Pixel based skin colour classification exploiting explicit skin cluster definition methods

F. Gasparini, S. Corchs and R. Schettini
DISCo, Università degli Studi di Milano-Bicocca
Via Bicocca degli Arcimboldi 8, 20126 Milano, (ITALY)
Corresponding author: R. Schettini (schettini@disco.unimib.it)

ABSTRACT

In this paper we examine the performance, on a large and heterogeneous image database, of various skin detectors based on explicit colour skin cluster definition, coupled with a cast remover to see whether, and to what degree, the effectiveness of classification is improved, regardless of the strategy adopted. We also evaluate the hypothesis that a combination of some of the skin detection algorithms studied could ensure a more accurate classification than any of the algorithms provides individually. Different combination rules have been investigated. All the experiments have been performed on the Compaq skin database. The results are evaluated in terms of both recall (the ratio between the number of skin pixels correctly classified and the total number of actual skin pixels), and precision (the ratio between the number of skin pixels correctly classified and the total number of pixels labelled as skin pixels by the detection method employed).

1. INTRODUCTION

The detection of skin regions in colour images is a preliminary step in several applications, such as image and video classification and retrieval in multimedia databases, semantic filtering of web contents (through the definition of medium-level features), human motion detection, human computer interaction and video-surveillance. It would also be useful in image processing algorithms, as well as in intelligent scanners, digital cameras, photocopiers, and printers. Many different methods for discriminating between skin pixels and non-skin pixels are available in the literature. These can be grouped in three types of skin modelling: parametric, nonparametric and explicit skin cluster definition methods. The parametric Gaussian models[1] assume that skin colour distribution can be modelled by an elliptical Gaussian joint probability density function. Nonparametric skin modelling methods estimate skin colour distribution from the training data without deriving an explicit model of skin colour: an example is the histogram-based[2] nonparametric skin model. The simplest and often applied method is to build what is called an "explicit skin cluster" classifier which expressly defines the boundaries of the skin cluster in certain colour spaces[3-9]. The underlying hypothesis of methods based on explicit skin clustering is that skin pixels exhibit similar colour coordinates in a properly chosen colour space. The main difficulty in achieving high skin recognition rates, and producing the smallest possible number of false positive pixels, is that of defining accurate cluster boundaries through simple, often heuristically chosen, decision rules.

We analyse here the performance of different explicit skin cluster methods available in the literature. Skin detection, just like any other colour-based feature computation, is not completely reliable when acquisition conditions and imaging devices are not known a priori, or are not carefully controlled. Images being processed may also pass through a number of stages between entry and skin classification and these may introduce significant colour distortion. While in many cases the human observer will still be able to recognize the skin colours in the scene, we can only guess to what extent the existing algorithms can perform the same task. The problem is particularly acute when the images to be processed are collected from many different sources (the web, for example). Therefore, we examine here the performance of a selection of skin cluster strategies when coupled with a cast remover[10] to see whether, and to what degree, the effectiveness of classification is improved, regardless of the strategy adopted. Moreover, we evaluate the hypothesis that some combined form of these methods, could produce a more accurate classification than any of the algorithms provides individually.

All the experiments have been performed on the Compaq skin database[11], and the results are evaluated in terms of both recall (the ratio between the number of skin pixels correctly classified and the total number of actual skin pixels), and precision (the ratio between the number of skin pixels correctly classified and the total number of pixels labelled as skin pixels by the detection method employed).

2. METHOD

2.1 Methods of skin detection

The methods considered in this paper separate skin and non skin colours using a piecewise linear decision boundary. These explicit skin cluster methods propose a set of fixed skin thresholds in a given colour space. Some colour spaces permit searching skin colour pixels in the 2D chromatic space, reducing lighting variation dependence, others, such as the RGB space, address the lighting problem by introducing different rules depending on the illumination condition (uniform daylight or flashlight). Working within different colour spaces, we have implemented the seven different algorithms analysed in this paper. We refer to these methods using the name of the colour space adopted: RGB[3], YCbCr1[4], YCbCr2[5], HSI[6], HSV1[7], HSV2[8] and rgb[9].

2.2 Method of cast removal

Traditional methods of cast removal do not discriminate between images with true cast and those with predominant colours, but are applied in the same way to all images. This may result in an undesirable distortion of the chromatic content with respect to the original scene. To avoid this problem we have used a reliable and rapid method for classifying and removing a colour cast in a digital image without any a priori knowledge of its semantic content[10]. A multi-step algorithm classifies the input images as i) no-cast images; ii) evident cast images; iii) ambiguous cast images (images with feeble cast, or for which whether or not the cast exists is a subjective opinion), iv) images with a predominant colour that must be preserved, v) unclassifiable images. Classification makes it possible to discriminate between images requiring colour correction and those in which the chromaticity must, instead, be preserved. If an evident or ambiguous cast is found, a cast remover step, which is a modified version of the white balance algorithm, is then applied. The whole analysis is performed by simple image statistics on the thumbnail image. Since the colour correction is calibrated on the type of the cast, a wrong choice of the region to be whitened is less likely, and even ambiguous images can be processed without colour distortion.

2.3 Combining algorithms

The combination of classifiers has been widely proposed as a method to improve the efficiency and accuracy achieved by a single classifier[12]. Some combinations, such as those based on Bayes' theory, assume that the classifiers are statistically independent. Voting methods are another type of combination, they do not require independence and simply count the number of classifiers that agree in their decision and decide the class to which the input pattern belongs accordingly. Among the possible variations of this idea, we have considered here the unanimity and majority vote rules. We have also proposed a third combination rule, which evaluates the skin map as the difference between the skin map produced by the algorithm with the greatest recall among the seven methods considered (here HSI) and a non-skin map, obtained as the intersection of all the non-skin maps (i.e the map complementary to the skin map) obtained by the remaining six algorithms.

3. RESULTS AND CONCLUSIONS

We have evaluated the performance of each of the seven explicit skin cluster methods before and after cast removal and the performance of the combination methods proposed versus that of each of the single skin detector algorithms.

In order to quantify these performances, true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) have been computed for 2000 images taken from the Compaq skin database[11]. All these quantities refer to absolute values: for example TP represents the total number of skin pixels identified by the skin detector. The 2000 images correspond to 21.983.935 skin pixels and 145.275.330 non-skin pixels.

The performance of the algorithms before and after the application of the cast remover is compared in Table 1 using the following two measures: 1) recall, TP/(TP+FN), the ratio between the number of skin pixels correctly classified and the total number of actual skin pixels and 2) precision, TP/(TP+FP), the ratio between the number of skin pixels correctly classified and the total number of pixels labelled as skin pixels by the considered skin detection method. The corresponding values for the combined algorithms are shown in Table 2.

Table 1: Comparison of the performance of the single skin detectors before and after the application of the cast remover.

Classifier	**Recall (%)** Before cast removal	**Recall (%)** After cast removal	**Precision (%)** Before cast removal	**Precision (%)** After cast removal
RGB	89	88	37	41
HSV2	74	75	42	44
HSI	93	92	35	38
HSV1	46	45	54	58
YCbCr1	91	91	30	32
YCbCr2	76	77	32	34
rgb	42	40	36	38

Table 2: Comparison of the performance of the combined algorithms before and after the application of the cast remover.

Classifier	**Recall (%)** Before cast removal	**Recall (%)** After cast removal	**Precision (%)** Before cast removal	**Precision (%)** After cast removal
Majority	81	81	42	44
Unanimity	30	30	57	60
Our proposal	93	92	37	39

The results of these experiments are plotted in Figure 1, in terms of recall versus precision. On the left is shown the comparison of the seven methods before (empty symbols) and after (full symbols) cast removal, and on the right the results, before cast removal only, of the combining rules with respect to the single methods.

From the analysis of Table 1 and Table 2 and Figure 1, we can argue that:

- The performance of the methods studied indicates that different methods may be either more precision- or more recall-oriented. The qualification of "best method" is therefore application-dependent.
- With respect to the application of the cast remover:
  1. The performance of the seven methods considered always improves in terms of precision after application of the cast remover.
  2. In some cases, recall also increases, while in others it is almost the same, or not significantly less.
- With respect to the combining rules:
  1. The unanimity rule is precision-oriented, but it significantly reduces the true positive.
  2. The majority vote rule produces a high value of recall (81%) with a significantly high value of precision (42%), showing a good compromise between correct and incorrect classifications.
  3. The method we have proposed is recall-oriented; it produces high values of true positives, with a slight increase in precision. Preserving the true positives was in fact our objective in choosing the method with the highest recall to generate the skin map to be corrected by the non-skin map.

  As with the single methods, the most convenient combination rule can be chosen, according to the particular application needs (recall vs precision).

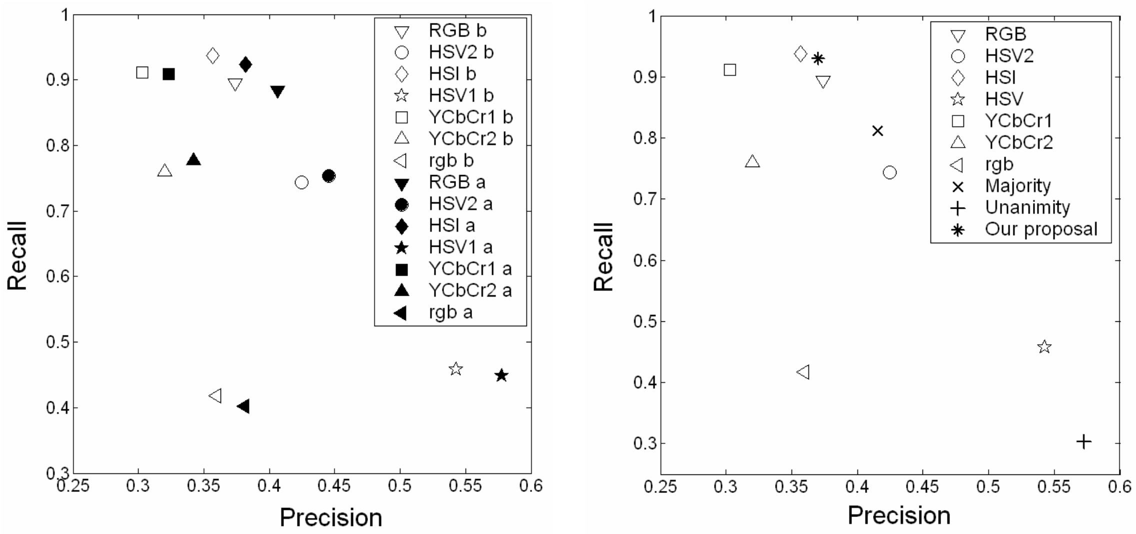

Figure 1: Results of the experiments in terms of recall versus precision. On the left, the comparison of the seven methods before (b, empty symbols) and after (a, full symbols) cast removal. On the right, the results of the combining rules with respect to the single methods, before cast removal.

References

1. M.-H Yang and N. Ahuja, “Gaussian Mixture Model for Human Skin Colour and its Applications in Image and Video Databases”, SPIE/EI&T Storage and Retrieval for Image and Video Databases (San Jose, January 1999), pp. 458-466.
2. M. Jones, J. Rehg, “Statistical Color Models with Application to Skin Detection”, IEEE Conference on Computer Vision and Pattern Recognition, CVPR '99 (1999), pp. 274-280.
3. J. Kovac, P. Peer and F. Solina, “2D versus 3D colour space face detection”, 4th EURASIP Conference on Video/Image Processing and Multimedia Communications (Croatia, 2003), pp. 449-454.
4. D. Chai and K. N. Ngan, "Face segmentation using skin colour map in videophone applications", IEEE Transactions on Circuits and Systems for Video Technology 9, 551-564 (1999).
5. R. Hsu, M. Abdel-Mottaleb and A. K. Jain, “Face detection in colour images”, IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 696-706 (2002).
6. I-S. Hsieh, K-C. Fan, and C. Lin, “A statistic approach to the detection of human faces in colour nature scene”, Pattern Recognition 35, 1583-1596 (2002).
7. S. Tsekeridou and I. Pitas, “Facial feature extraction in frontal views using biometric analogies", Proc. of the IX European Signal Processing Conference, vol. I, 315-318 (1998).
8. C.Garcia and G. Tziritas, “Face detection using quantized skin colour regions merging and wavelet packet analysis”, IEEE Transaction on Multimedia 1, 264-277 (1999).
9. G. Gomez and E. F. Morales, “Automatic feature construction and a simple rule induction algorithm for skin detection, Proc. Of the ICML workshop on Machine Learning in Computer Vision, A. Sowmya, T. Zrimec (eds), 31-38 (2002).
10. F. Gasparini and R. Schettini, “Colour Balancing of Digital Photos Using Simple Image Statistics”, Pattern Recognition 37, 1201-1217 (2004).
11. Compaq Cambridge Research Lab image-database. M. Jones and J. Rehg, “Statistical Colour Models with Application to Skin Colour Detection”. Compaq Cambridge Research Lab Technical Report CRL 98/11 (1998).
12. J. Kittler, M. Hatef, R.P.W. Duin and J.Matas, “On Combining Classifiers”, IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 226-239 (1998).