Real-time Facial Feature Extraction Using Statistical Shape Model and Haar-Wavelet Based Feature Search

Fei Zuo
Eindhoven University of Technology
Electrical Engineering Depart.
5600MB Eindhoven, The Netherlands
F.Zuo@tue.nl

Peter H. N. de With
LogicaCMG / Eindhoven Univ. of Technol.
P.O.Box 7089, 5605JB Eindhoven, NL
peter.de.with@logicacmg.com

Abstract

We propose a fast facial feature extraction technique for an embedded face recognition system. The novel key element is a combination of a statistical shape model and an application of a Haar-wavelet based feature matching. Our statistical face model is based on the Active Shape Model (ASM). However, ASM lacks robustness to illumination changes and it has a limited convergence area. Instead of a 1-D profile analysis, we propose a 2-D texture pattern search-and-fitting scheme, which provides more robustness and faster convergence than conventional ASM. Furthermore, we employ Haar-wavelets to model local facial textures, which yields two improvements: faster processing and more robustness with respect to low-quality images. Our proposed approach shows good results dealing with test face images, which are quite dissimilar with the faces used for statistical training. The convergence area of our proposed method almost quadruples compared to ASM, and the extraction accuracy is also improved. The total processing requires 30 - 70 ms, which is comparable to ASM, but faster than the Active Appearance Model (AAM).

1 Introduction

Accurate facial feature extraction is important for face alignment, which is an indispensable processing step between face detection and recognition. Our aim is to build a feature-extraction system that can be used for face recognition in embedded and/or consumer applications. This imposes specific requirements to the algorithm in addition to extraction accuracy, such as real-time performance under varying imaging conditions and robustness with low-cost imaging hardware.

In an earlier feature-extraction research, Yuille et al. [8] use parameterized deformable templates to extract eyes and mouth. However, it is computationally expensive and the convergence is not guaranteed. We have reported at an earlier stage [9] a facial feature localization technique, initializing from multiple starting points. Unfortunately, it can only provide partial feature descriptions (e.g. iris center). Recently, the Active Shape Model (ASM) [1] and Active Appearance Model (AAM) [2] have gained significant attention as two effective techniques for feature extraction. The ASM exploits statistical shape variations from a set of training faces and deforms a model shape to fit with a real image by adjusting the position of each feature point so that the model features match with the local grayscale profile from the real image. This simple fast approach has one major drawback: it only exploits 1-D information along the feature profile, which leads to convergence to local minima and a relatively small convergence area. Furthermore, it is sensitive to variations in face images which are captured in a different environment from the training images. The AAM incorporates global facial texture modelling and explicitly minimizes the texture matching error during the model fitting. Although it gives a better match with the image texture, it is slower than ASM and more sensitive to texture variations under different illumination conditions.

This paper attempts to solve the above problems. The key to our solution consists of two aspects. Our first contribution is the exploitation of 2-D texture information around each feature during the local fitting. This significantly increases the convergence area of our shape model and it improves the robustness of the algorithm. Our second contribution is the use of Haar-wavelets for modelling local feature patterns. The feature search is carried out in the Haar-wavelet domain, which offers high processing speed and robustness with low-quality images.

The remainder of the paper is organized as follows. Section 2 briefly outlines the traditional ASM algorithm. Section 3 discusses in detail our proposed fast 2-D Haar-wavelet based feature search. Section 4 gives the experimental results, and Section 5 presents the conclusions.
2 Active Shape Model (ASM)

The ASM exploits the statistical properties of face shapes, where each face shape is represented by a vector containing the positions of a defined set of feature points. An example of a manually annotated face shape is depicted in Fig. 1. A statistical shape model is built from the face shapes, i.e. the vectors of a set of labelled face images. These shape vectors are aligned to a common coordinate system and a PCA (Principal Component Analysis) is subsequently performed to obtain the major shape variation directions. The fitting of an initialized model and a new face image can be carried out by an iterative process, where each feature point is adjusted by searching for a best-fit neighboring point. The shape variations are constrained during the deformation process to maintain a plausible shape.

The ASM uses the grayscale profile which is perpendicular to the model boundary as the basic feature to perform the local feature match. The local profiles extracted from training images are modelled as forming a multivariate Gaussian distribution. In practice, the normalized first derivatives of these profiles are used to compensate for global illumination changes. The fitting function is formulated by calculating the Mahalanobis distance between the current profile and the average profile corresponding with the same feature point.

The profile-based local feature search is relatively fast, but it only exploits 1-D profile information in a limited search area, so that it can easily converge to wrong positions. This occurs more often when the new face image differs significantly from the training images in illumination and imaging quality. Furthermore, the capture range (or correct convergence area) of ASM is limited, and inappropriate initializations of the model will quite likely lead to convergence failures. The original paper proposes a multi-resolution search as a solution for this, but the construction of multiple image pyramids would significantly increase the computational complexity. In the following section, we propose a novel fast model fitting procedure using two-dimensional Haar-wavelet features. It offers a wider capture range and higher accuracy than ASM, while still keeping the computation cost at a low level.

3 Model fitting using 2-D Haar features

We extract an $N \times N$ square block around each feature point from the image, which contains richer texture information than a 1-D profile and provides more reliable local pattern information. In our proposal, these local facial features are modelled using Haar-wavelets, a variant of which has been successfully applied in face detection [7, 4]. A closer examination of the local feature blocks in face images shows that they usually contain relatively simple patterns having strong contrast. The 2-D basis images of Haar-wavelets match very well with these patterns, so that it is attractive to exploit them for efficient signal representation. Furthermore, the simplicity of Haar wavelets supports the implementation of real-time facial feature extraction.

3.1 Illumination correction

The local appearance of the feature is usually uniformly affected by illumination. For each feature block $B$ with pixels $P(x, y)$, we reduce the illumination interference by correcting its mean $\mu_B$ and variance $\sigma_B$, hence

\[
P_N(x, y) = (P(x, y) - \mu_B)/\sigma_B. \tag{1}\]

The correction has been proven to be quite effective in our experiments. In the sequel, we will efficiently combine the above normalization with the implementation of the Haar-wavelet transform.

3.2 Fast computation of Haar features

Haar-wavelet decomposition mainly involves summations of pixel sub-blocks (see the Haar basis images shown in Fig. 2), which can be efficiently computed by using two auxiliary ‘integral images’ [7, 4].

The integral image $I$ of image $P$ is defined as:

\[
I_P(u, v) = \sum_{x=1}^{u} \sum_{y=1}^{v} P(x, y). \tag{2}\]
For a block $B$ with its top-left corner $(x_1, y_1)$ and bottom-right corner $(x_2, y_2)$, the summation of the pixel intensity in this block can be computed as:

$$S(B) = I(x_1, y_1) + I(x_2, y_2) - I(x_2, y_1) - I(x_1, y_2).$$  \hfill (3)

A fast algorithm [7] can be applied to obtain the integral image of a given image in only one pass over the image. Similarly, a square integral image $I_q$ can be obtained by replacing $P(x, y)$ in Equation (2) by $P^2(x, y)$, which facilitates fast computation of the block variance $\sigma_B$.

For a given pixel feature block $B$, the corresponding Haar-wavelet coefficient $H(u, v)$ can be computed by

$$H(u, v) = \sum_{i=1}^{N_B} (Sgn(B_i) \sum_{x=1}^{M_B} \sum_{y=1}^{M_B} [B_i(x, y) - \mu_B])$$

$$= \frac{1}{N(u,v) \sigma_B} \sum_{i=1}^{N_B} [Sgn(B_i) \cdot S(B_i)].$$ \hfill (4)

Note that Equation (4) already incorporates the illumination correction from Equation (1). In the above, $B_i$ refers to sub-blocks corresponding to non-zero coefficient areas in the basis images (see Fig. 2). The number of these sub-blocks inside block $B$ is denoted as $N_B$. $Sgn(B_i)$ refers to the sign of the coefficient part corresponding to sub-block $B_i$, while the size of sub-block $B_i$ is $M_B \times M_B$. For $2^n \times 2^n$ blocks, $N_B$ can be 1, 2 or 4. Since coefficient $H(0, 0)$ only contains the average intensity value of the block, it is zero for all illumination-corrected block images and can be ignored during the matching. For all remaining basis images, the total area of $+1$ signed sub-blocks are equal to the area of $-1$ signed sub-blocks. $N(u, v)$ is the normalization factor for coefficient $H(u, v)$. The summation of sub-block data can be efficiently computed using Equation (3), so that the term $S(B_i)$ occurs in the second expression of Equation (4). Since $N_B \leq 4$, the computation of one Haar coefficient involves at most 9 table lookups of the integral images (see Equation (2)).

![Figure 3. Haar representation power.](image)

### 3.3 Haar-wavelet based local feature search

In our proposed approach, we search for the local optimal feature point by matching the current feature block $B$ with a corresponding average block $\overline{B}$ obtained from the training face images. The comparison is carried out in the Haar-wavelet domain, using the following fitting function:

$$F(B, \overline{B}) = \sum_{i=1}^{K} [H_i(B) - H_i(\overline{B})]^2,$$ \hfill (5)

where $H_i(B)$ gives the $i$-th (in zigzag order) Haar coefficient of block $B$. It can be seen from Fig. 3 that Haar-wavelet coefficients represent a highly compact description of the local facial features. For most feature points, less than 4% of the total coefficients are required to retain up to 95% of the feature energy. Thus, in Equation (5), the number of selected coefficients $K \ll N$ and $K$ can be variable to adapt to different feature structures. This significantly increases the search speed of the algorithm and reduces the interference of the image noise.

In ASM, the local optimal position is searched along the direction perpendicular to the model contour. We found this insufficient because the best nearby point is not always located on the search line. In our implementation, we adopt a search trace composed of eight radial lines originating from the current model point. This `star-like` search trace provides a more reliable convergence than the single 1-D profile search.

To further accelerate the fitting process, a multi-stage coarse-to-fine search can be used. The first-stage search uses large steps to quickly find an approximated convergence position, while the following stages take finer steps to refine the fitting accuracy.

### 4 Experimental results

To evaluate the algorithm performance, we implemented our proposed algorithm and ASM for comparison. The statistical face shape model was built from 37 annotated face images, each consisting of 58 feature points [6]. The test set consisted of face images from the BioID face database [3], Alex database [5] and our own HN2R database. These databases contained various faces taken under different conditions from the training images, which can verify the generic applicability of the algorithm. The size of the faces were scaled to roughly the same size ($300 \times 300$ pixels). For all the tests, we performed the single resolution search without using the image pyramids. Our proposed method used $32 \times 32$ pixel feature blocks and the profile length was also set to 32 pixels in ASM. The search ranges were equal for both algorithms for justified comparisons.

### 4.1 Convergence test

For each test face, we position the gravity center of the initial shape to $(x_r + \Delta x, y_r + \Delta y)$, where $(x_r, y_r)$ is the
reference gravity center. If the gravity center displacement between the fitted shape and the reference shape is $D_c$, the algorithm is considered converged for all situations with $D_c < 8$. Fig. 4 portrays the convergence test results. It can be seen that the convergence area of our proposal is almost four times as large as the area covered by ASM.

4.2 Accuracy test

In order to measure the fitting accuracy, we compute the average point-to-point distance between the fitted shape and the manually labelled shape. To reduce the error that may be introduced by manual labelling, we use a normalized point-to-point distance metric, in which the feature points along a certain contour are equally spaced. The evaluation results are shown in Fig. 5. The fitting process is originated by initializing the model with displacement $\Delta g$ ($\Delta g \in [-12, +12]$) to the reference position. We can see that our proposed approach achieves more accurate results than ASM. Up to 90% of the test cases achieve an accuracy of less than 9 pixels, while in ASM, only 60% of the test cases achieve the same accuracy.

![Figure 5. Accuracy test.](image)

Our proposed method takes 30 - 70 ms to process one test face image (Pentium-IV desktop, 1.7 GHz), which is comparable to ASM. Fig. 6 shows two examples. It can be seen that our proposed feature extraction is able to achieve correct convergence, even when the model is poorly initialized. The fitting procedure usually requires less than five iterations.

![Figure 6. Feature extraction examples.](image)

5 Conclusions

In this paper, we implemented a significantly improved facial feature extraction by employing two aspects. First, we used 2-D texture blocks as basis patterns for local features, which clearly outperforms ASM with larger convergence areas. Second, we employed the Haar-wavelets for efficiently representing the local features, thereby facilitating fast processing and increasing robustness for noisy, low-quality images. Our proposed feature extraction technique converges accurately, especially with face images which are quite different from the training faces. The processing takes only 30 - 70 ms on a Pentium-IV PC (comparable to ASM). The proposed feature-extraction technique is part of a prototype real-time face recognition system for customized consumer applications.

References