Grassmann manifold based shape matching and retrieval under partial occlusions

Chenxi Li*a,b, Zelin Shi², Yunpeng Liu³, Baoshu Xu⁴

²Shenyang Institute of Automation, Chinese Academy of Science, Shenyang 110016, China;
³University of Chinese Academy of Science, Beijing 100049, China

ABSTRACT

Shape matching and recognition is a challenging task due to geometric distortions and occlusions. A novel shape matching approach based on Grassmann manifold is proposed that affine transformations and partial occlusions are both considered. An affine invariant Grassmann shape descriptor is employed which projects one shape contour to a point on Grassmann manifold and gives the similarity measurement between two contours based on the geodesic distance on the manifold. At first, shape contours are parameterized by affine length and accordingly divided into local affine-invariant shape segments, which are represented by the Grassmann shape descriptor, according to their curvature scale space images. Then the Smith-Waterman algorithm is employed to find the common parts of two shapes’ segment sequences, and get the local similarity of shapes. The global similarity is given by the found common parts, and finally the shape recognition accomplished by the weighted sum of local similarity and global similarity. The robustness of the Grassmann shape descriptor is analyzed through subspace perturbation analysis theory. Retrieval experiments show that our approach is effective and robust under affine transformations and partial occlusions.

Keywords: Grassmann manifold, Partial shape matching, contour matching, partial occlusion, affine invariant

1. INTRODUCTION

Shape analysis plays an important role in computer vision and image analysis. Shape matching is its center problem which is difficult for two well-known factors. Firstly, planar shapes in images taken from different viewpoints of the same object suffer from geometric distortions. Secondly, objects may be partially occluded by other objects.

Geometric distortions can be locally approximated by affine transformations. Therefore, great research efforts have been devoted to developing affine-invariant shape description. For example, it was shown that the CSS representation¹ is robust to affine transformations. Fourier descriptors is generalized to affine invariant Fourier descriptors². Besides affine invariants mentioned above, another different way of treating affine-invariance is normalization, usually by whitening³,⁴. Shape contexts⁵ (SC) is also a popular shape descriptor which is generalized to inner distance shape contexts⁶ (IDSC), which is shown to be robust to deformations and partial occlusions. Shapes were projected into points of the Grassmann manifold giving a natural representation of shapes⁷, which is completely affine-invariant. All methods mentioned above are global descriptors of planar shapes and effective when objects are completely visible.

Many shape matching methods have been proposed to handle partial occlusion cases, while most of them are only invariant to similarity transforms and rare methods take into account affine transformations. Zhang and Tang⁸ defined a local affine-invariant descriptor for each point of a contour based on the Grassmann shape descriptor. However, the necessary setting of the number of the points of the local shape segment limits the effectiveness of their method. An affine-invariant contour descriptor⁹ based on affine-invariant signature, is proposed for occluded object recognition. But this descriptor is sensitive to noise since it uses high order derivatives.

*lichenxi@sia.cn
In this paper, we propose a new affine-invariant shape matching method under partial occlusions based on Grassmann shape descriptor. Our method is motivated by the work in In our method, we employ Grassmann shape descriptor, and measure the similarity between shapes from local and global aspects, while the work in considers only local aspect. We evaluate the performance of our proposed method by comparing with the methods proposed in and .

The rest of this paper is organized as follows. Section 2 gives the description of the Grassmann shape descriptor and its theoretical robustness analysis. In section 3, the partial shape matching algorithm is presented. Experimental results are given in section 4. Conclusion is remarked in section 5.

2. GRASSMANN SHAPE DESCRIPTOR AND ITS ROBUSTNESS ANALYSIS

2.1 Grassmann shape descriptor

Consider a parametric vector equation for a planar contour:

$$\Gamma(u) = (x(u), y(u))$$

(1)

where $u$ is an arbitrary parameter. In order to weaken the effect of affine deformation on the shape representation, a contour is often parameterized by affine length which has the following definition:

$$\tau(u) = \left[ (x'(t)y''(t) - x''(t)y'(t)) \right]^{1/3} dt$$

(2)

where $x'$ and $x''$ represent the first and second order derivatives of $x$. Derivatives of $y$ are analogous.

For a planar shape $X$ in an image, its contour can be extracted by an edge detector, and resampled uniformly using affine length to $n$ ordered points, which are denoted as $\{s_1, s_2, \cdots, s_n\}$, where $s_i$ is a $1 \times 2$ vector, representing the coordinate of the $i$-th point of the contour. The configuration matrix of shape $X$ is denoted as $S_x = [s_1^T, s_2^T, \cdots, s_n^T]^T$. Note that, we don’t consider the singular shape which is just a straight line, i.e. we have $\text{Rank}(S_x) = 2$. We use the centered configuration matrix $\bar{S}_x$ to instead $S_x$ by introducing the centering matrix and have

$$\bar{S}_x = (I - \frac{1}{n}1_n \cdot 1_n^T) \cdot S_x$$

(3)

where $I$ is $n \times n$ identity matrix and $1_n = [1, \cdots, 1]^T \in \mathbb{R}^n$.

For two shape contours $A$ and $B$, we assume their numbers of points are both $n$, and they are related by an affine transformation. Their centered configuration matrices are denoted by $\bar{S}_a$ and $\bar{S}_b$ respectively, then $\bar{S}_a = \bar{S}_b \cdot M$, where $M$ is a $2 \times 2$ non-singular matrix representing an affine transformation. So the column space $\text{col}(\bar{S}_a)$ of $\bar{S}_a$ and $\text{col}(\bar{S}_b)$ of $\bar{S}_b$ are identical, i.e., $\text{col}(\bar{S}_a) = \text{col}(\bar{S}_b)$.

Therefore, the column space of centered configuration matrix of one shape can be used as the affine-invariant shape descriptor, which is called as the Grassmann shape descriptor in this paper. The Grassmann manifold $Gr(2, n)$ is the affine-invariant shape space, since the column space of a centered configuration matrix of one shape is a point on it.

The Grassmann manifold $Gr(2, n)$ is a Riemann manifold, on which the geodesic distance can be defined as

$$d_G(\text{col}(\bar{S}_a), \text{col}(\bar{S}_b)) = \sqrt{\text{trace}(\Theta^T \Theta)}$$

(4)

where $\Theta$ is the principal angle matrix between $\text{col}(\bar{S}_a)$ and $\text{col}(\bar{S}_b)$. By performing singular value decomposition, $\Theta$ can be computed as $\Theta = \arccos(C)$, where $UCV^T = \text{svd}(\bar{S}_a^T \bar{S}_b)$.
The maximum $D$ of geodesic distance on $Gr(2,n)$ is $\sqrt{2}\pi/2$. For convenience, we use the normalized geodesic distance, which is denoted by $d_n(\tilde{S}_a,\tilde{S}_b)$, to instead the previous definition and we have

$$d_n(\tilde{S}_a,\tilde{S}_b) = \frac{1}{D} d_n(\text{col}(\tilde{S}_a),\text{col}(\tilde{S}_b))$$  (5)

The geodesic distance between two points of $Gr(2,n)$ gives the difference between the corresponding shapes and is used to define the similarity between two contours.

### 2.2 The robustness of Grassmann shape descriptor

In practice, the process of image acquisition and digitalization can cause deviation in shape representations. A good shape representation should have the property that the similarity measure doesn’t change too much if the deviation is small. The following theorem from subspace perturbation analysis theory reveals the perturbation bound of principal angles.

**Theorem 1**. Let $A, A' \in \mathbb{C}^{m \times p}, B, B' \in \mathbb{C}^{n \times q}$, $p \geq q$, and $\sigma(A,B) = \{\cos \theta_k\}_{k=1}^q = \{c_k\}_{k=1}^q$, $0 \leq \theta_1 \leq \cdots \leq \theta_q \leq \frac{\pi}{2}$. Suppose $\sigma(A',B') = \{\cos \theta_k'\}_{k=1}^q = \{c_k'\}_{k=1}^q$, $0 \leq \theta_1' \leq \cdots \leq \theta_q' \leq \frac{\pi}{2}$. Then

$$\sqrt{\sum_{k=1}^q (\theta_k' - \theta_k)^2} \leq \frac{\pi}{2} (\delta_e(A,A') + \delta_e(B,B'))$$  (6)

where $\delta_e(X,X') = 2\|X\|_F \cdot \|X'\|_F \cdot \frac{\|X - X'\|_F}{\|X\|_F}$, $\|\cdot\|_F$ is the Frobenius norm, $\|\cdot\|_2$ is the spectral norm and $X^\dagger$ is the Moore-Penrose pseudo-inverse of $X$.

Suppose $X$ is the configuration matrix of a given planar shape, and $X' = X + \Delta X$, i.e., $X'$ is the perturbed configuration matrix of the given shape. In theorem 1, we set $X = A$, $X' = B = B' = A'$, and $p = q = 2$ (implying planar contour). Then we can deduce the robustness of Grassmann shape descriptor in the following corollary.

**Corollary 1**: Let $X$ and $X'$ are configuration matrices of a given shape and its perturbed shape respectively, and $X' = X + \Delta X$, then

$$d_n(X,X') \leq \|\Delta X\|_F / \|X\|_F$$  (7)

We can conclude that the similarity measure using Grassmann shape descriptor is continuously dependent on the configuration matrix of one shape and the robustness of the shape descriptor is guaranteed theoretically.

### 3. SHAPE MATCHING ALGORITHM

This section describes how to employ the Grassmann shape descriptor into a partial shape matching strategy. At first, contours are divided into affine-invariant segment sequences according to their CSS images. Then segment sequences are aligned by the Smith-Waterman algorithm. The similarity of different shapes is defined from local and global aspect.

#### 3.1 Contour segmentation

Dealing with the partial shape matching problem needs local information of contours. Before the matching stage, we divide each contour into segments according to their CSS images. This segmentation approach is not only affine-invariant but also robust to noise.

The details of CSS image construction are not introduced here due to the space limitation. It should be mentioned that, a maximum is considered as noise if it is less than $1/5$ of the largest maximum. An example of segmentation can be found...
After segmentation, a shape \( A \) can be expressed as a segment sequence \( A = \{a_1, \cdots, a_M\} \), where \( M \) is the number of segments in the sequence.

### 3.2 Smith-Waterman alignment

The Smith-Waterman algorithm (or SW algorithm) was proposed for identification of common molecular sequences\(^{13}\). It is introduced to solve computer vision problems recently\(^{4}\).

For two shape contours \( A \) and \( B \), their segment sequences are \( A = \{a_1, \cdots, a_M\} \) and \( B = \{b_1, \cdots, b_N\} \) respectively. The segmental similarity between \( a_i \) and \( b_j \) is \( w(i, j) \) which can be calculated as in section 3.3. The gap penalty is \( w(i, -) = w(-, j) = -1 \), where minus “−” represents the insertion of a gap into the aligned sequences.

The details of the SW algorithm can be found in\(^{11}\). Through a trace-back procedure, we can find out the optimal alignment of the two segment sequences. If one of two shapes, for example \( A \), is closed, then the circular shift operator\(^{4}\) is applied to the segments sequence \( A \) to find the best alignment. If the shape is open, then its first segment and last segment are removed from the segment sequence as they may not be destroyed by occlusions.

### 3.3 Shape similarity calculation

For shape \( A = \{a_1, \cdots, a_M\} \) and \( B = \{b_1, \cdots, b_N\} \), we define the similarity between them from local and global aspect based on the Grassmann shape descriptor.

#### a) Local similarity

The segmental similarity between \( a_i \) of sequence \( A \) and \( b_j \) of sequence \( B \) is defined as

\[
w(i, j) = \begin{cases} 
1, & d_N(a_i, b_j) \leq \text{thresh}_{\text{local}} \\
8, & (d_N(a_i, b_j) > \text{thresh}_{\text{local}})
\end{cases}
\]

where \( d_N(a_i, b_j) \) is the normalized Grassmann geodesic distance explained in section 2.1, and \( \varepsilon \) (which is set to 0.001) is a small number avoiding the denominator Grassmann geodesic distance being zero. If the geodesic distance is larger than a given threshold \( \text{thresh}_{\text{local}} \), we set \( w(i, j) \) to -1, which means that these two segments are impossible derived from the same object. If the geodesic distance is smaller than \( \text{thresh}_{\text{local}} \), we set the maximum similarity between segments is the minimum between 8 and \( 1/(d_N(a_i, b_j) + \varepsilon) \) to improve the stability of the algorithm. In our experiments, the threshold \( \text{thresh}_{\text{local}} \) is set to 0.333.

By performing the SW algorithm, we can obtain the best alignment, and an example is shown in Figure 1(a). Suppose that \( A' = \{a_1, \cdots, a_l\} \) and \( B' = \{b_1, \cdots, b_j\} \) are the final matched segment sequences between shape \( A \) and \( B \), then the local similarity between the two shapes is defined as

\[
S_{\text{local}} = \sum_{k=1}^{l} w(i_k, j_k).
\]

#### b) Global similarity

Even though two shapes’ local similarity is high, they still may be derived from different objects. So we also consider the global similarity. To compute the global similarity, we only choose those matched segmental pairs whose segmental similarity are bigger than a given threshold \( \text{thresh}_{\text{global}} \) in order to eliminate the effect of incorrect matched pairs. In our experiments, the threshold \( \text{thresh}_{\text{global}} \) is set to 4. Supposing the segments chosen to compute the global similarity compose shape \( \tilde{A} \) and \( \tilde{B} \) respectively, we define the global similarity between shape \( A \) and \( B \) as

\[
S_{\text{global}} = \frac{1}{d_N(\tilde{A}, \tilde{B}) + \varepsilon}
\]
where \( d_{\text{Gr}}(\mathbf{A}, \mathbf{B}) \) is the normalized Grassmann geodesic distance, and \( \varepsilon \) (which is set to 0.001) is a small number, in case the denominator is zero. Figure 1 (b) is an example corresponding to Figure 1 (a).

The similarity of shape \( A \) and \( B \) is defined as

\[
Sim(A, B) = \omega_1 S_{\text{local}} + \omega_2 S_{\text{global}}
\]

(11)

where \( \omega_1 \) and \( \omega_2 \) are the weight of local similarity and global similarity respectively. In our experiments, \( \omega_1 \) and \( \omega_2 \) are both set to 0.5.

![Figure 1. (a) The best alignment by the Smith-Waterman algorithm; (b) Contours chosen to calculate global similarity.](image)

**4. EXPERIMENTAL RESULTS**

We evaluate the proposed algorithm on the MPEG-7 Part-B shape database. We choose 37 classes from the MPEG-7 Part-B shape database, as shown in Figure 2. In our shape dataset, each class has 14 images, among which the first image is chosen from MPEG-7 Part-B shape database, while the rest images are distorted by affine transformations. The parameters of affine transformations have 13 groups\(^\text{12}\). Therefore, we obtained 518 images in the dataset. The samples of affine distorted shape for class ‘horse’ is shown in Figure 3.

The proposed algorithm is compared with the Whiten method\(^4\) and IDSC method\(^6\) through retrieval rate test and robustness test. The Whiten method is proposed for partial shape matching and recognition under affine transforms, while the IDSC method is robust to deformations and occlusions. So these two methods are suitable for comparison.

![Figure 2. The selected shapes from MPEG-7 Part-B database](image)

![Figure 3. The samples of affine distorted shape for class ‘horse’. The first one is the original shape, while the rest are the affine distorted shapes.](image)

**4.1 RETRIEVAL RATE TEST**

We used the 37 original images as query set. For simulating partial occlusions, a part of each query contour points ranging from 0 to 30% were removed. We use Bulls-eye rate to measure the retrieval rate of each method. For each image in the template set, it is matched with all images in the dataset and the top 28 most similar candidates are counted. There are at most 14 of the 28 candidates that are correct hits. The retrieval rate is the ratio of the number of the correct hits to the highest possible correct hits (which is \(14 \times 37\)). The comparison of recognition rate between our method and the other two methods under different level of occlusions is shown in Figure 4. It can be seen clearly that our method outperforms the other two algorithms.
4.2 Robustness test

In this experiment, we investigated the robustness of our method. For a given query, 10% of landmark points were removed to simulate occlusion, and Gaussian noises of different levels (with standard deviation ranging from 0 to 6 pixels in 0.5 pixel increments) were added separately to the coordinates of the query. The comparison of recognition rate between our method and the other two methods under different noise levels is shown in Figure 5. It can been seen that our method is more robust than the Whiten algorithm especially when the standard deviation ranges from 0 to 3 pixels, while The IDSC method as a statistical method has the best anti-noise performance.

![Figure 4. Recognition rate under different occlusions](image1.png)

![Figure 5. Comparison under different noise levels](image2.png)

5. CONCLUSION

We proposed a new partial shape matching and recognition approach that is affine-invariant, including translation, rotation, scaling and shearing occasions. The robustness is tested by experiments as well. The future work includes introducing Bayesian framework into the proposed method to improve the accuracy and efficiency.

Acknowledgment

This work has been funded by Key Laboratory of Opto-Electronic Information Processing Foundation of Chinese Academy of Science (Y0A410140).

REFERENCES