A NEW COLOUR SPACE FOR SKIN TONE DETECTION

Abbas Cheddad, Joan Condell, Kevin Curran and Paul Mc Kevitt
School of Computing and Intelligent Systems, Faculty of Computing and Engineering
University of Ulster at Magee, BT48 7JL, Northern Ireland, United Kingdom
Email: cheddad-a@email.ulster.ac.uk

ABSTRACT

The majority of existing methods have one thing in common which is the de-correlation of luminance from the considered colour channels. It is believed that the luminance is underestimated here since it is seen as the least contributing colour component to skin colour detection. This work questions this claim by showing that luminance can be useful in separating skin and non-skin clusters. To this end, this work uses a new colour space which contains error signals derived from differentiating the grayscale map and the non-encoded-red grayscale version. The advantages of this approach are the reduction of space dimensionality from 3D to 1D space and the construction of a rapid classifier necessary for real time applications. This method is meant to assist digital image steganography to orient the embedding process since skin information is deemed to be psycho-visually redundant.

Keywords: Luminance; colour transform; skin tone detection; steganography.

1. INTRODUCTION AND BACKGROUND

Detecting human skin tone is of utmost importance in numerous applications such as, motion analysis and tracking, video surveillance, face and gesture recognition, human computer interaction, image and video indexing and retrieval, image editing, vehicle drivers' drowsiness detection, real time gait and gesture recognition and steganography. Detecting human skin tone is regarded as a two-class classification problem, and took a considerable amount of attention from researchers in recent years [1, 2] especially those who deal with biometrics and computer vision aspects. Challenges facing biometrics researchers and particularly those who are dealing with skin tone detection include choosing a colour space, generating the skin model and processing the obtained regions to fit any specific application. Modelling skin colour necessitates the identification of a suitable colour space and the careful setting of rules for cropping clusters associated with skin colour. Despite a lot of work which tackled this problem, unfortunately most tend to put the illumination channel in the "non useful" zone and therefore act instead on colour transformation spaces that de-correlate luminance and chrominance components from an *RGB* image. It is important to note that illumination and luminance are defined slightly differently as they depend on each other. As this may cause confusion, for simplicity, this work will refer to both of them as the function of response to incident light flux or the brightness. Abadpour and Kasaei [3] concluded that "*in the YUV, YIQ, and YCbCr colour spaces, removing the illumination related component (Y) increases the performance of skin detection process*". Others [4, 5] were in favour of dropping luminance prior to any processing as they were convinced that the mixing of chrominance and luminance data makes *RGB* basis marred and not a very favourable choice for colour analysis and colour based recognition. Therefore, luminance and chrominance have been always difficult to tease apart unless the *RGB* components are transformed into other colour spaces. Comprehensive literatures exist which discuss in depth the different colour spaces and their performance [3, 6, 7]. Albiol *et al.* [8] show in their work that choosing colour space has no implication on the detection given an optimum skin detector is used, in other words all colour spaces perform the same. Shin *et al.* [9] argue and question the benefit of colour transformation for skin tone detection. This work goes a step further and shows that the abandoned luminance component indeed carries considerable information on skin tone. Many colour spaces used for skin detection are simply linear transforms from *RGB* and as such share all the shortcomings of *RGB*.

Colour transformations are of paramount importance in computer vision. There exist several colour spaces but the native representation of colour images is the *RGB* colour space which describes the world view in three colour matrices. The luminance is present in this space and thus various transforms are meant to extract it out. The *Y*, *Cb* and *Cr* components refer to Luminance, Chromatic blue and Chromatic red respectively. This colour space is used extensively in video coding and compression, e.g., MPEG, and is perceptually uniform. Moreover, it provides an excellent space for luminance and chrominance separability. *Y* is an additive combination of RGB components and hence preserves the high frequency image contents; the subtraction of *Y* in Eq. 1 cancels out the high frequency (*Y*) [10]. Given the triplet *RGB*, the *YCbCr* transformation can be calculated using the following system (Note: the transformation formula for this colour space depends on the used recommendation):

$$\mathrm{YC_bC_r}: \begin{cases} \mathrm{Y} = 0.299\mathrm{R} + 0.587\mathrm{G} + 0.114\mathrm{B} \\ \mathrm{C_b} = 0.56(\mathrm{B} - \mathrm{Y}) \\ \mathrm{C_r} = 0.71(\mathrm{R} - \mathrm{Y}) \end{cases} \tag{1}$$

Hsu *et al*. [4] used *CbCr* for face detection in colour images. They developed a model where they noticed a concentration of human skin colour in *CbCr* space. These two components were calculated after performing a lighting compensation that used a "reference white" to normalise the colour appearance. They claimed that their algorithm detected fewer non face pixels and more skin-tone facial pixels. Unfortunately, the testing experiments that were carried out using their algorithm are not in reasonable agreement with this assertion. Some of such results are reported in this work. Similarly, Yun *et al.* [11] used Hsu's algorithm with an extra morphological step. Shin *et al.* [9] showed that the use of such colour space gives better skin detection results compared to seven other colour transformations.

It is well established that human visual system incorporates colour-opponency and so there is a strong perceptual relevance in this colour space [12]. The Log-Opponent (LO) uses the base 10 logarithm to convert *RGB* matrices, note that this system does not assume here a particular range for *RGB* values, into I, R_g, B_y as follows:

$$I = L(G), R_g = L(R) - L(G), B_y = L(B) - (L(R) + L(G))/2$$
$$\text{where}, L(x) = 105 + \log_{10}(x+1). \quad (2)$$

This method uses what is called hybrid colour spaces. The fundamental concept behind hybrid colour spaces is to combine different colour components from different colour spaces to increase the efficiency of colour components to discriminate colour data. Also the aim is to lessen the rate of correlation dependency between colour components [13]. Here two spaces are used, namely *IRGBy* and *HS* from the *HSV* (Hue, Saturation and Value) colour space. HS can be obtained by applying a non-linear transformation to the *RGB* colour primaries as shown in Eq. 3.

$$H = \begin{cases} h, B \le G \\ 2\pi - h, B > G \end{cases}$$
$$\text{where}, h = \cos^{-1} \frac{1/2[(R-G)+(R-B)]}{\sqrt{(R-G)^2 + (R-G)(G-B)}} \quad (3)$$
$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)}$$
$$V = \max(R,G,B)$$

A texture amplitude map is used to find regions of low texture information. The algorithm first locates images containing large areas whose colour and texture is appropriate for skin, and then segregates those regions with little texture. The texture amplitude map is generated from the matrix *I* by applying 2D median filters.

The third method is *N*-rules *RGB* (*NRGB*) classifier. This is a simple yet powerful method to construct a skin classifier directly from the *RGB* basis which sets a number of rules (*N*) for skin colour likelihood. Kovač *et al.* [14] state that *RGB* components must not be close together, e.g., luminance elimination. They utilized the following rules: An R, G, B pixel is classified as skin if and only if:

$$R > 95 \ \& \ G > 40 \ \& \ B > 20 \ \& \ \max(R, G, B) - \min(R, G, B) > 15 \ \& \ |R-G| > 15 \ \& \ R > G \ \& \ R > B \quad (4)$$

The following section will discuss the proposed method and compare it with methods available in the literature [4, 12, 14]. Due to space constraints no further discussion on these methods can be given; however, recent research has previously discussed such approaches [15, 16].

2. PROPOSED METHOD

Illumination is nicely smeared along *RGB* colours in any given colour image. Hence, its effect is scarcely distinguished here. There are different approaches to segregate such illumination. The utilized transformation matrix is: $\vec{\alpha} = [0.298936021293775390, \ 0.587043074451121360, \ 0.140209042551032500]^T$, where the superscript *T* denotes the transpose operator to allow for matrix multiplication. Let Ψ denote the 3D matrix containing the *RGB* vectors of the host image and let $x \in [1, 2, ..., n]$, n=W*H, W, H denote the width and height of the image, respectively.

Note that the proposed method acts on the *RGB* colours stored in double precision, i.e., linearly scaled to the interval [0 1]. The initial colour transformation is expressed as in Eq.5.

$$I(x) = (\Psi(r(x), g(x), b(x)) \otimes \vec{\alpha}) \quad (5)$$

where $\otimes$ represents the product operation. This reduces *RGB* colour representation from 3D to 1D colour space. The vector *I(x)* eliminates the hue and saturation information while retaining the luminance. It is therefore regarded formally as a grayscale colour. Next, the algorithm tries to obtain another version of the luminance but this time without taking the *R* vector into account (most of skin colour tends to cluster in the red channel). The discarding of red colour is deliberate, as in the final stage it will help calculating the error signal. Therefore, the new vector will have the largest elements taken from *G* or *B*:

$$\hat{I}(x) = \max_{x \in \{1,...,n\}} (G(x), B(x)) \quad (6)$$

Eq. 6 is actually a modification of the way *HSV* (Hue, Saturation and Value) computes the *V* values. The only difference is that the method does not include in this case the red component in the calculation. Then for any value of *x*, the error signal is derived from the calculation of element-wise subtraction of the matrices generated by Eq. 5 and Eq. 6 which can be defined as:

$$e(x) = I(x) - \hat{I}(x) \quad (7)$$

Note that $e(x)$ must employ neither truncation nor rounding. Creating a *skin probability map* (SPM) that uses an explicit threshold based skin cluster classifier which defines the lower and upper boundaries of the skin cluster is crucial to the success of the proposed technique. A collection of 147852 pixel samples was gathered from different skin regions exhibiting a range of races with extreme variation of lighting effect. After transformation using the proposed method, the projection of data admits a distribution that could be easily fit into a Gaussian curve using *Expectation Maximization* (EM) method which is an approximation of *Gaussian Mixture Models* (GMM). This experiment also points out that there are no other Gaussians hidden in the distribution. To identify the boundaries, some statistics need to be computed. Let μ and σ denote the mean and standard deviation of the above distribution, and let Δ_{left} and Δ_{right} denote the distances from μ on the left and right hand side respectively. The boundaries are determined based on Eq. 8.

$$\mu - (\Delta_{left} * \sigma) \approx 0.02511, \ \mu + (\Delta_{right} * \sigma) \approx 0.1177 \quad (8)$$

where Δ_{left} and Δ_{right} are chosen to be 1 and 3 sigma away from μ respectively to cover the majority of the area under the curve. Hence, the precise empirical rule set for this work is given in Eq. 9.

$$f_{skin}(x) = \begin{cases} 1 \text{ if } \quad 0.02511 <= e(x) <= 0.1177 \\ 0 \qquad \text{otherwise.} \end{cases} \quad (9)$$

It is claimed that, based on experiments on extensive data set, this rule pins down the optimum balanced solution. Even though the proposed algorithm adopts the inclusion of illumination the 3D projection of the three matrices $I, \hat{I}, e$ shows clearly the skin tone clusters around the boundaries mentioned in Eq. 9. The carried experiments defeat the claim reported previously in [4] showing the deficiency of using luminance in modelling skin tone colour. The hypothesis that this work supports is "*luminance inclusion does increase separability of skin and non-skin clusters*". In order to prove this argument, the proposed method is tested on different *RGB* images with different background and foreground complexities. Some images exposing uneven transition in illumination were selected to demonstrate the robustness of the algorithm. Figure 1 shows a test image from the Internet image database and the corresponding detected skin regions of each algorithm. As shown, the proposed algorithm is exquisitely insensitive to false alarms; therefore, it has the least false negative pixels compared to the other three methods, which renders the output cleaner in terms of noise interference. The supreme advantage that the proposed method offers is the reduction of dimensionality from 3D to 1D, which contributed enormously to the algorithm's speed as can be seen in Table 1. The proposed colour model and the developed classifier can cope with difficult cases encapsulating bad and uneven lighting distribution and shadow interferences. Consequently, these results respond evidently to those authors who arguably questioned the effectiveness of the use of illumination based on its inherent properties. The proposed algorithm outperforms both *YCbCr* and *NRGB* which have attracted many researchers to date. In addition to the arbitrary still images downloaded from the Internet, the algorithm was tested against a larger benchmark comprising 150 frames from the popular video "Suzie.avi". Depicted in Figure 2 is some hand labelled ground truth models and the corresponding performance of the proposed method. Figure 3 shows the graphical performance analysis of the proposed algorithm against those reported in this work. As can be seen the proposed method is very efficient as it preserves lower rates for the dual false ratios while securing a high detection rate among all methods.

4. CONCLUSIONS AND FUTURE WORK

This paper addresses a novel colour space where we believe human skin clusters can be well classified with carefully selected boundaries. We provided the detailed algorithm coupled with some experiments and results which are promising. Our test database consists of randomly collected images from the Internet, 150 frames from *Suzie.avi* movie and the first 20 frames from *Sharpness.wmv* (comes with Dell^{TM} package) which were hand labelled to generate quantitative measurement. Additionally, we have set in context and proved that our proposition is deemed true as our set of results agrees reasonably with the speculated hypothesis. Therefore, we consider that "luminance inclusion does increase separability of skin and non-skin clusters". Bear in mind that we are not relying solely on luminance. Future work will extend experiments to explore if skin colour detection can be improved in the reduced dimensionality space of wavelets. This work is incorporated into information hiding specifically steganography in video files to restrain permanently rotation and translation attacks (see the full project at: www.infm.ulst.ac.uk/~abbasc/).

REFERENCES

[1]M. Corey, F. Farzam and J.H. Chong, The effect of linearization of range in skin detection, in: Proceedings of IEEE International Conference on Information, Communications & Signal Processing, 10-13 December 2007, pp. 1-5.

[2]U.A. Khan, M.I. Cheema and N.M. Sheikh, Adaptive video encoding based on skin tone region detection, in: Proceedings of IEEE Students Conference, 16-17 August 2002, vol (1), pp. 129-34.

[3]A. Abadpour and S. Kasaei, Pixel-based skin detection for pornography filtering, Iranian Journal of Electrical & Electronic Engineering, 1(3)(2005) 21-41.

[4]R.L. Hsu, M. Abdel-Mottaleb and A.K. Jain, Face detection in color images, IEEE Trans. Pattern Analysis and Machine Intelligence, 24(5)(2002) 696-702.

[5]V. Vezhnevets, V. Sazonov and A. Andreeva, A Survey on pixel-based skin color detection techniques, in: Proc. Graphicon, Moscow, September 2003, pp. 85-92.

[6]J.B. Martinkauppi, M.N. Soriano and M.H. Laaksonen, Behavior of skin color under varying illumination seen by different cameras at different color spaces, in: Proc. of SPIE, Machine Vision Applications in Industrial Inspection IX, 2001, vol. 4301, pp. 102-113.

[7]S. L. Phung, A. Bouzerdoum and D. Chai, Skin segmentation using color pixel classification: analysis and comparison, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(1)(2005) 148-154.

[8]A. Albiol, L. Torres, and E.J. Delp, Optimum color spaces for skin detection, in: Proceedings of the IEEE International Conference on Image Processing, 2001, vol. 1, pp.122-124.

[9]M.C. Shin, K.I. Chang and L.V. Tsap, Does colorspace transformation make any difference on skin detection?, in: Proceedings of IEEE Workshop on Applications of Computer Vision, December 2002, pp. 275-279.

[10]N.X. Lian, V. Zagorodnov and Y.P. Tan, Image Denoising Using Optimal Color Space Projection, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 14-19 May 2006, vol.2, pp. 93-96.

[11]J.U. Yun, H.J. Lee, A.K. Paul and J.H. Baek, Robust Face Detection for Video Summary Using Illumination-Compensation and Morphological Processing, in: Proceedings of IEEE International Conference on Natural Computation, 24-27 August 2007, pp. 710-714.

[12]J. Berens and G.D. Finlayson, Log-opponent chromaticity coding of colour space, in: Proceedings of IEEE International Conference on Pattern Recognition, Barcelona, 2000, v.1, pp. 206-211.

[13]D. Forsyth and M. Fleck, Automatic Detection of Human Nudes, International Journal of Computer Vision, 32(1)(1999) 63-77.

[14]J. Kovač, P. Peer and F. Solina, Human Skin Colour Clustering for Face Detection, in: Proceedings of International Conference on Computer as a Tool, Slovenia, 22-24 September 2003, vol.2, pp. 144-148.

[15]A. Cheddad, J. Condell, K. Curran and P. Mc Kevitt, Method for Skin tone detection, United Kingdom Patent Application No. 0819982.0, University of Ulster.

[16]A. Cheddad, J. Condell, K. Curran and P. Mc Kevitt, A skin tone detection algorithm for use in steganography, in press, Signal Processing Journal, 2009, Elsevier.

APPENDIX

Fig 1. Performance analysis: (left to right) original image, outputs of [4], [12], [14] and of the proposed method respectively. Shown is a sample from the Internet database that appears in Table 1, where the image corresponds to image 2 in the table.

Table 1. Comparison of computational complexity of the proposed method against other methods [4], [12] and [14] on 12 images obtained from the Internet database of which a sample is shown in Fig 1. The algorithms were implemented in MATLAB.

Image #	Number of Pixels	Time elapsed in seconds [4]	[12]	[14]	Proposed
1	840450	0.5160	33.515	7.796	0.125
2	478518	0.4060	22.094	4.156	0.047
3	196608	0.2970	4.547	2.188	0.062
4	196608	0.3280	3.563	1.906	0.062
5	849162	0.5160	33.062	7.531	0.078
6	850545	0.6090	39	8.343	0.062
7	849162	0.6090	39.219	6.641	0.078
8	849162	0.5160	39.172	8.484	0.078
9	849162	0.6100	38.203	6	0.078
10	7750656	3.1720	> 600 *	54.86	0.562
11	982101	0.6410	79.469	7.297	0.078
12	21233664	9.3910	> 600 *	144	1.531

(*) the Log algorithm [12] did not converge for more than 10 min which forced us to halt its process.

Fig. 2. The first 4 frames *Suzie.avi*: (left) selected ground truth from our 150 manually cropped frames and (right) the proposed algorithm outputs.

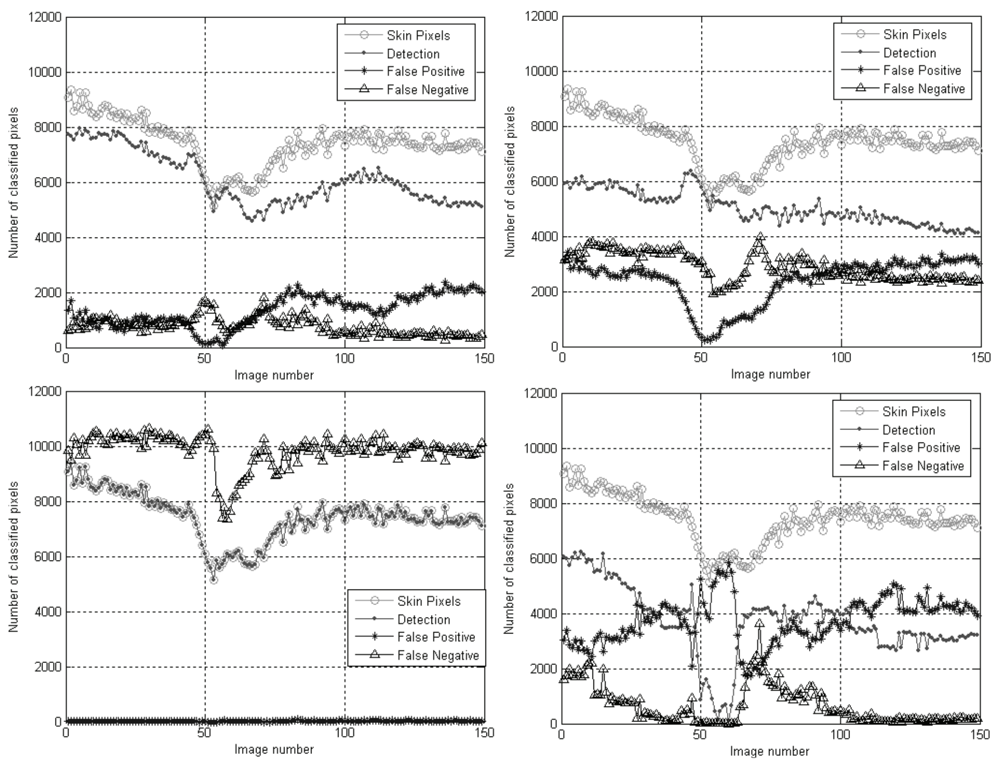

Fig. 3. Performance analysis on the entire 150 frames: (Top left to bottom right) our method, [14], [12] and [4] respectively.