Technion - Computer Science Department - Technical Report CIS-2005-01 - 2005

Expression-invariant representations for human faces

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

***Abstract*— An essential question in many fields that deal with the nature of facial appearance is what are the invariants of the human face under various expressions. That is, can someone's face have a unique description, that is independent of his or her facial expression. We previously suggested, without a proof, to treat facial expressions as isometries of deformable surfaces in the context of Riemannian geometry. Here we introduce, for the first time, an experiment that supports this claim. We also extend the model to handle topology-changing transformation, namely, expressions with open and closed mouth.**

***Index Terms*— Expression-invariant representation, multidimensional scaling, isometric embedding, face recognition, invariants.**

I. Introduction

THIS paper deals with the question of what are the invariants of the human face under various expressions. That is, is there a unique description to someone's face that does not change by his or her expression. Important applications of expression-invariant facial representation include the problem of face recognition in computer vision [1], [2], [3], texture mapping for facial animation in computer graphics [4], [5], emotion interpretation in psychology [6], and measurement of geometric parameters of the face in cosmetic surgery. The variability of the face appearance due to its non-rigid structure makes it a non-trivial task and challenges for a convenient model to analyze the nature of facial expressions.

Considering the face as a Riemannian two-dimensional manifold (surface) we perviously proposed to model facial expressions as *isometries*, since the facial skin can stretch only slightly [2]. Isometries preserve the geodesic distance (length of the shortest path) between any two points on the surface, and, roughly speaking, do not "stretch" or "tear" the surface. Facial expressions are therefore modelled as isometric surfaces, which are surfaces that can be obtained from some initial facial surface ("neutral expression") by means of an isometry.

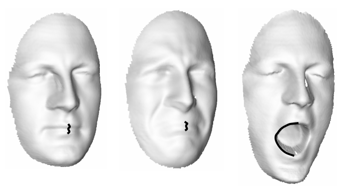
Indeed, the isometric model was shown to be useful for three-dimensional face recognition, and led to a method that can gracefully handle even extreme expressions [3]. However, so far, there was no solid experimental proof for this model, besides the face recognition results reported in [2], [3], [7] and the intuition that faces are "more isometric than rigid".

Moreover, the isometric model is limited by a tacit assumption that the *topology* of the facial surface is preserved. For example, facial expressions are not allowed to introduce arbitrary "holes" in the facial surface. This assumption is valid for most regions of the face, yet, the mouth cannot be treated by such a model. Opening the mouth changes the topology of the facial surface by virtually creating a "hole." Figure 1 demonstrates this particular property by showing one geodesic (shortest path) between two points on the upper and the lower lips. As long as the mouth is closed, the geodesic crosses the lips without any significant change of it length, even for extreme facial expressions. However, opening the mouth substantially changes the length of the geodesic. In this case, the geodesic between the two lips passes along the lip contour

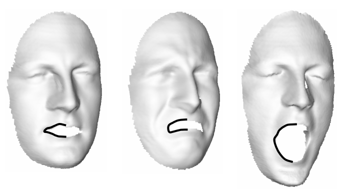
M. M. Bronstein, A. M. Bronstein and R. Kimmel are with the Department of Computer Science, Technion – Israel Institute of Technology, Haifa 32000, Israel (e-mail: bronstein@ieee.org; alexbron@ieee.org; ron@cs.technion.ac.il).

rather than across the lips.

Here, we suggest to overcome this flaw of the isometric model 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.



(a) Without topological constraint



(b) With topological constraint

Fig. 1. Illustration of the open mouth problem. (a) Geodesics on the facial surface with three facial expressions. An open mouth changes the topology of the surface, which leads to loosing the consistency of the geodesics. (b) the same face, in which the topology is constrained by cutting out the lips. Consistency of the geodesics is better preserved by this restriction.

Under the assumption of the topologically-constrained isometric model, finding an expression-invariant representation of the face is essentially equivalent to finding an isometry-invariant representation of a surface. Here, following [2], [3], we apply the method of bending invariant canonical forms, first introduced by Elad and Kimmel [8], [9]. According to this method, the Riemannian structure of the facial surface is converted into a Euclidean one by embedding the surface into a low-dimensional flat space and replacing the geodesic distances by Euclidean ones. The resulting representations ("canonical forms") can then be treated by conventional methods used for rigid surface matching.

The main emphasis of this paper is the validation of isometric model and its extended topologically-constrained version rather than extensive face recognition experiments. For the sake of completeness, we first briefly describe how to compute canonical forms of faces. Next, we provide an experimental validation of the topologically-constrained isometric model. Finally, we demonstrate how the proposed representation can be used to enhance our expression-invariant face recognition method and overcome the open mouth problem. More extensive face recognition results can be found in [7], [3].

II. Preprocessing

We assume that the 3D structure of the face is acquired by a range camera (3D scanner) and is given as a cloud of points that can be triangulated or represented as a parametric surface [10], [3]. Before computing the canonical form, the facial surface is preprocessed. The preprocessing includes three steps: (i) geometric processing of the facial surface (removal of acquisition artifacts, hole filling, selective smoothing); (ii) lip cropping; and (iii) facial contour cropping. For details on the first step see [3]. Lip cropping is performed by first segmenting the lips based on the 2D texture information, and then removing the corresponding points in the 3D data. This stage ensures that the topology of the facial surface is consistent, that is, the mouth is always "open."

Cropping the facial contour is performed using the *geodesic mask*. The key idea is locating invariant "source" points on the face (e.g. the tip of the nose) 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.

When the topologically-constrained isometric model is used (allowing both closed and open mouthes), two geodesic masks are computed: one around the nose tip and the nose apex

(Figure 2a); another one around the lips (Figure 2b). The radius of 30 mm is typically used for the lip mask. The two masks are then merged together (Figure 2c). In this way, we crop the facial surface in a geometrically consistent manner, which is insensitive to facial expressions (Figure 2). Fast Marching [11], a numerical algorithm for measuring geodesic distances on discrete surfaces, is used to compute the equidistant contours.

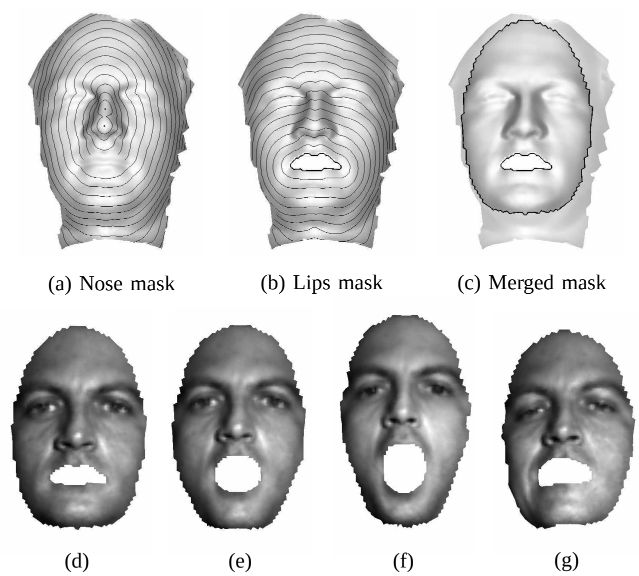


Fig. 2. Geodesic mask computation stages in case of the topologically-constrained isometric model assumption (open and closed mouth): geodesic circles around the nose and the nose apex (a); geodesic circles around the lips (b); the merged geodesic mask (c). Examples of the geodesic mask insensitivity to facial expressions (d)-(g).

III. Canonical forms of facial surfaces

According to the isometric model, facial surfaces under different facial expressions possess the same *intrinsic geometry*. Particularly, the geodesic distance $d(\mathbf{x}_1, \mathbf{x}_2)$ between any given two points $\mathbf{x}_1$ and $\mathbf{x}_2$ on the facial surface remains the same. Theoretically, the geodesic distances give a unique expression-invariant representation of an isometric face.

However, since working with a discrete set of surface samples $\mathbf{x}_1, ...\mathbf{x}_N$, there is neither guarantee that the surface is sampled at the same points, nor even that the number of points in two surfaces is the same. 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. Consequently, this make the use of such an invariant impractical.

An alternative proposed in [8] is to avoid dealing explicitly with the matrix of geodesic distances and find an representation of the original Riemannian surface as a submanifold of some convenient space (e.g. here we use a low-dimensional Euclidean space; other choices are possible as well [12], [13], [14]), such that the intrinsic geometry of the original surface is preserved. Such procedure is called *isometric embedding*, and is defied as a mapping $\varphi : (\mathcal{S}, d) \rightarrow (\mathbb{R}^m, d')$ that maps the surface samples $\mathbf{x}_1, ...\mathbf{x}_N$ into a set of points $\mathbf{x}'_1, ...\mathbf{x}'_N$ in a m-dimensional Euclidean space, such that the resulting distances $d'_{ij} = \|\mathbf{x}'_i - \mathbf{x}'_j\|_2$ equal the geodesic distances d_{ij} in some optimal way. The resulting set of points $\mathbf{x}'_i$ in the Euclidean space is called the *bending-invariant canonical form* of the face. Unlike the matrix $\mathbf{D}$, the canonical forms are invariant up to rotation, translation and reflection, and can be therefore treated by conventional algorithms used for rigid surface matching.

A. Multidimensional scaling

Unfortunately, since a "purely" isometric embedding (such that $d_{ij} = d'_{ij}$) does not exist in most cases, we have to turn to the "most isometric" embedding, the one that deforms the geodesic distances the least. The distortion introduced by such a mapping (*embedding error*) can be measured as a discrepancy between the geodesic and the resulting Euclidean distances. Finding the best approximate flat embedding is done by minimization of the embedding error. A family of algorithms used to carry out such an approximate flat embedding is usually referred to as *multidimensional scaling* (MDS) [15]. These algorithms differ in the choice of the embedding error criterion and the numerical method used for its minimization.

One straightforward possibility is to have the embedding error defined as a sum of squared differences

$$s(\mathbf{X}';\mathbf{D}) = \sum_{i>j}(d'_{ij}(\mathbf{X}') - d_{ij})^2, \quad (1)$$

where $\mathbf{X}' = (\mathbf{x}'_1, ..., \mathbf{x}'_N)$ is an $m \times N$ matrix representing the points in the embedding space, and $\mathbf{D}'$ is the matrix of mutual Euclidean distances depending on the points configuration $\mathbf{X}'$. The function (1) is referred to as *stress*. The MDS problem is posed as a least-squares (LS) problem and is known as LS MDS. We used the iterative SMACOF algorithm [15], [16] to compute the canonical forms.

When the embedding is performed into a space with $m \leq 3$ dimensions, the canonical form can be plotted as a surface. Figure 3 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.

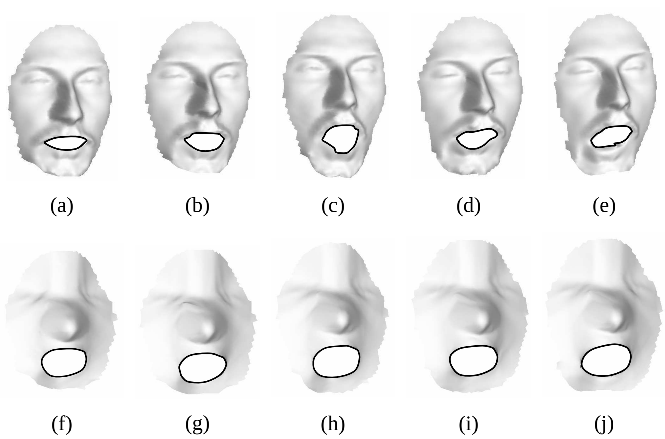


Fig. 3. Examples of topologically-constrained canonical forms (f)-(j) of faces with strong facial expressions with both open and closed mouth(a)-(e). Lip contours are emphasized.

IV. Results

In order to validate our approach, we performed two experiments. The first is a quantitative validation of the topologically-constrained isometric model. In the second experiment, we measure the sensitivity of canonical forms to facial expressions and compare them to straightforward rigid surface matching of the faces.

A. Isometric model validation

To validate the topologically-constrained isometric model, we track a set of feature points on the facial surface and measure how the distances between them change due to expressions, while preserving the topology of the surface. In this experiment, 133 white round markers (approximately 2 mm in diameter) were placed on the face of a subject as invariant fiducial points (Figure 4, left). The subject was asked to demonstrate weak, medium, and strong facial expressions, including open and closed mouth, a total of 16 faces (see some examples in Figure 4, right). The lips were cropped in all the surfaces.

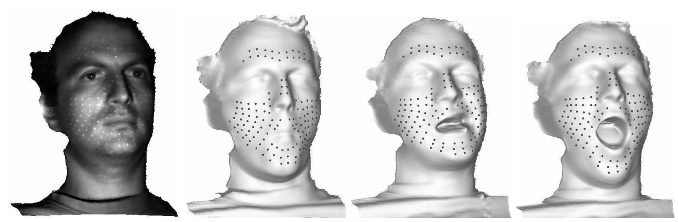


Fig. 4. Isometric model validation experiment. Left: facial image with the markers. Right: example of three facial expressions with marked reference points.

In order to quantify the changes of the distances due to facial expressions, we use two measures: the absolute and the relative error w.r.t the reference distances:

$$\begin{aligned} \epsilon_i^{\mathrm{abs}} &\equiv d_i - d_i^{\mathrm{ref}}; \\ \epsilon_i^{\mathrm{rel}} &\equiv \frac{d_i - d_i^{\mathrm{ref}}}{d_i^{\mathrm{ref}}}; \end{aligned} \quad (2)$$

$$\begin{aligned} \epsilon_i'^{\mathrm{abs}} &\equiv d'_i - d_i'^{\mathrm{ref}}; \\ \epsilon_i'^{\mathrm{rel}} &\equiv \frac{d'_i - d_i'^{\mathrm{ref}}}{d_i'^{\mathrm{ref}}}. \end{aligned} \quad (3)$$

Here d_i denotes the i-th geodesic distance and d_i^{ref} is the corresponding reference geodesic distance; d'_i and $d_i'^{\mathrm{ref}}$ denote the corresponding Euclidean distances. The reference distances were averaged on the neutral expressions.

Fig. 5. Definitions of distances in the isometric model validation experiment.

The distributions of $\epsilon^{\rm abs}$ and $\epsilon^{\rm rel}$ and their standard deviations are shown in Figure 6 and in Table I, respectively.

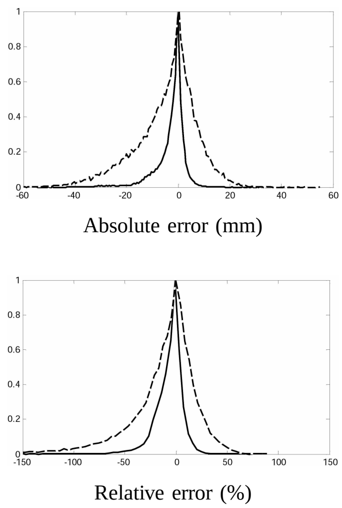


Fig. 6. Normalized histograms of the absolute and the relative error of the geodesic (solid) and the Euclidean (dotted) distances.

TABLE I

STANDARD DEVIATION OF ϵ^{abs} AND OF ϵ^{rel} FOR EUCLIDEAN AND GEODESIC DISTANCES.

	Euclidean	Geodesic
Absolute error standard dev.	12.03 mm	5.89 mm
Relative error standard dev.	39.6%	15.85%

We can conclude that the changes of the geodesic distances due to facial expressions are insignificant even for extreme expressions, which justifies our model. Moreover, the Euclidean distances were demonstrated to be much more sensitive to changes due to facial expressions compared to the geodesic ones. This observation will be reinforced by the following results, where we use our approach and straightforward rigid surface matching to measure similarity between faces.

B. *Sensitivity to facial expressions*

In the second experiment, we use a data set containing 102 instances of 6 subjects and one mannequin (`Eve`, see Figure 7). The human subjects appear in various facial expressions including both open and closed mouth (a total of 70 and 32 instances, respectively).

Expressions were classified into 7 types (neutral expression + 6 expressions with open or closed mouth) and into 3 strengths (weak, medium, strong, see Figure 9). A neutral expression is the natural posture of the face with a closed mouth; strong expressions are extreme postures that obviously rarely occur in real life. Expressions `open mouth`, `distorted` and `vowel 'u'` try to imitate facial appearance that can occur when the subject is speaking (Figure 8). Expressions `smile` and `sadness` appear both with closed and open mouth. Head rotations of up to about 10 degrees were allowed for in acquisition.

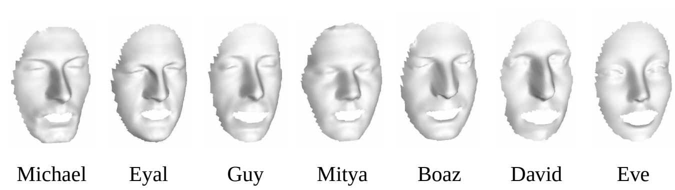


Fig. 7. Faces used in the experiment (shown with neutral expressions).

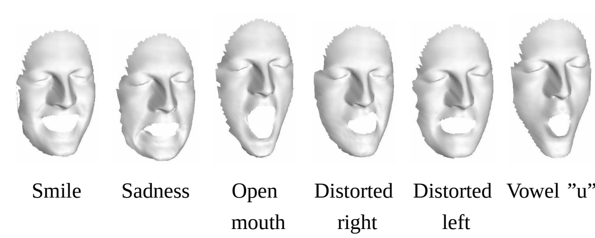


Fig. 8. Six representative facial expressions of subject Eyal.

Surface matching (both facial surfaces and canonical forms) was performed by comparing their moments [17]. The surface is described as a vector of its 56 high-order moments, and the Euclidean norm is used to measure the distance between these vectors [3].

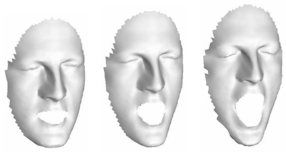


Fig. 9. Three degrees of the open mouth expression of subject Eyal.

Figure 10 depicts a low-dimensional representation (obtained by MDS) of the dissimilarities (distances) between faces. Each symbol on the plot represents a face; colors denote different subjects; the symbol's shape represents the facial expression and its size represents the expression strength. Ideally, clusters corresponding to each subject should be as tight as possible (meaning that the representation is insensitive to facial expressions) and as distant as possible from each other, which means that the representation allows us to discriminate between different subjects.

It can be seen that for rigid surface matching (Figure 10a) the clusters overlap, implying that variability due to facial expressions is larger than due to the subject's identity. On the other hand, using topologically-constrained canonical forms, we obtain tight and distinguishable clusters (Figure 10b).

Table II presents the values of the ratio of the maximum inter-cluster to minimum intra-cluster dissimilarity

$$\varsigma_k = \frac{\max_{i,j \in \mathcal{C}_k} \eta_{ij}}{\min_{i \notin \mathcal{C}_k, j \in \mathcal{C}_k} \eta_{ij}}, \quad (4)$$

($\mathcal{C}_k$ denotes indexes of k-th subject's faces, η_{ij} denotes dissimilarities between faces i and j) for facial and canonical surface matching. This criterion measures the tightness of each cluster of faces that belong to the same subject and its distance from other clusters. Ideally, ς_k should be zero. Our approach appears to outperform rigid facial surface matching by up to 790.3%.

V. Conclusions

We presented an empirical justification for treating human faces as deformable surfaces in the context of Riemannian

TABLE II

Description of the facial expressions in data set used in the test and the inter-cluster to intra-cluster dissimilarity ratio ς_k using rigid and canonical surface matching.

Subject	Color	Neut	Weak	Med	Str	Cl	Op	ς_k^{can}	ς_k^{orig}
Michael	red	2	2	10	6	6	14	2.55	17.10
Eyal	green	1	2	8	5	2	14	1.35	8.61
Guy	magenta	3	4	6	4	5	12	1.44	10.64
Mitya	yellow	2	9	7	5	6	17	2.40	14.77
Boaz	d. cyan	3	3	3	1	6	4	1.30	3.01
David	d. magenta	1	2	4	3	1	9	0.97	8.65
Eve	black	6	-	-	-	6	-	0.11	0.70

geometry. Facial expressions are thereby best modelled as isometries of the facial surface. As an example, we demonstrated (based on previous results) how an expression-invariant representation of the face can be given by a canonical form, which is an invariant of a given surface under isometries. We validated the isometry assumption of the human face in capturing facial expressions with both closed and open mouth, and thus in representing even extreme facial expressions. The resulting representation was demonstrated to be useful for 3D face recognition.

Acknowledgment

We are grateful to Eyal Gordon for help in data acquisition.

References

[1] Y. Tian, T. Kanade, and J. Cohn, *Handbook of face recognition*. Springer-Verlag, Heidelberg, 2003, ch. Facial expression analysis.

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

[3] ——, "Three-dimensional face recognition," Dept. of Computer Science, Technion, Israel, Tech. Rep. CIS-2004-04, 2004. [Online]. Available: http://visl.technion.ac.il/bron/michael

[4] F. I. Parke and K. Waters, *Computer Facial Animation*. AK Peters, 1994.

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

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

[7] M. M. Bronstein, “Three-dimensional face recognition,” Master’s thesis, Department of Computer Science, Technion - Israel Institute of Technology, 2004.

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

[9] ——, “On bending invariant signatures for surfaces,” *IEEE Trans. PAMI*, vol. 25, no. 10, pp. 1285–1295, 2003.

[10] 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, 2004.

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

[12] A. Elad and R. Kimmel, *Geometric methods in bio-medical image processing*. Berlin Heidelberg New York: Springer-Verlag, 2002, vol. 2191, ch. Spherical flattening of the cortex surface, pp. 77–89.

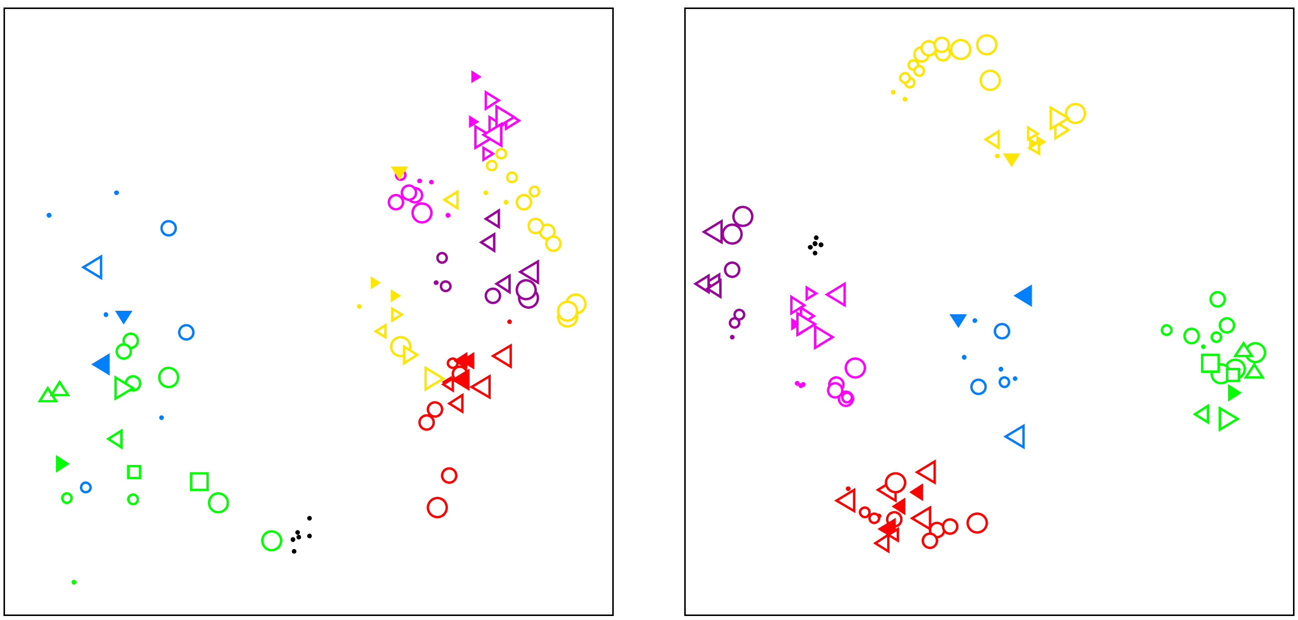
[13] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, “On isometric embedding of facial surfaces into $\mathbb{S}^3$,” in *Scale Space*, 2004, submitted.

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

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

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

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



(a) Rigid surface matching

(b) Canonical form matching

• Neutral ○ Open mouth ◁ Distorted left ▷ Distorted right
▽ Smile △ Sadness □ Vowel ”u”

Fig. 10. Low-dimensional visualization of dissimilarities between faces using rigid surface matching (left) and topologically-constrained canonical form (right). Colors represent different subjects. Symbols represent different facial expression with open mouth (empty) and closed (filled). The symbol size represents the strength of the facial expression. Some of the markers overlap and are invisible.