2011 Seventh International Conference on Intelligent Environments

A Concept Grounding Approach for Glove-Based Gesture Recognition

Yu Huang, Dorothy Monekosso, Hui Wang, Juan Carlos Augusto
School of Computing and Mathematics, University of Ulster, Jordanstown,
Co. Antrim, BT37 0QB, UK
{huang-y4, DN.Monekosso, h.wang, jc.augusto}@ulster.ac.uk

***Abstract*—Glove-based systems are an important option in the field of gesture recognition. They are designed to recognize meaningful expressions of hand motion. In our daily lives, we use our hands for interacting with the environment around us in many tasks. Our glove-based gesture recognition is focused on developing technologies for studying the motion and interaction with a data glove which can augment the capabilities of some users to perform some tasks. This idea is relevant to many research areas, for example: design and manufacturing, information visualization, robotics, sign language understanding, medicine and health Care.**

In this paper, we proposed a new concept grounding approach for glove-based gesture recognition. We record the data from finger sensors and then abstract and extract concepts from the data. This allow us to construct conceptual levels which we can use to study interaction and manipulation for users during their activities.

Keywords- gesture recognition; concept grounding; data glove; clustering ensembles

I. INTRODUCTION

The use of multimodel inputs such as facial expression recognition, audio expression recognition and hand gesture recognization provide interesting matural interfaces to intelligence environments. Hand gestures are the most frequently used in our daily lives, and glove-based systems play an important role in the field of gesture recognition. Gesture spotting implies determining the start and end points of a meaningful gesture pattern from a continuous stream of input signals. Four main approaches of hand gesture recognition have been used: HMMs (Hidden Markov Models)[1,2,3], Particle Filtering and Condensation Algorithm [4,5,6], FSMs (Finite State Machines)[7,8,9] and Connectionist Approach[10]. They have been widely used in many areas, such as design and manufacturing, information visualization, robotics, sign language understanding, medicine and health Care.

Tom Mitchell proposed a new notion called "robotics concept grounding"[11] in the robot learning area. We can import this notion into glove-based gesture recognition. Concept grounding can abstract and extract concept from the raw data, and then build the concept space, so that the learner can make a plan in a higher level (Figure 1). The framework uses generalisation and abstraction techniques to analysis the raw data in the state space and action space and generate some basic concepts which can abstract the raw data cluster. And then create the behaviours between state concepts and action concept. This paper aims to generate abstract concepts for glove-based gesture recognition.

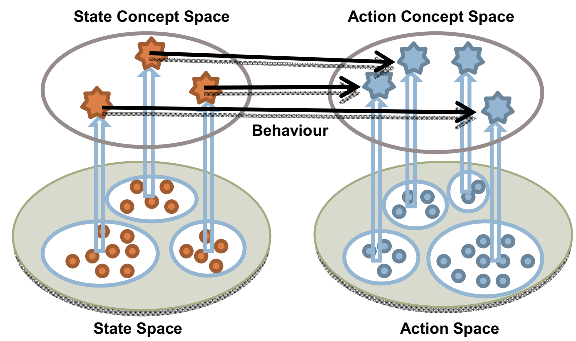


Figure 1. Concept Abstract & Extraction

Pejman Iravani[12] proposed a new concept grounding method based on Behaviour-based (BB) architectures[13] which have been successfully applied to some robotic applications. The main contribution of this method is the integration of techniques for generating grounded symbols from raw sensory-motor data and traditional behaviour-based architectures. He used HHI K-means to cluster the robot data and obtain the abstract concept, and then control the robot by using abstraction and generalisation techniques. But the function of clustering method (HHI K-means) is limited. It abstracts the concepts with the K-means cluster centers that can not describe the intent of actions in a noisy environment.

In fact, most concept formation models are conceptual clustering systems[14,15,16]. In most cases, the concepts can be represented as clusters. But the sensory data is usually very noisy. The sensor readings and actuator control are subject to inaccuracy due to noise. Although many clustering methods exist, no single algorithm can handle all kinds of dataset in noisy environment.

Each algorithm has its own approach to make a description of the number and shape of clusters based on their particular view of the data. It's difficult to choose an algorithm which would show a best performance in a particular data set. Thus we need to

978-0-7695-4452-6/11 $26.00 © 2011 IEEE
DOI 10.1109/IE.2011.51

employ fusion techniques to generate the suitable concepts for control. Clustering ensembles and multi-classifiers are effective data fusion methods in data mining domain, we can use these techniques in our research. In the previous work, we have proposed a novel method named CEMF[17], i.e., Clustering Ensembles based on Multi-classifier Fusion. In this paper, we used it to create abstract concept for glove-based gesture recognition.

In this paper, we proposed a new concept grounding approach for glove-based gesture recognition, The rest of the paper is organized as follows: Section 2 describes the basic idea of CEMF. The experimental results are reported in Section 3. And then section 4 provides conclusions and our future research work.

II. THE METHODOLOGY OF CEMF

A. *The framework of CEMF*

CEMF generates ***H*** different partitions using different clustering algorithms and matches clusters in different partitions in order to create the high confidence clusters which can be considered as abstract concepts. And then it constructs the MCSs(Multiple Classifier Systems) based on the performance of the classifiers in different subspace and choose the local optimum classifiers to classify the uncertain instances and new instances into different clusters(concepts). From the different purpose, in this paper we don't need to deal with the uncertain instances which could be considered as noisy data in some cases. As a recognition method, we can focus on how to classify the new instances.

First of all, we need to give some some definitions and notation. Let ***D*** be a gesture dataset, ***D***=$\{d_1, d_2, \dots, d_n\}$ with n instances. Each instance can be considered as a simple gesture. β is a new instance which is waiting to be classified. ***H*** is the number of partitions (i.e., ensemble size) , ***k*** is the number of clusters. Each cluster can be considered as a abstract concept here. Assume ***L*** is a set of cluster labels denoted as ***L***=$(l_1, l_2, \dots, l_k)$ and f is a maping function, it satisfies $f(d_i)\in$ ***L***. Assume ***Π***=$\{\pi_1, \pi_2,\dots,\pi_H\}$ is the set of ***H*** partitions, $\pi_i=\{C^i_1, C^i_2,\dots, C^i_n\}$ ($1\le i\le H$), where C^i_j is the j[th] cluster in the partition π_i. Let $\pi_i(d)$ be the cluster label of a instance d in partition π_i.

Assume ***S***=$\{s_1, s_2, \dots, s_n\}$ is the set of subspace, and Let ***S***(d) be the subspace for instance d. ***A***=$\{a_1, a_2, \dots, a_m\}$ is the set of classifier, m is the number of the classifiers which are used to build MCSs. ***θ*** is a predefined threshold of clustering consistency index. The detailed CEMF method is described as follows:

Input: ***D,H,k, Π, θ, A, L,*** β

Step1: Generate ***H*** different partitions on the dataset ***D***.

Step2: Match clusters in different partitions with Jaccard index method[18]

Step3: For every instance d in ***D***, calculate **C*I***(d) -the value of clustering consistency index[19]. After the clusters match, we can find that some instances are steadily assigned to the same cluster and some instances are not. So we use the ***CI*** function to select the the instances with the stable cluster assignment to create the high confidence clusters.

$$CI(d)=\frac{1}{H}\max\left\{\sum_{i=1}^{H}\delta(\pi_i(d),l)\right\}_{l\in L},$$

$$\delta(a,b)=\begin{cases}1, if\ a=b\\ 0, if\ a\neq b\end{cases}$$

$f(d)$ shows the cluster lable for d:

$$f(d)=\begin{cases}\arg\max\limits_{L}\left\{\sum_{i=1}^{H}\delta(\pi_i(d),l)\right\}_{l\in L}, if\ CI(d)\geq\theta\\ -1, if\ I(d)<\theta\end{cases}$$

Step4: For each instance d with $f(d)\neq-1$ in ***D***, divided d into new high confidence clusters $\boldsymbol{N_1}$, $\boldsymbol{N_2}$, …, $\boldsymbol{N_k}$ according to $f(d)$.

Step5: For i=1 to ***H*** do add $\boldsymbol{N_i}$ into the subspace-***S***(β).

Step6: For each instance d with $f(d)$= -1 in ***D***,
For i=1 to ***H*** do If $\delta(\pi_i(d),l_i)=1$ then
add $\boldsymbol{N_i}$ into the subspace-***S***(d).

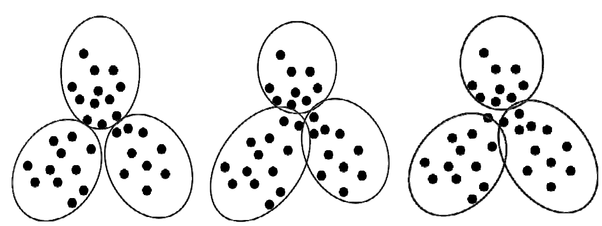
Step7: In the each subspace ***S***(d), Choose the best classifier α and then use α to classify d in ***S***(d).

$$\alpha(d)=\arg\max_{A}\left\{P(a)\right\}_{a\in A},$$

$P(a_i)$ is the classification accuracy of a_i in ***S***(d) with 10-fold cross validation.
And do the same process in the ***S***(β).

Step8: Ouput the cluster labels of instance d in ***D*** and the label of β**.**

B. *A graphic illustration*



a b c

Figure 2. The 3 different clustring results

In order to grasp the idea, we decribe the process by a graph in a 2-dimensional space. We use three different clustering methods to generate multiple partitions and get the 3 different clustering results(Figure 2). After that we match the clusters in different partitions and then create the 3 high confidence clusters with ***k***=3 and the threshold ***θ***=1. The result is showed in Figure 3.

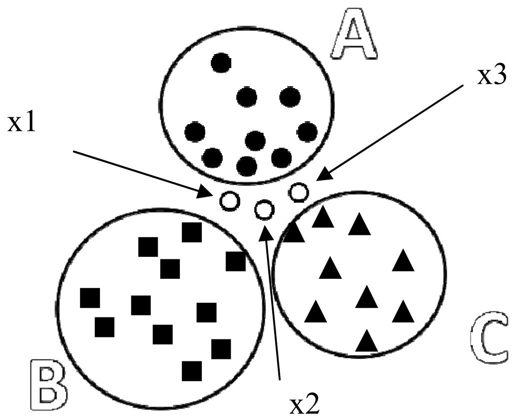


Figure 3. The 3 different clustring results

In Figure 3, there are 3 high confidence clusters which can be considered as abstract concepts. We give the concepts 3 class lables : A, B and C. But there are 3 instances(x1, x2, x3) in the uncertain area without the class lables. In the original CEMF method, we need to create the subspaces based on the multiple partitions for them. Figure 4 shows 3 subspaces. The instance x1 used to be covered by two clusters, so it is assigned to the subspace 1 which contains two classes (A and B). For the same reason, x2 and x3 are assigned to the supspace 2 and subspace 3 and then we build the MCSs to classify them. But in this paper we don't need to consider about the uncertain instance. Because our main purpose is how to classify the new gestures, instead of clustering all of the training data. We can build the MCSs with all the abstract concepts and classify the new instances(gestures) by the selected classifier based on dynamic classifier selection method.

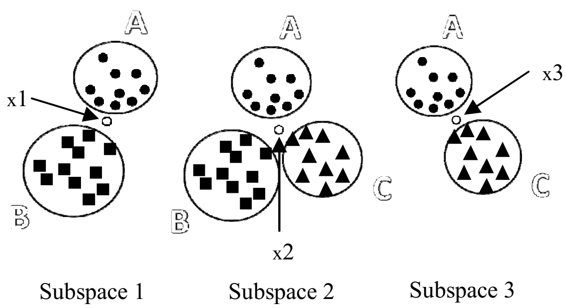


Figure 4. The 3 subspaces

III. Experiments and Evaluation

An experiment was designed to test the new method. We use a 5DT data glove (Figure 5.) to generate the gesture data. The 5 finger sensors in the glove can measure finger flexure for users. Figure 6 shows the data when we do the three gestures. Take the case of the first gesture, we bent three fingers(index finger, middle finger and ring finger), so that the glove received 3 flexure signals which were shown as waves. We can see that sometimes it is difficult to get perfect data for each fingers when we are using the data glove, because of the disturbance from the movements of other fingers. In the realtime environment, it could be considered as noisy data or uncertain instances in the previous algorithm.

Figure 5. The 5DT data glove

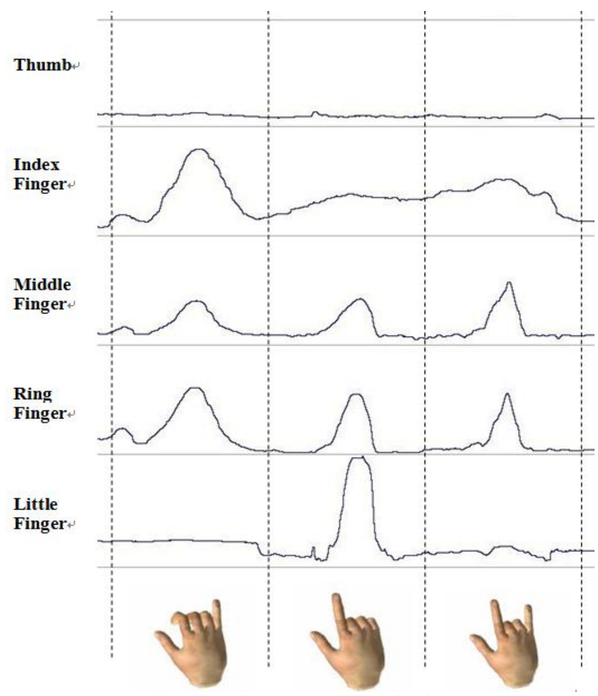


Figure 6. The 3 gestures and the data

There are 2^5 gestures with one hand, because each finger has two statuses. In this experiment, we separated the finger flexure time series and calculated the average value in every 1 second interval. So that the 5-dimension instances can be created for each gesture. We gathered the data 20 times for each gesture. And then a dataset was built with 640 instances.

In the experiment, we used the 10-fold cross validation method to compare the classification accuracy of our new method with 3 classical classification methods: SVM (Support Vector Machine), DT (Decision Tree), NB (Naive Bayes) and kNN (k Nearest Neighbour). And CG stands for the new concept grounding method in the following table. We employed these classificaion algorithms from the Weka Software package[20]. We used k-means to generate multiple partitions for the new method with random initialization of cluster centers and set k=32. The result shows that the performance of the new approach is better than other classical classification methods.

TABLE I. A COMPARISON OF FIVE METHODS

	Classification methods				
	SVM	DT	kNN	NB	CG
Classification Accuracy(%)	51.4	67.8	49.5	68.4	72.2

IV. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed a concept grounding approach for glove-based gesture recognition based on CEMF. We imported the idea of clustering fusion techniques to deal with the gesture recognition problem. The experiment result shows that it is an effective method in gesture recognition.

However, it is worth noticing that the new method will result in high computing cost because of the pre-modelling of clustering ensembles and dynamic classifier selection. After getting the abstract gesture concept, our future work focus on how to use association rule methods to generate the abstract behaviours from the gesture sequence.

References

[1] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE, vol. 77, no. 2, pp. 257–285, Feb. 1989.

[2] J. Weaver, T. Starner, and A. Pentland, "Real-time American Sign Language recognition using desk and wearable computer based video," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 12, pp. 1371–1378, Dec. 1998.

[3] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in timesequential images using hidden Markov model," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn., Champaign, IL, pp. 379–385, 1992.

[4] S. Arulapalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking," IEEE Trans. Signal Process., vol. 50, no. 2, pp. 174–188, Feb. 2001.

[5] C. Kwok, D. Fox, and M. Meila, "Real-time particle filters," Proc. IEEE, vol. 92, no. 3, pp. 469–484, Mar. 2004.

[6] A. Doucet, N. de Freitas, and N. Gordon, Eds., Sequential Monte Carlo in Practice. NewYork: Springer-Verlag, 2001.

[7] J. Davis and M. Shah, "Visual gesture recognition," Vis., Image Signal Process., vol. 141, pp. 101–106, 1994.

[8] M. Yeasin and S. Chaudhuri, "Visual understanding of dynamic hand gestures," Pattern Recogn., vol. 33, pp. 1805–1817, 2000.

[9] P. Hong, M. Turk, and T. S. Huang, "Gesture modeling and recognition using finite state machines," in Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recogn., Grenoble, France, pp. 410–415, Mar. 2000.

[10] M. S. Yang and N. Ahuja, "Recognizing hand gesture using motion trajectories," in Proc. IEEE CS Conf. Comput. Vis. Pattern Recogn., vol. 1, pp. 466–472, Jun.1998.

[11] T. Mitchell, *Machine Learning*, McGraw-Hill Science/Engineering/Math, 1997.

[12] P. Iravani, "Behaviour-based architecture for abstract learning and control," Towards Autonomous Robotic Systems (TAROS-04), 2004.

[13] M. J. Mataric, "Behavior-based control: Examples from navigation, learning, and group behavior, " Journal of Experimental and Theoretical Articial Intelligence, 9:323-0336, 1997.

[14] B. Mirkin, "Concept Learning and Feature Selection Based on Square-Error Clustering, " Machine Learning, vol. 35, pp. 25-39, 1999.

[15] M. Lebowitz, "Experiments with incremental concept formation: UNIMEM," Machine Learning, 2:103 - 138, 1997.

[16] D. H. Fisher, "Knowledge acquisition via incremental conceptual clustering," Technical Report 87-22 Department of Information and Computer Science, Univ of California Irvine, Irvine, CA, 1987.

[17] Y. Huang, D. Monekosso and H. Wang, "Clustering Ensembles Based on Multi-classifier Fusion," Intelligent Computing and Intelligent Systems (ICIS), pp.393 – 397, 2010.

[18] A. Fred, "Finding Consistent Clusters in Data Partitions," Multiple Classifier Systems, vol. 2096, pp. 309–318. Springer, Heidelberg , 2001.

[19] A. Topchy, B. Minaei-Bidgoli, A.K. Jain, W.F. Punch, "Adaptive Clustering Ensembles," Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), vol. 1, pp. 272–275, 2004.

[20] I.H.Witten and E.Frank, *Data Mining: Practical Machine Learning Tools with Java Implementations*, Morgan Kaufmann, San Francisco, 2000.