Performance Evaluation of Single and Multiple-Gaussian Models for Skin Color Modeling

TIBÉRIO S. CAETANO, SÍLVIA D. OLABARRIAGA, DANTE A. C. BARONE

Instituto de Informática, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Porto Alegre, RS, Brasil
{caetano,silvia,barone}@inf.ufrgs.br

Abstract. We present an experimental setup to evaluate the relative performance of single gaussian and mixture of gaussians models for skin color modeling. Firstly, a sample set of 1,120,000 skin pixels from a number of ethnic groups is selected and represented in the chromaticity space. In the following, parameter estimation for both the single gaussian and seven (with 2 to 8 gaussian components) gaussian mixture models is performed. For the mixture models, learning is carried out via the expectation-maximisation (EM) algorithm. In order to compare performances achieved by the 8 different models, we apply to each model a test set of 800 images - none from the training set. True skin regions, representing the ground truth, are manually selected, and false positive and true positive rates are computed for each value of a specific threshold. Finally, receiver operating characteristics (ROC) curves are plotted for each model, which make it possible to analyze and compare their relative performances. Results obtained show that, for medium to high true positive rates, mixture models (with 2 to 8 components) outperform the single gaussian model. Nevertheless, for low false positive rates, all the models behave similarly.

1 Introduction

A large number of applications require the location of people in digital images, such as face detection [3, 4, 5, 6], face and hand tracking [2, 7, 8] and gesture recognition [9]. As a strategy to reduce the search space for human targets, in several methods the image is initially segmented into "skin" and "non-skin" regions based on pixel colour. The goal is to detect people in the image in a faster and more accurate manner.

Several models for human skin colour have been proposed, most of them based on a single bivariate gaussian kernel to represent the skin cluster in some colour space (e.g. [2, 5, 6]). An alternative to the single-gaussian model is presented in [1], where the skin colour distribution is modelled with a *mixture* of bivariate gaussian components. A comparative study between the single-gaussian (SG) and the double-gaussian models has been presented there, but a more detailed analysis is lacking to understand how the model behaves when a higher number of gaussian clusters is used in the mixture.

This work proposes a performance evaluation technique to analyse the behaviour of models with respect to the number of gaussians. The analysis took into consideration models ranging from 1 to 8 gaussian components, which lead to two main conclusions. Firstly, skin colour mixture models clearly outperform the single gaussian model for medium to high true positive rates. In this case, however, no major differences are noticed among the performances of the 7 mixture models. Secondly, when low false-positive rates are required, all the 8 models tested perform quite similarly.

The paper is organized as follows. In section 2, the feature space for representing the human skin colour and the sample set used to train the models are presented. In section 3, the standard SG model is described. In section 4, we review our GM model approach as an alternative for SG. Section 5 reports the experiments applied to a test set, with models ranging from one to eight gaussian components. Finally, in section 7 we draw the conclusions and outline future work.

2 The Colour Representation and the Training Sample Set

2.1 The Colour Feature Representation

The colour space chosen to represent the human skin colour is the *chromaticity* space. The chromaticities r, g and b are defined as

$$r = \frac{R}{R+G+B}\ ,\quad g = \frac{G}{R+G+B}\ ,\quad b = \frac{B}{R+G+B}\ , \tag{1}$$

where R, G and B denote the red, green and blue components that describe the pixel colour. Here we only use r and g to describe the skin colour, given that the third component depends on the other two ($b = 1 - r - g$). The reason for using this colour space is due to evidences that human skin colour is more compactly represented in chromaticity space than in other colour spaces, such as RGB, HSI, SCT and YQQ [10]. In addition, the chromaticity space is normalized with respect to the illuminant, so it is more robust under lighting variations than other illuminant-dependent colour spaces.

2.2 The Training Sample Set

In order to cover a wide range of skin chromatic characteristics, a set of 1,120,000 skin pixels was collected from images of people from four different databases. These images cover a large spectrum of ethnic groups, such as Caucasian, African, Asian, and Hispanic, each group mostly represented in a separate database. In total, 550 images were collected both from random sites on the Internet and from the Web-available Stirling University face database [11].

The skin samples used to train the model were selected manually, avoiding areas of high luminance variation and highlights. Exactly a quarter of the samples (280,000 pixels) were collected from each of the four databases, aiming at avoiding bias to any ethnic group. Figure 1 shows the distribution of all skin samples in the chromaticity space used to train the models. Its visual inspection suggests that a bivariate gaussian may be a suitable model to fit the distribution, such as presented in section 3.

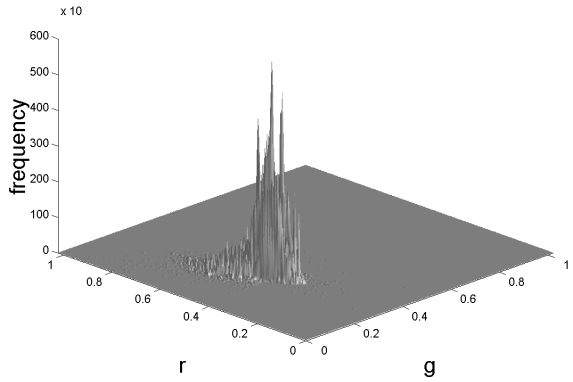

Figure 1: Distribution of the entire sample set of skin pixels in chromaticity space.

3 Single-Gaussian Model (SGM)

Results reported in the literature indicate that a single bivariate gaussian probability density function (*pdf*) can be used successfully as a model for the skin colour, even when multiple ethnic groups are considered [3, 4, 5, 6]. The model can be obtained via the maximum likelihood criterion, which looks for the set of parameters (mean vector and covariance matrix) that maximizes the likelihood function. The likelihood function for a multivariate gaussian *pdf* has a single maximum, and the estimates $\boldsymbol{\mu}$ and Σ for the mean vector and the covariance matrix are obtained analytically and have well-known values given by [12]

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k \tag{2}$$

$$\Sigma = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^t, \tag{3}$$

where $\boldsymbol{\mu}$ is the estimated mean vector, Σ is the estimated covariance matrix, n is the number of observations in the sample set, and $\mathbf{x}_k$ is the k^{th} observation. The resulting gaussian *pdf* that fits the data is then [12]

$$p(\mathbf{x}|\,\boldsymbol{\mu},\Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}(\det(\Sigma))^{\frac{1}{2}}} \times \exp\left(-\frac{1}{2}D^2\right) \tag{4}$$

where

$$D^2 = (\mathbf{x} - \boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x} - \boldsymbol{\mu})^t \tag{5}$$

is the square Mahalanobis distance and d is the dimensionality of the gaussian function ($d = 2$ in our particular case). Figure 2 shows a plot of the function estimated from the sample data set with the application of eq. (2), eq. (3) and eq. (4).

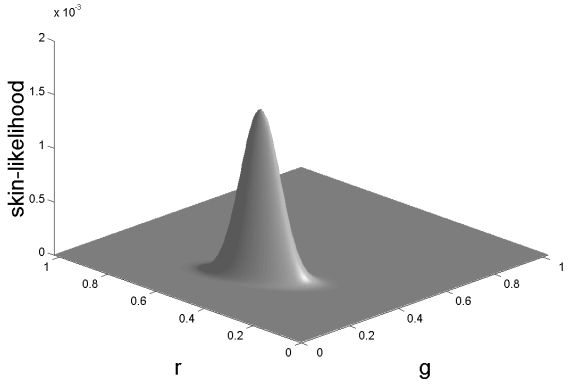

Figure 2: Estimated SGM for the skin colour sample set described in section 2.

4 Gaussian-Mixture Model (GMM)

As the training sample set histogram in fig. 1 shows, the skin distribution is clustered in a specific region of the colour space. We may ask, however, *how* well-clustered it is. In order to look for an answer to this question, we proposed in [1] a gaussian mixture approach to model the data. In that work, we analyse the behaviour of the single and double-gaussian models. Here, we improve on that previous work

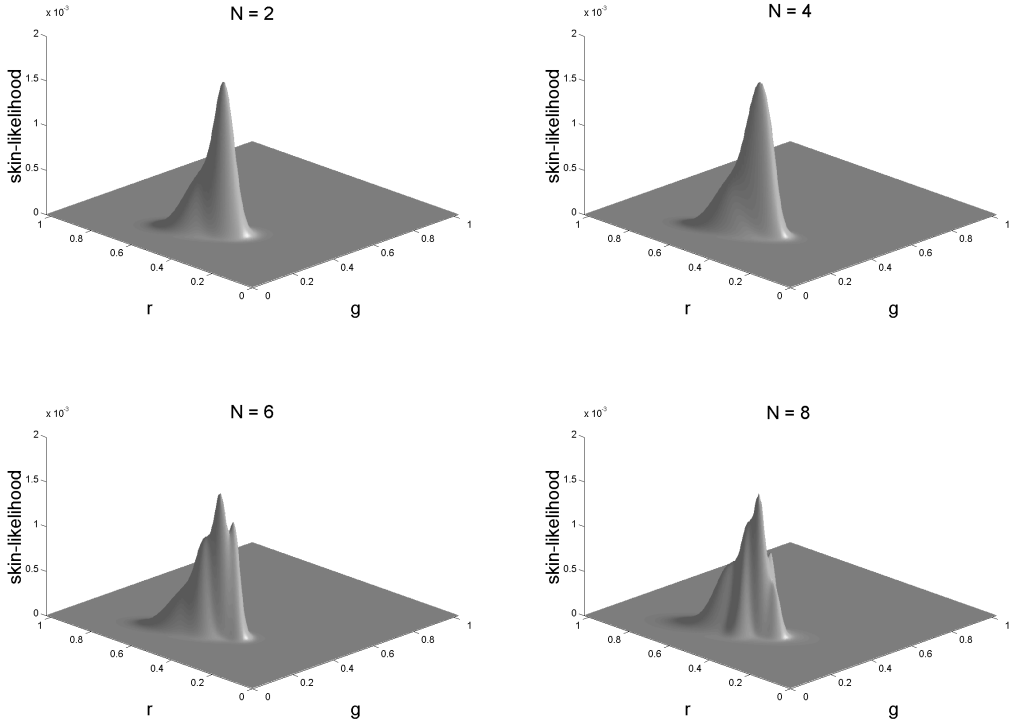

Figure 3: Estimated GMM's for the skin colour distribution described in section 2. From top to bottom and left to right, $N = 2, 4, 6$ and 8.

by generating several GMM's and by developing a strategy to compare their performances with respect to the SGM.

In this approach, we assume that the entire data set can be modelled by an N-gaussian mixture *pdf* given by [13]

$$p(\mathbf{x}|\Theta) = \sum_{l=1}^{N} p\left(\mathbf{x}|l, \Theta_l\right) P(l), \tag{6}$$

where N is the number of gaussians, $p\left(\mathbf{x}|l, \Theta\right)$ is the specific density function for gaussian l, $P\left(l\right)$ is the prior probability for gaussian l and Θ is the parameter vector containing the N mean vectors and the N covariance matrices. Following the maximum likelihood criterion [12], the goal here is to find the parameter vector Θ that maximizes the likelihood function. It is important to notice that, for a GMM, there is *no* analytic solution for the maximisation of the likelihood function [15]. It means that the optimal Θ must be estimated numerically.

The standard algorithm used to find maximum likelihood estimates for gaussian mixture models is the expectation maximisation (EM) procedure [14, 15]. The EM algorithm, when applied to this optimisation problem, results in a few simple updating rules for the parameters of the mixture. Executed as an iterative procedure, these rules guarantee that a local maximum of the likelihood function is reached and that the corresponding parameter vector Θ is found. The rules can be stated as [15]

$$P_l = \frac{1}{n} \sum_{i=1}^{n} p(l|\mathbf{x}_i,\ \boldsymbol{\mu}_l^{old}, \Sigma_l^{old}) \tag{7}$$

$$\boldsymbol{\mu}_l = \frac{\sum_{i=1}^{n} \mathbf{x}_i p(l|\mathbf{x}_i,\ \boldsymbol{\mu}_l^{old}, \Sigma_l^{old})}{\sum_{i=1}^{n} p(l|\mathbf{x}_i,\ \boldsymbol{\mu}_l^{old}, \Sigma_l^{old})} \tag{8}$$

$$\Sigma_l = \frac{\sum_{i=1}^{n} p(l|\mathbf{x}_i,\ \boldsymbol{\mu}_l^{old}, \Sigma_l^{old}) \left(\mathbf{x}_i - \boldsymbol{\mu}_l\right) \left(\mathbf{x}_i - \boldsymbol{\mu}_l\right)^t}{\sum_{i=1}^{n} p(l|\mathbf{x}_i,\ \boldsymbol{\mu}_l^{old}, \Sigma_l^{old})} \tag{9}$$

where n, $\boldsymbol{\mu}_l^{old}$, Σ_l^{old}, $\boldsymbol{\mu}_l$, Σ_l and P_l are, respectively, the sample size, the mean and covariance of the last iteration, the mean and covariance estimated for the present iteration and the prior probability estimated in the present iteration, all for a given gaussian l; $p(l|\mathbf{x}_i,\ \boldsymbol{\mu}_l^{old}, \Sigma_l^{old})$ is the probability that observation $\mathbf{x}_i$ belongs to gaussian l.

The above equations are used iteratively to update the parameters P_l, $\boldsymbol{\mu}_l^{old}$ and Σ_l^{old} for each gaussian l, until a

steady state or a maximum number of iterations is reached. Although the stopping criterion for the EM algorithm is a known problem [13, 15], for low dimensionalities (e.g. $d = 2$) convergence is usually fast. In our experiments we have not found problems concerning this issue.

An important issue is that, while there is only one maximum in the likelihood function for the SGM [12], the number of maxima in the likelihood function for a GMM seems to be unknown [13, 15]. Once the EM algorithm finds a *local* maximum, it is not guaranteed that, for a GMM, this is the best one. A possible technique to circumvent this difficulty is to initialise the algorithm multiple times with random initial parameters, measuring after each run the likelihood function. The chosen parameter vector Θ will be associated with the maximum value among all the maxima obtained. This exhaustive search increases the chances of finding the global maximum.

Figure 3 shows the estimated GMM's for $N = 2, 4, 6$ and 8. The GMM parameters for each N were obtained by selecting the result with highest likelihood among 1000 runs of the EM algorithm with random initialisation.

5 Experimental Setup

In this section we present the experimental setup for performance evaluation and comparison of the several GMM's with respect to each other and with respect to the SGM.

5.1 The Testing Sample Set

We have applied the eight models to a data set of 800 images containing equal amounts of people from each different ethnic group. In order to specify which pixels correspond to skin and which do not, we manually cropped the skin regions in those images. The resulting binary images constitute our *ground truth* – see an example in fig. 4.

5.2 The Skin Likelihood Image

The first step for measuring que performance of a given model with respect to the test set is to obtain the *skin likelihood images* (SLI) for all the images. The SLI is a greyscale representation of a given test image where the grey-level intensity of each pixel is proportional to the probability of this pixel belonging to skin, according to a given skin model m. The procedure for obtaining the skin likelihood images can be summarized as follows. Initially, for each image in the test set, the RGB vector for each pixel (x, y) in the image is obtained; then, the conversion from RGB into the rg representation is done. The resulting rg coordinates are then used to look up the specified model in order to obtain the respective probability. This procedure results in a greyscale image, where pixel brightness indicates how likely the pixel is to a skin one. Figure 5 shows an example of a test image and its respective SLI for the SGM and for the double-gaussian model (GMM with 2 gaussians).

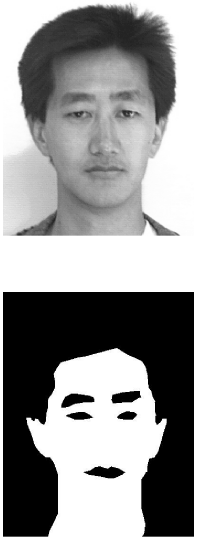

Figure 4: A sample image from the test set (top - original coloured) and its ground truth (bottom - binary).

5.3 Thresholding and ROC Curve Generation

Once the SLI for each image in the test set has been obtained, it is initiated the successive thresholding procedure. We start with a given small threshold, which is used to classify the SLI. All greyscale values that are above the threshold are considered skin; the others are considered non-skin. The outcome of this step is a binary classified image, whose pixels are assigned to one of the two classes (skin or non-skin). In the following, the segmented image is compared to the ground truth image. The true positive and false positive rates (TPR and FPR) of the classification process are then obtained by

$$TPR = \frac{TP}{S} \tag{10}$$

$$FPR = \frac{FP}{NS} \tag{11}$$

where TP is the number of true positives (pixels correctly assigned to the skin class), FP is the number of false positives (non-skin pixels wrongly assigned to the skin class), S is the total of skin pixels and NS the total of non-skin pixels. In our specific problem, as the test set is composed of 800 images, what is in fact done is to collect the *total* number of TP, FP, S and NS, covering the whole set of images. Just after that, eq. (10) and eq. (11) are computed and TPR and FPR obtained.

Figure 5: A sample test image (top - original coloured) and SLI's for the single-gaussian model (center) and for the double-gaussian model (bottom). The pixel intensity in each SLI is proportional to the probability of the given pixel to belong to skin, according to respective model.

As a result of these computations, we obtain, for each arbitrary threshold between the minimum and the maximum probability value of a given model m, a vector of measures $\mathbf{p} = (TPR, FPR)$ that expresses the performance of the model m in a specific operating point. For example, if two different models have the same FPR, the best will be the one that has the higher TPR. Conversely, if they have the same TPR, the best will be the one which has the lower FPR.

In this context, the well-known performance analysis technique based on Receiver Operating Characteristics (ROC) curves can be used. By applying K different thresholds, a set of K point vectors $\mathbf{p}_k = (TPR_k, FPR_k)$ can be obtained, which, when plotted, results in a ROC curve for the specific model with respect to the whole test set. Repeating the procedure for the M available models generates M ROC curves, which permits us to compare directly the relative performance of any two given models.

Figure 6 shows ROC curves for the single-gaussian model and the double-gaussian model applied to the image of fig. 5. It is possible to notice from the figure that the GMM with two gaussian components shows better performance for this specific image in practically all points of operation. For purposes of illustration, we have selected three of these points to analyse the relative behaviour of the two models.

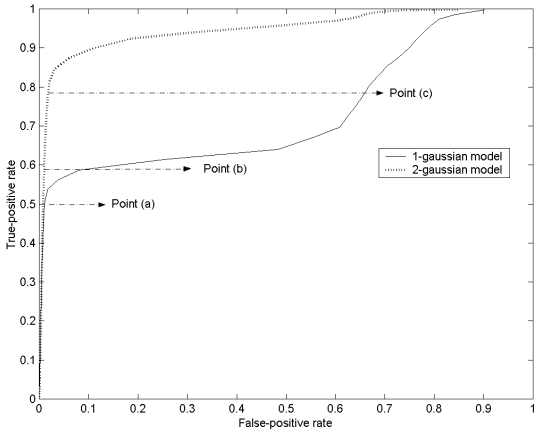

Figure 6: ROC curve for the SGM and the 2-gaussian models applied to the image in fig. 5.

The horizontal dashed arrows in fig. 6 indicate three major points chosen for illustrating the performance differences between the two models. In each point, the true positive rate is fixed and the corresponding difference in the false positive rate between the models is measured. It can be seen that there are no major dissimilarities at point (a): the difference between the false positive rate of the single-gaussian model (FPR_1) and that of the double-gaussian model (PFR_2) is minimal. At point (b), however, it is possible to see that FPR_1 is almost ten times FPR_2. A more extreme example takes place at point (c), where FPR_1 exceeds FPR_2 by more than 30 times.

Figure 7 shows a series of binary images obtained by thresholding at the different points shown in fig. 6.

6 Experiments and Results

The performance evaluation scheme described above was applied to the entire data set of 800 images containing people from a large spectrum of ethnic groups, such as Caucasian, African, Asian and Hispanic. Eight models were estimated from the data set, starting from the single-gaussian model until an 8-kernel gaussian mixture model. For each

Figure 7: Thresholded images for the single-gaussian model (top row) and the double-gaussian model (bottom row). First collumn: segmented images at point (a); second collumn: segmented images at point (b); third collumn: segmented images at point (c) of fig. 6. It can be seen that, despite the fact that true positive rates are the same in each collumn, the double-gaussian model has lower false positive rates.

model, a ROC curve was generated following the procedure outlined in section 5. Figure 8 presents a plot with the ROC curves for the 8 models.

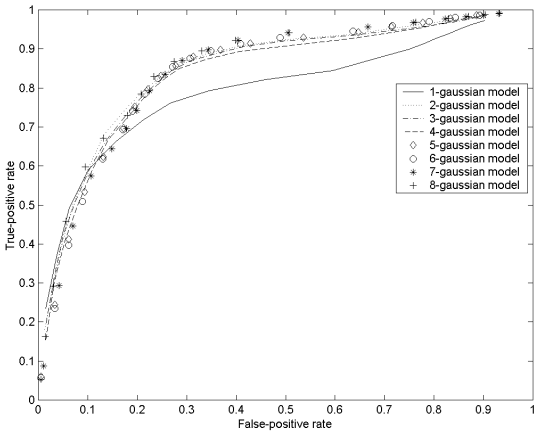

Figure 8: ROC curves for the SGM and for the several GMM's obtained by applying the models to the entire test set.

Some clear results are obtained from the plots. The most evident is that the SGM has poorer performance for medium to high true positive rates. Given that in the majority of practical applications the thresholds are set such that the true positive rate is kept at least in a medium range, it becomes clear that in most situations the GMM approach may outperform the SGM one. Nevertheless, it is important to notice that, in the low false positive region, all the models display a close behaviour.

Although it is clearly adequate to compare any GMM with the SGM, it is rather difficult to say which GMM is better. We believe that the differences among them are so subtle that it is inappropriate to judge their relative quality just by comparing the ROC curves. More exhaustive tests should be addressed in order to verify if there are indeed significant statistical differences among the different GMM's.

7 Conclusions

This work has presented a performance evaluation of single-gaussian and mixture of gaussians models for representing the human skin colour. An experimental setup was designed where a data set of skin pixels from several ethnic groups was used to train a single gaussian model and seven versions of gaussian mixture models. The eight models obtained were applied to a test set containing images different from those of the training set, but with people from the same four main ethnic groups. The problem was approached as a

classification task with two classes: “skin” and “non-skin”. The true skin regions for the whole data set (the ground truth) were manually selected. The performance of each model was measured in a process where true positive and false positive rates obtained for each classifier generated by a successive thresholding technique were combined to form Receiver Operating Characteristics curves. The analysis of those curves revealed two main conclusions. Firstly, GMM’s behave similarly over the whole range of the ROC curve. Secondly, although the performance of the SGM is similar to those of the GMM’s for low false positive rates, it is significantly decreased for high true positive rates. This conclusion suggest that skin color mixture models may be more appropriate than the single gaussian model when high correct detection rates are needed.

Acknowledgements

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for financial support.

References

[1] T.S. Caetano and D.A.C. Barone. *A probabilistic model for the human skin color*. IEEE International Conference on Image Analysis and Processing, ICIAP 2001, Palermo. IEEE Computer Society, Los Alamitos, (2001) 279–283.

[2] J. Yang and A. Waibel. *A real-time face tracker*. Proc. IEEE Workshop Applications of Computer Vision. Princeton, NJ, (1996) 42–147.

[3] J.G. Wang and E. Sung. *Frontal-view face detection and facial feature extraction using colour and morphological operators*. Pattern Recognition Letters **20** (1999) 1053–1068.

[4] Y. Wang and B. Yuan. *A novel approach for human face detection from colour images under complex background*. Pattern Recognition **34** (2001) 1983–1992.

[5] J. Cai and A. Goshtasby. *Detecting human faces in colour images*. Image and Vision Computing **18** (1999) 63–75.

[6] E. Saber and A.M. Tekalp. *Frontal-view face detection and facial feature extraction using colour, shape and symmetry based cost functions*. Pattern Recognition Letters **19** (1998) 669–680.

[7] K. Imagawa, S. Lu and S. Igi. *Colour-based hands tracking system for sign language recognition*. IEEE International Conference on Face and Gesture Recognition, Nara, Japan, 1998 462–467.

[8] J. Martin, V. Devin and J. Crowley. *Active hand tracking*. IEEE International Conference on Face and Gesture Recognition, Nara, Japan, 1998, 573–578.

[9] C. Wren, B. Clarkson and A. Pentland. *Understanding purposeful human motion*. IEEE International Conference on Face and Gesture Recognition, Grenoble, France, 2000, 378–383.

[10] E. Littman and H. Ritter. *Adaptive colour segmentation. A comparison of neural and statistical methods*. IEEE Transactions on Neural Networks **8** (1) (1997) 175–185.

[11] University of Stirling, Face Database: http://pics.psych.stir.ac.uk/cgi-bin/PICS/New/pics.cgi. November 2001.

[12] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, 1973.

[13] D. Titterington, A. Smith and U. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, Chichester (U.K.), 1985.

[14] A.P. Dempster, N.M. Laird and D.B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm*. Journal Royal Stat. Soc. **39** (1) (1977) 1–38.

[15] G. Mclachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, New York, 1997.