An Intelligent Multi-Gesture Spotting Robot to Assist Persons with Disabilities

Amarjot Singh, Devinder Kumar, Phani Srikanth, Srikrishna Karanam, and Niraj Acharya

Abstract—Sign language is used all over the world by the hearing-impaired and disabled to communicate with each other and the rest of the world. Sign language is more than just moving fingers or hands; it is a viable and visible language in which gestures and facial expressions play a very important role. These signs can be effectively used as far as communication with humans is concerned but with respect to communication with machines better methodologies and algorithms have to be developed. This paper presents a system which can be used by disabled people (who can communicate only with sign languages) to communicate with machines in order to employ it to their daily life. The system focuses on pose estimation of gestures used by physically disabled people to give suitable signals to the machines. Pose is generated using silhouettes followed by gesture recognition using Mahalanobis distance metric. The system further was to control a wireless robot from different gestures used by disabled people. The basic American sign language symbols I Hand, II Hand, Aboard, All Gone, etc were recognized and a specific signal with respect to each gesture is transmitted which controls the robot. The robot is made to perform forward, backward, left, right and stop actions for each gesture presented to the system by the disabled person. The present system can also be modeled to send wireless SMS or emails in worse situations. The system demonstrated potency in the estimation of motion and pose for interpreting the sign language by using the silhouettes of the pose. The proposed system recognizes sign language, thus providing disabled people a medium to communicate with machines, leading to simplicity in their day to day work.

Index Terms—Silhouette, motion segmentation, wireless robot, gesture recognition.

I. INTRODUCTION

Ign language is a combination of finger spelling, lip formations, and signs which are all heavily reliant on gestures and facial expressions to effectively convey meaning. Signs use visual imagery to convey ideas instead of single words. Moreover, signs are more reliant on gestures and facial expressions than finger spelling. It is therefore very important that the signer connects the correct facial expression with the particular sign. Hence, sign language plays a very important role in the life of disabled people to communicate with other human beings to accomplish their day to day requirements.

This field has seen an immense growth with the increasing development of computer vision in the form of various research papers and articles in journals [1], [2], [3].

With the advent of various affordable technologies there comes a need for enhancing the technologies being used to help the physically disabled people. Such systems in particular deal with certain parameters like sensing the human motion in certain area leading to estimation of the gesture. In addition to the technology described above, robots are vastly being used to assist and help disabled people using a number of other methodologies also like object recognition, machine learning etc. Robots are also used to improve the learning ability of the disabled person by recognizing and understanding the affective emotional state of a person in order to carry various activities such as game-play, learning and adapting to different situations. These activities are in sharp contrast with mainstream robotics applications where the emphasis is on removing the human involvement in activities in general.

Of the various existing algorithms for analyzing human motion, one method collects the optical flow over a certain area of interest in a particular sequence which can serve as an effective method for motion analysis. This method comprises of the representation of various patterns of motion using several layers of image silhouettes. The algorithm functions such that the arrival of any new frame minimizes the existing silhouettes to certain magnitude depending on the threshold value and overlaying the new silhouette that has been calibrated to maximum brightness. This methodology has been named as Motion History Image and has an advantage that a straddle of time ranging from certain frames to several seconds can be encrypted in a single image. Such generalization of the Motion History Image (MHI) in order to encode the time under observation in floating point data type is called as timed Motion History Image (tMHI)[4].

Recognition of pose in the current silhouette is achieved by the implementation of moment shape descriptors [5]. Normal optical flow is determined by a gradient of tMHI. The motion segmentation subjected to the object boundaries is performed in order to yield the orientation of motion and the magnitude of each region. The processing of the optical flow has been summarized in detail in Fig. 1. The conclusion to this procedure is the recognition of pose which is used in gesture recognition and object motion analysis.

The paper has been divided into six sections. The second section describes the methodologies involved in the silhouette generation followed by the in depth description of the gesture recognition methodology described in section III. Section IV explains the wireless robot, driven by the transmitted signal given by the person. Section V shows and explains the results for the experimentation while section VI elaborates the conclusion.

II. Silhouette Generation

There are multiple algorithms used for the generation of silhouettes like frame differing, colour histogram[6], back projection[7], stereo depth subtraction[8], etc. This paper uses the simplest method of background subtraction for the generation of silhouettes. A number of methods have been applied for back ground subtraction [9] but we will focus on a rather simpler methodology. This paper uses a naïve method for background subtraction. The pixels which are a set number of standard deviations from the mean RBG background constitute the foreground label. Noise is removed by the pixel dilation and region growing method. The silhouette is then extracted from the background. A major constraint in the usage of silhouettes is that no motion of the object under survey can be seen in the body region, for example if hands are moved in front of the human body they can't be extracted or differentiated from the obtained silhouette. The problem can be overcome by either the usage of multiple camera views simultaneously or by segmenting the flesh colored regions separately and overlaying them while crossing the foreground silhouette.
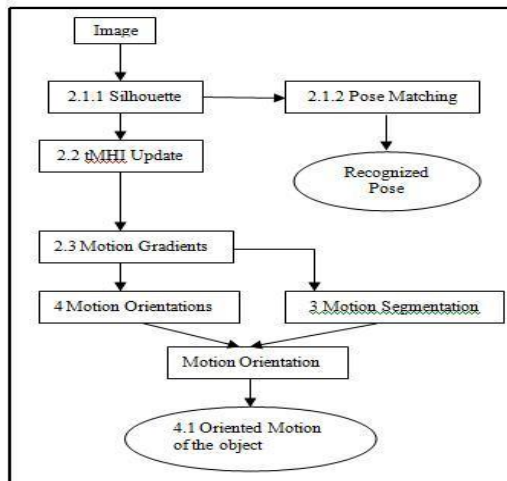

Fig. 1. Process flow chart.

III. Mahalanobis Match to Hu Moments for Silhouette Gesture Recognition

Gesture Recognition can be performed with the help of seven higher-order Hu-Moments [10] which provide shape differentiators that don't change with respect to translation and scale. The first 6 moments convert a particular shape while being obstinate towards any changes in translation of axes, degree of rotation and scale.

Mahalanobis distance metric becomes a necessity as the moments are of different orders [10]. The matching is based on a statistical measure of closeness to training examples as computing the distance between any two Hu-Moments is not feasible using Euclidean distance. The following equation can be used for computing the similarity

$$ma(x)(x \times m)^T \; P^1(x \times m)$$

where x denotes the moment feature vector, m denotes matrix: the mean of the training vectors while P^1 denotes the inverse covariance matrix for the training vectors.

IV. Wireless Robot

Once the gesture has been recognized, the wireless robot is controlled using a PC controlled transmitter and receiver. The transmitter is connected to the personal computer while the receiver is placed on the robot as shown in Fig. 3 (c). Once the gesture has been recognized, a signal transmitted to the robot is received by the receiver and the robot acts accordingly. Each gesture is assigned a particular signal leading to a specific action performed by the robot.
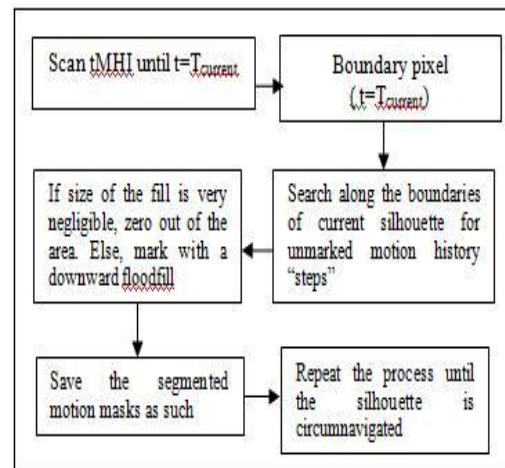

Fig. 2. Process flow chart for motion segmentation.

The signal is transmitted using a transmitter (Fig. 4 (a)) of HT12E encoder which encodes the signal being transmitted to the receiver (Fig. 4 (b)). The transmitter is connected to the PC using MCT2E integrated circuit chips. The information to be sent is encoded on a 12 bit signal sent over a serial channel. The 12 bit signal is a combination of 8 address bits and 4 data bits. The information sent by the transmitter is received by the HT 12D connected to the receiver. The detailed circuit diagram of the transmitter as well as receiver is shown in Fig. 4 (c) and Fig. 4 (d). The information is analyzed as 8 address bits and 4 data bits. If the address bits of the receiver and the address bits sent by the receiver are same then the data bits are further processed which leads to the movement of the receiver.

V. Results

The simulations enable us to investigate the capabilities of our system to demonstrate its utility for the hearing impaired and disabled people in their day to day life. The section elaborates and explains its capabilities through the results obtained for the gestures analysed. The simulations were carried out on a Pentium core 2 duo 1.83 GHz machine. The experimentation was to test the system's capability for the physically disabled people in their daily life. The results obtained after testing the system for 5 different stationary gestures were studied and analysed. In order to increase the efficiency of the system the gestures shown are coated with zinc powder provide better grayscale variation from the back ground as shown in Fig. 3. The system recognizes the sign using Mahalanobis distance. These gestures transmit signals with the help of a transmitter I.C., which can be further used to control the wireless robot. Later this can be used by the hearing-impaired to control equipments used in

their day-to-day life.

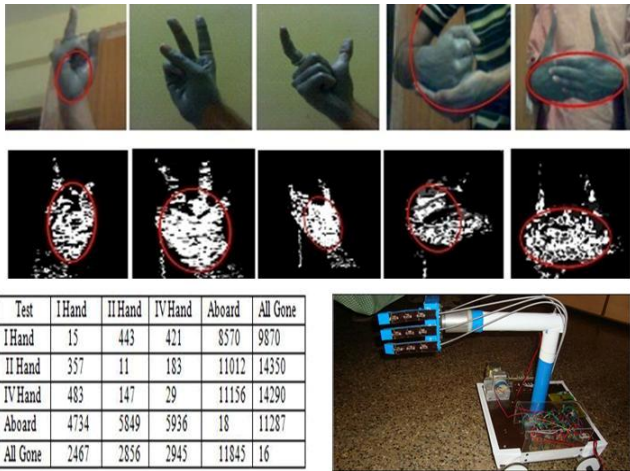| Test	I Hand	II Hand	IV Hand	Aboard	All Gone
I Hand	15	443	421	8570	9870
II Hand	357	11	183	11012	14350
IV Hand	483	147	29	11156	14290
Aboard	4734	5849	5936	18	11287
All Gone	2467	2856	2945	11845	16

Fig. 3. (a) Contour generation of gestures from video sequence (b) Discrimination results of pose recognition on distance grounds (c) Wireless robot
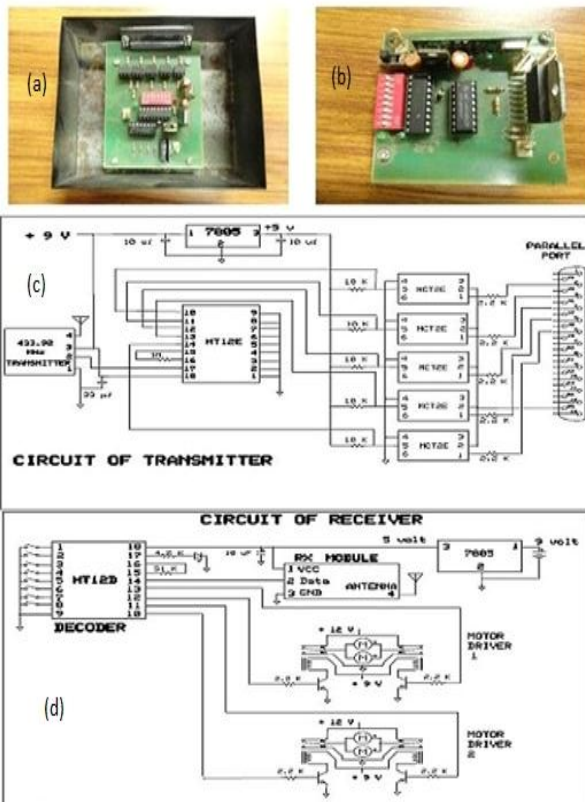

Fig. 4. (a) Transmitter circuit of wireless robot (b) Receiver circuit of wireless robot (c) Datasheet of transmitter circuit of wireless robot (d) Datasheet of receiver circuit of wireless robot

We describe the working and functions of our system by assuming a real world situation of disabled person seated on a chair. We make use of a digital camera to provide the input to the system in order to recognize the gesture. The system is tested mainly on 5 stationary gestures mainly I hand, II hand, IV hand, abandon and all gone. The system is trained in the initial step by presenting the gestures resulting into a particular signal for each gesture. We place the camera keeping in view of his disability. For example, if we have a person with paralysed hands and can give gestures only with his eyes, we place the camera such that it receives the gestures from his eyes. This input or the sequence of images taken are sent to a processing device for which we make use of a microcontroller. These input images are processed to generate silhouettes from which the timed Motion History Image (tMHI)[4] is generated as shown in Fig. 3(a). These images are used to compare with a pre-existing database which we need to create. This database contains all the gestures that the person can make and also the corresponding commands. These commands explain the taskhe want to undertake or his need. These commands can be shown on a display connected to the same chair so that any person can understand his requirements. The results for the recognition matrix generated are shown in Fig. 3(b). The gesture recognized by the system is further used to move the robot in a particular direction as specified by the gesture. I hand was used to move the robot forward, II hand was used to move backward while the right turn was take by IV hand and left turn was worked out by abandon. All gone gesture was used to stop the robot. The signals are transmitted by the I.C.'s mentioned above.

VI. CONCLUSION

The section gives a concise summary with applications related to the proposed system described above. This system makes use of the basic image segmentation and gesture recognition algorithms to operate the wireless robot using the gestures used by disabled people. The system is based on a novel system utilizing normal optical flow motion segmentation based on the tMHI that segments motion into regions which are meaningfully connected to movements of the object of interest. The system recognizes the gesture presented to the system on the basis of Mahalanobis distance which further controls the movement of the robot. This system can be further modified in future to operate equipment used in day-to-day life with the help of gestures. The system can be a boon for the physically and mentally challenged people. The system can be used by them to send signal to distant locations in case of an emergency or chaotic condition or any it can be used to control equipment by disabled people to make their daily life simple. The disabled people can use the system to send message to the people who are outside the home or can give signal to a particular machine to perform a particular task. To address this problem, we will send them SMS or E-Mails using the embedded web server which is connected to the World Wide Web. By making use of IP cameras, the family members can view the disabled person remotely. On the whole this system has great applications for the disabled people.

VII. FUTURE WORK

The aim in future would be to implement this system on a larger scale. An enhancement in the system can be to send wireless signals to the nearby shopping store depending upon the gesture showed by the person. This will make the disabled people more independent.

REFERENCES

[1] T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desktop and wearable computer based video," *Technical Report, Perceptual Computing, MIT Media Laboratory*, pp. 466, July 1998.

[2] T. Starner and A. Pentland, "Real-time American sign language recognition from video using hidden markov models," *Technical Report, MIT Media Lab, Perceptual Computing Group*, pp. 375, 1995.

[3] C. Vogler and S. Goldenstein, "Analysis of facial expressions in american sign language," in *Proc of the 3rd Intl. Conf. on Universal Access in Human-Computer Interaction (UAHCI)*, 2005.

[4] J. Davis, "Recognizing movement using motion histograms," *MIT Media Lab Technical Report*, pp. 487, March 1999.

[5] M. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Information Theory*, vol. IT-8, no. 2, 1962.

[6] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 22, pp.747-757, 2002.

[7] H. Tan, E. Viscito, E. Delp, and J. Allebach, "Inspection of machine parts by backprojection reconstruction," *IEEE int Conf. on Robotics and Automation. Proc*. pp. 503-508, 1987

[8] Y. Liu, G. Chen, N. Max, C. Hofsetz, and P. Mcguinness, "Visual hull rendering with multi-view stereo refinement," in *Proc of WSCG' 2004*. pp. 261-268, 2004.

[9] A. Elgammal, D. Harwood, and L. Davis. (1999). Non-parametric model for background subtraction. *IEEE FRAME-RATE Workshop*. [Online]. Available: http://www.eecs.lehigh.edu/FRAME/.

[10] C. Therrien, "Decision estimation and classification," *John Wiley and Sons, Inc.*, 1989.