Human Activity Recognition Using SVM Classifier

Meghana M N
Assistant Professor
Dept. ECE, ATMECE

Sushmitha Gowda K R
Assistant Professor
Dept. ECE, ATMECE

Abstract: In video surveillance moving object detection and recognition is the important research area of computer vision. Detection and recognition of moving is not easy task as continuous deformation of objects takes place during movement. Any moving objects has several attributes in temporal and spatial spaces. In spatial space object vary in size where as in temporal space it vary in moving speed. This work mainly focuses on multiple human detection and activity recognition. Multiple human video datasets are considered and in order to detect and track multiple human. Background subtraction technique is used for detecting moving multiple humans. Histogram of Oriented Gradient feature descriptor is used to extract features. For human activity recognition Support Vector Machine classifier is used.

Keywords: computer vision; activity detection; background subtraction; HOG descriptor; SVM classifier

I. INTRODUCTION

Over the recent years, detecting, tracking and action recognition of human beings in a video system is attracting more attention due to its wide range of applications in abnormal event detection, human gait characterization, person counting in a dense crowd, person identification, gender classification, fall detection for elderly people, etc.

Activity detection, tracking and recognition are very demanding research area in computer vision and human computer interaction. In recent time, recognition of human activity from video has drawn attention because of its growing need in various real life environments such as surveillance and security, entertainment and gaming, healthcare, etc. Human movement can be analyzed at different forms; an action is an atomic movement that focuses on small body parts such as waving hands or stretching arms. Activity consists of multiple action primitives and describes the whole body (such as walking or jumping). Interaction is used for two or more humans/objects activities such as carrying a mobile or pointing to a gun. Finally, group activity refers to groups of people with a number of objects such as a group of students or groups of people fighting. Although action and activity are semantically different, but in most cases they are the same. Their database is also applicable for any of them. For this reason, in many cases these two words will be considered interchangeably. The objective of this work is to describe human activity rather than other forms.

In general, the recognition of human activity from video involves different steps, such as, acquisition of input video and extraction of the frames, background subtraction of human body, motion detection tracking and finally, the recognition of the activity. Generally, human activity recognition from video starts with the acquisition of the input video and preprocessing such as frame extraction. Background subtraction includes subtraction of foreground object that is human from the background. The background subtraction step is followed by the motion detection in which the basic idea is to detect the moving objects in the frame and tracking is mainly used to track the detected human. Finally, particular activity are estimated and activity recognition algorithms are used to identify and investigate the actions. Fig. 1 shows the generalized block diagram of human tracking and activity recognition.

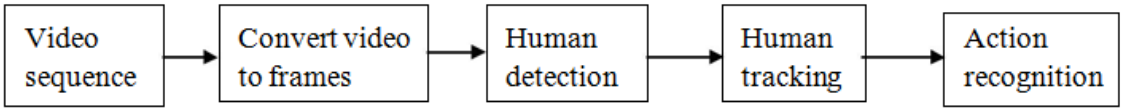

Fig. 1 Block diagram of Human Activity Recognition

i. Video Input

A recognition system obtains the environment information from visual inputs such as cameras. These inputs can be either a still image or a sequence of images.

ii. Detection

The human action and activity recognition systems often begin with human detection. The task of this step is to separate human from different images. This step usually involves image acquisition and background subtraction. First, in image acquisition operations input video is converted into image frames and then these images can be used toward the background subtraction which mainly focuses on detecting the human.

iii. Tracking

Object identification and tracking are two important parts of this system. At this step, the direction are obtained by following human's moving path during the sequential frames of video. The object tracking methods in this case are divided into 2 categories. Region-based tracking and Contour-based tracking.

iv. Recognition

Human activity recognition is used to detect the action and activity.

II. PREVIOUS SURVEY

In this section an overview of latest development of human activity analysis has been shown.

Author	**Year**	**Description**
Ahmad Jalal [1]	2014	Human tracking and activity recognition system which mainly uses body joints features for recognition.
Chunfeng Yuan [2]	2013	Human tracking and activity recognition system utilizes bag-of-visual-words (BOVW) approaches for recognition.
Christian Thurau [3]	2012	Histogram of Oriented Gradients is used since the HOG-descriptor allows for robust detection of humans in images and the set of action primitives to find humans via a nearest-neighbor search.
Caroline Rougier [4]	2011	A shape matching technique is used to track the person's silhouette along the video sequence.

III. BACKGROUND STUDY

A. *Background Subtraction*

To detect moving human in a video surveillance background subtraction is the foremost step. This technique detects moving human by taking difference between the current image frame and background image frame. The current pixel image frame is denoted by P[$C(X,Y)$] and background image pixel frame

denoted by P[B(X, Y)]. The current pixel is subtracted with the background pixel in order to detect moving human in video.

The mathematical representation is given as:

$$P\,[X, Y] = P\,[C(X, Y)\text{-}B(X, Y)] \quad (1)$$

This approach will work well for static background condition. A threshold T is used on this difference image in order to improve the subtraction.

The mathematical form for thresholding is written as:

$$P[C(X,Y)]–P[B(X,Y)]>T \quad (2)$$

Thresholding is the commonly used technique that computes a region as a set of pixels. The subtraction value obtained by taking difference between two images must be greater than a threshold value T, then the foreground image will be extracted. If the difference value is less then threshold T then no foreground image will be detected.

Background subtraction technique with current pixel P[C(X,Y)] and background pixel P[B(X,Y)] with thresholding technique T is depicted in Fig. 2.

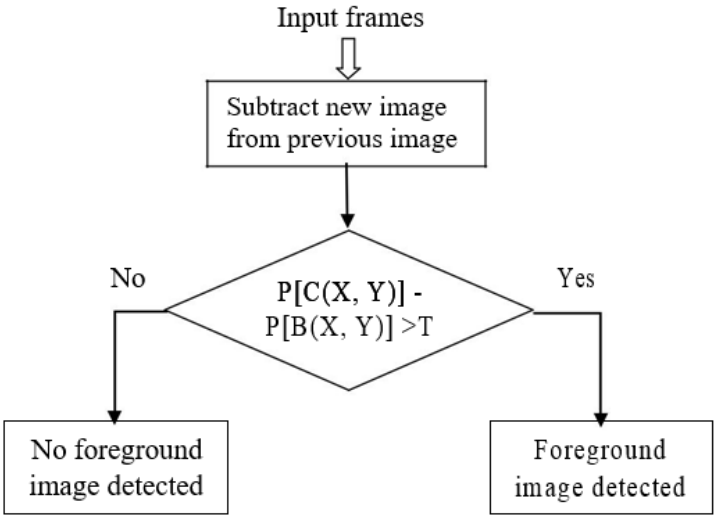

Fig. 2 Background subtraction technique

B. *Histogram of Oriented Gradient descriptor*

Histogram of Oriented Gradients (HOG) is a technique used to extract feature from each frames called images. From each binary pixel images HOG features are extracted.

The algorithm of implementing HOG feature extraction is as follows:

1. The entire image pixels are segmented into small connected parts called as cells. HOG directions is founded for all the pixels in connected path.

2. Discretize all cell into corresponding orientation bins for $[0^0,180^0]$ periods depending on the gradient orientation and all pixels inside cell gives a weighted vote to its angular bin.

3. Grouping the obtained cells into large interlinked blocks and normalize this obtained gradient strengths which represents the histogram for each blocks.

4. The histogram obtained for each blocks represents the descriptor.

C. SVM Classifier

Support Vector Machine (SVM) is a supervised learning model which is associated with the learning algorithm and it is used to analyze the data required for classification. SVMs are based on the decision plane concept that describes decision boundaries. The decision plane separates a set of objects which belongs to different classes. It accomplishes the task of classification by creating hyperplanes in a multi-dimensional space. SVM supports classification as well as regression tasks and can handle multiple categorical and continuous variables. For categorical variables, a dummy variable is constructed with values as either 0 or 1. SVM uses iterative training algorithm to develop an optimal hyperplane, which is used to minimize an error function.

IV. METHODOLOGY

The flowchart of human activity recognition is shown in Fig. 3. The foremost step in multiple human detection and HAR is acquiring the video datasets wherein different activities are performed. Frame extraction is required as videos cannot be processed directly. Later, background subtraction technique is used to find the moving human. Once after noise removal is done the gray scale images will be converted to binary images of 0s and 1s, where binary 1 is used for representing human region which is filled with white color and apart from moving human region binary 0 is used which represents absence of humans.

Once after detecting human the next stage is recognizing their activities by training and testing phase. In training phase the video dataset will be loaded followed by frames extraction. Training folder is created which contains the frames belonging to particular activities. HOG feature extraction technique is used for extraction of features and the features are trained with SVM classifier. In testing phase, the testing video is loaded and tested for different activities.

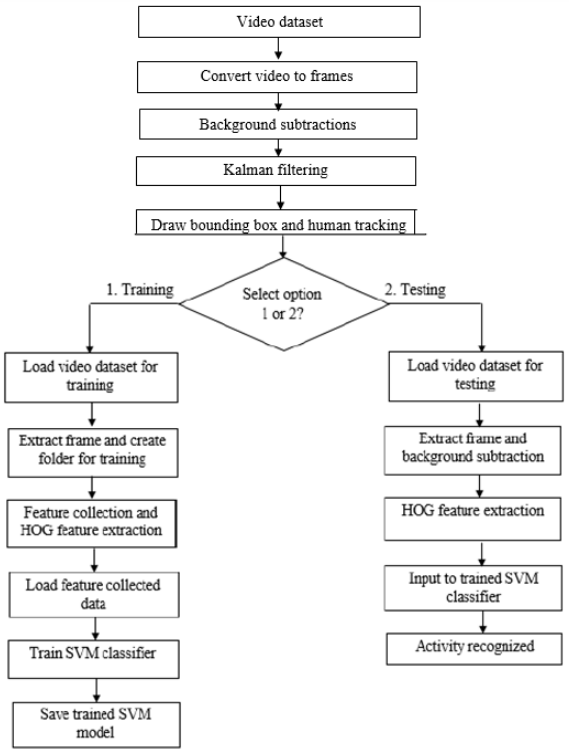

Fig. 3 Flow chart of human activity recognition

V. EXPIREMENTS AND RESULTS

This algorithm is based on Weizmann dataset for human activity recognition. Weizmann dataset comprised of 5 different activities like walk, run, sideways, forward jump and skip. The recorded video is of duration 5 seconds and memory size is about 6.23MB with a frame rate of 25 frames per second. Figure shows the experimental results of human activity recognition using Weizmann dataset performing activities like walk, run, jump and side jump as shown in Fig. 4.

Fig. 4 Recognized activities in Weizmann dataset

VI. CONCLUSION AND FUTURE WORK

The proposed work gives a solution for human detection and activity recognition. Although many works have been carried out, the proposed work provides excellent results for various kinds of video datasets considered. The human detection using background subtraction for static video gives effective result and with HOG feature extraction and SVM classifier recognition of human activities provides good recognition result with less minimum number of false detections.

Future work aimed towards minimizing the false detection like shadow and other reflections of human, it can also incorporate HAR for moving background scenario and recognizing various activity like talking, eating, etc. which can be done by understanding the body pose of each individual human in the video.

REFERENCES

[1] Ahmad Jalal, S. Kamal, Daijin Kim, "Human Activity tracking and recognition using joint features and self-organized map," IEEE Trans. on, Vol. 27, No. 12, pp. 1896-1909, 2014.

[2] Chunfeng Yuan, Xi Li, Weiming Hu, Hanzi Wang, "Human Action tracking and recognition using Pyramid Vocabulary Tree," IEEE Trans. On pp. 1-8, 2013.

[3] Christian Thurau, "Behavior Histograms for Action Recognition and Human Detection," Proceedings of the Inter. Conf. on Advanced Video and Signal based Surveillance, pp. 377-382, 2012.

[4] Caroline Rougier, Jean Meunier, Alain St-Arnaud, Jacqueline Rousseau, "Robust video surveillance for fall detection based on human shape deformation," IEEE Trans. on, Vol. 18, No. 8, pp. 1128-1139, 2011