Anomaly Detection in Human Activity

Saurabh Singh, Shinjan Sengar, Shivam Mishra, Manoj Kumar
Department of Information Technology JSS Academy of Technical Education,
NOIDA, India

Abstract- Anomaly detection in Human Activity from video is a crucial research topic nowadays because of its wide variety of applications such as security monitoring, video surveillance and healthcare applications. The method is used to distinguish between normal and abnormal human activities from a video. Videos are taken as inputs and then each video is converted into frames. Two-dimensional visual saliency map is created from color frames with the help of Image Signature [1]. Harris-Stephens algorithm is used to detect interest points from the saliency map. Histogram of Oriented Gradients (HOG) is used as a feature descriptor. Finally, Decision Tree (DT) is used for classification of activities into two classes: normal and abnormal.

Keywords- Anomaly detection, Saliency map, Image Signature, Harris-Stephens algorithm, Histogram of Oriented Gradients (HOG), Decision Tree (DT).

I. Introduction

Recognizing behavior of a human from videos is a challenging problem in many application areas, such as video surveillance, security and health care involving computer vision and machine learning applications. To monitor the behavior, daily activities and any other information of elderly, a close observation is necessary to protect them. With increased population of older adults in society, there is an urgent need for assistive technologies in the home. Older adults face many difficulties while undergoing their daily activities because of age-related changes. Thus, to take care of older adults, it is very important to know if any unusual activities are there. Abnormal activities of human are still difficult to recognize because they are not able to predict it prior and those types of strange events does not occur frequently. It is important to identify an emerging medical condition prior to it gets critical. So, it becomes necessary to observe the activities of daily livings seek for abnormal behavior in daily life. In this paper, the work is focused on to detect unusual or abnormal activities instead of considering regular activity recognition. “Abnormal activities” can be defined as “activities which are infrequent and not predicted in advance”.

In this work, a system is proposed to distinguish between normal and abnormal activities which uses 2D visual saliency map from color video sequences. Image Signature is used to create the saliency map. We show, within the theoretical framework of sparse signal mixing, that this quantity spatially approximates the foreground of an image. We experimentally investigate whether this approximate foreground overlaps with visually conspicuous image locations by developing a saliency algorithm based on the image signature. UR fall detection [2] dataset is used to evaluate the performance. Harris-Stephens algorithm is used detect the in from the saliency map. Histogram of Oriented Gradients (HOG) feature descriptor is used to extract the features from the interest point. Lastly, K-Nearest Neighbors classifier is used to classify the activities into two classes i.e. normal and abnormal. We will use ensemble technique to maximize our accuracy.

The rest of the paper is organized as follows: In Section 2 work related to abnormal activity detection is discussed. The proposed methodology and architecture is in Section 3. In Section 4, experimental results are presented with all the necessary details and discussion. Finally, the conclusion is provided in Section 5.

II. Related work

People working in the field of machine learning and computer vision, gives much attention to human activity recognition. Human performs large number of activities which can be classified into normal and abnormal activities. Various approaches have been proposed, for both crowded and noncrowded scenes.

They can be broadly categorized according to the type of scene representation adopted. One very popular category is based on trajectory modeling. It comprises tracking each object in the scene and learning models for the resulting object tracks. Both operations are quite difficult on densely crowded scenes, for which these approaches are not very promising.

Holistic image processing short-circuits the need for segmentation, key-point matching, and other local operations. Bolstered by growing general interest in large-scale image retrieval systems, holistic image descriptors have become a topic of intense study in the computer vision literature. GIST [3] is an excellent example of such an algorithm in this field. Other holistic scene models focus on the separation of foreground and background. For example, Candes et al. [4] introduced a sparse matrix factorization model. A more relevant study comes from Hou and Zhang [5], motivated by Oppenheim et al.'s early discovery [6], [7]. They found that the residual Fourier amplitude spectrum, the difference between the original Fourier amplitude spectrum and its smoothed copy, could be used to form a saliency map. The residual retains more high-frequency information than low, where the smoothed copy is similar to the original. The image signature, in comparison, discards amplitude information across the entire frequency spectrum, storing only the sign of each DCT component, equivalent to phase for a Fourier decomposition. The image signature is thus very compact, with a single bit per component, and as we shall show in the remainder of this paper, possesses important properties related to the foreground of an image.

These systems are mainly classified into single layered and hierarchical approaches [2]. Single layered techniques are used to represent the activity directly based on image sequences and further classified into methods such as space-time and sequential approaches. Space-time approaches consider the activity in 3D volume and represent it in space-time features from given video sequences. These approaches are again divided based on features used from 3D volume [8]. Sequential approaches describe the activity in sequence of observations and can be further classified depending on the technique they use for recognition such as Exemplar or state-based recognition [9].

Hierarchical techniques used for recognizing complex activities by analyzing the video into various feature descriptors [10]. Hierarchical approaches are categorized into statistical, syntactic and description-based methodologies. Statistical strategies used to represent the high-level human activities by concatenating state-based models hierarchically. Hidden Markov Model (HMM) is an example of such type of approach. Syntactic models use grammar-based syntax and model the activities as strings of symbols [11]. Description based models describe the sub-event of activities to represent human activities. Abnormality or anomaly can be realized by using normal activities and depends on the approaches used to classify them. Three broad categories such as supervised, semi-supervised and un-supervised are used to construct the model [12].

The training data of normal and abnormal behavior is provided in case of supervised approaches to detect anomalous or unusual data. Semi-supervised approaches use only normal behavior to train the model and detect abnormality either automatically or through the training process. Un-supervised approaches do not need any kind of training. These approaches work on some rule base or the conditions describing distinction between classes.

III. Proposed Methodology and Architecture

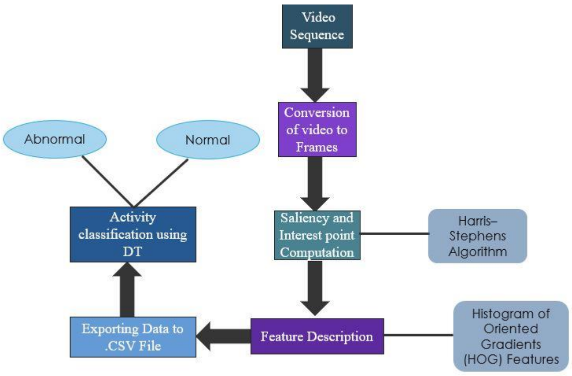

Fig 1: Architecture of proposed system

In proposed system Input videos are converted into frames. Further, we detect salient regions from each frame. The saliency of each frame is computed using DCT-based image signatures [1]. Once we find the salient image, next task is to find interest points those effectively contribute in describing the motion of the object. We use Harris-Stephens algorithm to detect the interest points. To remove or minimize unwanted interest points we introduced intermediate step of saliency computation before interest point detection. Further, based on interest points, feature descriptor is formed for each frame in a video. The activity in the video is represented using a feature vector. Finally, DT is used for classification of activities into two classes: normal and abnormal.

The detailed methodology is described in the following text.

Fig 2: RGB frame from one of the video

A. Saliency Computation

A part of an image that catches the attention of a viewer as it stands out from its neighborhood is called salient region. An object or a pixel in an image is referred as salient depending on the measure or quality by which it is distinguished from its surrounding. With respect to the application under consideration, we need to detect salient object; that is the moving foreground object in video under consideration. Hou et al. [1] used a binary, and holistic image descriptor called the "image signature" to highlight salient regions in the image. It is defined as the sign function of the Discrete Cosine Transform (DCT) of an image. We have used this method to compute salient image for further processing.

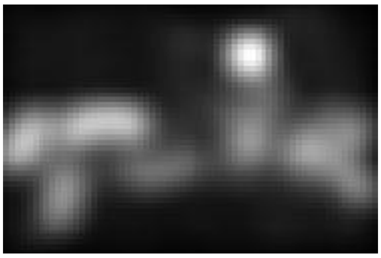

Fig 3: Saliency Map of Frame

B. Interest Point Detection

At this point we used the Harris-Stephens algorithm to detect the interest points.

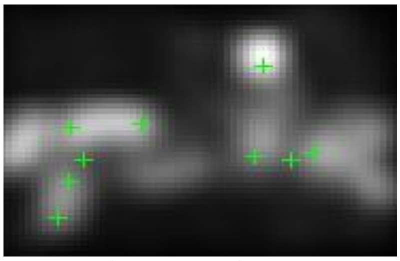

Fig 4: Saliency Map with Interest Points

C. Feature Description

We use Histogram of Oriented Gradients (HOG) as a feature descriptor in our proposed system. The features are returned in a 1-by-N vector, where N is the HOG feature length. The returned features encode local shape information from regions within an image.

D. Training and Testing

The proposed model is trained and tested using Decision Tree (DT).

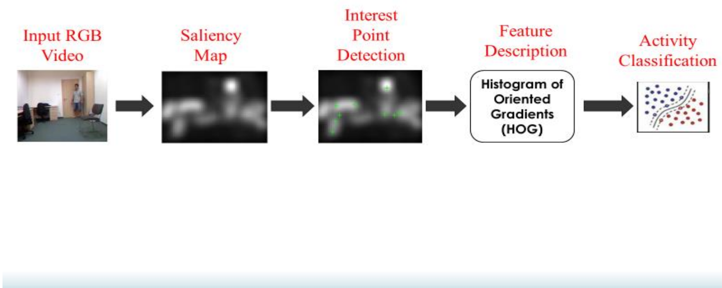

Fig 5: Block diagram for activity recognition

IV. Results

A. Dataset

The dataset consists of 63 (30 fall activities + 33 activities of daily living). The proposed system uses only RGB videos for processing. Here, fall is considered as abnormal activity and activities of daily living are considered as normal activities to evaluate the performance.

Normal activities such as walk, sit down, bend, lye on a bed and abnormal activities such as fall while sitting on a chair, fall while walking has been performed by five subjects.

B. Result

Experiments are performed with dataset. The results of intermediate steps as per methodology are shown in Figs. 2 ,3 and 4. Decision Tree classifier is used for classification

For this purpose, we partition the dataset with 85% samples used for training while remaining 15% used for testing. For the dataset accuracy of 90% is obtained.

V. Conclusion

This work proposes the method for abnormal activity detection in the home environment. Here we consider normal activities as daily activities of the person such as sitting on a chair, lying on the bed, reading, writing, picking up the fallen object, tightening shoelace, sweeping, cleaning, etc. Abnormal activities include forward fall, backward fall, fall from standing position, fall from a chair, etc.

The proposed system aims at detecting abnormal activities in the home environment. This is an attempt towards the aim of building a support system assisting elderly people living alone. The system involves first saliency computation from frames followed by feature extraction and description from frame. Finally, Decision Tree is used as a classifier to recognize whether the activity is normal or abnormal. Our aim is to report abnormal activity of elderly people correctly, we found the results of our system highly encouraging. The dataset is partitioned with 85% samples used for training and 15% used for testing. For dataset the system is giving 90% accuracy. Also, the attempts will be made to improve the efficiency of the system so as to get better computational cost.

VI. References

[1] Xiaodi Hou, Jonathan Harel, and Christof Koch, "Image Signature: Highlighting Sparse Salient regions", 2012.

[2] Smriti H Bhandari and Navnee S. Babar, "Abnormal Activity Recognition Using Saliency and Spatio-Temporal Interest Point Detector", 2018. no. 3, pp. 145-175, 2001.

[3] H. Zhou, H. Friedman, and R. von der Heydt, "Coding of Border Ownership in Monkey Visual Cortex", 2000.

[4] E. Cande`s, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis", 2009.

[5] X. Hou and L. Zhang, "Saliency Detection: A Spectral Residual Approach", Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2007.

[6] A. Oppenheim and J. Lim, "The Importance of Phase in Signals," Proc. IEEE, vol. 69, no. 5, pp. 529-541, May 1981.

[7] M. Hayes, J. Lim, and A. Oppenheim, "Signal Reconstruction from Phase or Magnitude," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 28, no. 6, pp. 672-680, 1980.

[8] Wang, H., Kläser, A., Schmid, C., Liu, C.L.: Dense trajectories and motion boundary descriptors for action recognition,2013.

[9] Aggarwal, J.K., Ryoo, M.S.: Human activity analysis: a review, 2011.

[10] Jhuang, H., Serre, T., Wolf, L., Poggio, T.: A biologically inspired system for action Recognition,2007.

[11] Tsai, W., Fu, K.S.: Attributed grammar-a tool for combining syntactic and statistical approaches to pattern recognition, 1980.

[12]. Brax, C., Niklasson, L., Smedberg, M.: Finding behavioural anomalies in public areas using Video surveillance data, 2008.

[13] Liming Chen, Member, IEEE, Chris Nugent, Member, IEEE, and George Okeyo, Member, IEEE, "An Ontology-Based Hybrid Approach to Activity Modeling for Smart Homes, IEEE transactions on humanmachine systems, vol. 44, no. 1, February 2014.

[14] Bingbing Ni, Yong Pei, Pierre Moulin, Fellow, IEEE, and Shuicheng Yan, Senior Member, IEEE, Multilevel Depth and Image Fusion for Human Activity Detection, IEEE transactions on cybernetics, vol. 43, no. 5, October 2013.

[15] Megha D Bengalur, Human Activity Recognition using Body pose features and support vector machine, International Conference on Advances in Computing, Communications and Informatics (ICACCI) 2013, IEEE Conference.

[16] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos. Anomaly detection in crowded scenes. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2010.