Journal of Korea Multimedia Society Vol. 14, No. 4, April 2011(pp. 504-516)

Hand Gesture Recognition using Optical Flow Field Segmentation and Boundary Complexity Comparison based on Hidden Markov Models

Sang-Yun Park†, Eung-Joo Lee††

ABSTRACT

In this paper, we will present a method to detect human hand and recognize hand gesture. For detecting the hand region, we use the feature of human skin color and hand feature (with boundary complexity) to detect the hand region from the input image; and use algorithm of optical flow to track the hand movement. Hand gesture recognition is composed of two parts: 1. Posture recognition and 2. Motion recognition, for describing the hand posture feature, we employ the Fourier descriptor method because it's rotation invariant. And we employ PCA method to extract the feature among gesture frames sequences. The HMM method will finally be used to recognize these feature to make a final decision of a hand gesture. Through the experiment, we can see that our proposed method can achieve 99% recognition rate at environment with simple background and no face region together, and reduce to 89.5% at the environment with complex background and with face region. These results can illustrate that the proposed algorithm can be applied as a production.

Key words: Hand Gesture Recognition, Posture Recognition, Optical Flow, Boundary Complexity, Fourier Descriptor, PCA, HMM

1. INTRODUCTION

With the development of ubiquitous computing, computer is becoming more and more important in our daily life. Computer applications require more and more unrestricted interaction between human and computers, which is great challenge to traditional input devices such as mouse, keyboard or pen etc. Hand gesture is frequently used in people's daily life. It's also an important component of body languages in linguistics. Compared with those devices mentioned above, hand gestures are more natural in interaction. The use of hand as a device for human-computer interaction (HCI) makes HCI easy [1,2].

The key problem in gesture interaction is how to make hand gestures understood by computers. Extra sensors and instruments, such as data gloves, may be easy to collect hand configuration and movement. However, these equipments are usually expensive and bring much cumbersome experience to users. Vision based gesture interaction have many appealing characteristics. The prominent one is that it realizes a natural interaction between human and computers independent of external dedicated devices. In general, the literatures of hand gesture recognition fall into two categories: learning based and model based method.

※ Corresponding Author : Sang-Yun Park, Address : (617-736) 100 silladaehak-gil Sasang-gu, Busan, Korea, TEL : +82-51-999-5355, FAX : +82-51-999-5184, E-mail : sypark@silla.ac.kr

Receipt date : Oct. 6, 2010, Revision date : Jan. 11, 2011
Approval date : Mar. 2, 2011

† Department of Information Communication Engineering, TongMyong University, Korea
†† Department of Information Communication Engineering, TongMyong University, Korea
(E-mail: ejlee@tu.ac.kr)

※ This research was supported by the MKE(Ministry of Knowledge and Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)(NIPA-2011-(C-1090-1121-0006))

Learning based method get classifier or detector with machine learning from the training data which is constructed by multi-cue features with plenty of sample images. One and Bowden [3] distinguished hand shapes with boosted classifier tree and obtained fairly good results. However, their method is time consuming and unpractical for interactive interfaces. Kolsch [4] used fanned boosting detection for classification and got nearly real time results, while the detector's training is computational expensive. What's more, the detector makes constraints on the resolution and aspect ratio of gesture template. As for the model based method, there are some researchers who have achieved satisfactory accuracy in hand gesture recognition. Lars and Lindberg used scale-space color features to recognize hand gestures [5]. In their method, gesture recognition is based on a stable hand gesture representation generated by a hierarchy of feature detection. This hand gesture representation is effective in recognition. Although the authors show nearly real-time application on a dual-processor computer, the computation costs expensively as feature detector and descriptor involve a great deal of large-scale Gaussian convolution.

In this paper, we will develop a real-time hand tracking and gesture recognition method which is robust and reliable in complex background. The tracking method is based on optical flow, the recognition method is based on HMM. The excellent feature of hand detection based on optical flow is that the algorithm is insensitivity to the static background but only sensitive to the moving object region. But if there are more than one object moving in the scene, we can't judge the hand part just based on the optical flow, but the boundary complexity can represent the shape edge feature, it is immune to shape rotation. The approach of our algorithm can be listed in detail as following:

1. Calculate the optical flow of input video flowing
2. Detect the skin region in regions which created the optical flow
3. Scanning the region with the maximum boundary complexity. (The region can be considered as the open hand region)
4. Then for recording the gesture sequence, the Fourier descriptor will be used to describe the shape of hand, and the PCA method will be employed to extract the principle component of hand optical flow field.
5. Finally, the HMM method will be used to recognize the feature vector. The output of these hidden models is the gesture that we want to recognize.

Our paper is organized as following: Section 2 will discuss the optical flow calculation method for detecting the moving object; Section 3 will discuss the color based skin region segmentation; Section 4 will discuss the boundary complexity calculation for confirming the hand region. Section 5 will discuss the hand gesture feature extracting method. Section 6 will discuss the hand gesture recognition method based on HMM. Finally, we give the experiment results and conclusion.

2. OPTICAL FLOW CALCULATION FOR DETECTING MOVING OBJECT

2.1 Optical Flow Field Calculation

In our system, the motion of the object provides important and useful information for object localization and extraction. To find the movement information, we assume that the input gesture is non-stationary. When objects move in the spatial-time space (an image sequence), we use the motion detector to detect the optical flow field of object. And for calculating the optical flow, we use the method which is proposed by Nagel [6,7] to calculate the flow velocity (displacement vector between frames). According to the method, the flow velocity of perpendicular direction to gray scale conversion line should change in small steps. This means that, we can set a pattern smoothing restrict to the basic of optical flow equation, so this can reduce the over smoothed to the edge of image when using the optical flow calculation equation [8, 9]. We define the two adjacent frames illumination energy as: $E(x, y, t_1)$ and $E(x, y, t_2)$, in the paper, we use $E(x, y, t_1)$ and $E(x, y, t_2)$ as $E1(X)$ and $E2(X)$, and here $X=(x,y)$. The basic of optical flow equation is as following:

$$(\nabla E)^T \bullet \mu + E_t = 0 \tag{1}$$

Here, $\nabla E = (E_x, E_y)^T$, E_x, E_y, E_t is partial derivative in point (x, y), and $\mu = (u, v)^T$, u, v is displacement between time $\Delta t = t_2 - t_1$ in x, y directions respectively. According to the "pattern smoothing restrict", Nagel suggest that constraint condition can be applied to u and v as following:

$$trace((\nabla u)^T W(\nabla v)) = \min \tag{2}$$

And here, W is weighted matrix, $W = F / trace(F)$, it compensates the changed in moving field according to the changed of gray value in the video image. And F is as following:

$$F = \left| \begin{vmatrix} E1_y \\ -E1_x \end{vmatrix} \begin{vmatrix} E1_y \\ -E1_x \end{vmatrix}^T + b^2 \begin{vmatrix} E1_{yy} & -E1_{xy} \\ -E1_{xy} & E1_{xx} \end{vmatrix} \begin{vmatrix} E1_{yy} & -E1_{xy} \\ -E1_{xy} & E1_{xx} \end{vmatrix}^T \right| \tag{3}$$

So based on Equation (1) and (3), the calculation of optical flow field can be converted to solving the equation as following:

$$\iint dxdy((E2(x) - E1(x-u))^2 + \alpha^2 trace((\nabla u)^T W(\nabla v))) = \min \tag{4}$$

a is used to control the smoothing level, according solving the equation, we can get the BSOR result for each point in the image.

2.2 Optical Flow Field Segmentation based on C-Mean Cluster Method

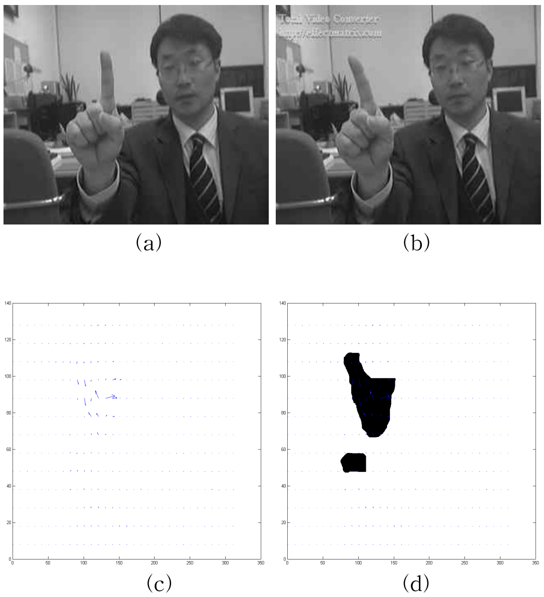
C-mean cluster method is method for calculating dynamic cluster by using error sum of squares. It is defined as following:

$$J_c = \sum_{j=1}^{c} \sum_{k=1}^{n_j} \left\| x_k - m_j \right\|^2 \tag{5}$$

$$m_j = \frac{1}{n_j} \sum_{k=1}^{n_j} x_k \quad j = 1, 2, \ldots, c \tag{6}$$

Here, x_k is the sample in set $X = \{x_1, x_2, \ldots, x_n\}$, the sample set will be converged to c separated subspace, they include $n_1, \cdots, n_c$ samples. m_j is the sample mean value of j^{th} subspace.

For separating the optical flow field, we define the sample is the point position in the optical flow field. The error sum of squares will use the Euclidean distance. The C-Means method will optimize the cluster result by iterative procedures. It will make the criterion function J_c to get the minimal value and get c classes. Then we will compare the sample number in each class, if the sample number is too few, it can be removed. If there are more than one classes, we can consider that there are several moving objects, and they can be separated.



(a) (b) (c) (d)

Fig. 1. Optical Flow Field Calculation and Segmentation Results: (a). Reference Frame; (b). Current Frame; (c). Optical Flow Field Calculated by (a) and (b)and (d) Optical Flow Field Segmentation Result.

3. SKIN REGION SEGMENTATION

We know that in the world, there exist three races of human: Whites, Negro and Yellow race, each race has a different skin color. In order to separate them, we should select a suitable color space which can linearly describe them. And in the color space, each skin color pixel can be strongly clustered together for the purpose of separating them early. So which color space will be the suitable for this will be the first problem we must solve. Through observing skin color feature, we can find that the RGB color space is not adequate for categorizing skin color because it has the three elements (R, G, B) that involve both chroma and luminance. The color component of different skin colors was found to be categorized in a small area of the chromatic color space. Although skin color varies extremely, they differ much less in chroma than in luminance.

Through the front comparison, we can convert the regular RGB image into YC_bC_r color model. And before the converting, to remove luminance from the RGB color space, we adapt the normalized RGB [10] representation that is obtained by a simple process:

$$\begin{aligned} red &= \frac{RED}{(RED+GREEN+BLUE)} \\ green &= \frac{GREEN}{(RED+GREEN+BLUE)} \\ blue &= \frac{BLUE}{(RED+GREEN+BLUE)} \end{aligned} \tag{7}$$

The advantage of this representation is that it can avoid the change of surface orientation relatively to the light source. Then the output result of the normalization process will be translated into YC_bC_r color space by using equation 8~10.

$$Y = -0.299 * R + 0.587 * G + 0.114 * B \tag{8}$$

$$Cr = (R - Y) * 0.713 + delta \tag{9}$$

$$Cb = (B - Y) * 0.564 + delta \tag{10}$$

The threshold value can be computed by using the C_r image, which we can apply the equation 4 to the image,

$$Cr_{Th} = w * \sum_{i=1}^{255} Hist(i) \tag{11}$$

Then according to the threshold, we will get the binary image of skin region, as result shown in Fig 2.

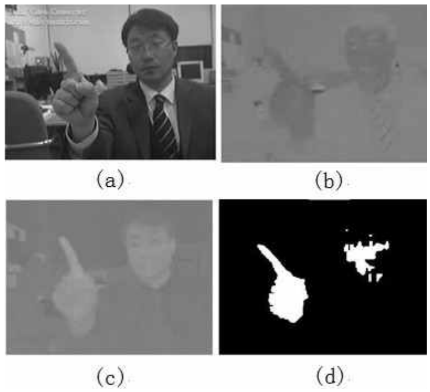


Fig. 2. Skin Color in YCbCr Space and Detection Result: (a) Skin Color Y Image; (b) Skin Color Cb Image; (c) Skin Color Cr Image and (d) Binary Detection Result.

By combining the optical flow field segmentation result and the skin region information, we can get the moving object with skin color, in other words the human motion can be tracked by using the logic 'AND' to combine the two types of information, that is:

$$Hm_i(x, y) = OF_i(x, y) \wedge S_i(x, y) \tag{12}$$

Here $OF_i(x, y)$, $S_i(x,y)$, indicate the optical flow, and skin color region. The $Hm_i(x, y)$ as human motion features can be extracted. The results shows in Fig 3.

Fig. 3. The Combined Region *Hmi(x, y)*.

4. REGION BOUNDARY COMPLEXITY CALCULATION

The combined skin regions usually include face region and other parts of body. In the hand gesture recognition system, what we needed is the hand region, so in this part we will utilize the boundary complexity method [11] to separate the hand region from other parts. As we know that, hand is composed of 5 spindling fingers. The hand region will have a more complex edge compare with other skin regions when it is open. The feature can make hand region separation more possible. Considering this, we propose the boundary energy estimation method which can measure the object edge boundary energy in an image. The basic foundation of the boundary energy estimation theory will be described in section 4.1.

4.1 Foundation of Edge Boundary Energy Estimation Method

Area and perimeters are two important attribute of geometric form, the one can not fill the needs for separating the geometric forms. Such that usually some geometric forms with the same area, but do not have the same perimeter or opposite. For separating this, a feature using to describe geometric form is called circularity. It has the minimum value when the geometric form is a circle. The value of circularity describes the complexity of edge. *A* common form of the circularity is described as following:

$$C = P^2 / A \tag{13}$$

P is perimeter, and *A* is area of object region, it will be 4π if the object is a circle and will be larger if the object is with more complex edge. The circularity (*C*) and edge complexity contain some relationships. One of them is the boundary energy [12]. We assume that an object with the perimeter of P, p is distance from one edge point to a start point. At any point, the edge will have an instantaneous curvature radius *r(p)*. It is radius of the circle which is tangent to the edge. So the curvature function at point p is as following:

$$K(p) = 1/r(p) \tag{14}$$

K(p) is a periodic function with a period of *P*. We can calculate the average energy of a unit length of boundary by using the following equation:

$$E = \frac{1}{P}\int_0^p \left|K(p)\right|^2 dp \tag{15}$$

For a given area, a circle will have the minimum boundary energy.

$$E_0 = (\frac{2\pi}{P})^2 = (\frac{1}{R})^2 \tag{16}$$

Here, *R* is radius of the circle. The boundary energy can describe the complexity of edge more similar with human perception.

4.2 Applying on Hand Region Classification

As analysis above, hand region is an object with a considerable region and complex edge. So we can enhance both of the two features by using the following equation:

$$\omega = a^{E*A} (a > 1) \tag{17}$$

Here, *a* is the enhancing coefficient, we can use it to adjust the result. Thus for each candidate skin region, we will calculate the ω value. The result is as shown in Fig 4. And finally we select the MAX region which can be considered as the hand region.

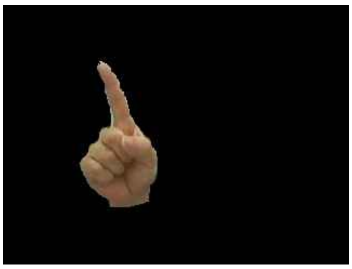


Fig. 4. Detected Hand Region.

5. HAND GESTURE FEATURE SEQUENCE EXTRACTION

5.1 Feature Selection for Hand Posture Description

For the posture of hand, they will be further converted to symbols which are the basic elements of the HMM. Effective and accurate feature vectors play a crucial role in model generation of the HMM. For selecting good features, the following criteria are considered useful: (1) Features should be preferably independent on rotation, translation and scaling. (2) Features should be easily computable. (3) Features should be chosen so that they do not replicate each other. This criterion ensures efficient utilization of information content of the feature vector. The features obtainable from the image of hand posture are spatial and temporal. To extract the shape features, we choose the FD to describe the hand shape. To associate with the hand motion, we can extract all the features to describe a hand gesture.

5.1.1 Fourier Descriptor

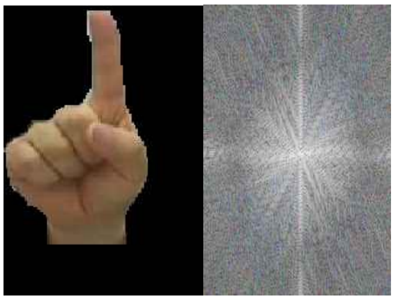
We may describe the hand shape by their features in the frequency domain, rather than those in the spatial domain. The local feature property of the node is represented by its Fourier Descriptors (FD) [13-15]. To assume the hand-shape is described by external boundary points, {*x(m), (m)*}, then we may use the FD representation for boundary description. To represent the boundary points, we may find the Fourier series of *x(m)* and *y(m)*, which are defined as *a(n)* and *b(n)*. For a closed boundary, this representation is called FD. The elements of the vector are derived as $S(n) = r(n)/r(1)$ where, $r(n) = [(a(n))^2 + (b(n))^2]^{1/2}$, n=1,2,⋯ Using of FD vectors of dimension 10 for hand written digit recognition is sufficient [14]. Here we assume that the local variation of hand-shape is smooth so that the high order terms of its FD are not necessary. So using 22 harmonics of the FD's is enough to describe the macroscopic information of the hand figures.

The advantage of using the FD is due to its size-invariant properties. For different scaled objects, only the magnitudes of their FD coefficients are changed by the same factor.

(a)

(b)

(c)



(d)

Fig. 5. Illustration the Invariant Properties of Fourier Descriptor: (a). Original Image and Fourier Descriptor (b). Rotation 90 and Fourier Descriptor (c). Zoom in 12:10 and Fourier Descriptor (d). Zoom out 8:10 and Fourier Descriptor.

5.1.2 Motion Analysis

For the moving hand, we will analyze the optical flow field formed by hand gesture. Through analyzing the optical flow field, we can extract the feature of hand gesture and recognize it. But the dimension of optical flow field is too high to calculate. Also, the noise signal is mixed within the optical flow field data, so we can't use the extracted data directly, we should firstly reduce the dimension and remove the noise by using the PCA method. The steps are as following:

1. For each hand gesture, totally within 2 seconds, 60 frames, we will get the optical flow feature as 32 * 20, because it has two components x and y. So for each frame, there will be 32 * 20 * 2 dimension as a column vector, for each hand gesture, there will be 60 optical flow field sequences, 60 column vectors.
2. We will apply the PCA method to extract the frontal 40 dimension eigen vectors.
3. In the recognition stage, we will project each optical flow into feature space to get the feature coefficient as the observation vector.

6. HAND GESTURE RECOGNITION USING HMMS

HMMs have been widely and successfully used in speech recognition and hand writing recognition [16]. Consequently, they seem to be effective for visual recognition of complex, structured hand gestures such as sign language recognition [17, 18]. A hidden Markov model (HMM) is a statistical model in which the being modeled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters from the observable parameters. The extracted model parameters can then be used to perform further analysis, for example, pattern recognition applications. An HMM can be considered as the simplest dynamic Bayesian network. Hidden Markov models are especially known for their application in temporal pattern recognition such as speech, handwriting, gesture recognition part-of-speech tagging, musical score following, partial discharges and bio-informatics.

We use HMMs to recognize different gestures because of their simplicity and reliability. The HMM use only three parameters: the initial state probability vector, the state-transition probability matrix, and the observable symbol probability matrix. Analysis of dynamic images naturally will yield more accurate recognition than that of a single static image. Gestures are recognized in the context of entire image sequences of non constant lengths. Using an HMM for gesture recognition is advantageous because it is analogous to human performance which is a doubly stochastic process, involving a hidden immeasurable human mental state and a measurable, observable human action.

Fig. 6 shows the architecture of Hidden Markov Model:

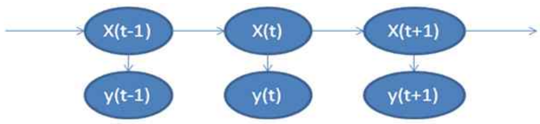


Fig. 6. Hidden Markov Model Architecture.

The diagram above shows the general architecture of a Hidden Markov Model (HMM). Each oval shape represents a random variable that can adopt a number of values. The random variable x(t) is the hidden state at time t. The random variable y(t) is the observation at time t. The arrow in the diagram denotes conditional dependencies.

The gesture recognition model designed in the work is a circular network of isolated gesture HMMs. Basically the network models the entire sequence of human subject motion. Non-gesture patterns are explained by the filler or garbage model denoted by "F" in the picture. There is only one type of non-gesture and thus only one filler HMM currently. The structure of the network can be described by the following generative rule:

1. <Gesture Motion> := <F> (<G><F>)*
2. <G> : = <G0> || <G1> || <G2> || ……. || <G9>
3. <F> : = <F1> || …… || <Fk>

Where <Gi> represents a gesture model for hand i and <Fi> represents a filler model where i=1, ….,k and k is the number of filler models. According to the above definition, we can build a network of HMMs as following:

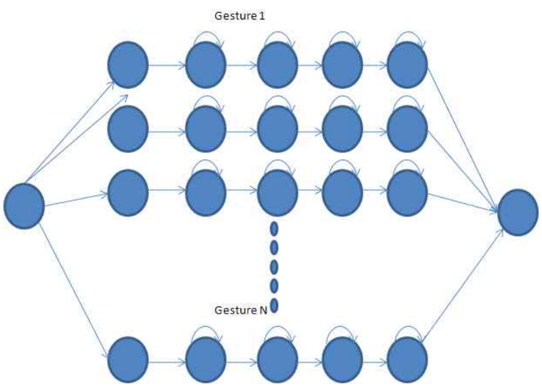


Fig. 7. HMM network for gesture recognition

Based on the features which are extracted from hand region and tracking process, we create a vector and use it as the input data of Hidden Markov Model. The output of these hidden models is the gesture that we want to recognize. The symbol generation process is illustrated in Fig. 8.

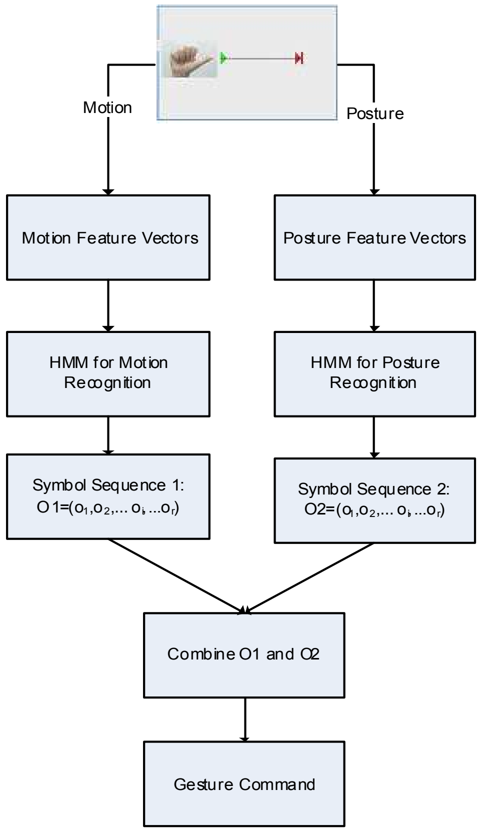


Fig. 8. HMM Model for Hand Gesture Recognition.

7. EXPERIMENT RESULTS

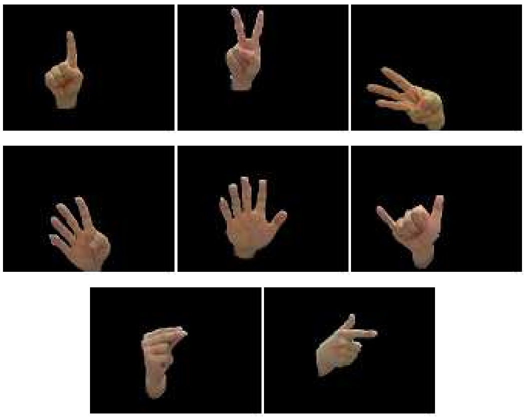
In the Experiment, we use a single hand to make hand gesture. And for evaluating the hand gesture tracking and recognition algorithm, we build the hand gesture database which contains 8 kinds of hand gesture performed 3 times by 20 different individuals. So there are 60 different image sequences capture for each hand gesture, and totally 480 image sequences are used for training. The size of each gray-level image frame is 320*240, its frame rate is 30 frames/sec, and each gesture-making takes about one second. Each gesture is standing before any stationary background with normal lighting. The proposed real-time tracking system can track and identify the moving objects in front of a stationary background.

Hand region detection will be disturbed by the regions with similar skin color. Such as face region, or any region closed to red. A correct detection can be considered as that the result with a correct palm center position, accurate hand edge information and accurate finger detection. The detection result can be shown in Fig. 9.

And a testing result in the hand posture database is concluded in Table 1. CD and MD denote the correct and missed detections respectively, while DR is the detection rate.

The same gesture made by different individuals may look different because of different hand-shapes and gesture speed. To design a robust rec-

(a)



(b)

Fig. 9. Hand Detection Result for Hand Gesture Database (a) Source Images and (b) Detection Result.

Table 1. Hand Region Detection Result Analysis

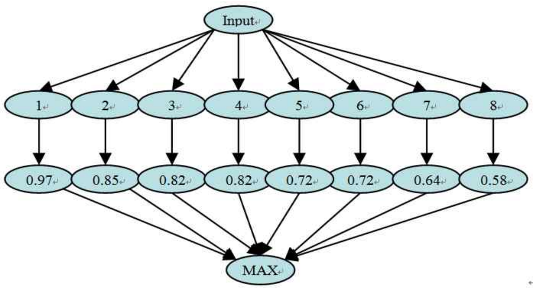
Background Complexity	Test Image	CD	MD	DR
Simple without Face Region	200	198	2	99%
Simple with Face Region	200	192	8	96%
Complex without Face Region	200	186	14	93%
Complex with Face Region	200	179	21	89.5%

ognition system, the training data are selected to cover all possible hand-shapes for each individual. Before using HMMs for training or recognition process, vector sequence is preprocessed to an observable symbol sequence O. The codebooks are created based on their corresponding training data. The codebook size M, which is power of 2, is chosen by experiments.

Totally 480 images sequences are collected for 8 different gestures, thus each kind of gesture with 60 sequences in average, in training phase and other 480 sequences are collected for test as shown in Table 1. The hand posture recognition rate of using training data for testing is 98%. The recognition rate of using testing data is 90.5%. And for hand motion, we extract the PCA feature of optical flow field for getting a precise description of motion. The recognition rate for hand gesture is about 93% in real-time. So the final hand gesture recognition rate is 84.16%.

Fig. 10 shows the results of the gesture recognition of the 1^{st}gesture in our vocabulary. Fig. 10(a) shows the sequence of observation symbols which is input to the hmm. Fig. 10(b) shows the output of the maximum likelihood of each HMM applied to the testing sequence. There are totally 8 HMMs in the recognition system of which first HMM generates the largest maximum likelihood. In our ex-

(a)



(b)

Fig. 10. Experiment results of recognizing the gesture from a sequence of frames: (a). Observation Symbols Sequences and (b) the Matching Result.

periments, we have tested 8 different gestures from different signers, some gestures are not precise, and the recognition rate drops to 70%. We find that our recognition system is insensitive to size and rotation, for small objects and for large objects, it can still effectively identify the correct gesture. However, the system can even make error recognition result because we don't have enough training data to make a good estimate of the HMM model parameters.

We have integrated the algorithm to a system. The system is composed of 3 parts:

1. MP3 player: in this part, one can use her/his hand to select the music which we want to listen, and use open hand posture to start the player. As seen at Fig. 11.
2. Hand gesture keyboard: in this part, one can use her/his hand to type the key in the screen, the figure can be shown as Fig. 12.

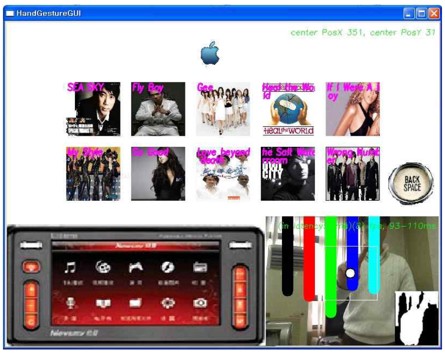


Fig. 11. Using Hand Gesture to Control MP3 Player.

Fig. 12. Using Hand Gesture to Control Keyboard.

3. Hand gesture game: in this part, we have developed a Bonga game which is based on hand gesture. As shown in Fig. 13.

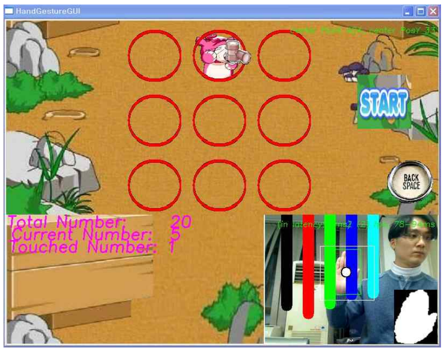


Fig. 13. Using Hand Gesture to Control Bonga game.

The main page is shown as Fig. 14.

Fig. 14 Main Page of Hand Gesture Control System.

In order to evaluate the performance of our proposed algorithm, we use the classical method in OpenCv (HandVu) for comparison.

The handVu hand gesture detection and recognition system which is used in OpenCV. The system use the pyramid L-K Optical flow algorithm

(a) (b) (c)

Fig. 15. Hand Posture in HandVu System

to track the hand, and use the Template based method to recognize the hand posture. In the system, it defines three kinds of hand posture: open hand, close hand and victory posture show as following:

The hand gesture is started when user with hand open posture, and stopped when user with hand close posture. The victory posture is used to make confirm command.

The following figure shows the tracking accuracy of two kinds of system testing in 100 hand moving videos.

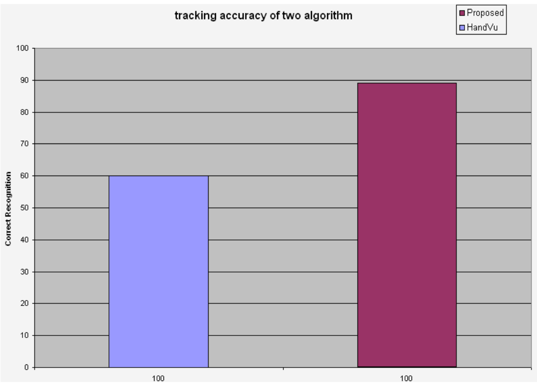


Fig. 16. Tracking Test in 100 Hand Moving Videos.

And the following figure shows the recognition rate of hand posture in histogram. The testing database contains 3 kinds of hand posture: open, close, victory, each posture with 100 testing images.

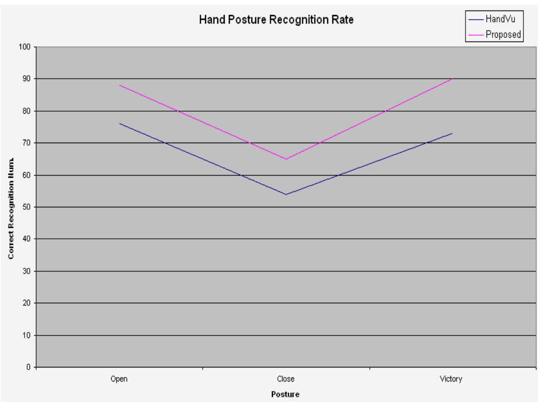


Fig. 17. Histogram of Hand Posture Recognition Rate.

From the results, we can conclude that for both hand detection and hand gesture recognition method. Our methods performed much better than the HandVu algorithm. That because HandVu gives a better hand detection algorithm which is based on optical flow method, but the method for extracting hand feature is so lack. It just use the skin color. So this may make some confused with face region, but in our method, we utilize both the skin color and the hand geometric information (hand region boundary energy) to avoiding the problem. And for recognizing hand posture, the HandVu method just use the mask pattern for each hand posture, it is very lack. But in our method, we use the feature based on FFT which can describe shape feature very strongly, and it is validated to shape rotation and transposition. Also for recognizing the hand posture, we use the artificial neural network to recognize the hand posture. All of these combination make the performance of our algorithm very strongly. But in Figure 17 we can also see that, the correct recognition number of hand close posture is down much than the other algorithm, that because the closing hand does not have complex boundary, so that make the feature based on boundary energy very weak. So it makes the recognition of closing hand leak. In our future recognition, it is a main field to be improved.

8. CONCLUSION

In this paper, we have presented a method to track and recognize the hand gesture by using optical flow and HMM algorithm. Since the moving hand can make an optical flow field, so based on the optical flow field and hand feature, we can detect the moving hand region. we use the region boundary complexity calculation to confirm the hand feature because of the complexity of hand boundary. And for describing the hand shape feature, we use the Fourier descriptor, for describing the hand motion we use PCA to compress the dimension of optical flow feature. We apply this system to recognize the single gesture to control IPTV controller panel. In the experiments, we test in different environment with different complex background. The experiment result shows that the system can get an accurate result with 96%. If we want to add a new gesture into the system, we only need to re-train a HMM for the new gesture.

REFERENCES

[1] Bai yu and Eung-Joo Lee, "The hand mouse: Hand detection and hand tracking," International Conference on Multimedia, Information Technology and its Application 2009(MITA 2009), pp. 244-245, 19-21 Aug. 2009.

[2] Bai yu, Sang-Yun Park, and Eung-Joo Lee, "Hand Posture Recognition Based on SVM Using on Mobile Phone," Conference on Korea Multimedia Society 2009, Vol.12, No.2, pp. 619-620, Nov. 2009.

[3] Eng-Jon Ong and Richard Bowden, "A boosted classifier tree for hand shape detection," in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Seoul, Korea, pp. 889-894, May 2004.

[4] Mathias Kolsch and Matthew Turk, "Robust hand detection," in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Seoul, Korea, pp. 614-619, May 2004.

[5] Lars Bretzner, Ivan Laptev, and Tony Lindeberg, "Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering," in Proceedings of Int. Conference on Automatic Face and Gesture Recognition. Washington D.C., pp. 423-428, May 2002.

[6] P. Viola and M Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of Computer Vision and Pattern Recognition. Hawaii, U.S. pp. 511-518, 2001.

[7] Nagel H.H, "Displacement Vectors Derived from Second-order Intensity Variations in Image Sequences," [J]. *Computer Vision Graph Image Process*, pp. 85-117, 1983.

[8] Nagel H.H and Enkermann W, "An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences," [J]. *IEEE Trans*, PAMI-5, pp. 565-593, 1986.

[9] HornB K.P and Schunck B.G, "Determining Optical Flow," [J]. *Artificial Intelligence*, No.17, pp. 185-203, 1981.

[10] P. Peer, J. Kovac, and F. Solina, "Human skin color clustering for face detection," In submitted to EUROCON 2003-International Conference on Computer as a Tool, 2003.

[11] Sang-Yun Park and Eung-Joo Lee, "Hand Gesture Recognition Algorithm Robust to Complex Image," *Journal of Korea Multimedia Society* 2010, Vol.13, No.7, pp. 1000-1015, July. 2010.

[12] W. Sharbek and A. Koshan, "Color Image Segmentation - a survey-." Tech. Rep. Institute for Technical Informatics. Technical University of Berlin, October, 1994.

[13] E. Person and K.S. Fu, "Shape Discrimination using Fourier Descriptor," *IEEE Transactions SMC,* Vol. No.3, pp. 170-179, 1977.

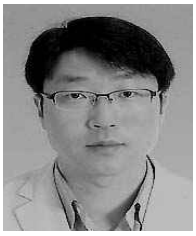
[14] D. Shridhar and A. Badreldin, "High-Accuracy Character Recognition Algorithm using Fourier and Topology Descriptor," *Pattern Recognition* Vol.17 pp. 515-524, 1984.

[15] Li wei and Eung-Joo Lee, "An IP-TV Remote Controller Based on Hand Posture Recognition Using Fourier Descriptor," Conference on Korea Multimedia Society 2009, Vol.12, No.2, pp. 510-511, Nov. 2009.

[16] A. Kundu and Y. He, P. Bahl, "Recognition of Handwritten Word: First and Second Order Hidden Markov Model Based Approach," *Pattern Recognition* Vol.22, No.3, pp. 283-301, 1989.

[17] Y. Yamato, J. Ohya. and K. Ishii, "Recognizing

Human Action in Time-Sequential Images Using Hidden Markov Models," Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp. 379-385, 1992.
[18] J. Schlenzig, E. Hunter. and R. Jain, "Recursive Identification of Gesture Inputs Using Hidden Markov Models," Proceedings Second Annual Conference on Applications of Computer Vision, pp. 184-194, 1994.



Sang-Yun Park

received his B.S. in Physics, M.S. in multimedia Application from Kyungsung University, Korea, in 1992, 1998, respectively. And now(2006-2011) he is a Ph.D student of Tongmyoung University in Korea.
His main research interests include in computer vision and biometrics.

Eung-Joo Lee

received his B.S., M.S. and Ph.D in Electronic Engineering from Kyungpook National University, Korea, in 1990, 1992, 1996, respectively.
In March 1997, he joined the Department of Information Communication Engineering of Tongmyoung University, Busan, Korea, as a professor. From 2005 to July 2006, he was a visiting professor in the Department of Computer and Information Engineering, Dalian Polytech University, China.
His main research interests are in image processing, computer vision, and Biometrics.