On the detection of low-resolution skin regions in surveillance images

Nils Janssen, Neil Robertson

On the detection of low-resolution skin regions in surveillance images

Nils Janssen and Neil Robertson
Joint Research Institute in Signal and Image Processing
Heriot-Watt University, Edinburgh, UK
{nj35,n.m.robertson}@hw.ac.uk

ABSTRACT
This paper presents a study into the detection of skin regions in images where faces may be low-resolution. We focus on surveillance footage and no assumptions are made about fine facial features being visible. This type of data presents the further challenge of changes in appearance of skin regions due to changes in both lighting and resolution. We investigate the use of common colour spaces (YIQ, YCbCr, HSV and RGB) and the effects of histogram size (number of bins) and dimensions (number of channels) by comparing the results to ground-truthed data. Error is measured per-pixel using the raw likelihood weights. We first use a non-parametric classification scheme based on a histogram similarity measure - the Battacharyya coefficient. Comprehensive results indicate that the YIQ colour space with 16 histogram bins gives the most accurate performance over a wide range of imaging conditions for non-parametric skin classification. We then compare this non-parametric method to established classification techniques. These are: Gaussian, Bayesian and Thresholding methods. We demonstrate better performance of the non-parametric approach vs. the other methods. We also show results of face-detection via a simple aspect-ratio.

1 Introduction

In this paper we address the problem of detecting skin-colour regions in video. This is quite simple in controlled situations (e.g. in the lab) and over short periods of time. It is a very challenging problem in outdoor environments where analysis may take place over days, perhaps weeks. Skin is primarily of interest because surveillance systems are interested in humans, in particular faces. Further if one considers that surveillance footage can often present very low resolution faces at varying degrees of occlusion, it is clear how challenging and interesting this problem is. We address the problem of skin-colour detection from that standpoint. In the data we analyse the resolution of faces ranges from roughly 15×20 to 30×30 pixels.

But why do we need another study on skin detection using colour? For two reasons. First, despite the obvious drawbacks involved in the use of colour in computer vision applications (e.g. lighting changes, camera gain etc.) it remains a powerful descriptor for image interpretation and a cue to which the human visual system is well-tuned and, consequently, cameras have been built to acquire. Since the data is available, colour persists to be used in many tracking, gesture and recognition applications [2, 17, 22]. However the limits of its use have not been investigated for the surveillance domain.

The second reason for this paper is that skin detection is directly related to face detection in challenging images such as surveillance footage. Frontal face detection, even in low resolution data has made considerable progress due to the Viola-Jones classifier [21] (notwithstanding the notable work by others e.g. [11, 16]). However, it appears that the Viola-Jones approach can fail at 35^o pan angles because the very features which are required to discriminate faces from other objects are obscured [14]. In particular, skin detection remains an important problem as this characterises humans quite distinctly. More importantly the face provides rich information about intention which is a vital cue for future intelligent surveillance systems. Predictions using gaze-direction information has only recently been addressed [19] but it is significant in terms of the interpretation of dynamic human motion.

Multi-dimensional histograms are an intuitive way to represent colour distributions. But this leads us to the question, *how much information is required?* Do we need 10 bins or 1000 bins or will parameterising be more efficient? If a higher resolution is chosen, what is the effect on the accuracy and speed of computation of metrics which rely on histogram matching? These are the questions motivating this paper.

Of course, there are many studies published on colour-spaces but these are not conclusive and there is some disagreeement in the literature [5, 10, 12]. This paper does not presume to be the definitive study on this subject but we do address the limits of skin detection for the low-resolution

Figure 1. Accurate skin detection in low-resolution data is difficult. When large changes in scale take place via zoom or motion, inaccuracies occur due to an extremely low number of true skin pixels available for classification. Face recognition often fails when fine features cannot be seen.

face which has not been done before. Although the data used in this paper is not *necessarily* non-frontal, the approach makes no assumptions about fine image features such as those used in HCI applications [13] or Viola-Jones [21] and this represents a contribution to the field of visual surveillance.

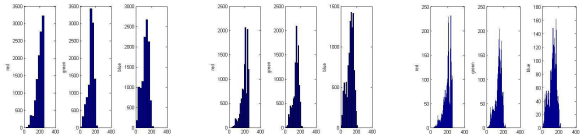
2 Contributions of this work

- For non-parametric skin classification over a range of lighting condistions and scales we compare the performance of the common colour-spaces, the effect number of histogram bins and computation times.

- We do this comparison to ascertain *per-pixel* accuracy as opposed to the *per-face* classification commonly found in the literature e.g. [20]. (Note that face detection is not the primary goal of this paper.)

- We compare a number of classification techniques: thresholding, parametric, Bayesian and non-parametric classification for the case of detecting skin in images where the faces are low-resolution.

Figure 2. A sample of our skin-colour data from the UCD Colour Database.



(a) 10 bins (b) 20 bins (c) 200 bins

Figure 3. Histograms of the RGB colour space illustrating the effect of choosing more (right) or fewer bins (left).

3 Colour spaces for non-parametric skin classification

Different papers comparing colour models for classification or skin-detection are presented but the results themselves give no definite solution to this problem. While several authors claimed that there is no optimal colour space for a certain classifier [4] other researchers have a different opinion. Albiol suggests that for each colour space an optimal skin classifier exists [1] i.e. there must be for each classifier at least one colour space which outperforms the others. Generally there is no optimal colour model to detect skin found in the literature. Generally, colour space transformation shall increase the distance between skin and non-skin pixels or bundle the skin colour characteristics.

To estimate the probability density function of skin colour in the colour spaces histograms are computed and normalized. The RGB and YCbCr colour spaces work with a vector for each pixel containing unsigned integers with 8 bit to specify the colour values. However the HSV colour space depicts the colour of a pixel with double values in the interval [0,1] while the colour channels in the YUV colour space are split into different vectors for each colour channel. Therefore different bin "edges" for the histograms must be used. The number of bins can be adjusted to compare the performance of the skin classifier to assess the dependency of skin classification on the number of bins. RGB histograms of the skin database are shown in Figure 3. Some examples from our skin database is shown in Figure 2. A

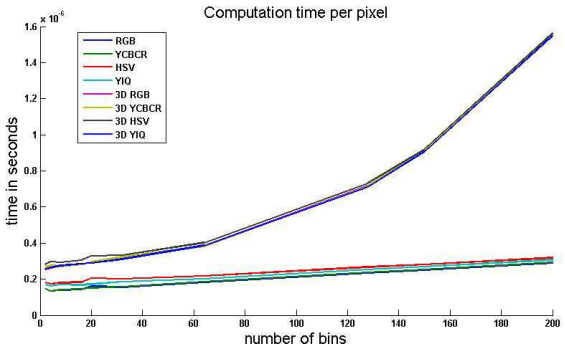


Figure 4. Accumulated computation time per pixel showing the time saving for 1-D vs. 3-D histograms.

comparison of separate histograms for each colour channel and combined 3 dimensional histogram is provided in this work. In particular we compare the YIQ, YCbCr, RGB and HSV colour spaces. We compare 1-D and 3-D histograms. In the YIQ and YCbCr colour spaces the histogram is only two dimensional because the Y component is neglected but in the further description they are also referred as "three dimensional" to distinguish between combined and separated histogram methods. 1-D histograms are obtained by multiplying the probabilities of each channel of the colour space being drawn from the skin histogram. 3-D histograms retain the independence of the channels. 1-D histograms have obvious computational time savings, as shown in Figure 3 whereas 3-D histograms assumes dependence of one channel on the other, which is more realistic.

The histograms show, on inspection, that the Cb, Cr, I, Q and H components of the respective colour spaces are highly discriminative for skin colour or colour itself but these are not shown here.

3.1 Skin probability calculation

The assumption that there exists a very small region of a colour space to represent skin has not been successful in the more complex imaging circumstances such as those encountered in low-resolution surveillance images. We use a selection of the whole UCD skin database drawn from 11045 examples to obtain a more accurate representation. Instead of simply picking out the value of each bin to calculate the probability of skin another assumption is made to make the classifier more robust. The Bhattacharyya coefficient was introduced as a similarity measure between two histograms i.e. between the histogram of the region in the new, unclassified, image and that of the training data. After linearizion a weighting factor for each pixel can be determined. This approach is typically used in the mean shift tracking algorithm [7] and essentially tells us how likely it is that each pixel is drawn from the skin database. This weighting factor is used to detect skin pixels as well and is shown in equation 7 below. We now describe the computation of the weighting factor.

A kernel is used because pixels close to the boundaries of the window are more likely to be part of the background or occluded. For the K classes of the models the probability estimations are computed with

$$\hat{q}_d = C \sum_{n=1}^{N_w} k\left(\|x_n^\star\|^2\right) \delta\left[b(x_n^\star) - d\right] \quad \text{for d=1} \cdots \mathbf{D} \quad (1)$$

and

$$\hat{p}_d = C \sum_{n=1}^{N_w} k\left(\|x_n\|^2\right) \delta\left[b(x_n) - d\right] \quad \text{for d=1} \cdots \mathbf{D} \quad (2)$$

where C is a normalization constant providing that the sum over p and q is one, x_n are the pixels of the candidate window, $x_n^\star$ are the pixels of the target and N is the number of all pixels in the window. Because in this work the window of target model and the candidate model is the same the same kernel function can be used for both models. With this model representation a comparison of candidate and target can be done. The mentioned Bhattacharyya coefficient

$$\rho\left[\hat{p}(y_t), \hat{q}\right] = \sum_{k=1}^{K} \sqrt{\hat{p}_k(y_t) \cdot \hat{q}_k} \quad (3)$$

is linearized round the starting point of the tracking y_0, hence

$$\rho\left[\hat{p}(y_t), \hat{q}\right] \approx \frac{1}{2} \sum_{k=1}^{K} \sqrt{\hat{p}_k(y_0)\, \hat{q}_k} + \frac{1}{2} \sum_{k=1}^{K} \hat{p}_k(y_t) \sqrt{\frac{\hat{q}_k}{\hat{p}_k(y_0)}} \quad (4)$$

where y_0 is the starting point. With the candidate model the equation can be written as

$$\rho\left[\hat{p}(y_t), \hat{q}\right] \approx \frac{1}{2} \sum_{k=1}^{K} \sqrt{\hat{p}_k(y_0)\, \hat{q}_k} + \frac{C_h}{2} \sum_{n=1}^{N_h} \omega_n k\left(\left\| \frac{y_t - x_n}{h} \right\|^2\right) \quad (5)$$

with

$$\omega_n = \sum_{k=1}^{K} \sqrt{\frac{\hat{q}_k}{\hat{p}_k(y_0)}} \delta\left[b(x_n) - d\right] \quad (6)$$

The second summand is the weighted density estimation. The first summand is a constant and the maximization of the Bhattacharyya coefficient is just dependent on the second summand. As the most similar object in the proximity of the start location is searched, the maximum of the weighted density is searched.

In this paper the target histogram is the skin pdf and the candidate histogram is the input picture itself. For each pixel this weighting factor represents the probability for skin after normalization. If an input picture has many pixels in a colour which is actual assigned with a high skin probability the division is leading to a reduction in this probability. This can reduce the misclassification if skin coloured objects appear. The classifier calculates element wise the ratio of the database and input histogram.

$$w_k = \sqrt{\frac{q_k}{p_k}} \tag{7}$$

where q is the colour histogram of the database and p the histogram of the input picture or image window. To compute the probability for the separated histograms each input pixel is assigned to one bin for each colour channel. The assigned entries of the weighting factors w_u are multiplied afterwards to obtain the skin probability. For the 3-D histograms there is only one weight for each colour bin which represents the probability. Using this way of computation the values in the resulting probability map are not distributed between 0 and 1 and must be normalized to represent a true probability.

Histogram equalization is used to normalize the computed map. Histogram equalization was developed to enhance the contrast of an image if the sensor provides just data in a part of the possible interval. With this normalization the values are then distributed over the interval [0,1] i.e.

$$s_k = \sum_{j=0}^{k} \frac{n_j}{n} \tag{8}$$

where s is the new gray level, k a particular gray level, n_k the number of times the kth gray level appears and n is the number of all pixels. With this method the interval $[0, 1]$ is not used optimally. This depends on the high amount of zeros in the calculated probability picture. The lowest value a pixel in the normalized picture can have is the probability of the first gray level of the original picture. Because of the high amount of zeros in the probability map the lowest assigned value in the normalized map is to high. To distribute the gray scale over the whole interval we apply

$$v' = \frac{v - min}{max - min} \cdot (max_{norm} - min_{norm}) + min_{norm} \tag{9}$$

hence if, $max_{norm} = 1$ and $min_{norm} = 0$, then

$$v' = \frac{v}{max - min} \tag{10}$$

In each of the calculated probability maps a 0 will probably appear which means that the normalization is simply a division through the largest value of the map. With this probability map a comparison of number of bins and used colour spaces is performed.

(a) Test image (b) Skin pixels

Figure 5. Hand-labelled ground truth skin pixels. Note the very low-resolution faces in the background.

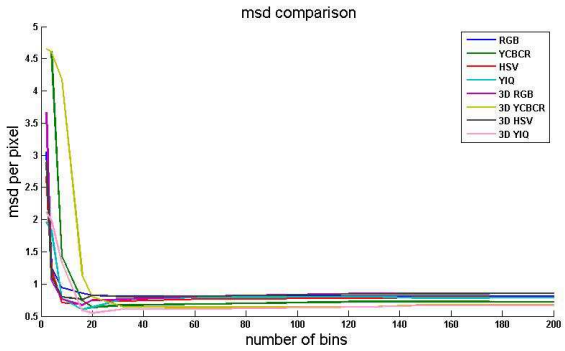


Figure 6. Computation results of the msd and computation time for a varied bin number of the histograms

3.2 Performance of the colour spaces with varying numbers of histogram bins

We used 37 new test images drawn from challenging surveillance-style scenarios. These images contain significant changes in lighting and resolution which causes the appearance to alter. Moreover there are varying degrees of head pan meaning classifiers based on face features (eyes, nose etc.) would fail. Ground truth skin pixels are obtained by hand. An example is shown in Figure 5. We compute the mean squared difference of the probabilities of skin in the test image using the approach described above from the ground truth. The results are shown in Figure 6 for all of

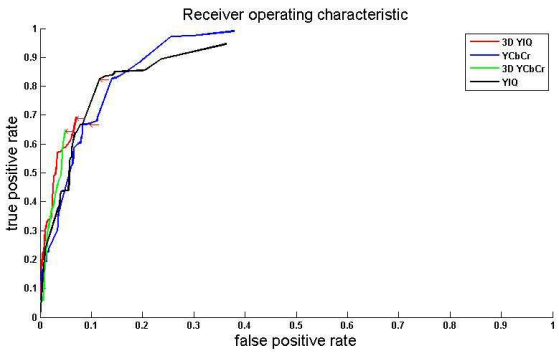


Figure 7. Receiver operating characteristic curve of the 4 best methods with a varying threshold to compute the binary image. Red arrows indicate the chosen threshold at 0.02.

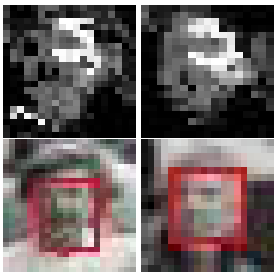


Figure 8. Skin pixel probability weights from the low-resolution faces in Figure 5 (row 1). Segmented faces using size constraint (row 2).

the 1-D and 3-D histograms with the number of bins ranging from 1 to 200. For each of the test pictures the 3D algorithms took the highest computation time which is a big drawback for a real-time application. Each pixel of the input picture has to be assigned in this histogram to obtain the weighting factors.

Due to the large number of results over this range, the four optimal results are taken to find the best working threshold for all methods using a receiver operating characteristic (ROC) curve which depicts the true positive rate over the false positive rate for different thresholds. Optimal values are in this kind of diagram in the upper left corner, the optimal case is the corner itself. The tested methods to generate the ROC curve are: 3-D YIQ (16 bins); 1-D YCbCr (20 bins); 3D RGB with 16 bins; YIQ (16 bins). The results are shown in Figure 7. The sum of mean square differences per pixel over the test dataset is shown in Table 1. For low-resolution faces a result is shown in Figure 8 .

The use of chroma channels outperforms the use of all

	3D YIQ	YCbCr	3D RGB	YIQ
sum msd	0.59	0.64	0.67	0.60

Table 1. Sum of mean square differences per pixel for the test database

channels. An increase of the histogram bin number over 40 respectively 128 gives no better performance for skin classification. On the other hand for the introduced classifier there is an optimum for a certain number of histogram bins which is in the region 10 to 20 (see Figure 6). We show the result of the optimum colour-space and histogram resolution applied to the text image of Figure

3.3 Choosing a threshold for skin vs. non-skin

To detect the areas on the picture which belong to skin and to separate the actual face region a binary map is computed representing skin (1) and non-skin (0) regions. Therefore an appropriate threshold for the continuous values of the probability map has to be chosen. It seems to be essential to find an adaptive threshold value which varies with the input picture. On the other hand the algorithm must not be forced to detected a face on every picture. Possibilities for adaptive thresholds are the mean, the median or Otsu's method [15].

To investigate the threshold value the probability maps were transformed to binary images by 1000 thresholds in the interval [0,1]. The results are compared with the ground truth afterwards. For comparison the ROC curve is depicted in Figure 7. The ROC curve depicts the over all pixels of the test database accumulated false positive rate and true positive rate for different thresholds. It is useful to choose a threshold with a very low false positive rate to discriminate between different objects in the further investigations.

As we saw before from the computed graphs the YIQ colour space with three independent one dimensional histograms is the most accurate over the test database. The best threshold empirically evaluated over the entire test database is a constant threshold of 0.02. The points in the ROC curve shown in Figure 7 for the constant threshold $T = 0.02$ are marked with red arrows (hence the seemingly incomplete ROC curve). A static threshold is a good solution for this classification method if the adaption by the weighting factor is taken into account. If less skin coloured pixels appear in an input figure the probability will increase. Compared to other classifiers the adaption to the input is therefore simply another step of the process.

Using a simple aspect-ratio filter to discriminiate between faces and other skin regions we apply the non-parametric approach to face detection. Illustrative results

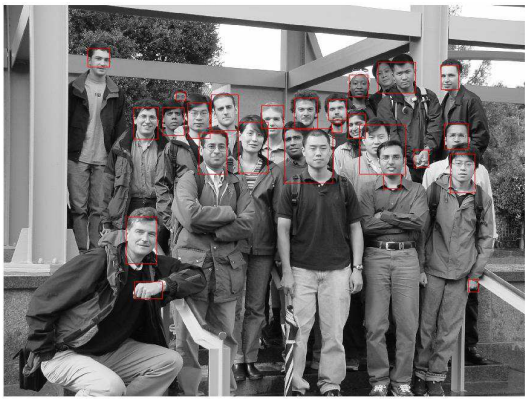


Figure 9. Results of the face detection based on the non-parametric classification of skin. Note a hand is detected due to the orientation and aspect ratio constraint placed on skin blobs.

are shown in figures 8 and 9.

4 Comparison with other classification techniques

Many alternative skin classifiers are available. Here we compare the performance of the non-parametric technique described above with a simple colour-thresholding method, a Gaussian and, finally, a Bayesian classifier.

4.1 Colour thresholding classifier

To provide a comparison of the non-parametric skin classification method another very simple classifier is explained here. Each input pixel is assigned as 1 or 0 depending on its position in the colour space. Just the pixels within a special two dimensional interval are assigned to one, all others are 0. To get a decisive interval this method has to work with a colour space where the skin colour is bundled to obtain sharp thresholds. Because of the expected shape of the histograms it is reasonable to use the YCbCr colour space for the thresholding method. For the implementation just the chroma channels Cb and Cr are used to remain independent of changes in lighting. The threshold intervals are [100:125] for the Cb channel and [144:166]. These are automatically chosen by evaluating the YCbCr space of the skin exmple database.

This classification method can be expressed as the assumption of a unity distribution within the rectangle determined by its boundaries. Every pixel within this rectangle is detected as skin.

4.2 Gaussian classifier

The Gaussian classifier used in Chang and Robles work [6] uses the YCbCr colour space for face detection. Under the assumption that human skin colour does not show big variations in the chromatic components even for different skin types Chang and Robles used the chroma channels Cb and Cr to compute the probability for detected skin, the luminance channel Y is neglected as before. A Gaussian model is used to describe the features of skin on the colour space. From the skin database the mean values of the colour channels Cb and Cr and the covariance matrix of the values are computed to obtain a two dimensional gaussian model of skin colour. The probability of skin for each input pixel is then computed with the formula

$$P(r,b) = exp\left[-0.5\left(\begin{bmatrix} b \\ r \end{bmatrix} - \begin{bmatrix} m_b \\ m_r \end{bmatrix}\right)^T C^{-1}\left(\begin{bmatrix} b \\ r \end{bmatrix} - \begin{bmatrix} m_b \\ m_r \end{bmatrix}\right)\right] \tag{11}$$

where b and r are the input colour channels Cb or Cr of the considered pixel, C is the covariance matrix of the two colour channels and m_b and m_r are the means of the colour channels. The calculation of the probability in this way delivers already a value between 0 and 1 which can directly be shown as a gray scale picture. Adaptive thresholding gives the binary picture. To distinguish between faces and other skin parts all skin coloured areas are evaluated by their height to width ratio which is assumed to be between 0.8 and 1.6. Furthermore the cross correlation to a template face is computed and must be higher than 0.6 to acknowledge a face. Another condition for recognized faces is that a least one hole occurs in a skin coloured area. The results for the true positive and false negative per-pixel classification rate are shown in table 2. Note that the authors claim 76% classification rate for face detection in a test set of 30 images. Errors or misclassifcations occured due to the number of holes or skin parts which did not belong to the face which is not an assumption we make in low-resolution data. The result computed from our test database compared to the ground truth is shown in Table 2.

Both classifiers achieve significantly poorer results than the non-parametric classifier for a chosen threshold of 0.2, as shown in Table 2 The true positive rate implies that the algorithm is working but could be explained due to optimization of the algorithm for this particular test database, too. Although the Gaussian classifier was much slower than the histogram classifier, the thresholding classifier has advantage in processing time.

Classifier	true positive rate	false positive rate
Gaussian	0.352	0.058
Colour Threshold	0.408	0.048
Proposed	0.685	0.064

Table 2. Results of Gaussian, Threshold and introduced classifiers computed from the test dataset.

4.3 Bayesian classifier

Bayesian methods for detecting faces in images have been presented [4]. This approach clearly requires expanding our exemplar set by including *non-skin* examples. An edge detector is used to discriminate joint skin coloured areas in pictures using *a priori* knowledge of the searched parts of faces. Phung *et al.* investigated the performance of the Bayesian classifier on different colour spaces and compared the result with alternative classifiers [4]. This method is claimed to detect different skin colours and demonstrate robustness to varying lighting conditions. However it should be noted that Phung *et al.* focused their research on daylight and indoor lighting not the range of lighting conditions expected in the surveillance domain that we have considered here.

The Bayesian decision rule for minimum cost is used to discriminate skin and non skin pixels by colour. First the probability density functions of skin and non skin pixels are estimated with histograms. Then the decision rule for the classifier is given by the following:

$$\frac{p(c \mid skin)}{p(c \mid nonskin)} \geq \theta \tag{12}$$

where $p(c \mid skin)$ and $p(c \mid nonskin)$ are the class conditional probability density functions obtained from the histograms and θ is the decision threshold. We follow Phung and used a low value for $\theta = 0.8$ which was derived experimentally in their work.

After the skin colour detection, edge detection is applied to alleviate the effect of false detected skin pixels. Sobel and Canny edge detectors are then used to find edges and separate the skin areas. The Sobel detector generates a gradient picture of the input picture and works especially well with inhomogeneous skin-coloured blocks. This is important to discriminate neighboring skin coloured areas. However the Canny detector uses in its implementation the Sobel operator to detect gradients and provides post-processing to detect just sharp edges. The breadth of this edge is only one pixel. In this way the different coloured regions are found. Because of the knowledge that skin has a high homogeneity in its colour all areas with a high standard deviation σ are removed for further consideration. Furthermore all areas are removed which deviate to much from the average size of the regions. The Bayesian classifer with a low decision threshold detected all skin pixels but also added background pixel with skin colour to the first binary map representing skin. Due to the edge and post processing these false detected pixels could be removed or at least discriminated to true skin parts. The result is shown in Figure 10 which gives a comparison with the other methods.

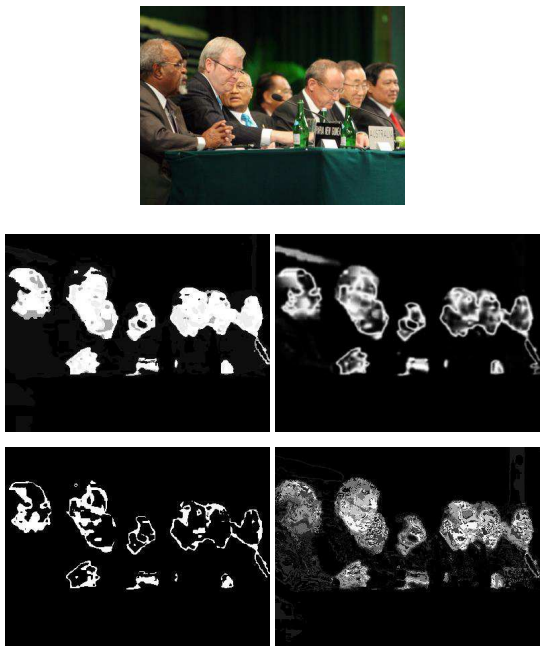


Figure 10. Skin detection results from other classifiers compared to the non-parametric approach. (First row) input image. (Black and white set, clockwise from top left) the non-parametric approach; Gaussian classification; Colour thresholding and Bayesian. White pixels equal higher likelihood of skin.

5 Conclusion and further work

We have demonstrated that colour classification can be used in surveillance imagery where no assumptions can made about the pose of the images person's head or the resolution of the face image. We have contributed a study on the effectiveness of various colour-spaces and the resolution of histograms in a non-parametric scheme. Finally we have shown that, compared to other fast and simple classification methods, the non-parametric scheme using the YIQ colour-space with 16 bins gives favourable performance. For face detection, thresholding and aspect-ratio can be used to give reasonable results. This approach could be used to boostrap

tracking and recognition algorithms.

For further work in this area research into face detection in low-resolution data needs to be done given the importance of person recognition for intelligent surveillance. The range of lighting conditions should be fully considered in the experimental validation, especially in outdoor situations. This requires a much larger training dataset than that used in this paper. Further the colour normalisation used here may need to be expanded to model intrinsic colour, illumination and reflectance.

References

[1] A. Albiol, L. Torres and E.J. Delp, Optimum colour spaces for skin detection, Int. Conf. Image Processing, 2001.

[2] L. Bretzner, I. Laptev and T. Lindeberg. Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. In Proc. Int. Conf. on Automatic Face and Gesture Recognition, Washington D.C, 2002

[3] EC-funded CAVIAR project/IST 2001 37540, http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/.

[4] S.L. Phung, A. Bouzerdoum, and D. Chai, Skin Segmentation Using colour And Edge Information, Proc. Int. Symposium on Signal Processing and its Applications, July 2003.

[5] D. Chai and K.N. Ngan, Locating facial region of a head-and-shoulders colour image Third IEEE International Conference on Automatic Face and Gesture Recognitions, Nara, Japan, pp. 124-129, April 1998.

[6] H. Chang and U. Robles, Face Detection, http://www-cs-students.stanford.edu/ robles/ee368/main.html

[7] D. Comaniciu, V. Ramesh and P. Meer, Real-Time tracking of Non-Rigid Objects using Mean Shift, IEEE Conf. Computer Vision and Pattern Recognition, 2000.

[8] P. Kakumanu, S. Makrogiannis and N. Bourbakis, A survey of skin-colour modeling and detection methods, Pattern Recognition, 40(3), pp 1106-1122, 2007.

[9] P. Kakumanu, S. Makrogiannis, N. Bourbakis, R. Bryll and S. Panchanathan, Image Chromatic Adaption using ANNs for Skin colour Adaption, Proc. IEEE Int. Conf. Tools with Artificial Intelligence, 2004.

[10] K. Hidai et al., Robust Face Detection against Brightness Fluctuation and Size Variation International Conference on Intelligent Robots and Systems, vol. 2 pp. 1379-1384, Japan, October 2000

[11] R. Hsu, M. Abdel-Mottaleb, and A.K. Jain, Face detection in color images, IEEE Trans. PAMI, vol. 24, no. 5, pp. 696-706, 2002.

[12] T.S. Jebara and A. Pentland, Parametrized Structure from Motion for 3D Adaptive Feedback Tracking of Faces Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, pp. 144-150

[13] Y. Matsumoto and A. Zelinsky, An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement Proceedings of IEEE Fourth International Conference on Face and Gesture Recognition, pp. 499-505, 2000.

[14] Mita et al, Discriminative Feature Co-Occurrence Selection for Object Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), July 2008.

[15] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Sys. Man. Cyber 1979, 9: 62-66.

[16] Y. Pang, Y. Yuan, and X. Li, Gabor-based region covariance matrices for face recognition, IEEE Transactions on Circuit Systems for Video Technology, vol. 18, no. 7, 2008.

[17] P. Prez, C. Hue, J. Vermaak and M. Gangnet, Colour-based probabilistic tracking. In Eur. Conf. on Computer Vision, 2002, LNCS 2350, Pages 661-675.

[18] P.W. Power and J.A. Schoonees, Understanding Background Mixture Models for Foreground Segmentation, Proc. Image and Vision Computing, 2002, pp. 267-271.

[19] N.M. Robertson and I.D. Reid, Estimating Gaze Direction from Low-Resolution Faces in Video Lecture Notes in Computer Science, 3952/2006, pp 402-415, Springer, ISBN 978-3-540-33834-5

[20] H. Schneiderman, Feature-Centric Evaluation for Efficient Cascaded Object Detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, June, 2004.

[21] P. Viola and M. Jones, Robust real-time face detection, Int. Journ. Computer Vision 57(2) 137-154, 2004.

[22] C. Yang, R. Duraiswami and L. Davis, Efficient Mean-Shift Tracking via a New Similarity Measure, Proc. Conf. Computer Vision and Pattern Recognition, 2005.