Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence

Towards Topological-Transformation Robust Shape Comparison: A Sparse Representation Based Manifold Embedding Approach

Longwen Gao and **Shuigeng Zhou**
Shanghai Key Lab of Intelligent Information Processing, and School of Computer Science
Fudan University, Shanghai 200433, China
{lwgao, sgzhou}@fudan.edu.cn

Abstract

Non-rigid shape comparison based on manifold embedding using Generalized Multidimensional Scaling (GMDS) has attracted much attention for its high accuracy. However, this method requires that shape surface is not elastic. In other words, it is sensitive to topological transformations such as stretching and compressing. To tackle this problem, we propose a new approach that constructs a high-dimensional space to embed the manifolds of shapes based on sparse representation, which is able to completely withstand rigid transformations and considerably tolerate topological transformations. Experiments on TOSCA shapes validate the proposed approach.

Introduction

Shape comparison is one of fundamental tasks of computer vision. It has been extensively studied in a variety of computer vision related subareas such as face recognition (Zhao et al. 2003; Bronstein, Bronstein, and Kimmel 2007), and some interdisciplinary areas such as computer games (Müller et al. 2005), artificial intelligence (Wolter and Latecki 2004), and biomedical image analysis (Hawkins, Skillman, and Nicholls 2007; Syeda-Mahmood, Beymer, and Wang 2007). In these contexts, shape comparison faces two challenges: less human interference and shape deformation. The first challenge requires the comparison process to be automatic and accurate. The second challenge requires the comparison method to be robust to different transformations, which is the focus of this study.

Automatically comparing 3-dimensional (3D) shapes while considering shape transformations has been studied in the literature. Transformations increase the difficulty of comparing or matching shapes. Most existing works dealing with rigid transformations match shapes by searching for a rigid transformation that maximizes the similarity between the shapes (Potmesil 1979). As non-rigid deformations are not easy to be formulated as rigid transformations, methods proposed for rigid transformations cannot work well with non-rigid deformations. One milestone of non-rigid shape comparison is to consider shapes as manifolds and embed them onto a plane or a sphere, and then compare the embedded shapes (Schwartz, Shaw, and Wolfson 1989; Mémoli and Sapiro 2005; Bronstein, Bronstein, and Kimmel 2007). Different from embedding methods, shape matching methods via minimum distortion correspondence finding search for a point correspondence that minimizes the distortion of the two shapes (Gold and Rangarajan 1996; Rodolà et al. 2013). Those methods above assume that the geodesic distance between any two points is unchanged. When comparing 3D shapes satisfying this assumption, e.g. the TOSCA dataset (Bronstein, Bronstein, and Kimmel 2006a; 2008), they can achieve a rather high accuracy (Elad and Kimmel 2003; Bronstein, Bronstein, and Kimmel 2007; Raviv et al. 2010). Despite the high accuracy, their shortcomings are obvious, since the assumption on geodesic distance can be easily broken by topological transformations such as stretching and compressing. Geodesic distance distortion will accumulate during computation and thus substantially affects the comparison result. When dealing with shapes made of elastic materials in real world, it is desirable that the method can tolerate or even resist topological transformations.

In this paper, we combine manifold embedding and sparse representation (SR) to construct a high-dimensional space for embedding shapes, instead of roughly embedding them into a 2-dimensional (2D) plane (Schwartz, Shaw, and Wolfson 1989) or a 3D sphere (Bronstein, Bronstein, and Kimmel 2007) and so on. The model of sparse representation (Yuan et al. 2006; Sprechmann, Bronstein, and Sapiro 2012) is naturally related to manifold structures, because it tends to select points in linear subspaces that construct the manifolds (Elhamifar and Vidal 2011). Furthermore, we ensure that the model selects only the neighbors and thus obtains the actual local structures. We show that the selection process of our model is an embedding, and by detecting and preserving only neighborhood information, our model can withstand any rigid transformation, and considerably tolerate topological transformations such as stretching and compressing. Contributions of this paper are as follows:

1. An improved selection model of sparse representation (SR) is proposed to effectively support shape comparison.
2. A high dimensional space for manifold embedding is constructed based on the improved SR model, which is proven to be robust against topological transformations.
3. An effective approximation of shape similarity is de-

veloped to efficiently support shape comparison.

4. Experiments on TOSCA shapes are conducted, which show that our approach outperforms both GMDS and the latest elastic net constraints based method.

Topological Transformation Robust Manifold Embedding

In this section we specify a neighborhood space for manifold embedding, whose dimensionality is the number of vertices of the original manifold. The embedding process here is called *manifold neighborhood representation* (MNR) (Elhamifar and Vidal 2011). For each vertex in the original manifold, this process selects its neighbors in a low-dimensional affine subspace and assigns each of them a real value. In what follows, we first detail the manifold neighborhood representation process, and then present some propositions that reveal its properties such as topological transformation robustness.

Manifold neighborhood representation

As in (Elhamifar and Vidal 2011), given a set S of N 3D points sampled from a manifold, we try to select a few neighbors for each of them. For a point $i \in S$ whose coordinate vector is $\boldsymbol{X}^i$, we assign a real number $\boldsymbol{C}^i_j$ to each other point j ($j \neq i$), so that the bigger the absolute value is, the better it fits in the local affine subspace of point i. We specify $\boldsymbol{C}^i_i = 0$. Then we can get a coefficient vector $\boldsymbol{C}^i = [\boldsymbol{C}^i_1, \cdots, \boldsymbol{C}^i_{i-1}, 0, \boldsymbol{C}^i_{i+1}, \cdots, \boldsymbol{C}^i_N]^\top \in \mathcal{R}^{N\times 1}$. Here, we use subscripts and superscripts to denote row and column indices of a matrix, respectively.

The *neighborhood selection process* is an optimization problem that minimizes the following objective function with respect to the coefficient vector $\boldsymbol{C}^i$:

$$\begin{aligned} &\min_{\boldsymbol{C}^i} \lambda \left\|\boldsymbol{Q}(i)\boldsymbol{C}^i\right\|_1 + \frac{1}{2}\left\|\boldsymbol{Y}(i)\boldsymbol{C}^i\right\|_2^2 \\ &s.t.\ \mathbf{1}^\top \boldsymbol{C}^i = 1, \end{aligned} \tag{1}$$

where $\boldsymbol{Y}(i)$ is a column-normalized matrix, whose column vectors are evaluated as follows:

$$\begin{aligned} &\boldsymbol{Y}(i)^j = \frac{\boldsymbol{X}^j - \boldsymbol{X}^i}{\left\|\boldsymbol{X}^j - \boldsymbol{X}^i\right\|_2}, j \neq i, \\ &\boldsymbol{Y}(i)^i = \mathbf{0}. \end{aligned} \tag{2}$$

$\boldsymbol{Q}(i)$ is a positive-definite diagonal matrix whose diagonal elements are the distances from point i to the other points. Here, we use geodesic distance for simplicity. So $\boldsymbol{Q}(i)^k_k$ means the geodesic distance from point k to point i.

In Eq. (1), the first part is a regularization term to ensure that bigger absolute coefficients are assigned to neighbors closer to a given point, and the second part measures the regression error to ensure that the selected neighbors fit in as much as possible the low-dimensional affine subspace of the given point. An assumption behind Eq. (1) is that the sampling density is high enough to ensure exactly one single subspace of any point of a manifold.

Note that the work (Elhamifar and Vidal 2011) is doing clustering, its model should be able to distinguish manifolds lying closely in the same space. So it requires a rather high sampling density, and does not concern outliers. However, in shape comparison, manifolds are embedded one by one, and noise points such as outliers should be taken into consideration. Thus, we do not require that each point must be inside the convex hull of its neighbors. The reasons are: first, we need not worry about the interference of some close points in another manifold, and second, representation without the sum constraint is more general and can handle outliers. Therefore, we propose a new model by throwing away the sum constraint as follows:

$$\min_{\boldsymbol{C}^i} \lambda \left\|\boldsymbol{Q}(i)\boldsymbol{C}^i\right\|_1 + \frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{C}^i - \boldsymbol{X}^i\right\|_2^2. \tag{3}$$

Here, we also omit the multipliers for all dimensions of $\boldsymbol{C}^i$ in the regression error term.

Considering the large number of variables in Eq. (3), we use the Limited-memory BFGS (LBFGS) algorithm (Liu and Nocedal 1989) to solve the optimization problem. Let f be the objective function in the form of Eq. (3), we define $\boldsymbol{s}_{k-1} = [\boldsymbol{C}^i]_k - [\boldsymbol{C}^i]_{k-1}$, $\boldsymbol{y}_{k-1} = \nabla f([\boldsymbol{C}^i]_k) - \nabla f([\boldsymbol{C}^i]_{k-1})$, $\rho_k = 1/\boldsymbol{y}_k^\top \boldsymbol{s}_k$ and $\boldsymbol{V}_k = \boldsymbol{I} - \rho_k \boldsymbol{y}_k \boldsymbol{s}_k^T$. The solving process follows the recursive formula in the inverse BFGS (Dennis and Schnabel 1983):

$$[\boldsymbol{C}^i]_{k+1} = [\boldsymbol{C}^i]_k + \alpha_k \boldsymbol{d}_k, \tag{4}$$

where α_k satisfies the Wolfe conditions (Wolfe 1969), and the search direction $\boldsymbol{d}_k$ is defined as

$$\boldsymbol{d}_k = -\boldsymbol{H}_k \nabla f(\beta_k). \tag{5}$$

We start with the symmetric and positive definite matrix $\boldsymbol{H}_0$ as an approximation of the inverse Hessian of f, and iteratively compute $\boldsymbol{H}_{k+1}$ by applying m updates to $\boldsymbol{H}_0$ in LBFGS as follows:

$$\begin{aligned} &\boldsymbol{H}_{k+1} = (\boldsymbol{V}_k^\top \cdots \boldsymbol{V}_{k-\hat{m}}^\top)\boldsymbol{H}_0(\boldsymbol{V}_{k-\hat{m}} \cdots \boldsymbol{V}_k) \\ &+ \rho_{k-\hat{m}}(\boldsymbol{V}_k^\top \cdots \boldsymbol{V}_{k-\hat{m}+1}^\top)\boldsymbol{s}_{k-\hat{m}}\boldsymbol{s}_{k-\hat{m}}^\top(\boldsymbol{V}_{k-\hat{m}+1} \cdots \boldsymbol{V}_k) \\ &+ \rho_{k-\hat{m}+1}(\boldsymbol{V}_k^\top \cdots \boldsymbol{V}_{k-\hat{m}+2}^\top)\boldsymbol{s}_{k-\hat{m}+1}\boldsymbol{s}_{k-\hat{m}+1}^\top(\boldsymbol{V}_{k-\hat{m}+2} \cdots \boldsymbol{V}_k) \\ &+ \cdots \\ &+ \rho_k \boldsymbol{s}_k \boldsymbol{s}_k^\top. \end{aligned} \tag{6}$$

Above, $\hat{m} = \min\{k, m-1\}$, and m is the maximum number of stored pairwise updates $\{\boldsymbol{y}_j, \boldsymbol{s}_j\}_{j=k-\hat{m}}^k$, which is set to 100 by default in our experiments.

Properties

Although some existing works use sparse representation to detect neighborhood, this work is the first to use sparse representation for manifold embedding. For example, in (Elhamifar and Vidal 2011) the authors used sparse representation only for neighborhood detection, then a normal embedding method was applied to the graph constructed with the neighborhood information. However, in this work, the model in Eq. (3) can be seen as an embedding, namely a *structure-preservation and injective mapping* from Euclidean space

$\mathcal{X}$ (where the original manifold lies) to coefficient space $\mathcal{C}$ (where the image manifold lies). In what follows, we present some properties of the model in Eq. (3), under the assumption that sampling density is high enough to maintain the local affine subspace of each point of a given manifold.

Proposition 1. *Given points $\boldsymbol{X}^i$ and $\boldsymbol{X}^j$ in $\mathcal{X}$, and their corresponding coefficient vectors $\boldsymbol{C}^i$ and $\boldsymbol{C}^j$ in $\mathcal{C}$, if $\boldsymbol{X}^j$ is in the neighborhood of $\boldsymbol{X}^i$, then $\boldsymbol{C}^j$ is in the corresponding neighborhood of $\boldsymbol{C}^i$ and $\boldsymbol{C}^j \neq \boldsymbol{C}^i$. And if there exists a point set S and another point $j \notin S$, such that $\boldsymbol{X}^j$ can be linearly represented by points $\{\boldsymbol{X}^i\}_{i\in S}$, then $\boldsymbol{C}^j$ can be linearly represented by the corresponding coefficient vectors $\{\boldsymbol{C}^i\}_{i\in S}$.*

Proof. Since manifolds are metric spaces, the neighborhood of $\boldsymbol{X}^i$ can be seen as a metric ball B_ε. Thus $\left\|\boldsymbol{X}^i - \boldsymbol{X}^j\right\| \leqslant \varepsilon$. Now we try to prove that $\left\|\boldsymbol{C}^i - \boldsymbol{C}^j\right\|$ is non-zero and bounded so that $\boldsymbol{C}^j$ is in a metric ball of $\boldsymbol{C}^i$.

According to the sampling density assumption, we can rewrite the model (3) into a constrained formula:

$$\begin{aligned} &\min_{\boldsymbol{C}^i} \sum_{k\neq i} \boldsymbol{Q}(i)_k^k \left|\boldsymbol{C}_k^i\right| \\ &s.t.\ \boldsymbol{X}\boldsymbol{C}^i = \boldsymbol{X}^i. \end{aligned} \tag{7}$$

Obviously, different right-hand vector $\boldsymbol{X}^i$ results in different optimal solutions, otherwise the constraint will not be satisfied.

This constrained optimization problem (7) can be further formulated into a linear optimization by introducing new variables $\boldsymbol{z}^+ - \boldsymbol{z}^- = \boldsymbol{C}^i$ (Bertsimas and Tsitsiklis 1997):

$$\begin{aligned} &\min_{\boldsymbol{z}^+,\boldsymbol{z}^-} \sum_{k\neq i} \boldsymbol{Q}(i)_k^k (\boldsymbol{z}_k^+ + \boldsymbol{z}_k^-) \\ &s.t.\ \boldsymbol{X}\boldsymbol{z}^+ - \boldsymbol{X}\boldsymbol{z}^- = \boldsymbol{X}^i \\ &\quad \boldsymbol{z}^+, \boldsymbol{z}^- \geqslant \boldsymbol{0}. \end{aligned} \tag{8}$$

Assume that we reach an optimal solution $\boldsymbol{C}^i = \boldsymbol{z}^+ - \boldsymbol{z}^-$ of the problem (8) for point i. To deal with point j, we need to change the right-hand vector $\boldsymbol{X}^i$ to $\boldsymbol{X}^j$ and the cost vector $diag\{\boldsymbol{Q}(i)\}$ to $diag\{\boldsymbol{Q}(j)\}$. By the triangle inequality of geodesic distance we get:

$$\boldsymbol{Q}(i)_k^k - \left\|\boldsymbol{X}^i - \boldsymbol{X}^j\right\| \leqslant \boldsymbol{Q}(j)_k^k \leqslant \boldsymbol{Q}(i)_k^k + \left\|\boldsymbol{X}^i - \boldsymbol{X}^j\right\|. \tag{9}$$

Since $\left\|\boldsymbol{X}^i - \boldsymbol{X}^j\right\| \leqslant \varepsilon$, we have:

$$\left|\boldsymbol{Q}(i)_k^k - \boldsymbol{Q}(j)_k^k\right| \leqslant \left\|\boldsymbol{X}^i - \boldsymbol{X}^j\right\| \leqslant \varepsilon. \tag{10}$$

That is, the right-hand vector and the cost vector do not change much. By sensitivity analysis (Bertsimas and Tsitsiklis 1997), if ε is small enough, the basis matrix of this linear optimization problem (8) will not change, and therefore the difference between optimal solutions $\boldsymbol{C}^i$ and $\boldsymbol{C}^j$ is small.

The proof of the second part of this proposition is straightforward. Assume

$$\boldsymbol{X}^j = \sum_{i\in S} \alpha_i \boldsymbol{X}^i, \tag{11}$$

for each point $i \in S$, we multiply its constrained equation in convex optimization (7) by α_i and sum them up, we have

$$\sum_{i\in S} \alpha_i \boldsymbol{X}\boldsymbol{C}^i = \sum_{i\in S} \alpha_i \boldsymbol{X}^i = \boldsymbol{X}^j. \tag{12}$$

By extracting the common factor $\boldsymbol{X}$ from the left part and replacing the right part with $\boldsymbol{X}\boldsymbol{C}^j$, we have

$$\boldsymbol{X}(\sum_{i\in S} \alpha_i \boldsymbol{C}^i) = \boldsymbol{X}\boldsymbol{C}^j. \tag{13}$$

Thus the proposition is proved. □

The proposition above indicates that the representation process in Eq. (3) is an injective and structure-preservation (*i.e.* preserving the structure of manifold) mapping, in other words, an embedding that embeds a manifold into a high dimensional coefficient space by sparse representation. One advantage of such embedding is that it keeps the relative spatial location information. So we have the following Proposition 2.

Proposition 2. *The embedded image of the original manifold is invariant with respect to (w.r.t.) rigid transformations such as rotation and translation, i.e., the model output $\boldsymbol{C}$ is invariant w.r.t. any rigid transformation $\boldsymbol{T} \in \mathcal{R}^{3\times 3}$ on $\boldsymbol{X}$.*

Proof. Checking model (7), we are to prove that both the cost vector $diag\{\boldsymbol{Q}(i)\}$ and the constraint do not change under any rigid transformation.

First, by definition, the distance between any two points stays the same under any rigid transformation, so we can conclude that the cost vector $\boldsymbol{Q}(i)_k^k$ stays the same for any i and k based on our definition of $\boldsymbol{Q}(i)$.

Second, upon a rigid transformation, the optimization constraint in Equation (7) becomes

$$\boldsymbol{T}\boldsymbol{X}\boldsymbol{C}^i = \boldsymbol{T}\boldsymbol{X}^i. \tag{14}$$

This constraint equation always holds, given any non-zero rigid deformation $\boldsymbol{T}$ and any feasible solution of the original model (7).

Thus, the model output $\boldsymbol{C}$ will not change upon any rigid transformation. □

The next proposition reveals the model's another advantage: robust against topological transformations. Here, topological transformations mean any spatial transformation of points such that their neighborhoods keep unchanged. As we address points on a manifold, we specify that the "neighborhood" of a point spans its local affine subspace. Assume that space $\mathcal{T}$ consists of manifolds topologically transformed from an original manifold $\boldsymbol{X}$, and space $\mathcal{A}$ consists of the corresponding output matrices (coefficient matrices) of model (3). Since identical transformation is a trivial topological transformation, we have $\boldsymbol{X} \in \mathcal{T}$ and its corresponding model output matrix $\boldsymbol{C} \in \mathcal{A}$.

Proposition 3. *If a topologically transformed manifold $\boldsymbol{Y} \in \mathcal{T}$ is in the neighborhood of $\boldsymbol{X}$, its model output matrix $\boldsymbol{D}$ is in the corresponding neighborhood of $\boldsymbol{C} \in \mathcal{A}$.*

Proof. Consider the neighborhood of $\boldsymbol{X}$ to be a metric ball B_ε. Thus we have $\|\boldsymbol{X}-\boldsymbol{Y}\| \leqslant \varepsilon$. Now we try to prove that $\|\boldsymbol{C}-\boldsymbol{D}\|$ is bounded, i.e., $\boldsymbol{D}$ is in a metric ball of $\boldsymbol{C}$.

For simplicity, we assume that after the transformation, only one point's spatial location changes (say point l). Thus,

$$\|\boldsymbol{Y}-\boldsymbol{X}\| = \left\|\boldsymbol{Y}^l - \boldsymbol{X}^l\right\|. \tag{15}$$

Based on the sampling density assumption, the optimization problem could be rewritten into the following linear optimization:

$$\begin{aligned} &\min_{\boldsymbol{z}^+,\boldsymbol{z}^-} \sum_{k\neq i} \boldsymbol{Q}(i)^k_k(\boldsymbol{z}^+_k + \boldsymbol{z}^-_k) \\ &s.t. \sum_{j\neq i,l} \boldsymbol{X}^j(\boldsymbol{z}^+_j - \boldsymbol{z}^-_j) + \boldsymbol{Y}^l(\boldsymbol{z}^+_l - \boldsymbol{z}^-_l) = \boldsymbol{X}^i, \\ &\quad \boldsymbol{z}^+, \boldsymbol{z}^- \geqslant \boldsymbol{0}. \end{aligned} \tag{16}$$

Above, the right-hand vector does not change, but one column of the left-hand matrix changes. According to the notations of linear optimization, we define the associated basis matrix as $\boldsymbol{B}$, and let $\boldsymbol{c}$ be a cost vector and $\boldsymbol{c}_B$ be the cost vector of the basic variables. For each variable, the reduced cost $\bar{\boldsymbol{c}}$ can be written as

$$\bar{\boldsymbol{c}}_k = \boldsymbol{c}_k - \boldsymbol{c}_B^\top \boldsymbol{b}^{-1}\boldsymbol{X}^k. \tag{17}$$

If the variable corresponding to point l is nonbasic, as long as the condition $\bar{\boldsymbol{c}}_l \geq 0$ holds, this variable can not be brought into the basis (Bertsimas and Tsitsiklis 1997). Replacing $\boldsymbol{X}^l$ with $\boldsymbol{Y}^l$, we have

$$\begin{aligned} 0 &\leqslant \boldsymbol{c}_l - \boldsymbol{c}_B^\top \boldsymbol{b}^{-1}\boldsymbol{Y}^l \\ &= \boldsymbol{c}_l - \boldsymbol{c}_B^\top \boldsymbol{b}^{-1}(\boldsymbol{X}^l + \boldsymbol{Y}^l - \boldsymbol{X}^l) \\ &= \bar{\boldsymbol{c}}_l - \boldsymbol{c}_B^\top \boldsymbol{b}^{-1}(\boldsymbol{Y}^l - \boldsymbol{X}^l). \end{aligned} \tag{18}$$

That is,

$$\boldsymbol{c}_B^\top \boldsymbol{b}^{-1}(\boldsymbol{Y}^l - \boldsymbol{X}^l) \leqslant \bar{\boldsymbol{c}}_l. \tag{19}$$

Thus, if we impose a bound on ε such that $\left\|\boldsymbol{Y}^l - \boldsymbol{X}^l\right\|$ is very small, the above inequality can be satisfied, and therefore the basis is still optimal.

Similarly, for the case that the variable corresponding to point l is basic, there exists an interval that guarantees an invariant basis (Bertsimas and Tsitsiklis 1997). According to the triangle inequality of geodesic distance and Eq. (15), we have

$$\left\|[\boldsymbol{Q}(i)^l_l]_{new} - [\boldsymbol{Q}(i)^l_l]_{old}\right\| \leqslant \left\|\boldsymbol{Y}^l - \boldsymbol{X}^l\right\| = \|\boldsymbol{Y}-\boldsymbol{X}\| \leqslant \varepsilon. \tag{20}$$

So this proposition is proved by sensitivity analysis (Bertsimas and Tsitsiklis 1997). □

Distance between Embedded Manifolds

The model above — using sparse representation to model the local affine subspace, can be used in the framework of manifold embedding based shape comparison as an embedding that selects a topological transformation robust manifold space, or used directly as a local shape descriptor in similarity calculation. Considering the high time cost of the manifold embedding framework, here we propose an efficient method to accurately approximate the distance between two manifolds.

Distance Definition

In last section, we consider only one manifold that is sampled into N points. In addition, now we consider another manifold that is also sampled into N points. Both are subject to transformations. We are to measure the distance between the two manifolds. The idea is that if there exists a perfect matching between the two sets of sampled points of the two manifolds, the distance between the two manifolds is defined as the $\mathcal{L}_{2,1}$-norm of the two corresponding model output matrices:

$$dist(\boldsymbol{C},\boldsymbol{D}) = \left\|\boldsymbol{P}C\boldsymbol{P}^\top - D\right\|_{2,1}, \tag{21}$$

where $\boldsymbol{C}$ and $\boldsymbol{D}$ are the model output matrices, the columns of which correspond to the points of the manifolds. $\boldsymbol{P}C\boldsymbol{P}^\top$ is a transformation of C to make the points in C be arranged in the order similar to that of the corresponding points in D. Here, the row transformation operator $\boldsymbol{P}$ can be represented as the product of some elementary row operations (Meyer 2000), each of which swaps two rows of matrix $\boldsymbol{C}$, that is,

$$\boldsymbol{P} = \boldsymbol{P}(i_1,j_1)\boldsymbol{P}(i_2,j_2)\cdots\boldsymbol{P}(i_n,j_n). \tag{22}$$

The following proposition reveals that the distance in Eq. (21) is well defined.

Proposition 4. *$dist$ defined in Eq. (21) is a distance.*

Proof. Here we prove only the triangle inequality since the other two properties of distance are obvious.

Given three output matrices $\boldsymbol{C}$, $\boldsymbol{D}$ and $\boldsymbol{E}$, we have the operator $\boldsymbol{P}_1$ from $\boldsymbol{C}$ to $\boldsymbol{D}$ and the operator $\boldsymbol{P}_2$ from $\boldsymbol{D}$ to $\boldsymbol{E}$. Thus the operator from $\boldsymbol{C}$ to $\boldsymbol{E}$ is

$$\boldsymbol{P}_3 = \boldsymbol{P}_2\boldsymbol{P}_1. \tag{23}$$

By the definition in Eq. (21), we have

$$\begin{aligned} dist(C,E) &= \left\|\boldsymbol{P}_2\boldsymbol{P}_1 C(\boldsymbol{P}_2\boldsymbol{P}_1)^\top - E\right\|_{2,1} \\ &= \left\|\boldsymbol{P}_2\boldsymbol{P}_1 C\boldsymbol{P}_1^\top\boldsymbol{P}_2^\top - \boldsymbol{P}_2 D\boldsymbol{P}_2^\top + \boldsymbol{P}_2 D\boldsymbol{P}_2^\top - E\right\|_{2,1} \\ &\leqslant \left\|\boldsymbol{P}_2(\boldsymbol{P}_1 C\boldsymbol{P}_1^\top - D)\boldsymbol{P}_2^\top\right\|_{2,1} + \left\|\boldsymbol{P}_2 D\boldsymbol{P}_2^\top - E\right\|_{2,1} \end{aligned} \tag{24}$$

Since elementary row operations do not change the $\mathcal{L}_{2,1}$-norm of a matrix, we have,

$$\left\|\boldsymbol{P}_2(\boldsymbol{P}_1 C\boldsymbol{P}_1^\top - D)\boldsymbol{P}_2^\top\right\|_{2,1} = \left\|\boldsymbol{P}_1 C\boldsymbol{P}_1^\top - D\right\|_{2,1}. \tag{25}$$

Applying the equality above to Eq (24), we have

$$\begin{aligned} dist(C,E) &\leqslant \left\|\boldsymbol{P}_1 C\boldsymbol{P}_1^\top - D\right\|_{2,1} + \left\|\boldsymbol{P}_2 D\boldsymbol{P}_2^\top - E\right\|_{2,1} \\ &= dist(\boldsymbol{C},\boldsymbol{D}) + dist(\boldsymbol{D},\boldsymbol{E}). \end{aligned} \tag{26}$$

Thus the triangle inequality is proved. □

Above, we assume that the point matching between the two manifolds is known. But finding a dense correspondence is technically difficult and time-consuming, which usually involves human labors for constructing an initial matching. For simplicity, we relax the distance definition to an approximate version as follows:

$$dist_{approx}(\boldsymbol{C},\boldsymbol{D}) = \min_{\boldsymbol{P}} \left\|\boldsymbol{P}C\boldsymbol{P}^\top - D\right\|_{2,1}. \tag{27}$$

Note that the relaxed definition is no longer a strict distance. And to compute the approximate distance, we have to find the row operator $\boldsymbol{P}$ by solving Eq. (27), which may not indicate a perfect correspondence between the two sets of sampled points of the two manifolds.

Optimization problem solving

Solving the optimization problem in Eq. (27) is actually to find a pairwise point matching of two manifolds. However in this equation, the operator P acts on both rows and columns (corresponding to $\boldsymbol{P}$ and $\boldsymbol{P}^\top$ respectively), which differs from the common bipartite graph matching problem. So we first transform the optimization problem in Eq. (27) to a bipartite graph matching problem.

We decompose the two matrices $\boldsymbol{C}$ and $\boldsymbol{D}$ using QR decomposition (Golub and Van Loan 2012), so we have

$$\boldsymbol{C} = \boldsymbol{Q}_1\boldsymbol{R}_1, \boldsymbol{D} = \boldsymbol{Q}_2\boldsymbol{R}_2. \tag{28}$$

By the property of QR decomposition (Stewart 2000), row operator $\boldsymbol{P}$ on $\boldsymbol{C}$ acts on rows of $\boldsymbol{Q}_1$, and column operator $\boldsymbol{P}^\top$ acts on columns of $\boldsymbol{R}_1$. Considering the uniqueness of the QR decomposition result, the optimal $\boldsymbol{P}$ satisfies the following two conditions,

$$\begin{aligned} \boldsymbol{P} &= \arg\min_{\boldsymbol{P}} \|\boldsymbol{P}\boldsymbol{Q}_1 - \boldsymbol{Q}_2\|_{2,1}, \\ \boldsymbol{P} &= \arg\min_{\boldsymbol{P}} \left\|\boldsymbol{R}_1\boldsymbol{P}^\top - \boldsymbol{R}_2\right\|_{2,1}. \end{aligned} \tag{29}$$

Minimizing either of them can lead to the optimal $\boldsymbol{P}$. But for a better numerical result, we combine the two conditions together by defining the edge weight between point i in the first manifold and point j in the second manifold as follows:

$$w_{ij} = \left\|\boldsymbol{Q}_{1i} - \boldsymbol{Q}_{2j}\right\|_{2,1}^2 + \left\|\boldsymbol{R}_1^i - \boldsymbol{R}_2^j\right\|_{2,1}^2. \tag{30}$$

Thus we construct a bipartite graph with N points on each side and $N \times N$ edges.

Now we are to find a best matching of the bipartite graph above. Since N can be thousands, it is too expensive to use the Hungarian algorithm (Kuhn 1955). Here we simply use a greedy algorithm that adjusts the initial matching iteratively to reduce the cost function as much as possible.

Performance Evaluation

The 3D shapes used in experiments are extracted from the TOSCA dataset (Bronstein, Bronstein, and Kimmel 2006a; 2008), including images of various gestures of human, cat, dog, centaur and so on. We use the simplest human body named David as the reference (also called the standard shape), and match it with other shapes of different gestures from a cat, a centaur, David, a dog, a male body named Michael (almost the same with David) and a female body named Victoria. All are shown in Fig. 1.

We compare our approach with the manifold embedding method using GMDS (Bronstein, Bronstein, and Kimmel 2006b) and the latest shape matching method using *elastic net constraints* (Rodolà et al. 2013). For simplicity, we denote the two compared methods as GMDS and ENC (the abbreviation of *elastic net constraints*) in the sequel. In our approach, we use the distance defined in Eq. (27) to measure performance. For GMDS, we embed one manifold into another and then compare the two manifolds. We use the maximum distortion of geodesic distance as an indicator of the distance between the two manifolds (Raviv et al. 2010). The smaller the indicator value is, the more similar the two manifolds are. For ENC, we set the tradeoff parameter α=0.65 and use the adopted Gromov-Wasserstein metric as the distortion (Rodolà et al. 2013). Our goal is to check how distortion (for GMDS and ENC) and distance (for our approach) change in response to different shape transformations.

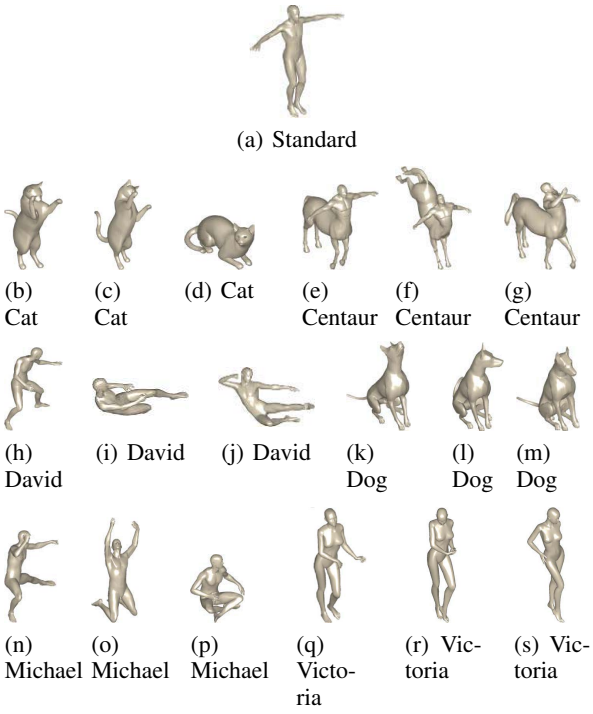

Figure 1: Standard human body (a) and the other 6 shapes with different gestures (b)∼(s).

First, we test some common and regular transformations, including scaling and stretching. Even for such transformations, GMDS and ENC can not tolerate because the geodesic distance between any two points is stretched under such deformations. However, stretching and scaling invariance is desirable in shape recognition as human eye is quite robust against such transformations. For scaling transformation, we set the *scale* value from 1.00 (no scaling) to 1.30 (30% scaling) and the results are presented in the 1st column of Fig. 2. For stretching transformation, we assume that stretching is formed by two forces acting on these shapes along the z-axis in two opposite directions. We also set the *scale* value from 1.00 (no stretching) to 1.30 (30% stretching), the results are presented in the 2nd column of Fig. 2. As the scale of transformation increases, distortions of GMDS and ENC change substantially, which means that GMDS and ENC is highly sensitive to scaling and stretching. For different shapes, distortion change of GMDS seems randomly, while distortion

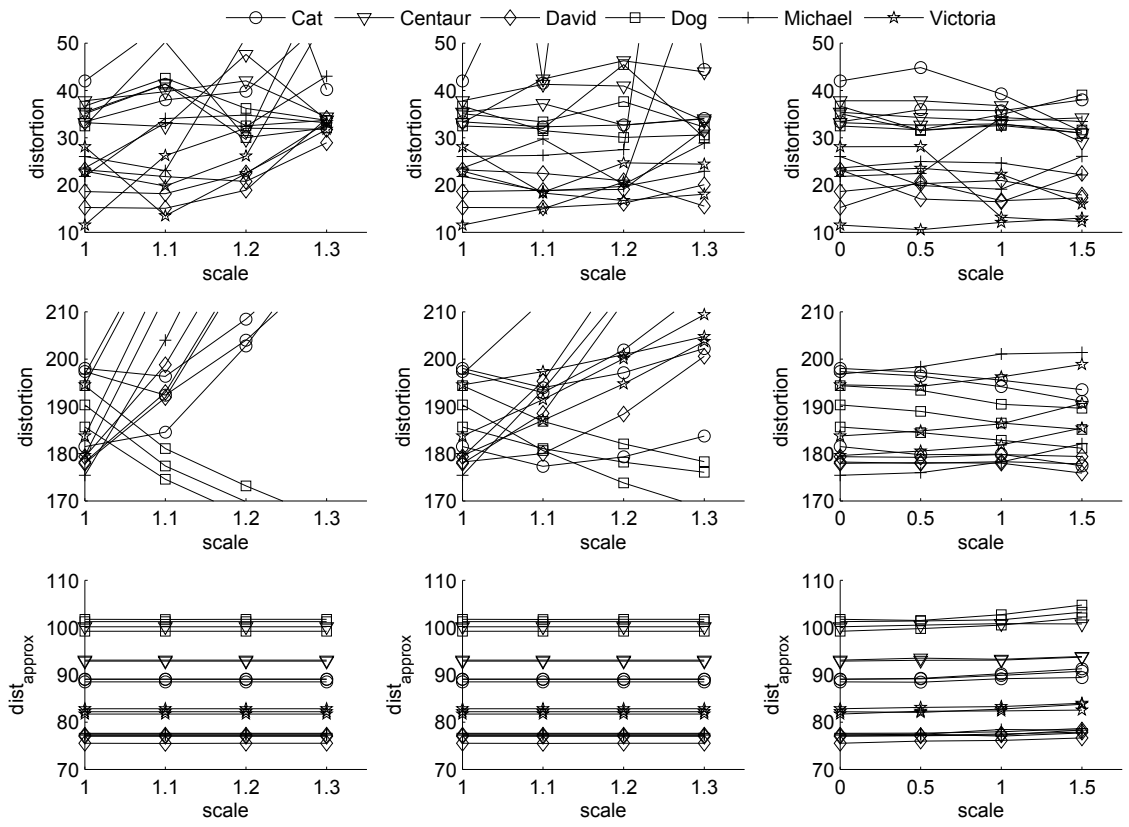

Figure 2: Robustness comparison over shapes under scaling, stretching and random topological transformations. The 1st, 2nd and 3rd rows correspond to the experimental results of GMDS, ENC and our approach, respectively; The 1st, 2nd and 3rd columns correspond to the experimental results of scaling, stretching and random topological transformations, respectively. In the figures of GMDS, we show only distortion values between 10 and 50, and in the figures of ENC, we show only distortion values between 170 and 210. Distortion values beyond these ranges are not displayed.

change of ENC shows some trends: except dog shapes that decrease monotonically, the other shapes increase monotonically or first go down and then go up. On the contrary, the distance of our approach is robust against the two kinds of transformations. Furthermore, except that the curve of one Centaur shape mixes with those of the dog shapes and Michael's curves mix with David's curves, the five clusters of curves clearly correspond to the five types of shapes: cat, Centaur, dog, David and Victoria.

Then, we test topological transformation robustness by imposing the shapes with topological transformation noise. For each shape to be matched, we repeat the following steps 3400 times to manually construct disturbed shapes: 1) randomly select a point on the manifold, 2) move the point along a randomly-selected direction up to a randomly-generated distance between 0 and 1.5. Experimental results are shown in the 3rd column of of Fig. 2. As we can see, the distance of our approach is stable while the distortion of GMDS is sensitive to the disturbance. However, in this case, ENC works better than GMDS, this is because that ENC extends the absolute criterion of distortion in GMDS to a relaxed proximity between shapes (Rodolà et al. 2013). Nevertheless, distortions defined by geodesic distance in GMDS and ENC are not robust to topological transformations. In summary, our approach outperforms both GMDS and ENC in resistance to the interference of different transformations.

Conclusion

In this paper, we propose a novel embedding via sparse representation that preserves local affine subspace and is robust against topological transformations. In addition to theoretical proofs, we also conduct experiments to examine the robustness of our approach under different transformations including scaling, stretching and random topological transformations. Experimental results have shown that our approach is effective and significantly outperforms the embedding method using GMDS and the lastest shape matching method using elastic net constraints in resistance to regular and topological transformations.

Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC) under grant No. 61272380, and the Innovation Research Program of Shanghai Municipal Education Commission under grant No. 13ZZ003.

Shuigeng Zhou is the corresponding author of this paper. All correspondence should be addressed to him at the School of Computer Science, Fudan Universitry, 220 Handan Road, Shanghai 200433, China.

References

Bertsimas, D., and Tsitsiklis, J. N. 1997. *Introduction to linear optimization*. Athena Scientific Belmont.

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

Bronstein, A. M.; Bronstein, M. M.; and Kimmel, R. 2006b. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *Proceedings of the National Academy of Sciences of the United States of America* 103(5):1168–1172.

Bronstein, A. M.; Bronstein, M. M.; and Kimmel, R. 2007. Expression-invariant representations of faces. *IEEE Transactions on Image Processing* 16(1):188–197.

Bronstein, A. M.; Bronstein, M. M.; and Kimmel, R. 2008. *Numerical geometry of non-rigid shapes*. Springer.

Dennis, J. J. E., and Schnabel, R. B. 1983. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16. SIAM.

Elad, A., and Kimmel, R. 2003. On bending invariant signatures for surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(10):1285–1295.

Elhamifar, E., and Vidal, R. 2011. Sparse manifold clustering and embedding. In *Advances in Neural Information Processing Systems*, 55–63.

Gold, S., and Rangarajan, A. 1996. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(4):377–388.

Golub, G. H., and Van Loan, C. F. 2012. *Matrix computations*, volume 3. JHU Press.

Hawkins, P. C.; Skillman, A. G.; and Nicholls, A. 2007. Comparison of shape-matching and docking as virtual screening tools. *Journal of Medicinal Chemistry* 50(1):74–82.

Kuhn, H. W. 1955. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2(1-2):83–97.

Liu, D. C., and Nocedal, J. 1989. On the limited memory bfgs method for large scale optimization. *Mathematical programming* 45(1-3):503–528.

Mémoli, F., and Sapiro, G. 2005. A theoretical and computational framework for isometry invariant recognition of point cloud data. *Found. Comput. Math.* 5:313–346.

Meyer, C. 2000. *Matrix analysis and applied linear algebra book and solutions manual*, volume 2. SIAM.

Müller, M.; Heidelberger, B.; Teschner, M.; and Gross, M. 2005. Meshless deformations based on shape matching. In *ACM Transactions on Graphics (TOG)*, volume 24, 471–478. ACM.

Potmesil, M. 1979. Generation of 3d surface descriptions from images of pattern-illuminated objects. In *Proc. IEEE Conf. Pattern Recognition and Image Processing*, 553–559.

Raviv, D.; Bronstein, A. M.; Bronstein, M. M.; and Kimmel, R. 2010. Full and partial symmetries of non-rigid shapes. *International Journal of Computer Vision* 89(1):18–39.

Rodolà, E.; Torsello, A.; Harada, T.; Kuniyoshi, Y.; and Cremers, D. 2013. Elastic net constraints for shape matching. In *Proceedings of 2013 IEEE International Conference on Computer Vision (ICCV'13)*.

Schwartz, E. L.; Shaw, A.; and Wolfson, E. 1989. A numerical solution to the generalized mapmaker's problem: flattening nonconvex polyhedral surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(9):1005–1008.

Sprechmann, P.; Bronstein, A. M.; and Sapiro, G. 2012. Learning robust low-rank representations. *arXiv preprint arXiv:1209.6393*.

Stewart, G. 2000. The decompositional approach to matrix computation. *Computing in Science & Engineering* 2(1):50–59.

Syeda-Mahmood, T.; Beymer, D.; and Wang, F. 2007. Shape-based matching of ecg recordings. In *Proceedings of 29th Annual International Conference of Engineering in Medicine and Biology Society (EMBS 2007)*, 2012–2018. IEEE.

Wolfe, P. 1969. Convergence conditions for ascent methods. *Society for Industrial and Applied Mathematics Review* 11(2):226–235.

Wolter, D., and Latecki, L. J. 2004. Shape matching for robot mapping. In *PRICAI 2004: Trends in Artificial Intelligence*. Springer. 693–702.

Yuan, M.; Yuan, M.; Lin, Y.; and Lin, Y. 2006. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B* 68:49–67.

Zhao, W.; Chellappa, R.; Phillips, P. J.; and Rosenfeld, A. 2003. Face recognition: A literature survey. *ACM Computing Surveys (CSUR)* 35(4):399–458.