Background Information Fusion and its Application in Video Target Tracking

Yuxi Chen
Chongzhao Han, Xin Kang
School of Electronic & Information Engr.
Xi'an Jiaotong University
Xi'an, Shaanxi Prov.
710049, China
chenyx@xauat.edu.cn

Mingjun Wang
Yuxi Chen
Network & Information Center
Xi'an U. of Arch.& Tech.
Xi'an, Shaanxi Prov.
710055, China
wmj@xauat.edu.cn

Abstract– *A Background Information Fusion (BIF) algorithm and its application in video target tracking is proposed in the paper. The BIF algorithm is based on the fact that there are redundant information between different frames. Unlike most of available tracking algorithms based on target features, which focus on the problems of what features the target have, and how does these features change when they moving in a specific scene, The BIF based Target Tracking focus more on the background in which the target existed than available algorithms. We proposed BIF based two step target extraction and tracking algorithm in this paper. At first step, BIF algorithm focus on recovering an intact background from a frame sequences, at the second step it extract target by background differencing algorithm. These two steps algorithms can eliminate the most of the difficulties and challenges in moving target extraction from time-varied background. Our experiments proved that BIF based tracking algorithm is stable, feasible, and effective.*

Keywords Background Information Fusion (BIF), BIF Assumption, Hit, NoHit, NoHitCount, NoHitFrameCount, DitherThreshold

1 Introduction

What is BIF?

BIF refer to Background Information Fusion, which is an innovative idea in video target tracking.

Why BIF needed?

We know that mankind vision differed from machine vision in many aspects, the most important difference in moving target tracking is how to extract object from video scene. Mankind use previous knowledge about the scene and target include:

- The background knowledge– color module, 3D shape module, shadow knowledge, lighting module, color distribution, grain distribution, etc.
- The target knowledge— the target 3D shape, color module, grain distribution of target and the variance of target's features
- The common knowledge— mankind visual experiences learned from long period time of life experiences, the skills and abilities to notice, see, tracking, understanding, and perception to arbitrary scene and target.

Compared with mankind vision, machine vision have many limitations– machine can not understand what is *foreground* and what is *background*, what is *target* and what is not. It can hardly understand target's variance in 3D shape, color, grain, position and deformation. Machine does not have enough *memory and skills* learned from its past visual experiences, which is very important for mankind to understand a scene and tracking of the target.

Recent studies in video target tracking try to solve the challenges of tracking by techniques described as below

- Background differencing algorithm— the simplest algorithm in extracting target from scene[1], but it has a assumption which is also a fatal limitation of the algorithm— the scene should have a stable or fixed background (in color, lightness, grain distribution, etc.), and the observer (camera) should not move in position. This assumption can not be satisfied in many practical video surveillance application, because the scene background is time-varied or arbitrary.
- Feature based tracking algorithm— feature based tracking algorithm has an assumption that the moving targets have some fixed features, by which we can distinct them from the scene background. The features include characteristic color, grain distribution, feature point, feature line, corner point, flex point [2,8], etc. In fact, these features only keep stable in a limited conditions, if the target varied in color, positon,2D shape, etc. caused by moving, the target features may difficulty to be detected and the performance of algorithm decrease significantly.
- Optical flow techniques [3], Temporal disturbances techniques [10], Integration of Temporal Variations techniques [4, 5], Temporal Coherence[3], Dynamic Template [6],Perspective Transform based techniques[7], Active Contours based techniques [9],etc. Optical flow techniques have a large computation cost and is difficult to be realized in hardware, whereas Temporal disturbances and Temporal related techniques assume that the motion of an object from frame to frame does not greatly exceed the it's dimension, this is not always hold in vehicle tracking application .

Now, let's come back to the question, *why BIF needed?* We hold an idea that background information is the most important knowledge in video target tracking, and that, the scene is just the synthesis (fusion result) of the background information and the moving target information, so we need BIF to recover an intact background for target extraction and tracking.

We know that most of previous studies in machine vision and video surveillance pay much attention on target itself, less on the background in which the targets existed. For several reasons the target based tracking algorithm is complicated and time consuming, one is that target often varied in positions, shape, color, lightness, shadow, and that a group of targets may have complicated relationship between each other, *e.g.* occlusion, relative velocity change, relative position change, the combination of these changes are usually very complicated, so the target based tracking is very complicated and poor performance in complicated environment.

Our innovation— the Background Information Fusion (BIF) based tracking algorithm, put emphasis on the background knowledge acquiring, recovery, and utilization. If the background information is a known condition, the previous complicated target extraction procedure will become a simply subtraction of the known-background from the scene.

As to video surveillance applications, BIF based tracking algorithm assumes that:

The scene background is larger in size than targets, slower in speed than targets, and has a lower variation rate in color, lightness, etc .than target. (1)

We call the above assumption by BIF Assumption.

These conditions in the BIF Assumption can be satisfied in most of practical applications, such as safe guarding, vehicle monitoring, industrial production line, we take a road vehicle monitoring application as an example in this paper.

The paper is organized as follows; first we introduce some basic concepts and definitions in Background Information Fusion algorithm in *Sec* 2. Then the BIF based vehicle extraction, tracking algorithm in *Sec* 3, *Sec* 4 presents some experiments result by BIF based tracking algorithm. Finally, in *Sec* 5 we give a brief summarize to BIF algorithm.

2 Background Information Fusion (BIF) algorithm

At first we introduce some definitions and basic concepts in Background Information Fusion (BIF).

2.1 Hit and NoHit

To give a clear explanation to these concepts, let's see Fig.1. to Fig.3. (*Frame k to Frame k+2).*

Fig.1 to Fig.3 shows three continuous frames in a video sequence. According to *BIF assumption in condition* (1), video scene background in surveillance is lager in size than vehicle; it is almost still and has little change in color during a short period of time. The intact background is always unknown since it is stained by moving vehicles frequently, and it also changes slowly with the atmosphere condition and environmental changes. To simplify the description, we use a group of schematic diagrams showed in Fig.4 to make it clear.

Fig.1 frame k

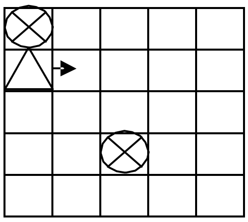
Fig.2 frame k+1

Fig.3 frame k+2

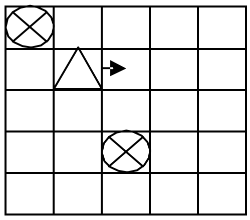
. ⊗ mark in Fig.4 stands for background image point which is not stained by target, □ mark stands for background point which is stained by target ,or target point which is varied on the background image.

A *Hit* is defined as an event a background point is stained by a target point. (2)

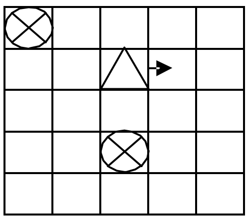
A NoHit is defined as an event that a background point do not stained by target color in a frame, or keep nearly the same color between several frames. (3)



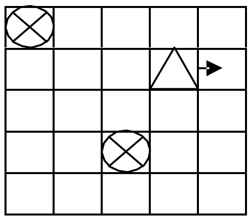
(a) Frame k-1



(b) Frame k



(c) Frame k+1



(d) Frame k+2

Fig.4 Schematic Diagram of Hit and NoHit(a ~ c)

Since vehicle has random movement on road, the Hit point and NoHit point is arbitrary.

2.2 NoHitCount

NoHiCount is defined as the continuous NoHit count number in a specific point. This parameter is very important in defining of a background point color. As shown in Fig.4, the top left point NoHitCount is :

$$NoHitCount(x,y) = NoHitCount(1,1) = 4 \quad (4)$$

2.3 NoHitFrameCount

NoHitFrameCount is defined as the number of NoHitCount from the frame *k* to *k+n,* if there is Hit event occurs from *k* frame to *k+n* frame, the NoHitFrameCount is defined as the maximum NoHitCount number:

$$NoHitFrameCount(x,y,k) = \max[NoHitCount_f(x,y)], \quad where \quad f = k \ldots k+n \quad (5)$$

2.4 AssureCount

AssureCount is defined as the minimum NoHitFrameCount number by which we can make sure of the background color in a specific background point . (6)

2.5 DitherThreshold

The DitherThreshold is defined as the color variance range in each point caused by environment slowly changing and the camera instability. In many industrial applications, the background environment is instable and camera is not still, so we use DitherThreshold parameter to describe environment noise. (7)

3 The Background Information Fusion (BIF) algorithm and BIF based Tracking algorithm

3.1 Background Information Fusion(BIF) algorithm

The primary objective of BIF algorithm is to recover an intact background by a sequence of frames. A video surveillance background can be divided into many pixel points, each point have two invariable properties in certain period of time. We call them *Time-Invariance and Spatial-Invariance* property. (8)

The *Time-Invariance* means that the background color at a specific position does not change with time, or between frames; The *Spatial-Invariance* means that a fixed point has a fixed background color regardless how many times it is *Hit* by target movement.

Based on the definitions and concepts above, we arrive to the algorithm of BIF. Suppose we study a frame image pixel by pixel, a background point color can be recovered by algorithm described by Eq.9.

Where *Color*(*x*,*y*,*k*) is position (*x*,*y*) color in frame *k*, and *BackgroundColor*(*x*,*y*) is the background color in position (x,y).

$$\begin{array}{l} for\ \ k=1\ to\ totalFrames \\ \ if \\ \quad NoHitFrameCount(x,y,k) > AssureCount \\ \ Then \\ \qquad BackgroundColor(x,y) = Color(x,y,k) \\ End \end{array} \quad (9)$$

If the DitherThreshold is taken into consideration, the NoHitCount should have tolerance in the fusion procedure; this tolerance range is described by DitherThreshold:

The Background Information Fusion Algorithm can be described by Eq. 10

$$\begin{array}{l} for\ \ k=1\ \ to\ \ totalFrames \\ \quad if\ \ Color(x,y,k) < Color(x,y,k+n) + DitherThreshold \\ \qquad and \\ \quad Color(x,y,k) > Color(x,y,k+n) - DitherThreshold \\ \quad then \\ \qquad NoHitCount(x,y) = NoHitCount(x,y)+1 \\ \quad end \\ end \end{array} \quad (10)$$

According to the above rules we can fuse the data from frames sequence to recover an intact background, if each background point color is known, the intact background can be recovered by combination of all the known point. An example of BIF result will be given n *Sec* 4.

3.2 Target extraction and tracking

After BIF, the scene background is a known condition, the target extraction procedure will become a simply subtraction of the intact background from the scene, this is the *background differencing algorithm*. The residual is the extraction result of target.

BIF based algorithm assumes that Background Information change slowly than the targets' information, as defined in Assumption (1), so it's not necessary doing BIF computation in each frame. In road vehicle monitoring application the scene background changes slower (mainly caused by intensity of illumination change in different time of a day and weather condition change) than vehicle's movement, the intervals we need do the next BIF computation is decided by environment condition changing rate, In our experiment we choose the intervals as 10 minutes. If the video capture frame rate is 15 fps (frame per second), then the intervals equals to 9000 frames. This significantly decrease the computation cost because in every 9000 frames we only need do BIF computation once (BIF need about 15 to 30 frames, or 1 to 2 second to to perform computation, as detailed discussed in *Sec* 4.2).

After *background differencing* (or background subtraction), the target is extracted frame by frame.

The *tracking algorithms* used in BIF based tracking can be based on available techniques, such as *background differencing* [1, 11], *optical flow techniques* [3, 4], *dynamic template* [6], *temporal coherence based techniques* [3], *perspective transform based techniques* [7], *Feature point based techniques*[8], *active contours techniques*[9], etc. Here we do not discuss them in detail. We only focus on the BIF procedure and just present a tracking result in *Sec* 4, the result is based on the combination of BIF and temporal coherence based techniques.

4 BIF examples and parameter discussion

4.1 Example: Fusion of a frame sequence to recover an intact background

In the viewpoint of BIF based method, the realistic video surveillance background is assumed to be intact, but in practical applications, the scene background is frequently *Hit* by target. Redundant information provided by frame sequence can help us remove the *Hit* point in the scene.

Fig.5 shows the fusion result of a frame sequence.

4.2 BIF Parameter discussion

4.2.1 *AssureCount*

The AssureCount is the most important parameter in BIF algorithm.

On one hand, if it is set to a small number, that means the fusion decision speed is fast, the smaller the AssureCount, the fewer the frames required to decide the background color of a specific position.

Fig. 5 BIF result by fusion of different frames

If it is set to a large number, that means the decision speed is slow.

On the other hand, small AssureCount means BIF need a small number of frames to decide a background point color, this would lead to low assurance rate of the fused result. If the AssureCount is set to a large number, the fusion result is credible. But too large a number of AssureCount will make it hard for fusion procedure to draw a conclusion, because the background is frequently *Hit* by target. To determine a reasonable parameter range, we proposed the following equation:

Suppose the vehicle velocity is V1 to V2, in urban area, it is limited to maximum 80km/h.

Suppose the distance between vehicle heads is ranged from B1 to B2, *e.g.* from 5m to 25m

Suppose video capture frame rate is 15 fps.

The AssureCount range can be determined by:

$$\frac{B1}{V2*1000/60*60*15} \quad , \quad \frac{B2}{V1*1000/60*60*15}$$

That is:

$$\frac{54*B1}{V2} < AssureCount < \frac{54*B2}{V1} \qquad (11)$$

if V1=30km/h, V2=80km/h, B1=5m,B2=25m,then the AssureCount ranged is 3.375 to 45 frame.

If AssureCount is outside of the range, BIF is instable or impossible, generally, we choose the smallest integer in the range as AssureCount parameter, here we choose 4, because 3.375<4<45.

4.2.2 DitherThreshold

DitherThreshold is another important parameter in BIF algorithm. DitherThreshold is determined by the environment variance and camera instability factor. If it is set to a very small number, *e.g.* set to zero, the NoHitFrameCount in some position may not be greater than the AssureCount forever, as shown in Eq. (9); this will cause undetermined point in background; If it is set to a very large number, The NoHitFrameCount can easily be greater than AssureCount, this will cause false determination of background color. From the experiences got from many of our experiments, we suggest it should be chosen to be 4 to 6 if video sequence is in 256 gray mode.

4.3 Target extraction and tracking result

The target extraction result based on BIF is shown in Fig.6, and tracking result based on *Background Differencing* and *Temporal Coherence based tracking algorithm* [4] is shown in Fig.7

Fig.6 target extraction result

5 Summarize to BIF based tracking

BIF is not only a method, but also a new viewpoint when tackle the challenges and difficulties in video target extraction. Most of available tracking techniques are target

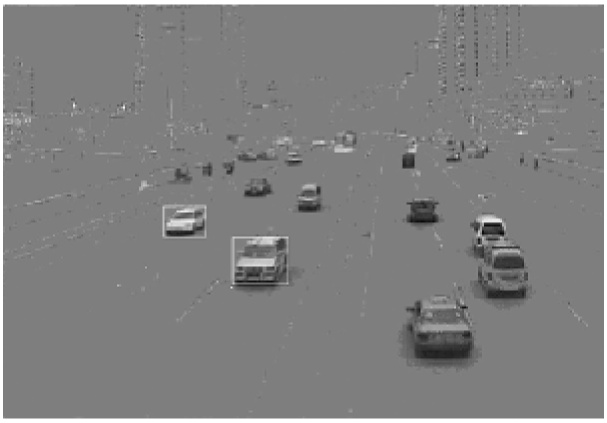


Fig.7 Tracking result based on Background differencing and temporal coherence of target

based, they focus on the problem of what features the target have, and how does these features change with target movement, these algorithms generally have a comparatively high computation cost and poor performance if the target features, such as edges, grains, lighting, 2D shape, feature point, feature line, etc. change significantly caused by targets' complicated movement.

The major innovations of our work are:

(1) Proposed a BIF idea in target tracking

(2) Proposed basic concepts and definitions in BIF, such as Hit, NoHit, NoHitCount, NoHitFrameCount,

AssureCount, DitherThreshold, BIF Assumption

(3) Proposed a frame sequence based image data fusion algorithm

(4) Proposed a two step tracking method in target tracking. At first step, algorithm focus on recovering *an intact background*, at the second step it extract target by *background differencing* algorithm. These two steps eliminate the most of the difficulties and challenges in extracting of the moving target from time-varied background.

Our experiments had proved that this algorithm is stable, feasible, and effective.

Reference

[1] Haritaoglu, D. Harwood, and L.S. Davis, A real-time system for detecting and tracking people..*J. ECCV98*, 21/2, 1998.

[2] Fusiello, E. Tommasini and Roberto, Improving feature tracking with robust statistics, .*J. Pattern Analysis & Applications*, Vol.2, 312-320.1999

[3] R.P.Wildes and L.Wixson,Detecting salient motion using spatiotemporal filters and optical flow. *J. DARPA Image Understanding Workshop*, Vol.1, 349–356, November 1998.

[4] Cohen and I. Herlin, Optical flow and Phase Portrait Methods for Environmental Satellite Image Sequences, *J. ECCV*, 141–150, Cambridge, April 1996.

[5] Isaac Cohen, Detecting and Tracking Moving Objects for Video Surveillance. *IEEE Proc. Computer Vision and Pattern Recognition*, 23-25, Jun.1999.

[6] Isaac Cohen and G´erard Medioni, Detecting and Tracking Moving Objects for Video Surveillance, *IEEE Proc. Computer Vision and Pattern Recognition Jun.* 23-25, 1999.

[7] M. Irani, P. Anandan, and S. Hsu, Mosaic based Representation of Video Sequences and their Applications. *IEEE. Int J. ICCV*, 605–611, Cambridge, Massachusetts, June 1995.

[8] Fusiello,E.Tommasini and V.Roberto, Improving Feature Tracking with Robust Statistics, *J. Pattern Analysis & Applications*,Vol.2, 312-320.1999.

[9] M.Isard and A.Blake. Contour Tracking by Stochastic Propagation of Conditional Density, *Proc, European Conference on Computer Vision* 343-356, Freiburg,Germany,1996

[10] Gilad Halevi and Daphna Weinshall, Motion of Disturbances: Detection and Tracking of multi-Body non-Rigid Motion , Machine Vision and Applications Vol 11,122-137, 1999.