International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

Study of Different Action Recognition Techniques

Nikita Pachpute[1], Priya Kamate[2], Gayatri Walmik[3], Shivani Jadhav[4]

IT Department, MIT College of Engineering, Pune , Maharashtra, India

---

Abstract - *In this paper we propose three approaches for human action recognition from input video stream:1) action recognition from silhouette images and 2) action recognition by using hidden markov model(HMM). 3) action recognition by using depth map resources (HDMM). First approach makes use of spatio-temporal body parts movement (STBPM) for extracting spatial and temporal features from silhouette image and RAC classifier for classifying human action based on extracted features. Second approach makes use of hidden markov model (HMM) which represents process as a set of states. Third approach use rotation and class score fusion.*

Key Words: Silhouette image, STBPM, RAC, HMM,HDMM, Activity recognition

1.INTRODUCTION

Now-a-days Human action recognition plays very important role in various applications such as human–compute interaction, human activities analysis, and real time surveillance systems. In daily life human performs different activities which are composed of several subtasks. These day to day activities include walking, running, dancing etc. These activities may be performed at different times and may have chronological relationship with each other. Motive for writing this paper is to recognize these activities and analyze human behavior. Recognizing human action is now becoming the topic of interest in the field of video surveillance. This paper describes two different approaches for action recognition. The contribution of this work consists of two parts. Firstly we extract the features of skeleton and then we classify the activity based on those features.

Three main complexity issues, as mentioned in [1], are generally present in any Human Action Recognition technique are - i) Environmental complexity: The process of Human Action Recognition depends on the quality of the video, which differs due to the environmental condition of the scene elements and makes the procedure more complex. This type of complexity includes occlusions, clutter, interaction among multiple objects, changing of illuminations etc. ii) Acquisition complexity: Besides environmental condition, the quality of the video also depends on video acquisition, which varies with respect to view point, movement of the camera etc. iii) Human action complexity: In general sense, human actions are of varied in nature, hence exact determination of human action is a complicated task. Presence of multiple human entities makes any Human Action Recognition technique more complex.

To handle those three issues mentioned above, some constrains have been made in [2].this constraints are like as i) use the videos, which contain human silhouettes only and didn't consider any silhouette extraction techniques also. Video So, the environmental and acquisition complicacies are avoided. ii) To reduce the complexity, the proposed work considers videos containing only one human object in each of the frame. iii)consider that the head should be in the upper portion of the body and the body should not be upside down.

2.RELATED WORK

Computer-vision-based human motion analysis has become an active research area. It is strongly driven by many promising applications such as smart surveillance, virtual reality, advanced user interface, etc. Recent technical developments have strongly demonstrated that visual systems can successfully deal with complex human movements. Human Action Recognition task mainly classified in three phases according to [5] as follows i)motion segmentation ii)tracking iii) Behavior understanding. In motion segmentation the image is divided into regions as foreground and background. Foreground image should be an object such as person, cars, animal etc Motion segmentation separates foreground images from background images. Object tracking is comes after the motion segmentation. Tracking is a particularly important issue in human motion analysis since it serves as a means to prepare data for pose estimation and action recognition. By using Tracking we can estimate the pose of the object and recognize the action of object. This can be done on the basis of points, lines, blobs etc. After the tracking we go for the Behavior understanding. Behavior understanding contains action recognition and description.

3.ACTION RECOGNITION FROM SILHOUETTE IMAGES

1) Silhouette Image: Silhouette image is the image of a person, animal, object or scene usually represented as a solid shape of single color. Silhouette image has only two color levels and it provides shape based information. Interior of silhouette image is featureless.

Fig 1. Example of silhouette images

For action recognition purpose silhouette image is divided into three horizontal levels and two vertical levels viz: Horizontal levels are categorized as: 1. Upper Level (UL): It includes top of the body part. Upper level contains body parts such as head and hand (if someone lifts his/her hands) 2. Middle Level (ML): It contains the middle part of body and hands. 3. Lower Level (LL): It contains lower region of body including legs.

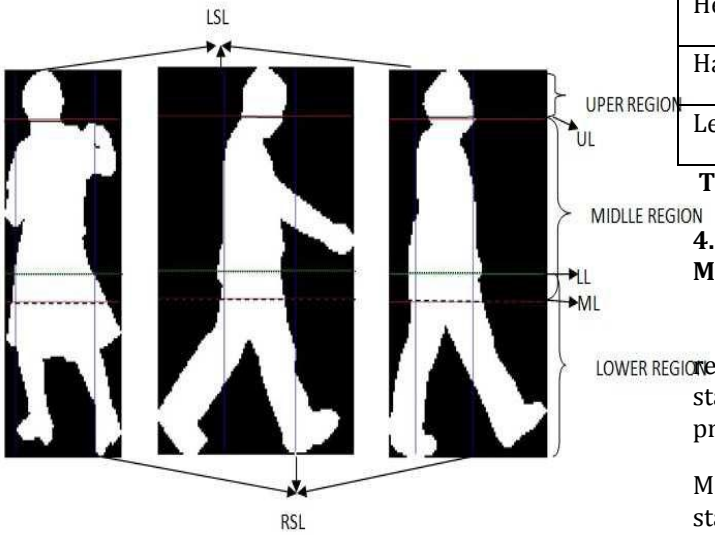

Fig 2. Action units of human body parts

Action recognition can be performed in two major steps: Foreground extraction and Feature extraction and action classification. A. Foreground extraction- This process is also called as background subtraction. In this process background is eliminated from input video. Generally for action recognition purpose information of background is not required so for reducing computational overhead background needs to be subtracted and only foreground image needs to be extracted. This process gives silhouette image as output. B. Feature extraction and action classification- Human body parts which perform action are called as action units (AU). For feature extraction movements of action units are analyzed. This can be done by comparing two adjacent frames of input video.

2) Spatio-Temporal Body Parts Movements (STBPM) - This is the feature vector with dimensions $k \times n$ where k represents total number of frames to be analyzed and n represents number of features to be extracted. STBPM holds different states of action units. It gives location of 1) HEAD 2) HAND 3) LEG and 4) BMC (Body Mass Centre) .STBPM also stores information about: I.HA (Head Angle): It is the angle between head and shoulder. II. SA (Stride Angle): It is the angle between two legs III. MD (Moving Direction): It is the direction of movement. IV. BR (Body Ratio): It is the ratio of width and height of the body. V. BB (Bounding Box): It is the imaginary box constructed around silhouette image.

3) Rule Action Classifier (RAC): RAC does the job of recognizing actions based on extracted features. It uses series of condition rules to categorize actions. Action units and their action region play a very important role in deciding action. Activity of any action unit is described by its action region i.e. UL, ML, LL etc.

Acting Components	Upper Level	Middle Level	Lower Level
Head	Yes	No	No
Hands	No	Yes	No
Legs	No	No	Yes

Table 1. Action Components and Action Units for Walking

4. ACTION RECOGNITION BY USING HIDDEN MARKOV MODEL (HMM)

HMM is the state model based approach for action recognition in which human activities are described as a statistical model with a set of states. HMM assumes that the process to be analyzed is a markov process.

Markov process: Markov process is a process whose next state does not depend on its previous state but it depends on its current state. Markov process is also called as stochastic process. States of Markov process are further categorized as 1) Observed state and 2) Unobserved state.

Unobserved state: Unobserved state is a state which is not visible directly but output depends on this state.

Elements of HMM:

$\lambda = \{N, M, A, B, \pi\}$

N=Number of states of Markov process

M=Amount of discrete output o/p operations

A=Transition probability matrix

B=Emission probability of o/p symbol per state

Π=Probability of starting

In HMM decision of making transition from one state to another state is taken on the basis of transition probability matrix A.

Skeleton Features: Skeleton is represented by set of Euclidean distances between joints. Upper body is represented as set of Euclidean distances between 1) left hand – head 2) left hand – left shoulder 3) left hand – left heap 4) left elbow – torso 5) right hand – head 6) right hand – right shoulder 7) right hand – right heap 4) right elbow – torso.

For describing pose of a skeleton/human bodies following set of Euclidean distances are used: 1) Left hand – head 2) Left hand – neck 3) Left hand – left shoulder4) Left elbow – torso 5) Left foot – heap 6) Left foot – neck 7) Left food – head 8) Left knee – torso 9) Right hand – head 10) Right hand – neck 11) Right hand – left shoulder 12) Right elbow – torso 13) Right foot – heap 14) Right foot – neck 15) Right food – head 16) Right- knee – torso

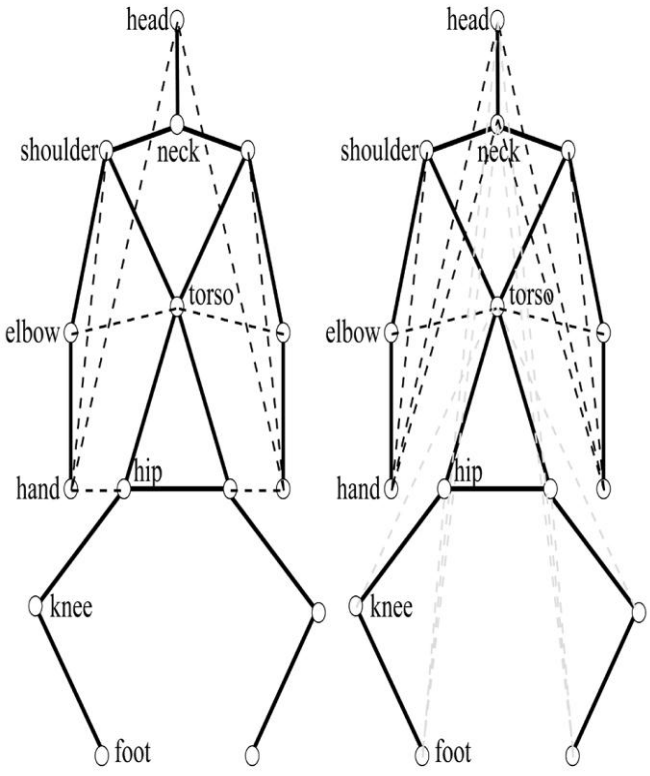

Fig 3. Skeleton Features for Pose Description

Observations for HMM are computed by calculating difference between features of 2 adjacent frames. Based on observations activities are further categorized as periodic activities and non periodic activities.

Periodic activities: Activities which repeat themselves after a particular period of time. For example, Walking.

Non-periodic activities: Activities which may or may not repeat themselves after a particular period of time. For example, stand up, sit down.

For processing activities motion data is segmented into following 3 sections:

Anticipation: Set of motion data when motion is about to start

Action: Motion data contains the poses when describing motion

Reaction: Set of motion data which describes recovery from an action to neutral position.

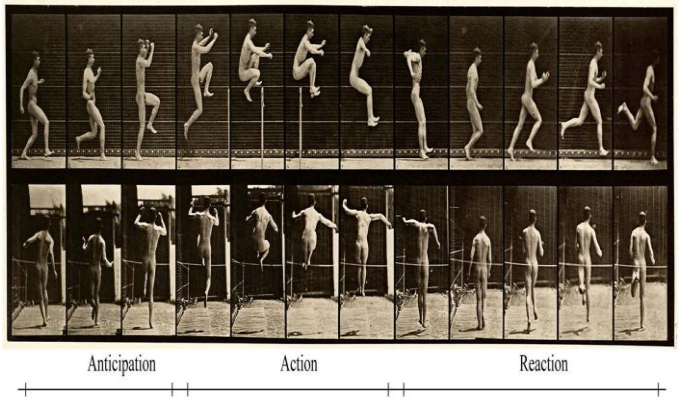

Fig 4. Segmentation of Motion Data

This segmented motion data is then passed as an input to the recognition model. And the activity which matches more accurately with the motion data is taken as the result of the action recognition process.

5. ACTION RECOGNITION BY USING DEPTH MAP RESOURCES

1) Depth map: Depth map is an image formed of gray pixels defined by 0-255 values. It captures 3D structure and shape related information of an image. It plays an important role in extracting spatio-temporal features of an image.

Action recognition by using depth maps can be done in following 3 steps:

Step 1: we propose a new architecture of HDMM and convolution networks.

Step 2: we propose architecture for recognizing view variant actions.

Step 3: we generate a large dataset by combining current features to evaluate the stability of proposed method.

2) Hierarchical Depth Motion Maps (HDMM): HDMM is an architecture which does image classification and artificial enlarging of training data for action recognition purpose.

For extracting motion features of an object difference between depth maps of consecutive frames is projected on three orthogonal Cartesian planes indicating front view, top view, and side view of an object. For recognizing actions with respect to viewpoint variations we rotate the depth data according to the view angle. We generate 3 $HDMM_s$ and 3 convolution networks for 3 different planes.

3) Rotation: Rotation needs to be performed on depth data for recognizing view invariant actions.

Co-ordinates of rotated objects can be written as $R=R_yR_z[X, Y, Z, 1]^T$ where R_y=Rotation factor for rotation in y direction and R_z=Rotation factor in z direction. Here we consider that if an object wants to move from point p0-pd in a plane, first it moves from p0-p1 through an angle Θ and then it moves from p1-pd through an angle β. While generating HDMM from depth data we generate 3 HDMMs and 3 convolution networks for 3 different planes. We project the 3D depth frame on 3 orthogonal Cartesian planes including front view, top view, and side view. Projected frame is denoted by map_p={f, s, t} . In order to retain subtle information we calculate difference between motion energy of sub-sampled frames. Motion energy is stacked across the entire depth sequence. After constructing motion maps we construct RGB image from motion maps.

4) Class score fusion: Given an input depth sequence for testing we consider the motion maps with temporal scaling without any rotation. The averaged scores of n scales for each test sample are calculated as the final score of this test sample in one channel of 3ConvNets. The final class scores for a test sample are the averages of the outputs of the three ConvNets.

6. CONCLUSIONS

In this paper we have studied two different approaches for action recognition. First approach makes use of silhouette images, STBPM, and RAC. Second approach makes use of skeleton features and HMM for recognizing actions.

ACKNOWLEDGEMENT

We feel fortunate enough to get this opportunity which will definitely strengthen our personality in every way. I am profoundly grateful to Prof. Aditi Jahagirdar for her expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

We take this momentous opportunity to express our heartfelt gratitude, ineptness and regards for her invaluable advice and wholehearted cooperation without which this project would not have seen the light of day.

REFERENCES

[1] Mishra, Miss Sapana K., Faizpur JTMCOE, and K. S. Bhagat. "*A Survey on Human Motion Detection and Surveillance.*" International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 4, April 2015 .

[2] Maity, Satyabrata, Debotosh Bhattacharjee, and Amlan Chakrabarti. "A Novel Approach for Human Action Recognition from Silhouette Images" arXiv preprint arXiv:1510.04437 (2015).

[3] Figueroa- Angulo, Jose Israel, et al. "Compound Hidden Markov Model for Activity Labelling." *International Journal of Intelligence Science* 5.05 (2015): 177.

[4] Afsar, Palwasha, Paulo Cortez, and Henrique Santos. "An integrated system for human action recognition from video using hidden Markov model." Institute of Electrical and Electronics Engineers, 2015.

[5] Wang, Pichao, et al. "Deep convolutional neural networks for action recognition using depth map sequences." *arXiv preprint arXiv:1501.04686*(2015).