

Real-time Tracking in Image Sequences based-on Parameters Updating with Temporal and Spatial Neighborhoods Mixture Gaussian Model

Hu Haibo, Zhao Hong

Abstract—Gaussian mixture background model is widely used in moving target detection of the image sequences. However, traditional Gaussian mixture background model usually considers the time continuity of the pixels, and establishes background through statistical distribution of pixels without taking into account the pixels' spatial similarity, which will cause noise, imperfection and other problems. This paper proposes a new Gaussian mixture modeling approach, which combines the color and gradient of the spatial information, and integrates the spatial information of the pixel sequences to establish Gaussian mixture background. The experimental results show that the movement background can be extracted accurately and efficiently, and the algorithm is more robust, and can work in real time in tracking applications.

Keywords—Gaussian mixture model, real-time tracking, sequence image, gradient.

I. INTRODUCTION

MOVING object tracking and analysis are widely used in image processing, computer vision, pattern recognition, and so on. How to design an effective and efficient tracking algorithm has become a hot research topic. Moving object detection is a core part of the whole tracking problem, and the main approaches are frame difference, optical flow and background subtraction.

Frame difference approach [2] compares the grayscale or gradient information between different frames, which has strong adaptability for dynamic scenarios, but can not extract full moving foreground in case that the object moves slowly or adjacent frames overlap with each other.

Optical flow approach [3] has high detection accuracy through combining time and space information, which can detect the moving object even though the camera is moving; however, its high computational complexity makes it difficult to realize this algorithm, and affects real-time moving object detection.

Background subtraction approach carries out statistics for the video sequences and obtains a robust scenario ground through subtracting background image from current image.

Hu Haibo is with the School of Software engineering, Chongqing University, 400030 P.R. China (phone: 86-23-65111025; fax: 86-23-65111025; e-mail: hbhu@cqu.edu.cn).

Zhao Hong is with the School of Software engineering, Chongqing University, 400030 P.R. China (e-mail: h.zhao@cqu.edu.cn).

This algorithm generally can provide the most complete characteristic data, work very fast, and meet requirements of real-time system. However, dynamic scenario changes will result in great difference between the extracted moving object and the target, therefore, the background needs to be updated continuously.

With many merits like simple, high real-time, etc., the background subtraction approaches have been extensively researched and widely applied. In the background subtraction approach, established background is expected to adapt to light changes, overcome the target occlusion and shadows, capture multiple moving targets, and recognize slow-moving targets, and capture the sudden intrusion and loss of the objects. Pfander et al [1] use the single Gaussian model to simulate background, and establish tracking system for human beings, this approach has a very good tracking performance in indoor scenarios, but it can not work efficiently in complex multi-peak problems, such as swinging leaves, sparkling lake, waving flags and so on.

Stauffer et al [4] use Gaussian mixture background model which is based on the mixture modeling of pixel sequences, and fully considers the time continuity of the pixels, but does not combine the information in the spatial neighborhood of the pixels, resulting in accumulated incorrect detection and increased noise. Elgammal et al [5] propose a non-parametric Gaussian core model for static scenarios, and divides the background models into long-time background model and short-time background model. As two different approaches are used when updating the parameter, slow movement impact in [4] can be solved, and a lookup table is used to simplify calculation. However, the relationship between the pixel and its neighboring pixels is not considered, incorrect detection continues to accumulate. Javed et al [6] use a Gaussian mixture model that combines color information and gradient information, obtains contour points of the moving objects through gradient model, fully employs gradient model robustness to noise, and accurately extracts the moving objects through block processing and combining color information, but it is very complex in computation, does not consider spatial coherence, and can not detect effectively and the noise effect still exists. Zhang et al [7] establish a new moving foreground detection algorithm for dynamic background subtraction by combining pixel spatial information. This approach considers pixel spatial information, and is able to detect moving

foreground in moving screens, however, as it does not take pixel time continuity into account, detection errors still exist.

According to the shortcomings of the above approaches, a new spatial Gaussian gradient color model based on Gaussian mixture background is proposed in this paper. The algorithm, which has been well applied in moving tracking, obtains moving object through background modeling that combines spatial gradient model and the Gaussian spatial color model. The new algorithm can not only effectively extract accurate motion background, but also has good real-time performance and robustness.

II. BACKGROUND MODELING

A. Image Sequence Pretreatment

For a particular pixel in the video image, pixel sequences are formed in time domain, which can be described as:

$$\{X_1, X_2, \dots, X_t\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\} \quad (1)$$

Where I denotes a particular image,
 t is a concrete frame, and,
 X_t is the pixel value in a frame.

Because there are often many noise points in the image, the foreground detection result is not satisfactory. Therefore, Gaussian filtering, median filtering and other denoising algorithms have been proposed to filter the noise, but Gaussian smoothed filter will blur image details when filtering the noise, and makes it difficult to detect certain small moving objects; median filter is mainly applied to filter "Salt and Pepper" noise with a limited role. In this paper, a simple three-order filter operator in (2) is adopted to filter most noise while preserving edge details.

$$sme(x) = \frac{1}{9} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 2 & 2 \end{bmatrix} \quad (2)$$

B. Parameter Space-time Neighborhood Method for Gaussian Mixture Background Model

Stauffer [2] proposes a Gaussian mixture model for background modeling, which is as follows:

$$P(X_t) = \sum_{i=1}^K w_{i,t} * \eta(x_t, u_{i,t}, \Sigma_{i,t}) \quad (3)$$

Where K is the number of Gaussian distribution, $w_{i,t}$ is the weight, $u_{i,t}$ is the i^{th} mean value of time t , $\Sigma_{i,t}$ is the i^{th} covariance of time t , η is Gaussian density function that is defined as (4),

$$\eta(x_t, u, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t - u)^T \Sigma^{-1} (x_t - u)} \quad (4)$$

With the color independence assumption, covariance can be defined as $\Sigma_{k,t} = \sigma_k^2 I$ (σ_k^2 is the covariance of k^{th} Gaussian distribution). If a pixel value x meets the condition $|x_t - u_{k,t}| < 2.5 * \sigma$, then x_t matches k^{th} Gaussian distribution at time t . Assign $M_{k,t} = 1$, other parameters are updated according to the following formula:

$$w_{k,t} = (1 - \alpha)w_{k,t-1} + \alpha(M_{k,t}) \quad (5)$$

$$u_t = I(x, y, t) = \begin{cases} (1 - \rho)u_{t-1} + \rho X_t, & \text{if } (M_{i,t} = 1) \\ u_{t-1}, & \text{else} \end{cases} \quad (6)$$

$$\sigma_t^2 = \begin{cases} (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - u_t)^T (X_t - u_t), & \text{if } (M_{i,t} = 1) \\ \sigma_{t-1}^2, & \text{else} \end{cases} \quad (7)$$

$$\rho = \alpha \eta(x_t | u_k, \sigma_k) \quad (8)$$

If a pixel does not match all of the distributions, we can take it as foreground point, and then create a new model to update the original one, take the recent distribution of pixel value as the new mean, and initialize a large variance and small weight. To simplify the calculation, this paper uses the method in [7] to build a lookup table: as for the updated hybrid model, if a pixel always match one of the K distributions with time, that is, $M_{k,t} = 1$, w will increase and σ will decrease continuously, you can obtain Gaussian background model and detected foreground by sorting w/σ and re-initialize w based on (2).

$$B = \arg \min_b \left(\sum_{k=1}^b w_k > T \right) \quad (9)$$

Where $1 \leq b \leq K$, and T reflects the threshold that the model has created. If T is too small, the model often uses a single distribution, and the model often uses the model of the best distribution; If T is too large, the model often uses multiple distributions, which is not robust to shivering leaves and Lake Scenario and so on.

In [4], a color information based Gaussian mixture background model is proposed, which fully considers the time continuity of the pixels, but ignores the spatial coherence. However, a pixel and its neighborhood pixels are not independent; their relations are shown as the velocity of moving objects and the similarity of color values. If a pixel is detected as background point, and the neighborhood pixels are foreground, the detection reliability needs to be considered. In

the Gaussian mixture model, as incorrect detection is not considered, which will result in incorrect detection accumulation, and affect moving target detection results. Therefore, this paper establishes a Gaussian mixture model by adding spatial information. Some spatial information is shown in Fig. 1, where black points represent foreground and white points represent background, and nine points above are the pixel distribution categories of the previous frame.

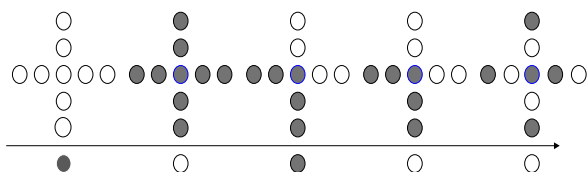


Fig. 1 The relation between pixels in the spatial and temporal neighborhood system

For the moving objects detected in the previous and posterior frames, because the relationship between spatial pixels is ignored, many isolated noise points are regarded as background, or some foreground points are taken as background. To fix such errors and prevent error accumulation, some spatial information is used in this paper. In Gaussian mixture modeling process, if a pixel does not match any distributions, then take it as a foreground point, and the matched points are called background points and denoted as follows: if pixel $I(x, y, t)$ is foreground, then it can be denoted as $I=1$, where $I \in PF$; if pixel $I(x, y, t)$ is background, then it can be denoted as $I=0$, where $I \in PB$, background and foreground set meet the condition $PB \cap PF = \emptyset$. In the current frame, they are updated in the following conditions:

If the test results of a pixel are different in the previous frame and current frame, that is, $I(x, y, t) * I(x, y, t-1) = 0$, the method used to update is:

$$R_1 = I(x-2, y, t-1) + I(x-1, y, t-1) + I(x+1, y, t-1) + I(x+2, y, t-1) \quad (10)$$

$$R_2 = I(x, y-2, t-1) + I(x, y-1, t-1) + I(x, y+1, t-1) + I(x, y+2, t-1) \quad (11)$$

$$I'(x, y, t) = \begin{cases} 1, & \text{if } (R_1 > thresh1) \& (R_2 > thresh2) \\ 0, & \text{else} \end{cases} \quad (12)$$

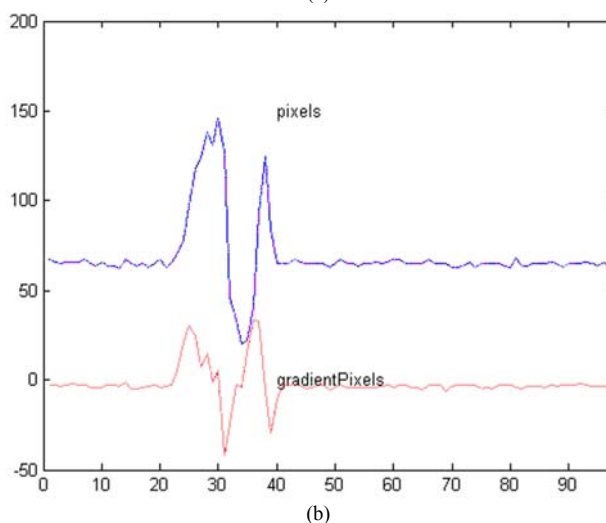
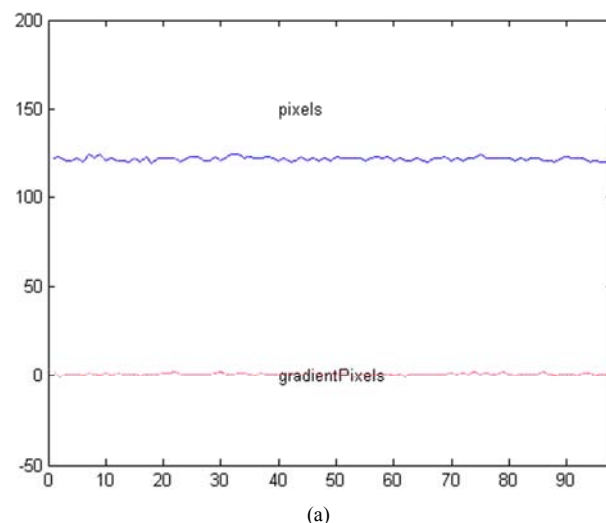
If the test results of a pixel are the same in the previous frame and current frame that is, $I(x, y, t) * I(x, y, t-1) = 1$, the method used to update is:

$$I'(x, y, t) = \begin{cases} 1 & \text{if } ((R_1 + R_2) > thresh3 \& I(x, y, t) = 0) \\ 0 & \text{if } ((R_1 + R_2) < thresh4 \& I(x, y, t) = 1) \end{cases} \quad (13)$$

Where $I(x, y, t-1)$, $I(x, y, t)$ and $I'(x, y, t)$ denote the pixel values detected in $t-1$ and t frame, as well as the updated t frame respectively, and thresholds are set to $thresh1=2$, $thresh2=2$, $thresh3=6$ and $thresh4=2$. Thus, incorrect detection is corrected, and the spatial adjacent pixels have a strong similarity. If the foreground point of a pixel is detected as background point when updating the model, use $u_{k,t} = I(x, y, t)$, $\sigma_t = 15$, $w_{k,t} = \frac{1}{\|I(x, y, t) - u_{k,t}\|^2}$ to update the model; If the background point of a pixel is detected as foreground point when updating the model, use $u_{k,t} = u_{k,t-1}$, $w_{k,t} = (1-\alpha)w_{k,t-1}$, $\sigma_{k,t} = \|I(x, y, t) - u_{k,t}\|^2 - C$ to update the model, where C is a constant.

C. Gradient Gaussian Mixture Background Model

Although gradient Gaussian mixture background model can remove most of the isolated noise points, as well as the holes in the connected parts, it does not have perfect moving object boundaries. As the gradient reflects the boundary information of moving objects, experiments show that it is also a Gaussian distribution. The statistics of the different types of pixels in gradient sequence and time sequence are shown in Fig. 2.



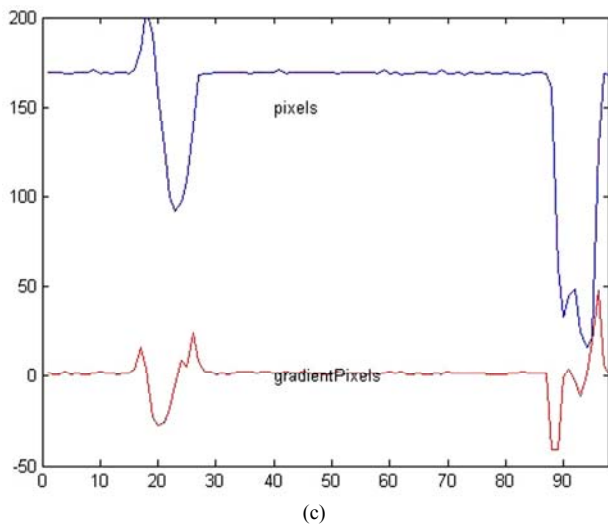


Fig. 2 Comparison of gradient & pixel sequence for moving object

A Gaussian mixture model for gradient sequence is established in this paper, which can detect the contour of a moving object, and obtain the optimized foreground by combining aforementioned spatial color model.

$$f_{x,y} \sim \frac{1}{n(2\pi)^{\frac{d}{2}}} \sum_{i=1}^K \|\Sigma_x + \Sigma_y\| \exp\left(\frac{1}{2}(\mu - \mu_i)^T (\Sigma_x + \Sigma_y)^{-1} (\mu - \mu_i)\right) \quad (14)$$

Where $\mu_{x,y}$ and $\sigma_{x,y}$ denote the mean and variance of gradient sequence. When $|g_{i,j} - \mu_{x,y}| \geq \beta \rho_{x,y}$, and parameter $\beta=1.9$, the current pixel is recorded as foreground, otherwise, it is recorded as background. In this paper, Gaussian spatial pixel model does not need to be updated, but only the mean value and the variance of the gradient model need to be updated.

$$\mu_{x,y} = \mu_{x,y} + \frac{1}{M} (g_{i,j} - \mu_{x,y}) \quad (15)$$

$$\sigma_{x,y}^2 = \sigma_{x,y}^2 + \frac{1}{M} ((g_{i,j} - \mu_{x,y})^T (g_{i,j} - \mu_{x,y}) - \sigma_{x,y}^2) \quad (16)$$

Carry out morphological erosion and dilation for the moving foreground obtained from Gaussian gradient model, thus some of the non-moving object contours are removed. Next, integrate spatial information with the gradient information to optimize the boundary of Gaussian space-time model, and ensure that Gaussian mixture background model with more correcting nature, moving objects with more perfect connectivity and more significant resistance to noise. In the experiment, Gaussian mixture background model establishes different moving boundaries for application scenarios. We can obtain a better result when the matched threshold is $gradient_threshold = 8$ through a large number of experiments.

III. TRACKING APPLICATION

In a tracking system, detection foreground is often affected by light and shadow, and detected as a foreground, which will affect tracking results, so it is essential to find an effective way to remove the shadow. Current shadow removal methods are mainly based on geometric model and shadow feature. Geometric method is based on prior information, such as moving target illumination, to remove the shadow, but it is usually used in fixed application scenarios; shadow feature based approach is based on the geometric characteristics of the shadow, such as brightness, color, texture information to differentiate shadow, background and foreground. When background subtraction approach is used, shadow is usually cast. As RGB color space is element correlated, shadow influence to the image can not be reflected. In this paper, the threshold is adjusted based on the method in [11], and normalized new color space is used to replace traditional color space to remove shadow. Experiments show that the new algorithm can work effectively.

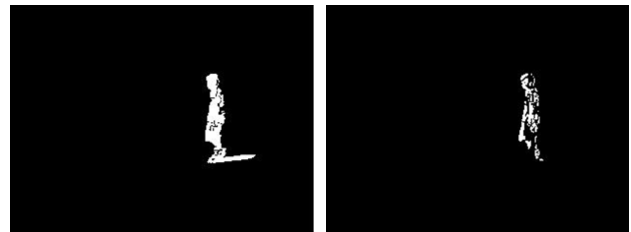


Fig.3 Foreground image with shadow pruned

The new algorithm is mainly used in movement tracking. Firstly, convert the given video or captured video to image sequences, filter with the filtering algorithm to smooth the image and reduce noise, then use the improved Gaussian mixture background proposed in this paper to simulate background, or detect the moving foreground. Exterminate optimized foreground from aforementioned shadow by using morphological filter for detected foreground; carry out statistical analysis for the moving objects for detected foreground after shadow detection, and establish Kalman tracking model relating to the data. To solve the complicate problems caused by multiple moving objects and occluders, a Kalman tracking model for data association is proposed by using multiple hypothesis approach in [5], and Kalman model is used to predict the moving object's position of the next frame and search for the optimal value by combining Mean Shift Algorithm, finally, carry out target behavior analysis for tracking objects.

IV. EXPERIMENT RESULTS AND ANALYSIS

To demonstrate the effectiveness of the algorithm, continuous detection and tracking simulations are carried out for sequence image. Test samples are taken from a 320×240 AVI video shot in a road scenario, and the frame rate is 30fps.

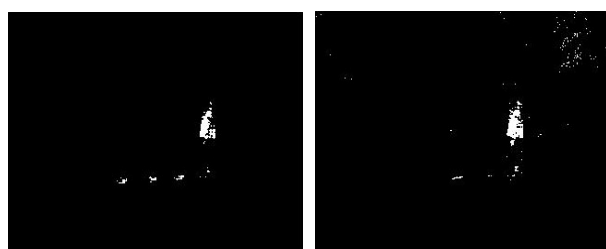
Preprocess the image first by using the three-order filter operator in (2) to filter most noise while preserving edge

details, as shown in Fig. 4, (a) is the background extracted by using the filtering and smoothing algorithm proposed in this paper, and (b) is the result that is not filtered.

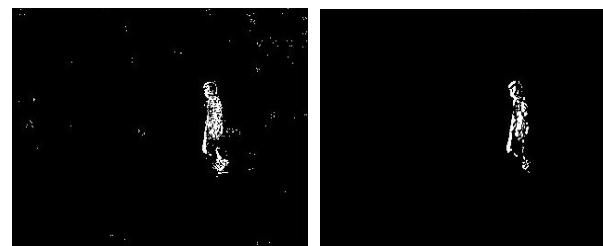


(a) with filter (b) without filter
Fig. 4 Comparison of foreground images

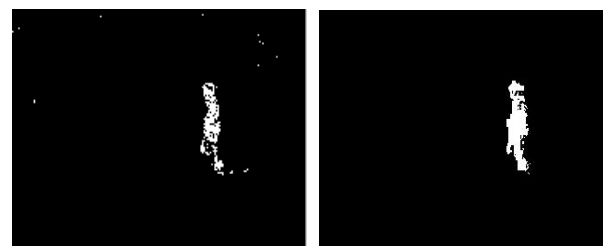
Thus different methods are used to establish background model, and extract the moving objects. It can be easily found that the method used in this paper has better real-time performance and robustness by comparing the different methods, the experiment results are shown in Fig. 5: (a) is a background obtained by calculating the mean value of 30 frames, then subtracting the third frame to obtain foreground, this approach has more noise. (b) is the background obtained by calculating the median value of 80 frames, then subtracting the third frame to obtain foreground, and the differential threshold used is 15. The theoretical analysis and experimental results show that median background approach can detect perfect foreground more easily than mean value approach, but it can not handle the situations that the moving object is similar to the background pixel values. (c) uses a single Gaussian background model, and utilizes the previous 10 frames to establish the model, and then update, where the Gaussian probability threshold used is 0.1. As shown in the upper right corner of the figure, this approach can not work effectively for scenarios like shivering leaves and so on. (d) uses a multiple Gaussian background model to obtain the foreground, and utilizes the previous 10 frames to establish the background model, Gaussian distributions number $K=3$, $D=2.5$, the variance $\sigma=6$, background estimation threshold $T=1.9 \cdot \sigma$. As time changes, incorrect detections will continue to accumulate, and the impact of noise points will be increased. (e) uses a nonparametric Gaussian core model, this approach can work more effectively, and fill the foreground holes extracted by Gaussian mixture model. However, as the spatial information is not integrated into the model, and the treatments for the boundaries are not enough, the foreground will still be affected by the noises, (f) is the result of this algorithm.



(a) mean value method (b) median method

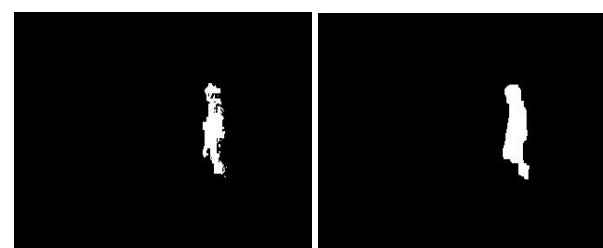


(c) single Gaussian background (d) multi-Gaussian background



(e) Non-parametric background (f) parametric background
Fig. 5 Methods comparison for foreground detection

In our experiments, the selected Gaussian distribution numbers in the color Gaussian background are $K=3$, $D=1.9$ and $\sigma=6$. There are often some glitches in the moving foreground obtained, but the detected boundaries become more perfect by using the gradient model, and a perfect foreground object obtained by combining the color information, gradient and spatial morphological operations, as shown in Fig. 6.



(a) without gradient model (b) with gradient model
Fig. 6 Foreground image obtained with gradient

Carry out post-processing for the moving targets obtained from above experiments, and then use Kalman filtering algorithm mentioned above to track the moving objects. Since this paper is for real-time systems, original Kalman algorithm is improved to solve the complicate problems caused by multiple moving objects and occluders, and a variety of assumptions are added in this paper. As shown in Figure 7, experimental results from left to right are moving scenario, background, foreground and the tracking effect, we can easily find that the new algorithm can work more effectively. Red line denotes the moving object region detected from the foreground, and green line is the result obtained by using the Kalman tracking. Experiment I is the 46th frame of a video, which reflects the tracking effect of the new algorithm when the color of moving object and background is very close. Experiment II is the 452nd frame of a video, the result shows that tiny object can still be effectively detected and tracked.

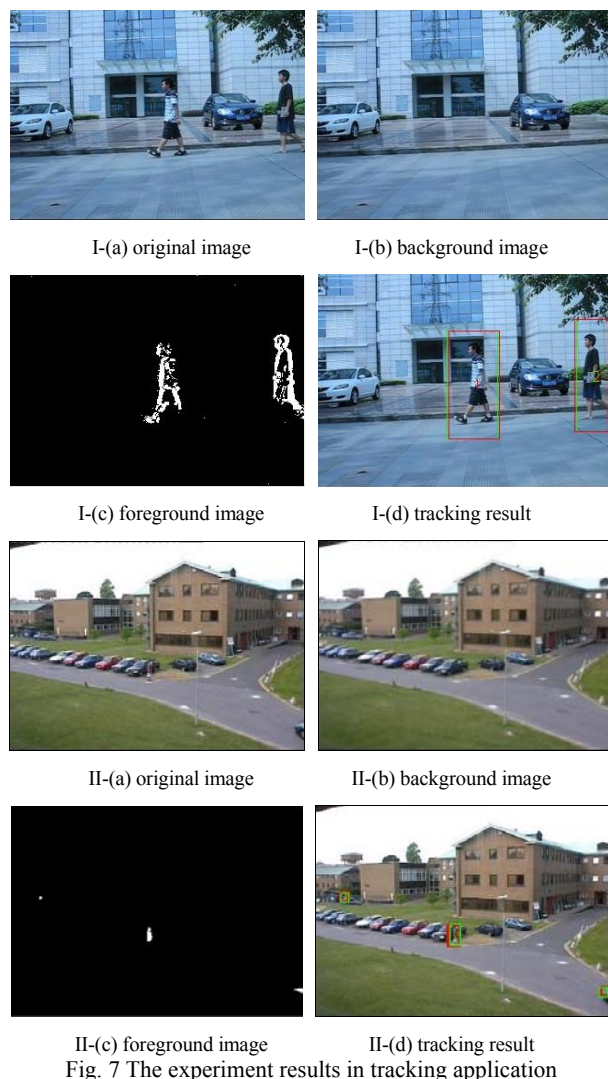


Fig. 7 The experiment results in tracking application

V.CONCLUSION

As the traditional Gaussian mixture model ignores pixel spatial relations, a new Gaussian mixture model is proposed in this paper. The new algorithm combines the spatial information and gradient information of the image to establish Gaussian mixture model, thus combines Kalman to establish moving tracking relating to the data. The model effectively solved the influence of the lighting change, multi-target movement and the disappearing, mixing and shading of movement objects to the tracking effects. The new algorithm is simple, fast, and consistent with the real-time tracking requirements in video surveillance systems and other practical applications. Experimental results prove the effectiveness and robustness of the algorithm.

REFERENCES

- [1] Wren, R. Christopher, A. Azarbayejani, T. Darrell and A. Pentland, "Pfinder: real-time tracking of the human body", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(7), pp.780-785.
- [2] A. Lipton, H. Fujiyoshi and R. Patil, "Moving target classification and tracking from real-time video," in Proc. of IEEE Workshop Application of Computer Vision, Los Alamitos, CA, Oct. 1998.
- [3] A. Shafie, F. Hafiz, and M. Ali, "Motion detection techniques using optical flow," *World Academy of Science. Engineering and Technology*, 2009, vol.56, pp.559-561.
- [4] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in Proc of IEEE conference on Computer Vision and Pattern Recognition, 1999, pp.246-252.
- [5] A. Elgammal, D. Harwood and L. Davis, "Non-parametric model for background subtraction," *ECCV 2000 in computer vision*, 2000, pp.751-767.
- [6] O. Javed, K. Shafique and M. Shah, "A hierarchical approach to robust background subtraction using color and gradient information". in Proc. Of IEEE Workshop Motion Video Computing, Dec. 2002, pp.22-27.
- [7] S. Zhang et al, "Dynamic background modeling and subtraction using spatio-temporal local binary patterns," in Proc. of IEEE Int. Conf. on Image Processing, 2008, pp.1556-1559.
- [8] C. Ridder, O. Munkelt and H. Kirchner, "Adaptive background estimation and foreground detection using Kalman-filtering," In Proc. of International Conference on Recent Advances in Mechatronics, ICRAM'95, 1995, pp.193-199.
- [9] Stauffer C, Grimson W. E. L. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 747-757.
- [10] K. Toyama, J. Krumm, B. Brumitt and B. Meyers. "Wallflower: principles and practice of background maintenance," in Proc. of the Seventh IEEE International Conference on Computer, IEEE CS, 1999, pp.255-261.
- [11] Wren, Christopher R, Ali Azarbayejani, Trevor Darrell and Alex Pentland. Pfinder: Real-Time Tracking of the Human Body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, July 1997, 19(7): 780-785.
- [12] D. Kim and K. Hong, "Practical background estimation for mosaic blending with patch-based Markov random fields," *Pattern Recognition*, 2008, 41(7), pp.2145-2155.
- [13] T. Chalidabhongse, K. Kim, D. Harwood and L. Davis. A perturbation method for evaluating background subtraction algorithms. In Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005, pp.15-16.
- [14] R. Cucchiara, C. Grana, M. Piccardi and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, pp.1337-1342.
- [15] X. Gao, T. Boulton, F. Coetzee and V. Ramesh, "Error analysis of background subtraction," In Proc. of International Conference on Computer Vision and Pattern Recognition, 2000, pp.503-510.
- [16] M. Heikkila and M. Pietikainen. "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 2006, 28(4), pp.657-662.
- [17] J. Pilet, C. Strecha and P. Fua, "Making Background Subtraction Robust to Sudden Illumination Changes," In Proc. Of European Conf. on Computer Vision, 2008, pp.567-580.
- [18] M. Law, M. Figueiredo and A. Jain, "Simultaneous feature selection and clustering using a mixture model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2004, 26(19), pp.1154-1166.