2001

Skin colour based face detection

Son Lam Phung
Edith Cowan University

Douglas K. Chai
Edith Cowan University

Abdesselam Bouzerdoum
Edith Cowan University

10.1109/ANZIIS.2001.974071
This is an Author's Accepted Manuscript of: Phung, S. L., Chai, D. K., & Bouzerdoum, A. (2001). Skin colour based face detection. Proceedings of 7th Australian and New Zealand Intelligent Information Systems Conference. (pp. 0). Australia. IEEE. Available here

Skin Colour Based Face Detection

Son Lam Phung, Douglas Chai and Abdesselam Bouzerdoum
Edith Cowan University
100 Joondalup Drive, Joondalup
Western Australia 6027
{s.phung, d.chai, a.bouzerdoum}@ecu.edu.au

Abstract. This paper describes a new approach to face detection. A colour input image is first processed using neural-networks to detect skin regions in the image. Each neural network separates skin and non-skin pixels on the basis of chrominance information. The skin-colour classifier employs the committee machine technique, which improves skin colour detection by combining the classification results of a set of multilayer perceptrons (MLPs). The skin colour classifier achieves a classification rate of 84% compared to 81% for the best individual MLP classifier. The output of the committee machine is processed by a 2D smoothing filter before being converted into a binary map using a threshold. Finally, several post-processing techniques based on shape and luminance features are proposed for rejecting non-facial regions.

1 Introduction

Face detection, which aims to detect the presence and subsequently the position of a face or faces in an image, is an important task in automated facial image analysis. The result of face detection enables tasks such as face recognition and facial expression analysis to be performed on focused image areas with irrelevant data such as the background being excluded.

A major challenge in face detection is to cope with a wide range of variations in the human face pattern that are due to factors such as different lighting, face orientation, face size, facial expression and people ethnicity. The presence of complex background or extra facial features such as glasses, beard, and moustache also adds to the complexity of the face detection problem.

Many approaches to face detection have been proposed. More well known approaches in recent years are those that use neural networks to detect face-like patterns in the input image [1-2, 9]. An input image is divided into non-overlapping windows of small size and each window is classified as belonging to face or non-face by some form of neural networks. Scale invariance is achieved by repeatedly reducing the window size and merging detection results across different scales, which are computation intensive. Other approaches to face detection include eigenfaces and wavelet packet analysis. The eigenfaces method considers an input pattern as a face if it is sufficiently close to the face subspace, which is spanned by a set of "face-like" vectors. 2D wavelet packet analysis is employed in [5] to extract texture features stored in the intensity plane (Y) of the image, which are then classified into face or non-face using the Bhattacharrya distance.

In this paper, we propose an approach to face detection that first applies a chrominance-based human skin colour model to detect skin regions in an image. Then several heuristic post-processing techniques based on shape and image luminance are used to refine the skin detection result by removing non-facial regions (see the block diagram in Figure 1). The rest of the paper is organised as follows. The major steps of the new face detection approach are described in Section 2. Experimental results of face detection using the proposed approach are reported in Section 3. Concluding remarks are given in Section 4.

2 Face Detection Algorithm

2.1 Neural-Network Based Skin Colour Model

Detecting human skin region is often the first step in many face detection approaches, which works based on the observation that the human skin colour is often distinctive from the colour of other objects and the background scene. The major advantages of this technique are speed and shape-invariance. The entire image can be segmented pixel-wise in real-time into skin and non-skin regions without taking into account spatial inter-pixel dependence. In addition, colour-based segmentation is not affected by changes in size and orientation of faces.

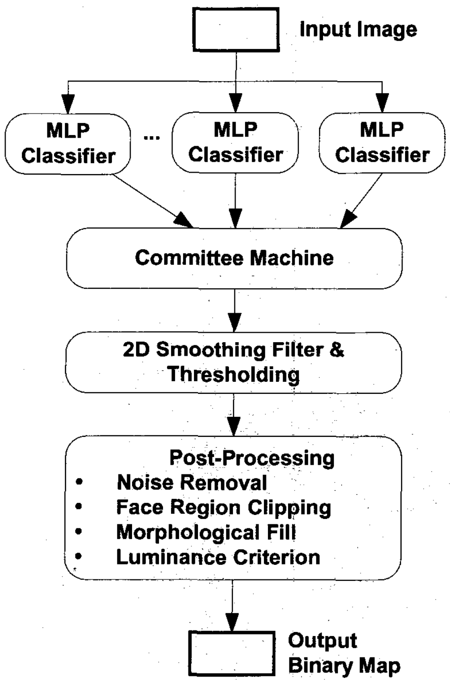


Figure 1: Block diagram of the proposed approach to face detection.

However, for this technique to work, an accurate model of the human skin colour is needed. The model must provide a comprehensive coverage of different types of the human skin colours (*e.g.* white, black, yellow) and remain compact in order to reduce the probability of false detection. Several human skin colour models exist in literature: fixed-range skin colour map (*i.e.* rectangular model) [3], Bayesian-classifier [8], Gaussian models [7], elliptic model [6], and fuzzy model [4]. In [10], we proposed the use of artificial neural network (ANN) as an adaptive classifier of skin and non-skin colour pixels. This section describes the neural-network-based skin colour classifier and new post-processing techniques to refine the neural network output.

The neural network classifies colour pixels into skin or non-skin colours based on the pixel chrominance. The rationale for this technique is that the skin colours for people of different ethnicity all share a marked similarity in chrominance characteristics, which can be employed to distinguish skin and non-skin colours. In comparison, there is a strong variance in the luminance of skin colours.

The colour space used in this work is YCbCr, which is defined in CCIR Recommendation 601-2. This colour space is widely used in still-image and video coding standards such as JPEG, MPEG and H.263. However, the approach described here is also applicable to other colour spaces that have separate luminance and chrominance signals. With the YCbCr space, luminance is stored in Y component, and chrominance is stored in Cb and Cr components. Conversion between YCbCr and RGB colour format is as follows:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256}\begin{bmatrix} 65.738 & 129.057 & 25.064 \\ -37.945 & -74.494 & 112.439 \\ 112.439 & -94.154 & -18.285 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \frac{1}{256}\begin{bmatrix} 298.082 & 0 & 408.583 \\ 298.082 & -100.291 & -208.120 \\ 298.082 & -516.411 & 0 \end{bmatrix}\begin{bmatrix} Y-16 \\ Cb-128 \\ Cr-128 \end{bmatrix} \quad (1)$$

Shown in Figure 2 is the joint distribution of chrominance values for skin colour pixels in Cb-Cr plane. The result was obtained by first manually segmenting skin regions in 100 colour images that contain people of different skin colours under normal lighting; and then normalizing the count for each (Cb, Cr) pair against the total number of skin colour pixels. First and second order statistics for chrominance and luminance of skin colours are shown in Table 1. Note that skin colour pixels distribute in a small region in Cb-Cr plane, albeit the boundary of the region is quite irregular.

Table 1: Statistics of skin colours

	Y	Cb	Cr
Mean	156.5	111.6	149.5
Std Dev	31.1	9.3	8.4
Min	50	74	130
Max	230	135	179
Median	158	112	150
Interquartiles	[137,179]	[108,118]	[144,154]

In our approach, the neural network classifier is a multilayer perceptron (MLP). The network accepts two inputs which are Cb and Cr values of an input pixel and is trained to generate an output between 0 and 1 indicating the probability of the colour pixel being skin colour (0: non-skin colour, 1: skin colour). The training data can be obtained by manually selecting skin region and non-skin region patches from images of people of different skin colours.

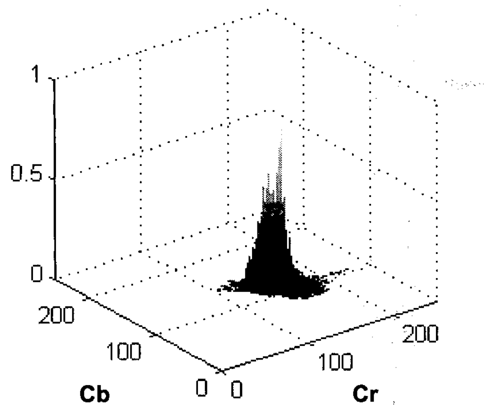


Figure 2: Distribution of skin colour pixels in Cb-Cr plane.

Committee Machine

Training of neural network often starts with the initialisation of network parameters (*i.e.* weights and biases) to small random values. Because of this stochastic nature of training, the performance of a trained network may deviate from that of another trained network even when both have the same structure (*i.e.* number of layers, number of neurons per layer, activation functions). This is particularly the case when there are more than one set of network parameters that fit the training data.

The committee machine (CM) technique improves the stability of the model, by fusing the results of a number of independent neural network classifiers, which can be different in structure or trained with different training data. A committee machine is characterised by the number of independent classifiers N and an arbitration scheme f used to determine the final outcome. Let $o_1, o_2, \ldots, o_N$ be the outputs of the independent classifiers. The output of the committee machine is determined as follows:

$$c = f(o_1, o_2, \ldots, o_N) \qquad (2)$$

Several arbitration schemes are possible ranging from mean selector, median selector and majority voting to the more sophisticated scheme that employs a dedicated neural network. The arbitration scheme used in our work is the median selector:

$$\{s_1, s_2, \ldots, s_N\} = \{o_1, o_2, \ldots, o_N\},$$

$$s1 \le s2 \le \ldots \le s_N$$

$$f(o_1, o_2, \ldots, o_N) = s_{\lceil \frac{N+1}{2} \rceil} \qquad (3)$$

Spatial Filtering

The probability of a pixel being skin colour is closely related to those probabilities of its neighbouring pixels. This fact can be utilised to refine, through spatial filtering, the results of skin-colour detection in the previous step.

Let **C** be the skin-colour probability matrix generated by the committee machine for a given input image. A convolution kernel **K** is applied on matrix **C** to generate the filtered output matrix **P**:

$$p(i,j) = \sum_{(p,q)\in N_{ij}} c(p,q)\, k(i-p, j-q) \qquad (4)$$

For a given pixel (i, j), the convolution kernel K is centred about it and the summation is for all pixels in its neighbourhood N_{ij}, which has the same size as that of the convolution kernel **K**. Here, we use a common convolution kernel: the averaging filter whose effect is illustrated in Figure 6c.

The output of the skin-colour classifier, after thresholding, is a binary map whose entries indicate skin (referred hereafter as *object* region) or non-skin (non-object region). For face detection purpose, even with a highly accurate colour classifier, two sources of errors must be addressed. First, some facial regions (*e.g.* eyes) do not have skin colours. Second, other body regions and background objects may have skin appearance. The next section presents several post-processing techniques for the face detection problem.

2.2 Heuristic Post-Processing Techniques For Face Detection

Noise Removal

All objects in the binary map whose area compared to that of the largest object falls below a threshold are removed. Our experiments have shown that this eliminates effectively a large number of false detections because they are often scattered and have relatively small area. Within individual object, all cross sections (horizontal or vertical) that are relatively short compared to the corresponding longest cross-section (horizontal or vertical) are removed. Used in conjunction with the first technique, this will remove "spikes" connected to the facial regions.

Face region clipping through eye/eyebrow and mouth detection

Because eye colour is invariantly different from skin colour, the eyes are represented in the binary map as two *non-object* regions. Eyebrow colour is either similar to or different from skin colour depending on people ethnicity. However, due to the close proximity between the eye and the eyebrow, the following method is sufficient to locate the eyes/eyebrows in the binary map:

(*i*) For each object in the binary map, define two search regions, which are the top-left and top-right quarters of the object (Figure 3).
(*ii*) An eye/eyebrow region is detected as the largest non-object region in a search region. The presence of eyes/eyebrows in the search regions may be used to verify that the object is a face candidate.
(*iii*). The coordinates [e_x, e_y] of eye/eyebrow are estimated as the *centre-of-mass* of an eye/eyebrow region (*e.g.* by averaging horizontal and vertical coordinates of all pixels in the eye/eyebrow region).

We also carry the above steps to search for the mouth in the bottom half of the object. The computed coordinates of the eyes (e_x_1, e_y_1), (e_x_2, e_y_2) and the mouth (m_x, m_y) are used to clip the object so that only the face region is selected and other detected body regions such as neck are removed (see Figure 3). Note that, depending on the face orientation, e_y_1 may not be the same as e_y_2. In estimating the clipping region, we approximate the eye vertical coordinate e_y as the mean of e_y_1 and e_y_2.

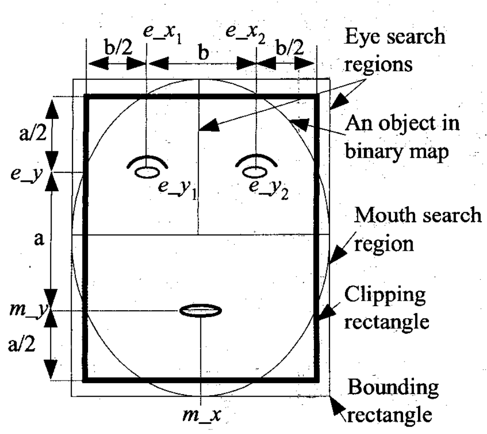


Figure 3: Face region clipping.

Morphological Operations

Morphological dilation and erosion operations are performed on the binary map to fill "holes" within each object region. These "holes" result from facial regions without skin colour appearance. After this step, the binary map will indicate the location of faces in the input image.

Luminance Criterion

The human face is a 3D object whose regions not only reflect light differently but also cast shadows onto other regions. In addition, regions such as eyes, eyebrows with different colours from the skin are present. Together, these lead to a high degree of variance in intensity of the face image. In comparison, the colours of non-face objects detected in the previous step tend to show stronger homogeneity. This suggests the following technique to remove the remaining non-faces, which is the last step in our face detection approach:

(*i*) For each object i in the binary map, read the corresponding region in the original colour image and compute the standard deviation s_i of all intensity values of the region.
(*ii*) Find the maximum standard deviation:

$$s_{max} = \max(s_1, s_2, \ldots, s_N) \quad (5)$$

(*iii*) An object j is removed if its intensity variation measure is small compared to s_{max}:

$$s_j/s_{max} < \theta_s \quad (6)$$

3 Experimental Results

A committee machine consisting of 15 multilayer perceptrons was used. These networks were divided into three equal groups each having 5 networks. Networks in the first group had a size of 2-8-1 (*i.e.* 2 inputs, a hidden layer of 8 neurons, 1 output), the second group 2-5-3-1 and the last group 2-4-4-1. Training was done using the Levenberg-Marquardt algorithm. Training data were prepared by using a custom program to manually select skin and non-skin rectangular patches from a training set of 100 CIF-sized (352×288) colour images. The training data collected were processed and subsampled giving a training set of near 1,300 samples.

The skin colour decision boundary in Cb-Cr plane generated by this committee machine is shown in Figure 4. The result was obtained by feeding into the committee machine all points in the (8-bit) Cb-Cr plane.

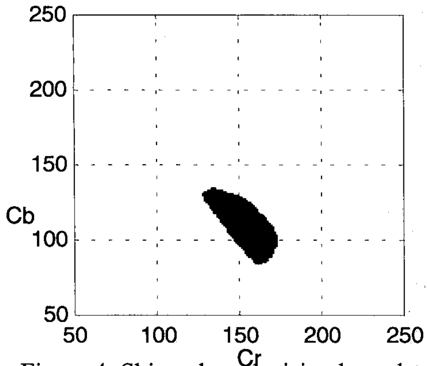


Figure 4: Skin colour decision boundary generated by a committee machine.

Table 2: Skin/non-skin classification comparison.

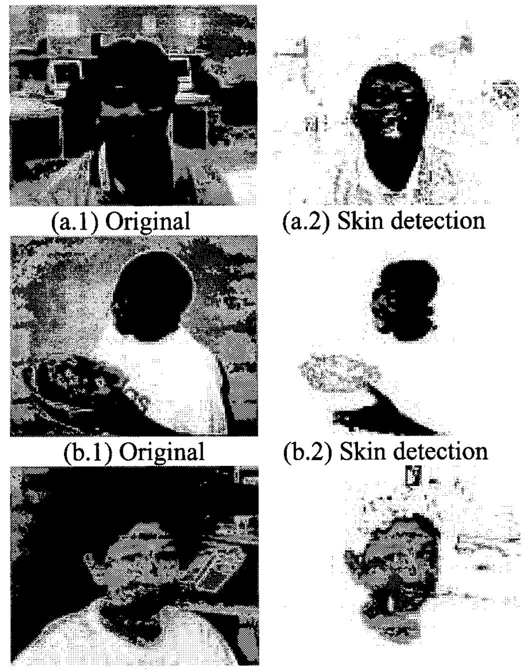
Methods	Classification Rate (%)
Best individual MLP	81.0
Committee Machine	84.0*
Bayesian Classifier [8]	75.8
Elliptic Model [6]	68.0
Gaussian Model [7]	68.0
Rectangular Model [3]	65.7

* False detection=11.9%, False dismissal=4.1%

The skin colour classifier was tested against a comprehensive test set of 1,250,000 samples. These samples included pixels manually extracted from the skin regions and from the background of a test set of 100 CIF-sized images. Comparison of the proposed method and others are shown in Table 2. All skin detection methods shown in the table were implemented and then tested on the same test set. Descriptions of the last four methods in the table can be found in the corresponding references. It is fair to conclude that the committee machine approach offers improvement over the single MLP classifier approach and a better classification rate over other methods.

Figure 5 shows exemplar results when the skin colour classifier using the committee machine is applied to images of people of different skin colour types: yellow, black and white.

However, as seen in Table 2, the false detection rate (when a non-skin colour pixel is detected as skin colour) is still high. This can be attributed partially to the fact that a skin colour classifier cannot reject pixels in the background having skin-colour appearance.



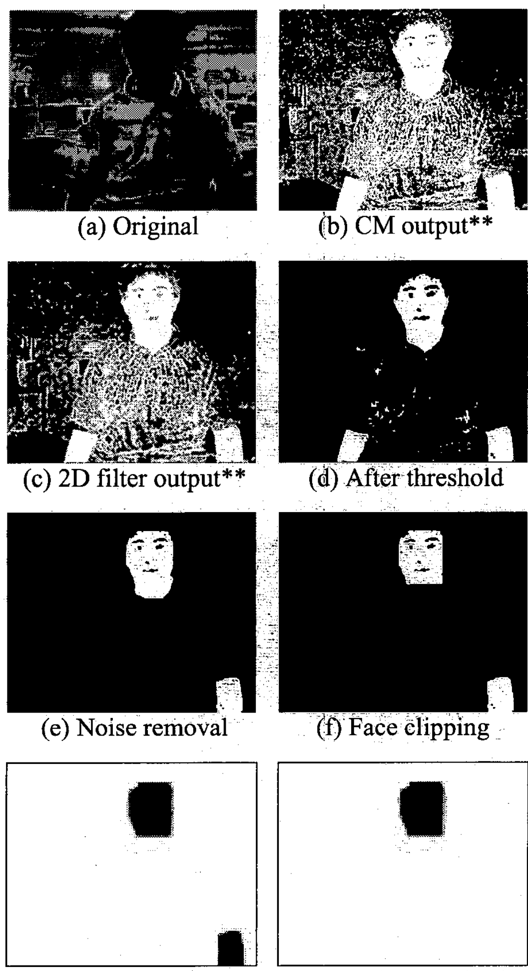
(a.1) Original (a.2) Skin detection
(b.1) Original (b.2) Skin detection
(c.1) Original (c.2) Skin detection

Figure 5: Skin detection for people of different skin colours: (a) yellow, (b) black (c) white.

The effects of individual stages in our proposed face detection method is illustrated in Figures 6 and 7. The skin threshold used is $\theta = 0.9$; the luminance threshold used is $\theta_S = 0.7$. Note that, the CM output and the 2D filter output can be considered as "skin probability" matrices. These matrices are scaled and displayed as gray-scale images for illustrative purpose.

4 Conclusion

A multi-stage face detection approach based on skin detection is presented. Combining the classification results of several neural network classifiers improves the skin detection. The result of skin detection is further refined using a number of post-processing techniques. These techniques reliably reduce the number of false detections in the previous stages by removing background of skin colour appearance, and adding facial pixels that do not have skin colour. Luminance and shape features are used to distinguish between face and non-face patterns. In further studies, more sophisticated classifier of face and non-face patterns will be investigated.



(a) Original (b) CM output**

(c) 2D filter output** (d) After threshold

(e) Noise removal (f) Face clipping

(g) Morphological fill (h) Luminance check

** Scaled to be displayed as gray-scale images.

Figure 6: Face detection process.

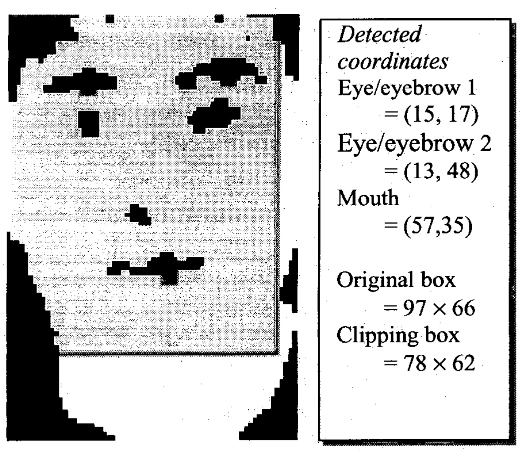


Figure 7: Face region clipping results (5.f above).

5 References

[1] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions on PAMI*, vol. 20, no. 1, pp. 23-38, January 1998.

[2] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Transactions on PAMI*, vol. 20, no. 1, pp. 39-51, January 1998.

[3] D. Chai and K. N. Ngan, "Face segmentation using skin color map in video phone applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 551-564, June 1999.

[4] H. Wu, Q. Chen, and M. Yachida, "Face detection from color images using a fuzzy pattern matching method," *IEEE Transactions on PAMI*, vol. 21, no. 6, pp. 557-563, June 1999.

[5] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 264-277, September 1999.

[6] N. Bojic and K. K. Pang, "Adaptive skin segmentation for head and shoulder video sequence," *Proceedings of Visual Communication and Image Processing 2000*, Perth, Australia, vol. 4067, pp. 704-711, June 2000.

[7] B. Menser and M. Wien, "Segmentation and tracking of facial regions in color image sequences," *Proceedings of Visual Communication and Image Processing 2000*, Perth, Australia, vol. 4067, pp. 731-740, June 2000.

[8] D. Chai and A. Bouzerdoum, "A Bayesian approach to skin color classification in YCbCr color space," *Proceedings of IEEE TENCON'2000*, Kualu Lumpur, Malaysia, vol. II, pp. 421-424, September 2000.

[9] R. Feraud, O. J. Bernier, J. E. Viallet, and M. Collobert, "A fast and accurate face detector based on neural networks," *IEEE Transactions on PAMI*, vol. 23, no. 1, pp. 42-53, January 2001.

[10] S. L. Phung, D. Chai, and A. Bouzerdoum, "A universal and robust human skin color model using neural networks," *Proceedings of International Joint Conference on Neural Networks 2001*, Washington D.C., U.S.A., pp. 2844-2849, July 2001.