IEICE TRANS. FUNDAMENTALS, VOL.E91–A, NO.8 AUGUST 2008

PAPER Special Section on Signal Processing

A Passive 3D Face Recognition System and Its Performance Evaluation

Akihiro HAYASAKA[†a)], ***Nonmember***, **Takuma SHIBAHARA**[†], ***Student Member***, **Koichi ITO**[†b)], **Takafumi AOKI**[†], ***Members***, **Hiroshi NAKAJIMA**[††], ***Nonmember***, ***and*** **Koji KOBAYASHI**[††], ***Member***

SUMMARY This paper proposes a three-dimensional (3D) face recognition system using passive stereo vision. So far, the reported 3D face recognition techniques have used active 3D measurement methods to capture high-quality 3D facial information. However, active methods employ structured illumination (structure projection, phase shift, moire topography, etc.) or laser scanning, which is not desirable in many human recognition applications. Addressing this problem, we propose a face recognition system that uses (i) passive stereo vision to capture 3D facial information and (ii) 3D matching using an ICP (Iterative Closest Point) algorithm with its improvement techniques. Experimental evaluation demonstrates efficient recognition performance of the proposed system compared with an active 3D face recognition system and a passive 3D face recognition system employing the original ICP algorithm.

key words: *biometrics, face recognition, phase-only correlation, stereo vision, 3D measurement*

1. 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. However, face recognition techniques using 2D images are affected strongly by variations in pose and illumination. 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 [2], [3], since 3D data is not affected by translation and rotation and is immune to the effect of illumination variation. The 3D face recognition method acquires 3D facial 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 have used active 3D scanners 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, since active 3D measurement employs structured illumination (structure projection, phase shift, gray-code demodulation, etc.) or laser scanning. Addressing this problem, this paper proposes a 3D face recognition system that uses (i) passive stereo vision to capture 3D facial information, and (ii) 3D face matching based on an ICP (Iterative Closest Point) algorithm [4].

A major problem of using "passive" stereo vision system for 3D facial 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-Only Correlation (POC) technique [5]–[7] 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 50 cm distance. With the high-quality 3D facial data, we show that practical face matching for biometric authentication could be performed based on an ICP algorithm. Experimental evaluation demonstrates efficient performance of the proposed 3D face recognition system which is comparable with that of the active 3D face recognition system. This paper also proposes the improvement techniques for the ICP algorithm to reduce the computation time and improve the registration accuracy. Through a set of experiments, we show that the improved ICP algorithm exhibits better performance than the original ICP algorithm.

This paper is organized as follows: Sect. 2 presents an overview of the proposed 3D face recognition system. Section 3 describes a stereo correspondence technique using POC. Section 4 describes a 3D face matching algorithm. Section 5 presents a set of experiments for evaluating verification performance of the proposed system. In Sect. 6, we end with some conclusions.

2. High-Quality 3D Measurement System with Passive Stereo Vision

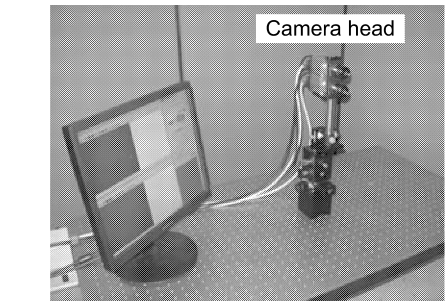
Figure 1 shows the developed high-quality passive 3D measurement system to capture the 3D facial data. The system has one stereo camera head which consists of a pair of two parallel cameras.

An important feature of the stereo camera head is that its baseline is designed as narrow as possible; the baseline

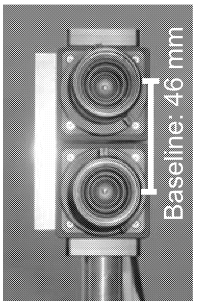
Manuscript received December 10, 2007.
Manuscript revised March 14, 2008.

[†]The authors are with the Department of Computer and Mathematical Sciences, Graduate School of Information Sciences, Tohoku University, Sendai-shi, 980-8579 Japan.
[††]The authors are with Yamatake Corporation, Fujisawa-shi, 251-8522 Japan.
a) E-mail: hayasaka@aoki.ecei.tohoku.ac.jp
b) E-mail: ito@aoki.ecei.tohoku.ac.jp

DOI: 10.1093/ietfec/e91–a.8.1974



(a)



[Specifications]
- Camera: Adimec-1000m/D
  10 bits digital resolution
  monochrome
- Image size: 1000x1000 pixels
- Lens: μTRON, FV1520
  15 mm focal length
- Image grabber: Coreco Imaging
  X64-CL-DUAL-32M
- Measurement range: 400 ~ 600 mm
- Lighting: Ambient light

(b)

Fig. 1 Passive 3D facial capture system: (a) system configuration and (b) close-up view of the camera head and system specification.

is 46 mm 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 next section describes our proposed technique.

3. High-Accuracy Stereo Correspondence Using Phase-Only Correlation

This paper 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. In general, the image matching methods using phase information in 2D Discrete Fourier Transform (2D DFT) exhibit better registration performance than the methods using SAD (Sum of Absolute Differences). The POC technique has been historically used to align and recognize images. The pioneering work for POC has been reported by Kuglin [8]. Then, the POC technique has been applied to estimate a sub-pixel translational displacement [5], [9]–[11] and similarity transformation parameters (translation, rotation and scaling) [5], [12], to recognize images [13]–[15], and to find correspondences between images [6], [7]. In the followings, we describe the high-accuracy image matching based on POC and its application to stereo correspondence problem.

3.1 Sub-Pixel Image Matching Technique

Consider two $N_1 \times N_2$ images, $f(n_1, n_2)$ and $g(n_1, n_2)$, where we assume that the index ranges are $n_1 = -M_1, \cdots, M_1$ ($M_1 > 0$) and $n_2 = -M_2, \cdots, M_2$ ($M_2 > 0$), and hence $N_1 = 2M_1 + 1$ and $N_2 = 2M_2 + 1$. Note that we assume here the sign symmetric index ranges $\{-M_1, \cdots, M_1\}$ and $\{-M_2, \cdots, M_2\}$ for mathematical simplicity. The discussion could be easily generalized to non-negative index ranges with power-of-two image size. Let $F(k_1, k_2)$ and $G(k_1, k_2)$ denote the 2D DFTs of the two images, which are given by

$$\begin{aligned} F(k_1, k_2) &= \sum_{n_1, n_2} f(n_1, n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2} \\ &= A_F(k_1, k_2) e^{j\theta_F(k_1, k_2)}, \end{aligned} \tag{1}$$

$$\begin{aligned} G(k_1, k_2) &= \sum_{n_1, n_2} g(n_1, n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2} \\ &= A_G(k_1, k_2) e^{j\theta_G(k_1, k_2)}, \end{aligned} \tag{2}$$

where $k_1 = -M_1, \cdots, M_1$, $k_2 = -M_2, \cdots, M_2$, $W_{N_1} = e^{-j\frac{2\pi}{N_1}}$, $W_{N_2} = e^{-j\frac{2\pi}{N_2}}$, and $\sum_{n_1, n_2}$ denotes $\sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2}$. $A_F(k_1, k_2)$ and $A_G(k_1, k_2)$ are amplitude components, and $\theta_F(k_1, k_2)$ and $\theta_G(k_1, k_2)$ are phase components. The cross-phase spectrum $R(k_1, k_2)$ is given by

$$R(k_1, k_2) = \frac{F(k_1, k_2)\overline{G(k_1, k_2)}}{|F(k_1, k_2)\overline{G(k_1, k_2)}|} = e^{j\theta(k_1, k_2)}, \tag{3}$$

where $\overline{G(k_1, k_2)}$ is the complex conjugate of $G(k_1, k_2)$ and $\theta(k_1, k_2)$ denotes the phase difference $\theta_F(k_1, k_2) - \theta_G(k_1, k_2)$. The POC function $r(n_1, n_2)$ is the 2D Inverse DFT (2D IDFT) of $R(k_1, k_2)$ and is given by

$$r(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1, k_2} R(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2}, \tag{4}$$

where $\sum_{k_1, k_2}$ denotes $\sum_{k_1=-M_1}^{M_1} \sum_{k_2=-M_2}^{M_2}$. When two images are similar, their POC function gives a distinct sharp peak. When two images are not similar, the peak drops significantly. The height of the peak gives a good similarity measure for image matching, and the location of the peak shows the translational displacement between the images.

In the following, we derive the analytical peak model for the POC function between common images that are minutely displaced with each other. Now consider $f_c(x_1, x_2)$ as a 2D image defined in continuous space with real-number indices x_1 and x_2. Let δ_1 and δ_2 represent sub-pixel displacements of $f_c(x_1, x_2)$ to x_1 and x_2 directions, respectively. So, the displaced image can be represented as $f_c(x_1 - \delta_1, x_2 - \delta_2)$. Assume that $f(n_1, n_2)$ and $g(n_1, n_2)$ are spatially sampled images of $f_c(x_1, x_2)$ and $f_c(x_1 - \delta_1, x_2 - \delta_2)$, and are defined as

$$\begin{aligned} f(n_1, n_2) &= f_c(x_1, x_2)|_{x_1 = n_1 T_1, x_2 = n_2 T_2}, \\ G(n_1, n_2) &= f_c(x_1 - \delta_1, x_2 - \delta_2)|_{x_1 = n_1 T_1, x_2 = n_2 T_2}, \end{aligned}$$

where T_1 and T_2 are the spatial sampling intervals, and

index ranges are given by $n_1 = -M_1, \cdots, M_1$ and $n_2 = -M_2, \cdots, M_2$. For simplicity, we assume $T_1 = T_2 = 1$. The cross-phase spectrum $R(k_1, k_2)$ and the POC function $r(n_1, n_2)$ between $f(n_1, n_2)$ and $g(n_1, n_2)$ will be given by

$$R(k_1, k_2) \simeq e^{j\frac{2\pi}{N_1}k_1\delta_1} e^{j\frac{2\pi}{N_2}k_2\delta_2}, \tag{5}$$

$$r(n_1, n_2) \simeq \frac{\alpha}{N_1 N_2} \frac{\sin\{\pi(n_1 + \delta_1)\}}{\sin\{\frac{\pi}{N_1}(n_1 + \delta_1)\}} \frac{\sin\{\pi(n_2 + \delta_2)\}}{\sin\{\frac{\pi}{N_2}(n_2 + \delta_2)\}}, \tag{6}$$

where $\alpha = 1$. The above equation represents the shape of the peak for the POC function between common images that are minutely displaced with each other. This equation gives a distinct sharp peak. (When $\delta_1 = \delta_2 = 0$, the POC function becomes the Kronecker delta function.) The peak position (δ_1, δ_2) of the POC function corresponds to the displacement between the two images. We can prove that the peak value α decreases (without changing the function shape itself), when small noise components are added to the original images. Hence, we assume $\alpha \leq 1$ in practice. For image matching task, we estimate the similarity between two images by the peak value α, and estimate the image displacement by the peak position (δ_1, δ_2).

In this paper, we also employ important techniques for high-accuracy sub-pixel image matching such as (i) function fitting for high-accuracy estimation of peak position, (ii) windowing to reduce boundary effects and (iii) spectral weighting technique to reduce aliasing and noise effects. See Refs. [6], [7] for detailed discussions of these techniques.

3.2 Correspondence Search with Sub-Pixel Accuracy

This section discusses a high-accuracy stereo correspondence algorithm based on the sub-pixel image matching mentioned above. Our algorithm employs (i) a coarse-to-fine strategy using image pyramids for robust correspondence search with POC-based block matching, and (ii) a sub-pixel window alignment technique for finding a pair of corresponding points with sub-pixel displacement accuracy. In the first stage, we estimate the stereo correspondence with pixel-accuracy using hierarchical POC-based block matching with coarse-to-fine strategy. The estimation error becomes less than 1 pixel for every corresponding point. The second stage is to recursively improve the sub-pixel accuracy of corresponding points by adjusting the location of the window function with sub-pixel accuracy. As a result, the coordinates of corresponding points are obtained with sub-pixel accuracy. Experimental evaluation shows that the displacement between corresponding points with 0.05-pixel accuracy when using 32×32-pixel matching window. See Refs. [6], [7] for detailed procedure of the POC-based correspondence search.

3.3 System Parameters and Other Techniques

The corresponding points between two stereo images obtained in the above algorithm are used to reconstruct the 3D information for face recognition. The listed below are major system parameters in our face recognition system and additional techniques to improve the accuracy of 3D facial data.

- **System parameters**
  Actual system parameters in our 3D facial capture system regarding stereo correspondence are summarized as follows: (i) matching block size is 33×33 pixels (weighted by 2D Hanning window), (ii) spectrum weighting function is 2D Gaussian with $\sigma = \sqrt{0.5}$, (iii) number of fitting points for sub-pixel disparity estimation is 5×5, (iv) number of layers for the coarse-to-fine search is 6, (v) maximum number of iterations for sub-pixel window alignment is 5, (vi) camera calibration is performed by Zhang's method [16] using a 20 mm × 20 mm checker pattern.
- **Techniques to eliminate wrong matches**
  In addition, we introduce some techniques to reduce the number of wrong matches in stereo correspondence. First, we use the peak value of POC function as a measure of reliability in local image matching around a correspondence candidate. When the normalized peak of POC function is below the specified value (0.3 in our system), the system rejects the candidate of the corresponding point. Also, we use epipolar geometric constraint to eliminate wrong matches in stereo correspondence.

These key features make possible to use the narrow-baseline stereo camera head shown in Fig. 1 (b) and to achieve fully automatic 3D reconstruction of a human face with very high accuracy. For the camera head, about 2,500–4,700 corresponding points are detected automatically, and hence the 3D facial information consists of 2,500–4,700 reconstructed points. The accuracy of the developed passive 3D capture system is evaluated through the experimental measurement of a reference planar object with wooden texture; the resulting RMS (Root Mean Square) error in measurement is 0.6 mm at a distance of 50 cm. Figures 2(a) and (b) show stereo images from the camera head and the reconstructed 3D information, respectively.

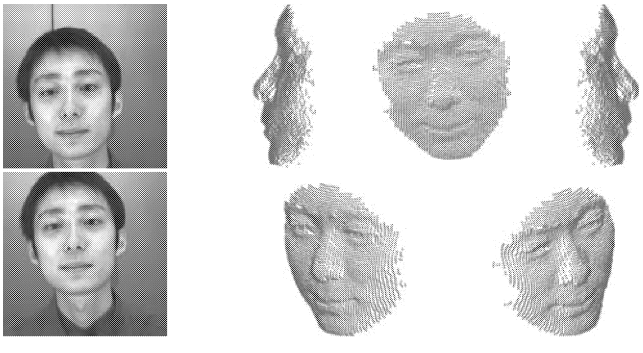


Fig. 2 Examples of 3D facial data captured by the proposed system: (a) stereo images captured by the proposed system, and (b) reconstructed 3D facial data viewed from different angles.

4. 3D Face Matching Algorithm

The face matching algorithm consists of the two steps: (i) align the 3D face surfaces with each other and (ii) evaluate their similarity based on some distance measure.

4.1 3D Face Alignment

As for the 3D face alignment, we have decided to use the simple 3D data registration algorithm, called an ICP (Iterative Closest Point) algorithm [4], since the quality of 3D information captured by our stereo vision system is sufficiently high. Let P be the set of 3D points of a face, and P' be the set of 3D points of another face. The ICP algorithm is summarized as follows.

Procedure: "ICP"
Input:

- the sets of 3D points P and P'

Output:

- the sets of aligned 3D points P and P'
- the transformation parameters $\boldsymbol{R}$ and $\boldsymbol{t}$

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

The original ICP algorithm described above may take much time to align 3D facial data sets. In order to reduce the computation time and improve the registration accuracy of the original ICP algorithm, we employ two techniques: (i) the coarse-to-fine strategy and (ii) the nose region extraction.

(i) Coarse-to-fine strategy

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 and the number of corresponding points gradually increases as the iteration step increases. The 3D point sampling is performed as follows.

Procedure: "3D point sampling"
Input:

- the set of 3D points P
- the parameter $a \geq 0$

Output:

- the set of sampled 3D points P_a

Step 1: Compute the Euclidean distance d^a_{ij} between p_i in P and the closest point p_j in P $(i \neq j)$.
Step 2: Compute the average value $\bar{d}^a$ of d^a_{ij}.
Step 3: Sample P with the sampling interval $\bar{d}^a \times a$. □

(ii) Nose region extraction

The facial images taken from the same person at the different timing are slightly changed by facial expression. The original ICP algorithm cannot handle the non-rigid body transformation such as facial expression changes. Addressing this problem, we use a nose region, which 3D structure is invariant to facial expression changes, to align 3D face surfaces accurately. The nose region extraction is performed as follows.

Procedure: "Nose region extraction"
Input:

- the set of 3D points P
- the parameter $b \geq 0$

Output:

- the set of extracted 3D points P_b

Step 1: Compute the center of gravity of P.
Step 2: Compute the Euclidean distance d^b_i between the center of gravity and p_i in P.
Step 3: Compute the average value $\bar{d}^b$ of d^b_i.
Step 4: Extract 3D points which the distance from the center of gravity is within $\bar{d}^b \times b$. □

In this paper, the parameters for the improvement techniques such as a, b and the number of layers are empirically determined. The number of layers is 4 (i.e., $s_{max} = 3$), and the parameters a for the 3D point sampling and b for the nose region extraction are set to $a = \{0.0, 1.0, 2.0, 3.0\}$ and $b = \{1.5, 1.5, 2.0, 3.0\}$ for each layer. These parameters have to be optimized depending on the individual 3D facial scanners. Figure 3 shows the 3D data set for each layer.

The following is the 3D face registration algorithm combining the ICP algorithm with the above two improvement techniques.

Procedure: "3D face registration"
Input:

- the set of 3D points of a face P
- the set of 3D points of another face P'
- the parameter sets a and b

Output:

- the sets of aligned 3D points P and P'

Step 1: Compute the center of gravity of each data set and align P and P' based on the center of gravity. Set the counter as $s = s_{max}$.
Step 2: Sample the points in P and P' and extract the nose region of each data set using "3D point sampling" and "Nose region extraction" to obtain P_{ab} and P'_{ab}, where the parameters for "3D point sampling" and "Nose region extraction" are $a(s)$ and $b(s)$, respectively.
Step 3: Align P_{ab} and P'_{ab} using "ICP" and transform the points in P' with $\boldsymbol{R}$ and $\boldsymbol{t}$.
Step 4: Decrement the counter by 1 as $s = s - 1$ and repeat from **Step 2** to **Step 3** while $s \geq 0$. □

Figure 4 shows examples of the 3D point registration

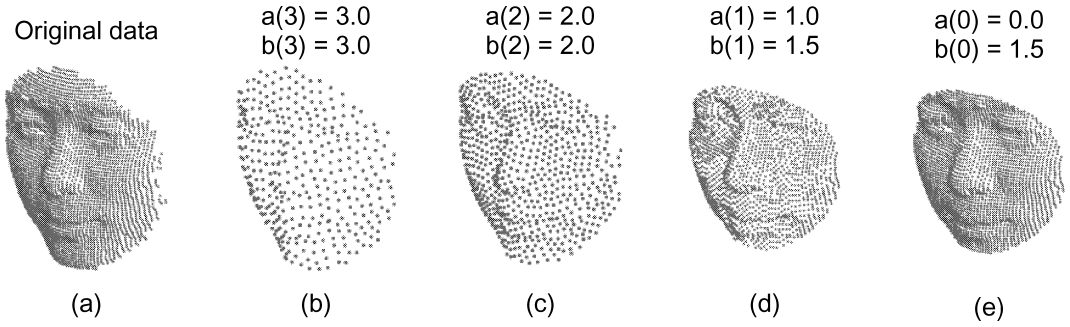


Fig. 3 3D data set for each layer: (a) original data set, (b) 3rd layer ($s = 3$), (c) 2nd layer ($s = 2$), (d) 1st layer ($s = 1$) and (e) 0th layer ($s = 0$).

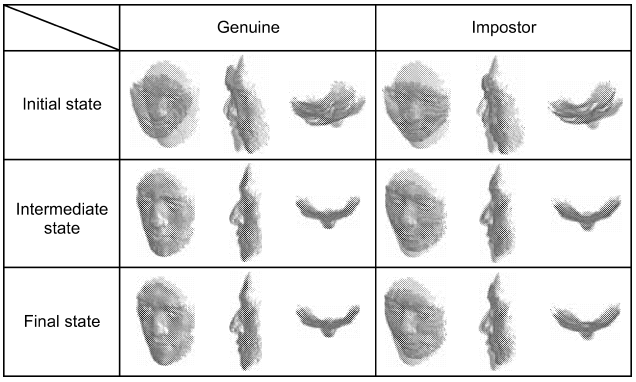


Fig. 4 3D point registration using ICP algorithm.

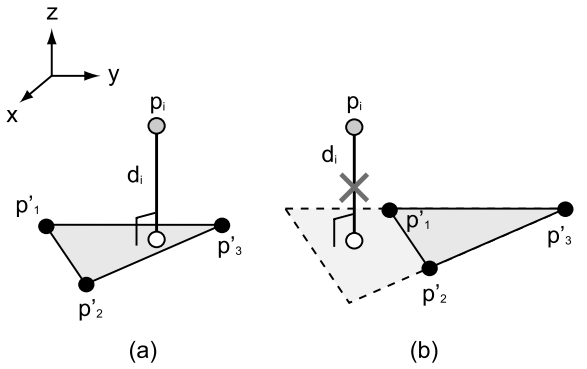


Fig. 5 Similarity evaluation: the orthogonal projection of p_i onto the plane of the three points is inside the triangular patch (a) and is not inside the triangular patch (b).

using the proposed 3D face registration algorithm. The genuine data sets are correctly aligned with each other, while the impostor data sets are not aligned even if the iteration step increases. Therefore, we can evaluate the similarity between the two data sets using the distance between aligned 3D points.

4.2 Similarity Evaluation

Dissimilarity between the two 3D facial data P and P' (normalized by the proposed registration algorithm) is evaluated by a simple point-to-plane distance. The similarity evaluation is performed as follows.

Procedure: "Distance measurement"

Input:

- the sets of aligned 3D points P and P'

Output:

- the average point-to-plane distance between P and P'

Step 1: For the point p_i in P, find the 3 points, p'_1, p'_2 and p'_3, in P' that are closest to p_i.
Step 2: Construct the triangle $\triangle p'_{123}$ using p'_1, p'_2 and p'_3.
Step 3: If the orthogonal projection of p_i onto the plane of $\triangle p'_{123}$ is inside, calculate the distance d_i between the point p_i and $\triangle p'_{123}$ as shown in Fig. 5(a). Otherwise, omit the point p_i for distance calculation as shown in Fig. 5(b).
Step 4: Repeat **Step 1** to **Step 3** for every point in P.
Step 5: Calculate an average of point-to-plane distances d_i.
□

If the distance is long, the pair of facial data is from different person. If short, it is from the same person.

5. Experiments and Discussion

In this section, we describe the performance evaluation of the proposed 3D face recognition system. We carry out two experiments: Experiment 1 is for evaluating the recognition performance of the active and passive 3D face recognition systems, and Experiment 2 is for evaluating the recognition performance of the original and proposed ICP algorithms.

5.1 Experiment 1: Comparison of the Active and Passive 3D Face Recognition Systems

The recognition performance is evaluated through an experimental matching of 23 subjects. In this experiment, 3 independent snapshots with neutral expression are captured by the face recognition system at different sessions for each subject, resulting in a total of 69 facial data. The recognition testing is done using a set of 2,346 ($= {}_{69}C_2$) pairs of facial data, including 69 ($= 23 \times {}_3C_2$) genuine attempts and 2,277 ($= {}_{69}C_2 - 69$) impostor attempts.

We compare two different 3D face recognition systems using active and passive 3D measurement. The active 3D

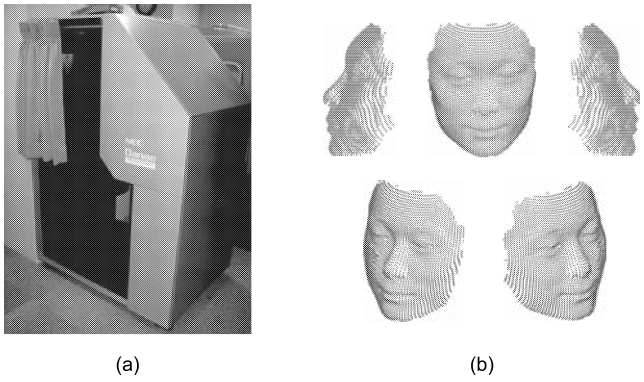


Fig. 6 Active 3D measurement system: (a) active 3D scanner "Danae 200" (made by NEC), and (b) examples of 3D facial data captured by Danae 200.

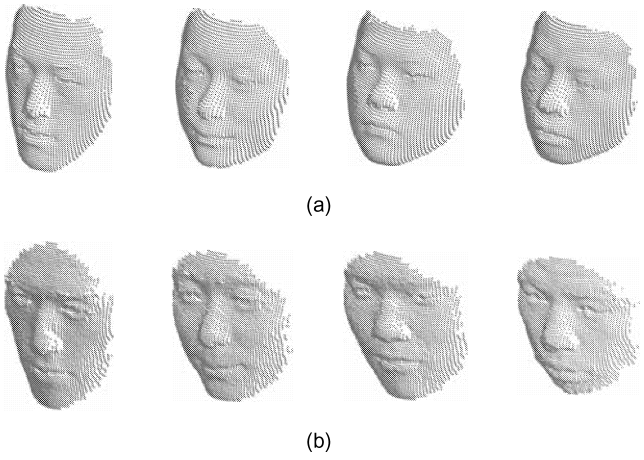


Fig. 7 Examples of 3D facial data: (a) 3D facial data captured by Danae 200, and (b) 3D facial data captured by the proposed system.

face recognition system consists of three steps: (i) capture the 3D facial data by using the active 3D capture system "Danae 200" made by NEC, (ii) align the 3D facial data using the 3D face registration algorithm, and (iii) evaluate the similarity based on distance between two data. Figure 6 shows Danae 200 and an example of 3D facial data captured by this system. The passive 3D face recognition system is described in the previous section. Figure 7 shows examples of 3D facial data captured by each system.

Figures 8(a) and (b) show the distance distributions of active and passive 3D face recognition systems, respectively, where the black bars indicate distance distributions for genuine attempts and the gray ones indicate distance distributions for impostor attempts. As for both cases, the distribution shows a good separation of genuine-matching and impostor-matching distances. Tables 1 and 2 summarize the average, the maximum and the minimum values of distances for active and passive 3D face recognition systems, respectively. As for the proposed system, a distance value within 0.79–0.92 mm 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 lower than the separation point then the two are deemed to be from the same individuals. As a result, the proposed system exhibits efficient recognition performance which is comparable with the active 3D face recognition system.

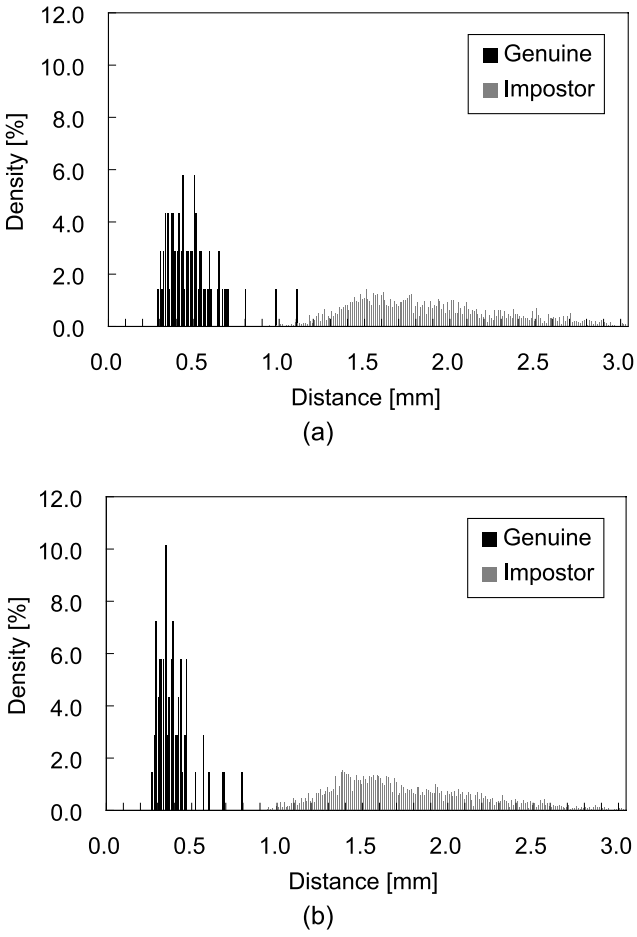


Fig. 8 Distribution of distances: (a) active 3D face recognition system, and (b) the proposed system.

Table 1 Average, maximum and minimum values of distances for genuine and impostor attempts (active 3D face recognition system).

	Ave. [mm]	Max. [mm]	Min. [mm]
Genuine	0.4777	1.0855	0.2815
Impostor	1.9288	3.9144	0.9293

Table 2 Average, maximum and minimum values of distances for genuine and impostor attempts (the proposed system).

	Ave. [mm]	Max. [mm]	Min. [mm]
Genuine	0.3872	0.7862	0.2687
Impostor	1.7403	3.9172	0.9213

5.2 Experiment 2: Comparison of the Original and Proposed ICP-Based 3D Face Recognition Algorithms

The recognition performance of the original ICP algorithm and the proposed ICP algorithm is evaluated through an ex-

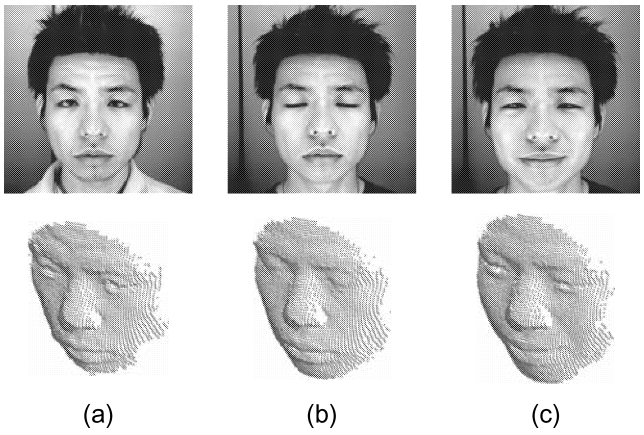


Fig. 9 Examples of 3D facial data for Experiment 2: facial data with (a) neutral expression, (b) closed eyes and (c) smile.

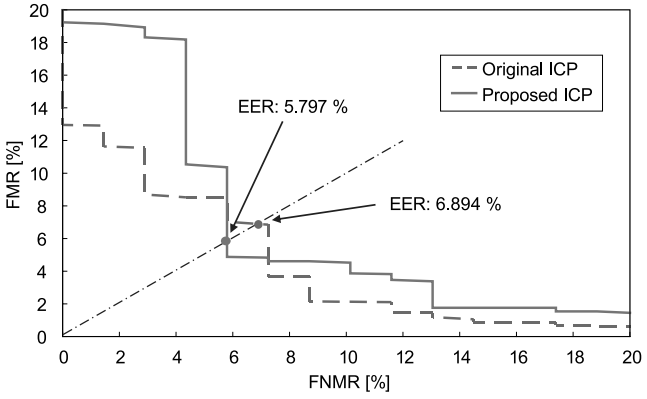


Fig. 10 ROC curve for Experiment 2.

perimental matching of 23 subjects. In this experiment, 3 independent snapshots with neutral expression, closed eyes and smile are captured by the passive 3D face recognition system at different sessions for each subject, resulting in a total of 69 facial data. Thus, the aim of this experiment is to evaluate the performance of the original and proposed ICP algorithms under varying the facial expression. Figure 9 shows examples of 3D facial data used in this experiment.

The performance is evaluated by the Receiver Operating Characteristic (ROC) curve, which illustrates the False Non-Match Rate (FNMR) against the False Match Rate (FMR) at different thresholds on the matching score. We first evaluate the FNMR for all the possible combination of genuine attempts; the number of attempts is 69 ($= 23 \times {}_3C_2$). Next, we evaluate the FMR for all the possible combination of impostor attempts; the number of attempts is 2,277 ($= {}_{69}C_2 - 69$). The performance is also evaluated by the Equal Error Rate (EER), which is defined as the error rate where the FNMR and the FMR are equal. Figure 10 shows the ROC curves for the two algorithms. The proposed algorithm exhibits higher performance, since its ROC curve is located at lower FNMR/FMR region than that of the original algorithm. The EER of the proposed algorithm is 5.89%, while the EER of the original algorithm is 6.89%. As is observed in the above experiments, the proposed algorithm exhibits better performance for 3D facial data with facial expression changes than the original algorithm.

The computation time of the proposed system is evaluated by using MATLAB 6.5.1 on Pentium4 3.20 GHz. The computation times for 3D face reconstruction and 3D face matching are 9.71 sec. and 6.59 sec., respectively. The total computation time is 16.30 sec. When employing only the original ICP algorithm, the computation time for 3D face matching is 34.13 sec. Thus, the use of the improvement techniques for the ICP algorithm makes possible to significantly reduce the computation time.

6. Conclusion

In this paper, we have proposed a face recognition system which uses (i) "passive" stereo vision to capture 3D facial information and (ii) 3D matching based on an ICP (Iterative Closest Point) algorithm. The result clearly demonstrates a potential possibility of creating a cost-effective and easy-to-use face recognition system applicable to a wide range of authentication applications. In our future work, we will develop the 3D face recognition system having facial expression analysis.

References

[1] S.Z. Li and A.K. Jain, Handbook of Face Recognition, Springer, 2005.
[2] A. Scheenstra, A. Ruifrok, and R. Veltkamp, "A survey of 3D face recognition methods," Audio- and Video-Based Biometric Person Authentication (AVBPA 2005), vol.3546, pp.891–899, July 2005.
[3] K.W. Bowyer, K. Chang, and P. Flynn, "A survey of 3D and multi-modal 3D+2D face recognition," Notre Dame Department of Computer Science and Engineering Technical Report, Jan. 2004.
[4] Z. Zhang, "Iterative point matching for registration of free-form curves," Technical Report RR-1658, INRIA-Sophia Antipolis, Valbonne Cedex, France, 1992.
[5] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," IEICE Trans. Fundamentals, vol.E86-A, no.8, pp.1925–1934, Aug. 2003.
[6] K. Takita, M.A. Muquit, T. Aoki, and T. Higuchi, "A sub-pixel correspondence search technique for computer vision applications," IEICE Trans. Fundamentals, vol.E87-A, no.8, pp.1913–1923, Aug. 2004.
[7] M.A. Muquit, T. Shibahara, and T. Aoki, "A high-accuracy passive 3D measurement system using phase-based image matching," IEICE Trans. Fundamentals, vol.E89-A, no.3, pp.686–697, March 2006.
[8] C.D. Kuglin and D.C. Hines, "The phase correlation image alignment method," Proc. Int. Conf. on Cybernetics and Society, pp.163–165, 1975.
[9] H. Shekarforoush, M. Berthod, and J. Zerubia, "Subpixel image registration by estimating the polyphase decomposition of the cross power spectrum," INRIA Technical Report, no.2707, Nov. 1995.
[10] H. Foroosh, J.B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," IEEE Trans. Image Process., vol.11, no.3, pp.188–200, March 2002.
[11] M. Balci and H. Foroosh, "Subpixel estimation of shifts directly in the Fourier domain," IEEE Trans. Image Process., vol.15, no.7, pp.1965–1972, July 2006.
[12] Q.S. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only

matched filtering of Fourier-Mellin transforms for image registration and recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.16, no.12, pp.1156–1168, Dec. 1994.
[13] J.L. Horner and P. Gianino, "Phase-only matched filtering," Appl. Opt., vol.23, pp.812–816, 1984.
[14] T. Nomura, K. Itoh, K. Matsuoka, and Y. Ichioka, "Binary fourier-phase only correlation," Opt. Lett., vol.15, pp.810–811, 1990.
[15] K. Ito, H. Nakajima, K. Kobayashi, T. Aoki, and T. Higuchi, "A fingerprint matching algorithm using phase-only correlation," IEICE Trans. Fundamentals, vol.E87-A, no.3, pp.682–691, March 2004.
[16] Z. Zhang, "A flexible new technique for camera calibration," Technical Report MSR-TR-98-71, Microsoft Research, Dec. 1998.

Akihiro Hayasaka received the B.E. degree in information engineering, and the M.S. degree in information sciences from Tohoku University, Sendai, Japan, in 2004 and 2006, respectively. He is currently working toward the Ph.D. degree. His research interest includes signal and image processing, and biometric authentication.

Takuma Shibahara received the B.S. degree in mathematical sciences from Yamagata University, Yamagata, Japan, in 2003, and the M.S. degree in information sciences from Tohoku University, Sendai, Japan, in 2005. He is currently working toward the Ph.D. degree. His research interest includes computer vision and image processing.

Koichi Ito received the B.E. degree in electronic engineering, and the M.S. and Ph.D. degree in information sciences from Tohoku University, Sendai, Japan, in 2000, 2002 and 2005, respectively. He is currently an Assistant Professor of the Graduate School of Information Sciences at Tohoku University. From 2004 to 2005, he was a Research Fellow of the Japan Society for the Promotion of Science. His research interest includes signal and image processing, and biometric authentication.

Takafumi Aoki received the B.E., M.E., and D.E. degrees in electronic engineering from Tohoku University, Sendai, Japan, in 1988, 1990, and 1992, respectively. He is currently a Professor of the Graduate School of Information Sciences at Tohoku University. For 1997–1999, he also joined the PRESTO project, Japan Science and Technology Corporation (JST). His research interests include theoretical aspects of computation, VLSI computing structures for signal and image processing, multiple-valued logic, and biomolecular computing. Dr. Aoki received the Outstanding Paper Award at the 1990, 2000, 2001 and 2006 IEEE International Symposiums on Multiple-Valued Logic, the Outstanding Transactions Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan in 1989 and 1997, the IEE Ambrose Fleming Premium Award in 1994, the IEICE Inose Award in 1997, the IEE Mountbatten Premium Award in 1999, the Best Paper Award at the 1999 IEEE International Symposium on Intelligent Signal Processing and Communication Systems, the IP Award at the 7th LSI IP Design Award in 2005, and the Best Paper Award at the 14th Workshop on Synthesis And System Integration of Mixed Information technologies.

Hiroshi Nakajima received the B.E. degree in electronic engineering from Tohoku University, Sendai, Japan, in 1990. He is currently with Systems Development Department, Yamatake Corporation, Fujisawa, Japan. His research interest includes biometric image processing.

Koji Kobayashi received the B.E. and M.E. degrees in electronic engineering from Tohoku University, Sendai, Japan, in 1976, and 1978, respectively. He is currently a general manager of Vision Sensing Department, Yamatake Corporation, Fujisama, Japan. His general interests includes real-time automation system architecture, network communication protocol LSI, biometric image processing, CMOS image sensor, and three-dimensional sensing.