

A NOVEL BACKGROUND SUBTRACTION ALGORITHM FOR PERSON TRACKING BASED ON K-NN

Asmaa AIT MOULAY ¹ and Aouatif AMINE ²

BOSS Team, Systems engineering laboratory, National school of applied
sciences, University Compus, Kenitra, MOROCCO

¹aitmoulay.asmaa@univ-ibntofail.ac.ma

²amine_aouatif@univ-ibntofail.ac.ma

ABSTRACT

Object tracking can be defined as the process of detecting an object of interest from a video scene and keeping track of its motion, orientation, occlusion etc. in order to extract useful information. It is indeed a challenging problem and it's an important task. Many researchers are getting attracted in the field of computer vision, specifically the field of object tracking in video surveillance. The main purpose of this paper is to give to the reader information of the present state of the art object tracking, together with presenting steps involved in Background Subtraction and their techniques. In related literature we found three main methods of object tracking: the first method is the optical flow; the second is related to the background subtraction, which is divided into two types presented in this paper, and the last one is temporal differencing. We present a novel approach to background subtraction that compare a current frame with the background model that we have set before, so we can classified each pixel of the image as a foreground or a background element, then comes the tracking step to present our object of interest, which is a person, by his centroid. The tracking step is divided into two different methods, the surface method and the K-NN method, both are explained in the paper. Our proposed method is implemented and evaluated using CAVIAR database.

KEYWORDS

Video Surveillance, Object Tracking, Feature Extraction, Background Subtraction, Big data.

1. INTRODUCTION

Video tracking can be defined as an action that can describe and highlights an object so that we can follow it in a sequence of frames. It's very useful for higher level applications that rely on visual input to have the ability to consistently detect and track object motion such as humans, vehicles...etc. Therefore many developments have been affected to the methods, algorithms and techniques that are used as well as being implemented in other different areas such as security, surveillance systems, robotics...etc. In various computer vision applications, background subtraction is an easy and quick way to locate moving objects in video shot with a static camera. Even if this method will appear easy, some videos generate some false positives. Like illumination changes, an animated background...etc. Also false negatives can happen too when a moving object have the same colours as the background ones. What makes the simple inter frame difference a weak solution, so in order to deal with those challenges, different background models

have been proposed. The goal in this paper is to review several state-of-the-art background subtraction models proposed in the past years of researches and to present our proposition using this technique.

Background subtraction is detecting moving objects in videos using static cameras. It's a basic task in many video analysis applications, such as intelligence video surveillance, people detection and tracking...etc. The idea in the approach of the background subtraction is to detect the moving object of interest from the difference between the current frame and a reference frame which is called the background model. This background model must have a good quality so it can represent a scene as well as to be adapted to the luminance changes so that when we subtract it with the current image we will obtain a good background subtraction results.

1.1. Motivation

Recent events, such as terrorist attacks or the large-scale acts of banditry have led to an increased demand for security in societies. Video surveillance systems called smart are dealing with the real time monitoring of mobile or immobile objects. The main objectives of these systems is to provide automatic interpretation of filmed scenes and to understand as well as to predict the actions and interactions of the objects observed. Video tracking helps us to map target objects in consecutive video frames, and since the demand of security in societies is growing it shall lead us to a need in surveillance application in many environments, particularly in the following domains:

1. Public safety and commercial sites: like supervising banks, shopping centres, airports, train stations, museums...etc.
2. automated video surveillance: designed to monitor the movement in an area to identify moving objects.
3. traffic monitoring: in case of breaks of traffic rules is monitored using cameras it will be tracked down easily.
4. information extraction: like to count people in public spaces, establishment of consumer profiles in malls...etc.
5. Military Application: border surveillance, monitoring secure web sites...etc.

2. MOVING OBJECT TRACKING

Object detection is an initial step of object tracking, it's a computer vision technology that deals with identifying instances of objects such as humans, vehicles, animals and other moving objects. Video sequence consist of several frame sequences which have certain surplus continuity. Frame sequences are extracted from the video sequence so we can detect moving objects.

Information about motions can be obtained through analysis and processing of frames at different periods of time. Therefore, two or more frames are acquired at different times and that contain information about relative motion between an imaging system and a scene. Video sequence analysis can be classified into three methods: optical flow method, background subtraction method, and temporal differencing method.

2.1. Optical flow

Optical flow is a vector-based method that matches points on objects over multiple frames in order to estimates motion in video. The motion field of frames is estimated to incorporate similar motion vectors into moving object. Solving transcendental equations is required in optical flow method. The calculation of this method is so complex and very sensitive to noise, which means that the amount of calculation is large. So this method is difficult to use in real time video because its performance and practicability is quite weak.[1]

To write the optical flow equation, we are going to suppose two assumptions: Assume that the apparent brightness of moving objects remains constant between frames.

And the image brightness is continuous and differentiable.

For the first assumption:

$$I(x(t+\Delta(t)),y(t+\Delta(t)),t+\Delta(t))=I(x(t),y(t),t) \quad (1)$$

The first order Taylor expansion of the left term is

$$\begin{aligned} I(x(t+\Delta(t)),y(t+\Delta(t)),t+\Delta(t)) &= I(x(t),y(t),t) + \frac{\partial I}{\partial x} \frac{\partial x}{\partial t} \Delta(t) + \frac{\partial I}{\partial y} \frac{\partial y}{\partial t} \Delta(t) + \frac{\partial I}{\partial t} \Delta(t) \\ &= I(x(t),y(t),t) + I_x u + I_y v + I_t \Delta(t) \end{aligned} \quad (2)$$

From (1) and (2), we derive the optical flow equation

$$I_x u + I_y v + I_t = 0 \quad (3)$$

Where I_t is the image difference between the two images

It provides a dense (point to point) pixel correspondence and Finding optic flow using edges has the advantage (over using two dimensional features) but it has also some disadvantages and problems and one problem common to almost all optical flow methods is the aperture problem, as one searches after displacement only in a small region near the previous location, one dimensional features can only be tracked in the normal direction of the object. Sources of other problems are illumination, numerical instabilities and occlusion. Illumination is a problem as motion is a geometric quantity while we are dealing with image intensity which depends on illumination. Numerical instability can be an important issue as many methods depend on numerical differentiation which is very noise sensitive. Occlusion is when a moving object occludes the background or another moving object, there are problems as patterns in the image appear and disappear.

2.3. Temporal differencing

Temporal differencing methodology is based on frame differences, it is the most simple and direct method. Moving objects are extracted according to the differences among two or three continuous frames. This method detects the moving regions and only objects marking relative motions. Moreover, given the fact that the time interval between two images is quite short, illumination changes have little influence on different images, which means the detection is effective and stable. Using frame differences can better adapt to environment in intensive fluctuation, and can easily detect those pixels causing images to change distinctly when the target moves. However, it is inadequate for dots with insignificantly changed pixels. Accordingly, the method is largely used in situations with comparatively simple background [2].

We can illustrate its equation as so :

$$|I(x,y,t) - I(x,y,t-1)| > Th \quad (4)$$

This algorithm present an image with white pixels in the spot where the current image I at the time t is different from the previous image $I(x, y, t - 1)$.

2.3. Background subtraction

Background subtraction is a widely used approach for detecting moving objects from a static camera. Objects can be detected by finding the difference between the current frame and

background frame. Background modelling can be categorised into two categories which are non-recursive and recursive techniques. For background estimation recursive technique includes frame differencing, median filter, linear predictive filter, and nonparametric model. Recursive technique is based on median filter, Kalman filter and mixture of Gaussian [1]. An image is divided into foreground and background in this method. The background is modelled, and the current frame and the background model are compared pixel by pixel. Those pixels accordance with the background model are labelled as the background, while others are labelled as the foreground. Background subtraction is a common method in moving object tracking algorithm, which is used more often in situations with relation to a background that is still. This method has low complexity. However, acquired background frames become sensitive to scene changes caused by illumination and external conditions as times goes on [2]. As the name suggests, background subtraction is the process of separating out foreground objects from the background in a sequence of a video frames, the equation that can explain this method would be like so:

$$|I(x,y,t) - B(x,y,t)| > Th \quad (5)$$

While I is the image at the time t and B is the background at the time t . We subtract the estimated background from the input frame and apply a threshold, Th , to the absolute difference to get the foreground mask. One of the reasons that led us to choose the background subtraction method is its speed, which is illustrated in the execution time.

These background methods are used to identify moving objects in a video, which is often the first step in complex systems such as activity recognition, object tracking, and motion capture. Needless to deny the advantages of extracting the moving objects which can improve the reliability of the system by reducing the search space, reducing processing needs, and allowing the use of simpler technics for the rest of the data extraction, objects also are allowed to become part of the background without destroying the existing background model and even a different threshold is selected for each pixel and they are adapting by time what makes this method very robust against movement that are part of background (moving branches of a tree). Every method has its own disadvantages, for the background subtraction methods, there are relatively many parameters and they should be selected intelligently which make the method more complicated and also cannot deal with sudden, drastic lightning changes, shadows and gives also false positives sometimes.

3. BACKGROUND SUBTRACTION

Various methods have been proposed in the literature to perform the task of detection object of interest. In applications using fixed cameras with static background invariant, the most common approach is the detection background subtraction. But background subtraction algorithms have to stand up with several challenges arising from the nature of video surveillance.

3.1. Challenges of background subtraction

There are so many different background subtraction challenges that exist in literature [16]. Here we list some of the remarkable challenges for this algorithm [7]:

Gradual illumination changes. It is desirable that background model adapts to gradual changes of the appearance of the environment. For example in outdoor settings, the light intensity typically varies during day.

Dynamic background. Some parts of the scenery may contain movement, but should be regarded as background, according to their relevance. Such movement can be periodical or irregular (e.g., traffic lights, waving trees).

Occlusion. Intentionally or not, some objects may poorly differ from the appearance of background, making correct classification difficult. This is especially important in surveillance applications.

Shadows. Shadows cast by foreground objects often complicate further processing steps subsequent to background subtraction. Overlapping shadows of foreground regions for example hinder their separation and classification. Hence, it is preferable to ignore these irrelevant regions.

Video noise. Video signal is generally superimposed by noise. Background subtraction approaches for video surveillance have to cope with such degraded signals affected by different types of noise, such as sensor noise or compression artefacts.

3.2. Background subtraction schema

In the literature, we can find so many methods in which the background subtraction algorithms are used, but most of them follow a simple flow diagram, which is shown in the figure 1. The four steps in a background subtraction algorithm are:

1. Pre-processing
2. Background modelling
3. Foreground detection
4. Data validation

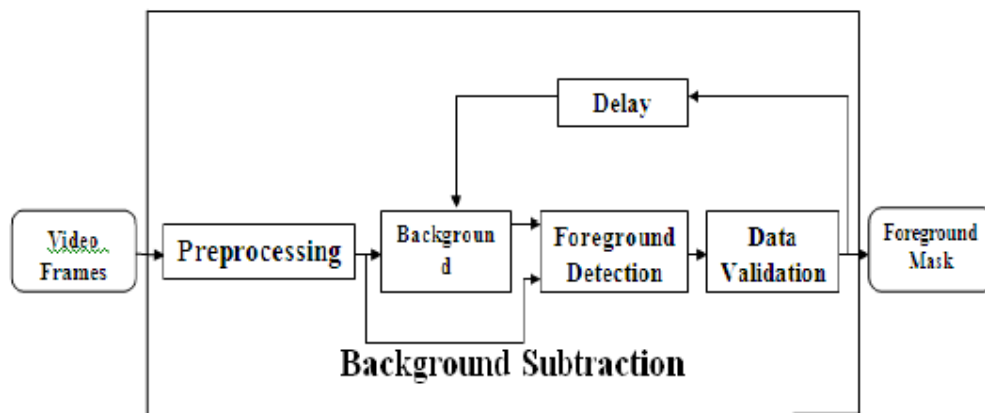


Figure 1. Flow diagram of a generic background subtraction algorithm

Pre-processing: This step consist of a collection of simple image processing tasks that change the raw input video into a format that can be processed by the next steps. In the early stage of processing, simple temporal or spatial smoothing are used to reduce noise such as rain and snow. In pre-processing, the data format used by the background subtraction algorithm is a very important key. Most of the algorithms handle luminance intensity, which is one scalar value per each pixel. However, colour image, in either RGB or HSV colour space, is becoming more popular in the background subtraction literature [8, 15]. These papers argue that colour is better than luminance at identifying objects in low-contrast areas and suppressing shadow cast by moving objects. In addition to colour, pixel-based image features such as spatial and temporal derivatives are sometimes used to incorporate edges and motion information. For example, intensity values and spatial derivatives can be combined to form a single state space for background tracking with the Kalman filter. Pless et al. combine both spatial and temporal derivatives to form a constant velocity background model for detecting speeding vehicles. The main drawback of adding colour or derived features in background modelling is the extra

complexity for model parameter estimation. The increase in complexity is often significant as most background modelling techniques maintain an independent model for each pixel.

Background Modelling: Background modelling is at the heart of any background subtraction algorithm. Much research has been devoted to developing a background model that is robust against environmental changes in the background, but sensitive enough to identify all moving objects of interest. We classify background modelling techniques into two broad categories, non-recursive and recursive. They are described in the following subsections.

- **Non-recursive techniques:** A non-recursive technique uses a sliding-window approach for background estimation. It stores a buffer of the previous L video frames, and estimates the background image based on the temporal variation of each pixel within the buffer [6].
- **Recursive Techniques:** Recursive techniques do not maintain a buffer for background estimation. Instead, they recursively update a single background model based on each input frame. As a result, input frames from distant past could have an effect on the current background model [6]. Recursive techniques require less storage not like non-recursive techniques, but any mistake or error in the background model can lead us to a much longer period of time.

Foreground detection: This step makes us able to compare the input video frame with the background model, and identifies foreground pixels from the input frame. There is some techniques that doesn't use the same image as a background model like the Mixture of Gaussian model (MoG), but the most of techniques use a single image as their background models, so the approach for foreground detection is to check whether the input pixel is different from the corresponding background estimate [17]:

$$|I_t(x,y) - B_t(x,y)| > T \quad (6)$$

Another popular foreground detection scheme is to threshold based on the normalized statistics:

$$\left(|I_t(x,y) - B_t(x,y) - \mu_d| / \partial_d \right) > T_s \quad (7)$$

Where μ_d and ∂_d are the mean and the standard deviation of $I_t(x,y) - B_t(x,y)$ for all special locations $(x; y)$. Most schemes determine the foreground threshold T or T_s experimentally [17].

Data Validation. To improve the candidate foreground mask based on information obtained from outside the background model, we define data validation process. All the background models have three main limitations: First, they ignore any correlation between neighbouring pixels; second, the rate of adaptation may not match the moving speed of the foreground objects; and third, non-stationary pixels from moving leaves or shadow cast by moving objects are easily mistaken as true foreground objects. When the background model adapts at a slower rate than the foreground scene, large areas of false foreground, commonly known as ghosts, often occur. Sophisticated vision techniques can also be used to validate foreground detection. Computing optical flow for candidate foreground regions can eliminate ghost objects as they have no motion. Colour segmentation can be used to grow foreground regions by assuming similar colour composition throughout the entire object [6].

4. EXPERIMENTAL RESULTS

In this section, we will track objects on a video of two seconds, this video has been taken from the CAVIAR DATA web site that contains a huge data of images and video clips with different scenarios of interest for the CAVIAR project using a wide angle lens along and across the hallway in a shopping centre in Lisbon. The resolution is half-resolution PAL standard (384x288 pixels) and compressed using MPEG2. The technique of background subtraction is used to better identify the objects in the surveillance videos. With this algorithm, we can get items to regions in

each image. We can continue to acquire the minimum bounding boxes of the objects and obtain the coordinates of the centroid of each object, which are then used to monitor the positions of objects through video frames.

4.1. Background generation

An image sequence of the background scene is first recorded. This sequence contains 68 images corresponding to two seconds with an acquisition rate of 34 frames per second.

4.2. Image segmentation

The detection image is the image resulting from comparing the current frame with the background model. Therefore it contains the elements of difference between these two images. Also, each pixel of the image is classified as part of the background or foreground element. The pixels classified in the first category will have a value of 0 (black) in a binary image of detection, and the other pixels will have a value of 1 (white).

4.3. Tracking

One of the most important issues in our subject is how an object is presented in a monitoring scenario. So an object can be defined as an object of interest for further analysis. In our case for example, we are tracking a person, so we have chosen our object which will be represented by his centroid.

- The surface method: in this method we tried to calculate the surface of each region in the binary image, then applied the code of centroid so it will associate the centroid to the largest region that represents for us the person to track.
- The K-NN method: the K-NN algorithm calculate the distance between a specific point with the others, and then find out the K nearest points to it, by then we calculate the sum of the distances related to a point so we can choose the smallest one.

The optical flow method is generally widely used for the detection and tracking of targets, this approach is an approximation of the movement of object tracked by estimating origins vectors of pixels in image sequences. Thus, the technique of calculating the optical flow is a differential method where we need it to estimate, at any point, the movement vector as a function of intensity changes of the pixel and its neighbours. These methods can accurately detect movement in the direction of the intensity gradient, but the movement that is tangent to the intensity of the gradient cannot be well represented, i.e. the optical flow methods are sensitive to illumination changes. Figure 4 shows a caption of a video tracking with the optical flow method.

To track a person on a video sequence, means that we need to trait every single image of the sequence, for example here we have 68 frames for a video of 2 seconds, so we need the whole process to be as fast as possible, because it will be applied to all the video sequence.

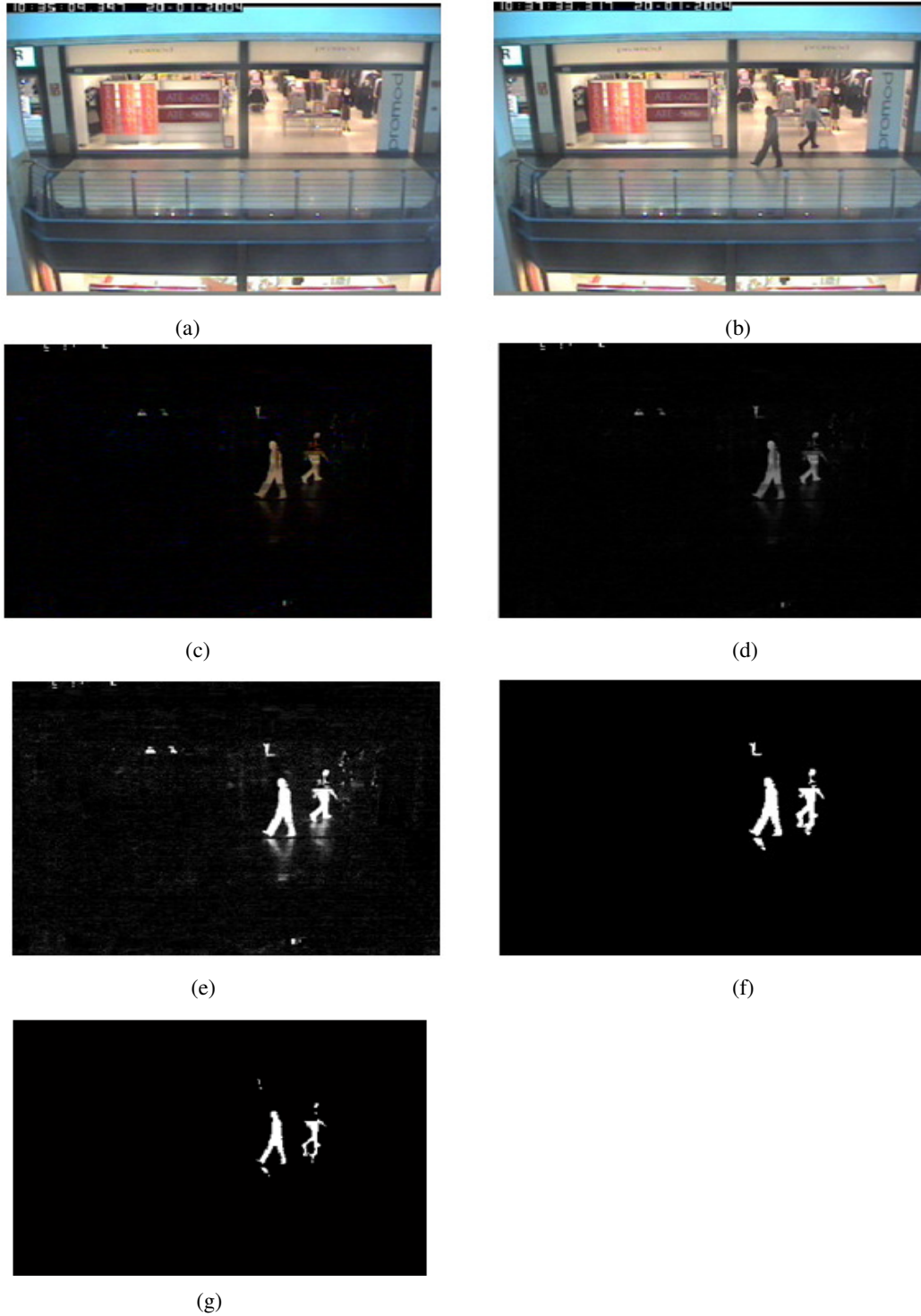


Figure 2. (a) The background model, (b) input image, (c) difference image between (a) and (b), (d) converted (c) to grayscale, (e) adjust intensity values of (d), (f) small Object removed from (e), (g) eroded image.

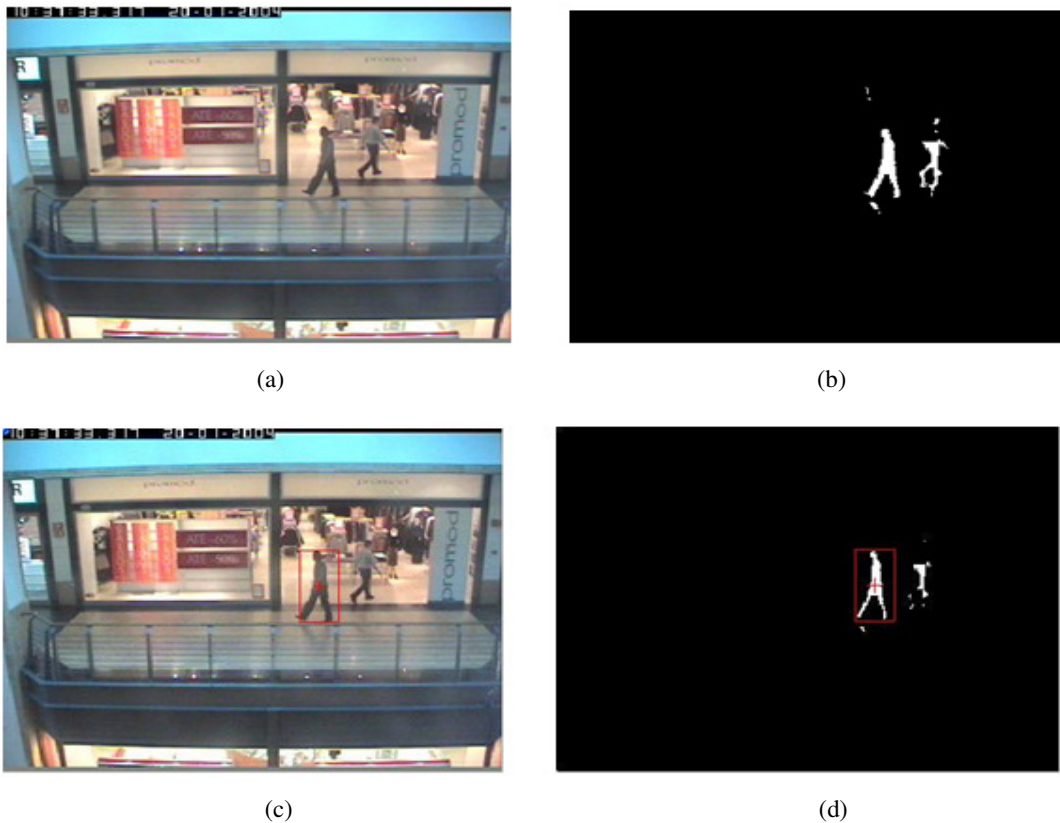


Figure 3. (a) And (c) tracking in a colour image, (b) and (d) tracking in a binary image

After executing the background subtraction method that we used, the total execution time calculated between converting the image to a grey scale and finding the centroid to track the person, was 3.797212 seconds per frame, which means $3.797212 \times 68 = 258.210416$ seconds, about 4.30 minutes for the whole sequence. While using the optical flow method (figure 4), the execution time was 425.59 seconds, about 7.1 minutes for the whole sequence. Therefore, the execution time is such an important task in video tracking, because here we're talking about a video of 2 seconds, but in real life, video surveillance stock huge quantities of videos and frames, which mean we need a very fast method to deal with this such a huge quantity of information.

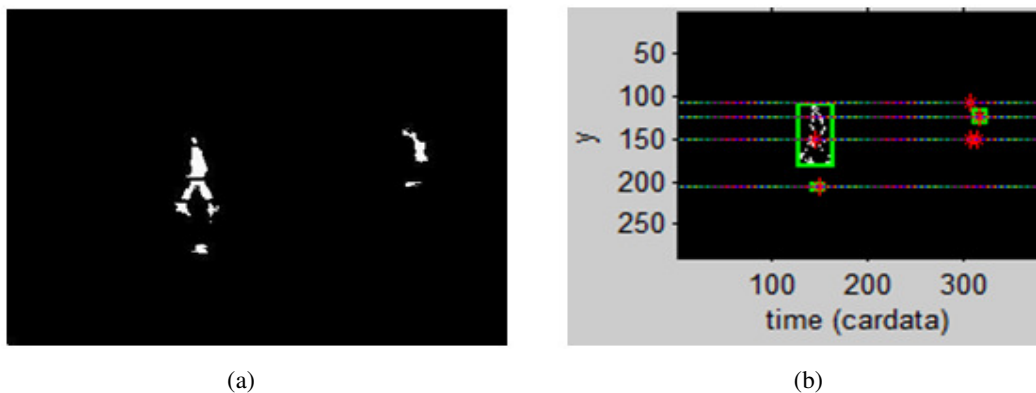


Figure 4. (a) A frame from the video, (b) the video tracking with the optical flow

3. CONCLUSION

In this paper, we introduced the object tracking and its importance now days, motivation and applications of this field and the classified methods of Video sequence analysis. We survey a number of background subtraction algorithms in the literature. We presented the challenges of this method and analyse it in pre-processing, background modelling, foreground detection, and data validation. Also, we have proposed our novel method using the background subtraction algorithm to extract object based on his surface information represented by centroid of this binary region, after that we used K-NN to track the target from frame to frame using the smallest distance between two detected regions represented by their centroids, then compare it with some other techniques.

ACKNOWLEDGEMENTS

This research was supported by the National school of applied sciences and the University of Ibn Tofail Kenitra and we are thankful that they have provided expertise that greatly assisted the research.

REFERENCES

- [1] Shipra Ojha and Sachin Sakhare, Image Processing Techniques for Object Tracking in Video Surveillance- A Survey, International Conference on Pervasive Computing, Department of Computer Engineering Vishwakarma Institute of Information Technology, india, 2015.
- [2] R.Venkatesan and A.Balaji Ganesh, Real Time Implementation On Moving Object Tracking And Recognition Using Matlab, Dept of TIFAC CORE Velammal engineering college Chennai, india, 2012.
- [3] Junzo Watada, Zalili Musa, Lakhmi C. Jain, and John Fulcher, Human Tracking: A State-of-Art Survey, Waseda University: Japan,University Malaysia Pahang: Malaysia, University of South Australia: Australia, University of Wollongong: Australia, 2010.
- [4] Patrick Dickinson and Andrew Hunter, Scene Modelling Using An Adaptive Mixture of Gaussians in Colour and Space, Department of Computing and Informatics, University of Lincoln, UK, 2005.
- [5] Alper Yilmaz, Omar Javed and Mubarak Shah, Object Tracking: A Survey, Ohio State University, University of Central Florida, USA, 2006.
- [6] Harsha Varwani, Heena Choithwani,Kajal Sahatiya, Shruti Gangan, Tina Gyanchandani and Dashrath Mane. Understanding various Techniques for Background Subtraction and Implementation of Shadow Detection., VES Institute of Technology, Chembur, 2013.
- [7] Sebastian Brutzer, Benjamin Hoferlin and Gunther Heidemann, Evaluation of Background Subtraction Techniques for Video Surveillance, Intelligent Systems Group, Universitat Stuttgart, germany, 2014.
- [8] C.Wren, A. Azabajejani, T. Darrel, and A. Pentland, P finder: Real-time tracking of the human body, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, pp.780-785. July 1997.
- [9] R. Cutler and L. Davis, view-based detection and analysis of periodic motion. Fourteenth International Conference on Pattern Recognition, vol.1, pp. 495-500, Brisbane, Australia, Aug 1998.
- [10] A. Elgammal, D. Harwood, and L. Davis, Non-parametric model for background subtraction in Proceedings of IEEE ICCV'99 Frame-rate workshop, Sept 1999.

- [11] C. Stauffer and W. Grimson, Learning patterns of activity using real-time tracking, in IEEE Trans on Pattern Analysis and Machine Intelligence, vol. 22, pp. 747-57, Aug 2000.
- [12] P. KaewTraKulPong and R. Bowden, An improved adaptive background mixture model for real-time tracking with shadow detection in Proceedings of the 2nd European Workshop on Advanced Video-Based Surveillance Systems, Sept. 2001.
- [13] M. Harville, A framework for high-level feedback to adaptive, per-pixel, mixture-of-Gaussian background models, , in Proceedings of the Seventh European Conference on Computer Vision, Part III, pp. 543-60: Copenhagen, Denmark, May 2002.
- [14] D. R. Magee, Tracking multiple vehicles using foreground, background, and motion models, , in Proceedings of the Statistical Methods in Video Processing Workshop, pp. 7-12: Copenhagen, Denmark, June 2002.
- [15] R. Cucchiara, M. Piccardi, and A. Prati Detecting moving objects, ghosts, and shadows in video streams, IEEE Transactions on Pattern Analysis and Machine Intelligence. vol.25, pp. 1337-1342, Oct 2003.
- [16] T. Bouwmans, F. El Baf, and B. Vachon, Background modeling using mixture of Gaussians for foreground detection-a survey, Recent Patents on Computer Science. Vol.1(Num. 3), pp. 219-237, 2008.
- [17] Sen-Ching S. Cheung and Chandrika Kamath, Robust techniques for background subtraction in urban traffic video, Center for Applied Scientific Computing: Lawrence Livermore National Laboratory, 2004.
- [18] C. Qin, H. R. Ren, C. C. Chang, Q. K. Chen, Novel Occlusion Object Removal with Inter-frame Editing and Texture Synthesis, Journal of Information Hiding and Multimedia Signal Processing, vol. 7, no. 2, pp. 386-398, 2016.
- [19] Jean-Philippe Jodoin, Guillaume-Alexandre Bilodeau, Nicolas Saunier, Background subtraction based on Local Shape, Ecole Polytechnique de Montreal P.O. Box 6079, Station Centre-ville, Montreal, (Quebec), Canada, H3C 3A7, 2012
- [20] <http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/>

AUTHORS

Ait Moulay Asmaa:

is a Phd student of computer engineering at ENSA of Kenitra, University of Ibn Tofail. She received her Master's degree in information systems security from the national school of applied sciences, and she is currently completing her researches on field of detecting and tracking individuals from multiple camera with the BOSS team in the LGS Laboratory, ENSA, Kenitra, Morocco.



Amine Aouatif:

Received her Master's degree (DESA) in Computer Sciences and Telecommunications engineering from the faculty of sciences, Mohammed V-Agdal University, Rabat, Morocco, in 2004. She remained at Computer Sciences and earned a PhD degree in 2009 for a dissertation titled "Feature Extraction and Selection for Dimensionality Reduction in Pattern Recognition and their Application in Face Recognition". Her research interests include but are not limited to dimensionality reduction, feature selection applied to face detection and recognition and driver hypo-vigilance. She joined the ENSA, Kenitra, Morocco, in 2010, as an assistant professor. Since November 2010, Aouatif Amine has been a Vice-Chair of IEEE Signal Processing Society Morocco Chapter. Since April 2015, Aouatif AMINE is the BOSS team header in the LGS Laboratory, ENSA, Kenitra, Morocco.

