A Method of Hand Gesture Recognition based on Multiple Sensors

Fan Wei, Chen Xiang, Wang Wen-hui, Zhang Xu ,
Yang Ji-hai
Electronic Sci. & Technol. Dept.
Univ. of Sci. & Technol. of China,
Hefei, PRC
fanwii@mail.ustc.edu.cn
xch@ustc.edu.cn

Vuokko Lantz
Multimodal Interaction
Nokia Research Center
Helsinki, Finland
Vuokko.Lantz@nokia.com

Wang Kong-qiao
Nokia Research Center
Nokia (China) Investment CO.,LTD.
Beijing, PRC
Kongqiao.Wang@nokia.com

***Abstract*—This paper presents a new method of gesture recognition based on multiple sensors fusion technique. Three kinds of sensors, namely surface Electromyography (sEMG) sensor, 3-axis accelerometer (ACC) and camera, are used together to capture the dynamic hand gesture firstly. Then four types of features are extracted from the three kinds of sensory data to depict the static hand posture and dynamic gesture trajectory characteristics of hand gesture. Finally decision-level multi-classifier fusion method is implemented for hand gesture pattern classification. Experimental results of 4 subjects demonstrate that each kind of sensor data has its advantages and disadvantages in representing hand gestures. And the proposed method could fuse effectively the complementary information from these three types of sensors for dynamic hand gesture recognition.**

Keywords- Gesture recognition; multiple sensors fusion

I. INTRODUCTION

Nowadays, hand gesture pattern recognition based on multi-sensor fusion technique has become a research branch in the field of HCI. For instance, Brashear et al. [2] combined the vision and the accelerometer to recognize the mobile sign language. Chen et al. [3] used accelerometer and sEMG sensors synchronously to detect hand movement information for hand gesture recognition. Zou et al. [7] took advantages of data-glove, vision device and elbow bend sensors for Chinese sign language gestures recognition task. Apart from the related work, we propose a novel approach which can fuse effectively complementary information from three types of sensors, namely sEMG sensor, a 3-axis accelerometer (ACC) and a web cam, for hand gesture recognition. And we also probe the advantages and disadvantages of these three sensors in representing hand gestures.

II. METHODS

As illustrated in Figure 1, which demonstrates the implementation procedure of our proposed method, hand gesture recognition consists of signal acquisition, data segmentation, feature extraction, and decision-level fusion classification.

In signals acquisition phase, three kinds of sensory signals are recorded synchronously firstly. Then sEMG signals are applied to extract the active segments relative to hand gesture action from the recorded signals [6]. Feature extraction for each gesture is conducted on its multi-stream active segments. And decision-level multi-classifier fusion method is used for features classification.

A. *Gesture feature extraction*

In our method, a hand gesture is depicted with static hand posture and dynamic gesture trajectory. And the features extraction are categorized into the following two parts

1) Static feature extraction from sEMG and vision signals

The EMG time-series in each segment is decomposed into several pieces with the 50% overlapped window of 128ms firstly, and then each channel data in these pieces is represented by the first three coefficients of the fourth Auto-Regressive (AR) model and the Mean Absolute Value. So each piece is converted into a $4n$-dimensional vector, where n is the number of channels, and each sEMG segment is represented by a sequence of all feature vectors of pieces [6].

The Polygon Fourier Descriptor (PFD) [5] is applied to describe the visual hand posture feature. As shown in Figure 2, the procedure of the PFD-based hand posture feature extraction can be conducted as the following steps: *a)* Convert the contour of the hand posture (Figure 2b) into polygon (Figure 2c) using Douglas-Peucker algorithm [4]. *b)* Calculate the centroid c (the cross in Figure 2b) of hand posture region. *c)* Draw the connecting lines which between the vertex v_k of the polygon and the centroid point and designate the longest one as the major axis. As shown in Figure 2d, v_kc is one of the connecting lines. *d)* Compute the normalized length l_k of the connecting line $|v_k c|$ according to formula (1), and structure the complex *f(k)* with the angle θ_k between the connecting line and the major axis and l_k as formula (2) shows.

This research was supported by National Nature Science Foundation of China (NSFC) under Grant 60703069, National High Technology of Research and Development Program of China (863 Program) under Grant 2009AA01Z322.

978-1-4244-4713-8/10/$25.00 ©2010 IEEE

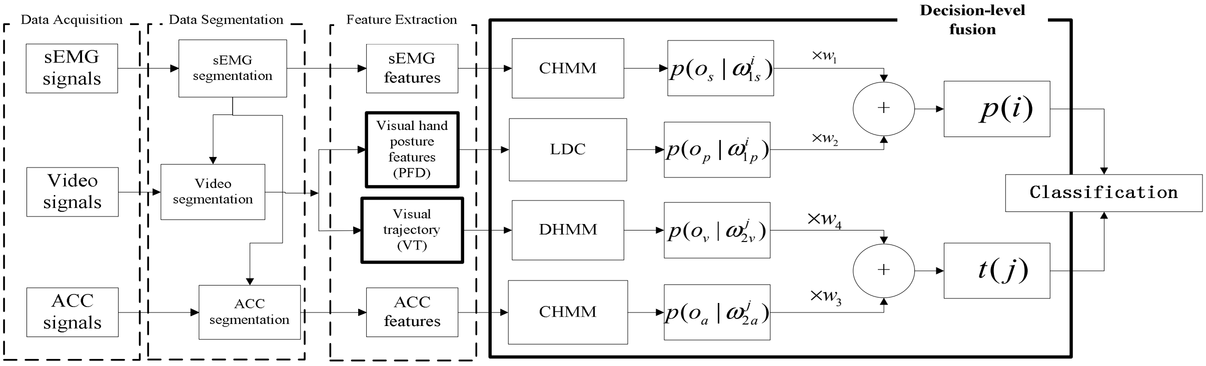

Figure 1. The implementation of gesture recognition based on multiple of sensors.

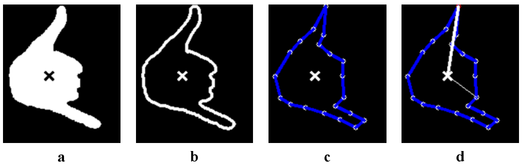

Figure 2. Illustration of PFD.

$$l_k = \frac{|v_k c|}{|major\ axis|} \tag{1}$$

$$f(k) = \theta_k + l_k j \tag{2}$$

e) Compute the Discrete Fourier Transform *F(n)* (namely PFD) of *f(k)* according to formula (3), where the *K* is the total number of the vertex.

$$F(n) = \frac{1}{K}\sum_{k=0}^{K-1}(\theta_k + l_k j)e^{-2\pi jnk/K} \tag{3}$$

2) Dynamic feature extraction

Dynamic characteristics of hand gestures are depicted by features extracted from acceleration signals and video. The measured three-dimensional ACC signals can directly represent the movement of hand when gesturing. ACC signals in each segment are normalized into [-1, 1] based on the min-max scaling method firstly, then the normalized signals is downsampled into 32 points as 3*32 ACC feature sequences. Thus eliminate the variations in gesture's scale and speed and improve the recognition accuracy of the gesture.

The visual trajectory is generated during tracking procedure. The CamShift algorithm [1] is applied to track the moving hand. For the simplicity, the frame differences between the successive frames and the skin color cue are used to acquire the initial location of the hand in vision field.

In order to obtain the 2D trajectory of the gesture, the centroids of the tracking object extracted using CamShift algorithm are recorded as the trajectory dots. Then the trajectory dots are converted into a sequence of directional code [8]. The angle *r* between the two adjacent points Dot1 (x_t, y_t) and Dot2 (x_{t-1}, y_{t-1}) is calculated according to formula (4) and encoded into a symbol number as figure 3 shows.

$$r = \arctan(\frac{y_t - y_{t-1}}{x_t - x_{t-1}}) \tag{4}$$

After this process, the trajectory dots are represented by a sequence symbol numbers.

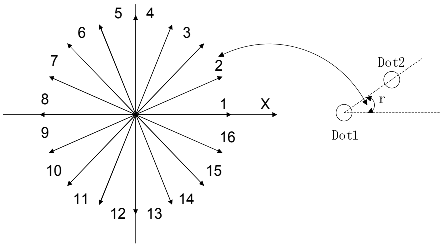

Figure 3. The encode of the direction angle.

B. *Decision-level Multi-classifier fusion*

Four kinds of features, namely sEMG features, PFD features, ACC features and visual trajectory (VT) features are used to represent different characteristics of hand gestures. In this article, the recognition of PFD features is resorted to Linear Discriminate Classifier (LDC) [5]. Continuous HMMs (CHMM) are used to model the sEMG and ACC, and the VT is structured by discrete HMM (DHMM).

Supposed a defined set contains C_1 static hand postures and C_2 dynamic trajectories, and the number of the hand gesture classes is $C_1 \times C_2$. The static hand posture set ω_1 and dynamic hand trajectory set ω_2 can be defined as formula (5) and (6) respectively,

$$\omega_1 = \begin{bmatrix} \omega_{1s}^1 \omega_{1s}^2 ... \omega_{1s}^{C_1} \\ \omega_{1p}^1 \omega_{1p}^2 ... \omega_{1p}^{C_1} \end{bmatrix} \tag{5}$$

where *s* represents the elements determined by sEMG features, and *p* represents the elements determined by PFD features.

$$\omega_2 = \begin{bmatrix} \omega_{2a}^1 \omega_{2a}^2 ... \omega_{2a}^{C_2} \\ \omega_{2v}^1 \omega_{2v}^2 ... \omega_{2v}^{C_2} \end{bmatrix} \tag{6}$$

Where *a* represents the elements determined by ACC features, and *v* represents the elements determined by VT features.

In the training phase, the feature sequences of sEMG and PFD are applied to train and establish the static hand posture CHMM model ω^i_{1s} and LDC model ω^i_{1p}, $i=1,2,...C_1$ respectively. Then ACC feature sequences and VT sequences are used to train and establish the dynamic hand trajectory CHMM models ω^j_{2a} and DHMM ω^j_{2v}, $j=1,2,...C_2$ respectively.

$$p(i) = w_1 p(o_s \mid \omega_{1s}^i) + w_2 p(o_p \mid \omega_{1p}^i) \tag{7}$$

$$t(j) = w_3 p(o_a \mid \omega_{2a}^j) + w_4 p(o_v \mid \omega_{2v}^j) \qquad (8)$$

Given O= [o_s o_p o_a o_v] (o_s, o_p, o_a, o_v are used respectively for representing sEMG, PFD, ACC and VT features) of an unknown hand gesture action, calculate the likelihood probabilities $p(o_s \mid \omega^i_{1s})$ that the sEMG features o_s subjected to ω^i_{1s} and $p(o_p \mid \omega^i_{1p})$ that the PFD features o_p subjected to ω^i_{1p} respectively, add the two kinds of probabilities with weight w_1 and w_2 as formula (7) shows, then find the class m which has the maximum weighted probability. The ACC features o_a and VT features o_v are processed as similar as sEMG, and then find the class n which has the maximum weighted probability of dynamic trajectory.

Finally, classify the unknown hand gesture action to the No. $m \times n$ class of the defined set.

III. Experiments

A. Hand gesture definition

In our study, a set with total twenty classes of hand gestures were defined for experiments. These gestures refer to four hand postures (Figure 4) and five hand movement trajectories (Figure 5). The four hand postures are indexed by T (extend the thumb), H (hand grasp), L (extend the little), S (extend both the thumb and the little) respectively, and the five trajectories are LTR (moving hand from left to right), UTD (moving hand from upper to down), CRW (drawing a circle with clock-wise direction), CRS (drawing a arc from left to right), and LRS (drawing a period of sin-wave from left to right).

B. Data Acquisition

A real-time platform was developed for three-sensor data acquisition. The hardware of the platform consists of 3-channel sEMG sensors, a 3-axis accelerometer, an off-the-shelf CMOS web cam, and a PC. The sEMG and ACC signals were acquired by Delsys Myomonitor IV sensor system, and digitized by a PCI Acquisition Card (PCI-6010 by NI) respectively. The web cam was connected to computer via USB. The real-time data acquisition software was also developed to acquire the data from these three kinds of sensors synchronously. The sample rate of the sEMG and ACC was set to 1 kHz and the web cam output the 640×480 images with the rate 25 frame/s.

The sEMG sensors were placed on the middle of forearm and the accelerometer was placed on the front-end of forearm, as shown in Figure 6. The web cam was placed in front of the subject about 1~1.5m. Four health subjects were recruited and each subject took part in five data acquisition trials in five different days. In each trial, twenty gestures were repeated 10 times respectively. In order to explore the practicability of gesture recognition based on multi-sensor fusion, no more restrictions were required for the subjects. The gestures were conducted depend on the subjects' own habits. And the video was collected under the complex background of our laboratory.

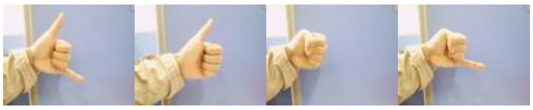

Figure 4. Four hand postures.

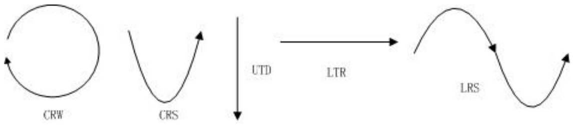

Figure 5. Five kinds of the dynamic trajectories.

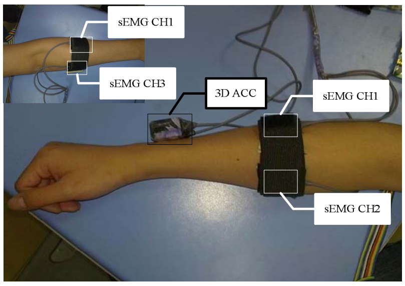

Figure 6. Sensor setup of data collection.

C. Experimental results

1) Hand posture and trajectory classification

The goal of this experiment is to explore the feasibility of hand postures classification using sEMG and PFD features for hand postures classification, and hand trajectories classification using ACC and VT features. For each subject, cross validation was implemented with data from four days for training and the left one for testing. Table Ⅰ to Table Ⅳ gives the recognition results of this experiment in the form of confusion matrixes. The confusion matrixes were obtained by summing up samples from 4subjects with five testing trials for each one.

From Table Ⅰ and Table Ⅱ, we can observe that the defined four hand postures can be effectively classified with sEMG features and PFD features. However, as shown in Table Ⅰ, it seemed the PFD features were better than sEMG features in discriminating four hand postures, especially in discriminating posture L and S. We also found that the significant individual differences among subjects' sEMG signals, and one of the 4 subjects got poor classification accuracy which resulted in the large Std.

From Table Ⅲ and Table Ⅳ which give the recognition results for five gesture trajectories, we can also observe that ACC features are more powerful than VT features. Since ACC features contain 3-axis information while only 2D information is extracted in VT features, we believe that the inherent difference between ACC and VT features result in this phenomenon.

2) Hand gesture classification

In order to evaluate the complementary functionality of different sensing techniques, seven types of experimental schemes, including sEMG, ACC, PFD+VT, PFD+ACC, sEMG+VT, sEMG+ACC, and sEMG+ACC+PFD+VT, were designed and implemented. The former three experimental schemes were conducted using only one kind of signal features, and the other four schemes applied different combinations of features. For the combination of four features, Table Ⅰ. Confusion Matrix of sEMG.

	H	L	S	T	Mean	Std
H	891	34	5	70	89.1%	21.3
L	2	774	178	46	77.4%	18.6
S	4	181	757	58	75.7%	11.3
T	57	31	53	859	85.9%	22.2

Table Ⅱ. Confusion Matrix of PFD.

	H	L	S	T	Mean	Std
H	867	42	67	24	86.7%	4.0
L	27	886	83	4	88.6%	4.9
S	1	24	950	25	95%	2.3
T	55	3	78	864	86.4%	7.3

Table Ⅲ. Confusion Matrix of ACC.

	CRS	CRW	LRS	LTR	UTD	Mean	Std
CRS	630	5	62	62	41	78.6%	15.7
CRW	39	649	52	11	49	81.1%	9.6
LRS	50	11	712	13	14	89.0%	7.0
LTR	39	8	8	725	20	90.6%	7.9
UTD	26	9	19	2	744	93.0%	7.5

Table Ⅳ. Confusion Matrix of VT.

	CRS	CRW	LRS	LTR	UTD	Mean	Std
CRS	532	213	15	15	25	66.5%	26.5
CRW	0	733	4	0	63	91.5%	13.1
LRS	53	200	522	0	25	65.3%	31.1
LTR	12	91	4	690	3	86.3%	10.3
UTD	65	58	27	0	650	81.3%	6.6

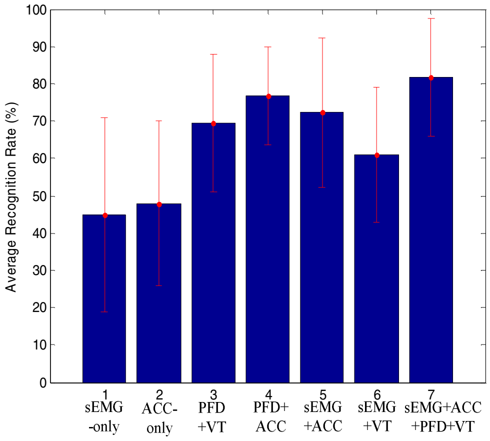

Figure7. The average recognition results of twenty gestures.

decision-level fusion classification method was applied and the weights were chosen to be 1:1 for the equal ability in recognition of these features. Cross validation experiments were conducted with four days' data for training and the left one for testing, and Figure 7 gives the average experimental results across subjects and all 20 hand gestures in the form of mean ± std. Following observations were made from Figure 7.

a) Since multiple sensors could acquire more complementary information than single sensor, the method based on multi-sensor fusion demonstrated good capability in hand gesture recognition. The combination of three kinds of sensors namely sEMG+ACC+PFD+VT scheme got the highest mean (81.7%) and low Std (15.8%).

b) Based on the experimental results of hand postures and trajectories classification, we found that PFD features are better than sEMG features in discriminating the four hand postures, and ACC features are more powerful than VT features for hand trajectories discrimination. The experimental results of the 20 hand gestures classification also support that conclusion.

c) Compared with sEMG sensor and 3-axis accelerometer, the camera showed good ability for hand gesture classification (69.4±18.4% for PFD+VT). This result is in accordance with our expectation because the vision signal was used both for features extraction of static hand postures and dynamic trajectories in our method.

IV. Conclusion and Future Work

This paper proposes a novel hand gesture recognition method, taking advantage of three kinds of sensory devices, includes sEMG sensors, a 3-axis accelerometer, and a web cam. Experimental results demonstrate that each kind of sensor has its advantage and disadvantage in representing hand gesture. And the proposed method can fuse effectively complementary information from these three types of sensors for dynamic hand gesture recognition.

However, as an exploring work for the application of hand gestures in HCI, there is no strict criterion for gestures performance during our experiment, so the hand gesture recognition rates seem to be lower than that of related literature [2, 3, 7]. The future work will focus on the advanced feature-level and decision-level information fusion algorithms for improving the accuracy of hand gesture recognition.

References

[1] G. R. Bradski and S. Clara, "Computer Vision Face Tracking For Use in a Perceptual User Interface," Intel Technology Journal, Vol. 2, 1998, pp.1-15.

[2] H. Brashear, T. Starner, P. Lukowicz, and H. Junker, "Using Multiple Sensors for Mobile Sign Language Recognition," In Proc. of the 7th IEEE Int. Sympos. Wearable Computers, 2003, pp 45-52.

[3] X. Chen, X. Zhang et al., "Hand Gesture Recognition Research Based on Surface Sensors and 2D-accelerometers," In Proc. of the 11th Int. Sympos. Wearable Computers, 2007, pp. 11-16.

[4] D. H. Douglas, and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," Canadian Cartographer, 1973, Vol 2(3).

[5] Wang Xiying, Doctor Thesis: Research on Key Issues of Vision-Based Interaction. 2007.

[6] X. Zhang, X. Chen et al., "Hand Gesture Recognition and Virtual Game Control Based on 3D Accdlerometer and EMG Sensors," IUI'09, 2009, pp. 401-405.

[7] W. Zou, K. Yuan et al., "A method for hand tracking and motion recognition in Chinese sign language," ICII 2001.

[8] Hyeon-Kyu Lee and J. H. Kim, "An HMM-Based Threshold Model Approach for Gesture Recognition," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 21(10), 1999, pp. 961-973.