Hand Gesture Interface Based on Improved Adaptive Hand Area Detection and Contour Signature

Lei Gu, Xiaoyang Yuan and Takeshi Ikenaga
Graduate School of Information, Production, System
Waseda University
Kitakyushu, Japan
korai@ruri.waseda.jp

Abstract—**HMD (head-mounted display) as a promising device is becoming more and more important in daily life. Many companies has been working on it for the next generation human-interface system. This paper presents a real-time hand gesture interface based on TSL (Hue, Saturation, Luminance) adaptive area detection and distance signature with single camera. First, apply self-adaptive skin color detection in TSL color space where skin color data can be clustered to segment hand area. Second, acquire the distance signatures from hand shape contours and obtain possible finger points which reduce the hand gesture recognition problem into finding peaks of one dimensional signature. Last, finger points are labeled by the information of signature. ROC (Receiver Operating Characteristic) Analysis shows the proposed hand area detection method always gives a result in feasible area (TPR>0.91, FPR<0.1) which is suitable for the following contour analysis, indicating that it's more stable and robust compared with other skin color based methods. The evaluation results show the potential of real-time on PC at around 10 fps.**

Keywords: hand area detection; adpaptive skin area detection; distance based gesture detection;

I. Introduction

In recent years, a new kind of important device HMD (Head-mounted display) which is derived from the virtual reality is showing growing importance and enlarging potential. The device should be a pair of glasses worn by a user which can show both the vision of computer-generated virtual image and real world behind the image. It provides with an easy and natural way of human computer interface that augments human's perception of reality and make the information about the surrounding real world of the user become interactive and digitally manipulable.

To put HMD into practical application of military business or video games purpose, hand gesture interface is one of critical technologies. This interface should track hand movement of a person and determine what gestures they are performing. To achieve this, a single camera as a kind of input device was thought to be not as effective as other types of devices such as stereo or depth-aware cameras in earlier years. But now, developers are challenging to enable hand gesture interface with single camera by improvement of algorithms or theories, since it doesn't need an expensive piece of hardware.

Analyzing gestures mainly has two approaches, 3D model-based approaches and 2D appearance based approaches. The 3D model-based approach need 3D spatial description of the hand and the appearance-based approach only needs 2D visual images on appearance of hands [1]. Since the input device is limited to single camera, appearance –based approaches should be used. Generally, the image processing of appearance-based hand gesture interface includes two main steps: first obtain the hand area silhouette and the hand contour, second calculate the finger points from the contours data. The accuracy and quality of every step will directly affects the final result of the whole interface.

Hand area searching can be reduced through the detection of skin color [2]. The simplest method is to define fixed decision boundaries in different components of a certain color space [3]. Decision rule is that any pixel values fall in boundaries defined will count as skin color otherwise not. Although this method has the merit of very low computations, it suffers from the problem of low accuracy due to the following reasons. 1) The distribution of skin color points should be similar to a multivariate Gaussian distribution [4] which has a smooth curved boundary. The fact that skin color distribution doesn't fit the model of boundary box in color space leads to the result of introduction of unexpected noised. 2) Different people will have different skin color distribution, which requires the detection system dynamically modify the decision rule to fit people.

Some adaptive methods [5, 6] have been proposed which to some extent, improve the detection system. However, some problem still exists if it's applied to HMD system. 1) In some detection systems, sampling is done by humans [6]. The skin color should self adaptive detected which can automatically modify the decision rule and to the detection without manual sampling. 2) Accuracy still needs improvement. New and different methods can be used to improve the accuracy. 2) Computation complexity should be controlled to meet the need of real-time. Too much calculation will make the hardware cost too high and is not suitable for the HMD application.

II. Proposed Hand Gesture Interface

This paper presents a method which consists of two main modules, improved TSL self-adaptive hand area detection and contour signature based finger detection.

978-1-4673-5082-2 ©2012 IEEE

Figure 1 is the flowchart of the proposed hand gesture interface system.

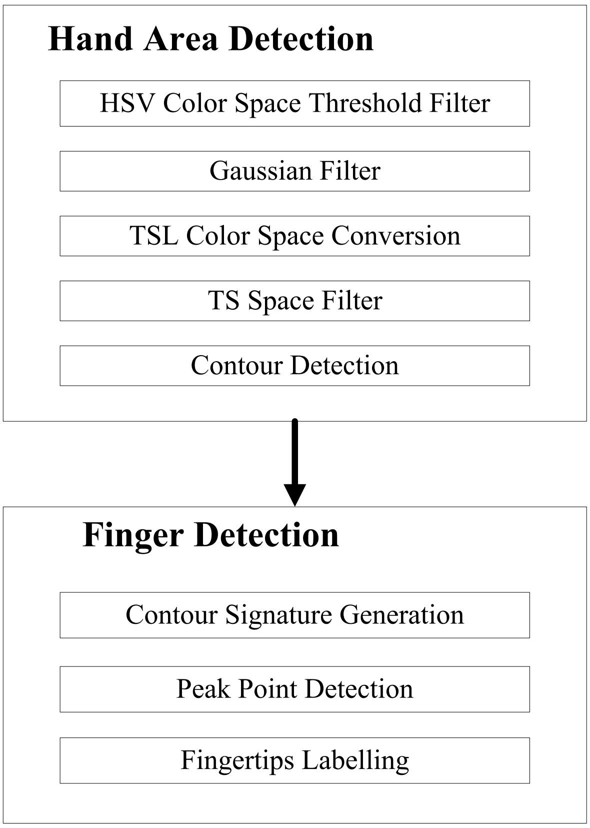


Figure 1. Flowchart of proposed hand gesture interface

Hand area detection is an important step. The usual way that setting lower and upper bounds in color space forming a threshold box in color space will be affected largely by noise and is not stable because the distribution of human hand in color space should be a cluster enveloped by a smooth continuous circle like shape and skin color of different people will have different distribution. By the introduction of TSL color space which is mainly used in face color detection before and self-adaptive method, the detection system is able to suppress the noise and to show a rather stable and accurate result of hand contours in real time.

Several approaches use contour curvatures to do the finger detection from hand contours. But curvature based detection has the problem that when the hand contour given is not ideal, fingertips cannot be correctly detected. The method proposed in this paper uses contour signature, the sequence of distances between contour points and moment, to calculate the finger tip points. The signature-based based finger detection will not be influenced by the size of hand. This approach show advantage of stability compared with curvature based finger detection approaches.

The rest of the paper is arranged as follows: Section III explains the improved TSL self-adaptive skin color area detection, Section IV about distance based fingers detection. Section V shows the experimental results. Finally Section VI presents the conclusion of work.

III. Improved TSL Self-Adaptive Hand Area Detection

Hand skin color of different people will differ slightly. Due to these differences, the general color space threshold may not be suitable to different people. In order to realize dynamic skin color detection, different detection rules should be taken. Thus, adaptive skin detection is needed.

A. Hand Area Dynamic Sampling

To crop the ROI(Region of Interest i.e. the sample skin) manually [7] doesn’t meet the need of real-time realization where hand position changes from time to time and consequently setting ROI manually becomes inconvenient. Therefore, dynamic sample method will automatically find ROI. By using threshold box in HSV color space hand area can be segmented roughly. The threshold box used in this paper is set as:

$$\begin{aligned} 100 \le H \le 120 \\ 70 \le S \le 150 \\ 103 \le V \le 255 \end{aligned} \tag{1}$$

The method takes strict restriction on the threshold to reduce the introduction of noise as much as possible since the following steps will try to expand sample area to the complete hand area.

General color information of sample points in hand area can be obtained. Many holes and noises points will appears on the cluster of points detected. Gaussian filter has been used to fill holes and to remove noises in order to enhance the sample cluster.

The kernel of filter uses a two dimensional Gaussian function.

$$G(x,y) = Ae^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

$$d(x,y) = \sum_{\substack{0 \le x' \le M \\ 0 \le y' \le N}} G(x',y') * s(x+x', y+y') \tag{3}$$

B. TSL Color Space Filter

TSL color space which composes of three components hue, saturation, and luminance, is proposed by Terrillon and Akamatsu in 2000 [8], and has been used for face detection. It also shows potential for hand skin color detection [2].

TSL color space try to make high difference between non-skin color and skin color value. Figure 2 shows TSL and HSV mapping image of same raw image. The TSL image shows high contrast between skin color and non-skin color so that hand area is more distinguishable compared with HSV image.

Mapped into TSL color space

Mapped into HSV color space

Figure 2. TSL(left) and HSV(right) color space image in gray scale

The transformation from RGB to TSL is as follows

$$L = 0.299r + 0..587g + 0.114b \tag{4}$$

$$S = \sqrt{\frac{9}{5}(r'^2 + g'^2)} * 255$$

$$T = \begin{cases} (\frac{\arctan(r'/g')}{2\pi} + \frac{1}{4}) * 255, g > 0 \\ (\frac{\arctan(r'/g')}{2\pi} + \frac{3}{4}) * 255, g > 0 \\ 0, others \end{cases}$$

Where:

$$r' = \frac{R}{R+G+B} - \frac{1}{3}$$
$$g' = \frac{G}{R+G+B} - \frac{1}{3} \tag{5}$$

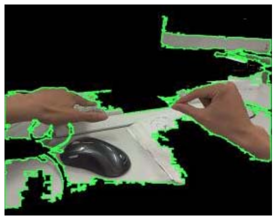
Since luminance component has little influence on the distribution shape and varies greatly when lighting effects changes. Further steps will only consider T and S components. The two dimensional graph of TS distribution of hand points can be established.

Contour detection and main part sift is applied on the TS graph to fill up the missing pixels in the cluster and remove extra pixels that doesn't belong to the cluster of hand color.

Sampled points sometimes may include unwanted noise points which would be amplified if the adaptive color redetection in next step is applied. TS space will give skin color point a high T(hue) value and give non skin color point a low T value. Thus, to remove the influence of noise points, a filter that cut off left part of the data on TS graph should be applied to it at first. Figure 3 shows a contrast between TS filtered image and unfiltered image.

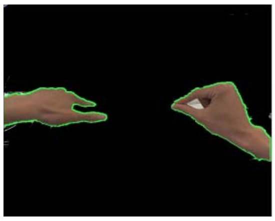
And then, a rule that decide whether or not a pixel can be classified as skin color should be applied. Instead of fitting sample points into a circle or box, contour detection is used to find a boundary that fit the cluster shape most. In that case noise parts are cut off in TS graph and only the main cluster in TS graph is remained.

Original TS graph



Result

Filtered TS graph



Result

Figure 3. Filtered and unfiltered TS graph and fetection results

Morphology closing has been used to remove extra pixels doesn't belong to the cluster and fill the holes in the cluster which enhanced the cluster graph

C. Hand Area Redetection

Detect the main area using obtained points in the step before. A binary mask picture can be drawn by rules that any pixel of which TS coordinate falls inside the boundary is considered as skin color pixel. Hand area can be segmented. Use contour detection again and leave the main part to segment the hand area.

IV. CONTOUR SIGNATURE BASED FINGERS DETECTION

A. Contour Signatures

To locate fingertips from the information of hand contours, a function to describe the contour should be used. An ideal one is a set function value that does not depend on geometrical transformations such as translation, rotation, size changes and reflection [9]. Distance based signature is translation, scale and rotation invariance [10]. By calculating contour signature, the feature of finger can be captured easily.

The contour is denoted by the set of points, ((x1, y1), (x2, y2), (x3, y3), ...,(xN,yN)) along the contour with N the total number of points of the contour. The hand contour signatures describing the contours can be constructed as follows: for each point on the contour, by calculating the distance from its coordinate to the moment of contour the signature, function $d(i) = d_i, i = 1,2...N$ is calculated where d_i is defined as:

$$d_i = \left\| (x_i, y_i) - \frac{\sum_{(x_j, y_j) \in C} (x_j, y_j)}{N} \right\| \tag{6}$$

Figure 4 shows the signature value 3D graph corresponding to contour points of a hand from different views. And Figure 5 shows the signature value in 2D graph. The peak value of curve indicates either the part of fingertips or the part arm. Therefore, fingertips can be distinguished clearly by peak detection.

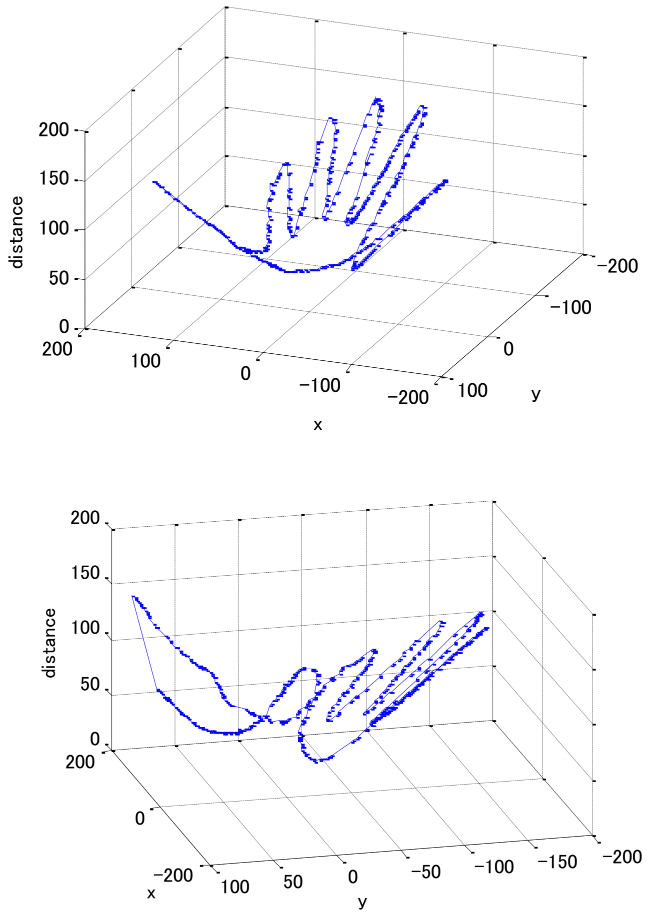


Figure 4. Signature values of points along the contour from different views

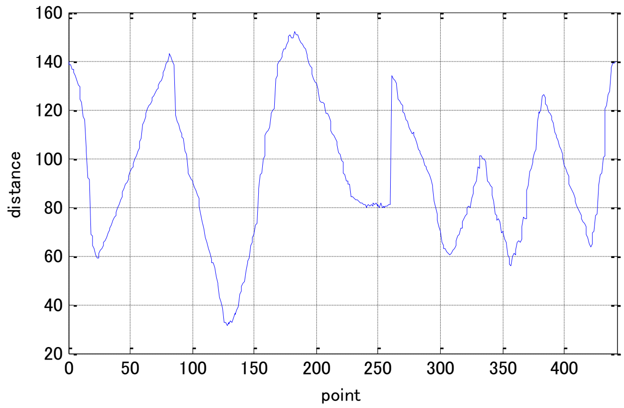


Figure 5. Data mapped into 2D graph

B. *Finger Detection*

The points around fingers will have a peak value. By taking use of this characteristic, fingertips position can be detected.

Before extracting peak points, point indicate arm which also has a peak value should be removed first. Following decision rule is used to judge whether or not a peak point is of arm.

$$(x-\bar{x}, y-\bar{y}) \bullet (\hat{x}-\bar{x}, \hat{y}-\bar{y}) < -threshold \quad (7)$$

Where $\bar{x}$ is the moment point of contours and $\hat{x}$ is the average point of peak points. When a point satisfies the condition above, it is judged as a non finger point.

In this paper, peak extraction is achieved by comparison between current point and several neighborhood points. If current point is larger than all the points compared, it's judged as finger point.

Compared with curvature based finger detection [1, 11], the result of proposed method is invariant with translation and size changes of hand shapes. It's more suitable for finger labeling.

In the coordinate system of OpenCV, the angle θ of a peak point is defined as

$$\theta = \begin{cases} 2\pi - \arccos(\frac{x}{x^2+y^2}), x \le 0 \\ \arccos(\frac{x}{x^2+y^2}), x > 0 \end{cases} \quad (8)$$

According to the angle obtained, peak points are labeled with numbers counterclockwise. Figure 6 is a illustration of labeling. Sometimes, the contour obtained has some noise and burrs, the peak point is detected incorrectly. To solve this problem, a merge processing is used as shown in Figure 6. When two points are too close (threshold is set as 8° in this paper), these points are merged as one point. Figure 7 shows the result image after labeling is applied.

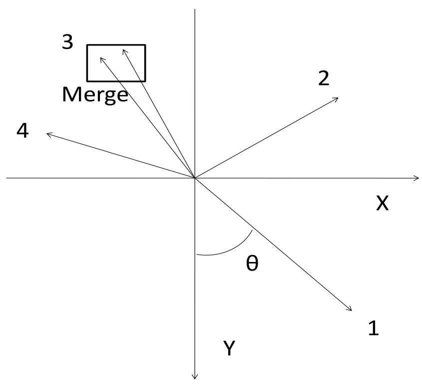


Figure 6. Fingertips labeling

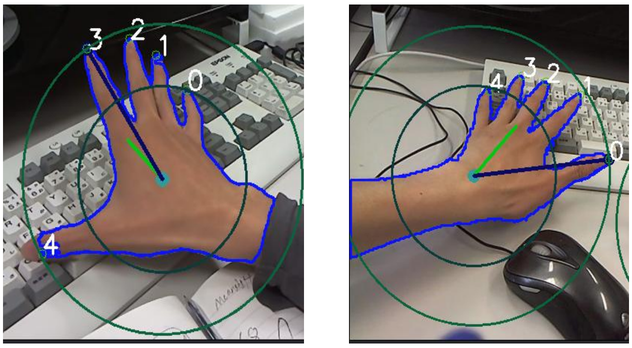


Figure 7. Result image after labeling

V. EXPERIMENTAL RESULT

Skin detection is a two-class prediction binary problem, in which the outcomes are labeled either as positive(skin color point) or negative(non skin point). Thus, to evaluate the performance of proposed adaptive skin color detection method and other methods, ROC(Receiver Operating Characteristic) can be used to analysis the results.

5 representative frames in video sequences of different hand gestures in different light situations are analyzed to show comparison between different methods. The features of these 5 pictures are as followings:

1	2	3	4	5
Two open hands	Noise area enclosed by fingers in strong light	One hand with finger stretched in weak light	Two hands in fist in medium light	Fingers partly stretched

In this method, TPR(True Positive Rate) and FPR(False Positive Rate) are used to measure the performance. The definitions are followings:

$$TPR = TP/(TP+FN)$$
$$FPR = FP/(FP+TN) \tag{9}$$

where TP is true positive points. FP is false positive points. TN is true negative points. FN is false negative points.

The image of actual value, the standard result, is obtained by manually deciding whether or not a pixel belong to hand area with human eyes.

TPR defines how many correct positive results occur among all positive samples available during the test. FPR, on the other hand, defines how many incorrect positive results occur among all negative samples available during the test.

Empirical evidence has shown that if TPR is too low the detected area will be incomplete and shattered, if FPR is too high, the detected area will be over-expanded, and the result will be good enough for further finger detection if and only if TPR is larger than 91% and FPR is less than 2%.

Table I and Table II are the TPR and FPR values respectively of 5 frame pictures. An item in bold print indicates that the contour obtained from this result is unacceptable for the following finger detection.

To clearly view the data cluster, data is plotted in TPR-FPR graph (Figure 8) where the feasible area is decided by TPR>0.91 and FPR<0.2.

TABLE I. TPR (TRUE POSITIVE RATE)

Test Image No. / Method	1	2	3	4	5
TSL	0.95432	0.95751	0.95721	0.97630	0.98978
TSL1	**0.87650**	0.95751	0.93286	0.97273	0.97631
TSL2	0.97829	0.99990	0.95721	0.99109	0.99549
TSL3	0.91508	**0.90316**	0.94792	0.97135	0.98265
TSL4	0.97998	0.99990	0.94833	0.99368	0.99633
HSV	0.99107	0.99964	0.93388	0.99681	0.99535
HSVBOX	**0.82175**	0.96733	**0.8017**	**0.81455**	0.94492

TABLE II. FPR (FALSE POSITIVE RATE)

Test Image No. / Method	1	2	3	4	5
TSL	0.01295	0.01775	0.00536	0.04826	0.03625
TSL1	0.00401	0.01605	0.00340	0.04672	0.00429
TSL2	0.01393	**0.41072**	0.00536	**0.34215**	0.03842
TSL3	0.00942	0.01336	0.00452	0.02426	0.02097
TSL4	0.01410	**0.41573**	0.00453	**0.35539**	0.04007
HSV	0.02517	**0.31792**	0.00444	**0.43456**	0.01164
HSVBOX	0.00269	0.02234	0.00161	0.00239	0.00101

TSL: Proposed methods

TSL1: proposed methods without Gaussian blur

TSL2: proposed methods without TSL filter

TSL3: proposed methods without morphology step

TSL4:proposed methods with part filter

HSV: adaptive color detection method in HSV color space

HSVBOX [3]: setting thresholds in HSV color space

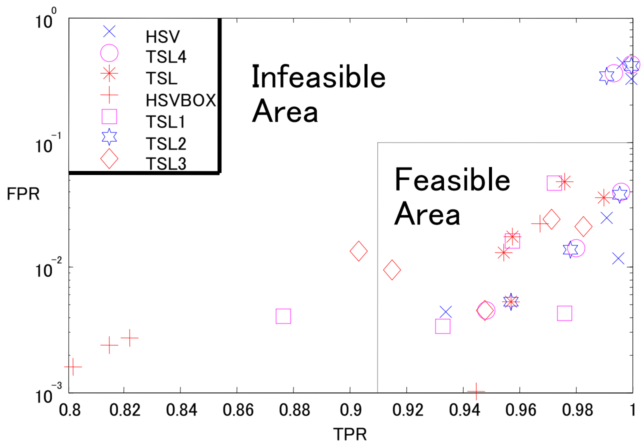
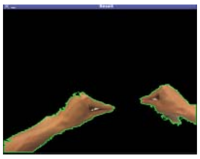
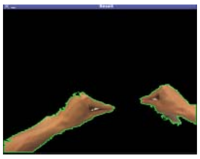


Figure 8. ROC distribution of different methods

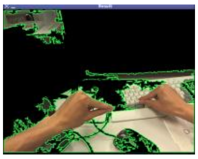
From the graph, it can be seen that the points of proposed method (TSL) almost stay in the feasible area while points of other methods is partly distributed in infeasible area.



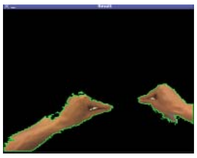
TSL



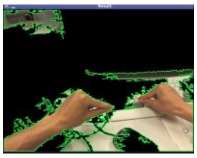
TSL1



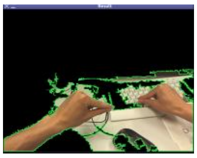
TSL2



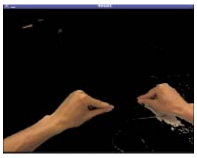
TSL3



TSL4



HSV



HSVBOX

Figure 9. Detection results comparison of different methods

In some cases, proposed method doesn't give the best result, but still stay in feasible area where the result is good enough for further analysis. On the other hand, other method although may shows better results in some cases but will give an unacceptable result in other cases which is unacceptable and impossible for further analysis. Figure 9 is an example of comparison with different methods.

According to the experimental results, compared with other methods, the proposed one in the paper shows strong stability and robustness since the results almost remain in the feasible region. This method could reduce the interference of noise regions.

The evaluation program is made on Ubuntu 11.04 Linux 2.6.38-14 with Eclipse SDK 3.5.2. CPU is Intel Core 2 Duo E8500 3.16GHz. OpenCV, Real time computer vision library, is used to provide with basic IO interface functions. Camera used is Logicool HD pro webcam C910 with resolution of 640*480. The frame rate is around 10 fps which can meet the need of real time.

VI. CONCLUSION

In this paper, a hand gesture interface has been presented. Proposed skin color detection method could reduce the interference of noise and is able to automatically be adaptive to different situations and different people compared with other methods. Proposed contour signatures based finger detection provides with a new way of locating fingertips from information of hand contour. Experimental result indicates that proposed hand gesture interface is capable of real-time video processing and can be used in HMD system.

Because of the complexity of TSL color space conversion, there is much potential for improvement and optimization of algorithm. Since information obtained from intra-frame is limited, inter-frame information can be used to improve this method.

ACKNOWLEDGMENT

This work was supported by KAKENHI (23300018).

REFERENCES

[1] Chin-Chen Chang, I-Yen Chen and Yea-Shuan Huang, Hand Pose Recognition Using Curvature Scale Space,Pattern Recognition, 2002. Proceedings. 16th International Conference on 2002, Volume 2 Page 386 – 389

[2] Wei Ren Tan, Chee Seng Chan, Pratheepan Yogarajah, and Joan Condell, A Fusion Approach for Efficient Human Skin Detection, Industrial Informatics, IEEE Transactions on Feb. 2012 Volume: 8 , Issue 1, Page 138 - 147

[3] Yoshimasa Iwasaki, Takeshi Ikenaga, Hand gesture interface for wearable hands-free digital camera, IMPS2011 on 26-28 Oct.

[4] J . Yang and A. Waibel. A real-time face tracker. In Proceedings of the Third Workshop on Applications of Workshop on Computer Vision (WACV'96), Sarasota, Florida, 1996.

[5] Ahmad Yahya Dawod, Junaidi Abdullah and Md.Jahangir Alam, Adaptive Skin Color Model for Hand Segmentation, 2010 International Conference on Computer Applications and Industrial Electronics (ICCAIE 2010), December 5-7, 2010, Kuala Lumpur, Malaysia

[6] Yingen Xiong, Bing Fang, and Francis Quekm, Extraction of Hand Gestures with Adaptive Skin Color Models and its Applications to Meeting Analysis, Multimedia, 2006. ISM'06. Eighth IEEE International Symposium on Dec. 2006

[7] Shu Mo, Shihai Cheng, and Xiaofen Xing, "Hand Gesture Segmentation Based on Improved Kalman Filter and TSL Skin Color Model," Multimedia Technology (ICMT), 2011 International Conference on. 26-28 July 2011

[8] Terrillon, Jean-Christophe; Akamatsu, Shigeru (2000), Comparative Performance of Different Chrominance Spaces for Color Segmentation, International Conference on Face and Gesture Recognition, pp. 54–61, retrieved 2008-02-10

[9] V.V. Kindratenko, "On Using Functions to Describe the Shape", Journal of Mathematical Imaging and Vision, 2003, vol. 18, number 3, pp. 225-245.

[10] Shi Wang, Hyongsuk Kim, and Hongxin Chen, Contour Tracking Using Centroid Distance Signature and Dynamic Programming Method, Computational Sciences and Optimization, CSO 2009. International Joint Conference on 24-26 April 2009, Volume 1 Page 274 – 278

[11] Seong-Pal Kang, Michal Tordon and Jayantha Katupitiya, Curvature based hand shape recognition for a virtual wheelchair control interface, Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on April 26-May 1, 2004V.V. Kindratenko, "On Using Functions to Describe the Shape", Journal of Mathematical Imaging and Vision, 2003, vol. 18, number 3, pp. 225-245.