Incremental Observable-Area Modeling for Cooperative Tracking

Norimichi Ukita Takashi Matsuyama
Department of Intelligence Science and Technology
Graduate School of Informatics, Kyoto University
Yoshidahonmachi, Sakyo, Kyoto, 606-8501 JAPAN
souhaku@vision.kuee.kyoto-u.ac.jp tm@i.kyoto-u.ac.jp

Abstract

In this paper, we propose an observable-area model of the scene for real-time cooperative object tracking by multiple cameras. The knowledge of partners' abilities is necessary for cooperative action whatever task is defined. In particular, for the tracking a moving object in the scene, every Active Vision Agent (AVA), a rational model of the network-connected computer with an active camera, should therefore know the area in the scene that is observable by each AVA. Each AVA should then decide its target object and gazing direction taking into account other AVAs' actions. To realize such a cooperative gazing, the system gathers all the observable-area information to incrementally generate the observable-area model at each frame during the tracking. Hence, the system cooperatively tracks the object by utilizing both the observable-area model and the object's motion estimated at each frame. Experimental results demonstrate the effectiveness of the cooperation among the AVAs with the help of the proposed observable-area model.

1. Introduction

This paper presents a real-time cooperative distributed vision (CDV, in short) system with multiple communicating *Active Vision Agents* (AVAs, in short)[1]. AVA is a rational model of the network-connected computer with an active camera. CDV system has many advantages (e.g. wide area observation, robustness by integrating multilateral information, flexibility of the system-organization, and compensation for troubles). In recent years, a number of related researches are, therefore, reported (see [2], for example). Various vision systems can be realized by employing CDV system. Above all, a object tracking system is one of the important basic technics for realizing applied systems (for example, autonomous surveillance and monitoring systems, ITS(Intelligent Transport System)).

In this paper, we put our focus upon the sharing knowledge of all the AVAs' abilities (i.e. observable area in the scene) for the efficient object tracking and scene observation. Our system incrementally acquire the observable-area information of each AVA, and enables AVAs to dynamically and appropriately changing their roles by taking into account all the AVAs' observable areas.

Experimental results demonstrate the effectiveness of the cooperation among AVAs with the proposed model.

2. Cooperative Tracking

2.1. Design of AVAs' behaviors while tracking

In our previous system(Sec. 5 in [1]), a group (we call an *agency*) of communicating AVAs cooperatively tracks a single object without being interfered by obstacles, where 1. each AVA possesses a Fixed-Viewpoint Pan-Tilt-Zoom (FV-PTZ, in short) camera, and 2. the external camera parameters (i.e. the 3D position of each camera) are calibrated.

The FV-PTZ camera allows us to generate background images taken with arbitrary combinations of the pan-tilt-zoom parameters from several images taken beforehand (Sec. 3 in [1]). An AVA can, therefore, detect an anomalous region by the background subtraction during widely observing the scene and adjusting the zoom. Thus the tracking by a single AVA is achievable by changing the gazing direction to the detected region in the image (Sec. 4 in [1])[1] .

With these resources, we designed the system as follows:

1. All AVAs search around for an object autonomously (Fig.1, 1.).
2. An AVA (denoted by AVA_m) detects an object, it regards this object as the target object and broadcasts the 3D view line (denoted by L_m) from the camera to the object. An agency is then formed(Fig.1, 2.).
3. Each AVA searches for the object along L_m after it receives the broadcasted message. If an AVA (denoted by AVA_w) detects an anomalous region, it replies to AVA_m the 3D view line (denoted by L_w) from the camera to the object (Fig.1, 3.).

[1] In order to apply this tracking method to the real-world system, the background subtraction method, that is robust for the illumination changes, the flicker of leaves and so on, is required. A lot of researches about this problem are reported (see [3], for example).

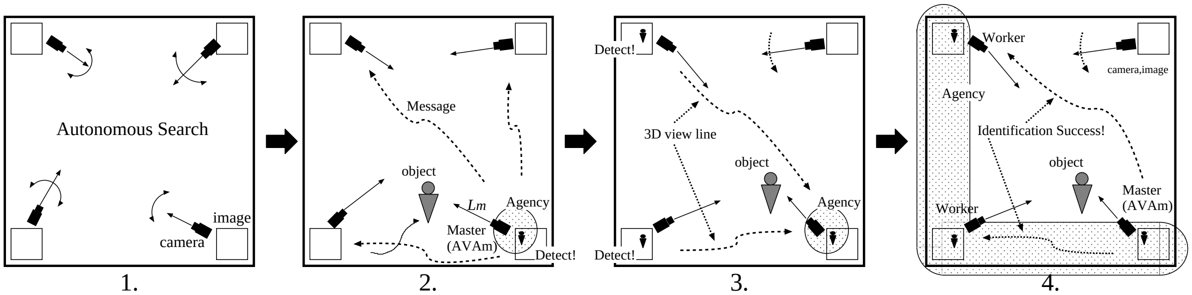

Figure 1. Object tracking system by multiple AVAs

4. AVA_m computes the distance between L_m and L_w. If the distance is less than the threshold, 1. the objects detected by AVA_m and AVA_w are considered as the same object, 2. the middle point between L_m and L_w is considered as the 3D position of the object. AVA_w is then joined into the agency, and engage in tracking the identified object.
5. AVA_m broadcasts the 3D position so that all AVAs gaze the identified object (Fig.1, 4.). The image-capturing timings are almost synchronized among all AVAs in one agency because the image-capturings are activated by the message from AVA_m.
6. Repetition of the object identification and the gaze navigation allows all AVAs to track the target object.
7. When all AVAs fail in tracking the object, they start searching for an object again.

AVA_m is called the master AVA and the other AVAs in the same agency are called the worker AVAs. The master authority is dynamically transfered to the worker AVA whose detected region of the target object is most reliable[2] .

In accordance with the AVA's behavior mentioned above, all AVAs cooperatively keep gazing the object even if the 3D geometric configurations of the scene is not known *a priori*.

2.2. Problem of Gaze Navigation

In our previous system, all AVAs can keep tracking an object of interest through the compulsory gaze navigation by the master AVA. All the AVA, therefore, keep obeying the gaze navigation even if the AVA cannot observe the target object due to obstacles. This problem is caused by lack of the information on the 3D geometric configurations of the scene. The AVA cannot, therefore, know whether the target object is interfered by the obstacle or the object detection is a failure when no anomalous region is detected in the observed image. The AVA that cannot observe the target object due to obstacles, however, should change its role for increasing the efficiency of the whole system. For example, the following functions can be considered for such an AVA.

[2] The reliability is determined by 1. the size of the detected region in the image, 2. the distance from the image center to the detected region in the image, and 3. the distance from the object to the camera in the scene.

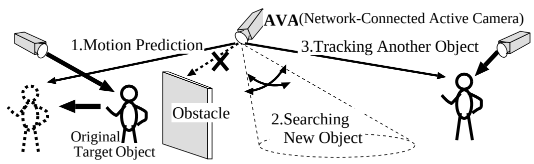

Figure 2. Dynamic role assignment to AVA

- The AVA predicts the position where the target object will appear within its observable area, then changes its gazing direction to ambush the object(Fig.2, 1.).
- The AVA gazes the area where none of the other AVAs observes to find another object(Fig.2, 2.).
- If the AVA can observe the target object of another agency[3] , it joins this agency(Fig.2, 3.).

To solve the gaze navigation problem, we expand our previous system to identify each AVA's visible/invisible area in the scene and employs this information for the gaze navigation. In the proposed system, all the visible/invisible information is gathered as the *observable-area model* of the scene, and the master AVA assigns the appropriate role to each AVA by referring the observable-area model.

3. Observable-Area Model

3.1. Data Structure of Observable-Area Model

We adopt the octree representation[4] for the data structure of the observable-area model. We have the following advantages in employing the octree representation:

- The octree representation allows us to reduce the amount of data, since the visible/invisible area usually masses in the scene.
- Easiness of resizing cubes in the octree allows us to localize the resolution of the observable-area model.

In each cube in the octree, three kinds of the visible/invisible labels are attached to each AVA.

UNDEFINE The system has not identified whether or not the AVA can observe the area.

VISIBLE The AVA can observe the area.

INVISIBLE The AVA cannot observe the area.

[3] In the multiple object tracking system, one agency is formed for each tracking target.

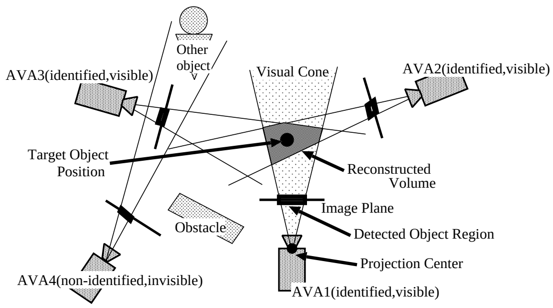

Figure 3. Reconstruction of the volume and generating the visible/invisible information

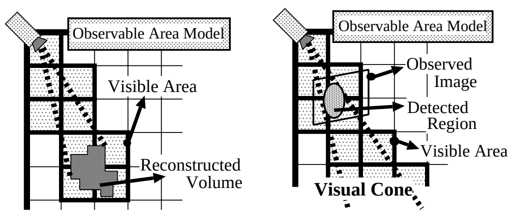

Figure 4. Visible area propagation (Left: CaseA, Right: Case B)

3.2. Generating Visible/Invisible Information

In our previous system[1], the object identification and 3D position reconstruction are realized by incorporating 3D view lines. In the proposed system, however, the master AVA computes the intersection of the visual cones, each of which is determined by the projection center of a camera and the detected region in the observed image. If the intersection exists, all the detected regions are considered as the same object. Moreover, the 3D position of the object can be obtained for the gaze navigation, since the computed intersection corresponds to the volume of the object.

If the image observed by AVA_w is actually used for the volume reconstruction, the system can then identify the area where the detected object exists to be visible from AVA_w. Otherwise, the area is identified to be invisible from AVA_w (Fig.3). The master AVA can, therefore, generate the visible/invisible information while tracking.

3.3. Generating Observable-Area Model

After a new visible/invisible information is obtained as mentioned in Sec.3.2, all the visible/invisible labels in the new information are respectively compared with those in the observable-area model to update the observable-area model. If the label in the new information is different from that in the cube whose position corresponds to the new information in the observable-area model, this cube is decomposed into octants. This decomposition is executed as long as the following two conditions are both satisfied.

1. The label in the new information and that in the observable-area model are different.
2. $\frac{distance}{focallength} < \frac{\mathrm{constant}}{2^{depth}}$ ($distance$ is the length from the camera to the area, $focallength$ is the focal length of the camera and $depth$ is the depth of the octree.)

Since the resolution of the reconstructed volume of the object depends on $focallength$ and $distance$, the number of decomposition is defined by the above inequality.

After the decomposition, each cube is given the visible/invisible label. To the three kinds of labels, the following order of priority is applied in case of substituting the label for the cube.

$$\mathrm{INVISIBLE} > \mathrm{VISIBLE} > \mathrm{UNDEFINE}$$

For example, VISIBLE and UNDEFINE are not recorded in the cube that has been identified as INVISIBLE. If all the labels of eight suboctants are the same, the suboctants are unified.

Moreover the visible information is propagated to facilitate generating the observable-area model. Two cases exist for the propagation of the visible area (Fig.4).

Case A The volume has been reconstructed.

Case B The volume has not been reconstructed.

In the case A, each cube corresponding to the reconstructed volume is identified to be visible. We can then identify the area between the object and the camera to be also visible from this camera if it can observe the object. The cubes between the object and the camera are ,therefore, updated as the visible area.

In the case B, the whole area included in the visual cone is updated as the temporary visible area except the area that has been already identified to be visible or invisible. Due to this function, there is a possibility of falsely attaching the temporary visible label to the area that is not yet estimated but actually invisible. To correct the observable-area model by the subsequent observation, the temporary visible label is updated when the area is identified as visible or invisible.

3.4. Control of Observable-Area Model

Since the observable-area model is made from the visible/invisible information which is generated by the master AVA, the master AVA can obtain the model by itself. Accordingly, the master AVA can refer the model immediately without increasing the load of the network. The master AVA, however, dynamically exchanges inside the agency for stabilizing the tracking. Thus, in proposed system, the master AVA delegates the following tasks to a model-control module, *observable-area model controller*.

- Keeping and updating of the observable-area model.
- Planning the appropriate roles of AVAs by referring the observable-area model.

This controller receives all the information from the master AVA at each frame to have the observable-area model.

The observable-area model controller assists the master AVA in improving the performance of the system.

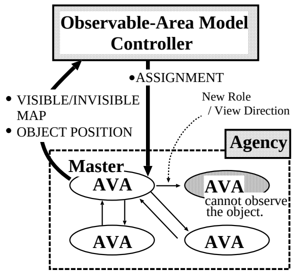

Figure 5. Message Flow

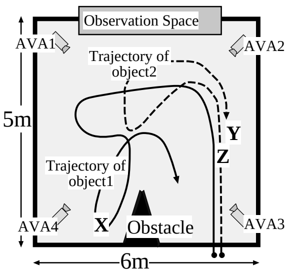

Figure 6. Experimental Environment

4. Cooperation with Observable-Area Model

This section addresses the communication protocol for 1. the updating of the observable-area model controller and 2. the role assignment to the AVA (Fig.5).

I. From AVA to Observable-Area Model Controller

At each frame, the master AVA transmits the following two messages to the observable-area model controller.

- **VISIBLE/INVISIBLE MAP :** The transmitted visible/invisible information for each AVA allows the observable-area model controller to update the observable-area model.
- **OBJECT POSITION :** The transmitted 3D position of the object allows the observable-area model controller to keep the object's motion trajectory.

II. From Observable-Area Model Controller to AVA

The observable-area model controller decides the role assignment to each AVA every after the updating of the model. If the model decides a new role assignment, the following message is transmitted to the master AVA.

- **ASSIGNMENT :** 1. The ID of the AVA that is assigned a new role, and 2. the details of the role, are included in the message.

The message is broadcasted because the master AVA dynamically exchanges, and accepted by only the master AVA. If master AVA receives ASSIGNMENT message and approve the effectiveness of the role assignment, the master AVA assigns the role to the AVA.

Since only the master AVA is allowed to order the worker AVAs, the system can avoid sending the different roles to one worker AVA at the same time.

5. Experimental Results

We experimented to verify the effectiveness of the proposed model for cooperative tracking. Our experimental results demonstrated the improvement in cooperation of the AVAs while tracking.

We conducted our experiments in the environment shown in Fig.6. Each AVA consists of a PC(PentiumIII 600MHz) and a FV-PTZ camera(SONY EVI-G20). In addition, all the PCs are connected by the Ethernet. Each camera is placed about 2.5m above the floor.

In this environment, object1 came into the observation space, and stayed for a while at the location X after moving along the trajectory. Note that AVA3 could not observe object1 when it was at the location X. Next, object2 moved along the trajectory and stopped at the location Y. After that, object1 started moving again.

Fig.7 shows an example of image sequences observed by AVA2 and AVA3. The size of each image is 320×240 $[pixel]$. The images from A1 to A14 and the images from B1 to B14 are respectively captured by AVA2 and AVA3 in the tracking without the observable-area model. The images from a1 to a14 and the images from b1 to b14, on the other hand, are respectively captured by AVA2 and AVA3 with the observable-area model. The enclosed regions with the white and black lines in the images indicate respectively the detected regions of object1 and object2. Each AVA captures the images at about 0.5 sec intervals on average.

Without the observable-area model, after searching (A1, A2 and B1, B2), AVAs detected object1 and regarded it as the target object, and then began to cooperatively track object1 (A3, A4 and B3, B4). However, AVA3 kept gazing at the direction of the 3D position of object1 transmitted from the master AVA, though AVA3 could not observe it due to the obstacle (from B5 to B13). The other AVAs, on the other hand, kept tracking object1 (from A5 to A14).

With the observable-area model, each AVA behaved in the same way as the cooperative tracking without the observable-area model (from a1 to a4 and from b1 to b4). AVA3 started, however, searching for another object (b7, b8) after AVA3 could not observe object1 (b5, b6). AVA3 then detected object2 at the location Z and started tracking it independently (b9, b10, b11, b12). Note that AVA2 modified the zoom parameter to acquire high resolution object images because each AVA could know that object1 stayed at the same place for a while through the message from the observable-area model controller (a9, a10). Note also that the gaze of AVA2 was directed toward object1 that was regarded as the target object even if AVA2 could also observe object2 (a11, a12). After that, when object1 started walking and became close to the area where AVA3 could observe, the observable-area model controller instructed the master AVA to change its gazing direction to object1 (b13). As a result, AVA3 changed its role, from independent tracking to cooperative tracking, and started again tracking object1 (b14). The visible/invisible area information for AVA3 is shown in Fig.8 where the two different views are illustrated. P and Q respectively indicate the visible and invisible area. Note that we see the obstacle area at R[4] .

[4] We can also estimate the 3D geometric information of the scene by integrating the observable-area information of each AVA.

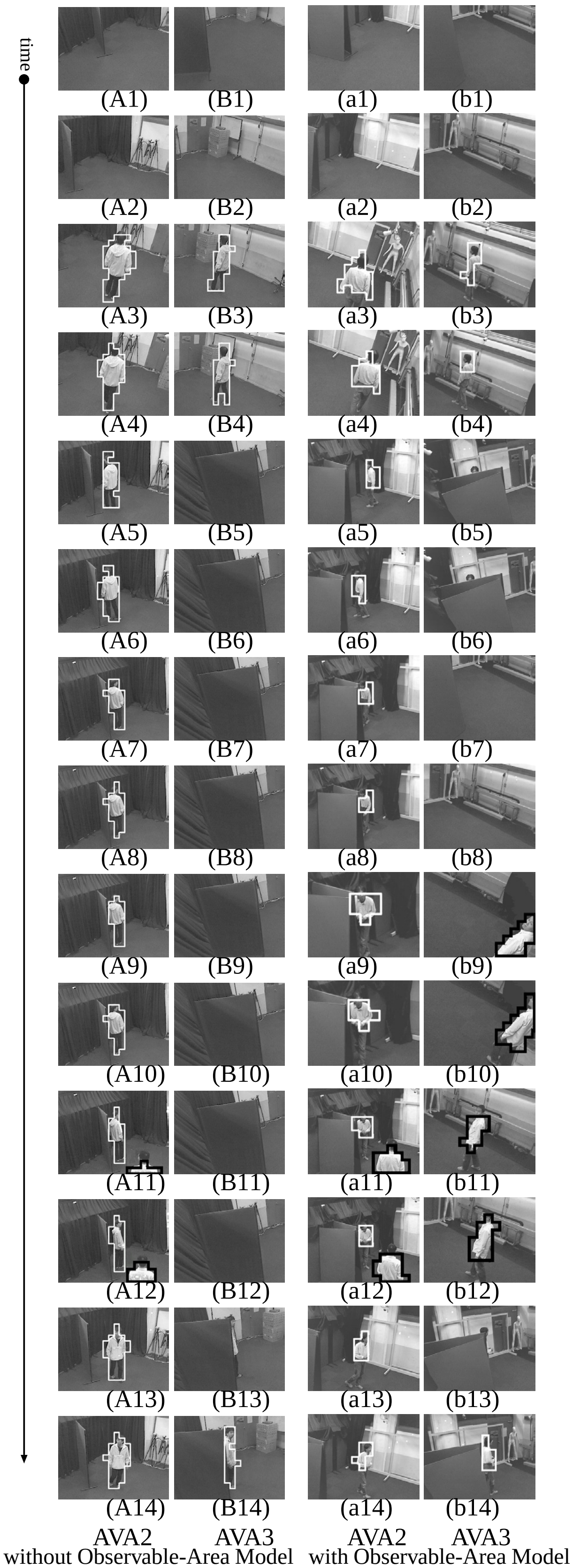

Figure 7. Partial Image sequences

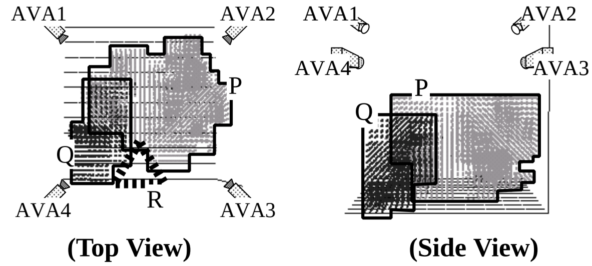

Figure 8. Acquired observable-area model

Comparing the experimental results of the two cases, we may conclude that the proposed model improves the effectiveness of cooperation in tracking.

6. Concluding Remarks

We proposed the incremental observable-area model for cooperative tracking. Our model allows a tracking system to assign the appropriate role to each AVA. We should note that the proposed model is evaluated by multi-agent systems with visual perception, however, the basic idea, i.e. the knowledge of partners' abilities are necessary for cooperation, is applicable to various types of multi-agent systems.

The proposed system is the expansion of our previous system[1] that tracks a single object. Thus the target object that is cooperatively tracked by multi-AVAs is a single, even if the AVA that is assigned a new role tracks another object independently. The flexible real-world system requires, however, the ability to track multi-target[5]. We are, therefore, now developing our system into multi-target tracking system by multi-agencies.

This work was supported by the Research for the Future Program of the Japan Society for the Promotion of Science (JSPS-RFTF96P00501).

References

[1] T. Matsuyama, “Cooperative Distributed Vision - Dynamic Integration of Visual Perception, Action and Communication -”, Proc. of Image Understanding Workshop, pp. 365-384, 1998.

[2] T. Kanade, *et al.*, “Cooperative Multisensor Video Surveillance”, Proc. of Image Understanding Workshop, pp. 3-10, 1997.

[3] C. Stauffer, E. Grimson, “Adaptive background mixture models for real-time tracking”, Proc. of CVPR99, Vol.II, pp.246-252, 1999.

[4] N. Ahuja, J. Veenstra, “Generating Octrees from Object Silhouettes in Orthographic Views”, IEEE Trans. on PAMI, Vol2, No.2, pp. 137-149, 1989.

[5] A. Nakazawa, H. Kato, S. Inokuchi., “Human Tracking Using Distributed Vision Systems”, Proc. of 14th ICPR, pp.593-596, 1998.