International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.9, (2016), pp.89-102
http://dx.doi.org/10.14257/ijsip.2016.9.9.09

Moving Targets Detection Based on Moving Saliency Calculation

Wang Hui-bin*, Du Min, Chen Zhe and Lu Miao

College of Computer and Information Engineering, Hohai University, Nanjing211100, China
hbwang@hhu.edu.cn

Abstract

Moving targets saliency extraction is the key technology in moving target detection, tracking and recognition systems. In order to solve the problems for moving targets detection tasks, such as the background interference, target occlusion and so on, a new moving target detection method by moving saliency calculation is proposed. By the Quaternion Fast Fourier Transform, the proposed method fuses four motion features as optical flows, direction vectors and frame difference. The moving saliency map, as the result, is generated. The experimental result demonstrates that our method has better ability to handle the problems in moving object detection, such as moving information loss in complex scene. The results with better accuracy score(Cgood=0.875， Cfalse=0.028) are given by our method.

Keywords: *moving target detection; moving saliency map extraction; Phase spectrum; motion feature fusion*

1. Introduction

Moving target detection is the key technology for intelligent video processing such as target tracking and target recognition. Classical methods mainly include frame difference method, background subtraction and optical flow method [1-6]. However, complex background and noise would affect the performance of methods by using the inter-frame motion characteristics in natural scene. Moreover, varied background information will significantly decrease the accuracy of target detection method based on background modeling, as high missing rate and misjudgment rate. Therefore, inhibiting the background information and maximally stretching the contrast of target-background is the key for moving target detection researches.

Due to the high background suppression and computational efficiency, the saliency based background suppression and target detection methods have been widespread concerned, such as Itti classical model proposed by Itti and Koch [7]. This model simulates human visual search process, normalizing variety of low-level features through central-surrounding operator to extract the local contrast of features. These features are merged into static saliency map. Xiaodi Hou [8] proposed a residual spectrum model to remove redundant background information and detect the proto objects by solving the residual spectral of images. However, the saliency based moving object detection methods are more widely applied to the single-frame static scene and static targets. For moving object detection, moving information should be considered to be fused into saliency calculation. Mahadevan [9] fused spatial information with moving information by the Itti model. Although it can be adapt to the shaking background, the large computational complexity generally blocks the application in practice. To solve this problem, Cui [10] proposed a saliency extraction method based on time slice spectral analysis. It can meet real-time requirements based on Residual spectral model. However, if the video sample is

---

*Corresponding Author

ISSN: 2005-4254 IJSIP

short and the moving information is missed, the extracted saliency is less and weak. It is difficult to distinguish between the target and background. Guo Chenlei [11] introduced the inter-frame difference feature into the PQFT model to detect moving targets. However, the inter-frame difference is only as the feature to compensate static characteristics such as color and gray-scale. Due to less moving information, test results include a large number of static salient features, such as texture background interferences. To solve this problem, it is necessary to largely enhance motion features for moving saliency extraction. In addition, the multi-features fusion process improves the robustness of moving saliency extraction.

In order to improve the capacity of the saliency based method to detect moving information and suppress background interference, this paper proposes a saliency calculation method by fusing multi motion features to detect moving targets.

2. Moving Saliency Calculation Model

Static saliency map is mainly obtained by calculating the intensity, color, orientation, *etc.*, The fusion result is not able to describe moving information between frames. For fusing moving information into the saliency calculation process, this paper establishes a novel Moving Saliency Map (MSM) based on the moving saliency calculation. Different from static saliency map, it generates saliency map only contain moving saliency for moving targets. This model contains 3 modules as feature fusion calculation, phase spectrum calculation and saliency calculation. This model uses four motion features as input. These motion features are fused to describe the moving properties. We use the phase spectrum calculation to analyze the statistical regularities in the frequency domain and obtain the changes in the time domain. The saliency calculation process analyses the frequency spectrum to restore the motion feature in the time domain and then obtain the moving saliency map. The Moving Saliency Calculation process is shown in Figure 1.

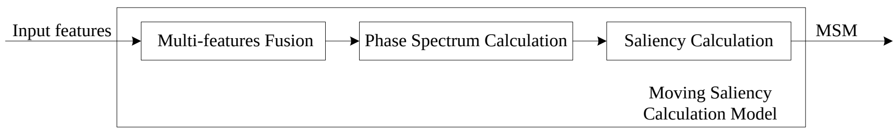


Figure 1. Moving Saliency Calculation Process

1.1. Multi-Features Fusion

Single motion feature can hardly describe the complete moving information of targets. Taking the multi-features fusion strategy into account, this paper selects three motion features as frame-difference, optical flow and direction vector as the input. The frame difference is a common motion feature extraction method, its advantage lies in the high computational efficiency and the results can reflect the differences between frames. However, due to the corresponding pixel subtraction strategy, the background noise is left in the result if the background changes or shakes. To solve this problem, we take the optical flow. The optical flow estimates the motion vectors of each pixel. By filtering the estimated motion vectors, it can suppress the background noise with small motion vectors in some extent. Finally, the direction vector suppresses the state variable noise in the background. This paper fuses above three motion features into a quaternion feature.

Assuming that transverse optical flow, longitudinal optical flow, direction vector and frame difference feature are $U(t), V(t), \theta(t), M(t)$ the quaternion feature set can be show as follows according to the quaternion formula [11]:

$$q(\mathrm{t}) = M(\mathrm{t}) + U(\mathrm{t})\mu_1 + V(\mathrm{t})\mu_2 + \theta(\mathrm{t})\mu_3 \tag{1}$$

Where $\mu_1^2 = -1$, $\mu_3 = \mu_1\mu_2$ and mutually orthogonal between μ_i 。

The advantage of using quaternion is that multi-features can be processed parallelly with high computational efficiency.

1.2. Phase Spectrum Calculation

For a static image, the phase spectrum curve can reflect the changes which called salient areas in original input signal.

Assuming that the input signal is $f(x)$, the amplitude and phase information given by the Fourier transform is $|F(\omega)|$ and $\theta(\omega)$ respectively, then phase spectrum can be represented as:

$$f_p(x) = IDFT(|F(\omega)| e^{j\theta(\omega)}) \tag{2}$$

If the input quaternion feature set is $q(t)$,then formula(1)can be represented as:

$$q(n,m) = f_1(n,m) + f_2(n,m)\mu_2 \tag{3}$$

Quaternion Fast Fourier Transform can be calculated as follows：

$$F_i[u,v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{1} \sum_{n=0}^{1} e^{-\mu_1 2\pi((mv/N)+(nu/N))} f_i(n,m) \tag{4}$$

Where n,m represent the time domain u,v represent the frequency domain. Assuming that the input image after the QFFT can be expressed as $Q(t)$, the phase spectrum can be calculated in polar coordinates by the following formula:

$$Q(t) = \|Q(\mathrm{t})\| e^{\mu\Phi(t)} \tag{5}$$

When $\|Q(\mathrm{t})\| = 1$, $Q(\mathrm{t})$ only contains the phase information in the frequency domain.

1.3. Saliency Calculation

By analyzing and calculating the spectrum information, we can extract the moving information in frequency domain. Then, this moving information in the frequency domain is required to be projected into the time domain, generating the moving saliency map. This paper transforms the motion information from the frequency domain into temporal domain by Quaternion Fast Fourier Inverse Transform. Assuming the image given by the phase spectrum calculation is as $F_i[u,v]$, then, QFFIT can be represented as:

$$f_i(n,m) = \frac{1}{\sqrt{MN}} \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} e^{\mu_1 2\pi((mv/M)+(nu/N))} F_i[u,v] \tag{6}$$

Assuming the phase spectrum after the QFFIT is $q'(t)$, the final fusion moving saliency map can be expressed as:

$$MSM(t) = g(t) * \|q'(t)\|^2 \tag{7}$$

Where $g(t)$ is the 2D Gaussian filter. $MSM(t)$ is the moving saliency map obtained by the phase spectrum method.

2. Target Detection Process Based on Moving Saliency Calculation

Moving saliency calculation based moving target detection process is shown in Figure 2. This process includes motion feature extraction, moving saliency calculation, additive correcting fusion.

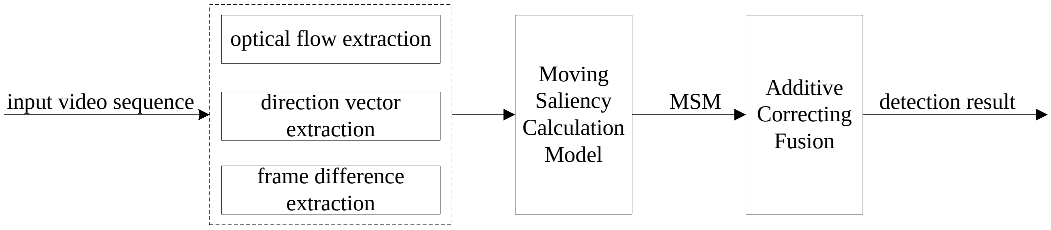


Figure 2. Moving Target Detection Process Diagram

2.1. Motion Feature Extraction

Motion feature selection and extraction is a key technology in our moving target detection method. This paper takes the optical flow, direction vector and frame difference as the input of model. The extraction method of these three motion features are as follows.

(1) Optical Flow

The optical flow feature represents the instantaneous rate of change in the gray value [12-13]. Gradient-based optical flow algorithm is used to get optical flows in x and y directions.

Assuming the gray value is $I(x,y,t)$ at time t . At time $t+\Delta t$, the gray value is $I(x+\Delta x, y+\Delta y, t+\Delta t)$. Since the gray value remained unchanged at time Δt , the optical flow constraint equation is described as follows [14]:

$$I(x,y,t) = I(x+\Delta x, y+\Delta y, t+\Delta t) \tag{8}$$

$$u^{(k+1)} = \overline{u}^{(k)} - I_x \frac{I_x \overline{u}^{(k)} + I_y \overline{v}^{(k)} + I_t}{\lambda^2 + I_x^2 + I_y^2} \tag{9}$$

$$v^{(k+1)} = \overline{v}^{(k)} - I_y \frac{I_x \overline{u}^{(k)} + I_y \overline{v}^{(k)} + I_t}{\lambda^2 + I_x^2 + I_y^2} \tag{10}$$

Where u, v represent optical flows in x and y directions.

(2) Direction Vector

Direction vector reflects the direction of the motion between frames, it is capable of describing the corresponding relationship between pixels of adjacent frames [15].

Assuming the motion vector component in the x axial is V_x^{motion} , y axis direction is V_y^{motion}, then the direction vectors can be obtained as follows:

$$\theta = arctg\left(\frac{V_y^{motion}}{V_x^{motion}}\right) \tag{11}$$

(3) Frame Difference

Difference algorithm is applied to two or more consecutive frames to obtain the frame diffidence vector as one of the input features.

$$M = \left| I(t) - I(t-\tau) \right| \tag{12}$$

Where $I(t)$ represents the gray value at time t , $I(t-\tau)$ represents the gray value after time τ . We can set $\tau = 1$.

2.2. Additive Correcting Fusion

Based on moving saliency calculation model and multi-features fusion method, the moving saliency map is given. However, when moving targets maintain static at a certain frame or several consecutive frames, the target will be regarded as background noise to be suppressed. Therefore, additive correcting fusion strategy is applied.

(1) Recognizing Temporal Static Target

Assuming the moving saliency map at time $k-1$ is $MSM(k-1)$,moving saliency map at time k is $MSM(k)$, the temporal static target can be recognized as follows:

$$M(k) = MSM(k) - MSM(k-1) \tag{13}$$

Where $M(k)$ is the subtractive result of moving saliency map. Assuming the threshold value is T , target can be recognized according to the following criteria:

Criterion 1 if $M(k) > T$ over 3 consecutive frames, then the target is disappeared, stopping testing.

Criterion 2 if $M(k) > T$ within 3 consecutive frames, then the target is temporal static target, correcting should be done to the result.

Criterion 3 if $M(k) \le T$, then the temporal static target is not exist.

(2) Additive Correcting Fusion Method

Assuming the moving saliency map at time $k-1$ is $MSM(k-1)$, at time k , moving target a has a temporal still and the moving saliency map is $MSM(k)$. Additive correcting fusion method is as follows:

$$MSM(k) = MSM(k) \oplus MSM(k-1) \tag{14}$$

Temporal static target can be detected by the fusion of moving saliency map in the time domain. It achieves the integration of static target detection and moving target detection and improves the accuracy of test result.

2.3. Algorithm Flow

Algorithm flow of our method is as follows:

Step1 Input images $I(t)$ and $I(t-\tau)$, getting 4 motion features according to Equation (9-12).

Step2 Fusing three motion features into quaternion feature set $q(t) = M(t) + U(t)\mu_1 + V(t)\mu_2 + \theta(t)\mu_3$ as input of model.

Step3 Using Equation (4-5) to get Phase Spectrum information.

Step4 Getting the moving saliency map by Equation (6-7).

Step5 Detecting the presence of temporal static target by using Equation (13), correcting the moving saliency map by using Equation (14).

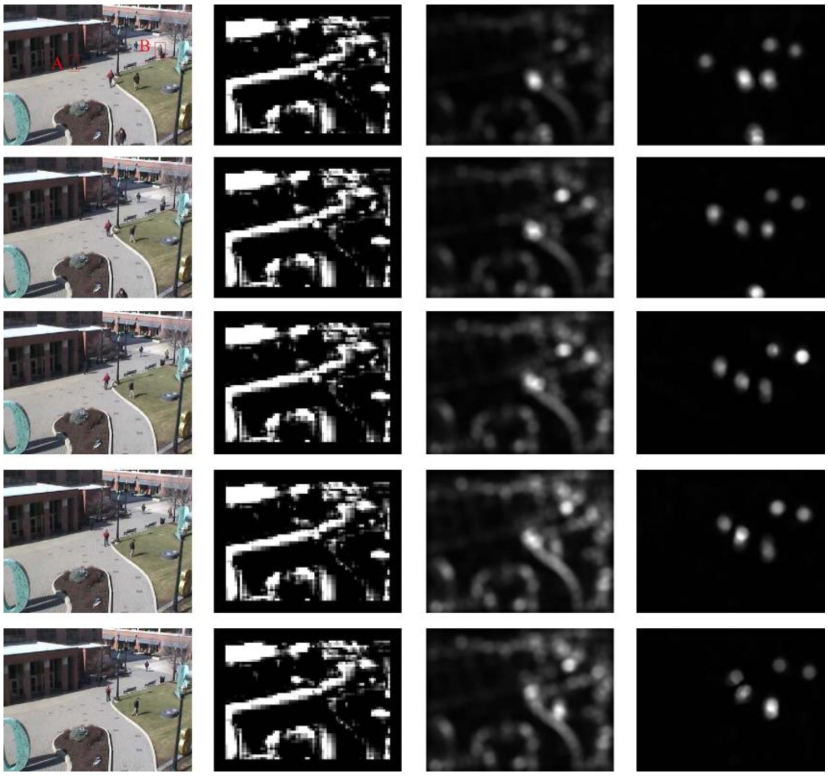
Step6 Getting the final detection result according to the moving saliency map.

3. Experimental Results and Analysis

In order to demonstrate the correctness and feasibility of our method, the paper selected four typical scenes to compare the performance of our method with traditional detection methods.

Experimental platform is PC with 2.5GHz of frequency, 4G of memory, the interval between frames is 0.3s.

3.1. Results Under Different Scenes

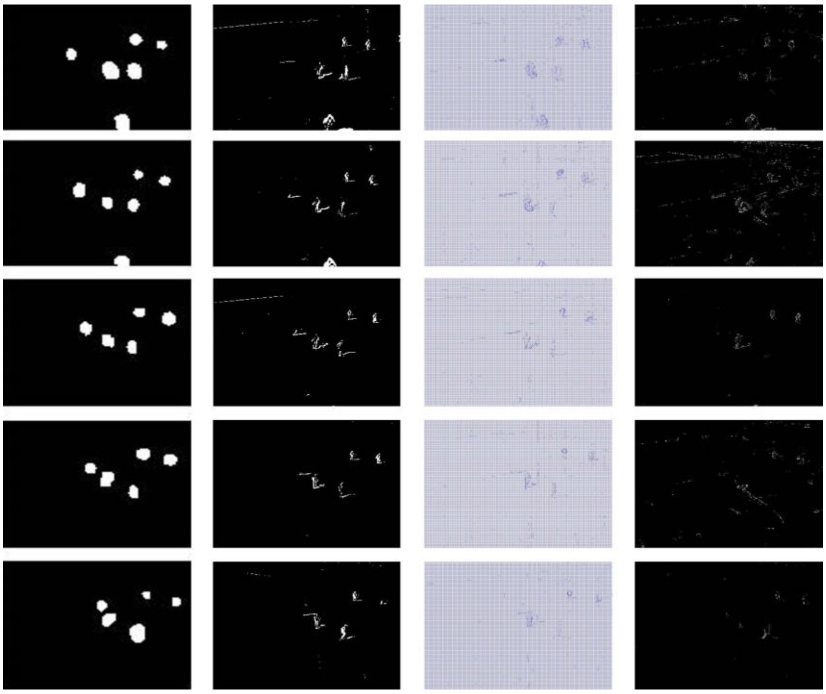


(a) video sequence (b) Itti Model (c) PQFT Model (d) our method

Figure 3. Multi-Target Saliency Detection Results

Figure 3, and 4, are test results given by target detection methods based on salience and traditional methods respectively. In this scene, there are many moving targets. This video is of 320 * 240 pixels. There are six small moving targets. Moreover the occlusion and the shadow exist in this scene. The target on the bottom disappears after a few frames.

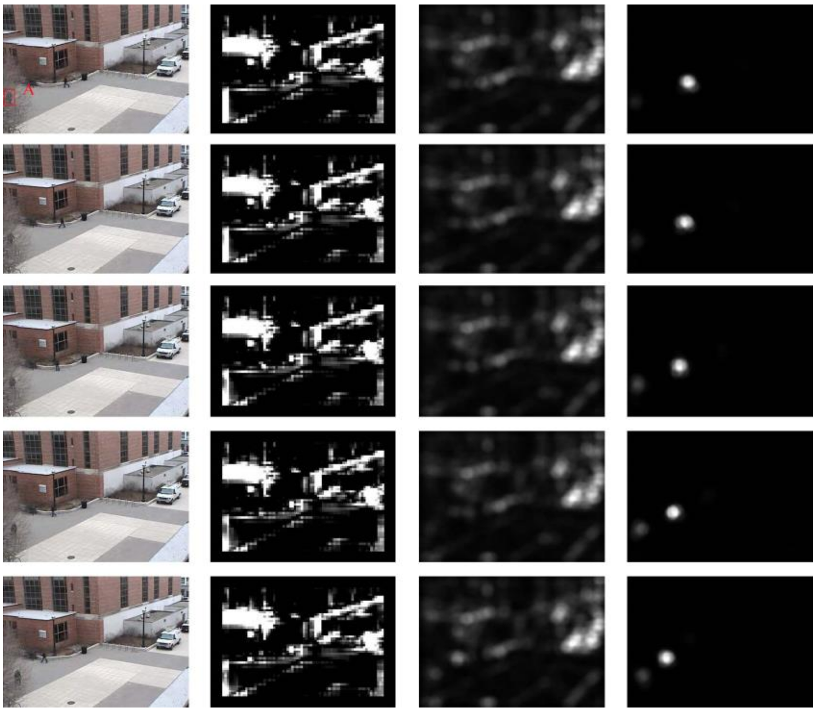
As we can see from the Figure 3, our method is able to accurately detect the target and the background noise is inhibited. Since the moving information is not introduced into the Itti model, the detect results include numerous background noise. By contrast, the moving information is introduced into the PQFT model for compensating the static features. Hence, parts of the moving target are detected by the PQFT, removing the influence of the salient static region.



(e)Our Method (f) Frame Difference (g) Optical Flow (h)GMM

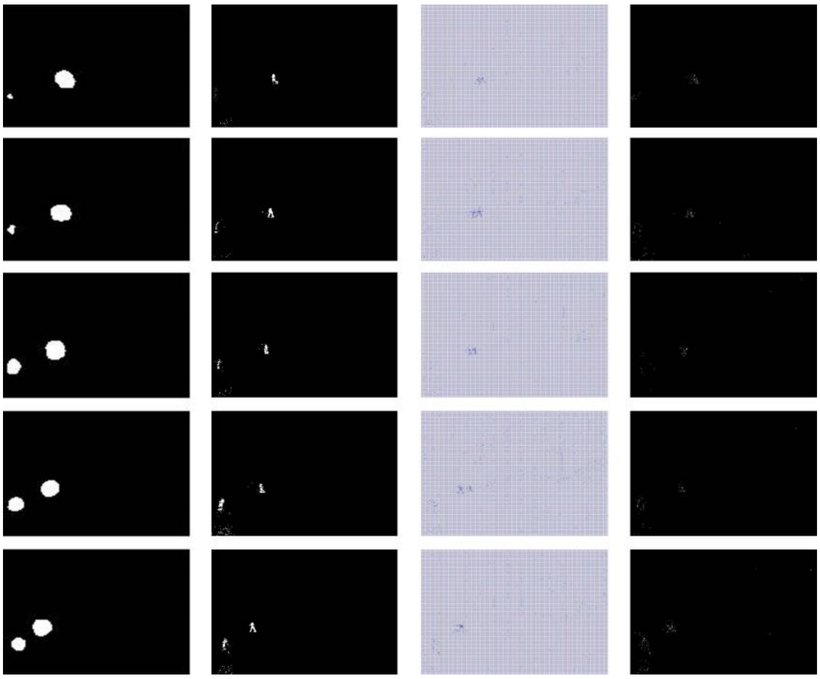
Figure 4. Multi-Target Detection Results

In particular, target A in Figure 3, has been obscured by shadows all the time. Only our method is able to detect the target A. Traditional algorithms as frame difference, optical flow and GMM method can detect only a small portion of the edge of the target B, since this target is blocked in the last frame. By contrast, the target B is completely detected by our method.



(a) Video Sequence (b) Itti Model (c) PQFT Model (d) Our Method

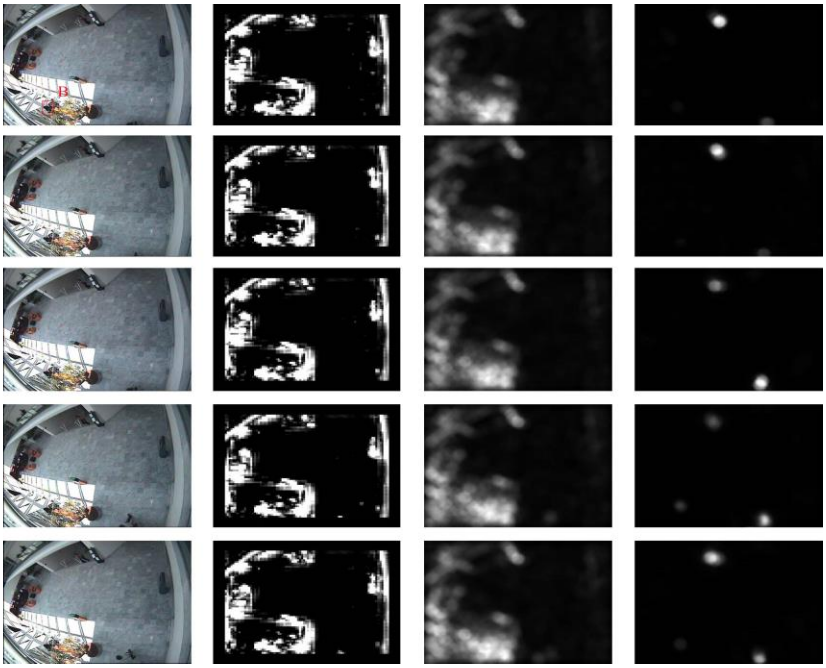
Figure 5. Saliency Detection Results Under the Occlus



(e) Our Method (f) Frame Difference (g) Optical Flow (h)GMM

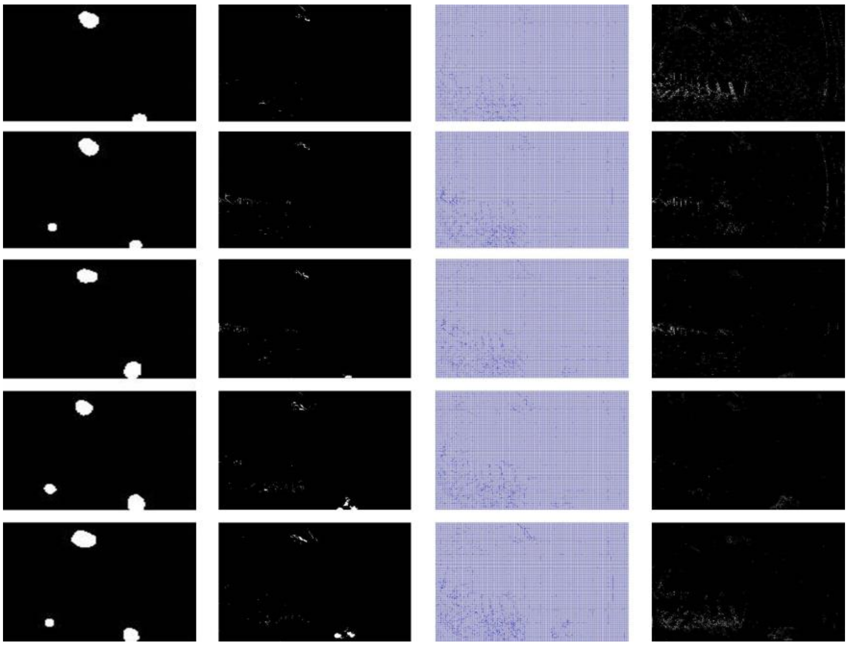
Figure 6. Detection Results Under the Occlusion

Figure 5, and 6, are test results obtained by target detection methods based on salience and traditional methods respectively. The video is of 320 * 240 pixels. Moving targets are pedestrian in left bottom. Target A is covered by tree at the first frame, we can hardly visually recognize the complete outline of the pedestrian. Although the moving saliency is small in our method, it can still detect the target successfully after binary processing. The other algorithms tend to remove it as the background noise and fail to test the target.



(a) Video Sequence (b) Itti Model (c)PQFT Model (d) Our Method

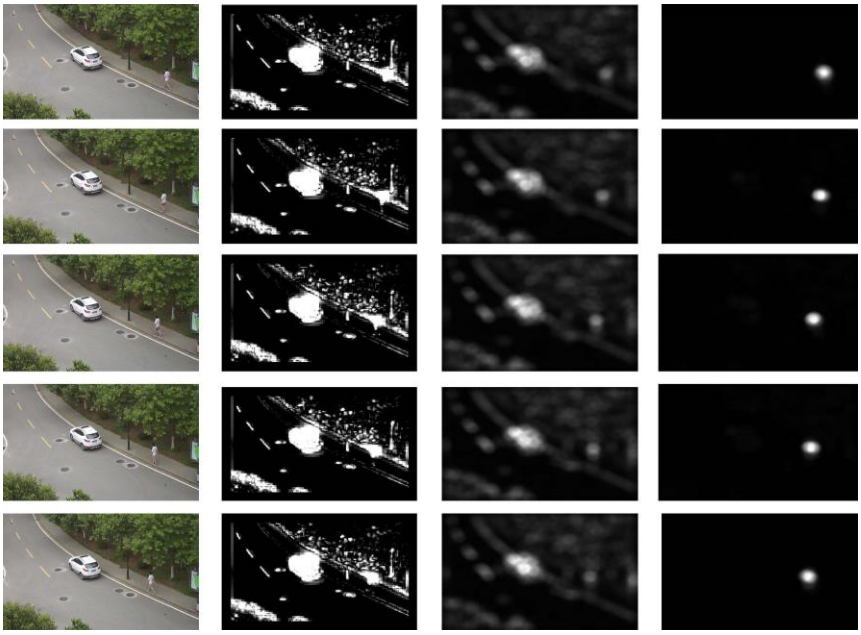
Figure 7. Saliency Detection Results Under Complex Scene



(e) Our Method (f) Frame Difference (g) Optical Flow (h)GMM

Figure 8. Detection Results Under the Complex Scene

Figure 7, and 8, are test results obtained by target detection methods based on salience and traditional methods. This scene includes complicated background. The video is of 320 * 240 pixels. There are 3 targets in the scene. The background in the scene is complex and its color is close to the targets. The target detection is also affected by the brightness change. Targets are small and are similar with the surrounding environment. Besides, target B is covered in the shadow. In this case, our algorithm can successfully detect the non-shaded targets and can detect target B partially. By contrast, all of the 3 targets are missed in the results given by other methods.



(a) Video Sequence (b) Itti Model (c) PQFT Model (d) Our Method

Figure 9. Saliency Detection Results with Slight Shake on Background

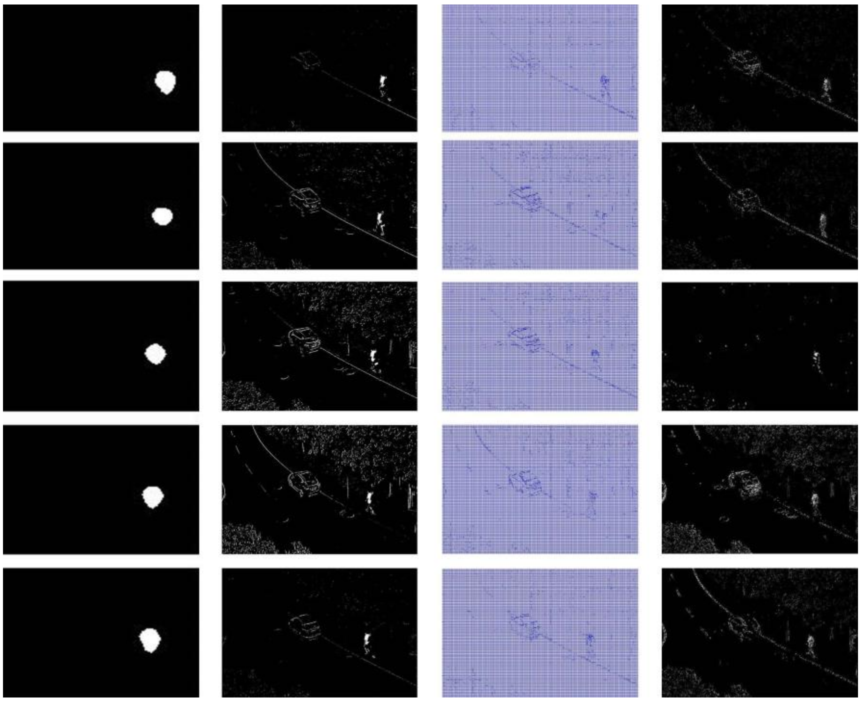


Figure 10. Detection Results with Slight Shake on Background

Figure 9, and 10, are test results obtained by target detection methods based on salience and traditional methods respectively. This scene includes shaking background. The video is of 1080 * 720 pixels. There is only one pedestrian in the scene. However, the shake camera generates the motion noise. We can see from (d) column in Figure 9, our moving saliency map is also affected by the background noise slightly. The Itti model and PQFT model mistake the static car as the moving targets. The same mistake is also given by frame difference method, optical flow method and GMM method. It can be seen from (e) in Figure 10, that our method can suppress the affection of car and detect the pedestrian successfully after binary processing.

3.2. Detecting Temporal Static Target

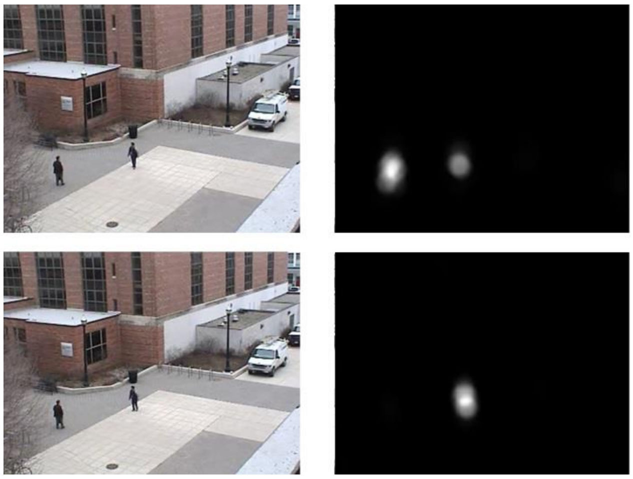
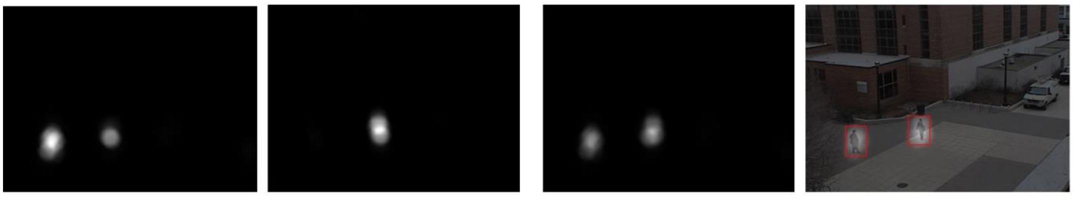


Figure 11. Detection Results with Short Static

It can be seen from Figure 11, that two targets were detected correctly in the previous frame by our method. However, in the next frame, only one target moved. Our test result showed only one target and mistake another target as the background noise.



(a) Previous MSM (b) Present MSM (c) Fused MSM (d) Test Result

Figure 12. Detection Results after Fusion Correction

As we can see in Figure 12, fused moving saliency map can not only detect the moving target, but also can detect the temporal static target accurately which achieves a combination of dynamic and static target detection.

Misjudgment of temporal static target may happen in the following two special conditions: one is the target still time is longer than the threshold; the other is that target is completely blocked within the set time.

3.3. Algorithm Performance Quantitative Evaluation

In order to evaluate the pros and cons of the proposed method, using four indexes to analyze and compare each algorithm used in the tests. Four indexes are Cgood, Cfalse, operation time t and target offset Pi.

$$C_{good} = \frac{card\left\{\Omega_{in} \cap \Omega_o\right\}}{card\left\{\Omega_o\right\}} \tag{15}$$

$$C_{false} = \frac{card\left\{\Omega_{in} \cap \Omega_b\right\}}{card\left\{\Omega_b\right\}} \tag{16}$$

Ω_{in} is the internal domain of extracted target, Ω_o is the real domain of the target and Ω_b is the background. C_{good} is the ratio that correctly extracted target region accounted for the real target area, C_{false} is the ratio that incorrectly extracted target region accounted for the background area.

Target offset refers to the average offset between the position of detected target and the actual target, it is calculated as follows:

$$P_i = \frac{\sum_{i=1}^{n} \left| X(i) - Y(i) \right|}{n} \tag{17}$$

Where n represents the number of pixels, $X(i)$ represents the pixel value of detected target position, $Y(i)$ represents the pixel value of real target position, Pi represents the average target offset.

Table 1. Performance Evaluation of Algorithms

Algorithms	average Cgood	average Cfalse	average operation time (s)	average target offset (pix)
frame difference	0.772	0.084	0.24	2.07
optical flow	0.703	0.102	2.63	1.63
mixture Gaussian	0.765	0.072	2.78	1.44
Itti model	0.261	0.626	3.04	10.63
PQFT model	0.554	0.507	0.45	4.25
Our method	0.875	0.028	0.62	2.78

It can be seen from Table 1, that correct extraction ratio of target area of our algorithm is the largest. Since the algorithm can suppress background noise and non-interested saliency static targets. Hence the error extraction ratio of the target area is lowest. In addition, the method has high computational efficiency. But we also found that due to regional average calculation strategy of target detection method based on saliency detection, the algorithm can’t get an accurate contour of target, so the average target offset is higher than the traditional algorithms. But compared with other methods based on salience, our method has been greatly improved.

4. Conclusion

This paper proposes a target detection method based on moving saliency calculation. The method extracts 3 motion features between adjacent frames of video sequence as quaternion input signal. The moving saliency map is generated by quaternion fast Fourier transform. The experiment result proved that our method can cope with background texture noise, object occlusion and short duration of the moving target and can accurately detect multiple moving targets with high accuracy. Fusion of the moving saliency maps can achieve the moving target detection in the time domain when having temporal static target.

Acknowledgments

I would like to express my gratitude to all those who have helped me during the writing of this thesis. I gratefully acknowledge the help of my supervisor Professor WANG Hui-bin. Also, I would like to thank CHEN Zhe, who kindly gave me a lot of suggestions. Last but not least, my gratitude also extends to my family who have been assisting, supporting and caring for me all of my life.

The work is supported by National Natural Science Foundation of China (No. 61374019, No. 61501173), Natural Science Foundation of Jiangsu Province (No. BK20130851, BK20150824), China Postdoctoral Science Foundation (No. 2015M581710), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-timetracking“, Computer Vision and Pattern Recognition, IEEE Computer Society Conference, **(1999)**.

[2] O. Barnich and M. V. Droogenbroeck, “ViBe: A universal background subtraction algorithm for video sequences“, Image Processing, IEEE Transactions on, **(2011)**.

[3] W. Fengchao, L. Xingtang and H. Shucai, “Target Fusion Detection with Multi-Feature Based on Fuzzy Evidence Theory“, Acta Optica Sinica, **(2010)**.

[4] J. Yang, Y. D. Sun, M. J. Wu and Q. N. Zhang, “Multi-Class Moving Target Detection with Gaussian Mixture Part Based Model“, IEEE International Conference on Consumer Electronics (ICCE), **(2014)**.

[5] C. Yin, R. Kan and G. Guohua, “Moving Object Detection Based on Improved Single Gaussian Background Model“, Chinese Journal of Lasers, **(2014)**.

[6] Z. Huixin, Q. Hanlin and L. Rui, “Multiscale Truncation for Dim and Small Target Background Suppression“, Acta Optica Sinica, **(2010)**.
[7] L. ITTI, C. KOCH and E. NIEBUR, “A model of saliency-based visual attention for rapid scene analysis“, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, **(1998)**, pp. 1254-1259.
[8] X. Hou and L. Zhang, “Saliency Detection: A Spectral Residual Approach“, Conference on Computer Vision and Pattern Recognition.Minnesota:IEEE, **(2007)**, pp. 1-8.
[9] V Mahadevan, Yahoo! Labs, “BangaloreBiologically Inspired Object Tracking Using Center-Surround Saliency Mechanisms“, Pattern Analysis and Machine Intelligence, IEEE Transactions on, **(2013)**.
[10] X. Cui, Q. Liu and D. Metaxas, “Temporal spectral residual: fast motion saliency detection“, In MM, **(2009)**, pp. 617–620.
[11] C. GUO, Q. MA and L. ZHANG, “Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform“, Conference on Computer Vision and Pattern Recognition.Alaska:IEEE, **(2008)**, pp. 1-8.
[12] Y. Niu, A. Dick and M. Brooks, “Locally Oriented Optical Flow Computation“, Image Processing, IEEE Transactions on, **(2012)**.
[13] M. Mueller, P. Karasev, I. Kolesov and A. Tannenbaum, “Optical Flow Estimation for Flame Detection in Videos“, Image Processing , IEEE Transactions on, **(2013)**.
[14] T. Wang and H. Snoussi, “Detection of Abnormal Visual Events via Global Optical Flow Orientation Histogram“, Information Forensics and Securit, IEEE Transactions on, **(2014)**.
[15] A. K. Jain and T. Q. Nguyen, “Discriminability Limits in Spatio-Temporal Stereo Block Matching“, Image Processing , IEEE Transactions on, **(2014)**.

