Isometric deformation invariant 3D shape recognition

Dirk Smeets *, Jeroen Hermans, Dirk Vandermeulen, Paul Suetens

K.U.Leuven, Faculty of Engineering, ESAT/PSI, IBBT-K.U.Leuven Future Health Department, Medical Imaging Research Center, University Hospitals Gasthuisberg, Herestraat 49 - bus 7003, B-3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 30 November 2010
Received in revised form 7 November 2011
Accepted 20 January 2012

Keywords:
Object recognition
Spectral decomposition
Isometric deformation
Face recognition
Three-dimensional face
Facial expression
Expression variation
SHREC

ABSTRACT

Intra-shape deformations complicate 3D shape recognition and therefore need proper modeling. Thereto, an isometric deformation model is used in this paper. The method proposed does not need explicit point correspondences for the comparison of 3D shapes. The geodesic distance matrix is used as an isometry-invariant shape representation. Two approaches are described to arrive at a sampling order invariant shape descriptor: the histogram of geodesic distance matrix values and the set of largest singular values of the geodesic distance matrix. Shape comparison is performed by comparison of the shape descriptors using the $\chi^2$-distance as dissimilarity measure. For object recognition, the results obtained demonstrate the singular value approach to outperform the histogram-based approach, as well as the state-of-the-art multidimensional scaling technique, the ICP baseline algorithm and other isometric deformation modeling methods found in literature. Using the TOSCA database, a rank-1 recognition rate of 100% is obtained for the identification scenario, while the verification experiments are characterized by a 1.58% equal error rate. External validation demonstrates that the singular value approach outperforms all other participants for the non-rigid object retrieval contests in SHREC 2010 as well as SHREC 2011. For 3D face recognition, the rank-1 recognition rate is 61.9% and the equal error rate is 11.8% on the BU-3DFE database. This decreased performance is attributed to the fact that the isometric deformation assumption only holds to a limited extent for facial expressions. This is also demonstrated in this paper.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

During the last decades, developments in 3D modeling and 3D capturing techniques caused an increased interest in the use of 3D objects for a number of applications, such as CAD/CAM, architecture, computer games, archaeology, medical applications and biometrics. In many of these fields, an important research problem is 3D shape retrieval, in which an object needs to be recognized from a large database of objects. One witness of the growing interest in 3D object retrieval is the yearly SHREC (Shape REtrieval Contest) contest [1], organized with the objective to evaluate the effectiveness of 3D shape retrieval algorithms.

The challenge of 3D shape recognition becomes even harder when intra-shape deformations are present in the database, as is often the case for articulating objects. Fig. 1, for example, shows some deformed shapes from the TOSCA database [2].

Three-dimensional face recognition is a domain of special interest within the 3D shape recognition research community. As a matter of fact, since the last decade of the 20th century, a shifting can be observed from 2D to 3D or 2D+3D face representations [3] for automated face recognition tasks. Three-dimensional face representations have the intrinsic advantage of not being affected by pose variations or varying lighting conditions. However, the intra-subject deformations still require special attention in automatic 3D face recognition. It has indeed been shown that, for instance, deformations due to cosmetic surgery negatively affect the performance of some state-of-the-art 3D face recognition methods [4]. However, most intra-subject deformations are due to natural intra-subject changes in facial shape, mainly changes in facial expressions. Those expression variations are reported as one of the main challenges of 3D face recognition research [5].

In this paper, a 3D shape recognition method that is invariant for isometric deformations is proposed, while not requiring explicit point correspondences for shape comparison. The 3D shapes are represented by an isometric deformation invariant matrix, i.e. a geodesic distance matrix (GDM). The latter contains the geodesic distance between each pair of points on the surface. This distance is the length of the shortest path on the object surface between two points on the object. Isometric deformations, by definition, leave these geodesic distances unchanged. The proposed method thus requires the intra-subject deformations to...
be near isometric, a hypothesis we need to verify as well. Using the geodesic distance matrix, two different shape descriptors are presented that are both invariant to the sampling order of points on the surfaces represented. As such, object recognition reduces to direct comparison of the shape descriptors without the need to establish explicit point correspondences. The first approach uses a simple histogram of the GDM values. The second shape descriptor is defined by the set of largest singular values of the GDM. It will be shown that this modal representation is an excellent shape descriptor.

This paper is organized as follows: after a discussion of related work in the remaining of this section, the 3D shape recognition method using our isometric deformation model is presented and compared with two other methods using an isometric deformation model in Section 2. The methods are validated for object recognition in Section 3.1 as well as for 3D face recognition in Section 3.2. The validity of the assumption that important intra-shape deformations are near isometric is investigated in Section 4. The recognition results are discussed and compared with other isometric deformation invariant 3D shape recognition methods in Section 5. Finally, Section 6 summarizes the main conclusions drawn from this research.

1.1. Related work

In literature, geodesic distance matrices have already been used to tackle 3D recognition problems involving non-rigid objects. Probably the best known of these contributions is the algorithm of Elad and Kimmel [6]. Here, the GDM is computed using the fast marching on triangulated domains (FMTDs) method. Subsequently, the GDM is processed using a multidimensional scaling (MDS) approach, converting non-rigid objects into their rigid invariant signature surfaces. These can be compared using standard algorithms for rigid matching. We will use an implementation of this method for comparison with our approaches.

The Geodesic Object Representation of Hamza and Krim [7] is another 3D object recognition method relying on geodesic distance matrices. In [7], GDMs are used to determine global geodesic shape functions. This global shape descriptor is defined in each point of the surface and measures the normalized accumulated squared geodesic distances to each other point on the surface. Using kernel density estimation (KDE), the global geodesic shape functions of a particular object are transformed into a geodesic shape distribution. For the actual recognition, these KDEs are compared using the Jensen–Shannon divergence.

Next, a similar method to our modal representation approach is the method shown in Jain and Zhang's work [8]. This method measures the inter-object distance by taking the $\gamma^2$-distance between the 20 largest eigenvalues of a weighted GDM of the object. In this paper, it will be shown that the weighting of the GDM has an adverse effect on the shape recognition performance.

Also in the field of expression-invariant 3D face recognition, the idea of using an isometric deformation model is not novel. Indeed, in 2003, the method of Elad and Kimmel [6] (which has been described above) was applied to 3D faces by Bronstein et al. [9]. The use of this method was justified by demonstrating that deformations due to facial expression variations are near isometric. An extension of this method is described in [10]. Instead of embedding the facial image into a Euclidean space, it is parameterized on a 2D sphere $\mathbb{S}^2$. In order to obtain a translation-, rotation- and scale-invariant shape descriptor, the nose tip is chosen to reside on the north pole of the sphere. The remaining degrees of freedom are eliminated by a spherical harmonic transform. The Euclidean norm between the transform coefficients is used as dissimilarity measure. Another extension is the partial embedding of one surface into another surface using generalized MDS (GMDS) [11,12]. GMDS maps the probe image on the model by minimizing the generalized stress, i.e. the weighted sum of differences between corresponding geodesic distances. The three-point geodesic distance approximation is developed for calculating the geodesic distance between points originally not on the model surface.

After the publication of the MDS method in [9], many researchers exploited the near invariance of geodesic distances during facial expression variations for expression invariant face recognition purposes. Berretti et al. [13] developed a method based on iso-geodesic stripes, i.e. sets of points with an equal normalized geodesic distance to a reference point located at the nose tip. The spatial relationship between intra-subject stripes is used to compare faces. Iso-geodesic curves form also the basis of the methods of Feng et al. [14] and Jahanbin et al. [15]. Similarly, Mpiiperis et al. [16] use a geodesic polar representation, in which each point on the face is characterized by the geodesic distance from the nose tip and the polar angle, compared to a reference path through the sillion. Ouji et al. [17] directly compare the geodesic distance from the nose tip to corresponding points. Correspondence is achieved with the iterative closest point (ICP) algorithm. Li and Zhang [18] and Gupta et al. [19] use pairwise geodesic distances between corresponding landmarks. Most of these methods require point correspondences at some stage during shape comparison. However, establishing these point correspondences is not a trivial task which is often performed iteratively introducing the risk of obtaining sub-optimal solutions. In contrast, the approaches proposed in this paper do not require explicit point correspondences at all.

2. Isometric deformation model

In mathematics, an isometry is a distance-preserving isomorphism between metric spaces. An isometric deformation of a submanifold $M$ in a Riemannian space $\mathcal{V}$ is therefore a deformation which preserves the length and energies of curves in $M$ [20]. For surfaces this means that isometries will bend the surface without stretching it. Isometric deformations, which reside in the field of differential geometry, have already been studied by C.F. Gauss at the beginning of the 19th century [21].
For recognition purposes, it is interesting to study geometric invariants during isometric deformations. According to Gauss’s Theorem Egregium, the Gaussian curvature $K$ of a surface, i.e. the product of the principal curvatures, is invariant under local isometry. However, as the computation of curvatures is sensitive to noise, the practical use of this first invariant is limited.

The first fundamental form of $\mathcal{M}$ is a more interesting geometric invariant for isometric deformation invariant 3D shape recognition, because its application is less sensitive to noise. In order to formalize this invariant, consider the regular patch (local surface) $f : Q \rightarrow \mathbb{R}^3$ mapping all $q = (u, v) \in Q$ to a point $x = f(q)$ on the 3D surface. The first fundamental form of $\mathcal{M}$ corresponds to the inner product of the tangent vectors and is explicitly given by the Riemannian metric $[22]$

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2,$$

with

$$E = \left[ \frac{\partial f}{\partial u} \right]^2,$$

$$F = \frac{\partial f}{\partial u} \cdot \frac{\partial f}{\partial v},$$

$$G = \left[ \frac{\partial f}{\partial v} \right]^2.$$

Integrating $ds$ over a path on the surface, determines the arc length of that curve. The shortest path between two points on the surface is called minimal geodesic and its length is called the geodesic distance. The geodesic distance $T_\mathbf{x}(x)$ between the 3D surface points $x = f(q)$ and $x_1 = f(q_1)$ can be obtained by solving the Eikonal equation

$$|\nabla T_\mathbf{x}(f(q))| = 1,$$

with boundary condition $T_\mathbf{x}(x_1) = 0$. In practice, geodesic distances are calculated by a fast marching algorithm for triangulated meshes $[23, 24]$.

2.1. Geodesic distance matrix as shape representation

Geodesic distance matrices (GDMs) are shape representations invariant for isometric object deformations. Hence, assuming isometric intra-shape deformations GDMs are appropriate to be used for shape descriptors.

A matrix $G = [g_{ij}]$ is regarded as GDM for a particular shape if its elements $g_{ij}$ correspond to the geodesic distances between the points $i$ and $j$ on the object’s surface. Using triangulated meshes as surface representations, GDMs can be computed with a fast marching algorithm $[24]$. As such, a GDM is obtained with a computational cost of $O(n^2)$, with $n$ the number of points on the surface. Fig. 2 contains an example 3D object and its associated GDM.

By definition, GDMs are symmetric and uniquely defined up to a random simultaneous permutation of their rows and columns due to the arbitrary sampling order of the surface points. Hence, $G$ and $G$ are equivalent GDMs if they are related by a permutation matrix $P$, according to

$$G' = PGP^T.$$  

Provided that two instances of the same object are represented by surface meshes containing an equal number of sufficiently dense sampled surface points, a one-to-one correspondence map (bijective map) can be assumed to exist between both surface representations. Hence, point correspondences are mathematically characterized by a permutation matrix and the geodesic distance matrices of these surface meshes are approximately related by Eq. (6). Hence, shape comparison reduces to verifying the extent to which Eq. (6) holds. However, in practice the point correspondences between the objects compared are generally not known. Since establishing explicit point correspondences between surfaces is far from trivial, this work proposes the use of shape descriptors extracted from GDMs that are invariant for simultaneous permutations of their rows and columns. Thereoto, two different methods are described in Sections 2.3 and 2.4. These algorithms will be compared with multidimensional scaling, a state-of-the-art technique for isometric deformation invariant recognition and described in Section 2.2.

The shape descriptors proposed in the following sections are not restricted to geodesic distance matrices only ($G_1 = [g_{ij}]$). Therefore they are also applied to other affinity matrices such as the squared GDM ($G_2 = [g_{ij}^2]$), the Gaussian weighted GDM ($G_3 = \exp(-g_{ij}^2/(2\sigma^2))$) and the increasing weighting function GDM ($G_4 = [1+(1/\sigma g_{ij})^{-1}]$, as defined in $[25]$).

2.2. Multidimensional scaling

Multidimensional scaling (MDS) is a technique that allows visualization of the proximity between points with respect to some kind of dissimilarity (distance) measure. Thereto, the original point configuration is mostly embedded in a lower dimensional space by a (approximately) distance preserving mapping. In the output space, Euclidean distances are generally used. In the method described in $[9]$, MDS is applied on the GDM in order to obtain a configuration of points where point-wise Euclidean distances approximately equal to the original point-wise geodesic distances. As an example, Fig. 3 shows the obtained 2D and 3D embedding after application of classical MDS on the...
object shown in Fig. 2(a). These point configurations are commonly referred to as canonical forms.

Classical MDS includes an eigendecomposition of the double-centered distance matrix

\[
- J = V \Lambda V^T,
\]

with \( J = I - \frac{1}{n} \mathbf{1} \mathbf{1}^T \) the centering matrix (\( I \) is the identity matrix, \( n \) the number of points on the surface and \( \mathbf{1} \) a \( n \times 1 \) vector with all entries equal to 1) and \( \Lambda \) the distance matrix containing squared geodesic distances. The coordinates of the canonical forms are then given by \( X = \mathbf{~V} \mathbf{~L}^{-1} / 2 \), with \( \mathbf{~L} \) a \( 2 \times 2 \) or \( 3 \times 3 \) diagonal matrix with the largest eigenvalues in order to obtain the 2D or 3D canonical form, respectively.

Because the GDM is invariant with respect to isometric transformations, the canonical forms of isometrically deformed objects have the same configuration in the embedding space. Therefore, objects can be compared by rigidly aligning their canonical forms and analyzing the registration error. Note that explicit point correspondences are required to obtain this rigid registration. Since establishing explicit point correspondences is far from trivial, the following sections propose shape descriptors which allow shape comparison without establishing point correspondences. As discussed in Section 2.1 and formalized in Eq. (6), these shape descriptors should therefore be invariant for simultaneous row and column permutations of the GDM.

2.3. Histogram representation

By transforming geodesic distance matrices into histograms of geodesic distances, a very simple shape descriptor is obtained that is invariant for simultaneous permutations of rows and columns of the GDMs. Many different possibilities can be thought of to construct these histograms. However, in this paper, only two of them are considered.

The first histogram-based shape descriptor is obtained by arranging all upper triangular values of the GDM into a histogram. The second one describes shapes using histograms of their mean geodesic distances per point (point-wise averaged (PWA) geodesic distances). As an example, both histogram representations for the object in Fig. 2 are shown in Fig. 4.

2.4. Modal representation

A second shape descriptor proposed uses a singular value decomposition (SVD) of a GDM. This elementary matrix operation transforms a GDM into a permutation-variant matrix of singular vectors and a permutation-invariant diagonal matrix. This is subsequently formalized.

Proof. Let \( P \) be a random permutation matrix, such that \( G' = P G P^T \) is a GDM with rows and columns permuted, and \( G = U \Sigma U^T \) a singular value decomposition of \( G \). Then

\[
G' = P G P^T = P U \Sigma (P U)^T. \tag{8}
\]

Because \( P U \) remains a unitary matrix and \( \Sigma \) is still a diagonal matrix with non-negative real numbers on the diagonal, the right hand side of (8) is a valid singular value decomposition of \( G' \). A common convention is to sort the singular values in non-increasing order. In this case, the diagonal matrix \( \Sigma \) is uniquely determined by \( G' \). Therefore, \( \Sigma = \Sigma' \), with \( \Sigma' \) the singular value matrix of \( G' \) according to \( G' = U \Sigma' U^T \). \( \square \)
Conceptually, the singular value decomposition separates the GDM into a diagonal matrix containing intrinsic shape information and a permutation-variant matrix containing rich feature vectors. While the former is used for shape comparison purposes, the latter could be used to establish point correspondences. For numerical reasons, only the k largest singular values are computed and retained as shape descriptors. As such, the computational complexity is limited to $O(k \cdot n^2)$, with n the dimension of the GDM. MATLAB 2009a is used for the calculation of eigenvalues.

2.5. Dissimilarity measures for shape comparison

In order to compare the shapes in a database, an appropriate dissimilarity measure has to be chosen to compare the associated shape descriptors of the two proposed approaches. In this work, the dissimilarity measures listed in Table 1 are examined. In this table, $H_{(j)}$ refers to the Shannon entropy. Moreover, $S = [s_1, s_2, \ldots, s_M]^T$ is an $M$-dimensional shape descriptor, with $j = 1, \ldots, N$ and $N$ the number of objects.

3. Recognition experiments and results

In this section, the approaches described in Section 2 are validated for general object recognition (Section 3.1) and for 3D face recognition (Section 3.2). Both for general object recognition and 3D face recognition, two types of recognition experiments are considered: identification and verification. These scenarios are shortly described in the following paragraphs, along with their commonly used performance measures.

The identification scenario involves a one-to-many comparison of objects. Specifically, a probe object is compared to many objects in a gallery. In this work it is assumed that each gallery contains exactly one example object per object class considered and that each probe object belongs to exactly one object class included in the gallery. Hence, each probe object matches with exactly one gallery object. Object identification involves retrieval of the matching gallery object for a given probe. Hence, using a shape dissimilarity measure, the gallery objects are ranked...
Therefore, we participated in the "SHREC 2010" and "SHREC 2011"Shape Retrieval Contest of Non-rigid 3D Models" challenge in 2010 [27] and the "SHREC 2011—Shape Retrieval Contest of Non-rigid 3D Watertight Meshes" in 2011 [28].

According to increasing shape dissimilarity with the probe object. Provided that an appropriate shape dissimilarity measure is used, the matching gallery object should reside on one of the first places in this shape dissimilarity ranking. The performance of a shape descriptor in the identification case can be characterized by the performance of a shape dissimilarity measure used. Plotting the rank N recognition rates for increasing N, results in the cumulative matching curve (CMC).

In the verification or authentication scenario a one-to-one object comparison is performed to decide whether two objects belong to the same or different object classes. Thereto a threshold on the shape dissimilarity measure is introduced. When the shape dissimilarity value is below or above this threshold, the objects are considered to belong to the same (genuine) or different (imposter) object classes respectively. The performance of a shape dissimilarity measure in the verification scenario is quantified by the receiving operator curve.

In the subsequent validation experiments, the methods proposed are compared with the state-of-the-art MDS technique as well as with a baseline recognition algorithm. The latter corresponds to the standard iterative closest point (ICP) algorithm [26]. This is a well-known and extensively used rigid shape registration algorithm that minimizes the sum of squared Euclidean distances between closest points. After rigid registration, the RMS registration error can be used as shape dissimilarity measure. Opposed to the algorithms of Section 2, the baseline ICP algorithm performs rigid shape comparison only.

### 3.1. Object recognition

To validate the different shape matching approaches for general 3D objects, we use the "TOSCA non-rigid world" database [2]. The TOSCA database is a clean dataset and is used to determine the influences of the different parameters. The best performing method on this database is then externally validated. Therefore, we participated in the "SHREC 2010—Shape Retrieval Contest of Non-rigid 3D Models" challenge in 2010 [27] and the "SHREC 2011—Shape Retrieval Contest of Non-rigid 3D Watertight Meshes" in 2011 [28].

#### 3.1.1. TOSCA non-rigid world

**Data and preprocessing:** The TOSCA database is intended as a benchmark dataset for the objective comparison of 3D object recognition methods and point matching algorithms involving non-rigid objects [2]. Thereto, it consists of various 3D non-rigid shapes in a variety of poses. In our experiments, we use 133 objects, including 9 cats, 11 dogs, 3 wolves, 17 horses, 21 gorillas, 1 shark, 6 centaurs, 6 seahorses, 24 female figures, and two different male figures, containing 15 and 20 poses. Each object contains approximately 3000 vertices. As the objects considered have closed surfaces, no region of interest needs to be determined. Moreover, because all objects are represented with approximately the same number of points, there is no need for re-sampling.

As mentioned in Section 3, both identification and verification experiments are conducted. For the identification experiment, the gallery is constituted of the first object of each class. This results in 122 probe objects to be compared with the gallery. Regarding the verification experiments, all objects are compared with each other resulting in $133 \times 132 = 8911$ different shape comparisons.

Before comparing the 3D object recognition methods with each other and with the baseline ICP method, the best set of influencing design parameters is determined for each method.

**Parameter and configuration tuning:** for the histogram approach of Section 2.3 the optimal number of histogram bins must be selected, while for the modal approach of Section 2.4 the optimal number of singular values must be determined. For both algorithms the best GDM weighting $G_i$ (Section 2.1) must be chosen as well as the dissimilarity measure $D_i$ used for shape comparison (Section 2.5). This parameter tuning is performed by conducting aforementioned identification and verification experiments for all possible combinations. The optimal values, $\eta$, along with the identification and verification results, are tabulated in Table 2. The meaning of the parameter $\eta$ depends on the method considered. For the histogram approach it corresponds to the number of bins, while for the modal representation $\eta$ represents the number of singular values.

As previously mentioned, the GDM-based object recognition methods proposed are compared with the state-of-the-art MDS approach. As to obtain a fair comparison, both the classical MDS (CMDS) and the least-squares MDS using the SMACOF algorithm as implemented by Bronstein [2] are considered. After

---

Fig. 6. The $R_{RR}$ (a) and EER (b) is plotted against the number of eigenvalues used in the shape descriptor. This illustrates the stability of the algorithm proposed w.r.t. the number of eigenvalues used.
determination of the 3D canonical shapes, these need to be aligned to enable shape comparison. Thereto, both the iterative closest point (ICP) algorithm [26] and moment matching (MM) are considered. The actual shape dissimilarity is expressed in terms of RMS registration error and the Euclidean distance between higher order moments when ICP or MM is used for object alignment respectively. The results of these MDS combinations are given in Table 3.

Results: After tuning each of these object recognition algorithms, their performance is compared for both experiments considered. Fig. 5(a) shows the CMC of the identification experiment conducted, while Fig. 5(b) displays the ROC for the verification experiment. Besides the GDM-based object recognition results, the recognition performance of the baseline ICP algorithms is included. As the rank-1 recognition rate ($R_{1\text{RR}}$) and the equal error rate ($EER$) are characteristic points on the CMC and ROC respectively, these values are listed in Table 4 for the different methods.

Both Fig. 5 and Table 4 clearly show the baseline rigid ICP algorithm performs considerably worse than any isometric shape comparison approach considered. Moreover, the modal representation outperforms all other methods described. As to illustrate the stability of its performance, Fig. 6 plots the $R_{1\text{RR}}$ (a) and the $EER$ (b) against the number of singular values used in the shape descriptor. Both identification and verification results remain stable when the number of singular values used is between 50 and 200.

Fig. 17 shows some qualitative results that illustrate the consistent correct recognition obtained with the modal representation.

Correspondences. As suggested in Section 2.4, point correspondences can be extracted from the eigenvector matrices, by finding the permutation from $U = PU$. However, since the eigenvectors are defined up to a sign, sign flips occur. Moreover, due to numerical issues eigenvector can change position. This is called eigenvector flip. To account for these sign and eigenvector flips Mateus et al. [29] proposes a histogram approach together with an EM-based matching framework. To demonstrate the proof-of-concept, we implemented the histogram approach that matches eigenvectors or their sign flipped version based on their histogram. After correction for sign and eigenvector flip, the permutation is found by matching the rows in $U$ and $U'$. An example of correspondences found by this approach is shown in Fig. 7.

![Fig. 7. An example of correspondences found by matching the eigenvector matrices after correction for sign and eigenvector flip.](image)

Table 5
Official shape retrieval results on the SHREC 2011 dataset [27].

<table>
<thead>
<tr>
<th>Authors</th>
<th>Method</th>
<th>NN (%)</th>
<th>FT (%)</th>
<th>ST (%)</th>
<th>E (%)</th>
<th>DCG (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smeets et al.</td>
<td>Modal repr.</td>
<td>99.5</td>
<td>78.8</td>
<td>94.4</td>
<td>68.0</td>
<td>96.1</td>
</tr>
<tr>
<td>Ohbuchi and Furuya</td>
<td>BF-DSIFT-E</td>
<td>98.0</td>
<td>76.6</td>
<td>89.2</td>
<td>64.5</td>
<td>94.1</td>
</tr>
<tr>
<td>Wuhrer and Shu</td>
<td>CF</td>
<td>98.0</td>
<td>63.5</td>
<td>78.0</td>
<td>55.3</td>
<td>87.8</td>
</tr>
</tbody>
</table>

![Fig. 8. Validation of the proposed method for the SHREC 2010 dataset. The CMC of (a) summarizes the results of the identification experiment, while the ROC in (b) shows the verification results.](image)

Please cite this article as: D. Smeets, et al., Isometric deformation invariant 3D shape recognition, Pattern Recognition (2012), doi:10.1016/j.patcog.2012.01.020
3.1.2. Shape retrieval contest 2010

Data and preprocessing: since the modal representation outperformed the other methods considered on the TOSCA database, this algorithm is also validated by participating in the “SHREC 2010—Shape Retrieval Contest of Non-rigid 3D Models” [27]. The used dataset is a subset of the McGill 3D Shape Benchmark [30], containing 200 non-rigid objects, including 20 ants, crabs, hands, humans, octopuses, pliers, snakes, spectacles, spiders and teddies each. The objective of this 3D Shape Retrieval Contest is to evaluate the effectiveness of 3D-shape retrieval algorithms for each. The objective of this 3D Shape Retrieval Contest is to evaluate the effectiveness of 3D-shape retrieval algorithms for non-rigid 3D objects [27]. Unlike the object instances of TOSCA, these are the same class as the probe. The first tier (FT) and second tier (ST) are the percentage of models belonging to the probe's class that appear within the top k ranked models, where k = n_c - 1 for the first tier and k = 2(n_c - 1) with n_c the size of the probe's class. The E-measure (E) is a composite measure of precision and recall. The discounted cumulative gain (DCG) is a statistic that measures the usefulness of a model based on its position in the ranked list. For the further details regarding the meaning of these measures, the reader is referred to [1,31].

Parameter and configuration tuning: As SHREC is intended as an external validation contest, no ground truth information was available for the competitors. Hence, an extensive parameter and configuration tuning of the method applied was not possible. Therefore, the modal shape matching approach used corresponds with the configuration that optimally performed on the TOSCA database. However, in order to account for the intrinsic shape differences within an object class, the number of singular values used is reduced to 19. This value was empirically determined.

Results: Fig. 8 gives the results of the validation and indicates an EER of 8.86% and a R_{RR} of 88.95%. This is a lower performance compared to the TOSCA database, which can be explained by the small intrinsic shape variations present between objects of the same class in this dataset.

In order to enable standardized external validation, SHREC prescribes the use of same predefined performance measures. For reasons of completeness, these are shown in Table 5. The nearest neighbor (NN) is defined as the percentage of closest matches that are the same class as the probe. The first tier (FT) and second tier (ST) are the percentage of models belonging to the probe's class that appear within the top k ranked models, where k = n_c - 1 for the first tier and k = 2(n_c - 1) with n_c the size of the probe's class. The E-measure (E) is a composite measure of precision and recall. The discounted cumulative gain (DCG) is a statistic that measures the usefulness of a model based on its position in the ranked list. For the further details regarding the meaning of these measures, the reader is referred to [1,31].

3.1.3. Shape retrieval contest 2011

Data and preprocessing: The modal representation is again validated by participating in the “SHREC 2011—Shape Retrieval Contest of Non-rigid 3D Watertight Meshes” [28]. This dataset is larger than the SHREC 2010 dataset, containing 600 non-rigid objects, recreated and modified from several publicly available databases such as the McGill database [30], TOSCA shapes [2] and the Princeton Shape Benchmark [31]. All meshes are down-sampled, keeping 3000 points. To compensate for scale differences, each object's GDM is normalized with the square root of the total surface area of the mesh (instead of the sum of all its elements).

Parameter and configuration tuning: also for SHREC 2011, no ground truth information was available for the competitors.

Table 7

<table>
<thead>
<tr>
<th>Experiment</th>
<th>R_{RR} (%)</th>
<th>EER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modal representation</td>
<td>61.87</td>
<td>11.84</td>
</tr>
<tr>
<td>ICP</td>
<td>71.63</td>
<td>13.23</td>
</tr>
</tbody>
</table>

Fig. 9. The surface of the left face (a) is cropped by only keeping vertices within a geodesic distance of 80 mm from the nose tip (b).

Fig. 10. Results of standard validation experiments on the BU-3DFE database with CMC (a) and ROC (b). Face recognition with a baseline algorithm (thin solid line) is compared to face recognition using modal representation (thick solid line).
Therefore, the modal shape matching approach used corresponds with the configuration that optimally performed on the TOSCA database. However, in order to account for the intrinsic shape differences within an object class, the number of singular values used is chosen to be 40.

Results: Table 6 shows the results of the SHREC 2011 contest, having strong field of participants. The modal representation provides the best results on all evaluation criteria.

3.2. Face recognition

3.2.1. BU-3DFE database

Data and preprocessing: In order to validate the performance of the modal shape matching method for face recognition in the presence of expression variations, it is applied to a subset of the Binghamton University 3D Facial Expression (BU-3DFE) database [32]. In this experiment, 900 different facial surfaces are considered, which originate from 100 subjects. While only 11% of these facial surfaces correspond to neutral expressions, the other 89% either have an angry or a sad expression, subdivided into four different levels of facial expression. Most facial surfaces had a closed mouth.

Because the modal shape matching approach assumes that each surface is constituted of corresponding points, the surfaces considered must represent the same facial area. As this prerequisite is not met for all faces in the dataset, some preprocessing steps are required. This involves detection of the nose tip and removal of all surface vertices with a geodesic distance above a certain threshold from the nose. The geodesic cropping is shown in Fig. 9(b) for a cutoff threshold of 80 mm.

In this paper, a very simple nose detection method is used. As for most faces in the dataset the gaze direction is approximately aligned with the z-axis, the nose tip mostly corresponds with the vertex with largest z-value. Only in 2% of the cases, a manual correction was needed. Although falling without the scope of this work, nose detection can be robustly automated with [33,34].

Finally, the surfaces are re-sampled as to be constituted by an equal number of points. For these experiments, 2000 facial vertices are randomly selected, as a trade-off between computational complexity and accuracy of the representation. The result is a normalized 3D face region that contains the same face area for every 3D surface represented by an equal number of points. To apply the modal shape matching approach proposed, a fast marching algorithm for meshes is used to compute the GDM for each face in the dataset.

Parameter and configuration tuning: Again, the best dissimilarity measure and the optimal number of eigenvalues is found by an extensive tuning. Using the Jensen–Shannon divergence ($D_{JS}$) with 189 eigenvalues gives the best results for the identification and verifications scenario.

Results: Fig. 10 shows the results of this approach for the identification and verification scenario for the optimal configuration, and the results of the baseline ICP algorithm. The equal error (EER) and rank-1 recognition rate ($R_{1}, RR_{1}$) are characteristic points on the ROC and CMC respectively and are listed in Table 7.

Both Fig. 10 and Table 7 show a similar performances for the baseline algorithm and the modal representation. To illustrate the stability of the performance of the modal representation, Fig. 11 plots the $R_{1}, RR_{1}$ (a) and the EER (b) against the number of singular values used in the shape descriptor. This experiment is repeated four times for subsets containing the neutral scan and all scans with a certain, manually annotated expression strength. The recognition performance clearly drops when emotions are more expressed (higher level of expression strength).

4. Validity of the isometric deformation assumption

The proposed methods require the intra-subject deformations to be near isometric. In this section, the validity of the isometric deformation assumption is examined for 3D articulating objects and 3D faces with expression variations.
4.1. 3D articulating objects

To examine the deviation to the isometric deformation assumption in a realistic situation when dealing with 3D articulating objects, we looked at the change in geodesic distance between four finger tips in three situations with different configuration of the hand (an articulating object), as shown in Fig. 12. This results in a mean coefficient of variation (CV) of 5.3% for the geodesic distances, while the CV for Euclidean distances is equal to 27.6%. This clearly indicates the benefit of using an isometric deformation model for recognition.

4.2. 3D faces

To validate the isometric deformation assumption for faces during expression variation, two experiments are setup.

Fig. 13. The geodesic distance is calculated from the nose tip to 21 other facial landmarks on the Bosphorus database [35] to examine its variation. The first six faces have different expressions, the last two have a neutral expression.

Fig. 14. The histogram of relative (a) and absolute (b) variation of geodesic and Euclidean distances between the nose tip and 21 other facial landmarks is calculated based on 303 expression scan of 80 subjects.

Fig. 15. The figure shows the setup of the experiment to validate the isometric deformation assumption for faces.
The first experiment examines the variation of the geodesic distance from the nose tip to 21 anatomical landmarks, as shown in Fig. 13, and compares the variation with the variation in Euclidean distance. Therefore, the Bosphorus database, including manually indicated landmarks, is used [35]. Most landmarks are located on the more rigid parts of the human face.

The results of the experiment are given in Fig. 14. The histogram of relative (Fig. 14(a)) and absolute (Fig. 14(b)) variation of geodesic and Euclidean distances between the nose tip and 21 other facial landmarks is calculated based on 303 expression scans of 80 subjects. Almost no difference between the distribution of geodesic and Euclidean distances can be noticed.

The second experiment considers the variation of geodesic distances between each pair of landmarks on the face, again for the Bosphorus database [35] with the landmarks mainly on the rigid part of the face, as well as for an own database with landmarks mainly on the non-rigid part of the face. For landmarks on the non-rigid part, a larger difference between variation of geodesic and Euclidean distances is expected.

On 171 expression scans of the Bosphorus database corresponding to 4 subjects, the mean coefficient of variation (CV) is 5.98% for the geodesic distances between all pairs of 22 landmarks, while the mean CV for the Euclidean distances is 6.16%.

To examine the variation for landmarks located on the less rigid parts of the face, we set up an experiment, in which 47 fiducials are manually indicated on subject’s face as shown in Fig. 15. Then, 18 3D scans are taken, 4 neutrals and the other with different (strong) expressions, using a structured light scanner. To compute the geodesic distances appropriately, the mouths need to be cut open. This is done by manually indicating eight points on the lips and computing the shortest geodesic path between the points in the correct order. The mesh is then cut along this path.

The results of this last experiment are shown as a normalized histogram of relative (Fig. 16(a)) and absolute (Fig. 16(b)) variation of geodesic and Euclidean distances using the distances on the average of all neutral scans as reference considering the distances between each pair of the 47 facial fiducials. The standard deviation of the relative variation of geodesic and Euclidean distances is 11.30% and 11.21% respectively. The variation, expressed as mean coefficient of variation (CV), for the geodesic distances is 8.25%, for the Euclidean distances 8.14%. The CV is higher than for the Bosphorus database because here more landmarks are chosen on the flexible part of the face.

4.3. Conclusion

The experiments, validating the isometric deformation assumption, clearly show that the assumption holds to a much smaller extent for 3D faces with expression variations compared to 3D articulating objects. Moreover, the stronger the expression, the more variation of geodesic distances can be noticed.

5. Discussion

The main advantage of the methods proposed in this paper is that, except for the preprocessing step in case of 3D face recognition, the algorithms do not determine any explicit point correspondence between the different surfaces. The spectral decomposition captures the intrinsic shape information in the diagonal matrix, independent of the sampling order and histograms are by definition sampling order independent. On the other hand, one could mention that it is hard to satisfy the condition that each point needs to have a corresponding point on each other facial surface of the same subject. However, when enough points on the same facial area are considered, there will always be a point close to the (virtual) corresponding point. The most important disadvantage of the proposed modal representation, however, is the sensitivity to topological changes. This occurs for example in non-rigid objects when two hands are merging, or in 3D faces when the mouth opens.

Object recognition experiments on the TOSCA shape database have clarified the outperformance of the approach with modal representation over the histogram approaches, the state-of-the-art multidimensional scaling technique and the baseline ICP algorithm, having a rank-1 recognition rate of 100% and an equal error rate of 1.58% using the 128 largest eigenvalues. This result proves the effectiveness of describing shapes that are represented by closed surfaces by the largest eigenvalues of an affinity matrix. For rigid objects a Euclidean distance matrix could suffice. However, when isometric deformations are present in the data, the geodesic distance matrix representation is preferable.

Although the algorithms are extensively compared to the state-of-the-art multidimensional scaling method and the ICP baseline algorithm, it is still useful to compare the results with other methods using an isometric deformation model, validated on the same database. Bronstein et al. [36] report an EER of 16.32% for the multidimensional scaling (MDS) algorithm and an EER of 15.49% for the generalized multidimensional scaling (GMDS)
algorithm, which is explained in Section 1.1 and in more detail in [11]. Although these methods also use the GDM representation, their performance is clearly lower. By applying MDS on this GDM, a canonical form in dimension $m$ is created. The canonical forms are compared using rigid registration. If the MDS is done by classical scaling, an eigenvalue decomposition is performed on the squared GDM after double-centering (see Eq. (7)). Due to the double-centering, the invariance of the diagonal matrix (with the eigenvalues) does not hold anymore when permuting the points on the surface. If $m=3$, only the three largest eigenvalues are taken and some intrinsic shape information is discarded, which might be another explanation for the lower performance.

Validation in the “SHREC 2010—Shape Retrieval Contest of Non-rigid 3D Models” [27] demonstrates an excellent performance of the modal representation approach on more noisy data. It outperforms the method of Ohbuchi and Furuya, which renders

![Fig. 17. Some visual results show the consistent correct recognition. The first column represent the probes, the second to fourth the three closest matches with their dissimilarity to the probe and the last column shows the correct match.](image-url)
range images from the 3D models. Recognition is done using a bag-of-features approach with SIFT features extracted from the range images. The singular value approach also outperforms the method of Wuhrer and Shu, which strongly resembles the least-squares MDS approach which is also validated here on the TOSCA method of Wuhrer and Shu, which strongly resembles the least-range images from the 3D models. Recognition is done using a and Godil constructs a canonical shape using MDS after which the performing approach (based on most evaluation criteria) of Lian performs better than all other participants. The second best recognition is done using a.

For 3D face recognition, the performances are clearly lower than for object recognition, with a rank-1 recognition rate of 61.9% and an equal error rate of 11.8%. The main reason for this discrepancy is probably the difference of validity of the isometric deformation assumption between articulating objects and 3D faces with the presence of expression variations. While the variation of Euclidean distances is much larger than the variation of geodesic distances for the articulating hand, no clear difference can be noticed for 3D faces with expressions. This last finding is in contradiction with the measurements by Bronstein et al. [38], where a clear discrepancy is noticed between variation of geo-desic and Euclidean distances. However, also in [38] a standard deviation of 15.85% on the relative variation is reported (we found 11.30%). Following [38], the standard deviation of the relative change in geodesic distance was found to be about 15%. Therefore, the isometric deformation model is an approximation. Moreover, without additional processing to disconnect upper and lower lip in the surface representation before determination of the geodesic distances, it is not valid for faces with open mouth. Also occlusions result in a miscalculation of geodesic distances. These aspects come down to the assumption that a bijective map that preserves distances on the surface exists between the surfaces. They can be seen as the general disadvantages of methods using an isometric deformation model. However other strategies for expression-invariant methods, not relying on the isometric assumption, have also disadvantages that are not present in methods using the isometric deformation model. Methods using statistical models to model expression variation always need a training stage to construct the model. Therefore, if no representative training data are used, the recognition performance will decrease. The region-based methods, that perform face recognition on the more rigid parts of the face, do not use all available information by throwing away those parts that are affected by expressions. This leads to loss of information that could be discriminative.

Due to the approximation of the isometry assumption, the facial area extracted by the geodesic cropping (described in Section 3.2.1) is only approximately the same for faces from the same subject. These differences in facial area clearly influence the global shape descriptor (modal representation) and can therefore be more decisive than subtle shape differences. Moreover, occlusions and missing data violate the assumption of the existence of a bijective map between the two surfaces and therefore deteriorate the modal representation too.

Again, it is interesting to compare the proposed modal representation approach with other expression invariant face recognition methods using an isometric deformation model and validated on the same database. Mpiperis et al. [16] developed a face recognition method, which is based on an isometric deformation model using the geodesic polar representation. Instead of calculating pairwise geodesic distances, geodesic distances from the nose tip to all other points are calculated to construct a geodesic polar parameterization. In [16], this method is compared with another method using an isometric deformation model developed by Bronstein et al. [38], Thereto, again a subset of the BU-3DFE database [32] of 1600 images is used. Mpiperis et al. reported an EER of 9.8% for their method and 15.4% for the method of [38]. A partial explanation for the better performance is the use of color information and a special method for handling the open mouth problem in [16]. It is again particularly interesting to compare the method with the related method described in [39] (which is the basis of [38]). Here, MDS is applied on the GDM representation of the face generating a 3D canonical form. Thus, only the three largest eigenvalues of the squared double-centered GDM are taken and some intrinsic shape information is discarded. This might by an explanation for the higher EER of 15.4% in [16] than the EER of 11.8% of the method that is described here.

6. Conclusions

In this work, we proposed two methods for isometric deformation invariant 3D shape recognition. Both are based on the geodesic distance matrix (GDM) as an isometric deformation invariant shape representation. Because this matrix is defined up to a random simultaneous permutation of rows and columns, two GDM permutation invariant approaches are proposed. The first calculates a histogram on the GDM, while the second computes the eigenvalues (singular values). The methods are validated for object and 3D face recognition in an identification and a verification scenario.

A better performance is noticed for the modal representation method for the object recognition scenario in comparison to the histogram method, the state-of-art MDS method and the baseline algorithm. For this experiment a rank-1 recognition rate of 100% and an equal error rate of 1.58% are reported for the 128 largest singular values. The method also outperforms other isometric deformation modeling methods found in literature. For 3D face recognition, the performance is lower probably because the isometric deformation assumption holds true to a much smaller extent.

The main advantage of the proposed methods is their independence to correspondence information. The singular value decomposition captures some intrinsic shape information in the diagonal matrix, independent of the sampling order while the histogram is sampling order independent by definition.

As future work, we propose to further exploit the modal decomposition method in order to obtain correspondences between different objects. This can be done using the eigenvectors or singular vectors based on the method of Brady and Shapiro [40]. To improve the performance for 3D face recognition in more realistic databases, such as FRGC v2 [41] or SHREC [42], mesh preprocessing algorithms need to be implemented to remove spikes and holes. Handling outliers and missing data requires intervention in the eigendecomposition of the distance matrix by giving more weight to inliers and overlapping regions, respectively.

Acknowledgments

This work is supported by the Flemish Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT Vlaanderen), the Research Program of the Fund for Scientific Research—Flanders (Belgium) (FWO) and the Research Fund K.U.Leuven.

Please cite this article as: D. Smeets, et al., Isometric deformation invariant 3D shape recognition, Pattern Recognition (2012), doi:10.1016/j.patcog.2012.01.020
References

[1] AIM@SHAPE. SHREC—3D shape retrieval contest. [http://www.aimatshape.net/event/SHREC/]


Paul Suetens is professor of Medical Imaging and Image Processing and chairman of the Medical Imaging Research Center at the University Hospital Leuven. He is also Head of the Division for Image and Speech Processing at the Department of Electrical Engineering of K.U.Leuven. He is author of more than 400 peer-reviewed papers in international journals and conference proceedings and author of the book “Fundamentals of Medical Imaging” (2 editions). He is involved in several graduate (Biomedical Engineering, Biomedical Sciences) and postgraduate programs (Advanced Medical Imaging, Biomedical Engineering). He is associate editor of IEEE Transactions of Medical Imaging, and has been a member of several research councils and commissions (POC, Research Council, IOF, FWO). He was chairman of the Board of Management of MEDICIM, a spin-off of the K.U.Leuven (Medicim was acquired by Nobel Biocare in 2008).

Dirk Vandermeulen obtained a Masters Degree in Computer Science in 1983 and a Ph.D. in Electrical Engineering in 1991, both at the Katholieke Universiteit Leuven, Belgium. His initial research concentrated on image processing for computer-assisted stereotactic neurosurgery. Some of this work was implemented in a commercially available stereotactic planning system. His work then moved to medical image analysis, still with a strong focus on neurosurgical and neurological applications. He became assistant professor in 1998 and associate professor in 2004 at the Division for Image and Speech Processing at the Department of Electrical Engineering, K.U.Leuven, where he supervises research work on model-based image analysis, with a strong emphasis on image registration based approaches. Since 2002 he is involved in applying computer vision and medical image analysis techniques to forensic imaging, more in particular cranio-facial reconstruction and biometric authentication. He published more than 70 journal papers and over a hundred conference articles.