https://doi.org/10.15407/scine21.05.062

SABELNIKOV, P. Yu. (https://orcid.org/0000-0003-4709-8693)

V.M. Glushkov Institute of Cybernetics
of the National Academy of the Sciences of Ukraine,
40, Akad. Glushkova Ave., Kyiv, 03187, Ukraine,
+380 44 526 2008, incyb@incyb.kiev.ua

AUTOMATED DETECTION OF FOREIGN OBJECTS IN VIDEO SCENES WITH DYNAMIC OBSTACLES

Introduction. *The automatic detection of foreign and abandoned objects plays a critical role in modern video surveillance systems, particularly in areas with high pedestrian traffic, where such capabilities are essential for the prevention and investigation of security threats and terrorist activities.*

Problem Statement. *Video surveillance systems generate vast volumes of data, making it impractical for human operators to monitor all video streams continuously. To mitigate this challenge, automated video analytics systems have been developed to support or replace manual observation. Since human operators may miss subtle changes in complex video scenes, there is a clear need for robust algorithms capable of real-time scene analysis.*

Purpose. *This study aims to develop methods for the automatic detection of foreign objects in dynamic video environments, enabling early threat identification and supporting security operations at public institutions and critical infrastructure facilities.*

Materials and Methods. *Experimental validation has been performed using the PETS 2006 benchmark dataset, which includes seven scenarios of increasing complexity and is widely used to evaluate algorithms for detecting abandoned and removed objects in public environments.*

Results. *A novel method has been developed that is resilient to changes in lighting conditions and minor camera displacements. Experimental studies have demonstrated that the proposed algorithm reliably detects foreign objects in dynamic video scenes under varying illumination and scene perturbations. The method has shown stable performance with minimal feature extraction, achieving consistent results on PETS 2006 video sequences.*

Conclusions. *The proposed approach enables real-time operation of video surveillance systems, significantly reducing the cognitive load on operators by automatically detecting potentially dangerous abandoned objects and issuing timely alerts, thereby enhancing situational awareness and public safety.*

Keywords: video systems, video surveillance systems, video fragment, video sequence.

Citation: Sabelnikov, P. Yu. (2025). Automated Detection of Foreign Objects in Video Scenes with Dynamic Obstacles. *Sci. innov.*, 21(5), 62—75. https://doi.org/10.15407/scine21.05.062

© Publisher PH "Akademperiodyka" of the NAS of Ukraine, 2025. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

At the current stage of technological advancement, **computer vision systems** have gained widespread application across various domains. Their primary purpose is to assist — or even replace — human operators in tasks related to the acquisition and analysis of visual information. The deployment of computer vision systems (CVS) in applied fields serves as one of the most prominent indicators of progress in high technologies and their integration into diverse sectors of human activity.

One of the most extensive areas of application for CVS is video surveillance systems, which have become an integral part of everyday life. An increasing number of public and private facilities are being equipped with video monitoring infrastructure. These systems generate vast volumes of video data, which shall be processed and filtered to detect potentially dangerous events or to extract relevant historical footage for investigation. It is therefore both practical and logical to transfer part of the operator's workload to automated video analytics systems, which analyze real-time data streams from multiple cameras, thereby reducing the burden of routine monitoring.

The core functions of video analytics include object detection, object tracking, object classification, object identification, situation detection, and behavior prediction. These functions contribute significantly to improving the efficiency and responsiveness of surveillance operations.

Extensive research and development in computer vision have been undertaken by laboratories worldwide. From 2012 to 2020, significant advancements have been reported by the School of Electrical and Computer Engineering at the University of Oklahoma (Norman, USA), the Video Processing Laboratory at the University of California (San Diego, USA), the Intelligent Systems Research Group at the Institute of Research and Technology (Madrid, Spain), and Microsoft Research Cambridge (UK).

In Ukraine, numerous institutions under the National Academy of Sciences (NAS) and the Ministry of Education and Science have contributed to the development of video computing systems. Key contributors include the International Research and Training Center for Information Technologies and Systems (NAS and MES of Ukraine, Dr. M. I. Shlezinger), Taras Shevchenko National University of Kyiv (Prof. Y. V. Krak), Donetsk National Technical University (Dr. Ye. O. Bashkov), and the H. E. Pukhov Institute of Modeling Problems in Power Engineering of the NAS of Ukraine (Dr. Yu. M. Hruts), among others.

Depending on the specific application goals, intelligent video data processing can fulfill one or more of the following functions:

- Object Detection. For initial detection, both software-based and hardware-based motion detectors are commonly used.
- Object Tracking. Modern software enables trajectory tracking of moving objects using either multiple static cameras or a combination of static and pan-tilt-zoom (PTZ) cameras. Analysis includes not only the direction of movement but also speed. This function is also used as a filter to avoid redundant event registration when the same object moves from one camera's field of view to another.
- Classification and Statistical Analysis. Advanced video surveillance systems can classify individual objects or events using statistical filters. For instance, a system can distinguish between a person, a group of people, or a vehicle based on size and shape features. With high accuracy, the system may infer gender, approximate age, or even the make of a vehicle.
- Identification. Object recognition is one of the most sophisticated and technically demanding tasks. It requires high-performance algorithms and reliable hardware. Video surveillance systems are often integrated with other access control technologies, such as electronic badges, magnetic cards, fingerprint readers, and similar biometric tools.
- Incident Detection. Real-time analysis of video streams is used to identify anomalous or hazardous events. Key indicators include crossing virtual boundaries, abrupt changes in body posture (e.g., falls or jumps), the occurrence of fire

or smoke, or the unauthorized entry of a vehicle into restricted zones such as parking areas.

In general, the process of automated intelligent video analytics in surveillance systems can be divided into the following key stages:

- foreground segmentation;
- detection and classification of moving objects;
- trajectory tracking of identified objects;
- recognition and classification of actions or behaviors of objects of interest.

To address the tasks at each stage, researchers have employed a range of techniques. For instance, foreground segmentation is often performed using background subtraction methods, probabilistic approaches, and mathematical models such as co-occurrence matrices, temporal differencing, and optical flow. Additionally, neural network-based methods for background modeling have gained increasing traction in recent years.

Despite the extensive attention paid to video analytics tasks, the problem of recognizing situations where critical objects disappear or hazardous (unauthorized) objects appear in a dynamic scene remains insufficiently explored. Addressing this challenge requires not merely detecting movement, but filtering out dynamic elements (e.g., pedestrian motion) to identify persistent changes in the background image, previously captured under stable conditions. Therefore, any solution shall also account for variations in lighting and slight shifts in fixed camera orientation over time.

The present study has focused on developing video stream processing methods to support this particular function of video analytics.

Detection of abandoned objects is a common application in surveillance systems installed in areas of high pedestrian density. It serves both to prevent terrorist threats and to investigate unusual or suspicious activity. This type of object detector also finds application in tasks such as monitoring parking occupancy and safeguarding valuable assets.

Modern video surveillance platforms offer a wide range of tools for context-aware and situational monitoring of diverse environments. In the context of anti-terrorism measures, abandoned object detection is widely implemented in video systems deployed in public spaces considered vulnerable — such as banks, shopping centers, markets, train stations, airports, and subways. The detector is designed to register objects that were previously in motion and have remained stationary for a predefined period of time, thus signaling potentially suspicious behavior.

The functionality for detecting disappearing or newly appeared objects has proven highly relevant for use in retail enterprises, business centers, transportation infrastructure, and access control points. In public locations such as train stations, airports, metro systems, shopping malls, and central squares, there is a persistent risk that unattended items — such as a suitcase or other suspicious object — may signal the potential onset of a terrorist act.

A substantial body of existing research has addressed the detection and tracking of moving objects in video scenes by analyzing data from stationary or omnidirectional surveillance cameras. These efforts often involve panoramic scene reconstruction and the application of background subtraction techniques. However, in the author's view, insufficient attention has been paid to the inverse problem — identifying scenarios where critical objects disappear unnoticed or hazardous items appear in dynamic video scenes.

Solving this problem requires a fundamentally different approach: rather than focusing on motion detection, the objective is to filter out rapid dynamic changes (such as pedestrian movement) and instead detect persistent alterations in the background image, previously captured under stable conditions. This approach is particularly sensitive to subtle and gradual threats, making it crucial for maintaining situational awareness and public safety.

Consequently, the development and investigation of new methods for automatic detection of extraneous objects in video scenes with dynamic interference is both timely and socially significant. The relevance of this work is further underscored by its implications for institutional and civilian safety, particularly under conditions of martial law or heightened national security threats.

The goal of this research has been to advance the level of automation in video surveillance systems by developing robust methods for detecting foreign or abandoned objects in the presence of dynamic obstacles. These methods are intended to prevent potential offenses and terrorist threats, as well as to support investigations of such incidents in critical infrastructure and public institutions.

The implementation of the theoretical and practical results in hardware and software systems is expected to be economically viable due to the high market demand and the low reproduction cost of intelligent algorithmic solutions. This contributes not only to improved security but also to cost-effective deployment at scale.

The research has been carried out within the framework of the Youth Research Program for 2023—2024, as mandated by the Presidium of the National Academy of Sciences of Ukraine in Order No. 321, dated June 19, 2023, titled *On the Results of the Research Project Competition for Young Scientists of the NAS of Ukraine in 2023*.

1. DETECTION OF ABANDONED OR MISSING OBJECTS USING BACKGROUND SUBTRACTION AND SCENE CHANGE TRACKING

1.1. Methods and Algorithms for Background Subtraction and Scene Change Detection. The core principle behind change detection algorithms in video surveillance lies in identifying objects that have either disappeared from or appeared within the monitored area — and remained in that altered state for a non-trivial period.

Numerous background subtraction methods have been developed [1], each with specific advantages and limitations in terms of computational efficiency, robustness, and real-time performance. A reliable background subtraction algorithm shall cope with changing lighting conditions, repetitive movements, and varying weather.

To ensure adaptability to scene changes — such as lighting variation, dynamic backgrounds (e.g., moving trees or water), or temporary static objects — most modern background subtraction algorithms are designed to be adaptive. However, a slow-moving or stationary object may, after several frames, be incorporated into the background model. This can lead to tracking errors where the object disintegrates into fragments or “ghost” artifacts appear in areas where an object was removed. Since tracking processes generally consider pixel clusters as discrete objects, this higher-level information can be leveraged to enhance the accuracy of the background subtraction stage.

Several **conventional methods** for background subtraction have been widely used in the literature. These include:

1. *Frame Differencing*. One of the simplest background subtraction methods, this approach treats the previous frame as the background model and detects changes by subtracting it from the current frame [4—5].

2. *Kalman Filtering*. Kalman filters are particularly effective when pixel intensity values follow a Gaussian distribution. Adaptive filters update the mean and variance of the background model, thereby compensating for lighting changes and integrating temporarily static objects. Kalman filtering for background estimation has been detailed in [5].

3. *Approximate Median Filtering*. Proposed by McFarlane and Schofield [2], this recursive filtering method estimates the temporal median of each pixel. It has been adopted in urban traffic monitoring applications due to its computational speed.

4. *Single Gaussian Filtering*. This involves averaging a sequence of frames to build a background model. Each new frame is subtracted from this average, and differences exceeding a predefined threshold are flagged. This method is among the simplest and most intuitive for background subtraction [6].

5. *MIN-MAX Filtering*. This technique uses three values — maximum, minimum, and maximum difference — to determine whether a pixel belongs to the background. Haritaoglu et al. [7] have proposed a variant capable of locally adapting to

noise. In their model, each background pixel is characterized by a maximum value Ms, minimum value ms, and a maximum difference Ds across consecutive frames.

In summary, these background subtraction techniques form the basis for many real-time object detection systems and serve as a first-level filter for identifying potential changes in the scene that may warrant further analysis or alert generation.

Modeling the background using a single reference image, as in traditional approaches, requires a strictly static background, free from noise and interference. Since such conditions cannot be guaranteed in real-world environments, many advanced models have been trained on a sequence of training frames. Researchers have developed statistical methods based on Gaussian Mixture Models (GMM), which model each pixel as a combination of several Gaussian distributions to better account for dynamic scene elements.

Non-parametric methods have also been explored, ranging in complexity from basic models to more advanced approaches such as Eigen-background subtraction. Studies have evaluated the accuracy of object detection across various techniques. It has been shown that methods operating on grayscale video, such as MinMax filtering, generally produce less accurate results than those using color video. In contrast, algorithms like Kernel Density Estimation (KDE) and GMM have demonstrated high robustness in noisy environments. However, the Eigen-based method, while effective, requires substantial memory resources, making it unsuitable for practical deployment in many real-time applications.

Boult et al. [8] have described a system capable of reliably detecting slow-moving objects, which are often problematic for traditional background subtraction algorithms. Several research efforts have explored the integration of feedback loops from object tracking modules to the background subtraction process to enhance accuracy.

Abbott et al. [9] have proposed a method to reduce the computational cost of visual tracking systems by using track state estimations to direct and constrain the image segmentation process, combining background subtraction with connected component analysis.

Harville [10] has utilized high-level feedback frames in person-detection and tracking applications to locally adjust background sensitivity, thereby improving detection in dynamic scenes.

The authors of [11] have proposed a unified framework for simultaneous object detection and tracking, where tracking outputs are fed back into the detection stage. This interaction between tracking and background subtraction has proven beneficial for detecting slow-moving or stationary objects, enhancing both the accuracy and stability of the video analytics process.

Venetsianer et al. [12] have investigated methods for transferring foreground objects to the background and vice versa. Jian Yao and Jean-Marc Odobez [13] have proposed a similar multilevel background modeling mechanism designed to retain information about stationary objects. Teicher et al. [14] have developed an approach combining background modeling and object tracking to prevent stationary objects from being absorbed into the background.

1.2. Factors Complicating Change Detection in Video Scenes. In many cases, the process of detecting changes within a video scene is complicated by several factors, including:

- Instability of lighting conditions during continuous video surveillance;
- Slight shifts in the orientation of the camera due to temperature fluctuations in the morning, afternoon, and evening (based on our observations, thermal expansion and contraction of materials can influence even stationary cameras installed indoors);
- Environmental disturbances such as wind gusts (causing camera vibration or movement of tree branches), snow, rain, sudden cloud cover, or wave motion on water surfaces.

This study focuses exclusively on the first two factors, as it is oriented toward the use of stationary surveillance cameras installed indoors.

RESEARCH CONCEPT, METHOD, AND ALGORITHM FOR DETECTING SCENE CHANGES AND IDENTIFYING FOREIGN OBJECTS

The central concept of the proposed research lies in the application of feature descriptors invariant to illumination changes and slight displacements. These descriptors are calculated for individual elements of the scene, into which both the reference background image and each incoming video frame are segmented. Temporal comparison of these descriptors — both short-term and long-term — enables the detection of objects that have appeared or disappeared in the scene, as well as the identification of locations where foreign objects have been introduced.

2.1. Method for Detecting Abandoned Objects and Its Stages. The proposed method involves the following sequence of procedures, implemented within algorithms used for experimental validation. The goal is to identify the most efficient variants of these algorithms by testing them on benchmark video sequences that include **scenarios of object appearance or disappearance** within the monitored area:

The main procedures are as follows:

1. *Conversion* of images from grayscale to halftone representation.
2. *Preprocessing of images* to filter noise and smooth sharp transitions at object boundaries, reducing sensitivity to illumination changes and shifts when comparing elements of the background image and frames from the video sequence.
3. *Calculation* of invariant characteristics of background image elements during its initialization or update, and of the current frame in the video sequence.
4. *Comparison* of element characteristics based on relative errors and classification of each element into one of the following groups: unchanged, short-term changes, long-term changes; logging of events (the duration threshold for changes is set by the operator).
5. *Optional recognition of detected foreign objects* (e.g., a person or animal), followed by exclusion of such classes from operator attention.
6. *Optional automatic background update* if the current frame contains only unchanged or long-term changed elements (configurable by the operator).

2.2. Calculation of Image Element Characteristics Using Geometric Moments. Moment-based characteristics and moment invariants, calculated on their basis, have been widely applied in many digital image processing tasks [15]. The key advantage of moment invariants lies in their insensitivity to image rotation, which makes them highly effective as descriptors for detecting and recognizing objects with unknown orientations in an image. Moreover, through relatively simple transformations, it is possible to derive features that are robust to various transformations, including geometric and photometric changes.

Notably, moment invariants are particularly useful within the "sliding window" technique framework for object detection and recognition, as they allow for the parallel-recursive computation of localized moment characteristics.

Additionally, moment-based features are naturally normalized with respect to illumination changes. To facilitate their calculation, the background image and the video stream frames are divided into rectangular elements of a given size, which are further subdivided into smaller regions (Fig. 1, *a*, *b*). For instance, in Fig. 1, *b*, each element is divided into four regions.

We assume that illumination changes affect an element and its subregions approximately uniformly and linearly, given their small size and the smoothing of pixel brightness by the Gaussian filter. Based on this assumption, normalization of the characteristics of elements and their subregions with respect to illumination variation is carried out. Since moments are computed as integral values over the regions and their elements, a slight shift of images in the video sequence relative to the background image does not significantly affect the integral characteristic values being calculated and compared.

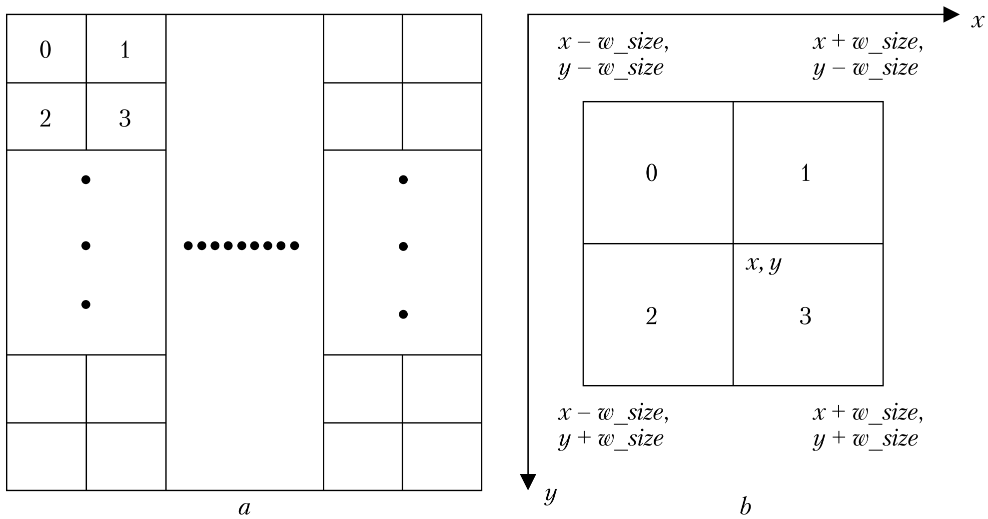

Fig. 1. The image field divided into elements (*a*), and a single element fragmented into 4 sections (*b*)

The geometric moments of image elements, with the coordinate origin set at the upper-left corner of the image, are computed as follows:

$$M_{x,y}^{l,k-l} = \sum_{i=x-w\_size}^{x+w\_size} \sum_{j=y-w\_size}^{y+w\_size} p_{i,j} \cdot i^l \cdot j^{k-l},\ l = (0,\ldots,k),$$

where k is the order of the geometric moments; x, y are the coordinates of the center of the element; w_size is the size of the element in both horizontal and vertical directions; $M_{x,y}^{l,k-l}$ are the geometric moments of the kth order, specifically: M^{00}, M^{01}, M^{10}, M^{02}, M^{20}, M^{11} for $k = 2$; $p_{i,j}$ is the pixel intensity at coordinates i, j.

Similarly, the geometric moments of the subregions of each image element can be computed in the same way.

Subsequently, transformations are performed to normalize the moments with respect to both translation and illumination variation. The transformations of the moments with respect to translation are defined as follows [16]:

$$\mu_{00} = M^{00},$$

$$\mu_{01} = M^{01} - \Delta y \cdot M^{01},$$

$$\mu_{10} = M^{10} - \Delta x \cdot M^{10},$$

$$\mu_{02} = M^{02} - 2 \cdot \Delta y \cdot M^{01} + \Delta y^2 \cdot M^{00},$$

$$\mu_{20} = M^{20} - 2 \cdot \Delta x \cdot M^{10} + \Delta x^2 \cdot M^{00},$$

$$\mu_{11} = M^{11} - \Delta y \cdot M^{10} - \Delta x \cdot M^{01} + \Delta x \cdot \Delta y \cdot M^{00},$$

where Δx, Δy are the distances from the top-left corner of the image to a given point (x, y) that is taken as the origin of the coordinate system, are denoted as $\Delta x = x$, $\Delta y = y$, since the top-left corner has coordinates (0, 0).

The translation is performed in such a way that the moments of the elements and their subregions are aligned with their respective centers.

The normalization of moments with respect to illumination, which affects the entire element and its subregions in approximately the same linear fashion, is carried out as follows:

$$\eta_{00}^{0,1,2,3} = \mu_{00}^{*0,1,2,3} / \mu_{00}^{*},$$

where $\mu_{00}^{*0,1,2,3}$, μ_{00}^{*} are the average brightness values of the pixels in the subregions and the element, $\eta_{00}^{0,1,2,3}$ are the normalized average brightness values of the subregions (the average brightness is calculated as the sum of the pixel brightness values within the subregions and the element, divided by the number of pixels in these areas, respectively); $\eta_{l,k-l} = \mu_{l,k-l} / \mu_{00}$ are the normalized moments of other orders of the image elements ($k = 2$, $l = 0,1,2$); $\eta_{l,k-l}^{0,1,2,3} = \mu_{l,k-l}^{0,1,2,3} / \mu_{00}^{0,1,2,3}$ are the normalized moments of other orders of the segments of the image elements ($k = 2$, $l = 0, 1, 2$).

It should be noted that the normalized moments $\eta_{1,0}$ and $\eta_{0,1}$ and $\eta_{1,0}^{0,1,2,3}$ and $\eta_{0,1}^{0,1,2,3}$ are the coordinates of the centers of gravity of the elements and their segments, respectively.

In this case, $f_0 = \eta_{1,0} \times \eta_{1,0} + \eta_{0,1} \times \eta_{0,1}$ and $f_0^{0,1,2,3} = \eta_{1,0}^{0,1,2,3} \times \eta_{1,0}^{0,1,2,3} + \eta_{0,1}^{0,1,2,3} \times \eta_{0,1}^{0,1,2,3}$ are the squares of the distances from the centers of the element and its segments to their respective centroids. In cases of significant shifts between the background image and the frames in the video sequence, alignment is recommended using the findings presented in [16]. If there is a difference in orientation, the approach described in [17] is advised. Alignment is performed by identifying and matching corresponding points in the images.

2.3. Comparison of Image Element Characteristics and Automatic Background Update. The background image is either pre-recorded or automatically selected from the initial frames of the video sequence, filtered, and stored by the operator. Importantly, it is not the image itself that is stored, but rather a matrix matching the size of the element grid. This matrix contains groups of calculated characteristics, the duration of changes detected in each element segment, and an indicator reflecting the absence of short-term changes across all segments of the image elements. The recommended set of characteristics for comparison, drawn from those described in the previous section, is determined through experimental studies as the most effective in terms of recognition accuracy and computational efficiency.

Each image from the video sequence is then processed in a similar manner. The difference lies in the fact that it is not necessary to store the characteristics of all image elements simultaneously. Instead, the image is scanned using a sliding window corresponding to the element size, and the characteristics of each element are computed individually and compared to the corresponding characteristics from the background image.

The complete set of element characteristics, calculated according to the algorithm provided in the previous section, for the background image (denoted by the subscript *fon*) and the analyzed image (denoted by the subscript *vid*), is as follows:

$$\eta_{00}^{0,1,2,3}\_fon,\ \eta_{l,k-l}^{0,1,2,3}\_fon,\ f_0^{0,1,2,3}\_fon,$$
$$\eta_{l,k-l}\_fon,\ f_0\_fon;$$
$$\eta_{00}^{0,1,2,3}\_vid,\ \eta_{l,k-l}^{0,1,2,3}\_vid,\ f_0^{0,1,2,3}\_vid,$$
$$\eta_{l,k-l}\_vid,\ f_0\_vid;$$

where $k = 2$, $l = 0, 1, 2$.

The characteristics make it possible to detect differences both in the image elements as a whole and in their individual subregions. In the following explanations, we will focus on identifying changes within the subregions of an element and will operate using the first group of characteristics, which are simple to compute and will be prioritized in experimental testing.

To identify differences, the relative errors of the characteristics of the i-th element in the background and analyzed images are calculated as Δ (*delta*), for example, for:

$$\eta_{00\_i}^{0}: delta_{00\_i}^{0} = 1 - \frac{\min(\eta_{00\_i}^{0}\_fon, \eta_{00\_i}^{0}\_vid)}{\max(\eta_{00\_i}^{0}\_fon, \eta_{00\_i}^{0}\_vid)},$$

and the errors are then compared with the maximum allowable error $\max\_delta_{00}$ for these characteristics:

$$if\,(delta_{00\_i}^{0} \geq \max\_delta_{00})$$
$$\{if\ (time_i^0 < \max\_time)\ time_i^0 = time_i^0 + 1;\}$$
$$\text{else}\ \{time_i^0 = 0;\}$$
$$if\,(delta_{00\_i}^{1} \geq \max\_delta_{00})$$
$$\{if\,(time_i^1 < \max\_time)\ time_i^1 = time_i^1 + 1\,;\}$$
$$\text{else}\ \{time_i^1 = 0\,;\}$$
$$if\,(delta_{00\_i}^{2} \geq \max\_delta_{00})$$
$$\{if\ (time_i^2 < \max\_time)\ time_i^2 = time_i^2 + 1\,;\}$$
$$\text{else}\ \{time_i^2 = 0\,;\}$$
$$if\,(delta_{00\_i}^{3} \geq \max\_delta_{00})$$
$$\{if\,(time_i^3 < \max\_time)\ time_i^3 = time_i^3 + 1\,;\}$$
$$\text{else}\ \{time_i^3 = 0\,;\},$$

where $\max\_time$, $time_i^{0,1,2,3}$ is the maximum duration and the time of change presence in the zeroth,

Fig. 2. Image from *a* video fragment illustrating a train station scenario: *a* — background image; *b* — a person has appeared but has not yet been marked as an abandoned object; *c* — the person has remained stationary and is marked as an abandoned object; *d* — recognition procedure performed, the object was identified as a person, the mark was removed; *e* — newly marked objects; *f* — objects with removed and retained markings

first, second, and third subregion of the element; $time_i^{0,1,2,3} = 0$ is the indicator of no change in the subregion of the element; $0 < time_i^{0,1,2,3} < \max\_time$ is the indicator of short-term changes in the subregion of the element allows filtering of potential dynamic interference; $time_i^{0,1,2,3} = \max\_time$ is the indicator of long-term changes in the subregion of the element indicates abandoned or missing objects in the video scene.

If the number of subregions with long-term changes within an element equals or exceeds G (where $G = 1...4$), that element is marked as part of an abandoned or missing object (for experiments, $G = 2$ has been used).

Automatic background updating, if this option is enabled by the operator, is performed in the absence of short-term changes in all subregions of the image elements.

2.4. Optional Feature for Classifying Abandoned Objects. A wide range of tools is available for object recognition. In this study, the YOLOv4 neural network [18] has been used in combination with the TensorFlow library [19], which provides extensive capabilities for image processing.

YOLOv4 is the most recent and advanced real-time object detector, delivering both high accuracy and fast performance. It supports the detection of approximately 70 object classes, including

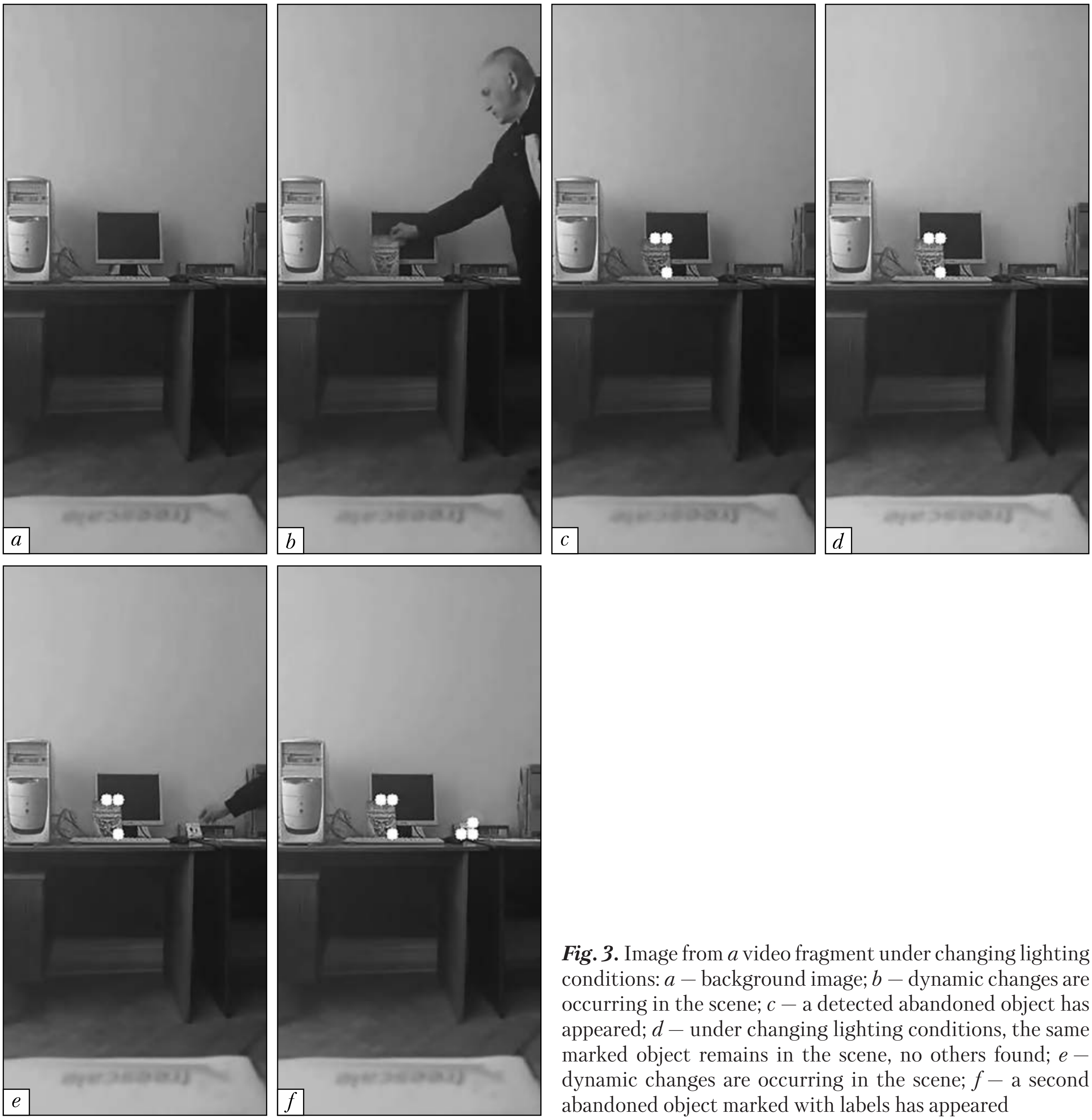

Fig. 3. Image from *a* video fragment under changing lighting conditions: *a* — background image; *b* — dynamic changes are occurring in the scene; *c* — a detected abandoned object has appeared; *d* — under changing lighting conditions, the same marked object remains in the scene, no others found; *e* — dynamic changes are occurring in the scene; *f* — a second abandoned object marked with labels has appeared

humans and animals, and can be applied to both image and video analysis, even under low-light conditions. The detector offers a variety of configuration options and parameters that allow tuning its sensitivity and detection speed based on specific task requirements. Due to its high efficiency and accuracy, YOLOv4 is widely used in fields such as video surveillance and the automotive industry.

TensorFlow is an open-source software library for machine learning, developed by Google, intended for constructing and training neural networks

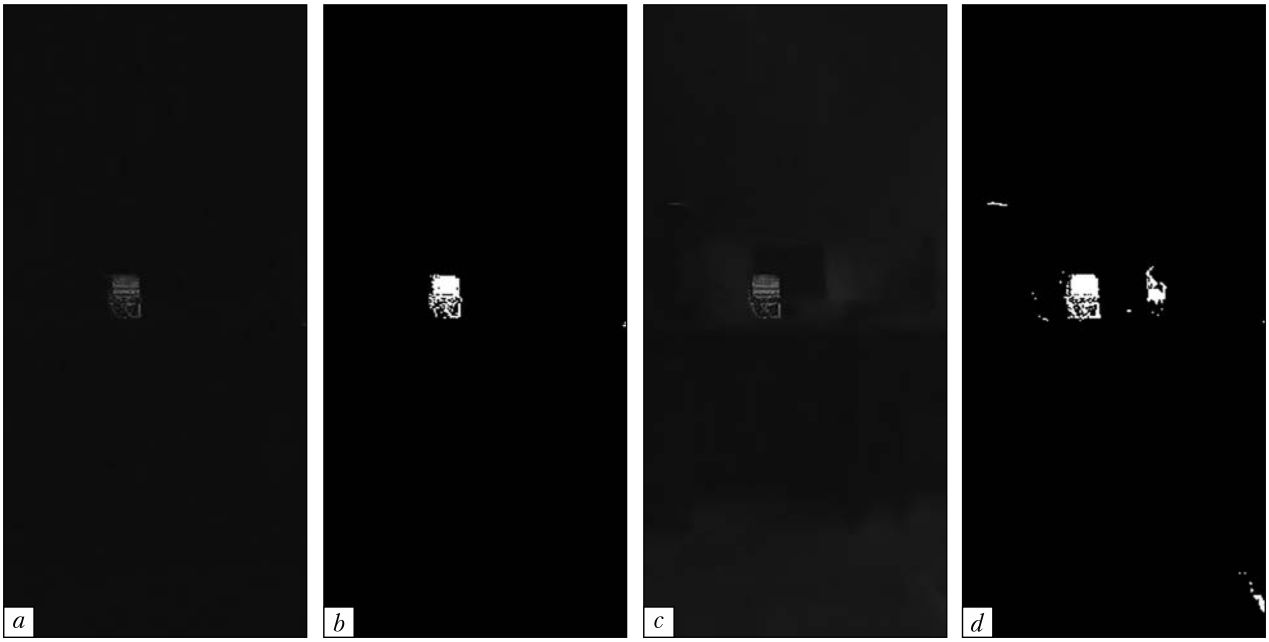

Fig. 4. Illustrations of background subtraction and threshold processing of obtained differences: *a* — result of background subtraction from the image shown in Fig. 3, *c; b* — result of background subtraction from the image shown in Fig. 3, *d*; *c* — result of constant threshold processing of the image shown in Fig. 4, *a; d* — result of constant threshold processing of the image shown in Fig. 4, *b*

capable of automatically locating and classifying objects at a level comparable to human perception. It is used both in research and in the development of Google's proprietary products. The primary API for working with the library is implemented in Python, with additional support for R, C#, C++, Haskell, Java, Go, JavaScript, and Swift. TensorFlow can operate on many parallel processors (CPU and GPU), leveraging the CUDA architecture for general-purpose GPU computing.

For the experimental studies, materials from PETS 2006 were used — a dataset specifically designed for testing algorithms for detecting abandoned and missing objects in public areas [20]. The ground-truth data for each video sequence includes information about the number of people and abandoned/missing objects present in the scene. The PETS 2006 dataset contains seven different scenarios with increasing scene complexity.

Approximately 14 fragments were analyzed, selected for their relevance to the detection of abandoned or missing objects. In all fragments, the method demonstrated stable performance. In some cases, small objects such as a pigeon that appeared and remained in place for a long time were detected; in order to classify such an object as abandoned, the threshold was set to $G = 1$. In addition, the method was tested under conditions of minor image shifts caused by temperature fluctuations and the resulting material expansion. The method was evaluated for a background image shift of 1 pixel, with results showing no deviation compared to the case with no shift. If larger shifts occur (greater than 1 pixel), image alignment based on distinctive points is required, as described in [16].

Since the analyzed image contains annotation marks after processing, for accurate human recognition the same image without markings is used.

Figure 4 shows the result of background subtraction and threshold processing of the obtained differences.

From Fig. 4, it is evident that simple background subtraction is significantly negatively affected by changes in lighting. When using a constant threshold, it was not possible to find a variant that would correctly highlight the object in Fig. 4, *d*. Applying automatic thresholding methods, such as Otsu's method, yields even worse results. Image shifts during simple background subtraction also emphasize object boundaries, which is a disadvantage.

Experimental results regarding the developed method for automatic detection of foreign objects in a video scene with dynamic disturbances indicate the following advantages:

- The algorithm demonstrated stable performance even with a minimal number of element characteristics on PETS 2006 video fragments;
- The developed method differs from traditional background subtraction methods in that it is not critical to lighting changes and minor image shifts;
- The proposed methods for calculating and comparing image element characteristics are time-efficient and allow real-time video data processing;
- Detection of small areas of short-term changes in the video scene enables effective filtering of dynamic disturbances;
- The method also proposes excluding detected humans or animals from consideration as abandoned objects.

The advancement of information technologies and image processing methods will enable the development and creation of more efficient and affordable video surveillance systems capable of operating in real time, relieving humans from exhausting video monitoring, and alerting them to important situations that may lead to undesirable consequences. This has significant social and economic importance, especially concerning the safety of enterprises and people, particularly under martial law conditions.

REFERENCES

1. Patryk, S., Vun, Yu. V., Komly, R. (2011). Performance in dicators for videotracking. *IETE Technical review*, 28(6), 493. https://doi.org/10.4103/0256-4602.90759. URL: https://www.researchgate.net/publication/272646487_Performance_Evaluation_Metrics_for_Video_Tracking (Last accessed: 20.06.2023).
2. Makfarlein, N. Dzh. B., Shofyld, K. P. (1995). Segmentation and tracking of pigletsinimages. *Machine vision and applications*, 8, 187—193. URL: https://link.springer.com/article/10.1007/BF01215814 (Last accessed: 20.06.2023).
3. Kukkyara, R., Hrana, K., Praty, A., Vetstsany, R (2005). Probabilistic pose classification for analyzing human behavior, transactions in systems. *IEEE Transactionson Systems, Man, and Cybernetics*, 35(1), 42—54. https://doi.org/10.1109/TSMCA.2004.838501
4. Huwer, S., Niemann, H. (2000). Adaptive change detection for real-time surveillanceapplications. *Proceedings Third IEEE International Workshop on Visual Surveillance*, 37—46. URL: https://ieeexplore.ieee.org/document/856856 (Last accessed: 20.06.2023).
5. Khalevy, H., Vainsholl, D. (1999). Disturbance motion: detecting and tracking non-rigid multi-body motion. *Machine Vision and Applications*, 11, 122—137. URL: https://link.springer.com/article/10.1007/s001380050096 (Last accessed: 20.06.2023).
6. Mokhamad, A. A. Kh., Osman, M. (2013). Adaptive median filter background subtraction technique using fuzzy logic. *International Conference on Computer, Electrical and Electronic Engineering (ICCEEE) (Khartoum, Sudan, 2013)*, 115—120. https://doi.org/10.1109/ICCEEE.2013.6633917
7. Haritaoglu, I., Harwood, D., Davis, L. S. (2000). W4: Real-time monitoring of people and their actions, pattern analysis and machine intelligence. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(22), 809—830. URL: https://3dvar.com/Haritaoglu2000W4.pdf (Last accessed: 20.06.2023).
8. Boult, T., Micheals, R. J., Gao, X., Eckmann, M. (2001). Into the woods: Visual surveillance of non-cooperative and camouflaged targets in compex outdoor settings. *Proceedings of the IEEE*, 89(10), 1382—1402. https://doi.org/10.1109/5.959337

9. Abbottand, R., Williams, L. (2009). Multiple target tracking with lazy background subtraction and connected components analysis. *Machine Vision and Applications*, 20, 93—101. https://doi.org/10.1007/s00138-007-0109-8
10. Harville, M. (2002). A framework for high-level feedback to adaptive, per-pixel, mixture-of-Gaussian background models. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (Eds.) Computer Vision — ECCV 2002. *Lecture Notes in Computer Science*, 2352, 28—31. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-47977-5_36
11. Wang, J. X., Bebis, G. N., Miller, R. (2006). Robust video-based surveillance by integrating target detection with tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (New York, NY, USA, 2006)*, 137—137. https://doi.org/10.1109/CVPRW.2006.180
12. Venetianer, P., Zhang, Z., Yin, W., Lipton, A. (2007). Stationary target detection using the object video surveillance system. *Advanced Video and Signal-based Surveillance*, 2007. https://doi.org/10.1109/AVSS.2007.4425317
13. Yaoand, J., Odobez, J.-M. (2007). Multi-Layer Background Subtraction Based on Color and Texture. *IEEE Conference on Computer Vision and Pattern Recognition* (*Minneapolis, MN, USA, 2007*), 1—8. https://doi.org/10.1109/CVPR.2007.383497
14. Taycher, L., Fisher, J. W., Darrell, T. (2005). Incorporating Object Tracking Feedback into Background Maintenance Framework, (*2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05), Volume 1 (Breckenridge, CO, USA, 2005)*, 120—125. https://doi.org/10.1109/ACVMOT.2005.63
15. Hlumov, N. Y., Miasnykov, V. V., Serheev V. V. (2010). *Detection and recognition of objects in images: electron. Textbook allowance*. Samara. https://repo.ssau.ru/handle/Uchebnye-posobiya/Obnaruzhenie-i-raspoznavanie-obektov-na-izobrazheniyah-Elektronnyi-resurs-elektron-ucheb-posobie-54634 (Last accessed: 20.06.2023).
16. Sabelnikov, P., Sabelnikov, Yu. (2023). Search for identical points in the inter-pixel space of videoimages. *Artificial Intelligence*, 28(1), 155—162. https://doi.org/10.15407/jai2023.01.155
17. Sabelnikov, P. Iu., Sabelnikov, Yu. A., Syniakov, M. M. (2023). Determining the orientation of circularareas in videoimages. *Artificial Intelligence*, 28(3), 131—138. https://doi.org/10.15407/jai2023.03.131 [in Ukrainian].
18. YOLOv4: Highspeed and accurate object detection. URL: https://docs.ultralytics.com/ru/models/yolov4/ (Last accessed: 12.06.2023).
19. Tensor Flow. URL: https://uk.wikipedia.org/wiki/TensorFlow (Last accessed: 23.12.2021).
20. Dataset PETS 2006. URL: https://www.researchgate.net/figure/PETS-2006-Dataset-Sequence-1-a-original-image-b-background-patches-that-can-not-be_fig3_236630140 (Last accessed: 23.12.2021).

Received 22.07.2024
Revised 03.01.2025
Accepted 09.01.2025

П.Ю. Сабельніков (https://orcid.org/0000-0003-4709-8693)

Інститут кібернетики ім. В.М. Глушкова НАН України,
просп. Академіка Глушкова, 40, Київ, 03187, Україна,
+380 44 526 2008, incyb@incyb.kiev.ua

МЕТОДИ АВТОМАТИЧНОГО ВИЯВЛЕННЯ СТОРОННІХ ОБ’ЄКТІВ У ВІДЕОСЦЕНІ З ДИНАМІЧНИМИ ПЕРЕШКОДАМИ

Вступ. Виявлення сторонніх об’єктів та залишених предметів часто застосовується у системах відеоспостереження, встановлених у місцях масового скупчення людей для запобігання терористичним актам або для їхнього розслідування.

Проблематика. Системи відеоспостереження генерують величезну кількість інформації, яка має бути відфільтрована для виявлення потенційно загрозливих ситуацій, і цілком логічно перекласти частину роботи людини-оператора на системи автоматичного аналізу інформації (відеоаналітики) від групи відеокамер, звільнивши таким чином людину від досить монотонної роботи. Оскільки оператор може не відслідкувати всі зміни у відеосцені, виникає потреба створення методів і алгоритмів автоматичного аналізу відеоінформації для спрощення роботи людини-оператора.

Мета. Розробка методів автоматичного виявлення сторонніх об’єктів у відеосцені з динамічними перешкодами для запобігання можливим правопорушенням або терористичним актам і розкриття цих подій в установах та підприємствах країни.

Матеріали й методи. Для експериментальних досліджень використано набір даних *PETS* 2006 (сім сценаріїв із зростаючою складністю сцени), спеціально створений для тестування алгоритмів детектування залишених та зниклих об’єктів у громадських місцях.

Результати. Розроблено метод, не критичний до зміни освітлення і незначного зсуву зображень, та проведено експериментальні дослідження щодо автоматичного виявлення сторонніх об’єктів у відеосцені з динамічними перешкодами. Відповідний алгоритм показав стійку роботу навіть за мінімальної кількості характеристик елементів на відеофрагментах із *PETS* 2006.

Висновки. Використання запропонованого методу дозволить системі відеоспостереження працювати в режимі реального часу, звільняючи оператора від виснажливого нагляду за відеосценою та сповіщаючи про виявлення залишених предметів, що можуть нести потенційну небезпеку.

Ключові слова: відеосистеми, системи відеоспостереження, відеофрагмент, відеопослідовність.