

# Behavior Analysis from Videos Using Motion based Feature Extraction

Priyanka Kokate , Surendra Gupta

Department of Computer Engineering Shri G.S. Institute of Tech. and Science 23, Sir Visvesvaraya Marg, Indore (MP)

Abstract— Now a days crimes are increasing with faster rate. According to reports published crimes in India have seen a spike of approximately 79.4 % in past 20 years. Therefore, an automated system should be there to detect the criminal activities at the time crime is taking place. For this, behavior of the person need to be analyzed, person with abnormal behavior should be categorized as suspicious object and should be signaled to human operator. Behavior analysis from the videos should be done to avoid crimes upto an extent. In this paper, we present an approach for human behavior analysis.

In this paper, we propose a novel algorithm based on the motion feature to detect abnormal behaviors in video based systems. Different from the previous work that uses independent local feature, the algorithm explores the global moving relation between the current behavior state and the previous behavior state.

*Index-terms* - Feature Extraction, Classification, Behavior, KNN(K nearest neighbor).

## I. INTRODUCTION

The foremost objective of Behavior Analysis is to define whether the behavior of a individual is normal or abnormal. Abnormal behavior is defined as the behavior which is diverges from what is likely as usual behavior. Abnormal behavior is every so often hard to define, for copious causes. First of all, people should reach a decision on what's normal. This can differ widely both inside and outside values.

The helpfulness of a human behavior study with growing crime rates in now a days, there is a conforming awareness for the need of detecting abnormal motion. Automation of abnormal human behavior exploration can play a substantial role in safety by reducing the interval taken to thwart undesirable events and recognizing them up through the suspicion stage only. With improvements in technology, processing of video have become more automated than manual. Behavior Analysis of human even though crucial and is highly challenging. But by automating the human behavior detection the risk can be avoided upto a certain extent.

The need for smart video surveillance systems is due to the large amount of highly security-sensitive fields such as department stores, banks, checkpoints, Airports, Parking lots etc. and the large number of videos collected from these sensitive areas by surveillance cameras. Maximum work of behavior analysis focuses on feature extraction and behavior description, which are closely connected.

In image processing and computer vision, a feature is a piece of information which is applicable for solving the computational work affiliated to a certain application. Features may be specific cognition in the image such as edges, points or objects. According to the features used for analysis, the behavior analysis methods can be classified into three kinds: spatial-based (such as shape), motion-based (such as trajectory) and spatial-temporal-based methods. Motion based features are time dependent features, they define relation between the frames for example frame at time  $t$  and time  $t+1$ . Motion Features captures the motion between the frames, identify hidden pattern between them.

If the evaluation procedure is tied to as task that is learning algorithm is used to classify the data, the feature extraction approach use motion based features to compute feature of the frames and the features are merged together to form a single feature vector for each individual video. Further this feature vectors are used to classify and predict the suspicious human behavior from videos.

Therefore, in this paper, a technique has been proposed for behavior analysis of human from videos. The basic idea of this paper is to reduce the efforts required to analyze the human behavior, which is done manually. To improve the prediction accuracy, we have used the motion feature extraction method. In addition, we reduce the number of feature space that is non motion based methods in which each frame is processed and a feature vector is generated corresponding to every single frame, while in motion based feature extraction each individual video have a single feature vector hence the feature space is equivalent to number of videos and also reduce unnecessary or noisy information. Besides, this improves the presentation of information pre- diction with speeding up the processing algorithm.

This paper is organized as follows. In section II an overview of work already done in this area with their advantages and disadvantages is discussed. In the section, III the proposed approach is presented. Section IV presents the details of experimental analysis of the result. Conclusion and future work is discussed in section V.

## II. RELATED WORK

This paper[1] present a system which takes as input a video stream obtained from an airborne moving platform and produces an analysis of the behavior of the moving objects in the scene. To achieve this functionality, there system relies on two modular blocks. The first one detects and tracks moving regions in the sequence. It uses a set of features at multiple scales to stabilize the image sequence, that is, to compensate for the motion of the observe, then extracts regions with residual motion and uses an attribute graph representation to infer their trajectories. The second module takes as input these trajectories, together with user-provided information in the form of geospatial context and goal context to instantiate likely scenarios. They present details of the system, together with results on a number of real video sequences and also provide a quantitative analysis of the results.

To bridge the gap between a high level, symbolic description of an activity and the information of signal level provided by the sensors, a hierarchy of entities was proposed. Involved entities in the recognition of activity are organized into abstraction of three levels : extraction of features, prop- erties of mobile object, and scenarios. Image processing the first module detects moving regions and computes features by the use of visual data from an individual frame. They had used eight features which include width, direction of motion, height, speed and the distance to a object referenced. Tracking module then tracks the regions which were detected can be correspond to noise, to a part of a mobile object, to one mobile object, or to a group of mobile objects. Image features are not interpreted and are imprecise. Hypotheses are generated by behavior analysis module to consider the tracked moving regions as mobile objects composed of one or several regions. Spatial and temporal properties of mobile objects are computed at several layers. Based on mobile object properties relative behavior of objects are analyzed.

The paper[3] provides localization and detection of anomalous behaviors in crowded places, and spatial and tem- poral joint feature extraction is proposed. Proposed method was based on how objects are appeared and movements of the objects, using a set of mixture of dynamic textures models. These models are used to implement 1) A center- surround discriminant saliency detector that produces spatial saliency scores, and 2) A model of normal behavior that is learned from training data and produces temporal saliency scores. On multiple scales Spatial and temporal anomaly maps are defined. The multiscale scores act as potentials of aconditional random field that guarantees global consistency of the anomaly judgments. A data set of densely crowded pedestrian walkways is introduced and used to evaluate the proposed anomaly detector. Experiments on this and other data sets show that the latter achieves state-of-the-art anomaly detection results.

This paper[10] present a novel statistical framework for modeling the local spatio-temporal motion pattern behavior of extremely crowded scenes. There key insight is to exploit the dense activity of the crowded scene by modeling the rich motion patterns in local areas, effectively capturing the underlying intrinsic structure they form in the video. In other words, they model the motion variation of local space-time volumes and their spatial-temporal statistical behaviors to characterize the overall behavior of the scene. They demonstrate that by capturing the steady- state motion behavior with these spatio-temporal motion pattern models,we can naturally detect unusual activity as statistical deviations. There experiments show that local spatio-temporal motion pattern modeling offers promising results in real-world scenes with complex activities that are hard for even human observers to analyze.

In paper[11] they propose a context-aware method to detect anomalies. By tracking all moving objects in the video, three different levels of spatiotemporal contexts are considered, i.e., point anomaly of a video object, sequential anomaly of an object trajectory, and co-occurrence anomaly of multiple video objects. A hierarchical data mining ap- proach is proposed. At each level, frequency-based analysis is performed to automatically discover regular rules of nor- mal events. Events deviating from these rules are identified as anomalies. The proposed method is computationally efficient and can infer complex rules. Experiments on real traffic video validate that the detected video anomalies are hazardous or illegal according to traffic regulations.

In paper[5] they propose a novel algorithm based on the acceleration feature to detect anomalous crowd behaviors in video surveillance systems. Due to the unstable optical flow resulting in the unstable speed, a new global acceleration feature is proposed, based on the gray-scale invariance of three adjacent frames. It can ensure the pixels matching and reflect the change of speed accurately. Furthermore, a detection algorithm is designed by acceleration computation with a foreground extraction step. The proposed algorithm is independent of the human detection and segmentation, so it is robust. For anomaly detection, this paper formulates the abnormal event detection as a two-classified problem, which is more robust than the statistic model-based methods, and this two-classified detection algorithm, which is based on the threshold analysis, detects

anomalous crowd behaviors in the current frame. Finally, apply the method to detect abnormal behaviors on several benchmark data sets, and show promising results.

### III. PROPOSED METHOD

The proposed approach is focused on classifying the behavior of the human as normal or abnormal accordingly from the videos and then predicting it as well.

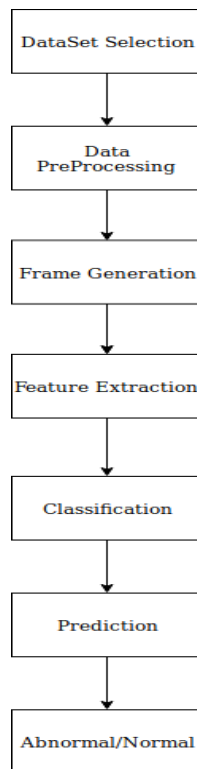
It consists of data preprocessing module, feature extraction module, classification module, prediction module. Architecture of the proposed system is illustrated in figure 1.

#### A. Data preprocessing Module and Frame Generation

The system takes videos dataset as an input. It read the videos from the dataset split every single video in the form of frames. Multiple frames are generated from single video.

#### B. Feature Extraction Module

Apply Feature Extraction on the given data to extract features of the frames generated.



**Fig. 1:** Architecture Diagram of Proposed Approach

- Apply Motion Feature extraction method to extract features.

#### C. CLASSIFICATION MODULE

- 1) Load the generated feature space.
- 2) Use KNN(K-Nearest Neighbour) for classification.

Predictions are made by KNN directly by using the training dataset. Predictions are made for a new instance (x) by searching through the entire training set for the K most similar instances (the neighbors) and summarizing the output variable for those K instances. For regression this might be the mean output variable, in classification this might be the mode (or most common) class value.

Distance measure is used to determine which of the K instances in the training dataset are most similar to a new instance. Most popular distance metric for real-valued input variable is Euclidean distance.

**Characteristics of KNN:**

- When KNN is used for classification, the output is calculated as the class with the highest frequency from the K-most similar instances.
- Each instance in essence votes for their class and the class with the most votes is taken as the prediction.
- Ties can be broken consistently by expanding K by 1 and looking at the class of the next most similar instance in the training dataset.

**Algorithm 1** Heuristic Algorithm

**Input:** Frames generated from individual video. **Output:** Extracted features and generated feature vector.

**Procedure:**

- 1: Loop
- 2: Calculate the intensity feature value of each attribute (96 attributes).
- 3: Find the average of each intensity attribute value corresponding to individual frame.
- 4: Subtract the average intensity value from each intensity feature value.
- 5: Square all the newly generated values after subtraction.
- 6: Now add all the newly generated intensity feature values.
- 7:

$$\sigma^2 = \frac{\sum(X - \mu)^2}{N} \tag{1}$$

$$= \frac{\sum(X^2 - 2\mu X + \mu^2)}{N} \tag{2}$$

$$= \frac{\sum X^2}{N} - \frac{2\mu \sum X}{N} + \frac{N\mu^2}{N} \tag{3}$$

$$= \frac{\sum X^2}{N} - 2\mu^2 + \mu^2 \tag{4}$$

$$= \frac{\sum X^2}{N} - \mu^2 \tag{5}$$

- 8: A single feature vector is generated corresponding each single intensity attribute and thus a single motion feature vector is generated corresponding every single video.

**Properties of KNN :**

- 1) Lazy Learning : Learning of the model is not required and all calculation and processing happens at the time of prediction. Hence, KNN is often called as a lazy learner.
- 2) Instance-Based Learning : Training instances are used for predictions are raw training instances. Hence KNN is often referred to as instance-based learning (where each training instance is a case from the problem domain).
- 3) Non-Parametric : No assumption is made by KNN about the data and functional form of the problems to be solved. Hence KNN is referred to as a non-parametric machine learning algorithm.

Classification module takes resultant feature vectors of videos as an input. Dataset is divided into two parts: training set and testing set for applying classification. Classifier is trained using training set and using testing set its accuracy is computed.

*D. Prediction Module*

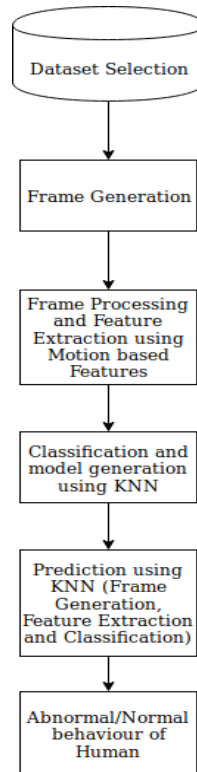
1. Load the model generated by the Classification module (trained model).
2. Test the model over the test data (videos that need to be classified).

*E. Performance Evaluation Module*

There are many parameters available which can be used to check performance of classification. The parameters used in the project to check classification performance are :

- Accuracy : It is the ratio of number of correct prediction to total no of prediction.
- Recall : Recall refers to the percentage of total relevant results correctly classified by your algorithm

- Precision : It is a measure of exactness. It can be defined as, from the samples labelled as positive, how many actually belongs to positive class.
- F-measure : It is the harmonic mean between recall and precision.



**Fig. 2:** Flow Diagram of Proposed Approach

#### IV. IMPLEMENTATION

Basic version or non motion based version is implemented first in which only intensity features of each frame is extracted, stored and further used as input to classification module. In the second version motion based features are extracted of each individual video and stored, which is further used as input to classification module.

#### V. TESTING AND RESULTS

Proposed approach is implemented using Python. It is evaluated on Fights and Riots datasets. The model is trained over 247 videos.

The implemented method provides 68% testing accuracy with dataset of Fights and 64% testing accuracy with dataset of Riots. The proposed method is tested over 180 videos of Fights and 100 videos of Riots.

**TABLE II: THE COMPARISON OF MOTION BASED AND NON MOTION BASED FEATURE EXTRACTION CLASSIFICATION**

K=3, Fights Data set

Parameters	Motion Based	Non Motion Based
Accuracy	0.68	0.42
Recall	0.75	0.66
Precision	0.95	0.19
F-measure	0.72	0.32

K=3, Riots Data set

Parameters	Motion Based	Non Motion Based
Accuracy	0.56	0.14
Recall	0.28	0.5
Precision	0.96	0.14
F-measure	0.72	0.25

**TABLE III: THE COMPARISON OF MOTION BASED AND NON MOTION BASED FEATURE EXTRACTION CLASSIFICATION**

K=5, Fights Data set

Parameters	Motion Based	Non Motion Based
Accuracy	0.62	0.35
Recall	0.65	0.56
Precision	0.84	0.13
F-measure	0.68	0.3

K=5, Riots Data set

Parameters	Motion Based	Non Motion Based
Accuracy	0.51	0.15
Recall	0.25	0.51
Precision	0.96	0.15
F-measure	0.67	0.25

**TABLE I: THE COMPARISON OF MOTION BASED AND NON MOTION BASED FEATURE EXTRACTION CLASSIFICATION**

K=1, Fights Data set

Parameters	Motion Based	Non Motion Based
Accuracy	0.7	0.41
Recall	0.8	0.66
Precision	0.95	0.18
F-measure	0.79	0.31

K=1, Riots Data set

Parameters	Motion Based	Non Motion Based
Accuracy	0.63	0.11
Recall	0.31	0.5
Precision	0.96	0.11
F-measure	0.77	0.21

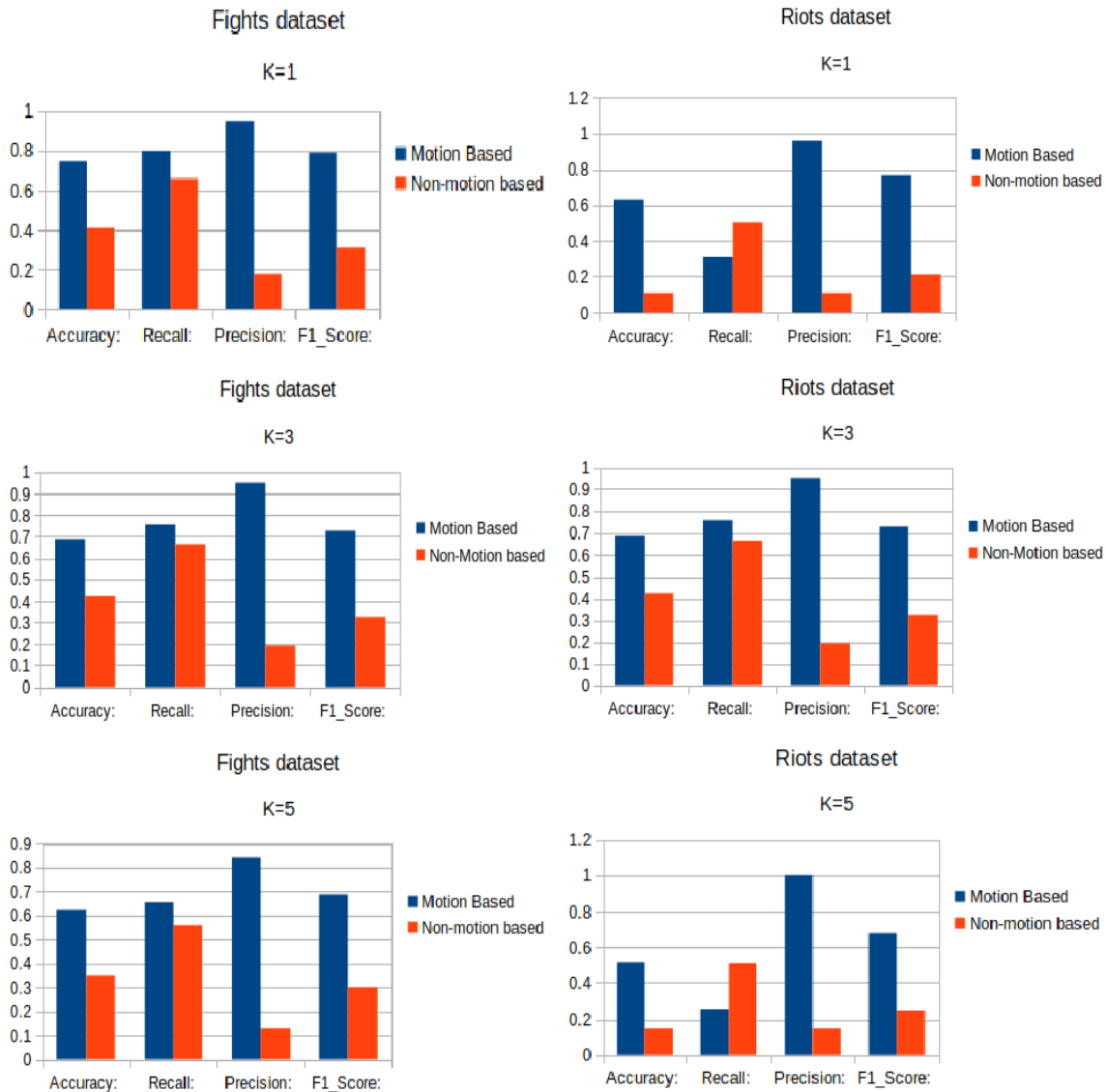


Fig. 3: Parameters comparison on Fights Dataset

Hence the above results shows that the above technique gives best results with hyper-parameter k=1 with both the datasets.

## VI. CONCLUSION

With the growing number of surveillance cameras installed in private and public areas, there has been a demand for the automatic and intelligent analysis of video sequences using computers. Unusual behavior detection in a crowded scene has recently been of great interest in the area of vision based surveillance. The outcome of the project is focused of predicting the behavior of the person in the video is normal or abnormal. Determining the abnormal videos will help to decrease the crime rate upto an extent. Expected result shows increase in classification accuracy of proposed model as compared to the base model. Accuracy is most significant in the field of crime detection and prevention to prevent major disaster upto an extent. Experimental results show that Motion based feature extraction, a feature extraction technique greatly boost the accuracy of classification and prediction. For machine learning problems Feature extraction, plays an important part in extraction of information from images "extracting from the raw data information that is most suitable for classification purposes, while minimizing the within class pattern variability and enhancing the between class pattern variability, has been proven to be effective and efficient in preparing high-dimensional data." [9] feature extraction is a special form of dimensionality reduction.

The main goal of feature extraction is to obtain the most relevant information from the original data and represent that information in a lower dimensionality space.

## VII. FUTURE WORK

The proposed method is applied to detect the abnormal behaviors involving changes in the motion. The proposed method can be accurately estimated for only highly distinguishable regions. And one of the major problems of the proposed method is a real-time performance. As part of the future work, the authors plan to study the factors affecting the speed of the algorithm, and improve the performance of the algorithm on the real-time and the accuracy. Accuracy of many classification algorithms depend on the behavior of the dataset being analyzed. More accurate the results are, more frequently we can detect the suspicious activity and thus disaster or crime can be avoided not completely but can be prevented upto an extent.

## REFERENCES

- [1] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia, 'Event detection and analysis from video streams', IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no.8, pp.844–851, Aug. 2001.
- [2] Paulo Vinicius Koerich Borges, Nicola Conci, and Andrea Cavallaro, 'Video-Based Human Behavior Understanding: A Survey', IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 23, NO. 11, NOVEMBER 2013.
- [3] Weixin Li, Student, Vijay Mahadevan and Nuno Vasconcelos, 'Anomaly Detection and Localization in Crowded Scene', IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 36, NO. 1, JANUARY 2014.
- [4] Oluwatoyin P. Popoola and Kejun Wang, 'Video-Based Abnormal Human Behavior Recognition', IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 42, NO. 6, NOVEMBER 2012.
- [5] Chunyu Chen, Yu Shao, and Xiaojun Bi, 'Detection of Anomalous Crowd Behavior Based on the Acceleration Feature', IEEE SENSORS JOURNAL, VOL. 15, NO. 12, DECEMBER 2015.
- [6] Kaiqi Huang, Shiquan Wang, Tieniu Tan and Stephen J. Maybank, 'Human Behavior Analysis Based on a New Motion Descriptor', IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 19, NO. 12, DECEMBER 2009.
- [7] Geetanjali Vinayak Kale, Varsha Hemant Patil, 'A Study of Vision based Human Motion Recognition and Analysis', International Journal of Ambient Computing and Intelligence Volume 7 Issue 2 July- December 2016.
- [8] Dong-Gyu Lee, Heung-Il Suk, Sung-Kee Park and Seong-Whan Lee, 'Motion Influence Map for Unusual Human Activity Detection and Localization in Crowded Scenes', IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 25, NO. 10, OCTOBER 2015.
- [9] Gaurav Kumar and Pradeep Kumar Bhatia, 'A Detailed Review of Feature Extraction in Image Processing Systems', 2014 Fourth International Conference on Advanced Computing and Communication Technologies.
- [10] Louis Kratz and Ko Nishino, 'Anomaly Detection in Extremely Crowded Scenes Using Spatio-Temporal Motion Pattern Models', 2009 IEEE Conference on Computer Vision and Pattern Recognition.
- [11] FanJiang, Junsoung Yuan, Sotirios A.Tsaftaris and Aggelos K.Katsaggelos 'Anomalous video event detection using spatiotemporal context', March 2011 Volume 115 Issue 3 Computer Vision and Image Under [12]standing.