2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks

Gesture Recognition From Indian Classical Dance Using Kinect Sensor

Sriparna Saha[1], Shreya Ghosh[2] , Amit Konar[1]
[1]*Electronics and Telecommunication Engineering Dept.*
[2]*School of Bioscience and Engineering*
Jadavpur University
Kolkata, India
{sahasriparna, shreyaghosh215}@gmail.com
konaramit@yahoo.co.in

Atulya K. Nagar
Mathematics and Computer Science Dept.
Liverpool Hope University
United Kingdom
nagara@hope.ac.uk

***Abstract*— This work proposes gesture recognition algorithm for Indian Classical Dance Style using Kinect sensor. This device generates the skeleton of human body from which twenty different junction 3-dimensional coordinates are obtained. Here we require only eleven coordinates for the proposed work. Basically six joints coordinates about right and left hands and five upper body joint coordinates are processed. A unique system of feature extraction have been used to distinguish between 'Anger', 'Fear', 'Happiness', 'Sadness' and 'Relaxation'. This system checks whether the emotion is positive or negative with its intensity information. A total of twenty three features have been extracted based on the distance between different parts of the upper human body, the velocity and acceleration generated along with the angle between different joints. The proposed algorithm gives a high recognition rate of 86.8% using SVM.**

Keywords- angle, feature extraction, gesture recognition, Kinect sensor, skeleton

I. INTRODUCTION

The rich Indian culture has given way to several Indian classical dances. Among the many Indian classical dance forms, some of the popular ones are Bharatnatyam, Odissi, Kuchipudi, Kathak etc. The dance form mainly used here is Odissi. However, the algorithm proposed in this paper also finds relevance in other Indian classical dance forms such as Bharatnatyam, Kuchipudi and Manipudi. Odissi has originated from Orissa, an Indian state. In Odissi, a combination of facial expression, body movements and feet movement is used to depict a specified feeling, also called 'Rasa' [1]. There are various forms of 'rasas' used in this dance like anger, fear, happiness, sadness, sarcasm, shyness etc. Another important aspect of Odissi is its body movement and its feet movement. The body movement involves the movement of torso from one side to another side, while keeping the hip in a stationary state.

With the rapid advancement of technology, usage of the internet has become an inevitable part of the human life. With the proposed algorithm, anybody in any part of the world can attempt to learn the gestures made use in the above mentioned forms of dance and can also evaluate the performance, via the internet. Thus this algorithm aims to serve as an e-learning program, which offers much more flexibility and accessibility to the common man. Moreover, it is also a fast and cost effective way of spreading the Indian classical dance forms worldwide.

In this paper, we have made use of a Kinect sensor which can detect skeleton of the performer while dancing, using 20 body joint co-ordinates [2]. The Kinect sensor detects the 3D image representation of an object [7]. It tracks the skeleton of the person standing in front of within a finite amount of distance using a set of visible and IR cameras [3]. For the recognition purpose, we are using SVM. Support Vector Machine (SVM) is a new way of data mining and pattern recognition [15]. SVM is based on Statistical theory [16] and Optimization Theory [17]. It has many applications in pattern classification such as military, economic and many other fields. It is basically a binary classifier, but multi-class classification is also possible using this, and it gives excellent result there also [18].

Human gesture is mainly described by hand movements [11]. In [4], depth image of hand gestures are taken for training of multi class support vector machine. In [5], Ballet poses are sequenced based on the skeleton obtained using motion capturing devices. Three angles of any leg are taken as the feature and the other leg is kept static. This paper neglects the difficulty arises due to movement of two legs at the same time. But in our paper we consider movement of two hands at the time as are able to identify gesture correctly with accuracy 86.8%. In [6], Seventeen different ballet postures are recognized using fuzzy image matching technique. Here the skeleton of the dancer is approximated using straight line and quadrants of the straight lines are calculated. Any straight line can belong to more than one quadrant and this belongingness of the straight line to a specific quadrant is calculated using fuzzy logic. This process is much more complex and time consuming with respect to our proposed algorithm. The usage of Kinect itself provides straight line approximated skeleton, so we can join twenty co-ordinates of the human body accordingly by straight line. When segmentation is used at the pre-processing stage [8], lighting condition of the background plays an important role for posture recognition.

978-0-7695-5042-8/13 $26.00 © 2013 IEEE
DOI 10.1109/CICSYN.2013.11

This problem can be neglected by the help of Kinect sensor [9].

In our experiment we have worked on recognizing five types of gestures. They are ‘Anger’, ‘Fear’, ‘Happiness’, ‘Sadness’, and ‘Relaxation’. While ‘anger’ is expressed by aggressive hand movements, ‘relaxation’ shows static hand postures but kept at a definite angle, as is discussed in this paper. A total of 23 features have been extracted for each video sequence for depicting a specified gesture. Among them, eight are from both of the hands and the rest are from head and the body. The feet information has not been taken into account in this as it is not so important in conveying gestures in an Indian classical dance forms. The time complexity of the proposed algorithm is 5.615 sec in an Intel Pentium Dual Core processor running Matlab R011b for a video consisting of 200 frames. Secondly the proposed work in our paper is applicable in different lighting conditions. In various illumination conditions, where human eye can detect object, Kinect sensor also detects the skeleton of the human body. Recognition rate is as high as 86.8% using SVM. So the work gives a high level of accuracy. This work is not only useful for Indian dance forms but also can be use for other international dance types for gesture recognition purpose. Moreover this work shows promise in the recognition of other complex emotions like disgust, shyness etc.

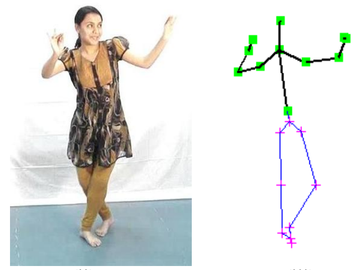
In this paper, we propose an algorithm for gesture recognition in Indian classical dance forms. While Section II gives an overview of Kinect Sensor. Section III explains the features extracted from the subject skeleton in order to design the proposed algorithm. Section IV contains the experimental results using SVM. Finally Section V concludes with future work.

II. Kinect Sensor

The Kinect is an upcoming technology which basically looks like a webcam as shown in fig. 1(i) [13]. It detects the 3D image representation of an object. It tracks the skeleton of the person standing in front of within a finite amount of distance. It has a set of visible IR and RGB cameras. The IR cameras are responsible for sensing the skeleton and hence the body postures irrespective of the color of the performer’s dress or distance from the camera. The Kinect sensor has the capability of capturing 3D motion, facial gesture and voice gestures with the help of the RGB camera, the depth sensor or the IR camera and the multi-array [13].

The green squares in fig. 1(iii) shows the body joints points required in the proposed work and these joints are connected using bold black lines while the other body joints are represented using red plus signs and joined via the thin blue lines. The actual RGB image for the skeleton is shown in fig. 1(ii).

(i)



(ii) (iii)

Figure 1. (i) Kinect sensor, (ii) RGB image of whole body, (iii) Skeleton image

III. Feature Extraction

The model shown in fig. 2 gives an idea about the classification of gestures based on type and intensity [12]. According to it, if ‘Happiness’ and ‘Relaxation’ are positive gestures then ‘Anger’, ‘Fear’ and ‘Sadness’ are negative gestures. On the other hand, while ‘Anger’, ‘Fear’ and ‘Happiness’ gestures indicate higher arousal or intensity, ‘Relaxation’ and ‘Sadness’ gestures show the opposite characteristics. The block diagram shown in fig. 3 provides a basic overview of Russell’s circumflex model [14], which is applied in the proposed algorithm for gesture recognition in Indian dance forms.

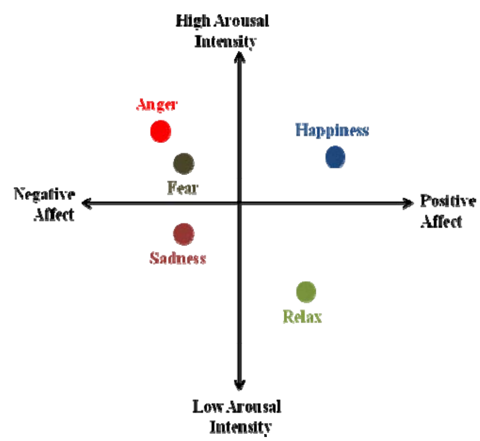


Figure 2. Model for Gesture Classification

The distance of hand with respect to spine for both hands corresponds to two features. The same holds true for the elbows. Then the velocity and acceleration for the hands and the elbows gives a total of sixteen features. The angle of the head with respect to the spine and the shoulder centre gives one feature while the angle between the shoulder, elbow and wrist corresponds to two features for both hands. Therefore, this gives a total of twenty three features.

A. *Distance of hand and elbow with respect to spine*

The Euclidean distance of hand and elbow is considered as a feature, which is represented by the distance value. In ‘anger’, hands move in a specified manner, while movement of hands in ‘sadness’ are completely different. Though there is a similarity in the movement of hands in ‘sadness’ and that in ‘fear’, these two gestures are distinguished based on other features, which are mentioned later on. ‘Anger’ is characterized by fast ‘to and fro’ movements of hands which are accompanied by vibration of the palm or the hands. Therefore, the distance of hands and elbow constantly increases and decreases with respect to the spine. This is clearly indicated in the plots obtained for anger, which shows ‘peaks’ ,as shown in fig. 9. In case of 'sadness' as well as 'fear' gesture, hands come closer to the body. Consequently the absolute distance of the hands and the elbows with respect to the spine decreases. This gradual decrease in amplitude is also indicated in fig. 10. In the second skeleton view of fig. 3s, frame no 55, the distance between hand left and spine is shown using red arrow line and for elbow right and spine is marked by violet arrow line. Specific pattern is obtained for different gestures [10]. The Euclidean distance is calculated based on the following formula

$$Dist=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2} \quad (1)$$

where Dist is the Euclidean distance between two points with co-ordinates (x1,y1,z1) and (x2,y2,z2) respectively.

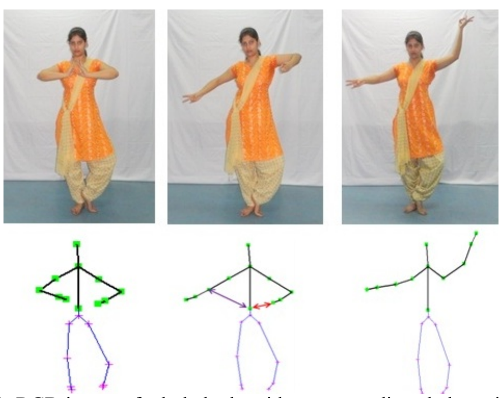


Figure 3. RGB image of whole body with corresponding skeleton image of ‘Anger’ gesture for frame no 25, 55 and 65 from left to right and top to below.

B. Calculation of velocity and acceleration for hand and elbow

It has already been mentioned that the hand movement in ‘anger’ is very fast and is accompanied by vibration of the palm or the hands. Therefore, the maximum velocities of the hands are much greater than that in any other gesture. For example, in ‘anger’, the maximum acceleration is always greater than the threshold. The threshold values are always greater than $20m/s^2$, $70m/s^2$ for the elbows and the hands respectively. Experimental results suggest that the maximum and minimum values for velocity and acceleration in sadness and fear are very low when compared to those for anger. Velocity of a joint is calculated using the displacement of the joint is two consecutive frames. Here the video is taken at 30 frames per sec, so the time becomes (1/30) sec. Acceleration is obtained by dividing the change in velocity again by (1/30) sec for two consecutive frames.

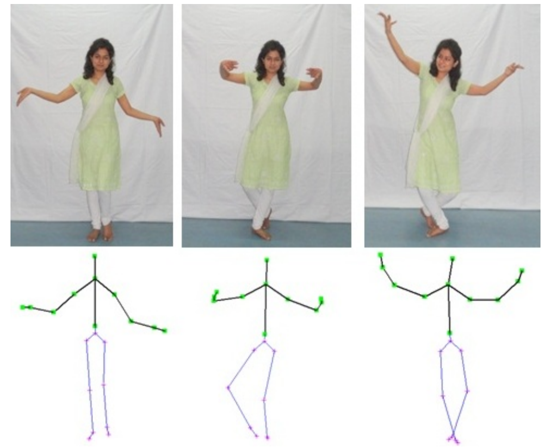


Figure 4. ‘Happiness’ gesture is explained using RGB image and skeleton structure for frame no 45, 65 and 105.

C. Calculation of angle between head, shoulder centre and spine

This feature is chosen in order to differentiate between ‘sadness’ and ‘fear’. Since, there is not much difference between the hand movements to express ‘sadness’ and ‘fear’ as in both the cases hands and elbow come closer to the body, the only way to distinguish these two gestures is by the calculation of the angle between head, shoulder centre and spine. In case of ‘fear’, this angle remains almost constant with a very small variation range of around 2 degrees to 5 degrees while in case of ‘sadness’ the variation range is as large as 25 degrees to 35 degrees. The second two skeletons of ‘fear’ and ‘sadness’ gestures, the angle between head, shoulder centre and spine is shown using red curved line in fig. 5 and 7 respectively. For five different emotions, the angle is plotted as shown in fig. 10. Here sampling is done to remove noise. The sampling rate is taken as 30 frames per sec which is again empirically chosen.

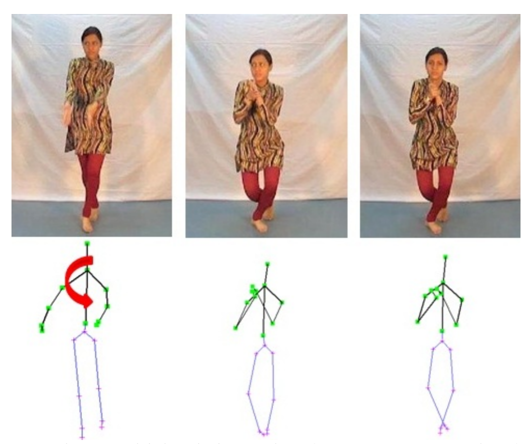


Figure 5. RGB image with its skeleton view for ‘Fear’ gesture using frame no 45, 85 and 140.

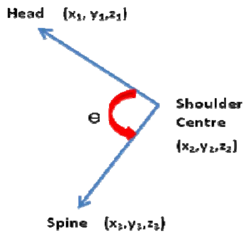


Figure 6. Angle between Head, Shoulder Centre and Spine

Let, the co-ordinates of head, shoulder centre and spine are (x_1, y_1, z_1), (x_2, y_2, z_2) and (x_3, y_3, z_3) respectively. The vectors formed by head, shoulder centre and spine, shoulder centre are

$$vec1 = (x_1 - x_2)\vec{i} + (y_1 - y_2)\vec{j} + (z_1 - z_2)\vec{k} \quad (2)$$

$$vec2 = (x_3 - x_2)\vec{i} + (y_3 - y_2)\vec{j} + (z_3 - z_2)\vec{k} \quad (3)$$

The angle between two vectors is calculated by the following equations

$$A = \mathrm{norm}(\mathrm{cross}(\mathrm{vec1}, \mathrm{vec2}), \mathrm{dot}(\mathrm{vec1}, \mathrm{vec2}) \quad (4)$$

$$\mathrm{angle} = \frac{\mathrm{atan2(A)}}{\mathrm{pi}} \times 180^0 \quad (5)$$

where norm function returns a matrix

atan2 denotes the arctangent by the equation described below whose range is $\left(-\pi/_2, +\pi/_2\right)$

$$\mathrm{atan}\,2(b,a) = \begin{cases} \arctan(b/_a) & a > 0 \\ \arctan(b/_a) + \pi & b \geq 0, a < 0 \\ \arctan(b/_a) - \pi & b < 0, a < 0 \\ +\pi/_2 & b > 0, a = 0 \\ -\pi/_2 & b < 0, a = 0 \\ undefined & b = 0, a = 0 \end{cases} \quad (6)$$

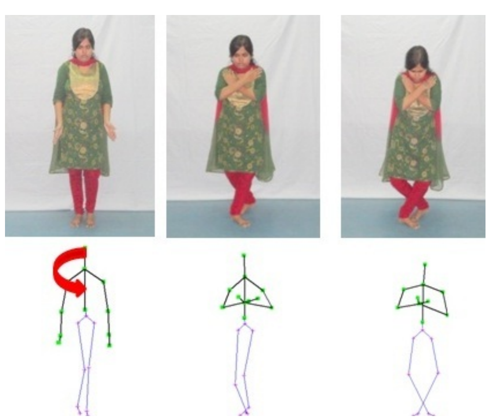


Figure 7. RGB image with its skeleton view for 'Sadness' gesture using frame no 45, 85 and 140.

D. *Calculation of angle between shoulder, elbow and wrist*

This feature is important as it distinguishes 'Relaxation' from the other gestures. Here we are calculating percentage of angle which lies in between 80 to 100 degree. In 'Relaxation' emotion the angle between shoulder centre, elbow and wrist is 90 degree. Since the structure of human body varies from person to person, this angle can be between 80 degrees to 90 degrees. This has been determined experimentally. A 10 degree range is given to both the sides. The percentage of angle means the number of angle whose value is in between 80 degree to 100 degree divided by the total number of angles. For each frame, we are calculating a different angle. So, the total number of angle is equal to the total no of frames. Therefore, the equation for calculating percentage of angle can written as

$$Percentage = \frac{Required\_frames}{Total\_no\_of\_frames} \quad (10)$$

where Required_frames are the no of frames for which the angle value of head, shoulder centre and spine is in between 80 to 100 degree.

For relaxation, this angle lies between 80 degrees and 100 degrees for above 90% frames unlike in other gestures. Here, the angle between shoulder, elbow and wrist is marked by red curved line for both the hands in fig. 8. The angle calculation is already explained in equ. 6.

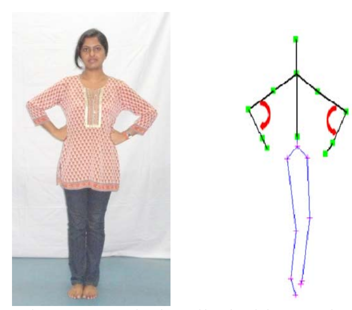


Figure 8. 'Relaxation' gesture is described with RGB image and skeleton view for frame no 25.

IV. EXPERIMENTAL RESULTS

Algorithm for the Proposed Work

Step 0 Create an initial database of skeletons for five emotions

BEGIN

Step 1 Determine the acceleration for hand and elbow for both the hands

Step 2

If

Acceleration for hand is greater than 70 m/s^2 and for elbow 20 m/s^2

Then the unknown emotion is 'Anger'.

Else

If

Step 3

Calculate the distance between spine and hand

Distance is decreasing continuously

Step 4

Examine the angle between head, shoulder centre and spine

If

Angle is decreasing continuously

Then the unknown emotion is 'Sadness'.

Else

Then the unknown emotion is 'Fear'.

End If

Else

If

Step 5

Determine the percentage of angle between elbow, shoulder centre and wrist

Percentage is greater than 90%

Then the unknown emotion is 'Relaxation'.

Else

Then the unknown emotion is 'Happiness'.

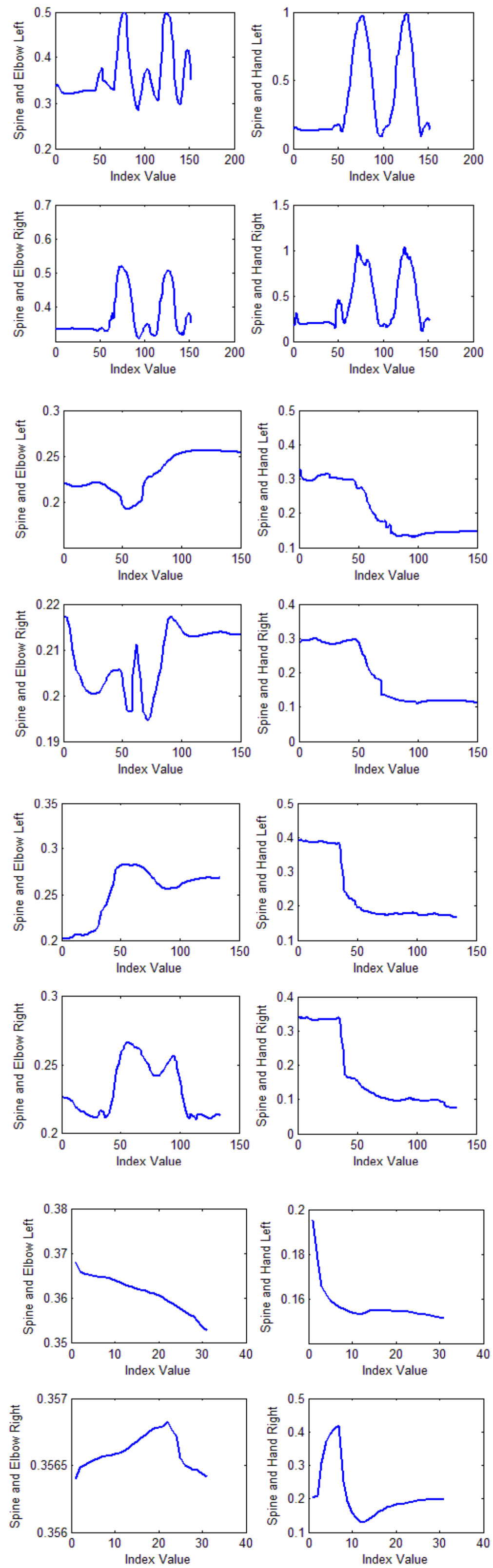
End If

End If

End If

END

Experimental results suggest that in the gesture 'Anger', the threshold values for maximum acceleration of hands and elbows is 70 m/s^2 and 20 m/s^2 respectively. For subject 1, the velocity of elbows ranges between a maximum of 1.1478 m/s and a minimum of 6.9001e-004m/s. For hands, this range lies in between 3.4501 m/s to 2.2606e-004 m/s. The gestures 'Sadness' and 'Fear' gives comparatively lower values of velocity. For 'Fear' it ranges between 0.158 m/s to 1.5827e-004 m/s while in 'Sadness' it is 0.214 m/s to 5.4002e-004 m/s, for elbows as well as hands. Fig. 9 shows the Absolute distance plots for the five gestures. Fig. 10 gives us the plots for the angle of head with respect to the head and the shoulder centre for all the gestures. Sampling rate for fig. 10 is 30 frames per second. The overall accuracy obtained is 86.8% using SVM. We have taken datasets from 10 subjects. From the datasets 80% data are used for training purpose using 'svmtrain' function and remaining data are used for classification using 'svmclassify' function. Computation time is 5.615 sec in an Intel Pentium Dual Core processor running Matlab R011b for a video consisting of 200 frames



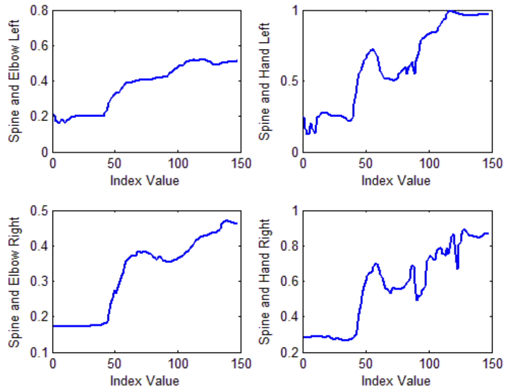


Figure 9. Four hand joint distance from spine are shown for 'Anger', 'Fear', 'Sadness', 'Relaxation', 'Happiness' gestures respectively.

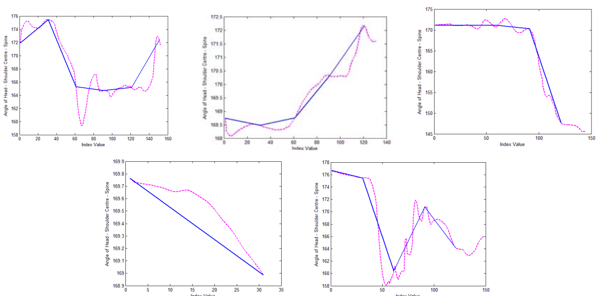


Figure 10. Angle of head with respect to shoulder centre and spine are shown for 'Anger', 'Fear', 'Sadness', 'Relax', 'Happiness' gestures respectively.

V. CONCLUSION AND FUTURE WORK

It is the first time that body gesture is recognized for Indian dance style through our proposed algorithm. The Kinect sensor is neither affected by the dress or skin color of the subject nor is it disturbed by the distance of the subject from the camera. This device is easily available with a low cost, so acceptance of this paper is wide. Moreover, this paper can be used for e-learning of Indian classical dance to teach novice dancers how to convey different gestures of the same. We have achieved overall accuracy rate of 86.8% with a very little computation time of 5.615 sec in an Intel Pentium Dual Core processor running Matlab R011b for a video consisting of 200 frames.

In this paper we have worked with five gestures. However, there are many other complex gestures as well. So the future scope of this experiment includes working with more complex gestures like 'Surprise', 'Disgust', 'Shyness' etc., which is currently in progress.

ACKNOWLEDGMENT

I would like to thank University Grant Commission, India, University of Potential Excellence Programme (Phase II) in Cognitive Science, Jadavpur University.

REFERENCES

[1] http://ccrtindia.gov.in/odissi.php

[2] K. Lai, J. Konrad, P. Ishwar, "A gesture-driven computer interface using Kinect, " IEEE Southwest Symposium on Image Analysis and Interpretation, 22-24 April 2012.

[3] F. Ryden, "Tech to the future making a "Kinection" with Haptic Interaction," IEEE Potentials, 2012.

[4] K. K. Biswas, S. K. Basu, "Gesture Recognition using Microsoft Kinect," Proceedings of the 5th International Conference on Automation, Robotics and Applications, Dec 6-8, 2011, Welington, New Zealand.

[5] A. LaViers, Y. Chen, C. Belta, and M. Egertedt, "Automatic Sequencing of Ballet Poses," IEEE Robotics & Automation Magazine, 13 Sep. 2011.

[6] S. Saha, A. Banerjee, S. Basu, A. Konar, and A. K. Nagar, "Fuzzy Image Matching for Posture Recognition in Ballet Dance," 2013 International Conference on Fuzzy Systems, (submitted for publication).

[7] W. Zeng, "Microsoft Kinect Sensor and Its Effect," IEEE Computer Society, 2012.

[8] G. Feng, Q. Lin, "Design of elder alarm system based on body posture recognition," International Conference on Anti-Counterfeiting Security and Identification in Communication (ASID), 18-20 July 2010.

[9] T. Gill, J. M. Keller, D. T. Anderson, R. H. Luke III, "A system for change detection and human recognition in voxel space using the Microsoft Kinect sensor," IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 11-13 Oct. 2011.

[10] Agnès Just, Yann.R and Sébastien.M, "Hand posture classification and recognition using the modified census transform," Proc of the IEEE Int. Conf. on automatic Face and Gesture Recognition, 2006.

[11] Clay Alexis. La branche émotion, un modèle conceptuel pour l'intégration de la reconnaissance multimodale d'émotions dans des applications interactives : application au mouvement et à la danse augmentée. Phd thesis, University of Boredeaux 1.2009.

[12] L. Cheng, Q. Sun, H. Su, Y. Cong, S. Zhao, "Design and Implementation of Human-Robot Interactive Demonstration System Based on Kinect," 24th Chinese Control and Decision Conference (CCDC), 23-25 May 2012.

[13] http://en.wikipedia.org/wiki/Kinect

[14] J. A. Russell, M. Lewicka, T. Niit, "A cross-cultural study of a circumplex model of affect," Journa; of Personality and Social Psychology, 1989, vol. 57, no. 5, 848-856.

[15] Cortes C, Vapnik V.Support Vector Networkers [J]. Machine Learning, 1995,20:273-295

[16] Vapnik V, Statistical leaming theory [M], New York: Wiley, 1998.

[17] Hu Yuquan: Cause of Operation Research [M] Beijing: Tsinghua University Press, 2003.

[18] F. F. Chamasemani, Y. P. Singh, "Multi-class Support Vector Machine (SVM) classifiers – An Application in Hypothyroid detection and Classification," Sixth International Conference on Bio-Inspired Computing: Theories and Applications, 2011.