Indian Sign Language Recognition Using SVM[1]

J. L. Raheja[a], A. Mishra[b], and A. Chaudhary[c]

[a] *Machine Vision Lab CSIR-CEERI, Pilani, India*
[b] *School of Instrumentation D.A.V.V. Indore, India*
[c] *Researcher, Pilani, India*
e-mail: jagdish@ceeri.ernet.in, anand.mishra2k88@gmail.com, dr.ankit@ieee.org

Abstract—Needs and new technologies always inspire people to make new ways to interact with machines. This interaction can be for a specific purpose or a framework which can be applied to many applications. Sign language recognition is a very important area where an easiness in interaction with human or machine will help a lot of people. At this time, India has 2.8M people who can't speak or can't hear properly. This paper targets Indian sign recognition area based on dynamic hand gesture recognition techniques in real-time scenario. The captured video was converted to HSV color space for pre-processing and then segmentation was done based on skin pixels. Also Depth information was used in parallel to get more accurate results. Hu-Moments and motion trajectory were extracted from the image frames and the classification of gestures was done by Support Vector Machine. The system was tested with webcam as well as with MS Kinect. This type of system would be helpful in teaching and communication of hearing impaired persons.

Keywords: hand gesture recognition, kinect, gesture classification, depth segmentation, support vector machine, sign language, dynamic gestures.

DOI: 10.1134/S1054661816020164

1. INTRODUCTION

Humans and Machines both have their capabilities and restrictions. There are some ways, in which they can interact with each other but Human has to use specific method to interact with machine. Natural way of interaction is still not very much in use. Many Artificial Intelligence algorithms are developed to do the same. Hand actions and natural language recognitions are some of these methods. Sign language recognition (SLR) comes under dynamic hand gesture recognition as few gestures can be combination of more than one static gesture. SLR is a very useful technology to help disabled people and to communicate with them. In the world, there are many sign languages which are generally used in specific country. Each sign language (SL) differs from the other based on spoken language and country's culture.

This paper discusses recognition of Indian Sign Language (ISL) which is mainly used in India, a country which population is nearly 18% of whole world population. There has been a lot of work in different SLR for other SL and they are at very advanced stage but ISL recognition systems are still in developing phase. A vision based technique has been used in this paper for hand sign recognition and support vector machine (SVM) was used to classify as a sign of ISL. Sign languages contain large number of signs which makes difficult to make a robust system to recognize all the sign. Also the similar appearing signs go wrongly recognized sometimes. The facility to recognize signs whether they are made of by single hand or both hands also has to be included in systems. This system would process signs of ISL from deaf persons and could be used to communicate with them. Section 2 describes the related work regarding while Section 3 proposes the algorithm of pre-processing. Section 4 evaluates the performance of proposed system and Section 5 gives the idea for future work.

[1] The article is published in the original.

Received September 1, 2015

2. RELATED WORK

Various approaches have been used to develop sign language systems. Sensor based systems and vision based systems are main classifications of these. In both approaches, the processing of data was done in similar style after collection it by different sources. Voice and gesture recognition based system is proposed by Bolt [1]. Hand Tracking based gesture recognition system is discussed in [2]. Freeman [3] tries to design a system for operating television by hand gestures. Posture and Gesture Interpretation systems [4, 5] are proposed using vision methods which were extended further. Different algorithms were used for extracting signs from images. For sign language, hand contour and Hu moments [6, 7] are extracted.

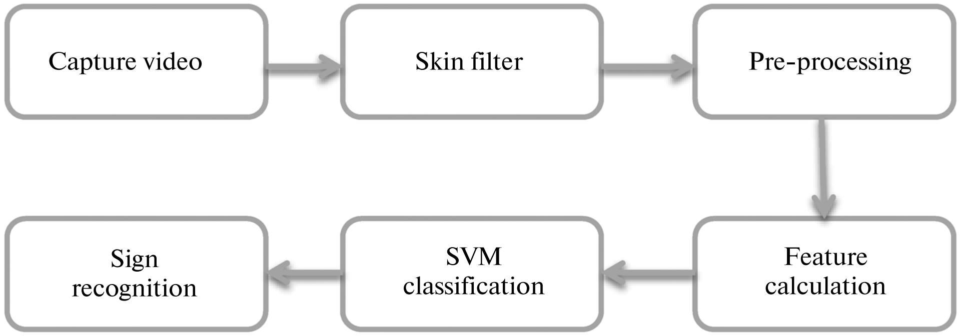

Fig. 1. Algorithm block diagram.

Hand features play an important role in hand recognition. Fingertips and center of palm were detected by many researchers [8–10]. Raheja [11] also did hand feature detection using MS Kinect. For tracking different methods have been applied like Mean Shift or Kalman Tracking [12, 13]. Different soft computing based methods have been used for gesture recognition like Neural Networks [6, 14], Hidden Markov Model [12] and Support Vector Machine [7, 15, 16]. A survey on soft computing based hand gesture recognition is available in [17].

3. GESTURE RECOGNITION

This section will discuss several steps, which were taken toward the hand gesture recognition for ISL recognition. The block diagram for the whole system is shown in Fig. 1. A real-time video of 30 frames per second was taken, which was analyzed frame by frame for dynamic gestures classification. The system uses skin filter to extract hand region from input video frame. The image frame was converted into HSV color space as it is less sensitive to light effects [18]. The ROI would be extracted from the captured video frame and smoothening and noise removal operation would be done.

The proper and efficient features need to be calculated because their selection would reflect efficiency of system. Before starting all these we have to create an empty feature vector to store these feature vectors where we will do feature vectors concatenation. To concatenate all feature vectors in a single row would result a new vector of desired length. The gesture length is need to be decided so that the other time warping algorithm's requirement would get reduced. This resulting single feature vector would be compared with vectors in database of ISL's signs. This comparison would be done using Support Vector Machine (SVM) classifier. The result of SVM would give a class of signs from database to which input sign gesture belongs. In the initial design, we assume that user will show only one hand at a time to the system. MATLAB with MS Kinect was used to capture video. For interfacing C++ with *mex* files were used. These *mex* initializes camera so that the depth and RGB images were extracted.

3.1. Binary Image Formation from Input Video Frame

We needed to convert, captured video frame to binary image so that it would help in future steps and fastening the system. This binary silhouette plays a vital role in proper calculation of required features. The input frame from video sequence was converted into HSV color space. The Hue of skin region is different from the background region; hence the hue factor was separated from HSV image. The Hue of skin region was determined from this image and an experimental threshold was applied on the result of that calculated hue value. This process is shown in Fig. 2 in different steps.

By experiments, 0.1 Hue value was calculated as threshold. The binary image after threshold operation contained noises because the input image has various objects whose hue is in range of human skin. But the noise pixels are smaller than hand region, so we can select largest connected region in image. Noise removal and smoothening was also done on this image. The depth image collected by Kinect was formed based on depth information [11]. The depth image was segmented based on closest object, which would be hand. The process is shown in Fig. 3. The depth of the hand would be used further in feature calculation.

3.2. Feature Extraction

To recognize hand in video frame, different features were tested for their optimality. The features to be calculated were decided according to type of gesture to recognize. For proposed sign recognition system, Hu-moments [6, 7], Number of hand fingers [9, 10], and tracked trajectory [12, 13] of hand centroid were tested. By experiments we found that only first two Hu-moments are to be used because others Hu-moments did not affect much for proposed sys-

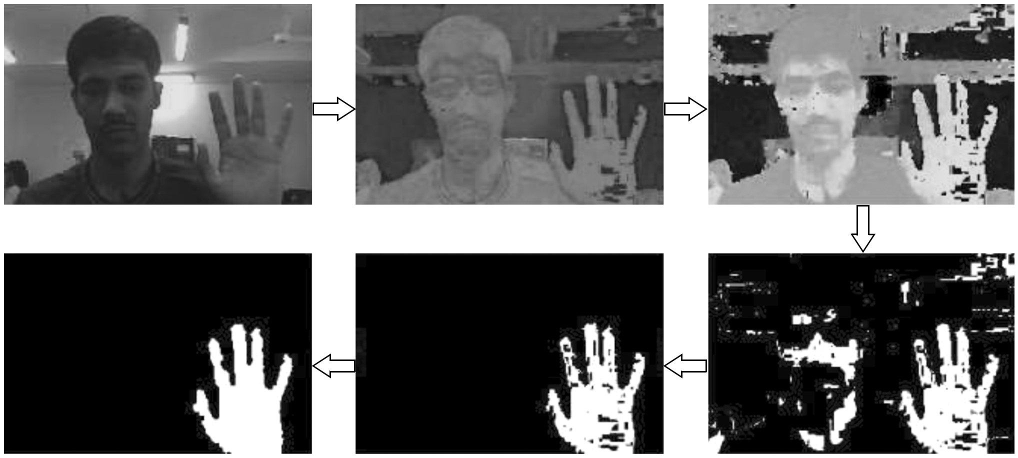

Fig. 2. Image pre-processing.

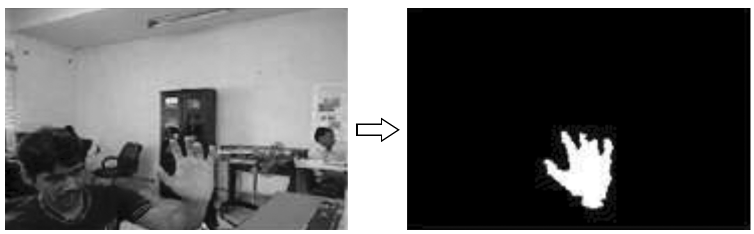

Fig. 3. Depth calculation for image frames.

tem. For the simplicity, we are taking only four signs into consideration for system testing. More signs could be added later. Four Indian sign language gestures tested were "A," "B," "C," and "Hello." These were selected because they are much different from each other.

Before starting whole process of image capturing, an empty feature vector was created to concatenate feature vectors which were calculated after every frame. This could be done until the length of feature vector is equal to length of hand gestures in pre-created database. The proposed system is tested for 4 gestures and its gesture length is fixed at 30 frames. For these 30 frames, the length of database and calculated feature vector is decided according to number of feature decided to extract. The whole process in described pictorially in Fig. 4.

Hu-moments. These are location, angle and shape invariant moments. If η_{pq} is $(p + q)^{th}$ order central moments which were collected after normalization then first two Hu-moments [19] are defined by Eqs. (1) and (2), respectively.

$$\Phi_1 = \eta_{20} + \eta_{02}, \quad (1)$$

$$\Phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2. \quad (2)$$

Only two moments are enough for proposed system because experiments showed that all others Hu-Moments were not giving much effect to testing of sign gesture used here.

Fingertips. Fingertips can be found using various algorithms contour analysis [9], K-Means algorithm [10], color based [19] are some of them. We used a spatial domain method for fingertip detection which gives number of fingers in input hand binary image. This was assumed that the used may not be having all five fingers due to some reasons which are not concerned by authors. The steps of fingertip detection algorithms are as follows.

Step 1. From input binary hand image, find centroid of image [8]. Take this as center of circular filter.

Step 2. Calculate radius from experiments for various area of hand. The general length of radius got from experiments is given by non-linear regression equation. This equation is given as follows.

$$\text{Radius} = 184.2644 * (\exp \hat{}\ 0.001548) * \text{depth}. \quad (3)$$

The depth in equation is the depth of centroid of hand which was calculated using MS Kinect depth image as discussed in section 3.2.

Step 3. Taking center and radius from previous steps, make a circular filter. Crop input hand image from this circular filter.

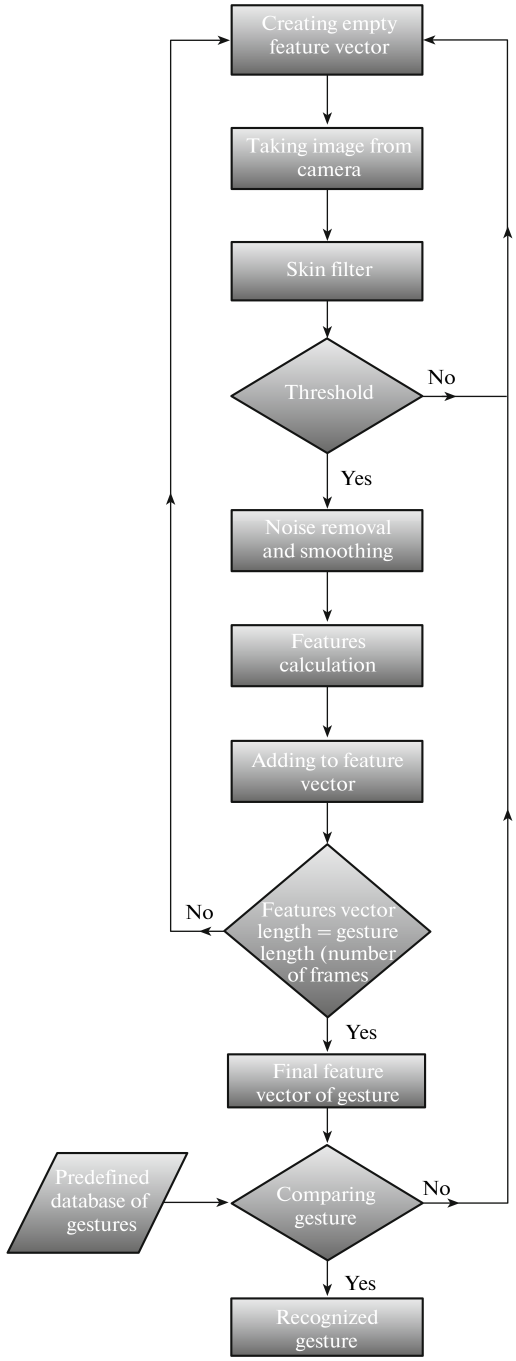

Fig. 4. Process diagram.

Step 4. Remove the wrist of hand if it remains. For this we have to know in which direction out hand facing in frame [8]. According to that remove the wrist of hand.

Step 5. The resultant binary image gives only the hand fingers of hand. Now find out the location of fingertips easily by finding extreme point of each finger.

The process of fingertips detection is shown in Fig. 5.

Trajectory tracking. The next feature to be extracted was the path of hand motion. This trajectory of hand should be tracked to avoid occlusion of hand with any other object or avoiding loss of hand from input image [11]. Tracking can be done easily using mean shift or Kalman tracking filter [12, 13]. This paper uses Kalman Filter because it was faster and efficient than mean shift. For proposed algorithm, the centroid of hand was already tracked and the differ-

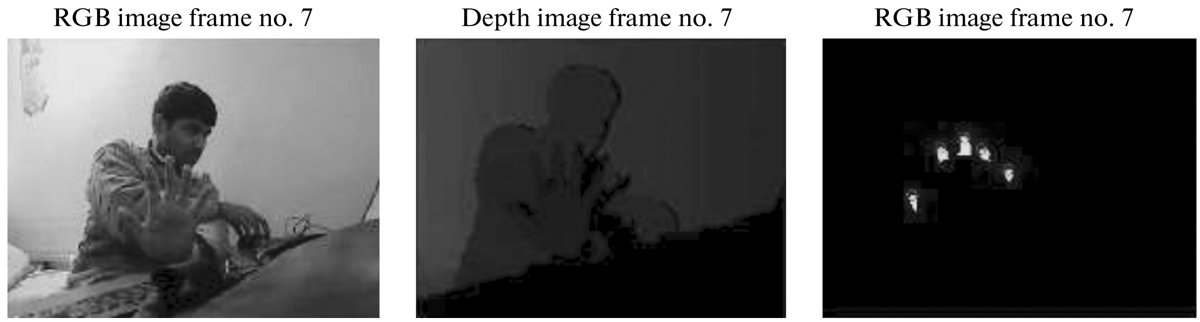

Fig. 5. Fingertip detection.

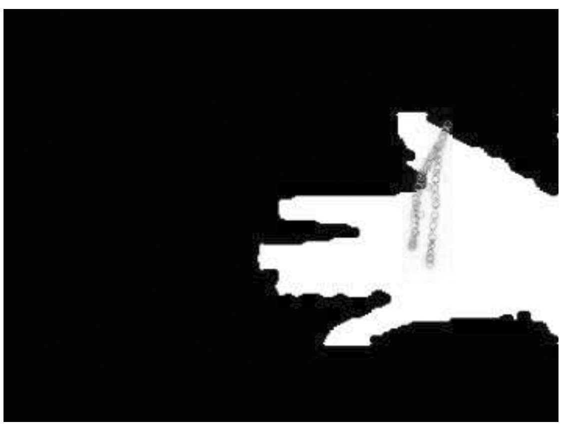

Fig. 6. Hand motion tracking.

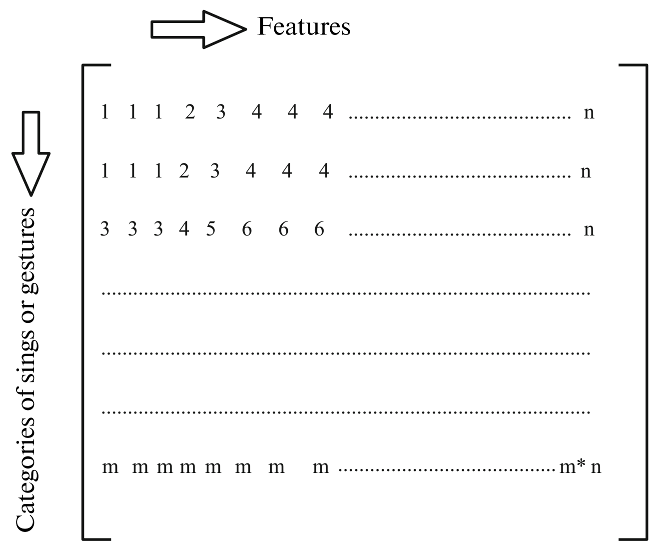

Fig. 7. Feature vector.

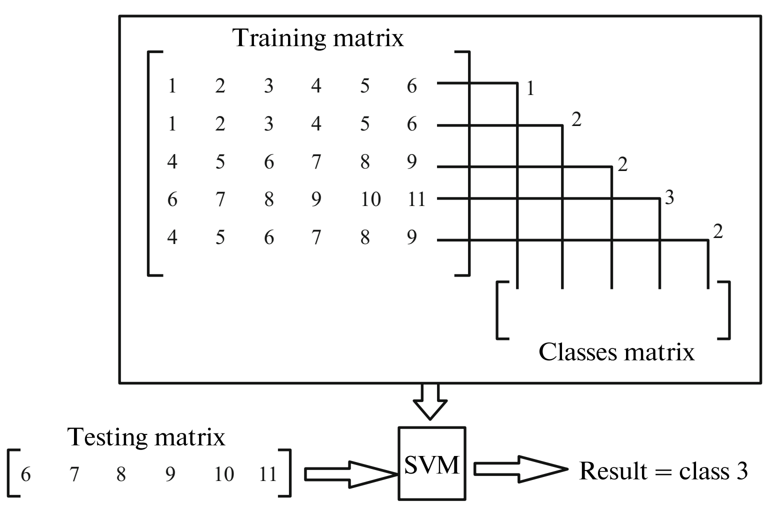

Fig. 8. SVM classification.

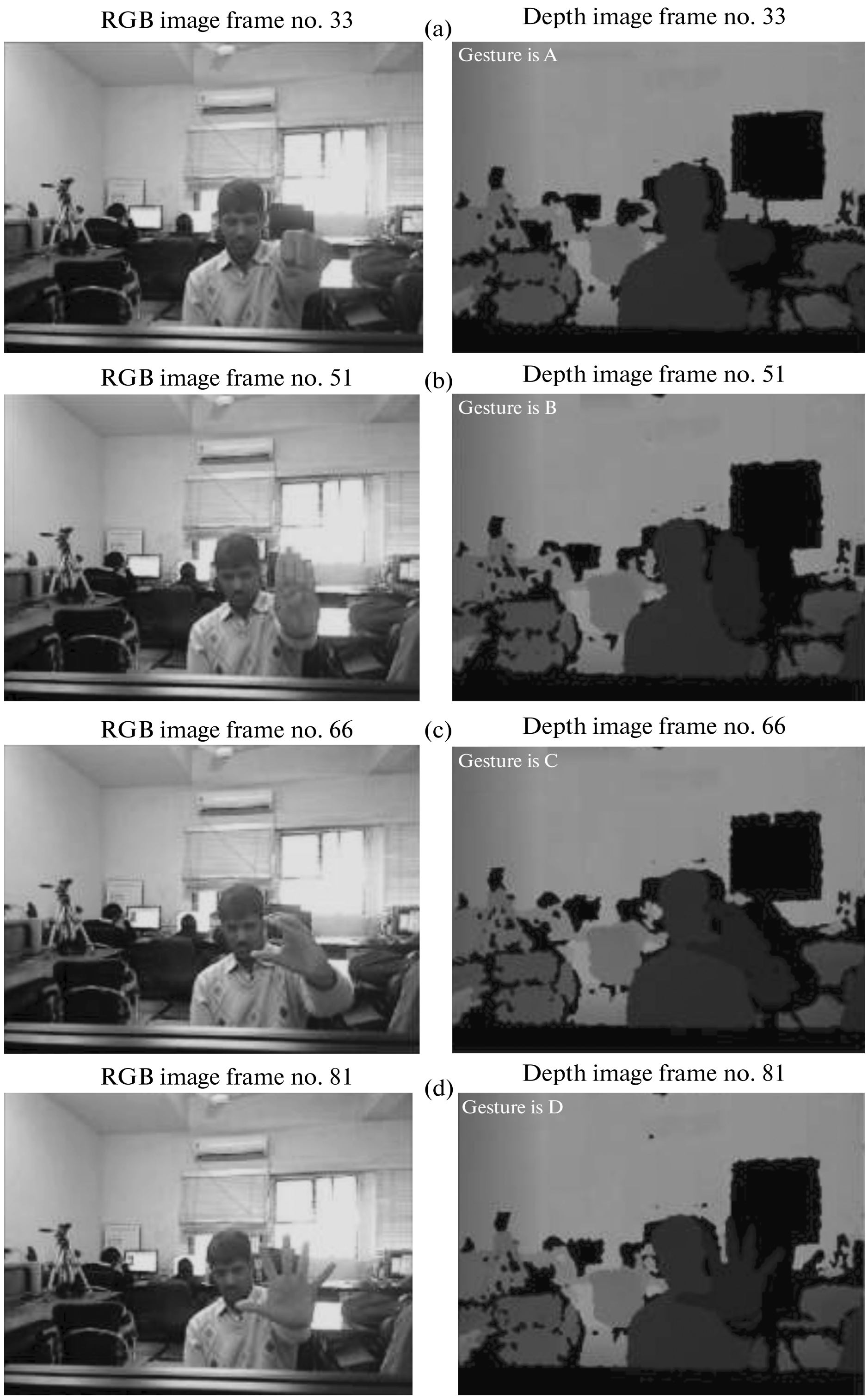

Fig. 9. (a) Gesture A, (b) gesture B, **(**c) gesture C, (d) gesture Hello.

ence between values of centroid co-ordinates of two successive frames was stored in the feature vector. Tracking in proposed system is shown in Fig. 6. Dark circles define actual centroid and light circles define tracked centroid.

3.3. Classification Using SVM

Once the feature vector was extracted, it would be concatenated with feature vector from other frames. This can be done until the decided number of frames

Confusion matrix

	A	B	C	Hello
A	20	0	0	0
B	0	20	0	0
C	2	0	18	0
Hello	0	0	0	20

is completed. Proposed system was set to 30 frames. This feature vector was classified using Multi-Class SVM from a set of pre-defined database of 4 sign gestures of "A," "B," "C," and "Hello." The proposed system uses only linear classifier for experiment. Also gesture length is fixed so that severe mismatching of time frames between testing and given feature vectors. The gesture database consists of feature vectors from all the sign gestures which are in the testing system. These are separated by arranging in different class to same signs. Out of these classes SVM Classifier select the class to which input sign gesture belongs. Figure 7 shows the feature vectors for signs.

The database feature vectors, the testing feature vectors and classes are shown in Fig. 8. SVM Classification of input sign vector using database vectors is visualized. In the training three classes of gesture, the testing gesture is belongs to the class 3. The rows 1 and 2 belong to class 1st. Rows 3 and 5 belong to class 2nd. The row 4 belongs to class 3rd. The testing matrix is most similar to class 3rd so after SVM classification, the result comes is class 3. This is the mechanism of system for multiclass SVM classification. The results of four gestures which were part of initial system, are shown in Figs. 9a–9d.

3.4. Performance

This section dedicated to the evaluation and performance of the proposed algorithm sequence. The support vector algorithm for gesture recognition was tested for ISL signs in the paper. The confusion matrix of the implemented system is shown in Table 1.

The proposed systems has 97.5% recognition rate for selected 4 signs of Sign Language. Out of 80 signs 78 are correctly classified.

CONCLUSIONS

The proposed system targets to facilitates disabled people who are not able to hear and there are not many people who can understand their sign language. In India, there are approximately 21 million disabled persons of some kind and out of that 2.8 million are who can't speak or can't hear. This large community would be benefitted by using such mobile systems which can convert their signs into speech and they would get served. The Indian sign language recognition system was implemented and 4 gestures were classified using SVM. Results were 97.5% accurate, which were encouraging. We will extend this work to include all signs to make a complete and robust system for mobile platform. This approach can also be used for implantation for other regional languages.

ACKNOWLEDGMENT

This research was carried out at Central Electronics Engineering Research Institute, Pilani, India as a part of "Supra Institutional Project on Technology Development for Smart Systems." Authors would like to thank to the Director, CEERI Pilani for his active encouragement and support.

REFERENCES

1. R. A. Bolt, "Put that there—voice and gesture at the graphics interface," in *Proc. 7th Annu. Conf. on Computer Graphics and Interactive Techniques* (ACM, New York, 1980), pp. 262–270.
2. X. Cao and R. Balakrishnan, "VisionWand: Interaction techniques for large displays using a passive wand tracked in 3D," in *Proc. 16th Annu. ACM Symp. on User Interface Software and Technology* (Vancouver, 2003), pp. 173–182.
3. W. T. freeman and C. D. Weissman, "Television control by hand gestures," in *Proc. IEEE Int. Workshop on Automatic Face and Gesture Recognition* (Zurich, June 1995).
4. I. V., Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," IEEE Trans. Pattern Anal. Mach. Intellig. **19** (7), 677–695 (1997).
5. A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," Comput. Vision Image Understand. **108**, 52–73 (2007).
6. Chih-Lyang Hwang and Hai-Wu Lee, "The command control by hand gesture with Hu and contour sequence moments and probability neural network," in *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics* (*SMC*) (Anchorage, 2011), pp. 2056–2061.
7. Liu Yun and Zhang Peng, "An automatic hand gesture recognition system based on Viola-Jones method and SVMs," in *Proc. 2nd Int. Workshop on Computer Science and Engineering* (Qingdao, 2009), pp. 72–76.
8. A. Chaudhary, J. L. Raheja, and K. Das, "A vision based real time system to control remote robotic hand fingers," in *Proc. IEEE Int. Conf. on Computer Control and Automation* (Jeju Island, May 1–3, 2011), pp. 118–122.
9. D. Lee and S. Lee, "Vision-based finger action recognition by angle detection and contour analysis," Electron. Telecommun. Res. Inst. J. **33** (3), 415–422 (2011).
10. Te-Hsiu Sun, "K-cosine corner detection," J.Comput. **3** (7), 16–22 (2008).
11. J. L. Raheja, A. Chaudhary, and K. Singal, "Tracking of fingertips and centre of palm using KINECT," in *Proc. 3rd IEEE Int. Conf. on Computational Intelligence, Mod-*

elling and Simulation (Langkawi, Sept. 20–22, 2011), pp. 248–252.

12. Nianjun Liu and B. C. Brian Lovell, "Gesture classification using hidden Markov models and Viterbi path counting," in *Proc. 7th Digital Image Computing: Techniques and Applications Conf.* (Sydney, Dec. 10–12, 2003).

13. M. Elmezain, A. Al-Hamadi, R. Niese, and B. Michaelis, "A robust method for hand tracking using mean-shift algorithm and Kalman filter in stereo color image sequences," Int. J. Inf. Commun. Eng. **6** (3), 131–135 (2010).

14. A. Chaudhary, M. B. L. Manasa, and J. L. Raheja, "Light invariant neuro-vision system for elder/sick people to express their needs into lingual description," in *Proc. Microsoft Research India's Annu. Research Symp. TechVista* (Kolkata, Jan. 20, 2012).

15. Yen-Ting Chen and Kuo-Tsung Tseng, "Multiple-angle hand gesture recognition by fusing SVM classifiers," in *Proc. 3rd Annu. IEEE Conf. on Automation Science and Engineering* (Scottsdale, AZ, Sept. 22–25, 2007), pp. 527–530.

16. G. R. Naik D. Kant, and K. Jayadeva, "Twin SVM for gesture classification using the surface electromyogram," IEEE Trans. Inf. Technol. Biomed. **14** (2), 301–308 (2010).

17. A. Chaudhary, J. L. Raheja, K. Das, and S. Raheja, *A Survey on Hand Gesture Recognition in Context of Soft Computing* (Springer, Berlin, Heidelberg, 2011), pp. 46–55.

18. J. L. Raheja, K. Das, and A. Chaudhary, "Fingertip detection: a fast method with natural hand," Int. J. Embedded Syst. Comput. Eng. **3** (2), 85–89 (2011).

19. M. Hu, "Visual pattern recognition by moment invariants," IRE Trans. Inf. Theory **8** (2), 179–187 (1962).

20. http://censusindia.gov.in/Census_And_You/disabled_population.aspx

Jagdish Lal Raheja received his Master of Technology from Indian Institute of Technology, Kharagpur, India and Ph.D. from Technical University of Munich, Germany. Currently he is Senior Principal Scientist at Digital Systems Group, Central Electronics Engineering Research Institute (CEERI), Pilani, India. He has been a DAAD fellow and visiting scientist to many countries. He has published more than 100 papers in International Journals and peer reviewed conferences. His areas of research interest are digital image processing, embedded systems and human computer interface. He is in the editorial board of several international journals and also reviewer for many reputed conferences. He has been Principal investigator in many important real life projects.

Anand Mishra is working toward his Master of Technology in Instrumentation Engineering from School of Instrumentation, Devi Ahilya Vishwavidyalaya, Indore, INDIA. His areas of research interest are digital image processing, artificial intelligence, VLSI systems, and digital communication.

Ankit Chaudhary received his Master of Engineering in Computer Science and Engineering from Birla Institute of Technology and Science, Pilani, and Ph.D. in Computer Vision, from Central Electronics Engineering Research Institute. His areas of research interest are computer vision, digital image processing, artificial intelligence and robotics. He is member of IEEE and also serves in the editorial board at many international journals.