International Journal of Engineering Research and Advanced Technology (IJERAT)

DOI: 10.31695/IJERAT.2022.8.8.3

E-ISSN : 2454-6135

Volume.8, No. 8
August - 2022

Subject Review: Image Processing Techniques for Object Tracking in Video Surveillance

Ahmed Othman Khalaf[1], Shaimaa Khudhair Salah[2]

[1,2]Assistant Professor

Department of Computer Science

Collage of Education,

University of Mustansiriyah

Iraq

ABSTRACT

The article is devoted to real-time algorithms for detecting events described by four scenarios: movement in a forbidden direction, being in a sterile zone, leaving (stealing) an object, throwing an object. The main idea of the algorithms is the analysis of the trajectories of moving objects, for which two different approaches are proposed in the article.

Key Words: Digital Image Processing, Video Analytics, Video Surveillance Systems, Object Detection, Object Tracking.

1. INTRODUCTION

One of the main trends in the video surveillance market is to reduce the cost of cameras while improving image quality. These factors lead to the spread of video surveillance systems. Surveillance video camera operators perform monotonous work for a long time, which causes the problem of fatigue and loss of concentration. For example, with continuous monitoring for 12 minutes, the operator begins to miss 45% of potentially disturbing events, and if the monitoring time is increased to 22 minutes, then the percentage of missing increases to 95 [1]. Therefore, there is a need for video analysis without direct human participation.

Video analytics is a technology based on digital image processing and machine vision methods for automated extraction of information from video materials. The scope of automated video surveillance systems is very extensive. They are used to monitor, protect and ensure the safety of critical facilities, which include transport infrastructure facilities (railway stations, ports and airports), engineering infrastructure facilities (power plants and water supply systems), etc. In addition, video analytics systems are used at commercial and private facilities (trade, office, warehouse).

Automated video surveillance is used to solve a wide range of problems, for example: detection of the intersection of a given line, fire detection, recognition of car numbers, counting the number of store visitors.

Transport Security and Rules for Mandatory Certification of Technical Means for Ensuring Transport Security” 1.

1. Movement in a forbidden direction - a scenario of a situation in a recorded scene, according to which the movement of an object in a forbidden direction relative to conditionally specified boundaries is considered alarming.

2. Sterile zone - the appearance of an object in the field of view of the camera, its intersection with a conditionally specified forbidden line or being in the forbidden zone. At the same time, the object in these two scenarios can be a person, a vehicle, or an animal.

3. Abandoned (disappeared) object - leaving objects in the camera's field of view (or a zone limited by conditional lines) or the disappearance of an object that was previously in the camera's field of view.

4. Thrown object - throwing an object into the camera's field of view.

2. OVERVIEW OF EXISTING ALGORITHMS

Among the tasks of digital image processing, two can be distinguished related to the movement of objects: the definition of moving objects and the construction of their trajectories. When solving these problems, two types of video images are distinguished: those taken with a fixed and moving camera, for which there are their own methods. Next, methods for detecting moving objects and constructing the trajectories of their movement will be considered. The considered algorithms use data obtained from a stationary camera.

2.1. Moving object detection

The task of detecting moving objects in a video image is one of the classic tasks of computer vision. The initial data of the problem is a sequence of images, the output data is the coordinates and sizes of detected objects. The traditional method for detecting motion in an image is to calculate the interframe difference, the absolute difference between two successive images from a video stream [2]. The result is an image with highlighted areas of motion. Most often, it is difficult to highlight the contours of an object in such an image; this can only be done if the object is solid and convex.

To detect moving objects, you can use the background subtraction method, which calculates the absolute difference between the current frame and a previously calculated background image. The efficiency of object detection largely depends on the method of calculating the background image, but even when using primitive methods, moving objects will be completely selected on the resulting difference image. There are many background calculation algorithms (for example, GMM [3], MOG [4]), which differ in time and quality of work.

Recently, machine learning methods have been used to select objects. To do this, you need to collect a training sample and adjust the classifier. To detect objects, the video image frame is divided into a set of smaller images, after which each of them is fed to the input of the classifier, which, based on the calculated features, will determine whether the object contains the image or not.

According to the described method, many systems have been built (for example, HOG + SVM [5], DPM [6], CNNs [7]), which show high performance. But the process of training and tuning these systems is quite laborious and requires a well-organized training set. For each type of object (for example, a person or a car) and camera viewing angle, it is necessary to collect an individual training set and adjust the algorithm. Only under these conditions, the systems demonstrate a high quality of object detection.

2.2.Object Tracking

The task of tracking objects is to build trajectories of movement of given objects based on a sequence of images. In this case, the sequence of images can be supplemented with various information: objects found on previous frames, the speed and direction of their movement.

The block matching algorithm [8] divides the previous and current frames into non-intersecting macroblocks and compares blocks from the previous frame with neighboring blocks on the current frame. The best match determines the macroblock's movement vector from one location to another. Thus, by calculating the movement of all macroblocks, the overall picture of the movement of objects in the frame is determined. The advantage of the method is the relatively high speed of operation, the disadvantage is the instability to rotations and other image distortions.

Some generalization is the method of calculating the optical flow [9], which finds the shift of pixels, which provides a more accurate determination of the direction of movement of the object. The optical flow can be dense or sparse. In the first case, the optical flow calculates the displacements of each pixel, in the second case, the selected set of pixels. The application of the method requires the fulfillment of two conditions: the intensity of the image of the object does not change over time, the nearest points of the object move at a similar speed.

Wavelet transform is used to track small objects on the scene [10]. The specified method finds an object regardless of whether it is moving or stationary, but the method is able to detect the movement of only one object.

An object can be tracked using its location and movement speed, for which the Kalman filter is used [11]. The algorithm consists of two stages: state prediction and correction of the predicted value. The advantage of the method is the ability to predict the trajectory of an object if it was blocked for a short time, for example, by another object; the disadvantage is the ability to track only trajectories without a sharp change in direction of movement.

Another method is to highlight the features of objects and pairwise comparison of features of objects from the previous and current frames. The speed and accuracy of the method is determined by the methods of calculating the features of objects and the metrics that determine their similarity. Color histograms [12], textures [13], and key points [14] are used as features.

2.3. General scheme of the algorithm

The algorithms proposed by us are a set of related modules. During operation, the module waits for input data, after which it performs calculations and generates output data, which will subsequently be passed to other modules. A set of modules can be

DOI : 10.31695/IJERAT.2022.8.8.3

configured separately for each of the cameras or a group of cameras, depending on the tasks to be solved. The application of this approach allows you to create an easily scalable and distributed system.

The moving object tracking algorithm developed by us consists of two modules: a moving object detector and an object tracking module. A separate module analyzes the movement of objects and generates a message about an event that is sent to the operator of the automated video surveillance system.

3. MOVING OBJECT DETECTOR

The detector of moving objects sequentially accepts one frame at an input, which is a matrix

F= ($F_{X,Y}$) of size W * H Matrix element - value

image intensity at the point (x, y).

At the output, the detector issues a list of detected objects - a structure containing the coordinates of the object in the image, its size and a set of features necessary to identify the object in the next frames. The object detector is based on the background subtraction method. Bi background update based on the i-th frame of the video stream:

going on

$$B_i = \begin{cases} F0, i = 0, \\ (1-\alpha)Bi - 1 + \alpha Fi, \ otherwise. \end{cases}$$

The parameter α is called the learning factor, it determines the influence of the current frame to update the background. The value of the parameter must be between 0 and 1. If set to 0, the background will not be updated, and if set to 1, the calculated background will match exactly. give with the previous frame.

During the experiments, it turned out that for some values of the learning coefficient the algorithm does not have time to take into account changes in the illumination of the frame, for example, in cases where when the sun hides behind clouds or sets below the horizon. And as the parameter increases against the calculated background, objects appear that have stopped for a short period of time. neither. For individual cases, it was possible to choose a parameter at which the described problems we are missing, however, of universal significance when the algorithm works correctly on a lot of videos, could not be found. To solve this problem, we proposed additionally calculate the background with a different learning coefficient $\alpha' > \alpha$, which we mean B'.

At the first stage, the background of the image is calculated. The first step is to calculate two images background Bi and Bi' with learning coefficients α and α' respectively. second step find areas with changed lighting, for which we calculate the difference

$$D' = | B_i - B_i' | ,$$

on which, in addition to the desired areas, stopped objects will be highlighted. A distinctive feature of areas with changed illumination is the sample variance of the pixel intensity close to zero, so the detection of such areas can be carried out as follows:

$$m'_{x,y} = \begin{cases} 1, & \sigma\left(\bigcup_{(x',y')\in N_q(x,y)} d'_{x',y'} \right) < \varepsilon, \\ 0, & \text{otherwise.} \end{cases}$$

где $N_q(x_0, y_0) = \{(x, y), \|x - x_0| \le q \text{ и } |y - y_0| \le q\}$ – point neighborhood (x_0, y_0) size q,

$$\sigma(P) = \frac{1}{|P|}\sum_{p\in P}(p)^2 - \left(\frac{1}{|P|}\sum_{p\in P} p\right)^2 \text{ – sample variance}$$

P. The third step you-

DOI : 10.31695/IJERAT.2022.8.8.3

International Journal of Engineering Research and Advanced Technology (IJERAT)

DOI: 10.31695/IJERAT.2022.8.8.3

E-ISSN : 2454-6135

Volume.8, No. 8
August - 2022

we count the background B as follows:

$$B=(1-M')B_i+M'F_i,$$

where $M'=(m'_{x,y})$, calculated at the previous step, and matrix multiplication is performed according to elementally.

At the second stage, objects are detected, for which the difference is calculated $D = |F_i - B|$ value and a binary image is built with selected areas of motion along the right

$$mx,y=\begin{cases} 1, d_{x,y} \geq \mathrm{t}, \\ 0, \ \text{otherwise} \end{cases}$$

Further, morphological transformations are applied to the resulting binary image - opening and closing. Contours are selected on the resulting binary image, the sizes and coordinates of objects are determined, and histograms of object images are calculated.

4. OBJECT TRACKING MODULE

The object tracking module takes the detected objects as input, and outputs information about their movement: the start and end time of tracking, the size of the object, location and direction of movement. The module works with a structure that we call the track. Track t is a list of objects t- Э to (O_1,..., O_n) found on the sequence solid frames containing the same set of physical objects.

Let's take an example. Let two people walk in the frame, and they correspond to tracks with identify by the fixators A and B. When the trajectories intersect, one person will overlap the other, because of which the detector will select one object. Such combined objects should keep in the new track, for example, with the identifier AB. After the closure, people will detected separately and two new tracks will be created with IDs A' and B' (Fig. 1).

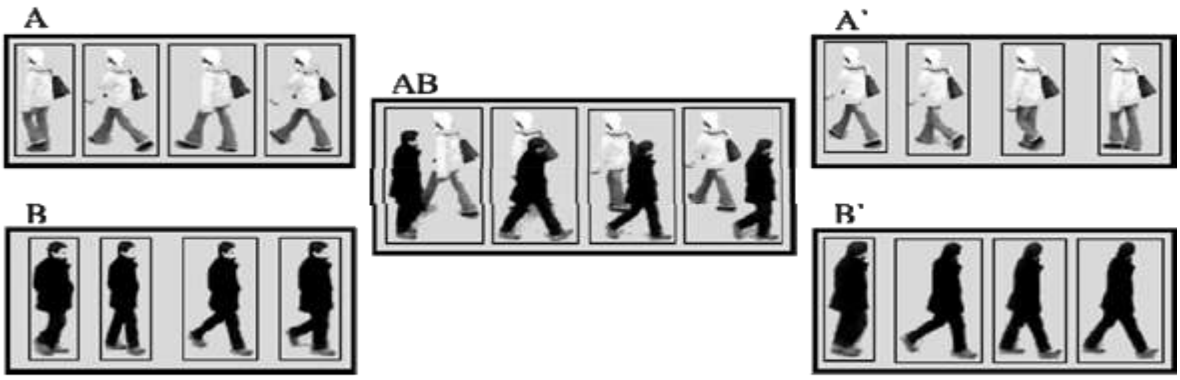

Fig. 1. Operation of the modules for detecting and tracking objects

At the first stage, the module creates new tracks and updates the existing ones by analyzing the distance between objects detected in the previous and current frames. Denote the set of previously discovered tracks T, c with every track t_i □ T can one-to-one match o_i is the object detected on the previous frame, a multiple of whose identity will be denoted O. Through O' we denote the set of objects found in the thick frame. Let's build a bipartite graph

$$G=(O,O',E),$$

Where: $E=\{(o_i,o'_j)\ o_i\in O,\ o'_j\in O',\ d(o_i,o'_j)<d_{\min}\}$, a $d(o_i,o'_j)$ and o'_j. We add an object o'_j to track t_i, if $(o_i,$ $o'_j)\in E$ and $\deg(o_i)=\deg(o'_j)=1$.

then we say that track ti , parent track t'_j , or t'_j is the child of the track t_i . For example, tracks A and B in Fig.1 are parents track AB, and tracks A' and B' as children of the track AB.

At the second stage, the split tracks are identified. Histogram analysis is performed to determine the similarity of tracks. The track is matched with a histogram equal to the sum of the histograms of the objects contained in the track. The similarity of two histograms is determined through the Bhattacharya distance:

$$d_B\left(H_p, H_c\right) = \sqrt{1 - \frac{\sum_i \sqrt{H_p(i) H_c(i)}}{\sqrt{\sum_i H_p(i) \sum_i H_c(i)}}}$$

Let track t have parents $p_1,..., p_n$ and children $c_1,...,c_m$. Then the matrix is built

$$S = \left(d_B\left(H_{p_i}, H_{c_j}\right)\right),$$

the rows of which correspond to the parents, the columns to the children. At the intersection of row and column tsa contains a number that characterizes the measure of the similarity of the tracks. Tracks P_i and C_j are similar if three conditions are met:

$$s_{i,j} \le \varepsilon,$$
$$\arg\min_{i'} s_{i'.j} = i,$$
$$\arg\min_{j'} s_{i.j'} = j.$$

4.1. Abandoned object detection module

During the experiments, it turned out that the tracking module described above does not detect fast moving (blurred) objects, so we proposed another method for detecting thrown objects.

The abandoned object detection algorithm relies on the following properties: the object is small; the object is compact, i.e. the object is close to a circle in shape; track the torus of motion is a parabola. Note that the trajectory of the thrown object remains parabolic even with perspective transformations of the image.

The first stage of the algorithm determines the areas of heavy movement, for which the inter-frame difference is calculated. The images of an abandoned object do not overlap on adjacent frames due to the high speed of movement, so such objects are well distinguished on the interframe difference.

Based on the small size of the thrown object and its compactness, the areas that do not satisfy these properties are filtered out on the interframe difference. After that, the centroids of the connected regions are searched for in the binary image.

At the second stage, the parabolic trajectory is searched. To do this, we use the RANSAC (Random Sample Consensus) algorithm presented by Fischler and Bolles in 1981 [15]. The algorithm is an iterative method for estimating the parameters of a mathematical model based on a data set containing a large number of outliers. In the problem of detecting an abandoned object, the mathematical model is a parabolic trajectory. The input data of the algorithm are the detected centroids. The task of the algorithm is to divide them into two sets: good points (not outliers) that satisfy the hidden model, and false points (outliers), noises - random inclusions in the original data.

The algorithm works in several iterations. The number of iterations is equal to the expected number of frames on which the thrown object moves, and two steps are performed at each iteration.

1. Hypothesis. A minimum set of samples (INR) is selected from the input data set. After that, the model parameters are calculated using only the INR elements. The INR power is defined as the smallest value sufficient to determine the model parameters. To calculate the parameters of a parabolic trajectory, a set of samples from three points is sufficient.

2. Testing the Hypothesis. In the second step, RANSAC checks which elements of the data set are consistent with the model defined by the parameters obtained in the first step. A set of such elements is called a consensus set. The result of the RANSAC algorithm is the model parameters and the consensus set, the elements of which satisfy the found model (Fig. 2). Moreover, if the power of the consensus set turned out to be less than a given value, then we consider that the throw event did not occur.

DOI : 10.31695/IJERAT.2022.8.8.3

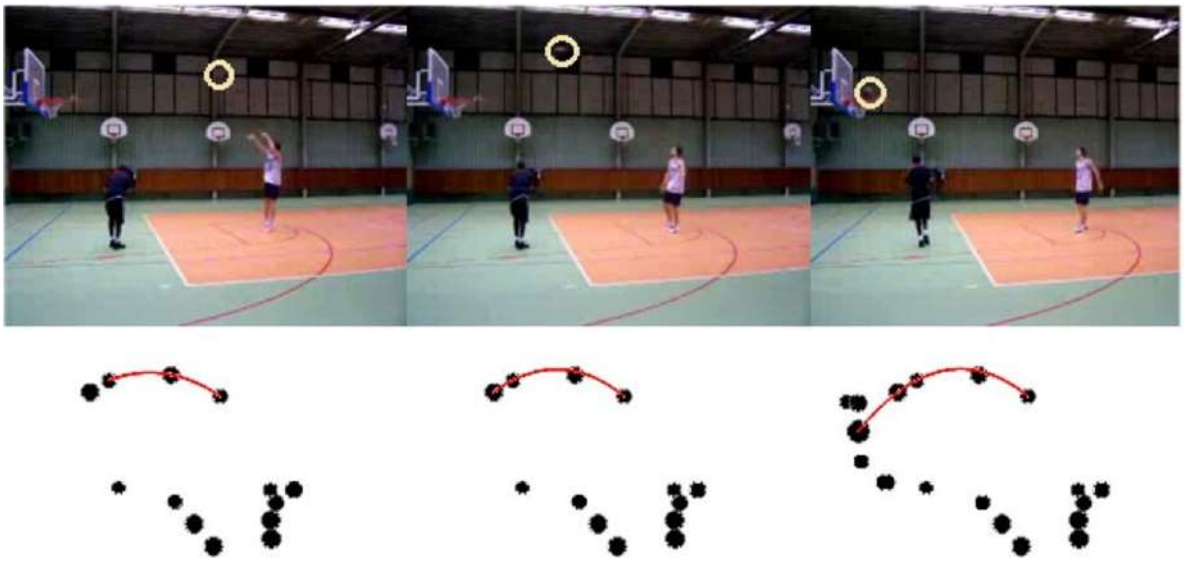

Fig.2. Consensus set of points and the calculated throw trajectory

5. TESTS AND RESULTS

The main parameters characterizing the operation of the proposed methods are:

1) Sensitivity - the proportion of true positive responses of the algorithm from the total number of events that needed to be detected.

2) Specificity - the proportion of true positive responses of the algorithm from the total number of responses.

3) Speed of work - the ratio of the number of frames to the time of their processing.

The algorithms were tested on video files containing 20 events of each type.

5.1 Test results

Table 1 Parameters Characterization

Event Type	Sensitivity	Specificity
Movement in a prohibited direction	0,95	0,86
Being in a sterile area	0,95	1,0
abandoned item	0,9	1,0
thrown object	0,8	0,94

The speed of operation was measured with all modules operating simultaneously. The input was a video stream with a resolution of 1920 × 1088 pixels, a rate of 25 frames per second. Computing resources: 4-core Intel Core i5 processor, 3200 GHz, 16 GB RAM. The processing speed is 75 frames per second, which is three times the speed of the video stream, the amount of memory consumed is 2 GB.

During testing, several problems were found. In rare cases, an object is divided into two (for example, the upper and lower halves of a person), the algorithm incorrectly identifies objects and incorrectly determines the direction of movement in cases where the trajectories of objects intersect.

The abandoned object detection algorithm in rare cases does not detect small thrown objects or objects that blend into the background. The algorithm is unable to detect multiple thrown objects moving at the same moment.

DOI : 10.31695/IJERAT.2022.8.8.3

6. CONCLUSION

As a result of the work, methods for detecting four types of alarm events were proposed: movement in a prohibited direction, being in a sterile zone, leaving an object, and throwing an object.

The first three events are recognized by the moving object tracking module, which uses our improved background subtraction method for object detection and analysis of distances and image histograms to match objects detected on different frames.

To recognize the "thrown object" event, a separate module has been developed that uses the interframe difference method to identify areas of motion and the RANSAC algorithm to search for the trajectories of a thrown object.

Further research will be aimed at solving the problems identified during testing and improving methods for detecting and tracking objects to increase the sensitivity and specificity of the algorithms we have developed. It is also planned to increase the speed of their work by optimizing and parallelizing calculations, in particular, transferring image processing operations to the computing power of graphics accelerators.

In addition to the tasks listed above, research will be aimed at developing a single database of detected objects that combines information from the object tracking module.

ACKNOWLEDGMENT

The authors would like to thank AL_Mustansiriyah University (www.uomusiriyah.edu.iq), Baghdad-Iraq for its support in the present work.

REFERENCES

1. Ainsworth T. Buyer Beware // Security Oz. 2002 Vol. 19. P. 18–26.

2. Single M. Motion Detection Based on Frame Difference Method International // Journal of Information & Computation Technology. 2014. Vol. 4. No. 15. P. 1559–1565.

3. Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction // IEEE Int. Conf. pattern recognition. 2004 Vol. 2. P. 28–31

4. Bouwmans T., El Baf F., Vachon B. Background Modeling using Mixture of Gaussians for Foreground Detection – A Survey // Recent Patents on Computer Science. 2008 Vol. 1. P. 219–237.

5. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005. P. 886–893.

6. Felzenszwalb P., Girshick R., McAllester D., Ramanan D. Object Detection with Discriminatively Trained Part Based Models // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010 Vol. 32. No. 9. P. 1627–1645.

7. Girshick R., Donahue J., Darrell T., Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation // IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 580–587.

8. Je C., Park H. M. Optimized Hierarchical Block Matching for Fast and Accurate Image Registration // Signal Processing: Image Communication. 2013. Vol. 28. No. 7. P. 779–791.

9. Aslani S., Mahdavi-Nasab H. Optical Flow Based Moving Object Detection and Tracking for Traffic Surveillance // International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. 2013. Vol. 7. No. 9. P. 1252–1256.

10. Zaveri M. A., Merchant S. N., Desai U. B. Small and Fast Moving Object Detection and Tracking in Sports Video Sequences // IEEE International Conference on Multimedia and Expo. 2004 Vol. 3. P. 1539-1542.

11. Comaniciu D., Ramesh V., Meer P. Kernel-based object tracking // IEEE Transactions on pattern analysis and machine intelligence. 2003 Vol. 25. No. 5. P. 564–577.

12. Chitaliya N. G., Trivedi A. I. Novel block matching algorithm using predictive motion vector for video object tracking based on color histogram // 3rd International Conference on Electronics Computer Technology. 2011 Vol. 5. P. 81–85.

13. Hingane P., Shirsat S. Object Tracking Using Joint Color-Texture Histogram // International Journal of Science and Research. 2013. P. 2603–2606

DOI : 10.31695/IJERAT.2022.8.8.3

14. Tissainayagama P., Suterb D. Object tracking in image sequences using point features // Pattern Recognition. 2005 Vol. 38. No. 1. P. 105–113.
15. Fischler M. A., Bolles R. C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography // Communications of the ACM 24. 1981. P. 381–395.

DOI : 10.31695/IJERAT.2022.8.8.3