Nonlinear Shape-Texture Manifold Learning

Xiaokan WANG†, Nonmember, Xia MAO†(a), Member, and Catalin-Daniel CALEANU††, Nonmember

SUMMARY For improving the nonlinear alignment performance of Active Appearance Models (AAM), we apply a variant of the nonlinear manifold learning algorithm, Local Linear Embedded, to model shape-texture manifold. Experiments show that our method maintains a lower alignment residual to some small scale movements compared with traditional AAM based on Principal Component Analysis (PCA) and makes a successful alignment to large scale motions when PCA-AAM failed.

key words: manifold learning, shape-texture manifold, local linear embedded

1. Introduction

Active Appearance Models (AAM) as a model-based method has proved itself competent for many computer vision tasks that require an accurate alignment between an input image and a template, such as face expression recognizing, gesture analyzing, and lip analyzing. With a basic assumption that all of possible deformations of an object distribute in a low dimension manifold, traditional AAM decomposes each image into two dependent shape and appearance manifold, and constructs them on a training set by Principal Component Analysis (PCA). Aligning an input is treated as minimizing the residual between shape-texture and appearance. In addition to the classical appearance based face analyzing techniques, e.g. Eigenfaces [1], AAM brings improvements in real-time implementation [2]–[4], significance appearance variation performance [5], illumination standardization and pose normalization [6], [7]. However, the drawbacks of PCA-AAM focus on two aspects: firstly, some researches point out the shape alterations of many objects belong to a nonlinear manifold [8], [9] which PCA could not approximate very well [10], especially when it comes to some large scale movements such as opening mouth [11]; secondly, the alignment residual of PCA-AAM is globally spread over the image, which could cause higher alignment residual [12].

To overcome these two limitations of PCA-AAM, we turn to nonlinearity manifold learning, a research fever recent years, as it was reported in [13] and [14]. Based on the conception of approximating global nonlinearity with local linearity, nonlinear manifold learning is applied in many signal processing and data mining tasks, such as classification [15], face detection [16], [17], and data presentation [18]. Some nonlinear technologies are also merged into AAM. De la Torre et al. utilize kernel PCA to improve the nonlinear capability of original PCA [10]. C Mario Cristoudias et al. model shape and texture appearance manifold with two nonlinear methods: a mixture Gaussians models and a nearest-neighbor [19]. In our research, Local Linear Embedded (LLE) is chosen to take the place of PCA as the manifold analyzing tool to model human face shape-texture manifold for its concise expression and nice performance in nonlinear analyzing. On a test dataset, containing 162 human frontal face images, our method maintains a lower alignment residual dealing with some small scale movements and makes a successful alignment to large scale one when PCA-AAM failed.

2. Active Appearance Models

For intuition and simple calculation, linear models of shape-texture and appearance manifold are widely used in human face analyzing such as detecting, tracing and recognizing. Shape information indicates a series of feature points on human face which could make an outline of face contours, some samples are given in Fig.1, and the texture is the piecewise presentation of image based on triangles formed by feature points. Appearance is a “shape free” representation of object based on average shape. The fitting process of AAM can be described as

\[
\min_{x \in \Omega} \left[ A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) - I(W(x; p)) \right]^2
\]

in which, all of the calculations are within \( x_0 \), the average shape of training set, and \( A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) \) is the appearance information, \( A_i, i = 0, 1, 2, \cdots \) are bases of appearance space. \( I(W(x; p)) \) is the shape-texture part, where

Fig. 1 Training images with feature points,
W(x; p) is a piecewise wrap function mapping points x in the template image to a new position \( x' = W(x; p) \) in the input image according to parameter p which is the coordinate of input image in the shape space.

\[
s = s_0 + \sum_{i=1}^{\infty} p_i s_i = s_0 + p^T S_{base}
\]

(2)

In other words, assume that \( S \subseteq R^D \) is the point set of shapes, where \( D = 2n \), n is the number of feature points. The task of manifold learning is to find a mapping function \( f \) under which \( S \) is mapped to \( P \subseteq R^d \), \( d \ll D \), i.e.

\[
f(s) = p, \ s \in S, \ p \in P
\]

(3)

PCA treats \( f \) as a linear function and get the inverse of \( f \) though Eq. (2). Iain Matthews and Simon Baker make a full statement about AAM; furthermore, they proposed a more effective algorithm using Inverse Compositional (IC) algorithm to make AAM achieve real-time alignment which we also utilize in our method [2]. Although PCA makes AAM a success, the alignment results, presented in the experimental part of our work, performed on our test set, considering large scale movements, are not satisfactory.

3. Nonlinear Shape-Texture Modeling

For improving the nonlinear alignment performance of AAM, we utilize LLE to model nonlinear shape-texture manifold, which is based on the assumption that a low dimensional presentation of a manifold remains a local linearity relationship [13]. That is to say, \( S_k = \{s_i \} \subseteq S, i = 1, 2, \cdots, k \) is the neighbors of \( s_0 \in S \), and there exists a linear relation in the neighborhood i.e. \( s_0 = \sum_{i=1}^{\infty} w_i s_i \). If \( P_k = \{ p_i \} \subseteq P, p_i = f(s_i), i = 1, 2, \cdots, k \) is the corresponding points in low dimension and \( p_0 = f(s_0) \), then

\[
p_0 = \sum_{i=1}^{\infty} w_i p_i
\]

(4)

Generally speaking, LLE calculates \( f(s_0) \) by the local linear relationship, i.e. \( w_i, i = 1, 2, \cdots, k \). According to other LLE researches, we use L2 distance and “K-neighbor” which chose the k most nearest points as neighbors [14].

Besides, in the step of updating parameter circularly during calculating Eq.(1), there is a need for mapping from shape-texture manifold space back to compute shapes, which can be realized conveniently by PCA. However, most manifold learning methods do not have a reconstructing algorithm. For LLE, fortunately, the original point in the high dimension space can be recovered under the invariance of neighborhood representing, based on which we propose a reconstruction strategy described as below.

Assume that \( p_0 \in P, P_k = \{ p_1, p_2, \cdots, p_k \} \) is neighborhood matrix of \( p_0 \). Local linear relationship can be expressed as \( p_0 = \sum_{i=1}^{k} w_i p_i = P_k W', \) where \( W' = (w_1', w_2', \cdots, w_k')^T \) is the coefficient matrix denoted the local linear relationship. \( S_k = \{ s_1, s_2, \cdots, s_k \} \) is the corresponding points matrix of \( P_k \), i.e. \( p_i = f(s_i), i = 1, 2, \cdots, k \). Our purpose is to find the corresponding point \( s_0 \in R^D \) to fulfill \( f(s_0) = p_0 \), which could be realized in two steps as follows.

(1) Calculate local linear relationship using \( W = P_k^+ p_0 \), where \( P_k^+ \) is the generalized inverses of \( P_k \);
(2) Get original shape \( s_0 \) corresponding to \( p_0 \) according to the relationship \( W' \), that is \( s_0 = S_k W' = S_k P_k^+ p_0 \).

From the reconstruction course above, we can see the reconstruction processing is just like an inverse processing of LLE. The convenient “reversibility” of LLE is part of the reason we choose it. A simple experiment is designed to test the ability of our reconstruct method. Choosing 58 face images with manual labeled feature points showed in Fig. 1, we use LLE to get a low dimension space and reconstruct each shape with the same dimension applying our strategy. Figure 2 shows the sum of reconstruction errors at different dimension compared with PCA, seen from which, LLE performs better than PCA as the dimension increases. Figure 3 shows the full version of our LLE-AAM combined with IC algorithm. We will demonstrate some experimental results in the next section.

4. Experiment Results

We choose 58 pictures with manually labeled feature points
(see Fig. 1) for both PCA-AAM and LLE-AAM training, and test our algorithms on a one person dataset including 162, 640 × 480 frontal pictures.

We use a tree-based 20 × 20 gentle Adaboost frontal face detector [20] to gain the face position in every training image. First of all, we build PCA-AAM using IC algorithm described above, and to our training set, the dimension of shape-texture manifold space is destined for 11 so as to retain 95% of the model variance. Then, for convenient contrast, we assume the same dimension and \( k = 11 \) to construct our LLE-AAM. After that, we evaluate the capability of LLE-AAM along with different neighborhood sizes.

Three groups of shape and appearance alignment results are displayed in Fig. 4. The consequence for group A is both well for PCA-AAM and LLE-AAM. However, Fig. 5 shows the appearance alignment residuals according to iterations, which indicates that LLE-AAM has a lower residual compared with PCA-AAM when they converge. Group B shows an interesting result that PCA-AAM outputs a bad alignment while LLE-AAM still performs well, though PCA-AAM does well in almost the same pose in our experiment. It is possibly because PCA-AAM converges to some local minimum of Eq. (1) rather than the global one. If we regard both poses in group A and B as small scale movements, the third group turns to a large scale one, a face with an open mouth. The differences between the two models in Group C is obvious and enable us to conclude that LLE-AAM outperforms PCA-AAM with respect to our training set. All of the PCA-AAM results to large scale movements in our test set like open mouth are not very good while LLE-AAM does well. Different from the analyzing for group B, we suppose that maybe there exist some points in shape-texture manifold which PCA-AAM could not reach.

An alignment residuals box plot is drawn in Fig. 6 (a) to show a quantitative difference between two models. Computing over 162 test images, LLE-AAM has lower average fitting error and narrower data distribution than PCA-AAM.

We impute the poor performance of PCA-AAM to the simple global linearity assumption of shape-texture manifold. Additional residual is introduced when nonlinearity manifold is modeled by global linearity method, which can be seen from the alignment residual boxplot of PCA-AAM and LLE-AAM. Besides, under a uniform coordinate base, some shapes could not be figured very accurately just like Group C showed.

Finally, we demonstrate the alignment residuals under different neighborhood sizes, and take \( k = 10 \) and \( k = 11 \) for example in Fig. 6 (b). When the size of neighborhood decreases, LLE-AAM performs poorer, and we have to get rid of some test images for which the result of \( k = 10 \) is too bad. This is possibly because the fewer neighbors can not express intrinsic property of the shape-texture manifold sufficiently.

5. Conclusion and Future Works

In this work, we have introduced a nonlinearity shape-texture model to improve the performance of AAM. Within our test dataset it is possible to conclude that our method has
lower fitting error than PCA-AAM to small degree of face transforming and makes more accurate alignments when it comes to large scales change of face. Future works may focus on the exploration of reducing the dimension of LLE for saving memory and increasing computing efficiency.

Acknowledgements

This work is supported by the National Nature Science Foundation of China (No.60873269), Bilateral Inter-Governmental S&T Cooperation between China and Romania (No.39-5) and UEFISCSU 599/19.01.2009.

References

[1] M. Turk and A. Pentland, “Eigenfaces for recognition,” J. Cognitive Neuroscience, vol.3, no.1, pp.71–86, 1991.
[2] I. Matthews and S. Baker, “Active appearance models revisited,” Int. J. Comput. Vis., vol.60, no.2, pp.135–164, 2004.
[3] S. Baker and I. Matthews, “Equivalence and efficiency of image alignment algorithms,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001.
[4] L. Matthews, T. Ishikawa, and S. Baker, “The template update problem,” IEEE Trans. Pattern Anal. Mach. Intell., vol.26, no.6, pp.810–815, 2004.
[5] G. Papandreou and P. Maragos, “Adaptive and constrained algorithms for inverse compositional active appearance model fitting,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2008.
[6] F. Kahraman, M. Gokmen, S. Darkner, and R. Larsen, “An active illumination and appearance (aia) model for face alignment,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–7, 2007.
[7] F. Kahraman, B. Kurt, and M. Gokmen, “Robust face alignment for illumination and pose invariant face recognition,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–7, 2007.
[8] V.I. Morariu and O.I. Camps, “Modeling correspondences for multicamera tracking using nonlinear manifold learning and target dynamics,” IEEE Conference on Computer Vision and Pattern Recognition, p.545-C552, 2006.
[9] C.S. Lee and A. Elgammal, “Nonlinear shape and appearance models for facial expression analysis and synthesis,” 18th International Conference on Pattern Recognition, p.313-C320, 2006.
[10] F. de La Torre and H.N. Minh, “Parameterized kernel principal component analysis: Theory and applications to supervised and unsupervised image alignment,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2008.
[11] F. Kahraman, B. Kurt, and M. Gokmen, “Robust face alignment for illumination and pose invariant face recognition,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–7, 2007.
[12] W. Hao, L. Xiaoming, and G. Doretto, “Face alignment via boosted ranking model,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2008.
[13] S.T. Roweis and L.K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” SCIENCE, vol.290, no.5500, pp.2323–2356, 2000.
[14] J.B. Tenenbaum, V. de Silva, and J.C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” SCIENCE, vol.290, no.5500, pp.2319–2323, 2000.
[15] O. Tuzel, F. Porikli, and P. Meer, “Pedestrian detection via classification on riemannian manifolds,” IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.10, pp.1713–1727, 2008.
[16] M.H. Yang, “Face recognition using extended isomap,” 2002 International Conference on Image Processing, p.117-C120, 2002.
[17] L. Wei-Kai and G. Medioni, “3d face tracking and expression inference from a 2d sequence using manifold learning,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2008.
[18] A. Elgammal and C.S. Lee, “Separating style and content on a nonlinear manifold,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004.
[19] C.M. Christoudias and T. Darrell, “On modelling nonlinear shape-and-texture appearance manifolds,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.2, pp.1067–1074, 2005.
[20] R. Lienhart, A. Kuranov, and V. Pisarevsky, “Empirical analysis of detection cascades of boosted classifiers for rapid object detection,” Lect. Notes Comput. Sci., pp.297–304, 2003.

Fig. 6  Boxplot of alignment residual. The horizontal lines of each box stand for the top quartile, median and bottom quartile alignment residual values in the test dataset, and the data outlier is labeled by the cross.