

Component-Based Active Appearance Models for Face Modelling

Cuiping Zhang and Fernand S. Cohen

Electrical and Computer Engineering Department,
Drexel University, Philadelphia PA 19104, USA
{zcp, fscohen}@cbis.ece.drexel.edu

Abstract. The Active Appearance Model (AAM) is a powerful tool for modelling a class of objects such as faces. However, it is common to see a far from optimal local alignment when attempting to model a face that is quite different from training faces. In this paper, we present a novel component-based AAM algorithm. By modelling three components inside the face area, then combining them with a global AAM, face alignment achieves both local as well as global optimality. We also utilize local projection models to locate face contour points. Compared to the original AAM, our experiment shows that this new algorithm is more accurate in shape localization as the decoupling allows more flexibility. Its insensitivity to different face background patterns is also clearly manifested.

1 Introduction

Face recognition has received a lot of attention in the past decades. Detecting a face and aligning its facial features are usually the first step, therefore crucial for most face applications. Among numerous approaches, the Active Appearance Model (AAM) [1] and the Active Shape Model (ASM) [2] are 2 popular generative models that share a lot in common. As a successor of the ASM, the AAM is computationally efficient and has been intensively studied by many researchers.

The AAM has several inherent drawbacks as a global appearance based model. First, It has a simple linear update rule stemming from a first order Taylor series approximation of an otherwise complex relationship between the model parameters and the global texture difference. Clearly, any factor that is part of the global texture will affect the AAM's performance (examples are the global illumination, partial occlusions, etc.). In a converged AAM, the local alignment results may need further refinement to meet the accuracy requirement of many applications. Secondly, the gradient descent information near the face contour seeps the background pattern in the training set. Hence, the AAM can't perform well for test face images with unseen backgrounds.

With all these problems associated with the AAM in mind, in this paper, we propose a component-based AAM that groups landmark points inside the face area into three natural components in addition to a globally defined AAM. The independence of the sub-AAMs leads to a more accurate local alignment

result. For the model points on a face contour, a strategy similar to the ASM is adopted. The ASM iteratively adjusts any model point along its normal direction so that an associated texture pattern is in accordance with a typical distribution. Our new method makes full use of what we already have during the AAM procedure and local projection models are built on a standard shape frame. The revised projection models, together with the component-based analysis improve the overall modelling performance, especially on the test set.

The paper is organized as follows: In section 2, the original AAM is briefly introduced. Section 3 presents the idea of the component-based AAM. In section 4, details about our local projection models are given. Section 5 presents our experiment results and discussions. Last section is the conclusion.

2 AAM Basic Idea

In the AAM, a face's shape is defined as a sequence of the coordinates of all landmark points. Let \mathbf{S}_0 be the mean shape of all training images. A shapeless texture vector is generated after warping the face patch inside the convex hull of all landmark points to the mean shape. Fig. 1(a) shows a face image overlapped with landmark points and the resulted shapeless texture is in Fig. 1(c).

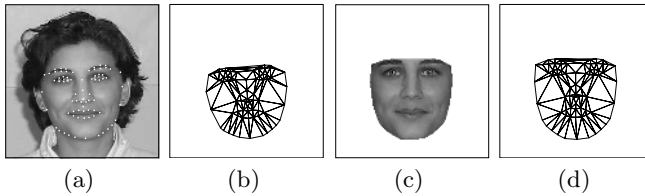


Fig. 1. (a)landmark points. (b)Face mesh. (c)Shapeless texture. (d)Base face mesh.

All raw shape vectors need to be aligned to a common coordinate system. This normalization introduces a similarity transformation between a original face vector on the image frame and its normalized one on the model frame. Similarly, All raw texture vectors also undergo offsetting and scaling operations for normalization purpose.

PCA is used to model the shape and texture variations. A normalized shape \mathbf{x} and texture \mathbf{g} can be formulated as $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_s \cdot \mathbf{b}_s$ and $\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \cdot \mathbf{b}_g$. The column vectors in the matrices \mathbf{P}_s and \mathbf{P}_g are the principal modes for the shape and texture variations of the training set. They span the shape and texture subspaces, respectively. The vectors \mathbf{b}_s and \mathbf{b}_g , as the projected coefficients in the subspaces, are named shape and texture parameters. They can be concatenated for further de-correlation in a mixed eigen-subspace and the projected coefficients \mathbf{c} encode both shape and texture information. The reconstruction of the shape vector \mathbf{x} and the texture vector \mathbf{g} from \mathbf{c} is straightforward and easy.

The complete appearance parameter set includes 4 similarity pose parameters Ψ due to the coordinate normalization and the mixed parameter vector \mathbf{c} , i.e., $\mathbf{p} = \{\Psi, \mathbf{c}\}$. Modelling an unknown face in a test image is a process searching for the optimal appearance parameter set that best describes the face. In an iterative search, let the texture residual $\mathbf{r}(\mathbf{p})$ (also referred as difference image) be the difference of the reconstructed model texture \mathbf{g}_m and the texture \mathbf{g}_s extracted from the test image, $\mathbf{r}(\mathbf{p}) = \mathbf{g}_s - \mathbf{g}_m$. The matching error is measured as the RMS of the texture residual $\mathbf{r}(\mathbf{p})$. The AAM assumes a linear relationship between $\mathbf{r}(\mathbf{p})$ and the update for the model parameters $\delta\mathbf{p}$: $\delta\mathbf{p} = -\mathbf{R} \cdot \mathbf{r}(\mathbf{p})$, where \mathbf{R} is a constant gradient descent matrix estimated from the training images [1]. As face image backgrounds are encoded, it is suggested to use a random background so that \mathbf{R} is irrelevant of the background patterns in the training set [1]. However, useful heuristic information for face contour is also lost as a result.

3 Component-Based AAM

Based on the fact that local shape depends only on local appearance pattern, we propose a Component-based AAM in an effort to gain better feature localization. The basic idea is to group landmark points to components and train the local models independently. To avoid possible confusion, we refer to the original AAM as the global AAM. Three components on the mean shape frame are highlighted in Fig. 2(a). Landmark points are naturally grouped to balance the added computational cost and algorithm efficiency. Columns 2(b) to 2(d) show the components of the person in Fig. 1(a). The top row shows local shapes and the bottom row shows warped shapeless textures.

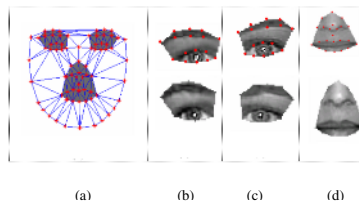


Fig. 2. (a) Definition for local components. (b) Left eyebrow and eye. (c) Right eyebrow and eye. (d) Nose and mouth.

Our component-based AAM is a combination of one global AAM and three sub-models. As part of the global face patch, all components are normalized to the same common coordinate system as that for the global face. This establishes clear correspondence between the global model and the sub-models. Not only all sub-models share the same 2D pose parameters as the whole face, but the component shapes, textures and texture residuals are just fixed entries in their counterparts of the global model. Sub-models are trained separately.

During the modelling process, the component-based AAM algorithm switches between the global model and the sub-models alternatively. After one iteration of

the global AAM, we have current estimates of: the global shape \mathbf{x} , the texture \mathbf{g} , the texture residual $\mathbf{r}(\mathbf{p})$ and the global matching error is e_0 . The various steps to model local components are detailed as follows(for i th sub-model, $i = 1$ to 3):

- **Global to local mapping:** Generate the sub-model shape \mathbf{x}_i , texture \mathbf{g}_i and texture residual $\mathbf{r}_i(\mathbf{p}_i)$ by looking up the fixed entries in \mathbf{x} , \mathbf{g} , and $\mathbf{r}(\mathbf{p})$. Project $\{\mathbf{x}_i, \mathbf{g}_i\}$ onto local subspaces, $p_i = \{\Psi, \mathbf{c}_i\}$.
- **Local AAM prediction:** Apply local AAM to obtain new sub-model shape vector \mathbf{x}'_i , texture vector \mathbf{g}'_i and local 2D pose Ψ_i .
- **Local to global mapping:** Use $\{\mathbf{x}'_i, \mathbf{g}'_i\}$ to update the corresponding entries of the global texture vector \mathbf{g} and component points on the image frame.
- **Decision making:** If the new global parameters lead to a smaller matching error, accept the update.

In summary, sub-models update component points independently. Meanwhile, they are united and confined within a global AAM. In this way, error propagation between local components is reduced and modelling ability is enhanced locally. In [3], sub-models are constructed to model vertebra. However, they basically repeat the same sub-model for a sequence of triplet vertebrae and propagate their results, therefore different from our approach.

4 Local Projection Models

When a test face is presented in a background unseen in the training set, the AAM often fails, especially for face contour points. Since landmark points on a face contour are usually the strongest local edge points, we developed a method similar to the ASM to complete our component-based AAM. The ASM moves a landmark point along local normal direction so that its profile conforms to a typical distribution. Instead of using the edge strength along the profile directly, we believe that edge information would be more prominent and stable after taking the local average. Further, we associate our local projection models with the triangulation result of the landmark points. Fig. 3(a) is the mesh of landmark points for the person in Fig. 1(a). Fig. 3(b) shows the mean shape.

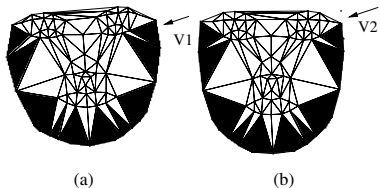


Fig. 3. Mesh definition. (a)Shape of the person in Fig. 1(a). (b)Base shape mesh.

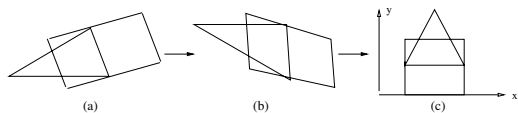


Fig. 4. Triangle-parallelgram pairs. a) Original image frame. b) Mean shape frame. c)Standard pair.

Triangles sitting on the face boundary form a special "face component". They are filled with black color. Their bottom sides form the face contour. Assume each black triangle is associated with a parallelogram with the bottom side of the triangle being the parallelogram's middle line. Our local projection models are built based on the analysis of the edge map inside these parallelograms. Fig. 4 illustrates how a triangle V_1 on the face image is transformed to V_2 on the base frame and subsequently to V_0 on a standard triangle which is an isosceles triangle. After these transformations, any projection along the face contour direction in the face image is now simplified to a summation along the x (or y) axis.

The piece-wise affine transform parameters between V_1 and V_2 are available in the basic AAM model fitting process. The transforms between V_0 and all the triangles of the base shape could be computed in advance. Clearly, with the help of the base shape and a standard triangle-parallelogram pair, local projection models can lock face contour points to the locally strongest edge points. It is much easier and robust compared to the ASM. The regions of interest for the local projection models are proportional to current face landmark points. Therefore there is no scaling problem at all.

5 Experiment Results and Discussion

Our face database includes 138 nearly frontal images from various different face databases[4][5][6]. All images were roughly resized and cropped to 256 by 256. We believe a blend face database is the best way to robust test. We sequentially picked 80 images to train the face shape subspace and rest of them as the test set.

We also tested on the Japanese Female Facial Expression database (JAFFE)[7], which contains 213 images of 7 facial expressions of 10 female models. The only pre-processing we conducted is to scale original 200 by 200 images to standard size of 256 by 256.

In an iterative realization, the global AAM is run first and when it fails to converge, local sub-models are launched, followed by the local projection models to lock the real face boundary. Search will stop when the stopping criterions are met.

To evaluate the fitting quality, we manually labelled all landmark points and created distance maps for all images. The model fitting quality is then measured by the average point to edge distance. Within the same framework, we tested and compared three different algorithms: the AAM search; the AAM with the component analysis (AAM_CA); the AAM with the component analysis and the local projection models (AAM_CA_LPM).

5.1 Component-Based AAM Search

Fig. 5 compares AAM and AAM_CA model fitting results. As expected, a converged global AAM usually couldn't achieve optimal local alignment. Better localization of the component feature points could be seen on the bottom row. Table 1 shows average point to edge errors for algorithms with and w/o the component analysis. Only face component points are considered.

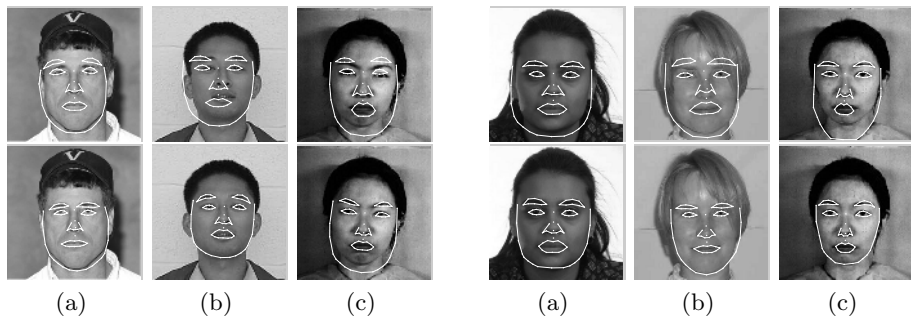


Fig. 5. AAM (top row) versus AAM_CA (bottom row). (a) Training set. (b) Test set. (c) JAFFE database. **Fig. 6.** AAM_CA (top row) versus AAM_CA_LPM (bottom row). (a) Training set. (b) Test set. (c) JAFFE database.

Table 1. Average error(contour excluded) **Table 2.** Average error(contour only)

Algorithms	Training	Test	JAFFE
AAM	2.0661	3.5513	3.1696
AAM_CA	1.8988	3.2429	2.9377

Algorithms	Training	Test	JAFFE
AAM_CA	3.5298	4.7153	7.4741
AAM_CA_LPM	3.2909	3.8430	4.1356

5.2 Face Contour Detection with Local Projection Models

We compared AAM_CA and AAM_CA_LPM model fitting results to show how the integration of local projection models can help to solve the boundary problem. Fig. 6 shows some examples. Table 2 compares the average point to edge errors. It is interesting to see in Fig. 5(b), boundary points are correctly aligned due to the component analysis. Also Fig. 6(b) has correct component points. Apparently the integration of the local AAM analysis and local projection models makes our fitting algorithm more accurate and robust.

Convergent rate curves are compared for different algorithms in Fig. 7. A good approximation of an error density function can be obtained from the histogram of the resulted point errors for all images. Given a number ε in x-axis,

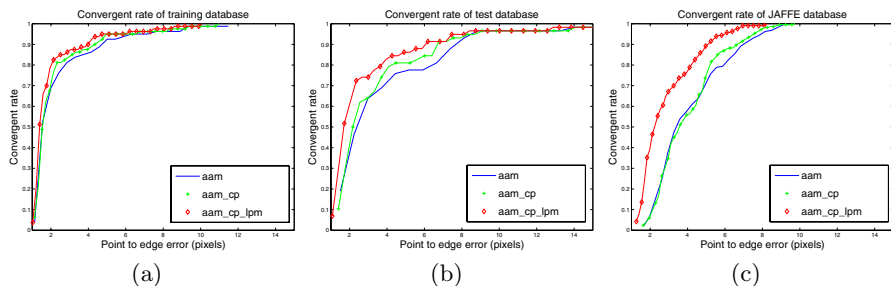


Fig. 7. Curves of convergent rate versus error threshold. (a) Training set. (b) Test set. (c) JAFFE database.

y-axis gives the percentage of images with errors smaller or equal to ε . Clearly AAM_CA_LPM has the best performance and the improvement is especially prominent for the JAFFE database.

6 Conclusion

In this paper, we proposed a component-based AAM algorithm to deal with the lack of the accuracy of feature localization in the original AAM. All component sub-models and the local projection models are tightly combined and smoothly interact with the global AAM model by sharing intermediate results. Robust and accurate face alignment makes it possible to extend the research to face recognition, 3D modelling etc. Extending our algorithm to images taken from different viewpoints is straightforward.

References

1. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. *IEEE Trans. PAMI* **23** (2001) 681–685
2. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models: Their training and application. *CVGIP: Imaging Understanding* **61** (1995) 38–59
3. Roberts, M., Cootes, T., Adams, J.: Linking sequences of active appearance sub-models via constraints: an application in automated vertebral morphometry. In: 14th British Machine Vision Conference. Volume 1. (2003) 349–358
4. Zhang, C., Cohen, F.: Face shape extraction and recognition using 3d morphing and distance mapping. In: 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France (2000)
5. Phillips, P., Moon, H., Rauss, P., Rizvi, S.: The feret evaluation methodology for face recognition algorithms. In: Proceedings of IEEE Computer Vision and Pattern Recognition. (1997) 137–143
6. Unknown: The psychological image collection at stirling.
(<http://pics.psych.stir.ac.uk/>)
7. Lyons, M.J., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan (1998) 200–205