Tracking Object by Logic Reasoning

Huibin Wang*, Junlei Shen, Jie Shen and Zhe Chen

College of Computer and Information Engineering,
Hohai University, 211100, Nanjing, China
hbwang@hhu.edu.cn

Abstract

A logic reasoning model based method is proposed in this paper for object tracking in compressed domain. First, the object is defined by the motion detector. Then, the supposed target trajectory and predicted position are estimated by the direction angle and the intensity within motion field. The possible conditions of lost, stop or occlusion are determined by logic and reasoning analysis and handled by the direction angle and the intensity in motion field. The outperformance of our method on reducing loss rate and enhancing the trajectory fit has been demonstrated by the experiments. The accurate object tracking result of the proposed method is presented as well.

Keywords: *Moving target tracking, logical reasoning, H.264 compressed domain*

1. Introduction

Moving object tracking is an important subject in image analysis and understanding, which is to identify the target and related them and then to achieve moving targets tracking in video sequences finally. Currently, most tracking methods are focused in pixel domain [1, 2, 3, 4]. Compared with these methods, the compressed domain approach can analyze the target during the decoding process and handle the problem of high time-delay. Therefore, these advantages in compressed domain can adapt well to some real-time and low-cost monitoring system. However, only a few (motion vectors, DCT coefficients, etc.) available information is left in compressed domain for tracking. Moreover we have to face the general problems in tracking process, such as loss, stop and occlusion. Hence, many challenges would present when tracking object in compressed domain.

In recent years, H.264 coding has become to be the major standard for video surveillance. There are some moving target tracking methods in H.264 compressed domain, for example, Xu et al. [6] used the macro-block threshold comparator to solve tracking loss problem in compressed domain; Christian Kas et al. [7] introduced a template matching method in compress domain; Liu et al. [8] proposed a method for merging unreliable trajectory in moving target tracking. While, most popular methods are still blocked with the bottleneck of occlusion.

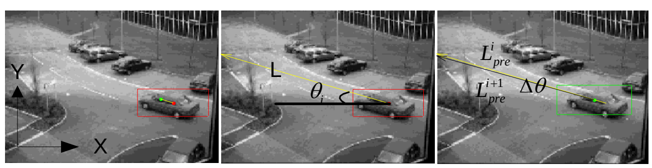
This paper presents a logic reasoning model for tracking moving targets in H.264 compressed domain. Firstly, a supposed trajectory directed by the motion vector is established. Then the average motion vector (AVMV) is used to predict the region of next frame. The direction angle and the intensity of motion filed within predict region is selected to evaluate the motion situation of the target, such as target loss, stop or occlusion. With the help of logical reasoning and screening mechanism, we can locate the target in the most possible region. Ultimately, this algorithm captures the accurate position by constantly correcting in H.264 compression domain.

2. The Mathematical Description of Logic Reasoning in Compressed Domain

According to the logical reasoning mechanism, in the problem of tracking target in compressed domain, the information such as the motion vectors can be regarded as the abstract concept which is able to be used in the position prediction, so as to deal with the problem of occlusion and loss in the tracking process.

Logical reasoning strategy mainly composes of three steps. First of all, collected information is conceptualized and abstracted. The situations we face in the target tracking are judged by the rules subsequently. At last, the previous judgment is reasoned to estimate the optimal position. If it is correct, we can get the real target location by deductive reasoning. While, if not (such as loss and halt), the results should be evaluated with non-deductive reasoning. After repeated judgment and identification, the correct region belonged to the target is obtained.

In compressed domain, the basic information element **MV** is displacement vector information, which is generated by motion estimation from one frame to the next. **MV** includes displacement information in two direction, that is the x and the y direction, respectively. We use $\mathbf{MV}_x$ to denote the motion vector in x direction, and $\mathbf{MV}_y$ in y direction, the angle between them is defined as θ.



(a) (b) (c)

Figure 1. The Target Information and Predicted Trajectory

The detector in compression domain is applied to detect the moving target region in the i th frame (Figure 1(a)). The red and green dots in Fig.1(a) are the centroid of the current target region and the predicted target region in next frame, respectively; and the yellow line describe the movement between successive frames.

2.1. Abstraction in Compressed Domain

As shown in Figure 1(a), the red dot (x_0^i, y_0^i) is the moving target's centroid in the current frame i, and the red rectangle (w_0^i, h_0^i) expresses the limits of the detection region. The yellow line is the average motion vector $\mathbf{AVMV}_i$ of the current target's region, which includes a pair of components, $\mathbf{AVMV}_i^x$, $\mathbf{AVMV}_i^y$. The direction angle θ_i can be calculated as $\theta_i = \mathbf{AVMV}_i^y / \mathbf{AVMV}_i^x$.

2.2. Assumption and Judgment in Compressed Domain

By judging the previous information, following hypotheses are built as: (i) If the target speed is constant or almost unchanged within a series of frames, the motion vector detected previous can be used to predict the target centroid in the next frame, regarding as the real position $(x_{0,pre}^{i+1}, y_{0,pre}^{i+1})$ in next frame(green point in Figure 1(a)). (ii) If the target remains unchanged in the whole movement process, the motion trajectory is L^i_{pre} (yellow arrow in Figure 1(b)). Thus, the prediction point $(x^{i+1}_{0,pre}, y^{i+1}_{0,pre})$ and the motion trajectory L^i_{pre} are computed as

$$(x^{i+1}_{0,pre}, y^{i+1}_{0,pre}) = (x^i_0 + \mathbf{AVMV}^x_i, y^i_0 + \mathbf{AVMV}^y_i) \quad (1)$$

$$L^i_{pre} = L((x^i_0, y^i_0), \theta_i, W, H) \quad (2)$$

Where, (x^i_0, y^i_0) is the centroid of moving target in frame i, $\mathbf{AVMV}^x_i$ and $\mathbf{AVMV}^y_i$ show the movement in line with the x and y axis in frame i, θ_i is the direction angle of the motion region in frame i, W, H are the width and height of the frame image.

2.3. Logical Reasoning in Compressed Domain

In order to decide the reasoning method, the predicted region of the target must be evaluated by analysing and abstract to the actual situation based on the logical reasoning analysis.

As shown in Figure 1(c), if the predicted region is correct, then the difference $\Delta\theta$ between the predicted angle θ_i and the direction angle θ^{i+1}_{pre} calculated by the average motion vector $\mathbf{AVMV}^{pre}_{i+1}$ is smaller than the threshold θ_0. The black line denotes the predicted trajectory L^{i+1}_{pre} in line with the predicted direction angle θ^{i+1}_{pre}.

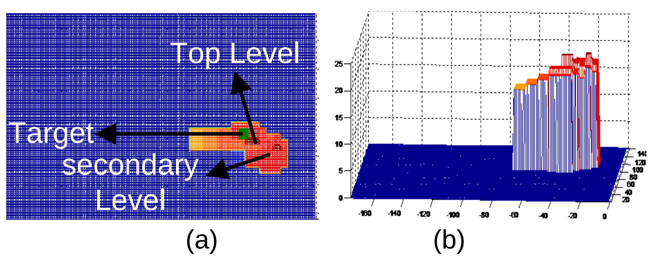


Figure 2. The Motion Filed Distribution in $i+1$th Frame. (a) Contour Distribution; (b) Intensity Distribution

As shown in Figure 2, the centroid of target region in $i+1$th frame is between the top and secondary level contour lines whose intensity can be roughly divided into 4 to 5 level, the darker the level line, the greater the strength of motion vector. If the level of the centroid is smaller than N, then $|\text{MV}(x^{i+1}_{0,pre}, y^{i+1}_{0,pre})| > \text{MV}_N$, and the predicted position is correct. With the above reasoning theory, this paper uses deductive reasoning method to get the correct region (L^{i+1}_{pre}, $\mathbf{AVMV}^{pre}_{i+1}$, θ^{i+1}_{pre}).

According to the analysis above, deductive reasoning method can be used in this condition. But in some cases, the moving direction changes suddenly, and the direction angle of the predicted field may deviate from the prior direction. Thus, we need to continuously suppose and correct the conditions to satisfy the requirements of deductive reasoning.

The strategy with many assumptions has the ability to solve two sorts of problems. The first one is abrupt moving direction change, which results in the deviation of the target from the predicted trajectory. The second one is the loss of moving target if it is stopped or occluded.

2.3.1. Abrupt moving direction change: Abrupt changes in the direction of motion may result in false moving target loss in compressed domain. The difference between the predicted centroid of $i+1$th $(x_{0,pre}^{i+1}, y_{0,pre}^{i+1})$ and the real position (x_0^{i+1}, y_0^{i+1}) makes the target lost.

In accordance with the logic theory, although the predicted target is lost, the target will not disappear, we can estimate the target with many assumptions. If any parts of the real target motion included in predicted region (Figure 3(a)), the algorithm assume that the real target region in frame i can be predicted according to the direction angle θ_{pre}^{i+1} and the deviated trajectory L_{rev}^{i} ,as

$$L_{rev}^{i} = L((x_0^i, y_0^i), (\theta_i + \theta_{pre}^{i+1}), W, H) \quad (3)$$

Where, (x_0^i, y_0^i) is the centroid of frame i, θ_i is the direction angle in frame i, θ_{pre}^{i+1} is the direction angle of the predicted region in frame $i+1$, W is the width of the video image. H is the height of the video.

If the intensity of the motion is not satisfied with the condition as $|\mathrm{MV}(x_{0,pre}^{i+1}, y_{0,pre}^{i+1})| > \mathrm{MV}_N$, we continually correct it until meeting the conditions.

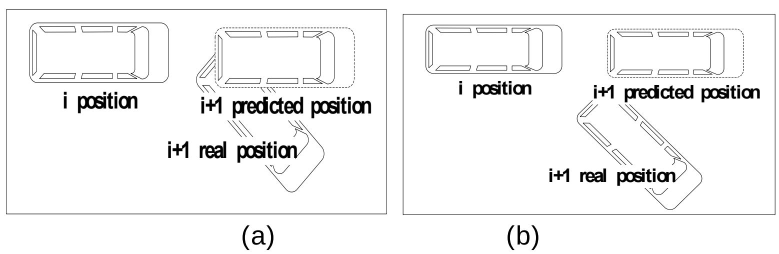


Figure 3. Two Different Situations within Motion Direction Change. (a) Parts of the Real Target Motion Filed include in Predicted Region; (b) the Real Target is not included in Predicted Region

If the real target is not included in the predicted region (Figure 3(b)), we choose θ to correct until satisfying the situation in Figure 3(a). Where, different θ has different means, such as $\theta = 0^0 \sim 90^0$ means turning right; $\theta = 135^0 \sim 225^0$ turning back; and $\theta = 270^0 \sim 360^0$ turning left.

2.3.2. The loss of the moving target: Real loss of moving target may appear under two situations that the target is stopped or occluded.

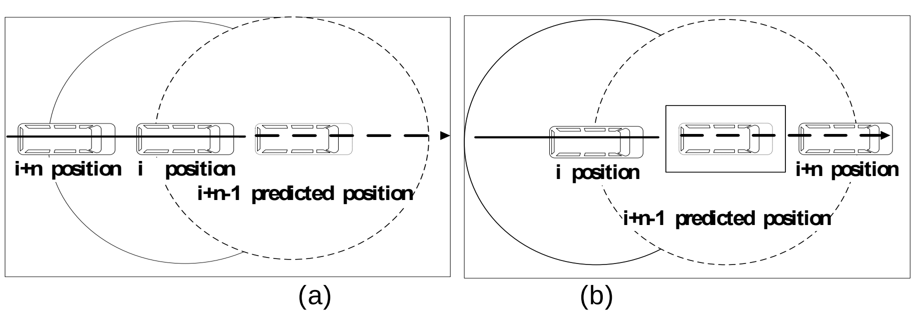


Figure 4. Two Situations in Frame $i+n$.(a) Target Stop; (b) Target Occlusion with Static Material

In this paper, the algorithm generates the centroid and trajectory and judges the two situations with many assumptions. This method generates the supposed centroid in the next frame until the target reappears in frame $i+n$.While, if the target stops, it must reappear in the predicted region of frame i (Figure 4(a)); if the situation is occluded by static material, it must reappear in the supposed predicted region of frame $i+n-1$ (Figure 4(b)). Where, the dashed circle in Figure 4 define the supposed predicted region, the circle denote the predicted region of frame i.

To the problem of multiple-target occlusion, it contains two situations, one is homonymous moving multiple-target occlusion; another is opposite moving multiple-target occlusion.

In the former case, limited to the information in compressed domain, we regard multiple-target as a single target. In this paper, we only intend to find a simple, fast and low–cost method, so this simplification is still feasible. In the latter case, the strength of the predicted region satisfies the strength condition, accordingly the direction should satisfy the direction condition in accordance with the logic theory. But it's not true. So there is only one possible situation left, that opposite moving multiple-target occlusion. In this situation, we regard the supposed centroid as the real position.

3. Moving Object Tracking Algorithm Based on Logical Reasoning in H.264 Compressed Domain

Moving target tracking algorithm based on logical reasoning in H.264 compression domain is shown as a diagram (Figure 5). Firstly, motion detector identifies the moving regions and abstracts the regional information into the descriptors in the tracking process. After extracting and judging the information of the predicted region, we can choose a rational reasoning way to capture the target. While, if the target is lost, halted or occluded, we update the model to satisfy the new conditions until capturing the target. By continuously judging, reasoning and supposing, this algorithm achieves the task of target tracking accurately.

The algorithm for tracking moving targets in H.264 compressed domain can be divided into five steps described as follow.

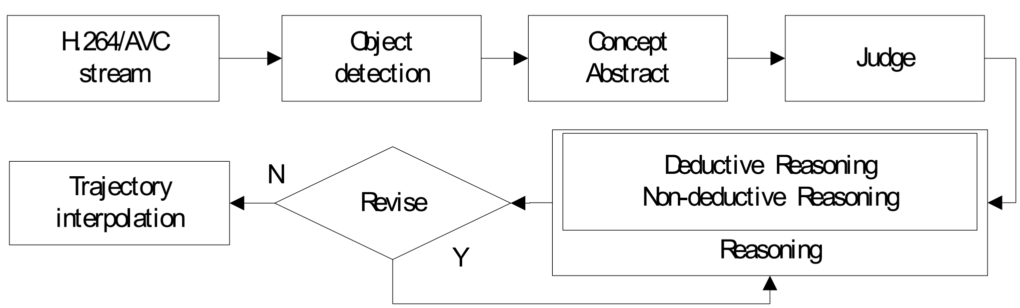


Figure 5. The Diagram of the Proposed Algorithm

step1, using the compressed domain detector to identify the target in frame i, and getting its position (x_0^i, y_0^i), the average motion vector $\mathbf{AVMV}_i$, the size (w_0^i, h_0^i) and the predicted trajectory L_{pre}^i;

step2, predicting the target in frame $i+1$, and extracting the average motion vector $\mathbf{AVMV}_{i+1}^{pre}$, the direction angle θ_{pre}^{i+1} and the centroid $(x_{0,pre}^{i+1}, y_{0,pre}^{i+1})$;

step3, calculating the difference $\Delta\theta$ between θ_{pre}^{i+1} and θ_i and the strength of the motion field in prediction region $|\mathrm{MV}(x_{0,pre}^{i+1}, y_{0,pre}^{i+1})|$, if $\Delta\theta < \theta_0$ and $|\mathrm{MV}(x_{0,pre}^{i+1}, y_{0,pre}^{i+1})| > \mathrm{MV}_N$, the target is correct; else correcting the prediction trajectory and getting the new trajectory L_{rev}^{i} until the conditions above is meet;

step4, if the correction rate is within the 360 degree, then the target is not lost, else the target is stopped or occluded. Accordingly, we locate the supposed position $(x_{0,pre}^{i+n}, y_{0,pre}^{i+n})_{pho}$ and trajectory $L_{pre,pho}^{i+n}$. If the target appears in the predicted region of (x_0^i, y_0^i) in frame $i+n+1$ ($(x_0^{i+n+1}, y_0^{i+n+1}) \in ((x_0^i, y_0^i), L_{pre}^i)$), and $\Delta\theta < \theta_0$, and $|\mathrm{MV}(x_{0,pre}^{i+1}, y_{0,pre}^{i+1})| > \mathrm{MV}_N$, then the target is stopped from frame $i+1$ to $i+n$ and (x_0^i, y_0^i) is the real position within these period; else opposite moving multiple-target occlusion appears, and $((x_{0,pre}^{i+1}, y_{0,pre}^{i+1})...(x_{0,pre}^{i+n}, y_{0,pre}^{i+n}))$ is the real position from frame $i+1$ to $i+n$; if the target arrives in the prediction region of $(x_{0,pre}^{i+n}, y_{0,pre}^{i+n})_{pho}$ in frame $i+n+1$ ($(x_0^{i+n}, y_0^{i+n}) \in ((x_{0,pre}^{i+n}, y_{0,pre}^{i+n})_{pho}, L_{pre,pho}^{i+n})$), then this situation is occluded by the static material, we can regard the missing position $((x_{0,pre}^{i+1}, y_{0,pre}^{i+1})_{pho}...(x_{0,pre}^{i+n}, y_{0,pre}^{i+n})_{pho})$ as the real position from frame $i+1$ to $i+n$;

step5, repeating step 2 to step 4 until that all the video frames have been detected. Consequently, the trajectory is established by linking all the centroid in each frame $((x_0^i, y_0^i), (x_0^{i+1}, y_0^{i+1}), ..., (x_0^{i+M}, y_0^{i+M}))$.

4. Experiment Results

In this paper, the motion field is extracted by H.264 encoding software of JM8.6 version. The based configuration is: Baseline Profile, IPPP...P, one reference frame, and the motion estimation search range is [-16, 16], the quantitative parameter is 28, the frame image size is 176 * 144, and the frame rate is 30 frames per second.

Figures 6, 7, 8 which show the experiments with a single target denote motion direction as almost no difference, a little difference and much difference respectively. Fig.9, 10 show the experiments in the case of multi-target with non-occlusion and occlusion respectively.

Figure 6. PETS-2000 Test Sequence, from Frame 124 to 224

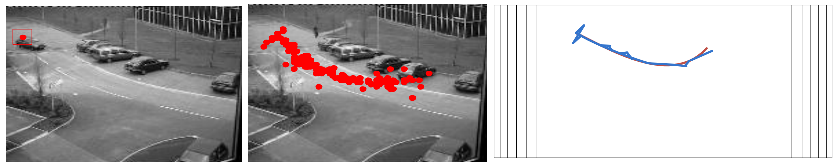


Figure 7. PETS-2000 Test Sequence, from Frame 390 to 670

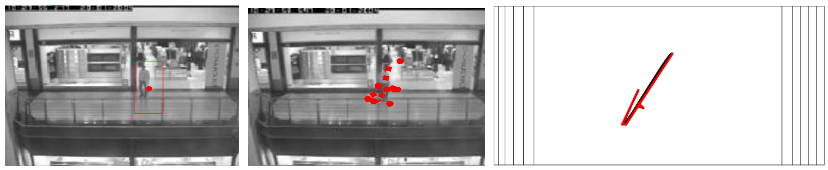


Figure 8. One Leave Shop Re-enter Test Sequence, from Frame 101 to 201

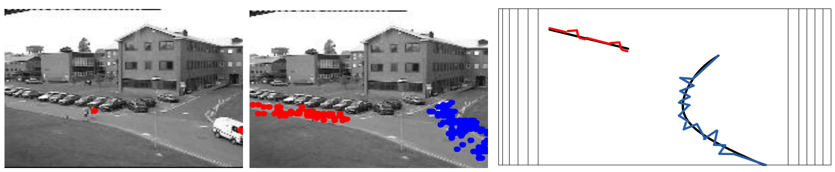


Figure 9. PETS-2001 Test Sequence, from Frame 1600 to 1900

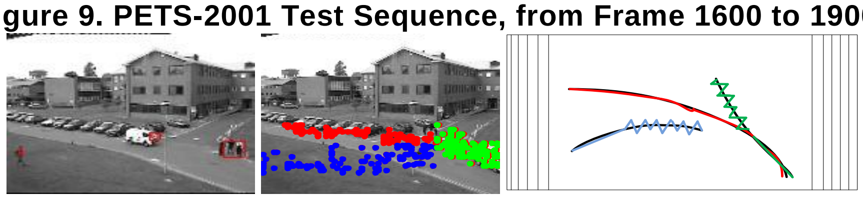


Figure 10. PETS-2001 Test Sequence, from Frame 725 to 1000

As shown in Figure 6, the target is moving from the right to the left side. The left figure shows the moving target region. The middle one shows the centroid point set. The right one shows the real trajectory and the trajectory originated from our algorithm. It can be seen that the generated trajectory fit the real one generally. Figure 7 is a video sequence that a car moves from the left side, slows down, turns right and stops at the right sides successively. In spite of the noise point existed in centroid set when the car slow down, generated trajectory can still follows the real position. Figure 8 shows a video sequence that a person goes out of the shop and then returns quickly. Although the target’s movement velocity is small; the camera is far from the scene and the loss rate is high the target can still be tracked accurately. So we can still obtain the approximate trajectory by this algorithm.

As shown in Figure 9, we are to address the problem of multiple-target without occlusion. The white vehicle in the image moved from right to the left slowly. In spite of the noise points, the tracking trajectory is consistent with the real one generally. Figure 10 shows a sample of multi-target occlusion situations. We use many assumptions to restore the lost centroid position. Because of the overlapped position, three pedestrians are seem as one target for testing in this paper.

Table 1. Experiment Results in Video Sequences

Frame Name	Frame(target number)	Loss Rate(error/total)	Goodness of Fit
PETS-2000	124~224(1)	(2/100)	0.9835
PETS-2000	390~670(1)	(15/280)	0.9826
One Leave Shop Re-enter	101~201(1)	(12/100)	0.9891
PETS-2001	1600~1900(2)	(3/300)	0.9901
PETS-2001	725~1000(3)	(17/275)	0.9844

Table 1 describes the loss rate and goodness of fit between the tracking trajectory and real trajectory from the above test videos. The goodness of fit can be evaluated as

$$\phi_j = 1 - \sqrt{\frac{1}{kWH}[\sum_{n=1}^{k}((x_0^{i+n}, y_0^{i+n}) - (x_{0,pre}^{i+n}, y_{0,pre}^{i+n}))^2]} \qquad (4)$$

Where, ϕ_j is goodness of fit of the j target. k is frame number. (x_0^i, y_0^i) is the real centroid of the i th frame. $(x_{0,pre}^i, y_{0,pre}^i)$ is the generated centroid of the i th frame. W and H is the width and height of the video image.

The data in Table 1 prove the outperformance of our algorithm. In the case of abrupt motion and occlusion, the target can still be located, and the full trajectory can also be extracted in real-time as well.

5. Conclusion

Based on the logical reasoning mechanism, an algorithm to track moving targets in H.264 compression domain is presented. This method detects the moving regions by a compressed domain detector at first and abstracts and conceptualizes the information within the region successively. It combined the compressed domain motion vectors with a simple logical reasoning mechanism to track moving targets. This research presents an alternate way for target tracking comparing with the methods worked in pixel domain. Experimental simulations further prove its better performance on the advantage in enhancing the precision of tracking in video surveillance system. In the future work, we are to explore a fused method for target tracking in compression domain, aiming to achieve higher tracking accuracy.

References

[1] S. W. Ha and Y. H. Moon, J. International Journal of Smart Home, 5, 4 **(2011)**.

[2] J. S. Park, J. H. Yoon and C. Kim, J. International Journal of Signal Processing, Image Processing and Pattern Recognition, 1, 1 **(2008)**.

[3] Z. Islam, C. M. Oh and C. W. Lee, J. IJSP. 2, 1 **(2009)**.

[4] S. Kim, S. Lee, S. Kim and J. Lee, J. International Journal of Advanced Science and Technology, 3 **(2009)**.

[5] E. Maggio, M. Tajo and A. Cavallaro, J. CSVT. 18, pp. 1016-1027, **(2008)**.

[6] S. Xu, G. Zhou and G. Zhu, "A Second-generation Video Compression Standard for Moving Object Extraction and Tracking Algorithm", Processing of the 14th National Conference on Image and Graphics, **(2008)**, pp. 447-451; Fu Jiang, China.

[7] C. Kas and H. Nicolas, J. PSIVT. 5414, pp. 318-329, **(2009)**.

[8] H. Liu, M. –T. Sun and R. –C. Wu, J. Journal of Visual Communication and Image Representation. 22, pp. 432-439 **(2011)**.