# Automated Target Tracking Activation Based on Motion Detection

Dejan Cizelj, Miloš Pavlović, *Member, IEEE*, Tomislav Unkašević, *Member, IEEE* and Zoran Banjac, *Member, IEEE*

***Abstract*—The activation of the target tracking process in surveillance imaging systems is usually manual, which require a constant user's observation of the scene. Especially, target selection and tracking activation is very challenging for fast moving targets. To overcome the problem of manual target tracking activation and to achieve a certain degree of automation, this paper presents a real-time algorithm for automated moving target tracking activation. The proposed algorithm is based on adaptive mixture of Gaussians background subtraction method and Kalman filtering. Experimental results using a short-wave infrared camera show that the proposed algorithm is efficient in activation of tracking with a precisely estimated target size, resulting in accurate target tracking by pan-tilt-zoom surveillance imaging system.**

***Index Terms*—Kalman filter, Motion Detection, Target Tracking Activation.**

## I. Introduction

Modern video surveillance systems have a great need for video analytics for timely response to situations that may pose a security risk. However, video surveillance systems usually require the user to be constantly present and observe the scene being recorded in order to respond in a timely manner to situations that may be risky. In other words, a frequent problem is the lack of intelligence for independent functioning without the presence of the user. In order to develop systems that can be said to be smart, which can see, somehow interpret, understand and react in a timely manner to different situations in the environment, they need to have the ability to detect and track targets of interest [1].

In most cases, video surveillance systems, in addition to their basic recording role, do not have the ability to track targets on the scene being observed. For those that have implemented video tracking algorithms, the activation of the tracking process is usually manual, where the user needs to select the target of interest for tracking by defining the bounding box around the target. For moving targets, especially if they move very fast through the camera's field of view (FOV), target selection and tracking activation is very challenging. Even a very experienced user has difficulties to accurately select a target that is moving very fast, which results in either unsuccessful activation of the video tracking process or tracking of an incorrectly selected target and thus poor tracking performance. To overcome the problem of activation for moving targets and achieve a certain degree of automation of this procedure, an algorithm for detection of moving targets and activation of tracking need to be implemented. Moreover, algorithm for moving targets detection can be useful for detecting targets intruding into the scene and be the essential step in analyzing the scene.

A static camera observing a scene is a common case of a surveillance system. The usual applicable assumption is that the images of the scene without the intruding objects exhibit some regular behavior that can be well described by a statistical model. With statistical model of the scene, an intruding target can be detected by identifying the parts of the image that do not fit the model. This process is known as background subtraction [2]. The adaptive mixture of Gaussian (MOG) [3] background subtraction method is implemented, as it is one of the fastest methods for motion detection with the possibility of real-time processing. Although MOG is efficient method for motion detection, due to the presence of noise and camera shake, false motions may be detected, or some motions may be missed due to insufficient contrast between the background and the moving target. In order to overcome these problems and fully automate the process of activating the tracking of an intruding moving target, it is necessary to establish a relationship between motion detections in successive frames of the video sequence. For this purpose, a Kalman filter [4] is implemented for the process of so-called tentative tracking in order to increase the certainty that the surveillance system will activate the tracking of a real moving target.

The paper is organized in the following manner: Section II describes the methodology of work, algorithm for the motion detection, as well as the Kalman filtering technique. Section III describes the proposed method for automated target tracking activation. In Section IV, experiments are described, and results are discussed. The conclusion is presented in Section V.

Dejan Cizelj is with the Singidunum University, Danijelova 32, 11000 Belgrade, Serbia (e-mail: dayantcizela@gmail.com).

Miloš Pavlović is with the School of Electrical Engineering, University of Belgrade, 73 Bulevar kralja Aleksandra, 11020 Belgrade, Serbia, and Vlatacom Institute of High Technologies, 5 Blvd. Milutina Milankovića, 11070 Belgrade, Serbia (e-mail: milos.pavlovic@vlatacom.com).

Tomislav Unkašević is with the Vlatacom Institute of High Technologies, Blvd. Milutina Milankovića 5, 11070 Belgrade, Serbia (e-mail: tomislav.unskasevic@vlatacom.com).

Zoran Banjac is with the Vlatacom Institute of High Technologies, Blvd. Milutina Milankovića 5, 11070 Belgrade, Serbia (e-mail: zoran.banjac@vlatacom.com).

 ISBN 978-86-7466-970-9

## II. METHODOLOGY

Fully automated tracking is of interest in this paper, which implies the complete absence of the user during the execution of the tracking algorithm. This is made possible by the existence of a target detector that is an integral part of the tracking algorithm [5], [6]. It mainly serves to initialize tracking or to help estimate the state of the target.

The goal of the paper is to maximally reduce the need for manual user activities, primarily in the tracking activation process, and to increase the precision of defining the bounding box around the target for tracking, which is especially challenging for targets that pass through the camera's FOV very fast. To achieve this, a real-time motion detector is needed, as well as a method to increase the certainty that tracking will be activated for a real moving target. The motion detection method, as well as the Kalman filter, as the main modules for the development of the algorithm for the automated moving target tracking activation, are described in the sequel.

### A. Motion Detection

Many approaches have been developed for motion detection in video sequences in the past years [7]. Some of the main directions in motion detection are based on background subtraction [2], [3], optical flow [8], and approaches based on deep learning [9].

As the goal is to timely detect an incoming moving target, and activate the tracking of that target, the motion detection method should have a low processing time. For this reason, module for motion detection in automated tracking activation algorithm is based on background subtraction method - adaptive mixture of Gaussians (MOG)-based background / foreground segmentation [3].

MOG uses a mixture of K Gaussian distributions to model each background pixel, where K is a number ranging from 3 to 5, with each Gaussian representing a different color. The weighting parameters of the mixture are utilized to determine the proportion of time each color is present in the scene. This method assumes that the background contains the B most probable colors, which are the colors that remain static for longer periods of time. Static, single-color areas tend to form tightly clustered points in color space, while moving targets create wider clusters due to different reflection surfaces during motion. The metric used to measure this is referred as the fitness value. In order to make the model adaptable to variations in illumination and to operate in real-time, an update approach is implemented based on selective updating. Each new pixel value is compared to the existing model components based on their fitness value. If a match is found, the first matched model component will be updated. In case of no match, a new Gaussian component will be introduced with the mean located at that pixel value and a large covariance matrix along with a small weighting parameter [3].

### B. Kalman filter

To establish a connection between motion detections on consecutive frames and to reduce the probability of tracking activation for false target, Kalman filter is used, which is one of the most important contributions to linear estimation theory [4].

A system is represented with the state-space model, which comprises of state and observation equations:

$$x_{k+1} = F_k x_k + G_k w_k \tag{1}$$

$$y_k = H_k x_k + v_k \tag{2}$$

where $x_k$ is the state vector, $y_k$ is the observation or measurement vector, $w_k$ is the state noise vector, and $v_k$ is the additive measurement noise vector, at the discrete time step indexed by $k$. Moreover, $F_k$ represents the state-transition matrix, $G_k$ is the state-noise matrix, and $H_k$ is the observation matrix. The state noise $w_k$ and the observation noise $v_k$ are assumed to be zero mean noises, and uncorrelated by itself and mutually:

$$E\left\{w_k\right\} = 0; E\left\{w_k w_k^T\right\} = Q_k \delta_{kj} \tag{3}$$

$$E\left\{v_k\right\} = 0; E\left\{v_k v_k^T\right\} = R_k \delta_{kj} \tag{4}$$

where $\delta_{kj}$ is the Kronecker's delta symbol ($\delta_{kj} = 0$ if $k \neq j$, and $\delta_{kk} = 1$), and $E\{\cdot\}$ being the mathematical expectation. $Q_k$ and $R_k$ present the state noise and the observation noise covariance matrices, respectively.

Kalman filter has recursive predictor-corrector structure, so the standard Kalman filter prediction and estimation equations are the following:

- Prediction step:

$$\hat{x}_{k|k-1} = F_{k-1}\hat{x}_{k-1|k-1} \tag{5}$$

$$P_{k|k-1} = F_{k-1}P_{k-1|k-1}F_{k-1}^T + G_{k-1}Q_{k-1}G_{k-1}^T \tag{6}$$

- Estimation step:

$$\varepsilon_k = y_k - \hat{x}_{k|k-1} \tag{7}$$

$$K_k = P_{k|k-1}H_k^T\left[H_k P_{k|k-1}H_k^T + R_k\right]^{-1} \tag{8}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \varepsilon_k \tag{9}$$

$$P_{k|k} = \left[I - K_k H_k\right]P_{k|k-1} \tag{10}$$

In (5) – (10), $\hat{x}_{k|k-1}$ is prediction of the present state, $x_k$, $P_{k|k-1}$ is the corresponding prediction error covariance matrix, $\varepsilon_k$ represents the measurement residual, $K_k$ is the Kalman gain, $P_{k|k}$ denotes the estimation error covariance matrix, and $I$ represents the identity matrix.

With its simple predictor-corrector structure, the Kalman filter is useful for applications in the detection of fast-moving targets, as it requires low computation time.

 ISBN 978-86-7466-970-9

## III. Automated Tracking Activation

The result of applying the MOG background subtraction algorithm is a binary image in which moving object are represented with white pixels while the background is represented with black. The white pixels in this image are the result of moving objects on the scene, but also the influence of the noise and camera's shake. Also, it is often the case that one moving object is represented by several clusters of white pixels. In order to overcome the problems of false movements and separated movements of the same object, morphological operations are applied to the obtained binary image. The procedure of morphological opening (erosion followed by dilation) is applied for removing small blobs presenting false movement, and then morphological closing (dilation followed by erosion) is applied for filling small holes between separated movement regions of the same object. In this way, the problem of false small motion detections is largely eliminated, and it is also achieved that one moving object is represented by one region of white pixels. After applied morphological operations, it is necessary to define closed contours of white pixels, and then rectangular bounding boxes around the obtained closed contours.

For the obtained bounding boxes around the detected moving objects, a comparison of their area with a defined threshold is applied. Only those whose area exceeds the defined threshold of 100 pixels should be kept in the set of moving objects in order to eliminate very small objects or detected movements that actually represent noise in the image.

In order to additionally reduce the probability of activating the tracking process for the detection of a false moving target, a tentative tracking of the detected motion is implemented. For the first moving target intruding into the scene, an instance of the Kalman filter is created, in charge of tracking the given detection through the next frames of the video sequence, and target existence period counter is initialized (*t_pred*). The implemented Kalman filter is a constant velocity model, and the state vector of the target is defined as [10]:

$$X = \begin{bmatrix} x_c \\ y_c \\ s \\ r \\ \dot{x}_c \\ \dot{y}_c \\ \dot{s} \end{bmatrix} \qquad (11)$$

where $x_c$ and $y_c$ represent the position of the detected moving target bounding box center in the horizontal and vertical directions, while $s$ represents the area of the bounding box, and $r$ is the ratio of the width and height of the bounding box. The last three states in the vector X represent the corresponding velocities. The model defined in this way is important in order to track the position of the moving target in the image as well as changes in the dimensions of the target over period of tentative tracking.

The bounding boxes around the newly detected motions in the next frame and the prediction of the position and size of the target from the previous frame based on the Kalman filter are then passed to the block for calculating the IOU (Intersection Over Union) [11] metric between the obtained bounding boxes. The IOU metric actually determines whether the newly discovered motion detection in the image is caused by a temporary tracked target. If the overlap according to the IOU metric between the motion detection and the Kalman filter prediction of the temporary tracked target exceeds a threshold of 0.5, the state of the temporary tracked target can be updated, and its update counter incremented (*t_update*). If the target state is updated in 5 frames, the tracking process by the surveillance system will be automatically activated for that target. If this condition is not met, the procedure continues until a real moving target is detected, or the process is reinitialized if tracking is not activated after the first 10 frames of tentative tracking. The block diagram of the proposed algorithm is shown in Fig. 1.

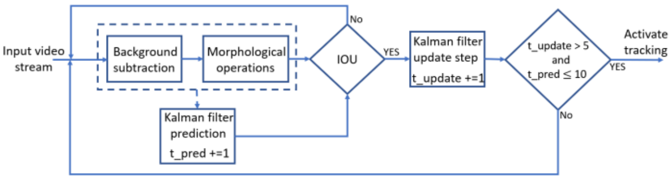



Fig. 1. Block diagram of the proposed tracking activation algorithm.

## IV. Experimental Work

The system setup used in this paper is based on vMSIS (Vlatacom Multi-Sensor Imaging System) [12], which is shown in Fig. 2. The system consists of three video channels: a channel operating in the visible range (FULL HD resolution: 1920x1080 pixels), short-wave infrared (SWIR) channel (resolution: 576×504), and mid-wave infrared (MWIR) channel (resolution: 640x480). The system is mounted on a pan-tilt positioner and equipped with a video tracking algorithm.

Fig. 2. Multi Sensor Imaging System on pan-tilt positioner, equipped with color, SWIR and thermal camera.

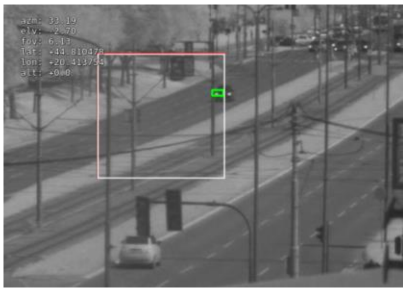
 ISBN 978-86-7466-970-9

The implemented video tracking method is based on the Kernelized Correlation Filter (KCF) algorithm [13]. Video tracker tracks the target of interest in the image and provides control signals to the pan-tilt positioner, so that the target remains in the camera's FOV. The input of the tracking algorithm is a bounding box with the coordinates of the center, as well as the height and width of the target in the image plane. Therefore, the output of the automated tracking activation algorithm is a bounding box defining the detected moving target. The size of the bounding box should be well estimated, because the extracted features of the tracked target will be incomplete if the size of the bounding box is smaller than the size of the target, and if the size of the bounding box is larger than the target, variable background features will be introduced into the target appearance model.

The proposed algorithm for automated target tracking activation is implemented in C++ programming language using the OpenCV library to perform most of the functions related to image processing. The algorithm is implemented on the NVIDIA Jetson TX2 within the vVSP (Vlatacom Video Signal Processing) platform [14], where it achieves an average processing speed of 68 FPS, which is enough for real-time operation.

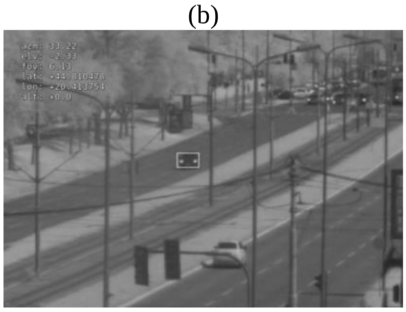
The experimental analysis was performed using the SWIR camera. Fig. 3 shows the automated activation of car tracking, as a typical fast-moving target, in the region of interest (ROI). It can be seen that the algorithm successfully detects the movements when the target enters the ROI, the detected movements are temporary tracked, and after determining that they are caused by the real target, the KCF tracker is activated with a precisely estimated size of the bounding box around the target. KCF tracker then generates the control signals to the pan-tilt positioner, brings the target to the center of the camera's FOV and enables long-term target tracking.

Fig. 4 presents objective performance of the algorithm on the created dataset of SWIR video sequences. The results are obtained from 100 target tracking activations. Target of interest for tracking is manually labeled with the corresponding bounding box on each frame which presents the moment of the tracking activation. The center of the bounding box is the position of the target for tracking, while the size of the target is represented with the bounding box width and height. The graph in Fig. 4 represents the accuracy for different values of the IOU metric between the estimated bounding box of the moving target and the ground-truth bounding box, at the moment of tracking activation. Accuracy is calculated as the percentage of target tracking activations in which the IOU between the estimated bounding box and the ground-truth bounding box is higher than the defined IOU threshold. The obtained results shows that the algorithm very well estimates the position and size of the target at the moment of tracking activation. The accuracy of the algorithm is most affected by the estimation of the target size. Size estimation is particularly challenging due to partial and full occlusions which cover segments of the target area, as well as target shadow which has the same movement dynamics as the target.



(a)

(b)



(c)

Fig. 3. Automated moving car tracking activation in ROI: a) detected motion of incoming target in ROI b) tentative tracking of detected moving target c) activated tracking of a moving target by the PTZ surveillance imaging system using KCF-based tracker.

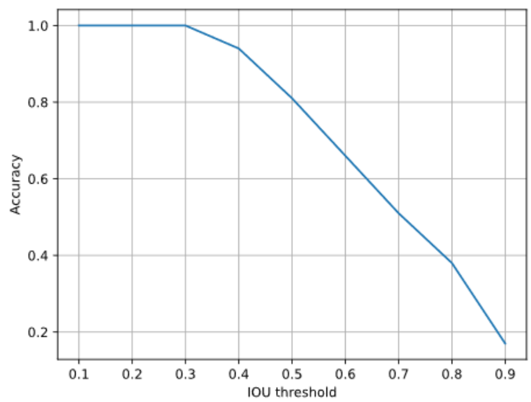



Fig. 4. Target tracking activation accuracy for different IOU thresholds.

## V. Conclusion

This paper presents an algorithm for automated tracking activation of moving targets in order to overcome the problem of manual tracking activation of fast-moving targets and achieve a certain degree of autonomy of surveillance imaging systems. The detection of moving targets is based on the background subtraction method, a mixture of Gaussian components, which is computationally very efficient and enables real-time operation on low processing power units. To reduce the probability of false detections and increase the precision of the target bounding box estimation, tentative motion tracking based on the Kalman filter was applied. Application of automated tracking activation algorithm significantly facilitates user activities, and easy and precise activation of the fast-moving targets tracking is achieved. Moreover, the proposed algorithm is of great importance for surveillance systems in conditions without significant background textures, where it can be applied for the detection of incoming targets and their automated tracking.

## ACKNOWLEDGMENT

This paper was supported by the Vlatacom Institute of High Technologies under project #157 VMSIS3_Advanced.

## REFERENCES

[1] A. Cavallaro and E. Maggio, “Video Tracking: Theory and Practice,” John Wiley & Sons, 2011.

[2] B. Garcia-Garcia, T. Bouwmans, and A. J. R. Silva, “Background Subtraction in Real Applications: Challenges, Current Models and Future Directions,” Computer Science Review, 35, 100204, 2020.

[3] P. KaewTraKulPong, and R. Bowden, “An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection,” Proc. of the 2nd European Workshop on Advanced Video-Based Surveillance System: Computer Vision and Distributed Processing, pp. 135-144, 2001.

[4] M.S. Grewal and A.P. Andrews, “Kalman filtering: Theory and Practice with Matlab,” John Wiley & Sons.

[5] M. Pavlović, N. Stojiljković, I. Gluvačević, M. Vučetić, and M. Perić, “Real-Time Moving Object of Interest Detection in Multi-Sensor Imaging System,” Proc. of 7th IcEtran 2020, Belgrade, Serbia, 2020.

[6] N. Wojke, A. Bewley, and D. Paulus, “Simple Online and Realtime Tracking with a Deep Association Metric,” Proc. of the IEEE International Conference on Image Processing (ICIP), pp. 3645-3649, 2017.

[7] M. N. Chapel, and T. Bouwmans, “Moving Objects Detection with a Moving Camera: A Comprehensive Review,” Computer Science Review, 38, 100310, 2020.

[8] S. S. Beauchemin, and J. L. Barron, “The Computation of Optical Flow,” ACM Computing Surveys (CSUR), vol. 27, no. 3, pp. 433-466, 1995.

[9] D. Sun, X. Yang, M. Y. Liu, and J. Kautz, “Models Matter, So Does Training: An Empirical Study of CNNs for Optical Flow Estimation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 6, pp. 1408-1423, 2018.

[10] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple Online and Realtime Tracking,” in Proc. of IEEE International Conference on Image Processing, pp. 3464-3468, IEEE, 2016.

[11] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized Intersection Over Union: A metric and a Loss for Bounding Box Regression,” in Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 658-666, 2019.

[12] Vlatacom Institute: Electro-optical systems VMSIS3, product brochures, available online: https://www.vlatacominstitute.com/electro-optical-systems, accessed on 19-April-2023.

[13] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “High-Speed Tracking with Kernelized Correlation Filters.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, pp. 583-596, 2014.

[14] N. Latinović, I. Popadić, B. Tomić, A. Simić, P. Milanović, S. Nijemčević, M. Perić, and M. Veinović, “Signal Processing Platform for Long-Range Multi-Spectral Electro-Optical Systems,” Sensors 2022, 22, 1294.

 ISBN 978-86-7466-970-9