Human Activity Recognition Using BPNN Classifier On HOG Features

Sonia,
School of Electronics and Electrical Engineering,
Lovely Professional University,
Punjab, India

Abstract: **With an increase in number of surveillance cameras being used everywhere, there is a need of automatic systems for analysis of human activities. Human Activity Recognition is an active research area in computer vision. Research in computer vision has always focused on developing the systems capable of operating automatically. Human activity recognition has many applications such as gesture recognition, robot learning, security systems etc. Human activity recognition is achieved by human detection, feature extraction and classification of activities. In this paper human activity recognition is performed by human detection using background subtraction, feature extraction using HOG and activity classification using BPNN.**

Keyword: Background subtraction, Histogram of Oriented Gradient features, Back Propagation Neural Network(BPNN) classifier

I. Introduction

Human Activity Recognition is the vast area dealing with human motion to understand human activities ranging from simple to complex. Due to prevailing security issues, cameras are being deployed everywhere. There arises a need to develop an intelligent system to recognize human activities and to take necessary action wherever necessary. It has various applications in different fields ranging from human computer interface to robot learning and control. Human activity recognition is done by three steps: human detection, feature extraction and activity classification. First operation is performed to obtain background by background modelling[2] by using some algorithm which can be by using mean filter, median filter etc. In second step, human detection is done. Many algorithms such as Background subtraction, Gaussian mixture model, optical flow, temporal differencing etc. are used for human detection. In third step, features[3] are extracted of the detected human being. Many algorithms such as HOG (Histogram of Oriented Gradients), SIFT (Scale Invariant Feature Transform), SURF (Speeded up Robust Features), Shape based features (eccentricity, centroid, area, orientation etc.), body modelling etc. Then human activities are classified by providing extracted features to the classifier. Classifiers such as BPNN (Back Propagation Neural Network), SOFM (Self Organizing Feature Map), SVM (Support Vector Machine), Bayesian, HMM (Hidden Markov Model) etc. So, with all the steps performed in sequential manner, human activity could be recognized out of a video

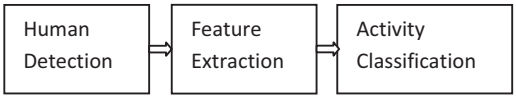

Figure 1. Block Diagram of Human Activity Recognition

The main focus of paper is to recognize human activities by background subtraction, HOG features and BPNN classifier.

II. Literature Review

The importance and emerging technologies in human activity analysis, representation and analysis has led to various researches which are described below:

Manoranjan Paul, Shah M E Haque and Subrata Chakrabort [1] proposed various methods for human detection in videos and its applications. Detection process is classified in two steps: object detection and object classification. Algorithms discussed for object detection are Background subtraction, optical flow and spatio-temporal filter. Object classification is achieved through Shape based method, Motion based method and Texture based method. The most accurate method for object detection is spatio-temporal filter as it considers human motion as whole and keeps into account the complete volume spanned by human motion. For object classification, Texture based methods are found best applicable as localized portions of an image are taken and gradient computation is performed on uniformly spaced cells. Many applications of human activity recognition and various datasets of the same are discussed.

978-1-5386-6483-4/18/$31.00 ©2018 IEEE
DOI 10.1109/ICICS.2018.00065

Rupali S.Rakibe and Bharati D. Patil[2] presented an idea for detection of moving objects based on background subtraction algorithm. Background subtraction gives the complete object information if background of the object is known. Frames are extracted out of the input video, Background could be obtained through mean filter or median filter, Background subtraction is done and hence a moving object is detected. Shape based methods are chosen to determine that moving object is human or not

Pedro Canotilho Ribeiro and Jos´e Santos-Victor[3] classified human activity recognition in 3 major parts: Data Modelling, Feature Selection and Classifier design. 16000 images of 5 different activities were taken for the experiment. Low level activities are considered such as active, inactive, running, walking and fighting. The most suitable and strong features are selected out of all the features so that a classifier could be trained. Features are taken in several sub-groups.

Tao Zhao, Member, IEEE, and Ram Nevatia, Fellow, IEEE[4] presented an approach for multiple people tracking using a single stationary camera. The algorithm is implemented by segmenting human motion into 2 parts: global motion and limb motion. For multiple people segmentation, first people are located by head top and then geometrical shadow analysis is performed. Human masks and shadows are eliminated to represent foreground. For tracking and segmentation of multiple humans, 3D shape model is used.

Javier Andreu and Plamen Angelov[5] discussed the need and different arising applications of human activity recognition systems. As different researchers develop different algorithms for the same, efficiency is measured for the all. System developed for human activity recognition should be able to tackle different levels of complexity as it should be able to differentiate between: Gestures, Individual actions, Human-Object interactions, Human- Human interactions and composite activities. Challenges faced in this context are: (1) To design an algorithm to extract optimal features (2) Real time and scalable systems (3) Low computational algorithms

III. Dataset used

Out of all the available datasets of KTH dataset, Weizmann dataset, PETS dataset and INRIA XMAS dataset, Weizmann dataset[10][1] is used. There are total of nine activities in the dataset which are run, bend, jump, double hand wave, single hand wave, jack, side walk, jack and skip.

Figure 2. Weizmann dataset

IV. Research Methodology

A. Background Subtraction

For recognition of human activities, background modelling is the first required step. Background modelling gives the background and areas of the image containing important information. Background estimation is performed using mean filter in which n different frames are taken. It is the best approach for finding the background of a moving object with a static background. Number of frames being used should not be too less as it may give improper background.

$$Bi(x,y,t) = \frac{1}{n}\sum_{i=0}^{n-1} Im(x,y,t-i) \quad (1)$$

where, *Bi*(x, y, t) = Background image estimated at time t,
n = Number of frames,
Im(x, y, t-i) = Image at time t-i.

After estimating the background of video, next step is to detect moving human out of that. Different algorithms of the type such as Gaussian Mixture Model, Temporal Differencing, Min Max method, Background estimation and others can be employed for the same. Background estimation is used as it will provide proper results for static background. In background subtraction, difference of current frame and background is taken in pixel by pixel form. When the image difference is above the threshold value, it is taken as foreground image.

$|I_{cf} - I_{bk}| > T$

Where, I_c = current image,

I_{bk} = background or reference image

T = Threshold value

After background subtraction, if output does not come properly, morphological operation can be applied for removing the erroneous blobs present.

B. Histogram of Oriented Gradients (HOG):

After detecting the presence of moving human in the video, next step is to extract the necessary features to describe the human motion and silhouette.

Histogram of oriented gradients (HOG) is the local feature descriptor [8] used for activity classification. The basic idea behind HOG is that local shape information can be described completely by intensity gradients or edge directions. HOG technique works by taking gradient portions of localized images. Algorithm is described as follows:

- Video is taken and frames (images) are separated out of it.
- Every image is divided into N block regions called "blocks".
- Local regions formed in previous step are divided into "cells".
- Within all the cells, histogram of edge orientations is accumulated.
- All the histograms are combined and used as feature descriptor.

HOG features are extracted in local regions with 16 × 16 pixels. Histogram of oriented gradients is calculated in each 4 × 4 local cells. Sobel filter has proved the best for calculating edge gradients and orientations. Each feature is defined by position of its cell $P(u_p, v_p, w_p, h_p)$, the parent local region position $L(u_l, v_l, w_l, h_l)$ and the orientation bin number k. So, each cell feature f is denoted by $f(P, L, k)$.

Step 1: In image S, gradient at the point (u, v) is found by convolving gradient operator with the image.

$$G_u(u,v) = [-1\ 0\ 1] * S(u,v) \quad (2)$$

$$Gv(u,v) = [-1\ 0\ 1]^T * S(u,v) \quad (3)$$

Step 2: The strength of gradient at the point (u, v) is given as:

$$G(u, v) = \sqrt{G_u(u,v)^2 + G_v(u,v)^2} \quad (4)$$

Step 3: The orientation of the edge at point (u, v) is given as:

$$\theta(u,v) = arctan\left[\frac{|G1(u,v)|}{|G2(u,v)|}\right] \quad (5)$$

Step 4: Orientation range is divided into k bins and value of k_{th} bin is denoted as:

$$B_k(u, v) = G(u, v), \text{ if } \theta(u,v) \in bin_k \quad (6)$$

Step 5: Then the feature value is defined as:

$$f(P, L, K) = \frac{\sum_{(u,v)\in C} B(u,v)}{\sum_{(u,v)\in B} G(u,v)} \quad (7)$$

So, every HOG feature can be found at different position of u and v.

C. BPNN Classifier:

The back propagation neural network is the learning algorithm used for classification purposes. BPNN [6] is a multi-layer neural network which consists of a hidden layer, an input layer and an output layer. Biased neurons are present in output layers and hidden layers. Activation function is responsible for biasing the neurons and it is differentiable. BPNN training is done in three parts: Feed-forwarding the input training pattern, error calculation and its back propagation, weight updating

1. Dataset is divided in to 2 parts: Training dataset and testing dataset.
2. Train the classifier using training dataset and test it using testing dataset.
3. Call the function which trains the classifier with the features of the training data and activity vector.
4. Required features are obtained for testing dataset by histogram of oriented gradients.
5. After training the classifier, predict the activities of testing data.
6. Plot the confusion matrix.

V. Results

Total of 9 activities comprise the activity video which is loaded in MATLAB Software. There are total of 360 frames with each activity comprising of 40 frames. Size of every image sequence in video was 144 x180 x3.

Background obtained is given below

Figure 3. Background Image

After obtaining background image, human is detected out of the video by using background subtraction. Bounding box represents the detected part and will move along as the human will move.

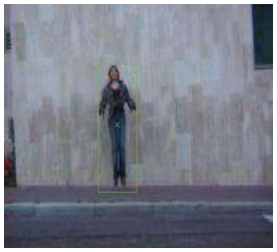

Figure 4. Human Detection for jump activity

HOG Feature Vector: Using HOG, total of 200 features are selected for detected human in the video. For 180 frames in the training set, a feature vector 180×200 is formed. For 180 frames in the testing set, a feature vector of 180×200 is formed.

Table 1. Confusion matrix for 9 classes

P / A	A1	A2	A3	A4	A5	A6	A7	A8	A9
A1	40	0	0	0	0	0	0	0	0
A2	0	39	1	0	0	0	0	0	0
A3	0	0	40	0	0	0	0	0	0
A4	0	0	0	40	0	0	0	0	0
A5	0	0	0	0	40	0	0	0	0
A6	0	0	0	0	0	40	0	0	0
A7	0	0	0	0	0	0	40	0	0
A8	0	0	0	0	0	0	0	40	0
A9	0	0	0	0	0	0	0	0	40
Total Accuracy = 99.7%									

A1, A2, A3, A4, A5, A6, A7, A8 and A9 represent Bend, Run, Jump, One Hand Wave, Two Hand Wave, Walk, Jack, Side Walk and Skip respectively. In Confusion Matrix P represents predicted activities and A represents actual activities.

VI. Conclusion

The proposed method for Human Activity Recognition will help to recognize the human activities. To avoid any confusion for the classifier to classify correctly, features are only extracted for the moving human. Human Activity Recognition is done using Weizmann dataset and accuracy of 99.7% is achieved. It works effectively well for extracted HOG features.

VII. Future work

In this paper, static background is taken in the dataset. It could also be implemented for dynamic background. Different classifiers could be used to avoid delay in the system's output.

VIII. References

[1] Manoranjan Paul, Shah M E Haque and Subrata Chakraborty,"Human detection in surveillance videos and its applications ",Paul et al. EURASIP Journal on Advances in Signal Processing2013,2013:176.

[2] Rupali S.Rakibe, Bharati D.Patil,"Background subtraction algorithm based human motion detection", International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013

[3] Pedro Canotilho Ribeiro & Jos´e Santos-Victor," Human Activity Recognition from Video: modeling, feature selection and classificationarchitecture",
International workshop on human activity recognition andmodelling.

[4] Tao Zhao, Member, IEEE, and Ram Nevatia, Fellow, IEEE," Tracking multiple humans in complex situations", IEEE transactions on pattern analysis and machine intelligence, VOL. 26,NO.9.

[5] Javier Andreu • Plamen Angelov," Towards generic human activity recognition for ubiquitousApplications", J Ambient Intell HumanComput(2013)4:155–156.

[6] Neural Network Toolbox, www.mathworks.in/product/neural-networks

[7] Xianping Huang, Lili Zheng, Ronhua Liang, Wanliang Wang, "Human Action Recognition Based on SVM using multiple features",International Conference on Artificial Intelligence and Softcomputing,vol.12,2012.

[8] Navneet Dalal and Bill Triggs, "Histograms of Oriented Gradients for Human Detection", CPVR 2005.

[9] Yong-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo,"A review on video based human activity recognition", Computers 2013, 2, 88–131 manuscripts; doi: 0.3390/computers2020088

[10] www.cs.utexas.edu