# A Background Reconstruction Algorithm based on Pixel Intensity Classification in Remote Video Surveillance System[1]


**Zhiqiang Hou**
Institute of Automation,
School of Electronics and
Information Engineering,
Xi'an Jiaotong University,
Xi'an, 710049,
P.R. China
Hou-zhq@sohu.com

**Chongzhao Han**
Institute of Automation,
School of Electronics and
Information Engineering,
Xi'an Jiaotong University,
Xi'an, 710049,
P.R. China
czhan@mail.xjtu.edu.cn



**Abstract**- *A background reconstruction algorithm based on pixel intensity classification is presented in this paper. Assumption that background pixel intensity appears in image sequence with the maximum probability is adopted. The pixel intensity values are classified based on the calculated pixel intensity difference between inter-frame, finally, the intensity value with the maximum frequency is selected as the background pixel intensity value. In proposed algorithm, the pre-training of no-moving object in background and the models of background and target aren't needed, and only one parameter is adjusted. Simulation results to real video surveillance sequences show that background can be reconstructed correctly, so target can be extracted perfectly and tracked successfully.*



**Keywords:** Background reconstruction, background subtraction, target tracking, image sequence analysis, video surveillance system


## 1. Introduction

In a remote video surveillance system, motion detection and segmentation are very important. Typically, this system consists of stationary camera with fixed focal length; hence, the background in the observed scenes is still. There are mainly three methods to detect and segment targets under these conditions: optical flow, frame difference and background subtraction.

Optical flow [1,2] can be used to detect and track targets without only a prior knowledge about background, this method can also be performed when the camera is moving. But optical flow method has a high complication and is sensitive to the noise. In general, a special hardware is needed by this method.

Frame difference [3,4] can detect targets in real time even though the environment is changing dynamically, but the segmentation of target is not integrated. An improved algorithm about frame difference method is presented in [5] that three-frame subtraction is used to extract the moving objects. This improvement can obtain a better result than two-frame difference, but the shortcomings of frame difference are keeping up.

Background subtraction [6-12] involves calculating a reference image, subtracting each new frame from this reference image and thresholding the result, finally, a binary segmentation of the image to differentiate the moving objects from the fixed background can be obtained. This method can get an integrated object; it is the easiest and the most effective method among the above three ones. The simplest way to perform background subtraction method is that a background image, which doesn't include any moving object and is known in advance, is chosen as a reference image, and then the current frame subtracts it. However, the background can always be changed, for instance, whatever lighting effects occur in the visual field, shadows, slow-moving objects, the season changes, objects being introduced or removed from the scene, and the camera jitters and so on; in these cases, the background must be adaptive refreshed. A standard form of adaptive background is a *time-averaged background image* (TABI), a background approximation is obtained by averaging a long time image sequences. Although this method is effective in situations where objects move continuously, it is not robust to scenes with many moving objects especially if they move slowly, and the foreground objects always can be blending into the background image.

Recently, a lot of work focus on adaptive background [6-12], these methods may be grouped into two species: one estimates background value using temporal smoothing [6-9],


[1] The work is supported by National Key Fundamental Research (973) Program in China, the project number is: 2001CB309403.

another chooses a single value from a set of past observations [10-12].

In the former methods, Ridder et al.[6] modeled each pixel with Kalman Filter which made their system more robust to lighting changes in the scene, but the method recovers slowly. Friedman and Russel[7] implemented a pixel-wise EM framework to detect vehicles, their method attempts to explicitly classify the pixel values into three separate, predetermined distributions corresponding to road, shadow and vehicle. Because background is complex, it is not enough to present background colors only using one distribution. Stauffer and Grimson[8] presented a generalization to the previous approach, the pixel value is modeled by a mixture of $K$ Gaussian distribution ($K$ is a small number from 3 to 5) to model variations in the background. Waterflower algorithm [9] is a three-component system to reconstruct background: the pixel-level, the region-level and the frame-level component, and a Wiener filter is used to update the background model. These methods often assume that an initial model can be obtained by using a short training sequence in which no foreground objects are present; actually it is difficult or impossible to control in public area. Another difficult of these methods is the background model is not enough to describe the real background, hence, the foreground objects are easily to be blended into background.

In the latter methods, the initial work was done by Long and Yang[10], they presented a adaptive smoothness method based on the assumption that intensity of background is stable for a long period, their method finds intervals of stable intensity and uses a heuristic which choose the longest, most stable interval as the one most likely to represent the background. However, for sequences in which foreground objects were stationary for a long period of time, many pixels are incorrectly classified. Gloyer et al.[11] assumed that background would be visible more than 50 percent of the time during the training sequence, they presented Median method that uses the median intensity value for each pixel as background pixel intensity value. Kornprobst et al.[12] assumed that background would be defined as the most often observed part over the sequence, they presented a approach to deal with the background reconstruction and motion segmentation based on *Partial Differential Equations* (PDE), the result is good, but their method is complex and the parameters are difficult to choose. These methods can reconstruct background even though foreground objects are present in the sequence, and can avoid blending. Because these methods reconstruct background from a past image sequence, they don't change faster than the former methods do, but make a delay, yet this doesn't have a bad effect on a long time video surveillance system.

Our algorithm belongs to the second methods. After analyzing these methods, the assumption in [12] is adopted that background would be the most often observed part over the sequence. According to this assumption, we presented a background reconstruction algorithm based on *Pixel Intensity Classification* (PIC). In this approach, pixel intensity difference between inter-frame is calculated, and the pixel intensity values are classified based on this difference, finally, the intensity value with the maximum frequency is selected as the background pixel intensity value. In proposed algorithm, the pre-training of no-moving object in background and the models of background and target aren't needed, and only one parameter is adjusted.

This paper is organized as follows: Section 2 describes the algorithm in detail. Section 3 is devoted to showing the simulation results; the comparison with the typical background reconstruction method is also given here. Finally, the conclusions are set out in Section 4.

## 2. Background reconstruction algorithm based on pixel intensity classification

### 2.1 Assumption

Motivated by the work of Kornprobst et al.[12], we adopted the assumption that background is most often visible in the image sequence, in other words, background pixels appear in the image sequence with the maximum frequency. Compared with the assumptions in [10] and [11], this assumption is more consistent with the real condition.

In [10], the authors assumed that intensity of background is stable for a long period, the longest stable interval is chosen as the one most likely to represent the background. But if foreground objects were stationary for a long period of time, the longest stable interval always represents foreground objects. In [11], the authors assumed that background can be seen more than 50 percent of the time, but if background can be seen less than 50 percent of the time, a wrong result can be obtained.

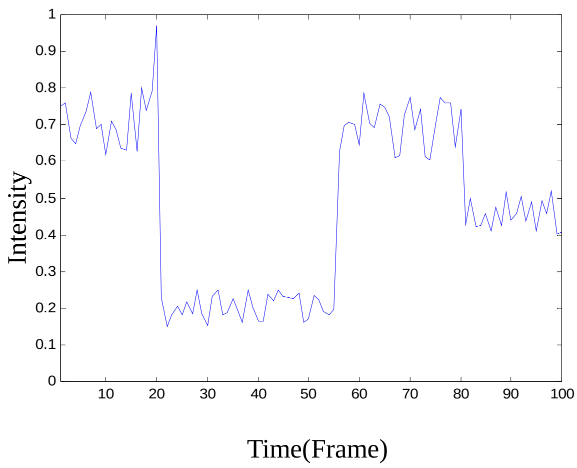



Fig.1. Example intensity history plot of a pixel in 100 frames.

A curve that represented a pixel changed in 100 frame is shown in Fig.1. It is known that curve from $1^{st}$ to $20^{th}$ frames corresponds to background, curve from $21^{st}$ to $55^{th}$ frames corresponds to a foreground object; curve from $56^{th}$ to $80^{th}$ frames corresponds to background and curve from $81^{st}$ to $100^{th}$ frames corresponds to another foreground object. If assumption in [10] is used, because curve from $21^{st}$ to $55^{th}$ frames is the longest interval, the first foreground object will be chosen as background. If assumption in [11] is used, because background can be visible in 45 percent, less than 50 percent of the time, the result is also incorrect. If we use the assumption in [12], background appears in 45 frames, the first foreground object appears in 35 frames and the second foreground object appears in 20 frames, the frequency of background is maximum, so the correct result can be obtained.

## 2. 2 Algorithm step

In [12], an algorithm based on PDE is presented according to the assumption that background is most often visible in the image sequence, however, this method is a very complex approach. Under this assumption, in this paper, we proposed another algorithm that classifies the pixel according to its intensity value in the image sequence, then the intensity value appeared in the image sequence with the highest frequency is chosen as background intensity value. The problem to be solved is how to classify the pixel in frames 1 to 20 and frames 56 to 80 as one group; classify the pixel in frames 21 to 55 as the second group and the pixel in frames 81 to 100 as the third group. Then the intensity value of pixel with the highest frequency is chosen as background intensity value.

Algorithm includes two stages. In the first stage, the intensity stable intervals are located. In the second stage, the average intensity value of each intensity stable interval is calculated, and the close results will be classified as the same group named intensity homogenous interval. After taking count of pixels of the same group, the intensity value appeared with the maximum frequency is chosen as the background intensity value.

Algorithm steps are described in the following:

**Step 1**: classify the intensity stable intervals;

Let ( $I_1, I_2, \ldots, I_M$) represent a image sequence, $N+1$ frames are selected with the same sample from these $M$ frames and marked as $(f_0, f_1, f_2, \ldots, f_N)$

$f_i(x, y)$ represents the intensity value of the point $(x, y)$ of the $i$th frame where $i = 0,1,2,\ldots, N$.

Then

$$a_j(x,y)=\begin{cases}1, & |f_j(x,y)-f_{j-1}(x,y)|>\xi \\ 0, & |f_j(x,y)-f_{j-1}(x,y)|\le\xi\end{cases} \quad j=1,2,..N \qquad (1)$$

where, $\xi$ is a threshold to determine whether the intensity value at the point $(x, y)$ changes or not. This parameter is the only one which need adjusted in our algorithm. Through experimentation, it is found that the result is insensitive to $\xi$ which can be flexibly selected. Suppose a image intensity is divided into 256 grade (8 bit), if $\xi$=10-25, the correct restoration background can be obtained, different $\xi$ only results in different running time, smaller $\xi$ needs longer running time.

In formula (1), if $a_j(x, y) = 1$, then intensity value of $f_j(x, y)$ is different from intensity value of $f_{j-1}(x, y)$, and they are in different intensity stable interval. If $a_j(x, y) = 0$, then intensity value of $f_j(x, y)$ is the same as intensity value of $f_{j-1}(x, y)$, and they are in the same intensity stable interval.

**Step 2**: calculate average intensity of each intensity stable interval;

Assume $p$ intensity stable intervals are obtained, let $(m_1, m_2, \ldots, m_p)$ represent the pixel's number of these $p$ intensity stable intervals, it can be written as

$$\sum_{s=1}^{p} m_s = N + 1;$$

Average intensity value of the $s$th intensity stable interval is

$$\bar{I}_s(x, y) = \frac{\sum_{j=b}^{b+m_s} f_j(x, y)}{m_s}$$

$$(if \quad s=1, then\ b=0; \quad if \quad s=2,3,\ldots, p; then\ b=m_1+m_2+\ldots+m_{s-1}) \qquad (2)$$

**Step 3**: classify intensity stable intervals with close average intensity values as the same group named intensity homogenous interval, and take count of pixels of intensity homogenous interval.

Select all intensity stable intervals with closed average intensity value from these $p$ intensity stable intervals, if

$$|\bar{I}_i(x, y) - \bar{I}_j(x, y)| < \frac{\xi}{2}$$

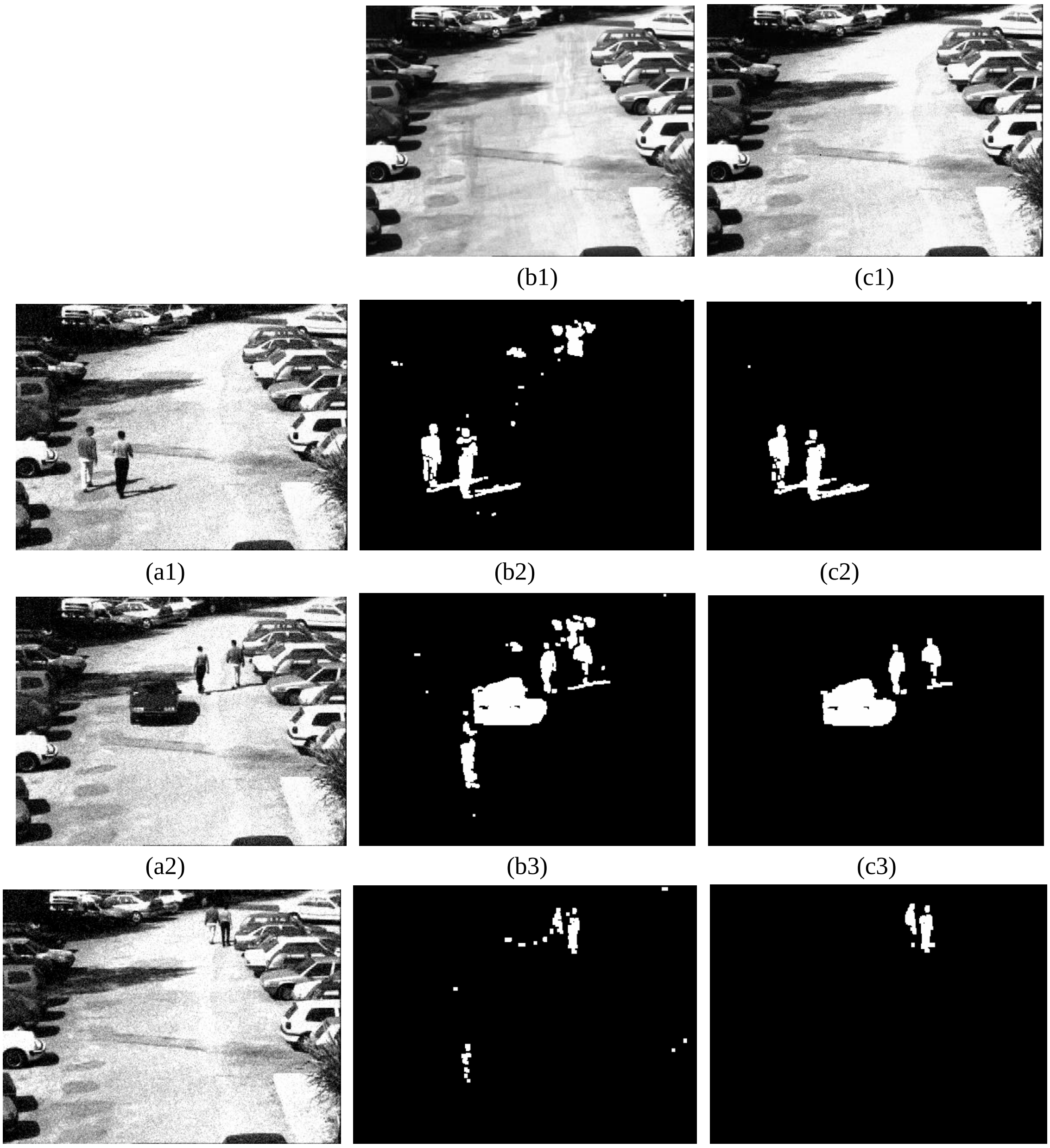
$$(i=1,2,\ldots,p; j=1,2,\ldots,p; and\ i \neq j) \qquad (3)$$

Then, the $i$th intensity stable interval and the $j$th intensity stable interval are intensity homogenous interval. In formula (3), $\xi$ is the same as formula (1).

Suppose there are $q$ intensity homogenous intervals classified, and the pixel numbers of these $q$ intensity homogenous intervals are marked as $(n_1, n_2, \ldots, n_q)$, that is

$$\sum_{t=1}^{q} n_t = N + 1;$$

If the $i$th and $j$th intensity stable intervals are classified as the $k$th intensity homogenous interval, then



(b1) (c1) (a1) (b2) (c2) (a2) (b3) (c3) (a3) (b4) (c4)

Fig.2. *Inria_1 Sequence*, (a) is the initial image, where (a1) is the 2nd frame, (a2) is the 8th frame and (a3) is the 11th frame. (b) is results by using TABI, where (b1) is the reconstruction background image, (b2), (b3) and (b4) are the results of motion detection and segmentation, (b2) corresponds to the 2nd frame, (b3) corresponds to the 8th frame and (b4) corresponds to the 11th frame. (c) is results by using our method, where (c1) is the reconstruction background image, (c2), (c3) and (c4) are the results of motion detection and segmentation, (c2) corresponds to the 2nd frame, (c3) corresponds to the 8th frame and (c4) corresponds to the 11th frame.

$$n_k = m_i + m_j;$$

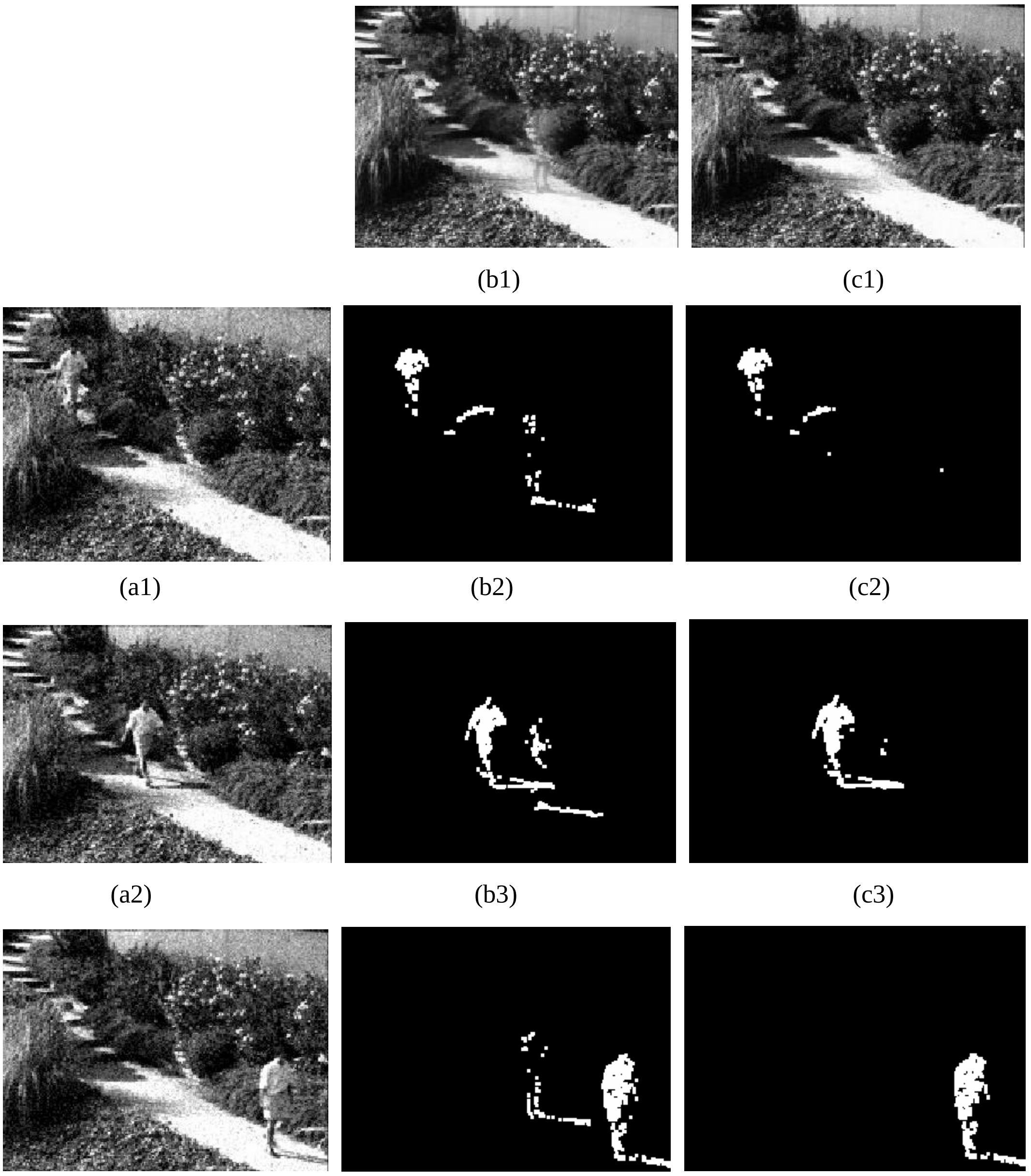
Average intensity value of the *k*th intensity homogenous interval is

$$\overline{w}_k(x,y) = \frac{\bar{l}_i(x,y) \times m_i + \bar{l}_j(x,y) \times m_j}{m_i + m_j} \quad (4)$$

Note that intensity homogenous interval can include more than two intensity stable intervals.

**Step 4**: choose intensity value with the maximum pixel number of intensity homogenous interval as background intensity value.

The maximum pixel number of intensity homogenous interval is marked as $n_{background}$ and the corresponding average intensity value is marked $w_{background}$, that is



(b1) (c1) (a1) (b2) (c2) (a2) (b3) (c3) (a3) (b4) (c4)

Fig. 3. *Inria_2 Sequence*. (a) is the initial image, where (a1) is the 30$^{th}$ frame, (a2) is the 50$^{th}$ frame and (a3) is the 112$^{th}$ frame. (b) is results by using TABI, where (b1) is the reconstruction background image, (b2), (b3) and (b4) are the results of motion detection and segmentation, (b2) corresponds to the 30$^{th}$ frame, (b3) corresponds to the 50$^{th}$ frame and (b4) corresponds to the 112$^{th}$ frame. (c) is results by using our method, where (c1) is the reconstruction background image, (c2), (c3) and (c4) are the results of motion detection and segmentation, (c2) corresponds to the 30$^{th}$ frame, (c3) corresponds to the 50$^{th}$ frame and (c4) corresponds to the 112$^{th}$ frame.

$$n_{background} = \max(n_1, n_2, \cdots, n_q) \qquad (5)$$

$$w_{background}(x,y) = \overline{w}_{background}(x,y) \qquad (6)$$

Then algorithm is finished.

## 3. Simulation results and comparisons

A lot of real video surveillance sequences were used to reconstruct background by using our algorithm. To compare with our approach, the results calculated by TABI are also given here. Two examples are shown here, the parameter $\xi$ is 10.

Fig.2 is Inria_1 sequence, there are 12 frames used to reconstruct background. Figs.2 (a1), (a2) and (a3) are the

2nd, 8th and 11th initial frames, respectively. Fig.2 (b1) is reconstruction background image by using TABI, it can be found that foreground objects are blended into background, Figs. (b2), (b3) and (b4) are motion detection results corresponding to the 2nd, 8th and 11th initial frames by using Fig.2 (b1). Because of blending, these results include some false objects. Fig. 2 (c1) is reconstruction background image by using our method, Figs. 2 (c2), (c3) and (c4) are motion detection results corresponding to the 2nd, 8th and 11th initial frames by using Fig.2(c1), it can be seen that correct results are obtained.

Fig.3 is Inria_2 sequence, there are 120 frames used to reconstruct background. In this sequence, there is one person walking from left to right, and standing a long time in the middle of scene. Figs.3 (a1), (a2) and (a3) are the 30th, 50th and 120th initial frames, respectively. Fig.3 (b1) is reconstruction background image by using TABI, it can be found that the person are blended into background in the middle of scene, Figs. 3(b2), (b3) and (b4) are motion detection results corresponding to the 30th, 50th and 112th initial frames by using Fig.3 (b1). As the same as Fig. 2, because of blending, these results include a false object. Fig. 3 (c1) is reconstruction background image by using our method, Figs. 3 (c2), (c3) and (c4) are motion detection results corresponding to the 30th, 50th and 112th initial frames by using Fig. 3(c1), it can be found that correct results are obtained.

## 4. Conclusion and future work

In this paper, a background reconstruction algorithm based on pixel intensity classification is presented. This method doesn't need making models for background and foreground. Background can be reconstructed even though there are moving objects in the scene, and blending is avoided effectively. Meanwhile, only one parameter, which can be chosen in a wide range, is needed to be determined in the algorithm. Simulation results show that the accurate background can be reconstructed by using our algorithm, hence, motion detection and segmentation can be performed correctly.

The algorithm is based on assumption that background is most often visible in the image sequence. If this assumption doesn't accord with the real condition, background doesn't reconstruct correctly. How to reconstruct background under this condition is our future work.

## References


[1] Horn B.K., and Schunk B. G., Determining optical flow, *Artificial Intelligence*, 17:185-203, 1981

[2] Smith S. M., and Brady J. M., ASSERT- 2: Real-time motion segmentation and shape tracking, *IEEE Trans. PAMI*, 1 7(7):814- 820, 1995

[3] Neri A., Colonnese S., Russo G., and Talone P., Automatic moving object and background separation, *Signal Processing*, 66(2):219-232, 1998

[4] Thomas M., and Ngan K. N., Automatic segmentation of moving objects for video object plane generation, *IEEE Trans. Circuits and Systems for Video Technology*, 8(5);525-538, 1998

[5] Jolly M-P. D., Lakshmanan S., and Jain A. K., Vehicle segmentation and classification using deformable templates, *IEEE Trans. PAMI*, 18(3):293-308, 1996

[6] Ridder C., Munkelt O., and Kirchner H., Adaptive background estimation and foreground detection using Kalman-filter, *Proc. of International Conf. on Recent Advances in Mechatronics*, ICRAM'95, UNESCO Chair on Mechatronics, pp:193-199, 1995

[7] Friedman N., and Russell S., Image segmentation in video sequences: A probabilistic approach, *Proc. of the 13th Conf. on Uncertainty in Artificial Intelligence* (UAI), San Francisco, Aug. 1-3, 1997

[8] Stauffer C., and Grimson W. E. L., Adaptive background mixture models for real-time tracking, in *Proc. of IEEE Computer Society Conf. on Comp. Vis. and Patt. Recg.*, Vol:2, pp:246-252,1999

[9] Toyama K., Krumm J., Brumitt B., and Meyers B., Wallflower: Principles and practice of background maintenance, *International Conf. on Computer Vision*, Kerkyra Greece, pp: 255-261, 1999

[10] Long W., and Yang Y., Stationary background generation: An alternative to the difference of two images, *Pattern Recognition*, 23(12):1351-1359, 1990

[11] Gloyer B., Aghajan H. K., Siu K. Y., and Kailath T., Video-based freeway monitoring system using recursive vehicle tracking, *In Proc. of IS& T-SPIE Symposium on Electronic Imaging: Image and Video Processing*, 1995

[12] Kornprobst P., Deriche R., and Aubert G., Image sequence analysis via partial difference equations, *Journal of Mathematical Imaging and Vision*, 11(1):5-26, 1999