Cooperative Distributed Registration for Face Recognition in Natural Environments

Takekazu KATO, Yasuhiro MUKAIGAWA and Takeshi SHAKUNAGA
Department of Information Technology, Okayama University
Tsushima naka 3-1-1, Okayama, 700-8530, Japan

ABSTRACT

This paper proposes a framework of the cooperative distributed registration for effective face recognition. In order to realize a robust recognition in natural environments, a variety of face images should be registered in a dictionary. Multiple distributed cameras cooperatively work as an agency for this purpose. A concept of dynamic assignment is introduced for realizing effective registration in the flexible agency configuration. Each camera is appropriately assigned to a person according to the variety of registered images. In order to verify the effectiveness of the cooperative distributed registration, both the cooperative registration and the uncooperative registration were compared on a set of simulation image streams. Experimental results show that the effective images can be registered by the cooperative distributed registration.

Keywords: cooperative distributed vision, face registration, face recognition, person observation, lighting condition.

1. INTRODUCTION

If a good technology is established for a person observation in a wide area such as the whole of a building, it can be applied not only to security uses but also to comfortable interfaces between humans and environments. It is important to robustly identify the person in natural environments to realize such a person observation. Face recognition seems a feasible way to the person identification because a face is an effective person identifier and it is easily observed by a camera.

A lot of methods have already been developed for face recognition [3, 5, 7, 8]. They recognize persons whose face images have already been registered in the dictionary. The learning images are supposed to be taken in the controlled environments. It is, however, difficult to apply the conventional methods for the face recognition in natural environments because the lighting condition cannot be controlled there.

In order to recognize a face, the effective registration is also important, especially in natural environments. If sufficient face images are registered, the face recognition is reduced to the conventional method. We propose a novel concept called the cooperative distributed registration to realize the effective registration. Multiple distributed cameras are used to acquire a variety of face views, and they cooperatively work for the effective registration. This paper discusses a framework of the cooperative distributed registration for the robust face recognition.

2. CONCEPTUAL ASPECT OF PERSON REGISTRATION

2.1. Person Registration for Recognition in Natural Environments

A view of face is considerably affected by lighting conditions and face directions. It is necessary to cope with a variety of views for the robust face recognition in natural environments. The subspace method[1] robustly recognizes a particular face in a variety of lighting conditions, when all the varieties can be registered in the dictionary. The subspace method is effective if target persons are specified in advance. The face images of unknown persons, however, cannot be registered in advance. In such a case, we should devise either the registration or the recognition because the sufficient dictionary cannot be prepared.

Intensities normalization in a local region is effective for the robust recognition in different lighting conditions[2]. It is, however, difficult to robustly recognize if the change of views is not small. Such a local normalization is not effective enough for our purpose.

In our problem, face recognition should be quickly accomplished when the target person moves. Registered persons and unregistered persons must be quickly distinguished while the person is being tracked. The subspace method seems to satisfy this requirement when a sufficient dictionary is made up. If the subspace method is used in the recognition phase, our problem can be reduced to registration of unknown faces which cover a variety of views.

2.2. Hierarchy of Registration

The registration methods are classified into three classes, the isolated registration, the distributed registration and the cooperative distributed registration. Characteristics of them are summarized as follows:

Isolated Registration: The face images are registered by a single camera. Since the direction of the registered face depends on a relative pose of the target person to the camera, a variety of views cannot be registered. Furthermore, if there are multiple persons in the scene, the camera cannot track them at the same time.

Distributed Registration: The face images are registered by the multiple distributed cameras. Each camera independently works. Even if there are multiple persons in the scene, each camera independently registers a target person. Since each camera is assigned to the particular person, the registered images are not expected to cover a variety of views.

Cooperative Distributed Registration: The face images are registered by the multiple distributed cameras. They cooperate in registering face images. Since each camera is dynamically assigned to the effective person, an efficient image set can be registered for each person.

3. IMPLEMENTATION OF COOPERATIVE DISTRIBUTED REGISTRATION

3.1. Overview of Registration and Recognition System

3.1.1. Components of the System

Our system comprises a set of agents and two special-purpose units.

Agent: An agent is an entity which consists of an active camera and a processor. The active camera is controlled by the processor. Each agent has ability to detect and to track a person, and can communicate with another via a network.

Detection and Tracking Agencies: An organized subset of agents is called an agency and it plays a central role in the system. Each agency comprises a special agent called an agency master, and zero or more agents called workers. Each agent should belong to an agency as the agency master or a worker. An agency master is defined as an agent which possesses a master privilege, and it works as a manager of the agency. That is, all the information on the agency members is maintained by the agency master. The information on the agency task is also maintained by the master.

There are two types of agencies in the system. The first type is called a detection agency, which has a task to detect a person in the territory of the agency. The second type is called a tracking agency, which has a task to track a particular person.

A tracking agency has a mode, which indicates a specific task of the agency. The mode is changed when the task is accomplished, but changes are not reversible. A tracking agency may change its members due to the mode change. Once the person goes out of the sight, the agency terminates and all the workers as well as the agency master are transferred to other agencies. The tasks of tracking agency are specified in the following three modes.

[T1] **Recognition Mode:** Track the target person in order to get one or more face images in an appropriate (frontal, left or right) direction for face recognition. Then ask the recognition unit to judge whether the face images can be identified with one of the registered persons. If the person is identified with someone in the data base, and sufficient face images are already registered for the person, the mode of the agency is changed to [T3]. Otherwise, the mode of the agency is changed to [T2].

[T2] **Registration Mode:** Track the target person in order to get a face image set for the registration of the target person in the three (frontal, left and right) directions. When the registered images are sufficient to robustly recognize in all the three directions, and the mode of the tracking agency is changed to [T3]. Otherwise, the tracking agency continues tracking in the registration mode.

[T3] **Tracking Mode:** Simply track the identified person. No face image is taken in this mode.

Registration and Recognition Units: The registration unit receives images from a tracking agency in the registration mode, and it creates a face dictionary for the recognition. The recognition unit receives a face image from a tracking agency in the recognition mode, and recognizes it with someone in the dictionary. Finally, the unit reports the recognition result to the tracking agency.

The implementation of the units should depend on the size of the system, performance of the processors, etc. The two units may be implemented within an agent when it is implemented on a high performance processor. On the other hand, they may be implemented as a server when the system is small or when agents are implemented on low performance processors.

3.1.2. Cooperative Behavior of the System

In the cooperative distributed vision system, the whole scene is cooperatively observed by the distributed cameras, even if each of them can observe just a small area. If all the agents directly communicate each other, a lot of communication is necessary. To reduce the communication cost, our system prepares two layers of cooperations; the first is a local cooperation within an agency and the other is a global cooperation between agencies.

Local Cooperation within an Agency: The task of an agency is accomplished by cooperation within the agency. The task depends on the agency type and its mode. In the detection agency, the observation area of each agent is allocated by the agency master. In the tracking agency, information on the target person is exchanged among the agency members. The tracking agency recognizes a face in the recognition mode, and registers a face image of the target person in the registration mode. The results of recognition and registration are collected to the agency master.

Global Cooperation between Agencies: The cooperation between agencies controls the whole system by dynamically changing the agency configuration. It is realized by the communication between the agency masters. The change of the agency configuration is classified as follows:

- **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.

- **Agent Transfer between Agencies**

  When an agency contains an unnecessary agent, the agency releases the agent to another agency. When an agency requires more agents, some agents are transferred from other agencies.

In the initial state, there are one or more detection agencies and no tracking agency in the system, and all agents are included in the detection agencies.

Figure 1 shows a typical change of the agency configuration. There are thee detection agencies. When an agent detects a person, a tracking agency is generated (Fig. 1(a)). Then, the tracking agency tracks the target person in the recognition mode. If no person is registered yet, the mode is immediately changed to the registration mode. In the registration mode, several agents join the tracking agency to accomplish the registration as soon as possible (Fig. 1(b)). When the registration is perfectly done, the mode is changed to the tracking mode, and the agency no longer requires a lot of workers for the task. If there is another tracking agency in the neighborhood, and it is in the registration or recognition mode, some workers are transferred to the tracking agency (Fig. 1(c)). Finally, when the target person goes out, the tracking agency terminates (Fig. 1(d)).

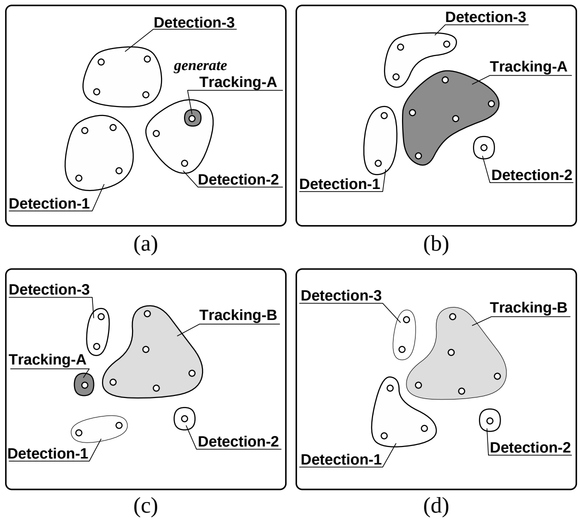

Figure 1. Typical change of the agency configuration

3.2. Fundamental Functions

3.2.1. Person Detection and Tracking

When a new person comes into the observation area, it is necessary to detect the person and to hand it over to the tracking agency. This task is accomplished by the distributed detection agents. Each agent independently detects a person within the observation area. Since it is unknown where and when a person appears, a detection agent has to quickly search a wide area without overlooking any person.

In our system, face region is detected by using color information. The colors of skin and hair are registered in advance, and a face region is detected by the combination of the both colors. This method is robust against a wide variety of lighting conditions, when colors are represented in hue and saturation. Combining this method and the moving object detection, an active vision agent detects a person in the natural environments.

A tracking agency has a task to track a particular person. Agents in a tracking agency should cooperatively track a target person. Matsuyama and Wada[4] proposed a cooperative object tracking method, and our system is based on it. Although their method focuses on the position of the target person, our system needs more information for the recognition and the registration. A required information in each mode is summarized as follows:

Recognition Mode: The agency extracts face images in the appropriate directions for the recognition. For this purpose, the 2D face position and the direction are necessary.

Registration Mode: The agency registers a face image set, and communicates with other agencies. In order to transfer an agent, it is necessary to communicate the 3D face position and the direction from the agent.

Tracking Mode: The agency tracks the identified person without any special jobs. Thus, the only 2D body position is necessary.

3.2.2. Face Extraction

A face image should be extracted both for registration and for recognition. In our system, the face image is extracted by combination of the eigenface method[5, 6, 8] and a facial structure analysis[7]. An eigenface is prepared for each direction (frontal, left and right) to cope with the pose change. To adjust to the scale change, each eigenface is converted in some scales. Thus, some candidates of face regions are picked up by finding the region whose DFFS (distance from feature space)[5] is smaller than a threshold.

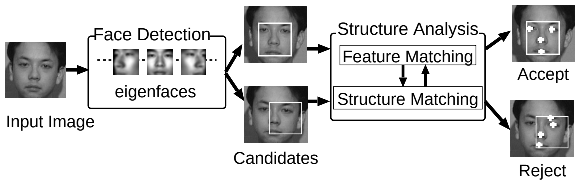

Figure 2. Face extraction and verification

The eigenface method is often unstable against the background changes as well as the direction and scale changes. The facial structure analysis is also used to make up for these weak points. It checks the candidates by the relative positions of face components such as eyes, a nose and a mouth. The most appropriate face region is finally selected as shown in Fig.2.

3.2.3. Face Recognition by Subspace Method

Since a face image changes due to a face pose and a lighting condition, face recognition should also cope with them. In order to adjust to the pose changes, face images are registered in each direction (frontal, left and right), separately. In order to adjust to the light changes, a subspace is created from registered images which are taken in different lighting conditions. The subspaces are created in the three directions for each person, and the recognition is accomplished for each direction.

A face image extracted by the tracking agency is recognized using the DFFS and DIFS (distance in feature space) for all the registered persons. If the smallest DFFS is more than a threshold, the face is judged to be unknown.

3.3. Framework for Cooperative Distributed Registration

3.3.1. Required Images for Robust Recognition

It is necessary to deal with both the lighting condition changes and face direction changes for carrying out the robust recognition in the natural environments. To cope with the lighting condition changes, the subspace should cover a variety of lighting conditions. To cope with the face direction changes, the subspace should be prepared for each direction. Therefore, the robust recognition is realized by creating the subspace which covers all the possible lighting conditions for each direction.

If each camera is assigned to a particular person, registered images hardly cover the possible conditions. In order to efficiently register face images of target persons moving freely in the observation area, tracking agencies had better dynamically exchange the target persons. The registered images should be estimated how well they cover the possible lighting conditions for this purpose. The agency configuration are dynamically changed according to the estimation.

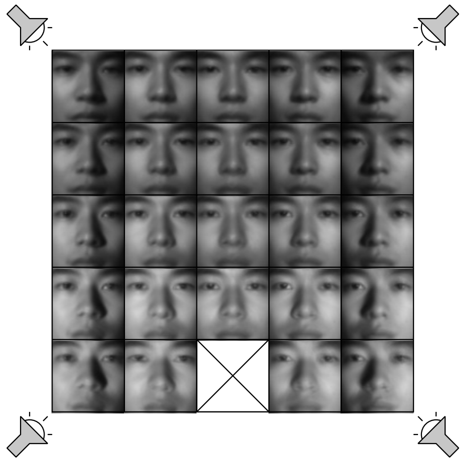

Figure 3. Average images in 24 lighting conditions

3.3.2. Variety Estimation of Lighting Conditions

The estimation function is defined to estimate how well the registered images cover the possible lighting conditions. Let us define the function using a subspace which is created from the registered images.

It is known that the DFFS is small for a face image of the target person in any lighting condition when the subspace has been created from sufficient images of the same person which can cover all the lighting conditions. The face images, however, cannot be prepared in all the lighting conditions when the target persons are unspecified in advance. The average face images are used instead of them, which were created in all the considered lighting conditions from face images over many persons.

The estimation function is defined as a maximum of the DFFS to the average images when the subspace was created from the registered images. If the estimation function is small, it suggests that registered images enough cover the lighting conditions.

Figure 3 shows the average face images for each lighting condition in the current implementation. The light source is set at 25 different directions (actually 24 images are used, because one light source is blocked up by the camera).

3.3.3. Dynamic Agency Configuration

The effective agency configuration depends on some factors such as the state of registered images, a location of target person and the lighting conditions. The dynamic configuration is realized by transferring an agent which does not contribute for the effective registration. The agent transfer is accomplished by the following process.

2. The registered image set is estimated in each direction. After another image is registered by an agent, the estimation is updated. The difference of estimation is regarded as a contribution of the agent.
3. If the contribution of the agent (Agent-A) in the agency (Agency-1) is smaller than a certain threshold, Agency-1 sends a request message for another target person to the other agencies. The additional information, which includes the 3D location and the observation range of the Agent-A, is also attached to the message.
4. When an agency receives this message, it calculates the face direction of their target person from Agent-A. If the agency does not have an agent which can register the direction and the estimation of the direction is smaller than a threshold, the agency sends a reply message to Agency-1 along with the estimation.
5. After Agency-1 receives reply messages from some agencies, Agency-1 selects an agency (Agency-2) which replied with the maximum estimation result. Agency-1 contracts with Agency-2 to transfer Agent-A. Finally, Agency-1 sends cancel messages to the other agencies.
6. Agent-A is transferred from Agency-1 to Agency-2. Agent-A is expected to contribute in Agency-2.

4. EXPERIMENTAL RESULTS

4.1. Relation between Estimation and Recognition Rate

A basic experiment is accomplished for checking the simultaneous distribution of the estimation and the recognition rate. First, face images were taken for 50 persons in 24 lighting conditions. A set of the following trials were made on the image set. In each trial, four images were selected at random for each person, and estimated as mentioned in 3.3.2. The rest 20 images were used for the recognition. In the experiment, 100 trials were made for each person.

Figure 4 shows the simultaneous distribution of the estimation and the recognition rate. The recognition rate closely correlates with the estimation of the registered images. We can see that the proposed estimation is effective to estimate a variety of the lighting conditions.

4.2. Registration

We show results of the registration by a single camera. A face region was detected and tracked with color information. Figure 5 shows an example of the face tracking. A rectangle with solid lines indicates a face region, and a rectangle with broken lines indicates a hair region. In order to keep the face region in the center of images, the pan-tilt-zoom camera (SONY EVI-G20) was automatically controlled. The tracking process was accomplished within 1/30 seconds per frame (video rate).

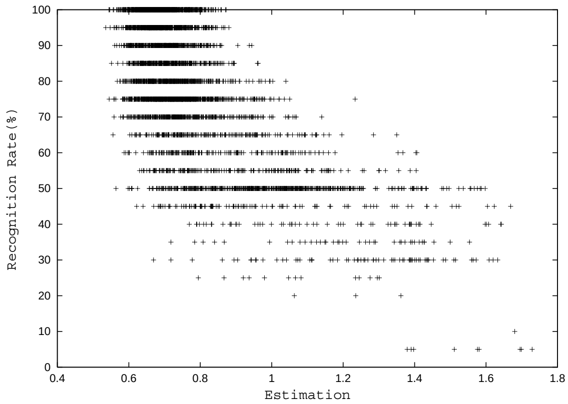

Figure 4. Simultaneous distribution of the estimation and the recognition rate

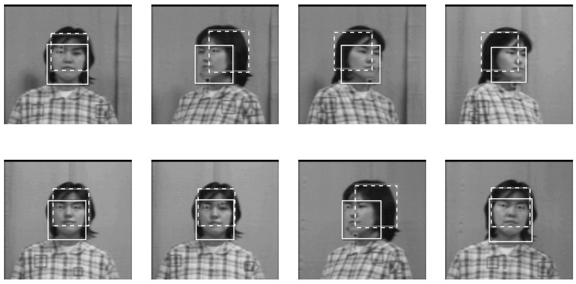

Figure 5. Result of tracking

Figure 6. Eigenfaces

Figure 7. Registered images

After a face image is taken, the face direction is roughly estimated by the eigenface method. The eigenspaces are constructed from the learning face images which were taken for 10 persons with 3 directions (frontal, left 45 degrees and right 45 degrees) in 24 lighting conditions. Each of the three eigenfaces is represented in 15 dimensions. Figure 6 shows the eigenfaces in the three directions.

Figure 7 shows the results of the face extraction with the eigenfaces. The label (Frontal, Left or Right) in each image indicates the estimated direction. We can see that the face was correctly detected and that the direction was correctly estimated. It takes about 5 seconds to extract a face. We are trying to implement

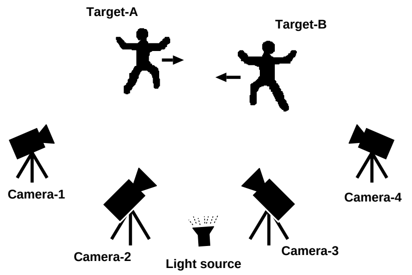

Figure 8. Configuration of cameras, persons, and the light source

Figure 9. An example of input images

this process on the parallel processor for the practical use.

4.3. Dynamic Agency Configuration

The effectiveness of the cooperative distributed registration is verified in the following experiment. Both the cooperative registration and the uncooperative registration were examined on a set of simulation image streams.

In the simulation streams, two persons (Target-A and Target-B) are registered by four cameras (Camera-1, Camera-2, Camera-3 and Camera-4) in one light source. Eight image streams are prepared for all the combinations of the persons and the cameras. Figure 8 shows a configuration of the cameras, the persons, and the light source. We assume that the persons are walking in the observation area without any occlusions. Each stream consists of 21 images. Figure 9 shows an example stream which corresponds to the combination of Camera-1 and Target-A.

In the initial state, two tracking agencies are created in the system. One agency includes two agents, Camera-1 and Camera-2, and the agency has a task to register Target-A. The other agency includes two agents, Camera-3 and Camera-4, and the agency has a task to register Target-B.

Uncooperative case: Figure 10 shows the registered images in the uncooperative registration. Since no agent is transferred to the other agency, the only partial set of images are registered. Figure 11 shows the changes of estimation. Note that the total of the final estimation is still large and that an effective image set is not registered because each agent is wastefully occupied by an agency.

Cooperative case: Figure 12 shows the registered images in the cooperative registration. Agents are dynamically transferred between the two agencies to reduce the estimations of the registered images. The numbers of images are almost the same in the three directions. Figure 13 shows that an effective image set is registered because each agent was suitably transferred between the two agencies.

5. CONCLUSIONS

We proposed a framework of the cooperative distributed registration which realizes the effective registration for the robust face recognition. Multiple distributed cameras cooperate each other for the effective registration. Each camera is dynamically assigned to the suitable person according to the variety estimation of the registered images. Experimental results show the effectiveness of our system. Now, we are going to implement the cooperative distributed system with more cameras for the practical use.

6. ACKNOWLEDGMENT

This work was supported by the Research for the Future Program of the Japan Society for the Promotion of Science, 'Cooperative Distributed Vision for Dynamic Three Dimensional Scene Understanding' Project (JSPS-RFTF96P00501).

References

[1] Ariki, Y., Ishikawa, N. and Sugiyama, Y., "Extraction and Recognition of Facial Regions by Subspace Method," Proc. ACCV'95, pp-III-7 38-III-742, 1995.

[2] Doi, M., Sato, K., and Chihara, K., "A Robust Face Identification against Lighting Fluctuation for Lock Control," Proc. FG'98, pp.42–47, 1998.

[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, 1990.

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

[5] 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, 1997.

[6] Pentland, A., Moghaddam, B. and Starner, T., "View-based and modular eigenspaces for face recognition, " Proc. CVPR'94, pp. 84–91, 1994.

[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, 1998.

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

Figure 10. Registered images without cooperation

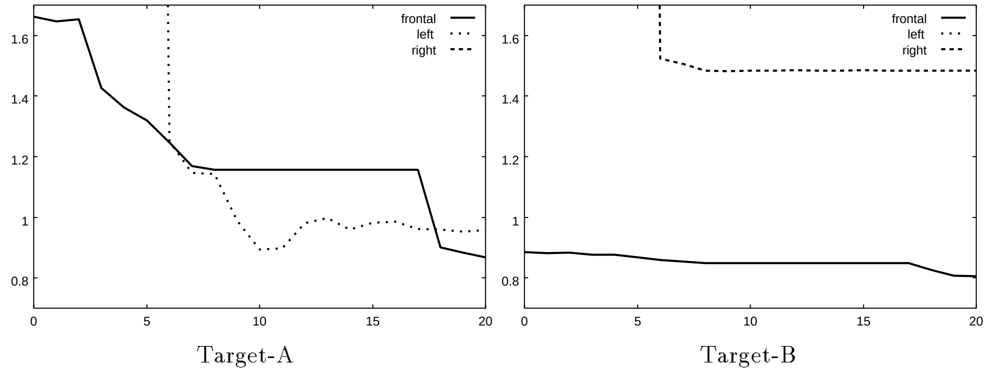

Figure 11. Changes of estimation for the three directions without cooperation

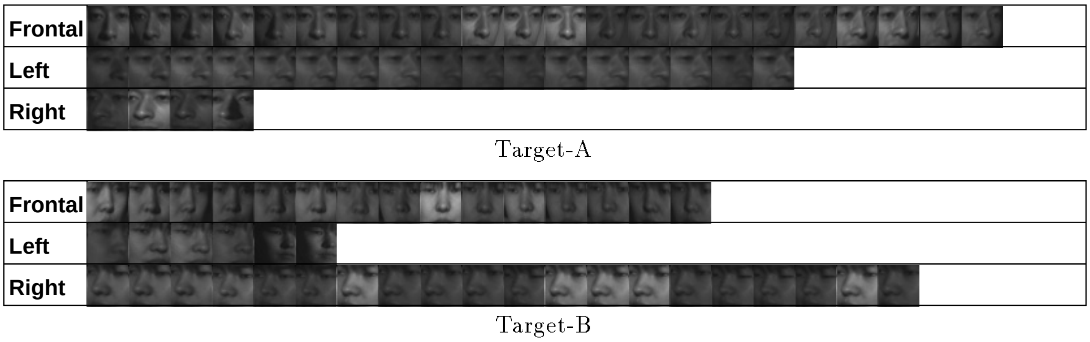

Figure 12. Registered images with cooperation

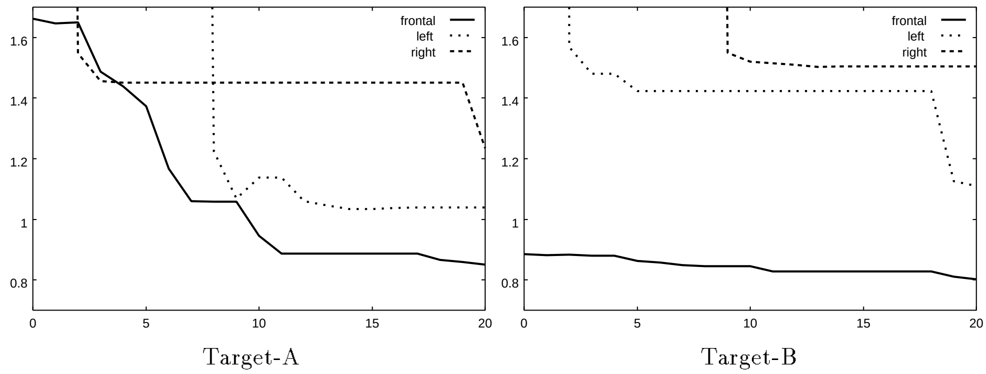

Figure 13. Changes of estimation for the three directions with cooperation