Multi-Scale Motion Attention Fusion Algorithm for Video Moving Target Detection

Liu Long¹, a and Fan Jiang¹, b
¹Xi’an University of Technology

-liulong@xaut.edu.cn ; bfanjiang@stu.xaut.edu.cn

Abstract: Motion Target detection is a key technology in video analysis, this paper propose a target detection algorithm based on multi-scale motion attention analysis, which provides a new method for motion target detection under global motion scene; Firstly, the noise of motion vector field is removed by filter, and according to the mechanism of visual attention, spatial-temporal motion attention model is built, then the trust degree of motion vector is suggested on the basis of validity analysis of motion vector, and decision fusion of multi-scale motion attention is carried by D-S theory to detect region of motion target; The test result of different video shows the algorithm is effectively for target detection under global motion scene, overcome the defect of traditional algorithm and has better robustness than other algorithms.

1. Introduction
Motion target detection is of important and practical value, which is an indispensable link for identifying a visual target, understanding the nature of its behavior and analyzing video information and is widely applied in civil and military areas; hence one of hot research topics in the field of video processing.

According to the adoption of static or motion video lens, motion target detection can be divided into two types: local detection of motion scene [1-2] and global detection of motion scene [3-10]. Recent years, the visual perception research has gradually taken in the fruits in the field of human psychology and physiology research, particularly those of visual attention research. Itti and Koch [11-12] came up with the concept of attention area and established the model of visual attention with static image features. Yu-Fei Ma [13] defines a model of motion attention comprehensively based on the motion vector energy of motion vector field, spatial and time coherence from MPEG code stream decompression. Zhiwen Yu, etc. [14] regionally classify the computation of visual attention regions in static image so that visual attention regions approximate target regions gradually, thus establishing the relationship of data-driven mechanism between visual attention and high-level semanteme. Junwei Han [15] classifies visual attention into static and motional attention. Yuming Fang, etc [16] put forward a combined top-down-and-bottom-up model of visual attention for detecting artificial moving targets. The model selects and fuses the characters of bottom-up attention of lightness and the top-down directional characters to obtain a target region.

The research reported in [13-16] all concern the computation of visual attention and the application of the mechanism of visual attention to target detecting and segmenting, but the methods may not be suitable for the target detect in global motion scene because of their limitations. An algorithm based on multi-scale motion attention fusion is put forward in the present paper, where a model of
time-and-space motional attention is defined on the basis of the mechanism of visual-motion-attention formation and the target region is obtained by fusing the information of multi-scale motion attention according to D-S evidence theory. The algorithm helps to get over the defect and limits of traditional algorithms and has better robustness and wide application.

2. The Model of Motion Attention
Noise can result in regional disorderly motion in motion vector field and it can be reduced to obtain neatly divided motion vector field through time-and-space filtering process.

In time dimension, the differential value of motion vector within the neighborhood is defined to describe time attention factor. The time attention factor of \( V_{k,i,j} \) is

\[
A^T_{k,i,j} = |\Delta V| = |V_{k,i,j} - V_{k-1,i,j}| 
\]

Where \( V_{k,i,j} \) and \( V_{k-1,i,j} \) stand respectively for the motion vectors at the coordinate \((i, j)\) within frame \(k\) and \(k-1\).

In spatial dimension, let \( MB_{k,i,j} \) be the macroblock at coordinate \((i, j)\) within \(k\)-frame, \(i\) and \(j\) respectively stand for its abscissa and ordinate; \( \Lambda_{k,i,j} \) stands for the set including macroblock \( MB_{k,i,j} \) and neighbouring macroblocks, then, the space motion attention factor of \( V_{k,i,j} \) is defined as

\[
A^S_{k,i,j} = a \cdot |V_{k,i,j} - u_{k,i,j}| + b \cdot P\log \frac{1}{p} 
\]

Where \( a \) and \( b \) are coefficient, \( V_{k,i,j} \) stands respectively for the motion vector at the coordinate \((i, j)\) within \(k\)-frame; \( V_{k,i,j} \) represents for the point set of vector, where

\[
\sum_{[i,j \in \Lambda_{k,i,j}]} V_{k,i,j} 
\]

\( u_{k,i,j} \) is the normal Gauss’ probability function of the present motion vector estimation.

Motion attention involves time and space factors and therefore, attention must be paid to the fusion of time and space attention in setting up a model of motion attention. A model of motion attention may be defined as a linear-fusion model of time and space attention factors, whose formula is

\[
A_{k,i,j} = \alpha \cdot A^T_{k,i,j} + \beta \cdot A^S_{k,i,j} 
\]

Where \( \alpha \) and \( \beta \) are the coefficient for positive value.

3. Multi-scale Fusion of Motion Attention
According to optical flow equation, the closer the directional derivative in certain direction of image is to a gradient value in an ideal state, the more reliable and credible the estimation result will be. Therefore, directional derivative may be used to measure the credibility of motion vector of optical flow estimation.

Owing to the dependence of visual sense on observation angle, the fusion of multi-scale motion attention will cause the picture of significant motion attention to approximate to its actual state, highlighting the motion contrast between lens and a target to create favourable conditions to obtain the target region. Based on D-S evidence theory, motion attention and motion vector credibility are used to define the measure formula of target pixel block, the strategy fusion is carried out with motion attention of different spatial scale as multi-sources for the fusion so that the position where a target lies is obtained through bilateral processing.
The fusion result is, in fact, mainly made up of the boundaries or the internal local region of a moving target. To determine the region where a target is, the region of significant attention is binarized with threshold, then the median point of the target is computed.

4. Experimental Analysis
The high resolution (1920 × 1080) global motion video data from Taiwan DINGJI information picture Co., Ltd. are adopted in the experiment and for the convenience of multi-scale operation, the resolution is reduced to 1024 × 1024. The experiment is done on a Dell computer (Core 2.0GHz, 1GB RAM) and the programming environment is Matlab2010. In the experimental analysis, the algorithms----Globe Motion Compensation Based(GMC) [4], Motion Attention Based(MA) [13] Globe Motion Compensation and Visual Attention Based(GMC-VAB) [15]----and the algorithm proposed in this paper Multi-Scale Motion Attention Fusion Based (MSMAFB) are compared with the focus on the effect of detecting motion target in a global motion scene. The parameter configuration of GMC-VA and MSMAF is shown in Table 1 and 2 and the experimental result is shown in Figure 1. To examine the computing performance, the average consumption time of the results from 50 series experiments is worked out before the algorithm programs are optimized. The average consumption time is listed in Table 3.

Table 1. Parameter Configuration of GMC-VA

| Parameter Configuration | Static Attention Parameter | Motion Attention Parameter | Attention Model Parameter | Target Segmentation Parameter |
|-------------------------|-----------------------------|-----------------------------|---------------------------|-------------------------------|
| τ = 50                  | σ = 1                       | δ = 0.9                     | η = 0.8                   |

Table 2. Parameter Configuration of MSMAF

| Parameter Configuration | Time Filtering Parameter | Time Filtering Parameter | Attention Model Parameter | Credibility Function Parameter | Scale Layers |
|-------------------------|--------------------------|--------------------------|----------------------------|--------------------------------|---------------|
| δ = 0.5                 | Median filter            | α = β = 1                | γ = 2, ϵ = 1               | H₀ = 0.2                      | 3             |

Table 3. Computation Time Comparison ( s/t )

| Series     | GMC   | MA    | GMC-VA | MSMAF |
|------------|-------|-------|--------|-------|
| Hummingbird| 143.96| 85.68 | 161.37 | 123.46|
| Leopard    | 140.12| 85.87 | 168.78 | 126.75|
| Aircraft   | 148.49| 85.36 | 159.90 | 118.27|
| Horse      | 160.78| 86.29 | 157.60 | 116.59|
Altogether eight global motion scene sequences are adopted in the paper, for each of which 10 time-point frames are tested. In the following part, the dynamic characteristics of the eight global motion scene sequences are introduced. “Hummingbird” sequence: the target moves intensely against a background of strong brightness change, the hummingbird keeps pecking at the food on the desk with the wings flapping at high speed and the lens moves slightly with the changing position of the bird. “Leopard” sequence: The target moves slowly against a background of yellow grassland of relatively large texture correlation, the leopard moves to the left and the lens moves slowly correspondingly. “Aircraft” sequence: The target slides speedily in the air with a background of distant vision of objects of complicated texture on the land, the aircraft flies to the top of the image and the lens moves with it. “Horse” sequence: The target moves intensely against a background of trees and grass of flat changing texture, the horse runs to the left and the lens moves quickly to the left.

As is shown in Figure 1, the effect of detecting moving target in global motion scene from the algorithm GMC and GMC-VA is not as satisfactory as that from MSMAF; The computation result of the algorithm MA can not accurately reflect the significant features of actual motion in the scene. When global movement is intense or a background is complicated, the target can hardly be kept effectively with the background removed by the algorithm GMC, which can be seen in “Horse” sequence, but the algorithm GMC can give acceptable effect of target detecting through setting the threshold when the lens moves slowly while the target moves intensely, which can be seen in “Hummingbird”, “Aircraft” and “Leopard” sequence. Under the same conditions, global compensation and static attention are fused in the algorithm GMC-VA so that it produces better effect than GMC, but its effect is not so good in certain cases of motion target detecting, for example, the target region cannot be accurately located in “Hummingbird”, “Aircraft” sequence. The algorithm MA can not give accurate computation of motion significant features in global motion scene and is obviously not suitable for motion target detect. The above discussed algorithms are mainly influenced by the following two factors: (1) Inaccurate estimation of global motion resulting in unsatisfactory detect effect; (2) Noise interference and optical estimation error resulting in inaccurate computation of motion attention. As the experiment results show, by the algorithm MSMAF put forward in the paper, a comparatively accurate motion region in global motion scene can be obtained through reasonably defining motion attention model and multi-scale fusion, hence better detect effect. Besides, Table 3 indicates that the algorithm MSMAF consumes less time than similar algorithms in the light of operation efficiency.

5. Conclusion
With regard to the limitations of current algorithms for the target detect in global motion scene, a target detection algorithm based on multi-scale motion attention fusion is put forward in the paper, offering a new valuable idea for target-detect research. The model of motion attention is defined on the basis of the mechanism of motion attention formation, and multi-scale spatial motion attention is fused according to D-S evidence theory. The proposed algorithm not only breaks through the
limitations of traditional algorithms, but also has, as the experiment results show, better robustness and wider application for target detect in global motion scene.

**Acknowledgments**

Thanks to the funding of the National Natural Science Foundation of China (61673318, 61703331, 61502385)

**References**

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

[2] KaewTraKulPong P, Bowden R, An improved adaptive background mixture model for real-time tracking with shadow detection, in *2nd European Workshop on Advanced Video-based Surveillance Systems, Kingston upon Thames*, 2001.

[3] B K P Horn and B G Schunck, “Determining optical flow,” *Artif. Intel.*, vol. 17, pp. 185-203, 1981.

[4] Bin Qi, Mohammed Ghazal, Aishy Amer. Robust Globa Motion Estimation Oriented to Video Object Segmentation. *IEEE Transactions on Image Processing*, Vol.17, No.6, PP.958-967, 2008.

[5] Yue-Meng Chen. A Joint Approach to Global Motion Estimation and Motion Segmentation from A Coarsely Sampled Motion Vector Field. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.21, No.9, 2011.

[6] Arvanitidou, M G; Glantz, A; Krutz, A; Sikora, T; Mrak, M; Kondoz, A. Global motion estimation using variable block sizes and its application to object segmentation. *Image Analysis for Multimedia Interactive Services*, 2009. WIAMIS ’09. pp.173-176, 2009.

[7] L Cloutier, A Mitiche, and P Bouthemy, “Segmentation and estimation of image motion by a robust method,” in *Proc. IEEE ICIP*, Nov. 1994, pp. 805–809.

[8] M M Chang, A M Tekalp, and M I Sezan, “Simultaneous motion estimation and segmentation,” *IEEE Trans. Image Process.*, vol. 6, no.9, pp. 1326–1333, Sep. 1997.

[9] F Dufaux and J Konrad. Efficient, Robust and Fast Global Motion Estimation for Video Coding. *IEEE Transactions on Image Processing*, vol. 9, no.3, pp. 497–501, 2000.

[10] Y-P Tan, D D Saur, S R Kulkarni, and P J Ramadge. Rapid Estimation of Camera Motion from Compressed Video with Application to Video Annotation. *IEEE Transactions on Circuits Systems Video Technology*, Vol.10, No.1, pp. 133–146, 2000.

[11] Itti L, Koch C Computational Modeling of Visual Attention. *Nature Reviews Neuroscience*, Vol.2, No.3, PP.193-203, 2001.

[12] Itti L, Koch C Niebur E A Model of Saliency-based Visual Attention for Rapid Scene Analysis. *IEEE Trans on Pattern Analysis and Machine Intelligence*. Vol.20, No.11, PP.1254-1259, 1998.

[13] Yu-Fei Ma, Xian-Sheng Hua, Lie Lu. A Generic Framework of User Attention Model and Its Application in Video Summarization. *IEEE Transaction on Multimedia*, 2005, vol. 7, no.5, pp.907-919.

[14] Zhiwen Yu, Hau-San Wong A Rule Based Technique for Extraction of Visual Attention Regions Based on Real-Time Clustering [J]. *IEEE transaction on multimedia*, 2007, vol. 9, no.4, pp.766-784.

[15] Junwei Han, Object Segmentation from Consumer Video: A Unified Framework Based on Visual Attention. *IEEE Transactions on Consumer Electronics*, Vol.55, No.3, PP.1597-1605, 2009.

[16] Yuming Fang, Weisi Lin, Chiew Tong Lau, and Bu-Sung Lee A, Visual Attention Model Combining Top-Down and Bottom-Up Mechanisms for Salient Object Detection[C]. *ICASSP* 2011, pp.1293-1296, 2011.