2010 International Conference on Measuring Technology and Mechatronics Automation

A Novel Method on Moving-Objects Detection Based on Background Subtraction and Three Frames Differencing

Lian Xiaofeng, Zhang Tao, Liu Zaiwen

Beijing Technology and Business University, Beijing, 100048, China
lianxf@th.btbu.edu.cn

Abstract—**This paper proposes a new method for moving-objects detection based on fusion of background subtraction and an improved three frames differencing. In the method, the adaptive background model is build by Gaussian model for each pixel in the image sequences, and combined with temporal differencing method to update the selective background, and simultaneously use background subtraction method to extract movement areas from the background model. Next, adopt the median filter and mathematical morphology operation to eliminate noise and the small areas of non-human motion parts. Finally obtain the complete reliable moving-objects. The experimental results show that the proposed method has high accuracy, and can meet the needs of real-time. So it can be applied in visual surveillance system effectively.**

Keywords: Moving-object detection; Background subtraction; Three frames differencing; Background updating

I. INTRODUCTION

Moving-object detection is one of the basic and vital portions in visual surveillance system with the purpose of subtracting interesting targeted area and locating the moving-objects from video flows (image sequences). It's important to segment the motion areas effectively for recognition, tracking as well as behavior understanding. However, due to the diversity of applications in the reality, such as light changes, camera shake, self- and mutual-occlusion of the moving-objects, the accuracy and rapidity of the moving-objects detection is challenged with a variety of complications. At present, almost all the moving-objects detection algorithms have been based on a certain degree of constraints and assumptions, therefore it remains to be one of utmost research topics in the field of computer vision applications. In general, the moving-objects detection methods have the following three original ways: background subtraction, inter-frame differencing and optical flow. Hereinto, *Background Subtraction*[1] uses difference image between current frame and background reference frame to threshold and segment the motion areas. The outstanding feature of the method is that it can obtain a relatively complete moving-object, but it's sensitive to the dynamic environment; *Inter-Frame Differencing*[2] usually differences the grey-level of the adjacent two or three images, and then obtains the motion region by thresholding. This method realizes simply, operates quickly and has strong adaptability for the dynamic scene etc, but the integrity of the moving-objects extracted is dissatisfied, and easy to produce the cavities in interior. *Optical Flow*[3] reflects the light changing tendency of every pixel in scene. Without any priori information, it can detect an independent moving-object. The method, however, can't meet the need of real-time because the heavy computation load and need relevant hardware support. What is worse, it's easily affected by noise.

In view of the shortcomings of traditional methods, the combination of the background subtraction and three frame differencing proposed in the paper can overcome the setbacks and flaws in single method effectively, and makes full use of their advantages. In the new method, firstly we select an image as background frame, and build the adaptive background model for each pixel in the image sequence by Gaussian model, then uses the information obtained by the temporal difference method to update the selective background, and simultaneously using background subtraction method to extract motion areas from the background model. Integration the two prospects for object identification, together with utilization of the median filter and mathematical morphology operation eliminating noise and the small area of non-human motion parts. Finally we can obtain the complete reliable moving-objects, which is convenient for the following recognition and tracking.

II. MOVING-OBJECT DETECTION BASED ON FUSION OF BACKGROUND SUBTRACTION AND COTERMINOUS FRAMES DIFFERENCING

Moving-object detection mainly contains two parts: the segment and classification of moving-objects. Firstly, subtracting the motion regions from the image sequences, and then classifying those regions to obtain the human body's part. The algorithm flow is shown as Fig.1.

A. Image Pre-Processing

In the acquisition and transmission process, the image will become blurring and rough due to the transmitting channels and sampling systems as well as the shaking of cameras and disturbing of surrounding noise. To ensure the input date to be a continual visual flow, the image should be carried on smoothed processing. This method carries on the

Supported by National Key Technology R&D Program (2007BAK361307-05); Excellent Talents Foundation of the Organization Department of Beijing Municipal Committee(D2009D005003000001) and the Research Foundation for Youth Scholars of Beijing Technology and Business University.

978-0-7695-3962-1/10 $26.00 © 2010 IEEE
DOI 10.1109/ICMTMA.2010.132

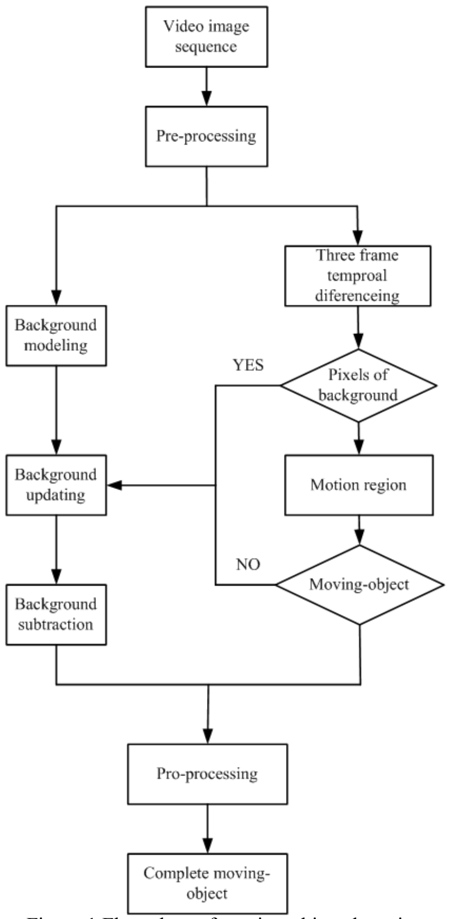

Figure.1 Flow chart of moving-object detection

spatial median filter to smooth the input video data, which can not only restrain the random pulsing noise and salt-pepper noise, but detect the edge of the moving-object as well. In our experiment selecting the 5×5 function window, and taking the point (x, y) as the central function window to calculate the mean value, then assigning this value to the corresponding pixel in the center of template.

Pre-processing also means a proper transformation of the visual data to meet the needs of algorithm format. Because the grey level is a key factor to judge the background and moving-object, furthermore, taking into account to reduce the complexity of the algorithm, we process the color image according to Eq. (1)

$$f(x,y,z) = 0.299 f(x,y,z)_R + 0.587 f(x,y,z)_G + 0.144 f(x,y,z)_B \tag{1}$$

where, $f(x,y,z)_R$, $f(x,y,z)_G$, $f(x,y,z)_B$ denote the RGB component of point (x, y) respectively.

B. Background Modeling

Background subtraction is one of the most common methods to segment moving-object, which includes background modeling and updating two steps. In this method, the detection of moving-object is based on the reference of the background image whose accuracy determines the correctness and reliability of the detection as well as the complexity. Meanwhile, since usually the scene is dynamic, background models need to be updated promptly to adapt to the changes in the external environment and reflect accurately the current state of background.

Taking into account the pixels of background region in adjacent frames change slowly and the difference was not significant in a certain period of time, while the corresponding pixels of motion regions varied widely. In addition, due to the influence of noise in the course of video sequence acquisition, the brightness and grey level of each pixel in image take a mean value as baselines and vibrate randomly within a certain deviation nearby it. Therefore, assuming that the grey level of each pixel varies independently and randomly in successive video frames, and the grey level set of pixels from the same location obeys the Gaussian probability distribution respectively. According to Eq.(2), using Gaussian model to modeling for each pixel of the selected background frames[4], and make use of Gaussian function fitting to the next frame to extract the background image.

$$P\left(X_{(i,j,t)}\right) = \frac{1}{\sqrt{2\pi}\sigma_{(i,j)}} \exp\left(-\frac{\left[X_{(i,j,t)} - \mu_{(i,j)}\right]^2}{2\sigma_{(i,j)}^2}\right) \tag{2}$$

where, $\mu_{(i,j)}$, $\sigma_{(i,j)}$ represent the mean and standard deviation respectively, $X_{(i,j,t)}$ is a set of all the pixel whose coordinate is (i, j) in the t frame of the image sequence.

At the beginning of modeling, the distribution of each pixel is unknown. Taking grey level of pixel in the background frame as $\mu_{(i,j)}$, and assigning the standard deviation $\sigma_{(i,j)}$ to zero to initialize the model. The actual background model is achieved through learning the new available frames and updating the parameters of Gaussian model corresponding to each pixel with taking into account the past background information. So the background model build by this method has good adaptability to the light changes. However, the pixels belongings to moving-object were also computed when calculate the mean of background frame, therefore the background cannot be updated timely. This paper presents a new method of selective background updating based on the coterminous frames differencing. Only the pixels belonging to static objects were updated to the background with a certain updating rate through cumulative effect of multi-frame, and then can effectively extract the integrated motion areas from the video sequences.

C. Background Updating Based on Three Frames Differencing

Temporal difference can detect the relatively changed information in the successive frames. Therefore, it's difficult to detect the overlaps in the moving-object and always cause some cavities inside the moving entity. In addition, the gray changing region detected by the temporal difference method contains the real moving-object and the background changing area which occluded by moving-object in the previous frame, thus causing the detected objects larger than it's real, and even overlaps. Using three frames temporal differencing method to search for the overlapping parts, then can obtain the moving-object and the information of background updating. Compared with the traditional method, the improved method has a significant adaptability to speed of moving-object[5]. The structure of three frame temporal differencing method is showed as Fig.2.

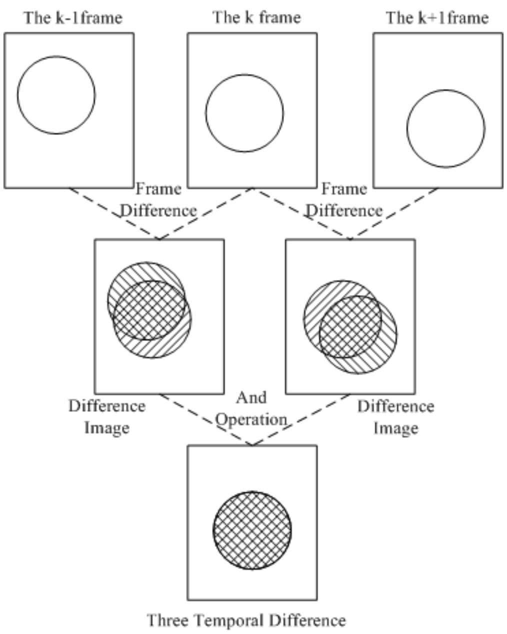

Figure.2 Three frame temporal differencing algorithm

Reading three consecutive frames $I_{k-1}(i,j)$ 、 $I_k(i,j)$ 、 $I_{k+1}(i,j)$ respectively, solving the grey-scale difference between two consecutive images $IZ_{(k,k-1)} = |I_{k-1} - I_k|$, $IZ_{(k+1,k)} = |I_k - I_{k+1}|$ separately, the rules are as follows:

if $IZ_{(k,k-1)}$ *and* $IZ_{(k+1,k)} < Th + \frac{1}{N}\sum |I_k(i,j) - I_{k+1}(i,j)|$,

then $B_1(i,j) = I_{k+1}(i,j)$

else $M_1(i,j) = I_{k+1}(i,j)$

Where, Th is the grey threshold, the additional item $\frac{1}{N}\sum |I_k(i,j) - I_{k+1}(i,j)|$ denotes image changes in overall light, so that the threshold Th can adapt to the changes of environment light. N is the number of pixels in the image, $B_1(i,j)$ 、 $M_1(i,j)$ represents separately the background region and motion region after determined by the difference. According to Eq. (3), (4) to update the pixels of $B_1(i,j)$:

$$\mu_{k+1} = (1-\alpha)\cdot\mu_k + \alpha\cdot x \tag{3}$$

$$\sigma^2 = (1-\alpha)\cdot\sigma^2 + \alpha\cdot d(i,j,t)^2 \tag{4}$$

While charring on further processes for the motion region segment $M_1(i,j)$, and fitting the pixels whose belong to it with their respective Gaussian models. In this paper, using the grey level $X(i,j,t)$ of pixel to measure the characteristics of this point, if $P(X(i,j,t)) \le T_p$ (T_p is probability threshold), then considered to be the background pixel, otherwise be the moving-object in scene. In practical applications, using the probability threshold instead of the equal threshold, take as $d(i,j,t) = |x(i,j,t) - \mu(i,j)|$, and then can set the appropriate threshold for the corresponding foreground detection to be $\sigma(i,j)$, the rules are as follows:

if $d(i,j,t) \le \sigma(i,j)$ *is the background region, then denote as* $B_2(i,j)$ *;*

else the pixel $d(i,j,t)$ *is the moving-object in scene, then denote as* $M(i,j)$.

Taking into account $B_2(i,j)$ was occluded by the moving-object in previous frame, then according to Eq.(5), (6) to include it in the background. The corresponding pixels in motion region $M(i,j)$ are not updated.

$$\mu_{k+1} = (1-\beta)\cdot\mu_k + \beta\cdot x \tag{5}$$

$$\sigma^2 = (1-\beta)\cdot\sigma^2 + \beta\cdot d(i,j,t)^2 \tag{6}$$

Where $\alpha,\beta \in (0,1)$ is the updating rate, x denotes the grey-scale of pixels in a new frame. Here, α,β are different, in addition, the points in $B_2(i,j)$ are belong to the region occluded by the moving-object in previous frame, and appear again in the current frame, so β is larger.

Experimental result show that the reliable background model $B(i,j)$ can be got by the strategy, even if the moving-object exist and motion, Figure 3(b) shows the accurate background image obtained by the background modeling and updating method proposed in the paper.

(a) Primitive Image

(b) Background Image

Figure.3 Result of background modeling

D. Moving-Object Extraction

The foreground regions in video images can be obtained by thresholding the absolute differences between the pixels of background image $B(i,j)$ and those of primitive video image $I_k(i,j)$. In this paper, we adopt the discrimination method based on relative difference which can enhance the detection effect of moving-object in shadow regions.

$$\Delta I_k(i,j) = \left| I_k(i,j) - B(i,j) \right| \tag{7}$$

$$DB_k(i,j) = \begin{cases} 1, & \Delta I_k(i,j)/B(i,j) \geq Th' \\ 0, & otherwise \end{cases} \tag{8}$$

In order to enhance the detection effect, threshold selection is the key factor for binarization processing of grey image. The experimental results show that the suitable threshold range is $15 < Th' < 20$ in which can obtain the more accurate motion regions, simultaneously can interference from the changes of external. Where, the white regions regard as the moving-object and the black express the static region in binary image $DB_i(i,j)$. Because the three frame temporal differencing method has strong adaptability to light while background subtraction can make up the disadvantage of interior cavity in the moving-object extracted by traditional temporal differencing. Therefore, we can get the complete moving-object segment from the background images through merging the motion region obtained by those two methods, as the following Eq.(9) shows.

$$Re_k(i,j) = \begin{cases} 255, & DB_k(i,j) \cup M(i,j) = 1 \\ 0, & otherwise \end{cases} \tag{9}$$

where, $Re_k(i,j)$ represents the final extracted moving-object based on the above method.

E. Post-Processing

There are still a lot of tiny residual noises except those moving-objects in the image. The moving-objects may also be incomplete, even have some few cavities which not be detected. In addition, some pixels on the background can be mistaken as moving-objects due to the slight disturbance on the background and noise effect. In order to eliminate these influences such as cavities, noises and the small area of non-human motion parts, first of all, using the median filter as well as the erosion and dilation operations of mathematical morphology to process the difference image which obtain in the previous step, and then filling the internal cavities to remove the tiny noise and small non-human moving parts in the images, finally obtaining the complete reliable moving-objects[6]. In our experiments, however, the shape of segment motion regions mostly is distorted as well as several objects are adhered which caused by the disturbance of shadows, and which not avail to further processing, such as recognition and tracking and so on. In this paper we can achieve a preferable performance in eliminating the shadows based on the method proposed in literature [7].

III. Experimental Results and Analysis

In our experiment, we test the accuracy and real-time performance of the algorithm through analyzing and processing the video streams captured by camera real-time. The environment of our experiment is Pentium 4 2.5GHz CPU and the 256M memory within Windows XP. The algorithm are test in the video sequence images which frame rate is 15FPS and size is 320×240, and experimental results demonstrate that the approach is effectiveness and meets the need of real-time processing. Some experimental results are shown as follows.

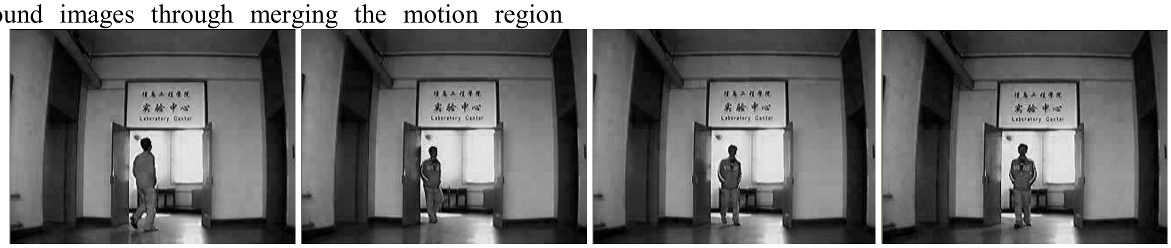

Figure.4 Primitive image sequence

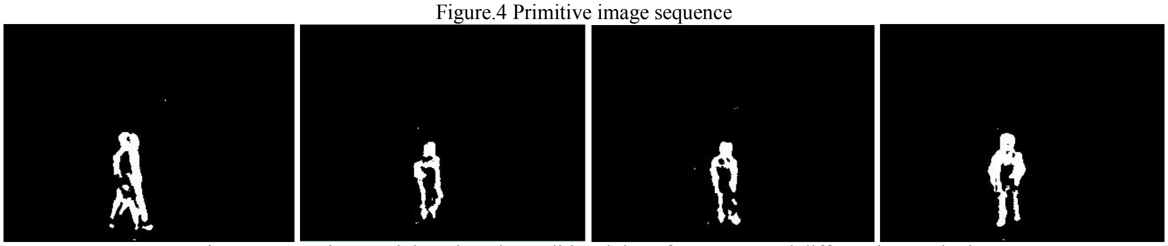

Figure.5 Detection result based on the traditional three frame temporal differencing method

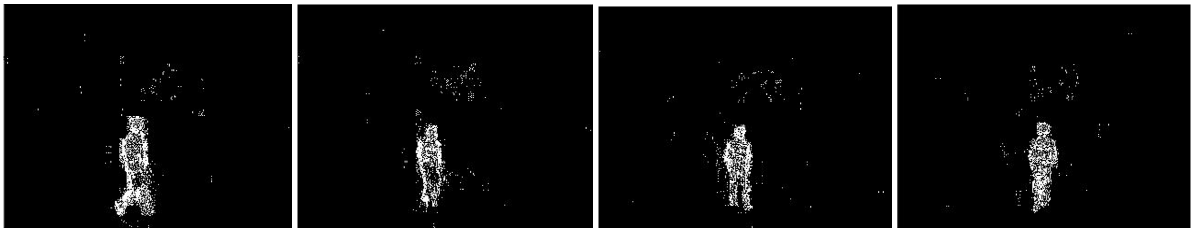

Figure.6 Motion detection image based on the proposed algorithm

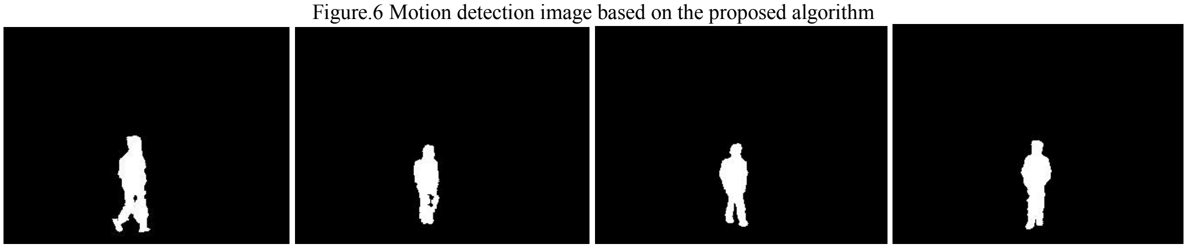

Figure.7 Final detection result after the post-processing

In Figure 4, there are four images randomly sampled from the video sequences. Figure 5 show the detection results based on the traditional three frames temporal differencing method, in which include the majority of moving-object, but the phenomenon of ghost and cavities are very obvious as the object moves too fast or too slowly which is occluded in the previous frame. The results shown in Figure 6 not only contain the non-detection moving-object based on background subtraction, but also improve the effect of traditional method, and thus ensuring the correct update of background images. The result images, however, contain a lot of small patches which don't belong to the moving-object due to noise and background disturbances. At last, the final detection results based on the new proposed method are shown in Figure 7, in which effectively suppress the impacts of noise and background disturbance by deleting the non-motion patches and thus accurately detecting the moving-objects in the video image.

IV. CONCLUSION

In view of the influence of light variety on the stability of the visual surveillance system, we proposed a new method based on fusion of background subtraction and an improved coterminous frames differencing for moving-object detection and adaptive background updating with the purpose of subtracting moving-objects in the complicated situations. From the real test in the image sequence, we can find the new method proves the effective ability which overcome the poor adaptability of background subtraction, and simultaneously made up the disadvantage of the cavity in the interior of the moving-objects based on the traditional temporal differencing. In addition, the new method effectively suppresses the impacts caused by situations changing, backgrounds disturbing and the shadows and so on. Together with measures to reduce misjudgment, the new method proves to be applicable to visual surveillance system in that it dramatically decreases the misjudgments caused by noise and significantly improves the accuracy and fastness of detection.

However, the new method also has not yet solved some problems, for example, the effect of background update is poor while the grey level between the moving-objects and the background are close or light has changed suddenly, so it is very difficult to detect the moving-objects. Moreover, the method cannot solve the problem of self- and mutual-occlusion and therefore it is necessary to develop a better model to deal with the corresponding problems accurately for the different parts of a moving-object, which will be explored in the further research.

ACKNOWLEDGMENT

We would like to thank all the members of the ICMCN research group at BTBU for their fruitful insights and commentary. A special thanks goes to Pro.Liu for his supervision in the research process.

REFERENCES

[1] N.T.Siebel, S.Maybank. "The ADVISOR visual surveillance system". In Proceedings of the ECCV 2004 Workshop Applications of Computer Vision. Prague, Czech Republic, 2004, pp:103-111.

[2] Collins Retal. "A system for video surveillance and monitoring: VSAM final report". Carnegie Mellon University: Technical Report CMU, Pittsburgh USA, 2000.

[3] Barron J, Fleet D, Beauchem in S. "Performance of optical flow techniques". International Journal of Computer Vision, Vol.12 (1), pp:42-77, 1994.

[4] C.Stauffer, W.E.L. Grimson. "Learning Patterns of Activity Using Real-Time Tracking". IEEE Transactions on Pattern Analysis arid Machine Intelligence, Vol.22(8), pp:747-757, 2000.

[5] ZHOU Xihan, LIU Bo, ZHOU Heqin. "A Motion Detection Algorithm Based on Background Subtraction and Symmetrical Differencing". Computer Simulation.Vol. 22(4), pp:117-119, 2005.

[6] YANG Li, LI Yushan, et al. "Detection Contours of Multiple Moving Objects with Complex Background". Institute of Electronics, Chinese Academy of Sciences, 2005, 2(2): 306-309..

[7] Cucchiara R, Grana C, Piccardi M, et al. "Improving shadow suppression in moving object detection with HSV color information". In Proceedings of IEEE Intelligent Transportation Systems conference. Oakland, CA, USA, 2001: 334-339.