Chapter 5

STORY OF CINDERELLA:
Biometrics and isometry-invariant distances

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

Abstract: In this chapter, we address the question of what are the facial measures one could use in order to distinguish between people. Our starting point is the fact that the expressions of our face can, in most cases, be modeled as isometries, which we validate empirically. Then, based on this observation, we introduce a technique that enables us to distinguish between people based on the intrinsic geometry of their faces. We provide empirical evidence that the proposed geometric measures are invariant to facial expressions and relate our findings to the broad context of biometric methods, ranging from modern face recognition technologies to fairy tales and biblical stories.

Key words: biometrics, isometry, face recognition, facial expression, multidimensional scaling, intrinsic geometry.

1. INTRODUCTION

Most of us are familiar with the story of the handsome prince who declares that he will marry the girl whose foot fits into the glass slipper she lost at his palace. The prince finally finds Cinderella by matching the slipper to her foot, and they live happily ever after. This is how Charles Perrault's story ends. One of the stepsisters of the German Cinderella, according to the Brothers Grimm version, succeeds in fitting her foot into the slipper by cutting off a toe, in an attempt to create a fake biometric signature. Cinderella's stepsister was not the first – it appears that identity frauds date back to biblical times. In Genesis, Jacob stole his father's blessing, the privilege of the elder, by pretending to be his firstborn brother Esau. By hand

scan Isaac wrongly verified his sons Esau's identity, since smooth-skinned Jacob wrapped kidskin around his hands to pose as his brother. Face recognition, another biometric technology, Little Red Riding Hood make the unavoidable conclusion that it was the wolf she was talking to rather than her grandmother.

With this example, we leave the fairy-tale world and enter into the realm of modern biometric technologies. In this chapter, we focus on the problem of three dimensional face recognition, though the approach we introduce is general and can be applied to any non-rigid surface comparison problems under reasonable assumptions. The main question we will try to answer is the facial measures we could use in order to distinguish between people.

2.   UBIQUITOUS ISOMETRIES

Recently, a team of French surgeons has reconstructed the face of a woman by transplanting donor tissues. This remarkable operation raised controversial questions regarding the identity of the patient: will she recover the lost identity or look like the donor? Apparently, the lady's face has preserved its original features: though the skin tone may have changed, the geometry of the patient's face remained (at least partialluy) more or less intact. The reason is that the rigid structure of the scull was unaltered, which preserved the form of the overlaying tissues. Obviously, the geometry of the face reflects important information; uniquely describing our identity.

At this point, the term "geometry" requires a more accurate definition. Due to the fact that Nature provides us with rich facial expressions, our face undergoes complicated non-rigid deformations. Emotions may drastically change the way the facial surface is embedded in the ambient three-dimensional Euclidean space. Such changes are called *extrinsic*. Clearly, the extrinsic geometry is not preserved by facial expressions. Yet, if we restrict our measurements of distance to the facial surface, we notice that distances measured on the surface (that is, the lengths of the shortest paths on the surface, referred to as geodesic distances) remain almost unaltered. This happens due to the fact that our skin and underlying tissues exhibit only slight elasticity: they can be bent but not too much stretched.

In order to validate this claim, we marked tens of fiducial points on a subject's face and scanned its geometry under various, strong and weak, expressions (see Fig. *5-1*). We then compared the absolute change in both geodesic and Euclidean distances between the points. Figure *5-2* demonstrates the result of this experiment. Although there is some change in the geodesic distances between corresponding points in different expressions, considering the 3D scanner accuracy, these distances are

approximately preserved. Geodesic distances exhibit smaller variability compared to Euclidean ones, as depicted in Fig. *5-2*.

Geometric quantities that can be expressed in terms of geodesic distances are referred to as the *intrinsic geometry* and appear to be insensitive to facial expressions. Consequently, facial expressions can be modeled as near-isometric deformations of the face, i.e. such deformations that approximately preserve the distances on the surface. Stated differently, the intrinsic geometry reflects the subject's identity, whereas the extrinsic geometry is the result of the facial expression.
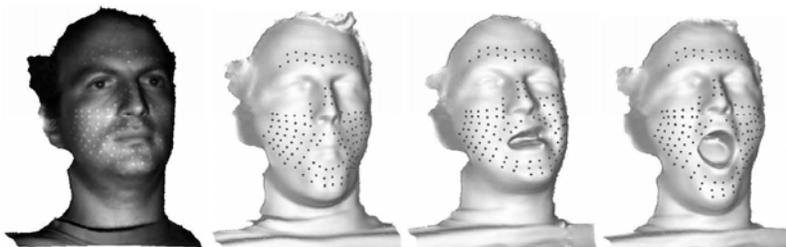

Figure 5-1. Isometric model validation experiment. Left: facial image with the markers. Right: example of one moderate and two strong facial expressions with marked reference points.
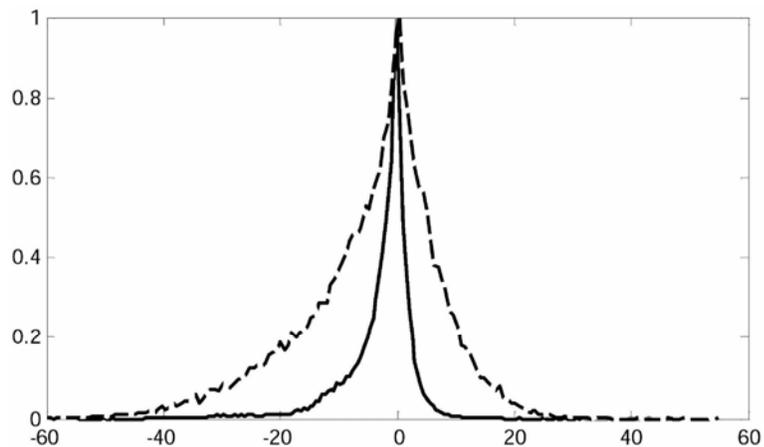

Figure 5-2. Histogram of geodesic distance deviation from the isometric model (solid); for comparison, a histogram for the Euclidean distances is shown (dashed).

In order to compare between faces in a way insensitive to facial expressions, we have to find invariants that uniquely represent the intrinsic geometry of a surface, without being affected by its extrinsic geometry. We model the two faces that we would like to compare as smooth Riemannian surfaces S and Q, with the geodesic distances, d_S and d_Q, respectively. In what follows, we devise computational methods for the comparison of intrinsic geometric properties of two faces.

3.   FLAT EMBEDDING AND CANONICAL FORMS

Our first attempt of expression-invariant face recognition was based on replacing the intrinsic geometry of the surface by a Euclidean one by a process known as *isometric embedding*. The first use of this ides in computer vision dates back to Schwartz *et al.*[1], who tried to analyze the brain cortical surface by embedding it into the plane. Revisiting the ingredients of this method, Zigelman *et al.*[2] introduced a texture mapping procedure, in which the geodesic distances are computed with a numerically consistent efficient scheme[3].

Embedding into the plane can be thought of as an invariant parameterization of the surface. However, in order to compare between intrinsic properties of two surfaces, the dimension of the embedding space has to be at least three, or, using a fancier terminology, the *co-dimension* has to be at least one. Elad and Kimmel[4] proposed to embed the metric structure of the two dimensional surface S into $\mathbf{R}^n$, where $n \geq 3$. Formally, we try to find a map $\varphi : S \to \mathbf{R}^n$, such that $d_{\mathbf{R}^n}(\varphi(s_1), \varphi(s_2)) = d_S(s_1, s_2)$ for every $s_1, s_2 \in S$. Such a map is called isometric embedding. However, for a general non-flat surface, a truly isometric embedding usually does not exist; all we can find is a *minimum-distortion embedding*.

Practically, the surface is sampled at a set of m points $\{s_1, ..., s_m\}$, and we find a configuration of points $\{x_1, ..., x_m\}$ in $\mathbf{R}^n$ by solving the following optimization problem,

$$\{x_1, ..., x_m\} = \arg\min_{x_1, ..., x_m} \sum_{i < j} (d_{\mathbf{R}^n}(x_i, x_j) - d_S(s_i, s_j))^2. \qquad (1)$$

Here, $x_i = \varphi(s_i)$ are the images of the samples of S under the embedding φ. We try to find such a configuration of points that the Euclidean distances $d_{\mathbf{R}^n}$ between each pair of image points is as close as possible to their corresponding original geodesic distances d_S. A numerical procedure

solving the above optimization problem is known as *multidimensional scaling* (MDS).

$\{x_1,...,x_m\}$ can be thought of as an approximation of the intrinsic properties of S. We call it the *canonical form* of the surface. The comparison of canonical forms is a significantly simpler task than comparison of the intrinsic geometries of the non-rigid surfaces themselves. Indeed, for canonical forms there is no difference between extrinsic and intrinsic geometries. Unlike the rich class of non-rigid isometric deformations the original surfaces can undergo, the only degrees of freedom for the canonical forms are the rigid transformations (translation, rotation and reflection), which can be easily solved for using efficient rigid surface matching methods such as the iterative closest point (ICP) algorithm[6, 7] or moments signatures[4,5]. The latter method is especially attractive, since it produces a simple signature describing the geometry of the canonical form that can be efficiently compared to a large data base of signatures representing other faces. Canonical forms cast the original problem of non-rigid surface matching to the simpler problem of rigid surface comparison. Figure *5-3* depicts faces with various expressions embedded into $\mathbf{R}^3$ by the described procedure. Note how even strong expressions of the same subject have just little influence on the canonical forms.

Based on this approach, we built a prototype face recognition system that achieved sufficient accuracy to tell apart identical twins, even in the presence of extreme facial expressions[8]. Nevertheless, the canonical form approach is limited in two aspects. First, the inevitable distortion introduced by the embedding sets an artificial threshold to the sensitivity of the method. Second, in order to perform an accurate matching, the support of the surfaces S and Q must be the same. For that purpose, a pre-processing by means of a consistent cropping of S and Q is required. Changing the surface support regions would generally yield different canonical forms, an undesired property, since in many practical applications, matching of partially missing or partially overlapping surfaces is required.

4. SPHERICAL CANONICAL FORMS

In order to reduce the distortion of embedding facial surfaces into a Euclidean space, we should search for better spaces than the Euclidean ones. A simple space with non-Euclidean geometry, in which the geodesic distances are given analytically is the *n*-dimensional sphere $\mathbf{S}^n$. There exist almost straightforward generalizations of the MDS methods suitable for embedding into $\mathbf{S}^n$. Given the control over the sphere radius R, the

spherical geometry constitutes a richer choice, since it includes the Euclidean case at the limit $R \to \infty$. Once embedded into a sphere, the spherical canonical forms have to be matched. For that goal we developed various tools that can be found in Bronstein *et al.*[9]
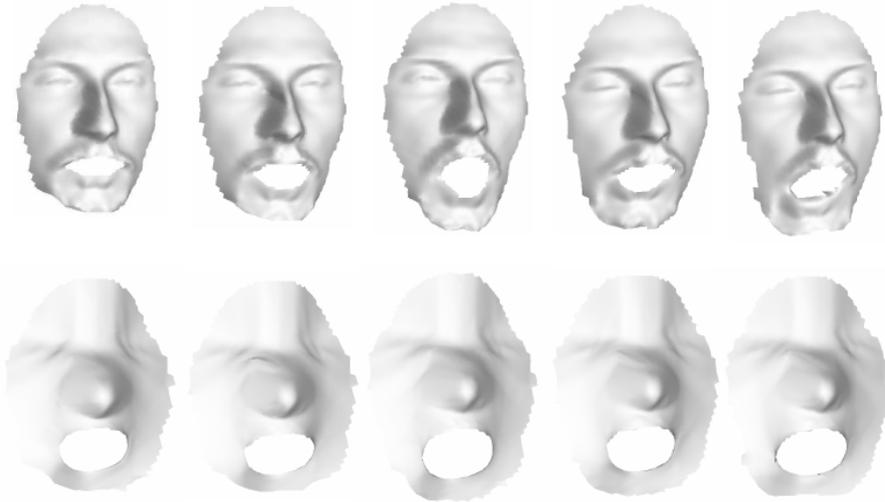

Figure 5-3. Facial expressions and the corresponding canonical forms.

Figure *5-4* demonstrates that embedding into spherical spaces has smaller distortions for some range of radii similar to the radius of an average human face. Moreover, the recognition rates exhibit a clear correlation with the embedding distortion: the lower is the distortion; the more accurate is the recognition. This gives an empirical justification to the pursuit of better embedding spaces.

Although there is an improvement in the recognition rates, the spherical embedding is not the end of our journey. We are still occupied with the problem of partial matching and that of the unavoidable embedding distortions even when selecting a somewhat more favorable embedding space.

5.   GENERALIZED MULTIDIMENSIONAL SCALING

Replacing the Euclidean geometry of the embedding space by the spherical one usually leads to smaller metric distortions and, consequently, to better isometry-invariant representation of surfaces, while maintaining

practically the same computational complexity compared to the Euclidean MDS algorithm. Nevertheless, spherical embedding cannot completely avoid the distortion.
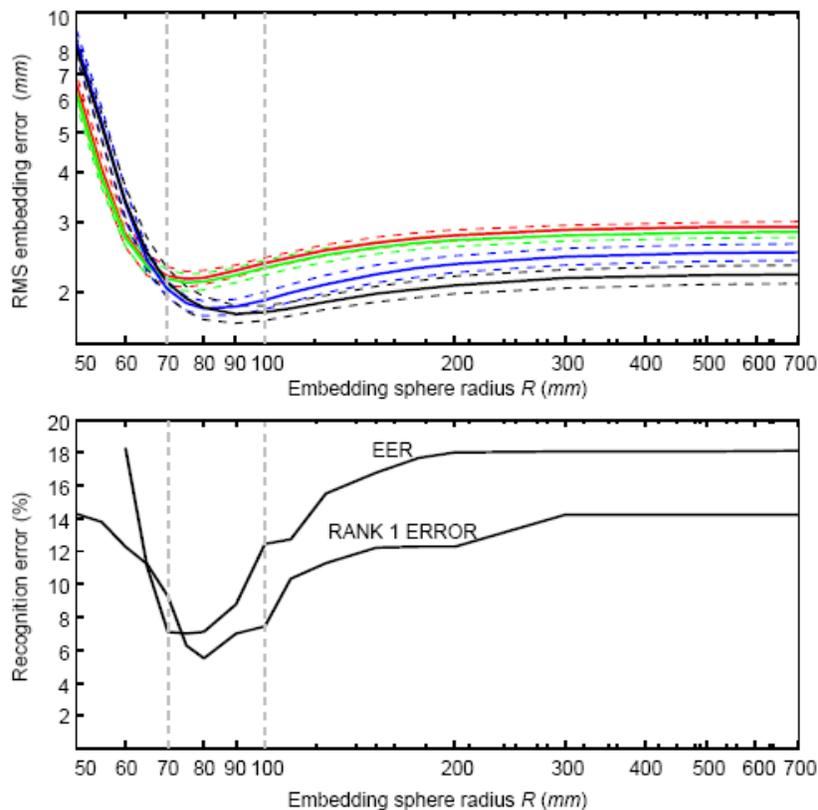

Figure 5-4. First row: embedding error versus the embedding sphere radius for four different subjects (colors denote different subjects, dashed lines indicate 95% confidence intervals). Second row: Equal-error and rank-1 error rates versus the embedding sphere radius. The asymptote $R \rightarrow \infty$ corresponds to embedding into $\mathbf{S}^n$.

It is, however, possible to completely avoid the need of intermediate space by choosing one of the surfaces, say Q, as the embedding space. In other words, we would like to embed S directly into Q. The embedding can be achieved by solving an MDS-like problem,

$$\{q_1,...,q_m\} = \arg \min_{q_1,...,q_m \in Q} \sum_{i>j} (d_S(s_i,s_j) - d_Q(q_i,q_j))^2, \qquad (2)$$

that we term the *generalized multidimensional scaling* or GMDS for short [10, 11]. As in the standard MDS procedure, we find a configuration of points $\{q_1,...,q_m\}$, on the surface Q that represent the intrinsic geometry of S as accurately as possible. The points q_i are the images of s_i under the embedding $\varphi: S \to Q$. The minimum achievable value of the cost function in (2) quantifies how much the metric of S has to be distorted in order to fit into Q. If the two surfaces are isometric, such an embedding will be distortion-less; otherwise, the distortion will measure the dissimilarity between S and Q. This dissimilarity is related to the *Gromov-Hausdorff distance*, first used in the context of the surface matching problem by Mémoli and Sapiro[12].

So far, the embedding distortion has been an enemy that was likely to lower the sensitivity of the canonical form method; now it has become a friend that tells us how different the surfaces are. For this reason, GMDS is essentially the best non-Euclidean embedding, in the sense that it allows to completely avoid unnecessary representation errors stemming from embedding into an intermediate space. Strictly speaking, we do not use canonical forms anymore; the measure of similarity between two surfaces is obtained from the solution of the embedding problem itself.

Another important advantage of GMDS is that it allows for local distortion analysis. Indeed, defining the *local distortion* as

$$\sigma_i = \sum_j \left| d_S(s_i,s_j) - d_Q(q_i,q_j) \right|^2, \qquad (3)$$

we create a map $\sigma: S \to Q$ quantifying the magnitude of the change the metric of S undergoes in every point in order to be embedded into Q (Figure 5-5). Practically, it allows us to determine how much two faces are dissimilar, and also identify the regions with the largest dissimilarities. Last, GMDS enables *partial matching* between non-rigid surfaces, that is, matching a part of S to Q. Partial matching is of paramount importance in practical applications, where due to the limitations of physical acquisition devices, parts of the facial surface may be occluded.

Although GMDS looks like a powerful instrument for isometry-invariant surface matching, there is some cost for its advantages. First, unlike the Euclidean or the spherical cases, we gave up the luxury of computing the distance in the embedding space analytically. Nevertheless, geodesic distances on arbitrarily complex surfaces can be efficiently approximated.[10, 11] The overall complexity of GMDS is comparable to that of the standard

MDS algorithms. Another, shortcoming stems from the fact that every time we need to compare between two faces, we have to solve a new embedding problem. This makes one-to-many comparison scenarios with large databases improbable. An hierarchical matching strategy or a combination of GMDS with the canonical form approach provide some remedy to this difficulty.
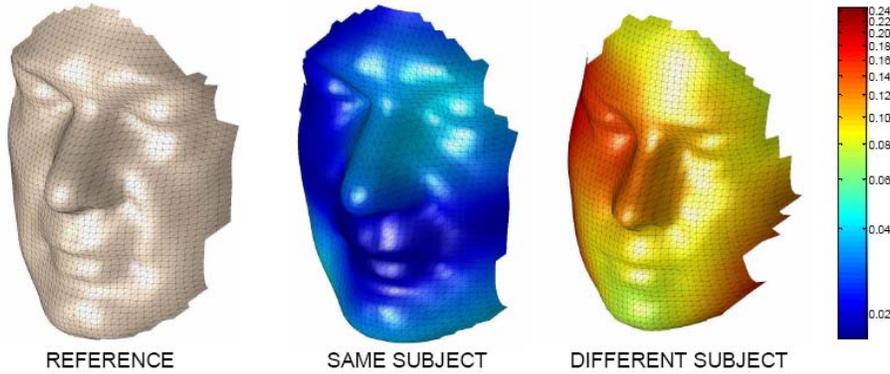

REFERENCE   SAME SUBJECT   DIFFERENT SUBJECT

Figure 5-5. Local distortion map obtained by embedding two faces of two different subjects into a face of a reference subject.

6.   COMPARISON OF PHOTOMETRIC PROPERTIES

So far, we focused our attention on the recognition of the intrinsic facial geometry that appeared to be insensitive to expressions. However, our face is also endowed with photometric characteristics that provide additional useful information for the recognition task. A somewhat simplified model incorporating both geometric and photometric properties of the face consists of a non-rigid surface S and a scalar field $\rho : S \to [0,1]$ associated with it. The scalar field ρ measures the *reflectance coefficient* or *albedo* at each point on the surface, that is, the fraction of the incident light reflected by the surface. If the acquisition device is capable of sensing multiple color channels, ρ can be replaced by a vector field (for example, $\rho : S \to [0,1]^3$ in case of a standard tri-chromatic camera) measuring the reflectance coefficient for different light wave lengths. Using the computer graphics jargon, ρ is the *texture* of S.

In practice, the albedo cannot be directly measured by a camera; what we observe is the *brightness* of the surface, or, in simple words, the amount of radiation scattered from it in the camera direction. However, it appears that our skin behaves approximately like a diffusive reflector, which means that its apparent brightness is roughly the same regardless of the observer's viewing direction. This fact allows using the Lambertian reflection law to estimate the reflectance coefficient ρ given the surface normal field. Clearly, such information is unavailable in two-dimensional face recognition methods, which are based on the brightness image of the face.

In this setting, the problem of expression-invariant face recognition aims at measuring the similarity of two faces, based on the similarity of their intrinsic geometries (S, d_S) and (Q, d_Q), and their photometric properties, ρ_S and ρ_Q. However, in order to be able to compare between ρ_S and ρ_Q, we have to bring them first to some common coordinates system, in which the facial features coincide.

Following the steps we took for comparing the intrinsic geometry, let us briefly go through the same evolution for the texture. Common coordinates can first be found by a common parameterization of S and Q into some planar domain. Such a parameterization should be invariant to facial expressions, which according to our isometric model is not influenced by the extrinsic geometry of the surface. After the surfaces S and Q are re-parameterized, ρ_S and ρ_Q can be represented in the common parameterization domain that makes the comparison trivial using standard image matching techniques. Expression-invariant comparison of the photometric properties of the faces therefore reduces to finding an isometry-invariant "canonical" parameterization of the facial surfaces.

The simplest way to construct such a parameterization is by embedding the surface into $\mathbf{R}^2$ using an MDS algorithm. The problem is very similar to the computation of the canonical form, except that now the embedding space, serving as the parameterization domain, is restricted to be two-dimensional. We refer to such an embedding as *zero co-dimensional*. As the embedding is based on the intrinsic geometry only, such a parameterization will be invariant to isometries, and consequently, the reflectance image in the embedding space will be insensitive to facial expressions. We term such an image the *canonical image* of the face. However, recall that the embedding into $\mathbf{R}^2$ is defined up to rigid isometry, implying that the canonical images can be computed up to planar rotation, translation and reflection, which has to be resolved. Also, the inevitable distortion of the metric introduced by the embedding into a plane makes the canonical image only approximately invariant.

A partial fix for the latter problem comes from non-Euclidean embedding, for example, into the two-dimensional sphere $\mathbf{S}^2$. Since a face is more similar to a sphere than to a plane, spherical embedding produces canonical images with lower distortion. A clear relation between better representation and better recognition is observed again[11]. Another advantage of the spherical embedding is that the obtained spherical canonical images (Figure *5-6*) can be represented using a signature of the spherical harmonic coefficients, that are known to be invariant to rigid isometries on $\mathbf{S}^2$. A property analogous to the translation invariance of the magnitude in the Fourier transform.
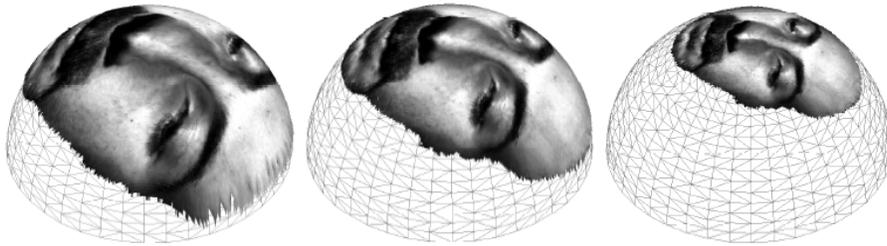

Figure 5-6. Canonical images of a face in $\mathbf{S}^2$ with different radii.

All the approaches described so far provide only approximate isometry-invariance, since a fixed embedding space implies necessarily embedding distortion. As an alternative, we can resort yet again to using the GMDS for embedding S into Q. In addition to quantifying the similarity of the two intrinsic geometries, the minimum-distortion embedding $\varphi: S \to Q$ would also bring ρ_S and ρ_Q to the same coordinates system in Q. The photometric distance between ρ_Q and $\rho_S \circ \varphi$, measured either locally or globally, provides additional information about the similarity of the two faces. Such an approach is inherently metric distortion-free and naturally allows for partial comparison of both photometric and geometric information.

7. CONCLUSIONS

We started with fairy tales, and like most fairy-tales, we are at the happy ending part of our story. We hope the reader found the plot illuminating and rewarding. We first claimed that our face can be described by the isometric

model and validated this claim empirically. We studied a simple isometry invariant signature obtained by replacing the intrinsic geometry by a Euclidean one. We applied this process in a prototype face recognition system, which was able to distinguish between identical twins (the first two authors).

This was just the beginning of the journey for us. Soon after, we noted that embedding into non-Euclidean spaces provides smaller embedding errors and consequently better recognition rates. In both cases, the numerical tools we used for the embedding are members of the well known family of multidimensional scaling algorithms. The next step was to eliminate the embedding error altogether by actually harnessing it to our goal. We utilized the minimal distortion of embedding one surface into another as a measure of similarity between the two surfaces. The new numerical tool, used for solving the embedding problem, is a generalized MDS procedure.

In this chapter, we looked through a keyhole to the world of metric geometry, where objects are non-rigid and isometries introduce new challenges into our life. This is a new playground for engineers and it conceals numerous enigmae and problems waiting to be solved. Our goal was to let the reader touch some of these problems. The methods we discussed could one day help each of us could find his own Cinderella, or at least buy slippers of the right shape...

REFERENCES

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

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

3 R. Kimmel and J. A. Sethian, "Computing geodesic paths on manifolds," *PNAS*, 95(15):8431–8435, 1998.

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

5 M. Elad, A. Tal and S. Ar, "Content based retrieval of VRML objects-an iterative and interactive approach," *EG Multimedia*, 39:97–108, 2001.

6 Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," *Proc. IEEE Conference on Robotics and Automation*, 1991.

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

8 A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *International Journal of Computer Vision (IJCV)*, 64(1):5–30, August 2005.

9 A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant representations of faces," *IEEE Trans. Imag. Proc.*, 16(1):188-197, 2007.

10 A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching," *PNAS*, 103:1168–1172, 2006.

11 A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Efficient computation of isometry-invariant distances between surfaces," *SIAM Journal on Scientific Computing*, 28(5):1812-1836, 2006.

12 F. Mémoli and G. Sapiro, "A theoretical and computational framework for isometry invariant recognition of point cloud data," *Foundations of Computational Mathematics*, 5(3):313-347, 2005.

13 A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant face recognition via spherical embedding," *Proc. IEEE International Conf. Image Processing (ICIP)*, Vol. 3, 756–759, 2005.