Automatic human face detection for home surveillance application

Automatic Human Face Detection for Home Surveillance Applications*

Fei Zuo[1] and Peter H.N. de With[2]

[1] Eindhoven University of Technology, EESI, 5600MB Eindhoven, the Netherlands
F.Zuo@tue.nl
[2] CMG Eindhoven / Eindhoven University of Technology, the Netherlands
peter.de.with@cmg.nl

Abstract. This paper concentrates on exploiting human face information for surveillance applications in a home environment. The system features real-time human face detection and facial feature identification. It is our aim to insert the results in a video-security system architecture, where MPEG-4 coding techniques enable low bit-rate video transmission over a home network environment. The processing contains the following essential elements: (1) skin-color segmentation, (2) histogram analysis for facial feature detection, and (3) probability-based confidence value evaluation of facial features. We have tested our sequence of processing algorithms on a set of video sequences. The experimental results show that our approach offers near real-time processing speed with good detection capability, but the robustness needs further improvement.

1 Introduction

Video surveillance systems have become popular in many professional applications. The small inexpensive cameras for multimedia computing enable the introduction of video surveillance applications in a consumer home environment. In the home environment, video surveillance is attractive for new applications when it is combined with face recognition. Such a system can automatically perform tasks such as identifying persons approaching the back door for security purposes. The required processing task of such a application usually involves a wide variety of video processing algorithms ranging from motion segmentation to pattern recognition.

A typical home video surveillance system may contain the following video analyzing tasks.

1. Moving person segmentation for identifying approaching persons. Motion information is analyzed here to separate moving targets from their background.

* This work is supported by the HomeNet2Run project, which is running under the ITEA research programme within the EUREKA framework.

2. Face detection. When the person gets near enough, his/her face can be detected and extracted. The corresponding facial features are extracted and/or coded at the same time for later processing.
3. Face recognition. Face recognition distinguishes different persons and decides about the person's identification.

Among these tasks, face detection is an indispensable and important step because it usually sets the input for the face recognition modules. The performance of most face recognition algorithms relies on an accurate locating of human face regions. However, face detection has remained to be a difficult problem because human face varies strongly in its appearance. The challenges associated with face detection can be attributed to the following factors: pose, presence or absence of structural components, facial expression, occlusion, image orientation and imaging conditions [3].

A wide variety of methods for face detection have been proposed in the past decade. Extended surveys of these methods can be found in [3][4]. These methods range from simple edge-based algorithms to complex high-level approaches utilizing advanced pattern recognition techniques. Among these approaches, skin color has been proved to be an effective facial feature for locating facial regions. In addition, skin-color-based methods usually achieve much higher speed than other computation-expensive approaches and are especially suitable for real-time applications such as video surveillance. Several studies [1][2][3] have revealed that human skin color forms a highly condensed cluster in certain color spaces, and its distribution can be characterized as a multivariate Gaussian distribution. However, skin color detection methods are prone to detection errors, especially in cases where human faces are immersed in a skin-color-like background. Some recent skin-color based methods use geometric analysis or motion information for face verification. However, these methods generally cannot achieve robustness due to the following reasons:

1. Most verification processes only simply make a true-or-false decision for a candidate skin-color blob, therefore, they are prone to make erroneous decisions when faces and their background are contained in the same skin-color region.
2. In geometry-based verification, a set of rules defining a face are usually formed according to heuristics or anthropometric measurements. Due to the variation of human faces (e.g. race, sexe, and age), these rules should be well defined. The definition of a 'perfect' threshold for classification of faces and non-faces has been proved not feasible.

In this paper, we propose a skin-color based face detection which invokes a further extraction process. The extraction process uses *confidence values* for some facial features. The confidence value of each facial feature is based on a probabilistic analysis of facial geometry. The proposed approach selects the most probable facial feature combination to form a complete facial region. Our approach has the following advantages over aforementioned skin-color-based methods:

1. It can extract the face region from a large skin-colored area by verifying the existence of facial features such as mouth and eyes.
2. It avoids a strict true-or-false decision by incorporating probabilistic metrics. A set of facial features is considered to have a certain probability to represent a human face. If one set of facial features achieve a sufficiently high confidence value, the corresponding region is regarded as a real face region.

The paper is structured as follows. In section 2, we give a brief introduction of the proposed method and its processing flow. Section 3 illustrates in more detail the key algorithms, i.e. the skin-color segmentation and probability-based facial region evaluation. Section 4 presents the evaluation results of experiments using our proposal. Section 5 presents conclusions.

2 Sequence of Processing Algorithms for Face Detection

In this section, we present an overview of all processing tasks that are performed for our face detection proposal. The purpose of this section is to briefly discuss the individual steps in the processing. In the next section, we will discuss details of the most important steps. Fig. 1 visualizes the sequence of processing steps.

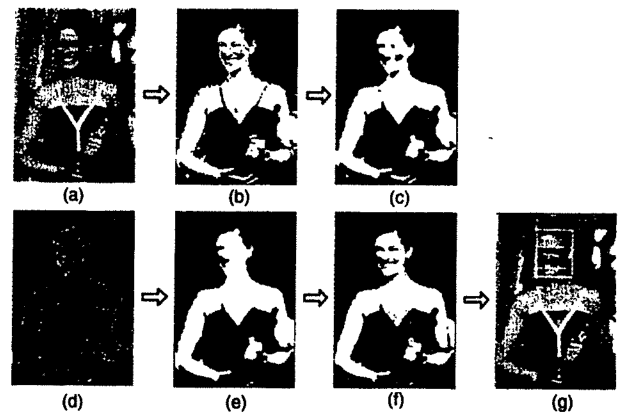


Fig. 1. The sequence of processing algorithms for the proposed face detection method (a) Original image frame (b) Skin color segmentation (c) Filtering (d) Contouring (e) Region forming (f) Thresholded image for histogram analysis (g) Final result.

Each frame of the original video sequence (Fig. 1(a)) is first segmented into areas based on skin-color detection (Fig. 1(b)) [1]. A binary majority filter is then applied to smooth the segmentation results and it also removes minor objects and it fills holes in the object (Fig. 1(c)). The contours of the skin-colored areas are then traced (Fig. 1(d)) and candidate regions that may contain faces are recognized and labeled (Fig. 1(e)). Each candidate region is converted to a binary image using a threshold based on both luminance and chrominance information (Fig. 1(f)). The binary image emphasizes some of the salient facial features.

After these pre-processing steps, we use a probability-based face verification algorithm to further extract facial regions from the candidate regions. In this algorithm, possible prominent facial feature regions (mouth and eyes) are first located, using histogram analysis. Then so-called individual local confidence values are assigned to each feature. All possible combinations of facial features are then formed to generate a face candidate set. For each combination, the global relationship between constituting feature regions is examined to yield a global confidence value. The candidate combination with the largest comprehensive confidence value is considered the most-likely facial region. After such a facial region is determined, post-processing eliminates unlikely candidates based on shape analysis (Fig. 1(g)). The rectangular bounding box indicates the face region, while the enclosed upper and lower horizontal lines represent the eye region and mouth region, respectively.

3 Key Algorithms for Face Detection

3.1 Skin Color Segmentation

We apply the research result from [1]because it uses two signal components only which supports real-time applications. We analyzed a sample set of facial pixels taken from people of different race, gender, age and under various lighting conditions. The skin color is mostly concentrated in two normalized signal components $r = \frac{R}{R+G+B}$ and $g = \frac{G}{R+G+B}$. The combination (r, g) yields a bivariate Gaussian distribution, which is plotted in Fig. 2.

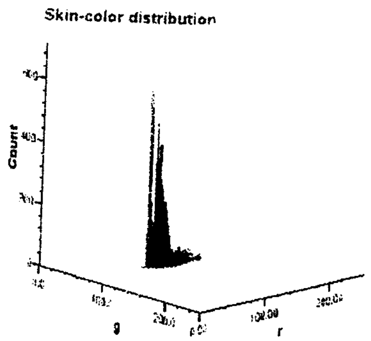


Fig. 2. Bivariate Gaussian distribution of skin color showing a narrow concentration ($\mu_r = 110.4157$, $\mu_g = 80.4125$, $\sigma_r^2 = 154.9733$, $\sigma_g^2 = 20.1298$, $\sigma_{gr} = \sigma_{rg} = -18.5518$).

The mean values of r and g ($\mu_r = \Sigma_{i=1}^{n} r_i$, $\mu_g = \Sigma_{j=1}^{n} g_j$) are computed. Also the covariance term σ_{gr} and the signal variances σ_r^2 and σ_g^2 are evaluated for a rough segmentation of the original image frame into possible facial areas.

3.2 Probability-based Facial Region Verification

Facial features have been widely used in facial processing literature for classifying human faces. However, facial feature appearance and distribution vary considerably among different people and under different conditions. Some of the facial features can not be explicitly distinguished from image deteriorations resulting from local poor video quality (e.g. a dark noisy area). Some of the facial features can be easily mixed up like eyebrows and eyes. For this reason, we focus on first selecting the most prominent facial features - eyes and mouth. We assume that these features are always present in every face image. At the same time, the detection of less prominent features, like nostrils and eyebrows, can contribute positively to a confident identification of prominent facial features.

The algorithm starts with a binary image which may contain a human face. First the horizontal histogram is generated from the input image [6], from which second derivatives of local maxima are located as reference lines. These reference lines correspond to some comparatively 'flat' areas having no significant dark regions, i.e. potential cheek lines. The local minima above the reference line and the local minima below the reference line form two candidate sets: the eye region candidate set $ES_k = \{e_i \mid i = 1, m\}$ and the mouth region candidate set $MS_k = \{m_j \mid j = 1, n\}$(see Fig. 3). The algorithm iterates over the whole facial region to get all reference lines and all possible eye and mouth candidates: $ES = \bigcup ES_k$ and $MS = \bigcup MS_k$.

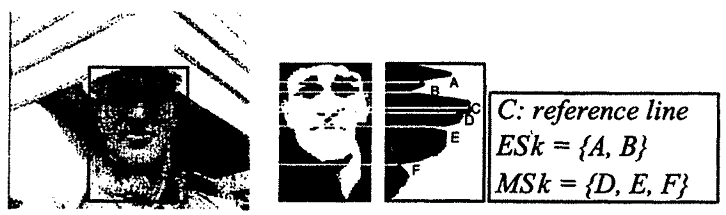


Fig. 3. Face detection and subsequent facial histogram analysis.

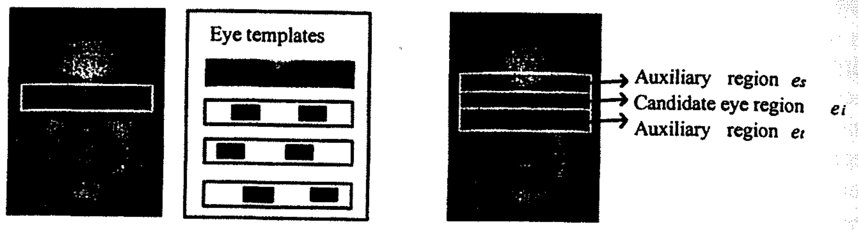
For each $e_i \in ES$, we define the probability of being the real eye region as a local confidence function:

$$LC(e_i, EYE) = \sum_{s=1}^{m} w_s P_s \ . \qquad (1)$$

The term w_s is a weighting factor and P_s represents the confidence value based on different assessment criteria that are explained below. Two criteria are of significant importance for their contribution to the confidence function.

1. A principal criterion is the matching probability of e_i fitting to a pre-defined eye region template (see Fig. 3(a)). A series of templates are used in this stage to accommodate for various eye positions. The probability value indicates how well e_i fits to one of these available templates.

2. Another important criterion considers the possible existence of less prominent features like nostrils and eyebrows, for increasing the reliability of our decision. The detection of less prominent features in the correct positions relative to e_i increases the probability that e_i is a real eye region. In Fig. 3(b), auxiliary regions e_s and e_t are introduced for this purpose, where e_s may contain the eyebrows and e_t may contain the cheek. The presence of low-intensity pixels in e_s and high-intensity pixels in e_t gives e_i a higher confidence value. This is especially useful when a significant eyebrow area appears in the face.



(a) (b)

Fig. 4. Local confidence evaluation

Similar processing is also applied to evaluate the mouth region. After local confidence evaluation is examined, the individual candidate features of ES and MS are extended with their corresponding confidence values, as follows:

$$ESC = \{(e_i, c_{e_i}) \mid e_i \in ES\},\ \ MSC = \{(m_j, c_{m_j}) \mid m_j \in MS\},$$

where c_{e_i} represents the confidence value of e_i and c_{m_j} represents the confidence value of m_j.

The examination of only local features and their relations is not sufficient for a final judgment of a facial region. A global confidence evaluation process is applied to assess the overall probability for a face candidate. We define a face candidate as a combination of facial feature regions. For each face candidate $(e_i, m_j) \in ESC \times MSC$, a global confidence function is evaluated as follows:

$$GC(e_i, m_j) = \sum_{t=1}^{n} w_t Q_t \ . \qquad (2)$$

The term w_t is a weighting factor and Q_t represents the global confidence value according to global assessment criteria. These criteria are based on geometric relations between different features. For example, one of the criteria is based on the statistical analysis of the ratio between face width at the eye level d_e and the distance between eye and mouth d_v. The probability density function of this parameter for a large group of people can be characterized as a normal

distribution. Therefore, the confidence value Q_t in our proposal is based on the probability evaluation of the specific ratio d_e/d_v.

The final confidence evaluation function of a face candidate combines equations (1) and (2) into:

$$Confidence(e_i, m_j) = w_g GC(e_i, m_j) + w_e LC(e_i) + w_m LC(m_j) \ . \qquad (3)$$

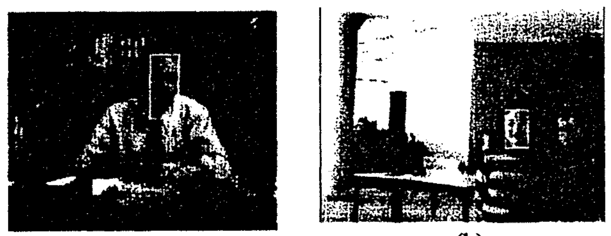
The variables w_g, w_e, w_m are weighting factors. The face candidate with the highest confidence value is selected as the most probable facial region. A post-processing step is then performed to make a final judgment using a shape analysis of the selected region.

4 Experimental Results

To evaluate the efficiency of our proposed technique, we tested our method on several video sequences including both standard testing sequences and self-made sequences. The system achieves an average of 10.7 frames/sec processing speed on a Pentium IV desktop (Intel 1.7G Pentium IV processor, 256M memory, Red Hat Linux 7.2, gcc) with a resolution of 320 pixels by 240 lines, which is sufficient for real-time video surveillance applications.

Example frames from the testing video sequences are shown in Fig. 5. In sequence Salesman (Fig. 5(a)) which contains relatively detailed background, the system identifies the face region correctly in 90% of the frames. We recorded a special experimental indoor sequence (Fig. 5(b)) because it contains skin-colored background objects such as cabinets, tables and parts of the person's clothes. The system successfully identifies the face region in 95% of the frames. The experiments prove that our approach is quite attractive for real-time video surveillance applications.

However, initially the proposed method was prone to make false detections when dealing with more detailed background objects, and it showed low robustness against poorly-illuminated environments. These problems were solved by incorporating an image correction pre-processing step and a refinement of the previously mentioned confidence evaluation criteria.



(a) (b)

Fig. 5. Detection results on different video sequences

5 Conclusions

In this paper, we have proposed a fast human face detection technique for home video surveillance applications. The proposal first uses a skin-color model to roughly segment the possible facial areas. Afterwards, a probability-based evaluation process is applied to further extract facial features from candidate regions. In this step, prominent feature regions such as eyes and mouth are extracted based on confidence value evaluation. The experimental results show that a 90~95% detection accuracy is achieved. As there is no training procedure involved in the processing, the proposed technique is suitable for real-time video processing.

Our proposal can be further refined by incorporating more advanced normalization and criterion selection. Motion information can also be included to improve the overall detection accuracy. The robustness of the system with respect to poor illumination and rotation of the faces still need further study.

References

1. Yang, J., Lu, W., Waibel, A.: Skin-Color Modelling and Adaptation. Technical Report CMU-CS-97-146. School of Computer Science, Carnegie Mellon University, Pittsburgh (1997)
2. Sobottka, K., Pitas, Ioannis.: Extraction of Facial Regions and Features using Color and Shape Information. In Proc. 13th International Conference on Pattern Recognition. Vienna Austria (1996) 421–425
3. Yang, M-H., Kriegman, D., Ahuja, N.: Detecting Faces in Images: A Survey. in IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol 24, no. 1. (2002) 34–58
4. Hjelmaas, E.: Face Detection: A Survey. in Computer Vision and Image Understanding, Vol 83. (2001) 236–274
5. Yang, M-H., Ahuja, N.: Detecting Human Faces in Color Images. in Proc. International Conference on Image Processing, Vol 1. 1. (1998) 127–130
6. Kotropoulos, C., Pitas, I.: Rule-based Face Detection in Frontal Views. in Proc. International Conference on Acoustics, Speech and Signal Processing. Munich Germany (1997) 127–541