Journal of Digital Contents Society Vol. 16 No. 6 Dec. 2015(pp. 901-908)
http://dx.doi.org/10.9728/dcs.2015.16.6.901

투시적 깊이를 활용한 중첩된 객체의 관계추적

박화진*

요 약

스토킹과 같은 장시간 동안의 이상행위를 추적하기 위해선 네트워크로 연결된 다중 CCTV환경하에서 객체간의 관계를 지속적으로 추적하는 시스템이 매우 필요하다. 그러나 추적과정에서 자주 발생하는 객체의 겹침문제가 해결되지 않는다면 객체 추적이 중단되거나 다른 객체로 대체되는 등의 치명적인 오류가 발생할 가능성이 농후하다. 본 연구는 기 설치된 CCTV를 최대한 활용하기 위해 투시적 투영깊이 및 객체특성을 활용하여 겹침문제를 해결함으로써 중첩된 객체 관계를 지속적으로 추적가능하게 한다. 객체간 겹침문제 뿐만 아니라 배경에 포함된 객체 즉 벽이나 기둥 등의 객체와의 겹침문제도 함께 다룬다.

키워드 : 객체관계추적, 겹침객체, 투시적 투영깊이, 다중 CCTV, 이상 행위 감시 시스템

Relation Tracking of Occluded objects using a Perspective Depth

Hwa-Jin Park*

Abstract

Networked multiple CCTV systems are required to effectively trace down long-term abnormal behaviors, such as stalking. However, the occluding event, which often takes place during tracking, may result in critical errors of cessation of tracing, or tracking wrong objects. Thus, utilizing installed regular CCTVs, this study aims to trace the relation tracking in a continuous manner by recognizing distinctive features of each object and its perspective projection depth to address the problem with occluded objects. In addition, this study covers occlusion event between the stationary background objects, such as street lights, or walls, and the targeted object.

Keywords : object relation tracking, occluded objects, perspective projection depth, multi-CCTV, abnormal behavior surveillance system

1. Introduction

Study of object tracking has been of high priority in the area of visual recognition. The early research began with object recognition, which made it possible to track objects and recognize the movements of the objects by way of analyzing traces of the objects, and now focuses on prediction of objects' movements. In this process of study, the problem of occluded objects has been tried to solve in various ways. To recognize the depth, one of the solutions provided recently focuses on the functional aspect of apparatus, not on software. In other words, depth is measured using a stereo camera or a function of multi-view of a single camera, and the occluded objects are identified based on the

※ Corresponding Author : Hwa-Jin Park
Received: December 10, 2015
Revised: December 25, 2015
Accepted: December 29, 2015
* Sookmyung Women's University
Tel: +82-2-710-9204, Fax: +82-2-710-9704
email: phj2000@sm.ac.kr

depth of objects.

This paper studies object relation tracking that has been advanced in the area of object tracking study and focuses on the methodologies of continually tracking the relation of occluded objects. The object relation tracking is a process of calculating the relations among different objects, and the relations are measured in a continuous manner to selectively track objects that meet certain criteria. Motivation of this study on the object relation tracking is to develop an abnormal behavior monitoring system. Especially we are focusing on stalking, one of the abnormal behaviors, so that the system can prevent threatening behaviors by giving a warning to a potential victim after recognizing a stalker in advance. Multiple cameras identify passersby and recognize a staker who presents behaviors associated with stalking other objects by calculating relations of objects within a certain distance in a continuous manner[1].

However, if an object is occluded with another object, the recognition information is lost. Thus, the calculation of the relation tracking for the object can be halted, as the object is recognized as a different object, even though the object remains the same. Furthermore, we have to consider other scenarios to effectively identify stalking behaviors, including an object being occluded behind walls, or streetlight, which can be used by stalkers to conceal their 0bodies.

This paper applies occlusion solution methodologies to various occlusion cases using projective depth, in order to recognize the occluded objects and perform the continuous relation tracking methodologies for separated objects.

This paper consists of as follows. Chapter 2 summarizes related works and the existing methodologies to occlusion event. Chapter 3 explains the process of a stalking detection system. Chapter 4 suggests a method for applying the perspective depth to separate the occluded objects. Chapter 5 describes the implementation and its analysis. Finally, Chapter 6 provides results and future study.

2. Relating Study

2.1 Occlusion handling

The occlusion situation can frequently take place during tracking. Without solving the overlapping problem, it would be impossible to effectively track objects.

The situations of the overlapping can be categorized as self occlusion –a part of oneself is overlapped with another part of oneself, object occlusion –more than two objects are overlapped, and background occlusion –backgrounds such as streetlight, or walls, are overlapped with an object.

A recent study defined the problem of overlapping as 7 status and categorized the existing solution methodologies[4].

In order to solve the occlusion problem, the study has developed into the following methods: depth analysis, a fusion method, and an optimal camera placement method[3]. When two objects are overlapped, if the depth is shorter, it means that the object is closer to the camera. The depth analysis uses such phenomenon. The fusion method tracks characteristics and movements of objects and predicts the next movements, based on the information, which makes it possible to track locations of the overlapped objects. Lastly, uses of multiple cameras and placing cameras in multiple locations are noted in the optimal camera placement method. For example, when cameras are attached to the ceiling at an angle of 360˚, the occlusion does not occur, as it films from the above, although slight visual distortion is inevitable. If more than two cameras are used, the value of object depth can be easily measured.

Recently cameras, including infrared cameras or stereo cameras, that have a function of calculating the object depth, are developed in the market. But, most of the CCTVs on the road are not equipped with functionality of measuring the object depth. Therefore, it would take enormous time and costs to replace all of them to resolve the problems pertinent to occlusion. So, this paper suggests fluoroscopy projective depth method which is compatible with the existing regular CCTVs.

2.2 Perspective projection depth

The transforming process of a 3-dimensional object is called projection which consists of parallel projection and a perspective projection. Perspective projection, which is more realistic than parallel projection, is a method of projecting to human vision or camera images. A vanishing point denotes a point in which a parallel line merges from the optic angle. Vanishing point can be classified with 1-point and 2-point perspective views based on the number of the points. Projection depth indicates the distance between the optic point in the line and the vanishing point. In a projective space, most affine transformations, usually invariant in 3-dimensional Euclidean space, are not invariant, but a cross-ratio. In other words, the relative depth remains constant because of the cross ratio of four collinear points a, b, c, and d[2].

cr (a, b, c, d) = ratio(a, b, d) / ratio(a, c, d)

3. Inter-Object Relation tracking

3.1 necessity

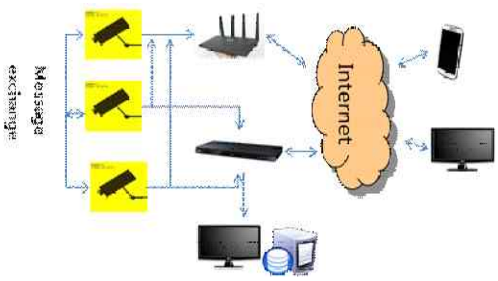
Although most abnormal behaviors presented by under two individuals could be successfully detected by the abnormal behavior detection system, detection of behaviors like stalking that requires long time observation through multiple CCTVs is on the early stage of study. It is because behaviors like stalking are secretly proceeded and can be hard to detect. Stalkers can deceive the target object and CCTV by changing their clothes frequently or hiding themselves. Furthermore, stalkers' behaviors are irregular and difficult to set certain pattern.

Despite of the aforementioned challenges, the ultimate goal of this study is to identify the stalker within that time frame by the fact that a stalker follows his target for a long period of time. By doing so, inter-object relation tracking has been developed to identify such abnormal activities.

3.2 Multi-CCTV surveillance system Enviroment

Network based system environment is required for long-term monitoring system to trace objects for long time, at least 30 minutes. A graphic model that portrays the networked system is as followed in (Figure 1).

(Figure 1) Diagram of networked surveillance system



The classified Single CCTV module and multiple CCTV module according to abnormal activities are employed to detect them efficiently. (Figure 2) shows the flow of information pertinent to the objects, shared among CCTVs. Surveillance control server consists of MC-ABM (Multi Camera-Abnomal Behavior Module), MC-IMM (Multi Camera-Abnormal Information Manage Module), and cameras consist of ORM(object recognition module) and SC-ABM (single Camera-Abnormal Behav ior Module)[6].

3.3 Inter-Object Relation

Relation of inter-object are defined as the following 3 functions.

- Member (n, m, t, min, max) = {m| n∈ image(t), min<dist(t),(n, m)<max}, n & m are blobs, t is a frame, dist is distance.
- Following (n, m) = {m|∀m ∈ member (n, m, t, min, max), $(n \cdot \vec{m})<0$, m is ahead of n}
- Followed (n, m) = {m|∀m ∈ member (n, m, t, min, max), $(\vec{n} \cdot \vec{m})<0$, n is ahead of m}

Member m refers to any objects within the certain range from one object. Following(n, m) is defined as n is following m, and Followed(n,m) is defined as object n is followed by object m. Although the definitions of Following(n,m) and Followed(m,n) may be interchangeable at a certain condition, in most of cases in which there are multiple objects m, Thus, Following(n,m) is not equal to Followed(m, n).

Information related to particular features of objects, locations, colors of clothes (upper, lower) and bags, heights, speed of movement, and pathways are recognized and saved in the surveillance system. Then, the relation information is saved in the format based on the equations.

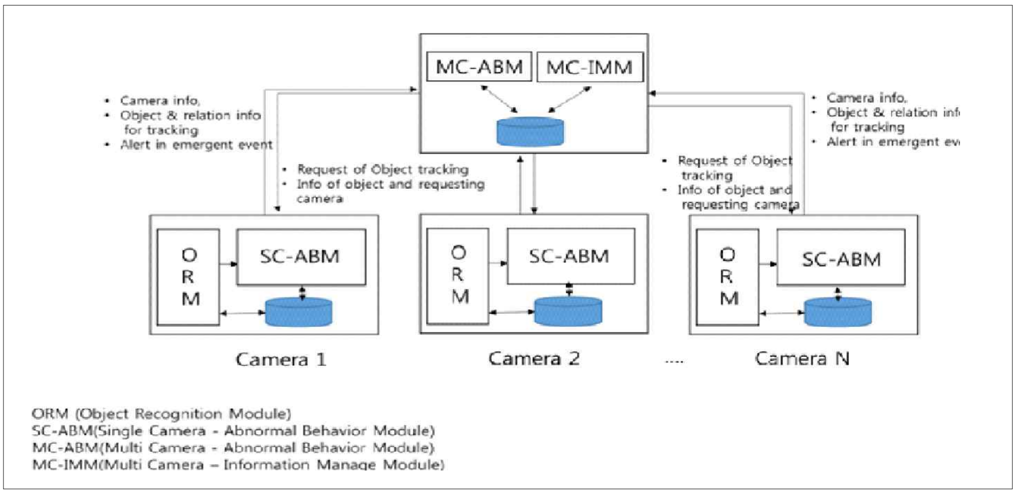
$$Relation(n,m) = \begin{cases} 0, & if\ n = m \\ \ , & if\ (n \cdot m) < 0 \\ FD, & if\ n\ is\ followed\ by\ m \\ FW, & if\ n\ is\ folloing\ m \end{cases}$$

In addition, an entering time and an outgoing time of each object from each image are recorded for tracing identical objects between CCTVs.

3.4 Detecting Process of Stalking

The following steps are criteria for identifying stalking behaviors.

(Figure 2) Tracking Module structure



(1)Find out whether there is an object who comes close by above 90 percentages, using member which has group characteristics data saved for at least 30 minutes for each individual object.

(2)Find out whether there is an object who comes close by above 90 percentages, using “followed” relation which has group characteristics data saved for at least 30 minutes for each individual object.

(3)If there is an object in (1) and (2), find out whether he or she has identical tracks with individual objects and if one does, it is considered as stalking.

4. Occluded object's relation tracking

4.1 Occluded object recogntion

Occlusion of objects results in a recognition error of the objects as new objects at the detecting phase. Therefore, it is necessary for the system to perceive the occlusion event, and after the perception, the occluded objects need to be separated to a few independent objects. Occlusion can be detected rather easily. Moving object can be recognized through the means background_diffimage. If any significant changes in the size of the object are detected, then the system perceives the change as the indication of an occlusion event.

The overlapped objects are differentiated respectively, and then each object is compared and identified to the objects from the previous images. This process is called Separation of Occluded Object. In order to differentiate the overlapped objects, this study utilizes foot based perspective projection depth method. Furthermore, to effectively identify objects even in the case of discontinued tracking due to overlapping of objects behind street lights, or walls, the system predicts an estimated route based on the previous information.

4.2 Foot based perspective projection depth

The perspective projection of a 3-dimensional object is transformed according to the length of a focal point, the height, and angle of a camera. This can be seen in the matrix of intrinsic parameter. We will use the depth of position of object's feet that is projected to the images in order to distinguish the occlusion of each object. It is also assumed that a human stands straight up and walks, and the ground is flat. The position of feet is defined as a value of a medium point of a bounding box bottom. The depth of feet is replaced with the value of y by the cross ratio. The horizontal line with a vanishing point can be gotten by re-scaling of the value of y[12].

4.3 Occlusion occurring situation

Whether there happens an occlusion can be checked by comparing the number of blobs and the size of its corresponding blob of its current and previous images. Under the assumption that one unit of an object consists of a blob, the next step will be one of several possible cases. n(po) indicates number of objects in previous image while n(co) refers to number of objects in current image in <Table 1>.

<Table 1> Occlusion occurring situation

comparison of n(po) and n(co)	possible cases	occl-usi-on
<	• a new object enters from a border	
	• a new object appears	

	from somewhere in an image not an border	
	• an occluded object in previous is separated	
>	• an object goes out through a border	
	• an occlusion occurs.	Y
	• an object hides himself behind objects in backgound like a wall or a pillar.	Y
=	• no change occurred regarding objects	
	• an object disappears and another object appears at the same time	Y

4.4 Occluded objects relation tracking

The occluded objects are categorized based on the types of overlapping objects, which can be either moving objects, or stationary background objects. Moreover, based on the area of occlusion determines whether it is partial, or complete occlusion.

If partial occlusion takes place between an object and a stationary background object, then the sizes of the blobs get smaller and smaller until complete occlusion occurs, and the blobs disappear from the screen. In this case, the object is still in the center of the camera view, and from the point of disappearance initiates relation tracking. When the occluded objects are separated from the background objects, the system can be able to track the objects without any error.

On the other hand, when partial occlusion takes place between mobile objects, the system separates each overlapping object using perspective depth information. The system continues tracking the relation of the separated objects. In case the object overlaps with moving objects, partial and complete occlusion object relation trackings are simultaneously carried out. After detecting occlusion of inter-objects in one of cases in <Table 1>, object relation tracking of these objects will be performed in the following three steps.

1) Check if there are any overlapped objects using bounding box information.
2) Entry of objects of a current image after checking the depth information of overlapped objects, separating the overlapped objects and calculating positions of current frame referring to previous velocity.
3) Perform tracking of object relation to all objects including the separated ones.

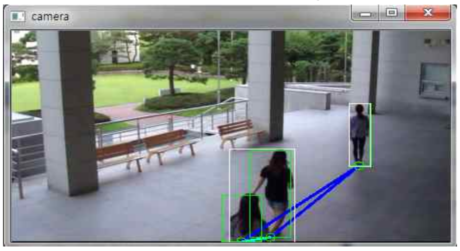
5. Implementation and Discussion

(Figure 3) demonstrates the situation before resolving the occlusion effect in which the object was surrounded by the white box. In this situation, two objects are overlapped and considered as one unit, bounded with a white rectangular box. Here, two objects are separated using the difference in perspective depth, and, as a result, two objects are bound by green box respectively and identified as distinct objects. The feet of objects are indicated by a circle underneath the rectangular box and the depth value is calculated using equation, max(depth)-y. In (Figure 3), the blue line indicates his followers and the object walking ahead. The target is denoted by a circle at the end of the blue of the object, indicating a situation in which he is being followed. Additionally, the occluded objects are regarded as separate objects and the relation between objects is maintained. Images of (Figure 3) (a) and (b) illustrates the occlusion occurring and separation and relation tracking between objects. The relation is shown by a directed straight line in (Figure 3) (a)and (b). Blue line shows its followers of the person walking ahead. (Figure 3) (c) and

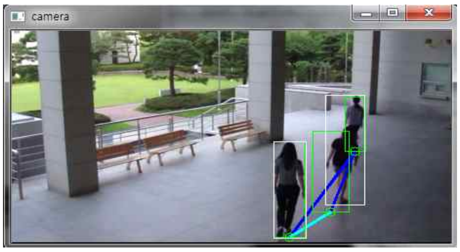
(d) shows the contour image of (a) and (b) respectively.

In this study, the relation has been established based on the locations of objects' feet. In other words,

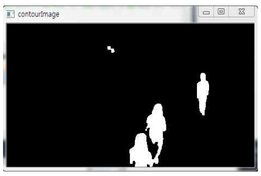
(Figure 3) Occluded object's relation in original images and Contour Images



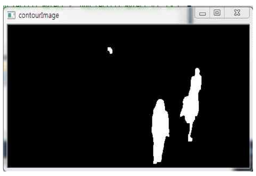
(a) Original Image 1



(b) Original Image 2



(c) contour image of (a)



(d) Contour Image of (b)

6. Conclusion and Future Research

In order to trace long-term abnormal activities such as stalking, we need a system that can track objects' relation continuously in multiple cameras and provide a solution of the problems pertinent to occlusions of objects. Therefore, this study employs a foot based perspective projection depth to differentiate the occluded objects respectively. The tracking of each separated object is performed along multiple cameras. This approach suggests solutions to prevent critical errors resulted from the occlusion effect. The resultant implementation is applied to the only single camera even though it works under a networked multi-camera surveillance system because the overlapping problem arises only in a single camera. However, those effects of the separated object are very crucial across multiple cameras. The future study is to establish a networked multi-CCTV surveillance system to trace objects' relations even in the presence of occlusion effect.

References

[1] B Y Lee, L H Liew,W S Cheah, Y C Wang, "Occlusion handling in videos object tracking: A survey," IOP Confernce series: Earth and Environment Science, vol 18, 2014

[2] G. Farin, "Curves and Surfaces for CAGD," Morgan-Kaufmann, 2002

[3] D. Greenhill, J. Renno, J. Orwell, G.A. Jones, "Occlusion Analysis: Learning and Utilising Depth Maps in Object Tracking," Image Vision Comput., vol.26, no. 3, pp. 430-441, March 2008

[4] P.Guha, A.Mukerjee and V.K. Subramanian, "Formulation, Detection and Application of Occlusion States (Oc-7) in the Context of Multiple Object Tracking," in Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE international Conf. on Klagenfurt, 2011

[5] J.Krumm,S.Harris,B.Meyers,B.Brumitt,M.Hale,S.Shafer, "Multi-Camera multi-person tracking for easy living," third IEEE International Workshop on Visual Surveillance, 2000

[6] H.J.Park, "Monitoring system for an abnormal behaviors by object's tracking," Journal of Digital Contents Society, vol.14, no.4, pp.589- 596, 2013.12

[7] A. Saha, J.Mulherjee, and S.Sural, "New pixel-decimation patterns for block matching in motion estimation," Signal Processing Image Communication, vol.23, no.10,pp.725-738, 2008

[8] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," IEEE Trans. Patt. Analy. Mach. Intel, Vol.25, pp.564-575, 2003.

[9] M. Black and A. Jepson, "Eigentracking: Robust matching and tracking of articulated objects using a view-based representation," Int. J. Comput. Vision Vol. 26 No.1, pp.63-84, 1998.

[10] Huttenlocher. D, Noh. J, and Rucklidge. W.,"Tracking nonrigid objects in complex scenes," IEEE International Conference on Computer Vision (ICCV), pp. 93 - 101, 1993.

[11] Kang. J, Cohen. I,and Medioni. G, "Object reacquisition using geometric invariant appearance model," In International Conference on Pattern Recongnition (ICPR), pp.759 - 762, 2004.

[12] H. Park, "Occlusion Handling Based on Projective Depth in Object Tracking", 5th International Conference on IT Convergence and Security (ICITCS) 2015, pp. 1-2, doi:10.1109/ICITCS.2015.7293000

박 화 진

1989년 : 숙명여자대학교 대학원 (전산학석사)
1997년 : 미 아리조나주립대(공학박사)
1998년 : 삼성 SDS 선임연구원
2000~현재 : 숙명여자대학교 멀티미디어과학과 교수
관심분야 : 컴퓨터 그래픽, 가상현실, 영상인식