

Research Article

An Improved AAM Method for Extracting Human Facial Features

Tao Zhou,¹ Xiao-Jun Wu,¹ Tao Wu,¹ and Zhen-Hua Feng^{1,2}

¹ School of IoT Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

² CVSSP, University of Surrey, Guildford GU2 7XH, UK

Correspondence should be addressed to Xiao-Jun Wu, wu_xiaojun@yahoo.com.cn

Received 7 July 2011; Accepted 24 October 2011

Academic Editor: Wenyu Sun

Copyright © 2012 Tao Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Active appearance model is a statistically parametrical model, which is widely used to extract human facial features and recognition. However, intensity values used in original AAM cannot provide enough information for image texture, which will lead to a larger error or a failure fitting of AAM. In order to overcome these defects and improve the fitting performance of AAM model, an improved texture representation is proposed in this paper. Firstly, translation invariant wavelet transform is performed on face images and then image structure is represented using the measure which is obtained by fusing the low-frequency coefficients with edge intensity. Experimental results show that the improved algorithm can increase the accuracy of the AAM fitting and express more information for structures of edge and texture.

1. Introduction

Facial features extraction is an essential technique for image analysis and interpretation, which is usually used for face recognition, face reconstruction, and expression analysis [1, 2]. In recent years, finding a precise and efficient method for extracting facial features has gained more concern in pattern recognition. Wiskott et al. [3] utilized Gabor filters for face recognition and proposed the concept of elastic graph for facial features extraction. Beinglass and Wolfson [4] proposed a new extraction method in using generalized hough transform by each clear feature point being as the reference point of graph. Then some new methods based on symmetry transform and multiscale symmetry transform were applied for extracting facial features.

Currently, statistics-based flexible models, especially the active appearance model (AAM) [5] proposed by Cootes et al. have gained great success in this area. AAM is an efficient fitting algorithm for extracting features. The goal of the AAM search is to find the model parameters that generate a synthetic image as close as possible to a given input image and to

use the resulting AAM parameters for interpretation. Fitting the model and the target image is treated as a nonlinear optimization problem. Therefore, the fitting task requires a huge amount of computation when the standard nonlinear optimization techniques such as the gradient descent method are used. Unfortunately, these algorithms were inefficient because of the huge complexity. Matthews and Baker [6] combined AAM with the Lucas-Kanade algorithm [7] to build an efficient AAM fitting method. The method can be implemented very efficiently because the Hessian matrix and the Jacobian matrix can be precomputed.

However, the original formulation of the AAM using image intensity values to model the texture variations are very sensitive to changes in imaging conditions such as lighting and camera parameters. Many researchers pointed out the drawback and have paid much attention to the texture representation. Wolstenholme and Taylor [8] introduced wavelet-based method to represent the texture using as few dimensions as possible. This method can efficiently decrease image dimensions and keep fitting accuracy. Gao et al. [9] utilized Gabor filters to represent the image texture, the Gabor representations lead to more accurate fitting. In addition, other researchers proposed some measures which include the direction of gradient, the edge, and the corner of training images at each pixel. Cootes and Taylor [10] proposed to model the local orientation of structures using the appearance of edge strength. Kittipanya ngam and Cootes [11] proposed representation of the image structure in a region which can improve fitting performance. In a word, all above methods can improve the fitting accuracy of AAM.

All the above improved AAM algorithms considered the texture and local structure separately, we now consider to propose a new representation which combines the edge intensity and the texture information. The edge intensity of image reflects the discontinuity of gray-value, where it is less sensitive to imaging conditions. In this paper, translation invariant wavelet transform is performed on face images and then image texture representation is obtained by the low-frequency coefficients, which are fused with edge intensity. Experimental results obtained on IMM database and Weizmann database show that the improved algorithm can express more information for structures of edge and texture and thus leading to more accurate fitting.

The remainder of this paper is organized as follows. Section 2 concisely introduces the backgrounds of original AAM and translation invariant wavelet transform. Then, we show an improved AAM algorithm in Section 3. Experimental results are presented in Section 4. Conclusions are drawn in Section 5.

2. Background

2.1. Active Appearance Model (AAM)

AAM is a statistical model including two aspects: model building and fitting algorithm. AAM learns characteristics of objects by building a compact statistical model. The statistical model, which represents shape and texture variations of objects is obtained from applying principal component analysis (PCA) to a set of labeled data.

In AAM, an object is described by a set of landmarks which indicate important positions at the boundary of the object. Landmarks are labeled manually on each object in the training data. A shape \mathbf{s} is defined as coordinates of v vertices:

$$\mathbf{s} = (x_1, y_1, \dots, x_i, y_i, \dots, x_v, y_v)^T, \quad (2.1)$$

where (x_i, y_i) is the coordinate of the i -th landmark in the object and v is the number of the landmarks in an image. Given a set of labeled shapes, we first align these shapes into unified framework by procrustes analysis method. Then, PCA is applied to extract shape eigenvectors and a statistical shape model is constructed as

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^n \mathbf{s}_i p_i, \quad (2.2)$$

where \mathbf{s}_0 is the mean shape, \mathbf{s}_i is the i -th shape eigenvector corresponding to the i -th eigenvalue computed by PCA, and p_i is the i -th shape parameter. The above equation allows us to generate a new shape by varying shape parameters.

For each training face image, the number of pixels inside shape contour is different, and the corresponding relationship of pixels among different training images is uncertain. In order to obtain the required texture information between landmarks, a piece-wise affine warp based on the Delaunay triangulation of the mean shape is used. We obtain a texture-vector set including texture information of each training images by warping the labeled shapes onto the mean shape. Then, the texture-vector set is normalized to avoid these influences from global changes in pixel intensities and so on. Next, PCA is applied to the normalized data, we can easily derive the statistical model of appearance which is represented as follows:

$$\mathbf{A} = \mathbf{A}_0 + \sum_{i=1}^m \mathbf{A}_i \lambda_i, \quad (2.3)$$

where \mathbf{A}_0 is the mean appearance, the appearance vectors \mathbf{A}_i ($i = 1, 2, \dots, m$) are the m eigenvectors corresponding to the m largest eigenvalues, and the coefficients λ_i ($i = 1, 2, \dots, m$) are the appearance parameters.

AAM based on inverse compositional algorithm is an efficient fitting algorithm, evolved from Lucas-Kanade algorithm. The goal of Lucas-Kanade algorithm is to find the location of a constant template image in an input image, minimizing the error between template image and input image as follows:

$$\sum_x [I(W(x; p)) - A_0(x)]^2, \quad (2.4)$$

where $A_0(x)$ is template image and $I(W(x; p))$ is input image. In Lucas-Kanade algorithm, with respect to Δp and then updating the warp using

$$p \leftarrow p + \Delta p, \quad (2.5)$$

the inverse compositional algorithm is different from Lucas-Kanade algorithm, updating the warp using (2.6) and minimizing (2.7):

$$W(x; p) \leftarrow W(x; p) \circ W(x; \Delta p)^{-1}, \quad (2.6)$$

$$\sum_x [I(W(x; p)) - A_0(W(x; \Delta p))]^2, \quad (2.7)$$

where \circ is compositional operation.

Taking the Taylor series expansion of (2.7) gives

$$\sum_x \left[I(W(x;p)) - A_0(W(x;0)) - \nabla A_0 \frac{\partial W}{\partial p} \Delta p \right]^2, \quad (2.8)$$

where $A_0(W(x;0))$ is texture information of mean shape. Minimizing (2.8) is a least squared problem, the solution after taking the derivative will obtain the following:

$$\Delta p = H^{-1} \sum_x \left[\nabla A_0 \frac{\partial W}{\partial p} \right]^T [I(W(x;p)) - A_0(x)], \quad (2.9)$$

where H is Hessian matrix with A_0 and I :

$$H = \sum_x \left[\nabla A_0 \frac{\partial W}{\partial p} \right]^T \left[\nabla A_0 \frac{\partial W}{\partial p} \right]. \quad (2.10)$$

In AAM-based inverse compositional algorithm, since the template A_0 is constant and the Jacobian matrix, the Hessian matrix are nothing to do with location x variables, the relevant computation can be moved to a precomputation step and computed only once. So it is an efficient fitting algorithm than traditional AAM.

2.2. Translation Invariant Wavelet Transform (TIWT)

Wavelet transform is a powerful tool in the analysis of functions and applications in image processing. Via this method, characteristics of a function are separated into wavelet coefficients at different scales. Kourosh and Hamid [12] proposed to utilize different scales energy to depict texture of image via wavelet transform. This method would not be complete to characterize texture features of image because of its rotation invariance. Xiong et al. [13] constructed a translation and scale-invariant wavelet transform. One disadvantage of the classical wavelet translation is that it is not translation invariant, thus, it will lead to filtering result exists a visual distortion. To overcome these defects, Berkner [14] proposed a new approach based on the translation invariant wavelet transform (TIWT). The result is a transform that is not orthogonal anymore and contains redundant information.

The TIWT function representation of an image texture is obtained by convolving TIWT functions with an image. After convolving with TIWT functions, there are many components (images) for an image. Each component is the magnitude part of the output, which is obtained by convolving the image with a TIWT function. As an example, a face image is decomposed into 3 scales via TIWT with harr base. There are 10 components (images) for the image and the example of the convolving result is shown in Figure 1.

From Figure 1, we can see that each component has the same dimensions with the original image. The last component represents low-frequency coefficients and other all components represent high frequency coefficients. The low-frequency coefficients contain most of the energy in the image, which is also the most important visual component. The results show that the low-frequency coefficients represent the approximate weight of the image and contain contour information of the image, while the high-frequency coefficients represent

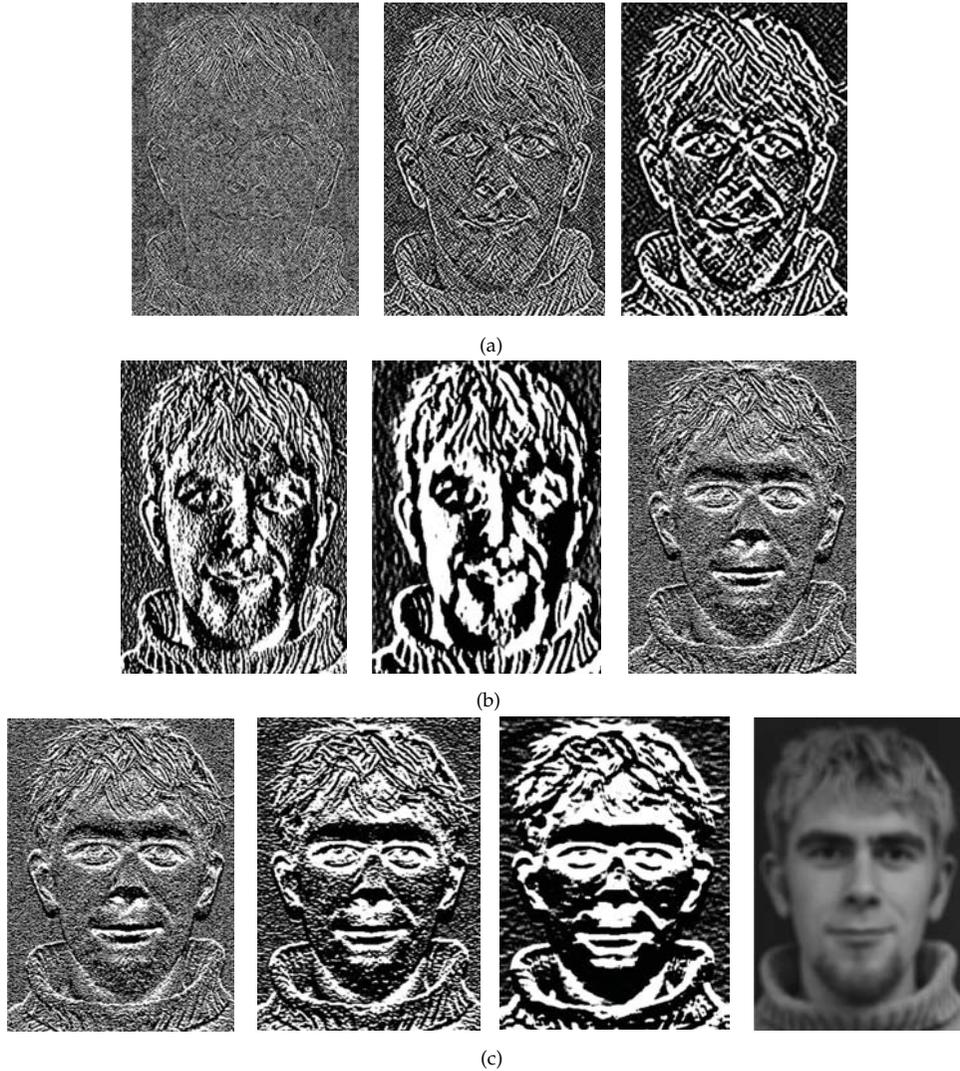


Figure 1: TIWT-based texture by convolving an image with TIWT functions, which contain 3 different scales. The results include row (a): first scale; row (b): second scale; row (c): third scale.

the details of the image. The low-frequency coefficients can gain much clear image representation than the high-frequency coefficients.

3. Improved AAM Algorithm

Translation invariant wavelet transform-based texture can provide enough information for model building. Rather than represent the image structure using intensity values, we use a measure which combines texture information and edge intensity. This improved measure is less sensitive to imaging conditions. In this section, the improved AAM method will be illustrated in detail.



Figure 2: Texture representation. (left) The low-frequency coefficients; (right) The low-frequency coefficients fused by edge intensity.

Given a face image in training set, we can get the low-frequency coefficients of each image via translation invariant wavelet transform. Then, the low-frequency coefficients are fused with edge intensity. We use the method based on edge intensity which was proposed in [15] to fuse the low-frequency coefficients. Assume $C(i, j)$ to represent the low-frequency coefficients obtained via translation invariant wavelet transform, the edge intensity $E(i, j)$ can be computed as follows:

$$E(i, j) = (C(i, j) * M1)^2(i, j) + (C(i, j) * M2)^2(i, j) + (C(i, j) * M3)^2(i, j),$$

$$M1 = \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}, \quad M2 = \begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix}, \quad M3 = \begin{bmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{bmatrix}, \quad (3.1)$$

where $*$ represents convolution, $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$, M and N are the image dimensions. From (3.1), we can see that the edge intensity reflects the edge information of image on horizontal, vertical, and diagonal direction. Edge intensity-based texture representations tend to be less sensitive to imaging conditions. Figure 2 shows texture representation which contains the low-frequency coefficients and the low-frequency coefficients fused by edge intensity.

We represent the image structure using the low-frequency coefficients fused by edge intensity. To build a statistical shape model is similar to the original AAM. Then, to build an appearance model, the texture information is sampled using the low-frequency coefficients fused by edge intensity. All other calculations remain the same as the original AAM.

When fitting model, the same approach as building an appearance model is used to sample the texture of the target image. Then, to find the best model parameters of the most similar AAM instance with the target image, we can extract precise facial features when model converged to the target image.

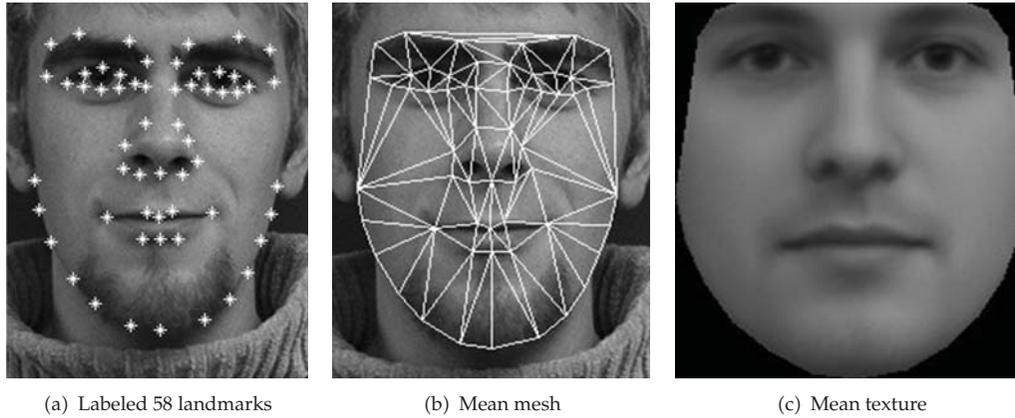


Figure 3: Models of shape and texture on IMM database.

4. Experimental Results

4.1. Database

We implement the following experiments on two databases: IMM face database [16] and Weizmann face database [17]. IMM face database consists of 240 annotated images of 40 different human faces, and each image is labeled 58 landmarks manually. Weizmann face consists of 1260 images, which has 28 subjects under 5 different poses, 3 illuminations, and 3 expressions. Each image is labeled 56 landmarks manually.

4.2. Fitting Accuracy Comparison

In order to verify the fitting performance of the improved AAM, experimental results will be obtained on the above-mentioned databases. Translation invariant wavelet transform based on texture representations are used to build the appearance model. Figure 3 shows models of shape and texture on IMM database. Figure 4 shows fitting comparison between AAM and improved AAM on IMM database. From Figure 4, we can see that the improved AAM obtains more accurate fitting result in the facial contour than original AAM. This is because that the improved method adds edge intensity and provides more information of training images.

In order to examine the fitting accuracy of different AAM methods, we compute the mean displacement errors E between hand-labeled landmarks and fitting model points, the mean displacement can be computed as follows:

$$E = \frac{1}{M} \sum_{i=1}^M \left(\frac{1}{N} \sum_{j=1}^N (\text{dist}(p_{ij} - p'_{ij})) \right), \quad (4.1)$$

where M is the number of test face images, N is the number of labeled landmarks, p_{ij} is the labeled coordinate, and p'_{ij} is the coordinate after model fitting. $\text{dist}(p_{ij} - p'_{ij})$ represents the Euclidean distance between two points. Figure 5 shows mean displacements of varying number of iterations for different AAM models on IMM database. Figure 6 shows mean

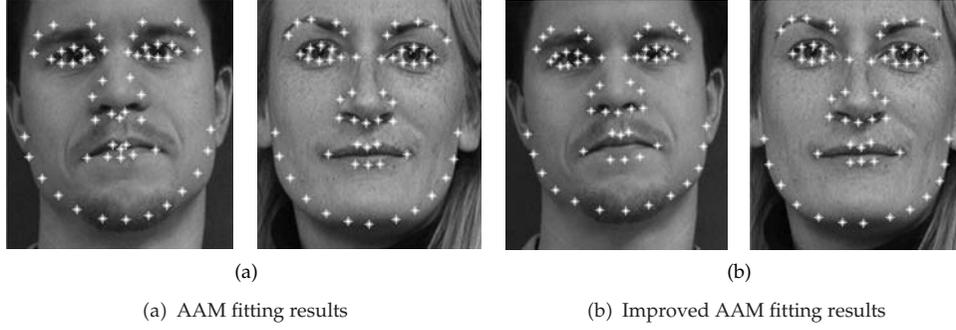


Figure 4: Fitting comparison between AAM and improved AAM on IMM database.

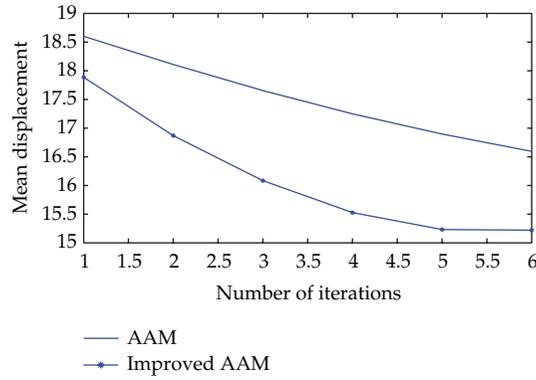


Figure 5: Mean displacements versus number of iterations for different AAM models on IMM database.

displacements of varying number of iterations for different AAM models on Weizmann database. From Figure 5, we can see that improved AAM has much lower mean displacement than that of original AAM. Figure 6 tells us the same story. The precision of feature extraction using improved AAM has been improved, and the fitting speed has been improved.

We can also calculate the overall improvement percentage of our improved AAM to the original AAM by

$$I = \frac{E_{\text{original-AAM}} - E_{\text{improved-AAM}}}{E_{\text{original-AAM}}} \times 100\%. \quad (4.2)$$

Table 1 shows fitting comparison between the original AAM and improved AAM on IMM database. Table 2 shows fitting comparison between the original AAM and improved AAM on Weizmann database. From the above two tables, fitting results show that improved AAM has a lower mean displacement than the original AAM, with the overall improvement about 6.73% and 8.28% in two different databases, respectively, compared with original AAM. Because a new texture representation is used in improved AAM, the improved method can provide more texture information for model fitting than intensity values, especially the new texture presentation is fused with edge intensity, which is less sensitive to imaging conditions. So this improved method can improve the fitting accuracy and the fitting speed has been improved.

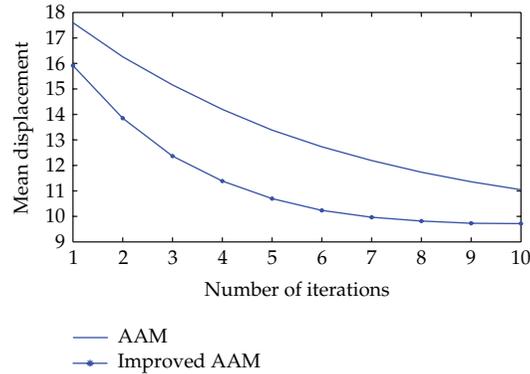


Figure 6: Mean displacements versus number of iterations for different AAM models on Weizmann database.

Table 1: Fitting results comparison of two different AAM models on IMM database.

Algorithm	Mean displacement	Improvement percentage
AAM	10.4169	
Improved AAM	9.7154	6.73%

Table 2: Fitting results comparison of two different AAM models on Weizmann database.

Algorithm	Mean displacement	Improvement percentage
AAM	16.5952	
Improved AAM	15.2211	8.28%

5. Conclusions

As an important method for statistical modeling, active appearance model in the area of human facial features extraction has a wide range of applications. However, intensity values used in original AAM cannot provide enough information for model fitting. To overcome these defects, a new texture representation is proposed in this paper. We propose to utilize translation invariant wavelet transform to obtain the low-frequency coefficients, then, the low-frequency coefficients are fused with edge intensity, the edge intensity will be expressed as texture information for each pixel. The new texture representation is less sensitive to imaging conditions. Experimental results show that the improved algorithm improves the accuracy and exhibits a better fitting performance.

Acknowledgments

This work was supported in part by the following projects: Key Grant Project of Chinese Ministry of Education. (Grant no. 311024), National Natural Science Foundation of China (Grant nos. 60973094, 61103128), and Fundamental Research Funds for the Central Universities (Grant no. JUSRP31103).

References

- [1] W.-C. Kao, M.-C. Hsu, and Y.-Y. Yang, "Local contrast enhancement and adaptive feature extraction for illumination-invariant face recognition," *Pattern Recognition*, vol. 43, no. 5, pp. 1736–1747, 2010.
- [2] Y. Tong, Y. Wang, Z. Zhu, and Q. Ji, "Robust facial feature tracking under varying face pose and facial expression," *Pattern Recognition*, vol. 40, no. 11, pp. 3195–3208, 2007.
- [3] L. Wiskott, J. M. Fellous, N. Krüger, and C. D. von Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.
- [4] Beinglass and Haim J. Wolfson, "Articulated object recognition, or: how to generalize the generalized Hough transform," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '91)*, pp. 461–466, Maui, Hawaii, USA, 1991.
- [5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [6] J. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.
- [7] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: a unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.
- [8] C. B. H. Wolstenholme and C. J. Taylor, "Wavelet compression of active appearance models," in *Proceedings of the 2nd International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 1679, pp. 544–554, 1999.
- [9] X. Gao, Y. Su, X. Li, and D. Tao, "Gabor texture in active appearance models," *Neurocomputing*, 2009.
- [10] T. F. Cootes and C. J. Taylor, "On representing edge structure for model matching," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1114–1119, 2001.
- [11] P. Kittipanya ngam and T. F. Cootes, "The effect of texture representations on AAM performance," in *Proceedings of the 18th International Conference on Pattern Recognition, (ICPR '06)*, vol. 2, pp. 328–331, 2006.
- [12] J. K. Kourosch and S. Z. Hamid, "Rotation-invariant multiresolution texture analysis using radon and wavelet transforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 590–593, 2003.
- [13] H. Xiong, T. Zhang, and Y. S. Moon, "A translation- and scale-invariant adaptive wavelet transform," *IEEE Transactions on Image Processing*, vol. 9, no. 12, pp. 2100–2108, 2000.
- [14] K. Berkner, "smoothness estimates for soft-threshold denoising via translation-invariant wavelet transforms," *Applied and Computational Harmonic Analysis*, vol. 12, no. 1, pp. 1–24, 2002.
- [15] R. Chao, K. Zhang, and Y. J. Li, "Image fusion algorithm using wavelet transform," *Acta Electronica Sinica*, vol. 32, no. 5, pp. 750–753, 2004.
- [16] M. B. Stegmann, "IMM database," <http://www2.imm.dtu.dk/~aam/datasets/datasets.html>.
- [17] The Weizmann Face Database, <http://www.wisdom.weizmann.ac.il/~vision/>.