Real-Time and Accurate Segmentation of Moving Objects in Dynamic Scene

Tao Yang
College of Automatic Control
Northwestern Polytechnical University
Xi'an China 710072
yangtaonwpu@msn.com

Stan Z.Li
Microsoft Research Asia
Beijing China 100080
szli@microsoft.com

Quan Pan, Jing Li
College of Automatic Control
Northwestern Polytechnical University
Xi'an China 710072
quanpan@nwpu.edu.cn,
Jingli_xiao@msn.com

ABSTRACT
Fast and accurate segmentation of moving objects in video sequences is a basic task in many computer vision and video analysis applications. It has a critical impact on the performance of object tracking and classification and activity analysis. This paper presents effective methods for solving this problem. Firstly, a fast and efficient algorithm is presented for background update to handle various sources of scene changes, including ghosts, left objects, camera shaking, and abrupt illumination changes. This is done by analyzing properties of object motion in image pixels and temporal frames, and combining both levels of constraints. Moreover, the algorithm does not need training sequence. Secondly, a real-time and accurate moving object segmentation algorithm is presented for moving object localization. Here, a novel filtering method is presented based on multiple scale and fast connected blob extraction. An intelligent video surveillance system is developed to test the performance of the algorithms. Experiments are performed using long video sequences under different conditions indoor and outdoor. The results show that the proposed algorithm is effective and efficient in real-time and accurate background update and moving object segmentation.

Categories and Subject Descriptors
I.4 [**Image Processing And Computer Vision**]: Segmentation—*pixel classification*

General Terms
Algorithms

Keywords
Background modeling, foreground segmentation, video surveillance, video processing.

VSSN'04, October 15, 2004, New York, New York, USA.

1. INTRODUCTION[1]
Segmentation of moving object in video sequences is a basic task in many computer vision and video analysis applications, for instance, video surveillance [1,2,3], indexing for multimedia [4,5,6], people detection and tracking [1,2], perceptual human-computer interface. Accurate moving object segmentation will greatly improve the performance of object tracking, recognition, classification and activity analysis. The most common approaches to identifying the moving objects are optical flow [7] and background subtraction [1,2,3,8,9,10,11] based ones. Optical flow can be used to detect independently moving objects in the presence of camera motion. However most optical flow based methods are computationally complex and cannot be applied to full-frame video streams in real-time without specialize hardware. Even though many background subtraction algorithms have been proposed in the literature, the problem of moving object segmentation in dynamic scene is still far from being completely solved. To design the moving segmentation algorithm for a real-time surveillance system, several serious problems have to be concerned, The first one comes from fast and accurate background modeling and maintenance, in the presence of many serious problems such as ghost, left object, uncertainty camera shaking, abrupt illumination changes would bring great challenges to this problem. Another problem is how to achieve an accurate segmentation result in real time. Chris Stauffer [7] deal with motion segmentation problem based on an adaptive background subtraction method by modeling pixels as a mixture of Gaussians and uses an on-line approximation to update the model. Several improvements on Gaussian mixture modeling have been made by P.KaewTraKulPong [9,10]. Robert [1] presents a three-frame differencing operation is performed to determine regions of legitimate motion, followed by adaptive background subtraction to extract the entire moving region. Toyama [11] developed a called three level processing (Wallflower) approach for background maintenance.

This paper addresses the problem of real time moving object segmentation in dynamic scene. Rather than relying upon the distribution of the pixel value [7,9,10], a two level background update fusion algorithm is presented based on analyzing properties of object motion in image pixels and temporal frames.

[1] The work presented in this paper was performed at Microsoft Research Asia. T. Yang, Q. Pan, and J. Li were also sponsored by the foundation of National Laboratory of Pattern Recognition and the National Natural Science Foundation (#60372085) of China.

The basic idea of this pixel level background update algorithm comes from an assumption that the pixel value in the moving object's position changes faster than those in the real background. Fortunately, this is a valid assumption in most application fields such as traffic video analysis, people detection and tracking in intelligent meeting, outdoor security surveillance in residential area, parking lot and entrance. Under this assumption, we can distinguish the foreground and background accurately by a simple frame-to-frame difference method, which could only detect the fast changes of pixel.

However, this simple method will fail when the inside color of object is uniform. In this situation, pixel values do not vary within the object. To deal with this problem, we present a dynamic matrix to analyzing the changes detection result of the frame-to-frame difference method, where the motion state of each pixel is stored in the matrix. Only those pixels whose value do not change much can be updated into the background.

Although the pixel level background update method could deal with many serious problems mentioned above, it still has a drawback in that it only considers each individual pixel while ignoring the motion information contained in the frame. To solve this problem, a frame level update method is presented. When the motion in the frame is small (less than a threshold), we can make decision that there are no moving objects in the frame and all those stable pixels will be updated into the background without analyzing the dynamic matrix in the pixel level. With the help of frame level update method, our system could make fast adaptation to abrupt background changing such as abrupt camera shaking and illumination changing.

In order to get real-time and accurate object segmentation result, an "OR" logic is used to get the subtraction image after the background subtraction between the input frame and the reference background model in R,G,B channels, so as to avoid losing object information. However, lot of noise would remain due to the use of the "OR" logic. Here, a novel filtering method is presented based on fast-connected blob extraction. Blob with small size will be considered as noise. In particular, before the extraction of connected blobs, we perform down sampling on the background subtracted image first, so as to get rid of small noise and increase the speed of the blob extraction step. After the filtering, the object position will be transformed to the original background image to produce accurate segmentation result.

An intelligent video surveillance system is developed to test the performance of the algorithms. Experiments are performed using video sequences under different conditions indoor and outdoor. The results show that the proposed algorithm is effective and efficient in accurate background update and moving object segmentation in dynamic scene. In particular, it is highly computationally cost effective and accurate and thus provides enough time for further target tracking, classification.

The remainder of this paper is organized as follows. Section 2 outlines the algorithms. Section 3 describes the background update algorithm. Section 4 explains the real time moving object segmentation algorithm. Section 5 and 6 contain the experimental results and discussion of future extensions.

2. OUTLINE OF THE ALGORITHM

The flow diagram of the algorithm is shown in Figure 1. There are five major parts: pixel level motion detection, frame level motion detection, background update, background subtraction and object segmentation. Pixel level motion detection identifies each pixel's changing character over a period of time by frame-to-frame difference and analyzes the dynamic matrix presented in this paper. Frame level motion detection focuses on the motion pixels ratio in the current frame. Fusing the detection result of both pixel and frame level, the background update model will maintain the suitable background model under different conditions. In background subtraction step, each video frame is compared against the reference background model, pixels in the current frame that deviate significantly from the background will be detected. After the real time object segmentation based on connected blob extraction and image down sampling, the moving object positions will be gained and transformed to the original subtraction image to get the accurate final segmentation results.

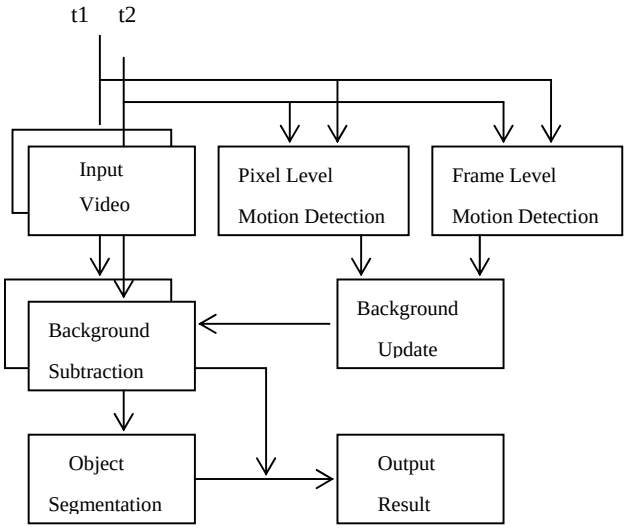

Figure 1. The block diagram of the algorithm.

3. BACKGROUND UPDATE

Background subtraction, the process of subtracting the current image from a reference one, is a simple and fast way to obtain the moving object in the foreground region and has been employed by many surveillance systems [1]. The first step of moving object detection is to acquiring the reference image. The most straightforward approach would be to simply set up the camera, empty the scene of any moving objects, and take a snapshot. Although this approach is simple, it is always impractical in real scene. For instance, it is difficult to empty a scene, the illumination can change over time, and the camera position can drift. A more practical approach is background update model that can adapt to a slowly changing background in real-time.

Given a new pixel sample, there are two alternative mechanisms to update the background model. Firstly, selective update: add the new sample to the model only if it is classified as a background sample. Secondly, blind update: just add the new sample to the model. The first enhance detection of the targets, since target pixels are not added to the background. This involves an update decision: we have to decide if each pixel value belongs to the background or not. The problem of this approach is that any incorrect detection decision will result in persistent incorrect detection later, which is a dead lock situation. The second approach does not suffer from this deadlock situation since it does

not involves any update decisions, however, it allows intensity values that do not belong to the background to be added to the model, which leads to bad detection of targets as they erroneously become part of the model.

Consider the time-varying value of a pixel of a video sequence, one popular method which belongs to the selective update mechanisms is to model the probability of observing the current pixel value as a mixture of K Gaussian distribution [8] , here more than one Gaussian is a possibility for bimodal scenes such as a tree swaying in the wind and or a flashing light, for each pixel in a frame of video, the method has to consider the last N values taken by the pixel, find the K Gaussian and weights that best fit this sample of N values using an algorithm such as K-Means or Expectation Maximization(EM), in the resulting differencing image, any value larger than three standard deviations from the mean is considered foreground. Although the performance of K Gaussian background model [8] is satisfied in theory, the proceeding algorithm is too computationally intensive to real-time use, especially the step of fitting K Gaussians to the data for each pixel and every frame, what's more, it will cost long time to adapt the abrupt scene changing such as camera shaking, illumination changing and new left object in the scene.

As we have mentioned above, the pixel value in the moving object's position changes faster than those in the real background. Under this assumption, different to the statistic-based methods above, we take the moving character as the main feature to distinguish whether it is foreground or not. Here the moving character includes two levels, the pixels level and the frame level. The pixel level makes the preliminary classifications of foreground versus background and also handles adaptation to changing background. All pixel level processing happens at each pixel independently and ignores information observed at other pixels. The fame level addresses fast background update when no moving object in the field of view and it makes the method robust even in abrupt camera shaking or illumination changing. With the help of the two level background update processing steps, our method is accurate and fast to scene-changing problem.

3.1 Pixel level

The pixel level processing step analyzes a dynamic matrix $D(k)$ to make a decision whether this pixel belongs to foreground or not. Let $I(k)$ denotes the input frame at time k, and the subscript i, j of $I_{i,j}(k)$ represent the pixel position. Equation (1) and (2) show the expression of frame-to-frame difference image $F(k)$ and the dynamic matrix $D(k)$ at time k .

$$F_{i,j}(k) = \begin{cases} 0 & \left|I_{i,j}(k) - I_{i,j}(k-\gamma)\right| \le Tf \\ 1 & otherwise \end{cases} \quad (1)$$

$$D_{i,j}(k) = \begin{cases} D_{i,j}(k-1) - 1 & F_{i,j}(t) = 0, D_{i,j}(k-1) \ne 0 \\ \lambda & F_{i,j}(t) \ne 0 \end{cases} \quad (2)$$

Where γ represent the interval time between the current frame and the old one, Tf is the threshold to make a decision whether the pixel is changing at time k or not, and λ is the time length to record the pixel's moving state, once the $D_{i,j}(k)$ equates to zero, the pixel will be updated into the background with a linear model (3).

$$B_{i,j}(k) = \alpha \cdot I_{i,j}(k) + (1-\alpha) \cdot B_{i,j}(k-1) \quad (3)$$

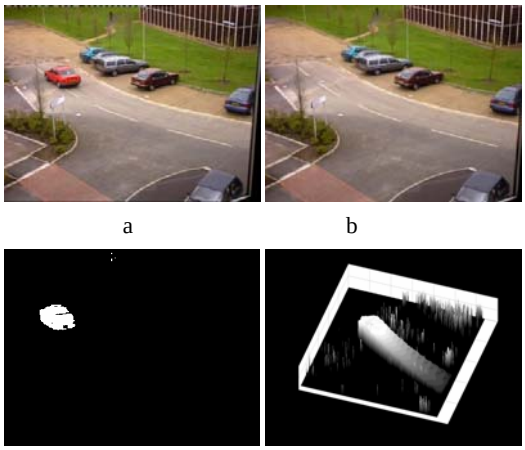

Figure 2. Dynamic matrix. a) Input video b) Background c) Subtraction Image d) Dynamic matrix

where $B(k)$ is the background image at time k and α is the weight of input frame. We update background only when the pixels value do not change much for a period of time. One advantage of this background update mechanism is that it makes the selective pixel update step more reliable. Another advantage is that it could handle fast and accurate background modeling in many serious situations such as ghost, object starts moving or stopping, illumination changing, camera shaking and so on. In addition, it is not sensitive to noises caused by slow background changing because usually the real background pixels do not change much in a short period of time and will be updated into the background quickly, as a result, in the following background subtraction step, they will not be detected as the foreground.

Figure 2 shows an example of the dynamic matrix. Figure 2(a) is the current input video. Figure 2(b) and (c) show the background and subtraction image separately. Figure 2(d) displays the dynamic matrix $D_{i,j}(k)$ and it is easy for us to see the motion of each pixel over the past several frames from it. The colors of the pixels in Figure 2(d) are in direct proportion to the value of the dynamic matrix. The white pixel in Figure 2(d) represents the value of the matrix at that point equates to λ , and the black pixels means the matrix value is zero, which can be updated into the background. The gray area clearly shows the changing of the value on the trajectory of the red vehicle.

The main advantage of this dynamic matrix technique is that it reduces the uncertainty of a pixel update step. Contrast to the statistic based background model, this approach does not need to empty a scene to get the initial background and can be run in real time. In particular, it is robust even in the dark or the high illumination part of the scene, which is full of noises and may not be modeled by Gaussian distribution.

3.2 Frame level

The frame level background update mechanism utilizes the moving character of the whole image ν (4) to achieve fast background update.

$$\nu = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m} F_{i,j}(k)}{m \times n} \quad (4)$$

where m, n represent the width and height of the image. In our experiments, once ν is less than 0.001, we will make a decision that no moving object in the current image and update all the stable pixels in the current frame to the background immediately using (3). What's more, the update weight of input frame is larger than the old background in frame level because if there is no moving object in the entire frame, it will be more likely to be the background in our definition of background, which takes the motion as the main difference of foreground and background.

Our approach to frame level can make full use of the moving information of the input video and update the background in time. For instance, when a stable object moving out of the field of view, it will cost the pixel level λ frame to deal with the ghost, however once the object is out of the image, the frame level will update the background immediately, the same thing also happens when the camera shaking or abrupt illumination changing, which is a seriously problem for many background model and always cost long time to recover from it, in our approach, the background will be updated in real-time.

4. REAL TIME MOVING OBJECT SEGMENTATION ALGORITHM

In order to get real-time and accurate object segmentation result, an "OR" logic is used to compute the background subtraction image between the input frame and the reference background model in R,G,B channels (5), so as to avoid losing object information. In equation (5), $S(k)$ denotes the subtraction image at time k, Ts is the subtraction threshold, and $I.R(k)$, $I.G(k)$, $I.B(k)$, $B.R(k)$ $B.G(k)$ $B.B(k)$ represent R,G,B pixel value of input frame and background model separately.

$$S_{i,j}(k) = \begin{cases} 1 & \begin{array}{l} \left|I.R_{i,j}(k) - B.R_{i,j}(k)\right| > Ts, or \\ \left|I.G_{i,j}(k) - B.G_{i,j}(k)\right| > Ts, or \\ \left|I.B_{i,j}(k) - B.B_{i,j}(k)\right| > Ts \end{array} \\ 0 & otherwise \end{cases} \quad (5)$$

However, lot of noise would remain due to the use of the "OR" logic and will consequently affect the accuracy of the segmentation. Fortunately, in most application fields, the noises are always random and the size of the noises blob is relatively small and it can be removed through connected area analysis. Figure 3 shows the diagram of the algorithm. Firstly, we down sampling the background subtraction image , the main purposes of this step including two aspects: filtering small noises and decreasing the computational burden. Then, a fast blob labeling algorithm is used to label the connected pixels with same ID. After that, the morphology-filtering step will get rid of little blobs according to their size. Since segmentation of moving object in video sequence is a basic and critical task in many video analysis applications, an accurate segmentation will do great help to the further analyses, for instance, target classification and recognition.

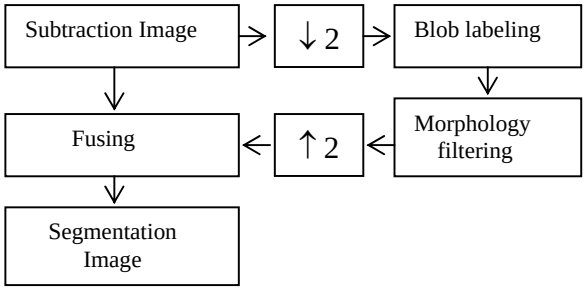

Figure 3. The block diagram of the object segmentation algorithm.

Thus, instead of using the filtered image, in which the contour of the target is smoothed and hard to be recognized, we fuse the target position information from the filtered image and the subtraction image to get the final segmentation image.

5. EXPERIMENTS

We have developed a real time intelligent surveillance system based on the presented algorithms. The system is implemented on standard PC hardware (Pentium IV at 1.3GHz). The video image size is 320x240 (24 bits per pixel). The system is tested in typical indoor and outdoor environments for handling ghost situation, background modeling, abrupt large area changes, left object, and object detection and tracking. The image sequences in those examples are captured by Sony dcr-pc9e at 25fps and the average frame rate of our system is 14.5fps. The following presents results.

5.1 Ghost

This situation is due to a background change caused by an object that is previously part of the background but now starts its motion. Figure 4 measures reactivity in a limit condition when the background reflects changes from a car that starts its motion after having previously been part of the background. While the car is parked, it is included in the background image. At frame #10, (Figure 4, first column), it starts reversing. Until frame #97 (Figure 4, second column), the moving object still substantially covers the area where it was stopped, preventing separation from its forming ghost. However, after a few frames (frame #150, frame #210), the correct background update and the correct segmentation can be achieved with our approach (Figure 4, fourth column).

5.2 Auto Background Modeling

In many real application fields, it is usually difficult to get a clear background as the training sequences. For example, in traffic surveillance sequence almost every frame contains moving objects. In the content-based video retrieval application, it's necessary to extract meaningful video objects from scenes to enable object-based representation of video content, thus the automated background creation is quite necessary. Figure 5 shows one of our

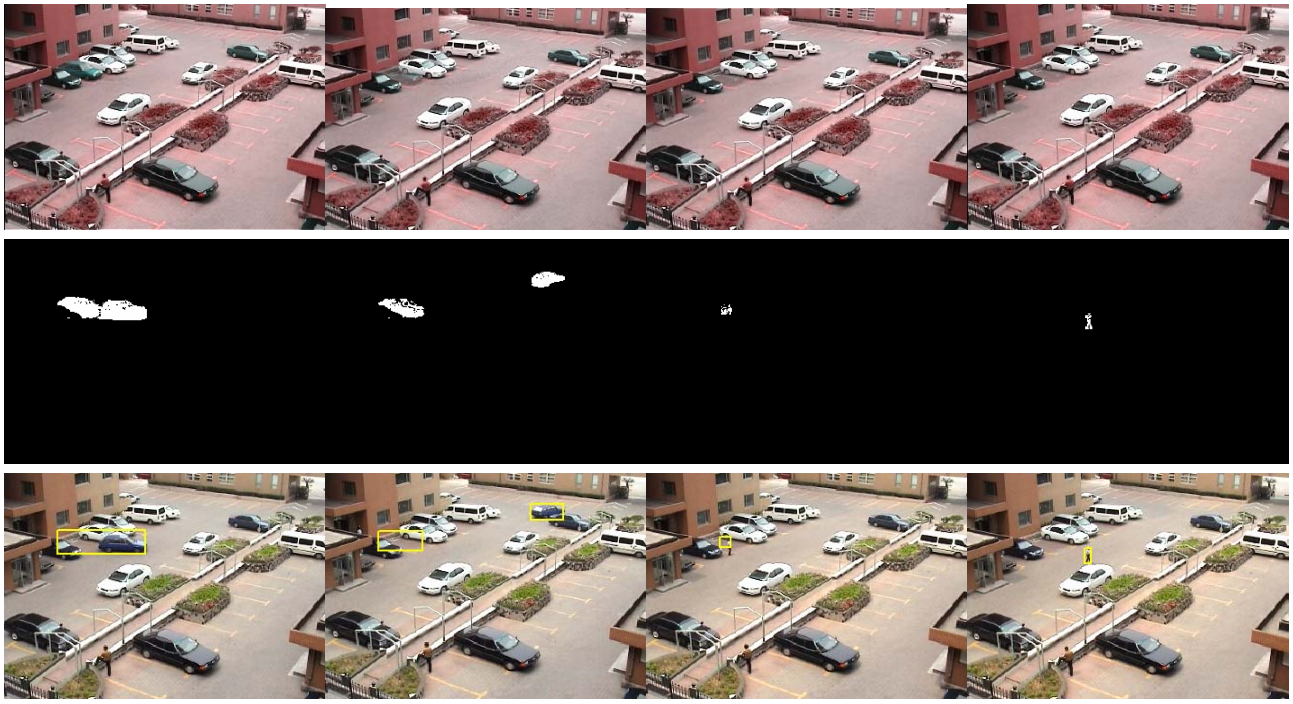

Figure 4. The reactivity of background model when ghost happens. These images are captured at frame #10, frame #97, frame#150, frame #210 separately. The first row contains the background model. The second row contains the moving object segmentatin result. The third row shows the objects' position boxes calculated from the blob labeling results.

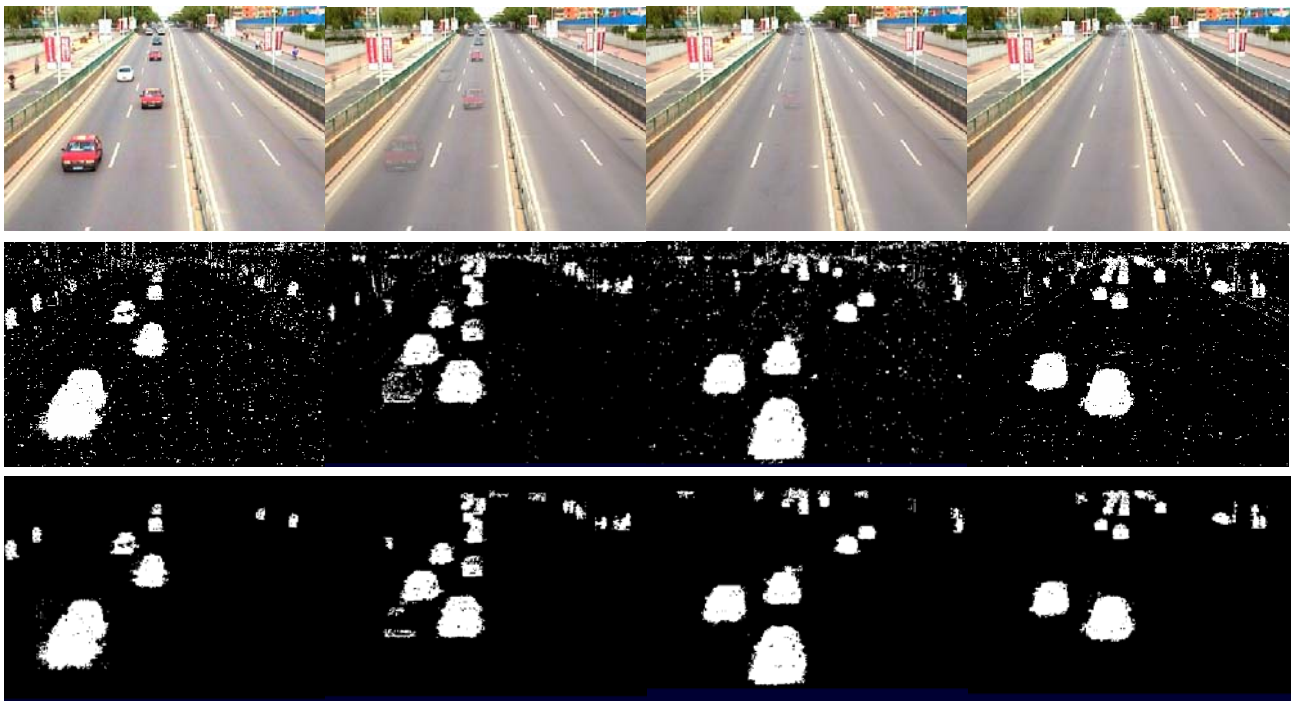

Figure 5. Vehicle segmentation for traffic video surveillance . These images are captured at frame #1, frame #78, frame#161, frame #236 separately. The first row contains the background model. The second row contains the subtraction image. The third row contains the moving object segmentation result. The fourth row shows the objects' position boxes calculated from the blob labeling results.

(a) **(b)**

Figure 6. Abrupt large area changing . These images are captured at frame #1, frame #30. (a) The first column contains the background in abrupt camera shaking. The second column contains the subtraction image. (b) The first column contains the background in abrupt illumination changing. The second column contains the subtraction image.

experiment results in traffic surveillance application. When we begin to detect traffic parameters at frame #1, the background contains many moving vehicles, several frames later, the pixel level update approach modeling the correct background and accurate segmentation can be achieved, frame #78,161and 236 show the background update process (Figure 5, first row)of our approach. It is hard for those statistic methods to achieve this result without background training sequences. If the value of the detected foreground points is used in the static update, but with a limited weight, the update will be very slow. Contrast to the original background subtraction image (Figure 5, second row), the objects' appearance in the final segmentation result (Figure 5, third row) of our method is accurately preserved.

5.3 Abrupt Large Area Change

Pixel based background update methods a long time to adopt to this change, however, through analyzing the motion of the entire frame, the frame level background update methods will model the accurate background with little delay. In Figure 6(a), an abrupt camera shaking happens at frame #1, at this movement, subtraction image clearly shows that the majority of the image area changed, (Figure 6(a), second column, upper image). As we have mentioned above, the weight of the input frame in frame level is larger than the background model, thus once the camera stops moving, the correct and fast background update can be achieved by our frame level update approach. The same correct background update result can be found in Figure 6(b), which shows real-time background update under the abrupt illumination changing.

5.4 Left Object

"Left Object" means a new object is left in the scene or an old object is placed in different places. Figure 7 shows a person moves chair to a new place. Before the chair is moved, the person is detected correctly at frame#1, (Figure 7, second column), at frame #13, the chair is placed to a new place, and a ghost is detected at the following several frames. Few frames later, our pixel level update approach correctly select and update the moved red chair and the ghost area into the background model while segment the moving person accurately at the same time (frame #98, #189). In Figure 8, a person left a white bag in the scene (frame #540), after several frames, the bag is updated into the background and the moving person can be detected correctly (frame#619, second column). In particular, through analyzing the pixel value difference between the new background pixel and the old one in our background update step, the new object can be classified easily. The green area in the background (frame#619, first column) represents the left bag, and it will be helpful for application like abandoned bag detection for security surveillance.

5.5 Object Detection and Tracking

Figure 9 shows some scenes and multiple objects detection and tracking results. Except the last one which comes from the PETS test data (Figure 9, right, lower image), all the other test

sequences are captured in the real scene. Based on the object segmentation result, kalman filter and data association methods are used to estimate the motion parameters of each target. In particular, although the targets speed of those scenes are quite different, our system performed efficiently and accurately with little parameters adjusting.

6. CONCLUSION

This paper has presented a real-time and accurate algorithm for moving object segmentation in dynamic scene. The algorithm has the unique characteristic of explicitly addressing various difficult situations such as ghosts, automatic background modeling, left object, uncertainty camera shaking , and abrupt illumination changes. Further it avoids problems caused by undesired background modification. Based on this algorithm, an intelligent video surveillance system has been developed and experiment results proved that this system performed robustly in quite different indoor and outdoor environment. Moreover, based on the moving object algorithm of this paper, which is highly computationally cost effective, the system can perform in real-time even on common PC.

Bowman, B., Debray, S. K., and Peterson, L. L. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst., 15,* 5 (Nov. 1993), 795-825.

7. REFERENCES

[1] Collins, Lipton, Kanade, Fujiyoshi, Duggins, Tsin, Tolliver, Enomoto, and Hasegawa. A System for Video Surveillance and Monitoring.*VSAM Final Report*, *Technical report* CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May, 2000.

[2] Haritaoglu, I. Harwood, D., Davis, L.S.. W4: real-time surveillance of people and their activities**.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 22 , Issue: 8 , Aug. 2000, 809–830.

[3] Pakorn KaewTrakulPong, Richard Bowden. A real time adaptive visual surveillance system for tracking low resolution colour targets in dynamically chaning scenes. *Image and Vision Computing* 21,2003, 913-929.

[4] Cucchiara, R., Grana, C., Piccardi, M., Prati, A..Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 25 , Issue: 10 , Oct. 2003 ,1337-1342.

[5] Xu, H., Younis, A.A., Kabuka, M.R..Automatic Moving Object Extraction for Content-Based Applications. *IEEE Transactions on Circuits and Systems for Video Technology*,Volume: 14 , Issue: 6 , June 2004 ,796 – 812.

[6] Shu-Ching Chen, Mei-Ling Shyu; Peeta, S., Chengcui Zhang. Learning-based spatio-temporal vehicle tracking and indexing for transportation multimedia database systems. *IEEE Transactions on Intelligent Transportation Systems* , Volume: 4 , Issue: 3 , Sept. 2003,154–167.

[7] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1): 42–77, 1994.

[8] Stauffer, C, Grimson, W.E.L. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Volume: 22 , Issue: 8 , Aug. 2000,747–757.

[9] P.KawTraKulPong, R.Bowden. An improved adaptive background mixture model for real-time tracking with shadow detection. *In Second European Workshop on Advanced Video-based Surveillance System*s,2001.

[10] Liyuan Li, Weimin Huang, Gu, I.Y.H., Qi Tian. Foreground object detection in changing background based on color co-occurrence statistics. *Proceedings of Sixth IEEE Workshop on Applications of Computer Vision* , 3-4 Dec. 2002, 269–274.

[11] K. Toyama, J. Krumm, B. Brumitt, and B.Meyers . Wallflower: Principles and practice of background maintenance. *Proceedings of IEEE International Conference on Computer Vision*, 255-261,1999.

Figure 7. Left object, a person move a red chair into different place. These images are captured at frame #1, frame #13,frame #98, frame #189. The first column contains the background model. The second column contains the bounding box of the detected moving object.

Figure 8. Left object, a person left a bag in the scene. These images are captured at frame #540, frame #619. The first column contains the background model. The green area represents the left object. The second column contains the bounding box of the detected moving object.

Figure 9. Real time target detection and tracking results.