EXPRESSION-INVARIANT FACE RECOGNITION VIA SPHERICAL EMBEDDING

*Alexander M. Bronstein, Michael M. Bronstein, Ron Kimmel**

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

ABSTRACT

Recently, it was proven empirically that facial expressions can be modelled as isometries, that is, geodesic distances on the facial surface were shown to be significantly less sensitive to facial expressions compared to Euclidean ones. Based on this assumption, the 3DFACE face recognition system was built. The system efficiently computes expression invariant signatures based on isometry-invariant representation of the facial surface. One of the crucial steps in the recognition system was embedding of the face geometric structure into a Euclidean (flat) space. Here, we propose to replace the flat embedding by a spherical one to construct isometric invariant representations of the facial image. We refer to these new invariants as *spherical canonical images*. Compared to its Euclidean counterpart, spherical embedding leads to notably smaller metric distortion. We demonstrate experimentally that representations with lower embedding error lead to better recognition. In order to efficiently compute the invariants we introduce a dissimilarity measure between the spherical canonical images based on the spherical harmonic transform.

1. INTRODUCTION: THE ISOMETRIC MODEL OF FACIAL EXPRESSIONS

We start by briefly reviewing the ideas behind our three-dimensional expression-invariant face recognition system. We refer to [1, 2, 3, 4] for a detailed introduction and presentation of the 3DFACE system and computational methods. Here, we focus on the isometric model of human facial expressions and the face recognition method based on this model.

A face can be thought of as a complete compact smooth two-dimensional Riemannian manifold (surface) $(\mathcal{S}, g)$ with a Riemannian metric g, endowed with some property field (e.g., the scalar field $\rho : \mathcal{S} \mapsto [0, 1]$ representing the gray-scale albedo of the face). Practically, range and intensity imaging devices can provide a finite set of points $\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_N \in \mathcal{S}\}$ obtained by the sampling of $\mathcal{S}$, and the corresponding reflectance value r at these points.

Most existing face recognition algorithms are based on the reflectance image only (2D face recognition). Two-dimensional data, however, is sensitive to various *external* factors influencing the reflectance image such as illumination conditions, head orientation, and the use of make up. In addition, facial expressions (which can be considered *internal* factors) affect the reflectance image as well (see an example in Figure 1, first row). A recent trend in face recognition is the attempt to use the 3D data (geometry of the face), sometimes combined with the reflectance image. Yet, while the geometry of the face is practically invariant to most external factors, it is still affected by facial expressions.

In [1], we have presented an expression-invariant face recognition method, based on the conjecture that facial expressions can be modelled as *isometries* of the facial surface. Formally, this means that a facial expression is a diffeomorphism $f : (\mathcal{S}, g) \mapsto (\mathcal{Q}, h)$ between two Riemannian surfaces, which preserves the *geodesic distances*, that is, for all $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in \mathcal{S}$, and $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \mathcal{Q}$ where $f : \boldsymbol{\xi}_i \mapsto \boldsymbol{\eta}_i$,

$$d_{\mathcal{S}}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = d_{\mathcal{Q}}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2), \tag{1}$$

where $d_{\mathcal{S}}$ and $d_{\mathcal{Q}}$ denote the geodesic distances induced by the Riemannian metrics g and h, respectively. The surfaces $(\mathcal{S}, g)$ and $(\mathcal{Q}, h)$ are called *isometric*. Although in reality strong facial expressions are not strictly isometric, this model is still far better than the common practice of regarding facial surfaces as rigid objects. An experimental validation of the isometric model is available in [4].

2. ISOMETRY-INVARIANT CANONICAL FORMS

Under the assumption of the isometric model, it is clear that in order to obtain an expression-invariant (isometry-invariant) representation of the face, one has to get rid of its *extrinsic geometry*, that is, the way the surface $\mathcal{S}$ is immersed into the ambient three-dimensional Euclidean space, keeping only the *intrinsic geometry*, that is, the geometry *on* the surface itself.

An obvious isometric invariant of the surface is the set of all the geodesic distances between its points. However, we should remember that only a *sampled* version of the surface $\mathcal{S}$ is available, and therefore in practice we have a *finite metric space* $(\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_N\}, \mathbf{D})$, where the matrix $\mathbf{D} = (d(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j))$ denotes the mutual geodesic distances between the points in $\mathcal{S}$[1]. There is no guarantee that different instances of the same facial surface are sampled at the same points, nor that the number of samples is the same. Moreover, even if the samples are the same, they can be ordered arbitrarily. This ambiguity makes impractical the use of $\mathbf{D}$ itself as an invariant representation.

An alternative proposed in [7] is to avoid dealing explicitly with the matrix of geodesic distances and represent the Riemannian surface as a subset of some convenient m-dimensional space $\mathcal{S}'^m$, such that the original intrinsic geometry is preserved. Such a procedure is called an *isometric embedding*, and allows to get rid of the extrinsic geometry, which does not exist in the new space. As the result of isometric embedding, the representations of all

*This research has been supported by the Israel Science Foundation (ISF) grant no. 738/04 and the Bar-Nir Bergreen Software Technology Center of Excellence – Faculty of Computer Science internal grant.

[1] In practice, $d(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)$ are neither available, yet they can be approximately computed from the point cloud $\{\mathbf{x}_n\}$. Here, a modified fast variation of the Fast Marching Method on parametric manifolds [5, 6] was used for this purpose.

the isometries of $\mathcal{S}$ are identical, up to the isometry group in $\mathcal{S}'^m$, which is usually easy to deal with, for example, in $\mathbb{R}^m$ all the possible isometries are rotations, translations, and reflections. Elad and Kimmel focused on embedding into $\mathbb{R}^m$ for $m > 2$; planar embedding was used beforehand in the analysis of cortical surfaces [8] and in texture mapping [9]. Yet, the embedding space $\mathcal{S}'^m$ does not necessarily has to be Euclidean and can be chosen completely to our discretion.

In the discrete setting, isometric embedding is a mapping between two finite metric spaces

$$\varphi : (\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_N\} \subset \mathcal{S}, \mathbf{D}) \rightarrow \left(\{\boldsymbol{\xi}'_1, ..., \boldsymbol{\xi}'_N\} \subset \mathcal{S}'^m, \mathbf{D}'\right),$$

such that for all $i, j = 1, ..., N$, $d'_{ij} = d_{ij}$. The matrices $\mathbf{D} = (d_{ij}) = (d(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j))$ and $\mathbf{D}' = (d'_{ij}) = (d'(\boldsymbol{\xi}'_i, \boldsymbol{\xi}'_j))$ denote the mutual geodesic distances between the points in the original and the embedding space, respectively. Following Elad and Kimmel, the image of $\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_N\}$ under φ is called the *canonical form* of $(\mathcal{S}, g)$ [7]. In general, such isometric embedding does not exist, and therefore one has to bear in mind that the canonical form is an *approximate* representation of the discrete surface. It is possible to find optimal canonical forms in sense of some metric distortion criterion. Again, the canonical form is uniquely defined up to any transformation in the embedding space that does not alter the distances (like translations, rotations, and reflections in an Euclidean embedding space).

The expression-invariant 3D face recognition method introduced in [1] is based on embedding facial surfaces into $\mathbb{R}^3$, followed by rigid comparison of the resulting canonical forms. However, a "geometry-only" approach does not make explicit use of the reflectance image, which may contain significant information and can help to discriminate between faces. A way to incorporate the facial image into the geometric framework is to perform the embedding of $(\mathcal{S}, g)$ with the associated reflectance field r into $\mathbb{R}^2$, exploiting ideas similar to those used in [9] for texture mapping [1, 2]. As a result, the reflectance image in the embedding space, which we term as the *canonical image* appears like a warped version of the original image. Since the Euclidean distance between the pixels in the canonical image approximates the geodesic distance between the corresponding points on the facial surface, such a representation is insensitive to facial expressions. If instead of the reflection image an estimate of the albedo (reflection coefficient) or an illumination-compensated image is used, the canonical image is also insensitive to illumination.

A disadvantage of the discussed methods stems from the translation, orientation, and reflection ambiguity of canonical forms in $\mathbb{R}^3$ and canonical images in $\mathbb{R}^2$. The possibility to perform precise alignment required for their comparison is, in practice, limited by the relatively small number of points participating in the embedding.

3. SPHERICAL CANONICAL IMAGES

This paper introduces improvements to the method of isometry-invariant canonical images, novel in three aspects. First, we propose to embed the facial image into a two- dimensional sphere $\mathbb{S}^2$ rather than a plane, which will be shown to produce lower embedding distortions. Second, we justify the need for searching embedding spaces "better" than the Euclidean one, where the embedding error is lower than that in $\mathbb{R}^2$ by showing that embedding error is related directly to recognition accuracy. Last, we take advantage of the fact that the new canonical image is defined on $\mathbb{S}^2$ and use spherical harmonics to measure dissimilarity between two images. Rotation and reflection invariance of spherical harmonics removes the embedding ambiguity and does not require alignment of the canonical images.

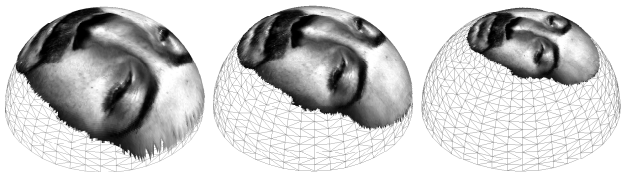

Fig. 2. Face embedded into $\mathbb{S}^2$ with different radii, from left to right: $R = 80mm, 100mm$ and $150mm$.

We parameterize $\mathbb{S}^2$ using two angles: the elevation $\theta^1 \in \left[-\frac{\pi}{2}, +\frac{\pi}{2}\right]$ measured from the azimuthal plane in the northern direction, and the azimuthal angle $\theta^2 \in [0, 2\pi)$. The geodesic distance between two points $\boldsymbol{\theta} = (\theta^1, \theta^2)$ and $\boldsymbol{\theta}' = (\theta'^1, \theta'^2)$ is given by

$$d_{\mathbb{S}^2}\left(\boldsymbol{\theta}, \boldsymbol{\theta}'\right) = R \cos^{-1}\left(\cos\theta^1 \cos\theta'^1 \cos(\theta^2 - \theta'^2) + \sin\theta^1 \sin\theta'^1\right), \quad (2)$$

where R is the sphere radius. The choice of R changes the curvature of the embedding space, thus influencing the embedding error. We will show that there exists an optimal value of R suitable for most human faces. For convenience, we will henceforth consider a unit sphere; different values of R will be achieved by scaling the input distance matrix $\mathbf{D}$.

The embedding procedure aims to find such a configuration of points $\{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_N\}$ on $\mathbb{S}^2$ that minimize some discrepancy measure (*stress*) between d_{ij} and $d'_{ij} = d_{\mathbb{S}^2}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)$. Among the variety of stress functions, see for example [10], we have chosen the raw stress

$$\epsilon(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_N; \mathbf{D}) = \sum_{i<j} \left(d_{ij} - d_{\mathbb{S}^2}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)\right)^2. \quad (3)$$

Gradient descent with backtracking line search was used for minimization. For reasons stated in the following, one of the points, say $\boldsymbol{\theta}_1$, was restricted to reside on the north pole of the sphere ($\theta_1^1 = \frac{\pi}{2}$). This point was chosen to be the nose tip and was determined as a local maximum of the Gaussian curvature on the facial surface.

The transformation $\varphi : \mathcal{S} \mapsto \mathbb{S}^2$ from the original manifold to the sphere can be regarded as a warping transformation, which maps the original facial image onto a portion of the sphere. The resulting image, $f : \mathbb{S}^2 \mapsto \mathbb{R}$, can be computed for any $\boldsymbol{\theta}$ by means of linear interpolation (see Figures 1–2). The spherical canonical image f is invariant to isometric deformations of the face (hence, insensitive to facial expressions) by definition of the embedding. Moreover, if an albedo estimate is used as the original image, f is also insensitive to illumination. Note that the spherical canonical image is not a fully invariant signature of the face, since fixing a single fiducial point on the pole still allows one degree of freedom of rotation and reflection about that point.

4. SPHERICAL HARMONIC SIGNATURES

We resort to the *spherical harmonic transform* in order to obtain a truly invariant signature of the face. A function $f \in \mathbb{L}^2\left(\mathbb{S}^2\right)$ can

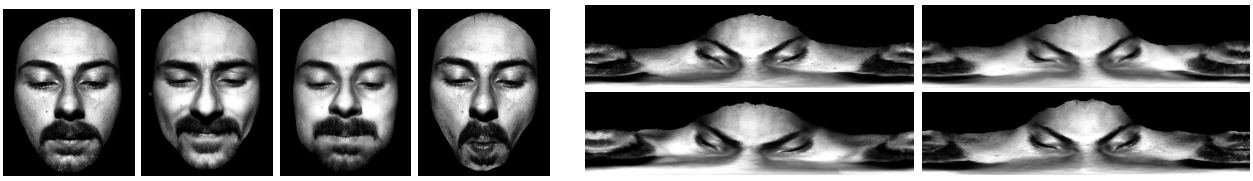

Fig. 1. Left: four representative facial expressions, from left to right: neutral, disgust, inflated cheeks, and deflated cheeks. Right: the same faces after embedding into $\mathbb{S}^2$ ($R = 100mm$), represented in the parametric plane $\theta^1 = 0 \div 90^\circ, \theta^2 = 0 \div 360^\circ$.

be expanded in the spherical harmonic basis with the coefficients

$$\begin{aligned} \hat{f}_{l,m} &= \langle f, Y_l^m \rangle \\ &= \int_0^{\pi} \int_0^{2\pi} f\left(\theta^1, \theta^2\right) \overline{Y_l^m\left(\theta^1, \theta^2\right)} d\theta^2 \cos(\theta^1) d\theta^1, \end{aligned} \tag{4}$$

for $l \in \mathbb{N} \cup \{0\}$ and $|m| \leq l$, where

$$Y_l^m\left(\theta^1, \theta^2\right) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m\left(\sin\theta^1\right) e^{\imath m \theta^2} \tag{5}$$

is the (l,m)-spherical harmonic and P_l^m is the associate Legendre function of degree l and order m. A discrete version of the spherical harmonic transform $\hat{f}_{l,m}$ can be carried out efficiently using the FFT [11].

A very handy property of spherical harmonics is that for every $\Delta\theta^2$, $f\left(\theta^1, \theta^2\right)$ and $f\left(\theta^1, \theta^2 \pm \Delta\theta^2\right)$ are transformed to two sets of coefficients, which differ only in the complex phase. Hence, the set of coefficients $c_{l,m} = \left|\hat{f}_{l,m}\right|$ removes the rotation and reflection ambiguities from the canonical image and defines an invariant signature of the face.

As a dissimilarity measure between two such signatures, we use the Euclidean norm

$$d_F\left(c_{l,m}, c'_{l,m}\right) = \sum_{l \geq 0, |m| \leq 0} \left(c_{l,m} - c'_{l,m}\right)^2. \tag{6}$$

Using basic properties of spherical harmonics, it is straightforward to show that such a measure is characterized by the *similarity* property, i.e.

$$d_F\left(c_{l,m}, c'_{l,m}\right) \leq \sum_{l \geq 0, |m| \leq 0} \left|\hat{f}_{l,m} - \hat{f}'_{l,m}\right|^2 = \left\|f - f'\right\|, \tag{7}$$

and since $d_F\left(c_{l,m}, c'_{l,m}\right)$ is invariant under any $\mathcal{R} \in \mathcal{G}$, where $\mathcal{G}$ is the group of rotations and reflections about the north pole on $\mathbb{S}^2$,

$$d_F\left(c_{l,m}, c'_{l,m}\right) \leq \min_{\mathcal{R} \in \mathcal{G}} \left\|f - \mathcal{R}f'\right\|. \tag{8}$$

In other words, similar f and f' (up some $\mathcal{R} \in \mathcal{G}$) will result in small dissimilarities in sense of d_F, whereas dissimilar f and f' will result in large values of d_F. In practice, different "frequencies" in the spherical harmonics domain typically have different weights. Therefore, a more sophisticated dissimilarity measure could be based on the weighted Euclidean norm. Optimal weights can be found from a training set e.g. by means of PCA.

A disadvantage of the proposed representation stems from the fact that it is invariant to azimuthal roto-reflection only, and not to the general one. This, in turn, requires the use of constrained embedding, which fixes the location of the nose tip, and thus relies on its faithful localization. The use of a non-constrained embedding is feasible in combination with a signature invariant under a general roto-reflection group on $\mathbb{S}^2$. Construction of such a signature is based on the fact that the subspace

$$V_l = \operatorname{span}\left\{Y_l^m : |m| \leq l\right\} \tag{9}$$

is closed under a roto-reflection group on $\mathbb{S}^2$ [12]. Hence, the signature of the form

$$c_l = \left\|\hat{f}_l\right\| = \left\|\left(\hat{f}_{l,-l}, ..., \hat{f}_{l,l}\right)\right\| \tag{10}$$

is invariant under general rotations and reflections. Dissimilarity between such signatures can be measured as in (6).

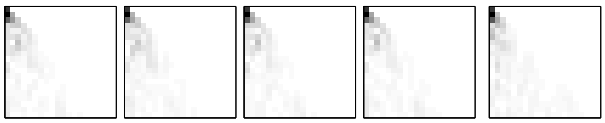

Fig. 3. Part of the spherical harmonics coefficients (magnitude only) of the four faces from Fig. 1 (four leftmost images), and of a distinct subject's face (rightmost image). Horizontal and vertical axes correspond to m and l, respectively; white indicates zero. Distances in terms of d_F from the leftmost face are indicated below the images.

5. NUMERICAL RESULTS

A data-set of 104 faces was used for the experiments. The set consisted of four subjects (two of which are identical twins) with extreme facial expressions. Each subject was acquired with five instances of neutral expression and three instances of smile, anger, surprise, inflated cheeks, deflated cheeks, and neutral expression with eyeglasses. The faces were preprocessed, and 500×500 matrices $\mathbf{D}$ of geodesic distances between points on the facial surfaces were computed. For further technical details on data acquisition, preprocessing and geodesic distance computation, the reader is referred to [1, 2, 4].

In the first experiment, the influence of the embedding sphere radius on the embedding error was tested. Figure 4 depicts the average RMS embedding error for each subject plotted as a function of R. Experiments on a larger variety of subjects allow to establish that embedding sphere radius yielding the minimum embedding error ranges from 90 to $100mm$, slightly depending on the subject.

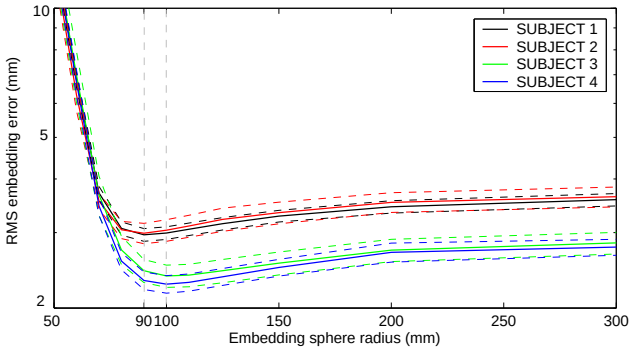

Fig. 4. Average embedding error vs. the embedding sphere radius for four different subjects. Dashed lines indicate 95% confidence intervals. $R = \infty$ stands for Euclidean embedding.

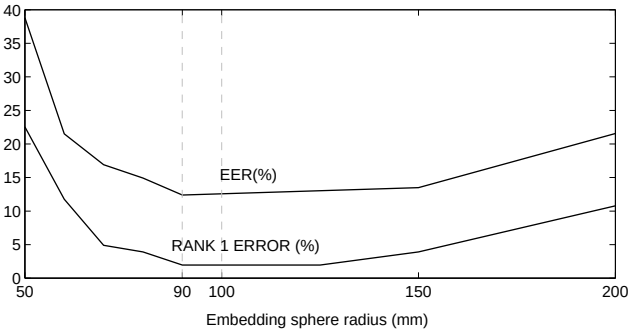

Fig. 5. EER and rank 1 error rate vs. the embedding sphere radius.

Intuition suggests that representation with lower distortion due to embedding should produce better recognition results. However, such a claim is by no means obvious and requires an experimental proof. In the second experiment, the fast spherical harmonic transform was applied to the data of the previous experiment[2] and Euclidean distances between the magnitudes of the spherical harmonic coefficients were used as a dissimilarity measure. Figure 5 presents the equal error rate (EER) and rank 1 recognition error as a function of the embedding sphere radius. The minimum EER of 12.39% is achieved at $R = 90mm$. At this embedding sphere radius, rank 1 error of 1.9608% is achieved, which remains nearly constant for $R = 90 \div 125mm$. We conclude that both recognition error measures achieve the minimum at embedding sphere radii that yield minimum embedding error.

6. DISCUSSION AND CONCLUSION

An extension of the 3DFACE system, which exploits the method of expression-invariant representation of faces was introduced. We proposed to replace the flat embedding by a spherical one. This modification was shown to be advantageous in the sense of embedding distortions. We also presented an experimental evidence that the spherical embedding has a direct consequence in terms of better recognition rates. The use of spherical canonical images allows us to perform matching in the spherical harmonic transform domain, which does not require preliminary alignment of the images.

Unlike the original methods using canonical forms and planar canonical images, which require a large amount of points for the alignment, the approach presented here appears to provide similar results with relatively sparser sampling of the facial surface. This allowed us to achieve real-time performance on a commodity AMD processor with SSE extensions.

[2]The images in the parametric plane were scaled in order to guarantee equal resolution for all embedding sphere radii. Histogram equalization was used to reduce sensitivity to illumination.

7. REFERENCES

[1] M. Bronstein A. Bronstein and R. Kimmel, "Expression-invariant 3d face recognition," in *Proc. AVBPA*. 2003, Lecture Notes on Computer Science, pp. 62–69, Springer.

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

[3] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," Tech. Rep. CIS-2004-04, Dept. of Computer Science, Technion, Israel, 2004, to appear in IJCV.

[4] M. M. Bronstein, A. M. Bronstein, and R. Kimmel, "Expression-invariant representation for human faces," *Trans. PAMI*, 2004, submitted.

[5] R. Kimmel and J. A. Sethian, "Computing geodesic paths on manifolds," *Proc. National Academy of Sciences*, vol. 95, no. 15, pp. 8431–8435, 1998.

[6] A. Spira and R. Kimmel, "An efficient solution to the eikonal equation on parametric manifolds," *Interfaces and Free Boundaries*, vol. 6, no. 3, pp. 315–327, September 2004.

[7] A. Elad and R. Kimmel, "On bending invariant signatures for surfaces," *IEEE Trans. PAMI*, vol. 25, no. 10, pp. 1285–1295, 2003.

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

[9] G. Zigelman, R. Kimmel, and N. Kiryati, "Texture mapping using surface flattening via multi-dimensional scaling," *IEEE Trans. Vis. and Comp. Graph.*, vol. 9, no. 2, pp. 198–207, 2002.

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

[11] P. Kostelec D. Healy Jr., D. Rockmore and S. Moore, "FFTs for the 2-sphere – improvements and variations," *The Journal of Fourier Analysis and Applications*, vol. 9, no. 4, pp. 341–385, 2003.

[12] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Proc. Eurographics Symp. on Geom. Proc.*, 2003.