iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.59849

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Crowd Anomaly Detection Using Machine Learning

Dr. Kavuri. K.S.V.A. Satheesh[1], Sk. Waaheda Zeenathul Quraan[2], G. Mercy Jasper[3], Ch. Sai Teja[4], Ch. Sai Vignesh[5], V. Lohith[6]

[1]*Associate Professor,* [2, 3, 4, 5, 6]*B.Tech Students Department of Artificial Intelligence and Data Science, Vasireddy Venkatadri Institute of Technology, Guntur*

Abstract: ***Surveillance videos have become crucial objects for ensuring the safety and security of crowded places such as concerts, subways, airports, and many others. In recent times, the installation of security cameras has rapidly increased in both public and private places. The complexity of finding the anomaly will increase drastically as the amount of footage increases and there may be occlusions for the analysis of the video, which consumes both time and may result in false detections. To overcome these drawbacks, we have produced an approach using machine learning algorithms. In this approach, we use the background subtractor for the elimination of the static things in the video frames and concentrate on the dynamic things. Secondly, we extract the spatio-temporal features from the individual frame to increase the efficiency of the detection. This approach is evaluated using the UCSD Dataset, which comprises 50 training and 48 testing samples. The effectiveness of this approach has been testified against metrics such as F1 score, Recall, Precision, and Accuracy***

Index Terms: Surveillance videos, machine learning, spatiotemporal features, background subtractor

I. INTRODUCTION

The biggest challenges of anomaly identification in crowds are false positives, background clutter, and occlusion. The goal of recent technological developments has been to increase crowd anomaly detection's efficacy. Machine learning has recently advanced quickly in addressing these difficulties with the use of algorithms that can discern between normal and aberrant behavior and evaluate complex data patterns.

We have mainly used machine learning methods, such as the k-neighbors and decision tree classifiers, to improve the ability to distinguish between normal and aberrant behavior. The uploaded video frames' attributes are utilized to train the decision trees. The characteristics include human activities, spatiotemporal elements, etc. As a result, by training the classifier on various kinds of video frames, we may raise its performance. We are utilizing the K-Neighbors classifier, which can recognize patterns in the frames that contradict each other as well as the commonalities between typical video frames.

Our method additionally makes use of the background subtractor, which distinguishes between dynamic and static items individually. It is crucial for crowd anomaly identification because, when we remove the static elements from the video frames, we may reduce the amount of time needed. There will be many video frames with static and dynamic patterns.

II. LITERATURE REVIEW

Recently, computer vision has become interested in identifying dense and crowded areas. Problems such as classifying videos accor ding to crowded groups [1], estimating crowd sizes [2],

and determining the targets of individuals in the crowd are research topics of science. The most popular are color flow as ARG ET AL or other types of spatioLemoryporal Adam. [3] Use histograms to check the potential of the visual field. Kim and Grauman [4] u se a combination of PCA models to model local optical properties and use Markov random fields (MRFs) to maintain global consist ency. Mehran et al. [5] are inspired by classical group behavior studies, which use concepts such as relationships to describe the beh avior of groups. Boiman and Irani [6] used spatiotemporal patches and reported areas as disparities that could not be reconstructed u sing data from previous frameworks. Spatiotemporal gradients, whose statistics are modeled using integrated HMMs, were proposed in [7] to detect anomalies in crowded conditions. We identify three characteristics that an agent should have: 1) a common model of the appearance and quality of the crowd model; 2) physical detection capabilities; and

3) spatial anomalies. We then propose to use dynamic texture (DT) as a representation [8]. We then introduce DT-

based spatial and temporal normality models. Temporal normality is modeled using a combination of DT [9] (MDT), and spatial nor mality is measured using an MDTbased discriminative saliency detector [10].

©IJRASET: All Rights are Reserved

Basic facts and performance evaluation procedures regarding adverse events are provided. Finally, the vulnerability detection algorit hm is tested against the previous algorithm to create a set of security measures against which future algorithms can be compared.

III. PROPOSED METHODOLOGY

The methods we have employed for our approach are described in this section. Our approach consists of four steps. First, the input video frames undergo data preparation. Second, the spatiotemporal and optical flow patterns are extracted from the preprocessed data. Third, the k-neighbor and decision tree classifiers serve as the foundation for the model's definition. Fourth, the model is evaluated by running the trained model against the testing video frames.

A. Data Preprocessing

The UCSD dataset, which was collected via a stationary camera with several camera angles, is the primary dataset that we used for our project. It is made up of the Ped1 and Ped2 datasets.

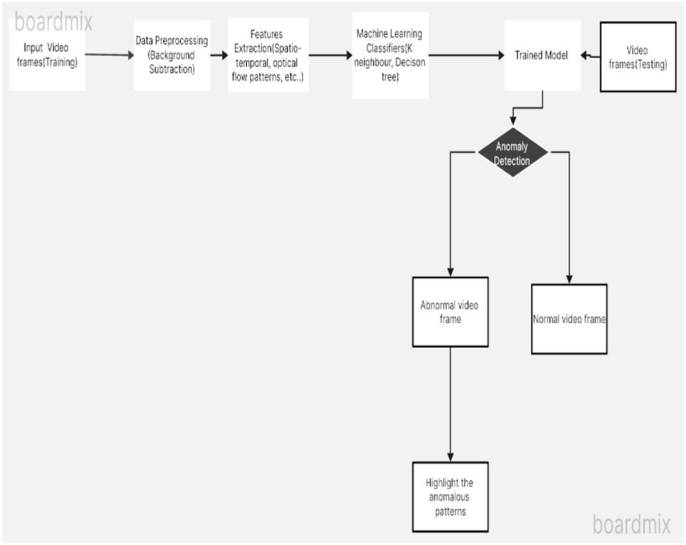


Figure –1: Methodology we followed for crowd anomaly detection

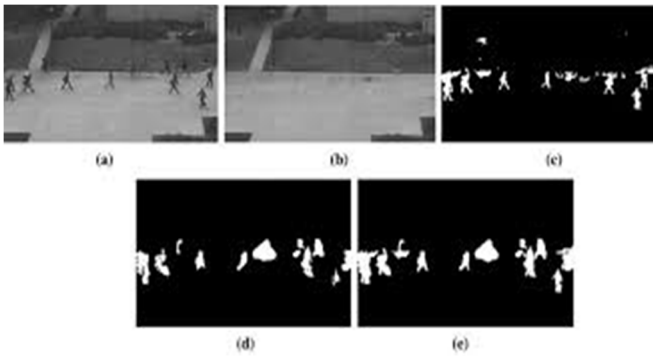


Figure 2: Images after background subtraction

B. Feature Extraction

Following the preprocessing stage, features including the optical flow, spatiotemporal, and histogram of oriented graphics (HOG) are extracted using the preprocessed data. The features that characterize optical flow calculate object motion between frames and produce an optical flow vector between the motions.

We can increase the range of genuine detections by using the spatio-temporal characteristics, which include information about each frame's spatial and temporal aspects. The process of extracting spatial characteristics from video frames involves the usage of HOG.

The OpenCV library is used for the feature extraction in our project which contains functions such as “cv2.calcOpticalFlowFarneBack()” for optical flow features and “cv2.HOGDescriptor()” for the HOG extraction.

C. *Defining Model*

In our model, we have used machine learning classifier models such as the decision tree classifier and k-neighbour classifier.

1) *Decision Tree Classifier:* Following the second stage of feature extraction, the retrieved features are fed into the classifier, which uses them to categorize the video frames and differentiate between normal and anomalous frames. While learning, the classifier differentiates the frames recursively according to their region, captured motion, and other factors, ultimately creating a decision tree. In the end, it has a decision tree with leaf nodes, each of which includes roots representing features and leaves representing flag values denoting normal or deviant behavior. For this distinction, the decision rules are the primary source of dependence for the decision tree classifier.
2) *K-Neighbours Classifier:* We have also employed k-neighbors in our project to improve our outcomes. K-Nearbors functions by identifying patterns in the video frames and calculating the distance for the "Euclidean Distance" feature. Any frame that is more than this Euclidean distance is regarded as abnormal.

We utilized the sci-kit-learn library, which has pre-defined models for the trees and neighbors, to create the classifier.

D. *Testing the model*

Using the preprocessed data from steps one through two, we trained our model in the third phase. We now use the testing movies in the dataset to test our trained model. The testing films are occasionally pre-processed, but they are also assessed against the model, and pre-process extraction must be done if necessary. The computation of accuracy and efficiency among several metrics, including accuracy, F1score, and recall, is highlighted in the results. Following testing, we can apply our model to video clips. When we provide our video sample, the model does the classification and, if any anomalies are found, the anomalous patterns are indicated as seen below with a white rectangular box.

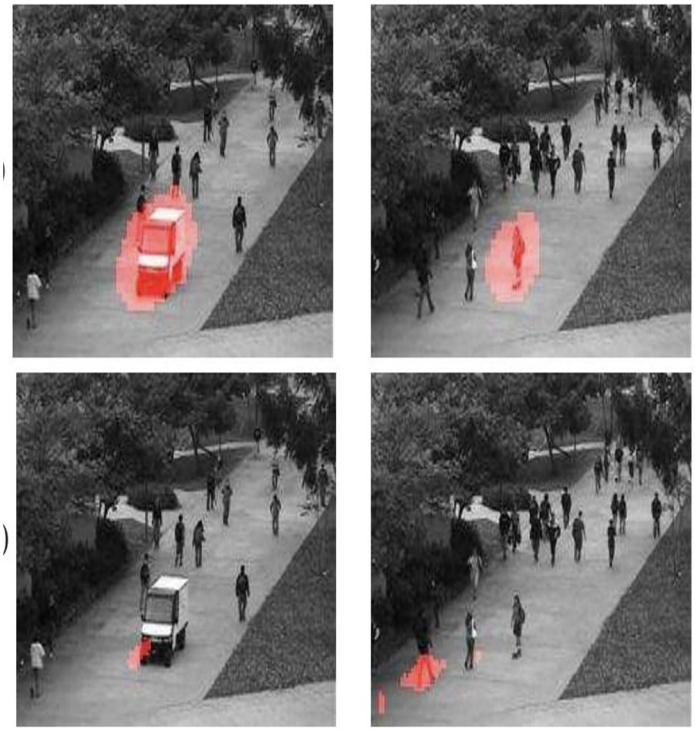


Figure 3 highlights the anomalous pattern in the video frames

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Using machine learning techniques like the k-neighbor and decision tree algorithms, which both rely on supervised learning, we have constructed our model. The UCSD Dataset contains labeled video clips, and the model is trained using these labeled datasets through N-supervised learning. When we input video frames, the model is supposed to identify any aberrant patterns in the footage and highlight them with a red rectangle box.

Evaluation:

We have evaluated our model against metrics such as recall, f1 score, and Precision. All these metrics have been evaluated both pixel-wise and frame-wise.

The maximum time consumption has been reduced by the usage of the background subtractor which decreases the pre-processing time by about 15%. The extraction of temporal and spatial features has increased accuracy by about 10%.

Evaluation Results	
Precision	0.791
Recall	0.157
F1	0.262
C	
Results pixel-wise	
EER	0.29
RD	0.70
True positive	1324.0
False Positive	2106.0
Precision	0.38
Recall	0.58
F1	0.46

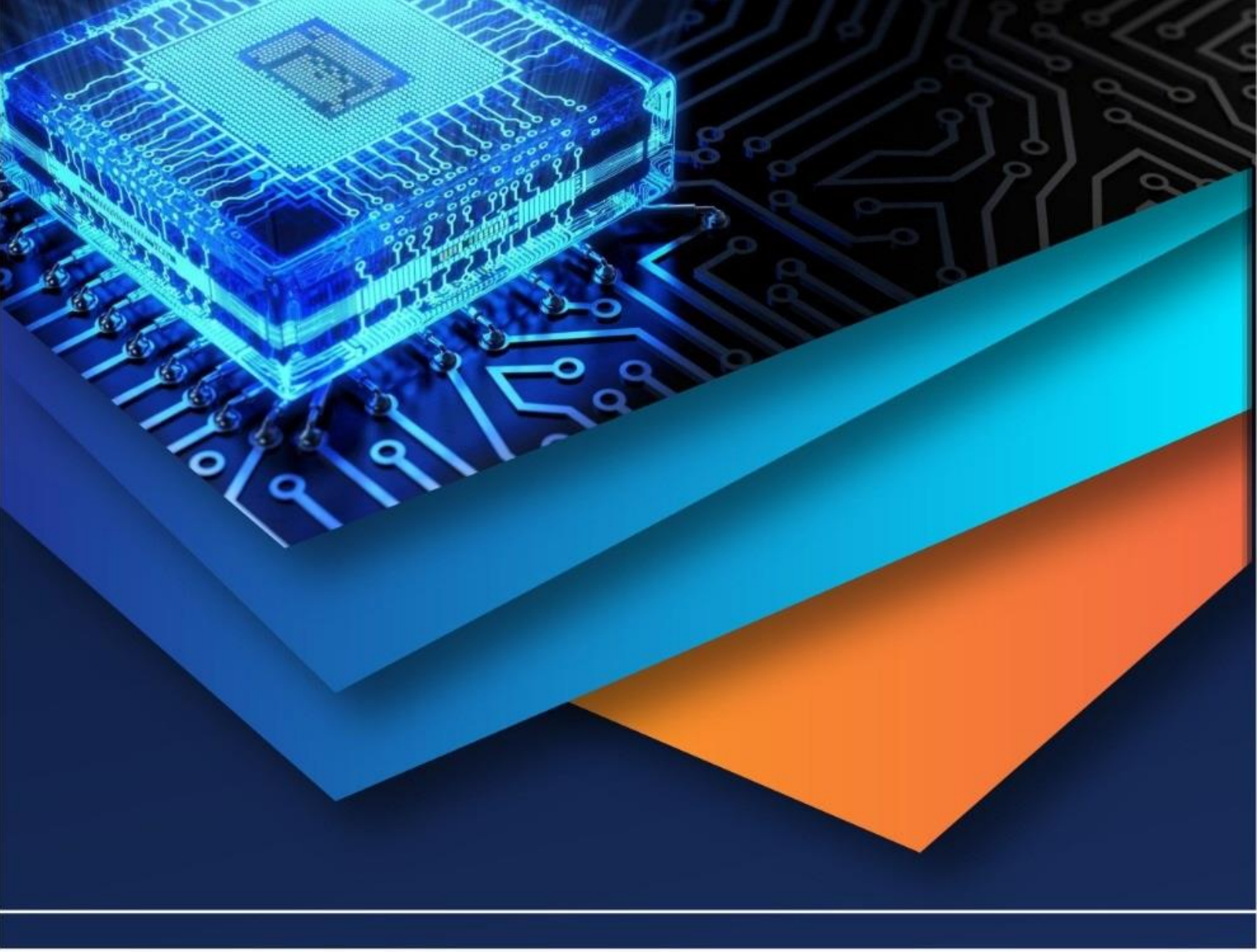
The above table shows the evaluation results of our model against the metrics pixel-wise.

V. CONCLUSION

We developed a model that makes use of machine learning techniques like decision trees and k-neighbor classifiers. We tested our technique on the UCSD dataset, which is a benchmark dataset. One of the main factors in the time consumption decrease has been the use of background subtractors. The classifiers perform tasks on the entire dataset, look for patterns in the video clips, and construct a decision tree with features at the root and final values representing normal and abnormal. We documented the outcomes of both pixel-wise and frame-wise evaluations of our model using measures including precision, recall, and F1 score.

REFERENCES

[1] S..Ali and. Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In CVPR, pages 1–6, 2007.
[2] D. Kong, D. Gray, and H. Tao. Counting pedestrians in crowds using viewpoint invariant training. In BMVC,2005.
[3] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. PAMI, 30(3):555–560, March 2008.
[4] J. Kim and K. Grauman. Observe locally, infer globally: A space-time mrf for detecting abnormal activities with incremental updates. In CVPR, pages 2921–2928, 2009.
[5] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using the social force model. In CVPR, pages 935–942, 2009. A. Adam, E. Rivlin, Shimshoni and D. Reinttiz "Robust Real-Time Unusual Event Detection", March 2008
[6] O. Boiman and M. Irani. Detecting irregularities in images and video. IJCV, 74(1):17–31, August 2007.
[7] L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In CVPR09, pages 1446–1453, 2009.
[8] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto. Dynamic textures. IJCV, 51(2):91–109, 2003.
[9] A. B. Chan and N. Vasconcelos. Modeling, clustering, and segmenting video with mixtures of dynamic textures. PAMI, 30(5):909–926, May 2008.
[10] D. Gao and N. Vasconcelos. Decision-theoretic saliency: computational principle, biological plausibility, and implications for neurophysiology and psychophysics. Neural Computation, 21:239–271, Jan 2009



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)