A Novel Approach to Gesture Recognition in Sign Language Applications Using AVL Tree and SVM

Sriparna Saha, Saurav Bhattacharya and Amit Konar

Abstract Body gesture is the most important way of non-verbal communication for deaf and dumb people. Thus, a novel sign language recognition procedure is presented here where the movements of hands play a pivotal role for such kind of communications. Microsoft's Kinect sensor is used to act as a medium to interpret such communication by tracking the movement of human body using 20 joints. A procedural approach has been developed to deal with unknown gesture recognition by generating in-order expression for AVL tree as a feature. Here, 12 gestures are taken into consideration, and for the classification purpose, kernel function-based support vector machine is employed with results to gesture recognition into an accuracy of 88.3%. The foremost goal is to develop an algorithm that act as a medium to human–computer interaction for deaf and dumb people. Here, the novelty lies in the fact that for gesture recognition in sign language interpretation, the whole body of the subject is represented using a hierarchical balanced tree (here AVL).

Keywords Gesture recognition · Sign language · AVL tree · Support vector machine · Kinect sensor

S. Saha (✉) · S. Bhattacharya · A. Konar
Electronics & Tele-Communication Engineering Department,
Jadavpur University, Kolkata, West Bengal, India
e-mail: sahasriparna@gmail.com

S. Bhattacharya
e-mail: saurav.mtechiar@gmail.com

A. Konar
e-mail: akonar@etce.jdvu.ac

P.K. Sa et al. (eds.), *Progress in Intelligent Computing Techniques: Theory, Practice, and Applications*, Advances in Intelligent Systems and Computing 518,
DOI 10.1007/978-981-10-3373-5_27

1 Introduction

Body gesture plays the most important and guiding aspects in non-verbal communication between two persons. Hence, the main principle lies in recognition of body gesture that imparts certain information as a sign language. For recognition of body gestures, Microsoft's Kinect [1] plays an important role as it aims to detect human body using 20 joints in 3D coordinate space.

Ren et al. [2] developed Finger's earth mover distance-based technique for hand gesture recognition. Li [3] exposed an approach to human gesture recognition where a specific depth information-based threshold value is assigned. The threshold value helps in detection of hand from body, and clustering is carried out by k-means clustering algorithm. Oszust and Wysocki [4] have proposed a method for Kinect sensor-based sign language recognition. Skeletal images of the body as well as shape and position of the hands are the variants of this work. In this paper, Polish Sign Language (PSL) words are studied using k-nearest neighbor (kNN) classifier. This work helps impaired people to hear and interact globally. Le et al. [5] approached to recognize human posture using Kinect for health monitoring framework. The experiment is being performed with four main postures of standing, sitting, bending and lying using support vector machine (SVM). Biswas and Basu [6] also proposed an approach for gesture recognition using Microsoft's Kinect where multiclass SVM is used to categorize eight gestures. Patsadu et al. [7] proposed a comparison study for human gesture recognition using Kinect. Here, backpropagation neural network (BPNN) is used as a classifier. There exists certain shortcoming as the author considered only few specified gesture, but all other possible gestures are not considered.

Proposed methodology is to process the skeleton obtained from Kinect sensor to produce feature in terms of in-order expression from an AVL tree [8]. For this purpose, weight adaptation is executed on the 3D coordinate value acquired for each joint. Moreover, the weight is empirically calculated in such a way such that the z coordinate gets the maximum importance over all other coordinate weightage values (i.e., x and y). Reason is that z coordinate basically deals with the depth value, which is ultimately the area of interest to our proposed work. The values obtained through this operation are basically 20 weighted values for 20 3D joints with respect to a frame. With this, a balanced AVL tree is framed. Since human body structure is balanced with head as a parent node and all other parts as its children nodes, thus in-order traversal to the tree structure is performed. With these features as inputs, SVM [9, 10] classifier based on kernel function is used for recognition of unknown gestures related to sign language with 88.3% accuracy.

2 Proposed Work

The work under taken can be segregated into four stages:

2.1 Detection of Human Body Gesture Using Kinect Sensor

For the implementation of our proposed work for gesture recognition in sign language applications, we have used Microsoft's Kinect sensor [1]. This device has the ability to represent the human body in skeletal form with 20 body joint coordinates in 3D space using software development kit (SDK) v. 1.6. From these 20 joints, we have studied that while recognizing the gestures, all the body joints do not have equal importance due to their motion while displaying the sign languages. Thus, based on this factor, we have differentiated these joints in two different groups namely static (s) and dynamic (d) joints. The s joints are those in which very little variation (nearly equal to zero) is observed in successive frames. Here, d joints are those where a large amount of variations are noticed in the consecutive frames. If there are total S and D number of joints, then

$$S + D = 20 \text{ where } 1 \leq s \leq S \text{ and } 1 \leq d \leq D \tag{1}$$

2.2 Weight Adaptation for Static and Dynamic Body Joints

Our aim is to draw a balanced binary tree from 20 joints where we have empirically adjusted weightage value to the 3D coordinates. The weightage value assigned to each coordinate is based on the contribution of that coordinate to impart fruitful information so as to recognize the translation of the joints from previous frames. Thus, appropriate weights need to be given to the x, y and z coordinate values, and weighted sum of these coordinates is taken as an input to form a node of the AVL tree. The adaptation of these weights for s and d joints is done differently, but a same constraint is followed such that the summation of the weights (respectively, w_x, w_y and w_z) for these three directions is

$$\left(w_x + w_y + w_z\right) \approx 1 \tag{2}$$

Suppose for s and d joints the corresponding 3D coordinate values are (x_s, y_s, z_s) and (x_d, y_d, z_d) correspondingly. Thus, the respective values given to form the nodes $(n_s \text{ and } n_d)$ of the AVL tree are

$$n_s \leftarrow \left(x_s \times w_x^s + y_s \times w_y^s + z_s \times w_z^s\right) \tag{3}$$

$$n_d \leftarrow \left(x_d \times w_x^d + y_d \times w_y^d + z_d \times w_z^d\right) \tag{4}$$

where superscript for weight w is given to indicate the type of the joints s and d.

2.3 Feature Extraction Using In-Order Expression from AVL Tree

For human body, the joints obtained using Kinect sensor are balanced from the laws of nature. So as to implement the structure of human body to a computer terminology, binary tree approach is implemented. Binary tree is an unbalanced tree since whenever a new data is being added to it, the height of left and right sub-trees of all the nodes differs. This disadvantage is overcome by introducing a balance factor to any node, which ultimately results into an AVL tree [8]. Here lies our novelty as we have taken the in-order expression for the AVL tree for each skeleton. The in-order expression is given priority when compared with its counter parts, e.g., post-order and pre-order due to the following reason. For instance, Kinect sensor defines shoulder left as a parent node followed by elbow left, wrist left, hand left as its components. In case of in-order evaluation, the same kind of definition is observed in which the parent node represents the vital components and its children are sub-part of the parent node. Thus, to represent the body joints as a balanced tree structure, we have implemented in-order traversal to form our feature space.

Suppose there are total t numbers of frames obtained to express a particular gesture g. Thus, for a particular frame $i(1 \leq i \leq t)$, the in-order expression is IO_i, which is comprised of the arrangement of 20 joints to form the balanced tree. The total feature space becomes,

$$IO = [IO_1\, IO_2\, \ldots\, IO_i\, \ldots\, IO_{t-1}\, IO_t]^T \tag{5}$$

2.4 Classification Using SVM

Support vector machine (SVM) computes as a non-probabilistic binary linear classifier that separates input vectors into two classes. In case of a linear SVM [11], it is performed by building a hyper-plane depending on the support vectors. This phenomenon is only applicable when the data are linearly separable. However, this condition is accomplished accounting a kernel function [9, 10]. The overall work flow is presented in Fig. 1.

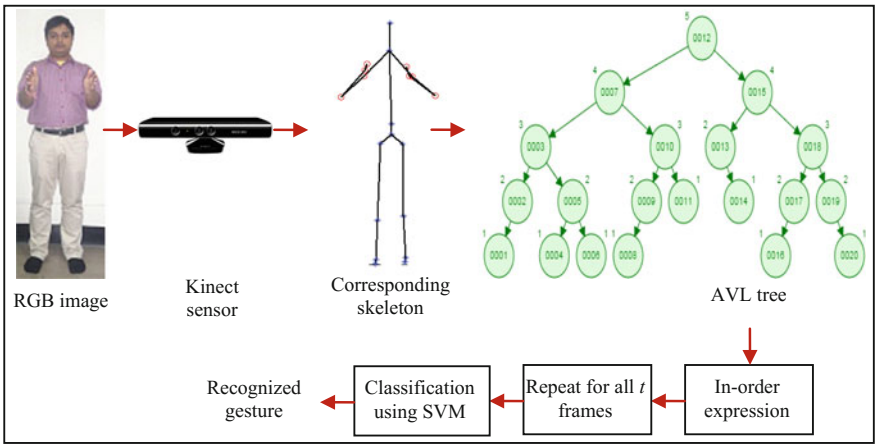

Fig. 1 Block diagram for gesture recognition in sign language applications using AVL tree and SVM

3 Experimental Results

The detailed results obtained while performing the experiment are given in this section. The 12 gestures related to sign language applications are: Move down to up, Move up to down, Move left to right, Move right to left, Pushing, Grabbing, Chirognomy, Waving, Quenelle, Self-clasping, Applause and Fist.

Each gesture is taken for 3-s duration, i.e., t is 90 (=3 s $\times$ 30 frames/s) frames. For the implementation of the proposed work, we have created three different datasets after accumulating data from three distinct age groups 25 $\pm$ 5 yrs, 30 $\pm$ 5 yrs and 35 $\pm$ 5 yrs.

The dimension of feature space IO for each gesture is 20 $\times$ 90. For the proposed work, the empirically calculated values are $s = 12$, $d = 8$, $w_x^s = w_y^s = w_z^s = 0.333$ and $w_x^d = 0.275$, $w_y^d = 0.275$, $w_z^d = 0.450$. For the unknown RGB and skeleton, images for frame number 30 are given in Fig. 2 and the in-order expression generated for the AVL tree (provided in Fig. 2 using OpenCV software) is IO_{54} = [0.803(ankle left) − 0.843(knee left) − 0.974(ankle right) − 1.043(knee right) − 1.149(hip left) − 1.031(hip right) − 1.049(hand right) − 1.050(foot left) − 1.237 (foot right) − 1.149(hip center) − 1.075(spine) − 1.041(shoulder left) − 0.748(wrist right) − 0.570(shoulder center) − 0.437(elbow left) − 0.393(wrist left) − 0.797 (elbow right) − 0.651(head) − 0.532(shoulder right) − 0.506(hand left)]. Here, the exact coordinate values obtained after weight adaptation with corresponding joint names are given. The unknown gesture is correctly recognized as 'Applause.'

The comparative framework includes k-nearest neighbor (kNN) [4], ensemble decision tree (EDT) and backpropagation neural network (BPNN) [7] with respect to precision, recall, accuracy, error rate and timing complexity. The analysis results

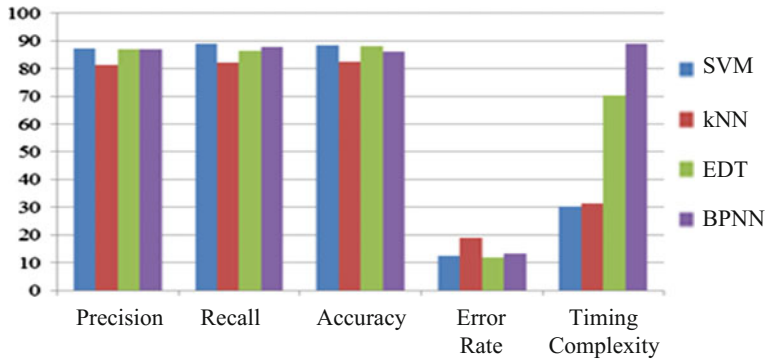

Fig. 2 Performance analysis results with respect to precision, recall, accuracy, error rate and timing complexity (in 10^{-1}s)

Table 1 Friedman test

Algorithm	Dataset 1	Dataset 2	Dataset 3	Average ranking	χ_F^2
SVM	1	2	1	1.3	8.2
kNN	4	4	4	4.0	
EDT	2	1	2	1.7	
LMA-NN	3	3	3	3.0	

are provided in Fig. 2, from where it is evident that SVM is the best choice for our proposed work.

The results obtained after Friedman test are provided in Table 1. The null hypothesis has been overruled, as ${\chi_F}^2 = 8.2 > 7.8$, the critical value of Chi-square distribution at probability of 0.05.

4 Conclusion

Kinect sensor-based sign language recognition dealing with both hand movement is developed in this work using AVL tree as a feature extraction procedure and SVM as a classifier. This framework yields a recognition rate of 88.3%. The data acquisition is performed by Kinect sensor for 3-s duration. The novelty of the framework lies in representing the complete human body joints in a form of balanced binary tree. Upon which in-order traversal is executed that ultimately yields as an input to SVM as a classifier. This paper has a pitfall that movement is only restricted to upper part of human body, as movement of whole body yields joints value that affects the balanced factor of AVL tree. Thus, our future advancement lies in dealing with movement of both upper and lower part of human body structure.

Acknowledgements The research work is supported by the University Grants Commission, India, University with Potential for Excellence Program (Phase II) in Cognitive Science, Jadavpur University and University Grants Commission (UGC) for providing fellowship to the first author.

References

1. K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.
2. Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *Multimedia, IEEE Trans.*, vol. 15, no. 5, pp. 1110–1120, 2013.
3. Y. Li, "Hand gesture recognition using Kinect," in *Software Engineering and Service Science (ICSESS), 2012 IEEE 3rd International Conference on*, 2012, pp. 196–199.
4. M. Oszust and M. Wysocki, "Recognition of signed expressions observed by Kinect Sensor," in *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*, 2013, pp. 220–225.
5. T.-L. Le, M.-Q. Nguyen, and T.-T.-M. Nguyen, "Human posture recognition using human skeleton provided by Kinect," in *Computing, Management and Telecommunications (ComManTel), 2013 International Conference on*, 2013, pp. 340–345.
6. K. K. Biswas and S. K. Basu, "Gesture recognition using microsoft kinect®," in *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on*, 2011, pp. 100–103.
7. O. Patsadu, C. Nukoolkit, and B. Watanapa, "Human gesture recognition using Kinect camera," in *Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on*, 2012, pp. 28–32.
8. R. W. Irving and L. Love, "The suffix binary search tree and suffix AVL tree," *J. Discret. Algorithms*, vol. 1, no. 5, pp. 387–408, 2003.
9. C. Cortes and V. Vapnik, "Support vector machine," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
10. J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.
11. T. M. Mitchell, "Machine learning and data mining," *Commun. ACM*, vol. 42, no. 11, pp. 30–36, 1999.