Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 03, March -2018

Analysis of Human Behaviour by Motion Segmentation

Human Behaviour Analysis

Shavet Bhagat[1],Dr.Sukhwinder Singh[2]

[1]*Electronics and Communication Engineering,Punjab Engineering College(deemed to be university),Chandigarh*
Electronics and Communication Engineering,Punjab Engineering College(deemed to be university),Chandigarh

Abstract—*Motion segmentation is considered as an eminent process with the perspective of computer vision techniques. It has been observed from last decades that a large amount of research work has been done to tackle the issue, though, most of the previous work still less competitive behind human perception. In this paper, the concept of motion segmentation is studied. A generalize process for detecting the human behaviour on the basis of the motion from the video is also discussed. The study also provides a review to the existing work of this domain.*

Keywords— *Motion Segmentation, Motion Analysis, Human Behaviour, Motion Detection.*

I. INTRODUCTION

Several multimedia applicationsare based on visual analysis of human actions and behaviour for better understanding such as human-machine interfaces, medicine, video surveillance and active assisted living.These applications provided basis for evolution of techniques for interpreting human behaviourthrough 2D videos captured via RGB cameras [1– 5]. Certain issues like occlusions, background clutter, color sensitivity associated with these advance methods. The scope for research work over human-activities and behaviour understanding techniques are exposed to new directions with the development of RGB-D sensors.Extensive research has been carried out by examining the data obtained by such cameras so as to compare its efficiency to RGB cameras [6-10]. The task of subtracting background and detecting people in foreground through complete knowledge regarding 3D architecture is facilitated by depth data.The methodologies implemented for operating such depth datamakes it capable of working in varying light and darkness. These depth sensors when combined with robust algorithms for recognizing patterns facilitates the representation of any human pose in each frame in terms of set of 3D joints [11].Employing 3D data obtained from systems that can capture human motions for analysis purpose have served as a profound research area over past few decades [12-14].Despite being highlyaccurate and robust its high expense may hinder it application in certain fields.The subject is required to bear some markers physically in order to represent human pose is another major drawback of such systems which obstructs its use by general public. Such issues associated with these systems served as a topic of interest for exploring data obtained from RGB-D.The temporal variability involved in these systems tends to increase its complexity as numbers of motions are essentially combined for featuring human behaviour.The requirement of motion analysis to be invariant to geometric transformations makes the job more complex.The loss of data or introduction of certain unwanted noise is accompanied by obstructing few components while representing human pose due to human interaction and object manipulations that describes human behaviour.

II. PHASES OF HUMAN BEHAVIOUR DETECTION FROM VIDEO

Techniques for analyzing human actions and behaviourundergo different phases such as initialization, tracking, poses and recognition.Initializationphaseis necessarily required forinitializing the required system for data processing by constructing appropriate model for it.

1. Initialization Phase

For the purpose of recognizing human behaviour and actions the initialization phase requires knowledge regarding composition of an individual as primary information. It needs approximation of shape, kinematic structure, beginning posture of the subject which is targeted for tracking. This knowledge required for accomplishing initialization is classified as kinematic structure, 3D shape, color appearance and body part approximation. Volumetric construction technique has been proposed in various proposed research works for initializing kinematic structure obtained from multiple-view image sequences. Automatic initialization of model appearance to a particular individual can be performed through such techniques.

2. Tracking Phase

Tracking phase involves two major processes for visual recognition of human behaviour. First, ground segmentation involves distinguishing the subject from its background in frame. The other process is called as temporal correspondence performs recognition of similar subjects from frame sequences.

3. Pose Estimation Phase

This phase deals with understanding and calculation of underlying kinematic structure of an individual. On the basis of usage of human model for estimating human pose, the algorithm employed can be categorized as Model free, Direct model use and Indirect model use. Model free does not use any particular model and only includes mapping of 2D sequences into 3D pose. Employing a specific model for within pose estimation is done in indirect model approach. Kinematic structure model involves the usage of explicit 3D geometric representation of human shape and structure to estimate pose.

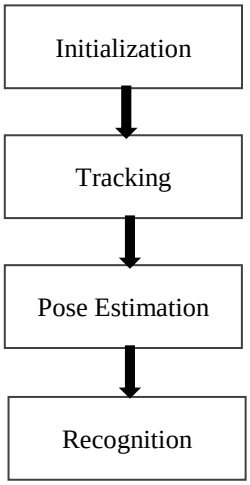

Figure 1 Different phases of human behaviour detection from video

4. Recognition

Various applicationssuch as surveillance, medical studies, robotics etc. involvesrecognition methodologies. Recognition is performed in different hierarchical levels like scene interpretation involves interpreting entire image with no distinction of objects. All human body parts are recognized in holistic recognition whereas action primitives involve recognition of actions is performed for interpreting its semantic depictions.

III. RELATED WORK

In recent years, recognition and understanding of human behaviour by analyzing depth data have attracted the interest of several research groups [15–18]. While some methods focus on the analysis of human motion in order to recognize human gestures or actions, other approaches try to model more complex behaviours (activities) including object interaction. These solutions focus on the analysis of short sequences, where one single behaviour is performed along the sequence.

However, additional challenges appear when several different behaviours are executed one after another over along sequence. In order to face these challenges, methods based on online detection have been proposed. Such methods can recognize behaviour before the end of their execution by analyzing short parts of the observed sequence.

The identification of number of behaviours within a single long sequence can be achieved via such schemes.The same thing is not involved in the methods that deal with direct analysis of whole sequence.The review of existing human behaviour analysis techniques that involved use depth data is given in the below section:

Techniques that are available for the purpose of gesture action recognition from RGB-D sensors are classified as skeleton-based, hybrid and depth map-based approaches.Shotton et al. [11] introduced the skeleton-based approach which gained huge popularity.No temporal information is required by skeleton-based techniques for accurately estimating the 3D pose of body joints in each depth maps. Yang and Tianin [19] proposed a technique which recognized human activitieson the basis of three characteristics of each joint obtained from pair wise differences of individual jointpositions that include initial, previous and present frames.A compact Eigen Joints representation of individual frame is performed by PCA which is followed by multi-class action classification by Naive-Bayes nearest-neighbor classifier.Luo et al.[20] proposed similar method which involved evaluation of pairwise differences just for current frame with respect to single reference point referred as hip point.The representation of such features was facilitated by dictionary learning method which was entirely based on geometry constrictions and group sparsity.Under SVM the sequences are categorized into distinct categories.Zanfir et al.[15] introduced the moving Pose feature which required capturing of human pose along with the speed and acceleration associated with body joints that lies within a small temporal window. The process of action recognition is accomplished by enhanced version of kNN classifier. Hong zhao et al. in [21] involved tracing of most prominent body parts in each action sequence via a part-based feature vector.Skeleton data can be represented by employing differential geometry.Vermulapalli and Chellappain [22] involved a single element for representing each skeleton on the Lie-groupwith corresponding curve on manifold.In[23], Slama et al. involved representing skeleton time series in terms of point on Grassman manifold. The Riemannian geometry of same manifold was used for classification purpose. In [24],Anirudh et al. employed Transport Square-Root Velocity Function for recognizing human actions for analyzing these actions which were considered as trajectories on Riemannian manifold.The whole set of depth image points were utilized for extracting significant descriptors. In [25], Yang et al. introduced a technique based on action dynamics which performed highlighting of regions where motion was traced by employing Depth Motion Maps. Depth Cuboid Similarity Feature [16], Random Occupancy Pattern [27], Spatio-Temporal Occupancy Pattern[26] introduced several other methodologies performed the feature extraction by decomposing 4D-space into spatio-temporal boxes.These extracted features were responsible for representing depth appearance in individual box that were from Spatio-Temporal Interest Points.In [28], Rahmani et al. presented a method for detecting key points. The Histogram of Principal Components was employed for defining point cloud within a range of volume.In [29], Oreifej and Liu employed polychoron vertices for quantizing the 4D space followed by distributing of

normal vectors for individual cells.For the purpose of featuring human action surface normal were utilized for defining both local motion and shape information.Althloothi et al. [31] introduced spherical harmonics representation along with 3D motion features for representing 3D features via kinematic architecture from skeleton.Multikernel learning based scheme combined both characteristics.In [32], Lu and Tang introduced Range-Sample which was a depth feature used for defining motion and shape geometry.

Interpreting the human interaction with its environment is much more difficult task which may not be addressed by simple understanding of human motion.Several hybrid solutionsoffered the use of depth maps for achieving modeling of scene objects and body skeleton for performing modeling of human motion.

In [33], Wang et al. presented the technique based on Local Occupancy Patterns for representing analyzed depth values corresponding to skeleton joints.Several other methods described human interaction with objects based on model spatio-temporal via Markov Random Field [17].

In [34], Wei et al. presented a graphical model which provided description regarding human actions by representing them in hierarchical structure which consisted of details pertaining to objects, human pose and their interactions.In [35], YuandLiu proposed use of middle level representation referred as orderlet for capturing significant depth features and skeleton. Due to short sliding window along a sequence the reviewed works mentioned earlier were having online action recognition abilities. The task of analyzing continuous depth sequences that involved numerous actions performed in successive manner was carried out by employing similar approach. Like in [18], Huang et al. employed Sequential Max Margin Event Detector algorithm for long sequences that consisted of numerous actions on line by successive discarding non-corresponding action classes.

IV. CONCLUSION AND FUTURE SCOPE

To sum up the concept, we can state that this study is a kind of evidence that proves the idea of how extensive the motion detection literature is and also authenticate that the research work in this field is still going on to resolve the certain issues. On the basis of the previous work that is discussed in this work, it can be said that the traditional motion detection mechanism lacks at various points like the suddenly halt of moving cameras, low quality of the video, etc.

Therefore, in future more research can be conducted by using prominent schemes in this field to overcome the pitfalls of traditional mechanisms.

REFERENCES

[1] Tian, L.Cao, Z.Liu, Z.Zhang, "Hierarchical filtered motion for action recognition in crowded videos", IEEE Trans. Syst. Man Cybern. PartC: Appl. Rev., Vol.42, Pp 313–323, 2012.

[2] B. Solmaz, B.E.Moore, M.Shah, "Identifying behaviours in crowd scenes using stability analysis for dynamical systems", IEEE Trans. Pattern Anal. Mach. Intell. 34, Pp 2064–2070, 2012.

[3] W.Ge, R.T.Collins, R.B.Ruback, "Vision-based analysis of small groups in pedestrian crowds", IEEE Trans. Pattern Anal. Mach. Intell. 34, Pp 1003–1016, 2012.

[4] O.Arandjelović, "Contextually learnt detection of unusual motion-based behaviour in crowded public spaces", International Symposium on Computer and Information Sciences, Pp.403–410, 2011.

[5] N. Buch, S.A.Velastin, J.Orwell, "A review of computer vision techniques for the analysis of urban traffic", IEEE Trans. Intell. Transp. Syst. 12, Pp. 920–939, 2011.

[6] S.Hadfield, R.Bowden, "Kinecting the dots: particle based scene flow from depth sensors", International Conference on Computer Vision (ICCV), Pp. 2290–2295, 2011.

[7] Z.Ren, J.Yuan, Z.Zhang, "Robust hand gesture recognition based on finger- earth mover's distance with a commodity depth camera", ACM International Conference on Multimedia, Pp.1093–1096, 2011.

[8] K.A.FunesMora, J.-M.Odobez, "Person independent 3D gaze estimation from remote RGB-D cameras", IEEE International Conference on Image Processing, Pp. 2787–2791, 2013.

[9] R.S.Ghiass, O.Arandjelović, D.Laurendeau, "Highly accurate and fully automatic head pose estimation from a low quality consumer-level RGB-D sensor", Workshop on Computational Models of Social Interactions: Human–Computer–Media Communication, Pp. 25–34, 2015.

[10] L. Sun, Z.Liu, M.-T.Sun, "Real time gaze estimation with a consumer depth camera", Inf. Sci.320, Pp. 346–360, 2015.

[11] J.Shotton, A.Fitzgibbon, M.Cook, T.Sharp, M.Finocchio, R.Moore, A.Kipman, A.Blake, "Real-time human pose recognition in parts from single depth images", IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Pp. 1–8, 2011.

[12] J.Barbic, A.Safonova, J.-Y.Pan, C.Faloutsos, J.K.Hodgins, N.S.Pollard, "Segmenting motion capture data into distinct behaviours", Graphics Interface, 2004.

[13] F.Zhou, F.Dela Torre, J.K.Hodgins, "Hierarchical aligned cluster analysis for temporal clustering of human motion", IEEE Trans. Pattern Anal. Mach. Intell. 35, Pp. 582–596, 2014.

[14] I. Kapsouras, N.Nikolaidis, "Action recognition on motion capture data using a dynemes and forward difference representation", J. Vis. Commun. Image Represent, Vol. 2, Pp. 1432–1445, 2014.

[15] M.Zanfir, M.Leordeanu, C.Sminchisescu, "The moving pose: an efficient 3D kinematics descriptor for low-latency action recognition and detection", IEEE International Conference on Computer Vision (ICCV), Pp. 2752–2759, 2013.
[16] L.Xia, J.K.Aggarwal, "Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera", CVPR Workshop on Human Activity Understanding from 3D Data, Pp. 2834–2841, 2013.
[17] H.S. Koppula, R.Gupta, A.Saxena, "Learning human activities and object affordances from RGB-Dvideos", Int. J. Robot. Res. 32, Pp. 951–970, 2013.
[18] D.Huang, Y.Wang, S.Yao, F.D.LaTorre, "Sequential max-margin event detectors", European Conference on Computer Vision (ECCV), Pp. 410–424, 2014.
[19] X.Yang, Y.Tian, "Eigen joints-based action recognition using naive- Bayes- nearest- neighbor", Workshop on Human Activity Understanding from 3D Data, Pp. 14–19, 2014.
[20] J.Luo, W.Wang, H.Qi, "Group sparsity and geometry constrained dictionary learning for action recognition from depth maps", IEEE International Conference on Computer Vision (ICCV), Pp.1809–1816, 2013.
[21] Hong Zhao. Chen,G. Wang, J.-H.Xue, L.He, "A novel hierarchical frame work for human action recognition", Pattern Recognit, Vol. 55, Pp. 148–159, 2016.
[22] R.Vemulapalli, F.Arrate, R.Chellappa, "Human action recognition by representing 3D skeletons as points in a Lie group", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Pp. 588–595, 2014.
[23] R. Slama, H.Wannous, M.Daoudi, A. Srivastava, "Accurate 3D action recognition using learning on the Grassmann manifold", PatternRecognit, Vol. 48, Pp. 556–567, 2017.
[24] R.Anirudh, P.Turaga, J.Su, A.Srivastava, "Elastic functional coding of human actions: from vector-fields to latent variables", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Pp. 3147–3155, 2015.
[25] X.Yang, C.Zhang, Y.Tian, "Recognizing actions using depth motion maps-based histograms of oriented gradients", ACM International Conference on Multimedia, Pp.1057–1060, 2012.
[26] A.W.Vieira, E.R.Nascimento, G.L.Oliveira, Z.Liu, M.F.Campos, " improvement of human action recognition from depth map sequences using space–time occupancy patterns", Pattern Recognit. Lett, Vol. 36, Pp. 221–227, 2014.
[27] J.Wang, Z.Liu, J.Chorowski, Z.Chen, Y.Wu, "Robust 3D action recognition with random occupancy patterns", European Conference on Computer Vision (ECCV), Pp.1–8, 2014.
[28] H.Rahmani, A.Mahmood, D.Q.Huynh, A.Mian, "Hopc: histogram of oriented principal components of 3D point clouds for action recognition", European Conference on Computer Vision (ECCV), Pp.742–757, 2014.
[29] O.Oreifej, Z.Liu, "HON 4D: histogram of oriented 4D normals for activity recognition from depth sequences", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Pp.716–723, 2013.
[30] X.Yang, Y.L.Tian, "Super normal vector for activity recognition using depth sequences", IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Pp.804–811, 2014.
[31] S. Althloothi, M.H.Mahoor, X.Zhang, R.M.Voyles, "Human activity recognition using multi-features and multiple kernel learning", Pattern Recognit, Vol. 47 Pp. 1800–1812, 2014.
[32] C.Lu, J.Jia, C.-K.Tang, "Range-sample depth feature for action recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Pp.772–779, 2014.
[33] J.Wang, Z.Liu, Y.Wu, J.Yuan, "Mining action let ensemble for action recognition with depth cameras", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Pp.1–8, 2012.
[34] P.Wei, Y.Zhao, N.Zheng, S.-C.Zhu, "Modeling 4d human–object interactions for event and object recognition", International Conference on Computer Vision (ICCV), Pp. 3272–3279, 2013.
[35] G.Yuand, Z.Liu, J.Yuan, "Discriminative order let mining for real-time recognition of human–object interaction", Asian Conference on Computer Vision (ACCV), Pp.50–65, 2014.

