World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:3, No:4, 2009

Performance of Histogram-Based Skin Colour Segmentation for Arms Detection in Human Motion Analysis Application

Rosalyn R. Porle, Ali Chekima, Farrah Wong, and G. Sainarayanan

Abstract**—**Arms detection is one of the fundamental problems in human motion analysis application. The arms are considered as the most challenging body part to be detected since its pose and speed varies in image sequences. Moreover, the arms are usually occluded with other body parts such as the head and torso. In this paper, histogram-based skin colour segmentation is proposed to detect the arms in image sequences. Six different colour spaces namely RGB, rgb, HSI, TSL, SCT and CIELAB are evaluated to determine the best colour space for this segmentation procedure. The evaluation is divided into three categories, which are single colour component, colour without luminance and colour with luminance. The performance is measured using True Positive (TP) and True Negative (TN) on 250 images with manual ground truth. The best colour is selected based on the highest TN value followed by the highest TP value.

Keywords**—**image colour analysis, image motion analysis, skin, wavelet transform.

I. Introduction

DETECTING arms for human motion analysis application is a challenging task as the arms tends to move farther, faster and more often than the other body parts. Silhouette and edge features may not be sufficient and effective for arms detection especially when the arms were occluded in the head or torso regions. Colour is a useful feature for arms detection. It provides computationally effective yet, robust information against rotation, scaling and partial occlusions [1]. Performing skin colour segmentation for arms detection in account of all variation such as race factors, illumination and blurriness due to fast movement is difficult and time consuming. Several issues have to be considered. This includes the choices of colour space to be implemented and how the skin colour distribution should be modelled. In human motion analysis, the skin colour template can be constructed specific to the human subject considered in the system [2]. Several researchers in human motion analysis field apply colour for performing arms or legs detection. In [3], the normalized RGB is used to segment the lower arms and upper-leg regions. A skin colour model is constructed from the pixels of skin colour cluster obtained from the k-means algorithm within the detected face region. The skin region is segmented using the learned colour histogram. In [2], the mean and standard deviation of RGB component in skin region are evaluated to segment the skin and non-skin regions. Thresholding method is used for the segmentation. In [4], the bivariate Gaussian (Hue and Saturation) is constructed from skin samples taken from natural images. A refined skin model is created based on a new bivariate that is learned from the inlier in region H.

In this paper, histogram-based skin colour segmentation is proposed for arms detection. A skin colour model is constructed from the skin colour pixels that are extracted within the face region. Histogram analysis is then performed to the skin colour model to obtain two threshold values. These threshold values are used to segment the image. Lastly, post processing is performed to eliminate other regions in the image except the arm regions. To select the best colour space for the histogram-based skin colour segmentation, six colour spaces are evaluated. These colour spaces are RGB, rgb, HSI, TSL, SCT and CIELAB. The performance evaluation is performed in three categories, which are single colour component, colour without luminance and colour with luminance. True Positive (TP) and True Negative (TN) are used as measurement metrics. The colour that provides the highest TN followed by TP value is selected as the best colour for the proposed segmentation method.

This paper is organized into six sections. The first section is the introduction. The second section describes the colour spaces that are used in this paper. The histogram-based skin colour segmentation procedure is explained in the third section. Later, in the fourth section, the post processing procedure is described. Experimental results and discussion are included in section five. Finally, overall research is concluded in section six.

R. R. Porle is with Computer Engineering Program, School of Engineering and Information Technology, Universiti Malaysia Sabah, 88999 Kota Kinabalu, Sabah, Malaysia (phone: +6088-320000 ext 3083; fax: +6088-320348; e-mail: rosalynrporle@yahoo.com.my).

A. Chekima, is with Computer Engineering Program, School of Engineering and Information Technology, Universiti Malaysia Sabah, 88999 Kota Kinabalu, Sabah, Malaysia (e-mail: chekima@ums.edu.my).

F. Wong is with the Electrical and Electronics Engineering Program, School of Engineering and Information Technology, Universiti Malaysia Sabah, 88999 Kota Kinabalu, Sabah, Malaysia (e-mail: farrah@ums.edu.my).

G. Sainarayanan is with the ICT Academy of Tamil Nadu, ELCOT Complex, 2-7 Developed Plots, Industrial Estate, Perungudi, 600096, Chennai, India (e-mail: sainarayanan@ictact.org).

II. Colour Spaces for Skin Colour Segmentation

The selection of colour space is crucial as it influences the performance of the segmentation methodology. There are many colour spaces available. In this paper, instead of randomly choosing a colour space, six colour spaces; RGB, rgb, HSI, TSL, SCT and CIELAB are examined. The description of each the colour space is presented as the following:

A. RGB

RGB colour space is considered as a basic colour space. It corresponds to the three primary colours which are the red, green and blue components. It is the most commonly used colour space for digital image storing and representation.

B. rgb

The rgb colour space is the normalized RGB, which is defined as

$$r_{rgb} = R/(R+G+B) \quad (1)$$

$$g_{rgb} = G/(R+G+B) \quad (2)$$

$$b_{rgb} = B/(R+G+B). \quad (3)$$

The b_{rgb} component is usually omitted in the skin segmentation as it does not hold any significant information. One of the advantages of this colour space is that it is invariant to lighting variation [5].

C. HSI

HSI is a colour space that describes colours as perceived by human beings [6]. HSI represents the Hue, Saturation and Intensity respectively. Hue defines the dominant colour such as red, green, purple and yellow of an area. Meanwhile, the saturation measures the colourfulness or whiteness in the perceived colour. Lastly, the intensity provides a measure of the brightness of the colour. The HSI components can be expressed by

$$H_{HSI} = \cos^{-1}\left[\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right] \quad (4)$$

$$S_{HSI} = 1-[3(\min(R,G,B))/(R+G+B)] \quad (5)$$

$$I_{HSI} = (R+G+B)/3 \quad (6)$$

D. TSL

TSL (Tint, Saturation and Luminance) is a perceptual colour that is similar to HSI colour space [7]. T_{TSL} describes the colour, S_{TSL} is the saturation and L_{TSL} is the luminance (for Gamma corrected RGB values). This is given as

$$T_{TSL} = \begin{cases} \frac{1}{2\pi}\tan^{-1}\left(\frac{r_{TSL}}{g_{TSL}}\right)+\frac{3}{4};\ g_{TSL}<0 \\ \frac{1}{2\pi}\tan^{-1}\left(\frac{r_{TSL}}{g_{TSL}}\right)+\frac{1}{4};\ g_{TSL}>0 \\ 0;\ g_{TSL}=0 \end{cases} \quad (7)$$

$$S_{TSL} = \sqrt{9/5\left(r_{TSL}^{\ 2}+g_{TSL}^{\ 2}\right)} \quad (8)$$

$$L_{TSL} = 0.299R+0.587G+0.114B \quad (9)$$

where

$$r_{TSL} = r_{rgb} - 1/3 \quad (10)$$

$$g_{TSL} = g_{rgb} - 1/3 \quad (11)$$

E. SCT

The Spherical Coordinate Transform (SCT) colour components are defined in (12) to (14). L_{SCT} is the intensity; A_{SCT} and B_{SCT} represent colours independently from intensity. If a particular colour is plotted as a point in RGB space, then the norm of the vector from the origin to the point is L_{SCT}, the angle between the vector and the blue axis is A_{SCT}, and the angle between the red axis and the projection of the vector onto the RG plane is B_{SCT} [8].

$$L_{SCT} = \sqrt{R^2+G^2+B^2} \quad (12)$$

$$A_{SCT} = \cos^{-1}(B/L_{SCT}) \quad (13)$$

$$B_{SCT} = \tan^{-1}(G/(L_{SCT}(\sin\angle A_{SCT}))) \quad (14)$$

F. CIELAB

The last colour space is CIELAB, which was adopted as an international standard in the 1970s, by CIE (International Commission on Illumination). The first colour component, L_{CIELAB}, is the lightness of colour. The A_{CIELAB} component represents its position between red/magenta and green. Meanwhile the B_{CIELAB} component represents its position between yellow and blue.

$$L_{CIELAB} = \begin{cases} 116(Y/Yn)^{1/3}-16 & ;Y/Yn>0.008856 \\ 903.3(Y/Yn) & ;\text{otherwise} \end{cases} \quad (15)$$

$$A_{CIELAB} = 500[f(X/Xn)-(Y/Yn)] \quad (16)$$

$$B_{CIELAB} = 200[f(Y/Yn)-(Z/Zn)] \quad (17)$$

where

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357800 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (18)$$

and

$$f(t) = \begin{cases} t^{1/3} & ; t > 0.008856 \\ 7.787\,t + (16/116) & ; \text{otherwise} \end{cases} \quad (19)$$

III. Histogram-based Skin Colour Segmentation

The flowchart of the histogram-based skin colour segmentation is illustrated in Fig. 1. Firstly, the colour space of input image is converted. Approximation image, which is extracted using Haar wavelet transform at level one [9], is used as input image. The wavelet transformation shrinks the images and at the same time smoothes the texture of the image. This can reduce the effect of lightness variations in the images.

The input images are in RGB colour format. The description of the desired colour was given in previous section. Then, the converted colour image is separated into three colour image components. For example the rgb colour image is separated into r_{rgb}, g_{rgb} and b_{rgb} colour image component respectively. The skin colour sample that is specific to the user in the image can be obtained from the head region. The location of the head region is estimated using pose estimation technique. Once the location is determined, the colour information within the head region is extracted. The colour information within the head region contains skin and non-skin colours. The non-skin colour is usually originated from the hair, eyes and the lips regions of the user. Extraction of the non-skin colour is time consuming. Moreover, it is difficult to distinguish the non-skin colour regions when the head is relatively small in the image. Due to these reasons, the non-skin colour in the skin colour sample can be ignored.

The extracted skin colour sample is analyzed using histogram distribution. From this histogram, the skin colour range for the user has to be determined. It is noted that each of the colour image components provides different range of histogram distribution. This is due to the small colour range in the skin colour sample. Therefore, each bin in the histogram is set to represent only one colour value. The occurrence in each bin is referred as frequency, f_c. Two threshold values are used to determine the skin colour range. The low threshold, T_{cl}, is the minimum colour value in the skin colour range. Meanwhile, the high threshold, T_{ch}, is the maximum colour value in the skin colour range. Minimum frequency, f_{cmin}, is used to calculate these threshold values. Let f_{cmax} be the maximum frequency and P_f is the frequency percentage. The range of P_f is set between 0 and 1. The minimum frequency is calculated as

$$f_{cmin} = f_{cmax} \times P_f \quad (20)$$

The skin colour range is assumed to be in the range where the frequency is more than the minimum frequency. If the histogram bin is denoted as H_b, the low threshold value is calculated as

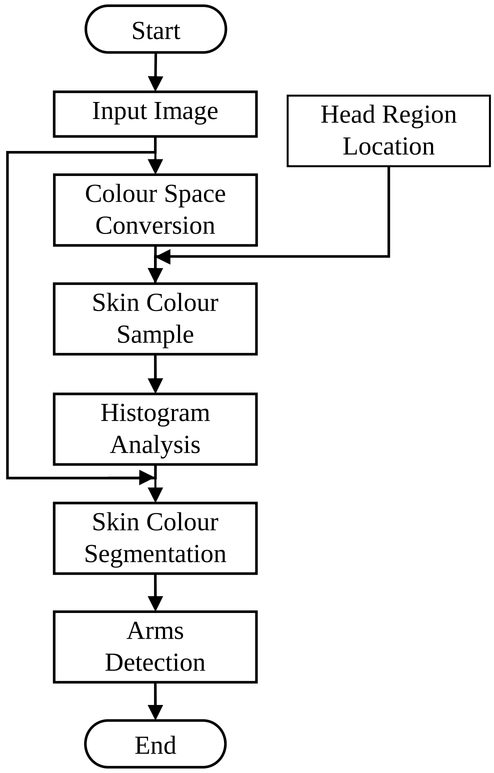

Fig. 1 Flowchart of the histogram-based skin colour segmentation

$$T_{cl} = \min\left[H_b\left(f_c > f_{cmin}\right)\right] \quad (21)$$

Meanwhile, the high threshold value is calculated as

$$T_{ch} = \max\left[H_b\left(f_c > f_{cmin}\right)\right] \quad (22)$$

After the threshold values are obtained, the colour image components can be segmented. The colour pixel in the colour image component is denoted as I_c. The colour pixel is detected as skin colour if its value lies between the low threshold and the high threshold values. If the colour pixel value is outside of this range, it is detected as non-skin colour. The colour pixel value is then set to 1 if it a skin colour. Otherwise, if it is a non-skin colour, the value is set to 0. This can be expressed as

$$I_c = \begin{cases} 1 & \text{if } T_{cl} \le I_c \le T_{ch} \\ 0 & \text{else} \end{cases} \quad (23)$$

After the colour image components are segmented, logical operator is used to combine these images. Two logical operators are suggested for the combination. The first logical operator is OR operator. For OR operator, the colour pixel of the combined image is assigned to 1 if any of the colour pixels, at the same location in the colour image components are detected as skin colour. Otherwise, the colour pixel of the combined image is assigned to 0 if none of the colour pixels, at the same location in the colour image components are detected as non-skin colour. The second logical operator is

AND operator. For AND operator, the colour pixel of the combined image is assigned to 1 if all the colour pixels, at the same location in the colour image components are detected as skin colour. Otherwise, the colour pixel of the combined image is assigned to 0 if any of the colour pixels, at the same location in the colour image components are detected as non-skin colour. The best logical operator for the colour image components combination will be determined in section IV.

IV. Experimental Results and Discussion

Six colour spaces are evaluated in this paper. They are RGB, rgb, HSI, TSL, SCT and CIELAB. The performances of these colour spaces are measured using True Positive (TP) and True Negative (TN). In the skin colour segmentation context, True Positive is defined as the number of correctly detected skin colour pixels divided by the total number of skin colour pixels. High TP value means that the colour has high skin colour detection performance. On the other hand, low TP value means that the colour has low performance in skin colour detection. The True Negative is defined as the number of correctly detected non-skin colour pixels divided by the total number of non-skin colour pixels. High TN value means that the colour has high performance in eliminating non-skin colour pixels. On the other hand, low TN value means that the colour has poor performance in eliminating non-skin colour pixels.

The performance measurement is performed in three categories. The first category is to evaluate the performances of each colour component. There are 18 colour components to be analyzed. The aim of this evaluation is to determine whether a single colour component can be used to segment the arms in the image. Another purpose is that the possibility of a single colour component to be combined with other colour component of different colour space. This combination may be performed if the three categories of the performance measurement did not produce satisfactory or higher performance. The second category is to evaluate the colour space without the luminance or intensity colour component. Many researchers suggested that omitting the luminance/intensity component from the colour space will improve the segmentation performance [10]. This possibility is also investigated in this paper. In this case, ten colour combinations are evaluated. Five colour combinations used OR operator and the other five colour combinations used AND operator. The last category is to evaluate the colour space with luminance or intensity colour component. For this category, 12 colour components combinations are evaluated.

The experimental database consists of 150 images. These images were extracted from 30 recorded videos in indoor environment with different user for each recorded video. For the evaluation purpose, most of the parameters are fixed. It is assumed that only P_f in (20) is adjustable. Ten different ratio values were selected (i.e. 0.1, 0.2, 0.3, …, 1) and from each ratio value, TP and TN are then calculated. Ground truth for test images were constructed manually. From these three categories, the best colour component or combination that produces the highest TN is selected as the best colour for the segmentation procedure. The results of the evaluation are presented as the following:

A. Single Colour Component

The first experiment evaluates the performance of 18 single colour components. Table I shows the performance result of 18 colour components. The best TN for each colour component is presented in the third column. The corresponding TP and P_f values are presented in the fourth and fifth column respectively. Three colour components have higher TN value, which are above 80 percent. They are H_{HSI}, T_{TSL} and A_{CIELAB}. By referring to the description of these colour components, they represent the colours (hue, tint and red/magenta and green) in the colour space respectively. The A_{CIELAB} colour component particularly has the highest TN value compared to the other colour components. This is because A_{CIELAB} only detect a specified colour range (red/magenta and green). This may not covers the colour range available in the background.

Other colour components such as R, G, B, r_{rgb} and g_{rgb} provide moderate performance. In this case TN value is around 70 to 80 percent. Meanwhile, the colour components such as b_{rgb}, I_{HSI}, L_{TSL}, L_{SCT}, A_{SCT} and B_{SCT} have performance that is around 60 to 70 percent of TN value. The remaining colour components, which are S_{HSI}, S_{TSL}, L_{CIELAB} and B_{CIELAB}, are the colour components that provide lowest TN performance.

B. Colour without Luminance Component

The second experiment evaluates the performance of ten colour combination without luminance or intensity component. In more detail; the rgb without b_{rgb} component, HSI without I_{HSI} component, TSL without L_{TSL} component,

TABLE I
The Performance of Single Colour Component

No.	Colour Component	True Negative	True Positive	*Pf*
1	R	73.87	57.46	0.6
2	G	73.51	56.43	0.6
3	B	71.55	59.73	0.6
4	r_{rgb}	74.21	51.87	0.6
5	g_{rgb}	70.72	51.08	0.5
6	b_{rgb}	65.71	55.14	0.6
7	H_{HSI}	85.48	53.85	0.5
8	S_{HSI}	57.58	50.98	0.6
9	I_{HSI}	67.30	60.97	0.6
10	T_{TSL}	82.85	52.73	0.5
11	S_{TSL}	51.71	43.81	0.7
12	L_{TSL}	68.82	58.58	0.6
13	L_{SCT}	66.56	60.87	0.6
14	A_{SCT}	61.43	52.51	0.6
15	B_{SCT}	60.05	65.32	0.6
16	L_{CIELAB}	58.53	57.74	0.6
17	A_{CIELAB}	95.99	58.59	0.1
18	B_{CIELAB}	50.43	49.81	0.5

SCT without L_{SCT} component and CIELAB without L_{CIELAB} component. Five colour combination used OR operator and the other five colour combination used AND operator. The result of colour combination without luminance component using OR and AND operators is shown in Table II. For OR operator, TN value for all combination is around 40 to 60 percent. Meanwhile, their respective TP values are around 50 to 70 percent. CIELAB recorded the highest performance with TN value of 58.83 percent for the OR operator. The second and third highest performances for the OR operator are HSI and SCT, with TN value of 56.00 and 55.41 percent respectively. For the AND operator, the TN value for all colour combination is more than 80 percent except SCT colour with 69.84 percent. Their corresponding TP values are around 50 percent. CIELAB has the highest TN with 97.71 percent for the AND operator. This is followed by rgb, HSI and TSL also when applying the OR operator. By comparing between the OR and AND operators, it is concluded that the AND operator performs better than the OR operator. Among all colour combination, CIELAB with AND operator achieved the highest performance.

C. Colour Space

The third experiment evaluates the performance of 12 colour spaces. Six colour combination used OR operator and the other six colour combination used AND operator. The result of colour space performance using OR and AND operators is shown in Table III. For OR operator, RGB colour space has the highest performance with TN value of 73.47 percent. This is followed by HSI, SCT, rgb, CIELAB and TSL colour spaces. All colour spaces except rgb have TP performances around 50 percent. The rgb colour space has the highest TP, which is around 66.43 percent. For the AND operator, the best colour space for this combination is CIELAB with TN and TP values of 97.68 and 55.07 percent respectively. The other colour spaces have TN values around 70 percent. By comparing these two operators, the AND operator gives better performance than OR operator. Among the colour spaces, CIELAB achieved highest performance in this category.

D. Overall result

The single colour component, colour space without luminance and colour space performances are compared to determine the best result for the histogram based skin colour segmentation. In all three categories CIELAB colour space has better performances than the other colour spaces. The CIELAB without luminance component using AND operator has the highest performance in eliminating non-skin pixels, which reached about 97.71 percent. This attribute is desirable in human motion analysis application as the occlusion and motion blur usually occurs during the video capture. The performance in detecting the skin pixel is moderate, which is around 55.01 percent. This is expected as the skin colour information in human's head region may not cover all skin colour range in the arms. The illumination changes and motion blur are also some of the factors that the colour in the arm region is slightly differ from the skin colour in the head region. In motion analysis application, the deformable arm regions can be solved by estimating or modelling the arms using robust technique.

TABLE II
THE PERFORMANCE OF COLOUR COMBINATION WITHOUT LUMINANCE COMPONENT

No.	Colour Combination	Operator	True Negative	True Positive	*Pf*
1	rgb	OR	46.37	73.15	0.8
2	HSI	OR	56.00	65.53	0.8
3	TSL	OR	52.99	55.75	0.9
4	SCT	OR	55.41	60.51	0.8
5	CIELAB	OR	58.83	52.39	0.8
6	rgb	AND	86.52	50.49	0.3
7	HSI	AND	83.94	58.08	0.3
8	TSL	AND	83.87	55.08	0.3
9	SCT	AND	69.84	58.56	0.4
10	CIELAB	AND	97.71	55.01	0.1

TABLE III
THE PERFORMANCE OF COLOUR COMBINATION WITH LUMINANCE COMPONENT

No.	Colour Component	Operator	True Negative	True Positive	*Pf*
1	RGB	OR	73.47	57.03	0.7
2	rgb	OR	58.55	66.43	0.7
3	HSI	OR	65.27	50.48	0.8
4	TSL	OR	53.91	53.36	0.8
5	SCT	OR	59.32	53.13	0.8
6	CIELAB	OR	55.70	55.38	0.6
7	RGB	AND	75.37	56.68	0.5
8	rgb	AND	78.49	55.83	0.3
9	HSI	AND	78.29	62.98	0.3
10	TSL	AND	76.63	60.17	0.3
11	SCT	AND	72.94	51.32	0.5
12	CIELAB	AND	97.68	55.07	0.1

V. CONCLUSION

Arms detection for human motion analysis application can be performed by evaluating the skin colour information in human's head region. The designed histogram-based skin colour segmentation computed the maximum colour frequency and two threshold values were derived based on the evaluated percentage of the maximum colour frequency. From the six colour spaces proposed for this algorithm, the CIELAB without luminance component using AND operator has achieved the highest TN value, thus, this colour is selected as the best colour to be used in the segmentation procedure.

REFERENCES

[1] P. Kakumanu, S. Makrogiannis and N. Bourbakis, "A survey of skin-color modelling and detection methods," *Pattern Recognition*, vol. 40, no. 3, pp. 1106–1122, March, 2007.
[2] J. Mulligan, "Upper Body Pose Estimation from Stereo and Hand-Face Tracking," in *Proc. 2nd Canadian Conf. on Computer and Robot Vision*, 2005, pp. 413–420.
[3] G. Hua, M. H. Yang and Y. Wu, "Learning to estimate human pose with data driven belief propagation," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, 2005, Vol. 2, pp. 747–754.
[4] A. S. Micilotta and R. Bowden, "View-based location and tracking of body parts for visual interaction," in *Proc. of British Machine Vision Conference.* 2004, vol. 2, pp. 849–858.

[5] S. Mallat, “Wavelets for a vision,” *Proc. of the IEEE*, 1996, vol. 84 (4), pp. 604–614.

[6] T. Acharya, and A. Ray, *Image Processing: Principles and Application*. John Willey & Sons, 2005.

[7] V. Vezhnevets, V. Sazonov and A. Andreeva, “A Survey on Pixel-Based Skin Color Detection Techniques,” in *Proc. of GRAPHICON*, Moscow, Russia, 2003, pp. 85–92.

[8] S. E. Umbaugh, *Computer imaging: digital image analysis and processing*. CRC Press, 2005.

[9] J. -C. Terrillon, M. N. Shirazi, H. Fukamachi and S. Akamatsu, “Comparative performance of different skin chrominance models andchrominance spaces for the automatic detection of human faces in color images,” *in Proc. of 4th IEEE International Conference on Automatic Face and Gesture Recognition,* Grenoble, France, 2000, pp. 54–61.

[10] S. Jayaram, S. Schmugge, M. C. Shin and L. V. Tsap, “Effect of colorspace transformation, the illuminance component, and color modeling on skin detection,” *in Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, vol. 2, pp. 813–81.