Face2Face: an isometric model for facial animation

Alexander M. Bronstein[1], Michael M. Bronstein[1], and Ron Kimmel[1]

Dept. of Computer Science, Technion – Israel Institute of Technology,
Haifa 32000, Israel
{alexbron, bronstein}@ieee.org, ron@cs.technion.ac.il

Abstract. A geometric framework for finding intrinsic correspondence between animated 3D faces is presented. We model facial expressions as isometries of the facial surface and find the correspondence between two faces as the minimum-distortion mapping. Generalized multidimensional scaling is used for this goal. We apply our approach to texture mapping onto 3D video, expression exaggeration and morphing between faces.

1 Introduction

Finding correspondence between human faces is a key problem in numerous problems on the border between computer graphics and computer vision, including: facial animation [1] and modelling [2–4], caricaturization and expression exaggeration [5], cross-parametrization [6, 7], texture mapping [6] and morphing [8, 9]. In the motion pictures industry, one of the challenges is the creation of visually-realistic animated human faces. The rapid development of 3D real-time video acquisition techniques [10] opens a new way to create a synthetic character, by scanning an actor and replacing his or her facial texture with a virtual one, automatically mapping a single image onto a 3D video sequence. We call the effect achieved in this way the "virtual makeup".

The common denominator of the above applications is the *correspondence problem*, i.e. the need to identify the same points in two different instances of a single face (e.g. deformed by facial expressions) or on two different faces. Specifically, we consider the problem of correspondence between 3D facial surfaces, which appears to be significantly harder than its 2D counterpart. Unlike synthetic face animation [1], where the correspondence between meshes and textures is known, in our case the 3D sequence is acquired by a range sensor and therefore, the correspondence is not readily available.

In 3D morphing, the correspondence is usually established by finding a common parametrization domain for the surfaces. Such parametrizations can be constructed using a set of fiducial points, which, in most cases, must be selected manually [9]. A parametrization of faces that is common to all expressions has

Fig. 1. Example of a 3D video sequence of an articulated face.

been proposed in [11, 4]. A hybrid method based on fitting 2D facial images to a deformable 3D model of the face was proposed in [2, 3]. In [12], it was empirically shown that natural facial expressions can be considered as isometries of the facial surface. Multidimensional scaling (MDS) [13] was then used to construct an intrinsic geometric representation of the face for expression-invariant face recognition. Here, we adopt the isometric model to establish correspondence by finding the "most isometric" mapping between two facial surfaces. Our approach is based on a numerical procedure similar to MDS, allowing to embed one surface into another. We refer to this method as the *generalized* MDS, or GMDS for short [14].

This paper consists of five sections. In Section 2, we present the isometric model of facial expressions. Section 3 describes the GMDS problem for finding correspondence between facial surfaces and deals with its numerical implementation. In Section 4, we show some applications of GMDS to a number of problems related to face animation. Section 5 concludes the paper.

2   Isometric model

Consider a 3D video sequence of an articulated face, acquired by real-time 3D scanner. We can think of the video as of a sequence of smooth compact connected two-dimensional Riemannian surfaces, denoted by $\{\mathcal{S}_0, \mathcal{S}_1, ...\}$. The *geodesic distances* (lengths of the shortest paths) $d_{\mathcal{S}_t} : \mathcal{S}_t \times \mathcal{S}_t \rightarrow \mathbb{R}$ on $\mathcal{S}_t$ are induced by the corresponding Riemannian metrics. These distances define the *intrinsic* geometry of the surface. The *extrinsic* geometry is captured by the vector field $\mathbf{s}_t : \mathcal{S}_t \mapsto \mathbb{R}^3$, representing the Euclidean coordinates of the surface points. We call $\mathcal{S}_0$ the *reference frame* or the *reference surface*.

Our goal is to find the *correspondence* between $\mathcal{S}_0$ and $\mathcal{S}_t$, represented by a bijective mapping $\varphi_t : \mathcal{S}_0 \rightarrow \mathcal{S}_t$. When only the geometry is available, this is a very challenging problem. Theoretically, the mappings $\{\varphi_1, \varphi_2, ...\}$ can be estimated by finding correspondence between some fiducial points or features [9]. Yet, the main limitation of feature-based approaches is the fact that they require a precise feature detector. Unfortunately, the number of features that can be robustly detected and tracked using facial surface geometry is usually small.

The geometry of the facial surface contains mostly low-frequency information, while feature detection usually requires high-frequency information. A few points such as the eyes and the nose tip, can be detected sufficiently accurately based on the surface curvature. This implies that the correspondence is available only between a sparse set of points. Alternatively, dense correspondence can be found using optical flow applied to the texture, as done by Blanz *et al.* [4]. However, this approach requires the texture information, which is not always available.

In [12], we showed empirically in the context of 3D face recognition that the deformations of a face due to natural expressions can be approximated by *isometries* (distance preserving transformations). Under this assumption, called here the *isometric model*, all instances of the facial surface in our video are isometric, i.e. there exists a sequence of bijective mappings $\{\varphi_1, \varphi_2, ...\}$; $\varphi_t : \mathcal{S}_0 \to \mathcal{S}_t$ such that

$$d_{\mathcal{S}_0}(s_1, s_2) = d_{\mathcal{S}_t}(\varphi_t(s_1), \varphi_t(s_2)), \tag{1}$$

for all $s_1, s_2 \in \mathcal{S}_0$. In practice, a genuine isometry between two surfaces does not exist, but can be approximated by finding a mapping that distorts the geodesic distances the least. Our claim is that such a near-isometric mapping establishes a correspondence between $\mathcal{S}_t$ and $\mathcal{S}_0$. In the following, we will write φ_t, implying the correspondence found in this manner.

Practice shows that the surfaces need not to be necessarily isometric in order for the minimum-distortion mapping to be a good correspondence. This is due to the fact that in a broad sense, all human faces have similar geometry. Thinking of two faces as of flexible rubber masks, the correspondence problem is that of putting one mask onto the other, while trying to stretch it as less as possible. It is obvious that in most cases, the geometric features (like nose, forehead, mouth, etc.) of the two masks will coincide. A recent breakthrough in surgical face transplantation reinforces this claim. Consequently, given two faces of different subjects, we can still use the same principle to find correspondence between them. We exemplify this idea in Section 4.3.

3 Generalized multidimensional scaling

Let us be given the reference frame $\mathcal{S}_0$ and another frame $\mathcal{S}_t$. Our goal is to find φ_t as the *most isometric mapping* between $\mathcal{S}_0$ and $\mathcal{S}_t$, i.e., a mapping that minimizes the distortion of the geodesic distances. The isometric model guarantees that there exists φ_t with zero or at least near-zero distortion. Since we deal with discrete surfaces, we assume $\mathcal{S}_t$ to be sampled at the points $\{s_1, ..., s_{N_t}\}$ and represented as a triangular mesh. For notation convenience, we write $\mathcal{S}_t$, intending its polyhedral approximation. We denote by $\mathbf{\Delta}_t = (d_{\mathcal{S}_t}(s_i, s_j))$ the matrix of all pairwise geodesic distances between the surface samples, computed numerically using, for example, the *fast marching method* (FMM) [15]. We are looking for a mapping $\varphi_t : \{s_1, ..., s_{N_0}\} \subset \mathcal{S}_0 \to \mathcal{S}_t$, such that $d_{\mathcal{S}_0}(s_i, s_j)$ is as close as possible to $d_{\mathcal{S}_t}(\varphi(s_i), \varphi(s_j))$ for all $(i, j) \in P \subseteq \{1, ..., N_0\} \times \{1, ..., N_0\}$ (some distances must be excluded; see Section 3.1). We refer to such φ_t as *partial embedding* of

$\mathcal{S}_0$ into $\mathcal{S}_t$. Note that $(\mathcal{S}_t, d_{\mathcal{S}_t})$ is assumed continuous here, as $\varphi_t(s_i)$ can be any point on the polyhedron $\mathcal{S}_t$, i.e., can fall between the samples. In practice, we have to approximate the values of $d_{\mathcal{S}_t}$ from $(\{s_1, ..., s_{N_t}\} \subset \mathcal{S}_t, \boldsymbol{\Delta}_t)$.

The partial embedding φ_t can be computed by minimizing the *generalized stress* [16],

$$\sigma(s_1', ..., s_{N_0}') = \sum_{i=1}^{N_0} \sum_{j=i+1}^{N_0} w_{ij} \left(d_{\mathcal{S}_t}(s_i', s_j') - d_{\mathcal{S}_0}(s_i, s_j) \right)^2. \qquad (2)$$

Here, $w_{ij} = 1$ if $(i, j) \in P$ and 0 otherwise, and P denotes the set of pairs of points which are included into the stress computation. The optimization is performed directly on the images $s_i' = \varphi_t(s_i)$, in an MDS-like spirit. The optimal solution

$$\{s_1', ..., s_{N_0}'\} = \underset{s_1', ..., s_{N_0}'}{\operatorname{argmin}} \sigma(s_1', ..., s_{N_0}'), \qquad (3)$$

establishes a correspondence between the given N_0 points $\{s_1, ..., s_{N_0}\} \subset \mathcal{S}_0$ and N_0 points $\{\varphi_t(s_1), ..., \varphi_t(s_{N_0})\}$ on the polyhedron $\mathcal{S}_t$. In this way, we obtain a correspondence between a dense set of points, since N_0 can be as large as necessary. This is opposed to methods based on fiducial points, where the number of points is usually limited. Also note that the mapping we find is $\{s_1, ..., s_{N_0}\} \to \{s_1', ..., s_{N_0}'\}$, and it will generally be bijective.

We refer to problem (3) as the GMDS (generalized MDS). It can be thought of as a generalization of MDS, in which the target Euclidean space is replaced with a general triangular mesh. Since s_i' may be arbitrary points between the samples of the polyhedron $\mathcal{S}_t$, the distances $d_{\mathcal{S}_t}$ between the vertices of $\mathcal{S}_t$ must be computed. We use the *three-point geodesic distance approximation*, a numerical procedure producing a computationally efficient $\mathcal{C}^1$-approximation for $d_{\mathcal{S}_t}$ and its derivatives, interpolating their values from the matrix $\boldsymbol{\Delta}_t$ of pairwise geodesic distances on $\mathcal{S}_t$ [16].

The numerical solution of the GMDS problem consists of bringing the stress (2) to a minimum over s_i' represented in some parameterization domain as vectors of coordinates $\mathbf{u}_i$. For example, if the surface $\mathcal{S}_t$ admits some global parameterization, say $[0, 1)^2 \mapsto \mathcal{S}_t$, every point on $\mathcal{S}_t$ can be represented by $\mathbf{u} \in [0, 1)^2$. Global parameterization is often readily available for objects acquired using many types of range scanners. Human faces usually fall into this category. [1] The minimization algorithm starts with some initial guess $\mathbf{u}_i^{(0)}$ of the points and proceeds by iteratively updating their locations, producing a decreasing sequence of stress values. In our implementation, we used a gradient descent algorithm safeguarded by inexact linesearch (Armijo rule) [18]. The complexity of the stress and its gradient computation is $\mathcal{O}(N_0^2)$. Since N_0 typically varies between tens to hundreds of points, GMDS is computationally efficient.

---

[1] For objects with more complicated topology, global parameterization may not exist; in this case, we represent a point on $\mathcal{S}_t$ by the triangle index m it and a vector $\mathbf{u}$ of *barycentric coordinates* [17] in the local coordinate system of that triangle.

Finally, we must note that GMDS is a non-convex optimization problem, like traditional MDS. Consequently, the use of convex optimization algorithms in this problem is liable to local converge [13]. Nevertheless, convex optimization is widely used in the MDS community if some precautions are taken in order to prevent convergence to local minima. Here, we use a multiscale optimization scheme that in practical applications shows good global convergence [16].

3.1 Selection of weights

Expressions with open mouth do not fit into the isometric model, in which we tacitly assumed a fixed topology of the surface. Opening the mouth creates a "hole" in the facial surface. Resolving this problem is possible imposing a topological constraint on the facial surface, for example, assuming the mouth to be always open [19]. This is achieved by essentially cutting off the lip contour in the reference frame $\mathcal{S}_0$, either automatically or manually (in practice, the lip detector does not have to be very accurate).

An important issue arising after such a processing is the inconsistency of minimal geodesics. Let $\mathcal{S}'_0$ denote the reference frame after lip cropping. We assume that the geodesic distances on $\mathcal{S}'_0$ are given by the restricted metric, $d_{\mathcal{S}'_0}(s_1, s_2) = d_{\mathcal{S}_0}|_{\mathcal{S}'_0}(s_1, s_2)$ (this notation implies that $d_{\mathcal{S}'_0}(s_1, s_2) = d_{\mathcal{S}_0}(s_1, s_2)$ for all $s_1, s_2 \in \mathcal{S}'_0$). However, $d_{\mathcal{S}'_0}$ is computed numerically on $\mathcal{S}'_0$ and can be inconsistent with $d_{\mathcal{S}_0}|_{\mathcal{S}'_0}$. Potentially, the problem arises with minimal geodesics that are close to the boundary $\partial \mathcal{S}'_0$. Such geodesics can be substantially different on $\mathcal{S}_0$ and $\mathcal{S}'_0$, and the corresponding distances are therefore inconsistent. In order to resolve this problem, define the set P of consistent distances, excluding every pair of points (s_i, s_j), for which the minimal geodesic passes through the cropped region $\mathcal{S}_0 \setminus \mathcal{S}'_0$. Particularly, we exclude in this way the distances that would have been measured S_0 across the lips on the original surface.

4   Applications

The knowledge of the intrinsic correspondence between two facial surfaces allows us to perform texture mapping onto all the frames of the video sequence. Moreover, we can also transform the extrinsic geometry of the faces, creating an interpolation or morphing effect between the 3D frames. Finally, the same approach can be applied to morphing between faces of different subjects.

4.1   Virtual makeup

Our first application is the "virtual makeup" – expression-invariant mapping of a single texture image onto a 3D video of an animated face. We first draw the texture (represented as the field $\alpha_0 : \mathcal{S}_0 \mapsto \mathbb{R}^3$, consisting of the R, G and B channels) on the reference frame $\mathcal{S}_0$. Next, using the correspondences, we map the texture onto the rest of the frames in the 3D video.

Fig. 2. Processing stages in the virtual makeup problem (left to right): reference surface; cropping and subsampling; texture mapping onto the reference surface; correspondence establishment using GMDS and texture mapping onto the target surface.

A scheme of the procedure is depicted in Figure 2. The reference surface $\mathcal{S}_0$ first undergoes cropping that removes the lips and leaves only the facial contour. The obtained region $\mathcal{S}_0'$ is subsampled using farthest point sampling, geodesic distance between the samples are computed using FMM [15]. The distances crossing the cropped lips region are assigned zero weights. Next, the texture α_0 is drawn on the reference surface. The points on $\mathcal{S}_0$ are then embedded into the target surface $\mathcal{S}_t$ using GMDS, which produces the correspondence φ_t. The mapping φ_t is used to interpolate the texture onto the surface $\mathcal{S}_t$, yielding a synthetic texture $\alpha_t = \alpha_0 \circ \varphi_t^{-1}$.

We tested our virtual makeup algorithm on a real 3D video sequence of a face, acquired by a structured light scanner at 640×480 spatial resolution, 3 frames per second (Figure 1). The lip contour in the reference frame was segmented manually. The cropped reference frame was sampled at 100 points; all the rest of the frames were sampled uniformly at about 3000 points. The surfaces were triangulated using Delaunay triangulation; then, the geodesic distances were computed using FMM [15]. The correspondence was found by embedding 100 points on $\mathcal{S}_0$ into $\mathcal{S}_t$ using a multiresolution optimization scheme, initialized with 8 points at the coarsest level. A MATLAB implementation of GMDS[2] was used. Figure 3 depicts a synthetic Shrek-like character, created from the video sequence by mapping a synthetic face texture image (drawn in Photoshop) using our algorithm. The faces produced in this way look real and the texture alignment is preserved even in case of strong facial expressions.

4.2   Expression interpolation and exaggeration

The correspondence found by means of GMDS can also be used to transform the extrinsic geometry of the surfaces. Let $\mathcal{S}_t$ and $\mathcal{S}_{t+1}$ be two adjacent frames in the 3D video, and let $\psi_t = \varphi_{t+1} \circ \varphi_t^{-1}$ be the correspondence between them. Let $\mathbf{s}_t : \mathcal{S}_t \mapsto \mathbb{R}^3$ and $\mathbf{s}_{t+1} : \mathcal{S}_{t+1} \mapsto \mathbb{R}^3$ denote the extrinsic Euclidean coordinates

---

[2] Codes and demos will be published on `http://tosca.cs.technion.ac.il`

Fig. 3. Virtual makeup: a few frames from the video sequence shown in Figure 1, with a Shrek texture image mapped using the correspondence established by GMDS.

of $\mathcal{S}_t$ and $\mathcal{S}_{t+1}$, respectively. The extrinsic geometry of the surfaces is assumed to be at least roughly aligned by means of a rigid (Euclidean) transformation. Three points are enough for such an alignment. In our case, this is a simple task since the correspondence is known.
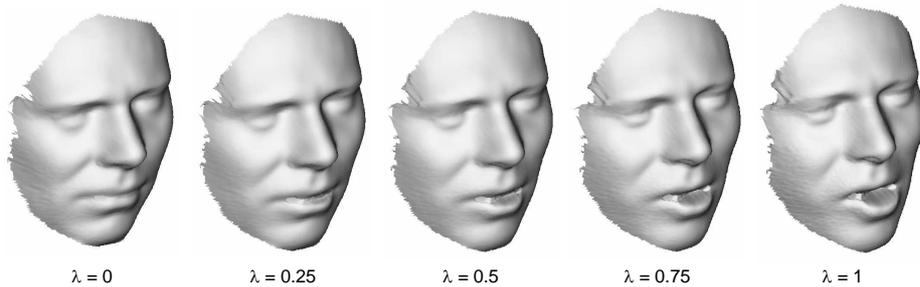
We define a new surface $\mathcal{S}_{t+\lambda}$ with extrinsic coordinate given by the following convex combination:

$$\mathbf{s}_{t+\lambda}(s) = \lambda \mathbf{s}_t(s) + (1 - \lambda)\mathbf{s}_{t+1}(\psi_t(s)), \tag{4}$$

for all $s \in \mathcal{S}_{t+\lambda}$ and $\lambda \in [0, 1]$. The corresponding texture $\alpha_{t+\lambda}$ is defined in a similar manner. Varying the value of λ continuously from 0 to 1, we create a natural interpolation between the frames $\mathcal{S}_t$ and $\mathcal{S}_{t+1}$. The synthetic surfaces obtained this way have a realistic look (Figure 4). Such an interpolation is useful, for example, as a method of temporal super-resolution of a 3D video. Allowing for $\lambda < 0$ or $\lambda > 1$, we can create a new, exaggerated facial expression (Figure 5).

4.3 Texture substitution and morphing between different faces

Relaxing the basic assumption of the isometric model, we can use GMDS in order to find the correspondence between two different faces. Though two different facial surfaces are not even approximately isometric, the minimum-distortion mapping appears to be a surprisingly good correspondence even in this case. In

Fig. 4. Expression interpolation between two frames in the video sequence (shown without texture to emphasize the natural look of the synthetic expressions).
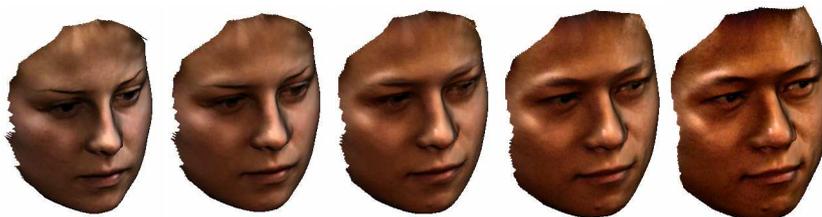

Fig. 5. Expression exaggeration. First row: original expressions. Second row: exaggerated expressions.

our example, as $\mathcal{S}_0$ and $\mathcal{S}_1$, we took a female and a male face from the Notre Dame database [20]. Each face was subsampled to approximately 3000 points and triangulated. The shapes were roughly aligned. Fifty points were taken on $\mathcal{S}_0$ and embedded into $\mathcal{S}_1$ using GMDS. The resulting correspondence φ_1 was then used to map the texture α_0 from $\mathcal{S}_0$ to $\mathcal{S}_1$. Figure 6 shows a synthetic face obtained by taking $\mathcal{S}_1$ with the texture $\tilde{\alpha}_1 = \alpha_0 \circ \varphi_1^{-1}$ (male geometry with a female texture). Figure 7 shows a morphing effect between $\mathcal{S}_0$ and $\mathcal{S}_1$, obtained by interpolating the extrinsic geometry and the texture according to (4).

Fig. 6. Texture substitution: GMDS is used to find the minumum-distortion mapping between face $\mathcal{S}_0$ and $\mathcal{S}_1$ (by embedding $\mathcal{S}_0$ into $\mathcal{S}_1$). Using this mapping as a correspondence, the texture α_0 is mapped onto $\mathcal{S}_1$.

Fig. 7. Morphing: the correspondence is used to transform the texture and the extrinsic geometry of $\mathcal{S}_0$ into the corresponding texture and extrinsic geometry of $\mathcal{S}_1$.

5 Conclusions

We presented an automatic geometric procedure for establishing dense correspondence between facial surfaces. Exploiting the empirical fact that facial expressions can be modelled as isometries, our approach is based on finding the minimum-distortion mapping between two surfaces. This mapping is computed by a procedure similar to multidimensional scaling (GMDS). The algorithm is computationally efficient, though currently not real-time. Our preliminary results show that near real-time performance can be achieved by exploiting multi-grid optimization [21] and implementation on graphics processors (GPU).

Unlike feature-based methods, our approach does not require feature detection and tracking. We find correspondence between an arbitrarily dense set of points, as opposed to feature-based methods, which are usually limited to a small set of fiducial points that can be robustly detected and tracked. Moreover, our approach is applicable when 2D information (texture) is not available. The proposed method is generic and has a wide range of uses in computer graphics and computer vision. We demonstrated some applications, including the "virtual makeup" by expression-invariant texture mapping onto an animated face, texture substitution and morphing.

References

1. Y. Lee, D. Terzopoulos, and K. Waters. Realistic modeling for facial animation. In *Proc. SIGGRAPH*, volume 16, pages 55–62, 1995.
2. T. Vetter and V. Blanz. A morphable model for the synthesis of 3D faces. In *Proc. SIGGRAPH*, 1999.
3. F. Pighin, R. Szeliski, and D. H. Salesin. Modeling and animating realistic faces from images. *IJCV*, 50(2):143–169, November 2002.
4. V. Blanz, C. Basso, T. Poggio, and T. Vetter. Reanimating faces in images and video. *Computer Graphics Forum*, 22(3):641–650, 2003.
5. S. E. Brennan. The caricature generator. *Leonardo*, 18:170–178, 1985.
6. 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.
7. K. Zhou, J. Snyder, B. Guo, and H.-Y. Shum. Iso-charts: Stretch-driven mesh parameterization using spectral analysis. In *Proc. ACM SGP*, pages 45–54, 2004.
8. M. Alexa. Merging polyhedral shapes with scattered features. *The Visual Computer*, 16:26–37, 2000.
9. V. Kraevoy, A. Sheffer, and C. Gotsman. Matchmaker: Constructing constrained texture maps. In *Proc. SIGGRAPH*, 2003.
10. P. S. Huang, C. P. Zhang, and F. P. Chiang. High speed 3-D shape measurement based on digital fringe projection. *Optical Engineering*, 42(1):163–168, 2003.
11. G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face recognition using active appearance models. In *Proc. ECCV*, 1998.
12. A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *IJCV*, 64(1):5–30, August 2005.
13. I. Borg and P. Groenen. *Modern multidimensional scaling - theory and applications*. Springer-Verlag, Berlin Heidelberg New York, 1997.
14. A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *Proc. National Academy of Sciences*, 103(5):1168–1172, January 2006.
15. R. Kimmel and J. A. Sethian. Computing geodesic on manifolds. In *Proc. National Academy of Science*, volume 95, pages 8431–8435, 1998.
16. A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Efficient computation of isometry-invariant distances between surfaces. Technical Report CIS-2006-02, Dept. of Computer Science, Technion, Israel, 2005.
17. M. S. Floater and K. Hormann. *Advances on Multiresolution in Geometric Modelling*, chapter Surface Parameterization: a Tutorial and Survey. Springer-Verlag, Heidelberg, 2004. To appear.
18. D. Bertsekas. *Nonlinear programming*. Atlanta Scientific, 2 edition, 1999.
19. A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Expression-invariant representations for human faces. Technical Report CIS-2005-01, Dept. of Computer Science, Technion, Israel, 2005.
20. K. Chang, K. W. Bowyer, and P. J. Flynn. Face recognition using 2D and 3D facial data. In *ACM Workshop on Multimodal User Authentication*, pages 25–32, 2003.
21. M. M. Bronstein, A. M. Bronstein, R. Kimmel, and I. Yavneh. Multigrid multidimensional scaling. *Numerical Linear Algebra with Applications (NLAA)*, 13:149–171, March-April 2006.