Expression-invariant three-dimensional face recognition

Alexander M. Bronstein
Email: alexbron@ieee.org

Michael M. Bronstein
bronstein@ieee.org

Ron Kimmel
ron@cs.technion.ac.il

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

One of the hardest problems in face recognition is dealing with facial expressions. Finding an expression-invariant representation of the face could be a remedy for this problem. We suggest treating faces as deformable surfaces in the context of Riemannian geometry, and propose to approximate facial expressions as isometries of the facial surface. This way, we can define geometric invariants of a given face under different expressions. One such invariant is constructed by isometrically embedding the facial surface structure into a low-dimensional flat space. Based on this approach, we built an accurate three-dimensional face recognition system that is able to distinguish between identical twins under various facial expressions. In this chapter we show how under the near-isometric model assumption, the difficult problem of face recognition in the presence of facial expressions can be solved in a relatively simple way.

0.1 Introduction

It is well-known that some characteristics or behavior patterns of the human body are strictly individual and can be observed in two different people with a very low probability – a few such examples include the DNA code, fingerprints, structure of retinal veins and iris, individual's written signature or face. The term *biometrics* refers to a variety of methods that attempt to uniquely identify a person according to a set of such features.

While many of today's biometric technologies are based on the discoveries of the last century (like the DNA, for example), some of them have been exploited from the dawn of the human civilization [17]. One of the oldest written testimonies of a biometric technology and the first identity theft dates back to biblical times, when Jacob fraudulently used the identity of his twin brother Esau to benefit from their father's blessing. The Genesis book describes a combination of hand scan and voice recognition that Isaac used to attempt to verify his son's identity, without knowing that the smooth-skinned Jacob had wrapped his hands in kidskin:

> "And Jacob went near unto Isaac his father; and he felt him, and said, 'The voice is Jacob's voice, but the hands are the hands of Esau'. And he recognized him not, because his hands were hairy, as his brother Esau's hands."

The false acceptance which resulted from this very inaccurate biometric test had historical consequences of unmatched proportions.

Face recognition is probably the most natural biometric method. The remarkable ability of the human vision to recognize faces is widely used for biometric authentication from prehistoric times. These days, almost every identification document contains a photograph of its bearer, which allows the respective officials to verify a person's identity by comparing his actual face with the one on the photo.

Unlike many other biometrics, face recognition does not require physical contact with the individual (like fingerprint recognition) or taking samples of the body (like DNA-based identification) or the individual's behavior (like signature recognition). For these reasons, face recognition is considered a natural, less intimidating, and widely accepted biometric identification method [4, 47], and as such, has the potential of becoming the leading biometric technology. The great technological challenge is to perform face recognition automatically, by means of computer algorithms that work without any

human intervention. This problem has been traditionally considered the realm of computer vision and pattern recognition. It is also believed to be one of the most difficult machine vision problems.

0.1.1 The problems of face recognition

The main difficulty of face recognition stems from the immense variability of the human face. The facial appearance depends heavily on the environmental factors, e.g. the lighting conditions, background scene and head pose. It also depends on the 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, e.g. 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. That is, unless the right measure is used to differ between individuals.

Theoretically, it is possible to recognize an individual's face reliably in different conditions, provided that the same person *has been previously observed in similar conditions*. However, the variety of images required to cover all the possible appearances of the face can be very large (see Figure 1). In practice, only a few observations of the face (and sometimes, even a single one) are available.

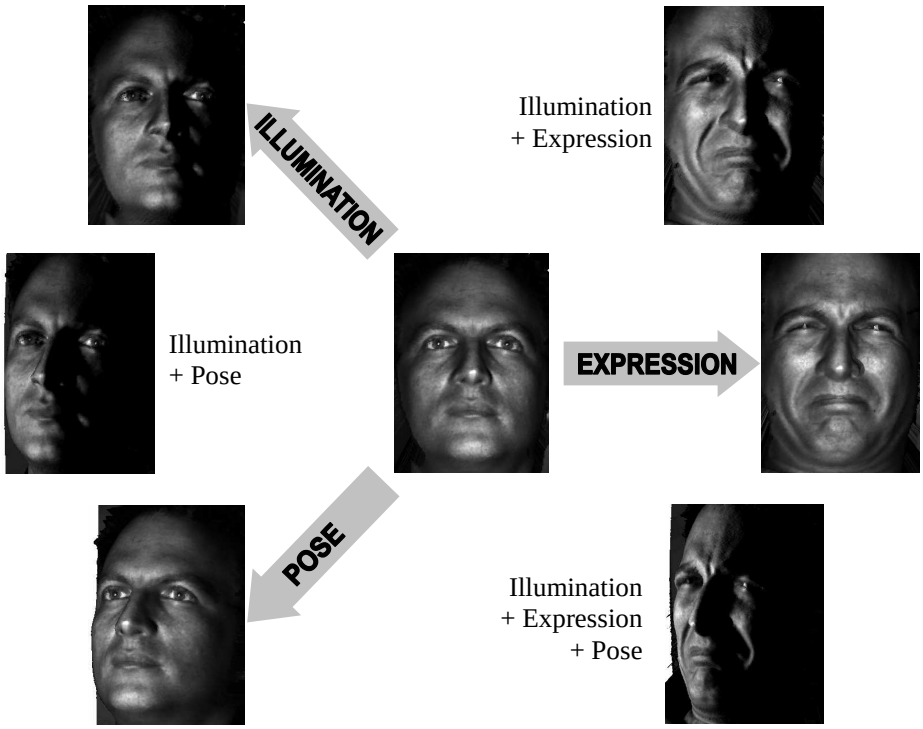


Figure 1: Illustration of some factors causing the variability of the facial image.

Broadly speaking, there are two basic alternatives in approaching this problem. One is to find features of the face that are not affected by the viewing conditions. Early face recognition algorithms [8, 37, 29] advocated this approach by finding a set of *fiducial points* (eyes, nose, mouth, etc.) and comparing their geometric relations (angles, lengths, and ratios). Unfortunately, there are only few such features that can be reliably extracted from a 2D facial image and would be insensitive to illumination, pose and expression variations [21].

The second alternative is to *generate* synthetic images of the face under 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 [36], or estimate a rough shape of the facial surface from a set of observations (e.g. using photometric stereo [27, 28]) in order

to synthesize new facial images and then apply standard face recognition methods like eigenfaces [49, 55] to the synthetic images.

Figure 2 shows a simple visual experiment that demonstrates the generative approach. We created synthetic faces of Osama Bin Laden (first row, right) and George Bush (second row, right) in different poses by mapping respective textures onto the facial surface of another subject (left). 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. 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 "lives" in a three-dimensional space, and using only its 2D projection can be misleading. Practically, one has the ability to draw any face on his own, so that he could essentially appear like any other person and deceive any 2D face recognition method.

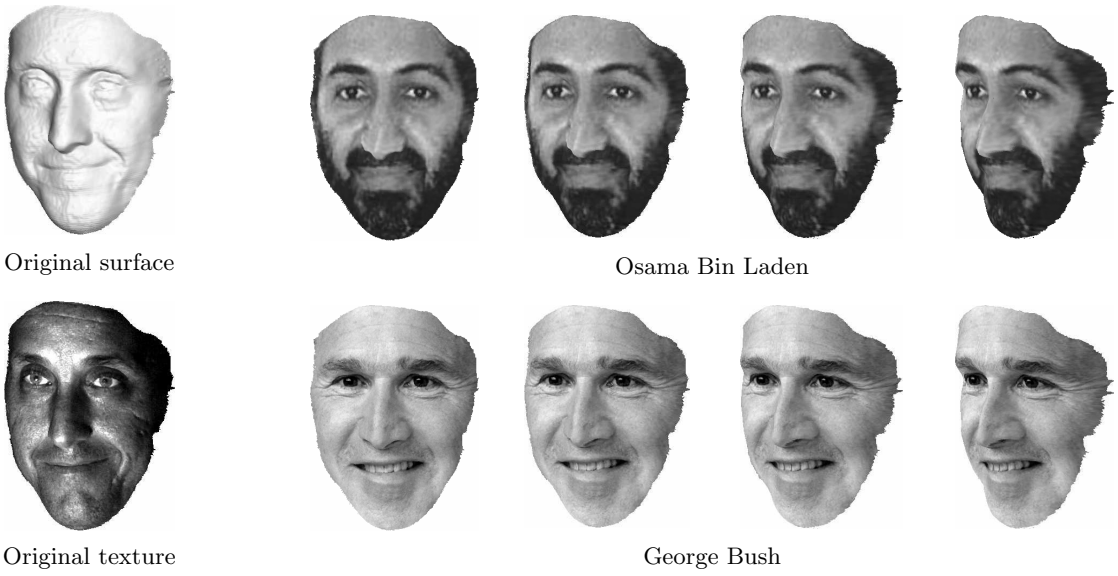


Figure 2: Simple texture mapping on the same facial surface can completely change the appearance of the 2D facial image and make the same face look like George Bush or Osama Bin Laden. This illustrates the fact that being able to change the texture of the face, e.g. by using make up, one can completely change his or her appearance as captured by a 2D image and disguise to another person. The 3D geometry of the face is the same in this example, and is more difficult to disguise.

0.1.2 A new dimension to 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, as 2D methods attempt to do. Three-dimensional facial geometry represents the internal anatomical structure of the face rather than its external appearance influenced by environmental factors. As a result, unlike the 2D facial image, 3D facial surface is insensitive to illumination, head pose [10], and cosmetics [41].

The main problem in 3D face recognition is how to find similarity between 3D facial surfaces. Earliest works on 3D face recognition did not use the whole facial surface, but a few profiles extracted from the 3D data [19, 46, 7, 32]. Attempts were made to extend conventional dimensionality reduction techniques (e.g. PCA) to range images or combination of intensity and range images [2, 35, 41, 20, 54]. Tsalakanidou et al. applied the hidden Markov model to depth and color images of the face [53]. Many academic (e.g. [42, 1]), as well as some commercial 3D face recognition algorithms treat faces as *rigid* surfaces by employing variants of rigid surface matching algorithms.

The intrinsic flaw of these approaches is their difficulty in handling deformations of the facial

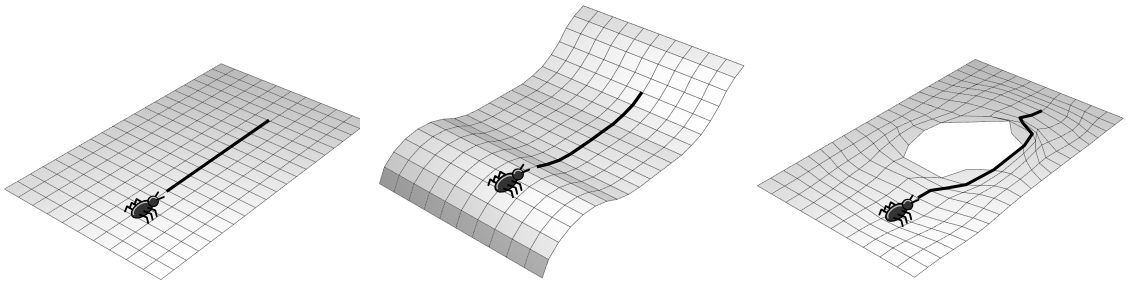
surface as the result of expressions. To date, only little research has been focused on trying to make face recognition deal with facial expressions. The majority of the papers, starting from the earliest publications on face recognition [8, 37, 29] and ending with the most recent results, address mainly the external factors like illumination, head pose, etc. [28] Moreover, though many authors mention the problem of facial expressions [31], a decent evaluation of currently available algorithms on a database of faces containing sufficiently large expression variability has never been done before [10].

In [13, 16, 14] we introduced an expression-invariant three-dimensional face recognition algorithm, on which the 3DFACE recognition system built at the Department of Computer Science, Technion, is based. Our approach uses a geometric model of facial expressions, which allowed us to build a representation of the face insensitive to expressions. This enabled us to successfully handle even extreme facial expressions.

0.2 Isometric model of facial expressions

In order to treat faces as deformable, non-rigid surfaces, we use the Riemannian geometry framework. We model faces as two-dimensional smooth connected compact Riemannian surfaces (manifolds). Broadly speaking, a Riemannian surface $\mathcal{S}$ can be described by a coordinate mapping $\mathcal{S} = \mathbf{x} : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$ from a domain U on a plane to the 3D Euclidean space and the *metric tensor* g, which is an intrinsic characteristic of the surface that allows us to measure local distances on $\mathcal{S}$ independently of the coordinates [40].

The deformations of the facial surface as the result of expressions can be expressed as a diffeomorphism $f : (\mathcal{S}, g) \rightarrow (\mathcal{S}', g')$ on the surface $\mathcal{S}$. Our observations show that in most parts of the human face the facial skin does not stretch significantly, therefore, we can model facial expressions as *isometries*. An experiment validating the isometric model is described in [18]. We showed that the isometric model faithfully describes most natural facial expression, and that such model is better than the rigid one. Isometric transformations preserve the *intrinsic geometry* of the surface. That is, the metric and consequently, the *geodesic distances* (the shortest paths between any two points on $\mathcal{S}$) remain invariant. To illustrate the idea of isometry, imagine a two-dimensional creature that lives on the surface (Figure 3). An isometry is a transformation that bends the surface such that the creature does not "feel" it. Faces in presence of facial expressions are thereby modelled as (approximately) isometric surfaces, that is, surfaces that can be obtained from some initial facial surface ("neutral expression") by means of an isometry. From the point of view of Riemannian geometry, such surfaces are indistinguishable as they have identical intrinsic geometry.



Original surface | Isometric transformation | Non-isometric transformation

Figure 3: Illustration of isometric and non-isometric transformations of a surface. Isometries do not change the intrinsic geometry of the surface, such that an imaginable creature living on the surface does not feel the transformation.

Isometry also tacitly implies that the *topology* of the facial surface is preserved. For example, expressions are not allowed to introduce "holes" in the facial surface (Figure 3, right). This assumption is valid for most regions of the face, yet, the *mouth* cannot be treated by the isometric model. Opening the mouth, for example, changes the topology of the facial surface by virtually creating a "hole." As a consequence, the isometric model is valid for facial expressions with either always open or always closed mouth.

This flaw of the isometric model can be dealt with by enforcing a fixed topology on the facial surface. For example, assuming that the mouth is always closed and thereby "gluing" the lips when the mouth is open; or, alternatively, assuming the mouth to be always open, and "disconnecting" the lips by introducing a cut in the surface when the mouth is closed. This new model, which we refer to as the *topologically-constrained isometric model*, is applicable to all facial expressions, including those with both open and closed mouth. The problem of open mouth in 3D face recognition is addressed in [18]. Here we assume that the mouth is always closed and thus limit our discussion to the isometric model.

0.3 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 another subject. Under the isometric model assumption, the problem is reduced to finding similarity between isometric surfaces.

Figure 4 illustrates the problem of isometric surface matching. The first row shows three isometries of the human hand (assume that the fingers do not touch each other, such that the topology is preserved), which, with a bit of imagination, look like a grenade, dog and cobra (second row). In other words, from the point of view of their *extrinsic* geometry, isometric surfaces can look different, while being just instances of the same surface. Deformations of the facial surface due to facial expressions are though not so extreme like those of a human hand, yet sufficiently significant to make the uncertainty region around the facial surface large enough that many other faces can fit within (see Figure 5).

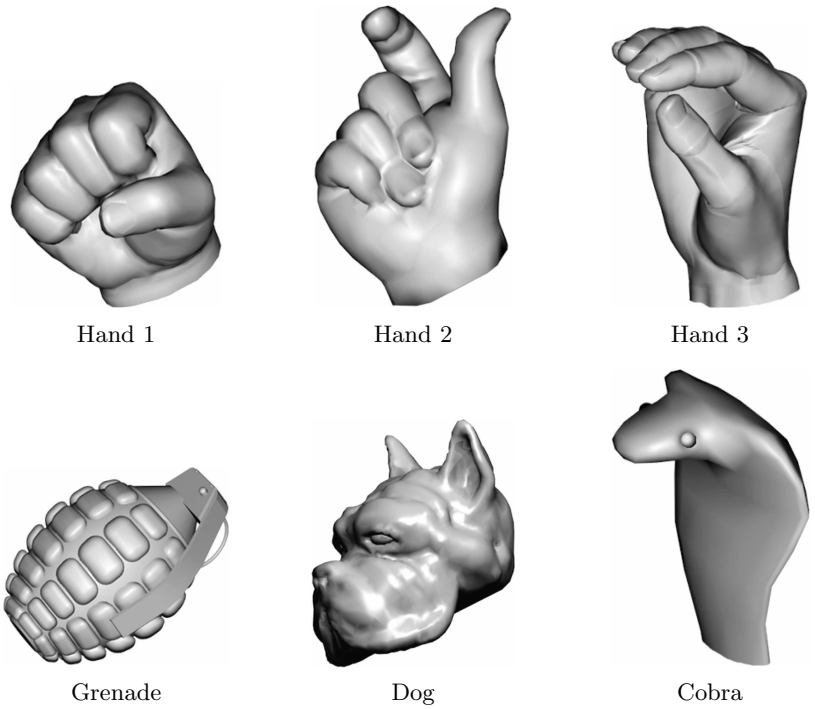
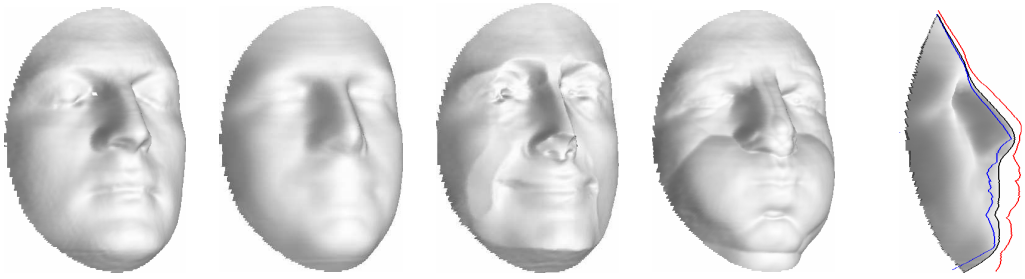


Figure 4: Illustration of the isometric surface matching problem. First row: isometries of a hand. Second row: different objects that resemble the hands if treated in a rigid way.

Lets assume that we have two instances of the same face differing one from another by a facial expression, and let $\mathcal{S}$ and $\mathcal{Q}$ denote the corresponding facial surfaces. According to our isometric model, there exists an isometric transformation $f(\mathcal{S}) = \mathcal{Q}$ that maps any point $\mathbf{x}$ on the surface $\mathcal{S}$ to a point $\mathbf{y}$ on the surface $\mathcal{Q}$. Since f is isometric, the geodesic distances are preserved, i.e. $d_{\mathcal{S}}(\mathbf{x}_1, \mathbf{x}_2) = d_{\mathcal{Q}}(\mathbf{y}_1, \mathbf{y}_2)$.

Theoretically, the geodesic distances give a unique expression-invariant representation of the face.



Neutral Average Min. envelope Max. envelope Profile

Figure 5: Variation of the facial surface due to facial expressions (left to right): neutral expression of subject Eyal, average facial expression, minimum and maximum envelopes, profile view showing the minimum and the maximum envelope.

However, since in practice the surfaces $\mathcal{S}$ and $\mathcal{Q}$ are represented by a discrete set of samples $\mathbf{x}_1, ...\mathbf{x}_{N_x}$ and $\mathbf{y}_1, ...\mathbf{y}_{N_y}$, respectively, there is neither guarantee that the surface is sampled at the same points, nor that the number of points in two surfaces is necessarily the same ($N_x \neq N_y$ in general). Moreover, even if the samples are the same, they can be ordered arbitrarily, and thus the matrix $\mathbf{D} = (d_{ij}) = (d(\mathbf{x}_i, \mathbf{x}_j))$ is invariant up to some permutation of the rows and columns. Therefore, though the matrix $\mathbf{D}$ can be considered as an invariant, making use of it has little practicality. Nevertheless, there have been very recent efforts to establish some theory about the properties of such matrices [43, 44].

The alternative proposed in [25, 26] is to avoid dealing explicitly with the matrix of geodesic distances and find an representation to the original Riemannian surface as a submanifold of some convenient space, with an effort to preserve (at least approximately) the intrinsic geometry of the surface. Such an approximation is called *isometric embedding*. Typically, a low-dimensional Euclidean space is used as the embedding space; the embedding in this case is called *flat*.

In our discrete setting, flat embedding is a mapping

$$\varphi : (\{\mathbf{x}_1, ..., \mathbf{x}_N\} \subset \mathcal{S}, d) \rightarrow (\{\mathbf{x}'_1, ...\mathbf{x}'_N\} \subset \mathbb{R}^m, d'), \tag{1}$$

that maps N samples $\mathbf{x}_1, ...\mathbf{x}_N$ of the surface $\mathcal{S}$ into a set of points $\mathbf{x}'_1, ...\mathbf{x}'_N$ in a m-dimensional Euclidean space, such that the resulting Euclidean distances $d'_{ij} = \|\mathbf{x}'_i - \mathbf{x}'_j\|_2$ approximate the original geodesic distances d_{ij} in an optimal way (here the matrices $\mathbf{D}$ and $\mathbf{D}'(\mathbf{X}')$ denote the mutual geodesic and Euclidean distances, respectively). The resulting set of points $\mathbf{x}'_1, ...\mathbf{x}'_N$ in the Euclidean space is called the *canonical form* of the facial surface [25, 26]. The canonical forms are defined up to a rotation, translation and reflection, and can be therefore treated by conventional algorithms used for rigid surface matching. Figure 6 shows an example of a deformable surface (human hand) undergoing isometric transformations, and the corresponding canonical forms of the hand.

We would like to find such a mapping φ that deforms the geodesic distances the least. Embedding error can be measured as a discrepancy $s(\mathbf{X}'; \mathbf{D})$ between the original geodesic and the resulting Euclidean distances. We use $\mathbf{X}' = (\mathbf{x}'_1, ..., \mathbf{x}'_N)$ to denote an $m \times N$ matrix representing the coordinates of the points in the embedding space.

0.3.1 Multidimensional scaling

Finding the best approximate flat embedding is possible by minimization of $s(\mathbf{X}'; \mathbf{D})$ with respect to $\mathbf{X}'$. A family of algorithms used to carry out such an approximate flat embedding is known as *multidimensional scaling* (MDS) [9]. These algorithms differ in the choice of the embedding error criterion and the numerical method used for its minimization.

One of the most straightforward possibilities is to have the metric distortion defined as a sum of squared differences

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

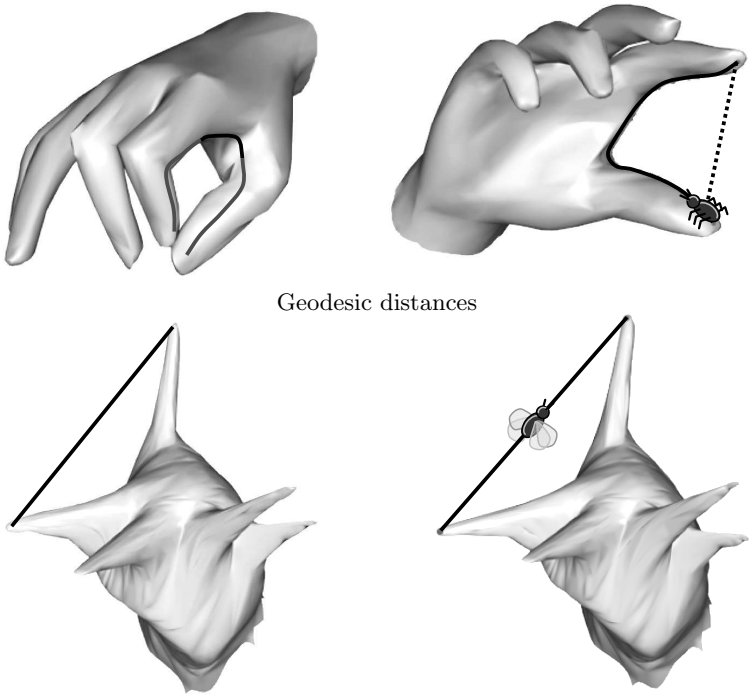


Figure 6: Illustration of the embedding problem and the canonical forms. First row: a Riemannian surface (hand) undergoing isometric transformations. Solid line shows the geodesic distance between two points on the surface, dotted line is the corresponding Euclidean distance. Second row: the hand surfaces embedded in a three-dimensional Euclidean space. The geodesic distances Euclidean ones.

and the MDS problem is posed as a least-squares problem (LS-MDS). Such embedding error criterion is called the *raw stress*. Since the stress is a non-convex function in $\mathbf{X}'$, standard convex optimization techniques do not guarantee convergence to the global minimum. Different techniques can be employed to prevent convergence to small local minima. An example is the iterative convex majorization algorithm (SMACOF, standing for Scaling by Majorization of a Convex Function) [9, 22], which is less sensitive to local minima.

An alternative to LS MDS is an algebraic embedding method due to Torgerson and Gower [52, 33] based on theoretical results of Eckman, Young, and Householder [23, 56], known as the *classical scaling*. Classical scaling works with the squared geodesic distances, which can be expressed as the Hadamard (coordinate-wise) product $\mathbf{\Delta} = \mathbf{D} \circ \mathbf{D}$. The matrix $\mathbf{\Delta}$ is first double-centered

$$\mathbf{B} = -\frac{1}{2}\mathbf{J}\mathbf{\Delta}\mathbf{J} \tag{3}$$

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

$$\mathbf{X}' = \mathbf{V}_+\mathbf{\Lambda}_+. \tag{4}$$

In practice, since we are usually interested in embedding into $\mathbb{R}^3$ or $\mathbb{R}^2$, no full eigendecomposition of $\mathbf{B}$ is needed – it is enough to find only the first three or even two eigenvectors. Arnoldi [3], Lanzcos or block-Lanzcos [30, 5] algorithms can be used to performs this task efficiently.[1]

[1]We thank Gene Golub and Michael Saunders (Stanford University) for their valuable comments on efficient numerical implementations of such eigendecomposition algorithms.

0.3.2 Canonical forms of facial surfaces

When embedding is performed into a space of dimension $m = 3$, the canonical form can be plotted as a surface. Figure 7 depicts canonical forms of a person's face with different facial expressions. It demonstrates that although the facial surface changes are substantial, the changes between the corresponding canonical forms are insignificant. Embedding into $\mathbb{R}^2$ is a special case – in this case, the codimension of the canonical form in the embedding space is zero. Such an embedding can be thought of as an intrinsic parametrization of the facial surface, which leads to a "warping" of the facial texture. This serves as a way of performing geometry-based registration of 2D facial images [12]. Flat embedding into $\mathbb{R}^2$ was previously used for cortical surface matching [48] in brain analysis, and adopted to texture mapping [57] in computer graphics.

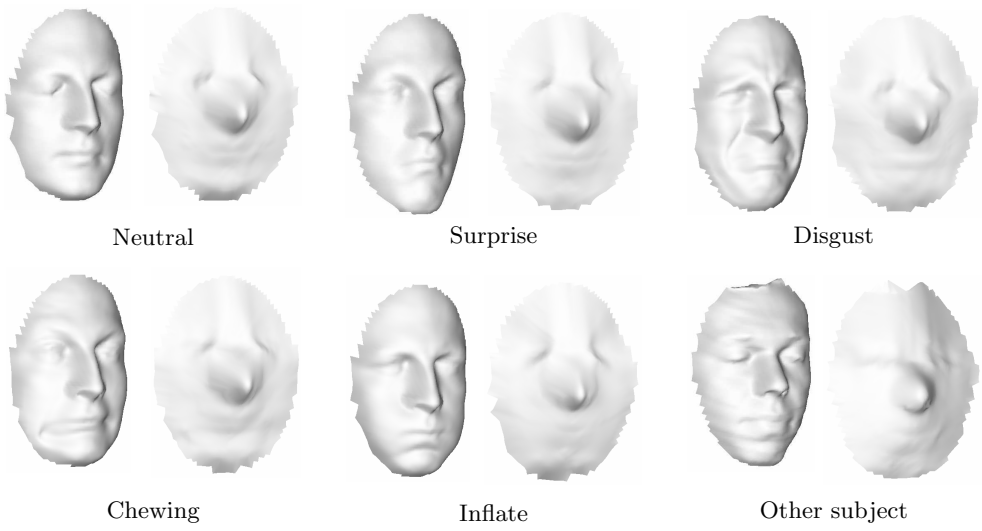


Figure 7: Examples of canonical forms of faces with strong facial expressions. For comparison, canonical form of a different subject is shown (second row, right).

0.4 The 3DFACE system

We designed a prototype of a fully-automatic 3D face recognition system based on the expression-invariant representation of facial surfaces. The 3DFACE system is shown in Figure 8. It can work both in one-to-one and one-to-many recognition modes. In one-to-one (verification) mode, the user swipes a magnetic card (5 in Figure 8) bearing his or her personal identification information. The system compares the subject's identity with the claimed one. In one-to-many (recognition) mode, the subject's identity is unknown *a priori* and it is searched for in a database of faces. In the current prototype, no automatic face detection and tracking is implemented. The monitor (3 in Figure 8) is used as a "virtual mirror" allowing the user to align himself relative to the camera.

Three-dimensional structure of the face is acquired by an active stereo time-multiplexed structured light range camera [11]. World coordinates of each point are computed by triangulation, based on the knowledge of the point location in the 2D coordinate system of the camera and the 1D coordinate system of the projector. The latter is inferred from a code which is projected in a form of light stripes onto the face of the subject using a DLP projector (1 in Figure 8). We use a 10-bit binary Gray code, which allows to obtain about 0.5 mm depth resolution with scan duration less than 200 msec. The scanner output is a cloud of 640×480 points.

The raw data of the scanner is first down-sampled to the resolution of 320×240, then undergoes initial cropping which roughly separates the facial contour from the background. In the topologically-constrained case, the lips are also cut off, hole filling (which removes acquisition spike-like artifacts) and selective smoothing by a Beltrami-like geometric filter [50, 38, 14].

The smoothed surface is resized again about 3 times in each axis, and then the facial contour is extracted by using the *geodesic mask*. The key idea is locating invariant "source" points on the face

Figure 8: The 3DFACE prototype system and its main components: DLP projector (1), digital camera (2), monitor (3), magnetic card reader (4), mounting (5) .

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. In the 3DFACE system, two fiducial points are used for the geodesic mask computation: the tip of the nose and the nose apex (the topmost point of the nose bone). These fiducial points are located automatically using a curvature-based 3D feature detector, similar to [45] (see details in [15]). The same feature detector is used to locate the left and the right eye.

The geodesic mask allows us to crop the facial surface in a geometrically consistent manner, insensitively to facial expressions. After cropping, the resulting surface contains between $2500-3000$ points. Computation of the geodesic distances is performed using the Fast Marching algorithm [39].

As the final stage, "canonization" of the facial surface is performed by computing the mutual geodesic distances between all the surface points and then applying MDS. We embed facial surfaces into $\mathbb{R}^3$ using LS MDS.

The 3D acquisition lasts less than 200 msec. The overall end-to-end processing time (including acquisition) is about 5 sec.

Since embedding is defined up to an Euclidean and reflection transformation, the canonical surface must 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 surface to resolve the translation ambiguity (here

$$\mu_{pqr} = \sum_{i=1}^{N} (x_i^1)^p (x_i^2)^q (x_i^3)^r \tag{5}$$

denotes the pqr-th moment); then, the mixed second-order moments $\mu_{110}, \mu_{011}, \mu_{101}$ are set to zero to resolve the rotation ambiguity. Finally, using the coordinate relations of three fiducial points on the face (two eyes and the nose tip), the reflection ambiguity is resolved.

0.4.1 Surface matching

The final stage of the face recognition algorithm is surface matching. Since the flattening compensates for the non-rigid isometries of the surface, standard rigid matching (see e.g. [34]) can be

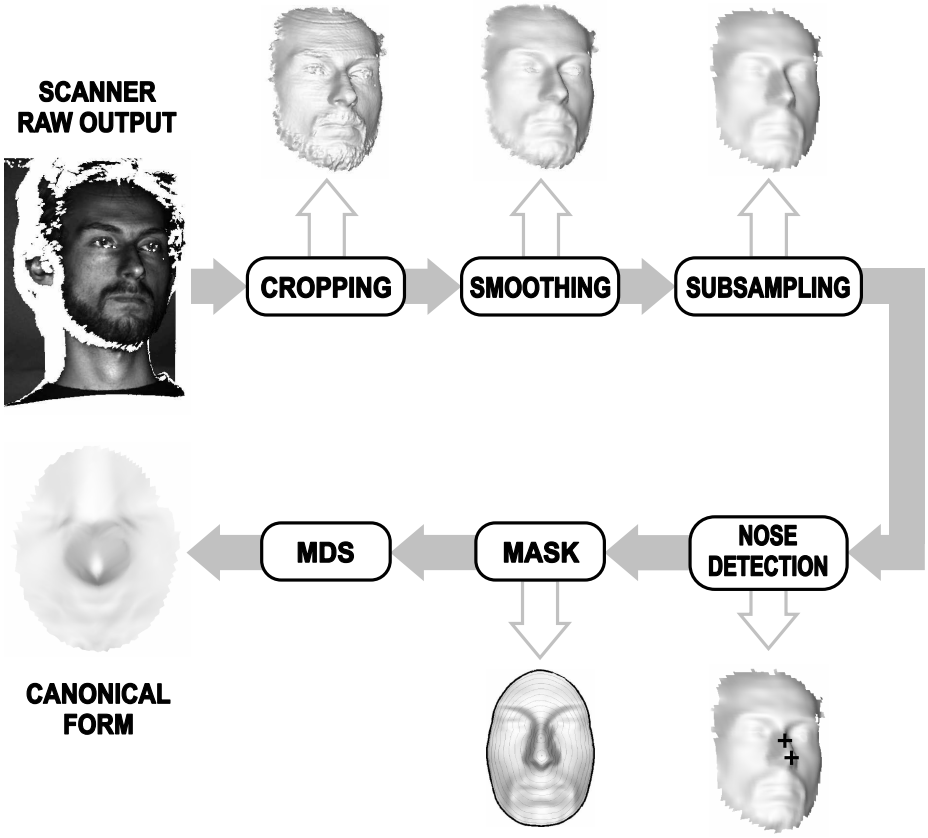


Figure 9: Scheme of preprocessing and canonization of the facial surface used in the 3DFACE system.

used for comparing the canonical surfaces. The standard choice in surface matching is the iterative closest point (ICP) method and its variants [6], 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 [51]. The main idea is to represent the surface by its moments μ_{pqr} up to some degree P, and compare the moments as vectors in an Euclidean space. Given two facial surface $\mathcal{S}$ and $\mathcal{Q}$ with the corresponding canonical forms $\mathbf{X}'$ and $\mathbf{Y}'$ we can define the distance between two faces as

$$d_{can}(\mathcal{S},\mathcal{Q}) = \sum_{p+q+r\leq P} (\mu_{pqr}^{\mathbf{X}'} - \mu_{pqr}^{\mathbf{Y}'})^2. \tag{6}$$

In [13, 12] we proposed to treat canonical forms as images. After alignment, both the canonical surface and the flattened albedo are interpolated on a Cartesian grid, producing two images. These images can be compared using standard techniques, e.g. applying eigendecomposition like in eigenfaces or eigenpictures. The obtained representation was called in [13] *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 as an additional classifier. Here, however, we focus on the 3D geometry, and in the following experiments use only the surface geometry ignoring the texture.

0.5 Results

In this section, we present experimental results evaluating the 3DFACE method. First, we perform a set of experiments, the goal of which is to test how well canonical forms can handle strong facial expressions. The data sets used in these experiments contain facial expressions with closed mouth only. An evaluation of our algorithm on a data set containing expressions with both open and closed

Table 1: Description of the Experiment I data. Asterisk denotes artificial subjects. Double asterisk denotes identical twins.

Subject	Color	Neutral	Weak	Medium	Strong	Total
Michael**	red	6	5	6	-	17
Alex**	blue	3	1	3	1	8
Eyal	green	4	1	7	9	21
Noam	yellow	3	-	-	7	10
Moran	magenta	4	-	4	10	18
Ian	orange	5	-	16	7	28
Ori	cyan	8	-	11	10	29
Eric	d. green	5	3	-	3	11
Susy	d. magenta	6	-	9	8	23
David	l. blue	5	2	6	5	18
Eve*	black	6	-	-	-	6
Benito*	grey	7	-	-	-	7
Liu*	l. grey	8	-	-	-	8

mouth can be found in [18]. Then, we provide a benchmark of 2D and 3D face recognition algorithms and compare them to our approach.

0.5.1 Sensitivity to facial expressions

In the first experiment, we studied the sensitivity of canonical forms to facial expressions. We used a data set containing 10 human and 3 artificial subjects. Subjects `Alex` (blue) and `Michael` (red) are identical twins. Each face in the data set appeared with a number of instances ($6-29$ instance per subject, a total of 204 instances) in a variety of facial expressions. The database is summarized in Table 1.

All expressions were conducted with a closed mouth and were classified into 10 types (neutral expression + 9 expressions) and into 3 strengths (weak, medium, strong). Neutral expressions are natural postures of the face, while strong expressions are exaggerated postures rarely encountered in everyday life (Figure 11, second row). The group of expressions including `smile`, `sadness`, `anger`, `surprise` and `disgust` are basic emotions according to Eckman [24]; the group `thinking`, `stress`, `grin` and `chewing` tries to imitate facial appearance that can occur in a natural environment; finally, expressions `inflate` and `deflate` result in the most significant deformation of the facial geometry, though rarely encountered (Figure 11, first row). Small head rotations (up to about 10 degrees) were allowed. Since the data was acquired on a course of several months, variations in illumination conditions, facial hair, etc. also present.

For reference, our approach was compared to rigid matching of facial surfaces. In both cases, the metric d_{can} based on moments of degree up to $P = 5$ (i.e. vectors of dimensionality 52), according to (5), was used.

The dissimilarities (distances) between different faces obtained by the metric d_{can} allow us to cluster together faces that belong to the same subjects. As a quantitative measure of the separation quality, we use the ratio of the maximum inter-cluster to minimum intra-cluster dissimilarity

$$\varsigma_k = \frac{\max_{i,j\in\mathcal{C}_k} \eta_{ij}}{\min_{i\notin\mathcal{C}_k, j\in\mathcal{C}_k} \eta_{ij}}, \tag{7}$$

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

$$\sigma_k = \sqrt{\frac{\frac{2}{|\mathcal{C}_k|^2-|\mathcal{C}_k|}\sum_{i,j\in\mathcal{C}_k,\ i>j} \eta_{ij}^2}{\frac{1}{|\mathcal{C}_k|(|\mathcal{C}|-|\mathcal{C}_k|)}\sum_{i\notin\mathcal{C}_k, j\in\mathcal{C}_k} \eta_{ij}^2}}, \tag{8}$$

($\mathcal{C}_k$ denotes indexes of k-th subject's faces, and η_{ij} denotes dissimilarities between faces i and j). This criterion measures how each cluster is tight and far from other clusters. Ideally, ς_k and σ_k should tend to zero.

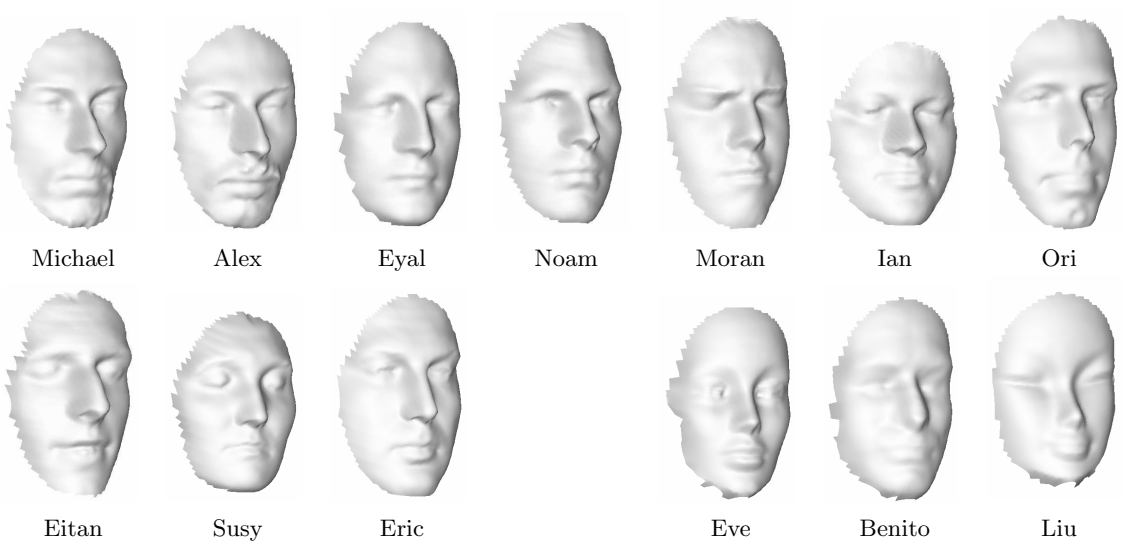


Figure 10: The subjects used in Experiment I (shown with neutral expressions). Second row right: three artificial subjects.

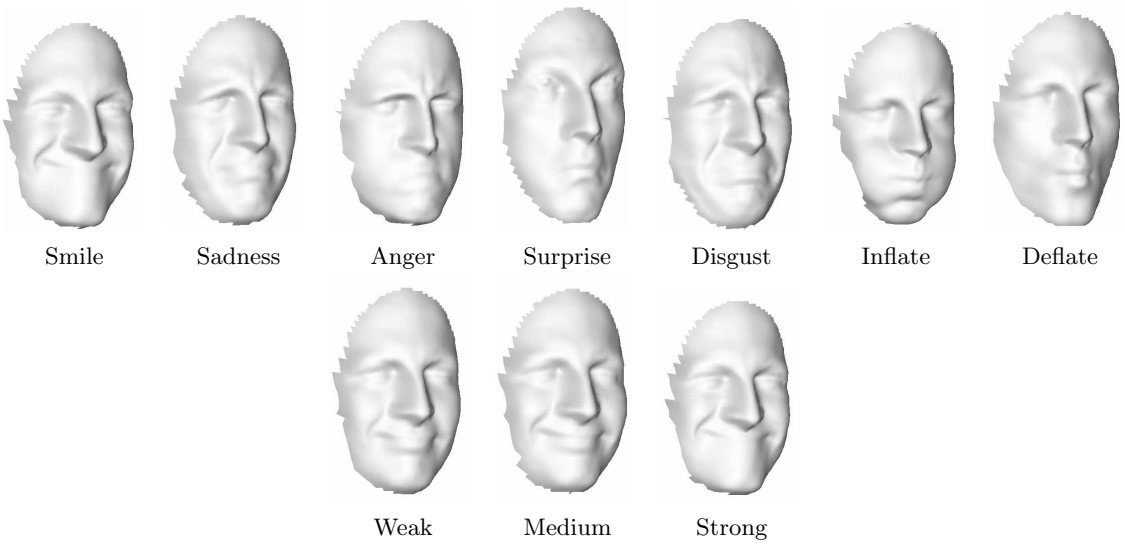


Figure 11: First row: seven representative facial expressions of subject Eyal in Experiment I. Second row: three degrees of the smile expression of the same subject.

Figure 12 depicts a three-dimensional visualization of the dissimilarities between faces, obtained by applying classical scaling to the dissimilarity matrix of faces. Each face on this plot is represented by a point; faces of different subjects are marked with different colors. The first row depicts the dissimilarities between faces with only neutral expressions. Faces of different subjects form tight clusters and are easily distinguishable. The advantage of canonical forms is not so apparent in this case. However, the picture changes drastically when we allow for facial expressions (Figure 12, 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.

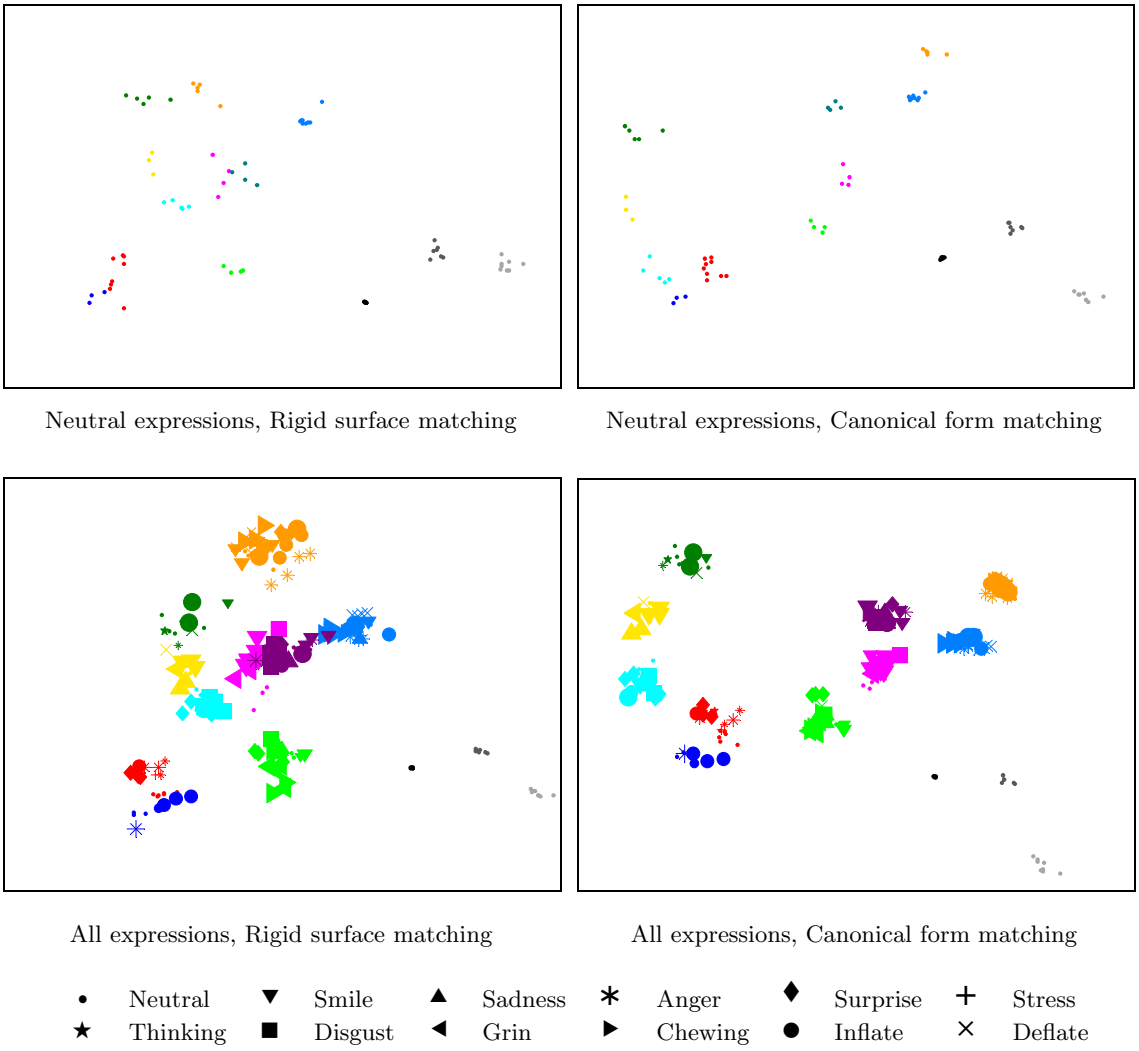


Figure 12: Low-dimensional visualization of dissimilarities between faces in Experiment I using original surface (left) and canonical form (right) matching. First row: neutral expressions only. Second row: all expressions. Colors represent different subject. Symbols represent different facial expressions. Symbol size represents the strength of the facial expression.

Figure 13 shows the separation quality criteria (ς_k and σ_k) for rigid and canonical surface matching. When only neutral expressions are used, canonical form matching outperform rigid surface matching on most subject in terms of ς_k and σ_k (by up to 68% in terms of ς_k and by up to 64% in terms of σ_k; slightly inferior performance in terms of ς_k is seen on artificial subject `Eve` and human subjects `Eyal`, `Noam` and `David`). 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 original surface matching by up to 304% in terms of ς_k and by up to 358% in terms of σ_k.

0.5.2 Comparison of algorithms

The goal of the second experiment is performing a benchmark of our method and comparing it to other face recognition algorithms. Faces from the probe database (30 different probe subjects in a variety of facial expression, total of 220 faces) were compared to a set of 65 gallery templates

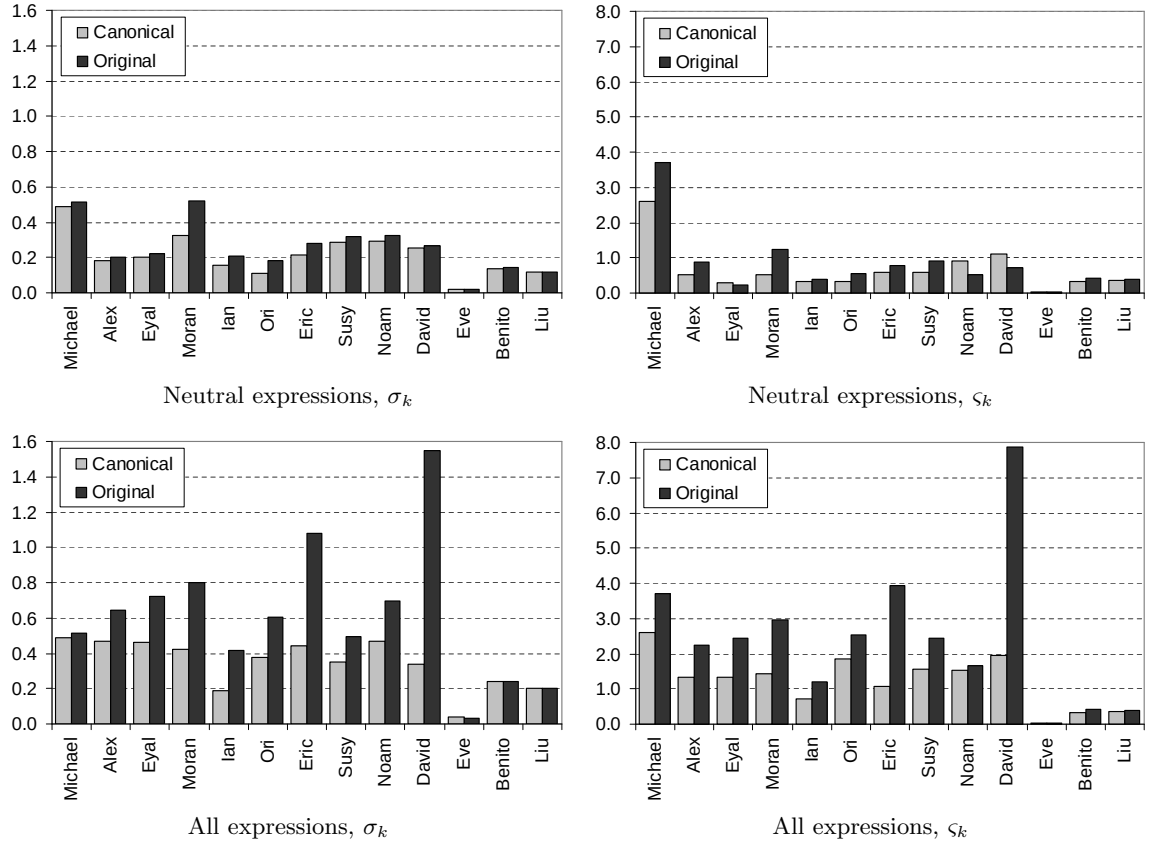


Figure 13: Separation quality criteria (ς_k and σ_k) using original (dark gray) and canonical (light gray) surface matching. The smaller ς_k and σ_k, the better is the separation quality.

(typically, two or three templates per subject was used). Only neutral expressions were used in the gallery.

Three algorithms were tested: canonical form matching, facial surface matching and 2D image-based eigenfaces. Eigenfaces 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 variability [28]).

Figure 14 (left) shows the cumulative match characteristic (CMC) curves of three algorithms compared in this experiment on full database with all facial expressions. Our approach results in rank 1 *zero* recognition error. Figure 14 (right) shows the receiver operation characteristic (ROC) curves. Our algorithm significantly outperforms both the rigid facial surface matching and the eigenfaces algorithm.

Figure 15 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 templates using eigenfaces, facial surface matching and canonical form matching.

First row in Figure 15 shows results typical for the described algorithms. Eigenfaces, being image-based, finds the subject `Ori 188` more similar to the reference subject `Moran 129` since they have the same facial expression (strong smile), though these are different subjects. Facial surface matching is confused by 3D features (outstanding inflated cheeks) that appear on the face of subject `Moran 129` due to the facial expression. These features are similar to the natural facial features (fat cheeks) of subject `Susy 276`. Finally, canonical surface matching finds a correct match (`Moran 114`), since flattening compensates for the distortion of the face of subject `Moran 129` due to smile.

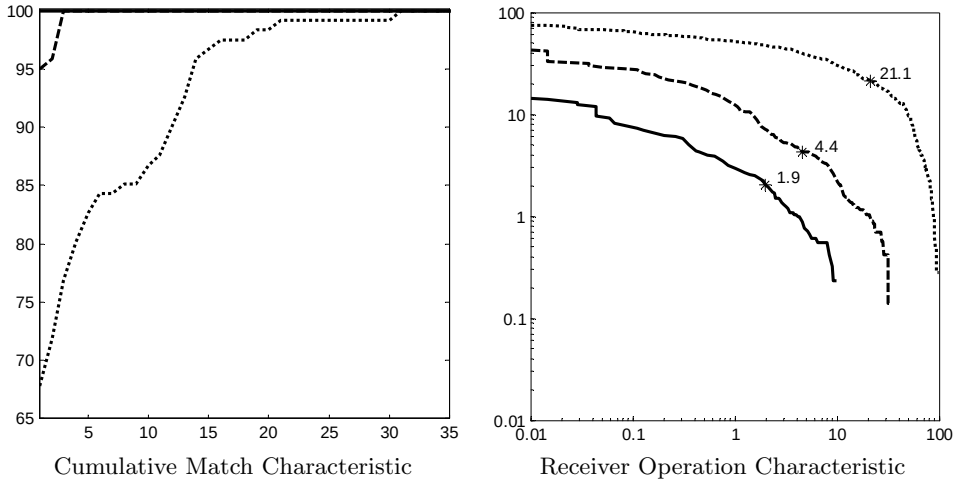


Figure 14: CMC (left) and ROC (right) curves of face recognition based on surface matching (dashed), canonical surface matching (solid) and eigenfaces (dotted). Obtained on database with all expressions. Star denotes equal error rate.

Second row in Figure 15 shows an example of identical twins recognition – the most challenging task for a face recognition algorithm. The eigenfaces algorithm resulted in 29.41% incorrect matches when enrolling `Michael` and 25% when enrolling `Alex`. Facial surface matching resulted in 17.64% and 0% wrong matches, respectively. Canonical form matching resulted in 0% recognition error for both twins.

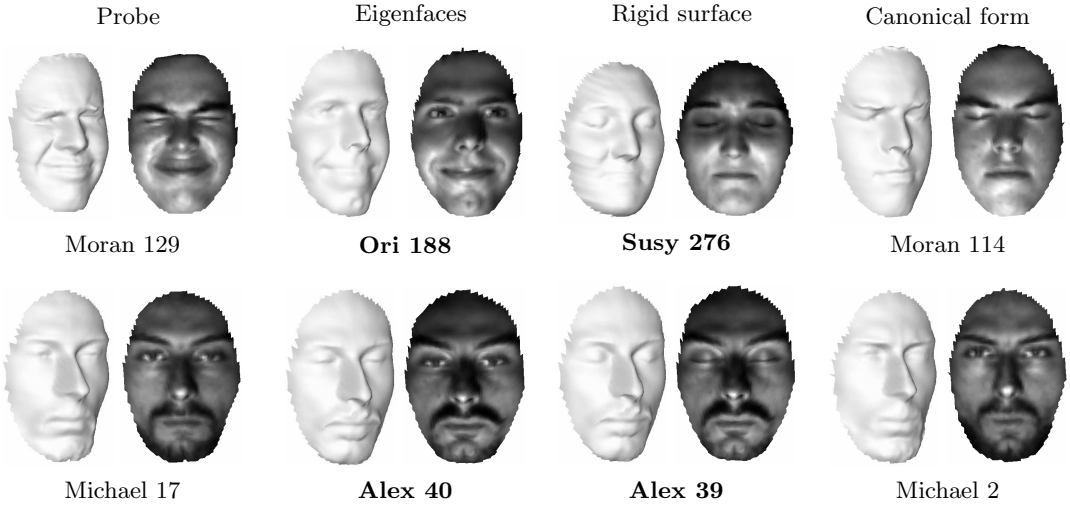


Figure 15: Example of recognition using different algorithms. First column shows the probe subject; second through fourth columns depict the closest (rank 1) matches found by the canonical form matching, facial surface matching and eigenfaces, respectively. Note that only the match using canonical form matching is correct. Numbers represent the subject's index in the database. Wrong matches are emphasized.

Comparing the canonical forms averaged on about 10 instances with different expressions for each of the twins, we found out a slight difference in the 3D geometry of the nose, which makes this distinction possible (Figure 16). Apparently, the difference is very subtle and is not distinct if using rigid surface matching, as the nose deforms quite significantly due to facial expressions.

Michael Alex Difference map

Figure 16: A pair of identical twins participating in the experiment (Alex and Michael), and the difference (blue shades represent the difference absolute value) computed in the canonical form domain and mapped onto the facial surface.

0.6 Conclusions

The geometric framework for 3D face recognition presented here provides a solution to a major problem in face recognition: sensitivity to facial expressions. Being an internal characteristic of the human face, facial expressions are harder to deal with compared to external factors like pose or lighting. This problem is especially acute when face recognition is performed in a natural environment.

Thinking of expressions as of approximated isometric transformations of a deformable facial surface allows to construct an expression-invariant representation of the face. Our approach outperforms other 3D recognition methods that treat the face as a rigid surface. The 3DFACE face recognition system prototype implementing our algorithm demonstrates high recognition accuracy and has the capability to distinguish between identical twins. It is now being evaluated for various industrial applications.

0.7 Acknowledgements

We are grateful to Gene Golub and Michael Saunders (Stanford University) for valuable notes on efficient implementation of eigendecomposition algorithms, to David Donoho (Stanford University) for pointing us to Eckman's publications on facial expressions, and to everyone who contributed their faces to our database.

This research was supported by the Israel Science Foundation (ISF), Grant No. 738/04 and the Bar Nir Bergreen Software Technology Center of Excellence (STL).

Bibliography

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

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

[3] W. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math. **9** (1951), 17–29.

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

[5] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, *Templates for the solution of algebraic eigenvalue problems: A practical guide*, third ed., SIAM, Philadelphia, 2000, Online: `http://www.cs.utk.edu/~dongarra/etemplates/index.html`.

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

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

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

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

[10] K. W. Bowyer, K. Chang, and P. Flynn, *A survey of 3D and multi-modal 3D+2D face recognition*, Dept. of computer science and electrical engineering technical report, University of Notre Dame, January 2004.

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

[12] A. M. Bronstein, M. M. Bronstein, E. Gordon, and R. Kimmel, *Fusion of 3D and 2D information in face recognition*, Proc. ICIP, 2004.

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

[14] ———, *Three-dimensional face recognition*, Tech. Report CIS-2004-04, Dept. of Computer Science, Technion, Israel, 2004.

[15] ———, *Three-dimensional face recognition*, IJCV (2005), to appear.

[16] A. M. Bronstein, M. M. Bronstein, R. Kimmel, and A. Spira, *Face recognition from facial surface metric*, Proc. ECCV, 2004.

[17] M. M. Bronstein and A. A. Bronstein, *Biometrics was no match for hair-rising tricks*, Nature **420** (2002), 739.

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

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

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

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

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

[23] C. Eckart and G. Young, *Approximation of one matrix by another of lower rank*, Psychometrika **1** (1936), 211–218.

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

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

[26] ———, *On bending invariant signatures for surfaces*, IEEE Trans. PAMI **25** (2003), no. 10, 1285–1295.

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

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

[29] A. Goldstein, L. Harmon, and A. Lesk, *Identification of human faces*, Proc. IEEE **59** (1971), no. 5, 748–760.

[30] G. H. Golub and C. F. van Loan, *Matrix computations*, third ed., The John Hopkins University Press, 1996.

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

[32] ———, *Face recognition from frontal and profile views*, Proc. Int'l Workshop on Face and Gesture Recognition, 1997, pp. 74–52.

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

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

[35] C. Hesher, A. Srivastava, and G. Erlebacher, *A novel technique for face recognition using range images*, Int'l Symp. Signal Processing and Its Applications, 2003.

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

[37] T. Kanade, *Picture processing by computer complex and recognition of human faces*, Technical report, Kyoto University, Dept. of Information Science, 1973.

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

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

[40] E. Kreyszig, *Differential geometry*, Dover Publications Inc., New York, 1991.

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

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

[43] F. Mémoli and G. Sapiro, *Comparing point clouds*, IMA preprint series 1978, University of Minnesota, Minneapolis, MN 55455, USA, April 2004.

[44] ———, *A theoretical and computational framework for isometry invariant recognition of point cloud data*, IMA preprint series 1980, University of Minnesota, Minneapolis, MN 55455, USA, June 2004.

[45] A. B. Moreno, A. Sanchez, J. Velez, and J. Diaz, *Face recognition using 3D surface-extracted descriptors*, Irish Machine Vision and Image Processing Conference, 2003.

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

[47] J. Ortega-Garcia, J. Bigun, D. Reynolds, and J. Gonzalez-Rodriguez, *Authentication gets personal with biometrics*, IEEE Signal Processing magazine **21** (2004), no. 2, 50–62.

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

[49] L. Sirovich and M. Kirby, *Low-dimensional procedure for the characterization of human faces*, JOSA A **2** (1987), 519–524.

[50] N. Sochen, R. Kimmel, and R. Malladi, *A general framework for low level vision*, IEEE Trans. Image Proc. **7** (1998), no. 3, 310–318.

[51] A. Tal, M. Elad, and S. Ar, *Content based retrieval of VRML objects - an iterative and interactive approach*, Eurographics Workshop in Multimedia, 2001.

[52] W. S. Torgerson, *Multidimensional scaling I - theory and methods*, Psychometrika **17** (1952), 401–419.

[53] F. Tsalakanidou, S. Malassiotis, and M. G. Strintzis, *Face localization and authentication using color and depth images*, IEEE Trans. Image Processing **14** (2005), no. 2.

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

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

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

[57] G. Zigelman, R. Kimmel, and N. Kiryati, *Texture mapping using surface flattening via multidimensional scaling*, IEEE Trans. Visualization and computer graphics **9** (2002), no. 2, 198–207.