

A Cost-Effective Method for Detecting and Tracking Moving Objects using Overlapping Methods

Yuanyuan ZHANG^{1*}

College of Economic and Management, North China Institute of Science and Technology
Langfang, Hebei 065201, China

Abstract—Overlay approaches for moving object detection and tracking have recently received attention as a crucial field for computer science and computer vision research. Using pixel overlap and visual attributes, these techniques enable the recognition and tracking of objects in movies or video data. Two color and edge features for the suggested method are presented in this article. The suggested approach uses the SED algorithm, and since the edges have a lower volume than the entire image, the processing process will be faster with the reduction of information. The characteristic of color is the HSV (hue, saturation and value) histogram because it is close to human vision. However, because the margins tidy up the shapes in the human eye, they contain important information. These concerns lead to the conclusion that the histogram of gradient angles based on regional binary patterns is the edge feature of the suggested system. There are two justifications for employing local binary patterns. First, the principal edges are emphasized by using local binary patterns. Another point is that the image produced by this method displays the image's texture; in other words, the shape's feature is taken from the context of the texture, which is regarded as a type of combination of features. Several criteria were evaluated in order to assess the suggested approach for tracking images in comparison to related systems; the most significant of these are the precision, recall, and similarity criteria. In comparison to other works, the findings for precision have generally increased accuracy by 25%, recall by 17%, and similarity by 12%.

Keywords—Tracking; moving object detection; image processing; binary patterns; HSV histogram

I. INTRODUCTION

Today, object identification, particularly machine vision and pattern recognition, is one of computer sciences most significant and broad study domains. Even after an object's appearance has changed, the human brain is still able to instantly recognize and categorize many different types of items [1]. Different characteristics, such as changes in brightness, state, texture, shape, and object occlusion, affect how well the human visual system can identify things [2]. Additionally, the human brain has the capacity to extrapolate its findings from a collection of things and reliably recognize ones it hasn't yet seen [3]. Scientists working in the fields of machine vision and computer science have used the study of the human brain's cognitive capacity to recognize and draw inspiration from it to design and present a variety of object recognition systems [4]. In order to organize visual information, these findings are also used in related applications like picture concept recovery and image indexing [5]. Visual characteristics are the primary source of information used by

conventional object recognition techniques to identify items in real-world photographs [6]. To a certain extent, visual characteristics, including color, form, texture, and picture edges, can compensate for changes in an object's appearance. Additionally, new methods have attempted to explain the interactions of objects in a scene or general statistical notions using conceptual features [7]. These methods for object identification applications increase recognition precision and clear up conflicts those traditional systems, with their reliance on visual features, experience [8]. There is a noticeable increase in image data globally, and that growth rate is accelerating daily. According to Information Trends, more than 1.1 trillion images were taken in 2016 with cameras and mobile devices [9]. The same forecast predicts that by 2020, this amount will have increased to \$1.4 trillion. Many of these photographs are made available online or through cloud services. In 2014, prominent websites like Instagram and Facebook had daily photo uploads of more than 1.8 billion [10]. Beyond consumer electronics, there are cameras everywhere that take pictures for automation. Traffic cameras and moving cars are both watching the road. Robots need to be able to see in order to categorize objects and get rid of trash intelligently. Engineers, medical professionals, and explorers of space all employ imaging instruments. We need to have an understanding of this data's contents in order to manage it successfully. A wide range of image-related tasks benefits from automated content processing [11]. This requires computer systems to bridge the "semantic gap" between surface pixel information stored in image files and how people perceive analogous images. Computer vision is applicable in this area. Images can contain objects that can be automatically found and recognized. One of the core issues with computer vision is what is known as object detection. We'll demonstrate that convolutional neural networks now offer the best method for detecting objects [12]. Examining and putting to the test convolutional object identification techniques is the main goal of our research. The goal of computer vision is to derive useful information from the content of digital photographs or videos. It's just straightforward image processing, which entails fiddling with visual data down to the pixel level. Applications of computer vision include traffic automation, picture classification, visual identification, image retrieval, augmented reality, machine vision, and reconstruction of 3D scenes from 2D images [13]. Image tracking is one way to detect moving objects. An image tracker is a system that sequentially tracks preset items in a series of photographs [14]. In other words, the tracking process is the act of estimating the temporal and spatial changes of the object or, more generally, the states of

the target object during the video sequence based on measurements and observations. The target object may be specified by detection algorithms or manually [15]. Using data from earlier frames and additional data, such as the target's movement model or appearance attributes, the tracking system calculates and looks for the target in the current frame. An innovative technique for tracking images is given in this study. Color and edge features, the SED technique, and local binary patterns have all been utilized in image analysis to extract information from images and moving objects. The application of this technique highlights how closely color qualities resemble human sight and how crucial the information at the margins is. This approach can also aid in developing algorithms with increased processing speed and accuracy for picture recognition. In general, these developments can advance the science of picture tracking and enhance the functionality of systems for image analysis and recognition. The writers' contribution to this study can be summed up as follows:

- Find objects in the image and background dataset.
- Image tracking using color and edge features.

The remainder of the essay is structured as follows: Section II highlights earlier research on object detection using various image processing methods. Section III contains the suggested concept and approach. Section IV discusses the outcomes of the evaluation and simulation and Section V contains the conclusion and recommendations for future work.

II. RELATED WORKS

The study gap that this research tries to solve is improving the efficiency and accuracy of moving object detection and tracking in computer vision and image processing. As mentioned in the introduction, to solve this gap, the proposed solution for tracking accuracy and detection combines color and edge features, uses SED algorithm and local binary patterns to increase detection and tracking of moving objects. The difficult task of identifying things in the image is one that numerous researchers are presently doing. The long-term objective of picture understanding is to recognize all objects in a general scene; however, this is still difficult due to factors like intra- and extra-class diversity, state and location, backdrop complexity, overlap, significant illumination variations, etc. Numerous publications published recently compared the effectiveness of various approaches' identification rates and mistake rates. However, a number of factors, including learning and execution times, the number of training samples, and the proportion of the mistake rate to the recognition rate are important when evaluating algorithms. The comparison is further complicated by the fact that researchers' definitions of recognition and error rate differ. The research in [16] proposes a new technique for employing a stereo camera to detect objects with many colors distributed unevenly in complicated backgrounds and then estimate the depth and form of the object. In this study, color saturation space is separated into fuzzy color histograms based on self-clustering in order to extract characteristics for object detection. A fuzzy color histogram is created for each window scan in a pyramid of graded images by adding the fuzzy degrees of all the pixels in each cluster. The right and left photos are initially segmented

using color space to identify the matched item region in the right image. The study in [17] provides an overview of current advances in the field of object recognition in remote-sensing photos. Many studies have been conducted to find things in aerial and satellite photos throughout the last few decades. They are separated into four primary groups in this article: machine learning-based approaches, knowledge-based methods, object identification methods based on object-based picture analysis, and methods based on template matching. This page explains the categorization of object recognition research. Depending on the format a user chooses, there are two additional types of matching methods: hard pattern matching and metamorphic pattern matching. In [18], a context-oriented salient Bayesian model is put out to address the problems of scale variance and detection ambiguity in small item identification. The sea is the object of this article's visual analysis. It is possible to understand that there is a link of reliance between the place and the scale at which things may occur by looking at the geometry of the camera in the image against the background of the sea and sky. The model described in this study is a universal model that may be applied to many contexts with various images to facilitate the object. The research in [19] presents a comprehensive review of the statistical learning-based representation of features in object recognition. This article compares the evaluation outcomes of object detection algorithms with various visual features and categorizes visual features based on differences in computations and visual attributes. Therefore, the objective of this review is to develop a thorough and complete plan for researchers. The portrayal of the influence of features is a concern when taking into account the demands of generic object recognition. To increase the representational strength of object recognition models, it is important to acquire extensive and powerful visual features. The display of comprehensive features can be effectively removed by combining various visual property aspects. An innovative component-based method for object detection on a two-dimensional image and its use as a visual landmark has been presented by researchers in [20]. Object recognition is a hybrid cryptographic system that makes it possible to keep track of the topology and use it to power the recognition procedure. As shown in the aforementioned study, it is challenging to separate the components of an item from a two-dimensional image; it is only logical for the image to represent either the upper or lower component. Determining the object and its representation from the existing image and the various numbers deduced from the pieces is the goal of this research. The study in [10] presents a brand-new energy function based on the autocorrelation function for active contour models, allowing for recognizing small objects against textured and chaotic backgrounds. To show the information of each area, the proposed method calculates picture characteristics for each pixel in the image domain using a combination of short-term autocorrelations. A novel energy function dubbed "normalized accumulated short-term autocorrelation" is introduced for the active contour based on the localized area using the collected features. Small items can be recognized in pictures with cluttered backgrounds and heterogeneous textures by decreasing this energy function. Researchers in [21] have presented a brand-new method for picture edge identification based on ACO. In the suggested

approach, coupled optimization techniques have been applied, which has aided in speeding up the process of solving optimization issues. In this method, artificial ants first produce a number of answers, the information from which is then useful for the genetic algorithm. These answers then serve as the initial population for the genetic algorithm, and the genetic algorithm then produces the subsequent population from these answers. A new edge identification technique for satellite pictures with low contrast has been proposed in another study [22]. It has been mentioned in this article that it is quite challenging to extract edges from satellite photos with low contrast, smoothness, and features. Therefore, the Sobel edge

detector performs preprocessing on the input image. Low-pass and high-pass filter operations are carried out on the normalized images after the input image has been normalized. The relaxation factor comes after the output of high-pass and low-pass filtering. Following this filtering process, the preprocessed image is subjected to the Sobel edge detection method for edge identification. By identifying the features that work best for identifying things in a picture, the structure of the object identification system can be reduced more successfully without the need to expand the training set, and a better outcome will likely be seen as a result. A comparison of earlier investigations is provided in Table I.

TABLE I. COMPARATIVE STUDY OF THE METHODS PRESENTED IN PREVIOUS STUDIES

Simplicity	Area	proposed method	Application	Ref.
Simple and easy to implement	Stereo camera	Fuzzy color histograms	Spatial	[14]
Simple method	Optical remote sensing images	Review of object recognition methods	conversion	[15]
Simple method	Infrared images	An outstanding text-oriented Bayesian model	conversion	[16]
Simple method	Decomposing images into different areas	Wavelet method	Spatial	[17]
The simple and fast method	Digital images	learning-statistics	conversion	[18]
The simplest and most suitable for dark images	driving	Image segmentation, color composition and topology	Spatial	[19]
Simple method	Identifying small objects	Contour model	Spatial	[8]
Simple method	Moving Pictures	Creating and testing hypothesis and verification step	Spatial	[20]
It provides a better visual effect	Digital images	meta-heuristic algorithm	Spatial	[21]
It provides a better signal than noise	Satellite Images	High-pass and low-pass filtering	Spatial	[22]

III. PROPOSED METHOD

The proposed system has two techniques for accuracy and speed, which have been reviewed and put into practice. Two qualities of color and edge are taken into consideration in the suggested strategy. The provided technique uses the SED algorithm, and since the edges have a lower volume than the entire image, with the reduction of information, the processing process will be faster, from the HSV histogram's characteristic of color to its nearness to human perception. Because shapes are cleaned up by their edges in human vision, other edges carry important information. These concerns lead to the conclusion that the histogram of gradient angles based on regional binary patterns is the edge feature of the suggested system. There are two justifications for employing local binary patterns. First, the principal edges are emphasized by using local binary patterns. Another point is that the image produced by this method displays the image's texture; in other words, the shape's feature is taken from the context of the texture, which is regarded as a type of combination of features. The proposed algorithm's flowchart is shown in Fig. 1.

A. Feature Extraction using SED Algorithm

The SED approach efficiently displays picture information and simultaneously extracts color and texture features. Five

structural formats are utilized to determine the texture perception from the original color image, separated into 72 colors in the HSV color space. A three-step process is employed to get the final descriptive image of the structural elements:

- With step 2, move the SED 2x2 from left to right and from top to bottom, starting at the reference point (0, 0).
- This value is saved if the structure's elements match the image's value (a match denotes that the value of the image in the related structure is equal).
- Then, the SED map is obtained, which has five $S_i(x, y) (1 \leq i \leq 5)$ structures and is distinguished by. When the components of the nondirectional representation structure are recognized, four additional structures in particular need to be found. Since there is no direction, every direction is feasible.
- The following equation, generated by combining five maps as in Eq. (1), illustrates the final SED map.

$$S(x, y) = \{(x, y) | S_1(x, y) \cup S_2(x, y) \cup S_3(x, y) \cup S_4(x, y) \cup S_5(x, y)\} \quad (1)$$

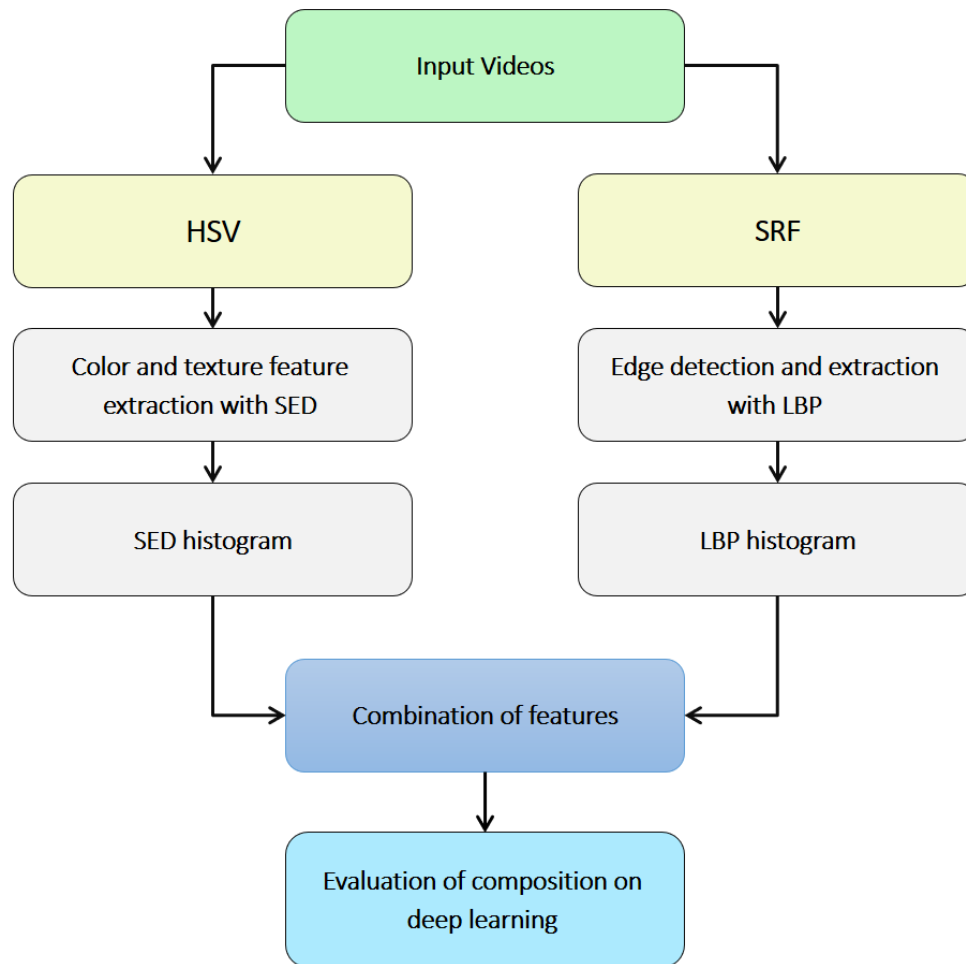


Fig. 1. Block diagram of the proposed method.

B. HSV Color Histogram

The use of color histograms, one of the most significant picture-based recovery approaches, is the most used technique for recovering the color of an image. The image is provided by this histogram, which is produced as a cumulative distribution of the sub-distances connected to each pixel. Defining the color space is the first step in producing a color histogram. Generally speaking, the RGB color space is present in a number of picture formats, including BMP, GIF, etc. We will have fewer calculations if we choose this space, but the main issue is the non-uniformity from the perspective of perception. Because of this, HSV space has been substituted for RGB space in the proposed method. The following is an example of an HSV space benefit. Its suitability with how humans perceives color is due to the independence of the color type and brightness components, which allows us to assign superiority to any one of the components (see Fig. 2). Based on category (H) or wavelength, level of color saturation (S), and level of brightness (V), the dimensions of this space constitute color. Conical in shape, the aforementioned space contains:

- The wavelength of a color is defined in the range $[0, 2\pi]$, where red is at an angle of 0 degrees, green is at an angle of $2\pi/3$, and blue is at an angle of $4\pi/3$, the final

angle. The wavelength of color is equal to the angle of the color in the section of the cone circle.

- At an angle of $\pi/2$, it changes back to red once more.
- The central axis of the cone corresponds to the brightness of the color.
- The amount of color saturation varies depending on how far a given point on the circle is from the center axis; the closer a color is to the axis, the less concentrated it is, and the more it resembles a gray hue.

Because it is more stable than changes in the direction of photographing an item or scene, H is the primary component employed in retrieval systems. It is, nevertheless, extremely sensitive to variations in light. After this space, it is linearly quantized to create the color histogram. Because H space is more significant than other components, the other two components are quantized into four intervals each, but this component is quantized into 16 intervals. By counting the points positioned in each interval and normalizing them to the overall number of picture points, the color histogram of the image is created. A 256-dimensional vector is produced after indexing each image using the HSV color histogram feature.

C. Texture Feature Extraction

The methods for texture-based recovery are based on the assessment of texture features in the images, such as texture contrast, roughness, texture direction, texture regularity or texture periodicity, and texture unpredictability.

In order to speed up processing, the input color photographs are converted to grayscale at this point. A Gaussian filter is then applied to the grayscale image to help remove any potential noise. Eq. (2) defines a Gaussian function as follows:

$$w(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

Then the image is averaged: Consider the noisy image $g(x,y)$, which is obtained from adding noise $n(x,y)$ with the original image $f(x,y)$.

$$g(x,y) = f(x,y) + n(x,y) \quad (3)$$

In this case, it can be easily shown with Eq. (4) that by averaging the noisy image of an object, a good approximation of the original image M can be obtained.

$$\check{g}(x, y) = \frac{1}{m} \sum_{i=1}^m g_i(x, y) \quad (4)$$

Many edge detection algorithms use the first derivative of illumination, which means that we operate with the original data's illuminance gradient. With this knowledge, the peaks of the brightness gradient can be looked for in an image. The second derivative of brightness intensity, which is actually the rate of change of the brightness intensity gradient and is the best for detecting textures, is the basis for some other edge detection algorithms. As a result, a brightness gradient can be seen on one side of the line and its opposite gradient on the other. As a result, we can anticipate a fairly significant shift in the illumination's gradient in intensity at the site of a line. You can look for the transition of gradient change zeros in the results to discover bright textures. If $I(x)$ is the light's intensity, then Eq. (5) [12] is as follows: To be able to extract texture features from photos is the aim. A multidimensional texture feature vector is believed to be the output of a color image, which is considered to be the input.

$$L(x, y) = \nabla^2 f(x, y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \quad (5)$$

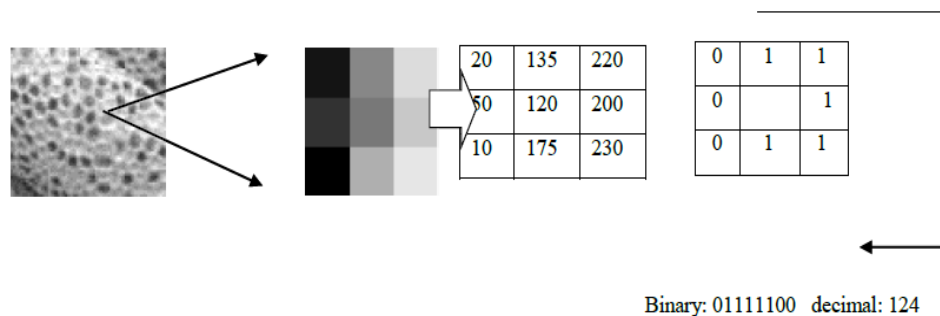


Fig. 3. How to calculate the LBP operator.

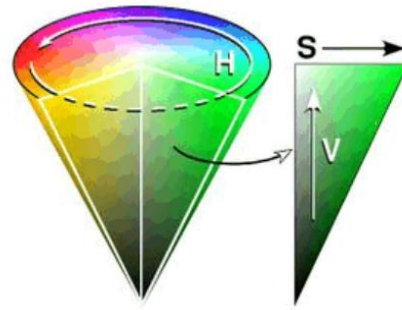


Fig. 2. HSV color space.

D. Edge Extraction using Local Binary Patterns

As a potent image texture descriptor, the primary LBP operator was first introduced in [23]. This operator creates a binary number for each individual pixel using the nearby 3x3 pixel labels. The value of nearby pixels is used to threshold these labels. The value of the center pixel is found. In this approach, a label of 1 is placed for pixels whose values are more than or equal to the value of the central pixel, and a label of 0 is placed for pixels whose values are lower than the value of the central pixel. They start to form. Fig. 3 illustrates how this operator functions.

The LBP operator's 3x3 neighborhood base is too small, which prevents it from dominating large-scale images. This operator, which is shown as LBPP, R, and can produce a maximum of p^2 different values according to p^2 of the binary pattern produced by P pixels on the neighborhood radius of R , was later proposed for this purpose with an extension of neighborhood size in [24]. Fig. 4 illustrates how to choose adjacent pixels in this kind of arrangement. It displays three different local binary radii.

Fig. 5 illustrates the unique significance and notion attached to the patterns created by LBP. This figure shows that LBP is capable of detecting points, edges in various directions, smooth areas, line ends, etc. The histogram of these labels is calculated after assigning LBP labels to pixels in order to create image texture. Numerous researches have concentrated on the use of LBP because of its ease of calculation and acceptable findings, and in this regard, new and varied variants of it have been suggested for usage in various circumstances [25].

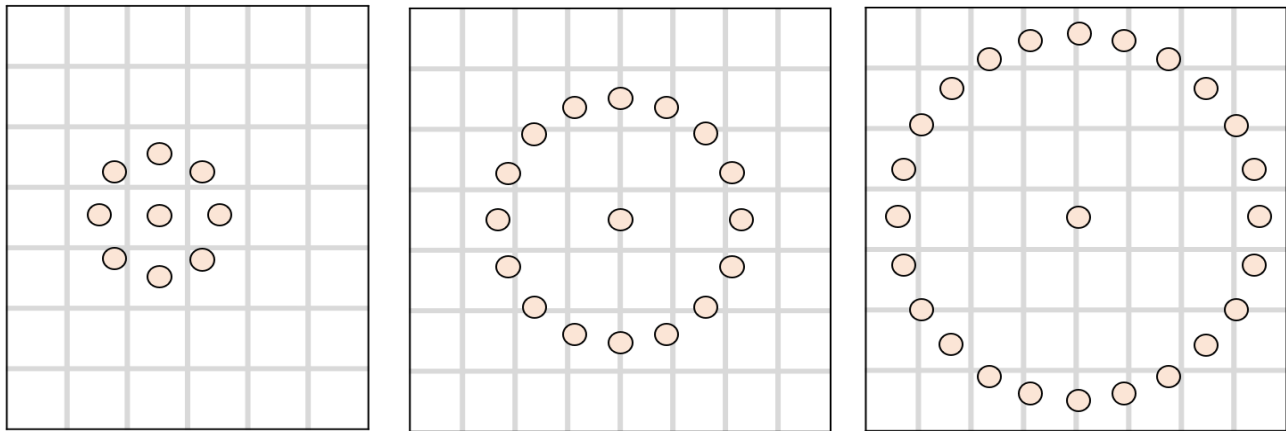


Fig. 4. LBP operator with different radii and number of neighborhoods.

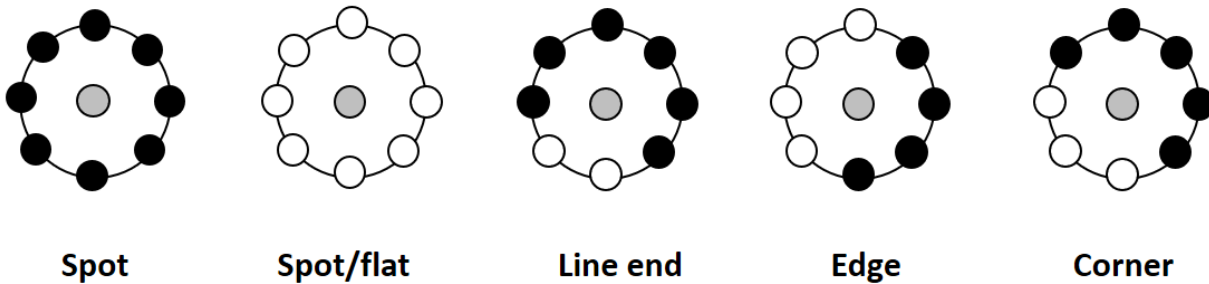


Fig. 5. Examples of the concept of patterns produced in LBP, the white and black circles represent the number one and zero, respectively.

After the pixels are labeled by the LBP operator, a histogram of the produced patterns is defined as Eq. (6).

$$H_i = \sum_{x,y} I(f_i(x,y)=i), \quad i=1, \dots, n-1 \quad (6)$$

where n is the number of patterns produced by the LBP operator and the function I is defined as Eq. (7).

$$I(A) = \begin{cases} 1 & A \text{ is True} \\ 0 & A \text{ is False} \end{cases} \quad (7)$$

Local binary patterns with radii of 1 and 8 neighbors are considered in the suggested strategy. Finding and obtaining most systems utilize the closest neighbor search to look for and locate related photos when searching for photographs. For each query image, the KNN classification algorithm examines the set of training images for related images. If the system employs n features, it will treat the images as vectors in the following n -dimensional space. K seeks the neighbor of the query image from the training photos, and among these neighbors, the class label that is in the majority is used as the category label of this image so that each image in this space is comparable to a point. The question foretells Finding and calculating a criterion for the similarity or distance between attributes in the data is the first task to be completed using k -NN. One of the challenges of this method is calculating the distance between the photos; if the distance between the images is not accurately determined, the algorithm will not function properly.

E. Combinations of Features

After the feature extraction phase, it should be decided how to combine them. There are other ways to combine the features, but the simplest is to arrange them in a row to create a

single matrix and then use other photos to calculate distance. This work has a number of issues. The majority of the features have different sizes, so when combined, the feature with the larger size will have a greater effect. However, larger size does not always imply higher efficiency or greater importance. It is also impossible to compare the effects of some features according to performance. It excels in other aspects. A more effective solution: weighting and distance calculation are done separately for each feature. There are two features here with the designations $F1$ and $F2$, respectively. The letters $w1$ and $w2$, respectively, designate the weights assigned to each attribute. In this manner, the effect of the features changes depending on the coefficient that $w2$ accepts, and the best weight can be attained with various variations. The edge feature's weight in the suggested technique is 0.4, while the color feature's weight is 0.6.

F. Detection Criteria with Deep Learning

Auto-indicators are one of the techniques of deep learning. An advanced form of artificial neural network called an auto-encoder is used to discover the best code. With this technique, you train an auto-encoder to reconstruct its input X rather than train the network to predict the target value Y along the input X . As a result, the output vectors will match the input vector's dimensions. Fig. 6 shows the general operation of an auto-encoder. The auto-encoder is improved during training by reducing the reconstruction error. It learned the appropriate code for the same feature [26].

The detection of moving targets in overlapping scenes will be carried out in this study with the use of an auto-encoder and a deep-learning technique. Fig. 7 illustrates the general

workflow of the method, which includes the following steps: (1) receiving the image; (2) filtering and enhancing the image; (3) modeling the background; (4) choosing one of the moving

images; (5) image training using deep learning; (6) target tracking with the aid of an auto-encoder and deep learning; (7) feature extraction; and (8) classification.

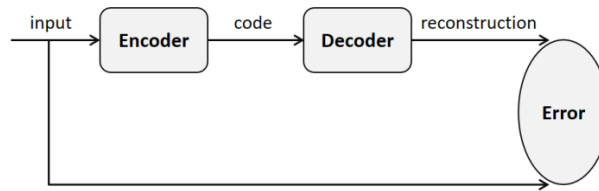


Fig. 6. General process of an auto-encoder.

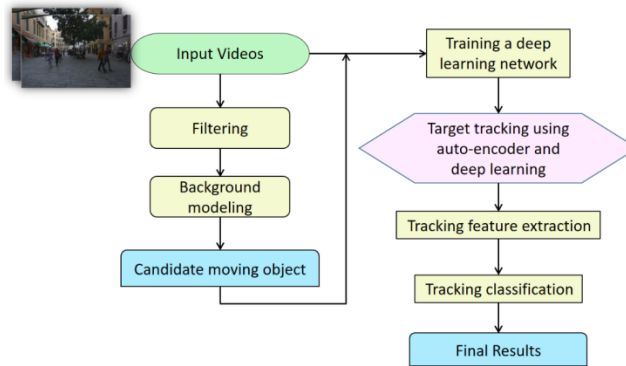


Fig. 7. Steps to perform work to detect moving targets.

The training photos, which contain the target images and their labels, are first fed into the deep learning algorithm, as can be seen in the flowchart of how the suggested approach is implemented. After training the network, it will be able to track the target in the test images. Higher-accuracy architecture will be selected while developing the algorithm's architecture [27]. The incoming moving images are first given a Gaussian filter, which suppresses some random noises that frequently emerge near object borders. In this step, several redundant and erroneous objects are eliminated from the results of the object detection process. A non-parametric background modeling approach [27, 28] is chosen after choosing the moving item in the second phase to produce the initial detection results of the moving object with high recall but low precision. Target tracking is recovered in the third stage by utilizing an auto-encoder and deep learning techniques that are derivations of a hybrid method. Then, characteristics are taken for each detector and applied to distinguish between actual moving targets and fake ones. The last stage can be utilized as a constraint to accurately inspect moving objects and obtain high accuracy and recall by combining the autoencoder and deep learning techniques.

IV. EVALUATION AND SIMULATION

Image tracking of moving targets in video images is particularly important in machine vision and widely used in robotics and automatic video surveillance. The importance of this issue lies in making the systems smarter and improving the accuracy and speed of the machine's decision-making by receiving visual information. In this research, the aim is to investigate image tracking algorithms based on deep learning

methods and provide a new tracking method based on deep learning for more accurate target tracking.

A. Data Sets

The dataset available at <https://motchallenge.net> will be used as the dataset of this research, which includes consecutive images of targets in different lighting conditions and with different overlaps. Fig. 8 shows an example of images from the dataset.

B. Comparison Methods

Deep learning methods and convolutional networks have had good results in tracking moving targets. For example, in study [10], a method based on convolutional neural networks is presented in study [10] to track targets in crowded and noisy environments with low contrast, provided that it is able to track moving targets using color. In research [16], a method based on deep learning to track targets is presented. In this method, first, the features of the target are learned by a very deep network, then these features are sent to the support vector machine, and finally, the support vector machine can identify the goals.

C. Precision Criteria

This criterion shows the type of accuracy; the higher it is, the better; this criterion shows the accuracy of the proposed method for positive cases and how accurate it was in identifying positive cases. This criterion is one of the most important accuracy criteria in detection algorithms and is calculated from Eq. (8).

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

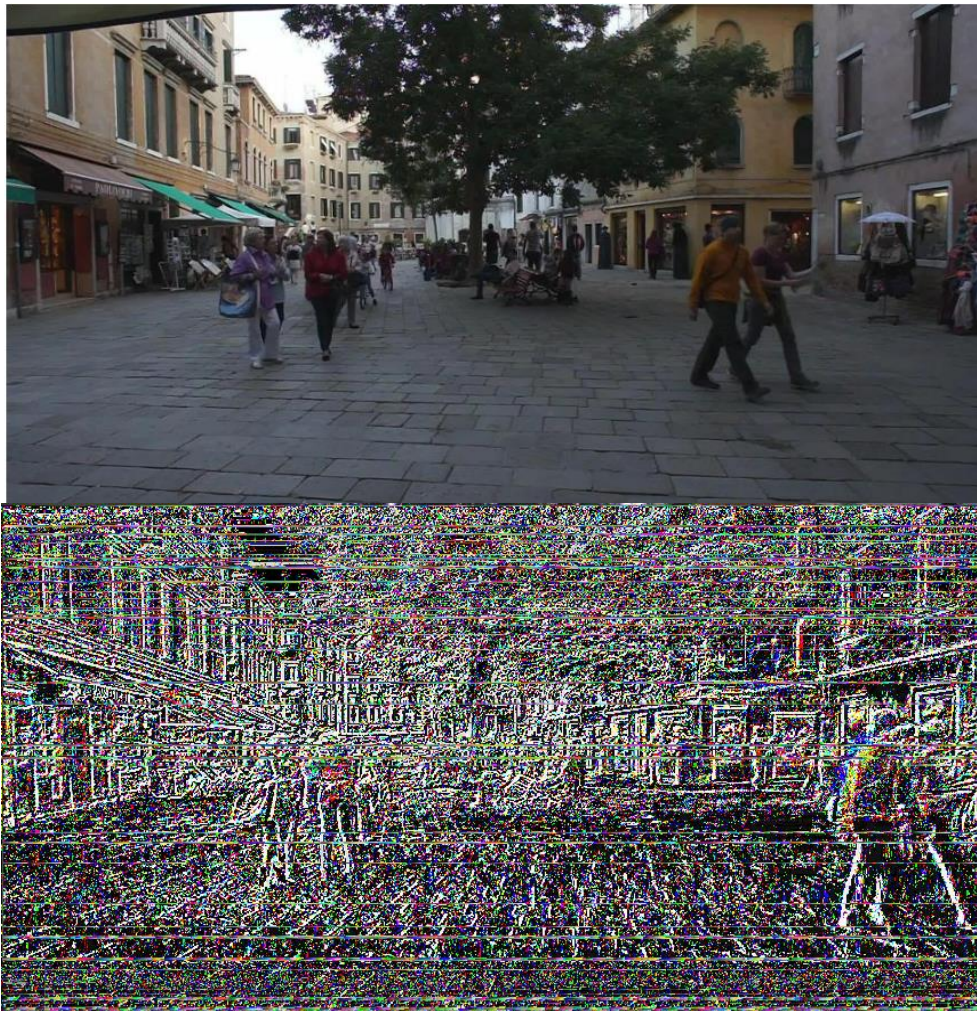


Fig. 8. An example of images from the dataset.

In this regard, TP is the number of data that are correctly recognized as positive, and FP is the number of data that is falsely recognized as positive. The simulation result based on the above criterion is shown in Fig. 9.

The simulation result shows that the proposed method performs better in this criterion and has been able to perform image recovery operations with higher accuracy. In Table II, the result of the Precision criterion can be seen in different simulation situations.

D. Recall Criteria

The next measure of accuracy is the Recall measure; this criterion checks the accuracy of the method in identifying negative states, that is, how accurate the method was in identifying negative states, which is calculated through Eq. (9).

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

In this regard, TP is the number of data that are correctly recognized as positive, and FN is the number of data that is falsely recognized as negative. The recall criterion is objectified according to the above scenario, and the result can be seen in Fig. 10.

The method presented in this research has a better performance in this criterion according to the investigation carried out by the simulator; in Table III, the result of this criterion can be seen in different simulation situations.

E. Accuracy Criteria

The next evaluation criterion is the accuracy criterion. This important criterion is calculated based on Eq. (10).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

In the above relation, the TP parameter represents the number of images that have been correctly retrieved; and the FP parameter also indicates the number of images that have a negative effect and the proposed models have predicted that sample negatively. FN represents the number of samples that have a negative effect on image recovery, and the proposed model has positively predicted these samples. TN represents the number of samples that have had positive effects on recovery, and the proposed method also predicts them to be negative in the recovery process; therefore, with the help of relation (10), the level of accuracy can be seen in Fig. 11.

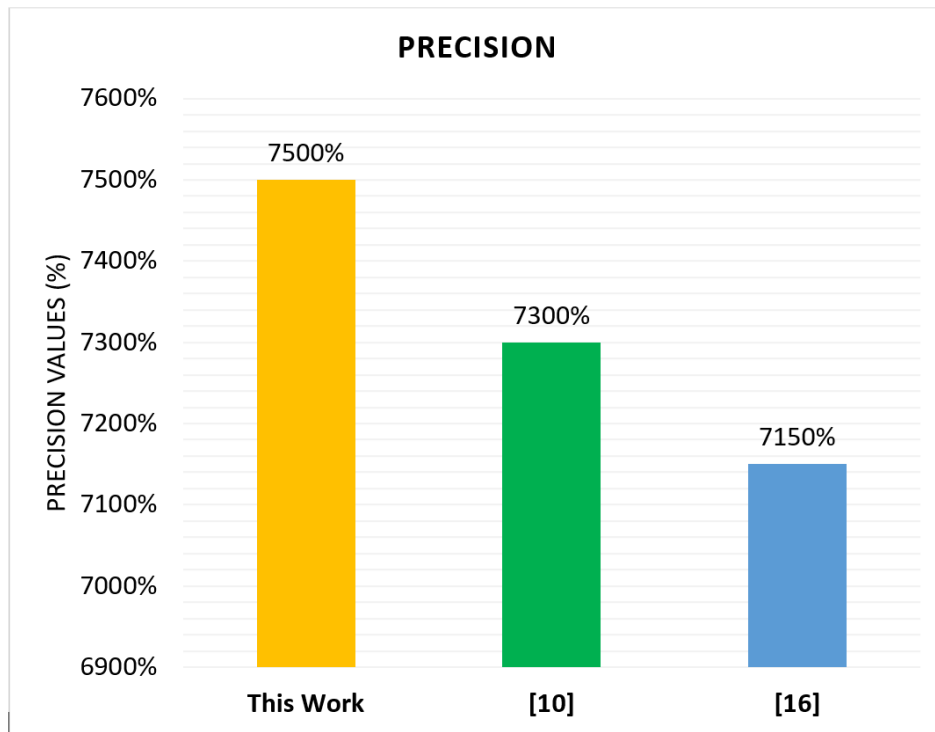


Fig. 9. Comparison of precision criteria.

TABLE II. COMPARISON OF PRECISION CRITERIA IN DIFFERENT SIMULATION SITUATIONS

	precision values				
	250	500	750	1000	Average
This Work	78.00%	71.00%	74.00%	77.00%	75.00%
[10]	70.00%	73.00%	74.00%	75.00%	73.00%
[16]	69.00%	71.00%	72.00%	74.00%	71.50%

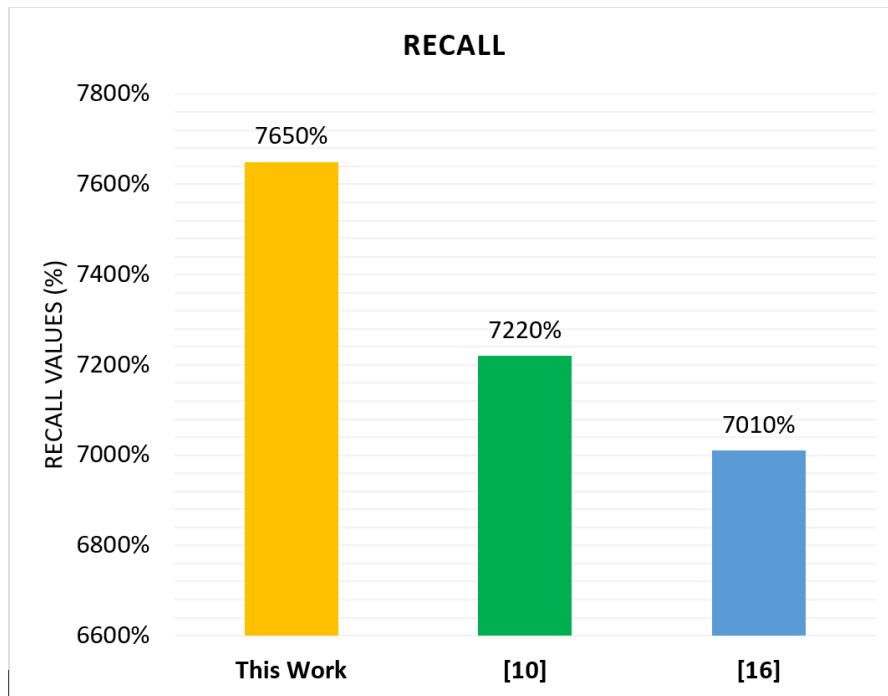


Fig. 10. Recall criterion.

TABLE III. COMPARISON OF RECALL CRITERIA IN DIFFERENT SIMULATION SITUATIONS

	Recall values				
	250	500	750	1000	Average
This Work	72.00%	75.00%	79.00%	83.00%	77.25%
[10]	68.00%	70.00%	73.00%	79.00%	72.50%
[16]	65.00%	68.00%	72.00%	77.00%	70.50%

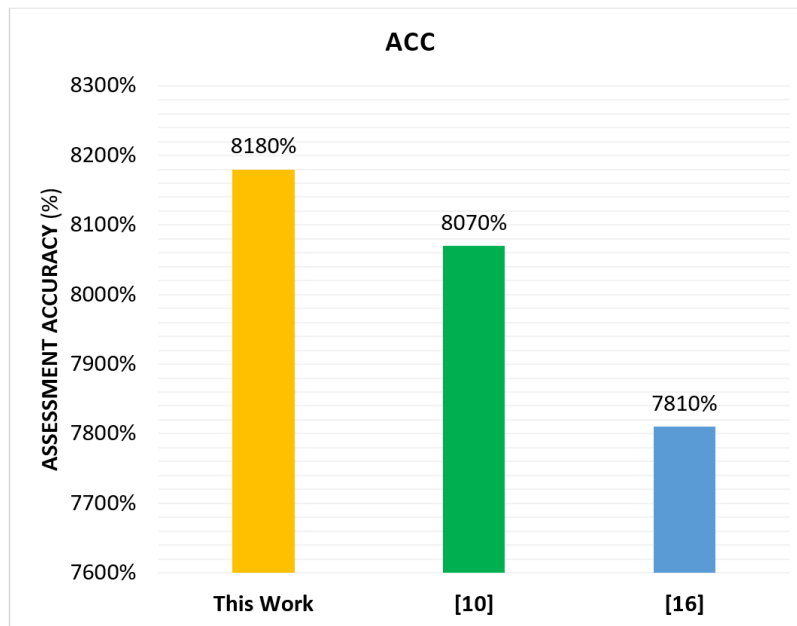


Fig. 11. Comparison of the accuracy of the proposed method with the other two methods.

TABLE IV. COMPARISON OF ACCURACY CRITERIA IN DIFFERENT SIMULATION SITUATIONS

	accuracy values				
	250	500	750	1000	Average
This Work	84.00%	82.00%	81.00%	80.00%	81.75%
[10]	82.00%	82.00%	80.00%	79.00%	80.75%
[16]	80.00%	81.00%	79.00%	78.00%	79.50%

According to the results obtained from Fig. 11, the proposed method has acceptable performance and has performed better than other methods. In Table IV, the result of this criterion can be seen in different simulation situations.

F. False Positive Rate (FPR) Criteria

The next evaluation is the false positive rate (FPR) criterion; this criterion checks how many percent of the images the system has retrieved incorrectly; the lower this criterion is,

the better. To evaluate this criterion, we performed the simulation in four stages, and the average result for all simulation models can be seen in Fig. 12. In Table V, the result of this criterion can be seen in different simulation situations.

G. FRR Criteria

The next evaluation is the false negative rate measure. This measure shows how many percent of the correct images the detection system misidentifies; it can be seen in Fig. 13.

TABLE V. COMPARISON OF AVERAGE FPR CRITERIA IN DIFFERENT SIMULATION SITUATIONS

	FPR values				
	250	500	750	1000	Average
This Work	6.00%	8.00%	12.00%	17.00%	10.75%
[10]	7.00%	12.00%	17.00%	26.00%	15.50%
[16]	8.00%	13.50%	19.00%	26.00%	16.63%

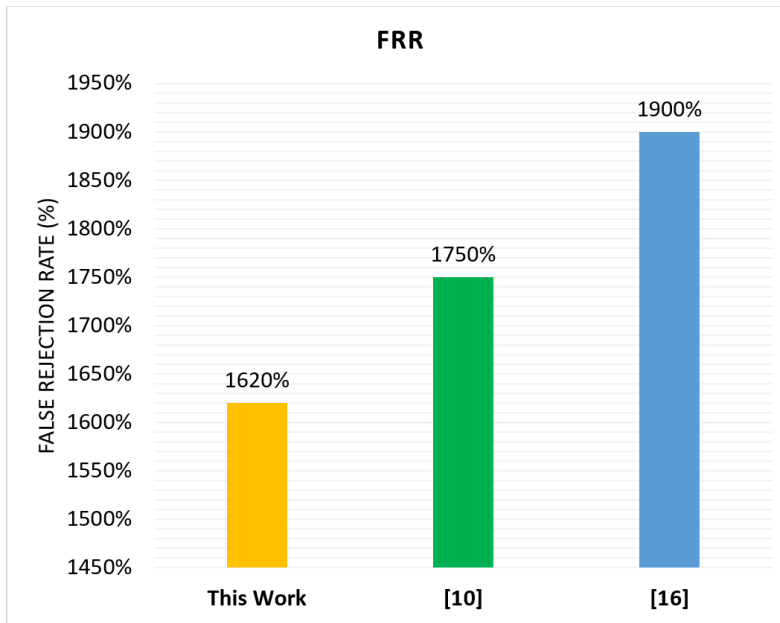


Fig. 12. Average FPR criterion.

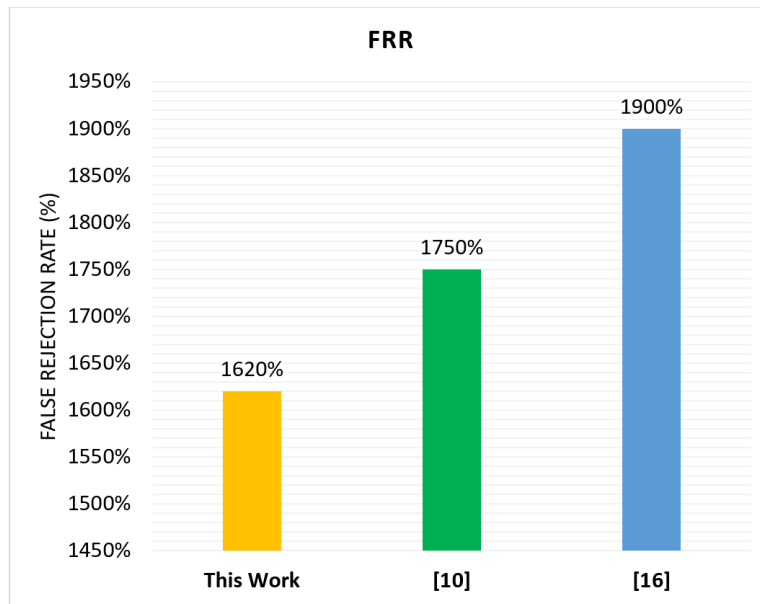


Fig. 13. The result of the FRR comparison.

TABLE VI. COMPARISON OF AVERAGE FRR CRITERIA IN DIFFERENT SIMULATION SITUATIONS

	FRR values				
	250	500	750	1000	Average
This Work	6.00%	13.00%	19.00%	27.00%	16.25%
[10]	7.00%	15.00%	19.00%	29.00%	17.50%
[16]	7.00%	16.00%	21.00%	32.00%	19.00%

In this simulation criterion, it shows that the proposed method has little improvement compared to the other method. In Table VI, the result of this criterion can be seen in different simulation situations.

H. Similarity Criterion

In this part, the proposed method is evaluated based on the similarity criterion in image recognition; in the scenario of this criterion, the degree of similarity of the output images of the program is evaluated, and the result is shown in Fig. 14.

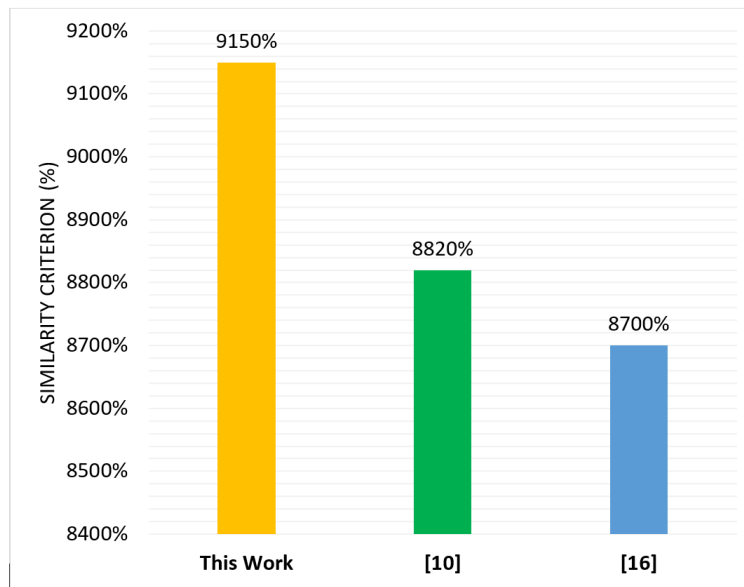


Fig. 14. Comparison of similarity criteria.

As the simulation results in Fig. 14 show, the proposed method performs better than the other two methods. The numerical result of this criterion in several stages of evaluation can be seen in Table VII.

1. LTDR Criteria

- LTDR (Label Tracking Detection Rate) criterion will be used to evaluate the presented algorithm. The LTDR measure determines the rate of assigning unique correct labels to targets and is defined as Eq. (11).

$$LTDR = \frac{1}{L} \sum_{i=0}^{L-1} \frac{TPM_i}{OAF_i} \quad (11)$$

In this regard, L is the total number of targets to be labeled; TPM_i is the number of frames in which the i_{th} target is

correctly labeled, and OAF_i is the total number of frames in which the i -th target is present. The value of this criterion is between zero and one, and the closer it is to one, it means that the detection algorithm has worked correctly. The evaluation result for this criterion can be seen in Fig. 15.

- Finally, Table VIII provides a summary of the accuracy ratings for this study. Higher recall, accuracy, and F-score have been attained using deep learning than with motion-based detection alone. These findings demonstrate how misdiagnosis brought on by inaccurate motion estimation can be supplemented by appearance information. Additionally, with a classification accuracy of over 96%, the deep learning approach may fully utilize the manually labeled training dataset.

TABLE VII. COMPARISON OF THE AVERAGE SIMILARITY CRITERION IN DIFFERENT SIMULATION SITUATIONS

	similarity values				
	250	500	750	1000	Average
This Work	88.00%	90.00%	93.00%	95.00%	91.50%
[10]	86.00%	87.00%	89.00%	91.00%	88.25%
[16]	84.00%	85.00%	89.00%	90.00%	87.00%

TABLE VIII. ACCURACY DETECTION BASED ON EVALUATION CRITERIA

	Using only the difference in motion	Deep learning by the appearance
Accuracy	0.854 ±0.12	0.914 ±0.02
Recall	0.752 ±0.14	0.769 ±0.06
Precision	0.609 ±0.10	0.664 ±0.10
False Positive Rate	0.781 ±0.09	0.813 ±0.08
false rejection rate	0.631 ±0.14	0.719 ±0.11
Similarity criterion	0.824 ±0.10	0.878 ±0.07

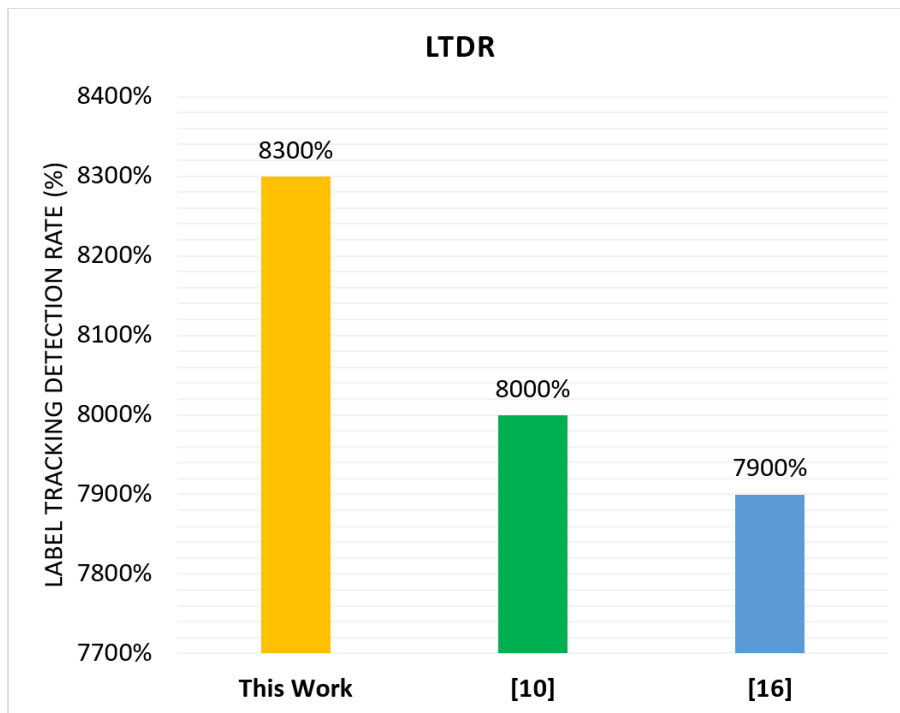


Fig. 15. LTDR criterion check.

In the continuation of this section, the studies related to the problems ahead and the goals of this research are as follows. Object tracking in computer vision, feature extraction and deep learning and convolutional neural networks were among the problems that were addressed in this research. Also, this article discusses the practical applications of object tracking in fields such as industrial, military and urban management. It highlights the importance of tracking moving objects in various scenarios including road control. The goals achieved in this research are: improvement in managing challenging situations, prediction of pattern change, efficiency and real-time processing and integration with autonomous systems. In summary, this research provides insights into the development and evaluation of a method for detecting and provides object tracking. This highlights the importance of accurate tracking in various applications and suggests potential directions for future research to improve the performance and applicability of the method in real-world scenarios.

V. CONCLUSION

A modern civilization needs an intelligent system to control its many management systems. The object tracking system is one of the crucial tools that are employed in the field of operational control in many locations, which makes the job of this field and its authorities very difficult. This goes back to the macro-policies of society in the field of social welfare. Using cutting-edge image processing methods, the technology described in this article can accurately and affordably detect moving objects and track them. In fact, the technique described in this study for detecting moving things, such as cars and items, is quicker and more accurate than the earlier methods. It has been demonstrated that the suggested moving object tracking system, combined with the background subtraction

algorithm, noise removal, filtering, and bubble routing, enables full automation when evaluating and identifying the moving item in the photos. More crucially, our system meets the standards established by decision-making bodies for road control for accuracy in detecting moving objects in road photographs.

In comparison to the tested samples, the test results show that the suggested method produces accurate and extremely dependable findings. The research's most significant applications largely concentrated on the precise position recognition of moving objects on photographs and their tracking using morphological analysis and bubble routing. Cameras should be positioned higher than the road surface in order to identify and track items like vehicles, but for all other uses, this is essentially unnecessary. Using a precise backdrop selection and processing each frame separately, this system accurately recovered pictures of moving objects. The wide applications of tracking moving objects are primarily in industrial, military, and urban management places, which can be, for example, the paths where moving objects are passing more frequently. As a result, according to these parameters, the proposed method can show different performances in different conditions. They even described the kinds of moving things in various photographs. Although the present study provides valuable insights into object detection and tracking, however, it has limitations including lack of real-world deployment, assumption of fixed cameras, performance on large-scale datasets, and limited discussion of handling. It is wrong to generalize to other types of objects.

Future research will focus on analyzing various visual patterns, such as grids, parallel lines, and dot matrices. Additionally, the moving object is more likely to shake since

the surface of the location where the video clip is captured may be uneven and bumpy. This shaking may move from left to right or up and down. As a result, the camera will record photos of the moving object in the route that is rotated and warped. Therefore, in order to be able to estimate and predict pattern change, robust and effective algorithms must be constructed in order to preserve the course of the pattern across successive frames. This is necessary to manage the change in the movement pattern of the valve that results from a natural impulse. The effectiveness and output of this algorithm can then be used to provide a more thorough description of the object's change in shape and pattern of movement.

REFERENCES

- [1] A. Mitrokhin, C. Fermüller, C. Parameshwara, and Y. Aloimonos, "Event-based moving object detection and tracking," in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018, pp. 1–9.
- [2] S. Mane and S. Mangale, "Moving object detection and tracking using convolutional neural networks," in 2018 second international conference on intelligent computing and control systems (ICICCS), IEEE, 2018, pp. 1809–1813.
- [3] P. K. Mishra and G. P. Saroha, "A study on video surveillance system for object detection and tracking," in 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 2016, pp. 221–226.
- [4] R. Verma, "A review of object detection and tracking methods," International Journal of Advance Engineering and Research Development, vol. 4, no. 10, pp. 569–578, 2017.
- [5] S. R. Balaji and S. Karthikeyan, "A survey on moving object tracking using image processing," in 2017 11th international conference on intelligent systems and control (ISCO), IEEE, 2017, pp. 469–474.
- [6] D. H. Ye, J. Li, Q. Chen, J. Wachs, and C. Bouman, "Deep learning for moving object detection and tracking from a single camera in unmanned aerial vehicles (UAVs)," Electronic Imaging, vol. 2018, no. 10, pp. 461–466, 2018.
- [7] A. Keivani, J.-R. Tapamo, and F. Ghayoor, "Motion-based moving object detection and tracking using automatic K-means," in 2017 IEEE AFRICON, IEEE, 2017, pp. 32–37.
- [8] Y. Ma, J. Anderson, S. Crouch, and J. Shan, "Moving object detection and tracking with doppler LiDAR," Remote Sens (Basel), vol. 11, no. 10, p. 1154, 2019.
- [9] Cao, Y., Xu, N., Wang, H., Zhao, X., & Ahmad, A. M. (2023). Neural networks-based adaptive tracking control for full-state constrained switched nonlinear systems with periodic disturbances and actuator saturation. International Journal of Systems Science, 1-16.
- [10] Wang, T., Zhang, L., Xu, N., & Alharbi, K. H. (2023). Adaptive critic learning for approximate optimal event-triggered tracking control of nonlinear systems with prescribed performances. International Journal of Control, 1-15.
- [11] Trik, M., Molk, A. M. N. G., Ghasemi, F., & Pouryeganeh, P. (2022). A hybrid selection strategy based on traffic analysis for improving performance in networks on chip. Journal of Sensors, 2022.
- [12] Trick, M., & Boukani, B. (2014). Placement algorithms and logic on logic (LOL) 3D integration. Journal of mathematics and computer science, 8(2), 128-136.
- [13] Yue, S., Niu, B., Wang, H., Zhang, L., & Ahmad, A. M. (2023). Hierarchical sliding mode-based adaptive fuzzy control for uncertain switched under-actuated nonlinear systems with input saturation and dead-zone. Robotic Intelligence and Automation.
- [14] Khezri, E., Zeinali, E., & Sargolzaey, H. (2022). A novel highway routing protocol in vehicular ad hoc networks using VMaSC-LTE and DBA-MAC protocols. Wireless Communications and Mobile Computing, 2022.
- [15] Zhao, H., Wang, H., Xu, N., Zhao, X., & Sharaf, S. (2023). Fuzzy approximation-based optimal consensus control for nonlinear multiagent systems via adaptive dynamic programming. Neurocomputing, 553, 126529.
- [16] M. Samiei, A. Hassani, S. Sarspy, I. E. Komari, M. Trik, and F. Hassanpour, "Classification of skin cancer stages using a AHP fuzzy technique within the context of big data healthcare," J Cancer Res Clin Oncol, pp. 1–15, 2023.
- [17] J. Sun, Y. Zhang, and M. Trik, "PBPHS: a profile-based predictive handover strategy for 5G networks," Cybern Syst, pp. 1–22, 2022.
- [18] M. Trik, H. Akhavan, A. M. Bidgoli, A. M. N. G. Molk, H. Vashani, and S. P. Mozaffari, "A new adaptive selection strategy for reducing latency in networks on chip," Integration, vol. 89, pp. 9–24, 2023.
- [19] Khezri, E., Zeinali, E., & Sargolzaey, H. (2023). SGHRP: Secure Greedy Highway Routing Protocol with authentication and increased privacy in vehicular ad hoc networks. Plos one, 18(4), e0282031.
- [20] M. G. Nayagam, K. Ramar, K. Venkatesh, and S. P. Raja, "Moving Object Detection and Tracking Algorithm Using Hybrid Decomposition Parallel Processing," Intelligent Automation & Soft Computing, vol. 33, no. 3, 2022.
- [21] Wang, Z., Jin, Z., Yang, Z., Zhao, W., & Trik, M. (2023). Increasing efficiency for routing in internet of things using Binary Gray Wolf Optimization and fuzzy logic. Journal of King Saud University-Computer and Information Sciences, 35(9), 101732.
- [22] Khezri, E., & Zeinali, E. (2021). A review on highway routing protocols in vehicular ad hoc networks. SN Computer Science, 2, 1-22.
- [23] Z. Bai, G. Wu, X. Qi, Y. Liu, K. Oguchi, and M. J. Barth, "Infrastructure-based object detection and tracking for cooperative driving automation: A survey," in 2022 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2022, pp. 1366–1373.
- [24] Cao, C., Wang, J., Kwok, D., Cui, F., Zhang, Z., Zhao, D., ... & Zou, Q. (2022). webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. Nucleic acids research, 50(D1), D1123-D1130.
- [25] Y. Guo et al., "The First Challenge on Moving Object Detection and Tracking in Satellite Videos: Methods and Results," in 2022 26th International Conference on Pattern Recognition (ICPR), IEEE, 2022, pp. 4981–4988.
- [26] Zhang, H., Zou, Q., Ju, Y., Song, C., & Chen, D. (2022). Distance-based support vector machine to predict DNA N6-methyladenine modification. Current Bioinformatics, 17(5), 473-482.
- [27] Z. Zhang, D. Cheng, X. Zhu, S. Lin, and J. Dai, "Integrated object detection and tracking with tracklet-conditioned detection," arXiv preprint arXiv:1811.11167, 2018.
- [28] M. Zohaib, M. Ahsan, M. Khan, and J. Iqbal, "A featureless approach for object detection and tracking in dynamic environments," PLoS One, vol. 18, no. 1, p. e0280476, 2023.