Robust Active Appearance Model Based Upon Multi-linear Analysis against Illumination Variation

Gyeong-Sic Jo, Hyeon-Joon Moon, and Yong-Guk Kim

School of Computer Engineering, Sejong University, Seoul, Korea
ykim@sejong.ac.kr

Abstract. Independent Active Appearance Model (AAM) has been widely used in face recognition, facial expression recognition, and iris recognition because of its good performance. It can also be used in real-time system application since its fitting speed is very fast. When the difference between the input image and the base appearance of AAM is small, the fitting is smooth. However, when the difference can be large because of illumination and/or pose variation in the input image, the fitting result is unsatisfactory. In this paper, we propose a robust AAM using multi-linear analysis, which can make an Eigen-mode within the tensor algebra framework. The Eigen-mode can represent the principal axes of variation across the order of tensor and it can apply to AAM for increasing robustness. In order to construct both of original AAM and the present AAM, we employ YALE data base, which consists of 10 subjects, 9 poses, and 64 Illumination variations. The advantage of YALE data base is that we can use the coordinate of landmarks, which are marked for train-set, with ground truth. Because when the subject and the pose were same, the location of face isalso same. We present how we construct the AAM and results show that the proposed AAM outperforms the original AAM.

Keywords: AAM, YALE data base, Multi-linear Analysis, Eigen-mode, Tensor.

1 Introduction

The Active Appearance Model (AAM) is a non-linear, generative, and parametric model for the certain visual phenomenon. And it is used for face modeling frequently as well as for other object modeling. The AAM is proposed in [1] firstly, and then improved in [2], which model shape and appearance separately. The AAM is computed by train-set, which consists of pair of images and land marks, which is marked manually by hand.

Generally the AAM fitting has performed successfully when error rate between input image and base appearance of the AAM is low. However, when error rate becomes high when illumination and 3D pose are changing, and its fitting result is unsatisfactory. In this paper, we propose a new AAM which contains Eigen-mode based upon multi-linear analysis. The multi-linear analysis is extension of Singular Value Decomposition(SVD) or PCA, and offers a unifying mathematical framework

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2009, LNCS 5615, pp. 667–673, 2009.

suitable for addressing a variety of computer vision problems. The multi-linear analysis builds subspaces of orders of the tensor and a core tensor. The advantage of multi-linear analysis is that the core tensor can transform subspace into Eigen-mode, which represent the principal axes of variation across the various mode (people, pose, illumination, and etc)[9]. In contrast, PCA basis vectors represent only the principal axes of variation across images. In other words, Eigen-mode covers the variation of each mode but PCA vectors cover all variations. We can build the AAM which includes not only the principal axes of variation across images but also variation across the various modes. To include the variation across the various modes, the AAM can contain the variations for several modes.

This paper is organized as follow. Section 2 and 3 explain AAM and multi-linear analysis. Then, in section 4, we describe the method how to apply multi-linear analysis to AAM. Finally, we are going to show our experimental results and summarize our work in sections 5.

2 Independent Active Appearance Models

Independent AAM models the shape and appearance separately [2]. The Shape of AAMs is defined by a mesh located at a particular the vertex location. Because AAM allows a linear shape variation, we can define the shape as follow:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{n} p_i \mathbf{s}_i. \tag{1}$$

In equation (1), the coefficients p_i indicate the shape parameters. $\mathbf{s}_0$ indicates a base shape, and $\mathbf{s}_i$ represent shape vectors. The shape vectors can be obtained by applying PCA to train-set after using Procrustes analysis in order to normalize the landmarks. The appearance of AAMs is defined within the base shape $\mathbf{s}_0$. This mean that pixels in image lie inside the base shape $\mathbf{s}_0$. AAMs allow a linear appearance variation. Therefore we can define the appearance as follow:

$$\boldsymbol{A}(\mathbf{x}) = \boldsymbol{A}_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \boldsymbol{A}_i(\mathbf{x}). \tag{2}$$

Where λ_i indicate the appearance parameters, $\boldsymbol{A}_i$ represent the appearance vectors, and $\boldsymbol{A}_0$ is a base appearance. After finding both the shape parameters and the appearance parameters, the AAMs instance can be generated by locating each pixel of appearance to the inner side of the current shape with piecewise affine warp. A model instance can be expressed as equation (3):

$$M\big(\mathbf{W}(\mathbf{x};\mathbf{p})\big) = \mathrm{A}(\mathbf{x}) \tag{3}$$

The parameters of both shape and appearance are obtained by a fitting algorithm.

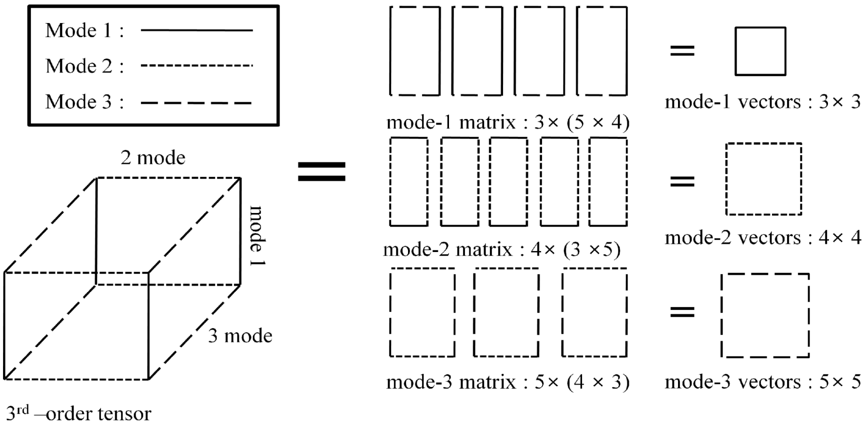

Fig. 1. Unfolding a 3^{rd} – order tensor of dimension 3x4x5

3 Multi-linear Analysis

3.1 Tensor Algebra

Multi-linear analysis is based on higher-order tensor. The tensor, well-known as n-way array or multidimensional matrix or n-mode matrix, is a higher order generalization of a vector and matrix. A higher order tensor N is could be given by $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$. Therefore the order of vector, matrix, and tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is 1^{st}, 2^{nd}, and N^{th} , respectively. In order to manipulate the tensor easily, we should unfold the tensor to matrix $A_{(n)} \in \mathbb{R}^{I_n \times (I_1, I_2, \ldots I_{n-1} I_{n+1}, \ldots I_N)}$ by stacking its mode-n vectors to column of the matrix. Figure.1 shows the unfolding process.

The mode-*n* product of a higher order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ by a matrix $M \in \mathbb{R}^{I_n \times J_n}$ is a tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N}$, which can be denoted by $\mathcal{B} = \mathcal{A} \times {}_n M$, and its entries are computed by

$$(\mathcal{A} \times {}_n M)_{i_1 \times \ldots \times i_{n-1} \times j_n \times i_{n+1} \times \ldots \times i_N} = \sum_{i_n} a_{i_1 \times \ldots \times i_{n-1} \times j_n \times i_{n+1} \times \ldots \times i_N} m_{j_n i_n}. \tag{4}$$

This mode-*n* product of tensor and matrix can be represented in terms of unfolded matrices,

$$B_{(n)} = MA_{(n)}. \tag{5}$$

3.2 Tensor Decomposing

In order to decompose the tensor, we employee Higher Order Singular Value Decomposition (HOSVD). HOSVD is an extension of SVD that expresses the tensor as the mode-*n* product of N-orthogonal spaces

$$\mathcal{D} = \mathcal{Z} \times_1 \mathrm{U}_1 \times_2 \dots \times_n \mathrm{U}_n \dots \times_N \mathrm{U}_N. \tag{6}$$

In equation (6), U_n are mode matrix that contains the orthonormal vectors spanning the column space of the matrix $\mathrm{D}_{(n)}$ which is result of unfolding tensor $\mathcal{D}$. Tensor $\mathcal{Z}$, called the core tensor, is analogous to the diagonal singular value matrix of conventional matrix SVD, but it is does not have a diagonal structure. The core tensor is in general a full tensor. The core tensor governs the interaction between the mode matrix U_n, where n is 1, 2, …, N. Procedure of tensor decomposition using HOSVD can be expressed as follows

- Compute the SVD of unfolded matrix $\mathrm{D}_{(n)}$ and set up matrix U_n with left singular matrix of SVD.
- Solve for the core tensor as follows

$$\mathcal{Z} = \mathcal{D} \times_1 \mathrm{U}_1^{\mathrm{T}} \times_2 \dots \times_n \mathrm{U}_n^{T} \dots \times_N \mathrm{T}_N^{T}. \tag{7}$$

4 Applying Multi-linear Analysis to AAMs

In section 2 and 3, we described AAMs and multi-linear analysis. Now, we explain how to apply multi-linear analysis to AAMs. Since, in independent AAMs, the appearance vectors of AAMs influence the fitting result poorly and the shape of AAMs is not influenced by changing illumination, we consider identify and pose for AAMs.

To build AAMs using multi-linear analysis, we construct a third-order tensor $\mathcal{D} \in \mathbb{R}^{\mathrm{I}\times\mathrm{J}\times\mathrm{K}}$ to represent identity, poses, and features. Using HOSVD, we can decompose the tensor $\mathcal{D}$ into three factors as follows

$$\mathcal{D} = \mathcal{Z} \times_1 \mathrm{U}_{id} \times_2 \mathrm{U}_p \times_3 \mathrm{U}_f, \tag{8}$$

where $\mathcal{Z}$ is the core tensor that governs the interaction among the three mode matrices(U_{id} , U_p, and U_f). Using core tensor and mode matrix U_f, we can build eigen-mode $\bar{\mathbf{s}}$ as

$$\bar{\mathbf{s}} = \mathcal{Z} \times_3 \mathrm{U}_f. \tag{9}$$

Since the AAM add Eigen-mode, we rewrite equation (1) as follow:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{n} p_i \mathbf{s}_i + \sum_{j=1}^{m} r_j \bar{\mathbf{s}}_j, \tag{10}$$

where r_j are parameters of Eigen-mode.

The advantage of the AAMs based upon Eigen-mode is that the shape is stable under higher error rate between base appearance and input image, which can be happened by changing illumination and pose because each Eigen-mode considers only each mode, not all train-set. Figure 2 compares the fitting results between the present AAM and the traditional AAM.

Fig. 2. The fitting results of the present AAM(bottom) and traditional AAM(top).

In Figure 2, the shape of traditional AAM(top) is not able to cover the darker region with the face. On the other hand, the shape of the present AAM(bottom) is covering the darker region very well with the face.

5 Experiments and Evaluation

We employ YALE face data base B[8], which is consisted of 10 subjects, 9 poses, and 64 Illuminations, for AAM training and experiment. In YALE face data base, when the subject and the pose are the same, the location of face is also same although there is changing illumination. This property allows that we use landmarks, marked for train-set, with ground truth, because the coordinates of landmarks is not changed in a category which is the same for the subject and the pose.

In order to build both of AAMs, we have constructed train-set which consists of images in 9 subjects, 9 poses, and 1 illumination and meshes made by marking 64 landmarks on each image. Ground truth was established by meshes for train-set and images have deferent subject and pose. Experiments were divided into two evaluations: one was a test about how speedily each AAMs ran fitting algorithm, and another was an evaluation about how correctly each model fitted for image.

5.1 Efficiency Comparison

The fitting speed of AAMs is important for applying AAM to real-time system. We have compared the fitting speed of both of AAMs, which is performed based on Quad core computer with CPU 2.4GHz and RAM 2GB. The fitting algorithm was run for 5 iterations. We measured the spent time for running fitting algorithm per iteration and all iteration. The traditional AAM used 11 parameters (4 global transform parameters and 7 local transform parameters), and our AAM used 18 parameters (4 global transform parameters, 7 local transform parameters, and 7 mode transform parameters).

Table 1. the speed of fitting algorithm for both AAMs

	1	2	3	4	5	Avg.
Traditional AAM	7ms	6ms	6ms	6ms	7ms	6.4ms
Present AAM	7ms	7ms	7ms	7ms	7ms	7ms

Table 1 illustrates that the elapsed times are similar, although our AAM used more parameters than traditional AAM.

5.2 Robustness Experiments

We have evaluated about how our AAM correctly fits for images under higher error rate between base appearance and input image. Our evaluation procedure can be expressed as follows:

- Dividing images into 4 categories. Each category consists of images which have the average pixel errors 40~49, 50~59, 60~69, and 70~79, respectively.
- Fitting for images, and then we measure coordinate errors between the ground truth and the fitted shape, per iteration.

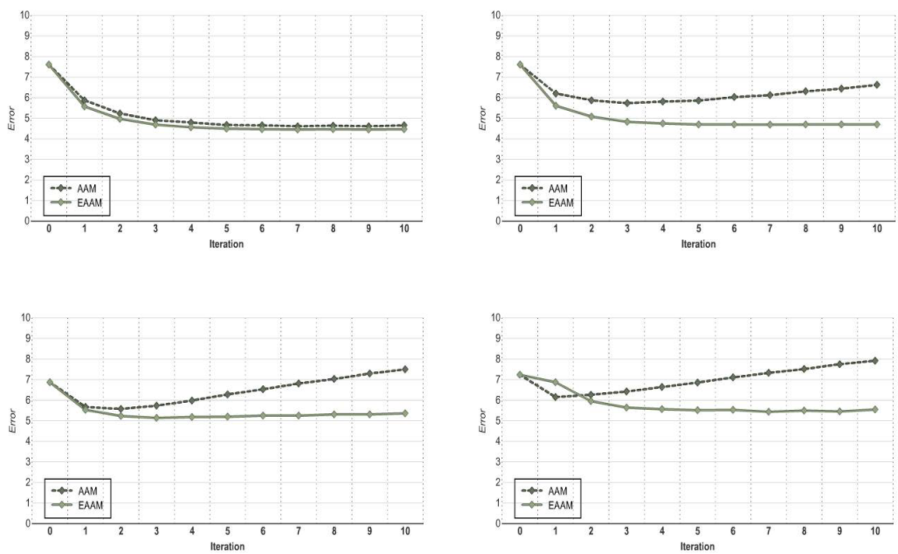

Fig. 3. Fitting error of both of AAMs. Each graph represents shape error per iteration under pixels error 40~49(top left), 50~59(top right), 60~69(bottom left), and 70~79(bottom right).

We employed L1 norm for measuring coordinate errors. In Figure 3, each graph represents the fitting error per iteration for categories. When average pixels error is increasing, the fitting error of traditional AAMs is also increasing, but our AAM is not increasing the fitting error.

6 Conclusion

In this paper, we proposed a AAM based upon Eigen-mode. In order to establish that AAM, we have built the Eigen-mode using multi-linear analysis, that employs HOSVD to decompose the tensor. We have shown that the present AAM has ability to fit for image speedily, even though parameters are increased. And it can fit for image under higher error rate. Since the present AAM is fast in fitting diverse images, it could be applied to any real-time systems. We plan to apply out AAM to real-time system to recognize face and facial expression tasks.

Acknowledgement

This work was supported by the Seoul R&BD Program (10581).

References

1. Edwards, G.J., Taylor, C.J., Cootes, T.F.: Interpreting Face Images Using Active Appearance Models. In: Proc. International Conference on Automatic Face and Gesture Recognition, pp. 300–305 (June 1998)
2. Matthews, I., Baker, S.: Active Appearance Models revisited. International Journal of Computer Vision, 135–164 (2004)
3. Cootes, T., Edwards, G., Taylor, C.: Active appearance Models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6) (2001)
4. Gross, R., Matthews, I., Baker, S.: Constructing and fitting active appearance models with occlusion. In: Proceedings of the IEEE Workshop on face processing in Video (June 2004)
5. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Congitive Neuroscienc 3(1), 71–86 (1991)
6. De Lathauwer, L., De Moor, B., Vandewalle, J.: A Multilinear Singular Value Decomposition. SIAM Journal of Matrix Analysis and Applications 21(4) (2000)
7. Vasilescu, M.A.O., Terzopoulos, D.: Multilinear analysis of image ensembles: TensorFaces. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 447–460. Springer, Heidelberg (2002)
8. Georghiades, A.S., Belhumeur, P.N., Kriegman, K.J.: From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. IEEE Trans. Pattern Analysis and Machine Intelligence 23(6), 643–660 (2001)