Illumination Invariant Face Alignment Using Multi-band Active Appearance Model

Fatih Kahraman[1] and Muhittin Gökmen[2]

[1] Istanbul Technical University, Institute of Informatics, Computer Science,
80626 Istanbul, Turkey
kahraman@be.itu.edu.tr
[2] Istanbul Technical University, Computer Engineering Department,
80626 Istanbul, Turkey
gokmen@ce.itu.edu.tr

Abstract. In this study, we present a new multi-band image representation for improving AAM segmentation accuracy for illumination invariant face alignment. AAM is known to be very sensitive to the illumination variations. We have shown that edges, originating from object boundaries are far less susceptible to illumination changes. Here, we propose a contour selector which mostly collects contours originating from boundaries of the face components (eyes, nose, chin, etc.) and eliminates the others arising from texture. Rather than representing the image using grey values, we use Hill, Hue and Grey value (HHG) for image representation. We demonstrate that HHG representation gives more accurate and reliable results as compared to image intensity alone under various lighting conditions.

1 Introduction

Face recognition is an active research topic in image processing and computer vision. Recent years have seen large efforts in searching for a face recognition system which is capable of working with images captured under various lighting conditions including different poses for apparent security reasons. Proposed algorithms could be summarized as sketch-based, feature-based and appearance-based. Most of the proposed algorithms either assume that there is a constant background or face image is already segmented from the background. Also, they require that the frontal view of the face is used under homogeneous illumination condition. Since the transformations (e.g., Fourier, Wavelet, and Karhunen-Loeve) used in these algorithms are linear, all nonlinear degradations on the input face image cause recognition performance to drop drastically [1].

There are several leading studies in transform-based approaches. Probably most well known is eigenface [2]. It has been known that eigenface is very sensitive to shape, pose, mimic and illumination changes. Ezzat and Poggio [3] developed a similar method which is capable of synthesizing face images from training face database. Unfortunately generalization performance of their approach is poor. In the work of Nastar et al. [5] shape and appearance model based on a 3D
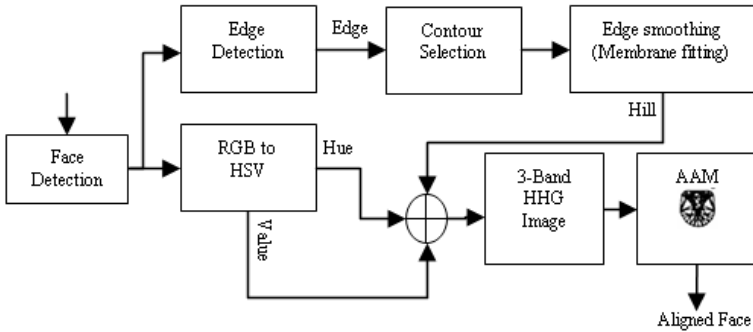
Fig. 1. Overview of the proposed method

appearance of grey level surface was used to synthesize human face. They tried to exactly synthesize face shape and texture by using three dimensional grey level surface model. Since the matching algorithm they offered gets easily trapped at local minimum, the use of method is very limited. Lanitis et al. [6] presented a similar method which models shape and normalized texture separately. Edwards improved this model further by exploiting the relationship between shape and texture. Edwards et al. [7] then offered a fast method to grap face by fitting the Active Shape Model (ASM) through least square minimization. The face is then normalized with respect to the fitted shape model in order to extract grey level texture. Although this approach yields satisfactory results, one cannot always obtain stable model parameters in the presence of variations [4]. This is due to the fact that the model does not completely utilize the face texture inside the face region.

In this study, Haar cascade classifier is used to detect face region from the input image using Haar wavelet features. Once the face region is detected, face components are extracted by applying AAM [15] [20] [4] [16]. Since AAM contains shape information, the pose and direction of the face can be directly estimated from the shape and then the face could be easily aligned to the mean shape. Our work is based on the use of Hill images obtained from selected edge segments together with modified Hue values as in [17], instead of grey values. The overview of the proposed method is depicted in Fig.1 The remainder of this paper is organized as follows: Section 2 describes how AAM is initialized by using Haar Cascade Classifier. Section 3 describes how salient face components are extracted. Details of the multi-band (HHG) representation are discussed in Section 4. Experimental results are presented in Section 5. Conclusions and future works are given in Section 6.

2   Face Detection and Model Initialization

The first step in our face alignment system is to detect face region from the still image. This is achieved by using Haar Cascade Classifier (HCC) which is scale invariant and can cope with pose variations to some degree. HCC has been

successfully used in many object recognition applications. The classifier was first proposed by Viola and Jones [10] [11] and then enhanced by Lienhart [12].

The classifier is trained by the images taken from face (positives) and non-face (negatives) regions of the size 20 by 20. After successful training, the classifier produces 1 where there exists a face and 0 otherwise [12]. HCC can only give a rough estimate of the face region.

We have used Active Appearance Model (AAM) to extract face landmark points. Initialization of AAM is very important for the convergence. Any improper initialization results in erroneous model convergence and consequently yields wrong landmarks. To avoid dependence on initialization, HCC result is used in order to initialize the AAM automatically (Fig.2). It helps both better initialization and fast convergence of the model.
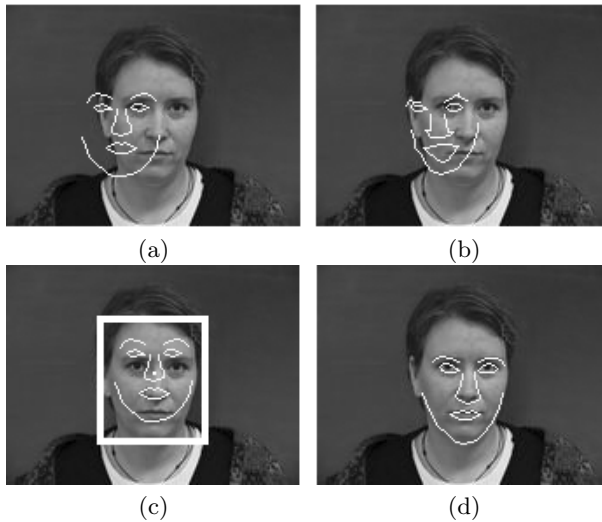

Fig. 2. AAM initialization: (a) Initialization of AAM by using mean shape without HCC, (b) AAM result for (a), (c) Initialization of AAM by using HCC, (d) AAM result of (c)

3   Face Modeling

In this section, we concentrate on face alignment problem. Face alignment is required to obtain high recognition rates in any face recognition system. Recent studies are centered on around this problem. Proposed solutions are generally model-based algorithms. Active Appearance Model (AAM) is the most promising algorithm among the model-based solutions. AAM can compute fast, deformable and robust matching between the model and the original image by utilizing the relationship between the shape and the texture. In this study, we use AAM to extract salient points on human face.

3.1   Active Appearance Model

Any human face can be synthesized from the model which is trained by the training database. The database is prepared by a human observer selecting the landmark points. Once the model is constructed, any given face can be mapped to the model space by minimizing the residual error between the original face image and the synthesized face image. The landmark points on the face images are selected by the human observer. Assume that the points belonging to the *ith* face image are denoted as $\{(S_i, T_i)\}$. Where S_i is a set of points containing shape information such that $S_i = \{(x_1, y_1), (x_2, y_2), \ldots, (x_K, y_K)\}$ and T_i contains texture information at S_i. AAM is obtained by applying principal component analysis on $\{(S_i, T_i)\}$

$$S = \bar{S} + P_s s \tag{1}$$
$$T = \bar{S} + P_t t \tag{2}$$

where $\bar{S}$ is called mean shape and $\bar{T}$ is called mean texture. s and t are the eigenvectors corresponding to the m largest eigenvalues. Any face image can be easily mapped to the model by multiplying the difference between the original image and the mean signal with the covariance

$$S = P_s^T (S - \bar{S}) \tag{3}$$
$$T = P_t^T (T - \bar{T}). \tag{4}$$

Any change on the shape leads to a change on texture since AAM model space is composed of the texture and shape subspaces [9]. Hence the appearance model (A) for any given image can be obtained by the formula

$$A = \begin{pmatrix} \Lambda s \\ t \end{pmatrix} \tag{5}$$

where Λ denotes diagonal shape weight matrix. Principal component analysis can be applied to A in order to reveal the relationship between shape and texture subspaces

$$A = P_a a \tag{6}$$

where a are the eigenvectors corresponding to the m largest eigenvalues.

4   Multi-band HHG Image Representation

In this section, we explain the basic components of our Hill, Hue and Grey value (HHG) based multi-band image representation.

4.1   Face Contour Detection

Classical AAM trained with raw RGB values fails at modeling the faces captured under different illuminations even if the face exists in the training database. This

is due to the fact that AAM makes use of texture resemblance in minimization. This makes AAM very sensitive to illumination variations. It suggests that a feature less sensitive to illumination changes is required rather than using texture alone. We have showed that contours originating from object boundaries are far less susceptible to illumination variations [19] [20]. Fortunately face components such as eyes, eyebrows, nose and lips correspond to object boundaries. Another observation is that contours arising from texture disappears at large scales. This shown in Fig.3(c). We utilized the Generalized Edge Detector (GED) [13] [14], which combines the most of the existing high performance edge detectors under a unified framework. Most important part of the contour extraction algorithm is to select the perceptually important contours among these contours obtained by tracing these edges. This is achieved by assigning a priority to each contour by simply calculating the weighted sum of the normalized contour length, the average contrast along normal direction and the average curvature. The priority assigned to the contour C_i is given by (7).

$$Priority(C_i) = w_{length} \cdot Length(C_i) + w_{contrast} \cdot Contrast(C_i) + w_{curvature} \cdot Curvature(C_i) \tag{7}$$

By means of our contour ordering we can obtain perceptually important contours mostly resulting from object boundaries by selecting only the leading contours in this order and omitting the others, as shown in Fig.3.
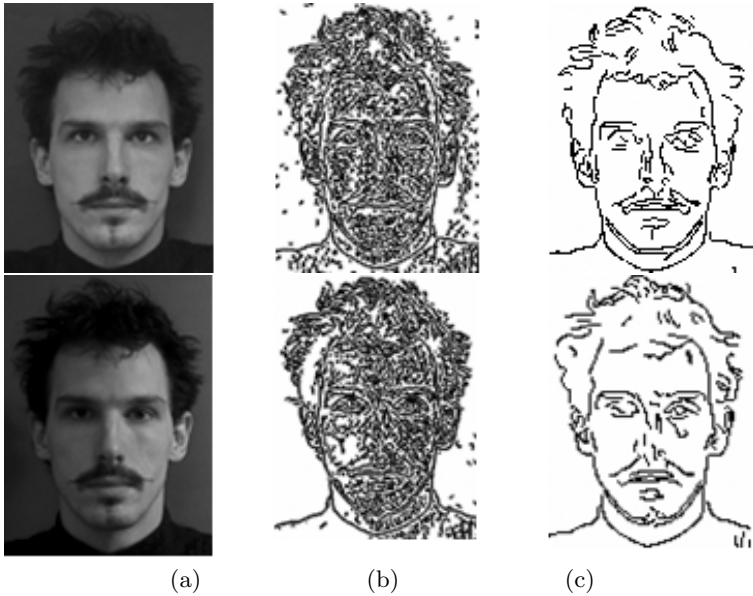

(a)   (b)   (c)

Fig. 3. Contour detection under different illumination: (a) Original image, (b) Detected edges at a small scale, (c) Selected contours

4.2   Membrane Fitting over Edges

By using an appropriate surface reconstruction algorithm, a dense image can be obtained from sparse edge map. To overcome the locality problem of edges, a membrane functional (9), can be applied to edge maps by minimizing.

$$E_m(x,y) = \int \int_\Omega (f - d)^2 dx dy + \lambda \int \int_\Omega (f_x^2 + f_y^2) dx dy \tag{8}$$

$$R_1(x,y;\lambda) = 1/2\lambda e^{-[(|x|+|y|)/\lambda]} \tag{9}$$

The spread edge profiles obtained by this membrane fitting gives rise to a dense image called "Hill" [19]. Hills have high values on boundary locations and decrease as we move apart from edges. It is known that, minimizing the membrane functional is equivalent to convolving the data with a first order regularization filter [14], which is shown in Fig.4(b). The details of the Hill representation are given in [19]. In our approach, first the goal oriented edge detection is applied by sorting edges and selecting meaningful contours describing a face. The detected contours are filtered with R-filter which has a very good localization performance while smoothing the edges. Resulting hill image is shown in Fig.4(c). Filtering the selected edges instead of using the contours alone improves the convergence of AAM under varying illumination conditions [21].
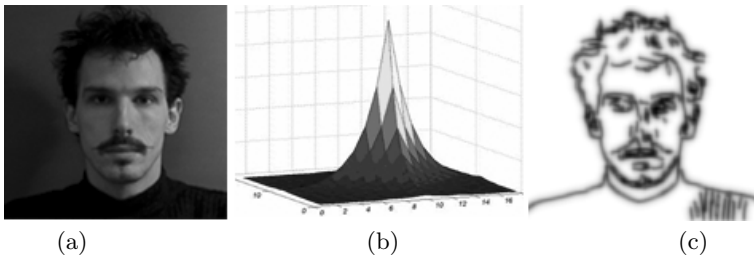

(a)   (b)   (c)

Fig. 4. Membrane fitting: (a) Original face image, (b) Regularization filter kernel, (c) Hill image

4.3   HSV Color Conversion

A disadvantage of the use of raw RGB values in AAM is its sensitivity to illumination changes. On the other hand, the HSV color space has some advantageous to expose the significant features for face recognition. It is well known that Hue band is less sensitive to illumination changes than raw intensity values. One way to increase tolerance toward intensity changes in image is to transform the RGB into a color space where intensity and chromaticity are separated and then use chromaticity part. Changes in illumination can have harsh inferences on the skin color distribution. These changes can be reduced in HSV color space. Thus we used Hue component for our multi-band image representation to get further
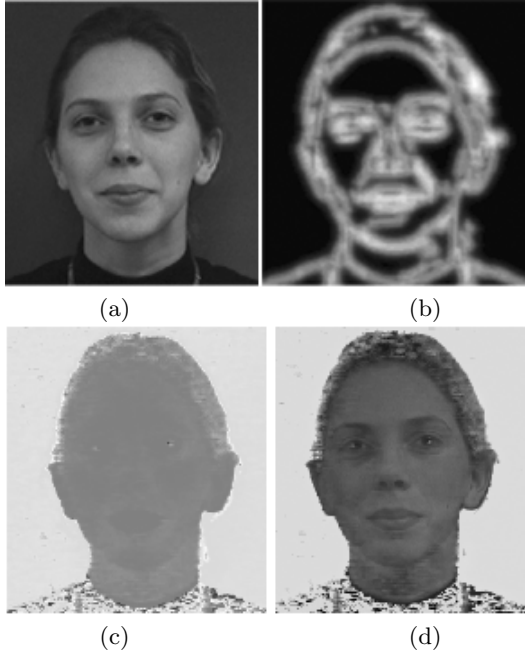
Fig. 5. Multi-band image representation: (a) Value band, (b) Hill band, (c) Hue band. (d) HHV image representation.

robustness to the illumination variation. The image, represented in RGB is converted to HSV. Hue is represented as an angle (0-360 degrees) and saturation as a real number between 0 and 1. We used modified version of Hue band for multi-band image representation as in [17]. Thus our multi-band model consists of the hill band, modified hue band and grey value band.

5   Experimental Results

In our experiments we have used IMM face database [18]. The image resolution is $320x240$. 74 color images which have full frontal face, neutral expression are selected and used in our experiments. 37 images with no spotlight are used for training and the remaining 37 images which have spot light added at the person's left side are used for testing. Face detection module gives a rough estimate of the face region. AAM performs the search in the region HCC suggests. Since the search starts at a point closer to the optimal solution, fast convergence is obtained and we escape from poor local minimum. Classical AAM yields satisfactory results provided that test images whose illumination is similar to the training images.

Since AAM is principal component analysis based approach, illumination changes cause AAM to match the model to wrong face components. Models based on raw intensity tend to be sensitive to changes in lighting conditions.

Thus models built based on one data set may not give good results on different face data set which is taken under different lighting condition.

Edge based representation tends to be less sensitive to lighting changes than raw intensity [20]. We experimentally prove that object boundaries are less sensitive to light conditions when compared to intensity. We obtain high accuracy segmentation results with AAM when multi-band HHG images are used. When classical AAM is used, it can be easily verified from the Fig.6(a) that segmented face components move towards the spot light direction. Fig.6(b) shows the segmentation result obtained with proposed algorithm where hills from selected contours are used. Since contours arising from object boundaries are less sensitive to illumination variations, the proposed AAM can still capture the face model accurately.

The accuracy of the segmentation is computed as the point-to-point (pt.-pt.) and point to curve (pt.-crv) distance measures. Pt.-pt measures Euclidean distance between corresponding landmarks of the model and the ground truth,
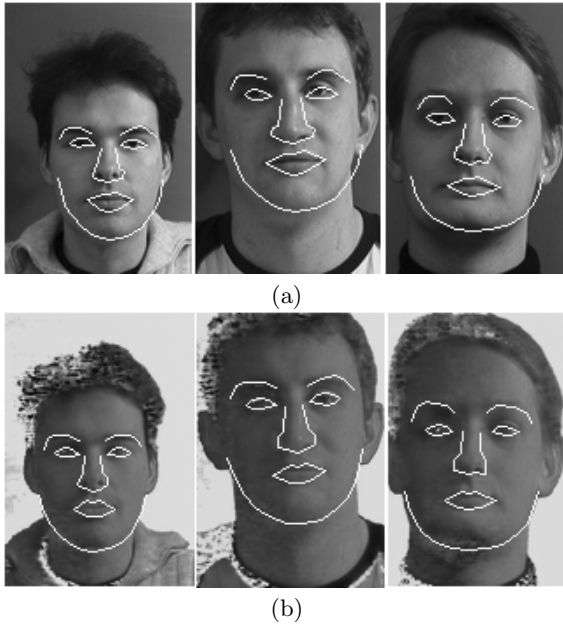

(a)

(b)

Fig. 6. Face segmentation results: Top:(a) Original AAM results, Bottom:(b) HHG based AAM results

Table 1. Face segmentation results for test images (320x240)

Method	Mean pt.-pt. Error	Mean pt.-crv. Error
Original AAM (RGB)	2.84 ± 0.78	1.35 ± 0.49
Stegmann's Method (VHE)	2.63 ± 0.64	1.27 ± 0.40
Proposed Method (HHG)	2.28 ± 0.10	1.00 ± 0.03

whereas pt.-crv. measures the shortest distance to the curve in a neighborhood of the corresponding ground truth landmark [17]. The comparative results are given in Table.1. It's seen that HHG outperforms both original AAM and the VHE multi-band based AAM with grey values, hue and edge strength.

6 Conclusion and Future Work

We proposed a novel multi-band image representation (HHG) for AAM. The HHG utilizes the hill images obtained by fitting a membrane to the selected contours detected by the Generalized Edge Detector (GED). We compared the performance of original AAM, VHE [17] based AAM and HHG based AAM. We have shown that HHG representation based AAM gives the best performance as compared to the original intensity and VHE based models. Our method can be thought as a extension of the Stegmann's paper [17]. HHG based AAM locates the points more accurately than the others for test images which taken under different illumination condition, because Hill images are obtained by smoothing the most prominent edge contours in the face image gathered by means of an efficient and effective edge ordering scheme. When compared to the original AAM and VHE, the proposed method is more accurate in segmentation, more robust to initialization and less sensitive to illumination condition. The performance of the proposed scheme on larger face databases is currently under investigation.

References

1. Xiaoguang L., Hsu R., Jain A. K. "Face Recognition with 3D Model-Based Synthesis," International Conference on Bioinformatics and its Applications, pp. 315–333, Florida, USA, 2004
2. Turk M.A. and Pentland A. P., "Face recognition using eigenfaces. In IEEE Computer Society Conference Computer Vision and Pattern Recognition, pp. 586–591, 1991.
3. Ezzat T. and Poggio T., "Facial Analysis and Synthesis Using Image-Based Models," In International Workshop on AutomaticFace and Gesture Recognition 1996, pp. 116–121, Killington, Vermont, 1996.
4. Ezzat Cootes T. F., Kittipanyangam P., "ComparingVariations on the Active Appearance Model Algorithm", Proc.BMVC2002, **Vol.2**, pp. 837–846.
5. Nastar C., Moghaddam B., and Pentland A., "Generalized Image Matching: Statistical Learning of Physically-Based Deformations," In European Conference on Computer Vision, **Vol.1**, pp. 589–598, Cambridge, UK, 1996.
6. Lanitis A., Taylor C., and Taylor T., "Automatic Interpretation and Coding of Face Images Using Flexible Models," IEEE Transactions on Pattern Analysis and Machine Intelligence, **19(7)**, pp. 743–756, 1997.

 7. Edwards G. J., Taylor C. J., and Cootes T., "Learning to Identify and Track Faces in Image Sequences," British Machine Vision Conference 1997, Colchester, UK, 1997.
 8. Ming Z., Stan Z. L., Chun C., "Subspace analysis and optimization for active appearance model based face alignment," The 6th International Conference on Automatic Face and Gesture Recognition. Seoul, Korea, May, 2004.
 9. Ahlberg J., "A System for Face Localization and Facial Feature Extraction," Report No. LiTH-ISY-R-2172, July 1999.
10. Viola P., and Jones M. J., "Rapid Object Detection using a Boosted Cascade of Simple Features," in Conference on Computer Vision and Pattern Recognition, 2001.
11. Viola P. and Jones M.J., "Robust real-time object detection", ICCV Workshop on Statistical and Computation Theories of Vision, 2001.
12. Lienhart R. and Maydt J. "An Extended Set of Haar-like Features for Rapid Object Detection," Proceedings of the ICIP'02, 2002.
13. Kurt B., Gökmen M., and Jain A.K. "Image Compression Based On Centipede Model," Proceedings of International Conference on Image Analysis and Processing, ICIAP'99, Vol.I, pp.303–310, 1997.
14. Gökmen M., Jain A.K., "$\lambda\tau$-Space Representation of Images and Generalized Edge Detection," IEEE Trans.on Pattern Analysis and Machine Intelligence, vol.19, No. 6, June 1997, pp.545-563.
15. Cootes T., Cooper D., Taylor C., and Graham J., "Active shape models - their training and application," Computer Vision and Image Understanding, 61(1), pp. 38–59, 1995.
16. Stegman M. B., "Analysis and segmentation of face images using point annotations and linear subspace techniques," Technical Report, DTU, 2002.
17. Stegmann M. B., Larsen R., Multi-band Modelling of Appearance, Image and Vision Computing, vol. 21(1), pp. 61-67, Elsevier Science, 2003.
18. Stegmann M. B., Ersboll B. K., and Larsen R.. FAME "A flexible appearance modeling environment," IEEE Trans. on Medical Imaging, 22(10):1319-1331, 2003.
19. Yilmaz A. and Gökmen M, "Eigenhills vs Eigenface and Eigenedge," Pattern Recognition, Vol.34, pp.181–184, 2001.
20. Cootes T. F. and Taylor C. J., "On representing edge structure for model matching," in Proc. IEEE Computer Vision and Pattern Recognition - CVPR, 2001, vol. 1, pp. 1114-1119.
21. Kahraman F., Kurt B. and Gökmen M., "Active Appearance Model Based Face Recognition," in Proc. IEEE 13th Signal Processing And Communications Applications Conference, 2005, vol. 1, pp. 209-212, Kayseri, Turkey.