DOI: http://doi.org/10.32792/utq.jceps.10.02.05

Detection and Classification of Moving Objects

Afaf J. Tuaimah[1], Proof Hazeem B. Taher[2]

[1]University of Thi-Qar – College of Sciences.
[2]University of Thi-Qar – College of Education for pure sciences.

Received 13/10/2019 Accepted 8/03/2020 Published 30/11/2020

Abstract:

Surveillance systems is the process of monitoring the behavior of people, objects or processes within systems for conformity to expected or desired standards in trusted systems for security. In tracking safety specific regions such as banks, roads, boundaries, forests and so on, video surveillance systems are becoming more common every day. As monitoring systems transition from analog to digital devices and numbers rise, there is a need to interpret the captured video automatically.

The objective of this paper is to design a system that can automatically able to detect objects in different scenes and then classify them. The proposed system consists of two stages: moving object detection and classification. The first stage consists of four steps: The first step is input video, the collection database is used in this paper where some videos are captured using fixed camera and other images are acquired from various internet sites. The second step is the pre-processing of frames using HSV color space, filtering them using Median and Gaussian filters. The third step is detection object using background subtraction method. The fourth step is feature extraction of object. In this paper, 21 features of shape features are extracted.

In the second stage of proposed system is classification the object in to (human and car) using K-nearest neighbor. Two different ratios of training/testing groups which are (50% - 50%) and (70% - 30%) are applied to the classifier. K-NN gives training accuracy 100 %. The accuracy of testing is 98.6111% and 97.7% when the ratio is (50% - 50%) and (70% - 30%) respectively. MATLAB R2018 is used to design the proposed system.

Keywords: surveillance, object detection, classification, K- nearest neighbor.

1.Introduction:

Surveillance systems is the process of monitoring the behavior of people, objects or processes within systems for conformity to expected or desired standards in trusted systems for security or social control [1]. Traditionally, video surveillance systems consist of cameras for monitoring screens. These systems are deployed to offer an overview of a big region to a limited number of operators. In particular, human operators handle online video outputs of the more sensitive fields and the remaining video outputs are registered for future use in the event of a diagnostic event. The number of monitoring devices is increasing, leading to new challenges, as the systems and operations behind these cameras are unable to deal with the fast rise in the number of video cameras and systems [2], [3]. In video surveillance applications, moving object detection in a video stream is an essential step [4]. It classifies the pixels either in the foreground or in the background[5]. Object detection is performed to verify the presence of objects in video and to identify that object accurately [6]. Many algorithms have been suggested for moving object detection. These include background subtraction, optical flow, temporal difference, and many other algorithms to detect moving items. From these, the most commonly used algorithm is background subtraction with many algorithms such as image difference, approximate average, Gaussian combination, and background of running Gaussian and Kernel density [7]. Object classification in video sequences is the topic of several global researches; it has a large range of applications in various areas such as biomedical imaging, video surveillance, car navigation, visual inspection, robot navigation and remote sensing. Classification of objects in videos is a complicated method that needs very robust and precise methods. The objective of classifying objects is to categorize different objects [8],[9].There are many algorithms can be used to classify the object, such as Nearest Neighbor, Super Vector Machine, Artificial Neural Network and etc.

2. Moving Object Detection and Classification:

First step in visual surveillance system includes moving object detection. Moving object detection is the process that segments the moving foreground object from the rest image. Successful segmentation of foreground object helps in the subsequent process such as object classification, personal identification, object tracking and activity recognition in the video. Motion segmentation is done mainly with background subtraction, temporal differencing, and optical flow. Out of the three methods, background subtraction is the most popular method for detecting moving regions in an image by taking the absolute difference between the current image and the reference background image. Figure (1) illustrates the basic flow diagram of the system [1]

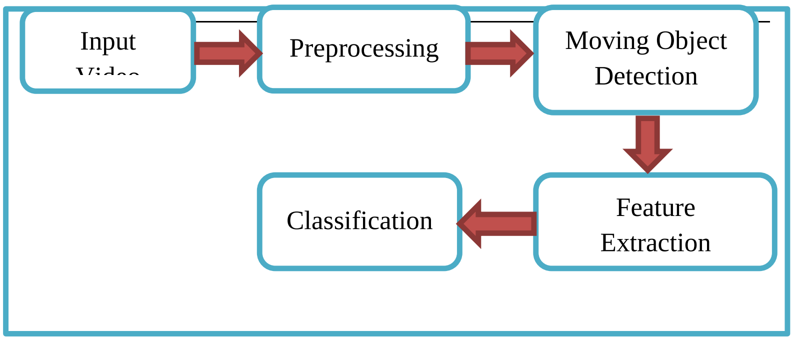


Figure (1): The basic flow diagram of object detection and classification.

3.Theoretical Background:

3.1. Segmentation:

Image segmentation is one of the most important processes for many applications. Image segmentation algorithms generally are based on one of two properties of intensity values: discontinuity and similarity. In the first category, the approach is to partition an image based on abrupt changes in intensity, such as edge in an image. The principal approaches in the second category are based on partitioning an image into regions that have some measure of homogeneity within themselves. Thresholding, region growing, and region splitting and merging are examples of methods in this category [13], [14]. Thresholding is segmentation technique that popular, fast, and computationally inexpensive. The pixels in this technique classified as belonging to either the set of object pixels or the set of background pixels as the intensity level function. Thresholding is done by assigning each pixel with intensity value T for extracting an object [11], [15]. A more principled approach to automatic threshold selection is given by Otsu's method. It selects the threshold which results in the tightest clustering of the two groups represented by the foreground and background pixels [16].

In image processing the mathematical morphology is used as a means to identify and extract meaningful image descriptors based on properties of shape within the image [16]. Morphological operators work with two images. The image being processed is referred to as the active image, and the other image is referred to as the structuring element (SE). The two principal morphological operations are dilation and erosion. These two basic operations can be combined into more complex sequences. The most useful of these for morphological filtering are called opening and closing [13].

First step for background subtraction is background modelling. It must sensitive enough to recognize moving objects. Background Modeling is to yield reference model [17]. It can be of two types firstly by considering first frame as the reference frame or background image. Secondly by considering average of 'n' frames as the background image. In this background subtraction method every pixel of on-going frame is subtracted with the pixels of the background image. The equations (4) and (5) shows the background subtraction method for first frame as the background image.

$$B(x,y) = I(x,y) \tag{4}$$

where $B(x,y)$ represents background image pixel by pixel. The background subtraction method splits the video frames into foreground and background object, where the foreground object is discovered by matching the current frames $I(x,y)$ through the background image $B(x,y)$. The equation used is:

$$C(x,y) = \begin{cases} 1 \;; & if \;\; B(x,y) - I(x,y) > T \\ 0 \;; & otherewise \end{cases} \tag{5}$$

where $C(x,y)$ is the foreground pixel, T is threshold value which can be set manually or can selected automatically as per video input. This method consumes less memory. Accuracy of detection is moderate. But it will not suit for multimodal backgrounds [18].

3.2. Features Extraction:

Feature extraction is a process that begins with feature selection. The choice of appropriate features depends on the particular image and the application at hand. Features are higher level representations of structure and shape. Shape features depend on a silhouette of the image object under consideration, so all that is needed is a binary which is as a mask of the image object [13], [19]. In this paper, the following features are used:

- The ratio of height to width of bounding box: The bounding rectangle is the smallest rectangle that contains the object region.

- Circularity: It is a scaled version of ratio of object's area to its perimeter such as in following equation:

$$Circ = 4\pi A/P^2 \tag{6}$$

where A is area of an object and P is perimeter of an object.

- The distance between extrema point of an object and it's centroid: Extrema point is (8×2) matrix that limits the extrema points in the region. The first column of the matrix contains the x- coordinates of eight points and the second column of the matrix contains the y- coordinates of eight points. The eight points are (right-top, right-bottom, left-top, left-bottom, top-right, top-left, bottom-right and bottom-left).

The centroid of an object represents the midpoint along each row and column axis corresponding to the middle based on the spatial distribution of pixels within the object.

The distance between extrema point of an object and it's centroid can be calculated using Euclidean distance.

- Eccentricity:

It is also called elongation, defined by the ratio of the bounding box of an object. This can be found by scanning the image and finding the minimum and maximum values on the row and columns where the object lies.

- Minor Axis Length: It is the length in pixels of the minor axis of the ellipse that has the same second moments as the region [14].
- Orientation of an object: It can be provided information about the object's orientation by the axis of least second moment. This axis corresponds to the line about which it takes the least amount of energy to spin an object of like shape. [13]:
- Invariant Moment: The concept of moments forms an important part of elementary probability theory [16]:

Moments can be extending to 2-D (and higher dimensional) functions. Thus, the $pqth$ moment of a 2-D density function $p(x, y)$ is given by

$$m_{pq} = \int_{\infty}^{\infty} \int_{\infty}^{\infty} x^p y^q \, p(x, y) dxdy \quad (7)$$

It can be calculated the moments of images, however, by replacing the density function with the intensity or binary image, in case binary image $I(x, y)$; the *moment of order* $(p + q)$ is as in following equation:

$$m_{pq} = \sum \sum x^p y^q I(x, y) \quad (8)$$

To make these moments translationally invariant, it can be enable to use the central moments defined by equation (9)

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y) \quad (9)$$

where $\bar{x} = \frac{m_{10}}{m_{00}}$ and $\bar{y} = \frac{m_{01}}{m_{00}}$

From normalized central moment, the invariant moment features can be created as following:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \quad (10)$$

where $\gamma = \frac{p+q}{2} + 1, for \ (p + q) = 2,3,4 \ldots$

The seven normalized invariant moments (or known as Hu moment) are used in this paper.

3. Object Classification:

Object classification is the categorization of object based on previously classes or types. Object classification is simply used to identify a class of a moving object which is visible in a video as a person or group of persons [21].

3.1. K-Nearest Neighbor Algorithm:

Nearest Neighbor method is the simplest algorithm for identifying a sample from the test set. The object of interest (a sample from the test set) is compared to every sample in the training set, using a distance measure, a similarity measure, or a combination of measures. The "unknown" object is then identified as belonging to the same class as the closest sample in the training set. This process is computationally intensive and not very robust. It can be made the Nearest Neighbor method more robust by selecting not just the closest sample in the training set, but by consideration of a group of close feature vectors. This is called the K-Nearest Neighbor method, which is one of the best text classification algorithms [13], [22]. K-Nearest Neighbor is a supervised learning algorithm where the result of new instance query is classified based on majority of K-Nearest Neighbor category. K-Nearest Neighbor algorithm used neighborhood classification as the prediction value of the new query instance. This algorithm predicts the test sample's category according to the k training samples which are the nearest neighbors to the test sample, and judge it to that category which has the largest category probability. In binary (two class) classification problems, it is helpful to choose k to be an odd number as this avoids tied votes [23].

4. The Proposed System:

The processed system is detection of moving object in video and classification it into (human or car). The proposed system consists of five phases which are data acquisition, preprocessing, object detection, features extraction and object classification as shown in Figure (2).

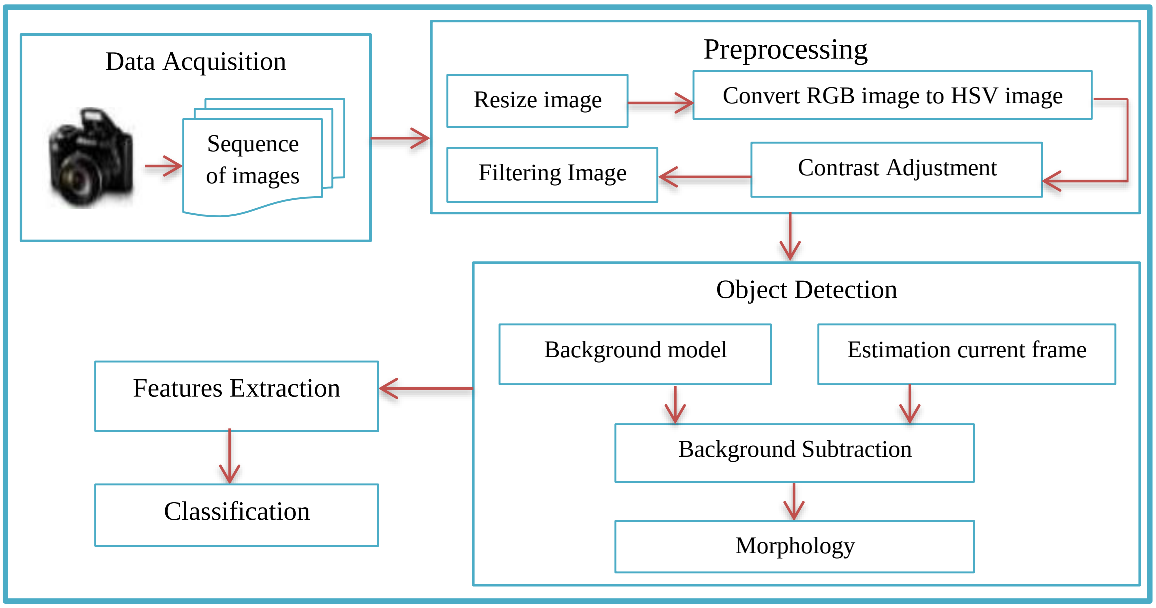


Figure (2): Block Diagram of proposed system

4. 1. Data Acquisition:

The first task in object detection algorithm is to get video feed from the camera. Read video can be denoted by constructing the structure to save the sequence of images. The structure contains data image, height and width of the frame, color map etc.

4.2. Preprocessing:

This phase is necessary to prepare the digital image for further processing. Preprocessing the input images can reduce noise and defects arising during the image capturing process. This phase includes: resize image, convert the resized image from RGB into HSV color space, applying contrast adjustment to HSV image and filtering image using Median and Gaussian filters.

4.3. Object Detection:

The system uses a background subtraction method for distinguishing foreground pixels which contain the moved objects from stationary background. The steps of object detection can be as follows:

4.3.1. Background model:

The first step in background subtraction is initialization the background model (reference image). There are many methods to estimate background model, first frame is selected as reference image.

4.3.2. Estimation Current Frame:

Since not all frames contain moving objects, the system needs to select current frame. It can be used the sum of absolute difference (SAD) between current frame and previous frame, as shown as in equations below:

$$AD = |I_k - I_{k-1}| \tag{11}$$

$$SAD = \sum_{i=1}^{Wframe} \sum_{j=1}^{Hframe} AD(i,j) \tag{12}$$

where AD is absolute difference, SAD is sum of absolute difference, I_k, I_{k-1} are the current frame and previous frame respectively. $Wframe, Hframe$ are width of frame and height of frame the total number of frames. After that, it can find max (SAD), where the index of max value corresponds to no. of current frame

---

4.3.3. Background Subtraction:

In this step the background model is subtracted from current frame, the resultant image (the difference image) is HSV image. The proposed system uses value channel (V) only, then this image (value channel of difference image) is compared with threshold value according to equation (5). The threshold value used is global threshold. Thresholding image is binary image, and then median filter is applied on binary image to reduce the noise in image.

4.3.4. Morphology:

The output of subtraction step generally contains noise therefore is not appropriate for further processing without special post-processing. Morphological operations, dilation and closing, are applied to the foreground image (thresholding difference image) in order to remove noise and fill holes that is caused by using the threshold value on difference image.

4.4. Feature Extraction:

In this paper the shape features are used that extract from silhouettes of the detected objects in binary images, and represented by calculate height to width ratio of bounding box which surrounds with detected objects as a first feature, (8-features) which extracted by calculating the distance between the 8 points (extrema point) with the center of object as a second feature until 9^{th} feature, the mean of these distance as 10^{th} feature, circularity as 11^{th} feature, eccentricity as 12^{th} feature, MinorAxisLength as 13^{th} feature orientation as 14^{th} feature and the seven invariant moment as the 15^{th} to 21^{th} feature.

4.5. Classification:

The goal of this step is to distinguish between object types (human or car), after obtaining features vector and storing it, it becomes ready to use in classification phase. Split data randomly into (training, testing) There are many types of classifiers. In our work, k- nearest neighborhood (K-NN) is used for categorizing the objects. The parameters of (K-NN) in this paper are (21) features as input, two output, (k=3), and Euclidean distance is used as a difference measure.

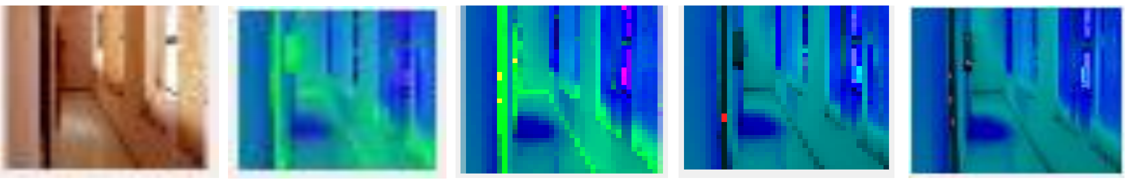
5. Experimental Results for the proposed system:

Table (1) explains the total number of data which are used in this paper.

Table (1): The number of data for each class

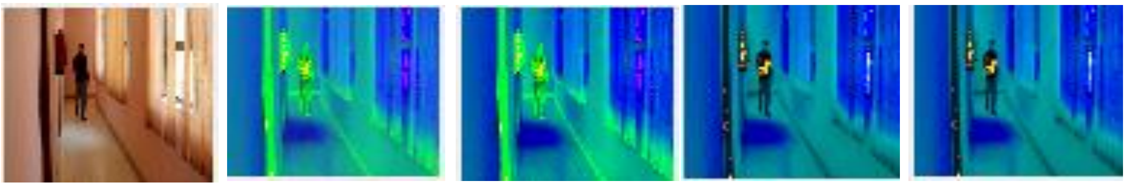
Data number		Total number of data
Data of human	Data of car	
80	64	144

Figures (4) and (5) show the results of preprocessing steps of background model and current frame for sample1 respectively.



(a) (b) (c) (d) (e)

Figure (4): Sample1 (background model) after applying the preprocessing steps: (**a**) RGB background model (**b**)Converting image (a) into HSV (**c**) Adjusting contrast, (**d**) Applying Median filter on (c), (**e**) Applying Gaussian filter on (**d**).



(a) (b) (c) (d) (e)

Figure (5): Sample1 (current frame) after applying the preprocessing steps: (**a**) RGB background model (**b**)Converting image (a) into HSV (**c**) Adjusting contrast, (**d**) Applying Median filter on (c), (**e**) Applying Gaussian filter on (**d**).

After the frames preprocessed, it became ready to use in object detection phase. Figure (6) shows the results of all steps of object detection for sample1.

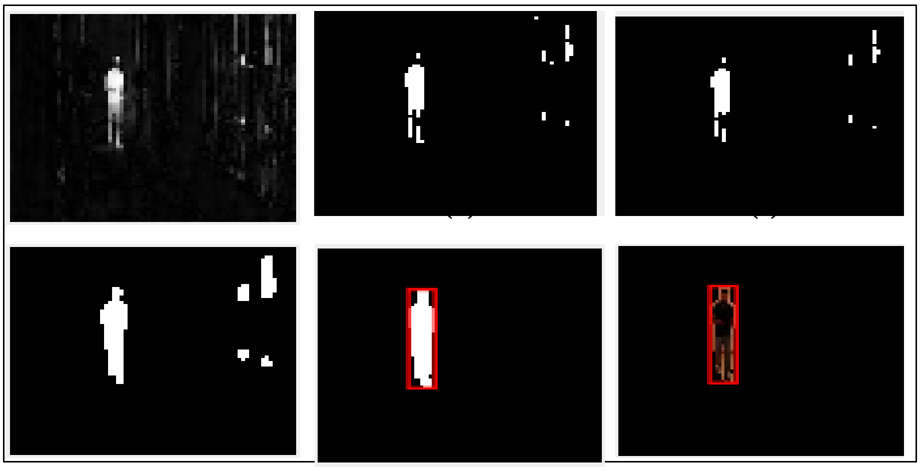


Figure (6): The results of object detection for sample1: (**a**) Result of subtracting images ,(**b**)Thresholding (a) (**c**)Applying Median filter on (**b**), (**d**)Apply morphological operation(dilation-closing-fill), (**e**) Remove isolated pixels **(f)** Multiplying (e) by RGB-current frame (image (a) of Figure (5).

After the moving objects are detected, its features must be extracted and set it to the features vector as shown in Table (2).

Table (1): Features vector of samples of human images and car images

Shape Features																					
samples	Height to Width Ratio	Distances								Mean of Distances	Circularity	Eccentricity	Minor Axis Length	Orientation	Moment1	Moment2	Moment3	Moment4	Moment5	Moment6	Moment7
		Distance1	Distance2	Distance3	Distance4	Distance5	Distance6	Distance7	Distance8												
Sample1 (human)	4.916667	0.489552	0.490056	0.16044	0.120098	0.512954	0.510787	0.155476	0.28021	0.339947	0.375512	0.983907	21.45887	-89.577	0.013648	0.000162	4.20E-08	5.89E-08	2.92E-15	7.43E-10	2.17E-16
Sample2 (car)	0.631313	0.54524	0.572617	0.771986	0.773764	0.643822	0.590478	0.877305	0.860346	0.704445	0.669932	0.790279	123.1333	6.850758	0.001377	5.48E-07	7.31E-11	2.40E-11	2.56E-22	6.28E-15	9.71E-22

These features are used to classify the objects into (human or car). K-nearest neighbor is used for classification. The database is separated randomly twice times with specific ratio at each time in order to form two different training/testing groups. These groups are group1 which denotes to (50% - 50%) and group2 which denotes to (70% - 30%).

class1: human, class2: car.

Tble (2): Number of data for each group

Groups	Training data		Testing data		Total data
	Class1	Class2	Class1	Class2	
Group1	40	32	40	32	144
Group2	56	44	24	20	144

Train to test ratio		50/50		70/30	
Class Name		Target Class		Target Class	
		Class1	Class2	Class1	Class2
Output Class	Class1	**40**	1	**24**	1
	Class2	0	**31**	0	**19**
Training accuracy		100%		100%	
Testing accuracy		98.6111%		97.7273%	

Table (3): classification accuracy of K-NN

The parameters of K-NN which are (k=3, distance metric: Euclidian), the accuracy of classification was as seen in Table (3). The number marked in bold font in this table represents the correct classification of each type in data set to corresponding class.

6. Conclusion:

In this paper, HSV color space is used, median and Gaussian are applying on all frames of video, then background subtraction method is used for detection single moving object in video stream, which is more appropriate for static indoor environments or for outdoor environments with limited motion, like parking lots, office etc. The system is designed for segmentation the foreground object from background, so it can be estimated the current frame which contains the moving object by calculate the sum of absolute difference and select large sum which corresponds to current frame. For classification, the system uses features shape of the objects. It is robust features. K-nearest neighbor is used for classification. The testing accuracy of K-NN was 98.6111% and 97.7% when the train to test ratio is (50% - 50%) and (70% - 30%) respectively. For future work, it can implement the proposed system in real time in order to use in security issues.

References:

1. Gupta, Pritee, Yashpal Singh, and Manoj Gupt. "Moving Object Detecti on Using Frame Difference, Background Subtraction and SOBS for Vi deo Surveillance Application." pp151-156 (2014).
2. Tuna, Hakan. Detection and classification of objects and texture. Master's thesis. bilkent university, 2009.
3. Sutor, Stephan R. "Large-scale high-performance video surveillance". University of Oulu Graduate School; University of Oulu, Faculty of Information Technology and Electrical Engineering, Department of Information Processing Science. Acta Univ. Oul. A 631, 2014.

4. Mashak, Saeed Vahabi, et al. "Background subtraction for object detection under varying environments." 2010 International Conference of Soft Computing and Pattern Recognition. IEEE, 2010.
5. Sharma, V., Nain, N., & Badal, T. "A survey on moving object detection methods in video surveillance". International Bulletin of mathematical research, 2(1), 219-218, 2015.
6. Patel, Hitesh A., and Darshak G. Thakore. "Moving object tracking using kalman filter." International Journal of Computer Science and Mobile Computing 2.4: 326-332. (2013)
7. Aldhaheri, Asim R., and Eran A. Edirisinghe. "Detection and Classification of a Moving Object in a Video Stream." Proc. of the Intl. Conf. on Advances in Computing and Information Technology-ACIT. 2014.
8. Setitra, Insaf. "Object classification in videos-an overview." Journal of automation and control engineering 1.2: 106-109. (2013)
9. Najva, N., and K. Edet Bijoy. "SIFT and tensor based object detection and classification in videos using deep neural networks." Procedia Computer Science 93: 351-358. (2016)
10. Panda, D. K. "Motion detection, object classification and tracking for visual surveillance application" (Doctoral dissertation). (2012).
11. Solomon, C., & Breckon, T. "Fundamentals of Digital Image Processing: A practical approach with examples in Matlab". John Wiley & Sons. (2011).
12. Umbaugh, S. E. "Digital image processing and analysis: human and computer vision applications with CVIPtools". CRC press. (2010).
13. Gonzalez, Rafael C., Steven L. Eddins, and Richard E. Woods. Digital image publishing using MATLAB. Prentice Hall, 2004.
14. Acharya, T., & Ray, A. K. (2005). Image processing: principles and applications. John Wiley & Sons.
15. Zhou, H., Wu, J., & Zhang, J. (2010). Digital Image Processing: Part II. Bookboon.
16. Dougherty, G. (2012). Pattern recognition and classification: an introduction. Springer Science & Business Media.
17. Kruthi, S., & Kamalakannan, J. (2018). Implementation of moving object detection and categorization from hevc compressed surveillance video. International Journal of Mechanical Engineering and Technology, 9(4), 35-42.
18. Shilpa, P. H., & Sunitha, M. R. (2016). A survey on moving object detection and tracking techniques. International Journal of Engineering and Computer Science, 5(5), 16376-16382.
19. Athanesious, J. J., & Suresh, P. (2012). Systematic survey on object tracking methods in video. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1(8), 242-247.
20. Michel Marie Deza and Elena Deza, “Encyclopedia of Distances”, Springer, pp. 94,2009.
21. Rahul Dutt Sharma1, Subham Kumar Gupta, A Survey on Moving Object Detection and Tracking Based On Background Subtraction. Oxford Journal of Intelligent Decision and Data Science 2018 SI. 1 (2018) 55-62
22. Jingwen Sun, Weixing Du, Niancai Shi, A Survey of kNN Algorithm, Information Engineering and Applied Computing (2018)

23. Kaghyan, Sahak, and Hakob Sarukhanyan. "Activity recognition using k-nearest neighbor algorithm on smartphone with tri-axial accelerometer." International Journal of Informatics Models and Analysis (IJIMA), ITHEA International Scientific Society, Bulgaria 1 (2012): 146-156.