I.J. Image, Graphics and Signal Processing, **2014, 2, 39-45**
Published Online January 2014 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2014.02.05

Autonomous Multiple Gesture Recognition System for Disabled People

Amarjot Singh[1], John Buonassisi[2], Sukriti Jain[3]
[1, 2] School of Engineering Science, Simon Fraser University, Burnaby, Canada.
[3]Dept of Electronics and Communication Engineering,
[3]Ambedkar Institute of Advanced Communication technologies and research, GGSIPU, India
[1] asa168@sfu.ca, [2] jab30@sfu.ca, [3]su_kriti24@yahoo.in

Abstract — The paper presents an intelligent multi gesture spotting system that can be used by disabled people to easily communicate with machines resulting into easement in day-to-day works. The system makes use of pose estimation for 10 signs used by hearing impaired people to communicate. Pose is extracted on the basis of silhouettes using timed motion history (tMHI) followed by gesture recognition with Hu-Moments. Signs involving motion are recognized with the help of optical flow. Based on the recognized gestures, particular instructions are sent to the robot connected to system resulting into an appropriate action/movement by the robot. The system is unique as it can act as a assisting device and can communicate in local as well as wide area to assist the disabled person.

Index Terms — Gesture Recognition, Motion Tracking, Robot, and Disability

I. INTRODUCTION

Sign language is a combination of finger spelling, lip formations, signs reliant to gestures and facial expressions used for communication. Signs use visual imagery to convey ideas instead of single words. The system plays a very important role in the life of disabled people as it is extensively used to communicate with other human beings and to perform day-to-day work.

A number of systems have been proposed in the past to assist disabled people in their daily needs [22]. In [19], the author used intentional head gestures to assist physically disabled people using computer access, in [20], MOVAID: a personal robot was designed to assist disabled and elderly people while in [21] the author designed a system for the assistance of the handicapped or old people. In addition to the technology described above, robots are vastly being used to assist and help disabled people [11]. In such settings, the intelligence of the disabled person is supplemented with the manipulation power of a robot in order to carry out activities such as elderly care [12], cooking [13], science experiments [14], and mixed manual/robot assembly and inspection [15]. Such systems in particular deal with certain parameters like sensing the human motion [23], [24] in certain area that further can help in estimating the gesture [10]. With the advent of various easily affordable technologies, there comes a need for enhancing the technologies to better assist the needy.

Out of the various algorithms that exist for analysing human motion [16], [17], [18], the system proposed in this paper makes use of optical flow over a certain region of interest in a particular sequence to recognize gestures. In this particular method, many different layers of silhouettes are used for representing various motion patterns. The algorithm functions in such a way 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 used to encode the time under observation in floating point data type is called as timed Motion History Image (tMHI) [4]. Further, the recognition of stationary pose in the current silhouette is achieved by moment shape descriptors [1].

In order to recognize gestures involving motion, normal optical flow is used. Normal optical flow is determined by gradient of tMHI. 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. Each step indicated in the figure corresponds to the section pertaining with relevant data to the respective step. As a result, recognition of various poses is obtained which are used in gesture recognition and object motion analysis. After the gesture recognition, appropriate predefined instructions are sent to the IP address of web server attached to the robot. Based on the instructions received, the robot takes necessary action/movement.

The system proposed in the paper is unique in multiple terms as (i) It can recognize both stationary and moving gestures (ii) The system can be used to operate local as well as remotely with the help of the web server. (iii) The system can also be used as a security device capable of sending signal to the security agency at the time of emergency using the webserver.

The paper has been divided into eight sections. The second section describes the methodologies involved in silhouette generation followed by the in depth description of silhouette recognition described in section three. Section four explains image segmentation involved to extract the object from the image. Fifth

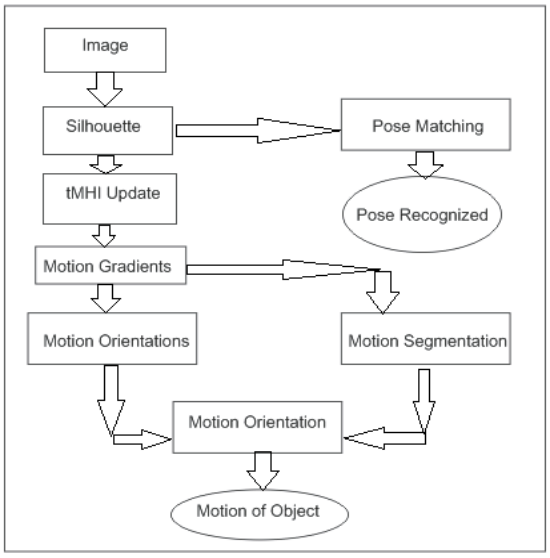

Figure 1. Process Flow Chart

section is used to describe now to match and identify different hand gestures, while the sixth section is used to describe the established system. Section seven shows and explains the results of experimentation. Finally, section eight vindicates the conclusion.

II. MOTION AND POSE ESTIMATION

The motion and pose estimation process is divided into three major steps (i) silhouette generation (ii) Computing motion history gradients. There are many algorithms that can be used for the generation of silhouettes like color histogram [7], back projection [8], stereo depth subtraction [9] etc. The paper uses a simple method for background subtraction for implementing silhouettes.

A. Silhouette Generation using Timed Motion History Images (tMHI)

A number of methods like mean shift estimation, histogram analysis, mixed Gaussian approach have been applied for back ground subtraction [2] 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 RGB background, constitute the foreground label. First, noise is removed by the pixel dilation and region growing method. The silhouette is further extracted from the background using tMHI. The algorithm functions in such a way 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. As mentioned earlier, we make use of floating point Motion History Image [4] wherein,

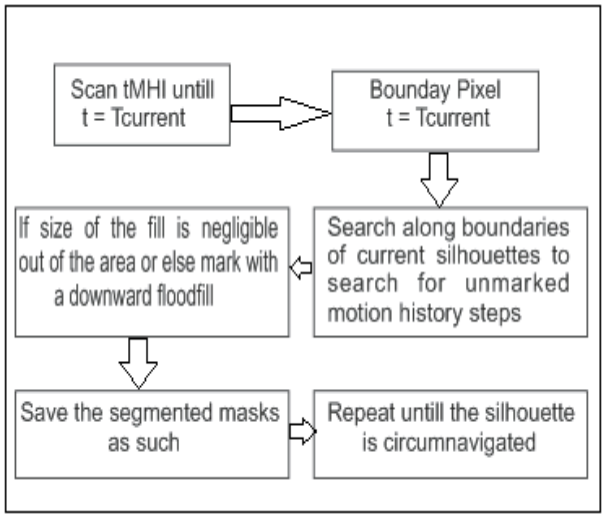

Figure 2. Process Flow chart for Motion Segmentation

new silhouette values are concatenated in with a floating point timestamp in the format: seconds.milliseconds. MHI is then updated as:

$$tMHI_{\alpha}(\bar{x},\bar{y}) = \begin{cases} \beta & if \quad current \quad silhouette \quad at(\bar{x},\bar{y}) \\ 0 & else \quad if \quad tMHI_{\alpha}(\bar{x},\bar{y}) < (\beta - \alpha) \end{cases} \quad (1)$$

Where β denotes the current timestamp, α the maximum time duration constant pertaining to the template. The advantage with this algorithm is that it makes our representation independent of the speed of the system or the frame rate such that irrespective of the capture rates, the same MHI area will be covered for a given gesture. Such representation is known as *timed Motion History Image* (tMHI).

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 overlay them while crossing the foreground silhouette.

B. Motion History Gradients

It is interesting to note that in an image, taking the gradients of tMHI, would yield the direction of the vectors pointing in the direction of the movement of the object. Also, each of these gradient vectors would be orthogonal to the boundaries of the motion of the object at each step, yielding a normal optical flow representation. Efficient evaluation of the gradients of tMHI is then carried out by convolution with separate Sobel filters in the coordinate direction thereby yielding the spatial derivatives $A_x(\bar{x},\bar{y})$ and $A_y(\bar{x},\bar{y})$. It is also noted that the calculation of the gradient information is to be restricted only to the locations within the tMHI.

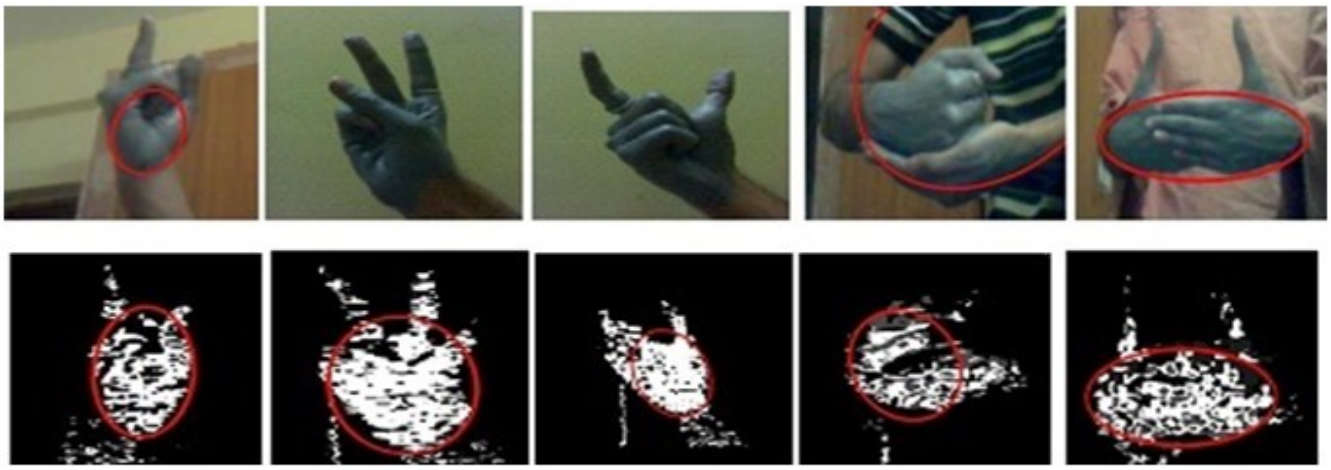

Fig. 3 Contour Generation of gestures from Video Sequence

Fig. 4 Gestures combined with motion

The application of the above method should not be carried out for the surrounding boundary of the tMHI as this would include non-silhouette pixels, which can lead to corruption of the results. It is also noted that the use of the gradients of the MHI pixels with a contrast either too low or too high within their neighborhood should be avoided. As a further step, we can extract the motion features of the object to varying scale. For example, a radial histogram of the motion orientation can be generated which can be used for direct recognitions [4]. We emphasize on a relatively simple method of global motion orientation as described in the next section.

III. GLOBAL GRADIENT ORIENTATION

In order to impart more importance to the most current motion within the given template, calculation of the global orientation should be weighted by normalized tMHI values [4]. This can be carried out by:

$$\bar{\theta} = \theta_{ref} + \frac{\sum_{\bar{x},\bar{y}} angDiff\,(\theta(\bar{x},\bar{y}),\theta_{ref}) \times (\beta,\alpha,tMHI_a(\bar{x},\bar{y}))}{\sum_{\bar{x},\bar{y}} norm(\beta,\alpha,tMHI_a(\bar{x},\bar{y}))} \quad (2)$$

where $\bar{\theta}$ denotes the global motion orientation, θ_{ref} the base reference angle, $\theta(\bar{x},\bar{y})$ the motion orientation map derived from gradient convolutions, $norm(\beta,\alpha,tMHI_a(\bar{x},\bar{y}))$ the normalized tMHI value and $angDiff\,(\theta(\bar{x},\bar{y}),\theta_{ref})$ the minimum signed angular difference of an orientation from the reference angle. The problems associated with the averaging circular distance measurements brings in the necessity for the usage of histogram based reference angle.

IV. MOTION SEGMENTATION

It is important for a segmentation scheme to know what is being segmented in a particular framework. One of the common methods used to extract objects is by collection of blobs of similar direction and motion [5]. However it doesn't guarantee that the motion corresponds to the actual movement of objects in a scene. The interest of the author is to group all the motion produced by the motion of the parts or the whole of the

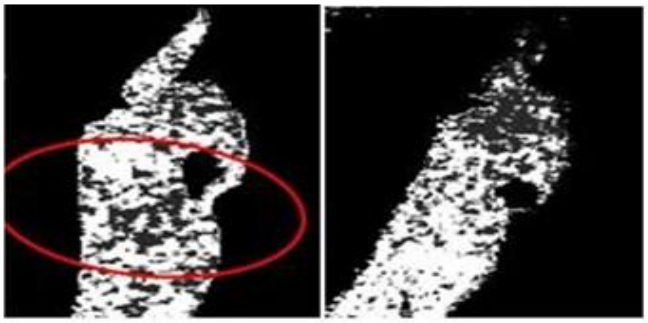

Fig.5 Silhouette for stationary gestures

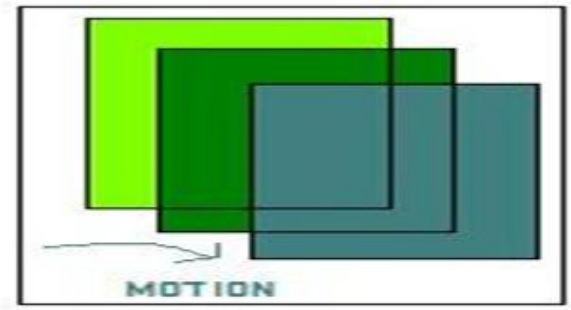

Fig.7 tMHI sequence for block motion

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. 6 Results of pose recognition on distance grounds

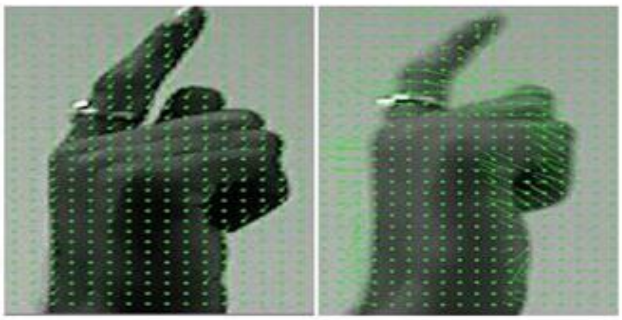

Fig.8 Global Motion Direction for Alive Gestures represented using Optical Flow

object of interest. This drawback can be overcome by marking the motion regions associated with the current silhouette with the help of downward stepping flood fill. The area of motion directly attached to the parts of the object of interest can be easily identified.

A. Motion associated with the objects

As per the construction, the silhouette with the most recent timestamp has the maximal values in the tMHI. The image is scanned until a similar value is found and then we walk along the silhouette to explore the areas of motion. Let ***dX*** be the threshold time difference. The algorithm given below describes the procedure for creating masks in order to segment the associated areas of motion. The block diagram for the algorithm is shown in Fig. 2. The step-by-step methodology is explained below:

1. Scan the tMHI until we find a pixel of the current timestamp. This is a boundary pixel of the most recent silhouette (Fig. 6.).
2. "Walk" around the boundary of the current silhouette region looking outside for recent (within ***dX***) unmarked motion history "steps". When a suitable step is found, mark it with a downward flood fill. If the size of the fill isn't big enough, zero out the area.
3. Store the segmented motion masks that were found.
4. Continue the boundary "walk" until the silhouette has been circumnavigated.

"Downfill" in this algorithm stands for the floodfills that fill the pixels with the same value or lower the pixels with a value lower than the current pixel being filled. The segmentation algorithm is a function of: (1) The maximum permissible limit of the downward step distance dX; (2) The minimum permissible limit of the downward flood fill.

The masks produced by the above algorithm are made use of in the selection of the valid motion history gradients. The segmented regions then are labelled with the respective weighted regional orientation as mentioned in Section 3. This ensures the connectivity of the motion regions with the objects as these masks derive directly from the past motion that spills from the current silhouette boundary of the object. The examples for the above principle are given in the below section.

TABLE 2. OPTICAL FLOW VARIATION OF EACH GESTURE FOR DIFFERENT TIME STAMPS

Gesture	|Θ| at T=1	|Θ| at T=5	|Θ| at T=20	|Θ| at T=40	|Θ| at T=60
Dig	0.28	1.44	8.92	12.64	16.73
Eat	0.9	1.78	5.67	11.67	13.44
Give	0.47	15.46	37.82	63.48	102.58
Walk	0.53	4.34	18.44	26.87	38.43
Alone	0.61	2.89	16.22	59.65	68.33

Figure 8. Microcontroller based Web server connected to the robot which is further connected to the laptop

B. *Motion Segmentation examples*

Motion segmentation using the method elaborated above, is demonstrated by giving motion gestures as input to the system as shown in Fig 8. Note that the small arrows correctly catch the finger motion. The movement of the hand is captured by optical flow tracking the fingers (Fig. 7). At left, hand has just been brought down as indicated by the large global motion arrow. Fig. 10 shows segmented motion and recognized pose for inputted to the system. Our motion segmentation method was implemented using OpenCV, an optimized open source computer vision library maintained by Intel Corporation [6].

V. MAHALANOBIS MATCH TO HU MOMENTS OF SILHOUETTE POSE

Recognition can be performed with the help of seven higher-order Hu moments [3], which provide shape differentiators that don't change with respect to translation and scale. The first 6 moments which convert a shape while being adamant towards any changes in translation of axes, degree of rotation and scale.

As the moments are of different orders, usage of Mahalanobis distance metric is a necessity [3]. 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.

$$ma(\bar{x}) = (\bar{x} - \bar{m})^T P^{-1} (\bar{x} - \bar{m}) \quad (3)$$

where x denotes the moment feature vector, m denotes the mean of the training vectors while K^{-1} denotes the inverse covariance matrix for the training vectors. Another approach to motion analysis and pose recognition makes use of the histograms of the segmented silhouette region.

VI. SYSTEM DESCRIPTION

The system comprises of a computing device (i.e. laptop), wireless robot, and AVR microcontroller based web server. The computing device can be any machine capable of running the presented algorithm of the paper for gesture recognition; in this case we have used a laptop. A wireless robot that is used to make the movements based on the instructions received. The third part is the AVR microcontroller based web server connected to wireless robot. AT mega 32 microcontroller is used in the web server having an Ethernet jack for Internet connectivity. The web server also has a LCD screen for displaying current IP address and various pins which are used for connecting the robot to web server.

First the computing device is used for gesture recognition. After the recognition of the gesture, the wireless robot is controlled using a microcontroller-based webserver (Fig 8) placed on the robot. Each gesture is assigned a particular signal that is transmitted to the robot using the Internet to the IP address of the web server connected to the robot. The web server passes the signal to robot and it acts accordingly. On identification of each gesture, the signal is transmitted using an LAN cable to the web server consisting of Ethernet port with a valid IP address that decodes the signal. The microcontroller based web server is

connected to the pc using LAN cable. The information to be sent is passed on to the web server. The data received by the web server is decoded and appropriate instructions are passed on the robot via the 4 ports to which the robot is connected. Finally the robot makes the appropriate action/movement based on the signal received from the web server's ports.

VII. EXPERIMENTATION RESULTS

The simulations enable us to investigate the capabilities of our system to demonstrate the utility for the hearing impaired in the day-to-day life. The section elaborates and explains in detail the results obtained from the simulations carried out on a Pentium Core 2 duo 1.83 Ghz machine. The experimentation was to test the system's capability for the hearing impaired in the daily life.

The system is tested mainly on 10 gestures I hand, II hand, IV hand, abandon, aboard, dig, eat, give, walk and alone taken from the standard library of American Sign Language (ASL). The testing set consists of 5 stationary gestures along with 5 gestures involving motion.

The five stationary gestures inputted to the system are segmented from the background as contours as shown in Fig. 5. The stationary gestures are easily extracted using variation in standard deviation of pixels of the moving body that is absent in the stationary background pixels. The gesture recognition is further carried out using mahabolis distance. The resultant matrix generated from recognition is shown in Table. 1.

The movement of the five gestures with motion is computed using optical flow computation as show in Fig. 6. Gestures are recognized using optical flow variation of each gesture for different time stamps as shown in Table. 2.

The gesture recognized by the system is further used to send appropriate instructions to web server attached to robot via the LAN port as show in Fig. 8. The robot moves in a particular direction as specified by the gesture signal/instruction forwarded by web server. In the present stage the five stationary gestures are used to control the robot. 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.

VIII. CONCLUSION

The system makes use of the gestures used by the hearing impaired to control the wireless robot. The system is based on a novel method of normal optical flow motion that segments motion into regions that are meaningfully connected to movements of the object of interest. The system recognizes the gesture presented to the system on the basis of mahabolis distance and optical flow computation for stationary and moving gestures respectively, which further used to controls the movement of the robot via a microcontroller-based webserver. This system can be further modified in future to operate the day-to-day equipment with the help of gestures. The system can be a boon for the physically and mentally challenged people.

REFERENCES

[1] Hu, M. Visual pattern recognition by moment invariants. IRE Trans. Information Theory, Vol. IT-8, Num. 2, 1962.

[2] Elgammal, A., Harwood, D. and Davis L. Non-parametric Model for Background Subtraction, IEEE FRAME-RATE Workshop, http://www.eecs.lehigh.edu/FRAME/. 1999.

[3] Therrien, C. Decision Estimation and Classification. John Wiley and Sons, Inc., 1989.

[4] Davis, J. Recognizing movement using motion histograms. MIT Media lab Technical Report #487, March 1999.

[5] Cuttler, R. and M. Turk. View-based interpretation of real-time optical flow for gesture recognition. Int. Conf. On Automatic Face and Gesture Recognition, page 416-421.

[6] Open Source Computer Vision Library (OpenCV) in C and optimized assembly modules on Intel's architecture can be downloaded from http://www.intel.com/research/mrl/research/cvlib.

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

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

[9] Yang Liu , George Chen , Nelson Max , Christian Hofsetz , Peter Mcguinness, "Visual Hull Rendering with Multi-view Stereo Refinement"WSCG, pp. 261-268, 2004.

[10] James M. Rehg and Takeo Kanade;" Visual tracking of high DOF articulated structures: An application to human hand tracking" ECCV 1994.

[11] Paolo Dario, Eugenio Guglielmelli, Cecilia Laschi, Giancarlo Teti "MOVAID: a personal robot in everyday life of disabled and elderly people" Journal of Technology and Disability, pp77-93, 1999.

[12] Balaguer, C.;Gimenez, A.;Jardon, A.;Cabas, R.;Correal, R. "Live experimentation of the service robot applications for elderly people care in home environments" Int conf. of Intelligent Robots and Systems, (IROS 2005). pp. 2345-2350, 2005.

[13] J.M. Noyes, R. Haigh, A.F.Starr "Automatic speech recognition for disabled people" journal of Applied Ergonomics, Vol 20, Issue 4, Pages 293-298 December 1989.

[14] Hayati, S.;Volpe, R.;Backes, P.;Balaram, J.;Welch, R.;Ivlev, R.;Tharp, G.;Peters, S.;Ohm, T.;Petras, R.;Laubach, S.;"The Rocky 7 rover: a Mars

sciencecraft prototype" Int conf. of Robotics and Automation, Vol 3 pp 2458 - 2464 1997.
[15] Abderrahim, M.;Balaguer, C.;Gimenez, A.;Pastor, J.M.;Padron, V.M. " ROMA: a climbing robot for inspection operations", proc. Int conf of Robotics and Automation, Vol 3 pp. 2303 – 2308, 1999.
[16] Ekinci M, Gedikli E. Silhouette based human motion detection and analysis for real-time automated video surveillance. Turk J Elec Engin, Vol. 13, pp. 199-229, 2005.
[17] C. Vogler, D. Metaxas, ASL recognition based on a coupling between HMMs and 3D motion analysis. Proc. of International Conference on Computer Vision, pp. 363-369, 1998.
[18] C. Bregler, Learning and recognizing human dynamics in video sequences. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition, pp. 568-574, 1997.
[19] Harwin WS, Jackson RD. Analysis of intentional head gestures to assist computer access by physically disabled people. J Biomed Eng, Vol. 12, pp. 193-8, 1990.
[20] P. Dario, E. Guglielmelli, B. Allotta, "MOVAID: a personnal robot in everyday life of disabled and elderly people", Techology and Disability Journal, n 10, IOS Press, pp. 77-93, 1999.
[21] P. Hoppenot, J. Boudy, J.L. Baldinger, F. Delavaux, E. Colle : "Assistance to the maintenance in residence of the handicapped or old people" - HuMaN 07, Tlemcen, 12th-14th march 2007.
[22] Amarjot Singh, Devinder Kumar, Phani Srikanth, Srikrishna Karanam, Niraj Acharya, "An Intelligent Multi-Gesture Spotting Robot to Assist Persons with Disabilities", in International Journal of Computer Theory and Engineering, Vol. 4, No. 6, pp. 998-1001, December, 2012.
[23] Devinder Kumar, Amarjot Singh, "Occluded Human Tracking and Identification using Image Annotation", in International Journal of Image, Graphics and Signal Processing, Vol.4, No.12, November 2012.
[24] Sridhar Bandaru, Amarjot Singh, "Advanced Mobile Surveillance System for Multiple People Tracking", in International Journal of Intelligent Systems and Applications, Vol. 5, pp. 76-84, May, 2013.

Amarjot Singh is a Graduate Student and Research Assistant in Laboratory for Robotic Vision at Simon Fraser University, Burnaby, Canada. Before joining SFU, he was a research engineer in Acoustic Research Laboratory at National University of Singapore. He graduated with a Bachelor of Technology in Electrical and Electronics Engineering from National Institute of Technology, Warangal (NITW), India. He has worked for multiple organizations in the past including INRIA, Sophia Antipolis, France, University of Bonn, Germany, IIT Kanpur, India, IISc Bangalore, India and DRDO, India. His research is focused on Image Segmentation, Motion Tracking and Computational Photography.

John Buonassisi is a undergraduate student at Simon Fraser University, Canada in his 5th year of Systems Engineering. Recently, John has completed an internship at the Robotic Vision Laboratory at Simon Fraser University and is currently finishing his undergraduate degree. John's research interests include machine learning, biomedical image processing, data compression, and object recognition.

Sukriti Jain is an Undergraduate researcher who has completed her Bachelors in Electronics and communication engineering from Ambedkar Institute of Advanced Communication technologies and research (AIACTR), GGSIPU, Delhi, India. She has also completed internships with National Thermal power co-operation, Delhi and Bharat Electronics Limited, Ghaziabad. Her research interests include signal processing, financial markets, bond portfolio management and Asset management.

