Fast Occluded Object Tracking Technique with Distance Evaluation

T.H. Tsai[1]**, Y. C. Liu**[2]**, T.M. Chen**[2]**, C.Y. Lin**[1]
[1]Department of Electrical Engineering, National Central University, Taiwan, ROC
[2]Department of Computer Science, National Taipei University of Education, Taiwan, ROC

Abstract

—Objects in the world exhibit complex interactions. When captured in a video sequence, some interactions manifest themselves as occlusions. A visual tracking system must be able to track objects which are partially or even fully occluded. Occlusion is a difficult problem in target tracking, especially when users need to track multiple targets simultaneously and maintain the target identities during tracking. With this complex design issue, in this paper we present a fast occluded object tracking technique. The purpose of this paper is to present a novel approach to cope with the occlusion problem explicitly. Our approach is by use of the occlusion handling with the measurement on object distance and power. Our system has the advantages of low cost and low complexity, and can be realized in real time system.

***Index Terms*—power, distance, occlusion, track, real time.**

1. INTRODUCTION

As an active research topic in computer vision, visual surveillance in dynamic scenes attempts to detect, recognize and track certain objects from image sequences, and more generally to understand and describe objects' behavior. The aim is to develop intelligent visual surveillance to replace the traditional passive video surveillance that is proving to be ineffective. The number of cameras needed exceeds the capability of human operators to monitor all of them. In short, the goal of visual surveillance is not only to put cameras in the place of human eyes but also to accomplish the entire surveillance task as automatically as possible.

Surveillance systems generally track moving objects from one frame to another in an image sequence. The tracking algorithms usually have considerable intersection with motion detection during processing. Tracking over time typically involves matching objects in consecutive frames using features such as points, lines or blobs. Tan delivered an overview paper

[1] on visual surveillance of object motion and behaviors, and he explained the relationship between segmentation, tracking and occlusion clearly. He divided tracking and surveillance system into single and multiple cameras. In practice, self-occlusion, and occlusions between different moving objects and the background are inevitable. For multiple object tracking, it is very important to maintain the tracking of moving objects before, during, and after object occlusions. The loss of a tracked object leads to the impossibility of analyzing its behavior.

The other surveillance systems attempt to track ofindividuals thanks to a cooperative approach between spatial detection and temporal tracking. Spatial detection is based on skin color classification and shape analysis by morphological tools. Temporal tracking is based on the analysis of the optical flow [2]. Most single camera methods cannot achieve the above in real time systems easily. Because they use a lot of objects' color information in pixels and their computational cost is too high. Only a few outstanding systems [3] can implement occlusion in a real time environment. In addition, they need extra memory space to store every object's color pixels information and their cost is still higher in hardware.

On the other hand, some existing systems focus on occlusion problems by using multiple cameras. Multiple camera-based visual surveillance systems can be extremely helpful because the surveillance area is expanded and multiple view information can overcome occlusion [1]. Other more advanced multi-camera surveillance systems need more beforehand preparation [4] [5], such as having to create an object model to scale in advance. This makes it harder to implement algorithm in an embedded system.

Inevitably, most of them need to use color or gray scale pixels information and extra memory space which also a bottleneck to perform in a real-time system. Actually, the low-cost and low-complexity problems have been indicated as the bottleneck for deploying surveillance systems in object-tracking, even in occlusion handling problems. In short, this will be less beneficial because the objects' color information not only needs extra computational costs, but also additional hardware costs.

Based on these observations, we propose a novel occlusion handling approach. Our system can keep computational complexity very low while still maintain a good tracking result. With the benefit of high definition on our method, the general surveillance systems even with a low cost camera are still workable. In addition, our method is optimized with high speedup ratio. This paper is organized as follows. The novel occlusion detection and tracking method is developed in Section II. In Section III, experimental results are provided. Finally, in Section IV, conclusive remarks are given.

2. THE PROPOSED ALGORITHM

The proposed tracking and occlusion handling method are based on constraining the speed and size changes of the object. In the following, we present occlusion detection and tracking method including occlusion handling which will be used under the above assumption.

2.1. OVERVIEW OF THE FRAMEWORK

In video sequences, we have to get video of interesting (VOI) region in every new frame. We use the Gaussian Mixture Model [6] to get every VOI. To simplify the computational complexity of the mixture models, grayscale images were used instead of the three independent color channels. After converting the original color images to gray scale, the algorithm determines the background and foreground blobs and outputs a binary representation of the foreground. Then we try to filter some of the noise (too small) below the threshold and labeling foreground with connected pixels. After labeling the foreground objects, every individual foreground blob without connected pixels has a unique number in the image. The block depicted by the dashed lines on the left part of Fig. 1 performs occlusion detection to find occlusion in advance. If no occlusion happens, we use the nearest neighbor data association to keep on tracking. Otherwise, the remainder of the proposed framework, which is indicated by dashed lines on the right of Fig. 1, deals with corresponding mislabeled objects. Then, we use the nearest neighbor data association to keep on tracking. Finally, we have to update object's information.

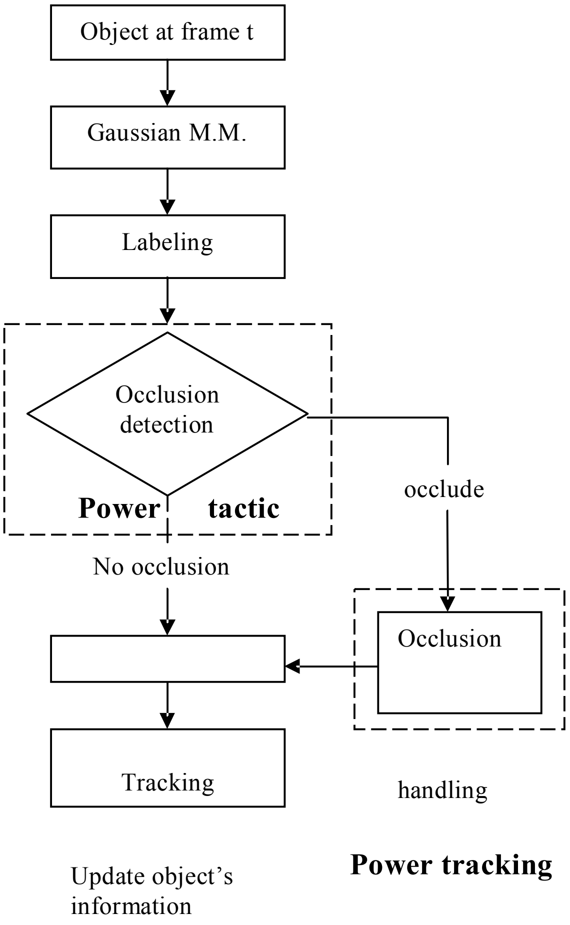


Fig. 1 System architecture consists of power tactic and power tracking method.

Mubarak used the method [7] which is used with both the distance between the objects and the change of the object size. Thonnat had been used the same method and proven to fail in their article [8]. I will describe their method clearly and shortly. They proposed an ambiguity matrix which also

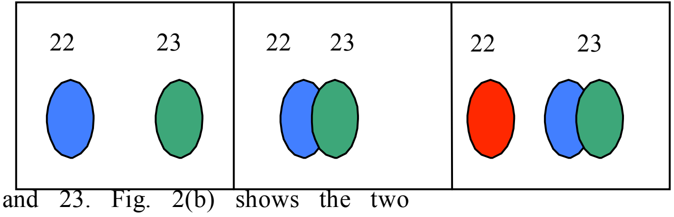
people being occluded into one person and they are still numbered 22 and 23 respectively. Unfortunately, Fig. 2(c) shows the leftist target (numbered as 22) is mislabeled because a third person appears and this person was not detected previously on Fig. 2(b), simultaneously merges with the two original people. Using distance and size to detect occlusion is easy, but it can't detect above popular case in a surveillance system. So we prove a novel method to detect occlusion in popular surveillance system.

In our paper, we assume that all ROI rest on the ground plane. Of course, this assumption may seem restrictive (cannot find people on the rooftops), because we want to set up the most popular surveillance scenes.

2.2. POWER CENTER POINT

The scene is a 3-D domain and the image is 2-D domain. After labeling the moving objects, it is very hard to piece together the original 3-D information, especially in occlusion status. This is a non-reversible process. Now we attempt to create a virtual-camera between 2-D and 3-D plane, it can play a bridge role to communicate them. Our critical idea is ROI who rest more close to the camera, it also have more priority to cover up other ROI. In the same way, it is like that prior ROI has more power to appear in the scene.

Under a zero-skew, unit aspect ratio perspective camera model, we illustrate the 3-D domain relationship on Fig. 3. Consider for example a coordinate system *(O, i, J, K)* [9] attached to a camera, whose origin *O* coincides with the lens, and vectors i and *J* to form a basis for a vector plane parallel to the image plane *Π'*, itself located at the positive distance *f*, from the lens along the vector *K*. The line perpendicular to *Π'* and passing through the lens is called the optical axis, and the point *C'* where it pierces *Π'* is called the image center. Draw a line, parallel to vector *J*, and perpendicular to *C'* and passing through *C'*, named it *CP*. Let *P* denote a scene point with coordinate *(x, y, z)* and *P'* denotes its image with coordinates *(x', y', z')*, so *(x' ,y', z')* = λ (x, y, z). Since *P'* lies in the image plane, we have *z' = f*. Since the three points *P, Q* and *P'* are collinear, we have *OP' = λOP* for some number λ, where *x' = f * x/z* and *y' = f * y/z*. When a camera is used with the above information, it will always remain at a roughly constant distance from the scene, we can go further and claim that all light rays parallel to the *K* axis and orthogonal to the image plane *Π'*. We make projection to the camera on line CP, but it does not belong to the image plane, we make a point and we name it Power Center Point (PCP).



and 23. Fig. 2(b) shows the two

(a) (b) (c)

Fig. 2 (a) Two persons are tracked (numbered 22 and 23) (b) shows that thanks to compound targets the tracked of the group and of each person are not lost. (c) the target(numbered 22) is mislabeled because the third person appear and simultaneously is merging with the first two persons.

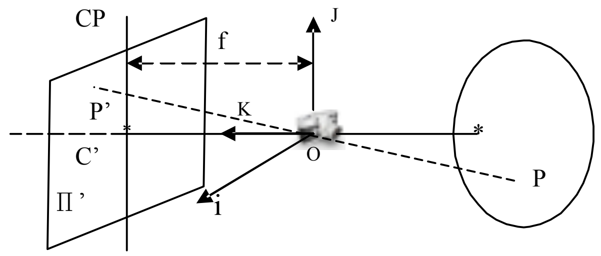


Fig.3. Scene plane is projected on image plane via camera

2.3. POWER TACTIC

According to Emmert's law, it helps us to get objects' relative position roughly. He used distance and size to judge which objects are occluded. We assume the objects' sizes are almost the same and occlusion is closely related to distance in one image situation. In real world, we use camera to monitor the scene for a while, so we also can use the scene temporal info. We combine PCP and the scene temporal info. to detect occlusion in the scene, and we name it Power Tactic (PT).

We illustrate PT more detail by Fig. 4. According to Emmert's law, *B* is in front of *A*. Draw the shortest line to connect *PCP* and *A*, *PCP* and *B*, and we name them *PA* and *PB* respectively. According to Pythagoras' Theorem, *PB* is shorter than *PA*. Similarly, *B* is also in front of *A*. Sometimes objects are too close or calculate error, we may judge inaccuracy. It is similar to nearsighted in human eyes. But we can get objects' relative position roughly. Furthermore, we must make sure the shortest distance between every object and PCP is unique. If objects have the same distance, we will predict their next position and recalculate theirs power again. It can help us get a permutation by every ROIs' power with our PT method. We can rank every object by its shortest distance between PCP, and then assign its power by this permutation in sequence. We have the following relationship:

$$\mathrm{argmax}_s^P (P = 1.....n) \qquad (1)$$

From equation 1, where *P* is the shortest distance between object and PCP ; *n* = number of objects; *s* = unchecked objects.

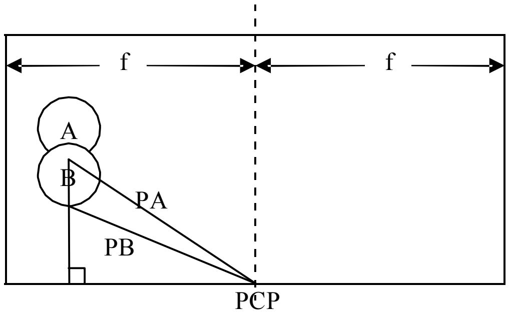


Fig.4. (b) the relationship between objects.

We also use comics in Fig. 5 to present ours PT why it can solve trap occlusion detection in other article we mention above [7][8].

Step 1) there are more than one ROI in frame F2, we calculate the objects' initial velocity and position. The threshold is closely related to object's velocity, in our experiment assign 30. We assume object's velocity doesn't change extremely. If distance is still higher than the threshold in frame F3, we do nothing change.

Step 2) If the distance is shorter than the threshold in frame F4 ~ F7, the occlusion will occur at any moment. We can use our method, PT, to prevent and detect occlusion in advance.

Step 3) Occlusion occurs in frame F5 and F6, ROI D appears at the same time. We use object tracking method (the feature-matching algorithm to make correspondence) to perform occlusion handling and we have to readjust every ROI's power until the distance is bigger than the threshold.

2.4. OBJECT TRACKING

Now, we can detect occlusion in the scene easily, and we use combine PT and general tracking method to propose a novel Power Tracking Method to perform object tracking.

Because we can use PT to prevent occlusion in advance, single object or multiple objects without having power themselves, it does need to use extra computation complexity to solve occlusion status. So we still can use the nearest-neighbor prediction and data association to perform tracking well. Once it starts to perform PT, each of video objects changes its status (initial status transform into power object) and their property is defined as follow:

$$VO_D^I\,(I = 1 \ldots N) = F(f, n, c, v, e, p, \;,\;) \qquad (2)$$

From equation 2, where VO is video object, *D* is discrete time, *N is* number of moving objects. And *F(.)* set of features which include frame number *(f)*, labeled number*(n)*, centroid *(c)*, velocity *(v)*, the position of the bottom of object (e), power *(p)*, occlude *(o)*.

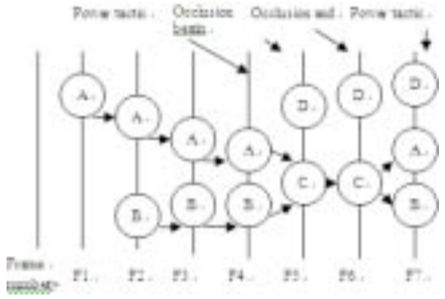
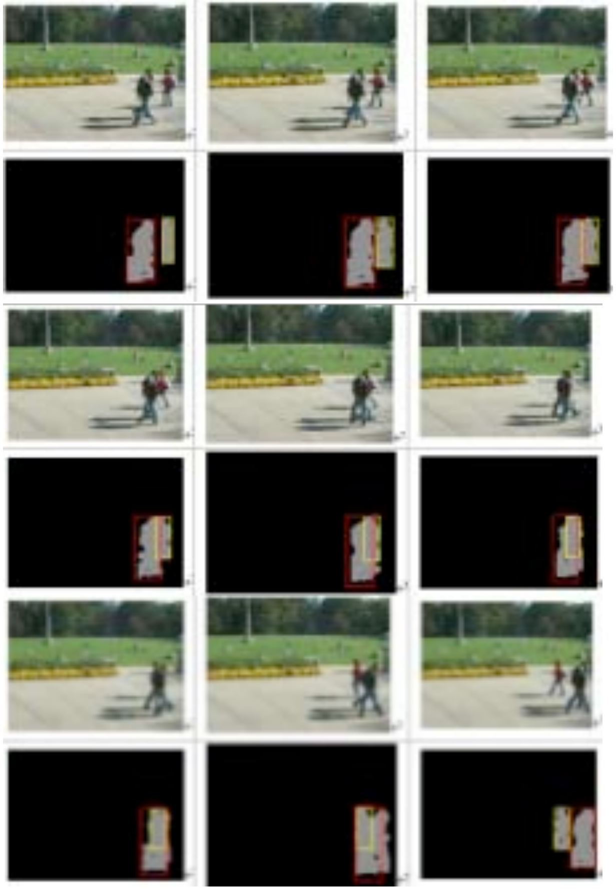


Fig. 5. F1 ~ F7 represent frame number. F5 and F6 occlusion occur and D appear at the same time. There are three objects in F5 and F6. F7 occlusion end.

3. EXPERIMENT RESULTS

In the framework of a video-surveillance application we have developed a system implementing in C++ language the tracking algorithm described in this paper and performs it on Intel 2.8G and windows XP. To validate our algorithm we use campus video sequences, including students walking in and out of the stationary camera's view. The sequence contains 300 frames. For a binary image frame at 320 x 240 resolution the total time of our proposed algorithm alone (tracking process alone) is around 0.4 second. Because our method is focus on tracking and occlusion handling, in order to achieve fair to compare. We only compare with other algorithm in the same field. We use Fig. 6 to illustrate the result in detail. We also list a compare table (table1) that they also have proposed the occlusion handling method before. Our method can perform in binary images and we only held some critical information in every object. Obviously, our method is faster than them very much. Furthermore, our method has a robust occlusion detect method that can help us save redundant calculation and perform occlusion handling in critical time. So our method is faster than other method at least thirty times.



(fig. 6.) (left to right and top to down) left two persons are original a group in the same direction and red clothes person in another direction. It should be labeled two objects (red and yellow). Even in fully occluded, we can track.

Tracking process with occlusion handling	
	Average running time
[15]	0.05 sec.

Table 1. Comparison results

4.CONCLUSION

In this paper, we describe the framework of our video surveillance system and provide the algorithm and implement results of our current work on occlusion handling. Our system work well in real-time. It can track moving objects even under occlusion accurately. And our algorithm can perform tracking well even use a single-number (black-and-white camera). We implement our algorithm in binary image, compare with other systems that use multiple-camera or color-image to perform occlusion handling, our system can save more money in hardware and complexity cost. Our system also use a simple and critical technique to deal with occlusion well

5. EFERENCES

[1] Weiming Hu, Tieniu Tan, *Fellow, IEEE*, Liang Wang, and Steve Maybank, “A survey on visual surveillance of object motion and behaviors,” IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 34, NO. 3, AUGUST 2004

[2] Casas, J.R.; Sitjes, A.P.; Folch, P.P “Mutual feedback scheme for face detection and tracking aimed at density estimation in demonstrations,” Vision, Image and Signal Processing, IEE Proceedings-Volume 152, Issue 3, 3 June 2005 Page(s):334 - 346

[3] Haritaoglu, I.; Harwood, D.; Davis, L.S. “W4 real-time surveillance of people and their activities,” Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 22, Issue 8, Aug. 2000 Page(s):809 – 830

[4] Mohan, A.; Papageorgiou, C.; Poggio, T.; **“**Example-based object detection in images by components,**”** Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 23, Issue 4, April 2001 Page(s):349 – 361

[5] Tao Zhao; Nevatia, R “Tracking multiple humans in complex situations,” Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 26, Issue 9, Sept. 2004 Page(s):1208 – 1221

[6] Stauffer, C.; Grimson, W.E.L “Learning patterns of activity using real-time tracking,” Pattern Analysis and Machine Intelligence, IEEE Transactions on , Volume: 22 , Issue: 8 , Aug. 2000 Pages:747 – 757

[7] Yilmaz, A.; Xin Li; Shah, M “Contour-based object tracking with occlusion handling in video acquired using mobile cameras,” Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 26, Issue 11, Nov. 2004 Page(s):1531 – 1536

[8] Bremond, F.; Thonnat, M “Tracking multiple nonrigid objects in video sequences,” Circuits and Systems for Video Technology, IEEE Transactions on Volume 8, Issue 5, Sept. 1998 Page(s):585 – 591

[9] David A. Forsyth, Jean Ponce, Computer Vision: A Modern Approach , Prentice Hall 2002