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 $\sim$0.6mm 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.

0-7803-9733-9/06/$20.00

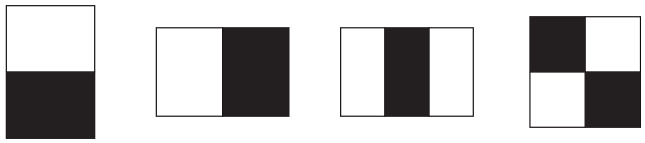

Fig. 1. Four types of binary rectangular filters used in this paper.

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 image by using the above algorithm.

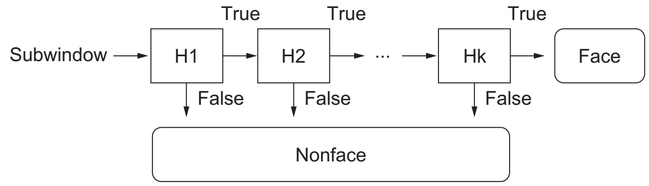

Fig. 3. A cascade of k-strong classifiers.

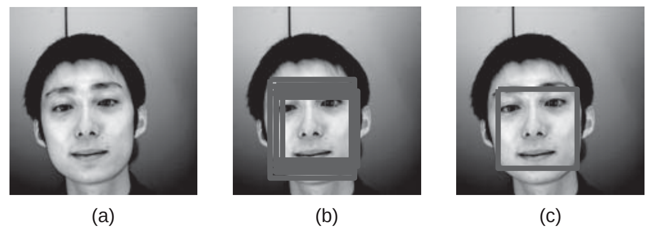

Fig. 4. Example of face detection: (a) original image, (b) detected face regions, and (c) combined face region.

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

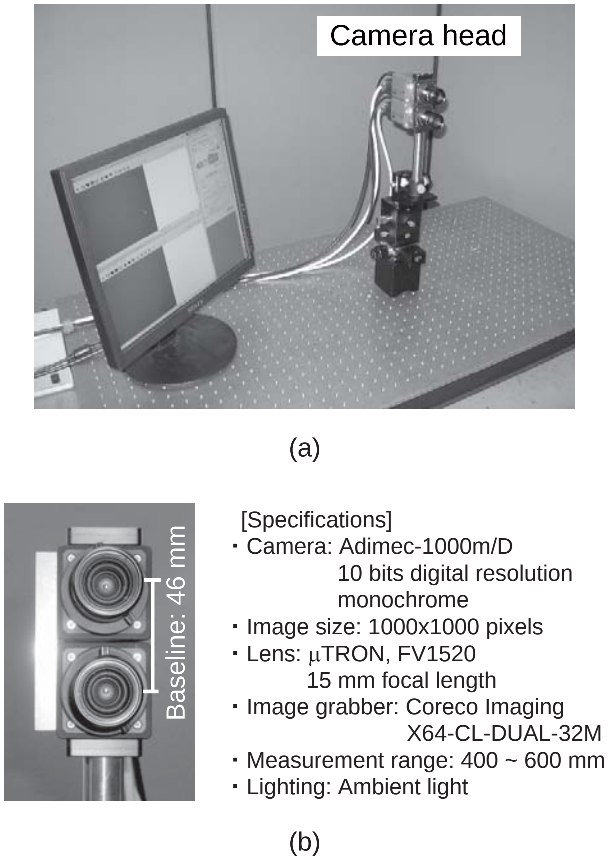

Fig. 5. Passive 3D facial capture system: (a) system configuration and (b) close-up view of the camera head and system specification.

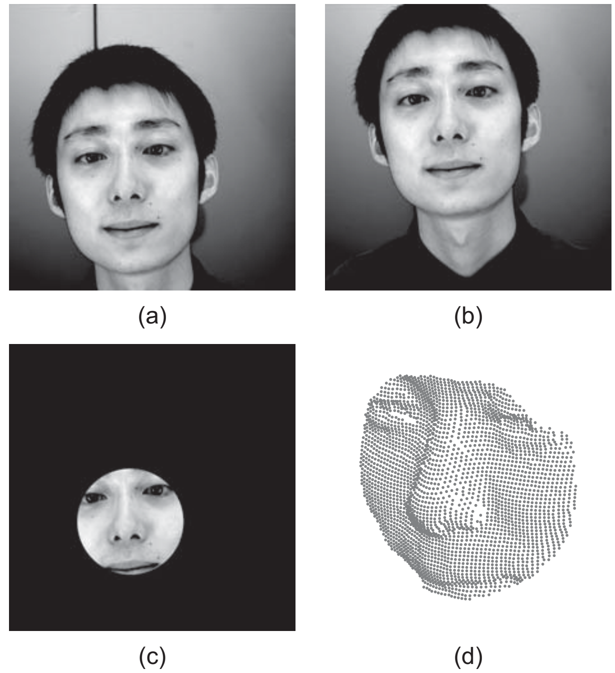

Fig. 6. Example of capturing 3D facial data by using the proposed system: (a), (b) stereo images, (c) detected face region, and (d) reconstructed 3D facial data.

system. Next, the face area of the image taken from the lower camera is automatically detected by using the AdaBoost-based face detection algorithm, and 3D facial data of its facial region is reconstructed as shown in Fig. 6. Finally, given a pair of 3D face data, face matching can be performed by using ICP algorithm.

The ICP algorithm consists of the two steps: (i) align the 3D face surfaces with each other (Figure 7) and (ii) evaluate their similarity based on some distance measure. As for the 3D face alignment in (i), we have decided to use a simple ICP (Iterative Closest Point) algorithm, since the quality of 3D information captured by our stereo vision system is sufficiently high [3].

Let M be the set of 3D points of a face, and M' be the set of 3D points of another face. The ICP algorithm is summarized as follows:

1) For every point m_i in M, find the closest point m'_i from M' as a corresponding point.
2) Based on the current correspondence, calculate the optimal transformation (i.e., rotation $\mathbf{R}$ and translation $\mathbf{t}$) between the two data sets M and M' using the least-square method.
3) Transform the points in M' with $\mathbf{R}$ and $\mathbf{t}$.
4) Repeat from step 1 to step 3 until convergence.

To accelerate the computation, we adopt the coarse-to-fine strategy in the above ICP procedure, where the initial alignment starts with fewer corresponding points (1/32 of the total points) and the number of corresponding points gradually increases as the iteration step increases.

Dissimilarity between the two 3D facial data M and M' (normalized by ICP) is evaluated by a simple point-to-plane distance. For every point m_i in M, we first find the three points in M' that are closest to m_i. Then, we evaluate the distance d_i between the point m_i and the triangular patch formed by the three points. If the orthogonal projection of m_i onto the plane of the three points is not inside the triangular patch, we omit the point m_i for distance calculation. The distance between the two facial data is defined as an average of individual point-to-plane distances d_i.

IV. Experiment and Discussion

In this section, we describe the performance evaluation of the proposed 3D face recognition system. The performance of the proposed 3D face recognition system was evaluated through an experimental matching of 22 subjects. In this experiment, 5 independent snapshots with neutral expression are captured at different sessions for each subject, resulting in a total 110 ($= 22 \times 5$) facial data. The recognition testing was done using a set of 5995 pairs of facial data, including 220 ($= 22 \times {}_5C_2$) genuine attempts and 5775 ($= {}_{110}C_2 - 220$) impostor attempts. In the face detection step, 200 face images and 1916 nonface images were used to train the classifiers for face detection (Fig. 8).

Figure 9 shows the distribution of distances for genuine attempts and impostor attempts. The distribution shows a

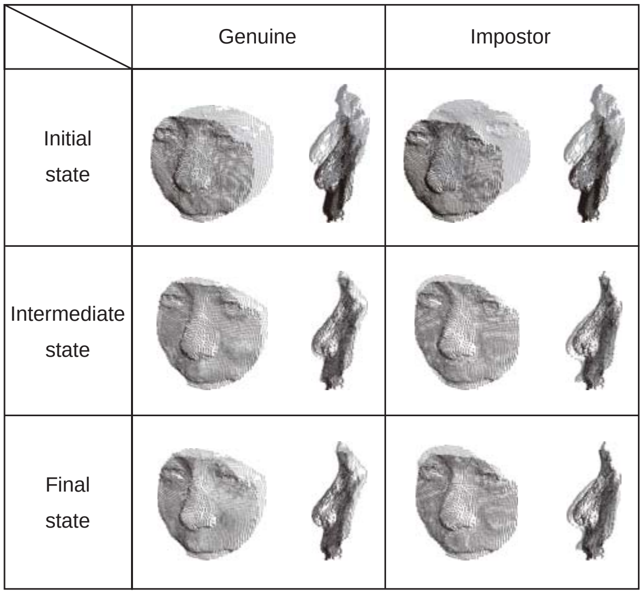

Fig. 7. 3D alignment by ICP algorithm.

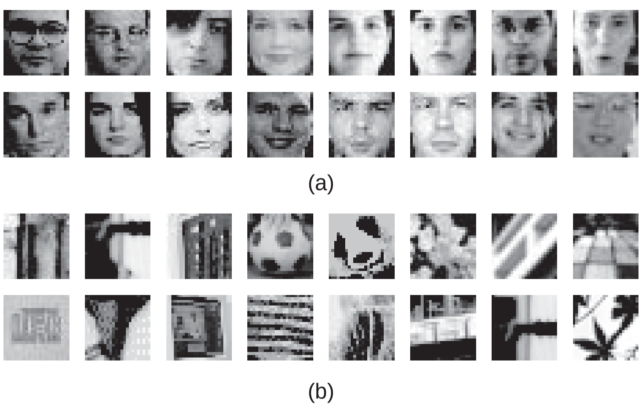

Fig. 8. Examples of face/nonface images using the training process of classifiers for face detection: (a) example of face images, and (b) example of nonface images.

good separation of genuine-matching and impostor-matching distances. Table I summarizes the average, the maximum and the minimum values of distances for genuine and impostor attempts. A distance value within 0.63 $\sim$ 0.64mm can be chosen as a separation point, so that if any two facial data generate a distance value greater than the separation point, they are deemed to be captured from different individuals. If two facial data generate a distance value less than the separation point then the two are deemed to be from the same individual. Thus, it was confirmed that the proposed 3D face recognition system is able to achieve an extremely high accuracy authentication by automatic operation completely. As is observed the experimental results, the recognition performance of the proposed system equals to that of the original system which the face area was specified manually [3].

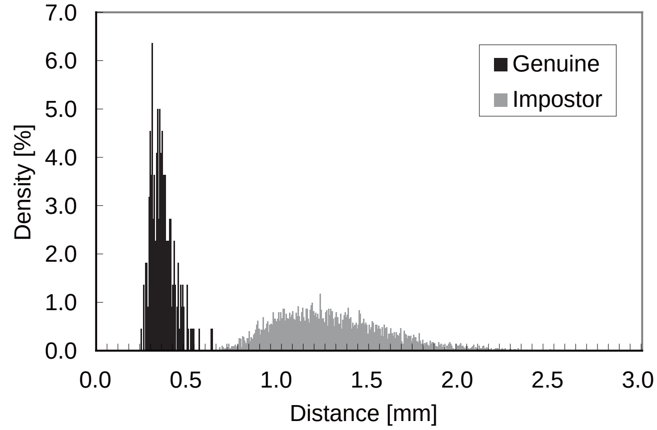

Fig. 9. Distribution of distances.

TABLE I
AVERAGE, MAXIMUM AND MINIMUM VALUES OF DISTANCES FOR GENUINE AND IMPOSTOR ATTEMPTS (THE VALUES IN PARENTHESIS ARE RESULTS OF THE ORIGINAL SYSTEM [3]).

	Average [mm]	Maximum [mm]	Minimum [mm]
Genuine	0.36 (0.41)	0.63 (0.73)	0.24 (0.23)
Impostor	1.30 (1.77)	2.76 (3.65)	0.64 (0.87)

V. CONCLUSION

In this paper, we have proposed a 3D face recognition system that use (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. Through experimental evaluations, it was shown that the proposed system can achieve automatic and high-accuracy authentication.

REFERENCES

[1] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*. Springer, 2005.

[2] P. Viola and M. Jones, "Robust real time object detection," *Proc. of 2nd Int. Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing and Sampling*, pp. 1–25, July 2001.

[3] N. Uchida, T. Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi, "3D face recognition using passive stereo vision," *Proc. of IEEE Int. Conf. on Image Processing 2005*, pp. II–950–II–953, Sept. 2005.

[4] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE Trans. on Fundamentals*, vol. E86-A, no. 8, pp. 1925–1934, Aug. 2003.

[5] M. A. Muquit, T. Shibahara, and T. Aoki, "A high-accuracy passive 3D measurement system using phase-based image matching," *IEICE Trans. on Fundamentals*, vol. E89-A, no. 3, Mar. 2006.

[6] K. Takita, M. A. Muquit, T. Aoki, and T. Higuchi, "A sub-pixel correspondence search technique for computer vision applications," *IEICE Trans. on Fundamentals*, vol. E87-A, no. 8, pp. 1913–1923, Aug. 2004.