

## A Block-Based Multi-Scale Background Extraction Algorithm

Seyed Hashem Davarpanah, Fatimah Khalid and Maryam Golchin  
Faculty of Computer Science and Information Technology,  
University Putra Malaysia, Serdang, Malaysia

---

**Abstract: Problem statement:** To extract the moving objects, vision-based surveillance systems subtract the current image from a predefined background image. The efficiency of these systems mainly depends on accuracy of the extracted background image. It should be able to adapt to the changes continuously. In addition, especially in real-time applications the time complexity of this adaptation is a critical matter. **Approach:** In this study, to extract an adaptive background, a combination of blocking and multi-scale methods is presented. Because of being less sensitive to local movements, block-based techniques are proper to control the non-stationary objects' movements, especially in outdoor applications. They can be useful to reduce the effect of these objects on the extracted background. We also used the blocking method to intelligently select the regions which the temporal filtering has to be applied on. In addition, an amended multi-scale algorithm is introduced. This algorithm is a hybrid algorithm, a combination of some nonparametric and parametric filters. It uses a nonparametric filter in the spatial domain to initiate two primary backgrounds. In continue two adapted two-dimensional filters will be used to extract the final background. **Results:** The qualitative and quantitative results of our experiments certify not only the quality of the final extracted background is acceptable, but also its time consumption is approximately half in compare to the similar methods. **Conclusion:** Using Multi scaling filtering and applying the filters just to some selected non-overlapped blocks reduce the time consumption of the extracting background algorithm.

**Key words:** Background extraction, background modeling, multistate modeling, block-based modeling, foreground

---

### INTRODUCTION

In Many of Intelligent Transportation Systems (ITSs) based image processing applications, the first step is accomplished by detecting changes in an image sequence. To identify the changes, a portion of the used methods employ algorithms based on inter-frame difference. But, this method always detects larger areas than the real moving objects. So, it cannot detect the immobile objects. In addition, it is liable to noise impact and it has low detection precision and low reliability (Xiaofei, 2009). As the second group, some researchers use algorithms which are based on the subtraction of an input frame from a reference one. The later one represents the background of the scene (Varadarajan *et al.*, 2009). Efficiency of this group's result mainly depends on accuracy of the extracted background image. In addition, it is impressed by the adaptability of the background image extracted in various situations. This point leads researchers to try to find a better approach to model, initialize, adapt and

subtract a more accurate background especially in real-time and outdoor environments.

The background objects are a combination of stationary objects such as traffic lights, or non-stationary objects such as wavering bushes. On the other hand, the background image pixels can be categorized to the static and dynamic pixels. The static pixels present the stationary objects and the dynamic ones belong to non-stationary objects (Li *et al.*, 2004). A proper background extraction algorithm should be able to handle both static and dynamic pixels.

The extracted features from an image sequence can be described as a combination of three characteristic types: spectral (Toure and Beiji, 2010), spatial and temporal features. Spectral features relate to gray-scale or color information, spatial features associate with gradient or local structure and temporal features present inter-frame changes at the pixel. These feature types or a combination of them is used in different methods to model the images.

Background extraction algorithms typically use techniques like image inpainting on a single image or

---

**Corresponding Author:** Seyed Hashem Davarpanah, Faculty of Computer Science and Information Technology,  
University Putra Malaysia, Serdang, Malaysia Tel: (+60-3)89471730 Fax: (+60-3)89466577

object-tracking techniques on a sequence of images 0. Image inpainting techniques can result in a degraded visual Quality since they try to form an estimate. But, object-tracking techniques including segmentation-based approaches and motion-based approaches do simultaneously foreground tracking and background Updating. These operations are usually expensive.

To compel the complexity problem, researches try to model frames using simpler ways. Some approaches treat a sequence as a set of single-pixel values changing over time. The statistics of pixel values are gathered separately and then are used to classify each pixel independently. This method suffers from noise of misclassified single pixels (Armanfard *et al.*, 2009a). On the other hand, region based models have been emerging recently. Since complexity, they exploit spatial information in a more meaningful way and are able to achieve superior segmentation than the methods relying solely on morphological post processing do. In addition they are less prone to segmentation errors caused by small movements of the camera. Finally, as the third group, in block-based approaches, an image is divided into overlapped or non-overlapped blocks. They are less sensitive to local movements.

In this study, we introduce a hybrid background extraction algorithm, a combination of an amended algorithm in the spatial domain, a nonparametric filter and a parametric filter method. This algorithm tries to model the background in the form of some blocks extracted from the traffic scene. In order to do it, multi-scaling technique and some different filters are utilized.

This research is based on the proposed algorithm by (Culibrk *et al.*, 2009). We enhance the time complexity problem their method faces with. Blocking the extracted temporal background, managing the blocks which spatial filters are applied to them and reducing the number of filters applied in this stage reduce the time needed to approximately half.

**Related works:** The researches having been done to model and adapt the background image can be categorized to two main groups; nonparametric and parametric algorithms. It is clear that there are some exceptions too. In continue some more significant samples of these two groups and their properties are highlighted.

Nonparametric algorithms category, the first main group of background extraction methods, is called in different resources as “Filter-based Methods”, “Low-pass filtering systems” or “Samples-based techniques”. These algorithms estimate, for each pixel or a group of them, the grey level intensity or the color value of the background scene from a sequence of the input images

without taking any predetermined form or assumption. This estimation is just constructed according to information derived from the data.

Nonparametric algorithms are non-expensive; since algorithms categorized in this group use some statistic filters which are applied to all sections of the images in the same way and with the same value of parameters, in compare to the second group’s algorithms, they consume less time and memory.

Researchers believe that this kind of algorithms is unable to cope with high frequency motion in the background 0. Especially in outdoor applications such as ITSs, this probably is so high that the scene or majority part of it changes completely; so, it is clear that researchers cannot rely on imprecise algorithms’ results.

As a simple nonparametric model, in order to estimate a new background, researchers construct a combination of the current background and one, two or more previous extracted backgrounds. The way frames combine together, the number and the composition of these frames and also the combination rate are investigated several times. The goal of these researches is improving the adaptability property of the final extracted background. Namely, Fuzzy logic and Cellular Automata were used by (Shakeri *et al.*, 2008) to decide how to combine frames together better. Xiaofei (2009) introduced background extraction system with self-adaptive update algorithm and putted forward an improved algorithm based on histogram statistic combining with multi-frame average. He claimed that his approach is so fast that it is suitable for DSP platforms. Barnich and van Droogenbroeck (2009) presented VIBE. He tried to find a way to combine frames randomly. Also, genetic algorithm was used by (Davarpanah and Fathy, 2004) to adapt the combination rate automatically.

The second main group to model and extract an adaptive background is parametric methods group. Parametric methods assume data are driven from a kind of probability distribution. They try to make inferences about the parameters of this distribution. It can be assumed that parametric methods make more assumptions than non-parametric ones. If these extra assumptions are correct, they can produce more accurate and precise estimates.

Drawback of parametric algorithms is their complexity. Since the values of model’s parameters should be updated continuously, execution of these algorithms consumes more time and memory. Researches try to reduce it by using different filters, various estimators and setting their parameters intelligently.

Stauffer and Grimson (1999) supposed a specific pixel may have values as a Gaussians distribution. In the Gaussian distribution model, the background of a scene each pixel can be modeled with a Gaussian distribution between 2 and 2.5 standard deviations (Peng and Peng, 2009). Later, Mixture of Gaussians (MoG) was introduced. According to MoG image is modeled in the form of combination of k separate Gaussian models (usually k is between 3 and 5). Researchers have being tried to set the parameters (mean, variance, weight and learning rate) for each Gaussian. They have represented various combinations of different number of Gaussian models to adapt to gradual background changes (Sheng and Cui, 2008).

Karmann *et al.* (1990) used Kalman filter to estimate the background. Their algorithm assumes the evolution of the background pixel intensity can be described by a finite-dimension dynamic system. Using Kalman filter to estimate and update the background can reduce the number operations required. But it cannot be suitable to handle the sudden background changes, sometimes it may result wrong detection or miss detection (Xiaofei, 2009).

First time, the use of Marr wavelet filter in background adapting was explained by Davies *et al.* (1998). In their proposed method, the filter was applied to three consecutive frames of the sequence, rather than to background reference image and the current frame, as is the common practice, the temporal filter used.

Elhabian *et al.* (2008) used Hidden Markov Model to model the pixel process. In their proposed Hidden Markov model various states of the model represent different states that might occur in the pixel process, such as background, foreground, shadows, day and night illumination. They also have claimed that the model can also be used to handle the sudden changes in illumination. To do so, they present the change from a status to another, such as the change from dark to light, day to night, indoor to outdoor, as the transition from state to state in the HMM.

Beside of two main groups for modeling the background, there are also some another endeavors which could not be categorized in these two groups. For example, several researches use key-points to model the background. Hamdoun and Moutarde (2009), instead of full video streams, just have used some interest point descriptors from some smart cameras. The main problem of using interest points is their instability at smaller scales, because they are more susceptible to changes in lighting and camera noise. Also, Zhong *et al.* (2009), in a same idea, has divided the image into some

patches and has represented each image patch as a Neighboring Image Patches Embedding (NIPE) vector. Local Binary Pattern, LBP, is used vastly for texture description. It has acceptable performance in texture classification, fabric defect detection and moving region detection. In this method, local features which consider the pixel's neighboring pixels are assigned as its texture features (Armanfard *et al.*, 2009b). LBP was used by (Heikkila and Pietaininen, 2006) to model the history of each pixel in any block. Zhong *et al.* (2009) has tried to reduce the size of the model.

Also, (Li *et al.*, 2004) a Bayesian framework is proposed. It incorporates spectral, spatial and temporal features to characterize the background appearance. In this method, to classify the background and foreground based on the statistics of principal features, a Bays decision rule is derived.

Finally, as another method to model the background, the dynamic features of non-stationary background objects are represented by the significant variation of accumulated local optical flows. The drawback of using optical flow is its complex formula and large calculation; for complex and rapid moving objects it causes poor results (Cui *et al.*, 2009).

**Proposed algorithm:** The proposed method is an improved version of an existent algorithm (Culibrk *et al.*, 2009). It is a hybrid algorithm. First, the proposed algorithm uses a nonparametric filter to initiate two primary backgrounds with different combination rates. In continue, we will combine these two temporary backgrounds to reach a new temporary one. In the next step, we will block the extracted background. Finally, by applying two 2-dimensional filters in spatial domain to some selected blocks, the background will be finalized. Details of the proposed algorithm are described more as follows. Also, its flowchart is illustrated in Fig. Fig. 1.

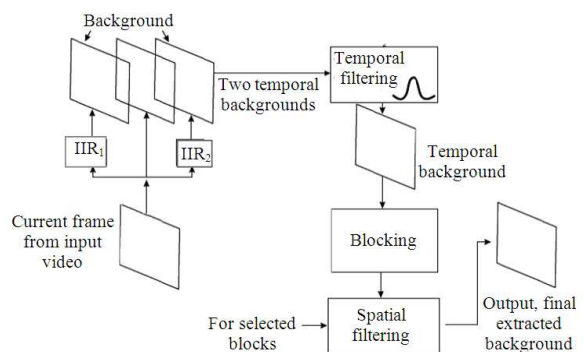


Fig. 1: Proposed algorithm

**Background model:** In this step two background images by using a nonparametric filter, actually an Infinite Impulse Response filter (IIR) and with two different values of combination rate parameter are calculated. This step passes two temporal background images to the next phase. IIR digital filters are commonly realized recursively by feeding back a weighted sum of present (current frame) and past (previous background) input values and adding these values to a weighted sum to construct the new background.

$$BK_{new} = (1 - \alpha) * BK_{old} + \alpha * Im g_{cur} \quad (1)$$

As shown in (1), the pace of adaptability of the calculated background depends on the value of  $\alpha$ . To provide better representation of background image and make the approach more resilient to effects such as those due to camera automatic gain adjustment, we executed this process two times and with two different values of learning rate ( $\alpha$ ). Finally, we had two primary backgrounds. To initialize the process, two learning rate parameters were assigned to 0.1 and 0.05, respectively. In addition, the backgrounds were initialized with the first frame of the data.

**Temporal filtering:** By using a one dimensional filter such as One-dimensional Mexican Filter and applying it to two primary backgrounds extracted in previous phase, we achieve a temporary background.

$$f(x) = \frac{2}{\sqrt{3}} \pi^{-\frac{1}{4}} (1 - (\zeta x)^2) \exp\left(-\frac{(\zeta x)^2}{2}\right) \quad (2)$$

First, according to (2) the requested filter is calculated. In that,  $x$  is Euclidean distance of the point from the center of filter and  $\zeta$  is scaling coefficient enabling the filter to be evaluated at sub-pixel ( $\zeta > 1$ ) level or stretched to cover a larger area ( $\zeta < 1$ ). In continue, the filter is applied to the sequence of two primary backgrounds extracted in the last stage and current frame between them.

Once the filter is applied, a Z-score test is used to detect the outliers in the frame. Z-score for each pixel is equal to division of distance between value of that pixel in the filtered image and mean value of pixels in the filtered image on Mean Absolute Distance (MAD). MAD is defined as follows:

$$MAD = \frac{\sum_{i=1}^N |fp_i - \mu|}{N} \quad (3)$$

Where:

- N = The number of pixels
- $\mu$  = The mean value of filtered image's pixels
- $fp_i$  = The value of  $i^{th}$  pixel of filtered image

In continue, the Z-score value of each pixel is compared to a user predefined threshold ( $\theta$ ). If it is less than  $\theta$ , it is discarded; the value of this pixel in the temporal background as the output of this stage will be zero.

**Blocking:** To compel the time complexity problem of the spatial filtering and reduce it, we think blocking is a suitable solution. In this step the temporal background is divided to B\*B non-overlapped blocks. For each block, number of moving pixels, the pixels with higher value than zero, is calculated. This amount is compared with a threshold value ( $\psi$ ). Blocks which have more values, will be adjusted with the special filters and the others will be just copied to the final background. Especially in the scenes traffic is not crowded, it is expected most of the blocks do not need to participate in the spatial filtering phase. Since vast majority of the time the algorithm needs relates to the applying spatial filters, this technique declines the time to half.

**Spatial filtering:** In addition, to smooth out the segmentation result, two dimensional filters are applied to each selected block in the last phase. Similar to the algorithm of Culibrk *et al.* (2009), two dimensional Negative Mexican Filter is used. While they apply the filter four times, our results show after applying twice of the filter over the square neighborhood 14 and 7 pixels wide respectively, the expected result are achieved.

After each time, the mean absolute difference of the filtered values is calculated, normalized and produced as the output of filtering phase. In the end of the second time, the values are again check with a threshold in the same manner as described in Temporal Filtering part of the article.

**Background extraction:** At the end, the final background is resulted from the combination of the current frame and the adapted background, the result of Background model phase with lower combination rate. This combination is done according to the values of the binary image pixels. Binary image is the output of the previous stage explaining the moving pixels. If the value of the binary pixel is one, the final background is assigned with the value of correspond pixel in the adapted background. Else, it will be replaced with corresponding pixel of current frame.

**MATERIALS AND METHODS**

In order to evaluate the performance of the proposed algorithm and compare its efficiency against the resource one, we implemented and tested both of them in the same conditions. They were implemented with MATLAB. They were run on a Pentium 3 GHz PC. The algorithms initialized with the values of Table 1. In this table, elements having more than one value are used more than one times in the algorithm. Also,  $\theta$  has two value sets. The first set is belonged to the reference algorithm and the second one is used in our algorithm. Their values are listed regarding to their use order in the algorithm. To have a fair experiment, we used the standard database ordered by PETS (IEEE International Workshop of Performance Evaluation of Tracking and Surveillance). Exactly, the PET2000 was used. This dataset can be found in <http://ftp.pets.rdg.ac.uk/>.

**RESULTS AND DISCUSSION**

In this research, the experiment was repeated several times with various frames and different number of them. Also, the experimental results were investigated quantitatively and qualitatively.

Sub-image (a) of Fig. 2 shows 100th frame of used sequential frames by the proposed algorithm. Its corresponding final background image extracted by the algorithm is shown in sub-image (b).

Table 1: List of parameters and their initial values

Parameter	Value	Parameter	Value	Parameter	Value
$\gamma$	10	$\alpha$	0.1, 0.05	$\psi$	30
$\theta$	5, 3.5, 2, 1.5	$\xi$	1, 0.5	B	23
M	288	N	384		

Table 2: The mean of quantitative factors in 100th frame

	Time (m sec)	RMS error (pixel)	FDR (pixel)
Proposed algorithm (Culibrk <i>et al.</i> , 2009)	83	2063	653
	142	2698	764



Fig. 2: Proposed algorithm. (a) 100th frame; (b) Final extracted background

These values for the reference algorithm are displayed in Fig. 3. The quality of subfigures shows that the performance of the proposed algorithm is acceptable.

To evaluate efficiency of the algorithm and compare it with the reference one quantitatively, three different scales were investigated:

- Just to compare the time complexity of the algorithms, their time consumptions were calculated. The averages of the time cost of two algorithms are shown in first column of Table 2. The expanded values are plotted in sub-image (a) of Fig. 4
- Root Mean Square error (RMS error) is another scale used. This value is calculated as follows:

$$\theta = \frac{1}{M * N} \sum_{x=1}^M \sum_{y=1}^N BG(x,y)^2 - STBG(x,y)^2 \quad (4)$$

Where, M and N are the dimensions of the image. And, STBG is the pure background frame. In this research, adapted background resulted of section A with lower combination rate was used as STBG.

This factor's averages in both algorithms are shown in second column of Table 2. Plot designed in sub-image (b) of Fig. 4 compares their amounts through different frames

- False detection rate (Al-Khateeb *et al.*, 2008) is the third quantitative factor used in this research. Actually, it is equal to the number of pixels which the difference between their values and the value of corresponding pixel in the adapted background calculated with the less combination rate is more than a user predefined threshold ( $\gamma$ ). The mean of acquired values in two algorithms are mentioned in the last column of Table 2. And their details are highlighted in last part of Fig. 4



Fig. 3: The Extracted background by reference 0 (a) 100th frame; (b) Final extracted background

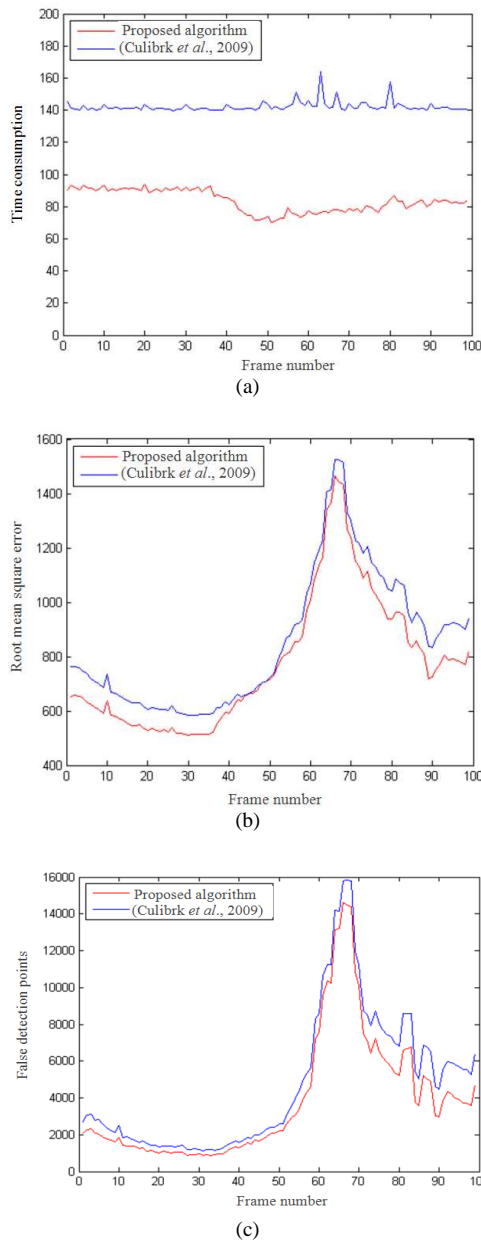


Fig. 4: (a) Time Consumption through the first 100 frames, (b) Root mean square error through the first 100 frames, (c) False detection rate through the first 100 frames

### CONCLUSION

Since the time complexity is a critical matter in the real time applications, our team concentrates on this matter. In this research, we presented an amended algorithm to extract the background which its time

consumption is approximately half in compare to similar endeavors. We first divided the image to some non-overlapping blocks. In continue we managed the number of blocks the algorithm should be applied to. The performance of the purposed algorithm was evaluated qualitatively and quantitatively. They were compared with the primary version of the algorithm's efficiency.

It seems it is not necessary to adapt the background in each frame. It causes the algorithm runs faster. As our future work, we are going to test it.

### REFERENCES

- Al-Khateeb, K.A.S., J.A.Y. Johari and W.F. Al-Khateeb, 2008. Dynamic traffic light sequence algorithm using RFID. *J. Comput. Sci.*, 4: 517-524. <http://www.scipub.org/fulltext/jcs/jcs47517-524.pdf>
- Armanfard, N., M. Komeili, M. Valizade, E. Kabir and S. Jalili, 2009a. A non-parametric pixel-based background modeling for dynamic scenes. *Proceedings of ACTEA*, July 15-17, IEEE Xplore Press, Zouk Mosbeh, pp: 369-373. DOI: 10.1109/ACTEA.2009.5227894
- Armanfard, N., M. Komeili, M. Valizadeh and E. Kabir, 2009b. Effective hierarchical background modeling and foreground detection in surveillance systems. *Proceedings of the 14th International CSI Computer Conference*, Oct. 20-21, IEEE Xplore Press, Tehran, pp: 164-169. DOI: 10.1109/CSICC.2009.5349318
- Barnich, O. and M. van Droogenbroeck, 2009. ViBE: A powerful random technique to estimate the background in video sequances. *Acoustics, Speech and Signal Processing*, Apr. 19-24, IEEE Xplore Press, Taipei, pp: 945-948. DOI: 10.1109/ICASSP.2009.4959741
- Culibrk, D., V. Crnojevic and B. Antic, 2009. Multiscale background modeling and segmentation. *Proceedings of the 16th International Conference on Digital Signal Processing*, July 5-7, IEEE Xplore Press, Santorini, Greece, pp: 1-6. DOI: 10.1109/ICDSP.2009.5201193
- Cui, B.X., S.M. Sun and Y. Duan, 2009. Vehicle detection based on adaptive background. *Proceedings of Second International Workshop on Knowledge Discovery and Data Mining*, Jan. 23-25, IEEE Xplore Press, Moscow, pp: 821-824. DOI: 10.1109/WKDD.2009.117
- Davarpanah, S.H. and M. Fathy, 2004. Using genetic algorithm in adaptive background extraction. *Proceedings of Intelligent Knowledge Systems*, Aug. 16-20, ASSOS, Troy, Turkey, pp: 1-1.

- Davies, D., P. Palmer and M. Mirmehdi, 1998. Detection and tracking of very small low contrast objects. Proceedings of the 9th British Machine Vision Conference, (BMVC'98), IEEE Xplore Press, Southampton, United Kingdom, pp: 599-608.
- Elhabian, S.Y., K.M. El-Sayed and S.H. Ahmed, 2008. Moving object detection in spatial domain using background removal techniques-state-of-art. Recent Patents Comput. Sci., 1: 32-54. <http://www.bentham.org/cseng/samples/cseng1-1/Elhabian.pdf>
- Hamdoun, O. and F. Moutarde, 2009. PhD forum: Keypoints-based background model and foreground pedestrians extraction for future smart cameras. Proceedings of the 3rd ACM/IEEE International Conference on Distributed Smart Cameras, Aug. 30-Sept. 2, IEEE Xplore Press, USA., pp: 1-2. DOI: 10.1109/ICDSC.2009.5289390
- Heikkila, M. and M. Pietaininen, 2006. A texture-based method for modeling the background and detecting moving objects. IEEE Trans. Patt. Anal. Mach. Intell., 28: 657-662. DOI: 10.1109/TPAMI.2006.68
- Karmann, K., A. Brandt and R. Gerl, 1990. Moving object segmentation based on adaptive reference images. Signal Process., 5: 951-954.
- Li, L., W. Huang, I.Y.H. Gu and Q.T. Yunlong, 2004. Statistical modeling of complex backgrounds for foreground object detection. IEEE Trans. Image Process., 13: 1459-1472. DOI: 10.1109/TIP.2004.836169
- Peng, J.W. and J.W. Peng, 2009. Real-time dynamic background segmentation based on a statistical approach. Proceedings of the IEEE International Conference on Networking, Sensing and Control, Mar. 26-29, IEEE Xplore Press, Okayama, Japan, pp: 404-409. DOI: 10.1109/ICNSC.2009.4919310
- Shakeri, M., H. Deldari, H. Foroughi, A. Saberi and A. Naseri, 2008. A novel fuzzy background subtraction method based on cellular automata. Proceeding of the 9th International Conference of Signal Processing, Oct. 26-29, IEEE Xplore Press, Beijing, pp: 899-902. DOI: 10.1109/ICOSP.2008.4697273
- Sheng, Z. and X. Cui, 2008. An adaptive learning rate GMM for background extraction. Proceedings of the International Conference on Computer Science and Software Engineering, Dec. 12-14, IEEE Xplore Press, Wuhan, Hubei, pp: 174-176.
- Stauffer, C. and W.E.L. Grimson, 1999. Adaptive background mixture models for real-time tracking. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 23-25, IEEE Xplore Press, Fort Collins, CO., USA., pp: 246-252. DOI: 10.1109/CVPR.1999.784637
- Toure, M.L. and Z. Beiji, 2010. Intelligent sensor for image control point of eigenfaces for face recognition. J. Comput. Sci., 6: 484-491. <http://www.scipub.org/fulltext/jcs/jcs65484-491.pdf>
- Varadarajan, S., L.J. Karam and D. Florencio, 2009. Background recovery from video sequences using motion parameters. Proceedings of the Acoustics, Speech and Signal Processing, Apr. 19-24, IEEE Xplore Press, Taipei, Taiwan, pp: 989-992.
- Xiaofei, L., 2009. A new fast hybrid algorithm of background extraction suitable for the DSP platform. Proceedings of the World Congress on Computer Science and Information Engineering, Mar. 31-Apr. 2, IEEE Xplore Press, Los Angeles, CA., pp: 1-4. DOI: 10.1109/CSIE.2009.406
- Zhong, B., H. Yao and S. Liu, 2009. Neighboring image patches embedding for background modeling. Proceedings of the Image Processing (ICIP), 16th IEEE International Conference on, Nov. 7-10, IEEE Xplore Press, Cairo, pp: 3209-3212. DOI: 10.1109/ICIP.2009.5414382