Proceedings of the 2004 IEEE
Conference on Cybernetics and Intelligent Systems
Singapore, 1-3 December, 2004

Facial Image Processing: An Overview

Douglas Chai
School of Engineering and Mathematics
Edith Cowan University
100 Joondalup Drive, Joondalup,
WA 6027, Perth, Australia

Kok Wai Wong
School of Computer Engineering
Nanyang Technological University
Nanyang Ave, Singapore 639798,
Singapore

Abstract – **Facial image processing is an area of research that holds an important key to future advances in intelligent human-to-computer and human-to-human systems. This paper presents an overview of this research. It also addresses some of the latest research directions and applications, as well as several important issues raised from recent studies.**

I. INTRODUCTION

Facial image processing is a specialized field of research in computer vision that deals with facial images. It involves the development of computer vision systems that can duplicate human vision in performing tasks such as finding and recognizing a person's face in a picture. Although such visual tasks seem to be effortless for a human to perform, it is very far from easy in terms of computer vision.

The development of such an intelligent system has been widely and actively studied in the field of computer vision and image processing for the past few decades. Moreover, the research activity in this area has intensified significantly in recent time as a result of increasing interests in facial biometrics, multimedia, videoconferencing and human-computer interaction. For example, face recognition surveillance systems have become more widespread and attracted more publicity following the September 11 terrorist attacks in the United States.

This paper presents an overview of the facial image processing research. In Section II, we report on some of the latest research directions and address several important issues raised from recent studies. Section III presents some of the applications of facial image processing, while concluding remarks can be found in Section IV.

II. FACIAL IMAGE PROCESSING

As depicted in Fig. 1, we consider the scope of facial image processing research to encompass three main elements, namely: facial image capture, analysis and compression. Among these, facial image analysis is so far the most actively studied as it is the core element of the facial image processing research. It involves research tasks such as the detection of human faces in images, the extraction of facial regions and

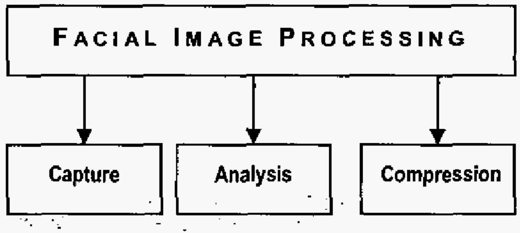

Fig. 1: The scope of facial image processing research.

their facial features, and the recognition of faces. On the other hand, facial image capture and compression attract much less attention as they play the "supporting roles". The former involves the acquisition of facial images in digital form, while the latter is concerned with the storage of facial images in compact form.

Although research into facial image analysis has been pursued at a feverish pace, there are still many problems yet to be fully and convincingly solved. This is because the level of difficulty of the problem depends highly on the complexity level of the image content and the working environment. Many existing methods only work well on simple input images with benign background and a certain view of a person's face. To cope with more complicated conditions, many more assumptions will have to be made, hence limiting their scope of application and their effectiveness.

As for facial image capture and compression, it is fair to say that researchers are currently contented with using generic methods to capture and compress the digital images of human faces, without much concern of their efficiency or any other functionalities. Although such an approach can fulfil their current objective, which is to solely capture and store images, it will most likely be inadequate in the future as the technology for facial image analysis continues to advance. For instance, the generation and storage of facial image data will become an important issue in the future as face recognition systems continue to gain greater popularity and usage, and their image databases continue to expand.

A. Facial Image Analysis

Facial image analysis is basically about *finding* and *recognizing* human faces and their facial features from visual

0-7803-8643-4/04/$20.00 © 2004 IEEE

scenes. Here we classify, as illustrated in Fig. 2, the current research activities of facial image analysis into three broad areas, namely: detection, extraction and recognition. These three are closely related research problems. They may be using similar means but they aim to achieve different end. Their objectives can be summarized as follows:

- The primary objective of face detection is to detect the presence of any human face in a given image. This can be achieved by identifying the visual attributes that distinguish the human faces from all other objects in the scene.
- Face extraction involves the separation of the facial region from the background scene so that the information about the face's location and its spatial extent are extracted.
- Face recognition is about finding the identity of a person by comparing his or her face against faces stored in a database of known identities.

As mentioned earlier, the level of difficulty of these problems depends highly on the complexity level of the image content and the working environment. The common factors to consider include:

- Unknown presence of single or multiple persons in a scene;
- Unknown size and position of the person's face;
- Variation in pose due to tilting and turning of the person's head;
- Variation in facial expression;
- Occlusion (i.e. faces that are partially hidden by other objects);
- Variation in lighting condition as well as level of contrast;
- Level of uniformity, structure and texture of the background scene.

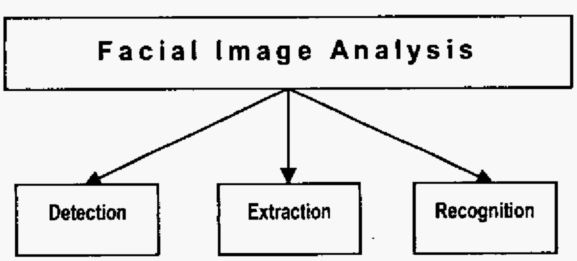

Fig. 2: Detection, extraction and recognition are three broad areas of facial image analysis research.

Current techniques used in facial image analysis can be broadly classified as follows:

- *Knowledge-based approach* – relies on rules that are derived from researcher's knowledge of what constitutes a typical face.
- *Feature-based approach* – involves detection and matching of invariant features of faces, such as the use of skin color information.
- *Template-based approach* – uses predefined or deformable face template to compute correlation between input and stored patterns.
- *Appearance-based approach* – is based on probabilistic framework, such as the eigenface method.

Comprehensive surveys of over 300 past and current research studies of facial image analysis can be found in [1-4].

Note that researchers in facial image processing now have more training data available to them than ever before. Prior to mid 1990's, the limited availability of facial image databases was restricting the study of data-driven learning algorithms, such as neural networks, support vector machines and distribution-based methods. Nowadays, however, researchers have collected over tens of thousands of facial image sets. For instance, Phung *et al.* [5] have accumulated over 12,000 different face patterns for training and testing algorithms. A sample of these patterns is given in Fig. 3.

B. Facial Image Capture

The image of a person's face must first be captured before any analysis can begin. Unlike past research studies that have

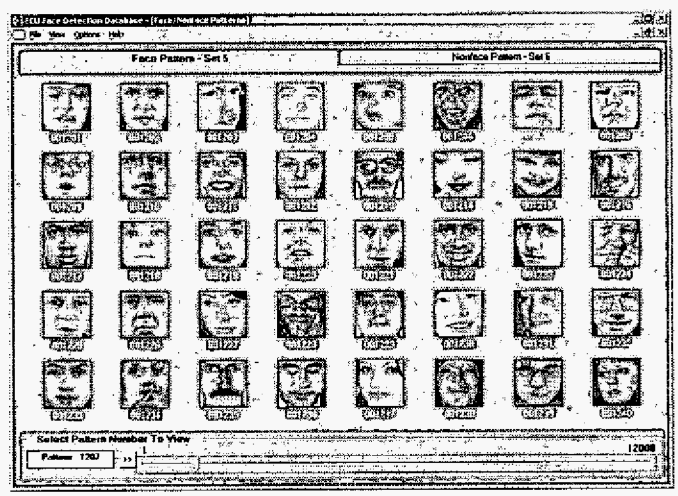

Fig. 3: A sample of face patterns in Phung *et al.*'s database.

focused on images taken in highly constrained environment, most of the current studies have shift their focus towards less controlled and more natural environment whereby images are taken with little or no constraint. One of the important factors is the lighting condition, which can be of different types of illumination (e.g. natural or artificial) as well as intensity and direction.

For instance, Fig. 4 shows two facial images of a same subject captured at different angles to the sun, in YCbCr format. By computing the normalized histograms of the two manually extracted facial regions, we found that although the distribution of Y values was altered, the characteristics of the Cb and Cr distributions remain mostly unchanged. This implies that there is a strong probability that face models derived from Cb and Cr values are robust against this particular type of illuminant changes.

In addition to coping with various lighting condition, an advanced facial image capturing system would have new functionalities such as face tracking, automatic mug-shot production, liveliness evaluation. Face tracking involves the task of following faces over time. For instance, a camera system that can always follow and centre on the subject will be a useful feature for video surveillance. In addition, the ability to capture facial images in certain head size and orientation will allow the automatic production of mug-shot. Finally, liveliness evaluation is about the capability of an image capturing system to distinguish between real life face and photo of a face.

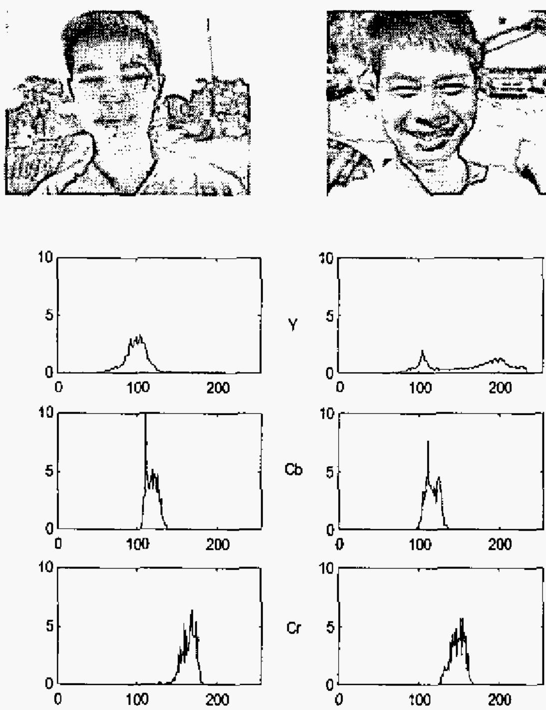

Fig. 4: Subject facing at different angles to the sun (above), and the respective normalized histograms of Y, Cb and Cr components of the facial regions (below).

C. Facial Image Compression

The study of facial image compression specializes in the image coding of facial image data. It usually involves the development of more efficient ways to compress facial image data for use in large-scale face recognition databases, multimedia content storage, and videoconferencing systems. Ideally, the objective is to achieve high compression ratio and still keep facial features at a recognisable quality, not only for human viewers but also for face recognition algorithm.

For example, using the region-of-interest (ROI) coding approach, important facial features can be treated with higher priority and quality than the non-essential or less-relevant background region. Fig. 5 depicts an image that has spatial variable quality. The facial region was encoded at a higher quality than the non-facial region by applying different encoding parameters to these two image regions.

The JPEG coding scheme, which uses discrete cosine transform (DCT), is a common format for storing images. However, its performance is not the best. The discrete wavelet transform (DWT) coding scheme coupled with bit-plane coding is more superior. A comparison study between these two schemes was carried out. The result, as presented in Fig. 6, shows that DWT significantly outperforms DCT at every tested bit rate, especially at the lower end of the bit rate spectrum. A sample output quality of these two coders is given in Fig. 7. The JPEG encoded image is badly degrades by blocking effects and it has lost all color information. In

Fig. 5: A ROI encoded image whereby the facial region is clearer and sharper than its background scene.

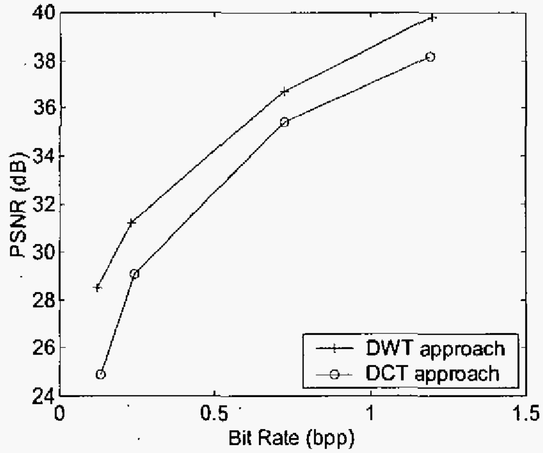

Fig. 6: Rate-distortion curves for DWT and DCT approaches.

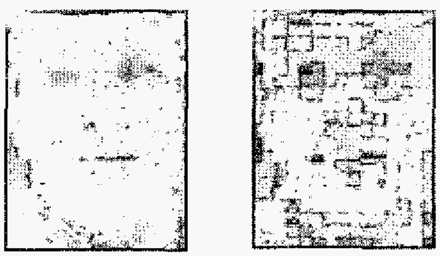

Fig. 7: Images encoded at 0.12 bits per pixel (bpp) by DWT-based coder (left) and JPEG coder (right).

contrast, the image produced by the DWT coder has a much better recognizable quality.

III. APPLICATIONS

Facial image processing technology can be applied to a myriad of intelligent systems that deal with image content of human faces. This includes areas such as human-computer interface study, object recognition and tracking, as well as image coding, retrieval, manipulation, enhancement, and modelling. Some specific examples of intelligent systems are discussed below:

- **Face recognition system.** Face recognition is a well-known biometric technology. It analyses the biological characteristics of facial structure taken through camera, and then crosschecks the data with those profiles stored on database. To *identify* criminals one-to-many matching is performed, whereas to *verify* the identity of a claimed person one-to-one matching is carried out; these are known as **face identification** and **face verification**, respectively. Such technology can be used in crowd surveillance, computer/building access control, or other security checkpoints.

- **Multimedia content.** Both MPEG-4 and MPEG-7 standards [6-8] were developed to address the recent growing activities in the digitization and integration of many media such as broadcasting, publishing, movies and communications into the so-called multimedia environment. Although its predecessors MPEG-1 and MPEG-2 are successful for their respective uses in VCD and DVD movies, the new MPEG-4 standard is set to revolutionize multimedia communications. MPEG-4 adopts an object-oriented approach whereby a video is composed of various audio-visual objects. Since human faces occur very frequently in video content, face detection becomes an important tool for generating the content-based image representation needed by an MPEG-4 system. In doing so, the extracted facial region can be encoded and decoded independently of other objects in the scene such as table, curtains, etc. Facial image processing technology is also useful for MPEG-7, which is a standard for description and search of audio and visual content. The ability to detect, recognise and track faces will help to structure the ever-growing multimedia databases for indexing and search.

- **Intelligent human-computer interaction.** Head tracking and gaze estimation techniques can be employed to design an intelligent human-computer interaction whereby computer users can navigate in 3D graphics scene and change camera viewpoint via head movement. In this computer vision approach, users do not need to wear any physical sensors such as goggles and helmets, and the control should be more intuitive and natural to users than keyboard and mouse interfaces. Importantly, this technology can help create alternative devices for the disabled.

- **Smart audio-visual recording system.** Face detection and tracking algorithms can be incorporated into an audio-visual recording system so that the speaker, say on a stage, can be located and tracked at all times. The location of the speaker will then steer the camera and the microphone array. In doing so, the speaker can move freely on stage while still being continuously framed by the camera. Furthermore, by tracking the speaker, the beam steering microphone array is able to pick up the speech produced by the moving speaker efficiently, as it can reduce competing acoustic signals from other sources.

- **Low bit-rate videoconferencing system.** During videoconferencing, viewers tend to direct their attention on the face of the opposite speaker. Thus the facial region is typically the most important image area. The image quality of this region of interest has a big impact on the overall perceptual quality of the videoconferencing images. Viewers' perception of quality is very subjective. Often, viewers perceive an improvement in the overall image quality when only the region of interest has improved. On the other hand, when coding artefacts occurred in the region of interest, they are more noticeable and annoying, and thus they degrade the overall perceptual quality. With this observation in mind, face detection can be integrated into the video coder so that the facial regions are treated with high priority and coded at higher quality [9]. Videoconferencing systems that operate in the low-bit rate environment, such as Internet or wireless videotelephony, will benefit the most from this development.

IV. Concluding Remark

Facial image processing field of research specializes in the processing of facial images. The current research directions are mainly focused on investigating advanced techniques for detecting, extracting and recognizing faces. However, as this technology matures, some of the focus will shift towards finding superior approaches to image capturing of human faces for subsequent analysis and/or storage, as well as more efficient ways to compress facial image data. The knowledge gained from this research will have applications in some of the world's fastest growing markets such as building/information security, videoconferencing and multimedia systems.

Acknowledgment

This work is supported by a grant from the Australian Research Council.

References

[1] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, vol. 25, no. 6, pp. 65-77, 1992.

[2] R. Chellappa, C. L. Wilson and S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705-740, 1995.

[3] E. Hjelmas and B. K. Low, "Face detection: a survey," *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236-274, Sep. 2001.

[4] M.-H. Yang, D. Kriegman and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, Jan. 2002.

[5] S. L. Phung, A. Bouzerdoum and D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press.

[6] ISO/IEC 14496-2, "MPEG-4 – Information Technology – Coding of audio-visual objects, Part 2: Visual," *ISO International Standard*, 2001.

[7] J. M. Martinez, "Overview of the MPEG-7 standard," *ISO/IEC JTC1/SC29/WG11 N4509*, Pattaya, Dec. 2001.

[8] R. Koenen, "From MPEG-1 too MPEG-21: creating an interoperable multimedia infrastructure," *ISO/IEC JTC1/SC29/WG11 N4518*, Pattaya, Dec. 2001.

[9] D. Chai and K. N. Ngan, "Coding area of interest with better quality," *IEEE International Workshop on Intelligent Signal Processing and Communication Systems*, Kuala Lumpur, Malaysia, pp. S20.3.1-S20.3.10, Nov. 1997.