Chapter 6
Face Recognition at a Distance: System Issues

Meng Ao, Dong Yi, Zhen Lei, and Stan Z. Li

Abstract Face recognition at a distance (FRAD) is one of the most challenging forms of face recognition applications. In this chapter, we analyze issues in FRAD system design, which are not addressed in near-distance face recognition, and present effective solutions for making FRAD systems for practical deployments. Evaluation of FRAD systems is discussed.

6.1 Introduction

Research and development of face recognition technologies and systems have been done extensively for decades. In terms of distance from user to the camera, face recognition systems can be categorized into near-distance (often used in cooperative applications), middle-distance, and far-distance ones. The latter cases are referred to as face recognition at a distance (FRAD).

According to the NIST's face recognition evaluation reports on FERET and FRGC tests [7] and other independent studies, the performance of many state-of-the-art face recognition methods deteriorates with changes in lighting, pose, and other factors. Those factors which can affect system performance are summarized into four types: (1) technology, (2) environment, (3) user, and (4) user–system interaction, shown in Table 6.1.

For near-distance face recognition, camera can easily capture high-resolution and stable face images, but in FRAD systems, the quality of face images become a big issue. The user–system interaction in middle to far face recognition systems are not so simple. To build a robust FRAD system, these issues should be solved: resolution, focus, interlace effect, and motion blur.

In FRAD systems, image sequence from a live video is usually used for tracking and identifying people of interesting. Video-based face recognition is a great challenge in face recognition area, which attracts many researchers attentions in recent

---

M. Ao (✉)
Center for Biometrics and Security Research and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science, Beijing 100190, China
e-mail: mao@cbsr.ia.ac.cn

Table 6.1 Performance affecting factors

Aspect	Factors
Technology	Dealing with face image quality, heterogeneous face images, and problems below
Environment	Lighting (indoor, outdoor)
User	Expression, facial hair, facial ware, aging
User–System	Pose (alignment between camera axis and facing direction), height

years [11]. McKenna et al. [6] modeled face eigenspace in video data via principal component analysis. Probabilistic vote approach is used to fuse the sequence information. Zhou et al. [12, 13] took advantage of the time and temporal information to improve the recognition performance. In [8], an active face tracking and recognition system is proposed, in which two cameras, a static and a PTZ, work cooperatively. The static camera is used to take image sequences for face tracking while the PTZ camera is used for face recognition. In this way, the system is supplied with high-quality images for face recognition since the PTZ camera can be adjusted to focus on the face to be recognized. However, the above recognition methods are initially developed to recognize one person in video sequence. Therefore, how to fuse the temporal and identity information for recognizing multi-faces in one scene is still an open problem to be studied.

This chapter is focused on issues in FRAD systems using video sequences. It is organized as follows. Section 6.2 provides an analysis of problems in FRAD systems. Section 6.3 presents solutions for making FRAD systems. Section 6.4 presents two examples of FRAD systems: the face verification system used in Beijing 2008 Olympic Games, and a system for watch-list face surveillance in subway. Finally, how to evaluate FRAD systems is discussed in Section 6.5.

6.2 Issues in Video-Based Face Recognition

6.2.1 Low Image Resolution

Low resolution is a difficult problem of face recognition at a distance; see Fig. 6.1. In this case, the view of the camera is usually wide and the proportion of the face in the whole image is small. So the facial image is always at low resolution which degenerates both the performances of face detection and the recognition engines.

While there is a long way to go to develop reliable algorithms to achieve good performance with low-resolution face images, using a high-definition camera is a current solution to this problem. However, a high-resolution image will decrease the speed of the face detection.

Fig. 6.1 High-resolution image (*left*) and low-resolution one (*right*)

6.2.2 Out of Focus

In the application of face recognition at a distance, the distance between the face and the camera is in a spacial extant. That means in the most cases the face is out of the focus of the lens which makes the face image blur; see Fig. 6.2.

Fig. 6.2 The face at the focus (*left*) and the face out of focus (*right*)

Although the focus is conceptually a point, physically the focus has a small extent, which is called the blur circle. This non-ideal focusing is caused by aberrations of the imaging optics. Aberrations tend to get worse as the aperture diameter increases. So using a small aperture lens can decrease the degree of the blur.

6.2.3 Interlace in Video Images

Interlace refers to the methods for painting a video image on an electronic display screen by scanning or displaying each line or row of pixels. This technique uses two fields to create a frame. One field contains all the odd lines of the image, the other contains all the even lines of the image. Because each frame of interlaced, video is composed of two fields that are captured at different moments in time; interlaced video frames exhibit motion artifacts if the faces are moving fast enough to be in different positions when each individual frame is captured. Interlace increases the difficulties to correctly detect and recognize the face image; see Fig. 6.3.

Fig. 6.3 The image captured by CCTV camera with interlace problem

To minimize the artifacts caused by interlaced video, a process called de-interlacing can be utilized. However, this process is not perfect, and it generally results in a lower resolution, particularly in areas with objects in motion. Using a progressive scan video system is the ultimate solution of this problem.

6.2.4 Motion Blur

Motion blur is a frequent phenomenon in digital image system. It may occur when the object is moving rapidly or the camera is shaking; see Fig. 6.4. To avoid the motion blur, the camera should use rapid exposures, which causes a new problem. When taking rapid exposures, the aperture stop should be increased. This makes conflict to the out-of-focus problem.

Fig. 6.4 When the exposure
of the camera (progressive
scan camera) is not rapid
enough, the motion blur
occurs

6.3 Making FRAD Systems

For a cooperative user system, the level of the cooperation of the users directly
determine the ultimate performance. How to make the users feel natural and barrier-
free is a problem of designing a face recognition system. A good design of the face
recognition system could substantially not only improves the practical application
performance, but also enhances the users' satisfaction with the system. The design
of a cooperative user system are mainly related to the following questions: how to
cover most of the user's height, how to get frontal face images, and how to capture
high-quality images.

For a non-cooperative user system, there are also some hints to increasing the per-
formance. To combine the tracking technology and recognition technology together
would get a better result. The system first tracks the person's face. Then the system
would get a series of images of a same person. Using these images to recognize
a person is easier than just using a single image. This method receives a higher
accuracy.

6.3.1 Cover Most Users' Heights

For a cooperative user system, different users with different heights brings a great
problem. How to make the vision of camera cover most of the users' height is an
important problem of designing the system. There are usually two solutions: using

Fig. 6.5 Cover most users' heights: (**a**) single camera scheme and (**b**) multi-camera scheme

(a)   (b)
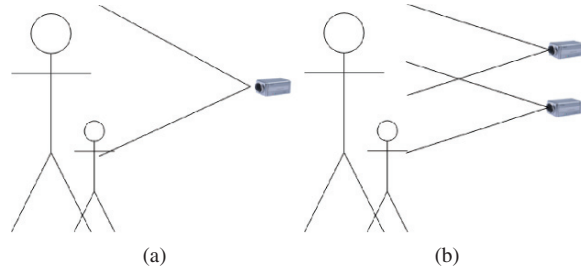
a single large vision camera and using multiple cameras; see Fig. 6.5. Both options have their pros and cons.

Using a single large vision camera is to directly cover most of the users' height. Staple camera's image aspect ratio is fixed, 4:3 or 16:9. Here, we rotate 90° to make the camera cover a higher field of vision. At this time the proportion of the face image is decreased due to the expansion of the vision of the camera. Take a 640×480 pixels camera as an example. If the camera is requested to cover the height of 1 m, the face image size is about 90 pixels with the eyes distance. Therefore, the scheme of using a single large vision camera always requests a high-resolution camera.

Using multiple cameras is to make each camera cover a different height. In multi-camera scheme, it is necessary to solve the problem how to use multi-images. There are usually two ways: merging the multi-images and using multi-images independently. Merging images brings a new preprocessing problem. When using multi-images independently, the multi-visions of the cameras should be overlapped in order that the face image is not cut into two images.

6.3.2 Capture Frontal Faces

The face algorithm gets the best results when the face images are frontal ones. How to make users to face the camera so as to capture the frontal face image is another system design problem. To achieve this purpose, we can place some devices to attract the attention of the user so that the system can capture the frontal face image. Placing a screen below the camera showing the images captured by the camera may be a good choice. The screen is able to attract the users' attention. Similar to the role of the mirror, most people would self-consciously watch the screen with their own image. As the distance between the screen and camera is close, watching the screen is nearly equal to watch the camera.

6.3.3 Capture High-Quality Images

The image quality is whether the clarity and exposure of the image meet the requirements. The main reason of blur is out of focus and the movement of the face and the exposure problem is mainly due to the changes of environmental light and the bright background light. To avoid out of focus, we can select a large depth of vision field lens and to avoid the motion blur, we can adjust the sensitivity of the camera and

Fig. 6.6 To dodge the sun
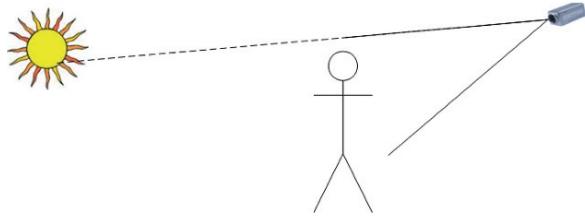
the speed of the shutter. When using analog video cameras, the speed of the capture device is another problem. When using high-resolution camera, the image captured by the decoding card will be jagged fuzzy because of the movement of the objects. One solution is to make the user keep stable during the recognition process. The auto-aperture lens is the only choice to solve the exposure problem caused by the change of the environment light. However, an auto-aperture lens camera captures the image with serious exposure problem in the case of bright background light, particularly when a strong illuminate such as the sun is in the field of camera vision. In order to avoid such a case, the camera should be placed at a high place; see Fig. 6.6.

6.4 Examples of FRAD Systems

6.4.1 Face Biometric for Beijing 2008 Olympic Games

The CBSR-AuthenMetric face recognition system is developed based on the above principles and has been used as a means of biometric verification in Beijing 2008 Olympic Games. This is the first time that a biometric is used for Olympic events; see Fig. 6.7(b). This system verifies in 1:1 mode the identities of the ticket holders (expectators) on the entry to the National stadium (Bird Nest). Every ticket holder is required to submit the registration form together with a 2 inch ID/passport photo attached. The face photos are scanned into the system. Every ticket is associated with a unique ID number. When the ticket is read in, the system takes the face images and compares them with the extracted face templates for the ID. The throughput for face verification (excluding walk-in and ticket reading times) is 2 seconds per person.

The system equipment consists of the following hardware parts: a CCTV camera, a PC system, a software system, a feedback LCD, and a standing casing. An industrial design of the system (with an RFID ticket reader incorporated) is shown in Fig. 6.7(a). The body–camera distance is about 1.5 m. The system should also take care of body height between 1.45 and 2.0 m.

The software system consists of three main modules: face detection, feature template extraction, and template matching. In the first, the input image is processed by AdaBoost and multi-block local binary pattern (MB-LBP)-based face detection [3, 10]. Effective MB-LBP and Gabor features are extracted and template matching classifiers are learned using statistical learning [1, 4]. The self-quotient image (SQI) technique [9] is used to deal with the illumination change problem.
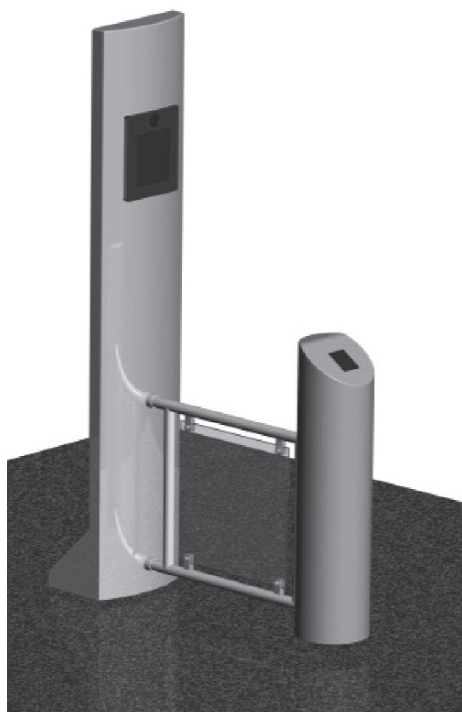
(a)

(b)

Fig. 6.7 Face verification used in Beijing 2008 Olympic Games. (**a**) The industrial design of the system. (**b**) On-site deployments and applications

The system has to deal with several technical challenges. It works outdoors between 3 p.m. to 8 p.m., so can face toward possible sunlight shed directly into the camera. This is the first challenge to the system. The second challenge is the non-standard photo images. Although requirements are specified as using 2 in ID/passport photos, some registrants use non-ID/passport photos, small photos, or unclear photos. The photo scanning process contains flaws: some scanned photos are out of focus, and some are scans of wrong parts of registration forms. Other changes are related to the coordination with other parts of the whole security system.

6.4.2 Face Surveillance

Face surveillance is a non-cooperative user application. In such settings, the system should be able to follow the faces when the people under tracking are not facing to the camera or when the people's state is not able to be recognized. In addition, mutual occlusions may occur as multiple faces are moving and interact with one another and some faces may disappear in several frames due to total occlusion. Moreover, the quality of the video is usually low because of the low resolution and object motion. Therefore, the system should track the faces and recognize the faces with a series of face images. This combination enhances the recognizing result. So the face tracking is necessary in such a task.

Figure 6.8 shows a method for incorporating face recognition into face tracking [5]. In the face tracking module, Gaussian mixture models (GMMs) are used to
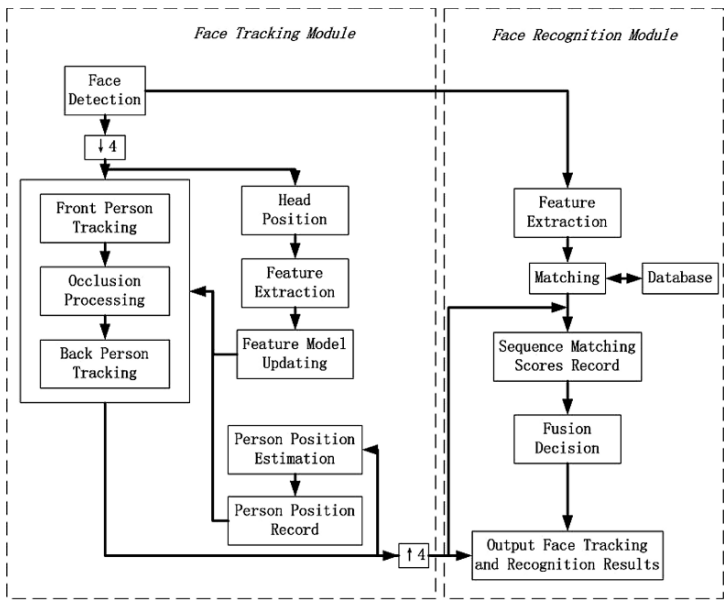

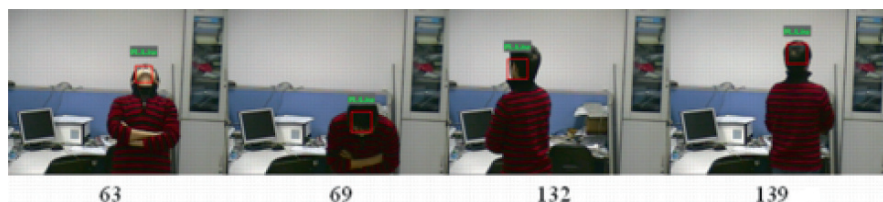**Fig. 6.8** Combining face tracking and face recognition

Fig. 6.9 Combining face tracking and face recognition can deal with largely rotated faces. From *left* to *right*: the face looking upward (frame 63), looking downward (frame 69), turning aside (frame 132), and turning back (frame 139)

represent the appearances of the tracked head and the upper body. Two GMMs are used to represent the appearance of each person. One model is applied to the head appearance to keep head tracking and to predict the head position in the next frame. The other is applied to that of the upper body to deal with occlusions. These two models are updated online.

The face recognition module uses an LBP and Adaboost method based on that in [2] to obtain identity matching scores for each frame. These matching scores are computed over time to obtain a score sequence. The matching scores are fused and used to help associate the tracked persons in consecutive frames, as well as to provide face recognition results. When the fused scores are very slow, the system will consider the corresponding persons have not been enrolled before. The recognition result can be shown on the tracked object; see Fig. 6.9.

This system is used in municipal subways for watch-list face surveillance. Subway scenes is often crowded and contains simultaneously moving objects including faces. Figure 6.10 shows a real scene. The cameras are fixed at the entrances and
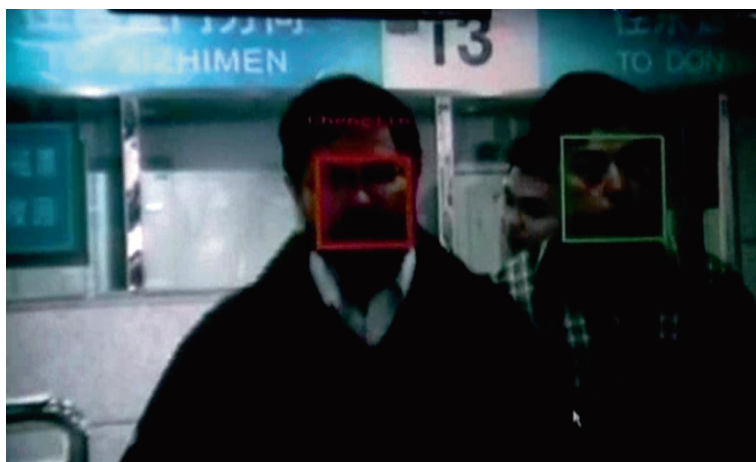

Fig. 6.10 Watch-list face surveillance at entrance of subways. A watch-list person is alerted by the *red* rectangle on the face

exits of the subway, where the people would face to the camera naturally. The system will automatically alarm when people in the watch-list appear in the field of view.

6.5 Evaluation of FRAD Systems

FRAD evaluations can be classified into three types: algorithm evaluation, application system evaluation, and application operational evaluation. Algorithm evaluation tests the performance using the data in a public database or a certain database for testing the accuracy of algorithms. Application system evaluation tests the face recognition system in the laboratory or a simulator environment. The face recognition system is constructed similarly to the real case. Some people test the system according to the process in real using. Application operational evaluation is to test the system in the real using. The system records the data in real using process for some time. The result of the testing is obtained by analyzing the log file of the system. These three types of system evaluations are ordered by the difficulty level increasing.

Algorithm evaluation of face recognition algorithm is a method which is used in a very wide range. In such an evaluation there are many well-known public database available such as used for FERET and FRGC. However, algorithm evaluation cannot be fully representative of the system in use as the final performance. There is a great distinction of the data between the real face recognition system and the database in personnel, the quality of the image shooting, the shooting environment, and the photography equipment used. So algorithm evaluation of the representative system can only be used for testing the performance of face recognition algorithm. Algorithm is the most crucial factor of a face recognition system performance, but not the only factor.

Application system evaluation is a mostly used method. In the simulation environment, the user tests the system in accordance with the real use of the system processes. In such a test, the simulated environment is different from the real environment in lighting condition and others. Also the users are different from the real users in experience, habit, and knowledge. As a result, the application system evaluation gives a result different from the real performance. And this result is always better than the real one.

Application operational evaluation is able to represent the real performance of the system. In real using process, the system records the data which are necessary in analysis. The result of the evaluation is given by analysis of the log file. The result of this evaluation matches the users' feeling.

Face recognition system performance does not entirely hinge on the performance of the algorithm. Only an algorithm performance testing is not enough for a face recognition system. The algorithm performance is merely one of the ultimate factors of the system performance. The different face recognition system using a same face recognition algorithm always gives different manifestations. How to increase the system performance without changing the face recognition algorithm is an important problem.

Proposed Questions and Exercises

- What hardware and software modules are needed for a general face recognition system and a video-based FRAD system?
- What are the main issues in FRAD? In what aspects is surveillance video-based FRAD more challenging than cooperative, near-distance face recognition?
- How would you propose solutions for dealing with these challenges?
- How face detection, tracking, and matching could be combined to deal with problems in FRAD?
- How do you expect camera properties and lens would affect the performance?
- How a multiple camera system could be used to deal with problems in FRAD?
- How would a super-resolution algorithm help solving the low-resolution problem?
- Implement a Matlab algorithm for de-interlacing.
- Implement a Matlab algorithm for de-blurring.
- What are criteria for performance evaluation of a FRAD system? Why is it more difficult than that of a face recognition algorithm engine?
- Assuming you are buying a FRAD system for a watch-list FRAD application, propose a protocol to test candidate products how it meets your requirements.

References

1. Z. Lei, R. Chu, R. He, and S. Z. Li. Face recognition by discriminant analysis with gabor tensor representation. In *Proceedings of IAPR International Conference on Biometric*, volume 4642/2007, Seoul, Korea, 2007.
2. S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(Special issu e on Biometrics: Progress and Directions), April 2007.
3. S. Z. Li and Z. Q. Zhang. FloatBoost learning and statistical face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1112–1123, September 2004.
4. S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *Proceedings of IAPR International Conference on Biometric*, volume 4642/2007, Seoul, Korea, 2007.
5. R. Liu, X. Gao, R. Chu, X. Zhu, and S. Z. Li1. Tracking and recognition of multiple faces at distances. In *Proceedings of IAPR International Conference on Biometric*, volume 4642/2007, Seoul, Korea, 2007.
6. S. Mckenna, S. Gong, and Y. Raja. Face recognition in dynamic scenes. In *Proceedings of British Machine Vision Conference*, pages 140–151. BMVA Press, 1997.
7. P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
8. S. Prince, J. Elder, Y. Hou, M. Sizinstev, and E. Olevsky. Towards face recognition at a distance. *Crime and Security, 2006. The Institution of Engineering and Technology Conference on*, pages 570–575, June 2006.
9. H. Wang, S. Z. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. *fg*, 0:819, 2004.
10. L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li. Face detection based on multi-block lbp representation. In *Proceedings of IAPR International Conference on Biometric*, volume 4642/2007, Seoul, Korea, 2007.

11. W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003.
12. S. Zhou, V. Krueger, and R. Chellappa. Face recognition from video: A condensation approach. *fg*, 0:0221, 2002.
13. S. Zhou, V. Krueger, and R. Chellappa. Face recognition from video: a condensation approach. *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 221–226, May 2002.