Occluded Parts Identification for Human Recognition System

Dr. Matheswari Rajamanickam

Assistant Professor, Department of Computer Science,
M.E.S College of Arts, Commerce and Science, Bangalore, Karnataka-560096.

Abstract: Object detection and tracking of moving objects play a vital role in digital image processing. Objects needs to be recognized at all the stages considering different factors like size, shape, color, dynamic background, cluttered environment, occlusion, optical flow and with fast moving objects. Humans with very little effort identifies number of objects in images even when they change their nature. Image objects changes in their viewpoints like different sizes / scale, when they are translated or rotated etc. Humans recognize objects even when they are partially obstructed from the view. Employing computers for these kind of tasks are more challenging. Therefore the proposed system combines appearance-based 3D object recognition systems with an algorithm for occluded parts identification. The method performs recognition of random objects in the presence of clutter and occlusion. Bounding box with patch-based framework for object detection is presented in this paper.

IndexTerms - Recognition based on appearance, Bounding box, Patch creation, Occlusion detection.

I. INTRODUCTION

The role of computer vision in object recognition has become essential application in day-to-day life. This involves identifying objects from the input video sequence. Recognition remains challenging in large part due to the significant variations exhibited by real-world images. Reliable object recognition algorithms must fulfill requirements like Reliability, Speed and Automation. There are number of techniques available for object recognition for computer vision. Object recognition algorithms need to be "trained" using digital images. A large number of images must be gathered in order to correctly classify new objects. Research in object recognition is increasingly concerned with the ability to recognize generic classes of objects rather than just specific instances. Numerous methods for object recognition have been developed over the last decades.

The learning module's input is an image, its output is a label, for either the class of the object in the image (is it a cat?) or its individual identity (is it my friend's face?). Learning module can be described as a binary classifier that gives an output of either "yes" or "no". The learning module is trained with a set of examples, which are a set of input–output pairs, that is, images previously labeled.

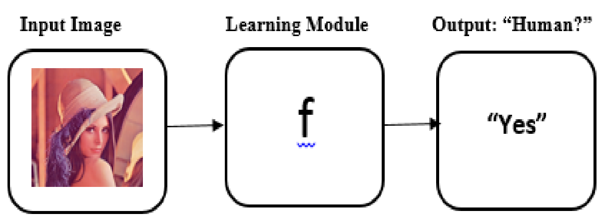


Fig.1. Object Recognition Algorithm - Learning Process

There are a number of techniques available for object recognition for computer vision and their features. Objects can be recognized based on different factors like,

- Appearance-based object recognition systems
- Model based object recognition
- Template based object recognition
- Part based object recognition
- Region based object recognition
- Contour-based object recognition

II. RELATED WORKS

An Active Contour models (ACM) was developed in [1] for detecting and tracking the object system. The association of distinct technique was also employed for detecting and tracking the object with different segmentation. However, the dynamic shadows were demanding issue in foreground segmentation. The objects can be merged or hidden by other objects with the size and shape of the objects also got distorted.

An Advanced Fuzzy Aggregation based Background Subtraction (AFABS) was designed in [2] for moving object detection in the presence of challenging dynamic background conditions. However, the visual display of the segmented images still causes demanding issues and therefore compromising the video quality.

The history of video-based object detection starts from detection of moving objects in videos captured by a stationary camera. Jain and Nagel [4] proposed the frame difference scheme to detect a foreground object. Wren et al. [5] proposed the use of a Gaussian model, Stauffer and Grimson [6] proposed a GMM-based approach, and Elgammal et al. [7] applied kernel density estimation for background modeling. Unfortunately, the above methods cannot serve well for scenarios in which the camera is moving (even with nominal motion). Recent researchers focus more on foreground object detection in videos captured by freely moving cameras.

In [8], Sheikh and Shah proposed to build foreground and background models using a joint representation of pixel color and spatial structures between them. In [3], Patwardhan et al. decomposed a scene and used maximum-likelihood estimation to assign pixels into layers. From their experimental results, only moving foreground objects with the average velocity up to 12–15 pixels per frame can be detected. As a result, their approach is only capable of handling videos captured by a camera with mild camera motions. In this paper, we address automatic video foreground object detection problems under arbitrary camera motion (e.g., panning, tilting, zooming, and translation). Prior methods focusing on this type of problem can be classified into two categories. The first category (e.g., Meng and Chang's method [9]) is to detect moving foreground object as the outliers, and thus to estimate the global motion of the camera. Tang et al. [10] proposed a parametric estimation method for detecting the moving objects.

III. RESEARCH METHODOLOGY

The research work is comprised of the following steps. Initially, the proposed system reads input video sequence and extracts the frame sequences. Further, moving objects are extracted using background subtraction algorithms and basic motion of objects are detection. Next, preprocessing for noise cancellation and object separation are performed. The research work employs patch based framework for occlusion detection and finally object tracking is carried out.

3.1. Appearance-based Object Recognition

Appearance-based object recognition systems are currently the most successful approach for dealing with 3D recognition of arbitrary objects in the presence of clutter and occlusion. For appearance-based models, only the appearance is used, which is usually captured by different two-dimensional views of the object-of-interest. Based on the applied features these methods can be sub-divided into two main classes, i.e., local and global approaches. A local feature is a property of an image (object) located on a single point or small region. It is a single piece of information describing a rather simple, but ideally distinctive property of the object's projection to the camera (image of the object). Examples for local features of an object are, e.g., the color, (mean) gradient or (mean) gray value of a pixel or small region. For object recognition tasks the local feature should be invariant to illumination changes, noise, scale changes and changes in viewing direction, but, in general, this cannot be reached due to the simple features itself. Thus, several features of a single point or distinguished region in various forms are combined and a more complex description of the image usually referred to as descriptor is obtained. A distinguished region is a connected part of an image showing a significant and interesting image property. It is usually determined by the application of a region of interest detector to the image. In contrast, global features try to cover the information content of the whole image or patch, i.e., all pixels are regarded. This varies from simple statistical measures (e.g., mean values or histograms of features) to more sophisticated dimensionality reduction techniques. Since the whole data is represented global methods allow to reconstruct the original image and thus provide in contrast to local approaches, robustness to some extent. Contrary, due to the local representation local methods can cope with partly occluded objects considerable considerably better.

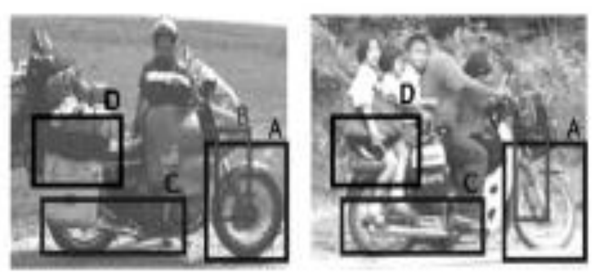
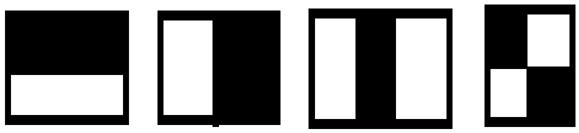


Fig. 2. Illustrations possible regions for histogram features.

3.2. Object Detection

As it is known, a video is a gathering of fundamental structural units, for example, scene, shot and edge. Objects (cars, pedestrians, etc.) are distinguished by the technique for Viola-Jones. This strategy permits the location of items for which learning was performed in [1, 2]. It was composed particularly with the end goal of face location and might be utilized for different sorts of articles. As an administered learning system, the strategy for Viola-Jones obliges hundreds to a great many samples of the located item to prepare a classifier. The classifier is then utilized as a part of a comprehensive quest for the item in all conceivable positions and sizes of the image to be prepared [3]. This system has the playing point of being compelling, fast. The system for Viola-Jones utilizes manufactured representations of pixel values: the pseudo-Haar characteristics. These attributes are controlled by the distinction of wholes of pixels of two or more contiguous rectangular areas (Fig. 3), for all positions in all scales and in a detection window.

The number of features may then be high. The best peculiarities are then chosen by a technique for boosting, which gives a "solid" classifier all the more by weighting classifiers "weak". The Viola-Jones algorithm uses the Haar-like features. The exhaustive search for an item is inside an image which can be measured in computing time.



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

Fig. 3. Examples of features used

Every classifier decides the vicinity or nonappearance of the item in the image. The least difficult and quickest classifiers are put in the first place, which rapidly disposes of numerous negative (Fig. 4).

In general, the technique for Viola-Jones gives great results in the Face Detection or different articles, with few false positives for a low figuring time, permitting the operation here progressively [15]. The recognition of different road moving objects is essential to reduce the impact of having an accident.

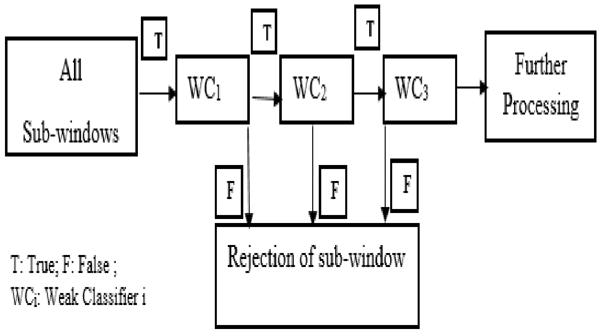


Fig. 4. Cascade of Classifers

We propose to detect and track human objects in video surveillance system in the presence of occlusions using a single stationary camera. We believed that this system is applicable in a wide variety of circumstances. The entire system for occlusion Recognition and Tracking of Occluded Objects in Dynamic Scenes detection is shown in fig1.First step is to capture the video and then it is converted in to frames for further processing. The background subtraction algorithm was performed in all image frames to extract the foreground object from background. After that, morphological operations were performed for noise cancellation. Then object separation is performed to extract foreground images. Then bounding box is created around each object for moving object tracking and the center of mass is calculated for occlusion detection. The boundary patches with fixed size is incorporated for detecting the occluded patches in each object by calculating the distance between the patches of objects in different direction.

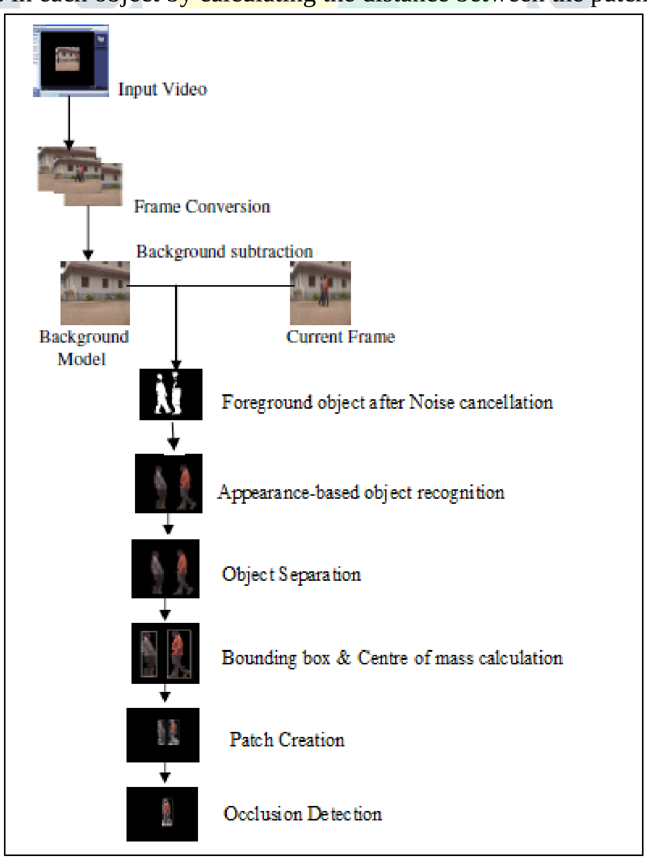


Fig. 5. Object Recognition, Occlusion Detection and tracking

3.3. Moving Object Extraction

In applications using fixed cameras with respect to the static background, a very common approach is to use background subtraction [12]. In various computer vision applications, the primary task is the moving object separation. One of the most common and efficient approach to focus moving objects is background subtraction. Here the static background model is subtracted from the current frames of video sequence. The process of segmenting the moving objects from the stationary scene is termed as background subtraction. The foreground objects are extracted from the background using different background subtraction methods. Background subtraction involves two distinct processes: Background modeling and foreground extraction. In background modelling [13], background is captured by a camera and it is periodically updated. To obtain good tracking of object in a scene, system must have good background estimation and background subtraction algorithm. The main advantage of background subtraction algorithm is that, it does not require a prior knowledge of shapes or movements of the moving objects. The challenges in the design of good background subtraction are; it cannot cope with illumination changes, the detection of unnecessary non stationary background object pixels and shadows. Another disadvantage of background subtraction methods is that, it cannot discriminate moving objects from the backgrounds when the background changes significantly. So the reference background image is updated periodically to adapt with dynamic scene changes. The noise in the background or in the foreground after background subtraction was efficiently removed with filters.

3.4. Background Subtraction Algorithms

Although different, most BS techniques share a common denominator: they make the assumption that the observed video sequence I is made of a static background B in front of which moving objects are observed. With the assumption that every moving object is made of a color (or a color distribution) different from the one observed in B, numerous BS methods can be summarized by the following formula:-

$$X_t(s) = \begin{cases} 1 & if \quad d(I_{s,t}, B_s) > T \\ 0 & otherwise \end{cases} \tag{1}$$

Where *T* is a threshold, Xt is the motion label field at time t (also called motion mask), d is the distance between Is,t, the color at time t and pixel s, and Bs, the background model at pixel s. The main difference between several BS methods is how B is modeled and which distance metric d they use. In the following subsection, various BS techniques are presented as well as their respective distance measure.

3.5. Basic Motion Detection

The easiest way to model the background B is through a single grayscale/color image void of moving objects. This image can be a picture taken in absence of motion or estimated via a temporal median filter [12, 17, 34]. In order to cope with illumination changes and background modifications, it can be iteratively updated as follows:

$$\mathrm{Bs,t+1} = (1-\alpha)\,\mathrm{Bs,t} + \alpha\,.\,\mathrm{Is,t} \tag{2}$$

where α is a constant whose value ranges between 0 and 1. With this simple background model, pixels corresponding to foreground moving objects can be detected by thresholding any of those distance functions:

$$\mathrm{d0} = |\mathrm{Is,t} - \mathrm{Bs,t}| \tag{3}$$

$$\mathrm{d1} = |\mathrm{IRs,t} - \mathrm{BRs,t}| + |\mathrm{IGs,t} - \mathrm{BGs,t}| + |\mathrm{IBs,t} - \mathrm{BBs,t}| \tag{4}$$

$$\mathrm{d2} = (\mathrm{IRs,t} - \mathrm{BRs,t})2 + (\mathrm{IGs,t} - \mathrm{BGs,t})2 + (\mathrm{IBs,t} - \mathrm{BBs,t})2 \tag{5}$$

$$\mathrm{d}\infty = \max\{|\mathrm{IRs,t} - \mathrm{BRs,t}|, |\mathrm{IGs,t} - \mathrm{BGs,t}|, |\mathrm{IBs,t} - \mathrm{BBs,t}|\} \tag{6}$$

Here R,G and B stand for the red, green and blue channels and d0 is a measure operating on grayscale images. Note that it is also possible to use the previous frame I_{t-1} as background image B. With this configuration though, motion detection becomes an inter-frame change detection process which is robust to illumination changes but suffers from a severe aperture problem since only parts of the moving objects are detected.

3.6. Noise Cancellation and Object Separation

The binary object images after background subtraction based on the GMM often have the discrete noises and holes in the object region [16], therefore the proposed system uses mathematical morphological processing operations such as erosion and dilation to remove the isolated noise and fill the hole in the object region. Object separation is performed in the proposed system after moving object extraction. The result of moving object extraction with morphological operation and object separated result are shown in fig 2.

3.7. Object Tracking

Object tracking is a pre-requisite for the important applications of computer vision. To detect and track people in a video sequence, group the pixels that represent individual people together and calculate the appropriate bounding box for each person. Tracked objects are labeled with numbers and surrounded by bounding box.

In the proposed system, to form bounding box, we need four points, starting position (x,y), length and breadth. Fig. 3 shows the object separated image with bounding box. The Algorithm 1 to create bounding box is as follows.

1. Let StX and StY be the Minimum value of row and column minus 0.5 gives starting position(x,y) respectively.
2. Bdth=Max(col)-Min(col)+1;
3. Ln=Max(row)-min(row)+1;
4. Boundingbox BB=[StX StY Bdth Ln];
5. Display(BB);

Algorithm. 1. Bounding box approach

3.8. Patch Based Framework

Patch is a small unit with same size and is created around each object .In the proposed system, the patch based frame works were used for identifying the patch which are going to be occluded. In this novel approach patches are created only at the boundary parts and check the distance (d) between the patch in two objects. If this distance is greater than the predefined threshold, we change the color of the corresponding patches in that objects. To create patches, first we create a mask with its size is equal to the size of the frame. Then, the edges of each object are finding out. From these edges each row and column values are stored in an array and then the size of patch is determined. In our work, the size of patch was selected as 4x4. By using the patch size and number of row values in the edges, we can find the total number of patches that can be created in each object. For creating patches, first we find a suitable location to fit the patch. For that, we search the position in all direction (top, bottom, left and right).If any patch is fit in any of the direction; the patches are placed in that direction one by one. After the creation of patches, we preserve the details of each patch and are called one patch set. The Patch set Pi has 3 parameters.

$$Pi = \{SI\ (R1, C1), PDI\ ((R1, C1), (R2, C2), ..(Ri, Ci)), IL\} \quad (7)$$

Where, SI (R1, C1) is the starting index of the first patch. The PDI is the patch data index and it represents the coordinates from starting point to the patch size-1. IL is the index label of each patch (i.e., up, down, left or right).

$$\mathrm{PDI}\ ((\mathrm{R1}, \mathrm{C1}), (\mathrm{R2}, \mathrm{C2}), \dots \dots \dots, (\mathrm{Ri}, \mathrm{Ci})) \quad (8)$$

After creating patches at the object boundary, the following steps are used to find the patches that make occlusion.

First, we assume a threshold value of distance between the patches. Then, take the starting index of all objects in the entire frames. After that, the staring index of the first patch in first object is subtracted with all the starting index of all patches in second object. If this difference is less than or equal to the predefined threshold value or greater than or equal to zero, then the direction of that patch (i.e., index label) is compared with the entire direction (Up, Down, Left and Right). At the time of matching, the color of the corresponding patches in all object in that direction is changed shown in fig.5.

3.9. Occlusion Detection

Centre of mass for all objects are calculated for occlusion detection. Using bounding box parameter in region props algorithm, we find the four extreme points of all object. From these points, select any two diagonal points and by using these values, calculate the center of mass point using the equation 9.

$$((\mathrm{C}(2) + (\mathrm{C}(2) + \mathrm{C}(4)))/2, \mathrm{C}(1) + (\mathrm{C}(1) + \mathrm{C}(3)))/2 \quad (9)$$

Where (C1, C2) is the starting co-ordinates, (C3, C4) is the co-ordinate value of column increment and row increment respectively. The Algorithm for occluded patch identification is stated below:-

1. First, set a threshold value.
2. Then, take the starting index of each object in all frames.
3. After that, the staring index of the first patch in first object is subtracted with all the starting index of the each patch in second object. If this difference is less than or equal to the predefined threshold value, and greater than or equal to zero, then the direction of that patch (i.e., index label) is compared with the entire direction(Up, Down, Left and Right).
4. At the time of matching, the color of the corresponding patches in that direction of each object will be changed. If the difference of that patches is not matched with the predefined threshold, the next patch in first object is taken and it is compared with all the patches in second object in the corresponding frame. This process is repeated and can find the patches which are going to be occluded.

IV. EXPERIMENTAL EVALUATION

The research work is implemented in MATLAB using Scene Background Initialization (SBI) dataset. Video frames ranging from 10 to 100 frames were taken as input to the proposed system. The evaluation of the proposed method is made by the calculation of the rate of good detections moving object (GDR) using the following formula.

Bad Detection Ratio = 100- GDR (10)

GDR = Number of detected moving object / Total moving object (11)

Table 1. Quantitative Evaluation using the sequences from SBI dataset

Video stream	Moving objects	Detection	GDR (%)	BDR (%)
Video 1	15	13	86.6	13.4
Video 2	3	3	100	0
Video 3	15	14	93.3	6.7
Video 4	20	19	95	5
Video 5	9	8	89	11
Video 6	4	4	100	0

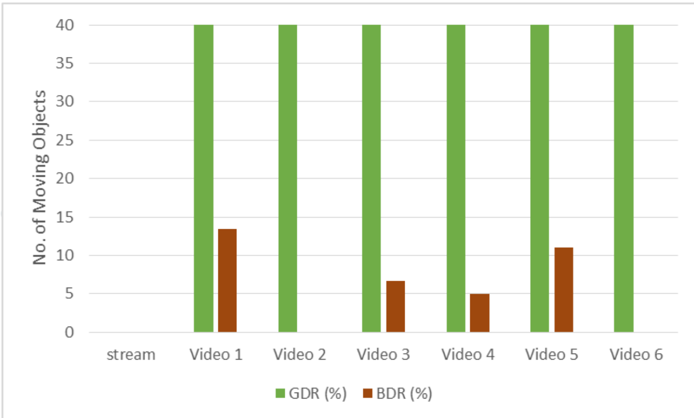


Fig. 6. Detection Rate using different video sequences

From Table 1 and Fig. 6, it is observed that the proposed method improved the detection rate of moving objects over different video streams. The new approach improves to results of the detection system in the presence of the occlusion's problem. This improved the results between 86 % up to 100 % of detection ratio (GDR).

V. CONCLUSION

Today video surveillance system is used in many sites for human detection and identification. In this paper, the Gaussian mixture model is used for motion detection. The detected objects are tracked using bounding boxes. The main innovation of this paper is occlusion detection in human parts. In the proposed work, a robust patch creation idea is used for occluded part identification. The occlusion detection by center of mass gives best result during human tracking. In the future works, the scope of present work can be extended to develop a simple model to detect and handle occlusion in infrared videos.

REFERENCES

[1]. A.S. Silva, F.M.Q. Severgnini, M.L. Oliveira, V. M. S. Mendes and Z. M. A. Peixoto and, "Object Tracking by Color and Active Contour Models Segmentation", IEEE Latin America Transactions, Volume 14, Issue No.3, March 2016, Pages 1488 - 1493.

[2]. Pojala Chiranjeevi and Somnath Sengupta, "Neighborhood Supported Model Level Fuzzy Aggregation for Moving Object Segmentation", IEEE Transactions on Image Processing, Volume 23, Issue No.2, February 2014, Pages 645 - 657.

[3]. K. Patwardhan, G. Sapiro, V. Morellas, Robust foreground detection in video using pixel layers, IEEE Trans. Pattern Anal. Mach. Intell. 30 (4) (2008) 746–751.

[4]. R.C. Jain, H.H. Nagel, On the analysis of accumulative difference pictures from image sequences of real world scenes, IEEE Trans. Pattern Anal. Mach. Intell. 1(2) (1979) 206–213.

[5]. C. Wren, A. Azarbayejani, T. Darell, A. Pentland, Pfinder: real-time tracking of human body, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 780–785.

[6]. C. Stauffer, W. Grimson, Learning patterns of activity using real time tracking, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 747–767.

[7]. A. Elgammal, R. Duraiswami, L.S. Davis, Efficient kernel density estimation using the fast gauss transform with applications to color modeling and tracking, IEEE Trans. Pattern Anal. Mach. Intell. 25 (11) (2003) 1499–1504.

[8]. Y. Sheikh, M. Shah, Bayesian modeling of dynamic scenes for object detection, IEEE Trans. Pattern Anal. Mach. Intell. 27 (11) (2005) 1778–1792.

[9]. J. Meng, S.F. Chang, CVEPS – a compressed video editing and parsing system, in: Proc. ACM Intl. Conf. Multimedia, 1997, pp. 43–53.

[10]. Y. Tan, S.R. Kulkarni, P.J. Ramadge, A new method for camera motion parameter estimation, Proc. IEEE Intl. Conf. Image Process. 1 (1995) 23–26.
[11]. Viola, P., Jones, J.: Robust real-time face detection. Comput. Vision Pattern Recogn. 57(2), 137–154 (2001)
[12]. T. Aach and A. Kaup. Bayesian algorithms for adaptive change detection in image sequences using markov random fields. *Signal Processing: Image Communication*, 7:147–160, 1995.
[13]. Y. Benezeth, B. Emile, and C. Rosenberger. Comparative study on foreground detection algorithms for human detection. *International Conference on Image and Graphics*, pages 661–666, 2007.
[14]. A.C. Bovik. *Handbook of Image and Video Processing*. Academic Press, Inc. Orlando, FL, USA, 2005.
[15]. Mignotte, M., Konrad. I.: Statistical background subtraction using spatial cues. Circuits Syst. Video Technol. IEEE Trans. 17(12), 1758–1763 (2007).
[16]. Neji, M., Wali, A., Alimi, A.M: Towards an intelligent information research system based on the human behavior: recognition of user emotional state. In: 12th IEEE/ACIS International Conference on Computer and Information Science, Japan (2013).
[17]. Ralph. O.M, Kong, S.G.: Senior member. In: Visual Analysis of Eye State and Head Pose for Driver Alertnes Monitoring, vol. 14, pp. 1462–1469. IEEE, USA (2013).
[18]. Wierwille, W.W, Ellsworth, L.A., Wreggit. S.S, Fairbanks. R.J and Kirn. C.L: Research on vehicle based driver status/performance monitoring: development, validation and refinement of algorithms for detection of driver drowsiness. National Highway Traffic Safety Administration, Technical report. DOT HS 808 247 (1994).
[19]. Marques, J.S., Jorge, P.M., Abrantes, A.J., Lemos, J.M.: Tracking groups of pedestrians in video sequences. In: IEEE International Conference on Pattern Recognition, vol. 9 (2003).
[20]. Chang, T.-H., Gong, S., Ong, E.-J.: Tracking multiple people under occlusion using multiple cameras. In: Proc. 11th British Machine Vision Conference (2000).