WAC 2002, Fifth Bianual World Automation Congress, June 9-13, 2002, Orlando, Florida

REAL-TIME HAND GESTURE TELEROBOTIC SYSTEM USING FUZZY C-MEANS CLUSTERING

JUAN WACHS, URI KARTOUN, HELMAN STERN,YAEL EDAN

Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Be'er-Sheeva, ISRAEL

ABSTRACT

This paper describes a teleoperation system in which an articulated robot performs a block pushing task based on hand gesture commands sent through the Internet. A Fuzzy C-Means clustering method is used to classify hand postures as "gesture commands". The fuzzy recognition system was tested using 20 trials each of a 12-gesture vocabulary. Results revealed an acceptance rate of 99.6% (percent of gestures with a sufficiently large membership value to belong to at least one of the designated classifications), and a recognition accuracy of 100% (the percent of accepted gestures classified correctly). Performance times to carry out the pushing task showed rapid learning, reaching standard times within 4 to 6 trials by an inexperienced operator.

KEYWORDS: gesture recognition, telerobotics, fuzzy c-means, hand gesture

INTRODUCTION

Hand gestures are one but a few of the methods used in telerobotic control [1]. This type of communication provides an expressive, natural and intuitive way for humans to control robotic systems. One benefit [2] of such a system is that it is a natural way to send geometrical information to the robot such as: left, right, etc. Gestures may represent a single command, a sequence of commands, a single word, or a phrase, and may be static or dynamic. Such systems should be accurate enough to provide the correct classification of hand gestures in a reasonable time [3].

Human-robot interaction using hand gestures provides a formidable challenge. This is because the environment contains a complex background, dynamic lighting conditions, a deformable hand shape, and a real-time execution requirement. There has recently been a growing interest in gesture recognition systems with a number of researchers providing some novel approaches, many of which are quite elaborate and require intensive computer resources. For example, Quek [4] develops a flow field computational algorithm. Koons, et. al. [5] describes an approach in which hand data is classified into features of posture, orientation and motion. In [6] color, motion, and tracking is used in a hand posture system for robot control. Classification is based on elastic graph matching. An excellent review of gesture modeling approaches is that of Huang, et.al. [7]. Of those mentioned our system is closest to that of the edge based technique to extract image parameters from simple silhouettes used by Segen [8].

In this paper we define a static hand gesture vocabulary for telerobotic control using a Fuzzy C-Means (FCM) recognition system. Although the speed of artificial neural network (ANN) classifiers allows real-time operation and comparable accuracy, a FCM is used because it requires smaller training sets and shorter training times. Moreover, it is compatible with future research needs to compare other systems based on cluster variances. ANNs preclude such comparisons, as cluster boundaries are transparent, except for very small networks. Our approach is implemented in a teleoperation client-server internet environment. The paper describes the gesture recognition algorithm and its evaluation.

SYSTEM ARCHITECTURE

To control a robot movement, the user evokes a gesture from a gesture vocabulary. The user lays his/her hand over a video imager, and a raw image is acquired. The gesture is classified using a recognition module based on the FCM algorithm [9] and is sent to the robot for execution. The components of the system consist of a five-degree of freedom robot, a PC system with a frame grabber, two USB cameras, and a video imager.

A set of recognized gestures is sent through the TCP/IP communication protocol to a distant robot PC server. The server is connected to the robot controller and two USB cameras, which continually capture the robot scene. Both side and front views of the scene are sent to client interface using the FTP protocol.

The hand gesture recognition system flow diagram is shown in Fig.1. Upon presentation of the robotic scene in the user's interface, a gesture G is selected from a gesture vocabulary {G1, G2,.., G12}. A vision system converts the captured image of the gesture into a feature string, which is subsequently recognized and sent to the robot PC server. After the robot executes the command, camera views of the robot environment are transmitted back to the interface.

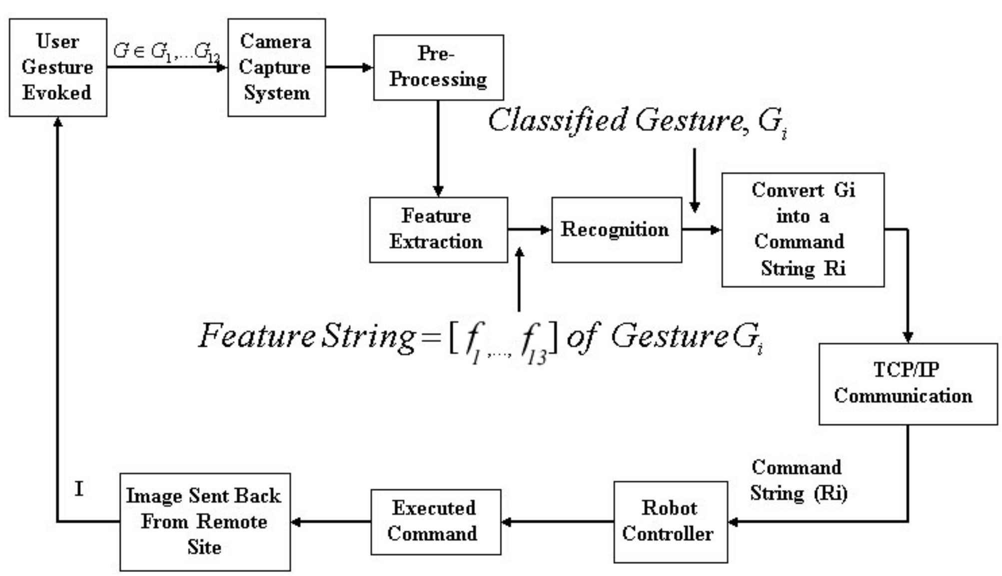


Fig. 1 System Flow Diagram

GESTURE CLASSIFICATION

Hand Gesture Language

A vocabulary of 12 static gesture poses was designed for robot control tasks (Fig. 2). The first six gestures of the vocabulary control the robot arm-using world coordinates. The forward and back hand gestures control the X-axis, the right and left hand gestures control the Y-axis, and the up and down hand gestures control the Z-axis of the robot arm. The Roll Right and Roll Left hand gestures rotate the wrist joint, and the Open Grip and Close Grip gestures control the robot gripper. The Stop hand gesture stops any action the robot performs. The Home gesture resets all robot joints in the home position. The gestures were selected to correspond to the robot teach pendant.

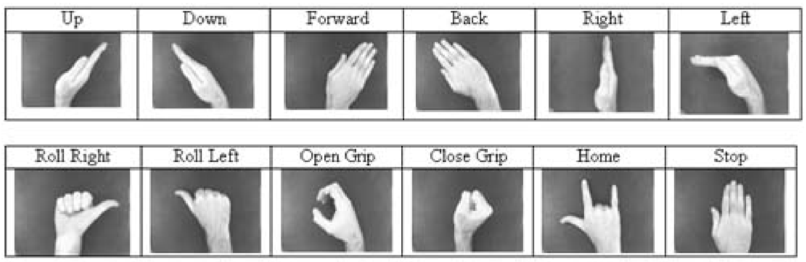


Fig. 2 Visual Gesture Recognition Languages

Fuzzy C-Means Clustering

Fuzzy C-Means Clustering algorithm (FCM) is an easily understood and fast computational algorithm described mathematically in [9]. Given a set of n data patterns, $X=x_1 ,...., x_k ,...., x_n$. ,the algorithm minimizes a weighted within group sum of squared error objective function, $J(U,V)$.

$$J(U,V)=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^{m}d^{2}(x_k,v_i) \tag{1}$$

Where x_k is the k-th p-dimensional data vector, v_i is the prototype of the center of cluster i, u_{ik} is the degree of membership of x_k in the i-th cluster, m is a weighting exponent on each fuzzy membership, $d(x_k,v_i)$ is a distance measure between data pattern x_k and cluster center v_i, n is the number of data patterns, and c is the number of clusters. The objective function $J(U,V)$ is minimized via an iterative process in which the degrees of membership u_{ik} and the cluster centers v_i are updated:

$$v_i=\frac{\sum_{k=1}^{n}u_{ik}^{m}X_k}{\sum_{k=1}^{n}u_{ik}^{m}}, \quad u_{ik}=1/(1+\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{ij}}\right)^{\frac{2}{m-1}}) \tag{2}$$

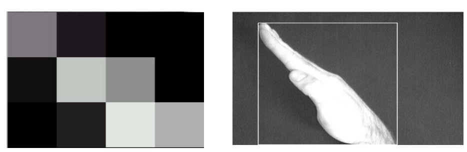
Where, the u_{ik} satisfies:

$$0 < \sum_{k=1}^{n} u_{ik} < n \forall i, \qquad u_{ik} \in [0,1], \sum_{i=1}^{c} u_{ik} = 1 \forall k \tag{3}$$

In the proposed methodology, the FCM algorithm is provided with a training set of gestures. Each gesture is represented by a feature vector. The set of feature vectors are clustered for subsequent use in a recognition system. Once the clusters have been created, they are labeled. (i.e, assigned a linguistic description of the gesture). This is conducted manually, hence the term unsupervised clustering. The cluster center, v_i is a prototype feature vector for cluster i, x_k is the feature vector of the k-th exemplar gesture in the training set, u_{ik} is the "degree of belonging" (membership value) of the k-th feature vector to cluster i, c is the number of gestures in the gesture set as well as the number of clusters, n is the number of images in the training set.

Preprocessing and Feature Extraction

Preprocessing of the image is simple and fast. It starts with segmentation of the hand from the background using an appropriate threshold value to obtain a black and white image. This is followed by constructing a bounding box around the segmented hand and performing a block partition. A feature vector of the image with 13 parameters is created. The first feature is the aspect ratio of the bounding box. The last 12 features are block mean gray values calculated from a 3 by 4 block partition of the image. The block mean gray values represent the average brightness of each block in the image. Fig. 3 illustrates a typical user gesture, its block mean gray values, and the resultant feature vector.



Feature Vector = [57 176 52 2 2 68 249 171 16 3 13 253 188]

Fig. 3 Illustration of a Feature Vector

Training Stage

The training stage involves running the Fuzzy C-Means algorithm for a set of exemplar hand gestures. Variations are incorporated by slightly varying the hand configuration for each gesture. At least 25 samples for each of the 12 hand gestures in the language comprise the training set. During the FCM clustering process, for each training image a membership vector is built and computed from (2). Equation (1), was used to empirically find the best value of parameter m.

Classification

Gestures performed by a user are classified using the highest membership value. In our case, if $x_{k'}$ is the feature vector of the current hand gesture image , its distance to each of the cluster centers v_i is determined and used in (2) to calculate the membership values $\{u_{ik'}\ \forall\ i=1,\dots,c\}$. The gesture is classified by finding: $u_{i'k'} =$ Max $\{u_{ik'}\ ;\ \forall\ i=1,\dots,c\}$. A further test is made before recognition is established. This test depends on a recognition threshold, τ. If $u_{i'k'} >= \tau$ is true, then the gesture is recognized as belonging to classification *i'*, otherwise it is rejected as all the membership values are too low.

TESTING AND TASK VALIDATION

Testing The Recognition System

An experiment was designed to test the recognition system. A single user presented twenty trials of each gesture to the gesture recognition system. Two measures of performance, based on a distinction between rejecting unclear gestures and the correct classification of clear gestures, were used in the evaluation. An additional test was undertaken to estimate speed recognition performance. A gesture k' whose entire set of membership values, $\{u_{ik'}\ ;\ \forall i=1,\dots,c\}$, is below the given recognition threshold value of τ is said to be rejected and therefore not classified. A gesture k' with at least one membership value above the threshold value τ, is said to be accepted and recognized as belonging to classification i', if $u_{i'k'} = Max\ \{u_{ik'};\ \forall i=1,\dots,c\}$. We can now define the two performance measures as:

(a) **Acceptance Rate** - The percent of gestures accepted. A gesture is rejected if it does not have a large enough membership value for any of the designated classifications.

(b) **Recognition Accuracy** - The percent of accepted gestures that are classified correctly.

Because the rejection rate and recognition accuracy are sensitive to the threshold, its value was determined empirically. The recognition threshold τ was varied over the interval [0, 1]. The acceptance rate was found to be concave decreasing (from 100%) in τ. The best value was determined as min (τ) such that, the recognition accuracy is 100% and occurred for $\tau = 0.3$.

Results indicate an acceptance rate of 99.6% (one unaccepted gesture out of 240), and a recognition accuracy of 100%. Thus, all accepted gestures were recognized correctly. Total recognition speed t was also estimated empirically using the following time parameters: t_c - camera capture and frame grabber time, t_p - preprocessing time, t_r - FCM recognition time. The mean times of 20 trials on a Pentium 400 MHz PC are: $t = t_c+t_p+t_r=.04ms+7.58ms+.81ms=8.43ms$. This time depends on the computer and frame grabber used, so comparison with other reported times is difficult. However, this time is sufficient to allow for real time operation. Kjeldsen [10] reports 2Hz for his pose real time recognition algorithm using a 100 MHz PC.

Case Study of a Robotic Task

A case study was conducted in which an operator using hand gestures controls a remote robot to push a wooden cube, located on a top of a pile into a container. An inexperienced operator performed ten identical experiments and the performance times for each were recorded. The learning curve of task completion times appears in Fig. 4. Standard times were reached after four to six trials.

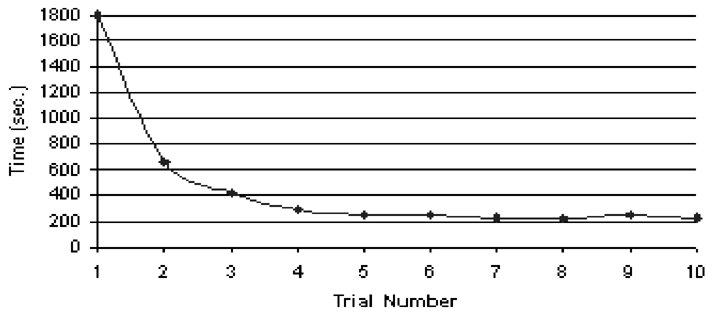


Fig. 4 Learning Curve of the Hand Gesture System

CONCLUSIONS

This project described a telerobotic gesture-based user interface system using visual recognition. Experimental results showed that the system satisfies the requirements for a robust and user-friendly input device. The Fuzzy C-Means algorithm provided enough speed and sufficient reliability to perform the desired tasks. The case study demonstrated the importance of latency for telerobotic systems. Although gestures were recognized quickly and sent in packet form, successful execution of the commands could not be verified until the image of the robot environment was received at the user interface. This resulted in an *overlapping effect* – the submission of a new gesture before complete information of robots current position was received. This is known as the problem of latency and future research will be directed toward its solution

ACKNOWLEDGEMENTS

This project was supported by the Ministry of Defense MAFAT Grant No. 2647, and partially supported by the Paul Ivanier Center for Robotics Research & Production Management, Ben-Gurion University of the Negev.

REFERENCES

1. Katkere A., Hunter E., Kuramura D., Schlenzig J., Moezzi S. and Jain R. "ROBOGEST: Telepresence Using Hand Gestures". Technical report VCL-94-104, University of California, San Diego. (1994).
2. Kortenkamp, D., Huber, E., and Bonasso, R. P. "Recognizing and Interpreting Gestures on a Mobile Robot". in AAAI96, (1996).
3. Pavlovic V, Sharma R., and Huang T. "Visual Interpretation of Hand Gestures for Human Computer Interaction: A Review". IEEE PAMI, Vol. 19. pp. 677-695. (1997).
4. F. K. H. Quek. Unencumbered Gestural Interaction. IEEE Multimedia. 36-47, Winter, (1996).
5. Koons D. B., et. al. "Integrating Simultaneous Input from Speech, Gaze, and Hand Gestures", 257-276 in Intelligent Multimedia Interfaces, M. T. Maybury (Ed), MIT Press, (1993).
6. Triesch J. and Malsburg C. V. D. "A Gesture Interface for Human-Robot Interaction". Proc. of 3th IEEE Intl. Conf. on Automatic Face and Gesture Recognition, pp. 546-551. (1998).
7. T. S. Huang, and V. I. Pavlovic. "Hand Gesture Modeling, Analysis,and Synthesis". Proc, of Intl. Conf. on Automatic Face and Gesture Recognition, (1995).
8. Segen J. "Controlling Computers with Gloveless Gestures".Proc.of Virtual Reality Systems, (1993)
9. Bezdek J. C. "Cluster Validity with Fuzzy Sets". Cybernetics. Vol. 3, No. 3, 58-73. (1973).
10. Kjeldsen F.C.M. "Visual Interpretation of Hand Gestures" PhD Dissertation, Columbia Univ. (1997)