Cooperative Distributed Tracking for Effective Face Registration

Takekazu KATO* Yasuhiro MUKAIGAWA* Takeshi SHAKUNAGA*

Department of Information Technology
Okayama University

Abstract

A three-layer framework is proposed for cooperative distributed face tracking. In this framework, multiple distributed cameras cooperate both for stable face tracking and for effective registrations. The framework consists of an independent task layer, a local cooperation layer and a global cooperation layer. In the independent task layer, each camera asynchronously works as an agent to track and to register a face. In the local cooperation layer, a set of agents forms an agency and it cooperatively tracks a particular person with estimating the 3D face location. A concept of asynchronous cooperative tracking is introduced for all the agents to independently track the target person and to utilize the integrated 3D information for the effective tracking. In the global cooperation layer, the assignments of agents to persons are dynamically changed in order to effectively track and register multiple persons. We call this mechanism a dynamic agency configuration. Each agent is transfered to a more contributory agency according to an evaluation of the registered face images. Experimental results show that each layer well worked in a real environment and that face images of multiple persons were effectively tracked and registered.

1 Introduction

A person observation is an important technology not only for security uses but also for comfortable interfaces between humans and environments. In order to observe a person in a wide area such as the whole of a building, it is important to robustly identify a person in various environments while the person is moving. A face image is a good person identifier because a face is easily observed by a camera as well as inherent to each person.

A lot of methods have already been proposed for the face recognition[3, 6, 8]. In these methods, persons can be well recognized if their face images have already been registered in the dictionary, and if the images are taken in enough various lighting conditions. It is, however, not easy to register enough images for each person if target persons cannot be specified in advance. The effective registration is very important to realize a robust recognition in such a real environment.

In this paper, we propose a three-layer framework of cooperative distributed tracking for the effective registration. The tracking system is based on cooperative distributed vision[5]. Multiple distributed cameras cooperatively track and register face images of the target person. The cooperative tracking is accomplished by the three-layer framework which consists of an independent task layer, a local cooperation layer and a global cooperation layer. In the independent task layer, each camera asynchronously works as an agent to track and register a human face. In the local cooperation layer, a set of agents forms an agency and the target person is stably tracked by the asynchronous cooperative tracking. In the global cooperation layer, the assignment of agents to persons is dynamically changed by dynamic agency configuration. The hierarchical layers realize both the stable tracking and the effective registration.

The rest of this paper is organized as follows: Section 2 describes an framework of the tracking system. Section 3, 4 and 5 present detail functions and experimental results of each layer, respectively. Finally, we make some conclusions in Section 6.

2 Framework of Cooperative Distributed Tracking

2.1 Agents and Agencies

The tracking system consists of multiple agents, where an agent is a model of an active camera with a processor. An organized set of agents is called an agency, and it plays a central role in this system. Definitions of agents and agencies are as follows.

An agent is an entity which comprises an active camera and a processor. It can detect and track a face from images taken by the camera, and controls the active camera according to the processing results. Although an agent independently works, it also can communicate with other agents via a network for the cooperations.

*Address: Tsushima naka 3-1-1, Okayama 700-8530, Japan
E-Mail:{kato,mukaigaw,shaku}@chino.it.okayama-u.ac.jp

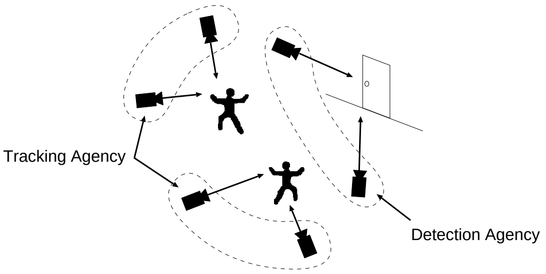

Figure 1: Detection agency and tracking agency

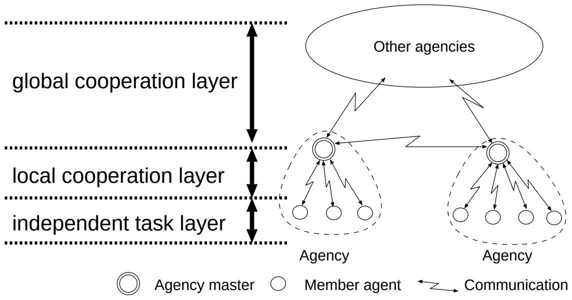

Figure 2: Three-layer framework

An agency is made up of a set of agents. Any agent always belongs to an agency as a member agent. There is one special agent called an agency master in each agency. The agency master manages all the member agents in the agency. It works as a delegate of the agency. That is, only agency masters can communicate with one another via a network.

There are two types of agencies in the system as shown in Fig.1. One is called a detection agency, which detects a new person in the territory of the agency. The other is called a tracking agency, which tracks a particular person.

2.2 Three-Layer Framework

In order to observe the whole scene and to register face images of multiple persons simultaneously, the distributed agents should cooperatively work because each agent can observe just a small area.

Frequent communication is required for realizing both the stable person tracking and the effective registration. If all the agents directly communicate with one another, it is hard to perfectly synchronize processes of all the agents because too much communication is necessary.

Perfectly synchronization is not required in our system which has a hierarchy of three-layers as shown in Fig.2. In principle, each agent asynchronously works in the independent task layer. The amount of communication can be reduced by combination of local and global cooperation layers.

Independent Task Layer:
In this layer, the agent extracts a face and tracks it for the registration. Since this task is independently done, it is not necessary to synchronize the processes of the agents in an agency. The detail functions of each agent are described in Section 3.

Local Cooperation Layer:
A task of an agency is basically accomplished in the local cooperation layer. The local cooperation mechanism depends on the agency type. In a detection agency, an agency master allocates an observation area to each member agent, and each member agent observes its territory. In a tracking agency, location information of the target person is shared among the member agents. The registered face images are taken in each agent and collected to the agency master. The details of cooperative functions within an agency are described in Section 4.

Global Cooperation Layer:
The whole system is controlled by communication among agency masters in the global cooperation layer. The system is controlled as follows:

1. An agency generation and termination. If a detection agency detects a new person, the agency generates a new tracking agency to track the person. If the target person goes out from the observation area, the tracking agency terminates.

2. The other is an agent transfer between agencies. When an agency requires more agents, some agents are transferred from another agency according to a protocol. If an agency contains an unnecessary agent, the agency releases the agent to another agency. This mechanism is called a dynamic agency configuration.

The details of cooperative functions among agencies are described in Section 5.

3 Independent Task Layer

3.1 Rapid and Accurate Face Extraction

Each agent in the tracking agency has to track a face region of the target person and to register the tracked face images. The face region should be rapidly extracted to keep the face region in the center of images. On the other hand, the face region should be accurately extracted for the stable registration. The accurate extraction is, however, time-consuming because the process needs a heavy computation. To cope with this problem, the face tracking is decomposed into two stage processes.

(1) Rapid Extraction for Tracking
In the first stage, the face region is rapidly and coarsely extracted by an incremental template matching. Once a face region is extracted, the previous face image is used as a new template. The camera is controlled due to the result.

Although the template matching can be rapidly accomplished, it cannot always provide an accurate result. When a scale of the face region drastically changes, the face region is sometimes extracted incorrectly. Furthermore, location errors of face regions are often accumulated during the tracking in a long image sequence. These errors are corrected in the next stage.

(2) Accurate Extraction for Registration

In the second stage, the face region is accurately extracted and the face direction is estimated for the stable registration. Average templates and the facial structure analysis[7] are used for the face extraction and the direction estimation. First, several candidates of the face region are detected using the average templates in the neighborhood of the result by the rapid extraction. Then the candidates are analyzed by the facial structure analysis.

The average templates are prepared by averaging face images of many persons in a variety of lighting conditions. To cope with the direction change, the templates are created for each direction (frontal, left and right). To cover the scale change, a set of templates is created in several sizes. Some candidates of the face regions are picked up by finding the region whose correlation to the template is larger than a threshold.

In the facial structure analysis, facial features are extracted and it is judged whether they consist in a face region. If a consistent set of facial features is found in a region, the region is selected as a candidate region. A face location, direction and scale are also provided as well as locations of facial features.

After the facial structure analysis, the template for the rapid extraction is updated to the current appearance. The successive update of the template is effective for suppressing accumulative errors caused by the incremental template matching.

3.2 Experimental results

Experimental results of the face tracking are provided in this section. In the experiment, a combination of a Pan-Tilt-Zoom controllable camera(SONY EVI-G20) and a PC(Dual PentiumIII 600MHz) is used as an agent. The rapid extraction was accomplished within a video rate (33ms). Results of the rapid and accurate extraction are shown in Fig.3. A rectangle with broken lines indicates a face region by the rapid extraction, while a rectangle with solid lines indicates a face region corrected by the accurate extraction. Four '+' marks indicate facial features extracted by the facial structure analysis. We can see that the locational errors were appropriately corrected by the accurate extraction.

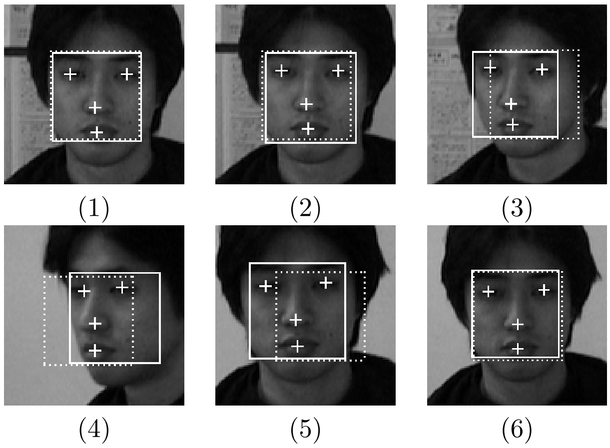

Figure 3: Results of the rapid and accurate extraction

4 Local Cooperation Layer

4.1 Asynchronous Cooperative Tracking

In principle, each agent autonomously tracks a face. The tracking result is sent from each agent to the agency master, and collected at the agency master to estimate the 3D location of the target person. The 3D information is broadcasted to all the member agents, and utilized by member agents for the effective tracking.

In ordinary stereoscopic algorithms, a synchronization of agent processes is necessary to estimate an accurate 3D face location. It is, however, difficult to perfectly synchronize the image capturing and image processing in a real environment. To cope with this problem, we propose an asynchronous cooperative tracking to stably track a particular person even if agents are not strictly synchronized. The 3D location information, which is managed by the agency master, is used as a weak constraint. That is, a 2D tracking result in each agent is prior to the 3D location information. The 3D information is substantially used for the verification whether the agent correctly tracks the person. It is also used for the recovery when the 2D tracking gets in failure.

(a) One successful agent

When only one agent is successfully tracking, a face location is constrained to a 3D line from the agent toward the face. Each member agent starts finding the face region along the corresponding epipolar line as shown in Fig.4(a). If a member agent finds a face, the agent starts tracking.

(b) Two or more successful agents

When two or more agents are successfully tracking, a 3D face location is estimated by finding an intersection of the 3D lines as shown in Fig.4(b). Since the two lines may not definitely intersect each other because of detection errors and a time lag, a least square method is used to decide an appropriate intersection.

The agency master estimates the 3D face location from tracking results of successful agents, and sends

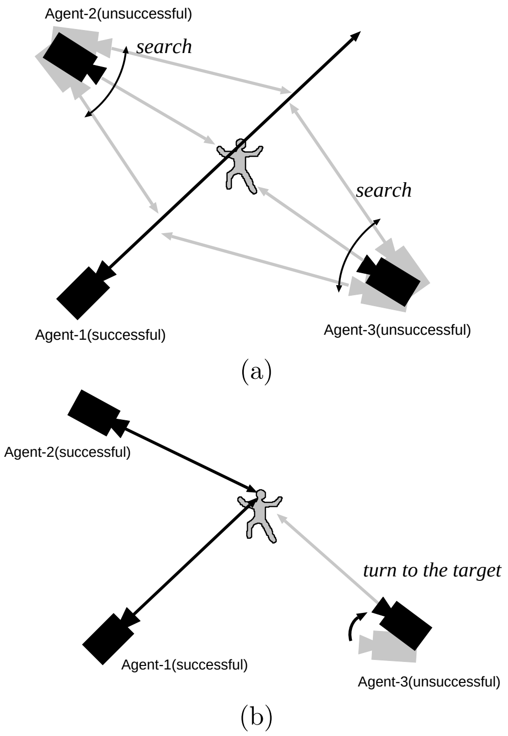

Figure 4: Verification and correction using 3D position of target person: (a)one agent is successfully tracking, (b) two or more agents are successfully tracking

the 3D face location to the member agents. Each successful agent checks a consistency of the 3D location and a tracked 2D location by the distance from the 3D location to the line of sight through the 2D location. If the distance is smaller than a threshold, these locations are regarded consistent. Then, the agent independently continues tracking. Otherwise, the tracking is unsuccessful.

Each unsuccessful agent turns a camera to the received 3D location. The agent tries to find a face region around the received location. If a face is found there, the agent starts tracking.

4.2 Experimental Results

Let us present experimental results of the asynchronous cooperative tracking. Four agents were used in this experiment. Figure 5 shows some examples of input image sequences of four agents. A rectangle of each image indicates the face region which is independently tracked by the agent. First, Agent-1 detected a new target person at t_1, and started tracking. Then, Agent-3 and 4 started finding the face along the lines indicated by broken lines. When Agent-4 found the face at t_4, the agency started the 3D tracking. Although Agent-3 failed in tracking at t_4, it was recovered before t_7. At t_9, Agent-1 and 4 lost the target person because he went out of the observation area, but Agent-2 and 3 still continued tracking. We can see that the agents cooperatively tracked the target person, and failures of the independent tracking were recovered.

Figure 5: Input images of four agents

Figure 6 shows locations of four agents and the target person. White marks indicate locations of the agents, and a black mark indicates an estimated location of the target person. Each line from the agent through the target person indicates the 3D line information which was sent to the agency master. We can see that the tracking was stably continued, even though the estimated location of the target person contained some errors.

5 Global Cooperation Layer

5.1 Dynamic Agency Configuration

A variety of face images should be registered in the dictionary for the robust recognition using the subspace method[1] because face views change according to a pose and a lighting condition. Assume that multiple persons are being tracked by some agents. If the assignments of agents to persons are fixed, a variety of face images cannot be registered because the agent tends to register almost the same images when the target person doesn't move. In order to effectively register a variety of face images, all the agents had better change their target persons if possible. If target persons can be exchanged between agents, more effective registration can be expected. To realize this, a concept of dynamic agency configuration is introduced.

The dynamic agency configuration allows the assignments of agents to the persons changed by trans-

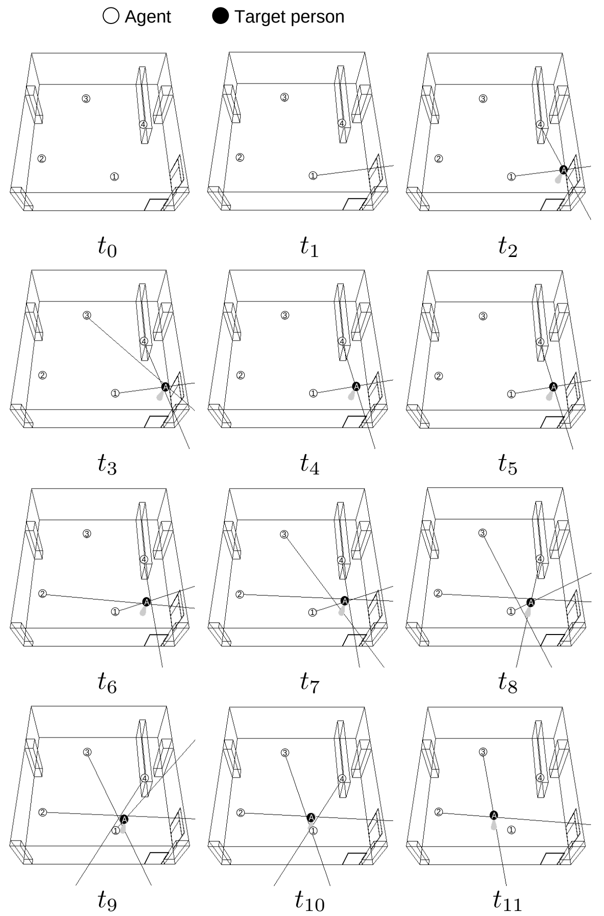

Figure 6: Results of the cooperative tracking in an agency

ferring an agent between agencies. The registered images are evaluated how well they cover a variety of face views for the dynamic agency configuration. When the evaluation is not sufficient, an agent transfer contract is made according to the agent transfer protocol. The detail functions are described in the following two sections.

5.2 Evaluation of the Registered Images

An estimation function is defined to evaluate an effectiveness of the registered images. Suppose that a subspace is created for a particular person. If the person is effectively registered, the DFFS (distance from feature space)[6] becomes small between the subspace and any faces of the person. On the other hand, the DFFS often becomes large if the subspace is imperfect. These facts mean that the DFFS indicates an effectiveness of the registered images.

However, face images cannot be prepared in all the lighting conditions in our problem because the target person cannot be specified in advance. Average face images are used instead of the target person. Although average face images are definitely different from those of a particular person, they keep the lighting informations in the average meaning. To create the average face images, face images are taken over many persons in many lighting conditions, and they are averaged in each lighting condition. Figure 7 shows the frontal average face images of all the lighting conditions in the current implementation. The light source is set at different locations in 25 lighting conditions (actually 24 images are used, because one light source is blocked up by the camera).

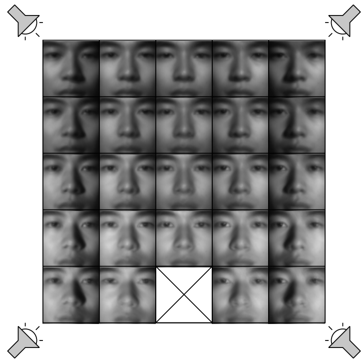

Figure 7: Frontal average images in all the lighting conditions

The estimation function is defined as the maximum DFFS between the subspace and average face images in all the lighting conditions. If the value of estimation function is small, it implies that an enough image set is already registered.

5.3 Agent Transfer Protocol

An agent transfer protocol is made up of some messages among agency masters. The agency master collects the information of the target person which includes 3D location and direction, and evaluates the estimation of registered images in the three directions. Agency masters send the messages to each other according to the information, and make a contract for the agent transfer. The protocol is defined with the following messages.

(a) "Request" message:

If the estimation in a direction is larger than a threshold, the agency master(sender agency master) sends a "Request" message with a 3D location, direction and estimation of the target person to the other agencies in order to get more agents.

(b) "Ack" message:

When an agency master(receiver agency master) receives a "Request" message, it finds out agents that can observe the target person in the requested direction if the agency contains enough member agents. If there are more candidates, contributions to the agency are compared, and the least contributor is selected, where the contribution of the agent is the estimation of the direction which is observed by the agent. If the agent is unsuccessfully tracking, the contribution is set to zero because it is the least contributor. If the estimation is smaller than the estimation included in the "Request" message, the agency master sends back an "Ack" message with the estimation.

(c) "Done" and "Cancel" messages:

When the sender agency master receives "Ack" messages, it selects an "Ack" message with the smallest estimation. The master send a "Done" message to make a contract with the sender of the "Ack" message. Then it also sends "Cancel" messages to the others. Finally, the agent is transferred.

5.4 Experimental Results

We show the experimental results of the multi-target tracking by dynamic agency configuration. In this experiment, eight agents were used to track the two target persons as shown in Fig.8. At t_1, the first target person(Target-A) was detected by Agent-1, and a new tracking agency(Agency-A) was generated. Target-A was tracked by Agency-A during t_1 - t_3. At t_4, the second target person(Target-B) was detected, and Agency-B was generated. We can see that the agency configuration was dynamically changed while the agencies continued tracking both A and B during t_6 - t_{11}.

6 Conclusion

This paper discussed cooperative distributed face tracking for the effective registration. A three-layer framework is proposed which consists of the independent task layer, the local cooperation layer and the global cooperation layer. We confirmed that each layer well worked in a real environment and that face images of multiple persons were effectively tracked and registered by the cooperation of agents.

Acknowledgment

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

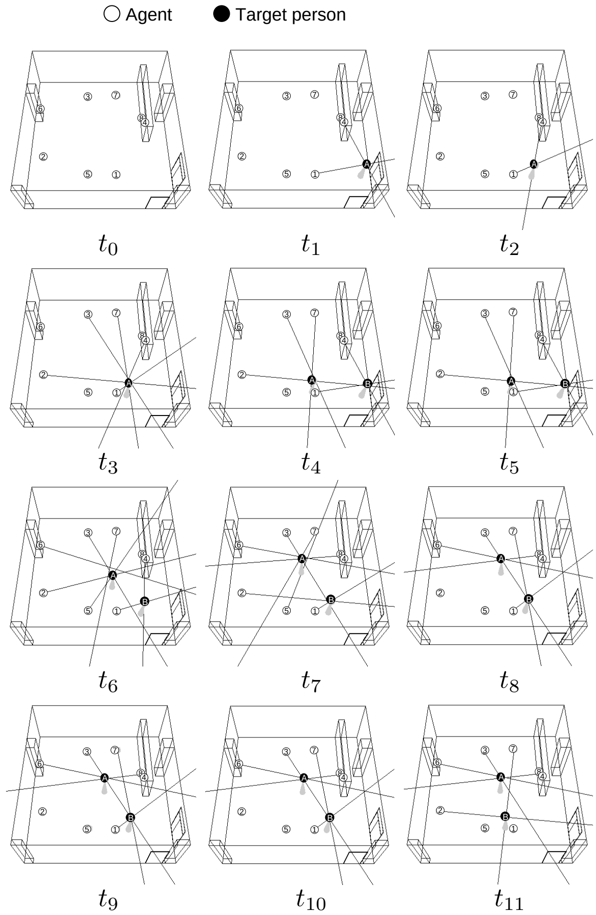

Figure 8: Agency configuration and locations of the target persons

References

[1] Ariki, Y., Ishikawa, N. and Sugiyama, Y., "Extraction and Recognition of Facial Regions by Subspace Method," Proc. ACCV'95, vol. III, pp 738–742, 1995.

[2] Haritaoglu, I., Cutler, R., Harwood, D. and Davis, L., "Backpack: Detection of People Carrying Objects Using Silhouettes," Proc. ICCV'99, pp.102–107, Kerkyra, Greece, Sept. 1999.

[3] Kirby, M. and Sirovich, L., "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp.103–108, Jan. 1990.

[4] Kato, T., Mukaigawa, Y. and Shakunaga, T., "Cooperative Distributed Registration for Face Recognition in Natural Environment," Proc. ACCV2000, pp.729–735, 2000.

[5] Matsuyama, T., "Cooperative Distributed Vision – Dynamic Integration of Visual Perception, Action, and Communication –," Proc. Image Understanding Workshop, pp. 365–384, Monterey, CA, Nov. 1998.

[6] Moghaddam, B. and Pentland, A., "Probabilistic Visual Learning for Object Representation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.19, no.7, pp.696–710, July 1997.

[7] Shakunaga, T., Ogawa, K. and Oki, S., "Integration of Eigentemplate and Structure Matching for Automatic Facial Feature Detection," Proc. FG'98, pp.94–99, Nara, Japan, April 1998.

[8] Turk, M. and Pentland, A., "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, vol.3, no.1, pp.71–86, Jan. 1991.