

Fusion of Information From Data-Glove and a Camera for Hand Gesture Recognition

Ramakant, Faculty

Rajiv Gandhi University of Knowledge & Technologies

RK Valley, India, 516329

Department of Electronics and Communication Engineering,

Abstract -- Hand gesture recognition system provides a natural, innovative and modern way of non-verbal communication. It has a wide area of applications in human robot interaction and sign language recognition. The intention of this project is to develop a novel approach for hand gesture recognition based on fusion of information from 5DT data-glove 14 ultra and a camera. In this project, artificial neural network is used to recognize the sensor values coming from the sensor gloves. These values are then categorized in to 16 hand gestures for single-handed gesture recognition. Parallel ANN is used for two-handed gesture recognition. The network was trained under very small mean-square error (MSE=0.001). The proposed implemented algorithm has been tested for single-handed, two-handed and American sign language recognition. It provides higher recognition rate as compared to the gesture recognition system which use either gloves or a camera (vision-based) alone.

Keywords: Computer vision, Back Propagation algorithm, Artificial neural network (ANN), 5DT Data-Glove Ultra 14, Hand gesture recognition, American sign language (ASL).

I. INTRODUCTION

Gesture recognition has emerged as one of the most important research area in the field of computer vision and pattern recognition. A hand gesture, which is the representation of ideas using unique hand shape or finger orientation, has the potential to interface with the computer system. Thus, it seems convenient that human-robot interfaces incorporate hand gesture recognition capabilities. For instance, I would like to have the possibility of transmitting simple orders to personal robots using hand gestures. The recognition of hand gestures requires both hand and single-handed gestures recognition.

Why fusion of information: 5DT data-glove can measure the movement of finger and abduction angle between the fingers. For American Sign Language recognition, orientation of the hand, movement of fingers and abduction angle between the fingers are required. In this case, I used the decision level fusion technique for fusion of information from data-glove and a camera. Hand orientation information is required for six gestures shown in Figure 1.1.

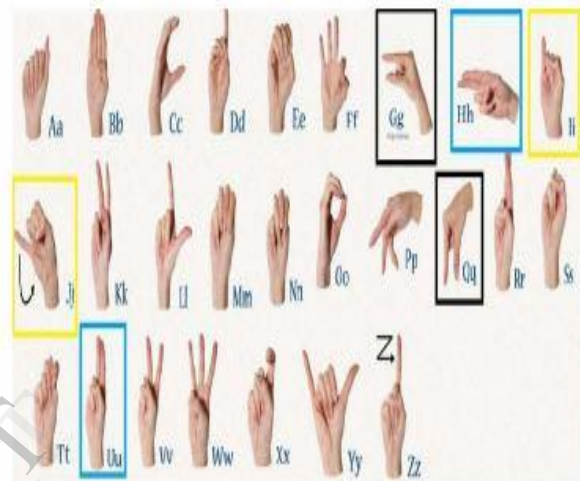


Figure 1.1: American Sign Language

I proposed a novel algorithm for hand gesture recognition based on fusion of information from data-glove and camera for all static hand gesture recognition with higher recognition rate as compare to vision based system. The Artificial Neural Network was trained under very small mean square error (MSE=0.001). It has been demonstrated that decisions from multiple classifiers can be combined to improve the overall classification performance by exploring the complementarities of source and pattern of two modalities. Here, the goal is to enhance classification through the fusion of the two decisions.

For our experiment, I chose a set of 16 static hand gestures. These 16 gestures comprised of binary open/close configuration of the fingers excluding the thumb.

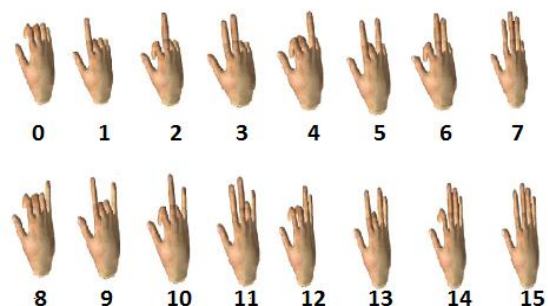


Figure 1.2: Sixteen static hand gestures

II. RELATED WORKS

Initially, most of the interaction techniques involving an electronic glove used static hand postures (static finger flexion), or static hand postures coupled with hand or head position and orientation. Takashi and Kishino [5] discuss a recognition of a hand-gesture based on self-organization using a data glove is presented. They have succeeded in recognizing 32 kinds of hand-shapes based on self-organization by measuring the angles of 10 joints of a hand using a Super Data-Glove. The number of samples in each gesture is 20, of which 10 samples are used for training and the rest are used for test. In this paper, they used only finger movement and abduction angle based gesture. They didn't consider the orientation of the hand. In [7], The Cyber-Glove glove has been used to measure the pose of a hand and provide data to a pose recognizer. Neural Network Toolbox is used to recognize the patterns that exist in the raw sensor data set of the data glove. Fifteen groups of test set of 100 hand gestures are chosen to test the neural network. Each group represents a sort of hand gestures. The neural network identifies gestures correctly 98 out of 100 hand gestures (98.0%). To test the generalization capability of the neural network, 50 hand gestures from 3 alien people are collected. The recognition rate is 92.0%. One problem encountered during the training of the neural network is premature saturation. In[9], Sign language recognition using 7-sensor glove of 5DT company is presented. Artificial neural networks are used to recognize the sensor values coming from the sensor glove. Recognition rate of 88% is obtained. One problem that was faced in the project that was some of alphabets involved static gestures and orientation of the hand. These gestures may not be recognized using this glove.

III. SINGLE-HANDED GESTURE RECOGNITION

3.1 Proposed real time data acquisition system and data analysis.

I use a 5DT data-glove 14 ultra, distributed by fifth dimension technology. It has 2 Sensors per finger and one sensor between the finger. Figure 3.1 shows the algorithm used for real time data acquisition. 5DT data-glove measures the angles of 14 joints of the hand: Thumb near joint, Thumb far joint, Thumb-Index abduction, Index near joint, Index far joint, Index-Middle abduction, Middle near joint, Middle far joint, Middle-Ring abduction, Ring near joint, Ring far joint, Ring-Little abduction, Little near joint, Little far joint. Movements of finger joints are measured at certain time interval by a 5DT data-glove. Data of one same hand gesture from three different people are shown in Figure 3.2.

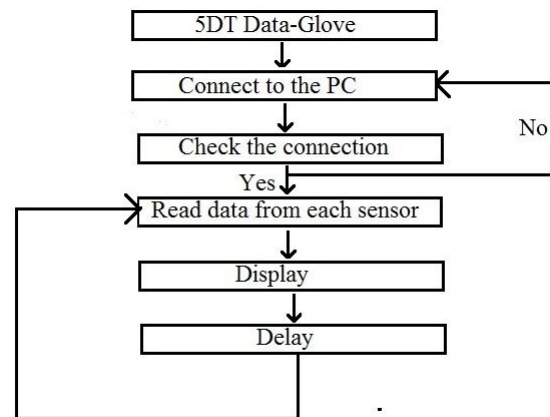


Figure 3.1: Proposed Real Time Data Acquisition System

According to the figure, I can find that the measured glove data have the same changing trends for a certain hand gesture, though different people have different hand shapes. A feed-forward neural network can represent an arbitrary functional mapping, so it is possible to use a back-propagation neural network to remember these trends or patterns.

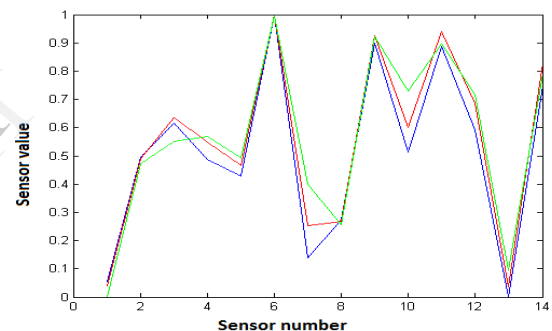


Figure 3.2: Single-handed gesture data at different time instants

3.2 Normalization of the Sensor Data

Normalization makes the data independent of the sensor range. With min/max normalization, the minimum and maximum value of a feature column are determined and taken as 0 and 1 respectively. Subsequently, all data in the column is normalized between 0 and 1. Equation shows the min/max normalization equation:

$$y_i = \left(\max_{target} - \min_{target} \right) + \left[\frac{x_i - \min_{value}}{\max_{value} - \min_{value}} \right]$$

Where y_i is the normalized value, \min_{target} and \max_{target} are 0 and 1 respectively. The values \min_{value} and \max_{value} are the respective min and max values of a column.

3.3 Neural network design and training

Neural network can easily model data which is too difficult to model with traditional methods such as inferential statistics or programming logic. I use neural network to recognize the patterns that exist in the sensor data set of the data glove. The selected design approach for

the neural network topology uses a feed-forward network with a single hidden layer. I use supervised learning algorithm for modifying the weights and biases of the neural network. The training method selected is back-propagation using an adaptive learning rate. The learning rate of standard BP neural network algorithm is a constant, which is usually between 0 and 1. As the learning rate of the adaptive learning adjustment method is fixed, when the training sequence close to the minimum of training, the change of the derivative between the error and the weight value is very small. As the learning rate is a constant, their product is also a constant that could not approach to the optimal weights value and the value of the error swings back and forth. As a result BP neural network algorithm cannot converge to the minimum value. Conventional adaptive learning rate adjustment Algorithm uses the method that whether the amendments of the weight value reduce the error function to determine the right direction. If the change of the weight value reduces the error, I could increase the learning rate. On the other hand, if the overshoot phenomenon occurs, it means the training has exceeded the minimum value; I should reduce the value of learning rate and minimize the value of training towards the approximation point. For adaptive learning the rule is:

$$\eta(t+1) = \begin{cases} \eta(t) & E(t+1) \leq E(t) \\ \eta_1 \eta(t) & E(t+1) > E(t) \end{cases}$$

Where η ($\eta = .045$) is the learning rate and η_1 ($\eta_1 = .80$) is constant. The mean square error of the BP neural network:

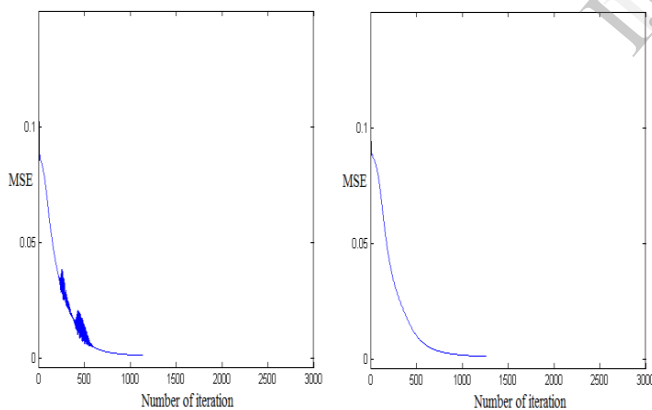


Figure 3.3: a) MSE at constant learning rate b) MSE at adaptive learning rate

3.4 Results

For single-handed gestures, I have considered 8 different classes (gestures) and 15 gestures/class are used for training and testing the proposed neural network recognizer. The neural network correctly identifies 117 gestures out of 120 hand gestures. The Recognition rate is almost 97.50 %. Again for single-handed gestures, I have considered 16 different classes (gestures) and 6 gestures are used for testing the proposed neural network recognizer. The neural network correctly identifies 94

gestures out of 96 hand gestures. The Recognition rate is almost 97.91 %.

| Gesture Patterns | Correct | Incorrect | Recognition using Data Glove | Recognition using Vision Based |
|------------------|---------|-----------|------------------------------|--------------------------------|
| TI | 15 | 0 | 100 | 93 |
| IM | 14 | 1 | 93 | 93 |
| IL | 15 | 0 | 100 | 94 |
| TIM | 15 | 0 | 100 | 92 |
| IMR | 15 | 0 | 100 | 95 |
| TIL | 15 | 0 | 100 | 93 |
| TIML | 14 | 1 | 93 | 92 |
| IMRL | 14 | 1 | 93 | 94 |
| Total = 120 | 117 | 3 | 97.50 | 93.25 |

Table 3.1: Comparison of gesture recognition using ANN for data-glove-based and vision-based recognition system [5]

I compared these 16 static hand gestures with the existing recognition system [3]. They collected the test set of 100 hand gestures to test the neural network. The neural network identifies correctly 98 out of 100 (98.0%). One problem encountered during the training of the neural network is premature saturation. To test the generalization capability of the neural network, 50 hand gestures from 3 alien people are collected.

| Gesture Pattern | Correct | Incorrect | Recognition with Calibration |
|-----------------|---------|-----------|------------------------------|
| 0 | 6 | 0 | 100 |
| 1 | 5 | 1 | 83 |
| 2 | 6 | 0 | 100 |
| 3 | 6 | 0 | 100 |
| 4 | 6 | 0 | 100 |
| 5 | 6 | 0 | 100 |
| 6 | 6 | 0 | 100 |
| 7 | 6 | 0 | 100 |
| 8 | 6 | 0 | 100 |
| 9 | 6 | 0 | 100 |
| 10 | 5 | 1 | 83 |
| 11 | 6 | 0 | 100 |
| 12 | 6 | 0 | 100 |
| 13 | 6 | 0 | 100 |
| 14 | 6 | 0 | 100 |
| 15 | 6 | 0 | 100 |
| Total = 96 | 94 | 2 | 97.91 |

Table 3.2: Comparison of sixteen static hand gesture recognition using ANN with existing method [3]

IV. TWO-HANDED GESTURE RECOGNITION

4.1 Two-Handed Real Time Data Acquisition System

Data from both the data gloves can be acquired on the time sharing basis. First, data from the left glove is acquired and after some time (sampling time) data from right hand glove is acquired. According to proposed system, I can connect both the data-glove to the USB port and the movements of finger joints are measured from each data-glove after a certain time interval. Figure 4.1 shows the algorithm used for real time data

acquisition.

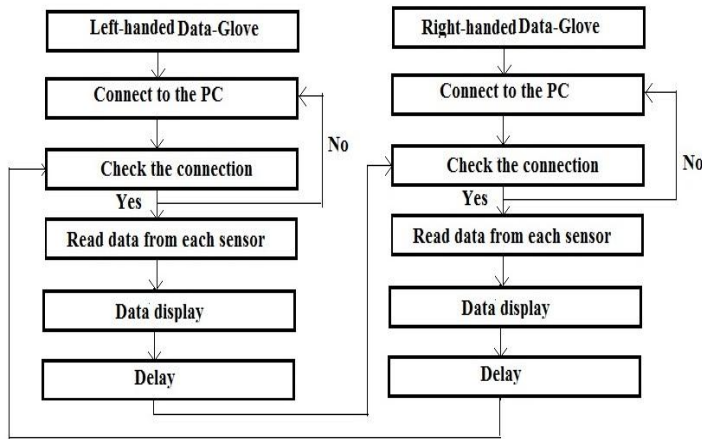


Figure 4.1: Data Acquisition system for left and right hand

For two-handed gesture recognition, I have selected a set of 16 static hand gesture (shown in figure 1.2) for both left hand and right hand. These 16 gestures comprised of binary open/close configuration of the fingers excluding the thumb. In these gestures, I have considered four fingers as a four bit binary number in which little finger is considered as MSB of the bit pattern.

4.2. Proposed Parallel Neural Network for Two-Handed Gesture Recognition

I used two neural networks for left hand and right hand. For two-handed real time gesture recognition; I collected the data from left-handed data-glove first and recognized this data using left-handed neural network. After a certain time interval, I collected data from right-handed data-glove and recognized using right-handed neural network. Combine the recognition result after the classification of gesture from both the neural network.

I considered both hand as a combination of 8-bit binary number. According to 8-bit binary number, I recognized 256 (from 0 to 255) gestures using both hand. By using single hand, I can recognize only 16 gestures.

4.3 Test Results for Two-Handed Gestures

I have considered 256 different classes (gestures) and 5 gestures/class are used for training and testing the proposed neural network recognizer.

| Left hand Gesture | Right hand Gesture | Two-handed gesture | Correct | Incorrect | Recognition |
|-------------------|--------------------|--------------------|---------|-----------|-------------|
| 0 | 0 | 0 | 5 | 0 | 100 |
| 0 | 1 | 1 | 4 | 1 | 80 |
| 0 | 15 | 15 | 5 | 0 | 100 |
| 1 | 0 | 16 | 5 | 0 | 100 |
| 1 | 15 | 31 | 5 | 0 | 100 |
| 2 | 0 | 32 | 5 | 0 | 100 |
| 2 | 15 | 47 | 5 | 0 | 100 |
| 3 | 0 | 48 | 5 | 0 | 100 |
| 3 | 15 | 63 | 5 | 0 | 100 |
| 7 | 0 | 112 | 5 | 0 | 100 |
| 7 | 15 | 127 | 5 | 0 | 100 |
| 8 | 0 | 128 | 4 | 1 | 80 |
| 8 | 8 | 136 | 4 | 1 | 80 |
| 14 | 0 | 224 | 5 | 0 | 110 |
| 14 | 15 | 239 | 5 | 0 | 100 |
| 15 | 0 | 240 | 5 | 0 | 100 |
| . | . | . | . | . | . |
| 15 | 14 | 254 | 5 | 0 | 100 |
| 15 | 15 | 255 | 5 | 0 | 100 |

Table 4.1: Two-handed gesture recognition performance

V. FUSION OF THE INFORMATION FROM DATA-GLOVES AND A CAMERA

As explained earlier, by using 5DT data-glove, I can measure only the finger movements and abduction angle between the fingers, but I cannot measure the orientation of the hand. Orientation of the hand, finger movements and abduction angle between the fingers are required to determine all the static hand gestures. For American Sign Language recognition, I can recognize all gestures except six letters (I and J, H and U, G and Q). Because, these letters are based on orientation of the hand, finger movements and abduction angle between the fingers. For recognition of these gestures, I proposed a new algorithm based on the fusion of data-glove and camera information. Six letters of American Sign Language are shown in Figure: 5.1. I can't recognize these six letters by using 5DT data-glove because, these letters are based on orientation of the hand. By using 5DT data-glove, I can measure the finger movements and abduction angle between the fingers. So, recognition of these letters, I can take the orientation information from the camera and local motions from the data-glove. At the final stage, decision level fusion technique is used for combination of data-glove and the camera information.



Figure 5.1: Gestures used for Fusion of Glove and Vision based approaches

5.1 Proposed System for Fusion of Information from Data-Gloves and a Camera

It has been demonstrated that, decisions from multiple classifiers can be combined to improve the overall classification performance by exploring the complementarities of source and pattern of two modalities. Here, the goal is to enhance classification through the fusion of the two decisions. Decision-level fusion optimally fuses two decisions according to the orientation of the hand (vision-based) and neural network classification for the data acquired by gloves. I can determine these six letters using vision-based technique but, the problem is that, I can't measure the finger movements and the abduction angle of the fingers accurately using a camera. 5DT data-glove provides accurate measurement of finger movement and abduction angle of the fingers. It is experimentally found that proposed system has higher recognition rate as compared to vision-based technique

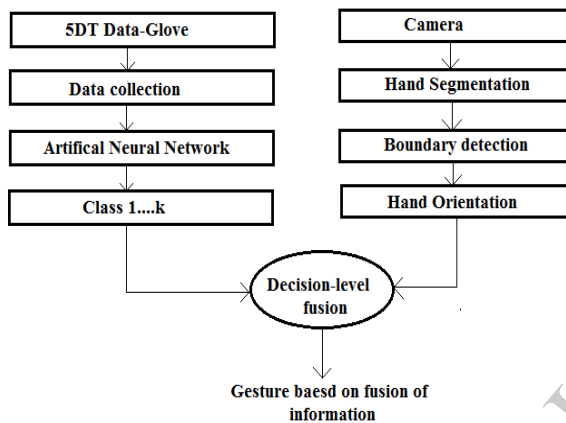


Figure 5.2: Proposed system for fusion of information from data-gloves and a camera

5.2 Segmentation of the Hand

Hand segmentation is the process which extracts the hand from the rest of the image. In this proposed method, RGB image is captured by the camera, but the RGB color space is more sensitive to different light condition so, I need to encode the RGB information in to HSV. HSV is a family of color space where, each pixel in the image will have three values: hue, saturation and value. Segmentation based on HSV requires a plain and uniform background. I concluded that hue, saturation and value thresholds can be used to extract the hand and its features. The hue threshold value must be greater than 0 and less than 140, The saturation threshold value must be greater than 0 and less than 220 whereas, the value threshold must be greater than 0 and less than 60. Thresholds have been implemented in our system by taking a hand gesture image after converting it from RGB to HSV. After that, I applied thresholds on it to produce a binary image (black and white image) where, all the pixels which satisfy the thresholds are colored white, whereas the rest of the image is colored with black.



Figure 5.3: Segmentation of Hand

If I take close look to the segmented image, I find that the segmentation is not perfectly done. Background may have some ones which is known as background noise and hand gesture may have some zeros that is known as gesture noise. These errors can lead to a problem in boundary detection of hand gesture; So, I need to remove these errors. A morphological filtering [4] approach has been applied using sequence of dilation and erosion to obtain a smooth, closed, and complete contour of a gesture.

5.3 Boundary and Orientation Detection

I implemented canny edge detection algorithm for finding the edges of the hand. After finding the edge of the hand in the image I proceed to the next step towards orientation detection. In this step, I identify either the hand is vertical or horizontal. I start the scanning the boundary matrix or edge of hand in binary image. Whenever I get x-boundary is equal to 1 for some time span at any place in the boundary matrix along with the increasing value of y-boundary, it will set as horizontal image. If I get y-boundary is equal to maximum of size of image with increasing value of x-boundary at any place in the boundary matrices. It will set as a vertical hand. At this stage, I have categorized hand in to two categories vertical and horizontal. By using fusion technique I can determine all the static hand gesture. Proposed technique is shown in Figure 5.2. By using the decision-level fusion, I can determine the difference between the orientation-based gestures.

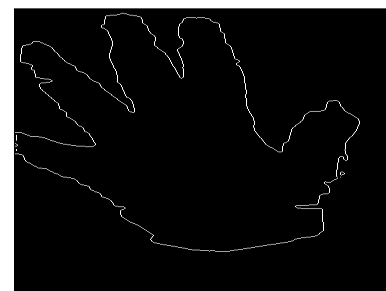


Figure 5.3: Boundary of the hand for orientation detection

In the American Sign Language, six gestures are based on the orientation of the hand as shown in figure 5.1. In this case, recognition performance of 98% is reported.

| Gesture | Correct | Incorrect | Recognition using data glove and a Camera (%) | Recognition using only Vision based (%) [10] |
|------------|---------|-----------|---|--|
| G | 15 | 0 | 100 | 90 |
| H | 14 | 1 | 93 | 100 |
| I | 15 | 0 | 100 | 70 |
| J | 15 | 0 | 100 | 100 |
| Q | 15 | 0 | 100 | 90 |
| U | 14 | 1 | 93 | 100 |
| Total = 90 | 88 | 2 | 98 | 92 |

5.4 Test Result for Static American Sign Language Recognition

I have collected a set of 15 gestures/class for training and testing the proposed system. Proposed algorithm using data-glove and a camera gives a higher recognition rate as compared to the systems which use only a camera (vision- based).

| Gesture | Correct | Incorrect | Recognition (%) |
|-------------|---------|-----------|-----------------|
| A | 15 | 0 | 100 |
| B | 15 | 0 | 100 |
| C | 15 | 0 | 100 |
| D | 15 | 0 | 100 |
| E | 14 | 1 | 93 |
| F | 15 | 0 | 100 |
| G | 15 | 0 | 100 |
| H | 14 | 1 | 93 |
| I | 15 | 0 | 100 |
| J | 15 | 0 | 100 |
| K | 15 | 0 | 100 |
| L | 15 | 0 | 93 |
| M | 13 | 2 | 86 |
| N | 12 | 3 | 80 |
| O | 14 | 1 | 93 |
| P | 15 | 0 | 100 |
| Q | 15 | 0 | 100 |
| R | 14 | 1 | 93 |
| S | 13 | 2 | 86 |
| T | 14 | 1 | 93 |
| U | 15 | 0 | 100 |
| V | 15 | 0 | 100 |
| W | 15 | 0 | 100 |
| X | 13 | 2 | 86 |
| Y | 15 | 0 | 100 |
| Z | 14 | 1 | 93 |
| Total = 390 | 375 | 15 | 96 |

VI. CONCLUSIONS

The proposed real time hand gesture recognition system provides an approach to recognize the single-handed and two-handed gesture recognition using 5DT data-glove and a camera. In case of single-handed, recognition performance of 97.5% and 97.9% is obtained and it is compared with vision-based and data-glove based systems. It shows that data-glove based approach gives higher recognition rate as compared to vision-based system. For two- handed gesture recognition, parallel artificial neural network is used. I considered two hands as a combination of 8 bit binary and recognized 256 gestures form 0 to 255.

A novel algorithm is proposed for hand gestures having different orientation using fusion of information from data-glove and a camera. It is seen that the recognition performance of the proposed system is better as compared to the systems which use either data glove or a camera (vision-based) for recognition.

REFERENCES

1. Laura Dipietrocc and Paolo Dario, "A Survey of Glove-Based System and Their Applicatio,," IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 38, pp. 461-482, 2008.
2. M. Ishikawa and H. Matsumura, "Recognition of Hand-Gesture Based on Self-Organization Using a Data-glove,," 6th International Conference on Neural Information Processing, vol. 2, pp. 739-745, 1999.
3. Deyou Xu, Wuyun Yao and Y. Zhang, "Hand Gesture Interaction for Virtual Training of SPG,," 16th Intelligent conference on Artificial Reality and Telexistence-Workshops, pp. 672-676, 2006.
4. S.A, Mehdi, Y.N Khan, "Sign language recognition using sensor gloves,," 9th Intelligent conference on Neural Information Processing, vol. 5, pp. 2204-2206, 2002.
5. M. K. Bhuyan, D. Neog, M. K. Kar, "Hand pose recognition using geometric features,," National Conference on Communications, pp. 1-5, 2011.
6. Fift Dimension Technologies, 5DT Data-Glove 14 Ultra. Retrieved from <http://www.5dt.com/product/pdataglove5u.html>, 2005.
7. R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms,," In ICML · S06: Proceedings of the 23rd international conference on Machine learning, pp. 161 } U168, New York, NY, USA, 2006.
8. Li Xiaoyuan, Qi Bin and Wang Lu, "A New Improved BP Neural Network Algorithm,," Second Intelligent conference on Intelligent Computation Technology and Automation, Vol. 1, pp. 19-22, 2009.
9. A. G. Blanco, M. D. Sonora, "Method to Approximate Initial Values for Training Lineal Neural Networks,," Electronics, Robotics and Automotive Mechanics Conference, pp. 443-446, 2008.
10. M. Panwar, P. S. Mehra, "Hand gesture recognition for human computer interaction,," International Conference on Image Information Processing (ICIIP), pp. 1-7, 2011.
11. Etsuko Ueda, Yoshio Matsumoto, Masakazu Imai and Tsukasa Ogasawara, "A Hand-Pose Estimation for Vision-Based Human Interfaces,," IEEE Trans. Indus-trial Electronics, vol. 50, pp. 676-684 , 2003.
12. J. M. Allen, P. K. Asselin and R. Foulds, "American Sign Language finger spelling recognition system,," IEEE 29th Annual, Proceedings of Bioengineering, pp. 285-286, 2003.
13. M. Minsky and S. A. papert, "Perceptron: An Introduction to Computing Ge-ometry,," MIT Press, Cambridge, MA, expanded edition, 1988.
14. L. B. Almeida, "Backpropagation in perception with feedback,," In R.Eckmillerand C. von der malsburg, editors, Neural computers, pp. 199-208, springer-verlag, berlin/heidelberg, 1988.