}_{\mathbf{\Delta}} &= \mathbf{B}_{\mathbf{\Delta}} + \delta\mathbf{B}_{\mathbf{\Delta}} = -\frac{1}{2}\mathbf{J}(\mathbf{\Delta} + \delta\mathbf{\Delta})\mathbf{J} \\ &= \mathbf{B}_{\mathbf{\Delta}} - \frac{1}{2}\mathbf{J}\delta\mathbf{\Delta}\mathbf{J}. \end{aligned}$$

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

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

Eigendecomposition of the perturbed double-centered matrix yields $\tilde{\mathbf{B}}_{\mathbf{\Delta}} = \tilde{\mathbf{V}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{V}}^T$, such that the perturbed canonical form is

$$\tilde{\mathbf{X}}' = \tilde{\mathbf{V}}_+^m (\tilde{\mathbf{\Lambda}}_+^m)^{1/2}. \tag{43}$$

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

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

where $\theta \, (\mathbf{v}_i, \tilde{\mathbf{v}}_i)$ is the acute angle between the vectors $\mathbf{v}_i$ and $\tilde{\mathbf{v}}_i$, and

$$\text{gap}(\mathbf{B}_{\mathbf{\Delta}}) = \min_{i \neq j} |\lambda_i - \lambda_j|, \tag{44}$$

is the spectral gap of the matrix $\mathbf{B}_{\mathbf{\Delta}}$. The $\text{gap}(\mathbf{B}_{\mathbf{\Delta}})$ is non-zero, since we assume that $\mathbf{B}_{\mathbf{\Delta}}$ has non-degenerate eigenvalues. Under a small perturbation, the order of the eigenvalues is preserved, i.e. $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \cdots \leq \tilde{\lambda}_N$; from Taylor expansion

$$\tilde{\lambda}_i^{1/2} - \lambda_i^{1/2} \approx \frac{(\tilde{\lambda}_i - \lambda_i)}{2\lambda_i},$$

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

$$\begin{aligned} \| \mathbf{v}_i - \tilde{\mathbf{v}}_i \|_2 &\approx \sin \theta \, (\mathbf{v}_i, \tilde{\mathbf{v}}_i) \approx \theta \, (\mathbf{v}_i, \tilde{\mathbf{v}}_i) \\ &< \frac{\max d_{ij}}{\text{gap}(\mathbf{B}_{\mathbf{\Delta}})} \sqrt{N}\epsilon \end{aligned}$$

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

$$\begin{aligned} \| \mathbf{x}_i' - \tilde{\mathbf{x}}_i' \|_2 &= \left\| \lambda_i^{1/2} \mathbf{v}_i - \tilde{\lambda}_i^{1/2} \tilde{\mathbf{v}}_i \right\|_2 \\ &\leq \lambda_i^{1/2} \| \mathbf{v}_i - \tilde{\mathbf{v}}_i \|_2 + \left| \lambda_i^{1/2} - \tilde{\lambda}_i^{1/2} \right| \| \tilde{\mathbf{v}}_i \|_2 \\ &\leq \lambda_i^{1/2} \| \mathbf{v}_i - \tilde{\mathbf{v}}_i \|_2 + \frac{\| \delta\mathbf{B}_{\mathbf{\Delta}} \|_2}{2\lambda_i^{1/2}} \\ &< \left(\frac{\lambda_i^{1/2}}{\text{gap}(\mathbf{B}_{\mathbf{\Delta}})} + \frac{1}{2\lambda_i^{1/2}} \right) \sqrt{N}\epsilon \max d_{ij}. \end{aligned}$$

Remark 3. For convenience, we defined our bounds in terms of spectral matrix norms. Bounds on other norms can be obtained by using appropriate norm inequalities.

Acknowledgments

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 Michael Elad for helpful discussions; to Michael Saunders and Gene Golub for advices on efficient eigendecomposition implementation, to Eyal Gordon for help in data acquisition and to all the people who agreed to contribute their faces to our 3D face database.

Notes

1. When the facial manifold is not given explicitly, and all we have is its gradients on a uniform grid, it is possible to smooth the gradient field instead of the surface, thus bypassing the surface reconstruction stage.
2. More recently, we studied zero-codimension embedding into $\mathbb{S}^2$ (Bronstein et al., 2005a).

References

Achermann, B. and Bunke, H. 2000. Classifying range images of human faces with Hausdorff distance. In *Proc. ICPR*, pp. 809–813.

Achermann, B., Jiang, X., and Bunke, H. 1997. Face recognition using range images. In *Int'l Conf. Virtual Systems and Multimedia*, pp. 129–136.

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.

Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., and van der Vorst, H. 2000. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*. Philadelphia: SIAM, third edition. Online: http://www.cs.utk.edu/dongarra/etemplates/index.html.

Besl, P.J. and McKay, N.D. 1992. A method for registration of 3D shapes. *IEEE Trans. PAMI*, 14:239–256.

Beumier, C. and Acheroy, M.P. 1998. Automatic face authentication from 3D surface. In *Proc. British Machine Vision Conf.*, pp. 449–458.

Bledsoe, W.W. 1966. Man-machine face recognition. Technical Report PRI 22, Panoramic Research Inc., Palo Alto (CA) USA.

Borg, I. and Groenen, P. 1997. *Modern Multidimensional Scaling—Theory and Applications*. Springer-Verlag: Berlin Heidelberg New York.

Bowyer, K.W., Chang, K., and Flynn, P. 2004. A survey of 3D and multi-modal 3D + 2D face recognition. Dept. of Computer Science and Electrical Engineering Technical Report, University of Notre Dame.

Bronstein, M.M. 2004. Three-dimensional face recognition. Master's Thesis, Department of Computer Science, Technion—Israel Institute of Technology.

Bronstein, A.M., Bronstein, M.M., Gordon, E., and Kimmel, R. 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, M.M., Gordon, E., and Kimmel, R. 2004a. Fusion of 3D and 2D information in face recognition. In *Proc. ICIP*, pp. 87–90.

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

Bronstein, A.M., Bronstein, M.M., and Kimmel, R. 2005a. Expression-invariant face recognition via spherical embedding. In *Proc. ICIP*, to appear.

Bronstein, A.M., Bronstein, M.M., and Kimmel, R. 2005b. On isometric embedding of facial surfaces into $\mathbb{S}^3$. In *Proc. Int'l Conf. Scale Space and PDE Methods in Computer Vision*, Lecture Notes in Comp. Science 3459, Springer, pp. 622–631.

Bronstein, A.M., Bronstein, M.M., Kimmel, R., and Spira, A. 2004b. Face recognition from facial surface metric. In *Proc. ECCV*, pp. 225–237.

Bronstein, M.M., Bronstein, A.M., Kimmel, R., and Yavneh, I. 2005c. A multigrid approach for multi-dimensional scaling. In *Proc. Copper Mountain Conf. Multigrid Methods* (submitted).

Bronstein, M.M., Bronstein, A.M., and Kimmel, R. 2005. Expression-invariant representations for human faces. Technical Report CIS-2005-01, Dept. of Computer Science, Technion, Israel.

Brunelli, R. and Poggio, T. 1993. Face recognition: Features vs. templates. *IEEE Trans. PAMI*, 15(10):1042–1053.

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

Cartoux, J.Y., La Preste, J.T., and Richetin, M. 1989. Face authentication or recognition by profile extraction from range images. In *Proc. Workshop on Interpretation of 3D Scenes*, pp. 194–199.

Chang, K., Bowyer, K., and Flynn, P. 2003. Face recognition using 2D and 3D facial data. In *Proc. Multimodal User Authentication Workshop*, pp. 25–32.

Cox, I., Ghosn, J., and Yianilos, P. 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, Amsterdam: North-Holland, pp. 133–145.

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

Eckart, C. and Young, G. 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*. Academic Press: New York.

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

Elad, A. and Kimmel, R. 2002. *Geometric Methods in Bio-Medical Image Processing*, Vol. 2191, Chapt. Spherical flattening of the cortex surface. Springer-Verlag: Berlin Heidelberg New York, pp. 77–89.

Elad, A. and Kimmel, R. 2003. On bending invariant signatures for surfaces. *IEEE Trans. PAMI*, 25(10):1285–1295.

Fleishman, S., Drori, I., and Cohen-Or, D. 2003. Bilateral mesh denoising. In *Proc. SIGGRAPH*, pp. 950–953.

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

Georghiades, A.S., Belhumeur, P.N., and Kriegman, D. 1998. Illumination cones for recognition under variable lighting: Faces. In *Proc. CVPR*, pp. 52–58.

Gheorghiades, A.S., Belhumeur, P.N., and Kriegman, D.J. 2001. From few to many: Illumination cone models for face recognition

under variable lighting and pose. *IEEE Trans. PAMI*, 23(6):643–660.

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

Golub, G.H. and Saunders, M. 2004. Personal communication.

Golub, G.H. and van Loan, C.F. 1996. *Matrix Computations*. 3rd ed. The John Hopkins University Press.

Gordon, G. 1992. Face recognition based on depth and curvature features. In *Proc. CVPR*, pp. 108–110.

Gordon, G. 1997. Face recognition from frontal and profile views. In *Proc. Int'l Workshop on Face and Gesture Recognition*, pp. 47–52.

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

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

Gruen, A. and Akca, D. 2004. Least squares 3D surface matching. In *Proc. ISPRS Working Group V/1 Panoramic Photogrammetry Workshop*, pp. 19–22.

Gudmundsson, S. 2004. An introduction to riemannian geometry (lecture notes). Online. Available at http://www.matematik.lu.se/matematiklu/personal/sigma/Riemann.pdf.

Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika* 33, pp. 469–506.

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

Hesher, C., Srivastava, A., and Erlebacher, G. 2003. A novel technique for face recognition using range images. In *Int'l Symp. Signal Processing and Its Applications*, Vol. 2, pp. 201–204.

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

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

Hugli, H. and Maitre, G. 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., Chen, C.-S., Hsieh, I.-B., and Fuh, C.-S. 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.

Kearsley, A., Tapia, R., and Trosset, M. 1998. The solution of the metric STRESS and SSTRESS problems in multidimensional scaling using Newton's method. *Computational Statistics*, 13(3):369–396.

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

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

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

Lee, J.C. and Milios, E. 1990. Matching range images of human faces. In *Proc. ICCV*, pp. 722–726.

Linial, N., London, E., and Rabinovich, Y. 1995. The geometry of graphs and some its algorithmic applications. *Combinatorica*, 15:333–344.

Mansfield, T., Kelly, G., Chandler, D., and Kane, J. 2001. Biometric product testing final report. Technical report, Centre for Mathematics and Scientific Computing, National Physical Laboratory, UK.

Mavridis, N., Tsalakanidou, F., Pantazis, D., Malassiotis, S., and Strintzis, M.G. 2001. 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*, pp. 157–160.

Medioni, G. and Waupotitsch, R. 2003. Face recognition and modeling in 3D. In *Proc. AMFG*, pp. 232–233.

Mémoli, F. and Sapiro, G. 2004. Comparing point clouds, IMA preprint series no. 1978, University of Minnesota.

Nagamine, T., Uemura, T., and Masuda, I. 1992. 3D facial image analysis for human identification. In *Proc. ICPR*, pp. 324–327.

Nash, S. 2000. A multigrid approach to discretized optimization problems. *Journal of Optimization Methods and Software*, 14:99–116.

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

Pentland, A., Moghaddam, B., and Starner, T. 1994. View-based and modular eigenspaces for face recognition. In *Proc. CVPR*, pp. 84–91.

Phillips, P.J., Grother, P., Michaels, R.J., Blackburn, D.M., Tabassi, E., and Bone, J. 2003. FRVT 2002: Overview and summary. Online. Available at www.frvt.org.

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

Schwartz, E.L., Shaw, A., and Wolfson, E. 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 Kirby, M. 1987. Low-dimensional procedure for the characterization of human faces. *JOSA A*, 2:519–524.

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

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

Spira, A. and Kimmel, R. 2004. An efficient solution to the Eikonal equation on parametric manifolds. *Interfaces and Free Boundaries*, 6(3):315–327.

Stewart, G.W. and Sun, J.-G. 1990. *Matrix Perturbation Theory*. Academic Press.

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

Tal, A., Elad, M., and Ar, S. 2001. Content based retrieval of VRML objects—an iterative and interactive approach. In *Proc. Eurographics Workshop on Multimedia*.

Tanaka, H.T., Ikeda, M., and Chiaki, H. 1998. Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition. In *Proc. Int'l Conf. Automated Face and Gesture Recognition*, pp. 372–377.

Tasdizen, T., Whitaker, R., Burchard, P., and Osher, S. 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.

Tsalakanidou, F., Tzocaras, D., and Strintzis, M. 2003. Use of depth and colour eigenfaces for face recognition. *Pattern Recognition Letters*, 24:1427–1435.

Tsitsiklis, J.N. 1995. Efficient algorithms for globally optimal trajectories. *IEEE Trans. Automatic Control*, 40(9):1528–1538.

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

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

Walter, J. and Ritter, H. 2002. On interactive visualization of high-dimensional data using the hyperbolic plane. In *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 123–131.

Wang, Y., Chua, C., and Ho, Y. 2002. Facial feature detection and face recognition from 2D and 3D images. *Pattern Recognition Letters*, 23:1191–1202.

Wiskott, L. 1995. *Labeled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis*, No. 53 in Reihe Physik. Thun—Frankfurt am Main: Verlag Harri Deutsch, PhD Thesis.

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

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

Zhang, Z.Y. 1994. Terative point matching for registration of free-form curves and surfaces. *IJCV*, 13:119–152.

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