VIDEO SURVEILLANCE USING DYNAMIC CONFIGURATION OF MUTIPLE ACTIVE CAMERAS

Nyoun Kim, Ig-jae Kim, Hyoung-gon Kim

Imaging Media Research Center, Korea Institute of Science and Technology
{nyoun, kij, hgk}@imrc.kist.re.kr

ABSTRACT

In this paper, we present a coordinated video surveillance system that can minimize the spatial limitation and can precisely extract the 3D position of objects. To do this, our system used an agent based system and also tracked the normalized object using active wide-baseline stereo method.

The system is composed of two parts: multiple camera agents (CAs) and a support module (SM). Each CA treats image processing and camera controlling. A SM performs a role that manages communication between CAs. Our proposed system extracts object positions independent of environment via the collaboration of CAs and a SM. Finally, through experimental results we show that the proposed system successfully tracks an object on real-time.

Index Terms— Surveillance, Stereo vision, Active vision, Cooperative systems,

1. INTRODUCTION

Video surveillance has been an active area of research. Many previous researchers have been concentrating on detection and tracking based on a security issue. The trend of researches has focused on other issues; autonomous system configuration, object identification and multimodal systems, which are corresponded with ubiquitous systems. According to this trend, there have been researches about the surveillance system using multiple active cameras (i.e. Pan-Tilt-Zoom camera). Consequently a coordinated surveillance system that includes all of these issues is necessary.

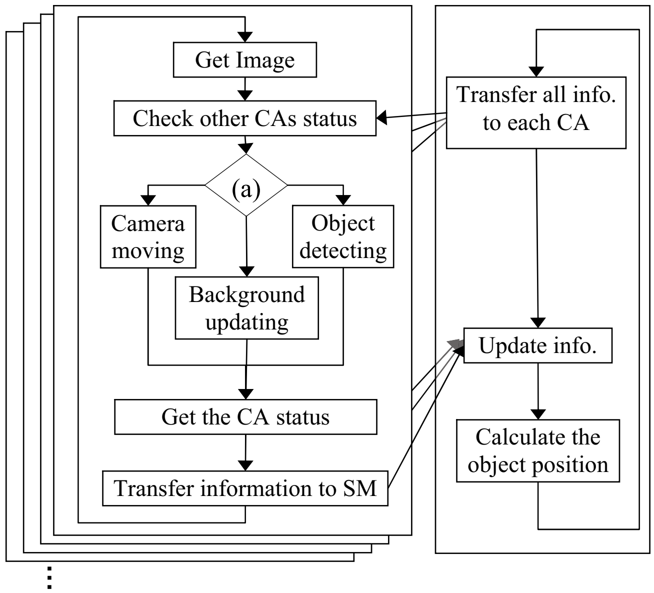
Matsuyama et al. [1] presented the architecture for a cooperative tracking system using multiple active cameras. They built the system that can track multi-targets by cooperation between agents, which include active cameras in the system. However, they did not try to overcome spatial limitation but only dealt with open spaces. Kettnaker et al. [2] devised the solution that is possible to track and detect the appearance/disappearance of objects through a multiple non-overlapping cameras setup in more complex environment (such as the one imposed by the corridors). Although they can track the prospective shape of object in a complex environment, they did not concentrate on positioning objects. Hampapur et al. [3] built the indoor surveillance system that processes object positioning and identification by combining active cameras with static cameras. In their work, however, they did not use an agent-based system that cameras communicate with each other, but a rule-based system that a camera management policy has to control multiple cameras. Furthermore, a number of cameras needed to be increased relative to the complexity of the environment since static cameras are used for objects detection/tracking while active cameras are only for identification.

In this paper, we propose a video surveillance system that extracts an object positioning using agent-based collaboration between multiple active cameras in a complex environment. We focus on the coordination of several issues which previous researchers have managed. Through our framework, moreover, an adjustable surveillance system can be established along with both the number of cameras and the complexity of the environment.

The rest of the paper is organized as follows. We will begin by presenting an overview of our system in Section 2. Section 3 will explain concretely how an agent-based camera extracts information. After describing the tasks of collaboration among the cameras in Section 4, experimental results from the real-integrated system follow in Section 5.

2. SYSTEM OVERVIEW

Each camera composes an independently working module—Camera Agent (CA). The CA is a chain of functions that processes image and converts it into worthy information for tracking using a corresponding active camera. After each CA finishes object detection and segmentation [4], it transmits current information to the other—Support Module (SM). SM is a medium that extracts 3D position based on the information from CAs using vision technique, and transfers the information to other CAs. This framework, therefore, is established by both several CAs and one SM, which communicate with one another. All CAs carry out three tasks such as; detecting and tracking an object in an arbitrary view, transmitting the processed information to the SM and controlling its camera Pan-Tilt-Zoom (PTZ) property. The CA



(b) Camera Agents (c) A Support Module

Fig. 1. Overview of our video surveillance system: (a) A Decision by both appearance and position of the object using a map of environment (b) the process of CAs (c) the process of a SM

has a map of environment. Whether the CA can detect the object or not is decided by the position of the object on the map. Likewise the position also decides whether the CA has to change to an optimized view.

There is a chain of communication every time an image is loaded, as shown in Fig. 1. An image is loaded to a CA, which receives both the status of other CAs and the position of object on the map from SM. The status means whether a background has been updated or not, whether a camera is changing its PTZ property or not and whether an object is detected or not. The CA decides what it has to process at the divergent point by the status information (see (a) in Fig. 1). If there is an optimally visible object in the camera view, the CA performs *object detection*. Otherwise, when the object is detected in a visible non-optimal view, the CA performs *camera moving*. In that case, the CA has to check whether other CA in relation of common detection is not moving. If an object is not detected in the map, the CA turns into *background updating* mode. After processing one of the three modes, the CA transmits its status into SM. As a result, other CAs decide what to do, with the status of this CA. Additionally, the CA transfers a head point of object to SM if the CA is in *object detection* mode. SM gathers head points from more than two CAs, and therefore it retrieves the 3D position of the object. Finally, SM updates the status of CAs and the system repeats this chain of functions in every image frame.

3. FUNCTIONALITY OF A CAMERA AGENT

In order to get image information, a CA converts a loaded image into an edge map by canny edge detection logic. It creates edge maps of both background image and current image. And through those maps, video-object-segmentation (VOS) is found [4]. A top-center point of VOS can be retrieved by histogram projection. The point is the basic information for the extraction of 3D-position of an object. And histograms that represent distance, angle and color of VOS are created using histogram-based-approach (HBA) [5]. They become identities of the object. Furthermore, they provide information that need for collaboration among CAs. The process of VOS is operated on real-time for every frame that the CA processes object detection. On the other hand, the process of HBA matching is operated once when other CA that hasn't seen the object starts to detect object. Thus, the system prevents an overload of computation.

For extracting 3D position, image information is combined with camera information. Now we describe the procedure of acquiring camera information. One CA solves camera calibration of each preset view in order to get the 3D position of an object in any observing view. This calibration task requires repetitive elaborate works. To minimize the task, we find a linear relationship between camera control parameters and calibrated intrinsic/extrinsic parameters of the camera. The orthogonal matrix triangulation (i.e. QR decomposition) separates intrinsic/extrinsic parameters of the projection matrix (p-matrix) previously calculated [6]. Our calibration method of an active camera using the orthogonal matrix triangulation is as follow.

Algorithm: active camera calibration

1. Get P_1 (p-matrix) from one PTZ view.
2. Move Pan (x degree) and get P_2 (p-matrix)
3. Compare P_1 with P_2.
4. In a state of fixed K (intrinsic parameter), solve relationship between R/T (rotation/translation) matrix and Pan/Tilt control of the camera.
5. Move Zoom and get P_3 (p-matrix)
6. Compare P_1 with P_3
7. In a state of Fixed R/T (extrinsic parameters), solve relationship between K matrix and Zoom control of the camera.

Finally, we can find a relationship between an active camera control and its P-matrix.

Camera calibrations of various preset views can be solved by interpolation of camera parameters.

4. COLLABORATION AMONG CAMERA AGENTS

In an arbitrary environment, CAs configure their setup dynamically by the appearance and the position of objects. The tasks that CAs accomplish by collaboration are described below.

- In initial state, the system decides a default view of each CA.

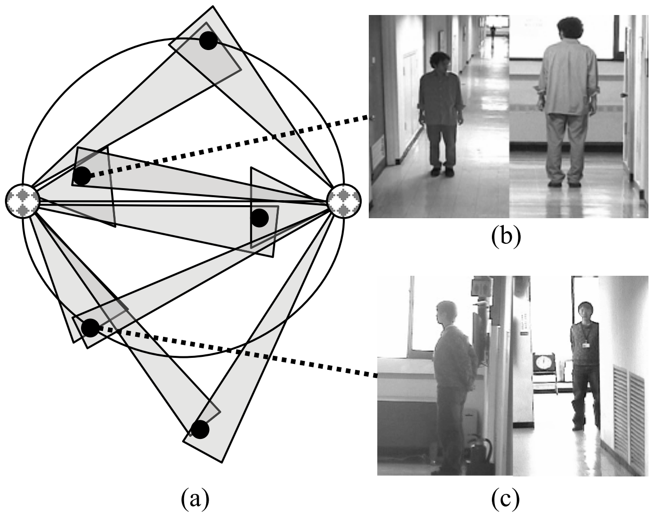


Fig. 2. (a) A Diagram of active wide-baseline stereo according to an object position: Black circles represent objects, circles with gray spots represent active cameras and rotated triangles represent active views of the camera. A short triangle means the camera is on zoom-out state and a long triangle means the camera is on zoom-in state. (b) An opposite view pair (c) A perpendicular view pair

- When one of CAs is out of work, the others are fit for an environment of surveillance except the excluded CA.
- This system extracts the 3D position of objects using dynamic stereo method.
- When the object is out of view in one CA, other CAs can follow the object by an object histogram matching.

In this section, several requirements for which collaboration is needed in this system are shown.

4.1. Initialize setup

The rule that decides the initial view of each active camera has to include a system environment. CAs can determine their initialization because they have map information of the environment. Let N_E be a number of entrances of an object in the map. Let N_C be a number of active cameras in this system. The condition to evaluate completeness of object detection can be expressed as:

$$N_E \leq N_C \tag{1}$$

If the condition is satisfied in this environment, the surveillance system can completely observe object appearance. In order to be a flexible system, however, even though the condition is not satisfied, the system tracks objects after initial detecting of one camera. Our system can maintain its stability in this way. Each CA corresponds with each area of object entrances in turn. No area is observed by more than two CAs until there remains no area of object entrances.

4.2. Background update

The fundamental task, called background update, saves background information in each observing view of the active camera. Because the saved background information varies depending upon the difference of lighting condition, each CA has to update its background views. As shown in Fig 1, a CA does not execute detecting/tracking object and updating background concurrently. At the time when all CAs don't detect objects, all CAs update their preset views in turn. By pre-defined order, each CA updates its preset views. While one CA updates an initial view that was decided in Section 4.1., other CAs observe own fixed view. When the CA, however, updates non-initial views, the system may not cope with the situation that a new object appears, because it is not in the optimized state. CAs repeat initialization except for one updating agent. If an object appears while the CA updates its background information, the CA stops updating and the surveillance system switches to object detection mode.

4.3. Wide baseline stereo

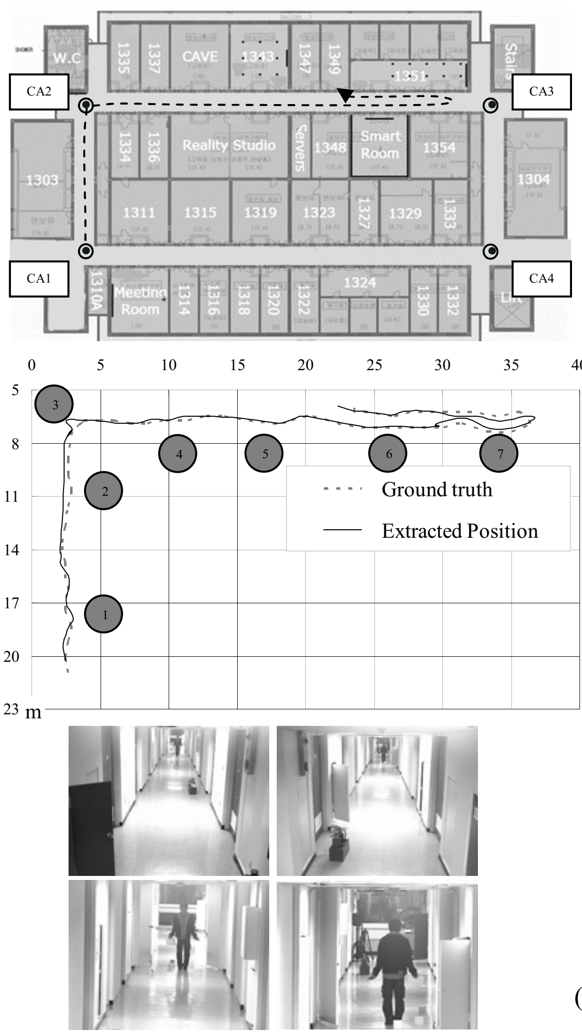
In the environment that includes an arbitrary corridor, for example, one CA can follow objects by changing its camera preset. However, the farther the object is from the camera the more difficult it is to extract the object position. In order to solve this problem, we propose an active wide-baseline stereo (WBS) that is applied for collaboration in a wide baseline view. Through active WBS, not only extracting position is more relevant but also observation is more robust because the size of object is normalized in any position of corridor. The maintenance of size in a wide baseline view is presented in Fig. 2. Relative to the object position between two CAs, they optimally alter their PTZ properties and apply the active WBS for extracting 3D position.

While a CA changes its PTZ property, it cannot detect an object. If two CAs in relation of WBS change their PTZ properties concurrently, the system cannot extract the position of objects in real-time. CAs, therefore, check their moving order and change their PTZ properties one by one. Each CA has a controlling function of camera moving and it knows the distance between the collaborating CAs and the object using the environmental map. After one CA with a short distance has changed its PTZ property, the others change. Therefore, the system avoids a simultaneous camera moving and it retrieves a continuous position of the object.

The SM receives top-center points of an object from two CAs in a relation of collaboration for extracting the object position, and it extracts a world coordinate of the current object using WBS [7].

5. EXPERIMENTAL RESULTS

In order to evaluate this proposed system, we set up a complex indoor environment around a research lab, as shown in Fig. 3. There are four corridors in the environment and networked active cameras are set at every corner that two corridors come across. Each camera is connected to a PC, which manages the camera and performs the role of CA.



(a)

(b)

(c)

Fig. 3. (a) A map of experimental environment (b) Comparison between an extracted position and a ground truth of the detected object; the circle numbers represent the index of the optimal setup for active WBS (c) Static WBS (top), Active WBS (bottom)

The image resolution of a camera is 640 x 480 with 10 fps in a detection mode and 30 fps in an update mode. (b) in Fig. 3 shows the result of tracking object after appearance. According to the position where the object is in a corridor, the observing CAs decide their optimal PTZ properties (Table 1). As a result, we can get the path corresponding with a ground truth. Since the average of error is 0.6m in the case of active WBS, the system is able to track the disappearance of an object such as entering a room (The distance between rooms is about 2m).

For comparison, we tested extracting the position using not only active WBS but also static WBS. The error of active WBS is smaller than the error of static WBS by far (Table 2). The reason is that to extract a head point from the image is difficult and to apply the static calibrated p-matrix is inaccurate.

Table 1. The zoom properties of cameras change by object position; both the number and the direction of each inequality-sign mean a degree of zoom in and a camera view. (see (b) in Fig. 3)

position	1	2	3	4	5	6	7
CA1	∨	∨∨	∨∨	∨∨	∨∨	∨∨	∨∨
CA2	∧∧	∧	∧	<	<<	<<<	<<<<
CA3	>>>>	>>>>	>>>>	>>>>	>>>	>>	>
CA4	>>>>	>>>>	>>>>	∨∨	∨∨	∨∨	∨∨

Table 2. Average errors of the distance between an extracting position and a ground truth according to the stereo type.

Stereo type	Short corridor (25m distance)		Long corridor (50m distance)	
Active WBS	Average error	0.697m	Average error	0.606m
	StdDev. of error	0.235m	StdDev. of error	0.433m
Static WBS	Average error	4.687m	Average error	619.997m
	StdDev. of error	1.734m	StdDev. of error	2514.224m

6. CONCLUSION

We have presented the system that detects and tracks an object by dynamic configuration between networked active cameras. The system makes use of active WBS for minimizing the spatial limitation and for tracking the normalized object. The system consists of independently working agents which collaborate with one another. The experimental result demonstrates that a cooperative active WBS is more robust than a static WBS in tracking objects. Although our system does not track multi-objects concurrently, we will complement the problem by identification of multi-objects. This system will also provide a framework of multimodal surveillance using precise position and of another surveillance issues such as identification and activity analysis.

7. REFERENCES

[1] N. Ukita and T. Matsuyama, "Real-time Multi-target Tracking by Cooperative Distributed Active Vision Agents," *in Proc. of AAMAS 2002,* pp. 829-838, 2002.
[2] V. Kettnaker and R. Zabih, "Bayesian multi-camera surveillance," *In Proc. of CVPR,* 1999.
[3] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, S. Pankanti, A. Senior, C. Shu, and Y. L. Tian, "Smart Video Surveillance", *IEEE Signal Processing Magazine,* Mar. 2005.
[4] C. Kim and J.N. Hwang, "Object-based video abstraction for video surveillance systems," *IEEE Trans. Circuits Systems Video Technology, 12 (12)*, pp. 1128-1138, Dec. 2002.
[5] E. Saykol and U. Gudukbay, "A histogram-based approach for object-based query-by-shape-and-color in image and video databases," *Image and Video Compution(23), No. 13*, pp. 1170-1180, 29 Nov. 2005.
[6] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing,* Cambridge University Press, 1988.
[7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.