2016 IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS), 16-19 October 2016, Abu Dhabi, UAE

Real Time Algorithm for Efficient HCI Employing Features Obtained From MYO Sensor

Ehab H. El-Shazly[1], Moataz M. Abdelwahab[1], Atsushi Shimada[2] and Rin-ichiro Taniguchi[2]
[1] Electronics and Communications Engineering Department
Egypt-Japan University of Science and Technology (E-JUST), Alexandria - Egypt
[2]Kyushu University,744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan
E-mail:{ehab.elshazly, moataz.abdelwahab}@ejust.edu.eg, atsushi@limu.ait.kyushu-u.ac.jp, rin@ait.kyushu-u.ac.jp

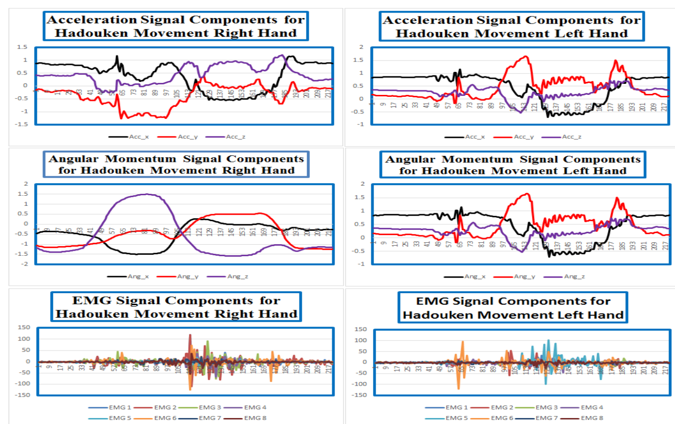
***Abstract*—This paper presents a new gesture recognition algorithm that uses different features obtained from MYO sensor. To preserve the spatial and temporal alignment for different features of each movement, Two Dimensional Principal Component Analysis 2DPCA is employed to obtain the dominant features by processing the obtained data in its 2D form. Canonical Correlation Analysis CCA is used to find a space where the projection of similar training/testing pairs become highly correlated. The testing sequences is matched to the training set that gives maximum correlation in the new space obtained by CCA. Two new data sets for common HCI applications (gaming and air writing) were collected at LIMU lab, Kyushu university and used to verify the efficiency of the proposed algorithm. Low processing complexity, efficient storage requirement, high accuracy and fast decision are factors that promotes our algorithm for real time implementation.**

I. Introduction

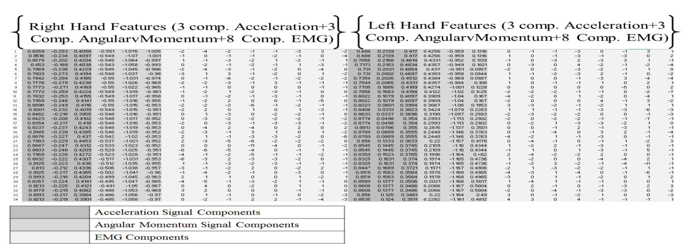
Recently, Human Computer Interaction (HCI) became very common, enabling different applications. Any interactive system depends on extracting discriminating features to train a recognition algorithm so that it can make true recognition decision during testing phase. Wearable devices of all kinds are becoming more popular. MYO sensor [1] developed by Thalmic Labs fitted around the upper part of the arm is able to detect muscle movements through capturing electromyography (EMG) data, acceleration as well as angular momentum. MYO armband consists of eight EMG-sensing modules, it senses the electrical impulses when the hands are moved, and transmit these data to computer using Bluetooth. The obtained features for each hand are, three components for acceleration, three components for angular momentum and 8 components for EMG. Figures 1 shows MYO sensor and sample of obtained data.

MYO sensor can measure acceleration from vibrations and the gravity so acceleration signal is more suitable for capturing noticeable, large-scale gestures [2], [3] which needs the movement of the whole arm. An accelerometer-based digital pen that uses acceleration signal was presented in [2] where time and frequency-domain features were extracted and kernel-based class separability was employed for selecting significant features. Linear discriminating analysis was employed for reducing the dimensions of features and the reduced features were sent to train probabilistic neural network PNN for recognition. Good recognition rate was achieved, but large training data and many pre-processing was required. In [3], the authors used measured acceleration and angular velocity signals to develop a smart assisted living SAIL system for the elderly and the disabled where five basic hand gestures were assigned to five commands. Their proposed algorithm used a NN segmentation module to realize gesture spotting and hierarchical hidden Markov model for classification. Coarse-grained classification was obtained from fusing the outputs of the NNs for the foot sensor and the waist sensor output. While

(a) MYO sensor.



(b) Different features obtained from MYO sensor for Hadouken movement.



(c) The obtained features arrangement for two arms.

Fig. 1: MYO sensor, obtained features and its arrangement.

fine-grained classification was done using the HMM module for strong displacement activities. Although The used data set in [3] was simple, reported accuracy still low.

EMG signals, which detect the activities of related muscles during gesture performance is more suitable for capturing fine movements [4], [5]. In [4], the authors reported a prototype of mobile-phone-based brain-muscle-computer interface for severely paralyzed persons. EMG activity on the surface of a single face muscle was measured with a standard electrode. The obtained analog signal was digitalized, split and filtered to extract total power within two separate frequency bands. After signal processing, the android phone sends commands to external devices via a bluetooth interface. The hardware architecture of this interface was presented, and the concept of using sEMG for fine details movement of the muscles was verified. In [5] the authors developed a Japanese text input method for mobile phones using surface electromyogram (sEMG). The proposed algorithm employed some signal processing techniques, generic dictionary and learning dictionary. sEMG was chosen, as the proposed application depends on the fine movement of the hand fingers.

In [6], the authors developed an algorithm that combine both acceleration and surface electromyographic (SEMG) signals for gestures recognition. The authors developed a prototype that was able to capture four sEMG signal and three acceleration signal components, also they collect a large data set containing 19 gestures varies between large scale and small scale gestures. Threshold segmentation technique was employed to detect the active segment in each movement. Different features and different classifiers for each gesture class were used. The reported algorithm required many processing to obtain the final feature descriptor, also it give response to each gesture instruction within 300 ms.

In our proposed algorithm, 2DPCA is employed to process the raw features from MYO sensor in its 2D form, maintaining only dominant features. CCA is employed to match the testing sequence to one of the obtained training sets depending on similarity measure. Our proposed algorithm can be adopted for either early gesture recognition or full gesture recognition at real time processing. This paper is organized as follows: Section II presents the proposed algorithm . Section III shows experimental results for street fighting game application. Section IV report the results obtained for air writing application. The conclusions are presented in section V.

II. Proposed Algorithm

The proposed algorithm basically consist of three building blocks, feature extraction, two Dimensional Principal Component Analysis (2DPCA) and Canonical Correlation Analysis (CCA). Figure 2 shows the proposed algorithm, first each recorded movement is stored in a matrix of size n by 28 where n is the number of frames per movement and 28 is the number of features vectors for both arms stacked together (6 vectors representing acceleration components of both hands, 6 vectors representing angular momentum and 16 vectors representing EMG).

To maintain the spatial and temporal relation between adjacent features, we process the features matrix in its 2D form.

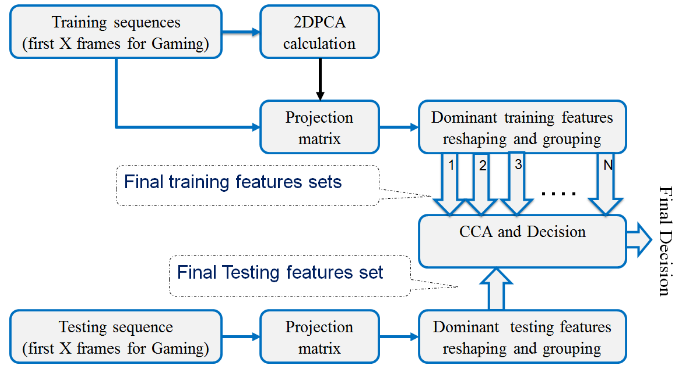


Fig. 2: Proposed algorithm.

2DPCA that maintains the spatial relation between adjacent components is preferred over traditional PCA to project the raw features matrix to the eigen space, where only the dominant eigen features are maintained, reshaped in to vectors. All eigen features vectors representing the same class are to be stacked together constructing the N final training sets to be stored and used during the testing phase, where N is the number of classes in the used data set. Testing sequence is projected to eigen space using the same projection matrix obtained in the training phase. The dominant features of the testing sequence is reshaped and matched to one of the obtained training sets using CCA. It transforms both testing set and each of the training sets to a new space where similar pairs become highly correlated. According to the application, we can use the first X frames for recognition in case of early recognition for gaming or we can use the whole resized sequence for air writing application. Detailed steps of the proposed algorithm are summarized as follows:

1) Read all training sequences to a 3D matrix M of size $(X * Y * K)$, where X represents window size used for early recognition or the unified size of all sequences for air writing, Y represents the number of used features' vectors (28 for both arms in gaming application and 14 for air writing application as gestures are performed using just one arm), and K represents the number of all training sequences.
2) The covariance matrix S, of size $(Y * Y)$, for the K training windows is calculated as follows:

$$S = \frac{1}{K}\sum_{J=1}^{J=K}\left(M_{X\times Y\times J} - \bar{A}\right)^{T}\left(M_{X\times Y\times J} - \bar{A}\right)$$

Where $\bar{A}$ is the average matrix of all K training windows of size$X$*Y.
3) A set of y eigen vectors, V_q of size (X x 1) corresponding to the dominant eigenvalues q, where q=1, 2,..., y, are obtained for matrix S.
4) Stack only dominant eigen vectors representing almost 99% of the data contained in the training data to one projection matrix $V = [V_1 V_2 V_3 ... V_y]$ of size $X * y$.
5) Project the transpose of each features sequence of size $Y * X$ using the obtained projection matrix to get a

projected features matrix of size $Y * y$.

6) Reshaped each projected features matrix in to one vector of size (Y*y by 1), group and stack all features of each class together. For any data set containing N classes and, in case of using J sequences for each class in the training phase, we will get N*J features vectors each of size (X*y).
7) Stack all the obtained features vectors from the same class together side by side to construct discriminating features sets. At the end of step 7, we will get N features set equal to the number of used classes, each of size (Y*y) by (J) representing training features sets or classes.
8) Testing sequence is first projected using the projection matrix V obtained in the training phase to yield a projected feature matrices of size (Y by y).
9) The obtained projected features matrix in step 8 is to be reshaped in to features vector each of size (Y*y by 1).
10) Canonical correlation is used to transform testing set and all training sets to a new space where similar training/testing pairs become highly correlated.

Canonical Correlation Analysis for classification depends on finding two sets of basis vectors such that the correlation between the projections of both the testing features vector and the stored training features sets on these basis vectors is maximized. Given two random vectors $f \in R^{m_1}$ and $g \in R^{m_2}$, a pair of transformation u and v called canonical transformation is found so that the correlation between $f^{'} = u^t f$ and $g^{'} = v^t g$ is maximized.

$$R = max_{u,v} \frac{E[f^{'}, g^{'}]}{\sqrt{E[f^{'2}]E[g^{'2}]}} = \frac{u^t C_{fg} v}{\sqrt{u^t C_{ff} u v^t C_{gg} v}} \quad (1)$$

R is the canonical correlation. where $E[h]$ is the expectation of h and C_{ff}, C_{fg} and C_{gg} are covariance matrices.

III. Gaming Application

For gaming, early recognition is essential which means that the algorithm is required to make recognition decision long before the movement ends using as minimum as possible number of frames. In that sense, sample movements from street fighting game [7] were collected to construct the used data set. Our newly collected data set consist of nine movements including Gaurd, Hadouken, Jump, Strong Punch, Weak Punch, Power Up, Strong Shoryken, Weak Shoryken and Squat. this data set was collected at LIMU lab, Kyushu university using two MYO sensor for both arms at recording rate of 50 Hz. Each movement was repeated 37 times with average number of frames per movement equal to 120 frame. 20 sequences were used as training set and the remaining 17 sequences used as testing set. This experiment was conducted to examine the efficiency of the proposed algorithm for gamin and its ability to discriminate between movements from the same class but performed with different velocities and strength. Only the first 45 frames were used for recognition as they represent almost one third the average sequence length. The used 45 frame window was picked up from 5 different positions along the sequence beginning, with a step of 10 frames. For example gesture $G1$ will give 5 sub windows $G1_1$ (from frame 1:45), $G1_2$ (from frame 11:55), ...and $G1_5$ (from frame 41:85). As some movements ends at only 85 frames, we pick training windows from only five positions and stopped at frame number 85. Segmentation step resulted in overall number of training windows of 900 (9 gestures * 5 sub windows * 20 repetition). we used the stacked acceleration and angular momentum components as features descriptor, as all the used gestures belong to large scale gestures, obtained accuracy are listed in table I.

TABLE I: Achieved accuracy for street fighting game

Used Features	ACC. + ANG.
Proposed Technique	98.2%
PCA+CCA	36.1%
2DPCA + NN	96.3%
2DPCA + Correlation	96.3%

The proposed algorithm was able to discriminate between challenging movements of the same trajectory with different strength. The experiment was conducted using Matlab 2015b running on machine with the following specification: Intel(R) Xeon(R) CPU E5-1620 V2@ 3.70 GHZ, with installed RAM 15.9 GB. The run time estimated to be 12.2 ms which is promising for real time implementation. To verify the efficiency of the proposed algorithm, we compare our results to other well known techniques (PCA and minimum distance classifiers). From table I, it is clear that 2DPCA outperforms the traditional PCA, that is because it maintains the spatial relation between adjacent features' components and temporal alignment of consecutive frames which is not the case with PCA. Also CCA outperforms traditional correlation or minimum distance classifier as it first transform testing set and all training sets to new space where similar pairs become highly correlated pairs and the training set that gives the highest CC value is recognized as the target set. The achieved excellent accuracy, fast run time and low computational complexity promotes our proposed algorithm for real time implementation. Most of the wrong classified testing sequences belongs to similar classes either the case of same class with different strength and speed or classes with similar trajectories.

IV. Air Writing Application

Air writing for human computer interaction was chosen to test our algorithm and a large data set containing 10 gestures representing digits from zero to nine was collected at LIMU lab, Kyushu university using one MYO sensor. We first align all sequences by clipping the beginning part that precede the active segment as it usually contains noise and do not have useful data. Simple threshold is used to detect the active segment position then, all obtained clipped sequences are to be resized to obtain equal length sequences. After resizing we directly follow the same mentioned steps in section II. Different experiments and different training/testing splitting scenarios were conducted.

A. Experiment 1

In this experiment, ten gestures representing digits from zero to nine as shown in figure 3 were recorded by three

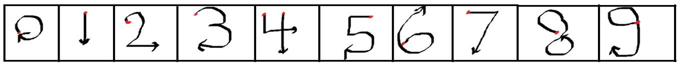


Fig. 3: Ten gestures air writing data set.

different performers, each gesture was repeated 20 times by the same person resulting in a total number of sequences equal 600. All the used ten gestures belongs to large scale gestures, so acceleration signal is the best choice to represent these gestures. Two training/testing splitting scenarios were tested. First, 40% of the sequences served as training set and 60% served as testing set. 8 repetitions for each person were picked out randomly to train the system and the remaining 12 repetitions were used for testing. 10 random runs were used to select the training and testing sub sets. The second splitting scenario was leave one person out, where all the sequences of one person were used as testing set and the sequences of the other two persons serves as training set. The obtained results listed in table II is the average results.

TABLE II: Achieved accuracy for experiment 1

Used Features	ACC.
40% Training, 60% Testing scenario	98.8%
Leave One Person Out scenario (LOPO)	90%

TABLE III: Run time and used Eigen vectors for experiment 1

Used Features	40% Training, 60% Testing Scenario	LOPO Scenario
Used Eigen Vectors	40	125
Run Time (msec)	3.5	11.5

Number of eigen vectors that contains most of the energy are used to construct the features vector. As shown in table III, the number of used eigen vectors changes according to energy distribution in each scenario resulting in changing the size of final descriptor and consequently the recognition time. The achieved real recognition time in millie seconds, simple processing steps and excellent accuracy verified the efficiency of the proposed algorithm for air writing application.

B. *Experiment 2*

In this experiment, a new edited data set was created from the recorded one. It contains 20 gestures representing digits consisting of two and three digits (0-9, 10, 21, 32, 43, 54, 65, 76, 87, 98, 100) as shown in figure 4. The same train/test splitting scenarios as experiment one were repeated. Listed results in table IV prove the efficiency of the proposed algorithm even when the number of used classes was doubled.

TABLE IV: Achieved accuracy for experiment 2

Used Features	ACC.
40% Training, 60% Testing scenario	98.7%
Leave One Person Out scenario	86.3%

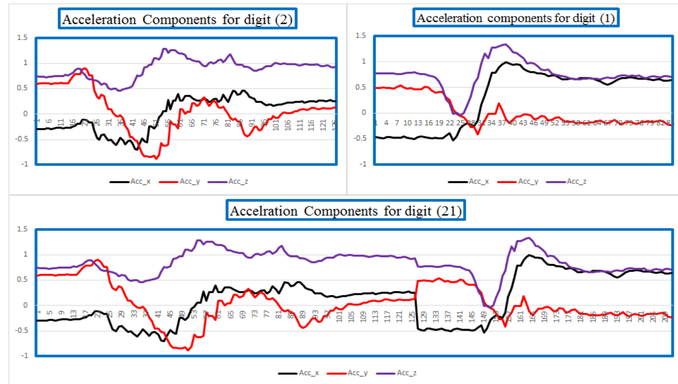


Fig. 4: Two digits gesture construction for experiment 2.

C. *Experiment 3*

In this experiment, we tested our algorithm for the data set collected in [6] for comparison. It was collected employing a prototype developed by the authors and consist of 10 digits performed by 20 performers (13 male and 7 females) each gesture was repeated 10 times resulting in overall 2000 sequences. In [6], the authors employed threshold classifier, Bayes linear classifier and an improved dynamic time-warping algorithm for classification. Various features and segmentation steps for both acceleration and sEMG signals were needed and a score-based classifier was used to find the final decision. Our algorithm uses only the raw acceleration signal as feature descriptor and achieved comparable results of 89.5 % using 40% training, 60% testing scenario .

V. Conclusion

In this paper a robust algorithm for HCI applications was presented and tested. Two newly collected data sets for two common applications (gaming and air writing) were collected using MYO sensor. Dominant features were extracted using 2DPCA, reshaped and classified using CCA. Different experiments verified the efficiency of the proposed algorithm for real time implementation as it achieved high accuracy, low computational complexity and fast recognition decision.

References

[1] http://www.thalmic.com/myo/techspecs, accessed August, 2015.

[2] J.WangandF.Chuang,"An accelerometer-based digital pen with a trajectory recognition algorithm for handwritten digit and gesture recognition," *IEEE Trans. Ind. Electron*, vol. 59, no. 7, pp. 2998-3007, Jul. 2012.

[3] C. Zhu and W. Sheng,"Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 3, pp. 569-573, May. 2011.

[4] S. Vernon and S. S. Joshi,"Brain-muscle-computer interface: Mobile-phone prototype development and testing," *IEEE Trans. Inform. Technol. Biomed.*, vol. 15, no. 4, pp. 531-538, Jul. 2011.

[5] A. Hatano, K. Araki, and M. Matsuhara,"A Japanese input method for mobile terminals using surface EMG signals," in Proc.22nd th Annu. Conf. Jpn.-Soc. Artif. Intell., 2009, pp. 5–14.

[6] Zhiyuan Lu, Xiang Chen, Qiang Li, Xu Zhang, and Ping Zhou,"A Hand Gesture Recognition Framework and Wearable Gesture-Based Interaction Prototype for Mobile Devices," *IEEE Transaction on Human-Machine Systems.*, vol. 44, no. 2, pp. 569-573, 2014.

[7] http://www.streetfighter.com/, accessed August, 2015.