Different Techniques of Object Detection and Tracking: In Video Monitoring System

Saurabh Patel and K Sampath Kumar

May 27, 2020

Different Techniques of Object Detection and Tracking: In Video Monitoring System

[1*]Saurabh Patel, [2]Dr. K Sampath Kumar

[1*]Department of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

[2]Department of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

[1*]saurabh1997patel@gmail.com, [2]k.sampath@galgotiasuniversity.edu.in

ABSTRACT

The paper includes the various methods which are related to object detection and tracking in live video surveillance to detect the object like the face or can be used to detect the people, cars in a security camera. These days we can easily find that people are following social distancing due to COVID -19. This paper point towards the various methods of detecting the object (classification) and tracking (GMM tracking). This paper points toward the detection of movable objects in the live video monitoring then tracking will track the moving object. Detecting a moving object is really a very big task and it the origin of the method. Object detection is really difficult to implement which depends upon the shape size and color of the object. In this paper, we will study the background subtraction using the pixel-based method, optical flow method, color-based method gradient-based method and frame differencing. We will also study tracking methods like kernel-based method silhouette-based method, and point-based methods.

KEYWORDS

Object detection, Object tracking, temporal frame differencing, Foreground detection method, point-based methods.

I. INTRODUCTION

Video Monitoring is a method of analyzing one scene or multiple scenes for a particular behavior that can be considered to be video sequences. It covers the area of computer vision. We can use video monitoring in security camera sports function, public transportation like airports buses, etc. especially for that bounded by the community spaces. It is the process of identification of particular cameras that can able to analyses the particular area. Video monitoring can be of 3 types fully autonomous [9], semi-autonomous, and manual. In fully autonomous [9], the system the image sequence input and rest process automatically. Semi-autonomous involves video monitoring with some human work and some automatically processing. Manual video monitoring includes the fully humanoid process. If it can be possible to track the security trends when occurred or about to occur in the future. By just watching the few video sequence it is easy to analyses the improper activities have happened. One such application if for live face detection [4]. Following a series of various images per second in a video, they try to analyses and learn the video sequence, properly let's say it takes two files one is a real video file and the other is fake. For an understanding of the system, it analyses all the files of the video sequence and tries to find which face is real or fake. Video monitoring has evolved a greater speed General surveillance is only using the CCTV for security. It access controls and alarms have too many similarities but when they used as an integrated system to keep the improper activities from out of the risk zone. That's why more and more healthcare organization continues to adapt to this kind of technologies. The most important thing in video monitoring is the texture and quality of the image received and this largely depends on the camera and equipment used. Video monitoring can work as an object tracking classification and detection. Monitoring the video for a long time by humans is impossible and not feasible that’s why autonomous can be of great help [8]. There are various object detection systems like person and anomaly identification. Intelligence visual monitoring (IVF) includes the autonomous systems that involve the process of analyzing the video sequence which is object behavior and then tries to detect it and track it to understand the importance of the scenes [10]. IVF focuses on wide-area monitoring and scene processing. It can be used to understand the behavior of the object.

II. OBJECT DETECTION

Object detection is the very first step to analyses the video. Our aim towards are object detection is to discover the finding of an object within the video frame or image. For the process to detect the object apart from the static background which can be the pixels that remain unchanged pixels throughout the process frame after frame.

2.1 Detection of an image and video

Photos can be of different format which contains .png, .jpeg, etc. have static pixel size but we cannot use the motion to detect the image input object but we must find new methods to detect for parsing of the input objects. When an image started with different positions, locations of the street the busy nature of the image makes it difficult for the system to process it. We can use edge detection that helps to determine the (obj.) input in the scene. Edges implement the object boundaries and find the grayscale labelling of the image. Edges not only helps in detecting the particular image but also to define the easiest way to interpret the most complicated image. For moving objects, every tracking methods need a detection algorithm for each frame or object to appear in the window. Object detection can also for people counting. Object monitoring can be used to detect the counts of people in exhausted places like malls, festivals, and cinemas.

These tend to more difficult as people move out of the frame perfectly and also because and people are moving objects and object is also used for industrial processing to identify products. Something always added moved and removed every day. The system can perform autonomous object numbering and object localization to improve accuracy [11]. It combines a variety of techniques to perceive the surrounding of techniques including the radar light GPS and also computer vision nowhere object detection comes in place the movement it senses a person on the road on its way the car automatically stops advance control interprets sensory details to find the appropriate paths as well as obstacles. Now our final use cases are most imp is security now it uses used in the banking industry as well as the phone industry or technological industry. It is used in the banking industry to detect Freud the fore try of notes and also theft now and one of the most important use is facial recognition which is popularly known as face unlock.

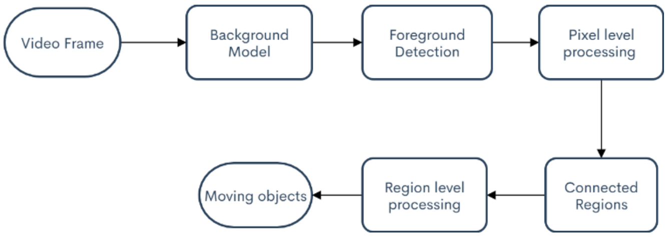

Figure 1: Object Detection System. [3]

2.2 Foreground detection

Foreground detection is used to detect foreground in real-life photos. The main method act as a separator of different video sequences. Every object detection first step is used to find the foreground object. By separating the foregrounds of the object it will help in object classification and understanding the behavior of the objects and also reduces the time complexity of the processing since the pixels can relate to the foreground object [1]. The initial step is the background scene initialization. There are various methods to design the background image. The background processing part of the system keeps it a compact and its coupling to a min so that the whole system can work efficiently [7]. The second procedure is to find the pixel by background model and present the image of the monitoring. The pixel method is based on background and tries to update at every movement when the image changes [4]. Also due to interference in the image also causes noise in the current pixel of the foreground. We can perform various methods to remove the noises like post-processing in foreground pixels [13]. Once we get the extracted noiseless pixel then we will calculate the bounded rectangles and component labelling algorithm. The regions may contain the broken regions due to the problems in the foreground segmentation. So small gaps cased due to environmental noise will be removed in his process known region level post-processing step and the final step is to extract the regions from the current image.

2.3 Pixel-based post-processing

The foreground detection may results in causes a lot of noises in the pixel. Basically it has many reasons. To remove permanently noises from the system we further post-process at the pixel level. There are various factors of the noises:
Camera Noise- Camera noise caused due to image accretion parts. This noise caused due to environmental accretion noise. The noise may produce due to the intensity of the pixels which due to the edge between the two different colored objects in the process tries to fit one color of the object in one pixel and another image in another pixel [12].

Reflected Noise- It is caused due to the light beam particles. When the light transfers from one location to another location. It causes diffraction in the background scene [12].

Colored object noise- When object color is the same due to the background reference makes extremely difficult to detect the foreground pixels [12].

When we adjust the pixel of the image file as mentioned to the nearer image it includes two methods of erosion and dilation. Dilation includes the addition of pixels and Erosion involves removal of the pixel and LPF which passes signals with a lower frequency than the selected one and weaken the signals with a frequency higher than the range to remove the disturbance from the image [3]. The main aim to remove the noisy foreground which doesn’t relate with main foreground regions, and used to remove disturbance from the background pixels near and outside of the object region. We used a low pass to remove the disturbance and blurring of the image. Blurring image is a pre-process in the image to smoothen the image file to delete the outlier’s pixels that may a disturbance in the image. Blurring is the part of the low pass filter. Let’s take a part named Gaussian blur [9], for example, let’s take an eye of a cat with a white rectangular shape. Now zoom in the eye of the cat as shown in the figure. When we apply the blur filter in the image we take one pixel at a time. The pixel we are taking is highlighted by red on the other side.

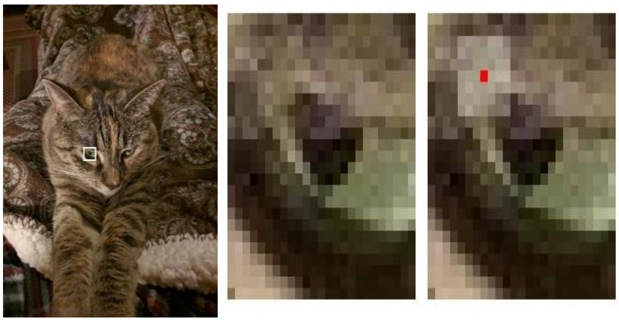

Figure 2: Pixel-based post-processing in the cat eye.

2.4 *Detecting Connected Regions*

After finding the foreground regions and applying the post-processing in the images to remove disturbance the major task is to detect changes in the image processing by grouping the connected regions and after finding the single region that relates to an object the boxes surfaces must be calculated.

2.5 *Region label post-processing*

As we remove disturbance from the pixels some AI very small regions remain left due to the inferior segmentation, to remove these types of regions we must remove the entrance of the foreground pixel. Once the regions are segmented we can remove the particular image from the process we can extract properties of the particular image. These properties are the CoM or known as Centroid and bounded regions of the connected components.

A. *Background subtraction using Alpha*

Object finding can only be made by building the depiction of the part of the image known as the background model is used to find the digressions in the model for each looming frame. Any notable mutation in the region from the model shows the movable object. When we combine the pixels the regions may undergo changes are pointed towards further processes. Using a connected component labelling by using in graph theory where subsets are uniquely connected. This process used in background subtraction. These methods were founded by Sliven and Heikkila. Right now the framework is started with some frames and these frames are updated regularly to adapt the dynamic changes. As the foreground pixel detected by removing the intensity value from the cut-off point of differences with changing entry value per pixel. The entry and reference background changes using foreground pixel information. It tries to find in the movable parts by removing the pixel of the image step by step from the recommendations of the background image that is created by balancing the image over time. The pixels where dissimilarities are above the entry value called as foreground. After that, we will analyses the geometrical structure by adding or removing the pixel as dilation and erosion are performed to remove the noises in the images [5].

We can mark pixels only as foreground if the condition of equality is not satisfied.

$$| I_t(x,y) - B_t(x, y)| > T \qquad (1)$$

Where T is the predefined entry value. The equation of the background image Bt changes using the 1st order reversal filter equation shown in the below:

$$B_{t+1} = \alpha I_t + (1 - \alpha) B_{t a} \qquad (2)$$

Where α is the adaption coefficient the idea is to find the new knowledge about the image is coming into the current scenario? After the rapid changes in the process scenes are updated to the background image.

B. *Temporal Frame Differencing*

This method is the subtraction of the two or more successive frames in the video format to extricate the movable areas. It is a highly complaisant process to undergo scenes in the rapid changes. Moreover, the process becomes a failure in extracting all the particular pixel of a foreground object particular when the object is in slow motion. When a foreground stops working temporal methods gets failed in finding a difference between the regular frames and losses in the objects. Suppose I want to predate the atmosphere for Friday and we have some model to find the atmosphere of Friday. In the std case, we will wait till Friday and then adjust all our models. However when today Thursday is then we have a very good idea about the atmosphere of Friday. Let’s take in(x) represents the grey level concentrated value at a particular position of the pixel is x and at a particular time instance n of video monitoring input sequence I which is in the range [1,255]. T is the entry value which is predefined. It developed by Lipton as two frames differencing told that when a pixel moves it can easily satisfy in [3].

$$| I_n(x) - I_{n-1}(x) | > T \qquad (3)$$

C. *Eigen Background Subtraction*

This method was founded by Oliver. He gives the Eigenvalues for the segmented moving objects. We can remove extensiveness using PCA. The subtraction method states that reduced spaces are formed using PCA should be represented as a static portion in the frame rest moving objects if the object moving in space. The main steps of their algorithm include [8]:

1. Compute C=AAT where C is the covariance matrix.
2. Using covariance matrix C the Eigenvalue L and eigenvector phi is calculated.
3. If M eigenvector has the largest value these values may retain.
4. For the new frame first projected onto the spaced by eigenvector and then the projectile coefficient of projectile and eigenvalues form the reconstructed from I’.
5. Compute I-I’ since the space formed due to the static and Eigenvalues.

III. OBJECT TRACKING

Object tracking is the activity of positioning the object over a movement of time on webcam. It has various uses which include video chat video monitoring video compression augmented reality traffic control and video formatting. Video tracking is a complex process because it takes a longer large amount of the data to process in the video. A further task is the object recognition techniques used for tracking which is a very complex algorithm. Tracking includes matched objects like motion location color texture shape etc. It may also be interpreted as the area it occupies by the object at a particular period of time. In the tracking method, the object can be represented as the shape like a rectangle circle and square etc. The shape used to define the limit of the motion [8]. For example, if an object is used as a point to represent then we can only use a translational model. In case of geometric shape representation like circle ellipse projectile transformation will occur. These methods represent the precision of the motion of the rigid body object. For a complaisant object, the contour is the most graphical representation and both parametric and non-parametric is used to define the motion of the object [14].
Different object tracking methods are:

3.1 Point-based tracking

The point tracking method is used to define as the detected objects as points. Tracking is used to associate the point throughout the frames. Association formed on the basis of one frame is based on another. This can be true when the complexities increase for the miscommunication in the detection. The point tracking methods divide into two methods statically and deterministic method and the biggest difference in these methods how they minimize the cost of problem-solving. This method requires an external mechanism to process it. The processing in there similarities of the object in the time frame t-1 to a particular object called the corresponding cost. The constraints listed down used as the combination in the corresponding cost [2]. Each tracking pointer should have three regions:

Region of quality
The quality region shows the object tracking which is trapped inside the square. This should round up a definite discernible object, and it should be clearly recognizable within the duration of the track, albeit there are changes in proper lighting, backgrounds, and measurable angle.

Region of finding

The region of finding try to search the features of the object. The image input we are tracking should always be within the area of the track but we shift the input imager from one location to another location the tracking should adjust according to the shape size and object. When the area of the track is smaller than tracking becomes at a faster rate but due to the small frame also causes the other region of dislocation may occur when leaving the region of finding.

Target Point

The target point is the point which we are tracking. It is the place where after effect or destination layer will be placed. This should be position at the center of finding.

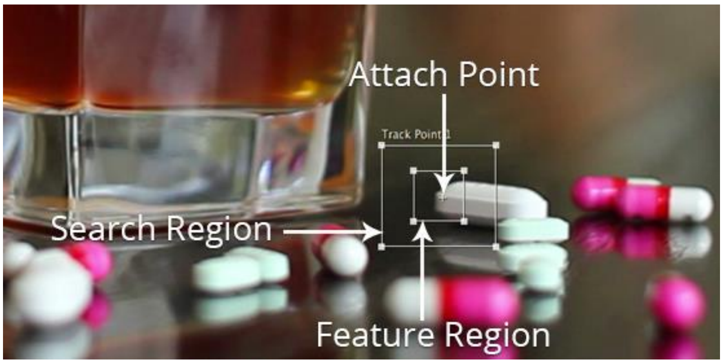

Figure 4: Tracking region on white pill

3.1.1 Deterministic method

When we develop the states for further processes it doesn't involve any randomness is called deterministic method. This means that by using initial state S0 always produce the result in any condition. When the deterministic method follows object tracking then object movement can be expected in trajectory motion [2].

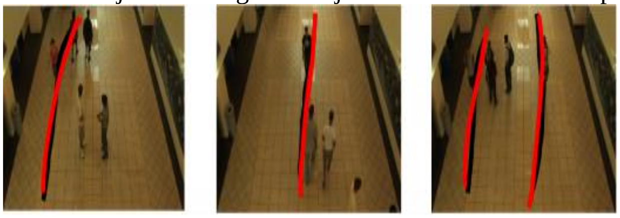

Figure 5: Black line showing the path on which people are moving and red line shows the path we estimated on which people will run.

3.1.2 Statistical Method

When unpredictable factors like a disturbance in measurement and unstable motion are very difficult to avoid. The static method take these problems as references while estimating the object state which makes them robust. This model considered more complex than the deterministic method [2].

3.2 Kernel-based tracking

It is a computation process that tries to restrict the image input and based on that try to maximize the similarity measure. Let's say we are taking a target frame at t-1 second and put in into the correlation filter then we target the preliminary stage of imager input at time t second. Now we will us the names of the tools as scale estimation which measures it with the response and weight after that we use the target the frame at time t now we use scene classification to find the similarity response holistic response and block response and finally, some adaptive updating mechanism includes 2 factors first is target future of the image template and the learning rate. We use different shapes like rectangle square ellipse to represent the tracking of the object. Moving object differs by its shape size location at each and every time. The shape which covers the input image is used as an identification of the object. Kernel tracking can be also called as appearance tracking because it covers all the features of the object. The kernel tracking is divided into 2 parts one is single view (known as template matching) and another one is multi view [2].

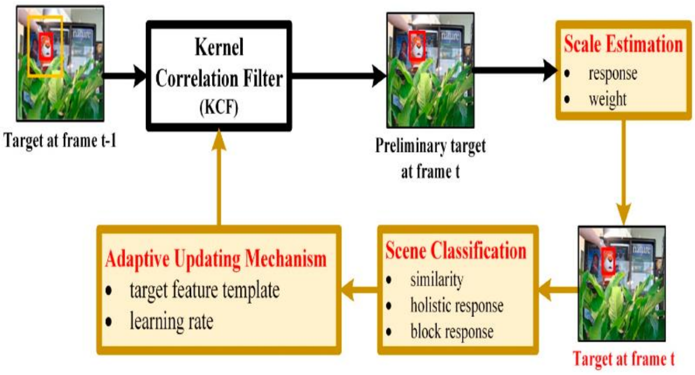

Figure 6: Kernel tracking in a frame

3.2.1 Template-based

Template matching matches an actual image patch over an input image the location of template match loc can be obtained by setting up the system object vision dot template matcher and using the steps here in the test here and the grayscale input image I grey and template let's move the Matlab let us load a mat file bike template which contains template image of a bike we can view it by using the "imshow" function. Now let's load an input bike image and view this as well recall the template matching need a grayscale image let us convert RGB image to greyscale image we can now set up the vison dot matcher system object. There is a metric property envisioned or template matter which allows us to compute to the difference between the original image pixel and corresponding image pixel by default this set to the sum of difference in the absolute where you sum the absolute values of the difference between the corresponding pixels some of the other options can be used are of squared differences and maximum absolute differences [2].

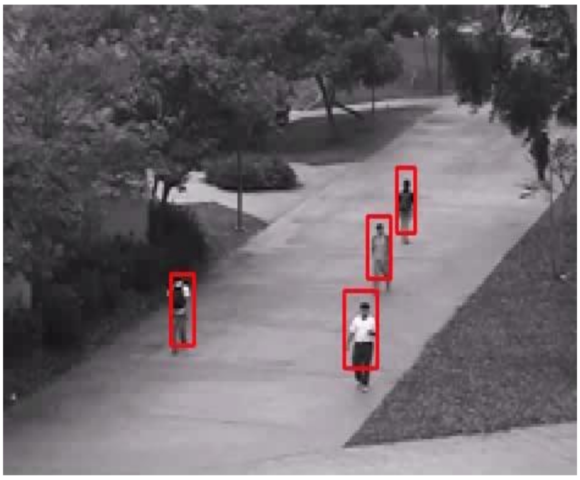

Figure 6: Tracking objects using geometrical shape

a) Single object tracking

The template approach is most commonly used for single object tracking. They have multiple features that we can use like color, shape, and size. The basic approach is to search the specific pattern in the image file hence we have to match the template to a specific portion of the image [2].

b) Multi-view object

When we process the multi-view method generates a problem that we have to represent the objects with the latest observation. This means the representation only considered the present for the visible view it, means only for one view. So if the tracking undergoes a major change from one position to another at a point of reference then this method becomes irrelevant and tracking may be lost. This problem can cause the blocking of the object and object night to leave the frame [2].

3.3 Silhouette-based tracking

This method is called as the region-based tracking. When we represent a complex shaped object using simple geometrical shapes will be deficient. If we represent the human body by a cylinder or skeleton model the tracking method will be wastage due to the bad object representing. With the silhouette method, this makes it possible to find the accurate space of the object. The silhouette can be used in autonomous gait identification which can be identified when the human body can be in motion. The machine includes the shape of the human body and the motion of movable objects. Identification of the people can become very complex when input image body position changes at each movement of time and motion. We can take gait motion into consideration that it consists of the human body position by considering the temporal differences. We have divided the model into three procedures: Human detection and tracking, extraction by features of an object, and classification of image input. For first step, we can take human detection and tracking, a method where we convert image input as a background modelling in which the position of the image. The process of background subtraction is used to convert the object detection in compartments now the second process is the person blob sequence in which we perform extraction and convert it into the 2D image and then again convert it into 1D normalization by contour unwrapping with respect to the CoM. The third process is to compute the eigenvalue in the classification phase [2].

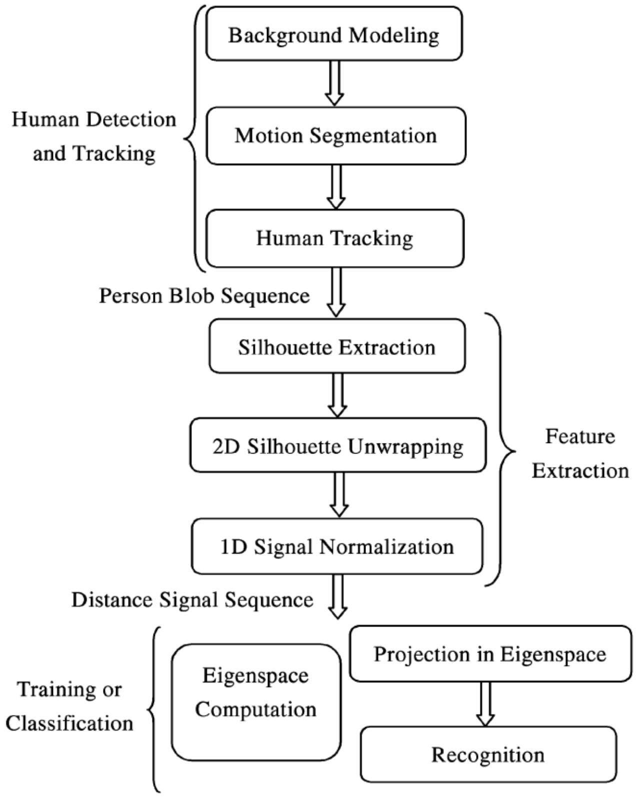

Figure 7: Silhouettes tracking in the model. [2]

3.3.1 Shape matching

This method used to generate a model from the past frames and used to in current or for future frames. In this method, we modeled the object in the form of histograms or edges or the combination of these two. To handle complex methods like rigid body changes in illumination and the viewpoint we update the model on each and every frame [2].

3.3.2 Contour evolution

These are called boundary tracking which uses the previous frame contour as the initialization of the present contour. This method uses the edge-based method which inconsiderate to illumination changes makes the algorithm strong. This approach is faster than the shape matching because the area of boundaries is very less than including the complete region. This method works on overlapping which means some parts of the object region must be taken over the object region of the previous frame [2].

IV. CONCLUSION

To analyze the static objects like image and refined information, image improvement object detection and tracking and understanding the nature of the object have been studied. In this paper, we have studied the various methods of object detection and tracking for video monitoring. We have described the background subtraction using alpha and temporal differencing. We have also discussed detection techniques. The failure of temporal difference occurs when the object is in slow motion or in uniform motion. When the object comes in the rest of motion the methods stops working and loses the path of the object. We have described background subtraction because it fast and efficient algorithm. This research paper gives you a brief idea about the important research regarding the future work of computer vision and its application. Here we studied about the tracking methods like point tracking which includes deterministic and statically methods in which Kalman filter and particle filter are the most popular used to track the object, kernel tracking which includes single view object and multi-view object and various other methods like mean shift method to find the targeted region of the object and silhouette tracking which includes shape matching (for recognition from past frames) and contour evolution (boundary method).

Funding:
This study was not funded by any Institution.

Conflict of Interest:
The authors declare that they have no conflict of interest.

REFERENCES

1. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. Int. J. Comput. Vision 1, 321–332, 1988.
2. W **Sanna A° gren,** Object tracking methods and their areas of application: A meta-analysis.
3. Kinjal A Joshi, Darshak G. Thakore, “A Survey on Moving Object Detection and Tracking in Video Surveillance System.” In International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012.
4. S. Zhu, and A. Yuille. Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. IEEE Trans. Patt. Analy. Mach. Intell. 18, 9, 884–900, 1996.
5. Elgammal, A. Duraiswami, R., Hairwood, D., Anddavis, L. 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proceedings of IEEE 90, 7, 1151–1163.
6. Christopher R. Wren, Ali J. Azarbayejani, Trevor Darrell, and Alex P.Pentland,”Pfinder: Real-Time Tracking of the Human Body” in IEEETransactions on Pattern Analysis and Machine Intelligence, July 1997, 19(7), pp. 780-785.
7. S. Y. Elhabian, K. M. El-Sayed, “Moving object detection in spatial domain using background removal techniques- state of the art”, Recent patents on computer science, Vol 1, pp 32-54, Apr, 2008.
8. Yilmaz, A., Javed, O., and Shah, M. 2006. Object tracking: A survey. ACM Comput. Surv. 38, 4, Article 13, December 2006.
9. In Su Kim, Hong Seok Choi, Kwang Moo Yi, Jin Young Choi, and Seong G. Kong. Intelligent Visual Surveillance - A Survey. International Journal of Control, Automation, and Systems (2010) 8(5):926-939.
10. A. M. McIvor. Background subtraction techniques. Proc. of Image and Vision Computing, 2000.
11. I. Haritaoglu, D. Harwood, and L. S. Davis, “W4: real-time surveillance of people and their activities,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 809-830, August 2000.
12. Elgammal, A., Duraiswami, R., Harwood, D., Anddavis, L. 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proceedings of IEEE 90, 7, 1151–1163.
13. WuU Z, and Leahy R. “An optimal graph theoretic approach to data clustering: Theory and its applications to image segmentation”. IEEE Trans. Patt. Analy. Mach. Intell. 1993.
14. ISARD, M. AND MACCORMICK, J. 2001. Bramble: A bayesian multiple-blob tracker. In IEEE International Conference on Computer Vision (ICCV). 34–41.