AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 3, ISSUE 1
ISSN: 2054-7412

Real Time Human Action Recognition using Kinematic State Model

[1]Geetanjali V. Kale and [2]Varsha H. Patil
Department of Computer engineering, MCERC, Nashik, Maharashtra, India
[1]gilkale@gmail.com, [2]varsha.patil@gmail.com

ABSTRACT

Human action recognition has tremendous applications in interdisciplinary domain and it's challenges kept researchers busy worldwide. Variety of applications gave rise to different representation and recognition methods. Posture can either be represented by shape features or skeleton features. We have represented action using sequence of postures and skeleton features are used for posture representation. Proposed work recognizes Yogasana from real time video. Yogasana is a type of exercise, in which specific sequence of the postures needs to be performed. Regular practice of it shows tremendous benefits in physiological and psychological disorders. System uses skeleton data of twenty human joints for representation of posture. Asanas are represented by kinematic state model using skeleton data provided by Kinect sensor. System is tested on 120 real time video sequences captured on four different subjects performing three asnas. Our system gave 96% recognition rate.

Keywords: human motion recognition, kinematic human model, Yogasana, Kinect.

1 Introduction

Human motion recognition from video or images is interdisciplinary domain having social, industrial and educational benefits. It is key requirement for many applications like smart surveillance, human computer interaction (HCI), behavioral biometric, motion analysis [1-4]. General steps in any human motion recognition applications are (a) track the human from given video, (b) represent human and their actions efficiently, (c) recognize the action or activity performed i.e. classify action in its correct class.

Tracking human from given video stream is low level vision problem. With the advances in domain video sensors like ASUS Xtion Pro Live and Microsoft Kinect tracks the human in video. Microsoft Kinect provides depth as well as human skeleton stream of video data along with RGB [5]. Kinect is originally designed for gaming, but, soon it became popular in Computer Vision community due to quality of its skeleton and depth data.

Interdisciplinary applications gave rise to different representation and recognition strategies. Broadly, human posture representation methods are classified as humanoid body model approaches and image based approaches [6]. Humanoid body model uses skeleton data and body kinematics. Image based approach uses body silhouettes or contours. Body silhouettes or contours can be extracted from either depth data or RGB stream. Image based approaches are very sensitive to change in human action, camera view and human size. Whereas, kinematic approach is camera view invariant. Proposed work uses kinematic data for posture representation.

DOI: 10.14738/aivp.31.1000
Publication Date: 27th February, 2015
URL: http://dx.doi.org/10.14738/aivp.31.1000

Section 2 discuss about previous work done in domain, section 3 explains proposed system along with methodology used, section 4 gives experimental details along with results and last section concludes paper with future scope.

2 Previous Work

Human motion recognition has a long history. Aggarwal and Ryoo, T. B. Moeslund et al., Poppe, and Pavan Turaga et al. have discussed insights of domain in detail [1]-[4]. This section discuss about work carried out on depth and skeleton data provided by Microsoft Kinect sensor after its launch in 2010. J. Han et.al. gave insights for researchers to exploit and improve computer vision algorithm using Kinect [13]. Authors also discussed object tracking and recognition, human activity analysis, hand gesture analysis, indoor 3D mapping using Kinect.

S. Nowozin and J. Shotton proposed concept of action points to recognize action from online low latency data [7]. Authors applied firing HMM and random forest direct classification on action points for low latency data. Results are analyzed on Weizmann repetitive actions, Kinect actions Dataset. The random forest recognition system based on the skeletal joint data has been successfully integrated into a game title that is currently being sold in retail stores. But proposed approach is suitable only for momentary and discrete movements. Assumptions do not fit for continuous data like walking or well defined steps in dance. M. Yang et al. has represented posture using angles made by body parts and achieved almost 100% results except for throw action [8]. X. Yang and Y. Tian used 3-D positions of body joints and combined action information using static posture, motion and offset [9]. Feature channel $f_c = [f_{cc},\ f_{cp},\ f_{ci}]$ where f_{cc} is posture feature, f_{cp} is motion feature and f_{ci} is offset feature, results in high dimensional data. They further used PCA to reduce redundancy and noise and also obtained EigenJoint representation for each frame. J. Sung et al. used a hierarchical maximum entropy Markov model (MEMM) for detection and recognition of twelve different activities performed by four people in different environments, such as a kitchen, a living room, an office, etc. [10]. Using RGBD data provided by Kinect sensor and achieved an average performance of 84.3% when the person was seen before in the training set (and 64.2% when the person was not seen before).

Z. Gao et al. have taken advantage of multi-modal information provided by Kinect and proposed multi-modality information collaborative representation and recognition (MMCRR) [11].

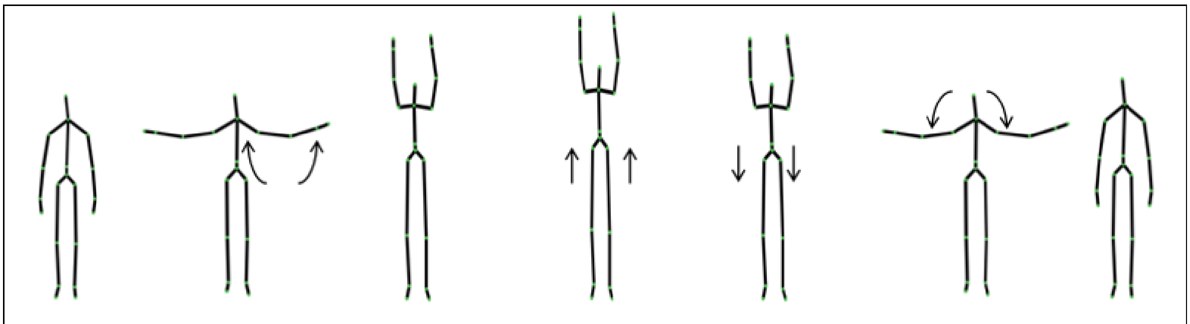
Major contour or silhouette based approaches fails to give anthropometric and view point invariance for action recognition. As persons shape changes with change in view point and size of person changes with change in anthropometry. 3-D kinematic approach can be used to overcome problem of view point invariance for recognition of human action. Proposed system uses kinematic approach and skeleton data obtained from the inexpensive Microsoft Kinect sensor. Our system recognizes performed *Yogasana* from given video. *Yogasana* is a type of exercise composed of sequence of postures. Regular practice of it shows tremendous benefits on health.

3 Proposed System

The present age of speed and competition has increased the stresses and strains resulting in an increasing prevalence of life style-related health problems such as hypertension, cardiac diseases, bronchial asthma, diabetes neurosis and depressive illness, low back pain [12][13]. With growing scientific evidence, Yogasana is emerging as an important health behavior-modifying practice to achieve good states of health. Regular practice demonstrated the beneficial effects on health behavior in many life style-related problems. Regular practice of Yogasana also reduces the stress, which in result improves the performance of students in examination, Yoga also helps for improving memory, intelligence and creativity [14].

Yoga is the name of one of the six orthodox philosophical schools and the Yoga school was founded by Patanjali. The 29th Sutra of the 2nd book of Patanjali gave concept of "Ashtanga Yoga" ("Eight-Limbed Yoga") [15]. Asana is third limb of "Ashtanga Yoga"that concentrates on practice of specific breathing techniques and postures. Proposed work concentrates only on postures and does not consider breathing technique.

In proposed work we have considered standing postures without twist action. Three types of Asanas used are (i) Samasthiti Tadasana,(ii)Urdhva Hastasana Tadasana, (iii) Veerbhadrasana. Detailed posture sequence of Urdhva Hastasana Tadasana is shown in Figure 1. Figure1 (a) shows skeleton representation of posture sequence to be followed in Urdhva Hastasana Tadasana and Fig1 (b) shows state transition of one pose to another sequentially. Postures can be viewed as states and Asana is represented by sequence states. Posture sequence considered in Urdhva Hastasana Tadasana is Samasthiti, hands apart, hands up, whole body up (stand on toes), body down (stand on foot), Hands Apart (intermediate posture considered for hands down), Hands down. In Samasthiti Sama means upright, straight, unmoved and Sthiti means standing still or steadiness. Samasthiti Tadasana has three key postures Samasthiti, body up and body down. Veerbhadrasana has seven key postures i.e. Samasthiti, legs apart, hands apart, knee bend, knee straight, hands down, legs near. Yogasana has standing, seating and lying down postures. It is transition of posture sequence from one state to another.



(a) Samasthiti	(b) Hands Apart	(c) Hands up	(d) Body up	(e) Body Down	(f) Hands Apart	(g) Hands Down

Fig 1(a): States using skeleton representation

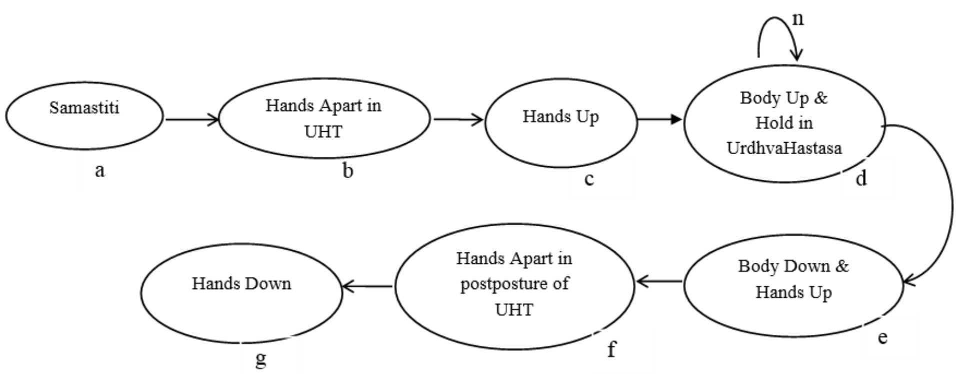


Figure 1(b): Sequence of State transition
Figure 1: Urdhva Hastasana Tadasana

$Yogasana = \{P_{1,} P_{2,} P_{3,} \dots \dots, P_n\}$, where P_i is posture in given Asana 1 ≤ i ≤ n, n is number of keyframes in particular *Asana*. Each posture is given by articulated human model represented using 20 joints, $P_i = \{J_{i,1}, J_{i,2}, J_{i,3}, \dots \dots, J_{i,20}\}$. Each joint can be described in space using three parameter triplet $\langle X_{ij,} Y_{ij,} Z_{ij} \rangle$, where 1 ≤ j ≤ 20 and i is posture 1 ≤ i ≤ n. $\therefore Posture\ P_i$ = $\{\{J^1_{i,x}, J^1_{i,y}, J^1_{i,z}\}\}, \dots, \{\{J^{20}_{i,x}, J^{20}_{i,y}, J^{20}_{i,z}\}\}$. Motion is represented by change in joint positions with respect to *Samasthiti*. Samasthiti is taken as reference posture for initialization$P_1 = \{\{J^1_{1,x}, J^1_{1,y}, J^1_{1,z}\}\}, \dots, \{\{J^{20}_{1,x}, J^{20}_{1,y}, J^{20}_{1,z}\}\}$. Difference in two postures can be given by, $diff(\partial P_i, \partial P_{i+1})$. This will result in vector $\{\partial J_1, \partial J_2, \partial J_3, \dots, \partial J_n\}$. Next posture / state in sequence is identified from difference of most informative joints of reference posture and recent posture.

$$\{\delta^k_{i,x}, \delta^k_{i,y}, \delta^k_{i,z}\} = diff\left(\{J^1_{1,x}, J^1_{1,y}, J^1_{1,z}\}, \{J^{20}_{1,x}, J^{20}_{1,y}, J^{20}_{1,z}\}\right)$$

Change will depend on type of action performed and sequence of postures.

4 Experimentation and Results

System is tested on three types of Asana - Samasthiti Tadasana, Urdhva Hastasana Tadasana, Veerbhadrasana (Warrior Pose - II). X and Y co-ordinates of twenty joint positions of human body obtained from Kinect are considered as feature for recognition. Joint positions considered are center, left and right point of hip and shoulder, left and right point of hand, wrist, elbow, knee, ankle, and foot, spine and head. All three asanas are tested on 4 female subjects and each asan is repeated 10 times i.e. 3×4×10=120 video sequences. Table 1 shows results obtained for three asanas repeated 40 times.

Table 1: Recognition results for three Asana

	Samasthiti Tadasana	Urdhva Hastasana Tadasana	Veerbhadra sana
Samasthiti Tadasana	39 (97.5%)	0	0
Urdhva Hastasana Tadasana	0	39 (97.5%)	0
Veerbhadrasana	0	0	38 (95%)

Our system shows average 96% recognition rate. It is observed that failure of recognition is due to non-identification of some intermediate posture in *asana.*

Fundamental requirement of any action recognition system or algorithm is minimize computational complexity as well as feature vector dimensionality. Our algorithm satisfies both, as we consider only most informative joints for action recognition.

Computational Complexity of each Asana is $3 \times \sum_{i=1}^{m} N_i$ where, m - number of postures in Yogasana and $N_i\ is$ number of joints changed in posture i (most informative joints)

5 Conclusion and Future Work

Many researchers are working worldwide on human motion recognition, but as of now, no off the shelf technology is available for it. This is due to different applications shows success for different representation and recognition methods. Proposed kinematic state model for identification of Yogasana from video sequence has shown good recognition rate. Rare failure is due to nonrecognition of intermediate states of posture. Model can be applied for different exercises like aerobics and dances like bharatnatyam, where action is sequence of specific postures.

In future, we will try to make recognition rate 100% and system can be applied for other applications.

AKNOWLEDGEMENT

Authors are thankful to Dr. Ruta Kulkarni of 'Yogapoint' for her guidance and students of Pune Institute of Computer Technology for performing Yogasana to create dataset. We are also thankful to anonymous reviewers for their time for review and expert comments on work.

REFERENCES

[1] J.K. Aggarwal and M.S. Ryoo, "Human activity analysis: a review", ACM Computing Surveys, vol. 43, No. 3, Article 16, 2011, pp. 16:1–16:43

[2] T. B. Moeslund et al., “A survey of advances in vision-based human motion capture and analysis”, Computer Vision and Image Understanding 104, 2006, pp. 90–126.

[3] R. Poppe, ”Vision-based human motion analysis: An overview”, Computer Vision and Image Understanding 108, 2007, pp 4–18.

[4] P. Turaga et.al, “Machine Recognition of Human Activities: A Survey”, IEEE Trans. on Circuits and Systems for video technology, vol. 18, no. 11, Nov. 2008, pp. 1473-1488.

[5] J. Han et al., “Enhanced Computer Vision with Microsoft Kinect Sensor: A Review “ , IEEE trans. on Cybernatics, Vol. 43 , no. 5, 2013, pp 1318-1343.

[6] G. V. Kale and V. H. Patil, "Vision Based Human Motion Representation and Recognition Methods a Survey", Proceedings of Third Post Graduate Conference on “Computer Engineering“ cPGCON 2014, Elsevier, Vol X, pp XX - XX.

[7] S. Nowozin and J. Shotton, “Action points: A representation for low-latency online human action recognition,” Microsoft Research Cambridge, Tech. Rep. MSR-TR-2012-68, 2012.

[8] M. Yang et al., “Human Action Recognition Based on Kinect”, Journal of Computational Information Systems 10: 12, 2014, pp 5347–5354.

[9] X Yang and Y Tian, “Effective 3D action recognition using EigenJoints”, Journal of Visual Communication and Image Recognition, 25 2014, pp 2–11.

[10] J. Sung et al., "Human Activity Detection from RGBD Images", AAAI Workshop on Pattern, Activity and Intent Recognition, July 2011.

[11] Zan Gao et.al, “Human Action Recognition Via Multi-modality Information“, J Electr Eng Technol Vol. 9, No. 2, 2013, pp 739-748.

[12] Sudheer Deshpande, Nagendra H R, Raghuram Nagarathna, "A randomized control trial of the effect of yoga on Gunas (personality) and Health in normal healthy volunteers" International Journal of Yoga Vol. 1:1, Jan-Jun-2008, pages:2-10

[13] K. Williamsa et.al.,"Effect of Iyengar yoga therapy for chronic low back pain", International Association for the Study of Pain, 115 (2005) 107–117.

[14] A. Kauts and N. Sharma, "Effect of yoga on academic performance in relation to stress," International Journal of Yoga Vol. 2:1 Jan-Jun-2009 Pages:39-43.

[15] B. K. S. Iyengar, "The Illustrated Light on Yoga", HarperCollins Publisher India, Tenth impression, published in 2005.