Skin Locus Based Skin Detection for Gesture Recognition

Muhammad Raza Ali[1]
alim@cs.man.ac.uk

Tim Morris[1]
http://www.cs.man.ac.uk/~tmorris

[1] Advanced Interface Group,
School of Computing,
University of Manchester, UK

Abstract

The paper describes ongoing research in computer vision based gestural interfaces. The aim of the research is to develop a novel, multi-user gesture recognition system that addresses key factors of human computer interaction; the latter is not fully addressed in computer vision research. The focus of current work has been on efficient and accurate skin colour detection.

Skin colour based segmentation for computer vision applications involving human users has received considerable attention over the years. One of the popular techniques is physics based skin colour locus in chromaticity space. The skin distribution in chromaticity space has shown robustness to changing illumination conditions. The paper proposes a technique based on skin samples for identifying skin locus, this does not require knowledge of camera parameters and correlated colour temperature of the illuminant. The paper also evaluates the potential for using skin locus based segmentation for an indoor gesture recognition system; previous studies have mainly focussed on face detection and face tracking.

1 Introduction

The increasingly diverse range of computerized systems has motivated researchers to develop novel ways of interaction. One of the active areas of research is computer vision based interfaces. The research is motivated by various advantages of computer vision based interfaces. These advanced interfaces offer a very natural and non-cumbersome interaction. Computer vision research in some application areas is well established such as teaching aids e.g. the concept of smart classroom [7] and multimedia presentations [3]. We are interested in employing vision based interfaces for much more advanced, real life applications with high cognitive workload e.g. a traffic monitoring control room. However, a new direction in computer vision based HCI research needs to be identified. Researchers in this area have focussed primarily on the accuracy of computer vision techniques and overlooked the need for addressing human factors when developing novel interaction techniques. Some studies in the last few years [6, 16] have pointed out this lack of focus and employed usability evaluation. However, if vision based interaction is to replace traditional methods for advanced, critical applications a more elaborate framework is required. This is the focus of our work in progress; developing a novel gesture recognition system where elaborate usability studies are part of the development process.

This will address key human computer interaction issues as part of the development process. The focus of current work has been on skin colour detection and using skin colour for background subtraction. Before we discuss the current work it would be useful to give a brief overview of the proposed system. Also, section 1.2 gives a brief background of skin colour based detection.

1.1 Overview of Research

A block diagram in figure1 gives an overview of our gesture recognition system. We have termed it the Gesture Recognition Module. Efficient skin detection is to be combined with optical flow in a particle filtering framework for a novel tracking mechanism based on skin similarity measure and optical flow magnitude. The tracking mechanism will be able to track multiple candidate (skin) regions and select the probable region for gesture classification. The system will support classifications of both static and dynamic gestures.

Traditionally, evaluation of a gesture recognition system similar to the one shown below is focussed on the accuracy and error rate of the proposed technique. The usability factors are mostly not taken into consideration. We plan to address this issue by focussing on two aspects of human computer interaction: a) at all key development stages we will have usability evaluation using standard and established methods e.g. cognitive walkthrough and Fitt's law, and b) measuring the cognitive workload experienced while using a vision based interface. This will help in creating a balanced evaluation framework for vision based interfaces.

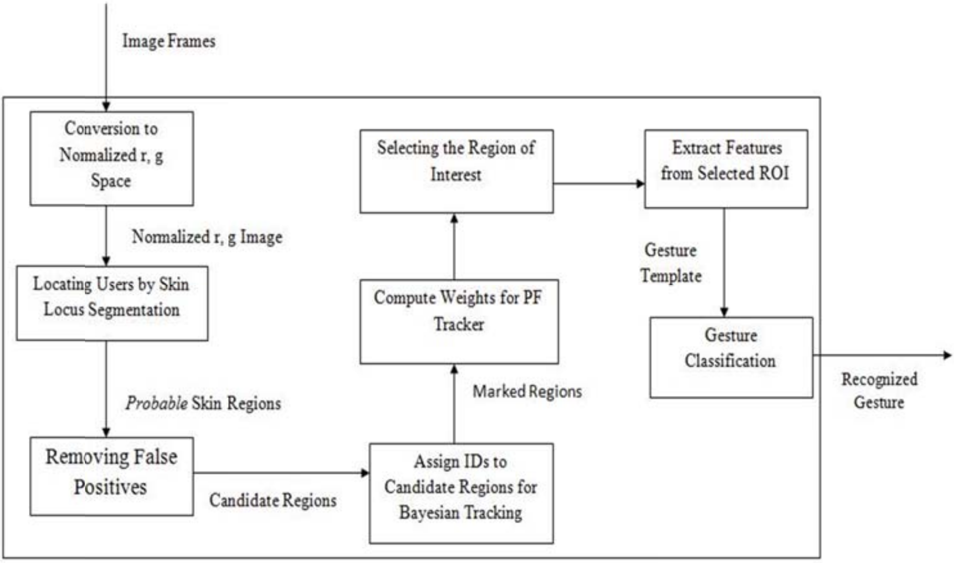


Figure 1: Gesture Recognition Module

1.2 Skin Colour Based Detection

Wide ranging computer vision applications involve human users either directly (human-computer interaction, human-robot interaction) or indirectly (surveillance). Skin colour is often used for detection and segmentation. Various techniques have been developed based on skin colour that aim to address *colour constancy*; a key problem in computer vision. Colour constancy is the apparent constant colour of an object under varying illumination conditions. Colour constancy has received considerable attention not only from computer vision researchers [1, 21] but researchers have also tried to understand colour constancy from physics [8] and optics points of view.

Skin colour based segmentation techniques can be broadly classified into four categories: a). Adaptive techniques [11] adapt to changing illumination conditions by adjusting a predefined threshold, b). Statistical colour models [13]: two popular approaches in this category are histogram colour models and Gaussian mixture models, c). Colour Correction Model: colour correction is performed in a colour space that explicitly separates luminance and chrominance. The aim is to reduce the dependency on the luminance component [19] and d). Skin Colour Locus Technique: this technique is based on the distribution of skin colour in chromaticity space. Once the skin distribution is determined, thresholds are set which classify an incoming pixel as skin or non-skin.

1.2.1 Chromaticity Space

The camera response to reflected light reduces the continuous light spectrum to a three dimensional space i.e. RGB. The RGB defines the *colour space*. Pascale [4] gives an excellent review of colour spaces. A colour signal comprises of luminance (intensity or value) and chrominance (colour information). It is often desirable to represent the colour information independent of intensity in a 2-dimensionsal space called a *chromaticity space*. Chromaticity space is advantageous not only in terms of dimensionality [21] but is more robust to illumination changes. Any colour space can be converted to a chromaticity space, and there are many options, for this study we use normalized red, green or [r, g] chromaticity space. The Red, Green and Blue channels are normalized as follows to remove the luminance component:

$$r = \frac{R}{R+G+B} \qquad\qquad g = \frac{G}{R+G+B} \qquad\qquad (1)$$

Choice of normalized [r, g] space can be justified by its accuracy in face detection [10]; it is second only to TSL (Tint, Saturation, Luminance). However, in terms of computational efficiency it has the highest throughput of any chromaticity space followed by normalized HS (hue-saturation) which achieves 66% throughput of the former [22].

1.2.2 Skin Colour Locus Technique

Skin locus depends upon of various factors; temperature of the light source, camera characteristics and skin reflectance. Previous studies have used representative conditions for determining the skin colour locus; however these studies have mainly concentrated on the physics of the skin distribution[15], and response of different cameras to similar conditions [2]. Soriano et. al. [14] is one of the very few studies that comprehensively evaluate skin locus based segmentation for tracking (faces only). They determined skin locus by taking pictures under four illumination conditions with known illuminant temperatures and camera calibrated for each condition.

This work describes a simple and effective method for determining the skin locus without the need for elaborate knowledge of camera parameters and CCT of light sources. The technique involves determining skin colour locus by skin samples. This work differs from previous related studies [14, 15] as skin locus is determined under everyday indoor illumination conditions. The motivation is an important characteristic of skin colour; its distribution occupies only a certain part of the chromaticity space. The simple technique described in this paper eradicates the need for specialized settings (dark room imaging, reflectors etc.) and tedious camera calibrations. We have identified thresholds that can be reliably used for an indoor computer vision application and can be easily fine-tuned to adjust to any changes. The results of evaluation using sign language and face recognition datasets show that the proposed technique can be used for skin detection for hand and facial gesture recognition. This paper also reports detailed skin detection results, for the skin locus technique on web images for the first time. The major drawback of this approach is that any background objects with colour similar to skin will be recognized as skin. We have shown that by combining optical flow with skin colour information false positives can be considerably reduced leading to accurate background subtraction.

2. Determining the Skin Colour Locus

This section outlines the details of our setup for acquiring images and describes the methodology for determining the skin locus.

2.1 Gestures & Image Acquisition

Three participants took part in the study, one each of Caucasian, East Asian and South Asian origin. Images of each participant were taken making three types of gestures (pointing, thumbs-up and stopping). Two other images were taken showing face and hands in different formations i.e. front and back of hands, and palm of hand with back of fist. In total 125 images were acquired and around 600 skin samples were manually taken from these images. The images are of size 640x480 pixels. This resolution is a good trade-off between computational efficiency and accurate skin detection. Images with small dimensions lead to poor skin detection (see 3.1). The size can be reduced for applications involving stationary users as we will see in section 4. The images were acquired at four different indoor locations over a period of five days. Two locations had fluorescent while two locations had incandescent lighting. For one location with each illumination type there was also contribution from natural light. Two cameras were used in the study; Samsung NV3 and the Canon EOS 400D. Some of the gestures are shown in figure 2.

Figure 2: Examples of Various Gestures

2.2 Skin Colour Distribution in Normalized [r, g] Space

Skin samples were taken from the acquired images and the skin colour distribution was determined. Samples were taken from the hands and face (forehead, middle part of the face, chin area and for some images from the cheek area as well). Skin distribution can vary due to individual characteristics of the body parts so it is important to consider these distributions.

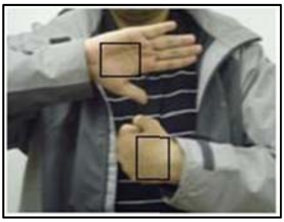


Figure 3: Skin Samples

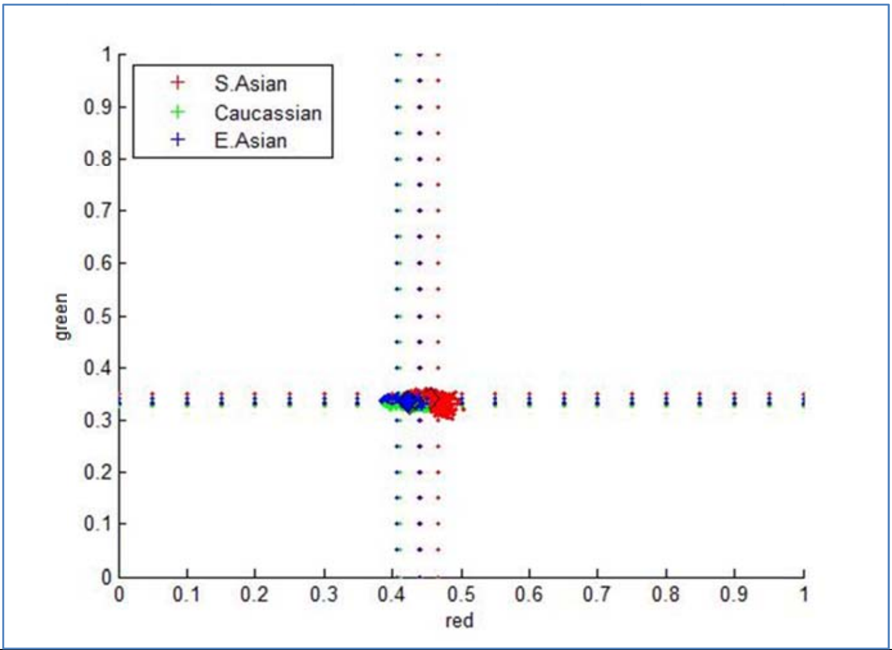
Scatter plots were generated from the extracted skin samples. Two types of experiments were conducted; to compare the skin colour distributions across skin types and compare the response of different cameras (un-calibrated) under similar conditions. Each image was taken under more than one white-balance settings. The first experiment was aimed at comparing the skin colour distribution for various ethnicities. An example scatter plot of the forehead region is shown in figure 4.These images were taken with the Samsung NV3 digital camera under fluorescent light with the auto-white balance feature enabled. The camera and illumination details are for information only. As mentioned above, camera and illuminant details do not determine the choice of thresholds. Figure 4 shows a significant overlap of skin colour distribution between ethnicities. Diamonds indicate the mean $\mu(r, g)$ values and the dotted lines indicate one standard deviation σ, for red and green chromaticity values. Despite slight difference in mean values there is a significant overlap of [r, g] values for different ethnicities. This trend was also observed for skin samples taken from participants' hands and other parts of the face for all conditions. For each image only three to five samples are needed for determining skin colour distribution for that image.

The second experiment involved comparing images from different cameras. This was important as we were interested in specifying credible threshold values that require little adjustment. The results of the experiment showed significant overlap between values in chromaticity space making it possible to specify thresholds. Even pictures that were taken at different locations under more or less similar conditions showed significant overlap.

More than 200 scatter plots were generated to compare skin distributions under different conditions and different camera settings. It is worth mentioning that procedure was adopted to enable us in specifying general thresholds for indoor applications requiring minimal adjustments. Otherwise, for a specific application setting only a handful of images and samples are sufficient for identifying the skin locus for those particular conditions.

2.3 Specifying the Thresholds

Chromaticity values do vary with change in illumination and camera settings but our experiments showed it is possible to define threshold values for an indoor gesture recognition system. Some gesture recognition systems may require zooming-in on the user or zooming-out; we have found that changing camera parameters such as focal length only slightly effect the distribution and new values still fall in the desired range. Taking advantage of significant overlap, thresholds for red and green chromaticity values were defined; the desired range for red lies between 0.40 and 0.65 inclusive, while green lies between 0.25 and 0.35 inclusive. Referring to the skin-locus identified by some important studies [15, 16] the upper red threshold, and green chromaticity ranges identified are *almost* the same as those identified in this study. Therefore, during evaluation in addition to skin detection results, we were also interested in investigating whether by varying the lower red threshold only; we can adjust to varying conditions.



Statistics:	S. Asian: μ=(0.4540, 0.3392) σ =(0.0133, 0.0090) Caucasian: μ=(0.4246, 0.3307) σ=(0.0135, 0.0045) E. Asian: μ=(0.4291, 0.3359) σ =(0.0150, 0.0055)

Figure 4: Scatter plot for the forehead region for three different skin types, with mean and standard deviation of [r, g] values.

3. Evaluation

The accuracy of skin detection was evaluated using two established sign language datasets. Sign language datasets provide good benchmarking data as they contain a variety of gestures and lighting conditions. Two face datasets and two datasets containing web images were also used in evaluation. No information regarding camera parameters or illumination conditions is available for these datasets.

The datasets used in the evaluation are: a). University of Athens Dataset. This dataset [9] contains images of 4 persons performing gestures of American Sign Language (ASL), b) DCU Gesture Image Dataset prepared by Coogan and Sutherland [20] contains 1120 images of a person for two lighting conditions (560 for each). Twenty occurrences for each of the 28 hand shapes were acquired from the user during four sessions over a period of 2 days, c) GTAV Face Database is a relatively new but very comprehensive database [5] that contains almost 1200 images of 44 subjects, around 27 images per person, d) UCD Face Database is a small database that contains 94 images that have been collected from various sources (available at http://ee.ucd.ie/validdb/datasets.html), e) CRL database [13] is a very comprehensive database of 4500 web images containing skin and the corresponding skin masks and f) IBTD dataset [18] containing 555 web images.

The skin detection results for sign language and face datasets are shown in Table 1. The segmentation was performed using the thresholds suggested in the previous section i.e. red[0.40, 0.65] and green[0.25, 0.35].

Dataset	Correct Detection Rate	False Positive Rate
University of Athens	75%	9%
DCU Gesture Image	96%	5%
GTAV Face Database	88%	11%
UCD Face Database	77%	1%

Table 1: Skin Detection Results

For University of Athens dataset the detection rate was over 80% for three subjects. However, for one subject many pixels were misclassified as non-skin (false negatives) with correct detection rate around 65%. Skin samples for the said subject were taken manually and the lower red threshold was adjusted to 0.38 raising the correct detection rate to around 85%. Good segmentation for web images in the UCD database is quite interesting considering the wide variety in sources and types of these images. To verify these results the proposed technique was evaluated on two datasets containing web images.

False positives for this technique depend upon the nature of background and other objects such as clothes, hair etc. Similar colour objects are likely to be detected as skin. In section 5 we discuss removing false positives by combining skin colour and optical flow information.

Dataset	Correct Detection Rate	False Positive Rate
CRL	72%	16%
IBTD	89%	21%

Table 2: Skin Detection Results for Web Images

3.1 Skin Detection in Web Images

The correct detection rate for our thresholds is low for the CRL dataset. It was noted that skin segmentation results were generally quite poor for images with smaller dimensions e.g. less than 100x80 pixels. However, by adjusting lower red threshold to 0.35 we increased the range of possible red chromaticity values. We achieved an improved correct detection rate of 89.5%. For the IBTD dataset, skin detection results with original thresholds were very encouraging: correct detection rate is 89% while the false positive rate is 21%. It is worth mentioning that very few images of IBTD dataset are of small dimensions. Correct detection rates for web images make skin-locus based segmentation worth investigating for applications such as objectionable image filtering [18]. The aforementioned databases can be useful for training a classifier for this purpose. Previous studies [2, 14] do not recommend using skin locus segmentation for web images due to the numerous sources from which these image are acquired but our results show that if image size is not too small good segmentation results can be achieved. As shown above, by adjusting thresholds we managed to improve results for CRL dataset which includes images of small dimension.

3.2 Comparison of Results

Skin detection results for the datasets listed in table 1 cannot be compared directly as these datasets are primarily used for face detection and sign language recognition. Studies using these datasets report face detection or gesture recognition results only, with skin detection as an intermediary step. We feel the need for a standardized skin detection dataset (prepared from gesture recognition point of view) similar to the KTH action recognition dataset, almost all studies in action recognition report results on the dataset. Results on web images can be compared as many studies have used CRL dataset. This survey [23] reports high detection rates on CRL dataset i.e. more than 93%. Our results for IBTD dataset are comparable with 90.04% detection rates reported by [12]. Despite these results, compared to the other techniques, computational efficiency of skin locus based skin detection makes it worth investigating for web based applications. These results can be improved by combining skin colour with other information e.g. text as suggested by [13].
Previous studies have not discussed any skin detection results on web images for skin locus technique.

4. Adjusting Thresholds

The ability to adapt to changing conditions is always desirable for skin segmentation. Automatic adjustment of thresholds can be easily accomplished while the system is online. Thresholds can be adjusted by following a simple two step procedure. We consider sequences from the German Finger Spelling Database [17]. This database contains sequences from two cameras at different angles, recorded on two different days under non-uniform natural light only. Settings differ from our training conditions especially for sequences recorded under bright natural light. In our experiments sunlight did not fall directly on the participant. This change can be adjusted to very easily. As in other datasets, no information is available regarding camera parameters. The first step is image or frame differencing between the background image and the image with skin followed by simple thresholding. The output of this stage is a skin segmented image, an example is shown in figure 5.

Frame 1 Frame 5 Segmented Image

Figure 5: Skin segmentation step

The second step simply computes the normalized red and green chromaticity values for the segmented region. Even if the segmentation is not perfect it would be good enough for computing the mean chromaticity values based on which the thresholds can be varied to adjust to new conditions. This process is efficient and can be included as a feature in a gesture recognition system. In exploiting the similarity of the skin colour distribution for various skin types only five videos (for bright daylight conditions) sequences were selected. Skin distributions for other conditions fell within our original thresholds. Based on the mean values of segmented regions, the lower red threshold only was adjusted i.e. lowered to 0.385. It is worth mentioning that this method can be used to update all thresholds adaptively (e.g. as in [19]) instead of using fixed thresholds, but our experiments have shown that just varying lower red threshold will suffice. Comparison of segmentation results can be seen in figure 6. Size of the acquired image can be reduced to 320x240 pixels for a desktop setting similar to this dataset.

(a) (b) (c)

Figure 6: (a) Input Image (b) Segmentation using original thresholds, red [0.40 0.65] green [0.25 0.35]. (c) Segmentation using updated thresholds, r [0.385 0.65] g [0.25 0.35].

5. Removing False Positives

All skin detection techniques suffer from the problem of detecting false positives. In addition to background objects, hair and clothes can also contribute to false positives. A trade-off between false positives and true positives is crucial and is application dependent. As our tracking mechanism will track skin regions, accurate skin detection is vital. During our experiments we noted that in an indoor setting sometimes large clusters of false positives are detected. These regions lead to erroneous background subtraction and cannot be reliably used for tracking and classification. To remove false positives we have combined skin colour information with optical flow magnitude. Background subtraction based on skin colour and optical flow has shown improved background subtraction results. Although the false positives are not removed completely, the size of these regions is now reduced greatly and they can be easily ignored as they are smaller than hands and face. Therefore, we have true skin regions as candidates for our tracker. Detailed experimental results will be reported elsewhere.

6. Conclusion

We have described current/completed work in our research in vision based HCI. The paper describes a simple yet effective technique for defining skin locus thresholds. The thresholds defined can be reliably used for an indoor gesture recognition system. We used only three skin types for training as consistency of skin colour distribution across all ethnicities is well established and was observed during evaluation. This shows the advantage of using skin colour information in applications involving human users. Despite our small training sample high detection rates were reported on datasets containing wider variety of skin types. We have shown that only one out of four thresholds mentioned in the paper needs to vary for the skin segmentation system to adapt. Combination of skin colour and optical flow has shown encouraging results for removing false positives. The authors plan to acquire more images to prepare a specialized dataset for evaluating skin detection techniques.

References:

[1] B. Funt, K. Barnard, and L. Martin. Is Machine Colour Constancy Good Enough? *In Proceedings of the 5th European Conference on Computer Vision ECCV'05*, pp: 445-459, 1998.

[2] B. Martinkauppi , M. Soriano and M. Laaksonen. Behaviour of skin color under varying illumination seen by different cameras in different color spaces. *In Proceedings of Machine Vision in Industrial Inspection*, SPIE Vol. 4301, 102-113, San Jose California, January, pp: 19-26 2001.

[3] C. Hardenberg and F. Bérard. Bare-Hand Human-Computer Interaction. *In Proceedings of the ACM workshop on Perceptive user interfaces*, pp: 1 – 8, 2001.

[4] D. Pascale. A Review of RGB Color Spaces from xyY to R'G'B', Published by the Babel Color Company. 5700 Hector Desloges, Montreal, Canada H1T 3Z6, 2003.

[5] F.Tarrés, A. Rama, "GTAV Face Database" available at: http://gps-tsc.upc.es/GTAV/ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm

[6] G. Iannizzotto, F.L. Rosa, C. Costanzo, and P. Lanzafame. A multimodal perceptual user interface for video-surveillance environments," *In Proceedings of International Conference of Multimodal Interfaces (ICMI'05)*, pp: 45-52, 2005.

[7] H. Ren and G. Xu. Human action recognition in smart classroom", *In the Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp: 417-422, 2000.

[8] H.Y. Chong, S.J. Gortler, and T.E. Zickler. The von Kries Hypothesis and a basis for colour constancy, *In Proceedings of 11th IEEE International Conference on Computer Vision*, 2007.

[9] I. Askaroglou , S. Tzikopoulos , D. Kosmopoulos and S. Theodoridis. Extraction of mid-level semantics from gesture videos using a Bayesian network. *1st Panhellenic Students Conference on Informatics*, Patras, Greece, 2007

[10] J.C. Terrillon and S. Akamatsu. Comparative performance of different chrominance spaces for colour segmentation and detection of human faces in complex scene images. *In Proceedings of the12th Conf. on Vision Interface*, pp: 180-187, 1999.

[11] K.M. Cho, J.H. Jang, and K.S. Hong. Adaptive skin-colour filter. *Pattern Recognition*, 34(5), pp: 1067-1073, 2003.

[12] M. Abdullah-Al-Wadud, O. Chae. Skin segmentation using colour distance map and water flow property, *The Fourth International Conference on Information Assurance and Security*, pp.83-88, 2008.

[13] M.J. Jones and J.M. Rehg. Statistical colour models with application to skin detection. *International Journal of Computer Vision,* 46(1), pp: 81-96, 2002.

[14] M. Soriano, B. Martinkauppi, S. Huovinen and M. Laaksonen. Adaptive skin colour modelling using the skin locus for selecting training pixels, *Pattern Recognition,*36(3), pp: 681-690, 2003.

[15] M. Storring, H.J. Andersen, and E. Granum. Skin colour detection under changing lighting conditions. *In Proceedings of Seventh Symposium Intelligent Robotics Systems*, pp. 187-195, 1999.

[16] N. Stefanov, A. Galata, R. Hubbold, Real-time hand tracking with variable-length Markov models of behaviour, *In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, 2005.

[17] P. Dreuw, T. Deselaers, D. Keysers, and H. Ney. Modelling image variability in appearance-based gesture recognition. *In ECCV Workshop on Statistical Methods in Multi-Image and Video Processing (ECCV-SMVP)*, pp: 7-18, Graz, Austria, May 2006.

[18] Q. Zhu, C.T. Wu, K.T Cheng and Y.L. Wu. An adaptive skin model and its application to objectionable image filtering. *ACM International Conference on Multimedia 2004 (ACM MM 2004)*, October 10-16, New York, USA, 2004.

[19] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in colour images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), pp: 696-706, 2002.

[20] T. Coogan, G. Awad, J. Han, and A. Sutherland. Real time hand gesture recognition including hand segmentation and tracking, *ISVC06*, pp: 495 – 504, 2006.

[21] V. Agarwal, A.V. Gribok, and M.A. Abidi. Machine learning approach to colour constancy. *Neural Networks Letter* , vol. 20, pp. 559-563, 2007.

[22] V. Chauhan and T. Morris. Facial feature tracking for cursor control. 12th SCIA, Bergen Norway, June 14-16, 2001.

[23] V. Vezhnevets, V. Sazonov, A. Andreeva. A survey on pixel-based skin colour detection techniques**,** GRAPHICON03, pp. 85-92, 2003.