Hand Gesture Recognition and Virtual Game Control Based on 3D Accelerometer and EMG Sensors

Zhang Xu, Chen Xiang, Wang Wen-hui, Yang Ji-hai
Electronic Sci. & Technol. Dept.
Univ. of Sci. & Technol. of China, Hefei; PRC
zhx90@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 describes a novel hand gesture recognition system that utilizes both multi-channel surface electromyogram (EMG) sensors and 3D accelerometer (ACC) to realize user-friendly interaction between human and computers. Signal segments of meaningful gestures are determined from the continuous EMG signal inputs. Multi-stream Hidden Markov Models consisting of EMG and ACC streams are utilized as decision fusion method to recognize hand gestures. This paper also presents a virtual Rubik's Cube game that is controlled by the hand gestures and is used for evaluating the performance of our hand gesture recognition system. For a set of 18 kinds of gestures, each trained with 10 repetitions, the average recognition accuracy was about 91.7% in real application. The proposed method facilitates intelligent and natural control based on gesture interaction.

Author Keywords:
Gesture recognition, human computer interaction, accelerometer, electromyogram.

ACM Classification Keywords:
H5.2. Information interfaces and presentation: Input devices and strategies (e.g., mouse, touchscreen).

INTRODUCTION
Hand gesture recognition provides an intelligent and natural way of human computer interaction (HCI). Its applications range from medical rehabilitation to consumer electronics control (e.g. mobile phone). In order to distinguish hand gestures, various kinds of sensing techniques are utilized to obtain signals for pattern recognition [5]. Acceleration-based and electromyogram-based techniques are two research branches in the field of hand gesture pattern recognition.

IUI'09, February 8–11, 2009, Sanibel Island, Florida, USA.

Acceleration-based (ACC-based) gesture control is usually studied as a supplementary interaction modality [6]. It is well suited to distinguish noticeable, larger scale gestures with different hand trajectories of forearm movements. With ACC-based techniques some subtle finger or hand movement may be ignored whereas electromyogram-based (EMG-based) gesture recognition techniques use multi-channel EMG signals which contain rich information about hand gestures of various size scales. Due to some problems inherent in the EMG measurements, including the separability and reproducibility of measurement, the size of discriminable hand gesture set is still limited to 4-8 classes [1,4].

In order to realize a natural and robust gesture-based HCI system, the selection of input hand gestures that are well discriminable from each other is of crucial importance. Considering the complementary features of ACC- and EMG-measurements, we believe that their combination will increase the number of discriminable hand, wrist and forearm gestures and the accuracy of the recognition system.

This paper describes the design of a novel hand gesture recognition system based on multi-channel EMG sensors and 3D accelerometer. Methods for hand gesture recognition based on EMG and ACC signals using multi-stream Hidden Markov Models are described first; and a control strategy for Virtual Rubik's Cube game is then proposed for the evaluation of the performance of the recognition methods. Next two kinds of testing schemes are implemented. The experimental results across five subjects show the great potential of our system for intelligent user interfaces. Finally, discussion and suggestions for future work are given together with conclusions.

RELATED WORK
Generally, existing sensing techniques for hand gesture recognition and interaction could be categorized into three main groups: vision-based, movement-based, and EMG-based techniques. Vision-based techniques can track and recognize hand gestures effectively [5]. At the same time, they are sensitive to user's circumstances such as background texture, color, and lighting. This limits their extensive application. Movement-based approach utilizes different sensors to

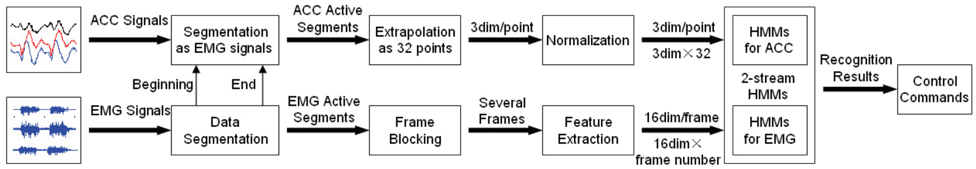

Figure 1: The structure of our hand gesture recognition system

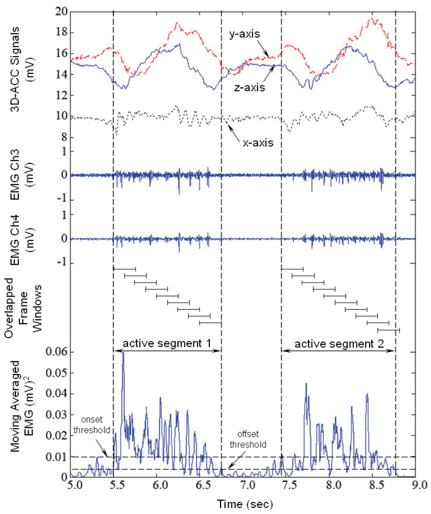

Figure 2. Illustration of data segmentation and EMG framing

measure movement. Glove-based gesture interaction is a typical movement-based technique and it achieves good performance especially in sign language recognition. In this approach user is required to wear a cumbersome data glove to capture hand and finger movement. This hinders the convenience and naturalness of HCI [5]. ACC-based approach is another popular movement-based technique. Accelerometers can be made easy to wear and are helpful in providing information about hand movements. The important advantage of EMG-based gesture interaction is its hands-free application. Yet, the current EMG-based HCI for fine control have a significant distance to the commercial applications [1,3].

Since each sensing technique has its own advances and capabilities, the multiple sensor fusion techniques can widen the spread of potential applications. For example, Brashear *et al.* [2] have built a lab-based sign language recognition system using both vision-based and accelerometer-based techniques. Their experimental results show that the combined sensing approach can improve recognition accuracy significantly. Sherrill *et al.* [8] have compared the performance of ACC-based and EMG-based techniques in the detection of functional motor activities for rehabilitation and provided evidence that the system based on combination of EMG and ACC signals can be built successfully. However, the ACC and EMG sensors fusion technique has not yet been applied to gesture-based interaction.

As for intelligent interaction, it is important to automatically specify the start and end points of a gesture action from continuous streams of input signals [5]. Yet, most of the previous work has taken this granted or have accomplished it manually. Mäntyjärvi *et al.* [6] utilized a sensing device, SoapBox, to realize ACC-based gesture interaction with fast and effortless customization. The gesture commands still had to be marked by pushing a button on SoapBox for temporal signal segmentation. Actually, it is easy and natural to detect muscle activation with EMG sensors, which indicate meaningful and intentional gestures. In our method, the start and end points of gestures are detected automatically by the intensity of EMG signals, and then both ACC and EMG segments are acquired for further processing.

METHOD

Hand Gesture Recognition Algorithms

Figure 1 shows the structure of our hand gesture recognition system using both multi-channel EMG and 3D ACC signals, which are recorded at 1kHz sampling rate. The processing of the two signal streams is carried out in following steps.

Data segmentation. The multi-channel signals recorded in the process of the hand gesture actions which represent meaningful hand gestures are called active segments. The gesture data segmentation procedure is very difficult due to *segmentation ambiguity* [5]. The EMG signal level represents directly the level of muscle activity. As hand movement switches from one gesture to another one, the corresponding muscles relax for a while, and the amplitude of EMG signal is momentarily very low. Thus the use of EMG signal intensity alleviates segmentation ambiguity and helps to implement data segmentation in multi-sensor system. In our method, the average signal of the multiple EMG channels is used for determining the boundaries of active segments. The segmentation is based on a moving average algorithm and thresholding. The ACC signal stream is segmented synchronously with the EMG signal stream.

Figure 2 illustrates the thresholding principles of the data segmentation method. The 3D ACC signal and two illustrational EMG signal channels are shown in Figure 2. First, the average value of the multi-channel EMG signal at time t is computed according to equation (1), where S is the number of EMG channels. Then, the moving average algorithm is applied with window size of W=60 sample points on the squared average EMG stream according to equation (2). Next, two thresholds, onset and offset threshold, are used for determining active segments. Typically, the offset threshold is lower than the onset threshold. The active segment begins when the moving averaged signal $EMG_{MA}(t)$ is above the onset threshold, and continues until all sample points in a 100ms time period have been below the offset threshold. The higher onset threshold helps to avoid false gesture detection whereas the lower offset threshold is for preventing the fragmentation of active segment as $EMG_{MA}(t)$ may oscillate around the onset offset during the gesture execution. As next step, segments whose lengths are less than a certain value (100ms) are abandoned as measurement noise. Finally, active gesture segments for both EMG and ACC signals are determined by the same beginning and ending points.

$$EMG_{averaged}(t) = \sum_{s=1}^{S} EMG_s(t) \quad \mathbf{(1)}$$

$$EMG_{MA}(t) = \frac{1}{W} \sum_{i=t-W+1}^{t} EMG^2_{averaged}(i) \quad \mathbf{(2)}$$

Feature Extraction. In active segments, the EMG stream is further blocked into frames with the length of 250ms at every 125ms utilizing overlapped windowing technique [1], as shown in Figure 2. Each frame in every EMG channel is filtered by Hamming window in order to minimize the signal discontinuities at the frame edges. Then, each windowed frame is converted into a parametric vector consisting of 3rd order Auto-regressive (AR) model coefficients and Mean Absolute Value (MAV) [1]. Hence, each frame of an n-channel EMG signal is presented by a 4n-dimensional feature vector, and the active EMG segments are represented by 4n-dimensional vector sequences of varying length.

Feature extraction of the 3D ACC stream in each active segment consists of two steps: scaling and extrapolation, with the method described in [6]. The amplitude of the 3D data in active segment is scaled using linear min-max scaling method. Then the scaled ACC active segment is linearly extrapolated to 32 points, so that the temporal lengths of all the 3D ACC data sequences are the same. These two steps normalize the variations in the scale and speed of gesture execution and thus improve the recognition of the type of the gesture. Normalized ACC active data segments are regarded as 3x32-dimesional feature vectors as such.

Hidden Markov Models for Recognition. The basic tool for the recognition of sequences of variable length is Hidden Markov Model (HMM). HMM represents a stochastic process that takes time series of observation data as input. The output of the HMM is the probability that the input data is generated by that model [5].

In our system, we utilize continuous density HMMs, where the observation data probability is modeled as a multivariate Gaussian distribution. Multi-stream HMMs [9] consisting of an EMG and ACC stream are proposed. The two-stream HMMs have the advantage that they can effectively combine EMG and ACC information, despite the time lengths of observation are different in the two streams. Each gesture class (or control command) ω_c is represented by two HMMs. There is a set of HMM pairs $\lambda_c=\{\lambda_c^{(E)}, \lambda_c^{(A)}\}$.

The logarithmic likelihoods $P(\boldsymbol{O}_t|\lambda_c)$ of a pair of EMG and ACC feature sequences $\mathbf{O}_t=\{\mathbf{O}_t^{(E)}, \mathbf{O}_t^{(A)}\}$ for the cth model is represented by the following expression [9]:

$$P(\mathbf{O}_t \mid \lambda_c) = \delta_E P\left(O_t^{(E)} \mid \lambda_c^{(E)}\right) + \delta_A P\left(O_t^{(A)} \mid \lambda_c^{(A)}\right) \quad 1 \le c \le C \quad \mathbf{(3)}$$

where t is the time, and $P(\boldsymbol{O}_t^{(E)}|\lambda_c^{(E)})$ and $P(\boldsymbol{O}_t^{(A)}|\lambda_c^{(A)})$ are logarithmic likelihoods for EMG feature sequence $\mathbf{O}_t^{(E)}=\{\mathbf{O}_1^{(E)}, \mathbf{O}_2^{(E)}, \ldots, \mathbf{O}_T^{(E)}\}$ and ACC feature sequence $\mathbf{O}_t^{(A)}=\{\mathbf{O}_1^{(A)}, \mathbf{O}_2^{(A)}, \ldots, \mathbf{O}_T^{(A)}\}$, respectively. δ_E and δ_A are EMG and ACC stream weight factors with the following restriction [9]:

$$\delta_E + \delta_A = 1,\ 0 \le \delta_E, \delta_A \le 1 \quad \mathbf{(4)}$$

The recognition result for an unknown gesture observation $\mathbf{O}_t$ is the class whose HMM pair achieves the highest combined likelihood for the given EMG and ACC observed feature sequences:

$$\hat{c} = \arg\max_c P(\mathbf{O}_t \mid \lambda_c), \text{ then } \mathbf{O}_t \in \omega_{\hat{c}} \quad \mathbf{(5)}$$

Virtual Game Control

The ACC and EMG signals are processed in real-time and the recognized gestures are translated into control commands in our interactive system. In order to assess the performance of our hand gesture recognition system and to create an enter-

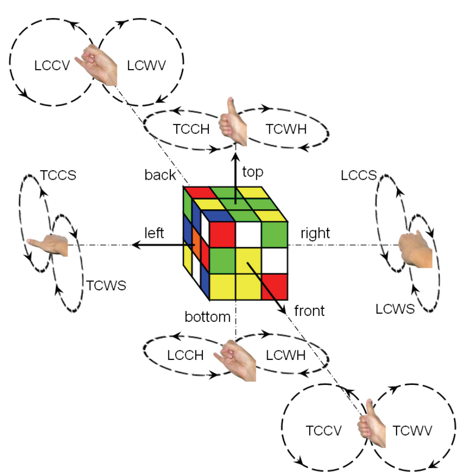

Figure 3: Twelve circular gestures to turn the six planar faces of the Cube

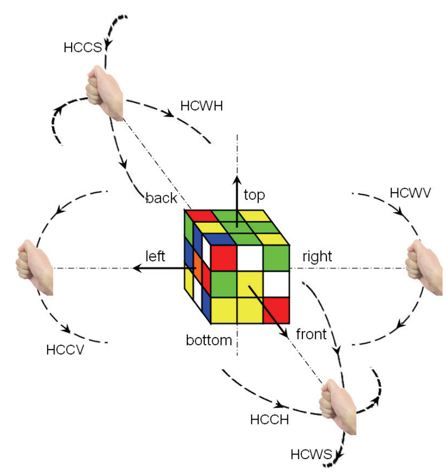

Figure 4: Six circular gestures to rotate the entire cube.

Finger action		Direction		Plane	
H	Hand grasp	CW	Clockwise	H	In left-front plane
T	Thumb	CC	Counter-Clockwise	V	In top-left plane
L	Little finger			S	In front-top plane

Table 1: Hand gesture name abbreviation

taining application of this novel HCI technique, Virtual Rubik's Cube game was built. It demonstrates the advantages of EMG and ACC combination by providing multiple degrees of freedom in control. Rubik's Cube is a mechanical puzzle invented in 1974. Please refer to [7] for more information about Rubik's Cube.

Utilizing the complementary characteristics of EMG and ACC signals, the set of defined hand gestures include both finger actions and circular hand movements of various orientations. Since any arbitrary transformation of the cube can be achieved by a series of steps of rotating the six external faces of the cube [7], we defined twelve circular gestures to rotate the six cube faces by 90 degrees clock-wise or counter-clockwise, as illustrated in Figure 3. Since the interface screen can only show three faces (e.g. the top, front and left as in Figure 3) of the cube at the time, six gestures with hand grasp (as shown in Figure 4) are proposed for rotating the entire cube by 90 degrees clockwise or counter-clockwise around three axes so that all six faces of the virtual cube can be brought into front view.

Each gesture defined is named by 4-letter abbreviation. These names indicate gesture meanings which are described in Table 1. With the Figure 3 and 4, it is intuitive to comprehend the gesture controls of Virtual Rubik's Cube. For example, gesture TCWH means thumb extension and hand circles drawn clock-wise in left-front plane. This gesture makes the topmost face of Virtual Rubik's Cube turn clockwise. Gesture HCCV means hand grasp and hands circles drawn counter-clockwise in top-left plane around front-axis. This gesture makes the entire cube rotate counter-clockwise around front-axis. With the gesture commands defined above, Virtual Rubik's Cube game can be played in a natural way. Figure 6 displays a subject playing with Virtual Rubik's Cube by using the hand gesture commands.

EXPERIMENTS

Experimental Setups

The EMG and ACC signal measurements were made with Delsys Myomonitor IV sensor system with inbuilt amplifier (60 dB). The sensor setup of four-channel EMG sensors and a 3D accelerometer are shown in Figure 5. In each EMG sensor, there are two silver bar-shaped electrodes with 10mm x 1mm contact dimension and 10mm electrode-to-electrode spacing. The four EMG sensors were attached to the inner side of a stretch belt, so the EMG sensors could be fixed conveniently in the middle of the subject's forearm. The 3D accelerometer (consisting of two mutually perpendicular 2D accelerometers) was placed on the back of forearm near the wrist. When user's palm was facing downwards, the three axes of accelerometer were pointing left, front, and top.

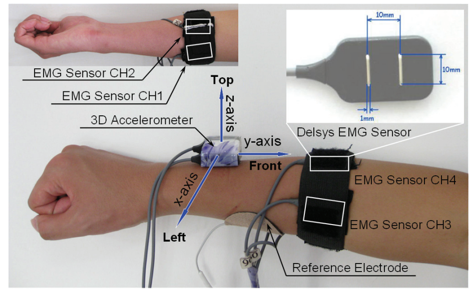

Figure 5: The sensor setup of four-channel EMG sensors and a 3D accelerometer.

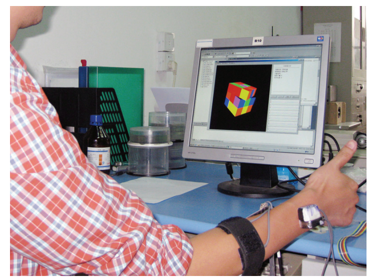

Figure 6: Subject performed hand gesture to control the Virtual Rubik's Cube.

Five subjects (2 males and 3 females) with ages ranging from 20 to 25 years participated into the hand gesture recognition experiments. These subjects have no history of neuromuscular or joint diseases and were informed of the associated risks and benefits specific to the study. Subjects signed an informed consent form prior to data collection and the experimental protocol was approved by the ethics committee of University of Science and Technology of China for human subjects. Each subject performed 18 hand gestures defined above in a sequence and in a way that felt natural to them. Each gesture was repeated 10 times for training the recognition system.

Left-to-right HMMs with five states were utilized in our system. The EMG and ACC HMMs were built independently. Both EMG and ACC HMMs were modeled using mixtures of three diagonal multivariate Gaussians distributions for each state. In the decision fusion step, the ACC and EMG stream weights were set equal: $\delta_E = \delta_A = 0.5$.

Two kinds of tests were implemented. In the first test, subjects performed gestures in order to confirm the validity of our hand gesture-based interaction method and to evaluate the system's recognition accuracy. In the second test, Virtual Rubik's Cube was scrambled in certain unsolved arrangements. Subjects tried to sort Virtual Rubik's Cube using the hand gesture commands. Since not all the subjects were familiar with the solution algorithms of Rubik's Cube, subjects were instructed how to solve the scrambled Rubik's Cube with 50 moves in advance. After that, they performed hand gesture commands step by step until the puzzle was solved.

Experimental Results

There were three experimental conditions in the first test: EMG-only, ACC-only and combination of EMG and ACC. Each gesture was performed more than 10 times for every test task. The average recognition accuracies for 18 gestures

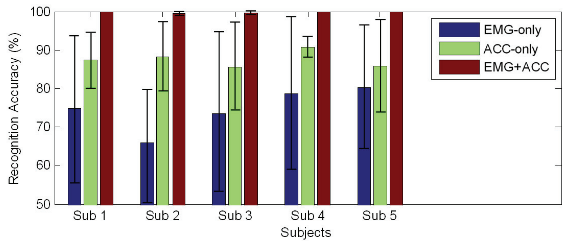

Figure 7: Average recognition accuracy (with standard deviation bars) for the five subjects

across five subjects are shown in Figure 7. The proposed method EMG+ACC achieved the highest accuracy, nearly 100%. Results for EMG-only condition were between 65.9-80.3% and for ACC-only condition between 85.5-90.7%. The large standard deviations of the accuracies for EMG-only and ACC-only indicate that several gestures are unclassifiable.

In the second test, every subject found it fun to play Virtual Rubik's Cube game. All the gesture commands were defined in pairs. If an occasional recognition error occurred, it seldom influenced the game: users could easily perform the gesture controlling the counter-action of the error command and continue to play. Table 2 shows the statistical results for five subjects. The recognition results achieved with our system were satisfactory as the overall accuracy was 91.7%.

The time delay between finished gesture command and system response as a cube action is less than 300ms. So the proposed system is capable for real-time operation [1]. The average input rate for gesture commands was about 16 per minute. These figures indicate that the proposed hand gesture-based interaction method is efficient.

Subject	Gesture Performed	Correctly Recognized	Accuracy (%)	Time Consumption	Commands per Minute
Sub 1	58	54	93.1	3' 28"	16.73
Sub 2	62	57	91.9	3' 55"	15.83
Sub 3	63	56	88.9	3' 43"	16.95
Sub 4	52	51	98.1	3'22"	15.44
Sub 5	66	58	87.9	4'11"	15.78
Overall	301	276	91.7	18'39"	16.14

Table 2: The statistical results for five subjects in puzzle solving mode

CONCLUSION AND FUTURE WORK

This paper has proposed a novel approach to hand gesture recognition which can be utilized in natural interaction between human and computers. The system combines information from a 3D accelerometer and multi-channel EMG sensors to achieve real-time hand gesture recognition using two-stream HMMs. Virtual Rubik's Cube game which is controlled by using hand gestures as input commands was introduced and implemented. Experiments were conducted to evaluate the performance of our gesture recognition system. For a set of 18 gestures, each trained with 10 repetitions, the overall recognition accuracy was about 91.7% in a real application. The proposed method facilitates quick and natural control in gesture-based interaction.

Our future work will focus on enhancing the robustness of the system and extending our methods to other types of applications, for example, to gesture-based mobile interfaces. We also intend to test the repeatability for multi-user and user-independent system. In addition, the design of tiny, wireless, and flexible sensors that are better suited for common users in real application is another goal of our research.

ACKNOWLEDGMENTS

This research is supported by National Natural Science Foundation of China under Grant Number: 60703069. We would also like to thank all the volunteers for their help in the experiments.

REFERENCES

1. Asghari Oskoei M. and Hu H. Myoelectric control systems—A survey. *Biomedical Signal Processing and Control*, Vol. 2(4), 2007, pp. 275-294.
2. Brashear H., Starner T., Lukowicz P. and Junker H. Using Multiple Sensors for Mobile Sign Language Recognition. In *Proc. 7th IEEE Int. Sympos. Wearable Computers*, 2003, pp. 45-52.
3. Costanza E., Inverso S.A., and Allen R. Toward subtle intimate interfaces for mobile devices using an EMG controller. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Portland, Oregon, USA, April 02-07, 2005, pp. 481-489.
4. Kim, J., Mastnik, S., and André, E. EMG-based hand gesture recognition for realtime biosignal interfacing. In *Proceedings of the 13th international Conference on intelligent User interfaces (IUI '08)*, New York, NY, pp. 30-39.
5. Mitra S. and Acharya T. Gesture Recognition: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 37(3), 2007, pp. 311-324.
6. Mäntyjärvi J., Kela J., Korpipää P., and Kallio S. Enabling fast and effortless customisation in accelerometer based gesture interaction. In: *Proc. 3rd Int. Conf. Mobile and ubiquitous multimedia*, New York, NY, USA, 2004, pp. 25-31.
7. Rubik's Cube http://en.wikipedia.org/wiki/Rubik's_cube
8. Sherrill D.M., Bonato P., De Luca C.J. A neural network approach to monitor motor activities. In *Proceedings of the Second Joint EMBS/BMES Conference*, vol.1, Houston, USA, 2002, pp. 52-53.
9. Tamura S., Iwano K., and Furui S. A stream-weight optimization method for audio-visual speech recognition using multi-stream HMMs. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 1, 2004, pp. I-857-60.