

# State-of-the-Art on Video-Based Face Recognition

**Yan Yan**

*Tsinghua University, Beijing, China*

**Yu-Jin Zhang**

*Tsinghua University, Beijing, China*

## INTRODUCTION

Over the past few years, face recognition has gained many interests. Face recognition has become a popular area of research in computer vision and pattern recognition. The problem attracts researchers from different disciplines such as image processing, pattern recognition, neural networks, computer vision, and computer graphics (Zhao, Chellappa, Rosenfeld & Phillips, 2003).

**Face recognition** is a typical computer vision problem. The goal of computer vision is to understand the images of scenes, locate and identify objects, determine their structures, spatial arrangements and relationship with other objects (Shah, 2002). The main task of face recognition is to locate and identify the identity of people in the scene. Face recognition is also a challenging **pattern recognition** problem. The number of training samples of each face class is usually so small that it is hard to learn the distribution of each class. In addition, the within-class difference may be sometimes larger than the between-class difference due to variations in illumination, pose, expression, age, etc.

The availability of the feasible technologies brings face recognition many potential applications, such as in face ID, access control, security, surveillance, smart cards, law enforcement, face databases, multimedia management, human computer interaction, etc (Li & Jain, 2005).

Traditional still image-based face recognition has achieved great success in constrained environments. However, once the conditions (including illumination, pose, expression, age) change too much, the performance declines dramatically. The recent FRVT2002 (Face Recognition Vendor Test 2002) (Phillips, Grother, Micheals, Blackburn, Tabassi & Bone 2003) shows that the recognition performance of face images captured in

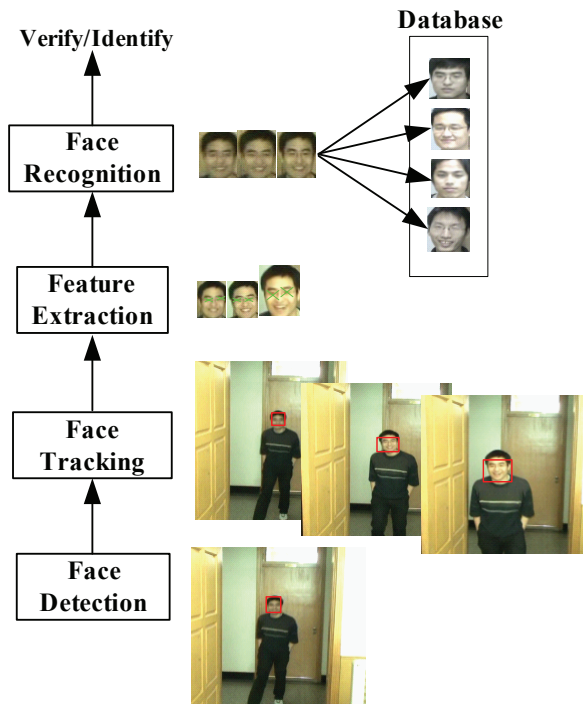
an outdoor environment and different days is still not satisfying. Current still image-based face recognition algorithms are even far away from the capability of human perception system (Zhao, Chellappa, Rosenfeld & Phillips, 2003). On the other hand, psychology and physiology studies have shown that motion can help people for better face recognition (Knight & Johnston, 1997; O'Toole, Roark & Abdi, 2002). Torres (2004) pointed out that traditional still image-based face recognition confronts great challenges and difficulties. There are two potential ways to solve it: video-based face recognition technology and multi-modal identification technology. During the past several years, many research efforts have been concentrated on video-based face recognition. Compared with still image-based face recognition, true **video-based face recognition** algorithms that use both spatial and temporal information started only a few years ago (Zhao, Chellappa, Rosenfeld & Phillips, 2003).

This article gives an overview of most existing methods in the field of video-based face recognition and analyses their respective pros and cons. First, a general statement of face recognition is given. Then, most existing methods for video-based face recognition are briefly reviewed. Some future trends and conclusions are given in the end.

## BACKGROUND

From a general point of view, a complete video-based face recognition system includes face detection module, face tracking module, feature extraction module and face recognition module. Face detection is at the bottom layer. The task of face detection is to determine the spatial position and pose of the face(s). Face tracking is at the middle layer. It follows the continuous change

Figure 1. A general framework of video-based face recognition system



of face position over time. Feature extraction is at a higher layer. Its task is to locate the position of facial features such as eye, nose, etc, and pull out related information. Face recognition module is at the top layer. The face recognition module identifies or verifies the input face(s), with the help of databases. Figure 1 gives the general framework of video-based face recognition system, with a flowchart and some examples.

In this article, the focus will be on the top layer of face recognition systems—face recognition module. The general statement of **face recognition** can be defined as: given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces (Zhao, Chellappa, Rosenfeld & Phillips, 2003). The still image-based face recognition usually refers to the process in which the input is a still image. On the other side, the video-based face recognition often refers to the process in which the input is a shot of video. The database can be also still image(s) or video. Therefore, according to different modalities of the input and database, four different scenarios of face recognition can be distinguished. Table 1 shows these four different scenarios of face recognition. **Video-**

**based face recognition** usually refers to both “Video - Image(s)” face recognition and “Video - Video” face recognition, that is, with video input.

Compared with still image-based face recognition, video-based face recognition can utilize the temporal and spatial information available in the video. It’s widely believed that video-based face recognition is more promising than still image-based face recognition. However, there also exist some difficulties in video-based face recognition, such as low-resolution face images, large variations of scale, illumination change, pose change, and occasionally occlusion in video. It is worth noting that if the time information of video is not considered, the video-based face recognition becomes the multiple-still-images input face recognition.

## VIDEO-BASED FACE RECOGNITION

According to the classification shown in Table 1, four scenarios of face recognition will be reviewed separately. The emphases will be put on “Video - Image(s)” face recognition and “Video - Video” face recognition. For simplicity, the position of the face in the video is assumed to be known in advance.

### “Image - Image(s)” Face Recognition

“Image - Image(s)” **face recognition** is the traditional still image-based face recognition. Numerous still image-based face recognition methods have been developed during the past few decades (Zhao, Chellappa, Rosenfeld & Phillips, 2003). Among them, global feature matching methods, such as Eigenface (Turk & Pentland, 1991), Fisherface (Belhumeur, Hespanha & Kriegman, 1997) and Bayesian (Moghaddam, Jebara & Pentland, 2000); and local feature matching methods, such as Elastic Bunch Graph Matching (EBGM) (Wiskott, Fellous, Krueger & Malsburg, 1997), are the widely used face recognition approaches. Recently, 3D deformable models (Banz & Vetter, 2003) and Local Binary Pattern (LBP) (Ahonen, Hadid & Pietikäinen, 2006) are the newly-emerging methods. Traditional still image-based face recognition has been widely used in biometric authentication, information security, etc.

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/state-art-video-based-face/10430](http://www.igi-global.com/chapter/state-art-video-based-face/10430)

## Related Content

---

### Supporting Empathetic Boundary Spanning in Participatory Workshops with Scenarios and Personas

Anna Salmi, Päivi Pöyry-Lassila and Juha Kronqvist (2012). *International Journal of Ambient Computing and Intelligence* (pp. 21-39).

[www.irma-international.org/article/supporting-empathetic-boundary-spanning-participatory/74368](http://www.irma-international.org/article/supporting-empathetic-boundary-spanning-participatory/74368)

### Adaptive AI in Metaverse Language Education

Rohit Yadav, Xuan-Hoa Nghiem and Mohit Yadav (2025). *Harnessing Chatbots for Effective Language Teaching* (pp. 49-74).

[www.irma-international.org/chapter/adaptive-ai-in-metaverse-language-education/375815](http://www.irma-international.org/chapter/adaptive-ai-in-metaverse-language-education/375815)

### Deontic Logic Based Ontology Alignment Technique for E-Learning

Lazarus Jegatha Deborah, Ramachandran Baskaran and Arputharaj Kannan (2012). *International Journal of Intelligent Information Technologies* (pp. 56-72).

[www.irma-international.org/article/deontic-logic-based-ontology-alignment/69390](http://www.irma-international.org/article/deontic-logic-based-ontology-alignment/69390)

### AI-Assisted MRI: Transforming Multiple Sclerosis Diagnostics, Treatment, and Prognosis

Alexander Zakharov, Alexander Kolsanov, Elena Khivintseva, Olga Pavlova, Vladimir . Kalinin, Nikolay Popov and Nadezhda Evgenievna Karpova (2025). *AI-Powered Systems for Healthcare Diagnostics and Treatment* (pp. 93-122).

[www.irma-international.org/chapter/ai-assisted-mri/376749](http://www.irma-international.org/chapter/ai-assisted-mri/376749)

### Supporting Empathetic Boundary Spanning in Participatory Workshops with Scenarios and Personas

Anna Salmi, Päivi Pöyry-Lassila and Juha Kronqvist (2012). *International Journal of Ambient Computing and Intelligence* (pp. 21-39).

[www.irma-international.org/article/supporting-empathetic-boundary-spanning-participatory/74368](http://www.irma-international.org/article/supporting-empathetic-boundary-spanning-participatory/74368)