**Proceedings of the 11th IASTED International Conference
Computer Graphics and Imaging (CGIM 2010)
February 17 - 19, 2010 Innsbruck, Austria**

A COLOUR BASED APPROACH FOR FACE SEGMENTATION FROM VIDEO IMAGES UNDER LOW LUMINANCE LEVELS

S. Anishenko[1,2], D. Shaposhnikov[1]
R. Comley[2], X. Gao[2*]

[1]A.B.Kogan Research Institute for Neurocybernetics, Southern Federal University, Rostov-on-Don, Russia.
sergey.anishenko@gmail.com

[2] School of Engineering and Information Sciences, Middlesex University, London, NW4 4BT, UK.
*Corresponding author, x.gao@mdx.ac.uk

ABSTRACT

For tracking head motions from a sequence of video images, segmentation of faces usually takes place first to locate head positions. Although a plethora of methods have been developed to segment faces, each approach has its own advantages and disadvantages depending on the properties of collected video clips, especially under low luminance level (i.e. lower than 10 cd/m$^{2)}$. This study extends a traditional approach based on skin colour to segment faces, and has arrived at a mixed-colour-space algorithm, which takes viewing illumination and post position into consideration and shows very promising segmentation results with fewer false regions. The procedure includes obtaining colour attributes that are selected from varying colour spaces/models in order to detect the regions of interest (ROIs), which maximises the characteristics of faces. Colour edges are then applied to classify ROI into faces.

KEY WORDS

Colour appearance model, video image processing, CIECAM, face segmentation, motion detection

1. Introduction

Face segmentation has been applied widely in many applications, including CCTV surveillance monitoring, face recognition, human computer interaction, and motion tracking, and yet still remains one of the challenges in the field of computer vision, especially when it comes to processing video images. This is because the change of lighting conditions, variation of poses and dynamics of backgrounds have all contributed to the dissimilar appearances a face may present in a clip of video sequences.

Many approaches have since been proposed and fall into two broad categories. The first group is based on modelling of face shape and appearance. Because face shapes may vary between people and change with pose positions [1], the success of this group of methods depends largely on the camera viewpoint and facial expression. On the other hand, the second group of algorithms take skin colours as a major clue [2] in the assumption that skin colour appears similar over the sequence of video images of interest. Although being very important information, because of its variation over the change of illuminations, colour usually works as a pre-processing step for reducing the number of false segments whereas the region of interests (ROI) have to be verified based on the other information [3].

Since each method is developed based on the properties associated with the collected video images, i.e. illumination, background, foreground, etc., this method usually does not work on the other images without substantial modifications.

The aims of this work are concerned with the feasibility of existing colour spaces and models of face segmentation under controlled varying illumination levels, and the applicability on classification of ROI, with the ultimate goal of developing a colour-based face segmentation algorithm.

2. Colour Spaces and Models

In this study, six colour spaces and one colour appearance model, which are frequently applied in image segmentations, have been investigated. They are RGB, HSV, HSL, CIEXYZ, CIELAB, CIELUV, and CIECAM02.

RGB (red, green and blue) is one of the most popular colour spaces for storing digital image data as the human visual system works to some extent in a similar way, i.e. all the other colours can be made from these three primary colours. Since most colour display devices (e.g. a colour monitor, a TV) use RGB to represent a colour image, applying RGB space therefore does not need any transformation [4].

However, the term ”skin colour” is not a physical property of an object, rather a perceptual phenomenon and therefore a subjective human concept. Significant perceptual non-uniformity makes RGB not a very favourable choice for colour analysis and colour based recognition algorithms. Therefore other colour spaces are introduced, such as HSL (hue, saturation and lightness) and HSV (hue, saturation and value), which are perceptually more uniform colour spaces than is RGB. Their intuitiveness (i.e. closer to human perception) of the colour space components, explicit discrimination between luminance and chrominance properties, and simple conversion from RGB, have led to these colour spaces being popular on segmentations based on skin colour [2].

To standardise colour spaces, the International Commission on Illumination (*Commission Internationale de l'Eclairage*) has introduced several spaces for different applications. The first one was introduced in 1931 when CIE defined the CIE*XYZ* colour space (also known as the tristimulus colour space) which is obtained based on the average observer. Each colour can be represented by three values of X, Y, and Z. It was claimed that CIEXY (normalized CIEXYZ) is one of the most efficient representations for skin segmentation because the skin chromaticities occupy a smaller area in the colour space [5]. But CIEXYZ is a non-uniform space, which has led CIE to recommend two other colour spaces for the measurement of colour differences. They are CIELAB and CIELUV. Colours in the CIELAB space are perceptually more uniformly spaced than are colours in the RGB or HSV spaces [6], enabling the use of a fixed colour distance in decision making over a wide range of colours [7].

As these CIE colour spaces are defined under fixed viewing conditions, i.e. D50 or D65 with constant surroundings, they don't account for the differences in colour appearance induced by the change of viewing environment. A human vision model is therefore recommended by CIE to predict a colour appearance under a wide range of viewing conditions, coining the name of colour appearance model and being built on colour vision theories. CIECAM02 can predict a colour as accurately as an average observer, which has been applied for segmentation of facial images captured under low level and varying illumination conditions [8] and traffic sign recognition [9].

CIECAM02 takes into account the tristimulus values (*X*, *Y*, and *Z*) of a stimulus, its background, its surround, the adapting stimulus, the luminance level, and other factors such as cognitive discounting of the illuminant. The output of colour appearance models includes mathematical correlates for perceptual attributes that are brightness (Q), lightness (J), colourfulness (M), chroma (C), saturation (s), and hue (h), which are calculated in the following formulas.

$$J = 100(\frac{A}{A_w})^{CZ}$$

$$C = t^{0.9}\left(\frac{J}{100}\right)^{0.5}(1.64 - 0.29^{n})^{0.73} \qquad (1)$$

$$h = tan^{-1}(\frac{b}{a})$$

$$M = C\,F_L^{0.25}$$

$$Q = \frac{4}{c}\sqrt{\frac{J}{100}}(A_w + 4)F_L^{0.25}$$

where

$$A = (2L'_a + M'_a + \frac{S'_a}{20} - 0.305)N_{bb}$$

$$s = 100\sqrt{\frac{M}{Q}} \qquad (2)$$

$$a = R'_a - \frac{12G'_a}{11} + \frac{B'_a}{11}$$

$$b = \left(\frac{1}{9}\right)(R'_a + G'_a - 2B'_a)$$

and R'_a, G'_a, B'_a are the post adaptation cone responses and A_W is the A value for reference white. Constants N_{bb}, N_{cb} are calculated as

$$N_{bb} = N_{cb} = 0.725(\frac{1}{n})^{0.2} \qquad (3)$$

where $n = \frac{Y_b}{Y_W}$, the *Y* values for the stimulus and reference white, respectively.

In this study all aforementioned (n=7) colour spaces are evaluated to model skin colour under varying lighting conditions.

3. Method

3.1 Image Collection

A group of 3 subjects have been employed to participate in the experiments. Each subject was videoed with varying pose positions, changing illumination levels, and altering distances between the camera and the subject. Illumination levels are set to be 6.4 cd/m^2 (bright as shown in Fig. 1.1), 1.4 cd/m^2 (dim as shown in Fig. 1.2), and 2.4 cd/m^2 (medium shown in Fig. 1.3). Under each illuminating condition three frames sequences (n=90) were captured (Fig. 1.a, 1.b, 1.c). The distance between the subject's head and the camera is equal to 1.08m. Before shooting, the cameras were calibrated as detailed in [10].

The position of a subject's head on an image is similar to the head position during a PET scanning procedure [8] as the main goal of the study is to monitor patent's movement while undergoing a Positron Emission Tomography (PET) scan. Video images have been captured using two off-the-shelf digital cameras, a Fujifilm and Sony. In this work the videos taken using the Fujifilm camera are applied as training datasets.

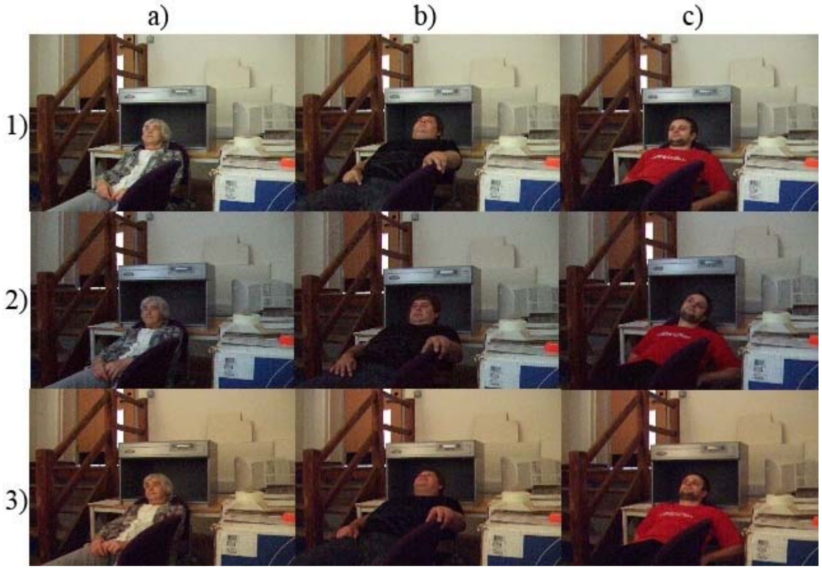

Figure 1. An example of training video frames.

3.2 Colour Elements for Skin Modelling

A facial area on each frame of the training video sequences is marked manually. Frames are then represented using seven colour models/spaces, as described above.

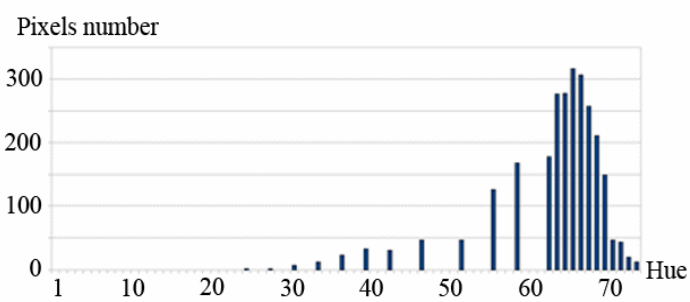

Figure 2. An example Hue (HLS) histogram in the face area.

The histogram of each colour element for each colour space/model is then calculated in the facial area and is illustrated in Figure 2. The maximum peak of the histogram is chosen as a skin colour attribute. Then the average value of each skin colour attribute is calculated independently for three training video sequences taken under varying viewing conditions. Five colour attributes with the minimum variations between skin colour attribute values under different lighting condition are then chosen to model skin colours, which are given in Table 1.

Table 1. Five skin colour attributes for CIECAM02 under three illuminations.

Illumination level / Colour parameter (space)	1 **(6.4 cd/m$^{2)}$**	2 **(1.4 cd/m^2)**	3 **(2.4 cd/m^2)**
Hue (HLS)	10.8±2.6	6.2±1.8	7.3±1.5
Achromatic response (CIECAM02)	26.5±3	23±2.7	24.4±2.9
Chroma (CIECAM02)	16±2.3	14.4±2.3	23.2±2.7
Colourfulness (CIECAM02)	11.9±1.7	10.6±1.6	16.7±2
Saturation (CIECAM02)	28.1±3.3	27.2±3.4	34.1±3.8

3.3 ROI Verification

Figure 3 shows the steps towards the segmentation of faces. After colour segmentation, a detected ROI needs to be verified (Fig 3.1, 3.2). As obtained in the experiment, a useful clue for verification is colour edges. In our work the Sobel operator is used to detect colour edges on all the frames that is demonstrated in Figure 3.3. By studying the edge density in facial areas, the empirical result is that an area should be classified as a face if:

$$0.15<D<0.5$$

where D is the edges density of brightness (CIECAM02).

3.4 Facial Image Segmentation

An algorithm for facial image segmentation based on mixed colour space has since been developed and evaluated, i.e. the mixed colour spaces are used in order to detect a skin colour pixel. The pixel is classified as skin if:

$$3 < H < 13 \text{ and}$$
$$18 < A < 28 \text{ and}$$
$$15 < M\text{-}S < 18 \text{ and}$$
$$10 < C\text{-}S < 13,$$

where H is Hue (HLS), A is Achromatic response (CIECAM02) as given in Eq. (2), C is Chroma (CIECAM02), M is Colourfulness (CIECAM02), and S is Saturation (CIECAM02). After the classification of skin colour pixels (Fig. 3.1), the marked pixels are grouped together into regions of interest using the nearest neighbourhood method (Fig. 3.2), arriving at a segmented ROI. Based on the colour information on colour edge densities, an ROI is classified as either a face or not (Fig. 3.4).

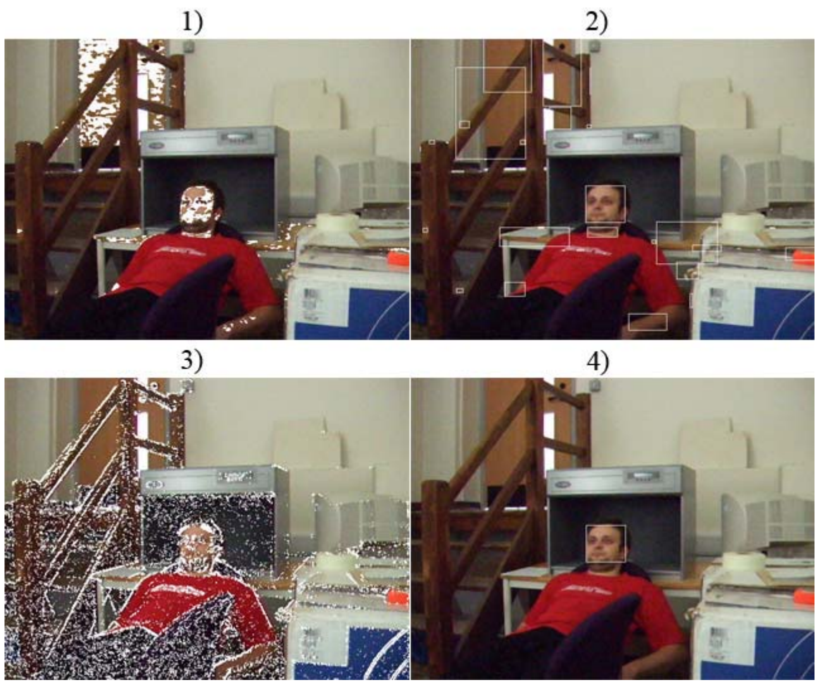

Figure 3. Procedures in segmenting: 1 – detected skin colour pixels, 2 – segmented areas, 3 – detected colour edges, 4 – verified areas.

In order to evaluate the developed algorithm for face segmentation two video sequences with unknown varying illumination levels and head pose have been collected using a web-camera as shown in Fig. 4.1 and the Fujifilm camera shown in Fig. 4.2. The face colour, head pose, and background are totally different from those in the training image set. Table 2 gives the result of the computer simulation of facial segmentation, where false positive regions include the segmented regions without any face content. On the other hand, false negative regions show the regions that are segmented that are not face regions. The results show that after verification using colour edge information, the face has been correctly identified, i.e. segmented with 0% false negative region for all the video images even though 11% of regions in Figure 4.2 have been falsely segmented.

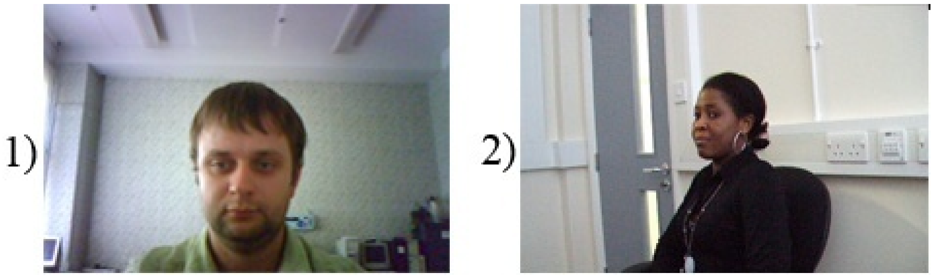

Figure 4. An example of testing video frames.

Table 2. List of results on the segmentation of faces.

Example of video / Result	**Fig 1.1**	**Fig 1.2**	**Fig 1.3**	**Fig 4.1**	**Fig 4.2**
Segmented (%)	100	100	100	100	100
False positive regions	240	206.7	302	23	3.3
False positive regions before verifying	3781	3059	3565	1524	440
False negative region (%)	0	0	0	0	11
False negative region before verifying (%)	0	0	0	0	0

4. Conclusion

In our study a new approach to perform face image segmentation based on skin colour is presented. A group of colour elements for modelling of skin pixels is chosen as training datasets based on three video sequences captured under certain lighting conditions. It is shown that based on the chosen attributes a face image can be segmented with very few false negative regions. Also a new approach to classify an ROI in images such as a face based on colour edge densities is presented.
Future work includes improvements in the described algorithm by adding a step to perform ROI classification more precisely. The developed software also needs to be optimized in order to perform real time segmentation of video frames. At present, the average time to process an image of size 640x480 is 3.24 sec when performed on a laptop with an Intel(R) Core(TM)2 Duo CPU at 2.2 GHz.

Acknowledgements

The work is supported in part by the Russian Foundation for Humanities N 09-06-95218 a/F and the BRIDGE Research Cooperation Award from the British Council, UK.

References

[1] Stan Z. Li & Anil K. Jain, Handbook of face recognition, ISBN 0-387-40595-X.

[2] Vezhnevets V., Sazonov V., Andreeva A., A Survey on Pixel-Based Skin Color Detection Techniques. Proc. Graphicon-2003, pp. 85-92, Moscow, Russia, Sept. 2003.

[3] R. Srikantaswamy, R. D. Sudhaker Samuel, A Novel Face Segmentation Algorithm from a Video Sequence for Real-Time Face Recognition, EURASIP Journal on Advances in Signal Processing, Volume 2007, Article ID 51648, doi:10.1155/2007/51648.

[4] Murad Al Haj, Andrew D. Bagdanov, Jordi Gonz`alez, and Xavier F. Roca. Robust and Efficient Multipose Face Detection Using Skin Color Segmentation. H. Araujo et al. (Eds.): IbPRIA 2009, LNCS 5524, pp152–159, 2009.

[5] J.C. Terrillon, Y. Niwa, and K. Yamamoto. On the selection of an efficient chrominance space for skin color-based image segmentation with an application to face detection. In Proceedings of International Conference on Quality Control by Artificial Vision, 2:409-414, 2001.

[6] J. Cai, A. Goshtasby. Detecting human faces in color images. Image and Vision Computing 18, pp63–75, 1999.

[7] Zarit, B. D., Super, B. J., and Quek, F. K. H. Comparison of five color models in skin pixel classification. ICCV'99 Int'l Workshop on recognition, analysis and tracking of faces and gestures in Real-Time systems, pp58–63, 1999.

[8] X.W. Gao, S. Anishenko, D. Shaposhnikov, L. Podladchikova, S. Batty, and J. Clark. "High-precision Detection of Facial Landmarks to Estimate Head Motions Based on Vision Models", Journal of Computer Science, v. 3, no. 7, pp.528-532, 2007.

[9] Gao X., Hong K., Podladchikova L., Shaposhnikov D. and Passmore P. Colour Appearance Based Approaches for Segmentation of Traffic Signs. // EURASIP Journal on Image and Video Processing. – 2008. - v.2008. - Article ID 386705. - doi:10.1155/2008/386705, - 7 pages.

[10] Anishenko S., Osinov V., Shaposhnikov D., Podladchikova L., Comley R., Sukholentsev K., Gao X. A Motion Correction System for Brain Tomography Based on Biologically Motivated Models. Proc. 7-th IEEE Int. Conf. On Cybernetic Intelligent Systems, pp. 32-36, London, UK, Sept. 2008.