www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 3 Issue 7 July, 2014 Page No. 7199-7202

Human Action Recognition Using Kinect

Niranjan Deokule[1], Geetanjali Kale[2]

[1]Department of Computer Engineering,
Pune Institute of Computer Technology
Pune, India,
niranjan.deokule@gmail.com

[2]Department of Computer Engineering,
Pune Institute of Computer Technology
Pune, India,
gilkale@gmail.com

Abstract: ***In this paper, a general approach for human action recognition is applied for classifying human movements into action classes. The propose method uses Kinect for capturing depth stream. The system performs preprocessing on depth information for reducing noisy pixels and getting depth information in appropriate format. The background subtraction method is used for extracting region of interest i.e. human. The system extracts contours of person. The Hu moments are extracted from contours of person for training action classifier. The Support Vector Machine (SVM) is used for classifying human activities.***

Keywords: Contours, Hu moments, Support Vector Machine.

1. Introduction

The human action recognition has gain much importance in computer-vision field such as visual surveillance, video retrieval, sports video analysis, human- computer interactions, Sign Language Recognition (Gesture recognition). The human action recognition is a challenging task due to large variations in people appearance, clothing, and motion performance. The recognizing of human actions is considered an important problem in the field of computer vision. The motivation behind choosing this topic is the large number of potential applications of human action recognition in areas of visual surveillance, video retrieval, sports video analysis, human- computer interactions.

The human action recognition is the process of recognizing the human movements or activities from images and videos based on the features extracted from the human body models. In this paper, we propose a human action recognition system that captures depth frames from Kinect camera. The depth information contains set of pixels indicating the distance of objects from camera in millimeters.

2. Literature Survey

In [1] Aggarwal and Cai, presented a review on human motion analysis which describes methods for body structure analysis, motion tracking and action recognition. In [2], an overview on markerless vision-based human motion analysis has been presented which focuses on human body modeling, pose estimation, and model-free approaches for human motion analysis. The paper [3] describes a comprehensive review of recent Kinect based computer vision algorithms and applications. This paper focuses on preprocessing, object tracking and recognition, human activity analysis, hand gesture recognition, indoor 3D mapping. In [4], a survey on different techniques and technologies used by Kinect and its applications in research and industry is presented. The brief description about the key components of Kinect and working principle of Kinect is presented in paper [5]. In [6], 2Dchamfer distance matching is used for human detection and contour based region growing algorithm is used for tracking the people in the frames. The motion capture and reconstruction is done in [7] by using depth information. They have converted depth data into human skeleton data i.e. Biovision Hierarchy (BVH) file. They have constructed virtual models of human bodies by using MakeHuman software.

The natural posture space is constructed by using K-nearest neighbor algorithm for accurately estimating the correct posture of the user in [10]. In [11], they have used body part labeling method for human posture tracking and Randomized Decision Forests for recognizing the human posture. A novel approach for human detection by using Kinect is presented in [12]. They have used pixel filtering and context filtering for preprocessing depth map and Support Vector Machine (SVM) for classification. Shotton, Cook, Sharp, have presented an algorithm for real time pose recognition in parts from single depth images in [13]. This algorithm works as a basic principle for working of Kinect and it is used in training the classifier for Kinect sensor. A randomized decision forest is used for training the classifier. The paper [14] focuses on a model free segmentation and tracking of objects by using Support Vector Machine (SVM). They have used rich feature set of optical flow, depth discontinuities, local image appearance, and surface normals. The paper [15] focuses on mutli-scenario

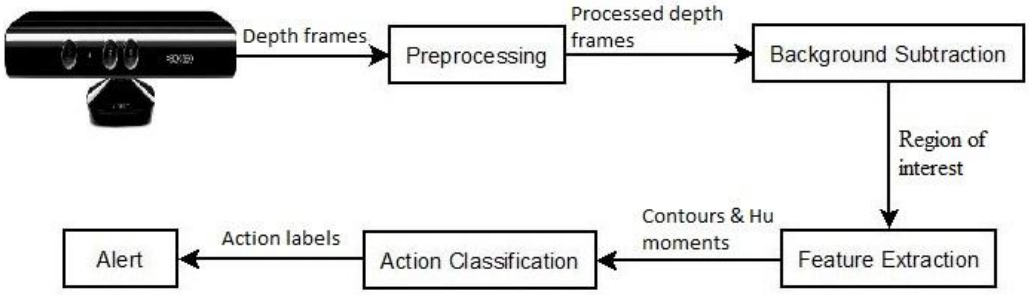

Figure 1: System Block Diagram

gesture recognition using Kinect. They have used contour tracking algorithm for hand detection and for gesture recognition, three layers of classifiers, finger counting, finger name collection, vector matching are used.

3. Proposed Method

This system takes input as depth frames from Kinect and classifies the human movements by using Support Vector Machine (SVM).
The proposed system has five stages.

1. Preprocessing
2. Background Subtraction
3. Feature Extraction
4. Action Classification

The system block diagram is shown in figure.1

3.1 Preprocessing

In [9], for preprocessing stage, they have used Z-score normalization on body joint positions and input vectors of twenty body joints positions for motion tracking. They have used back propagation neural network, support vector machine, decision tree, naïve bays for gesture recognition. This system takes input as depth frames captured from Kinect camera. The depth information contains information of distance between camera and objects. This depth data has some noisy pixels and if we use this depth data for further processes, then we will not get correct features. The system performs preprocessing by setting noisy pixels value as per the neighboring pixels.

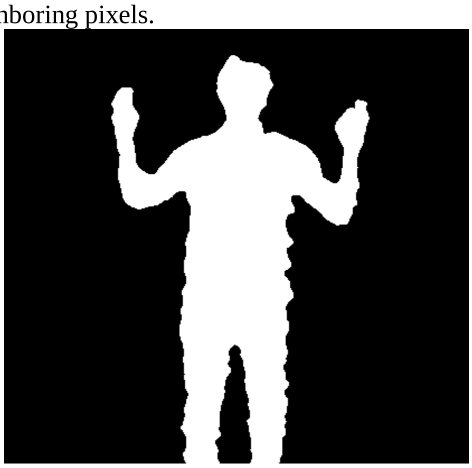

Figure 2: Region of interest (Human).

3.2 Background Subtraction

The background subtraction is common and widely used method for getting foreground mask from the images. The background subtraction method is used for getting region of interest i.e. person by subtracting current frames from the previous frames. Figure 2. Shows region of interest as human extracted from depth frames.

3.3 Feature Extraction

In [8], Hidden Markov model (HMM) is used for hand gesture recognition. They have used position coordinates of palm node, velocity of palm node as a feature for gesture analysis and recognition. The system performs feature extraction for extracting features from depth frames with region of interest. We have extracted contours and Hu moments from the depth frames. First contours are extracted from the depth frames, and then Hu moments are extracted from the contours of person. Figure 3. Shows contours extracted from region of interest.

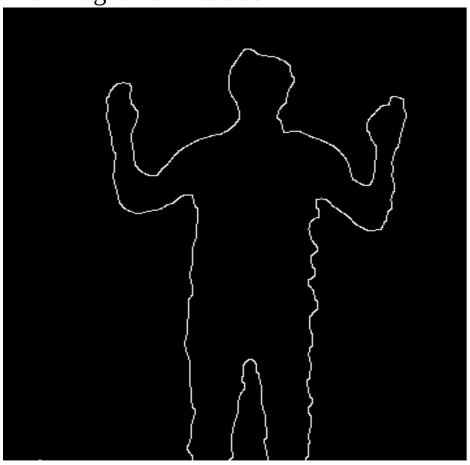

Figure 3: Contour extracted from region of interest

3.4 Action Classification

The action recognition stage is the final and most difficult stage in the system. This stage classifies the human movements as per the training features file. The extracted Hu moments are used for training the Support Vector Machine (SVM). Once the SVM model is trained, then it is used for classifying the human movements.

4. Mathematical Model

Let S be the system that describes method to detect suspicious activity by using depth information.

$S= \{I, O, F, Su, Fa, C \mid \phi_s\}$

Where,

I is the input to the system.

O is output of the system.

F is set of functions.

Su is success of system.

Fa is failure of system.

C is constraints of system.

- **Input:**

I is the input set such that I= {d1, d2,…,dn}- set of depth frames captured from Kinect.

- **Output:**

O=Notification or alert on detection of suspicious activity.

- **Functions:**

F is a set of functions where,

F = {f1, f2, f3, f4, f5}

1. f1= preprocessing();

f1 (d_i) = {pd_i}- set of processed depth frames.

This function takes depth frames captured by Kinect as input and provides the processed depth frames in order to minimize the noise in the depth pixels.

2. f2 = background_subtraction();

f2(pd_i) = {rd_i} – set of depth frames with region of interest as human.

The input to f2 function is set of processed depth frames from the f1 function. This function extracts region of interest as human from processed depth frames.

3. f3 = contour_extraction();

f3(rd_i) = $u_i(x,y)$ – set of contours.

The input to f3 function is set of depth frames with region of interest as human from the f2 function. This function extracts contours of human body in order to track the human movements in the next stage.

4. f4= moments_extraction();

f4(u_i) = {x_1, x_2, x_3, x_4, x_5, x_6, x_7} – set of Hu moments.

This function takes input as a set of contours from the f3 function and extracts Hu moments from contours.

5. f5= activity_recognition();

f5(X_i) = { ac_i} – action labels.

This function classifies the actions given by the f4 function into action labels.

- **Success:**

Su= {Generate notification or alert on detection of suspicious behavior}.

- **Failure:**

Fa= {fac,fr}.

fac =False Acceptance.

fr = False Rejection.

- **Constraints:**

1. Object should be in the Kinect range. (0.8meters - 5 meters).
2. Lighting conditions should be kept minimum.

5. Conclusion

In this paper, we proposed a novel approach for human action recognition by using depth information captured from Kinect. We have used contour and Hu moments extracted from contours as features for training action classifier. The Support Vector Machine (SVM) is used for classifying human movements into action classes.

References

[1] Aggarwal, J.K.; Cai, Q., "Human motion analysis: a review," Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE , vol., no., pp.90,102, 16 Jun 1997

[2] Poppe, Ronald. "Vision-based human motion analysis: An overview." Computer vision and image understanding 108.1 (2007): 4-18.

[3] Jungong Han; Ling Shao; Dong Xu; Shotton, J., "Enhanced Computer Vision With Microsoft Kinect Sensor: A Review," Cybernetics, IEEE Transactions on , vol.43, no.5, pp.1318,1334, Oct. 2013

[4] Cruz, L.; Lucio, D.; Velho, L., "Kinect and RGBD Images: Challenges and Applications," Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2012 25th SIBGRAPI Conference on , vol., no., pp.36,49, 22-25 Aug. 2012

[5] Zhang, Zhengyou. "Microsoft kinect sensor and its effect." Multimedia, IEEE 19.2 (2012): 4-10

[6] Lu Xia; Chia-Chih Chen; Aggarwal, J.K., "Human detection using depth information by Kinect," Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on , vol., no., pp.15,22, 20-25 June 2011

[7] Ming Zeng; Zhengcun Liu; Qinghao Meng; Zhengbiao Bai; Haiyan Jia, "Motion capture and reconstruction based on depth information using Kinect," Image and Signal Processing (CISP), 2012 5th International Congress on , vol., no., pp.1381,1385, 16-18 Oct. 2012

[8] Youwen Wang; Cheng Yang; Xiaoyu Wu; Shengmiao Xu; Hui Li, "Kinect Based Dynamic Hand Gesture Recognition Algorithm Research," Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on , vol.1, no., pp.274,279, 26-27 Aug. 2012

[9] Patsadu, O.; Nukoolkit, C.; Watanapa, B., "Human gesture recognition using Kinect camera," Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on , vol., no., pp.28,32, May 30 2012-June 1 2012

[10] Shum, H.P.H.; Ho, E.S.L.; Jiang, Y.; Takagi, S., "Real-Time Posture Reconstruction for Microsoft Kinect," Cybernetics, IEEE Transactions on , vol.43, no.5, pp.1357,1369, Oct. 2013

[11] Zheng Xiao; Fu Mengyin; Yang Yi; Lv Ningyi, "3D Human Postures Recognition Using Kinect," Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on , vol.1, no., pp.344,347, 26-27 Aug. 2012

[12] Yujie Shen; Zhonghua Hao; Pengfei Wang; Shiwei Ma; Wanquan Liu, "A Novel Human Detection Approach Based on Depth Map via Kinect," Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on , vol., no., pp.535,541, 23-28 June 2013s

[13] Shotton, J.; Fitzgibbon, A.; Cook, M.; Sharp, T.; Finocchio, M.; Moore, R.; Kipman, A.; Blake, A., "Real-time human pose recognition in parts from single depth images," Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on , vol., no., pp.1297,1304, 20-25 June 2013

[14] Teichman, A.; Lussier, J.T.; Thrun, S., "Learning to Segment and Track in RGBD," Automation Science and Engineering, IEEE Transactions on , vol.10, no.4, pp.841,852, Oct. 2013

[15] Yi Li, "Multi-scenario gesture recognition using Kinect," Computer Games (CGAMES), 2012 17th International Conference on , vol., no., pp.126,130, July 30 2012-Aug. 1 2012

[16] Microsoft Inc. Kinect for Windows: Develop What's Next, http://www.microsoft.com/enus/kinectforwindows/develop/, 2012.