

# COMPARING COLOR AND TEXTURE-BASED ALGORITHMS FOR HUMAN SKIN DETECTION

A. Conci, E. Nunes

*Instituto de Computação, Universidade Federal Fluminense, Niterói, Brazil*  
aconci@ic.uff.br

J.J. Pantrigo, A. Sanchez

*Departamento de Ciencias de la Computación, Universidad Rey Juan Carlos, 28933 Madrid, Spain*  
juanjose.pantrigo@urjc.es, angel.sanchez@urjc.es

**Keywords:** Skin detection, segmentation, pixel classification, color spaces, texture.

**Abstract:** Locating skin pixels in images or video sequences where people appear has many applications, specially those related to Human-Computer Interaction. Most work on skin detection is based on modelling the skin on different color spaces. This paper explores the use of texture as a descriptor for the extraction of skin pixels in images. For this aim, we analyzed and compared a proposed color-based skin detection algorithm (using RGB, HSV and YCbCr representation spaces) with a texture-based skin location algorithm which used a measure called Spectral Variation Coefficient (SVC) to evaluate region features. We showed the usefulness of each skin segmentation feature (color versus texture) under different experiments that compared the accuracy of our algorithms under the same set of hand segmented images.

## 1 INTRODUCTION

Skin segmentation has many important applications related to finding and analyzing people behaviour on images or video sequences. Some of these applications are visual tracking for surveillance, face detection, hand gesture recognition, searching and filtering image contents on the web and many others. In the case of Intelligent Human-Computer Interaction (IHCI), the capture and interpretation of the users motions and emotions is an important element to understand of his/her current cognitive state (Duric, ). This way the interface can interactively adapt to the user. To achieve this goal, the detection of a human in a video sequence through the skin color usually becomes a crucial preprocessing for tracking his/her movements, gestures and facial expressions.

This segmentation problem is stated classifying the pixels of an input image in two groups: skin and non-skin pixels. An important requirement for automatic (or semiautomatic) skin detection systems is a trade off between correct classification rate and response time. The common approach to skin detection is using the skin colour as feature. Different colour spaces and skin-colour classifiers have been reported in the literature (Phung et al., 2005). For

modelling the skin colour, most authors mainly use three types of colour representation spaces (Kakumanu et al., 2007): basic, perceptual and orthogonal spaces. The first group corresponds to most common used format to represent and store digital image, and it includes spaces like RGB or normalized RGB. The second group separates perceptual characteristics of the colours (such as hue, saturation and intensity). These characteristics are mixed in RGB, and this second group includes space like HSV, TSL or HSI. The last group reduces the redundancy in RGB channels by representing the colour with statistically independent components, and it includes spaces like YCbCr or YIQ. Among the skin classification techniques, some representative works use thresholds and histogram-based algorithms (Zarit et al., 1999), simple Gaussian and Gaussian mixture models (Terrillon et al., 2000), neural classifiers, Bayes classifier (Jones and Rehg, 1999), fuzzy methods (Soria-Frisch et al., 2007), and so on. However, skin colour detection presents several important problems under uncontrolled conditions: it is affected by the illumination, it can be confused in different colour spaces with the colour of metal or wood surfaces and also skin colour can change from a person to another (even with people of the same ethnicity). Texture is another very

important feature on image segmentation which, differently from colour, has hardly ever been used for skin identification. Texture based segmentation taking into account the positional relations among pixels in a region of the image; it considers not only one pixel but its behaviour in an area that is named texels. Since the colour skin segmentation can be perturbed by the above mentioned conditions, we have considered in this work combining it with texture as a complementary feature.

This paper aims to systematically compare different colour and texture-based skin-detection methods. The comparison is performed both qualitatively (visually) and quantitatively (using False Positive and False Negative classification error rates) for the same set of images. For the colour-based skin detection approach, we use three common spaces (one for each of the mentioned types of colour representations): RGB, HSV and YCbCr, respectively. We have used a simple pixel rule-based for the RGB space, and have proposed a new region-growing algorithm for skin detection which can be similarly employed in the HSV and YCbCr. For the texture-based skin detection approach, the Spectral Variation Coefficient (SVC) (Nunes and Conci, 2007) is applied to estimate skin region feature. Both approaches (colour and texture-based skin-detection) are computationally efficient and suited for real time applications.

The rest of the paper is organized as follows. Next Section describes the elements involved in the presented colour-based skin detection methods, as well as a global pseudo code of the approach. Section 3 describes a similar presentation for the texture-based approach. Experimental results for colour and texture-based skin detection methods are presented and compared in Section 4. Main conclusions and future work are outlined in Section 5.

## 2 COLOUR-BASED SKIN SEGMENTATION

As the human skin seems to have a characteristic range of colour, many skin detection approaches are based on classifying pixels using their colour (Forsyth and Ponce, 2003). A wide set of colour spaces have been considered to model the skin chrominance (Vezhnevets et al., 2003)(Kakumanu et al., 2007). However, according to (Albiol et al., ): “for all colour spaces their corresponding optimum skin detectors have the same performance since the separability in skin or not skin classes is independent of the colour space chosen”. In other words, the quality of a skin detection method is more dependent on the proposed

detection algorithm and less on the used colour space. In this section, in order to investigate this statement, three colour spaces are used to represent the human skin: RGB, HSV and YCbCr. For the case of basic RGB space, we applied the following simple explicit skin detection algorithm that can be found in (Kovac et al., 2003) and works on all the image pixels for uniform daylight illumination:

$$\begin{aligned} & (R > 95) \wedge (G > 40) \wedge (B > 20) \wedge \\ & (\max R, G, B - \min R, G, B > 15) \wedge \\ & (|R - G| > 15) \wedge (R > G) \wedge (R > B) \end{aligned} \quad (1)$$

where R, G and B represent the value of pixel in the respective RGB colour channel with values ranging from 0 to 255. For HSV and YCbCr spaces, a similar approach is followed. We retain luminance information by converting the image to gray levels. Next, we produce an initial over segmented image of regions by applying a morphological watershed method. For each produced watershed region their corresponding colour histograms for each of the three channels (in HSV or YCbCr spaces) are computed and compared using the Battachariyya distance (Kailath, 1967) to the corresponding previous trained skin histograms. If these histograms are similar, the region is considered as “skin region” and used as seed for a region growing algorithm applied on neighbour non-skin regions. Finally, the explicit skin detection algorithm in Fig. 1 is applied to each pixel of detected skin regions to discard False Positive (FP) skin regions when the percentage of skin pixels in a region is above an experimental threshold. The global proposed colour based skin detection method is outlined by the algorithm pseudo code shown in Fig. 1. Some important remarks on this algorithm are the following ones.

- To recognize the skin pixels in images, we performed a training stage on the system. For this task, we obtained different skins histograms using a set of 2,314 images where the skin was manually segmented (i.e. only skin pixels were considered). A histogram was created for each of the three channels in each considered color space (RGB, HSV and YCbCr, respectively), and its number of histogram bins was set to 10.
- The Battacharyya distance (Kailath, 1967) was used to compare the three histograms in a given color space of a test region with the corresponding ones of training skin regions histograms to decide if a region can be considered or not as skin region. This distance measures the similarity of two discrete probability distributions. Given the discrete probability distributions obtained from the corresponding histograms vectors, p and q respectively,

```

procedure ColorSkinDetection (imageIn, colorSpace, trainedSkinHistograms);
// Produces imageOut as a binary image where skin pixels are set to 1 and non-skin pixels are set to 0
begin
  if (colorSpace is ‘RGB’) then
    Compute imageOut from imageIn using explicit skin detection directly applied to pixels
  else // similar processing for ‘HSV’ and ‘YCbCr’ color spaces
    Retain luminance information from imageIn;
    Create an image of regions using watershed algorithm;
    for each “watershed region” do
      Compute histograms for the three channels in considered color space;
      Compare histograms with those of trained skin region histograms using Battacharyya distance;
      if (value  $\geq$  threshold) then Set region as “seed skin region” and use it for region growing method
    end for;
    while (set of “seed skin regions” is not empty) do
      Find “new skin regions” from seeds using region growing and Battacharyya distance;
      Apply explicit RGB skin detection on the pixels of “skin regions” to filter “true skin regions”;
      Compose binary skin image imageOut by joining the set of “true skin regions”
    end if
end

```

Figure 1: Pseudo code for the global proposed colour based skin detection method.

over the same domain  $X$ , the Battacharyya coefficient BC is defined as:

$$BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)} \quad (2)$$

where this coefficient represents a value between 0 and 1, and the value  $BC=1$  means the highest degree of similarity between both histograms.

### 3 TEXTURE-BASED SKIN SEGMENTATION

Many texture classification schemes have been used for grey level images. Texture image segmentation is common in analysis of medical images, remote sensing scene interpretation, industrial quality control inspection, document segmentation, image recovery in databases, visual recognition systems, etc.

In this paper, we use a segmentation method that considers all RGB bands or other colours spaces combinations. It also permits to difference among different colour and texture combinations in the same image. It is based on the Spectral Variation Coefficient (Nunes and Conci, 2007) to evaluate region features (it can be used for very small to very large regions) and it permits to obtain correct real-time texture boundaries. It also allows distinguishing among different textures with few changes on the same type

of patterns. The positional relation among the pixels on the texel (or texture element) are considered. The channels of the colour information for each pixel are combined in a new way by considering their mean and standard deviation. This scheme is computationally very efficient and it is suitable for real-time applications that combine colour and texture. It can be used for all type of texture because the texture rules of what will be identified are completely given by the used seed and are adapted to each situation. Additionally the k-means clustering technique is used to segment the regions according the SVC value for each texel band.

The first step in the calculation of the SVC for each texel is combining the information of the three original image color channels. The texel size is defined by the user or by the application. In our implementation it must have square shape, with  $M \times M$  pixels, ranging from  $M = 3$  to 21 and using odd  $M$  values (that is  $3 \times 3$ ,  $5 \times 5$ , ...,  $21 \times 21$ ). To show how the channel combination is performed, let us use a numerical example. Consider the  $7 \times 7$  texel on Fig. 2.a, represented by their intensity values for each pixel in channels R, G and B (see Figs. 2.b, 2.c and 2.d, respectively).

To achieve a better characterization of the texture variations in the texel, considering also the channel order, the value of the pixel intensity of each channel (R, G or B) is substituted by a new value considering the others two channels, as shown in Fig. 3, (i.e. for the first position of the matrix). Considering the data

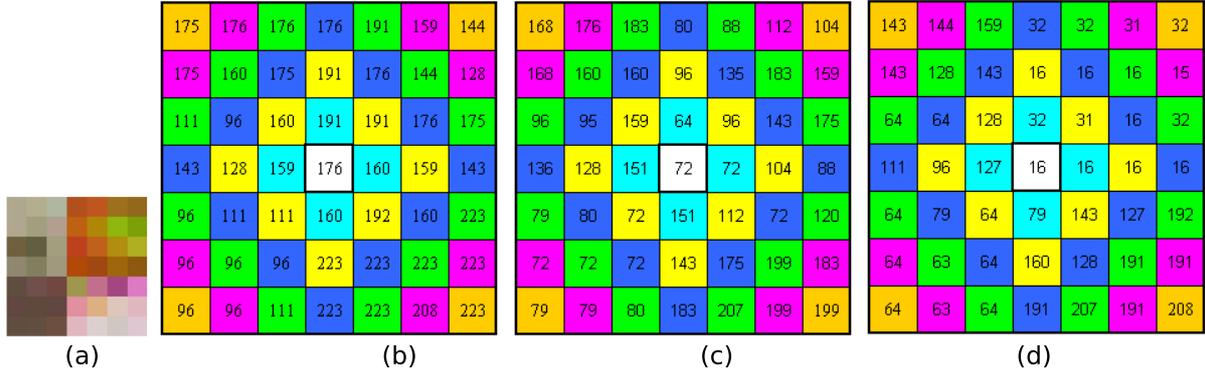


Figure 2: A 7x7 sample texel: (a) how it is originally seen by human eyes, (b) pixel intensities for Red, (c) Green and (d) Blue channels, respectively.

	R	G	B
$c=original\ value$	175	168	143
$a$	168	143	175
$b$	143	175	168
$x = \sqrt{a^2 + (b+1)^2}$	221,269	226,771	243,282
$y = \arctg(a/(b+1))$	0,862	0,682	0,803
$temp\ value = x \times y$	190,772	154,729	195,316
$c$	154,729	195,316	190,772
$d$	195,316	190,772	154,729
$x = \sqrt{c^2 + (d+1)^2}$	259,618	229,076	242,876
$y = \arctg(c/(d+1))$	0,740	0,823	0,630
$new\ value = x \times y$	192,040	188,594	152,902

Figure 3: Computed blended intensities in the R, G and B channels for a sample pixel.

of Fig. 2.b, the value 175 of the intensity of the first pixel in the channel R, will be changed by the result of a computation (illustrated by Fig. 3) which considers itself and the intensities of the other G and B channels (168 and 143, respectively). The new computed value will be 192,040.

The same procedure will be carried out for all pixels of the input image by substituting the original pixel intensity for its new computed blended channel value. This procedure was adopted, aiming distinction of the different combinations and order of each RGB channel in the texture on analysis.

The second step in the calculation of the SVC is to determine the average and the standard deviation values for each class of distances considering the used metric in the texel on analysis using the blended val-

ues. In this example of SVC computation, we are using the  $D_4$  metric, also known as Manhattan or city block distance (Sonka et al., 1999), to classify each pixel in the texel based on its distance to the central position. However, any other metric can be used in the implementation of the SVC procedure. Letting  $q(s,t)$  and  $p(x,y)$  be pixels on the  $(s,t)$  and  $(x,y)$  positions respectively, then  $D_4(p,q)$  is defined as:

$$D_4(p,q) = |x-s| + |y-t| \quad (3)$$

To exemplify the calculation, consider the 6 images in Fig. 4 that illustrates the levels of intensity of one of the 3 channels after blending (R for instance) for a texel of 7x7 pixels, where each classes at different  $D_4$  distance are represented in different colour (here represented values of distance from 1 to 6). The number of  $D_4$  distance classes under consideration is, of course, related to the texel size.

Finally, the SVC combines information from the space position of the pixels in the texture element (using the Fig. 4 distance classes) for each channel blended values through the mean  $md$  and the standard deviation  $sd$  inside each class of distances, thus obtaining:

$$SVD = \arctg\left(\frac{md}{sd+1}\right) \times \sqrt{md^2 + (sd+1)^2} \quad (4)$$

By continuing with this example and using the Figure 4 values, the SVC for each distance class associated with the R channel is computed and showed in Table 1.

The value of SVC for the Red channel SVCR, corresponding to the example pixel, is computed using the values of Table 1 and eq. 3, and SVCR=393,575. The SVC of the texel is also computed for the others G and B blended channels (in the same way as

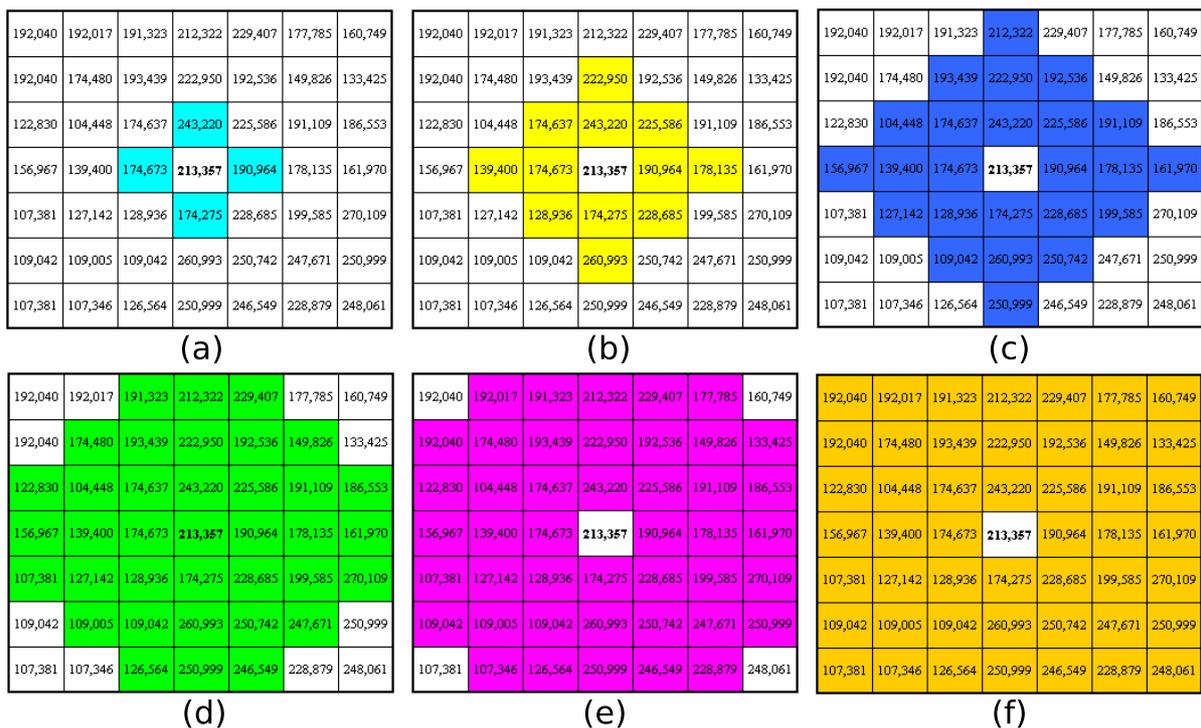


Figure 4: Six groups of pixels corresponding to the  $D_4$  metric distance classes on a  $7 \times 7$  texel on analysis: a)  $D_4 \leq 1$ , b)  $D_4 \leq 2$ , c)  $D_4 \leq 3$ , d)  $D_4 \leq 4$ , e)  $D_4 \leq 5$  and f)  $D_4 \leq 6$ .

Table 1: SVC computation for each distance class by considering the values of Figure 4.

Distance class	md	sd+1	SVC
1	195,783	29,203	281,627
2	195,205	40,312	272,505
3	187,198	44,971	257,026
4	184,846	49,282	250,654
5	182,863	49,760	247,336
6	182,380	49,973	246,467

shown in Fig 3 for the Red channel). This procedure can be adapted for other types of multichannel (or multiband) images. The value of the SVC in each channel of the color space defines a coordinate in the Euclidean space for the considered texel, characterizing a point in the three-dimensional space. Then, the samples of the training set are grouped by means of k-means clustering algorithm. The representative element of each cluster is its centroid, which has an average value for the SVC considered in each canal, relative to all the samples of cluster. From an initial estimation of the coordinates of the centroids, the algorithm computes the distance of each sample of the set of training skin centroids. This way, a region is considered as “skin region” and used as seed for a

region growing algorithm applied on neighbour non-skin regions if it is close to trained skin centroids. Finally, the explicit skin detection algorithm of equation 1 is applied to each pixel of detected skin regions to discard False Positive (FP) skin regions when the percentage of skin pixels in a region is above an experimental threshold. The proposed SVC-based algorithm for skin detection in the RGB color space for each texture region *texelln* is outlined by the pseudo code shown in Fig 5.

## 4 EXPERIMENTAL RESULTS

Initially, we tried to use the Compaq database (Jones and Rehg, 1999) for experiments. However, we visually checked that in many images of this database the manual skin segmentation was not correct. Therefore, our proposed color and texture approaches have been tested using other four color real images (shown by Fig. 6 where the skin was hand segmented). For each of these images, the corresponding number of pixels and percentages of skin and non-skin pixels are shown in Table 2. To compare the accuracy of proposed color and texture segmentation method for skin segmentation, a comparison

```

procedure SVCSkinDetection (texelIn) ;
// texelIn is a sample texture region where each pixel is represented by its Red, Green and Blue components
begin
  for each pixel and each channel of texelIn do
    Compute blending_texel value using texelIn; // as in Table 1
    Compute each distance class based on each pixel position in blending_texel; // using equation (3)
    for each channel of blending_texel do
      for each distance class of blending_texel do
        Compute mean md and the standard deviation sd values;
        Compute SVC of the distance class; // using equation (4)
      end for;
      Compute SVC of the color channel; // as in Table 2
    end for;
  return (SVCR, SVCG, SVCB) for each channel in RGB space for region texelIn
end

```

Figure 5: Pseudo code for the SVC-based skin detection method for each texture region *texelIn*.

of the corresponding classification errors: False Positives (FP) and False Negatives (FN) is presented.

Table 2: Some properties of the test images.

Image	Size (# pixels)	% skin pixels	% non-skin pixels
Javi	540,870	28.71	71.29
photo	307,200	30.42	69.58
Beckham	155,000	6.67	93.33
CarlosSainz	145,700	11.24	88.76

Some experiments were performed to compare the color and texture-based skin classification approaches. First, we analyzed and compared the effect of the tested color spaces (RGB, HSV and YCbCr, respectively) in skin classification using the algorithm presented in Section 2. This algorithm requires from three threshold parameters: *thRegion* that represents the percentage of pixels for a region to be “skin region” according to Battacharyya distance, *thNeighbour* that is the percentage of pixels for a neighbour of a “skin region” to be a new “skin region”, and *thRGB* that represents the percentage of skin pixels using eq. 1 in a region to be finally labelled as “skin region”. After some experimentation, we concluded that good parameter values were:  $[thRegion, thNeighbour, thRGB] = [0.4, 0.1, 0.5]$ . Figure 7 visually compares the segmentation results obtained by the color segmentation algorithm in the considered color spaces for these parameter values using CarlosSainz image. The result of the texture-based algorithm for the same test image is also shown in Figure 7. In general, the HSV color space produced better detection results than the other two used color spaces. Consequently, we have quantitatively com-

pared the FP and FN errors produced by our color algorithm in the HSV space with the corresponding ones produced by the texture segmentation method for the four test images. The corresponding results are presented in Tables 3 and 4, respectively. Table 3 corresponds to the results for the HSV color-based skin detection algorithm and Table 4 is referred to the results for the CVE texture-based skin segmentation algorithm. Both tables show the percentages of correctly detected skin pixels (the sum of percentages of TP and TN) and incorrectly detected skin pixels (the sum of percentages of FP and FN) for each respective algorithm and test image.

Table 3: Results for the HSV skin detection colour-based algorithm (where parameter values are  $[thRegion, thNeighbour, thRGB] = [0.4, 0.1, 0.5]$ ).

Image	javi	photo	Beckham	CSainz
% TP	27.92	29.61	3.57	6.72
% TN	70.16	68.38	93.12	84.78
% FP	1.13	1.20	0.21	3.98
% FN	0.79	0.81	3.10	4.52
% Suc.	98.08	97.99	96.69	91.50
% Err.	1.92	2.01	3.31	8.50

Some conclusions can be extracted from the experiments:

- For the considered test images the skin detection colour-based methods reduce in average a 22.4 % the False Positive (FP) error rate. This is mainly due to the final found skin region elimination, which discards some labelled skin regions using explicit pixel-based skin detection RGB. In this way, initial skin regions where the percentage of skin pixels is above an experimental threshold are



Figure 6: Considered test images: (a) Javi, (b) photo, (c) Beckham, and (d) Carlos Sainz.

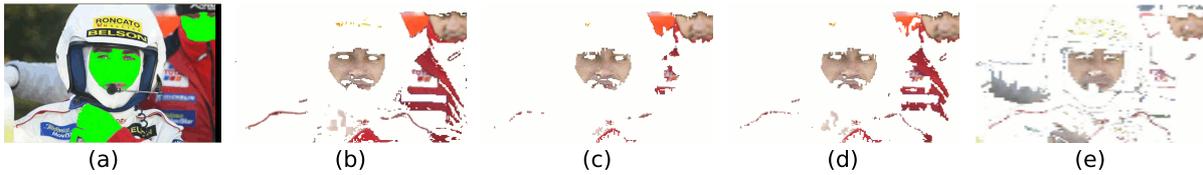


Figure 7: Visual results of detected skin for the three color space algorithms and the texture approach: (a) hand segmented skin, (b) RGB, (c) HSV, (d) YcbCr and (e) CVE texture.

Table 4: Results for the skin detection SVC texture-based algorithm (the number of classes in the application of the k-means algorithm varies between 5 and 10 in the experiments).

Image	javi	photo	Beckham	CSainz
% TP	28.53	29.45	6.03	7.83
% TN	69.31	68.48	89.57	87.20
% FP	1.98	1.10	3.76	1.56
% FN	0.18	0.97	0.64	3.41
% Suc.	97.84	97.93	95.60	95.03
% Err.	2.16	2.07	4.40	4.97

now labelled as “non-skin” regions.

- For the considered test images the skin detection texture-based methods reduce in average a 43.6% the False Negative (FN) error rate. This is due to ....
- By considering both FP and FN error rates, the skin detection texture-based approach reduces in average a 13.6% the misclassified skin pixels with respect to the colour-based approach (a 3.93% in of errors in the colour-based approaches versus a 3.4% in the texture-based method).
- The contour of segments region boundaries are also more properly identified using the texture-based algorithm.

## 5 CONCLUSION

This paper presented and compared both a color-based algorithm (using RGB, HSV and YCbCr representation spaces) and a texture-based algorithm (using the Spectral Variation Coefficient) for skin detection on color images. Although most work on skin detection is based on modelling the skin on different color spaces, we have explored the use of texture as a descriptor for the extraction of skin pixels in images. The accuracy provided by each segmentation feature-based algorithm (color versus texture) is shown under different hand-segmented images. The skin detection texture-based approach reduces in average a 13.6% the misclassified skin pixels with respect to the colour-based approach for the considered test images. A necessary future work is to validate the proposed algorithms using a standard skin database like the ECU dataset (Phung et al., 2005). This will permit to compare our recognition results with those presented by other authors for the same test images. Another improvement will consist in adapting our algorithms to detect skin in black or Asian people (not only white Caucasian).

## ACKNOWLEDGEMENTS

This research was partially supported by the Spanish Ministry of Education and Science Project no. TIN2005-08943-C02-02 (2005-2008).

## REFERENCES

- Albiol, A., Torres, L., and Delp, E. Optimum color spaces for skin detection. In *IEEE Int. Conf. on Image Processing*, volume 1.
- Duric, Z. e. a. Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. In *Proceedings of the IEEE*, volume 90.
- Forsyth, D. and Ponce, J. (2003). *Computer Vision: A Modern Approach*. Pearson, 1st edition.
- Jones, M. and Rehg, J. (1999). Statistical color models with application to skin detection. In *Proc. CVPR99*.
- Kailath, T. (1967). The divergence and bhattacharyya distance measures in signal selection. *IEEE Trans. on Comm. Tech.*, 15:52–60.
- Kakumanu, P., Makrogiannis, S., and Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40:1106–1122.
- Kovac, P., Peer, P., and Solina, F. (2003). Human skin colour clustering for face detection. In *EUROCON 2003*.
- Nunes, E. and Conci, A. (2007). Segmentao por textura e localizao do contorno de regies em imagens multibandas. *IEEE Latin America Transactions*, 5(3):185–192.
- Phung, S., Bouzerdoum, A., and Chai, D. (2005). Skin segmentation using color pixel classification: Analysis and comparison. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(1):148–154.
- Sonka, M., Hlavac, V., and Boyle, R. (1999). *Image Processing, Analysis, and Machine Vision*. PWS Publishing, 2nd edition.
- Soria-Frisch, A., Verschae, R., and Olano, A. (2007). Fuzzy fusion for skin detection. *Fuzzy Sets and Systems*, 158:325–336.
- Terrillon, J., Shirazi, M., Fukamachi, H., and Akamatsu, S. (2000). Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*.
- Vezhnevets, V., Sazonov, V., and Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. In *Proc. GRAPHICON03*.
- Zarit, B., Super, J., and Quek, F. (1999). Comparison of five color models in skin pixel classification. In *Proc. ICCV99*.