3D Face Recognition Using Passive Stereo Vision

Naohide Uchida, Takuma Shibahara, Takafumi Aoki* Hiroshi Nakajima, Koji Kobayashi

Graduate School of Information Sciences, Tohoku University
Sendai-shi 980-8579, Japan
*E-mail: aoki@ecei.tohoku.ac.jp

Yamatake Corporation
Isehara-shi 259-1195, Japan

Abstract— This paper proposes a face recognition system that uses (i) passive stereo vision to capture three-dimensional (3D) facial information and (ii) 3D matching using a simple ICP (Iterative Closest Point) algorithm. So far, the reported 3D face recognition techniques assume the use of active 3D measurement for 3D facial capture. However, active methods employ structured illumination (structure projection, phase shift, gray-code demodulation, etc.) or laser scanning, which is not desirable in many human recognition applications. A major problem of using passive stereo vision for 3D measurement is its low accuracy, and thus no passive methods for 3D face recognition have been reported previously. Addressing this problem, we have newly developed a high-accuracy 3D measurement system based on passive stereo vision, where phase-based image matching is employed for sub-pixel disparity estimation. This paper presents the first attempt to create a practical face recognition system based on fully passive 3D reconstruction.

I. Introduction

As one of the most important biometric authentication methods, 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 from 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]–[4], since 3D data is not affected by translation and rotation and is immune to the effect of illumination variation. The 3D face recognition research is, however, still weakly reported in the published literature. A main reason baffling the development lies in that 3D capture usually requires special expensive equipments.

So far, the reported 3D face recognition techniques tend to use active methods for 3D measurement to capture high-quality 3D facial information. However, active measurement employs structured illumination (structure projection, phase shift, gray-code demodulation, etc.) or laser scanning, which is not necessarily desirable in many cases of human recognition applications. This paper proposes a face recognition system that uses (i) "passive" stereo vision to capture 3D facial information and (ii) 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, and, with this reason, no practical approaches to passive 3D face recognition have been reported to the best of the authors' knowledge. Addressing this problem, we have newly developed a high-quality 3D facial capture system based on passive stereo vision, where a phase-based image matching technique [5],[6] is employed for sub-pixel disparity estimation. The developed 3D capture system can reconstruct 3D facial information with $\sim$0.6mm accuracy at 50cm distance. With the high-quality 3D facial data captured by the developed passive stereo vision system, our experimental observation clearly shows that a simple 3D registration scheme, called ICP (Iterative Closest Point) algorithm, can be used for practical face matching. This paper presents the first demonstration of a practical 3D face recognition system based on a passive 3D measurement system.

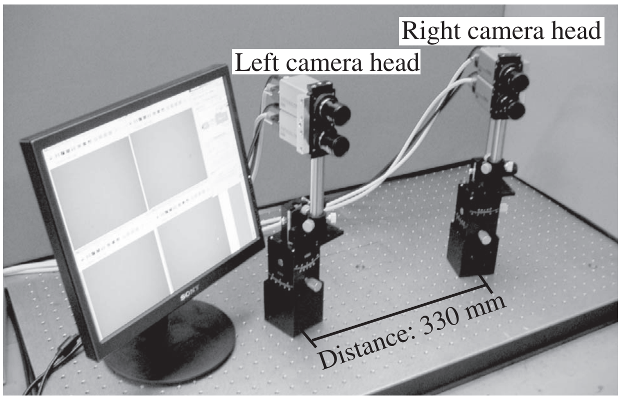
II. High-Quality 3D Measurement with Passive Stereo Vision

Figure 1 shows the developed high-quality passive 3D measurement system. The system consists of two stereo camera heads – the left camera head for capturing the left view of a face and the right camera head for right view. Each camera head 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 is 46mm limited simply by the size of the camera chassis.

In general, the following two features must be considered in designing the optimal camera configuration for face recognition:

- 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 data when compared with wide-baseline configuration.
- The wide-baseline camera configuration makes possible to achieve higher accuracy, but automatic stereo corre-

0-7803-9134-9/05/$20.00 ©2005 IEEE



(a)

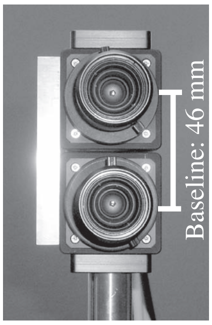


Image size: 1000 x 1000 pixels
CCD: Adimec-1000m/D
10 bits digital resolution
monochrome
Lens: TRON, FV1520
15 mm focal length
Image grabber: Coreco Imaging
X64-CL-DUAL-32M
Measurement range: 400 ~ 600 mm
Lighting: ambient light

(b)

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

spondence is very difficult and is limited only to a small number of edge points. This is almost unacceptable for practical 3D face recognition systems.

These are major reasons why the automatic passive stereo vision system could not be used for 3D face recognition.

We decided to use the narrow-baseline configuration shown in Figure 1 (b), where the problem of low accuracy in stereo correspondence must be overcome. Addressing this problem, we have newly developed a high-accuracy stereo correspondence technique using phase-based image matching (see our papers [5], [6] for detailed discussion on this technique). Listed below are key features of the developed technique.

• High-accuracy 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. In general, the image matching methods using phase information in 2D Discrete Fourier Transform (2D DFT) [7] exhibit better registration performance than the methods using SAD (Sum of Absolute Differences). In our work [5], we have proposed a high-accuracy image matching method employing (i) an analytical function fitting technique to estimate the sub-pixel position of the correlation peak, (ii) a windowing technique to eliminate the effect of periodicity in 2D DFT, and (iii) a spectrum weighting technique to reduce the effect of aliasing and noise. All these techniques of image matching are adopted in our 3D capture system. Our experimental observation shows that POC-based matching can estimate displacement between two images with 0.01-pixel accuracy when the image size is about 100×100 pixels.

• Sub-pixel correspondence search

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 degrade 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. Experimental evaluation shows that the developed method makes possible to estimate the displacement between corresponding points with approximately 0.05-pixel accuracy when using 11×11-pixel matching windows.

• 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 [8] using a 20mm$\times$20mm 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 reject 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 Figure 1 (b) and to achieve fully automatic 3D reconstruction of a human face with very high accuracy. For each camera head, about $2000 \sim 2500$ corresponding points are detected automatically, and hence the 3D facial information obtained from two camera heads consists of $4000 \sim 5000$ reconstructed points in total. The accuracy of

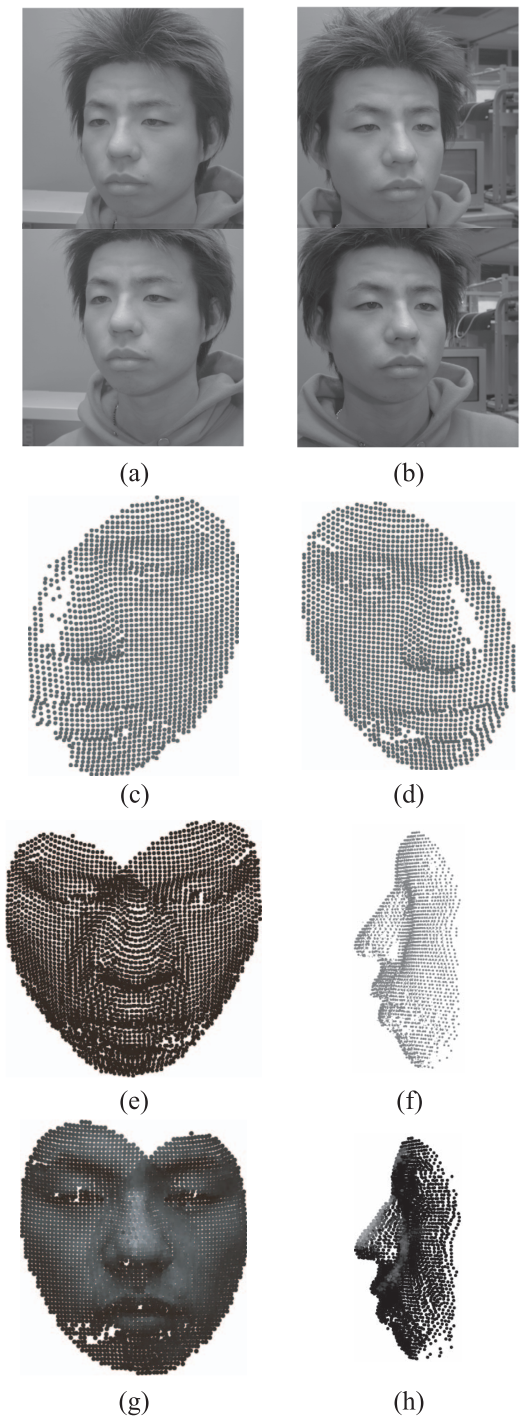


Fig. 2. 3D face data captured by the passive stereo vision system: (a) stereo images from the left camera head, (b) stereo images from the right camera head, (c) reconstructed 3D information (left camera head), (d) reconstructed 3D information (right camera head), (e) and (f) integrated 3D data (left camera head + right camera head), (g) and (h) texture mapped 3D data.

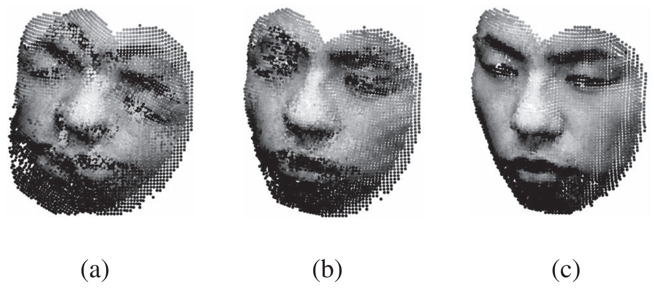


Fig. 3. 3D alignment by ICP: (a) initial position, (b) after 26 iterations, and (c) after 98 iterations.

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 0.6mm at a distance of 50cm.

Figure 2 (a) and (b) show stereo images from the left camera head and the right camera head, respectively. Figure 2 (c) and (d) show the reconstructed 3D information captured from the left camera head and that from the right camera head, respectively. Figure 2 (e) and (f) show the overall 3D data that integrates the left and right views (c) and (d), and Figure 2 (g) and (h) are the corresponding texture mapped 3D data.

III. 3D Face Matching

Given a pair of 3D face data, face matching can be performed by the two steps: (i) align the 3D face surfaces with each other (Figure 3) 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 [9], since the quality of 3D information captured by our stereo vision system is sufficiently high.

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.

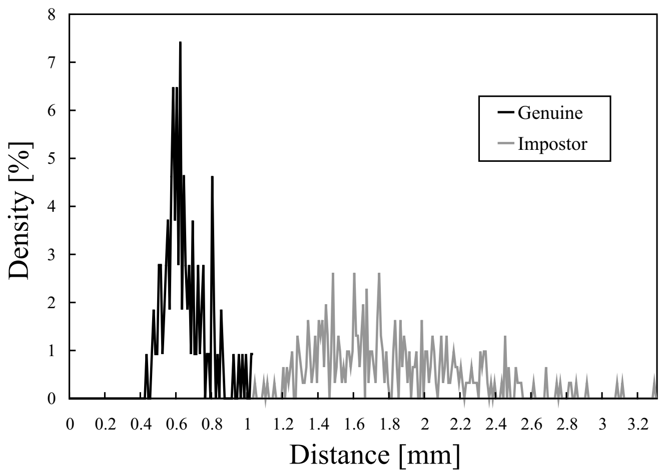


Fig. 4. Distribution of distances.

TABLE I. Average, maximum and minimum values of distances for genuine attempts and impostor attempts.

	Average	Max	Min
Genuine	0.661 mm	1.021 mm	0.429 mm
Impostor	1.839 mm	3.280 mm	1.034 mm

Dissimilarity between the two 3D facial data M and M' (normalized by ICP) is evaluated by a simple point-to-plane distance. We first reconstruct the surface of M' based on 2D Delaunay triangulation, and then calculate the distance d_i from the point $m_i (\in M)$ to the reconstructed surface of M' for every point in the overlapped facial region. The distance between the two facial data is defined as an average of individual point-to-plane distances d_i.

IV. Experiments and Discussion

The performance of the proposed passive 3D face recognition system was evaluated through an experimental matching of 18 subjects. In this experiment, 4 independent snapshots with neutral expression are captured at different sessions for each subject, resulting in a total of 72 facial data. The recognition testing was done using a set of 414 pairs of facial data, including 108 ($= 18 \times {}_4 C_2$) genuine attempts and 306 ($= 18 \times 17$) impostor attempts. Figure 4 shows the distribution of distances for genuine attempts and impostor attempts. The distribution shows a good separation of genuine-matching and impostor-matching distances. Table I summarizes the average, the maximum and the minimum values of distances for genuine attempts and impostor attempts. A distance value within $1.025 \sim 1.030$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 individual.

V. Conclusion

In this paper, we have proposed a face recognition system that uses (i) "passive" stereo vision to capture 3D facial information and (ii) 3D matching based on a simple ICP (Iterative Closest Point) algorithm. This is the first successful attempt to realize a practical face recognition system on the basis of a passive 3D measurement technique. The result clearly demonstrates a potential possibility of creating a cost-effective, easy-to-use face recognition system applicable to a wide range of authentication applications.

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, "Face recognition: A literature survey," *UMD CfAR Technical Report CAR-TR-948*, 2000.

[2] X. Lu and A. K. Jain, "Deformation analysis for 3D face matching," *Proc. 7th IEEE Workshop on Applications of Computer Vision*, pp. 99–104, Jan. 2005.

[3] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, Sept. 2003.

[4] C. Beumier and M. Acheroy, "Automatic 3D face authentication," *Image and Vision Computing*, vol. 18, no. 4, pp. 315–321, 2000.

[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 subpixel correspondence search technique for computer vision applications," *IEICE Trans. Fundamentals*, vol. E87-A, no. 8, pp. 1913–1923, Aug. 2004.

[7] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," *Proc. Int. Conf. on Cybernetics and Society*, pp. 163–165, 1975.

[8] Z. Zhang, "A flexible new technique for camera calibration," *Technical Report MSR-TR-98-71, Microsoft Research*, Dec. 1998.

[9] Z. Zhang, "Iterative point matching for registration of free-form curves," *Technical Report RR-1658, INRIA-Sophia Antipolis, Valbonne Cedex, France*, 1992.