AUTOMATIC DETECTION OF HUMAN FACES IN NATURAL SCENE IMAGES BY USE OF SKIN COLOUR AND EDGE DISTRIBUTION

Tsz Ying Lui, Ebroul Izquierdo

Department of Electronic Engineering
Queen Mary, University of London
email:{tsz–ying.lui, ebroul.izquierdo}@elec.qmul.ac.uk

ABSTRACT

We study different skin colour models and analysis the facial texture based on edge distribution to automatically detect and locate human faces in natural scene images. First colour segmentation is performed to obtain face candidates of an input image. Then the resulted binary image is grouped into clusters of connected pixels. Edge distribution is computed from each cluster and a SVM face classifier is trained for face detection. The experimental results show that the combination of skin segmentation and of edge distribution in detecting faces of various poses is efficient and effective.

1. INTRODUCTION

Face detection is currently a very active research area and the technology has come a long way since the survey of Chellappa et al. [6]. The main challenge in face detection is that human faces are highly dynamic and non–rigid objects. Moreover, different face poses, facial expressions and occlusions in images make the problem even more complex, if not unsolvable. To detect face, it means to local human faces in an image regardless of their locations, poses, scales and resolutions and lighting conditions. There are mainly two different approaches to face detection in the literature: feature–based approach that extracts facial components such as eyes, nose and mouth as cues to deduce the existence of a face; and the image–based approach that treats face detection as a recognition problem by training and learning. In the last couple of years, the image–based approach has shown great advances in algorithms and been proved to be very effective in dealing with complex environments such as low quality gray–scale images and cluttered backgrounds. However, some of the best algorithms are still too computationally expensive to be applicable for real–time processing [2].

To implement an efficient automatic face recognition system, the techniques used to detect and locate face in a scene have to be robust and with low system latency. Colour is a very powerful fundamental cue that can be used as the first step since colour segmentation is computational fast and relatively robust to changes in scale, in viewpoint and to shading. However, other low–level features are also necessary in order to discriminate true face regions from other face candidates. In this paper, we use a two–stage strategy for face detection in colour images. First, we use our prior knowledge on human skin colour to reduce the searching space for a face candidate. Then we train a face classifier by using SVM to detect the existence of a face. This paper organised as follows: in Section 2 a brief description on skin-colour segmentation is introduced. Section 3 presents the general theory of SVM and the facial features we used. Experimental results are given in Section 4 and follows with conclusion in Section 5.

2. SKIN COLOUR SEGMENTATION

The objective of skin modelling is to find a decision rule that could discriminate between skin and non–skin pixels. Although different models have been explored in the literature, most skin detection models use their own, publicly unavailable datasets, and in consequence it is difficult to evaluate the performance of different models. To have a fair comparison between different skin models, we studied four different skin detectors described in [8] and tested their performances under same conditions.

2.1. Explicitly Thresholding

The most easy way to build a skin classifier is to use explicit rule based on prior knowledge of human skin. The advantages of this method are its simplicity and fast. For normalized RGB colour space, we classify a pixel as skin pixel if it satisfies the following conditions:

$$\begin{gathered} R > 95,\ G > 40,\ B > 20,\ |R-G| > 15,\ R > G, \\ R > B,\ and\ \max\{R,G,B\} - \min\{R,G,B\} > 15. \end{gathered} \tag{1}$$

2.2. Bayes classifier

For Bayes classifier, we compute the probability $P(skin|c)$ of observing a skin pixel given a concrete colour value c based on a conditional probability $P(c|skin)$ – a probability of observing colour c, knowing that it is a skin pixel. By using the Bayes rule, we can compute this probability as:

$$P(skin|c) = \frac{P(c|skin)P(skin)}{P(c|skin)P(skin) + P(c|\neg skin)P(\neg skin)}. \tag{2}$$

The $P(c|skin)$ and $P(c|\neg skin)$ are directly computed from the sample population, whilst the prior probabilities $P(skin)$ and $P(\neg skin)$ are estimated from the overall number of skin and non–skin samples in the training set. From the above equation, the ratio of $P(skin|c)$ and $P(\neg skin|c)$ can be written as:

$$\frac{P(skin|c)}{P(\neg skin|c)} = \frac{P(c|skin)P(skin)}{P(c|\neg skin)P(\neg skin)}. \tag{3}$$

Comparing this ratio to a threshold Θ gives the skin/non–skin decision rule, which after some manipulations can be rewritten as:

$$\frac{P(c|skin)}{P(c|\neg skin)} > \Theta. \tag{4}$$

2.3. Gaussian Modeling

For a single Gaussian, the skin colour distribution can be modelled by a Gaussian joint probability density function (pdf) which is defined as:

$$P(c|skin) = \frac{1}{2\pi|\Sigma_s|^{1/2}} e^{-\frac{1}{2}(c-\mu_s)^T \Sigma_s^{-1}(c-\mu_s)} \tag{5}$$

where c is a colour vector and μ_s and Σ_s are the mean vector and covariance matrix respectively. The probability $P(c|skin)$ can be used directly as a measure of how likely a colour belongs to skin colour. For more complex skin distributions, the Gaussian mixture model (GMM) can be used, in which case the pdf is given by:

$$P(c|skin) = \sum_{i=1}^{k} \pi_i P_i(c|skin) \tag{6}$$

where k is the number of mixture components, π_i are the mixing parameters and $\pi_i P_i(c|skin)$ are each single Gaussian in the mixture model. The model parameters can be estimated by using the EM algorithm [4].

2.4. Elliptic Boundary Model

More recently, Lee and Yoo [5] argued that Gaussian model is not well enough to approximate the skin distribution because of the asymmetry of the skin cluster with respect to its density peak. They proposed an elliptical boundary model based on their observations that the skin cluster in approximately elliptic in shape. The elliptical boundary model is defined as:

$$\Phi(c) = (c-\phi)^T \Lambda^{-1} (c-\phi) \tag{7}$$

where the model parameters are estimated by as follows:

$$\phi = \frac{1}{n}\sum_{i=1}^{n} c_i; \quad \Lambda = \frac{1}{N}\sum_{i=1} nf_i(c_i-\mu)(c_i-\mu)^T \tag{8}$$

$$\mu = \frac{1}{N}\sum_{i=1} nf_i c_i; \quad N = \sum_{i=1} nf_i. \tag{9}$$

where n is the number of distinctive training colour vectors c_i of the training skin pixels, N is the total number of samples, and f_i is the number of skin samples of colour vector c_i. It should be noted that n not equals N. A pixel is classified as skin pixel if $\Phi(c) < \theta$, where θ is a threshold value.

For skin colour segmentation, we quantise the r and g colour components into 100×100 of binary value 0 or 1 in the normalised RGB colour space in which 1 corresponds to skin colour and 0 non-skin colour. 651,690 skin pixel samples and 4,325,376 of non–skin pixels have been collected as our training samples. In order to gather enough skin samples, we cropped face regions from images in Corel collection and images obtained from Internet with different lighting conditions and resolutions. Another 431,308 skin pixels and 3,342,356 non–skin pixels are collected in similar way as our testing set.

To our surprise as depicted in Table 1, the non–parametric models (empirically thresholding and Bayes SPM) are outperformed than parametric models (Gaussian mixture model and elliptical boundary model). Among them, the Bayes SPM performs best and follows by empriically thresholding. Although Gaussian mixtures and elliptical boundary model have high true positive, they come with high false positive as well. Once the potential skin regions have been detected, we apply 3×3 and 5×5 median filter to eliminate noises. Then the connected component algorithm in 8-connectivity is applied and regions smaller than a predefined threshold (area size of 30 pixels) are removed. Finally, holes in the regions are filled to produce a smoothly connected binary image. Once the image has been processed, it is passed to the face classifier for face detection. Fig. 1 shows the framework of the whole system.

Method	True Positive	False Positive
Empirically Thresholding	92.1%	15.55%
Bayes classifier	89.83%	7.6%
Gaussian Mixture model	92.12%	24.84%
Elliptical boundary model	90.34%	39.54%

Table 1. Performance of different skin detection models

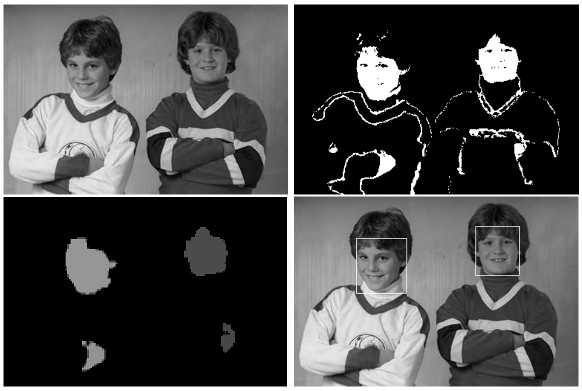

Fig. 1. From left to right and top to bottom are the original colour image, the binary segmented image, result of post–processed binary skin map and the detected faces, each bounded by a rectangle.

3. FACE DETECTION WITH SVM

Unlike many approaches that model the intensity level appearance of face by using exhaustive window searching, we looks for more computationally efficient approach. We extract the edge directional histogram (EDH) from face region to model the edge orientation distribution of face appearance. The edge information is extracted by using the Sobel operator from a 2–D array of pixels $I(x,y)$. The convolution of images with the two filter masks produce two edge maps $G_x(x,y)$ and $G_y(x,y)$ along the x and y direction respectively:

$$G_x(x,y) = K_x * I(x,y) \quad G_y(x,y) = K_y * I(x,y) \tag{10}$$

Fig. 2. Results of face detection.

where K_x and K_y are the two Sobel filter masks. The edge directional information $\Phi(x, y)$ is obtained by:

$$\Phi(x, y) = \arctan(\frac{G_y(x, y)}{G_x(x, y)}) + \frac{\pi}{2}. \quad (11)$$

The edge direction takes value from 0 to 2π but we quantized the value to 36 levels to produce a 36-bins EDH. By mapping the edge map with the binary map obtained from skin detection, the EDH is computed for each region and normalized with respect to the region area so that a feature vector of 36 values is obtained and used for classification.

After all the pre–processing steps, we employ the support vector machine (SVM) for training and classification. SVM is a maximum margin classification tool based on structural risk minimization principle [1]. The goal of SVM is to produce a model which predicts target class value of data instances in the testing set. Given a set of labelled pairs of training instances (x_i, y_i), $i = 1, \ldots m$ where $x_i \in R^n$, $y \in 1, -1$ and m is the number of samples, and a kernel function $K(x_i, x_j)$, the SVM is formed as solving the following optimization problem:

$$\overline{\alpha} = \arg\min \tfrac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j K(x_i,x_j) - \sum_{i=1}^{m}\alpha_i$$

$$\text{subject to} \begin{cases} 0 \le \alpha_i \le C, \quad i = 1,\dots m \\ \sum_{i=1}^{m}\alpha_i y_i = 0 \end{cases} \tag{12}$$

where C is the penalty parameter of the error term and $\overline{\alpha}_i$ are the support vectors. The classifier is given by:

$$f(x) = sign(\sum_{x_i \in \overline{\alpha}} \overline{\alpha}_i y_i K(x_i, x) + \overline{b}) \tag{13}$$

$$\overline{b} = \frac{1}{2}\sum_{x_i \in \overline{\alpha}} \overline{\alpha}_i y_i [K(x_r, x_i) + K(x_s, x_i)] \tag{14}$$

where x_r and x_s are different types of support vectors [7]. The SVM maps the training vectors x_i into a higher dimensional space by the kernel function and then finds a linear separating hyperplane with the maximal margin in this higher dimensional space. The four basic kernels are:

- linear: $K(x_i, x_j) = x_i^T x_j$
- polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \quad \gamma > 0$
- radial basis function (RBF): $K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2), \quad \gamma > 0$
- sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

where γ, r, and d are all kernel parameters.

4. EXPERIMENTAL RESULTS

Two tests are performed to evaluate the system performance. The first experiment is to test whether the EDH could well discriminate the face class from other non–face classes by using the FERET dataset. The FERET dataset contains 400 face images captured from 40 people while each subject have 10 images from different orientations. Feature vectors are extracted according to the procedures described before and SVM is trained for classification. We first train an initial face classifier from 200 face images picked randomly from the FERET dataset together with 100 randomly selected non–face images. Then by using the bootstrapping algorithm [3] we obtain another 200 non–face training samples and retrain our face classifier. We use the rest 200 FERET images and 300 non–face images that different from the training set for our first test. The RBF kernel with parameters $\gamma = 2000$ and $C = 100$ gives the best performance in which the recognition rates are 92.6% and 90.5% for the training and testing sets respectively. The result suggests that the human face occupies a tight cluster of edge distribution in the feature space which could be discriminated from other classes.

In the second experiment, the system is tested on real photos. Unfortunately, we have difficulties in collecting enough colour face images since most of the datasets are not publicly available. We cropped 80 face samples from Corel images with 200 FERET face images and 400 non–face samples to train the SVM classifier with the same procedures described before. 185 test images of various types are collected which contains 322 faces in total. 86.33% detection rate is achieved with 71 false detection. Most of the missed detection are due to the misclassification of skin pixels and occasionally face regions are too small and hence are eliminated during the post–processing step. The misclassification of skin pixels are mainly because of these pixels are exposed to too much light or are under shadows. Sometimes the face regions are connected to non–face regions caused by partial occlusions which change the edge distributions significantly and make the detection fails.

5. CONCLUSION

In this paper we propose a method of face detection by combining skin colour segmentation and learning the edge distribution of faces. Since most of the non–face candidates are filtered by the skin segmentation, the face detection procedure is greatly speed up. The preliminary results are very encouraging. However, two main errors have been encountered in this study. First, the robustness of skin segmentation has to be improved to tolerate larger variations of illumination changes. Second, misclassification is caused by partial occlusions that divided face into separate clusters or other objects connected to face which changes the edge distributions significantly. To tackle this problem, we plan to include split and merge algorithm into the skin segmentation step and train the classifier with bigger dataset with more side–view faces to increase the generalisation capacity of our algorithm.

6. ACKNOWLEDGEMENT

We would like to thank SCHEME for the financial support.

7. REFERENCES

[1] C. J. C. Burges, ”A Tutorial on Support Vector Machines for Pattern Recognition”, Knowledge Discovery and Data Mining, vol. 2(2), 1998.

[2] H. Ai, L. Ying and G. Xu, ”A subspace approach to face detection with support vector machines”, 16th Inte. Conf. on Pattern Recognition (ICPR’02), vol. 1, Quebec City, Canada, August 11–15 2002.

[3] H. A. Rowley, S. Baluja, and T. Kanade, ”Neural network-based face detection, IEEE Trans”, Pattern Anal. Mach. Intell. 20, 23-38, Jan 1998.

[4] J. Bilmes, ”A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models”, Technical Report ICSI–TR–97–021, University of Berkeley, 1998.

[5] J. Y. Lee and S. I. Yoo, ”An elliptical boundary model for skin color detection”, Proc. of the 2002 Inter. Conf. on Imaging Science, Systems, and Technology, Las Vegas, USA, Jun 2002.

[6] R. Chellappa, C. L.Wilson and S. Sirohey, ”Human and machine recognition of faces: A survey”, Proc. IEEE, vol. 83, no. 5, 1995.

[7] S. R. Gunn, ”Support Vector Machines for Classification and Regression”, Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.

[8] V. Vezhnevets, V. Sazonov and A. Andreeva, ”A Survey on Pixel-Based Skin Color Detection Techniques”, Graphicon 2003, Moscow, Russia, Sep 2003.