

Using Multi-Scale Filtering to Initialize a Background Extraction Model

Davarpanah, S.H., Fatimah Khalid,
N.A. Lili, S.S. Puteri and M. Golchin
Department of Multimedia,
Faculty of Computer Science and IT,
Universiti Putra Malaysia, Serdang, 43300, Malaysia

Abstract: Problem statement: Probability-based methods which usually work based on the saved history of each pixel are utilized severally in extracting a background image for moving detection systems. Probability-based methods suffer from a lack of information when the system first begins to work. The model should be initialized using an alternative accurate method. **Approach:** The use of a nonparametric filtering to calculate the most probable value for each pixel in the initialization phase can be useful. In this study a complete system to extract an adaptable gray scale background image is presented. It is a probability-based system and especially suitable for outdoor applications. The proposed method is initialized using a multi-scale filtering method. **Results:** The results of the experiments certify that not only the quality of the final extracted background is about 10% more accurate in comparison to four recent re-implemented methods, but also the time consumption of the extraction are acceptable. **Conclusion:** Using multi-scale filtering to initialize the background model and to extract the background using a probability-based method proposes an accurate and adaptable background extraction method which is able to handle sudden and large illumination changes.

Key words: Adaptive background extraction, background modelling, probability-based method, multi-scale filtering, non-parametric method, moving object detection, outdoor applications, history-based method

INTRODUCTION

In many visions-based surveillance systems such as Intelligent Transportation Systems, the first step is the detection of the changes in an image sequence. To identify changes, the first group uses methods which are based on inter-frame differences. These methods always detect larger moving object areas than the real. Thus they are not able to detect immobile objects accurately. In addition, they are sensitive to the existent noises. Thus, they have a low detection rate and low reliability (Xiaofei, 2009). In the second group of background extraction methods, the subtraction of the current captured frame from the background image as a reference image is used to detect the moving objects. The efficiency of this group mainly depends on the similarity of the extracted background image with the real background image. In addition, this approach is affected by the adaptability of the background image extracted in various situations. This problem prompted researchers to find a more performance-based 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 like traffic lights and non-stationary objects such as waving bushes. On the other hand, the background image pixels are categorized into static and dynamic pixels. The static pixels present the stationary objects and the dynamic pixels belong to the non-stationary objects (Li *et al.*, 2004). A proper background extraction algorithm should be able to handle both static and dynamic pixels.

The gray scale of a stationary pixel of each frame has a constant value; this constant is the intensity of that pixel. Gray values of the pixels in a motionless background image are not always the same. It is because of disturbances in lighting and intensity, cameras, atmosphere and the existence of moving objects for a short or a long period. So, the constant value allocated to each point may change in one or a few time slices. However, it seems that in a certain duration time, the value of each pixel in the background image of a video sequence is likely to be equal to the

Corresponding Author: Seyed Hashem Davarpanah, Department of Multimedia, Faculty of Computer Science and IT, Universiti Putra Malaysia, Serdang, Malaysia Tel: (+60-3)89471730 Fax: (+60-3)89466577

pixel with highest appearance frequency in corresponding scenes. History-based methods are based on this fact. They save the history of each pixel for a specific period of time and in the form of values allocated to one or a range of different gray scale values.

All the probability-based background extraction methods suffer from the same deficiency. If there are moving objects in the frames which participate in the initialization phase and they leave the scene after that, the probability based algorithms are not able to detect these movements. Therefore, they will appear as static objects in the extracted background for a long time following the initialization. This problem occurs since these methods simply rely on historical values. In our previous publication (Davaranah *et al.*, 2010), we introduced a method to update the history more accurately. It helped us to overcome the problem explained above.

MATERIALS AND METHODS

In this study, we introduce a complete background extraction system; a combination of a simple method to model the background, a new algorithm to initialize the background, a fast algorithm to update the system and to adapt the background, a fast technique to extract the binary objects, an enhanced median filter to reduce the amount of existent noise and an effective algorithm to construct the final background.

The difference of the proposed system in this article compared to our previous publication, (Davaranah *et al.*, 2010) is that in this study a new initialization method is introduced. This is done to manage the problem of the lack of information that probability-based methods inherently face within the first few frames.

The remaining part of this study is organized as follows: In the first section, a review of some more significant related works is presented. The second section gives an overview of the proposed system. The following section part provides the experimental results and the study is concluded in the last section.

Related works: Several studies have been undertaken over the last three decades to model, extract and adapt the background image for the last three decades. Speed, adaptability to periodic changes such as changes in sunlight, management of fast changes caused by moving objects entering the scene or leaving the scene are the most critical problems on these studies.

Previous methods have used nonparametric and parametric methods to model the changing behaviour of the scene, moving objects, background or some

sections within them. In the following, some of the significant samples of these studies and their properties are highlighted.

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 technique of combining the frames; the number of combined frames, the composition of these frames and the combination rates have been investigated several times.

The common goal of these previous studies is improving the adaptability of the final extracted background. Namely, Fuzzy logic and Cellular Automata are used to decide how to combine frames together more efficiency by (Shakeri *et al.*, 2008). Xiaofei (2009) introduced a background extraction system that is updated using a self-adaptive algorithm. He also presented an extended algorithm which is based on a combination of histogram statistics and multi-frame averages. The author has claimed that this approach is faster and suitable for platforms based on Digital Image Processing (DSP). Also, Barnich and Droogenbroeck (2009) presented the VIBE program to combine frames randomly.

Stauffer and Grimson (1999) supposed that a specific pixel may have values following a Gaussian distribution. In the Gaussian distribution model, the background is modeled as a set of pixels with a Gaussian distribution of 2 and 2.5 standard deviations for each pixel (Peng and Horng, 2009). Subsequently, a Mixture of Gaussians (MoG) was introduced. According to MoG, an image is modelled in the form of a combination of k separate Gaussian models (usually k is between 3 and 5). The parameters of the Gaussian models (mean, variance, weight and learning rate) are set. They have presented various combinations of a different number of Gaussian models which are used to adapt to the gradual background changes (Sheng and Cui, 2008).

Also, Local Binary Pattern (LBP) is used vastly as a texture descriptor. It has acceptable performance in fabric defect detection, texture classification, face recognition and moving region detection. In this method, local features which are considered as neighboring pixels are assigned to the texture features (Armanfard *et al.*, 2009a). LBP has been used by Heikkila and Pietakinen (2006) to model the history of each pixel in any block. Armanfard *et al.* (2009b) tried to reduce the size of the model.

In our own previous study, (Davaranah *et al.*, 2010), a combination of blocking and multi-scale methods were used. Block-based techniques are appropriate to control the movements of non-stationary objects because they are less sensitive to local movements, especially in outdoor applications. They can be useful to reduce the effect of

these objects on the extracted background. We also used a blocking method to select the regions in which temporal filtering has to be applied. This method is appropriate to extract moving objects when motion is not predictable. However its results are not stable. They are sensitive to any sudden motion.

Proposed system: The proposed system is a combination of the following subsystems:

- A probability-based method to model the background image
- An adaptive algorithm to initialize the background model
- An accurate method to extract moving objects based on the adaptive model resulted from the previous phases
- A post processing technique to reduce the noises existent in the binary image showing moving objects
- A simple method to extract the final background

To continue, each of these methods is described in detail below.

Before making a decision concerning the model of the system, it is necessary to know how to transform the visual data for input into the system. Gray levels are usually used by different researchers and it is a key factor to judge the extracted background and moving-objects. Furthermore, to reduce the complexity of the algorithm, we process a color image and convert it to gray levels using the same method as before in (Davarpanah *et al.*, 2010).

To initialize the system and to calculate a primary background, a new algorithm is presented. The proposed algorithm uses a distance classifier to classify the input image for each pixel until the F_n th frame. Then, the probability of the gray level of each pixel is calculated and a convergent value is used to extract the gray level of the background pixel whose probability is the maximum and greater than the convergent value. To update the history of each pixel and the cluster that its current value belongs to, a multi-scale nonparametric filtering technique is used. The background initialization algorithm is explained as in following. The pseudo code is also illustrated in Fig. 1.

- Step1: Capture the first input frame and initialize the model as explained before.
- Step2: While there is any pixel in the background (*Init_BG* in the pseudo code) not yet having a value, repeat steps 3-11.
- Step3: Capture the next input frame. (*Counter* is the number of grabbed frames).

```

BK_Initialization()
{
  Capture the first input frame and initialize the
  model(Calculate Mexican filter)
  While there is a pixel not assigned yet
  {
    Capture the current Image(Cur_Img)
    Calculate two primary backgrounds
    (Primary_BK1 and Primary_BK2)
    Apply Mexican filter to Primary_BK1 and
    Primary_BK2
    Calculate Z-score values and construct the
    Binary_Image
    For each pixel(i, j) from the current frame
    {
      If pixel(i, j) do not have any value
      {
        If(Binary_Image(i, j)=0)
          BK(i, j)=Cur_Img(i, j)
        Else
          BK(i, j)= Primary_BK2(i, j)
          Clu_no=Cur_Img(i, j)/Clu_size
          C(i, j, Clu_no)+=1
          M(i, j, Clu_no)= Weighted average of
          M(i, j, Clu_no) and BK(i, j)
          If( Counter >= F_n )
          {
            Pmax = Maximum probability value
            between all  $\mu/Counter$  values in
            pixel(i, j)
            If( Pmax >  $\theta$ ) Init_BK(i, j)=  $\mu_{(i,j),Pmax}$ 
          }
        }
      }
      Counter++
      If(Counter >= F_n)
         $\theta = 0.7 * \text{pow}(0.95, \text{Counter}-F_n)$ 
    }
  }
  Return(Init_BK)
}

```

Fig. 1: Pseudo code of the initialization algorithm

Step4: In this step, by using a nonparametric filter, actually an Infinite Impulse Response filter (IIR) and with two different values for the combination rate, two primary background images are calculated.

To realize the IIR filters, a weighted sum of the present (current frame) and past (previous primary background) input values are feeding back recursively and these values are added to a weighted sum to construct the new background image. This process is done using Eq. 1:

$$\text{Primary_BG}_{\text{new}} = (1-\alpha) \times \text{Primary_BG}_{\text{old}} + \alpha \times \text{Cur_img} \quad (1)$$

As shown in Eq. 1, the pace of adaptability of the calculated primary background depends on the value of α . we executed this process twice and with two different values for the learning rate (α). It is executed twice to achieve a more adaptive presentation and a more

resilient approach in front of effects such as those caused by camera automatic gain adjustment. Finally, we obtain two primary backgrounds. To initialize the process, two learning rate parameters are assigned as 0.1 and 0.05, respectively. In addition, the backgrounds are initialized with the first frame of the data.

Step5: A One-dimensional Mexican Filter is applied to two primary backgrounds extracted in the previous section (Eq. 2):

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

Where, x is the Euclidean distance between the pixel and the center of the filter. And F is a scaling coefficient which is used to enable the filter to be evaluated at the sub-pixel ($F > 1$) level or stretched to make able to cover a larger area ($F < 1$).

After applying the filter, a Z-score test is run. The Z - score is utilized to detect the outlines in the frame. The Z-score value for each pixel is calculated by dividing the distance between values of that pixel in the filtered image and a mean value of pixels in the filtered image over the Mean Absolute Distance (MAD) (Culibrk *et al.*, 2009). MAD is defined as Eq. 3:

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

where, N refers to the number of pixels, the mean value of pixels belonging to the filtered image is μ and fp_i is the value of i^{th} pixel of that image.

In the next step, the Z-score value of each pixel is compared to a user predefined threshold (θ). If it is less than θ , it is discarded; the value of this pixel in the binary image as the output of this stage will be zero.

Step 6: For each pixel of the captured frame, if it has not been assigned before, then do steps 7-10.

Step 7: The final background (in the pseudo code) is the result of the combination of the current frame and the adapted primary background, the result of step 4 with the upper combination rate. This combination is done according to the values of the binary image pixels:

$$\text{If (Bin_img (I, j) = 0)}$$

$$BK(I, j) = \text{Cur_img}(i, j)$$

Else

$$BK(i, j) = \text{Primary_BG2}$$

Step 8: Calculate the corresponding gray level slice for the fetched pixel according to Eq. 4:

$$\text{Clu}_{no(i,j)} = \left\lceil \frac{\text{Cur_Img}(i,j)}{\text{Clu_size}} \right\rceil + 1 \quad (4)$$

where, i and j are the coordinators of the corresponding point, Cur_Img refers to the most recently captured image and Clu_size is the size of each gray level slice.

Step 9: Classify the input pixel as a member of the Clu_no group. It is necessary to update the parameters of this group as in Eq. 5:

$$\begin{aligned} \mu(i, j, \text{Clu}_{no}) &= \mu(i, j, \text{Clu}_{no}) \times (i, j, \text{Clu}_{no}) + BK(i, j) \\ C(i, j, \text{Clu}_{no}) &= C(i, j, \text{Clu}_{no}) + 1 \end{aligned} \quad (5)$$

Step10: If the frame number (Counter) is less than F_n goes to step 2 else calculate the maximum cluster probability, $P_{max}(i, j)$, of the pixel using Eq. 6:

$$\begin{aligned} P(i, j, m) &= \frac{C(i, j, m)}{\text{Counter}} \quad \forall m \in [0..W-1] \\ p_{max}(i, j) &= \text{Max } p(i, j, m) \quad \forall m \in [0..W-1] \end{aligned} \quad (6)$$

Now, if ($P_{max}(i, j) > \theta$), pixel (i, j) of the background (Init_BG) is assigned with the mean value (μ) of the gray level range which has the highest probability ($P_{max}(i, j)$) in this pixel (Eq. 7). θ is the convergent value:

$$\begin{aligned} \text{Init_BG}(i, j) &= \mu(i, j, m) \\ \text{Where } P(i, j, m) &= P_{max}(i, j) \end{aligned} \quad (7)$$

Step11: To increase the speed of the process of extraction of the background pixels, the converged value, f , is updated as defined by Eq. 8. This process is done whenever the counter is equal or more than F_n :

$$\theta = \eta \times \omega^{(\text{Counter} - F_n)}, \text{counter} > F_n \quad (8)$$

where, η the initial convergent value of f and j is the weighting value of the initial convergent. Here, as in Chiu *et al.* (2010), we set η and ω to 0.7 and 0.95, respectively.

After initializing the system by capturing a new frame it should extract the corresponding background.

This is the same as in our earlier research (Davarpanah *et al.*, 2010).

In addition, a noise reduction process and final background extraction are completed based on Davarpanah *et al.* (2010).

RESULTS AND DISCUSSION

In order to demonstrate the effectiveness of the proposed system and compute its efficiency in comparison to the other related works, we implemented our algorithm and four algorithms from recent studies, namely Xiaofei (2009); Kong *et al.* (2007); Chiu *et al.* (2010); and Cui *et al.* (2009). All the algorithms were implemented using VC++ 2008. They were run on a Pentium 3 GHz PC under the Windows XP Operation System. To have a fair experiment, we tested all of these five systems under the same conditions. We used the standard database ordered by PETS (IEEE International Workshop on Performance Evaluation of Tracking and Surveillance). Actually, the PETS2000 (P0), PETS2001 during the day (PD) and PETS2001at night (PN) were used. These datasets can be found at <http://ftp.pets.rdg.ac.uk/>. In addition, some other data sets were used as well. We examined the systems using Campus (CM), Fountain (FT), Shopping Mall (SM), Bootstrap (BS), Escalator (ES), Hall Monitor (HM) and Lobby (LB). These data are accessible at http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html.

In this research, the experiments were repeated several times using various frames a different number of frames. Also, the experimental results were investigated quantitatively and qualitatively.

To evaluate the efficiency of our algorithm quantitatively compare it with the other algorithms, four different scales were utilized:

- The time consumptions of the algorithms were calculated to compare the time complexity. The averages of the time cost of the algorithms are shown in Fig. 2. As it is clear from this figure, the time consumption of Chiu *et al.* (2010) in comparison to the others is better. Actually, the time complexity of their main algorithm (Extraction Background Algorithm) is ON^2 versus the others where their time complexity is ON^3 due to the size of each dimension of the images
- The Root Mean Square error (RMS error) is another scale used. The averages of the factors in all the algorithms are illustrated in Fig. 3. As this figure shows, the resulting errors in the proposed algorithm are less than the others

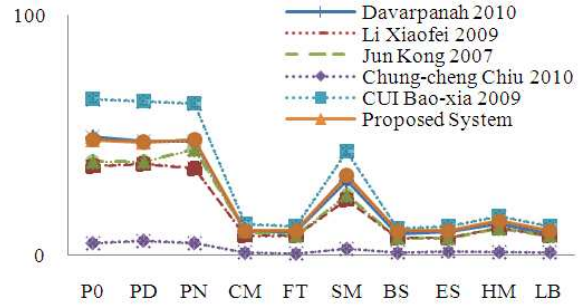


Fig. 2: Time consumption of the examined systems

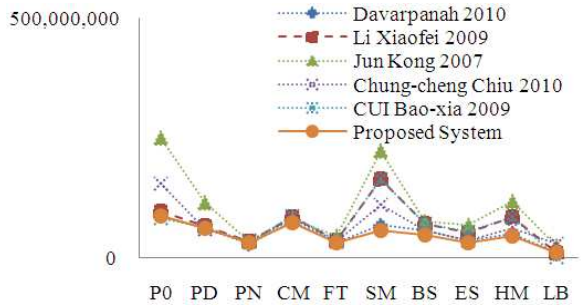


Fig. 3: Mean square error of different systems

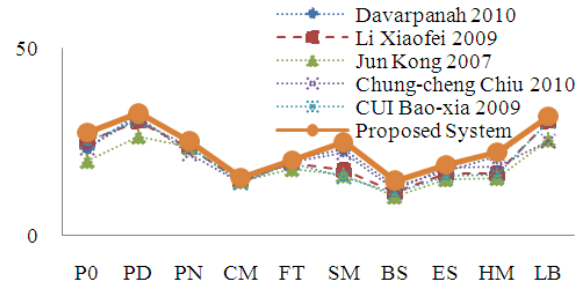


Fig. 4: Peak Signal to Noise Ratio of the systems

- Finally, as the last quantitative factor, Peak Signal to Noise Ratio (PSNR) is calculated. The value of the means of this scale for the six algorithms are plotted in Fig. 4. Since the PSNR indicates the power of the signal relative to noise, the higher values of this parameter are more valuable. The results show that the performance of our system in comparison to the others in terms of this factor is more acceptable

Figure 5 shows backgrounds extracted by two algorithms in a crowded situation and in different frames. In this test, our proposed system and Chung-Chiu (2010) are investigated. The test was executed on the Shopping Mall (SM) dataset. The results of our

experiments demonstrate that using the proposed initialization method is able to reduce the effect of a lack of a stationary object existing in the images participating in the initialization phase.

To do a qualitative test, we apply the proposed method and the other methods to four different data sets (PETS2000, PETS2001 at night, Shopping Mall and Bootstrap) separately. In each observation, the extracted binary image and the final extracted

background are saved. The binary image shows the moving objects are detected by the methods. Figure 6-9 illustrate the images achieved. These figures illustrate the Current frame (a), Extracted Final Background by Xiaofei (2009) (g), Extracted Binary Image and Extracted Final Background by the Proposed System (b1, b2), Davarpanah *et al.* (2010) (c1, c2), Kong *et al.* (2007) (d1, d2), Cui *et al.* (2009) (e1, e2) and Chiu *et al.* (2010) (f1, f2), respectively.

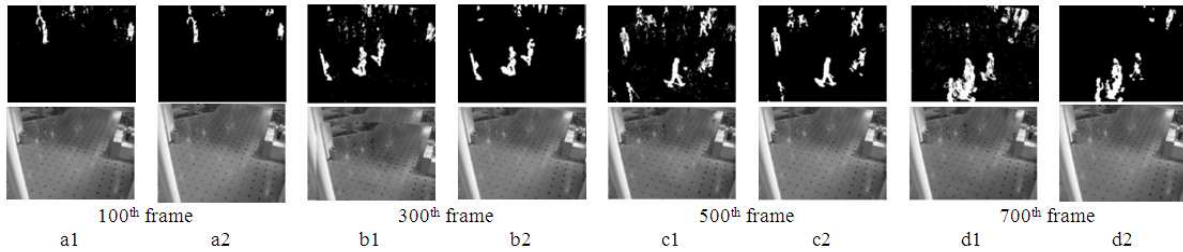


Fig. 5: Experimental Results (Extracted Binary Image and Background Image) of the proposed Method and Chiu *et al.* (2010) for the Shopping Mall (SM) data set. a1, b1, c1 and d1 are from Chiu *et al.* (2010) and a2, b2, c2 and d2 are achieved by the proposed algorithm

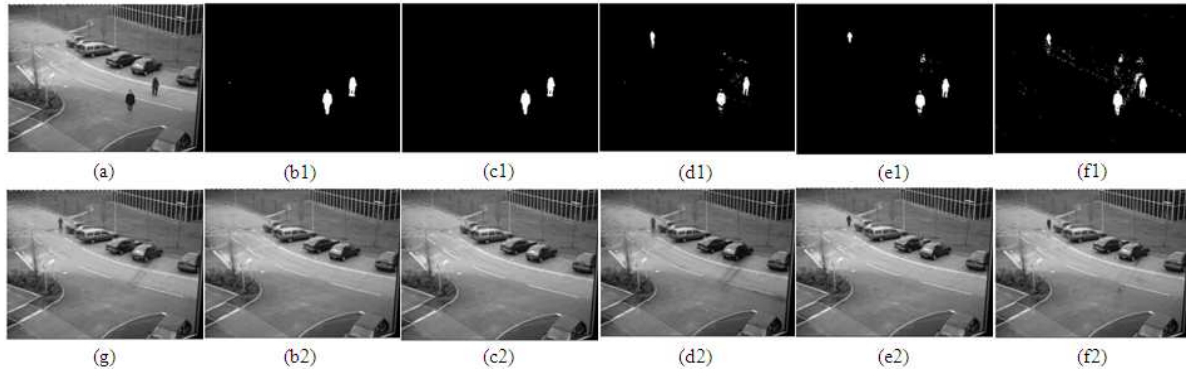


Fig. 6: Experimental Results of six algorithms in PETS2000 from the 780th frame to the 1230th.



Fig. 7: Experimental Results of six algorithms for the PETS2001 (Test data, Night) from the 100th frame to 600th

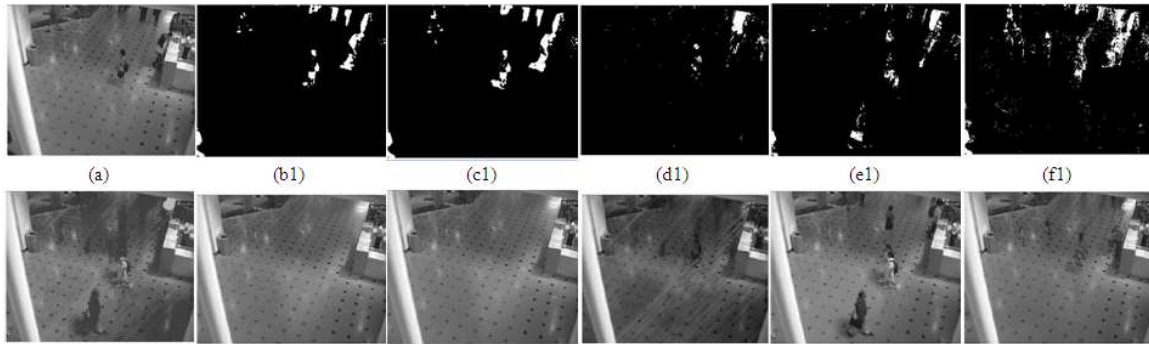


Fig. 8: Experimental Results for six algorithms for the Shopping Mall from 100th frame to the 1100th

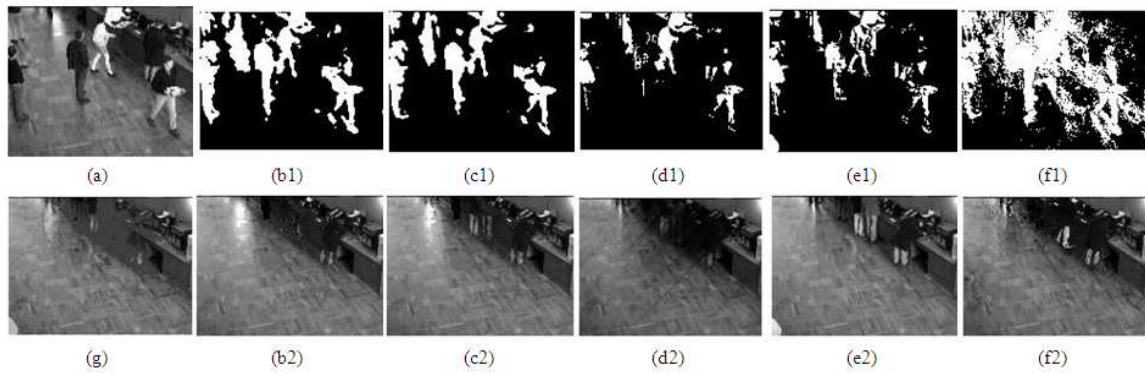


Fig. 9: Experimental Results of six algorithms for the Bootstrap from 200th frame to the 800th

CONCLUSION

In this study we introduce a new system which combines a probability-based technique and multi-scale filtering for model, to initialize and to extract an adaptable background image. Actually, the proposed system uses the history of each pixel to calculate the most probable value of that pixel in the new extracted background image. It also, in the initialization phase and to update the history of that pixel, uses a multiscale filter and applies a non-parametric filter to the current image.

Our various quantitative and qualitative experiments show that not only the accuracy of the extracted background is better for the proposed method than for the other methods investigated here, but also the time consumption is acceptable. The results illustrate that the problem of the existence of moving objects in frames used to initialize the model is solved by using the proposed initialization method.

In this research we save the history of each pixel separately. Using a blocking technique reduces the time complexity of the algorithm and strengthens it against noise caused by small motions. Also, it seems that it is not

necessary to adapt the background in each frame. Reducing such processing will result in executing the algorithm less time and thus reduce the execution time. As part of our future work, we intend to investigate these problems further.

REFERENCES

- Armanfard, N., M. Komeili, M. Valizade, E. Kabir and S. Jalili, 2009a. A non-parametric pixel-based background modeling for dynamic scenes. Proceedings of the International Conference on Advances in Computational Tools for Engineering Applications, Jul. 15-17, IEEE Xplore Press, Zouk Mosbeh, Lebanon, 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.V. Droogenbroeck, 2009. VIBE: A powerful random technique to estimate the background in video sequences. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Apr. 19-24, IEEE Xplore Press, Taipei, pp: 945-948. DOI: 10.1109/ICASSP.2009.4959741
- Chiu, C.C., M.Y. Ku and L.W. Liang, 2010. A robust object segmentation system using a probability-based background extraction algorithm. IEEE Trans. Circ. Syst. Video Technol., 20: 518-528. DOI: 10.1109/TCSVT.2009.2035843
- Cui, B.X., S.M. Sun and Y. Duan, 2009. Vehicle detection based on adaptive background. Proceedings of the 2nd International Workshop on Knowledge Discovery and Data Mining, Jan. 23-25, IEEE Xplore Press, Moscow, pp: 821-824. DOI: 10.1109/WKDD.2009.117
- Culibrk, D., V. Crnojevic and B. Antic, 2009. Multiscale background modelling and segmentation. Proceedings of the 16th International Conference on Digital Signal Processing, Jul. 5-7, IEEE Xplore Press, Santorini-Hellas, pp: 1-6. DOI: 10.1109/ICDSP.2009.5201193
- Davarpanah, S.H., F. Khalid and M. Golchin, 2010. A block-based multiscale background extraction algorithm. J. Comput. Sci., 6: 1445-1451. DOI: 10.3844/jcssp.2010.1450.1456
- Heikkila, M. and M. Pietakinen, 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
- Kong, J., Y. Zheng, Y. Lu and B. Zhang, 2007. A Novel background extraction and updating algorithm for vehicle detection and tracking. Proceedings of the 4th International Conference on Fuzzy Systems and Knowledge Discovery, Aug. 24-27, IEEE Xplore Press, Haikou, pp: 464-468. DOI: 10.1109/FSKD.2007.98
- Li, L., W. Huang, I.Y.H. Gu and Q. Tian, 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 W.B. Horng, 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 for urban traffic applications. Proceedings of the 9th International Conference on 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. DOI: 10.1109/CSSE.2008.963
- Stauffer, C. and W.E.L. Grimson, 1999. Adaptive background mixture models for real-time tracking. Proceedings of the IEEE Computer Society Conference on Computer Vision Pattern Recognition, Jun. 23-25, IEEE Xplore Press, Fort Collins, CO. DOI: 10.1109/CVPR.1999.784637
- Xiaofei, L., 2009. A new fast hybrid algorithm of background extraction suitable for the DSP platform. Proceedings of the WRI World Congress on Computer Science and Information Engineering, Mar. 31-Apr. 2, IEEE Xplore Press, CSIE, Los Angeles, CA., pp: 1-4. DOI: 10.1109/CSIE.2009.406