IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 4.819

Volume 11, Issue 2, November 2020

Monitoring Crowd Movement for Anomaly Detection Using Scale Invariant Feature Transform

Kaliram Perumal[1], Bharathi Subramaniam[2], Madhavi Nachimuthu[3], Gowrison Gengavel[4]
Assistant Professor, Department of ECE[1,3]
Assistant Professor(Sr), Department of ECE[2,4]
Institute of Road and Transport Technology, Erode, India

Abstract: *In order to detect the potentially dangerous arrears and the situation of a crowd in public security systems, the automated analysis of crowd monitoring using surveillance video is playing vital role***.***Even though many works are focused on the analysis related to the crowd behavior analysis, complexity in algorithm, real time working module and predefined rigid automatically selection rules are the major problems in the behavior analysis crowd detection. This work proposed a real time algorithm to detect the global anomalies in Scale Invariant Feature Transform(SIFT) based on holistic approach. Significantly deviation in the normal behavior from the previously stored data set,that is people running away from the crowd or suddenly gathering into a particular point were consider an the anomalies are the framework of the approach. The experimental result shows that, compared with the existing methods, the proposed method could able to run in real time and have less complexity in algorithm.*

Keywords: computational complexity, video surveillance, real-time global anomaly detection, SIFT, Anomaly detection

I. INTRODUCTION

Crowd analysis has become a significant interest in public and academic disciplines as a result of population development. Crowd assessments are used to create crowd control techniques for public space architecture, visual monitoring, public events as well as interactive environments that make places more comfortable to avoid crowd-related disaster.

Certain crowds are more difficult to study than others and how a crowd has broken up and observed is influenced by its psychological factor. Crowds may be informal, such as a group of pedestrians crossing the street or it may be formal such as those taking part in a marathon or a rally. It may be as passive as a crowd or as active and unpredicted as one. When the majority of the analyses are based on the main crowds, anomalies such as anyone blocking traffic or a vehicular passing through a group of walkers must be taken into accounts. Structured crowd, semi structured crowd and unstructured crowds are the three major styles of crowds that exists. The pedestrians in a structured crowd are identical to one another based on certain standard criteria called attributes and it is also known as organized crowd.A organized crowd could be seen in our daily lives in the form of a maven history of solders or an athlete running in a race. On the other hand, semi structured crowd are defined as a crowd in which pedestrians move in various direction at different times and finally the unstructured crowd is defined as a crowd with completely random irregular movement in which pedestrian movement is completely zigzag and can be clubbed or parameterized on a common scale.

The proposed approach focuses on crowd anomaly detection, taking into account irregular movement detection and surveillance video obtained from public space.

II. RELATED WORK

Anomaly detection attempts in recent years have mostly focused on hand crafted features and deep learning based approaches. In order to detect general anomalies, the hand-grafted function is commonly used in anomaly detection. The most widely used approach include mixture of dynamic texture models,social force models, optical flow based

DOI: 10.48175/IJARSCT-6558B

video descriptors, HMM on local spatial volumes and temporal analysis of gray level concurrence matrix.Despite their ability to identify general abnormalities, these techniques are unsuitable for practical applications due to their poor detection accuracy. Sparse representation and dictionary learning used sparse representation to learn the dictionary of normal behavior and reconstruction error to detect abnormal behavior.

Then, Lu et al. improved Zhao et al.'s approach by increasing the speed to 150 fps. Though, nothing in the dictionary was trained with abnormal events and was usually over complete.The expectation for anomaly detection cannot be guaranteed in real world conditions.

DeepLearning-Based Approaches. Recently, deep learning has attained significant success in object detection and video understanding. Thus, researchers attempted to detect anomalies in videos using deep learning. As a result of insufficient abnormal patterns, mostresearchersbuild a generative model of ordinary patterns and then detect anomalies based on the reconstruction errors of input data.AEs are commonly used to build generative models. An et al. used variational AEs for anomaly detection by evaluating reconstruction probability but not reconstruction errors. Dimokranitou et al. used adversarial AEs for robust anomaly detection. Raghavendra et al. used robust AEs to improve the generalization and induction of AEs in anomaly detection. Hasan et al. used convolutional AEs (CAEs) on stacked spatiotemporal frames to detect anomalies and better encode temporal variations. Chong et al. and Luo et al. proposed a similar architecture named LSTM based on CAEs (LSTM–CAEs); this approach can better capture temporal variants through LSTM. Except for AEs, GANs can also be used to build such a model. Schlegl et al. employed a GAN model for anomaly detection in medical images. Ravanbakhsh et al. proposed a cross-channel prediction conditional GAN architecture to fully utilize raw image pixels and corresponding optical flows. Compared with CAE-based anomaly detection approaches, the GAN-based anomaly detection approach reports large area under curve (AUC) and low equal error rate in benchmarking datasets Ped1 and Ped2. However, the preceding methods mentioned are easily influenced by backgrounds. Thus, the anomaly detection enactment is unsatisfactory.

Kumarganesh et al. (2018), suggested an ANFIS classifier approach used to recognize faulty portions from foundation images with an accuracy of 96.0% [6]. An ANFIS classifier technique was suggested by Kumarganesh et al. (2016) for the classification of imperfect portions from the basic images, and it achieved 97.63% Accuracy (Ac) [7]. Kumarganesh et al. (2014), proposed a JOSE-Encodingtechnique to achieved the detection accuracy level is 93% [8]. Kumarganesh et al. (2015) proposed a system using HEVC standard for competent health data and medicinal video information compression [9].

III. PROPOSED METHOD

The method proposed to detect the anomalies in the moving crowd using the Scale Invariant Feature Transform (SIFT).

Optical flow is a technique to notice moving objects during a sequence of frames and the vector position of picture elements is calculated and compared in sequence of frames for the pixel position. Normally the motion is pictured as the vector position of pixels.

The method of locating the moving object in a sequence of frames is performed by victimization the feature extraction of objects and police investigation the objects in a sequence of frames. By victimization, the position values of objects in each frame are able to calculate the position and rate of the moving object.

DOI: 10.48175/IJARSCT-6558B

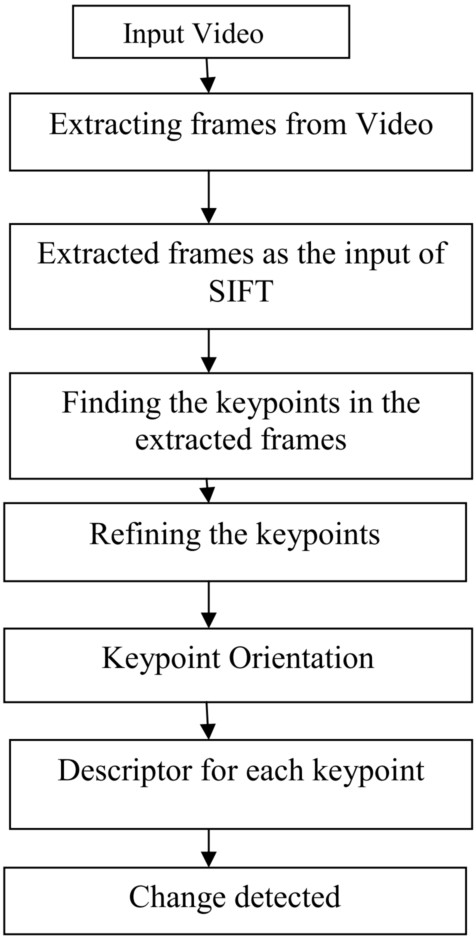

Input Video
Extracting frames from Video
Extracted frames as the input of SIFT
Finding the keypoints in the extracted frames
Refining the keypoints
Keypoint Orientation
Descriptor for each keypoint
Change detected

Fig.1 Flowchart for Growed movement for anomaly detection

IV. ALGORITHM USED

The scale-invariant feature transform (SIFT) is a feature detection algorithm in computer vision to detect and describe local features in images. SIFT robustly identify objects, even among the clutter and under partial obstruction, because the SIFT feature descriptor is invariant to uniform scaling, orientation, illumination changes and partially invariant to affine distortion.There are four steps in algorithm.

Step 1: Approximate Keypoint Location
Step 2: Refining Keypoints
Step 3: Assigning Orientation
Step 4: Descriptors for each Keypoint

4.1 Approximate Keypoint Location

The first stage is to construct a Gaussian "scale space" function from the input image. This is formed by convolution (filtering) of the original image with Gaussian functions by varying widths. The difference of Gaussian (DoG), $D(x, y, \sigma)$, is calculated as the difference between two filtered images, one with k multiplied by the scale of the other.

$$D(x, y, \sigma) = L(x, y, k\sigma)-L(x, y, \sigma)$$

The images of $L(x, y, \sigma)$, are produced from the convolution of Gaussian functions, $G(x, y, k\sigma)$, with an input image, $I(x, y)$.

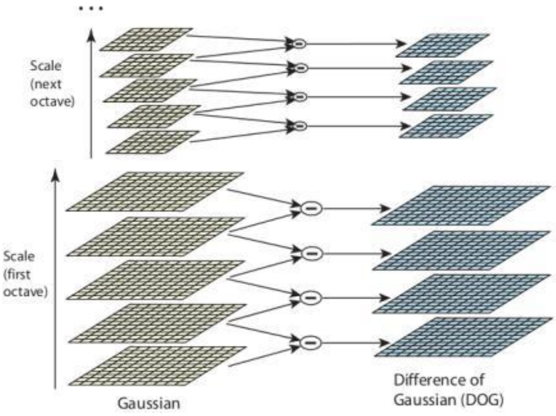
$$L(x, y, \sigma) = G(x, y, \sigma)*I(x, y)$$

First, the initial image, I, is convolved with a Gaussian function, G_0, of widthσ_0. Then it uses the blurred image, L_0, as the first image in the Gaussian pyramid and incrementally convolve it with a Gaussian, Gi, of width σi to create the i^{th}

DOI: 10.48175/IJARSCT-6558B

image in the image pyramid, which is equivalent to the original image filtered with a GaussianG_k of width $k\sigma_0$.The effect of convolving with two Gaussian functions of different widths is most easily found by converting to the Fourier domain, in which convolution becomes multiplication.

Next stage is to find the extrema points in the DOG pyramid. To detect the localmaxima and minima of D(x, y, σ), each point is compared with the pixels of all its 26 neighbors. If the value is the minimum or maximum then the point is an extrema. Further, it improves the localization of the keypoint to sub pixel accuracy, by using a second order Taylor series expansion.



4.2 Refining Keypoints

The second stage attempts to eliminate some points from the candidate list of keypoints by finding those that have low contrast or are poorly localized on an edge.To eliminate poorly localized extrema it uses the fact that in these cases there is a large principle curvature across the edge, but a small curvature in the perpendicular direction in the difference of Gaussian function. A 2x2 Hessian matrix, H, computed at the location and scale of the keypoint is used to find the curvature.

After that, to discard the keypoints with low contrast, the value of the second-order Taylor expansion {\displaystyleD ({\textbf {x}})} is computed at the offset {\displaystyle {\hat {\textbf {x}}}}. If this value is less than {\displaystyle 0.03}, the candidate keypoint is discarded. Otherwise it is kept, with final scale-space location {\displaystyle {\textbf {y}} + {\hat {\textbf {x}}}}, where {\displaystyle {\textbf {y}}} is the original location of the keypoint.

4.3 Eliminating edge responses

Finding these principal curvatures amounts to solving for the eigenvalues of the second-order Hessianmatrix, H:{\displaystyle{\textbf{H}}={\begin{bmatrix}D_{xx}&D_{xy}\\D_{xy}&D_{yy}\e The DoG function will have robust responses along edges, even if the candidate keypoint is not strong to small amounts of noise. Therefore, in order to improve the stability, it is necessary to eradicate the keypoints that have inadequate determined locations but have high edge responses. For the inadequate defined peaks in the DoG function, the principal curvature across the edge would be much larger than the principal curvature along it.

The eigenvalues of H are proportional to the principal curvatures of D. It turns out that the ratio of the two eigenvalues, say {\displaystyle \alpha} is the larger one, and {\displaystyle \beta} the smaller one, with ratio {\displaystyle r=\alpha /\beta}, is sufficient for SIFT's purposes. The trace of H, i.e., {\displaystyle D_{xx}+D_{yy}}, gives us the sum of the two eigenvalues, whereas its determinant, i.e., {\displaystyleD_{xx}D_{yy}-D_{xy}^{2}}, yields the product.Theratio {\displaystyle {\text{R}}=\operatorname {Tr} ({\textbf {H}})^{2}/\operatorname {Det} ({\textbf {H}})} can be shown to be equal to {\displaystyle(r+1)^{2}/r}, which depends only on the ratio of the eigenvalues rather than their individual values. R is minimum when the eigenvalues are equal to each other. Therefore, the higher absolute difference between two eigenvalues, which is equivalent to a higher absolute difference between two principal curvatures of D, higher the value of R. In that case, for some threshold eigenvalue ratio {\displaystyle r_{\text{th}}}, if R for a candidate keypoint is larger than {\displaystyle (r_{\text{th}}+1)^{2}/r_{\text{th}}}, that keypoint is poorly localized and hence rejected.

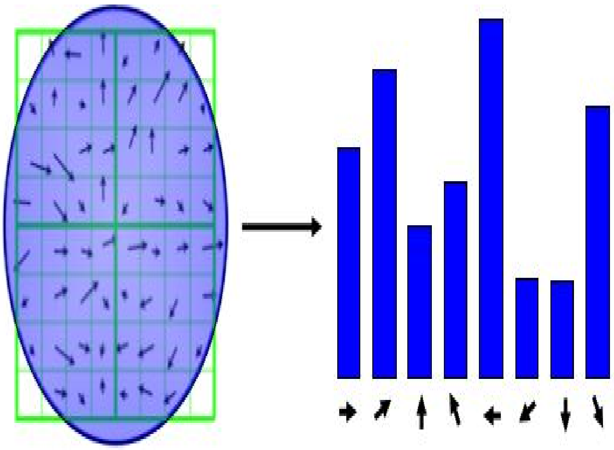
The new approach uses {\displaystyle r_{\text{th}}=10}. This processing step for suppressing responses at edges is a transfer of a corresponding approach in the Harris operator for corner detection. The difference is that the measure for thresholding is computed from the Hessian matrix instead of a second-moment matrix.

4.4 Assigning Orientation

In this step, each keypoint is assigned one or more orientations based on local image gradient directions. This is the key step in achieving invariance to rotation as the keypoint descriptor which can be represented relative to this orientation and therefore achieves invariance to image rotation.

First, the Gaussian-smoothed image {\displaystyle L\left(x,y,\sigma \right)} at the keypoint's scale {\displaystyle \sigma } is taken so that all computations are performed in a scale-invariant manner. For an image sample {\displaystyle L\left(x,y\right)} at scale {\displaystyle \sigma},the gradient magnitude, {\displaystyle m\left(x,y\right)}, and orientation, {\displaystyle \theta \left(x,y\right)}, are computed using pixel differences:{\displaystyle m\left(x,y\right)={\sqrt{\left(L\left(x+1,y\right)-L\left(x-1,y\right)\right)^{2}+\left(L\left(x,y+1\right)-L\left(x,y-1\right)\right)^{2}}}}

{\displaystyle\theta\left(x,y\right)=\mathrm {atan2} \left(L\left(x,y+1\right)-L\left(x,y-1\right),L\left(x+1,y\right)-L\left(x-1,y\right)\right)}. The magnitude and direction calculations for the gradient are done for each pixel in a neighboring region around the keypoint in the Gaussian-blurred image L. An orientation histogram with 36 bins is molded, with each bin covering 10 degrees. Each sample in the neighboring window added to a histogram bin is weighted by its gradient magnitude and by a Gaussian-weighted circular window with a {\displaystyle \sigma} that is 1.5 times that of the scale of the keypoint. The peaks in this histogram correspond to dominant orientations. Once the histogram is filled, the orientations corresponding to the highest peak and local peaks that are within 80% of the highest peaks are assigned to the keypoint. In the case of multiple orientations being assigned, an additional keypoint is created having the same location and scale as the original keypoint for each additional orientation.

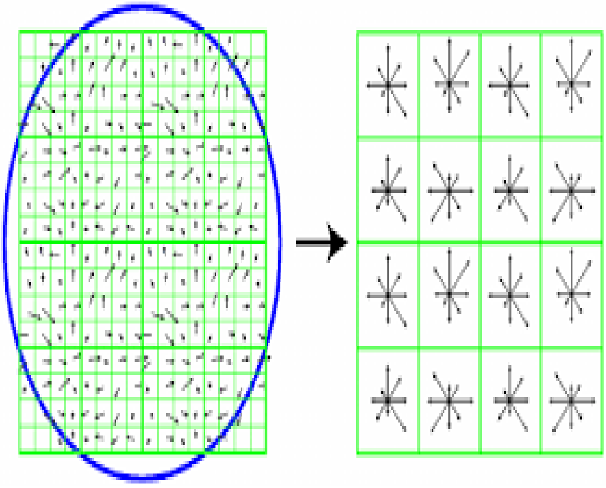


4.5 Descriptors for each Key point

The assignment orientation procedure found keypoint locations at particular scales and assigned orientations to them, which ensured invariance to image location, scale and rotation. Now we want to compute a descriptor vector for each keypoint such that the descriptor is highly distinctive and partially invariant to the remaining variations such as illumination, 3D viewpoint, etc. This step is performed on the image closest in scale to the keypoint's scale.

First a set of orientation histograms is created on 4×4 pixel neighborhoods with 8 bins each. These histograms are computed from magnitude and orientation values of samples in a 16×16 region around the keypoint such that every histogram contains samples from a 4×4 subregion of the original neighborhood region. The magnitudes are further weighted by a Gaussian function with {\displaystyle \sigma } which is equal to one half the width of the descriptor window. The descriptor then becomes a vector of all the values of these histograms. Since there are 4 × 4 = 16 histograms each with 8 bins the vector has 128 elements. This vector is then normalized to unit length in order to improve invariance to affine changes in illumination. To reduce the effects of non-linear illumination a threshold of 0.2 is applied and the vector is again normalized. The threshold process, also referred to as clamping, improves matching results even when non-linear illumination effects are not present. The threshold of 0.2 was empirically selected, and by replacing the fixed threshold with one systematically calculated to improve matching results.

Although the dimension of the descriptor, i.e. 128, seems high, descriptors with lower dimension than this don't perform as well across the range of matching tasksand the computational cost remains low due to the approximate BBF method used for finding the nearest neighbor. Longer descriptors continue to improve, but increasing to higher value which leads to increase in sensitivity to distortion and occlusion. It is also shown that feature matching accuracy is above 50% of viewpoint changes of up to 50 degrees.



Therefore, SIFT descriptors are invariant to minor affine changes. To test the distinctiveness of the SIFT descriptors, matching accuracy is also measured against varying number of keypoints in the testing database, and it is shown that matching accuracy decreases only very slightly for very large database sizes, thus indicating that SIFT features are highly distinctive.

V. CONCLUSION

The result of the proposed mechanism showed that it can able to detect the multiple moving objects in the dynamic video frame range used for experimentations. To detect the moving crowds, these frames are given as the input to the SIFT algorithm and it used to detect the abnormal activities in the moving crowd compared to the existing method efficiently.

REFERENCES

[1]. David G. Lowe (1999)."Object recognition from local scale- invariantfeatures", August 2002, Proceedings of the Seventh IEEE International Conference on Computer Vision pp 1-8

[2]. David G. Lowe (2001) “Local feature view clustering for 3D object recognition.” IEEE Conference on Computer Vision and Pattern Recognition,Kauai, Hawaii, 2001, pp. 682-688.

[3]. Laptev, Ivan &Lindeberg, Tony (2004), "Local descriptors for spatio-temporal recognition" .International Workshop on Spatial Coherence for Visual Motion Analysis SCVMA2004,Springer Lecture Notes in Computer Science, pp 91-103

[4]. Cui, Y.; Hasler, N.;;Thormaehlen, T.; Seidel, H.-P. (July 2009). "Scale Invariant Feature Transform with Irregular Orientation Histogram Binning" International Conference on Image Analysis and Recognition (ICIAR 2009).Springer.pp 258-267

[5]. Lindeberg, Tony (1998). "Feature detection with automatic scale election". International Journal of Computer Vision, Volume30, pages79–116.

[6]. Kumarganesh S, Suganthi M. An Enhanced Medical Diagnosis Sustainable System for Brain Tumor Detection and Segmentation using ANFIS Classifier. Current Medical Imaging Reviews 2018; 14(2): 271-279. DOI: 10.2174/1573405613666161216122938.

[7]. Kumarganesh S, Suganthi M. "An Efficient Approach for Brain Image (Tissue) Compression Based on the Position of the Brain Tumor" International Journal of Imaging System and Technology 2016; 26(4): 237-242. doi.org/10.1002/ima.22194.

[8]. Kumarganesh S, Suganthi M, "Efficient Lossless Medical Image Compression Technique for Real World Applications Using JOSE-Encoding" International Journal of Applied Engineering Research 2014; 9 (24): 24625-24640.

[9]. Kumarganesh S, M. Suganthi, 2015, "Efficient Medical Data and Medical Video Compression Using HEVC Standard" – International Journal of Advanced Science and Engineering ISSN: 2349-5359, Vol. 1 No. 3 PP 27-31.