IEICE TRANS. FUNDAMENTALS, VOL.E89–A, NO.3 MARCH 2006

PAPER Special Section on Multidimensional Signal Processing and Its Application

# A High-Accuracy Passive 3D Measurement System Using Phase-Based Image Matching

**Mohammad Abdul MUQUIT**[†a)], **Takuma SHIBAHARA**[†], *Nonmembers*, *and* **Takafumi AOKI**[†], *Member*

**SUMMARY** This paper presents a high-accuracy 3D (three-dimensional) measurement system using multi-camera passive stereo vision to reconstruct 3D surfaces of free form objects. The proposed system is based on an efficient stereo correspondence technique, which consists of (i) coarse-to-fine correspondence search, and (ii) outlier detection and correction, both employing phase-based image matching. The proposed sub-pixel correspondence search technique contributes to dense reconstruction of arbitrary-shaped 3D surfaces with high accuracy. The outlier detection and correction technique contributes to high reliability of reconstructed 3D points. Through a set of experiments, we show that the proposed system measures 3D surfaces of objects with sub-mm accuracy. Also, we demonstrate high-quality dense 3D reconstruction of a human face as a typical example of free form objects. The result suggests a potential possibility of our approach to be used in many computer vision applications.

***key words:*** *3D measurement, stereo vision, phase-based image matching, phase-only correlation, outlier detection*

## 1. Introduction

Recently the demand of high-accuracy 3D measurement is rapidly growing in a variety of computer vision applications, for instance, robot vision, human-computer interface, biometric authentication, etc. Existing 3D measurement techniques are classified into two major types—active and passive. In general, active measurement employs structure illumination (structure projection, phase shift, moire topography, etc.) or laser scanning, which is not necessarily desirable in many applications. On the other hand, passive 3D measurement techniques based on stereo vision have the advantages of simplicity and applicability, since such techniques require simple instrumentation. (See [1] for a good survey on this topic.) However, poor reconstruction quality still remains as a major issue for passive 3D measurement, due to the difficulty in finding accurate correspondence between stereo images; this problem is generally known as "correspondence problem" [2]. As a result, application of passive stereo vision to high-accuracy 3D measurement system for capturing 3D surfaces of free form objects is still weakly reported in the published literature. The objective of this paper is to implement a passive 3D measurement system, whose reconstruction accuracy is comparable with that of practical active 3D scanners based on structured light projection.

Manuscript received June 28, 2005.
Manuscript revised October 5, 2005.
Final manuscript received November 21, 2005.

[†]The authors are with Graduate School of Information Sciences, Tohoku University, Sendai-shi, 980-8579 Japan.
a) E-mail: mukit@aoki.ecei.tohoku.ac.jp

DOI: 10.1093/ietfec/e89–a.3.686

The overall accuracy of passive 3D measurement is mainly determined by (i) the baseline length between two cameras and (ii) the accuracy of estimated disparity between corresponding points [2]. Conventional approaches to passive 3D measurement employ wide-baseline camera pairs combined with feature-based correspondence matching [1]. However, in such approaches only a limited number of corresponding points can be used for 3D reconstruction. On the other hand, area-based correspondence matching (which must be combined with narrow-baseline stereo cameras to avoid projective distortion between stereo images) makes possible to increase the number of corresponding points. However, the accuracy of 3D measurement becomes severely restricted when the baseline is narrow [3]. In this paper, therefore, we focus on the techniques for high-accuracy stereo correspondence in order to overcome the limitation of measurement accuracy in narrow-baseline stereo vision.

The key idea in this paper is to employ phase-based image matching for high-accuracy stereo correspondence. Our experimental observation shows that the methods using phase-based image matching exhibit better registration performance than the methods using SAD (Sum of Absolute Differences) in general [4], [5]. In our previous work, we presented an application of phase-based image matching to a generic correspondence search problem [6], where a coarse-to-fine strategy combined with a sub-pixel window alignment technique is used to determine correspondence between two images with sub-pixel resolution.

The goal of this paper is to implement the phase-based correspondence search technique in a practical 3D measurement system, and to analyze its impact on the system's performance (i.e., reconstruction accuracy and reliability). We demonstrate that the use of phase-based correspondence search makes possible to achieve fully automatic high-accuracy 3D measurement with a narrow-baseline stereo vision system. Another contribution of this paper is a highly reliable technique for detecting and correcting outliers (wrong or unreliable corresponding points), which are generally caused by occlusion, image noise, photometric distortion, etc. [1], [7]. The proposed technique is based on Phase-Only Correlation (POC) function—a correlation function used in the phase-based image matching to evaluate similarity between two images. We found that the peak value of the POC function can be used as an efficient measure of reliability for stereo correspondence. We can easily detect outliers in corresponding points by finding the points

for which correlation peak value is less than a certain threshold. By correcting correspondence for the detected outliers, we can achieve high-quality dense reconstruction of 3D objects. Through a set of experiments, we show that the proposed system measures 3D surfaces of regular shaped objects (a solid plane and a solid sphere) with sub-mm accuracy. Also, we demonstrate high-quality dense 3D reconstruction of a human face as a typical example of free form objects, which suggests a potential possibility of our approach to be used in many computer vision applications.

## 2. High-Accuracy Stereo Correspondence Using Phase-Based Image Matching

In Sects. 2.1 and 2.2, we describe the high-accuracy image matching based on Phase-Only Correlation (POC) function†, and its application to stereo correspondence problem. (See our papers [5], [6] for earlier discussions on the proposed techniques.) Section 2.3 describes the outlier detection and correction technique using the POC function.

### 2.1 Phase-Based Image Matching

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 $n_1 = -M_1, \cdots, M_1$ and $n_2 = -M_2, \cdots, M_2$ for mathematical simplicity, and hence $N_1 = 2M_1 + 1$ and $N_2 = 2M_2 + 1$. The 2D Discrete Fourier Transforms (2D DFTs) of the two images are given by

$$\begin{aligned} F(k_1, k_2) &= \sum_{n_1 n_2} f(n_1, n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2} \\ &= A_F(k_1, k_2) e^{j\theta_F(k_1, k_2)}, \qquad (1) \\ G(k_1, k_2) &= \sum_{n_1 n_2} g(n_1, n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2} \\ &= A_G(k_1, k_2) e^{j\theta_G(k_1, k_2)}, \qquad (2) \end{aligned}$$

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 the operator $\sum_{n_1 n_2}$ 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 $e^{j\theta_F(k_1,k_2)}$ and $e^{j\theta_G(k_1,k_2)}$ are phase components.

The cross-phase spectrum (or normalized cross spectrum) $\hat{R}(k_1, k_2)$ is defined as

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

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

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

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 $f(n_1, n_2) = g(n_1, n_2)$, the POC function $\hat{r}(n_1, n_2)$ becomes the Kronecker delta function.) When two images are not similar, the peak drops significantly. The height of the peak can be used as a good similarity measure for image matching, and the location of the peak shows the translational displacement between the two images.

In the following, we derive the analytical peak model for the POC function between the same 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 $\delta_1$ and $\delta_2$ 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

$$f(n_1, n_2) = f_c(x_1, x_2)|_{x_1 = n_1 T_1, x_2 = n_2 T_2}, \qquad (5)$$

$$g(n_1, n_2) = f_c(x_1 - \delta_1, x_2 - \delta_2)|_{x_1 = n_1 T_1, x_2 = n_2 T_2}, \qquad (6)$$

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 $\hat{R}(k_1, k_2)$ and the POC function $\hat{r}(n_1, n_2)$ between $f(n_1, n_2)$ and $g(n_1, n_2)$ will be given by

$$\hat{R}(k_1, k_2) \simeq e^{j\frac{2\pi}{N_1} k_1 \delta_1} e^{j\frac{2\pi}{N_2} k_2 \delta_2}, \qquad (7)$$

$$\hat{r}(n_1, n_2) \simeq \frac{\alpha}{N_1 N_2} \frac{\sin\{\pi(n_1 + \delta_1)\}}{\sin\{\frac{\pi}{N_1}(n_1 + \delta_1)\}} \frac{\sin\{\pi(n_2 + \delta_2)\}}{\sin\{\frac{\pi}{N_2}(n_2 + \delta_2)\}}, \qquad (8)$$

where $\alpha = 1$. The above Eq. (8) represents the shape of the peak for the POC function between the same 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 $(\delta_1, \delta_2)$ of the POC function corresponds to the displacement between the two images. We can prove that the peak value $\alpha$ decreases (without changing the function shape itself), when small noise components are added to the original images. Hence, we assume $\alpha \leqq 1$ in practice. For image matching task, we estimate the similarity between two images by the peak value $\alpha$, and estimate the image displacement by the peak position $(\delta_1, \delta_2)$.

Listed below are important techniques for high-accuracy sub-pixel image matching.

**(i) Function fitting for high-accuracy estimation of peak position**

We use Eq. (8)—the closed-form peak model of the POC function—directly for estimating the peak position by function fitting. By calculating the POC function, we can obtain a data array of $\hat{r}(n_1, n_2)$ for each discrete index $(n_1, n_2)$. It is possible to find the location of the peak that may exist between image pixels by fitting the function Eq. (8) to the calculated data array around the correlation peak, where $\alpha$, $\delta_1$, and $\delta_2$ are fitting parameters.

**(ii) Windowing to reduce boundary effects**

†The POC function is sometimes called the "phase correlation function."

Due to the DFT's periodicity, an image can be considered to "wrap around" at an edge, and therefore discontinuities, which are not supposed to exist in real world, occur at every border in 2D DFT computation. We reduce the effect of discontinuity at image border by applying 2D window function to the input images. For this purpose, we use 2D Hanning window defined by

$$w(n_1, n_2) = \frac{1 + \cos\left(\frac{\pi n_1}{M_1}\right)}{2} \frac{1 + \left(\frac{\pi n_2}{M_2}\right)}{2}. \quad (9)$$

**(iii) Spectral weighting technique to reduce aliasing and noise effects**

For natural images, typically the high frequency components may have less reliability (low S/N ratio) compared with the low frequency components. We could improve the estimation accuracy by applying a low-pass-type weighting function $H(k_1, k_2)$ to $\hat{R}(k_1, k_2)$ in frequency domain and eliminating the high frequency components having low reliability. The useful weighting function is the DFT of 2D Gaussian function defined as

$$H(k_1, k_2) \simeq e^{-2\pi^2\sigma^2(k_1^2+k_2^2)}, \quad (10)$$

where $\sigma$ is a parameter that controls the pass-band width. When calculating the POC function, $\hat{R}(k_1, k_2)$ is multiplied by the weighting function $H(k_1, k_2)$ in frequency domain. Then, Eq. (8) is modified as

$$\begin{aligned} \hat{r}(n_1, n_2) &= \frac{1}{N_1 N_2} \sum_{k_1 k_2} \hat{R}(k_1, k_2) H(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2} \\ &\simeq \frac{\alpha}{2\pi\sigma^2} e^{-\{(n_1+\delta_1)^2+(n_2+\delta_2)^2\}/2\sigma^2}, \end{aligned} \quad (11)$$

where $\alpha$, $\delta_1$ and $\delta_2$ are fitting parameters for evaluating the correlation peak. That is, Eq. (11) should be employed for function fitting instead of Eq. (8) when using the spectral weighting technique.

All the above techniques for high-accuracy sub-pixel image matching are adopted in stereo correspondence search in our system. We use POC-based block matching of size $33 \times 33$ pixels, for which we can achieve about 0.05-pixel accuracy in displacement estimation.

### 2.2 Sub-Pixel Correspondence Search

This section discusses a high-accuracy stereo correspondence algorithm based on the sub-pixel image matching mentioned above. The algorithm described here is an improved version of the method reported in our previous paper [6]. The improved points are listed below:

• **Optimizing the shape of spectral weighting function**: We apply low-pass type spectral weighting function (Section 2.1(iii)) for better displacement estimation accuracy. In our previous paper [6], we used simple rectangular type weighting function. However, our experimental observation using the system proposed in this paper shows that Gaussian spectral weighting function provides better performance for measuring natural objects. Therefore, we implement a Gaussian version of the spectral weighting function given in Eq. (10).

• **Peak model of the POC function**: The peak model of the POC function used in the paper [6] was 2D periodic sinc function. Due to the modification in spectral weighting function (Eq. (10)) in this paper, we use the modified peak model of the POC function (the 2D IDFT of the Gaussian spectral weighting function), which is given by Eq. (11). Obviously, the modified peak model becomes the Gaussian function again and can be easily fitted to the discrete data array of the POC function by nonlinear least-square method. In our experiment, we use Levenberg-Marquardt algorithm for the least-square fitting. We empirically found that this peak model of the POC function ensures better correspondence accuracy with simple non-linear function fitting compared to that used in paper [6].

• **Sub-pixel window alignment by image shifting**: In our previous paper [6], the sub-pixel window alignment was done by shifting the center of the window function within a rectangular image block. In this paper, on the other hand, we shift the image to be extracted into the rectangular block with sub-pixel resolution, while keeping the center of the window function unchanged. We found the latter method provides better accuracy in sub-pixel correspondence estimation. This sub-pixel image shifting is done by rotating the phase component of the image block in frequency domain. In the previous version of window alignment, 5 iterations were needed. On the other hand, we found that 3 iterations are enough for the image shifting technique to be converged, which saves calculation time.

• **Procedure of sub-pixel window alignment**: In our previous paper [6], sub-pixel window alignment is done regarding both left and right images. In this paper, this alignment is done only regarding the right image. Thus the reference point in the left image remains unchanged in pixel level posiiton, and its corresponding point in the right image is estimated with sub-pixel accuracy. This reduces the computational complexity of correspondence search.

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 (i), we estimate the stereo correspondence with pixel-level accuracy using hierarchical POC-based block matching with coarse-to-fine strategy. Thus, the estimation error becomes less than 1 pixel for every corresponding point. The second stage (ii) of the algorithm is to recursively improve the sub-pixel accuracy of corresponding points by adjusting the location of the window function (Eq. (9)) with sub-pixel accuracy. As a result, the coordinates of corresponding points are obtained with sub-pixel accuracy.

Let $\boldsymbol{m} = (m_1, m_2)(\in \boldsymbol{Z}^2)$ be a coordinate vector of a reference pixel in the left image $I$ (i.e., the reference image), where $\boldsymbol{Z}$ is the set of integers. The problem of sub-pixel

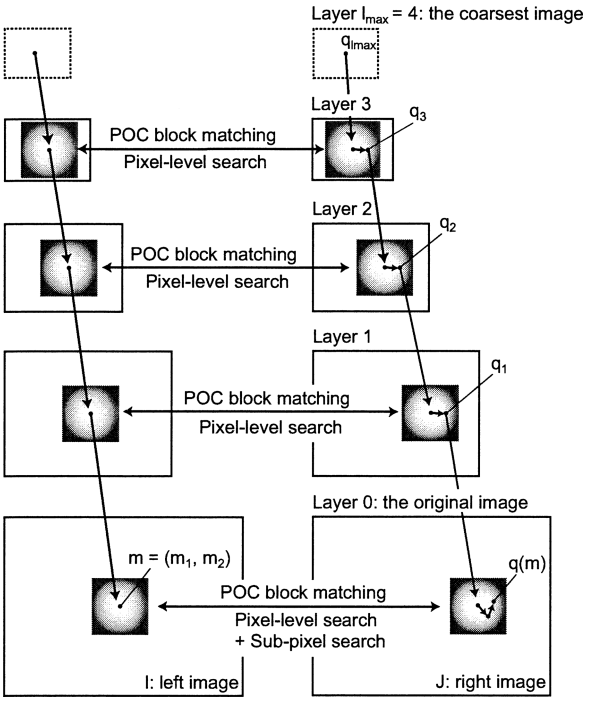


**Fig. 1** Sub-pixel correspondence search using a coarse-to-fine strategy (e.g., $l_{max} = 4$).

correspondence search is to find a real-number coordinate vector $\boldsymbol{q}(\boldsymbol{m}) = (q_1, q_2)(\in \boldsymbol{R}^2)$ in the right image $J$ that corresponds to the reference pixel $\boldsymbol{m}$ in $I$, where $\boldsymbol{R}$ is the set of real numbers, representing the coordinates in $J$ with sub-pixel accuracy. For convenience, we use the symbol $C_I^{int}(\subset \boldsymbol{Z}^2)$ to denote the set of all integer coordinate vectors (with pixel-level accuracy) in the reference image $I$, and the symbol $C_J^{real}(\subset \boldsymbol{R}^2)$ to denote the set of all real-number coordinate vectors (with sub-pixel accuracy) in $J$. Then, the problem is to find the set of corresponding points $\boldsymbol{q}(\boldsymbol{m}) \in C_J^{real}$ for all reference points $\boldsymbol{m} \in C_I^{int}$. Figure 1 shows an overview of the sub-pixel correspondence search algorithm. We give a simple description of the technique here:

**Procedure A**: "Sub-pixel correspondence search"

**Input**:

- the left image $I$ and the right image $J$
- the set of reference points $\boldsymbol{m} = (m_1, m_2) \in C_I^{int}$

**Output**:

- the corresponding points $\boldsymbol{q}(\boldsymbol{m}) \in C_J^{real}$ for all reference points $\boldsymbol{m} = (m_1, m_2) \in C_I^{int}$

**Procedure steps**:

**[Creation of coarse-to-fine image pyramids]**
**Step 1**: Let the images of layer 0 be given by $I_0 = I$ and $J_0 = J$. For $l = 1, 2, \cdots, l_{max}$, create the $l$–th layer images $I_l$ and $J_l$ (i.e., coarser versions of $I_0$ and $J_0$) by reducing the original images $I_0$ and $J_0$ with the scale factor $2^{-l}$.

For every reference point $\boldsymbol{m} = (m_1, m_2) \in C_I^{int}$, do the following **[Pixel-level estimation]** and **[Sub-pixel estimation]**.

**[Pixel-level estimation]**
**Step 2**: In the coarsest layer $l_{max}$, the location of the reference point $\boldsymbol{m} = (m_1, m_2)$ is mapped to the coordinates $(\lfloor 2^{-l_{max}} m_1 \rfloor, \lfloor 2^{-l_{max}} m_2 \rfloor)$. As an initial estimate of the corresponding point at the layer $l_{max}$, we simply assume that

$$\boldsymbol{q}_{l_{max}} = (\lfloor 2^{-l_{max}} m_1 \rfloor, \lfloor 2^{-l_{max}} m_2 \rfloor). \tag{12}$$

That is, we assume that the reference point and its corresponding point have the same coordinates at the coarsest image layer.
Let $l = l_{max} - 1$.
**Step 3**: In the $l$–th layer image $I_l$, the reference point is mapped to $(\lfloor 2^{-l} m_1 \rfloor, \lfloor 2^{-l} m_2 \rfloor)$. In $J_l$, on the other hand, $2\boldsymbol{q}_{l+1}$ gives an initial estimate for the corresponding point based on the result of upper layer estimation $\boldsymbol{q}_{l+1}$. Therefore, from $I_l$ and $J_l$, extract two image blocks (i.e., windowed sub-images) with their centers on $(\lfloor 2^{-l} m_1 \rfloor, \lfloor 2^{-l} m_2 \rfloor)$ and $2\boldsymbol{q}_{l+1}$, respectively. Estimate the displacement between the two image blocks with pixel accuracy using POC-based image matching, which is a simplified version of the matching algorithm described in Sect. 2.1. Let the estimated displacement vector be denoted by $\boldsymbol{\delta}_l = (\delta_{l1}, \delta_{l2})$. The $l$–th layer correspondence $\boldsymbol{q}_l$ is determined as follows:

$$\boldsymbol{q}_l = 2\boldsymbol{q}_{l+1} + \boldsymbol{\delta}_l. \tag{13}$$

**Step 4**: Decrement the counter by 1 as $l = l - 1$ and repeat from **Step 3** to **Step 4** while $l \geqq 0$.
**Step 5**: Let the pixel-level estimation of correspondence be given by $\boldsymbol{q}(\boldsymbol{m}) = \boldsymbol{q}_0$.

**[Sub-pixel estimation]**
**Step 6**: From the original images $I_0$ and $J_0$, extract two image blocks (i.e., windowed sub-images) with their centers on $\boldsymbol{m}$ and $\boldsymbol{q}(\boldsymbol{m})$, respectively. Estimate the displacement between the two blocks with sub-pixel accuracy using the POC-based image matching described in Sect. 2.1. Let the estimated displacement vector with sub-pixel accuracy be denoted by $\boldsymbol{\delta} = (\delta_1, \delta_2)$.
**Step 7**: Update the corresponding point as

$$\boldsymbol{q}(\boldsymbol{m}) = \boldsymbol{q}(\boldsymbol{m}) + \boldsymbol{\delta}, \tag{14}$$

and repeat **Step 6** and **Step 7** until $|\boldsymbol{\delta}|$ converges to small value. □

### 2.3 Outlier Detection and Correction

The sub-pixel correspondence search technique described in Sect. 2.2 determines corresponding point for any given reference point. The robustness of correspondence search is one of the most important characteristics of our phase-based approach, where a large number of corresponding points are

automatically detected without extracting image features. However, because of occlusion, image noise, projective distortion, etc., corresponding points for some reference points may not be estimated correctly. For such reference points, the proposed technique outputs wrong or unreliable corresponding points (generally known as outliers), which degrade the accuracy of measurement.

We propose an outlier detection technique using the peak value of the POC function as a measure of correspondence reliability. When the peak value of the POC function between the local image blocks, centered at the reference point $\boldsymbol{m}$ and at the corresponding point $\boldsymbol{q}(\boldsymbol{m})$, is below a certain threshold, $\boldsymbol{q}(\boldsymbol{m})$ is regarded as an outlier. This technique improves both reliability and accuracy of the overall 3D reconstruction.

We also implement an outlier correction technique in our system. Since our approach can detect a large number of high-accuracy reliable corresponding points (usually known as inliers), it is reasonable to assume a basic neighborhood constraint for natural object surfaces – neighboring points on natural object surfaces generally have smooth change in disparity [10]. Therefore, the true position of an outlier will have similar disparity to those of its neighboring points. The key idea is to assume a tentative disparity for an outlier; the tentative disparity is calculated by taking median of disparities of neighboring points. This tentative disparity is updated again with sub-pixel resolution by POC-based block matching. Procedure of the outlier detection and correction technique is given below:

**Procedure B**: "Outlier detection and correction":

**Input**:

- the left image $I$ and the right image $J$
- the corresponding points $\boldsymbol{q}(\boldsymbol{m}) \in C_J^{real}$ for all reference points $\boldsymbol{m} = (m_1, m_2) \in C_I^{int}$

**Output**:

- the corrected disparity vectors $\boldsymbol{d}_c(\boldsymbol{m}) \in \boldsymbol{R}^2$ for all reference points $\boldsymbol{m} = (m_1, m_2) \in C_I^{int}$, where we set $\boldsymbol{d}_c(\boldsymbol{m}) = (0, 0)$ for the outliers that cannot be corrected
- the corrected corresponding points $\boldsymbol{q}_c(\boldsymbol{m}) \in C_J^{real}$ for the reference points $\boldsymbol{m} = (m_1, m_2) \in C_I^{int}$ that have non-zero disparity $\boldsymbol{d}_c(\boldsymbol{m}) \neq (0, 0)$

**Procedure steps**:

**[Outlier detection]**
For every reference point $\boldsymbol{m} = (m_1, m_2) \in C_I^{int}$, do the following:
**Step 1**: Extract two image blocks from $I$ and $J$ such that the blocks have their centers on $\boldsymbol{m}$ and $\boldsymbol{q}(\boldsymbol{m})$, respectively. Estimate the peak value $\alpha$ of the POC function between the two image blocks as described in Sect. 2.1. Compare $\alpha$ with $\alpha_{th}$ to verify reliability of the correspondence between $\boldsymbol{m}$ and $\boldsymbol{q}(\boldsymbol{m})$, where $\alpha_{th}$ is the threshold value. If $\alpha \geqq \alpha_{th}$, then consider $\boldsymbol{q}(\boldsymbol{m})$ as an inlier, and set $\boldsymbol{q}_c(\boldsymbol{m})$ and its disparity vector $\boldsymbol{d}_c(\boldsymbol{m})$ as

$$\boldsymbol{q}_c(\boldsymbol{m}) = \boldsymbol{q}(\boldsymbol{m}), \tag{15}$$

$$\boldsymbol{d}_c(\boldsymbol{m}) = \boldsymbol{m} - \boldsymbol{q}(\boldsymbol{m}). \tag{16}$$

On the other hand, if $\alpha < \alpha_{th}$, then consider $\boldsymbol{q}(\boldsymbol{m})$ as an outlier, and give the outlier label as

$$\boldsymbol{d}_c(\boldsymbol{m}) = (0, 0). \tag{17}$$

(Here, we assume the use of parallel cameras in stereo vision, and hence a point with zero disparity has the physical meaning of point at infinity.)

**[Outlier correction]**
For every reference point $\boldsymbol{m} = (m_1, m_2) \in C_I^{int}$ that has been labeled as an outlier, i.e., $\boldsymbol{d}_c(\boldsymbol{m}) = (0, 0)$, do the following **Step 2** and **Step 3**:
**Step 2**: Consider $5 \times 5$ neighborhood points around $\boldsymbol{m}$. For these points, calculate median values of their horizontal and vertical disparities, which are denoted by $d_1^{med}$ and $d_2^{med}$, respectively. Using the median disparities, we calculate the tentative disparity vector $\boldsymbol{d}'(\boldsymbol{m})$ and the tentative corresponding point $\boldsymbol{q}'(\boldsymbol{m})$ as

$$\boldsymbol{d}'(\boldsymbol{m}) = (d_1^{med}, d_2^{med}), \tag{18}$$

$$\boldsymbol{q}'(\boldsymbol{m}) = \boldsymbol{m} - \boldsymbol{d}'(\boldsymbol{m}). \tag{19}$$

**Step 3**: Using $\boldsymbol{q}'(\boldsymbol{m})$ as an initial estimate of the correspondence, perform **[Sub-pixel estimation]** in **Procedure A**, and obtain an improved estimate for the corresponding point $\boldsymbol{q}'(\boldsymbol{m})$ with sub-pixel resolution. Let $\alpha'$ be the peak value of the POC function for the improved correspondence. If $\alpha' \geqq \alpha_{th}$, then the result of sub-pixel estimation is considered to be reliable, and the corrected corresponding point and disparity are set as

$$\boldsymbol{q}_c(\boldsymbol{m}) = \boldsymbol{q}'(\boldsymbol{m}), \tag{20}$$

$$\boldsymbol{d}_c(\boldsymbol{m}) = \boldsymbol{m} - \boldsymbol{q}'(\boldsymbol{m}). \tag{21}$$

Otherwise, if $\alpha' < \alpha_{th}$, consider the correspondence $\boldsymbol{q}'(\boldsymbol{m})$ as unreliable and the outlier label for $\boldsymbol{d}_c(\boldsymbol{m})$ remains unchanged as

$$\boldsymbol{d}_c(\boldsymbol{m}) = (0, 0). \tag{22}$$

For 3D reconstruction, we use only inlier correspondence pairs $(\boldsymbol{m}, \boldsymbol{q}_c(\boldsymbol{m}))$ for which $\boldsymbol{d}_c(\boldsymbol{m}) \neq (0, 0)$. □

Applying the outlier detection/correction procedure, we can obtain a set of high-accuracy corresponding points for 3D reconstruction. Note here that in conventional stereo vision systems, the use of epipolar constraint [8] is essential in correspondence search as well as outlier detection. The epipolar constraint is defined by the "fundamental matrix" $F$, which is esimated by camera calibration in advance. In our approach, on the other hand, correspondence search and outlier detection are done using the POC function without epipolar constraint, and hence we do not need to know the fundamental matrix of stereo cameras in advance. On the contratry, we can even use the high-accuracy corresponding

points to estimate the fundamental matrix itself. This property may be useful for such applications as 3D reconstruction from image sequences and other applications requiring self-calibration. See Appendix for further discussion.

## 3. Multi-Camera 3D Measurement System

We implement a multi-camera passive 3D measurement system based on the proposed correspondence search and outlier detection/correction techniques. In this section, the multi-camera system is presented in details, where Sects. 3.1 and 3.2 describe the system architecture and the procedure of 3D measurement, respectively.

### 3.1 System Architecture

Figure 2 shows the proposed multi-camera 3D measurement system consisting of six calibrated cameras into three pairs of camera heads (i.e., left, front and right camera heads). Here, correspondence is taken only between every two cameras of the same camera head, where the two cameras are parallel to each other with a very narrow baseline (around 50 mm).

In general, the following two features must be considered in designing the optimal camera configuration for a 3D measurement system for dense surface reconstruction:

- 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 correspondence is very difficult and is limited to a small number of edge points. This may be unacceptable in many practical applications of 3D measurement.

In our multi-camera system, we adopt narrow-baseline camera alignment, where the problem of low accuracy in 3D measurement is overcome by introducing the sub-pixel correspondence search technique. The use of phase-based image matching makes possible to achieve fully automatic high-accuracy 3D measurement with a narrow-baseline stereo vision system. This paper is the first demonstration of a passive 3D measurement system, whose reconstruction accuracy is comparable with that of practical active 3D scanners based on structured light projection.

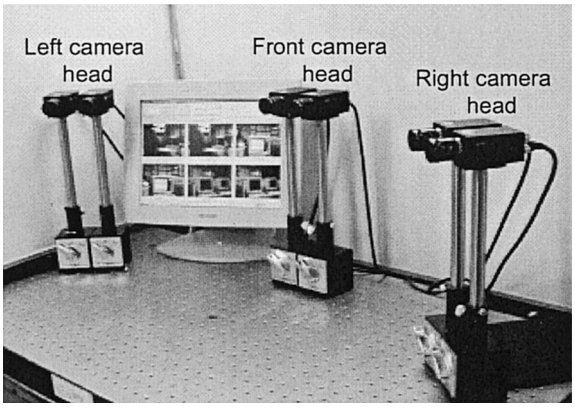


**Fig. 2** Multi-camera 3D measurement system.

In our system, we use simple off-the-shelf CCD cameras (JAI CVM10, 640 × 480 pixels, monochrome, 256 grey levels with a C-mount lens VCL-16WM), and a capture board (Coreco Imaging Technology, IC-PCI with AM-STD-RGB) for simultaneous imaging from the six cameras. Images are captured by the system in ambient light, and a volume of around 1500 (W)×1000 (H)×400 (D) mm$^3$ is usually adopted for measurement. Distance of target objects from the cameras is set around 800–1200 mm, and the camera focus are adjusted to the distance.

### 3.2 3D Measurement Procedure and System Parameters

The measurement procedure is divided into four steps as follows:

• **Camera calibration**: Camera calibration is done in order to determine the projective matrices, which consist of the basic camera parameters, such as the relative rotation/translation of cameras with respect to the world coordinate, focal lengths, image centers, etc. Such parameters are needed to calculate the 3D coordinates of a point [2], [8].

• **Correspondence search**: Correspondence search is done using Procedure A. The technique makes possible to search correspondence for any given reference point, and in our system we usually determine correspondence for every 5th point with respect to the horizontal and the vertical image coordinates. For POC-based block matching, we set the parameters as: (i) the block size is $N_1 \times N_2 = 33 \times 33$ pixels (weighted by 2D Hanning window), (ii) the spectral weighting function is 2D Gaussian function with $\sigma^2 = 0.5$, (iii) number of fitting points for the sub-pixel displacement estimation is $5 \times 5$, (iv) number of layers for the coarse-to-fine search is 5.

• **Outlier detection and correction**: Outlier detection and correction is done using Procedure B. The block size for POC-based block matching is same as in the correspondence search described above. The threshold $\alpha_{th}$ for the peak value of POC function is 0.3.

• **3D reconstruction**: The projective matrices of the cameras and the corresponding points are used to reconstruct the real-world 3D coordinates, where 4000 to 5000 points are reconstructed in our system.

In our system, we calibrate all the six cameras regarding one unique predefined world coordinate system. Therefore, the 3D points obtained by all the camera heads are reconstructed on that same predefined world coordinate, although the three camera heads (left, front and right) reconstruct 3D points independently. As a result, no further operation is needed to combine the 3D points. Please note that, the objective of using multiple cameras in our system is to cover larger area of the object surface for 3D reconstruction.

## 4. Experiments and Evaluations

In this section, we describe a set of experiments using simple and well characterized physical surfaces to evaluate the accuracy of the proposed 3D measurement system. In addition, a human face—a typical example of free form objects—is measured to demonstrate the system's capability of high-quality dense 3D reconstruction.

### 4.1 Impacts of Sub-Pixel Correspondence Search and Outlier Correction

We first evaluate the effects of the proposed techniques: the sub-pixel correspondence search (Procedure A) and the outlier detection and correction (Procedure B), on the quality of 3D measurement. For evaluation, we consider three different methods of 3D measurement:

- **Method I** employs a simplified version of Procedure A (where we skip the steps of [Sub-pixel estimation]), but does not employ Procedure B.
- **Method II** employs Procedure A, but does not employ Procedure B.
- **Method III** employs both Procedure A and Procedure B.

At first, we evaluate the accuracy of 3D reconstruction using two reference objects of geometrically regular shapes—a solid plane (a flat wooden board) of size $180 \times 150\,\mathrm{mm}^2$ and a solid sphere (a bowling ball) of radius 108.45 mm—both having sufficient machining accuracy. In order to evaluate measurement accuracy for the solid plane, we generate a best fitted plane for the measured points by the least-squares algorithm. Let $(x_i, y_i, z_i)$ be the reconstructed 3D points, where $i = 1, 2, \cdots, K$. The plane fitting is to minimize the following function:

$$P(a, b, c) = \sum_{i=1}^{K} (z_i - ax_i - by_i - c)^2, \tag{23}$$

where $(a, b, c)$ are fitting parameters. Accuracy of measurement is evaluated by the fitting error. Similar experiment is carried out using the solid sphere as a reference object. For sphere fitting, we minimize the following function:

$$S(c_1, c_2, c_3, r) = \sum_{i=1}^{K} \left( \sqrt{(x_i - c_1)^2 + (y_i - c_2)^2 + (z_i - c_3)^2} - r \right)^2, \tag{24}$$

where $(c_1, c_2, c_3, r)$ are fitting parameters.

We use here only front camera pair for 3D measurement, where the camera baseline is 50.84 mm (estimated by the camera calibration) and the distance between the camera pair and the reference objects is around 900 mm. Table 1 compares the errors in 3D measurement by the Method I, II and III, when we use the solid plane as a reference object. Table 2 summarizes the similar experiment, when we use the solid sphere as a reference object.

**Table 1** Errors [mm] in 3D measurement of a plane object.

| | RMS error | Maximum error |
|---|---|---|
| Method I | 0.87 | 13.93 |
| Method II | 0.61 | 12.81 |
| Method III | 0.42 | 1.23 |

**Table 2** Errors [mm] in 3D measurement of a sphere object.

| | RMS error | Maximum error |
|---|---|---|
| Method I | 1.59 | 20.19 |
| Method II | 0.63 | 18.52 |
| Method III | 0.55 | 4.12 |

These results show that the proposed sub-pixel correspondence technique contributes to reducing the RMS (Root Mean Square) error significantly, and the outlier correction technique is effective for reducing both the RMS error and the maximum error. Figure 3 shows the 3D surfaces of the plane object and the sphere object reconstructed by the Methods I, II and III, respectively, which clearly visualizes the significant impacts of the proposed technique. We can observe that the Method I tends to produce stepwise error in the 3D data, and even Method II produces scattered points (i.e., outliers). The Method III, on the other hand, successfully reconstructs the smooth surfaces of the reference objects.

Table 3 summarizes the total number of reference points for which correspondence search is carried out, the number of detected outliers, the number of corrected outliers, and the resulting number of reconstructed 3D points. For both plane and sphere objects, about 8–9% of the total reference points are classified into outliers, but around 80–90% of the outliers are corrected. Thus the outlier correction technique makes possible to increase the number of reconstructed 3D points, and thus the technique may be useful for many applications where dense reconstruction of 3D surfaces is necessary.

### 4.2 Multi-Camera System Performance

In this section, we use the three camera heads (i.e., left, front and right camera heads) simultaneously and evaluate the overall accuracy of 3D measurement by changing the position of the reference objects. At first, the object is placed around 900 mm away from the cameras, and images are captured by all the cameras. Then, a micro-stage (with $7\,\mu$m displacement error) is used to move the object 4 times, where each time the displacement is 5 mm and images are taken at each position. Thus, we have a set of reconstructed object surfaces at 5 different positions. Let us denote the 5 different positions of the object by P1, P2, P3, P4 and P5 in order, where the distance of every movement: P1 $\rightarrow$ P2, P2 $\rightarrow$ P3, P3 $\rightarrow$ P4 or P4 $\rightarrow$ P5 is 5 mm. Table 4 and Table 5 summarize the RMS fitting errors of the reconstructed object surfaces at the positions P1–P5, where RMS errors of the 3D data from the three camera heads are given in dif-

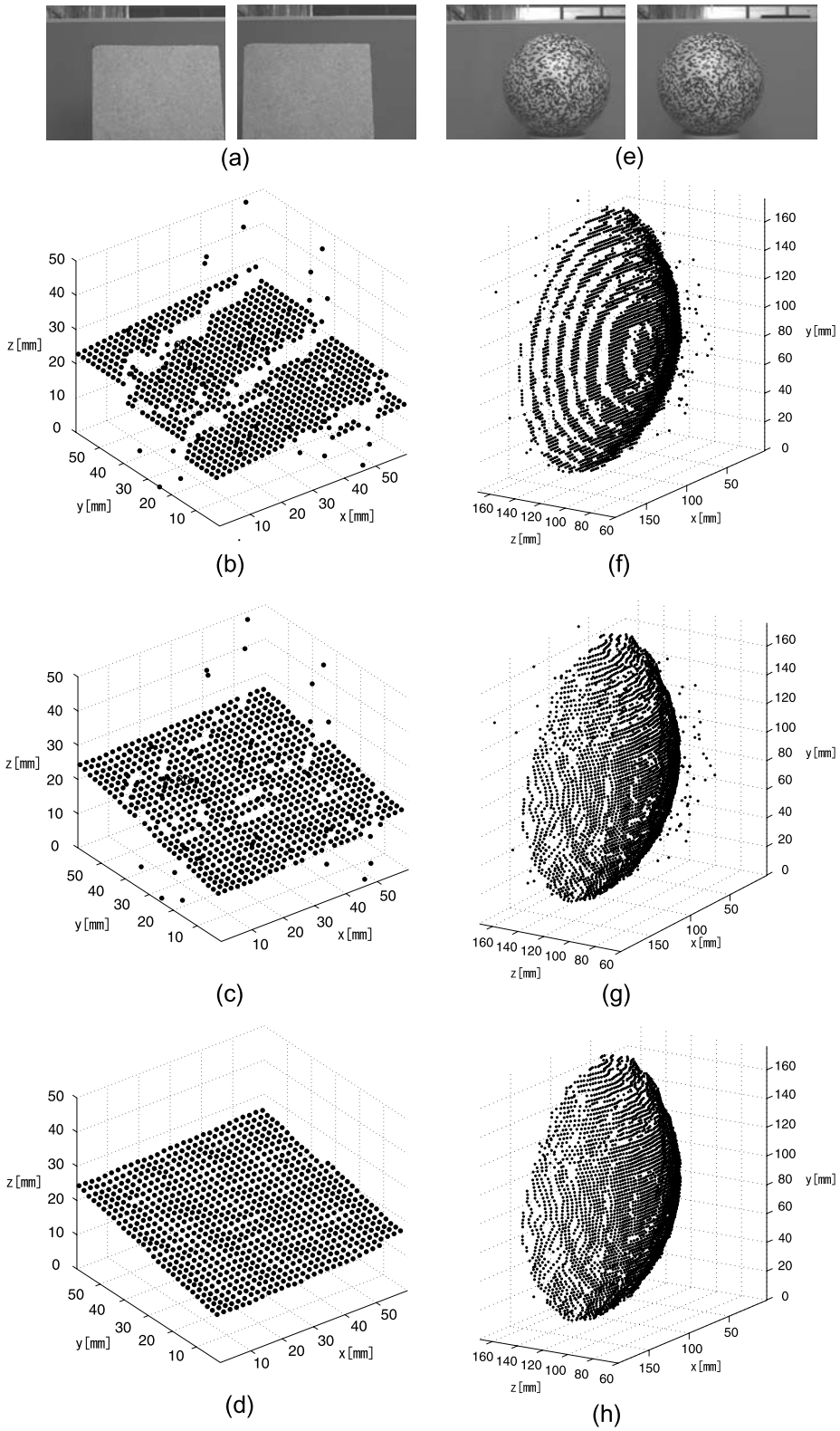


**Fig. 3** Impacts of sub-pixel correspondence and outlier correction: (a) 2D images of a plane object, (b) reconstructed plane using Method I, (c) reconstructed plane using Method II, (d) reconstructed plane using Method III, (e)–(h) similar images for a sphere object; here, a portion of the reconstructed object is presented for convenience in visualization.

**Table 3** Number of points in 3D reconstruction.

| | Plane | Sphere |
|---|---|---|
| # of points in correspondence search | 3898 | 4928 |
| # of detected outliers | 342 | 391 |
| # of corrected outliers | 309 | 312 |
| # of reconstructed points | 3865 | 4849 |

**Table 4** RMS errors [mm] in 3D measurement of a plane object at different positions.

| Position | Left camera | Front camera | Right camera | Combined |
|---|---|---|---|---|
| P1 | 0.35 | 0.42 | 0.35 | 0.40 |
| P2 | 0.35 | 0.46 | 0.35 | 0.40 |
| P3 | 0.38 | 0.43 | 0.36 | 0.40 |
| P4 | 0.36 | 0.43 | 0.37 | 0.40 |
| P5 | 0.38 | 0.41 | 0.36 | 0.39 |

**Table 5** RMS errors [mm] in 3D measurement of a sphere object at different positions.

| Position | Left camera | Front camera | Right camera | Combined |
|---|---|---|---|---|
| P1 | 0.64 | 0.55 | 0.58 | 0.65 |
| P2 | 0.54 | 0.53 | 0.57 | 0.65 |
| P3 | 0.54 | 0.60 | 0.54 | 0.65 |
| P4 | 0.67 | 0.58 | 0.58 | 0.64 |
| P5 | 0.70 | 0.47 | 0.56 | 0.70 |

**Table 6** Errors [mm] in 3D movement estimation for a plane object.

| Movement | Left camera | Front camera | Right camera | Combined |
|---|---|---|---|---|
| P1 → P2 | 0.06 | 0.06 | 0.00 | 0.05 |
| P2 → P3 | 0.08 | 0.08 | 0.02 | 0.05 |
| P3 → P4 | 0.03 | 0.09 | 0.01 | 0.03 |
| P4 → P5 | 0.09 | 0.10 | 0.05 | 0.10 |

ferent columns as well as the overall RMS error when all the 3D data are combined in a common world coordinate system. The RMS error ranges from 0.35 mm to 0.46 mm for the plane object, and 0.53 mm to 0.70 mm for the sphere object, respectively.

We also evaluate accuracy of 3D movement estimation by calculating the displacements of the reference object (the plane or sphere) for the movements: P1 → P2, P2 → P3, P3 → P4 and P4 → P5. We compare the estimated displacement with the actual displacement of the object surfaces (5 mm each time with 7 $\mu$m displacement error of the micro-stage). For the plane object, we calculate the distance between every two adjacent fitted planes, and evaluate the displacement error as shown in Table 6, where the error ranges from 0.00 mm to 0.10 mm. For the sphere object, we calculate the distance between the centers of every two adjacent fitted spheres, and evaluate the displacement error as shown in Table 7, where the error ranges from 0.01 mm to 0.11 mm.

All these experiment results show that our proposed system reconstructs 3D objects with around 0.5 mm error,

**Table 7** Errors [mm] in 3D movement estimation for a sphere object.

| Movement | Left camera | Front camera | Right camera | Combined |
|---|---|---|---|---|
| P1 → P2 | 0.10 | 0.04 | 0.01 | 0.10 |
| P2 → P3 | 0.03 | 0.08 | 0.06 | 0.08 |
| P3 → P4 | 0.07 | 0.03 | 0.09 | 0.01 |
| P4 → P5 | 0.11 | 0.10 | 0.03 | 0.02 |

where the distance between the object and the cameras is around 1 m. The accuracy is considered to be very high for passive 3D measurement system without using structured light projection or laser scanning.

Our case studies show that the measurement accuracy of active 3D scanners ranges from 0.05 mm to 1 mm in general. Typical examples of active scanners include Danae-S (1 mm) of NEC, FastSCAN (0.5–0.75 mm) of Polhemus, Aurora (1 mm) and Polaris (0.35 mm) of NDI, and VIVID 9i (0.05 mm) of Konica Minolta. The reason why the measurement accuracy varies within such a wide range is that systems using laser scanners (e.g., VIVID 9i) exhibit better performance compared to systems using structured light projection (e.g., Danae-S, Aurora, etc.). We conclude that our system's measurement accuracy (~ 0.5 mm) is comparable with that of active scanners using structured light projection.

### 4.3 3D Face Reconstruction

We demonstrate here the 3D measurement of a typical example of free form objects—a human face. Figure 4 compares the quality of 3D surfaces produced by the Methods I, II and III, where the 3D data from left, front and right camera heads are displayed independently. We can observe significant impacts of the sub-pixel correspondence search and outlier correction techniques on the quality of reconstructed surfaces. We cannot evaluate the accuracy of 3D reconstruction directly, since the precise dimensions of the face are not known. Therefore, we verify the reliability of the reconstruction by using epipolar constraint, where we evaluate the distance of every corresponding point from its epipolar line—the epipolar line is computed by using the fundamental matrix obtained in camera calibration. In an ideal situation, every corresponding point should be on its epipolar line and the distance should be zero. Table 8 summarizes the RMS errors in the evaluated distance for the two reference objects and the human face. As for the Method III, the RMS error is 0.16 pixels for the plane, 0.35 pixels for the sphere and 0.27 pixels for the face. Thus, we can conclude that the reconstructed 3D face has high accuracy comparable with the reconstructed reference objects.

Finally, Fig. 5 displays the combined 3D data of the human face from different view angles. To the best of the authors' knowledge, the quality of 3D reconstruction seems to be one of the best that is available with passive 3D measurement techniques reported to date. The result of this paper clearly suggests a potential possibility of our proposed approach to be widely used in many computer vision appli-

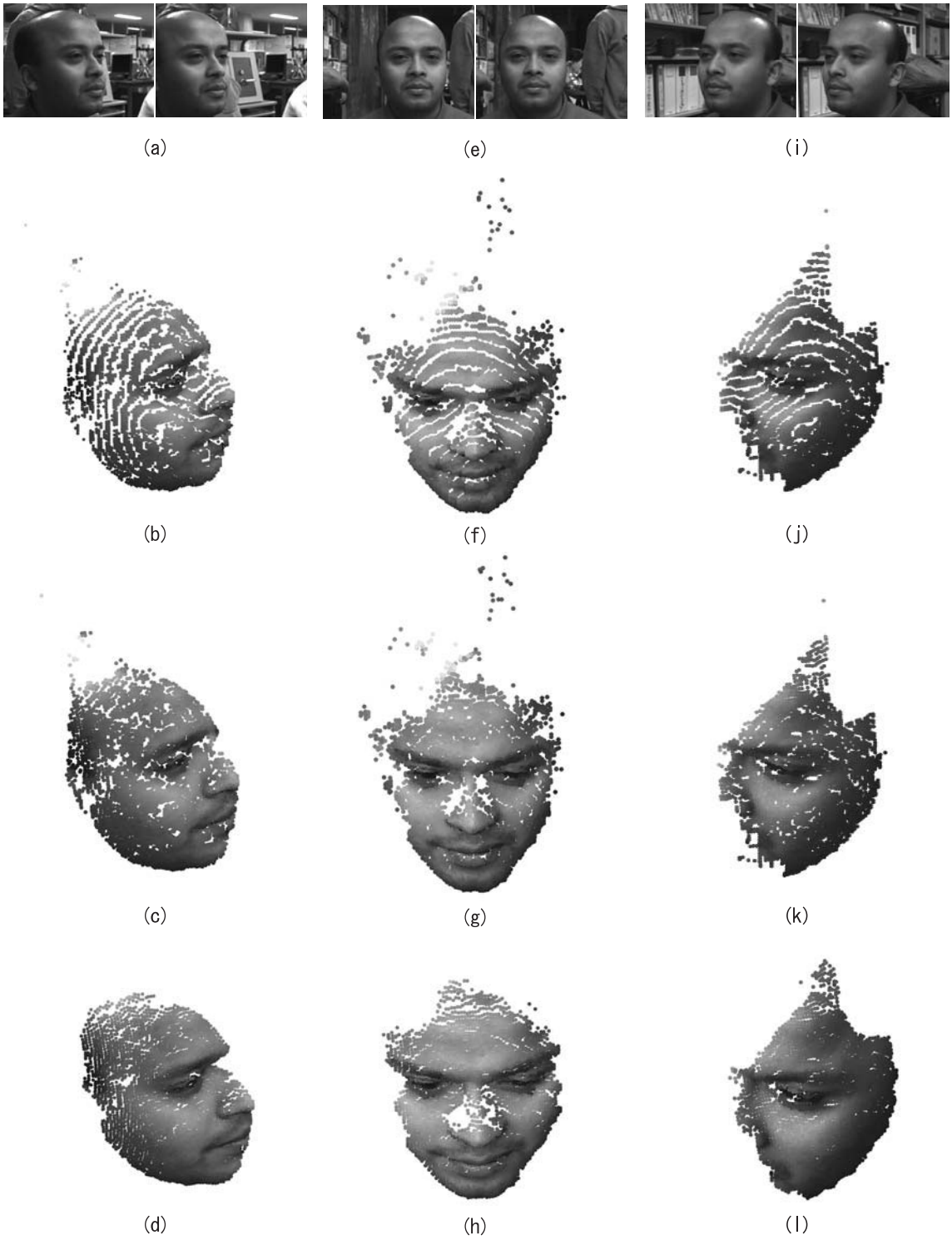


**Fig. 4** Impacts of sub-pixel correspondence search and outlier correction in 3D face reconstruction: (a) 2D images from the left camera head, (b) reconstructed 3D data using Method I for the left camera head, (c) reconstructed 3D data using Method II for the left camera head, (d) reconstructed 3D data using Method III for the left camera head, (e)–(h) similar images for the front camera head, (i)–(l) similar images for the right camera head.

**Table 8** RMS errors [pixel] in the distances between the corresponding points and their epipolar lines.

| | Plane | Sphere | Face |
|---|---|---|---|
| Method I | 0.33 | 0.65 | 0.79 |
| Method II | 0.25 | 0.64 | 0.48 |
| Method III | 0.16 | 0.35 | 0.27 |


**Fig. 5** Reconstructed 3D face data from different view angles.

cations, e.g., face recognition, biometrics, human interface, virtual reality, etc. In the conference paper [11], we presented an application of the proposed technique to 3D face recognition for biometric authentication.

At this moment, all the computations are done using MATLAB, where the reconstruction of 5000 points takes around 60 seconds. The bottleneck of the computation is the correspondence search process. However, the correspondence search can be performed for each pixel independently in our system, and thus, the computation time can be drastically reduced by introducing parallel processing. The goal of our current research project is to develop a real-time passive 3D capture system based on parallel DSP processors.

## 5. Conclusions

In this paper, we have proposed a high-accuracy multi-camera passive 3D measurement system, which employs (i) a phase-based sub-pixel correspondence search technique and (ii) an outlier detection and correction technique. We have successfully implemented a passive 3D measurement system with reconstruction accuracy comparable to practical 3D scanners using structured light projection. Through some experimental evaluations, we show that the system achieves sub-mm ($\sim 0.5$ mm) accuracy in 3D measurement, even with narrow baseline ($\sim 50$ mm) stereo camera heads. In addition, we show that the system performs dense reconstruction of free form objects with high quality, which reflects its potential possibilities in many computer vision applications. A main goal of our current research project is to develop a real-time passive 3D capture system based on the proposed approach.

### References

[1] M. Petrov, A. Talapov, T. Robertson, A. Lebedev, A. Zhilyaev, and L. Polonskiy, "Optical 3-D digitizers: Bringing life to the virtual world," IEEE Comput. Graph. Appl., vol.18, pp.28–37, May/June 1998.
[2] O.D. Faugeras, Three Dimensional Computer Vision—A Geometric Viewpoint, MIT Press, 1993.
[3] M. Okutomi and T. Kanade, "A multiple-baseline stereo," IEEE Trans. Pattern Anal. Mach. Intell., vol.15, no.4, pp.353–363, April 1993.
[4] C.D. Kuglin and D.C. Hines, "The phase correlation image alignment method," Proc. Int. Conf. Cybernetics and Society, pp.163–165, 1975.
[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] G. Tognola, M. Parazzini, P. Ravazzani, F. Grandori, and C. Svelto, "3-D acquisition and quantitative measurements of anatomical parts by optical scanning and image reconstruction from unorganized range data," IEEE Trans. Instrum. Meas., vol.52, no.5, pp.1665–1673, Oct. 2003.
[8] G. Xu and Z. Zhang, Epipolar Geometry in Stereo, Motion and Object Recognition, Kluwer Academic Publishers, 1996.
[9] K. Kanatani, "3-D interpretation of optical flow by renormalization," Int. J. Comput. Vis., vol.11, no.3, pp.267–282, 1993.
[10] I.J. Cox, S.L. Hingorani, S.B. Rao, and B.M. Maggs, "A maximum likelihood stereo algorithm," Computer Vision and Image Understanding, vol.63, no.3, pp.542– 567, 1996.

[11] N. Uchida, T. Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi, "3D face recognition using passive stereo vision," Proc. IEEE Int. Conf. Image Processing, 2005.
[12] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.
[13] M. Pollefeys, R. Koch, M. Vergauwen, and L.V. Gool, "Automated reconstruction of 3D scenes from sequences of images," ISPRS Journal of Photogrammetry and Remote Sensing, vol.55, no.4, pp.251–267, 2000.

## Appendix: Fundamental Matrix Estimation

The correspondence search technique introduced in this paper estimates correspondence with sub-pixel accuracy. This property can be utilized in estimating the fundamental matrix with high accuracy [9], [12]. In this paper, we show a simple comparison between the fundamental matrix $\boldsymbol{F}_p$ estimated from the corresponding points obtained by our proposed approach, and the fundamental matrix $\boldsymbol{F}_c$ obtained by the camera calibration parameters. For comparison, we estimate the error by calculating the distances of the corresponding points from their epipolar lines defined by $\boldsymbol{F}_p$ (or $\boldsymbol{F}_c$). Figure A· 1(a) and (b) show the error regarding $\boldsymbol{F}_p$ and $\boldsymbol{F}_c$, respectively. We found that in both cases the RMS error is in a similar level, i.e., 0.30 pixels for $\boldsymbol{F}_p$ and 0.27 pixels for $\boldsymbol{F}_c$. The fundamental matrix $F_p$ can be used for self-calibration [12], which is utilized in applications like 3D reconstruction from image sequences [13]. In addition, our method is applicable to high accuracy correspondence matching in multi-baseline stereo systems [3], [12]. However, one drawback of this method is that it can be applied only to narrow-baseline systems, and thus, the scope of its application is limited.

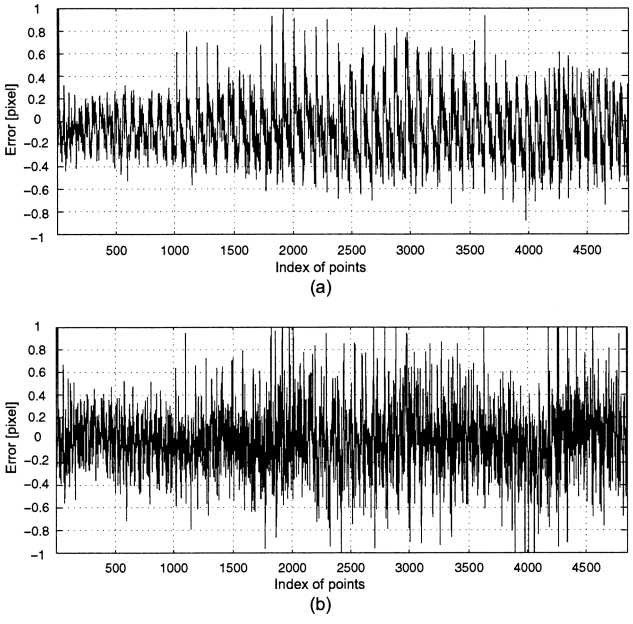


**Fig. A· 1** Comparison of the fundamental matrices $F_c$ (a) and $F_p$ (b) in terms of epipolar-line errors.

**Mohammad Abdul Muquit** received the B.E. degree in information engineering, and the M.S. degree in information sciences from Tohoku University, Sendai, Japan, in 2001 and 2003, respectively. He is currently working toward the Ph.D. degree. His research interests include computer vision and image processing. Mr. Muquit received the Student Award Best Paper Prize from IEEE Sendai Section in 2002.

**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.

**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 and 2001 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, and the IP Award at the 7th LSI IP Design Award in 2005.