

A 3D Face Recognition System Using Passive Stereo Vision and Its Performance Evaluation

Akihiro Hayasaka*, Takuma Shibahara*, Koichi Ito*, Takafumi Aoki*, Hiroshi Nakajima†, and Koji Kobayashi†

* Graduate School of Information Sciences, Tohoku University

6-6-05, Aramaki Aza Aoba, Sendai-shi 980-8579, Japan

Tel: +81-22-795-7169, Fax: +81-22-263-9308

E-mail: hayasaka@aoki.ecei.tohoku.ac.jp

† Yamatake Corporation, 54, Suzukawa, Isehara-shi 259-1195, Japan

Abstract—This paper proposes a three-dimensional (3D) face recognition system using passive stereo vision. So far, the reported 3D face recognition systems assume the use of active 3D measurement for 3D facial capture. However, active methods employ structured illumination or laser scanning, which is not desirable in many human recognition applications. Addressing this problem, we propose a 3D face recognition system that uses (i) AdaBoost-based face detection to automatically extract a face region from an image, (ii) passive stereo vision to capture 3D facial information, and (iii) 3D face matching based on a simple ICP (Iterative Closest Point) algorithm. Experimental evaluation demonstrates an efficient recognition performance of the proposed system.

I. INTRODUCTION

Biometric authentication has been receiving extensive attention over the past decade with increasing demands in automated personal identification. Among all the biometric techniques, face recognition has been an area of intense research [1].

Most of the reported approaches to automatic human face recognition use two-dimensional (2D) images [1]. However, face recognition techniques from 2D images are affected strongly by illumination and face varieties. The robust feature detection in 2D face images is still an open difficulty. Recently, the use of three-dimensional (3D) information has gained much attention, since 3D data is not affected by translation and rotation and is immune to the effect of illumination variation [1]. The 3D face recognition acquires facial 3D information (facial structure) obtained from the 3D scanner and then identifies a person by calculating the similarity between facial structures.

So far, the reported 3D face recognition systems usually use active 3D scanners which employ structured illumination (structure projection, phase shift, gray-code demodulation, etc.) or laser scanning to capture high-quality 3D facial structure. However, the use of active 3D scanners is not necessarily desirable in many cases of human recognition applications. Addressing this problem, this paper proposes a 3D face recognition system that uses (i) AdaBoost-based face detection [2] to automatically extract face region from an image, (ii) passive stereo vision to capture 3D facial information, and (iii) 3D face matching based on a simple ICP (Iterative Closest Point) algorithm.

A major problem of using “passive” stereo vision system for facial 3D measurement is its low quality and low accuracy of captured 3D information. Thus, no practical approaches to passive 3D face recognition have been reported to the best of the authors’ knowledge. Addressing this problem, we have developed a high-quality 3D facial capture system based on passive stereo vision. The key feature of this system is to employ a phase-based image matching technique [4], [6] for robust stereo correspondence and sub-pixel disparity estimation. The developed passive 3D capture system can reconstruct 3D facial information with ~ 0.6 mm accuracy at 50cm distance. In order to realize a fully-automatic 3D reconstruction from only a human face region, we employ AdaBoost-based face detection. With the high-quality 3D facial data, we show that practical face matching for biometric authentication could be performed based on a simple 3D registration scheme, called ICP algorithm. Experimental evaluation demonstrates an efficient performance of the proposed system.

II. FACE DETECTION

This section describes a face detection process which automatically extracts a human face region from a captured image.

The performance of the face recognition system is usually influenced by the reliability of face detection. Face detection can be performed based on several cues: skin color (for faces in color images and videos), motion (for faces in videos), facial/head shape, facial appearance, or a combination of these cues. The most successful face detection algorithm is based on the facial appearance. The procedure of appearance-based face detection is that an input image is scanned at all the possible locations and scaled by a subwindow, and then face detection is done by classifying the pattern in the subwindow as either face or nonface. The face/nonface classifier is learned from face and nonface training examples using statistical learning method. We use AdaBoost-based face detection method [2] since AdaBoost is the most successful learning method in terms of detection accuracy and processing speed. AdaBoost is one of the boosting method. The boosting algorithm generates different weak classifiers by changing example weights and then construct a strong classifier by combining weak classifiers.

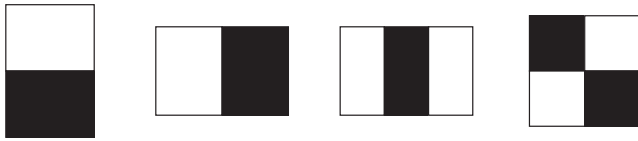


Fig. 1. Four types of binary rectangular filters used in this paper.



Fig. 2. Examples of selected binary rectangular filters.

AdaBoost-based face detection uses Haar-like features obtained by applying four kinds of binary rectangular windows (Fig. 1) to the image. Haar-like features are extracted from a square subwindow of an arbitrary position and size in the input image, and used for determining whether the subwindow includes a face or not. Since the binary rectangular filters are applied by changing aspect ratio, size and location inside the subwindow of size 24×24 , the total number of filters are tens of thousands. An effective rectangular filter for face detection is only a small part of them. Hence, an AdaBoost algorithm is used to select effective filters. Examples of selected binary rectangular filters are shown in Fig. 2. We generate the classifiers (*weak classifiers*) based on Haar-like features obtained from selected filters, where the weak classifier determines whether the subwindow includes a face or not. In the learning process of weak classifiers, we use a number of face and nonface images to perform effective learning. The final classifier (*strong classifier*) is generated by combining weak classifiers with weights of their reliability. A single strong classifier usually detects many nonface regions. In order to detect face regions correctly, the cascade connection of strong classifiers is used, where multiple strong classifiers are generated by changing face/nonface examples. Figure 3 shows an example of the cascade connection of k -strong classifiers. The performance of face detection is increased since the subwindow is recognized as a face only when passing through all the strong classifier as a face. This strategy can significantly reduce the computation time of face detection since most of subwindows are reject by the first strong classifier.

A face in an image may be detected several times at neighbor locations or on multiple scales as shown in Fig. 4 (b). Nonface regions may also be detected even if the cascade connection of strong classifiers is used. Addressing these problems, we use the fact that the correct regions are concentrated around the face region while the false regions are not. After detecting face regions, we select the correct regions and combine them to detect a correct face region. Figure 4 illustrates an example of face detection using the AdaBoost algorithm. The face region is automatically detected from the

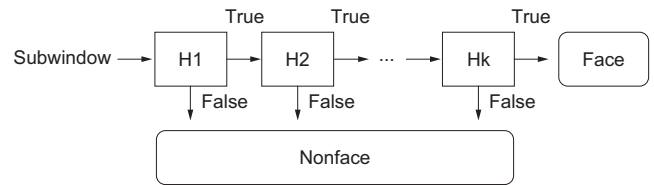


Fig. 3. A cascade of k -strong classifiers.

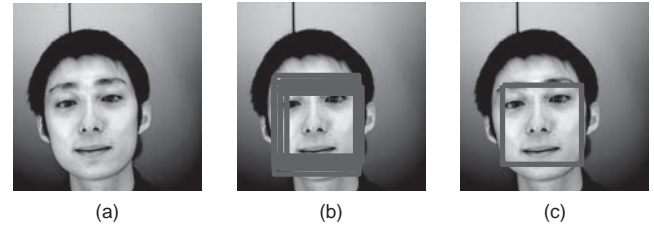


Fig. 4. Example of face detection: (a) original image, (b) detected face regions, and (c) combined face region.

image by using the above algorithm.

III. 3D FACE RECOGNITION SYSTEM

This section describes the 3D face recognition system proposed in this paper. Figure 5 shows the developed high-quality passive 3D measurement system to capture the 3D facial data. The system has one stereo camera head which of a pair of two parallel cameras. An important feature of the stereo camera head is that its baseline is designed as narrow as possible; the baseline is 46mm limited simply by the size of the camera chassis. The narrow-baseline camera configuration makes possible to find stereo correspondence automatically for every pixel, but a serious drawback is its low accuracy in the reconstructed 3D facial data when compared with wide-baseline configuration. Addressing this problem, we have developed a high-accuracy stereo correspondence technique using phase-based image matching. The proposed technique employs the Phase-Only Correlation (POC) function (which is sometimes called the “phase correlation function”) for sub-pixel image matching required in high-accuracy stereo correspondence (see our papers [4], [5] for detailed discussion on this technique). The stereo correspondence problem requires high-accuracy matching of smaller image blocks, such as 32×32 pixel blocks. However, the accuracy and robustness of phase-based image matching described above degrades significantly as the image size decreases. Addressing this problem, we have proposed some techniques to improve the registration accuracy for small image blocks [6]. Using the techniques, we have developed an efficient method of sub-pixel correspondence matching for our 3D facial capture system, which employs (i) a coarse-to-fine strategy using image pyramids for robust correspondence search and (ii) a sub-pixel window alignment technique for finding a pair of corresponding points with sub-pixel displacement accuracy.

The procedure of face recognition is summarized as follows. At first, we capture the stereo image by using the stereo vision

