

Global Consistent Shape Correspondence for Efficient and Effective Active Shape Models

Meng Wang

NYU Multimedia and Visual Computing Lab
Dept. of ECE, NYU Abu Dhabi, UAE
Dept. of CSE, NYU Tandon School of
Engineering, USA

Yi Fang*

NYU Multimedia and Visual Computing Lab
Dept. of ECE, NYU Abu Dhabi, UAE
Dept. of ECE, NYU Tandon School of
Engineering, USA

ABSTRACT

Finding the accurate corresponded landmarks from a collection of shape instances plays critical role in constructing active shape models (ASMs). We have developed a global consistent shape correspondence paradigm for efficient and effective active shape models to address challenging issues in statistical shape modelling. Specifically, in this paper, we developed techniques to perform a fast multiple shape matching to identify global consistent shape correspondence from a set of training shape instances via efficient low-rank recovery optimization. High quality ASMs can then be constructed based on the identified corresponded points. The entire process is unsupervised without manual annotation as well as free of selection of anatomically significant point. Experimental results on mobile hand image data demonstrate the superior performance of our proposed method over other state-of-the-art techniques like MDL in constructing active shape models.

Keywords

Active shape model; Shape correspondence; Joint shape matching

1. INTRODUCTION

1.1 Background

Active shape models (ASMs) are statistical models that are widely used to quantitatively represent various shape structures and their possible variations [6]. With the active shape models, a new shape can be generally reconstructed via a iterative fitting a process. The ASMs have been widely used in many computer vision applications such as object tracking and classification [7, 16].

*Corresponding author. Email: yfang@nyu.edu

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM '16, October 15-19, 2016, Amsterdam, Netherlands

© 2016 ACM. ISBN 978-1-4503-3603-1/16/10...\$15.00

DOI: <http://dx.doi.org/10.1145/2964284.2967283>

However, it is well known that the automatic identification of an accurate set of corresponded landmark points among training shape instances often poses challenges for existing techniques towards constructing active shape models. A small correspondence error in identified landmark points might significantly deteriorate the quality of constructed ASMs [8]. To address the correspondence challenges, most previous research firstly obtain the corresponded landmarks by manual annotation. Although the manually labelled landmarks can provide a decent set of correspondence points among shapes, it is a laborious process of manually annotating the shapes. In addition, due to the variety of individual backgrounds, the labelled landmarks are not always consistent each time and not be coincident with any anatomically significant point of a shape structure, resulting in an inefficient and ineffective construction of ASMs. Shape correspondence is therefore a well known challenging problem, especially for shape instances with large non-rigid shape variation and corrupted with great noise.

1.2 Related Works

There have been several prior works that were developed to address the shape correspondence challenge as discussed above [1, 2, 9, 10, 5, 18, 13]., We will briefly review two representative works, the minimum description length (MDL) method [10, 11] and the spherical harmonics descriptors method (SPHARM) [5]. One of widely used technique was developed based on MDL [10, 11] to establish shape correspondence through minimizing the errors measured by the required bit-length to transmit the constructed new shape model and all of the training shape models. Despite the effectiveness of MDL, it suffers the high complexity due to the nature of shape correspondence error measure and the local optimal solution if the training shape instances present a large structural variations and noise, leading to inconsistent and local shape correspondence among training instances. In [5], authors identify the shape correspondence using spherical parametrization of shape instances. In spite of the good performance in 3D shape correspondence, SPHARM requires expensive computation because of the conformal mapping transformation.

Recent work has indicated that optimizing the consistency of a collection of pairwise shape correspondences is able to give rise to a global consistent shape correspondence [15] among the population of shape instances. The hidden constraint used for joint multiple shape registration is the cycle consistency, that is, correct correspondences along different

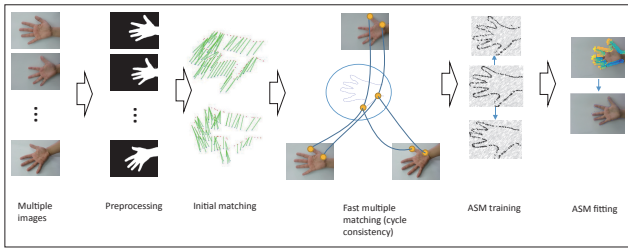


Figure 1: Pipeline of global consistent shape correspondence for active shape model.

walking paths of a collection of shapes should always be identical. By enforcing this constraint on multiple shape correspondence, most of the incorrect correspondences will be detected and replaced by correct maps along other paths, and missing maps will be filled up by the help of intermediate shapes. In addition, dissimilar shapes can be corresponded through the set of similar shapes between them. However, this constraint can not be modelled easily as the number of possible paths are often very large and the number of consistent cycles could be very small [20]. [15, 14] propose that the cycle consistency can be modelled as a positive semi-definite and low-rank optimization process for the global binary correspondence matrix. However, unfortunately the semi-definite programming (SDP) employed is not computational efficient when the matrix is large, which limits the application to a small scale collection of shapes instances.

1.3 Our method

In this paper, to efficiently find global consistent shape correspondence for better constructing active shape models with higher quality without manually annotating the landmark points, we introduce a fast low-rank matrix recovery optimization using the nuclear-norm relaxation for rank minimization via a fast alternating optimization to address the computational limitation of SDP in registering a large number of shapes of many sampling points from each shape [20]. There are three major attractive features in our proposed method:

- Global consistent shape correspondence: our method is able to find the corresponding landmarks that are globally consistent among all of training shape instances,
- Unsupervised learning: our method is able to automatically find anatomically significant corresponding landmarks without manual annotation,
- Efficient modelling: our method is able to quickly find consistent correspondence from a reasonable size of shape instances.

The pipeline of our method can be seen in Figure 1. In the first step the images are preprocessed to extract sample points on the contour. Second, an initial matching is performed based on the extracted descriptors of the sample points. Then the fast multiple matching are performed on the initial matching result to generate more consistent correspondences. Fourth, ASM is constructed using the well-corresponded training set. At last, the trained model is used

for analysis or fitting. The pipeline is applicable to different ASMs, for simplicity we take hand image processing as an example.

2. METHOD

2.1 Fast Multiple Shape Matching via Low-rank Recovery Optimization

2.1.1 Preprocessing

The purpose of this step is to extract the contour of the shape. Skin detection is an efficient way to extract the foreground when it comes to human skin such as hands. The specific skin color is defined so that the satisfying pixels can be picked up. The range of human skin can be adjusted according to the lighting conditions. The pixels just picked can be discontinuous and noisy, a morphological operation such as image close operation (dilate then erode) can be used to smooth and denoise the extracted foreground (remove isolated pixels, fill small holes). At last, the contour can be obtained from the extracted foreground with the isocontour geometry extraction method.

2.1.2 Initial Matching

Shape correspondences are normally calculated using shape descriptors as a similarity measure. A shape descriptor is a representation of structure information of the shape. For the hand image we are considering, shape context [3] descriptor is an efficient method. Shape context is calculated based on the context of sampling points, which is a robust and simple algorithm for finding correspondences between shapes. The extracted shape context descriptor can be used to calculate the similarity between points and perform matching task between contour images. Typical matching methods include greedy matching, Hungarian assignment, graph matching. As the number of initially sampled points is often not consistent, greedy matching performs better as we do not require a one-to-one matching.

2.1.3 Multiple Matching

Similar to fast optimization strategy used in [20, 19] for multi-view and multi-image matching, we introduce fast alternating minimization to solve the low-ranking matrix recovery optimization in order to handle practical problems with a large scale of shapes. As reported in [20], the fast alternating minimization used in [20] is about 20 times faster than the one used in [15] for joint registering 20 shapes with each shape of 500 samples. The computational time for the method in [15] exponentially increase as the size of shape descriptors increase. As we discussed before, the cycle consistency is an important constraint used in matching a set of shapes. It can be defined as:

$$\mathbf{X}_{ac} = \mathbf{X}_{ab}\mathbf{X}_{bc} \quad (1)$$

where a, b, c are any three shapes from the shape set, $\mathbf{X}_{ij} \in \{0, 1\}^{(m \times n)}$ is a binary matrix indicating whether a pair of vertices from two shapes i, j is a match, m, n are the number of sampled vertices in shape i, j .

The cycle consistency can be more concisely described by introducing a virtual set of unique vertices in the set of shapes. The virtual set of vertices can be utilized as intermediate vertices by many vertices in the same time. The

cycle consistency can be refined as:

$$\mathbf{X}_{ij} = \mathbf{A}_i \mathbf{A}_j^T, \mathbf{A}_i \in \{0, 1\}^{m \times k}, k \geq m \quad (2)$$

where \mathbf{A}_i denotes the mapping from shape i to the virtual set, m is the number of vertices in shape i .

All the \mathbf{X}_{ij} matrices can be tiled together into one giant matrix \mathbf{X} .

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1n} \\ \mathbf{X}_{21} & \mathbf{X}_{22} & \cdots & \mathbf{X}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{n1} & \mathbf{X}_{n2} & \cdots & \mathbf{X}_{nn} \end{pmatrix} \quad (3)$$

where n is the number of 3D shapes. Now \mathbf{X} can be denoted as $\mathbf{X} = \mathbf{A} \mathbf{A}^T$. From this equation, we can see that \mathbf{X} should be both positive semi-definite ($\mathbf{X} \geq 0$) and in low-rank ($rank(\mathbf{X}) \leq k$). Imposing a constraint on \mathbf{X} can avoid checking cycle-consistency over all the possible paths.

The cycle consistency has been formulated as an optimization problem to do matrix recovery. The overall cost function can be formulated as:

$$f(\mathbf{X}) = -\sum_{i=1}^n \sum_{j=1}^n \mathbf{S}_{ij} \cdot \mathbf{X}_{ij} + \lambda \|\mathbf{X}\|_* = -\mathbf{S} \cdot \mathbf{X} + \lambda \|\mathbf{X}\|_* \quad (4)$$

where \mathbf{S} is a global affinity matrix, $\lambda \|\mathbf{X}\|_*$ is the nuclear norm of \mathbf{X} . The nuclear term is the sum of singular values in \mathbf{X} , which is a relaxation for the rank of \mathbf{X} . Here we set $\mathbf{S} = \mathbf{M}$, where \mathbf{M} is the matching result from pair-wise correspondence, and \mathbf{M} is constructed from many pair-wise correspondences like Equation 3. It is proven that the positive semi-definite constraint can be fulfilled in the same time if λ is large enough [21]. The nuclear norm minimization problem is convex and the proximal method [17] and ADMM [4] can be used to solve the problem. A fast alternating [21] is proposed to solve this problem much more faster.

2.2 Active Shape Model

To build the ASM models, well-corresponded examples we obtained from multiple matching are used as a representative training set.

2.2.1 Alignment

In most cases among different shapes the position, scale, and rotation are different, so an alignment is a necessity to the training set. We make transformations (translation, scaling, and rotation) to each shape compared to the first shape such that they become as similar as possible. Then all the shapes are aligned to the mean shape. The Procrustes method is used to find the transformation that best aligned the shapes.

2.2.2 Applying Principal Component Analysis

A set of aligned vectors in the form of Eq 5 is obtained after shape alignment.

$$\mathbf{x} = [x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n] \quad (5)$$

where n is the number of sampled points in the matching, x_i, y_i are coordinates of point i . Now a shape can be represented by a vector in a $2n$ -dimensional space. As a result, the training set can be fully characterized by a distribution of m vectors, where m is the size of the training set. Principal Components Analysis (PCA) is then used to extract the directions of independent variation in the cloud of points. The most significant directions with largest eigenvalues are

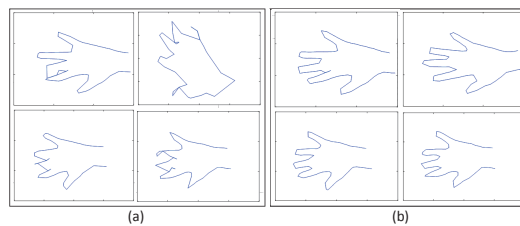


Figure 2: Results of multiple shape matching. First the corresponding points from the points in the first shape are plotted, then the points are connected according to the connection of corresponding points of the first shape. (a) Shapes generated from the initial matching. (b) Shapes generated from the refined multiple matching.

retained and the shape can be approximated without losing a lot of information by:

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P} \mathbf{b} \quad (6)$$

where $\mathbf{P} = \mathbf{p}_1 \mathbf{p}_2 \dots \mathbf{p}_t$ is the eigenvectors, $\mathbf{b} = [b_1 b_2 \dots b_t]$ is the coefficient. t is chosen so that 98% of the variance is represented (98% of the total eigenvalues are used). \mathbf{b} is a set of parameters to deform a shape and generate new plausible shapes.

2.2.3 Generate shapes

Suppose for model that the different parameters b_i are independent, gaussian distributed with zero mean, then 95% of the distribution of one parameter is covered in the range $[-2SD, 2SD]$, where SD is the standard deviation of the training vectors. By constricting b to $[-2\sqrt{\lambda_i}, 2\sqrt{\lambda_i}]$, where λ_i is the variance of the i th parameter, the shape generated is similar to the shapes in the training set. t is the number of modes needed to describe the shape properly.

2.2.4 Adjust ASM to fit a shape in the image

An ASM of the hand consists a mean shape $\bar{\mathbf{x}}$, a parameter vector \mathbf{b} , and eigenvectors. Through certain variations on the \mathbf{b} and transformation, we can generate a plausible shape that best fit a new input shape.

Plausible points selection. It is important to select plausible points that both keep the basic shape of the trained model and keep close to the new shape. For each point in the mean shape, the candidate points are selected along the boundary normal within a certain range to keep the basic shape. After inputting the new shape, candidates points that are the most similar in gradient textures to the point in the new shape are selected. The selected plausible points will form a new target shape.

Adjust the model parameters to best fit selected points. Given a new set of points \mathbf{x} , we can translate, rotate, scale and deform our model in order that it best fits \mathbf{x} . This can be done by applying a transformation to find the best alignment, and compute some new shape parameters \mathbf{b} to find the best deformation. Given the mean shape \mathbf{x} and the set of new points \mathbf{s} , the shape parameters can be obtained.

These two processes will be executed alternatively until there are no big updates on transformation and \mathbf{b} .

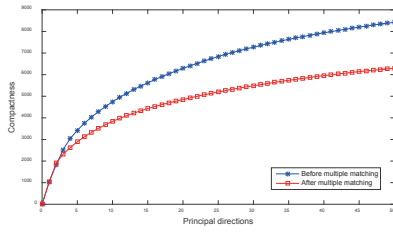


Figure 3: Quantitative evaluation of ASMs using compactness.

Table 1: Quantitative evaluation of ASMs using AIC

Method	AIC value
MDL	12.54
Shape Context	12.32
Our method	11.10

3. EXPERIMENT

3.1 Dataset and Settings

A set of experiments are performed to evaluate our method. The hand image data is selected from mobile hand images database (MOHI) [12]. MOHI is a database of hands created using mobile phones, which has modest qualities. 110 images are used to perform the multiple matching and active shape model training. The running time for 110 shapes with 200 samples per shape is about one hour, which is significantly (about 20 times) faster than other SDP-based matrix recovery method. For experiment purpose, equal number of points (200 points per shape) are sampled from the edge. In multiple matching, λ is set to 50.

3.2 Results

First a shape matching experiment is conducted before and after multiple matching to test the performance of multiple matching. As in Figure 2, the shapes are generated using the corresponding points to the first shape. The shapes in Figure 2(a) are generated using the correspondence of initial matching, which are problematic due to incorrect correspondences. We show matching result after multiple matching in Figure 2 (b), the shapes generated have much improved quality as the correspondence errors are largely decreased. As can be seen from the matching result, the multiple matching is effective in improving the correspondences.

We use two measures as evaluation criteria to evaluate our method quantitatively: compactness and Akaike’s Information Criterion (AIC). Compactness suggests the cumulative variance of the input vector. A good model is normally trained from data with as little variance as possible. Compactness is defined as:

$$compact(M) = \sum_{m=1}^M \lambda_m \quad (7)$$

where λ_m is the m_{th} largest eigenvalue. AIC is an evaluation criterion for evaluating and selecting models, which can also be used to evaluate the ASMs. We are using the form of Eq 8.

$$AIC = -\log\left(\frac{n}{R}\right) + 2\frac{k}{n} \quad (8)$$

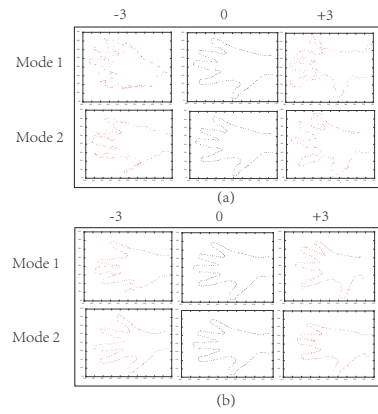


Figure 4: Comparisons of generated plausible shapes. (a)Shape context. (b)Our method of refined multiple matching. Shapes generated of first two modes varying from -3SD to 3SD are displayed.

where n is the size of the training set, R is the construction residual reconstructing using the first k principal directions. We set k to be 3 for all the methods. Smaller value of AIC indicates higher quality of the model. Quantitative results are shown in Table 1 and Figure 3. As we can see in Table 1, our method gives the best ASM among the three methods as indicated by the smallest AIC value. As we can see in Figure 3, our method gives a better performance as indicated by lower curve. The results quantitatively validate the superior performance of our proposed method over other techniques in constructing ASM.

To test the performance of the trained ASM, we also conduct an experiment on the plausible shapes generated from ASM. In Figure 4, plausible shapes generated are compared for ASMs based on correspondences before and after multiple matching. As indicated by Figure 4 (a) and (b), our method show great improvement in terms of the quality of generated plausible shapes. Shapes in (a) are inconsistent and many points of them does not lie on edge of hands. Shapes in (b) are more similar to hands contour edges in plausible variations from the mean shape. The global multiple shape matching has greatly improved the quality of ASM.

4. CONCLUSION

In this paper, we developed techniques to perform a fast multiple shape matching to identify global consistent shape correspondence from a set of training shape instances via efficient low-rank recovery optimization. High quality ASMs can then be constructed based on the identified corresponded points. The entire process is unsupervised without manual annotation as well as free of selection of anatomically significant point. Both quantitative and qualitative evaluation results on mobile hand image data demonstrate the superior performance of our proposed method in constructing active shape models.

5. ACKNOWLEDGMENTS

This material is partly based upon work supported by New York University Abu Dhabi institute (AD131 and REF131).

6. REFERENCES

- [1] X. Bai, S. Bai, Z. Zhu, and L. J. Latecki. 3d shape matching via two layer coding. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(12):2361–2373, 2015.
- [2] X. Bai, C. Rao, and X. Wang. Shape vocabulary: A robust and efficient shape representation for shape matching. *Image Processing, IEEE Transactions on*, 23(9):3935–3949, 2014.
- [3] S. Belongie, J. Malik, and J. Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *NIPS*, volume 2, page 3, 2000.
- [4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.
- [5] C. Brechbühler, G. Gerig, and O. Kübler. Parametrization of closed surfaces for 3-d shape description. *Computer vision and image understanding*, 61(2):154–170, 1995.
- [6] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):681–685, 2001.
- [7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.
- [8] P. Dalal, B. C. Munsell, S. Wang, J. Tang, K. Oliver, H. Ninomiya, X. Zhou, and H. Fujita. A fast 3d correspondence method for statistical shape modeling. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.
- [9] R. H. Davies, T. F. Cootes, C. J. Twining, and C. J. Taylor. An information theoretic approach to statistical shape modelling. In *BMVC*, pages 1–10, 2001.
- [10] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *Medical Imaging, IEEE Transactions on*, 21(5):525–537, 2002.
- [11] G. Gerig, M. Styner, D. Jones, D. Weinberger, and J. Lieberman. Shape analysis of brain ventricles using spharm. In *Mathematical Methods in Biomedical Image Analysis, 2001. MMBIA 2001. IEEE Workshop on*, pages 171–178. IEEE, 2001.
- [12] A. Hassanat, M. Al-Awadi, E. Btoush, A. Al-Btoush, G. Altarawneh, et al. New mobile phone and webcam hand images databases for personal authentication and identification. *Procedia Manufacturing*, 3:4060–4067, 2015.
- [13] A. Hill and C. J. Taylor. A method of non-rigid correspondence for automatic landmark identification. In *BMVC*, pages 1–10, 1996.
- [14] Q.-X. Huang, S. Flöry, N. Gelfand, M. Hofer, and H. Pottmann. Reassembling fractured objects by geometric matching. *ACM Transactions on Graphics (TOG)*, 25(3):569–578, 2006.
- [15] Q.-X. Huang and L. Guibas. Consistent shape maps via semidefinite programming. In *Computer Graphics Forum*, volume 32, pages 177–186. Wiley Online Library, 2013.
- [16] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 316–323. IEEE, 2000.
- [17] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):123–231, 2013.
- [18] H. H. Thodberg and H. Olafsdottir. Adding curvature to minimum description length shape models. In *British Machine Vision Conference, BMVC*, 2003.
- [19] J. Yan, J. Wang, H. Zha, X. Yang, and S. M. Chu. Multi-view point registration via alternating optimization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI’15*, pages 3834–3840. AAAI Press, 2015.
- [20] X. Zhou, M. Zhu, and K. Daniilidis. Multi-image matching via fast alternating minimization. *CoRR*, abs/1505.04845, 2015.
- [21] X. Zhou, M. Zhu, and K. Daniilidis. Multi-image matching via fast alternating minimization. *arXiv preprint arXiv:1505.04845*, 2015.