

Hand Gesture Recognition using Improved Hidden Markov Models

Wenkai Xu[†], Eung-Joo Lee^{**}

ABSTRACT

In this paper, an improved method of hand detecting and hand gesture recognition is proposed, it can be applied in different illumination condition and complex background. We use Adaptive Skin Threshold (AST) to detect the areas of hand. Then the result of hand detection is used to hand recognition through the improved HMM algorithm. At last, we design a simple program using the result of hand recognition for recognizing "stone, scissors, cloth" these three kinds of hand gesture. Experimental results had proved that the hand and gesture can be detected and recognized with high average recognition rate (92.41%) and better than some other methods such as syntactical analysis, neural based approach by using our approach.

Key words: Hand Detecting; Hand Gesture Recognition; YCbCr Skin Color Space; HMM

1. INTRODUCTION

The technology of hand gesture recognition has been a very important research topic of human-computer interaction(HCI). The main purpose of researching hand gesture recognition is to use this nature manner to implement human-computer interaction. With the rapid development of computer technology, the research on new human-computer interface, which suits human's communication custom, has obtained the great evolvement. The hand gesture acts as one of the common communication methods in human life, So the hand gesture recognition has become the research focus. Nevertheless, because of the variety, multi vocal, otherness in both temporal field

and spatial field uncertainty in vision and that the hand is complicated plasmodium, the research has become a multi-knowledge inter across problem filled with challenge.

Hand gesture recognition from visual images has a number of potential application in HCI (human computer interaction), machine vision, VR (virtual reality), machine control in the industry field, and so on. Most conventional approaches to hand gesture recognition have employed data gloves. But, for more natural interface, hand gesture must be recognized from visual images as in the communication between humans without using any external devices. Our research is intended to find a high efficiency approach to improve the algorithm of hand detecting and hand gesture recognition.

Our research is intended to find a better approach to detect the position of human face and hand. In the detected region, we are using the skin color model to detect the precise of positioning the face and hand location in YCrCb. After that, we extract the characteristic vector of several hand gestures by using improved HMM algorithm. The experimental results show that, the proposed

※ Corresponding Author : Eung-Joo Lee, Address : Dept. of I.C. Eng., Tongmyong University, 535 YongDang-Dong, Nam-Gu, Busan 608-711, Korea, TEL : +82-51-629-3700, FAX : +82-51-629-3729, E-mail : ejlee@tu.ac.kr
Receipt date : Mar. 23, 2011, Revision date : May 15, 2011
Approval date : June 1, 2011

[†] Dept. of Information Communication Engineering, Tongmyong University
(E-mail: xwk6298@hotmail.com)

^{**} Dept. of Information Communication Engineering, Tongmyong University

method of hand detection can detect the hand location regions quickly overall, and to a large extent suppress the background noise and the changes of background region, and the recognition rate of hand gesture is also quite satisfying.

2. HAND DETECTION ALGORITHM

2.1 Color Space Foundation

A color model is an abstract mathematical model describing the way colors can be represented as tuples of numbers, typically as three or four values or color components (e.g. RGB and CMYK are color models). However, a color model with no associated mapping function to an absolute color space is a more or less arbitrary color system with no connection to any globally-understood system of color interpretation.

Human skin color has been used and proved to be an effective feature in many applications from human face detection to hand tracking [1]. However, most studies use either simple thresholding or a single Gaussian distribution to characterize the properties of skin color. Although skin colors of different races fall into a small cluster in normalized RGB or HSV color space, we find that a single Gaussian distribution is neither sufficient to model human skin color nor effective in general applications. Although different people have different skin color, several studies have shown that the major difference lies largely in their intensity rather than their chrominance. Several color spaces have been utilized to label pixels as skin including RGB, normalized RGB, HSV (or HSI), YCbCr, YIQ, YES, CIE XYZ, and CIE LUV.

2.2 The YCbCr Color Model

Cathode ray tube displays are driven by red, green, and blue voltage signals, but these RGB signals are not efficient as a representation for storage and transmission, since they have a lot of mutual

redundancy[2]. YCbCr and Y'CbCr are a practical approximation to color processing and perceptual uniformity, where the primary colors corresponding roughly to red, green and blue are processed into perceptually meaningful information. By doing this, subsequent image/video processing, transmission and storage can do operations and introduce errors in perceptually meaningful ways. Y'CbCr is used to separate out a luma signal (Y') that can be stored with high resolution or transmitted at high bandwidth, and two chroma components (Cb and Cr) that can be bandwidth-reduced, subsampled, compressed, or otherwise treated separately for improved system efficiency. One practical example would be decreasing the bandwidth or resolution allocated to "color" compared to "black and white", since humans are more sensitive to the black-and-white information. The following formula (1) can express how to convert RGB color space to YCbCr color space:

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ Cb &= 0.564(B - Y) \\ Cr &= 0.713(R - Y) \end{aligned} \quad (1)$$

In human skin color, Cr is the most important component (we can find that in Fig. 1), so our method just cares about Cr component.

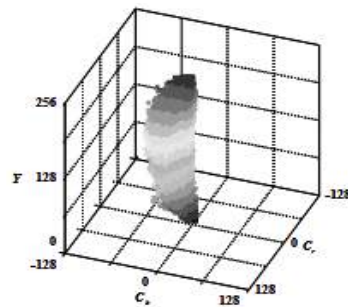


Fig. 1. Skin Color in YCbCr Color Space.

2.3 Hand Detection Based on Adaptive Skin Threshold (AST) segmentation

Threshold is very significant problem in image processing and pattern recognition [3]. However,

one threshold cannot give satisfactory result in yellow, black and white color. To overcome this shortcoming, we propose a improved method for skin detection which using Adaptive Skin Threshold (AST).

The covariance of face and another body part is very small, therefore, the extracted AST from face region becomes the best selection for hand region detection. We can express the relation between them as the formula (2) as followed:

$$Cr_{Thes} = \frac{1}{Fc_w \cdot Fc_h} \sum_{y=0}^{Fc_h} \sum_{x=0}^{Fc_w} C_r(x, y) \tag{2}$$

Here:

Fc_w means detecting face region width

Fc_h means detecting face region height

$Cr(x, y)$ means Cr value at (x, y) pixel

From the threshold extracted by using AST method of face region, we can use it for detecting hand region. Fig. 2 shows the proposed hand region detection flowchart. To begin with, we should do some preprocessing on the image we get through the camera. Firstly we convert the original image into binary and apply some morphology method such as Dilation and Erosion, used for noise removal and separate close white region. Based on

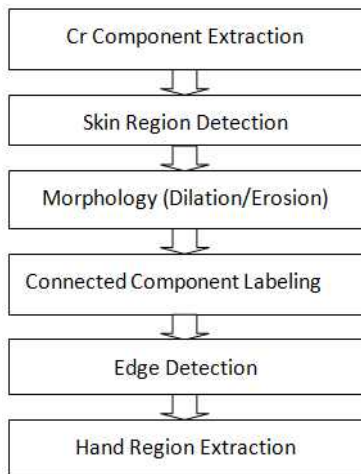


Fig. 2. Hand region detection flowchart.

the face region size, we can remove the small objects firstly and then find edge for remained skin region. For every hand region, it always includes edge not only on its boundary, but also within the contour. So, if any skin region doesn't satisfy this constraint, it will be removed. The biggest one of remained region is become hand region. The hand region detection process is shown as the following flowchart. Fig. 3 shows an example of hand detection

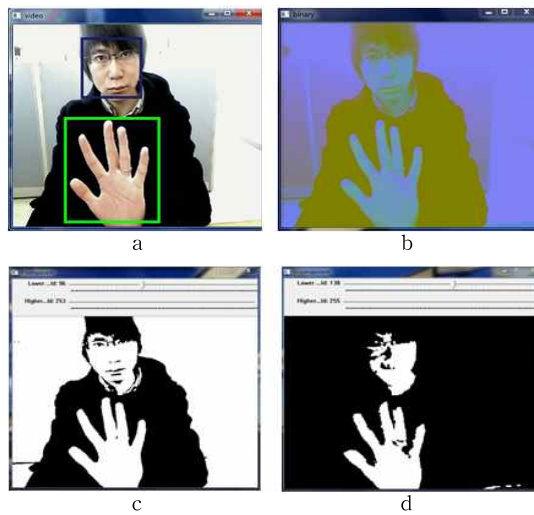


Fig. 3. (a) Original image, (b) Cr component in $YCbCr$ color space, (c) Extracted binary image using $YCbCr$, (d) Hand and face region

3. GESTURE RECOGNITION ALGORITHM BASED ON HMM

Hand gesture recognition can be classified into several categories as shown in Table 1 [4].

Up to the present, the level of recognition technology is far beyond that of human, therefore most researches have been done on the strong gesture grammar in the confined environment [5]. The basic recognition methods are mainly based on the statistical model of optical flow, HMM using dynamic feature, and the analysis of syntactical grammar.

Since the speech recognition field adopted

Table 1. Classification of Hand Gesture Recognition

Category	Type
Task	Pointing, tracking, gesture
Motion	Static, dynamic
Input information	Image by camera, electrical signal by data glove
Image feature	Natural image, marked image
Dimension of data	2-D, 3-D
Used hand	One-hand, two-hands
In view of psychology	Symbolic, kinetographic, iconic, baton, ideographic

HMMs with great success many years ago, these techniques are just now entering the computer vision area where time variation is significant. Hidden Markov models have intrinsic properties which make them attractive for gesture recognition, and also explicit segmentation is not necessary for either training or recognition.

A hidden Markov model is a collection of finite states connected by transitions [6]. Each state is characterized by two sets of probabilities: a transition probability, and either a discrete output probability distribution or a continuous output probability density function. HMM can be defined by: (1) A set of states $\{S\}$ with an initial state S_i and a final state S_f ; (2) The transition probability matrix, $A=\{a_{ij}\}$, where a_{ij} is the transition probability of taking the transition from state i to state j ; (3) The output probability matrix B . For a discrete HMM, $B=\{b_j(O_k)\}$, where O_k represents discrete observation symbol. For a continuous HMM, $B=\{b_j(x)\}$, where x represents continuous observation of K -dimensional random vectors. With the initial state distribution $\pi=\{\pi_i\}$, the compactly as $\lambda=(A, B, \pi)$. HMM can be based either on discrete observation probability distributions or continuous mixture probability density functions. In the discrete HMM, the discrete probability distributions are sufficiently powerful to characterize

random events with a reasonable parameters. The principal advantage of continuous HMM is the ability to directly model the parameters of a continuous signal. A semi-continuous HMM provides a framework for unifying the discrete and continuous HMMs. In our research, we only consider a discrete HMM. There are three basic problems that must be solved for the real application of HMM: 1) evaluation, 2) decoding, and 3) learning. The solution to these three problems are the Forward-Backward algorithm, the Viterbi algorithm, and the Baum-Welch algorithm. The evaluation problem is that given an observation sequence and a model, what is the probability that the observed sequence is generated by the model ($P(O|\lambda_i)$). If this can be calculated for all competing models for an observation sequence, then the model with highest probability can be selected as the 1st objective gesture model. In our research, the observation symbol O is represented as the sequence of O_i .

$$O=O_1, O_2, \dots, O_{t-1} \quad (3)$$

Where O_{t-1} is the 8 directional chain code between time $t-1$ and t . Therefore our problem is to find the gesture model with the highest probability on given O .

$$\text{Max } \{P(\lambda_i | O)\} \quad (4)$$

Where λ_i is the i^{th} vocabulary gesture. In this research, the gesture vocabulary consists of three gestures: $\lambda_1, \lambda_2, \lambda_{13}$ as shown in Fig. 4. This probability is not computable directly but using Bayes's Rule gives

$$P(\lambda_i | O)=P(O | \lambda_i)P(\lambda_i) / P(O) \quad (5)$$

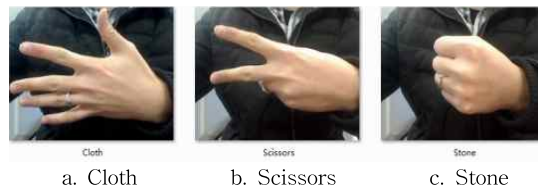


Fig. 4. Three hand gesture sample.

The initial topology for HMM can be determined by estimating how many different states are involved in specifying a gesture. While different topologies can be organized for each gesture, a 4-state HMM with skip transition is determined to be sufficient for our work as shown in Fig. 5.

The input information is a (x_i, y_i) center coordinate of a hand position on time series (t_i) , therefore the input data are $(x_1, y_1, t_1), (x_2, y_2, t_2) \dots (x_n, y_n, t_n)$. These are converted to $(O_1, O_2, \dots, O_{n-1})$, where O_i means the chain code between (x_i, y_i, t_i) and $(x_{i+1}, y_{i+1}, t_{i+1})$. The redundant trajectory made before and after the key gesture is eliminated by an interactive spotting.

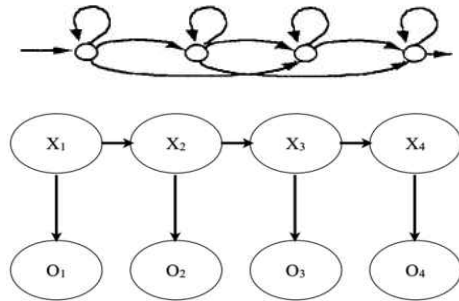


Fig. 5. The 4-state HMM used for recognition.

4. EXPERIMENTAL RESULTS

We design a simple program using the result of hand recognition for recognizing “stone, scissors, cloth” these three kinds of hand gestures. This system operated by hand gesture is running on the hardware environment of CPU Intel 2.3GHz Core2, a Logitech Web camera, and the software environment of Windows 7 and Visual Studio 2008 using OpenCV. Fig. 6 and Table 2 show the experimental results of the program.

These results show that the recognition results can be enhanced with the increased training. In real operating, the recognition results highly depend on the real gesture and at the same time all real data are new data. The described test was done on the batch processing condition of the accumulated data.

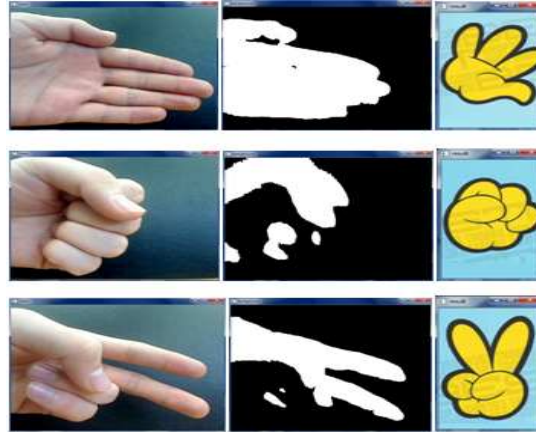


Fig. 6. Experimental Results.

Table 2. The Experimental Results

Type \ Results	Test Time	Recognition Time	Recognition Rate(%)
	342	321	93.86%
	350	319	91.14%
	347	320	92.22%

5. CONCLUSION

In this paper, a improve method of hand detecting and hand gesture recognition is proposed. We use Adaptive Skin Threshold (AST) to detect the areas of hand. Then the result of hand detection is used to hand recognition through the improved HMM algorithm. It can apply in different illumination condition and complex background. And the recognition rate of hand gesture is also quite satisfying.

ACKNOWLEDGEMENT

This work is financially supported by the Ministry of Education, Science and Technology

(MEST), the Ministry of Knowledge Economy (MKE) through the fostering project of HUNIC.

REFERENCE

- [1] Mubarak Shah. "Face Detection and Gesture Recognition for Human-Computer Interaction," Chapter 4, pp. 83.
- [2] Bai yu, "A New Method for Hand Tracking and Hand Gesture Recognition Using Improved CAMSHIFT and SVM Algorithm," Master's Dissertation, Tongmyong University Image Processing Lab, pp. 17, 2010.
- [3] N. A.Viet, "A Development of Non-touch Human Computer Interface Technology using Gesture Information," Master's Dissertation Tongmyong University Image Processing Lab, pp. 15, 2008.
- [4] B.W. Min, H.S. Yoon, and Toskiaki Ejima, "Hand Gesture Recognition Using Hidden Markov Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No.1, 1997.
- [5] T.E. Startner and A.Pentland, "Visual Recognition of American Sign Language using Hidden Markov Models," In Proceedings of the First International Workshop on Automatic Face and Gesture Recognition, pp. 189-194, 1995.
- [6] J. Yang, et al., "Human Action Learning Via Hidden Markov Model," *IEEE Trans. On Systems, Man, and Cybernetics*, Vol.27, No.1, pp. 34-44, January 1997.
- [7] Bai Yu, Sang-Yun Park, Yun-Sik Kim, In-Gab Jeong, Soo-Yol Ok, and Eung-Joo Lee, "Hand Tracking and Hand Gesture Recognition for Human Computer Interaction," *Journal of Korea Multimedia Society*, Vol.14, No.2, pp. 182-193, 2011.



Wenkai Xu

received his B. S. at Dalian Polytechnic University in China (2006 - 2010). From 2010 to now he is a M. S student of Tongmyong University, Korea. His main research tops are image processing, computer vision,

biometrics and hand recognition.



Eung-Joo Lee

received his B. S. , M. S. and Ph. D. in Electronic Engineering from Kyungpook National University, Korea, in 1990, 1992, and Aug. 1996, respectively. Since 1997 he has been with the Department of Information & Communications Engineering, Tongmyong University, Korea, where he is currently a professor. From 2000 to July 2002, he was a president of DigitalNetBank Inc.. From 2005 to July 2006, he was a visiting professor in the Department of Computer and Information Engineering, Dalian Polytechnic University, China. His main research interests includes biometrics, image processing, and computer vision.