EE368 Project
Face Detection and Gender Recognition

Michael Bax, Chunlei Liu, and Ping Li

26 May 2003

1 Introduction

The full face detection and gender recognition system described here is made up of a series of components connected in both serial and parallel, as illustrated in Figure 2. The successive stages are explained in detail in the body of this report.

2 Colour-based segmentation

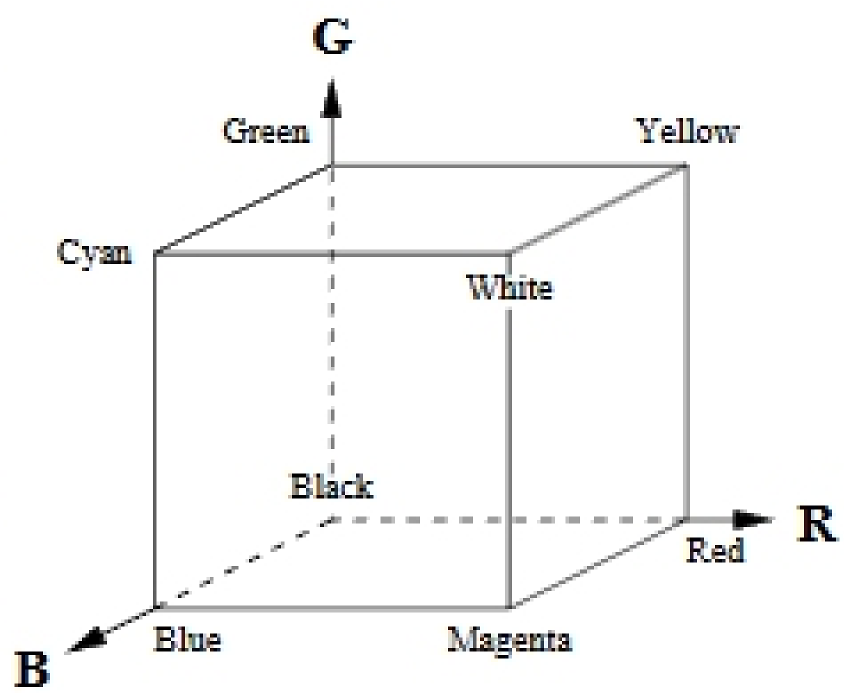
The first step in this face detection algorithm is that of colour segmentation. The goal is to remove the maximum number of non-face pixels from the images in order to narrow the focus to the remaining predominantly skin-coloured regions.

2.1 Colour space selection

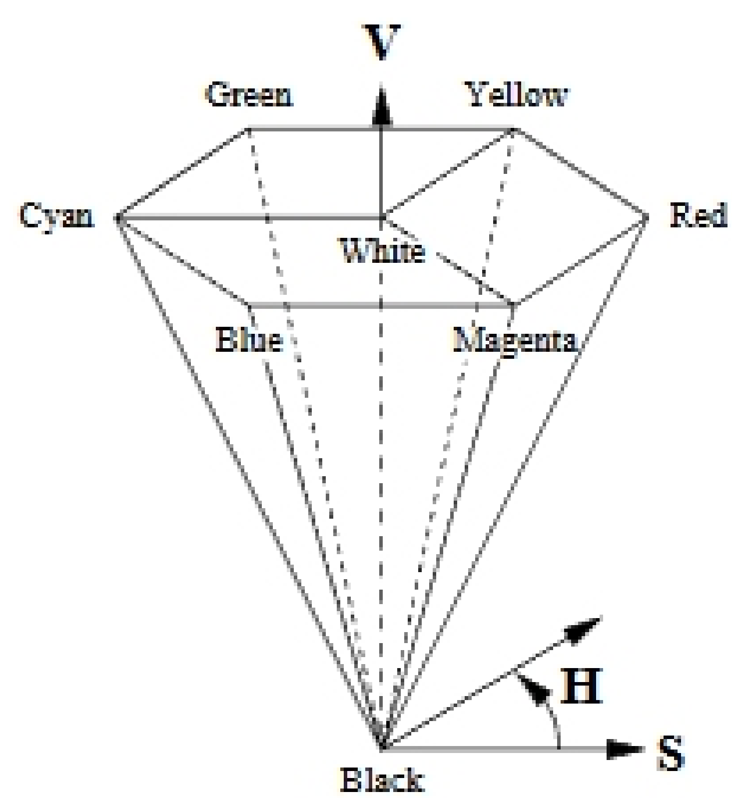
The first step in performing colour-based segmentation is choosing an appropriate colour space in which to operate from the wide variety of choices such as RGB, HSV, CMYK, YCbCr, etc [1]. Of these, RGB (red-green-blue) and HSV (hue-saturation-value) have been the most widely used. Figure 1 illustrates the geometries of the two spaces.

By way of example, HSV representation has certain advantages over RGB when it comes to face detection. As Garcia *et al* [2] note, skin colours are sensitive to the lighting condition. In the RGB space, each of the three components may exhibit substantial variation under different lighting environments. In HSV space, however, the hue and saturation components are virtually unchanged.

Figure 3 shows the histograms of RGB component values of both face and non-face pixels over all seven training input images. Similarly, Figure 4 shows the histograms of the same images in HSV space, where the S component and in particular the H component are well-clustered for face-pixels, while H and S are spread over a wide range for the remainder of the image. This observation favours using an HSV colour space if only a simple thresholding colour segmentation is desired.



(a) RGB model



(b) HSV model

Figure 1: The RGB and HSV colour models.

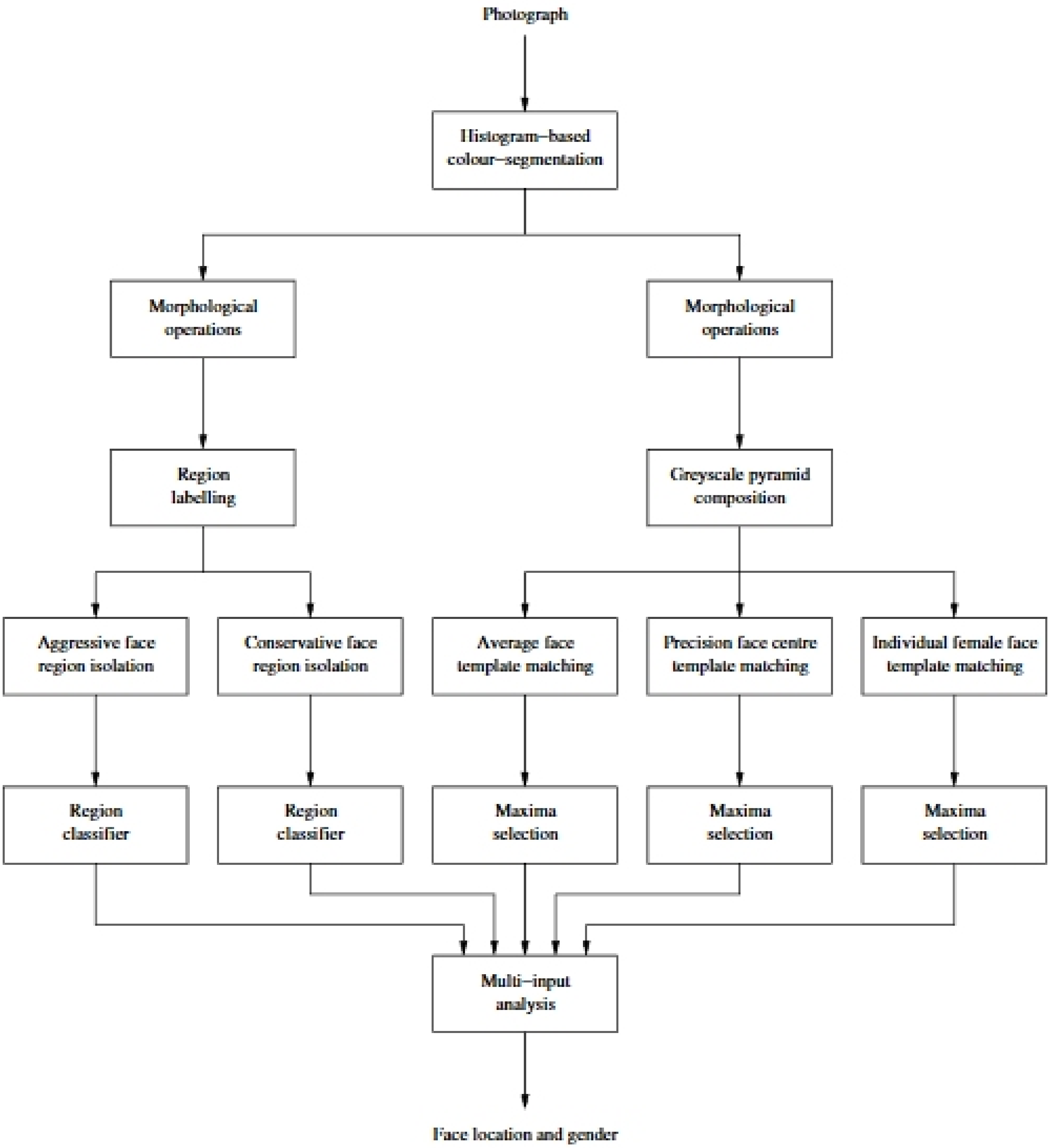


Figure 2: The face detection system schematic.

The same is true of the YCbCr colour-space, but the clustering does not quite as readily lend itself to projection onto the axial dimensions, and hence requires a somewhat more complicated 2D basis for segmentation.

Manual segmentations of this type suffer as a consequence of approximating a complex bounding region using simple geometric relations. Given that the unknown images were taken in similar lighting and with similar equipment to the training images, a probability-based segmentation may be used.

In this case, the choice of spatial domain is not as significant. However, the fundamental characteristics of a colour space can complicate processing — for example, the non-Cartesian nature of the HSV colour cone indicates that non-uniform quantization may be appropriate.

2.2 Probability-based classification

Due to the large number of pixels in the training images, there is enough data to create a reasonable estimate of the underlying probability density functions for both face and non-face skin colours. Let

$$f_{\mathbf{X}}(\mathbf{x}|\mathbf{X} \in \Phi)$$

be the colour space probability density function for a pixel vector $\mathbf{X}$ in the set of face pixels Φ; the vector components X_i represent the colour components — R, G and B in this case. Similarly, let

$$f_{\mathbf{X}}(\mathbf{x}|\mathbf{X} \notin \Phi)$$

be the probability density function for non-face skin pixels. These two density functions can be estimated from the empirical distribution of the pixels in the training images.

Conversely, let the probability that a given colour pixel is part of a face be

$$p_{\mathbf{X}}(\mathbf{X} \in \Phi|\mathbf{X} = \mathbf{x}).$$

The Bayesian formula [5] gives

$$\begin{aligned} R &= \frac{p_{\mathbf{X}}(\mathbf{X} \in \Phi|\mathbf{X} = \mathbf{x})}{p_{\mathbf{X}}(\mathbf{X} \notin \Phi|\mathbf{X} = \mathbf{x})} \\ &= \frac{f_{\mathbf{X}}(\mathbf{x}|\mathbf{X} \in \Phi)}{f_{\mathbf{X}}(\mathbf{x}|\mathbf{X} \notin \Phi)} \times \frac{\pi_\Phi}{\pi_{\bar{\Phi}}} \end{aligned} \tag{1}$$

where π_Φ and $\pi_{\bar{\Phi}}$ are the prior probabilities of a randomly-selected pixel falling in a face or the background, respectively. Without further information, the ratio of $\frac{\pi_\Phi}{\pi_{\bar{\Phi}}}$ can be estimated from the ratio of the total number of face pixels to the total number of non-face pixels.

For convenience, Equation 1 may be reformulated as

$$\begin{aligned} \log R &= \log \frac{p_{\mathbf{X}}(\mathbf{X} \in \Phi|\mathbf{X} = \mathbf{x})}{p_{\mathbf{X}}(\mathbf{X} \notin \Phi|\mathbf{X} = \mathbf{x})} \\ &= \log \frac{f_{\mathbf{X}}(\mathbf{x}|\mathbf{X} \in \Phi)}{f_{\mathbf{X}}(\mathbf{x}|X \notin \Phi)} + \log \frac{\pi_\Phi}{\pi_{\bar{\Phi}}}. \end{aligned} \tag{2}$$

If the prior information is known, the classification rule would normally be:

$$\begin{array}{ll} \mathbf{X} \in \Phi & R \geq R_0 = 1, \textit{i.e.} \log R \geq \log R_0 = 0 \\ \mathbf{X} \notin \Phi & \text{otherwise.} \end{array}$$

However, the prior information is not necessarily available; conversely, in many situations it is desirable to adjust the classification threshold R_0. For example, since in this approach colour-based segmentation is used for subsequent face detection, it is desirable to bias towards mis-classifying a non-face pixel as a face pixel rather than the reverse. Consideration of these factors yields the following classification rule

$$\begin{array}{ll} \mathbf{X} \in \Phi & \log \dfrac{f_{\mathbf{X}}(\mathbf{x}|\mathbf{X} \in \Phi)}{f_{\mathbf{X}}(\mathbf{x}|\mathbf{X} \notin \Phi)} + \alpha \geq 0 \\ \mathbf{X} \notin \Phi & \text{otherwise,} \end{array} \tag{3}$$

where

$$\alpha = \log \frac{\pi_\Phi}{\pi_{\bar{\Phi}}} - \log R_0 \tag{4}$$

is the undecided parameter to be chosen.

A range of values for α for image pixel classification in both RGB and HSV space were evaluated, the results of which are shown in Figure 5. The total classification error is composed of false positive error (mis-classifying non-face pixels as face pixels) and false negative error (mis-classifying face pixels as non-face pixels).

The two sub-figures are quite similar. The false negative error increases with α, while false positive error decreases with increasing α. The total error reaches minimum at $\alpha = 3$. This optimal value of α is supported by the observation that, over all 7 training images, the ratio of the total number of face pixels to non-face pixels is roughly $\frac{1}{16}$; $\log \frac{1}{16} = -2.77$.

Because this colour-based segmentation is only the first step in the larger face detection algorithm, it is best to retain as many face pixels as possible, minimizing false negative error (even at the cost of including additional non-face pixels), leading to the chosen value of $\alpha \approx 2$.

Although the performance differences between RGB and HSV colour space are quite small, RGB is nonetheless superior [3], and that was the colour space chosen for this algorithm.

In practice an empirical histogram derived from manual image segmentation is somewhat noisy. As Figure 6 shows, it has rough edges and isolated bins that are not likely features of the true distribution. In addition, pixel value noise in unknown images may cause marginal pixels to “skip” out of the histogram bin in which they should fall — abrupt discontinuities in the estimated histogram will exacerbate this problem.

The solution is to filter the empirical histogram to smooth out the transients locally, but without losing the