

# Target Detection and Extraction Based on Motion Attention Model

Long Liu <sup>a,\*</sup>, Qing Liu

Xi'an University of Technology, China

<sup>a</sup> Corresponding author: liulong@xaut.edu.cn

**Keywords:** water injection network; water injection network model; forward modeling and inversion iterative algorithm.

**Abstract.** In the paper, a new motion attention temporal spatial fusion model is constructed for motion object detection and extraction in view of the limitations of target detection and extraction method under global motion scene according to motion attention formation mechanism. In the algorithm, motion vector fields undergo superposition, filtering and other pretreatment firstly. Then, a motion attention fusion model is defined according to temporal - spatial change characteristics of motion vector. The model is adopted for detecting the motion object region. Finally, morphology and boundary following method are utilized for accurate extraction of the object region. The test results of many different global motion video scenes show that the algorithm has better accuracy and real - time performance than other similar algorithms.

## 1. Introduction

Motion object detection and extraction have been the research hot-spots in the field of video analysis with wide application value. The motion target detection and extraction can be broadly divided into two categories: firstly, the motion target is detected and extracted during stationary camera, namely local motion scene, secondly, the motion object is detected and extracted during moving camera, namely global motion scene. The motion target detection method is more mature under local motion scene. However, motion object detection and extraction are always difficult due to complexity of motion information under global motion scene.

A more typical method of studying motion attention is motion attention model related in literature [1] till present. However, it has the following two problems when it is used in object detection: (1) motion attention is defined from the perspective of statistics. The local motion contrast factors are not considered. (2) The temporal-spatial motion attention integration is not reasonable, and the average value is simply calculated on temporal-spatial motion attention only. Objects can not be accurately detected for the motion attention model in literature [1] due to the above problems.

## 2. Motion attention temporal-spatial fusion model

Motion vector field reflects the motion information in video information. Therefore, it is necessary to obtain motion attention effectively. Motion contrast is manifested as motion vector temporal-spatial distribution inconsistency in motion vector field from the perspective of motion vector calculation method. In the section, the motion attention temporal-spatial fusion model is defined according to temporal-spatial distribution characteristics of motion vector.

### 2.1 Motion vector field pre-processing

Motion vector field is estimated according to optical flow constraint equation. Motion vector field shows sparse and locally messy motion characteristics sometimes due to optical flow method defect since the motion characteristics of adjacent frames are not strong enough, and the video information has certain noise at the same time. In the paper, motion refers that vector field undergoes superposition and filtering. Motion vector accumulation process is shown as follows: set current frame as frame  $n$ , set center coordinate of macroblock as  $(k, l)$ , and express corresponding motion

vector as  $(v_x^{k,l}(n), v_y^{k,l}(n))$ , and motion vector accumulation with the former and later frames is calculated according to formula (1).

$$(v_x^{kl}, v_y^{kl}) = \sum_{i=n-c}^{i=n+c} (v_x^{kl}(i), v_y^{kl}(i)) \quad (1)$$

## 2.2 Calculation of temporal-spatial motion attention

In literature [1], Ma Yufei, etc. calculated motion vector strength. Temporal-spatial motion attention factor is defined according to entropy concept. The fused value of strength and temporal-spatial motion attention factor is adopted as the final motion attention saliency map. The method is mainly applied for video retrieval method. The temporal-spatial attention calculation depends on statistics of motion vector characteristics. In the paper, it is believed that motion attention is caused by motion contrast in temporal-spatial aspects. The temporal-spatial neighbourhood contrast of motion vector is larger, the factor causing motion attention is higher. The contrast is smaller, the factor causing motion attention is lower. The motion attention is directly obtained from motion contrast, and the motion attention should be defined mainly by temporal-spatial difference in motion vector local neighborhood.

In the paper, time-neighboring motion vector difference value is adopted for measuring motion vector time motion attention. It is defined as follows:

$$A_{k,i,j}^T = |\Delta V| = |\vec{V}_{k,i,j} - \vec{V}_{k-1,i,j}| \quad (2)$$

In the formula,  $\vec{V}_{k,i,j}$  and  $\vec{V}_{k-1,i,j}$  respectively represent motion vectors with coordinate position of  $(i, j)$  in frame  $k$  and frame  $k-1$ .

In spatial dimension, motion vector average value and motion vector difference value in 8 neighbourhoods are adopted for measuring spatial motion attention. It is assumed that  $MB_{k,i,j}$  is the macroblock in frame  $k$  with coordinate  $(i, j)$ ,  $i$  and  $j$  respectively represent horizontal-vertical coordinates of macroblock;  $S_{k,i,j}$  refers to the set containing macroblock  $MB_{k,i,j}$  and adjacent macroblock thereof. The spatial motion attention of  $\vec{V}_{k,i,j}$  is defined as follows:

$$A_{k,i,j}^S = |\Delta V| = |\vec{V}_{k,i,j} - \vec{u}_{k,i,j}| \quad (3)$$

In the formula:  $\vec{V}_{k,i,j}$  respectively represents motion vector in frame  $k$  with coordinate position of

$$(i, j), \bullet \text{ represents vector dot product, wherein } \vec{u}_{k,i,j} = \frac{\sum_{\{i,j\} \in S_{k,i,j}} \vec{V}_{k,i,j}}{8}.$$

## 2.3 Temporal-spatial motion attention fusion model

Since motion attention contains temporal and spatial factors, information fuse in two aspects of time attention and space attention should be considered for construction of motion attention model, Firstly, a linear fusion model with motion attention model of temporal-spatial attention factor is defined, and the most simplest linear combination is adopted as follows:

$$A_{k,i,j} = \alpha \cdot A_{k,i,j}^T + \beta \cdot A_{k,i,j}^S \quad (4)$$

Wherein,  $\alpha$  and  $\beta$  are positive value coefficients.

Formula (4) shows that linear operation has concise and efficient model composition. However, temporal-spatial attention influence contrast can not be reflected from the perspective of the influence of temporal-spatial motion change on attention, the influence weights of temporal-spatial attention on attention is different at different time points. In the paper, larger motion attention calculation value is regarded as the main factor of attention calculation. Bias factor  $\max(A_{k,i,j}^T, A_{k,i,j}^S) \cdot |A_{k,i,j}^T - A_{k,i,j}^S|$  is set,  $\max(A_{k,i,j}^T, A_{k,i,j}^S)$  indicates that  $A_{k,i,j}^T$  or  $A_{k,i,j}^S$  with larger value will play a dominant role in attention calculation, the difference value  $|A_{k,i,j}^T - A_{k,i,j}^S|$  of  $A_{k,i,j}^T$  and  $A_{k,i,j}^S$  is larger, and the weight is larger. Finally, the motion attention fusion model is defined as follows:

$$\begin{aligned}\tilde{A}_{k,i,j} &= A_{k,i,j} + \frac{1}{2} \delta \cdot \max(A_{k,i,j}^T, A_{k,i,j}^S) \cdot \sigma \\ &= \alpha \cdot A_{k,i,j}^T + \beta \cdot A_{k,i,j}^S + \frac{1}{2} \delta \cdot \max(A_{k,i,j}^T, A_{k,i,j}^S) \cdot |A_{k,i,j}^T - A_{k,i,j}^S|\end{aligned}\quad (5)$$

Wherein,  $\delta$  is a regulating factor, and positive value is adopted. The later part in formula (5) represents the influence bias of time and space on attention, and it reflects the influence weight of temporal-spatial attention on overall motion attention.

### 3. Motion object region precision extraction

#### 3.1 Determination of motion object region

In global motion scene, motion vector field will show local and temporary motion contrast due to interference and inaccurate estimation, which is not beneficial for judging whether the motion macroblock belongs to object region or not. Therefore, further treatment is also required for determining whether the motion macroblock belongs to object region or not.

Since the motion contrast caused by interference or inaccurate estimation of optical flow method is temporary, the motion contrast generated by object motion has certain continuity relatively. Therefore, time-neighboring average is firstly obtained for attention calculation of motion macroblock  $MB_{k,i,j}$ , thereby greatly lowering the misjudgement due to interference and inaccurate estimation, and average calculation is shown in formula (6). Then, the operator can judge whether motion macroblock  $MB_{k,i,j}$  belongs to object region (TR, Target Region) or not according to formula (7).

$$F_{k,i,j} = \frac{1}{n+1} \cdot \sum_{k=t-n}^{k=t} \tilde{A}_{k,i,j} \quad (6)$$

$$F_{k,i,j} = \begin{cases} \geq T & MB_{k,i,j} \in TR \\ < T & MB_{k,i,j} \notin TR \end{cases} \quad (7)$$

In the formula:  $n$  is an integer larger than zero,  $T$  represents positive number judging threshold, and  $MB_{k,i,j}$  refers to corresponding MacroBlock.

### 4. Experimental results

In the paper, different video scene sequences are adopted for testing. The detection algorithm flow based on motion attention fusion model is shown in figure 1. In the test, the results of global motion compensation test algorithm and motion attention model in literature [1] after binarization processing are compared with the algorithm in the paper. The above methods are respectively recorded as algorithm 1, algorithm 2 and algorithm 3. In the test process, parameter  $\delta$  and threshold value  $T$  are respectively selected as 0.9 and 5.6 according to experience, and Matlab 2010 programming language is selected.

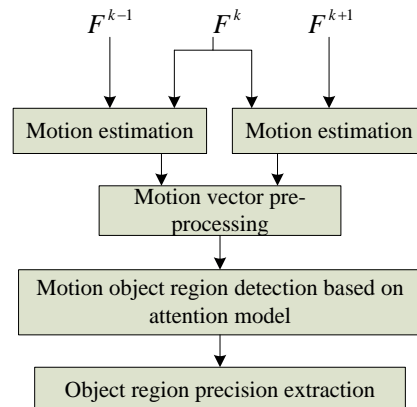


Figure 1. The Algorithm Flow of Motion Attention Fusion Model Based Moving Object Detection and Extraction

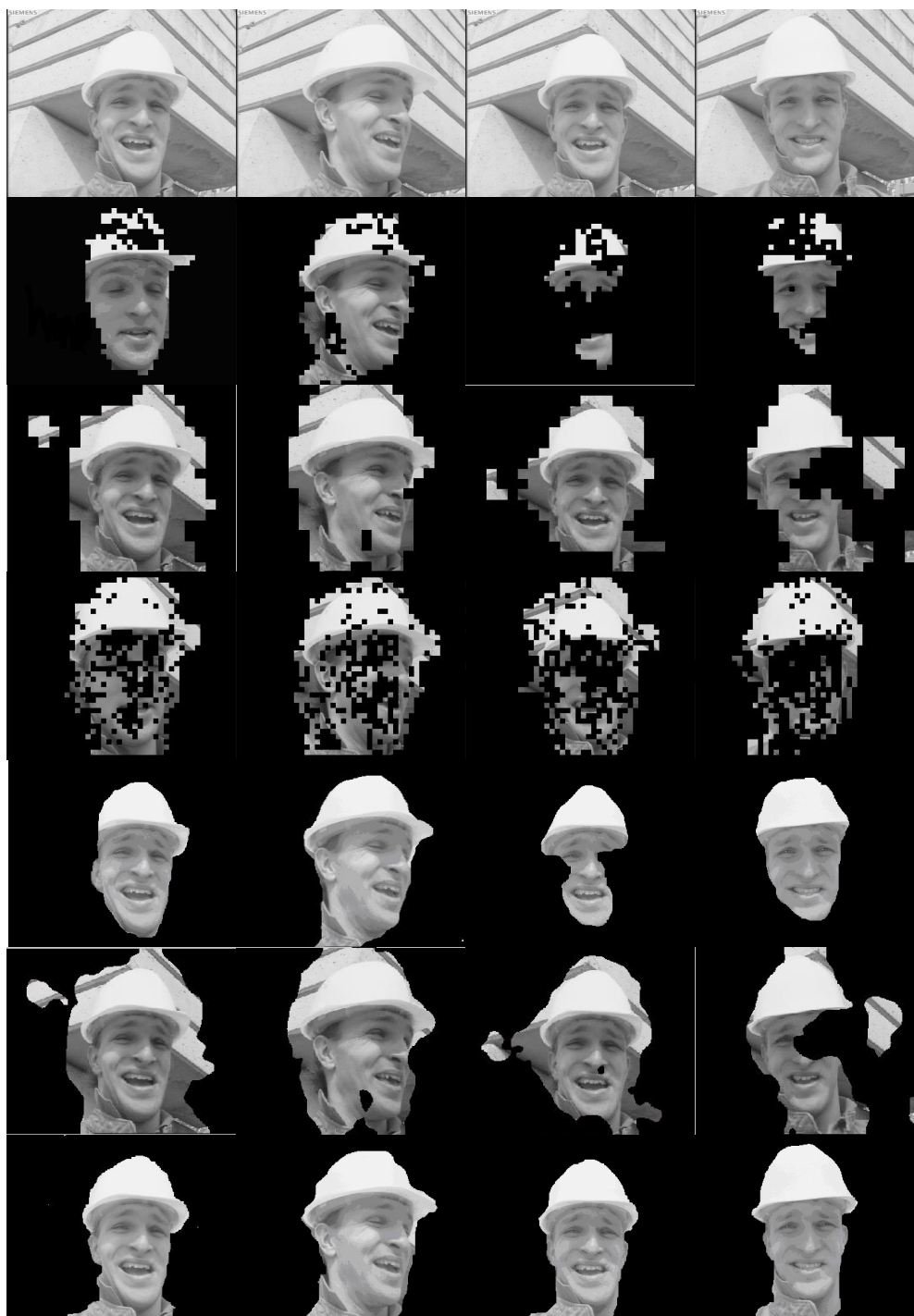


Figure 2. The Result of 'Foreman Video'

In the experiment, sequence 'Foreman' is selected for test. The above test sequence is global motion video scenes; 'Foreman' sequence has relatively large motion object, and stronger lens motion and object motion shaking.

Figure 1 respectively shows the test results of 'Foreman' (CIF) sequence frame 2, 12, 21 and 46, wherein line 1 to line 5 are respectively shown as follows: video sequence original image frame, algorithm 1 object preliminary test result, algorithm 2 object preliminary test result, algorithm 3 object preliminary test result, algorithm 1 final test result, algorithm 2 final test result and algorithm 3 final test result. The test results show that algorithm 1 is in the video sequence with larger motion object, such as 'Foreman' sequence. Since global motion estimation accuracy is not high, and the error of target detection and extraction is prominently increased compared with algorithm 3. Though

algorithm 2 is also is used for roughly calculating the object region, the result is worse than algorithm 3 (Table 1).

Table 1. The Time Cost Comparison of Moving Object Detection (ms/frame)

Test sequence	Format	Test frame number	Algorithm 1 average time loss	Algorithm 2 average time loss	Algorithm 3 average time loss
Foreman	CIF	Frame 1 to frame 125	350 ms/frame	156 ms/frame	81 ms/frame
	QCIF	Frame 1 to frame 125	127 ms/frame	89 ms/frame	37 ms/frame

In summary, motion objects can be detected and extracted effectively under global motion scene through utilizing the motion vector for temporal-spatial attention factor detection aiming at algorithm based on motion attention fusion model, thereby avoiding poor robustness and large operand due to global motion estimation, and improving testing accuracy and real-time performance. It is obvious that it has wide application value.

## 5. Conclusion

In the paper, a motion attention temporal-spatial fusion model aiming at target detection is proposed. Firstly, motion vector obtained according to optical flow method undergoes accumulation and filtering processing. The motion attention fusion model is defined according to temporal-spatial characteristics of motion vector. The model is adopted for testing motion object region in the motion vector field. Finally, the object region is precisely extracted. In the experiment, different video sequences are tested, it is proved that motion object test by the algorithm has better accuracy and real-time performance than other methods.

## Acknowledgments

This paper is supported by State Natural Sciences Foundation General Program (NO.61673318), State Natural Sciences Foundation Youth Project (NO.61703331) and Shaanxi Province Industrial Tackling Project (NO.2016GY-128).

## References

- [1] Chang Yulian, Gao Sheng, Guo Junzhong. Research on water injection network system model simplification technology and calculation method. *Acta Petrolei Sinica*, 2001,22(2):95-100.
- [2] Gao Sheng, Guo Junzhong, Chang Yulian. Research on mathematical model of oilfield water injection network system and its calculation method. *Drilling Technology*, 2001,24(5):54-56.
- [3] Cong Riqing. Water injection network system model calculation method. *Oil-Gasfield Surface Engineering*, 2004,23(8):13-14.
- [4] Hu Xiaoying, Liu Weiguo, Cao Xinzhai, Zhang Weiguang. A hydraulic calculation model based on oilfield water injection network. *Petroleum Instrument*, 2007,21(3):13-15.
- [5] Liu Yan, Wang Yongqing, Wei Guoqing. A method for solving the oilfield water injection network model. *West-China Exploration Engineering*, 2008:67-68.