A Sub-Pixel Correspondence Search Technique for Computer Vision Applications

Kenji TAKITA[†], *Student Member*, Mohammad Abdul MUQUIT[†a)], *Nonmember*, Takafumi AOKI[†], *Member*, and Tatsuo HIGUCHI[††], *Fellow*

SUMMARY   This paper presents a technique for high-accuracy correspondence search between two images using Phase-Only Correlation (POC) and its performance evaluation in a 3D measurement application. The proposed technique employs (i) a coarse-to-fine strategy using image pyramids for 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 proposed method makes possible to estimate the displacement between corresponding points with approximately 0.05-pixel accuracy when using 11×11-pixel matching windows. This paper also describes an application of the proposed technique to passive 3D measurement system.

key words: *phase-only correlation, phase correlation, image matching, sub-pixel matching, block matching*

1.   Introduction

Area-based image matching is an important fundamental task in a variety of image processing applications, such as stereo vision, motion analysis, image sequence analysis, etc. (see [1] for a good survey on this topic). Although in some applications pixel-level image matching may be adequate, image matching with sub-pixel accuracy is becoming essential in recent applications. In stereo vision, for example, if we can estimate the disparity between two stereo images with 1/10-pixel accuracy, we can then measure the distance of a target object with 10 times the accuracy of the stereo vision system using pixel-level image matching.

In response to this need, many methods have been developed to estimate the translational displacement between two images with high accuracy [1]–[4], and among them the methods using the *Phase-Only Correlation* (POC) function are highly effective due to their high accuracy and robust performance [5]. (Note that the POC function is sometimes called the "phase correlation function.") Our experimental observation shows that the methods using phase information in 2D Discrete Fourier Transform (2D DFT) exhibit better registration performance than the methods using SAD (Sum of Absolute Differences) in general. In our previous work, for example, we have proposed a high-accuracy image matching method based on the POC function [6], [7], which

employs (i) an analytical function fitting technique to estimate the 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. Through a set of experiments, we have demonstrated that this matching method can estimate the displacement between two images with 1/100-pixel accuracy when the image size is about 100×100 pixels.

In practice, there are many applications that require high-accuracy matching of smaller image blocks. Typical applications include stereo vision 3D measurement [8] and super-resolution image reconstruction [9]. These applications require high-accuracy image block matching for finding corresponding points in multiple images and for estimating the position of corresponding points with sub-pixel accuracy. However, the accuracy of the conventional matching methods, including our previously proposed method [7], degrades significantly as the image size decreases because the texture information contained in an image block decreases, correspondingly.

Addressing this problem, in this paper, we extend our previously proposed matching method [7] to improve the registration accuracy for small image blocks. The basic idea is to align the position of matching windows with sub-pixel accuracy in two image blocks. This *sub-pixel window alignment* technique is combined with a coarse-to-fine search technique to implement an efficient algorithm for finding the corresponding points in multiple images. Experimental evaluation shows that the proposed correspondence search algorithm makes possible to detect the position of corresponding points with approximately 0.05-pixel accuracy when using 11×11-pixel image blocks. This paper also describes an application of the proposed sub-pixel correspondence search algorithm to high-accuracy 3D measurement.

This paper is organized as follows. Section 2 summarizes the sub-pixel image matching method using the POC function. Section 3 proposes a new algorithm for finding correspondence between two images with very high accuracy. Section 4 shows the experimental evaluation of the matching accuracy of the proposed algorithm. Section 5 describes an application of the algorithm to high-accuracy 3D measurement and its performance evaluation. Section 6 concludes the paper.

2. High-Accuracy Image Matching Using POC

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 $N_2 = 2M_2 + 1$. Let $F(k_1, k_2)$ and $G(k_1, k_2)$ denote the 2D Discrete Fourier Transforms (2D DFTs) of the two images. $F(k_1, k_2)$ and $G(k_1, k_2)$ are given by

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

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

$$
\begin{aligned}
\hat{R}(k_1, k_2) &= \frac{F(k_1, k_2)\overline{G(k_1, k_2)}}{|F(k_1, k_2)\overline{G(k_1, k_2)}|} \\
&= e^{j\theta(k_1, k_2)}, \quad (3)
\end{aligned}
$$

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 Phase-Only Correlation (POC) function $\hat{r}(n_1, n_2)$ is the 2D Inverse Discrete Fourier Transform (2D IDFT) of $\hat{R}(k_1, k_2)$ and is given by

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

where $\sum_{k_1 k_2}$ denotes $\sum_{k_1 = -M_1}^{M_1} \sum_{k_2 = -M_2}^{M_2}$.

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 $n_2 = -M_2, \cdots, M_2$. 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

$$
\begin{aligned}
&\hat{r}(n_1, n_2) \\
&\cong \frac{\alpha}{N_1 N_2} \frac{\sin\{\pi(n_1 + \delta_1)\}}{\sin\{\frac{\pi}{N_1}(n_1 + \delta_1)\}} \frac{\sin\{\pi(n_2 + \delta_2)\}}{\sin\{\frac{\pi}{N_2}(n_2 + \delta_2)\}}, \quad (5)
\end{aligned}
$$

where $\alpha \leqq 1$.

We assume that $|F(k_1, k_2)| \neq 0$ and $|G(k_1, k_2)| \neq 0$ for all the frequency range in Eq. (3). When the images do not contain a certain frequency component, i.e., $|F(k_1, k_2)| = |G(k_1, k_2)| = 0$ for some (k_1, k_2), the phase $\theta(k_1, k_2)$ of this frequency component cannot be determined uniquely in Eq. (3) and its actual value depends on the implementation of the POC computation. Such indeterminate frequency components cause disturbance against the ideal peak model given by Eq. (5). In the extreme situation when the original image has no texture at all, its phase becomes indeterminate for all the frequency range. Assuming the indeterminate phase takes random value at every frequency component, no correlation peak could be observed. On the other hand, when the two images $f(n_1, n_2)$ and $g(n_1, n_2)$ are not the same, the phase $\theta(k_1, k_2)$ in Eq. (3) tends to take random values for every frequency component, and again no correlation peak appears. Thus, detecting the unreliable frequency components, where $|F(k_1, k_2)| \cong 0$ or $|G(k_1, k_2)| \cong 0$, is important for evaluating the overall reliability of POC-based image matching.

Listed below are important techniques for high-accuracy sub-pixel image matching [7].

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

We use Eq. (5)—the closed-form peak model of the POC function—directly for estimating the peak position by function fitting. By calculating the POC function for two images $f(n_1, n_2)$ and $g(n_1, n_2)$, we can obtain a data array of $\hat{r}(n_1, n_2)$ for each discrete index (n_1, n_2), where $n_1 = -M_1, \cdots, M_1$ and $n_2 = -M_2, \cdots, M_2$. It is possible to find the location of the peak that may exist between image pixels by fitting the function (5) to the calculated data array around the correlation peak, where α, δ_1, and δ_2 are fitting parameters. Figure 1 shows an example, where Eq. (5) is fitted to the data array of $\hat{r}(n_1, n_2)$.

(ii) Windowing to reduce boundary effects

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 edge in 2D DFT computation. We reduce the effect of dis-
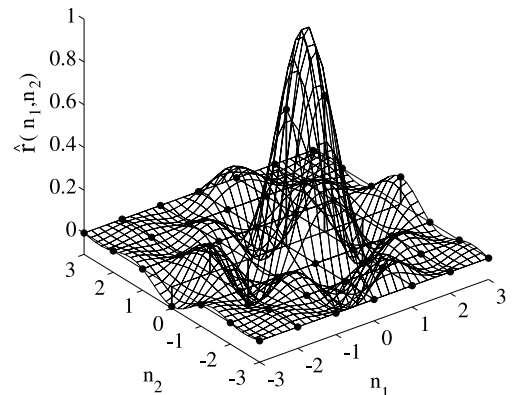

Fig. 1 Function fitting for estimating the peak position.

continuity at image border by applying 2D window function to the input images $f(n_1, n_2)$ and $g(n_1, n_2)$. For this purpose, we use 2D Hanning window defined as

$$w(n_1, n_2) = \frac{1 + \cos\left(\frac{\pi n_1}{M_1}\right)}{2} \frac{1 + \cos\left(\frac{\pi n_2}{M_2}\right)}{2}. \quad (6)$$

(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 simplest weighting function $H(k_1, k_2)$ is defined as

$$H(k_1, k_2) = \begin{cases} 1 & |k_1| \leqq U_1, |k_2| \leqq U_2 \\ 0 & \text{otherwise} \end{cases}, \quad (7)$$

where U_1 and U_2 are integers satisfying $0 \leqq U_1 \leqq M_1$ and $0 \leqq U_2 \leqq M_2$, respectively. The cross-phase spectrum $\hat{R}(k_1, k_2)$ is multiplied by the weighting function $H(n_1, n_2)$ when calculating the 2D IDFT. Then the modified $\hat{r}(n_1, n_2)$ will be given by

$$\begin{aligned} &\hat{r}(n_1, n_2) \\ &= \frac{1}{N_1 N_2} \sum_{k_1 k_2} \hat{R}(k_1, k_2) H(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2} \\ &\cong \frac{\alpha}{N_1 N_2} \frac{\sin\{\frac{V_1}{N_1}\pi(n_1 + \delta_1)\}}{\sin\{\frac{\pi}{N_1}(n_1 + \delta_1)\}} \frac{\sin\{\frac{V_2}{N_2}\pi(n_2 + \delta_2)\}}{\sin\{\frac{\pi}{N_2}(n_2 + \delta_2)\}}, \end{aligned}$$
$$(8)$$

where $V_1 = 2U_1 + 1$ and $V_2 = 2U_2 + 1$. When using the spectral weighting technique, Eq. (8) should be used for function fitting instead of Eq. (5).

3. Sub-Pixel Correspondence Search Algorithm

3.1 Algorithm Overview

The POC-based image matching technique described in the previous section is effective when the image size $N_1 \times N_2$ is relatively large [7]. We have demonstrated that the POC-based image matching can estimate the displacement between two images with 1/100-pixel accuracy when the image size is about 100×100 pixels.

In practice, there are many applications that require high-accuracy matching of small image blocks. Such applications include 3D measurement by stereo vision and super-resolution image reconstruction, where corresponding points in two images must be determined with high accuracy using block matching. In these applications, the accuracy of block matching (i.e., the accuracy of local displacement estimation) determines the total system performance directly.

However, the accuracy of the conventional matching methods, including our previously proposed method, degrades significantly as the image size decreases, since the texture information contained in an image block decreases. (See Sect. 4 for detailed experimental analysis.) Thus, it is difficult to implement a robust algorithm for high-accuracy correspondence search.

Addressing this problem, this section proposes a robust algorithm for finding the corresponding points between two images with high accuracy. The problem considered here can be stated as follows: Given an arbitrary pixel in a reference image, the problem is to find the position of the corresponding pixel in the second image with sub-pixel accuracy. The proposed algorithm combines the three basic techniques: (i) the POC-based image matching technique described in the previous section, (ii) a coarse-to-fine search technique for pixel-level correspondence estimation, and (iii) a sub-pixel window alignment technique for sub-pixel correspondence estimation.

Consider two images $I_0(n_1, n_2)$ and $J_0(n_1, n_2)$, which are taken at different times, or are taken from different sensors. We try to find a pair of matching points between the images. Let a pair of matching image points be $\boldsymbol{p}_0$ and $\boldsymbol{q}_0$, which exist in $I_0(n_1, n_2)$ and $J_0(n_1, n_2)$, respectively. Usually, one of the points, say $\boldsymbol{p}_0$, is specified in advance in the image $I_0(n_1, n_2)$. Then, the corresponding point $\boldsymbol{q}_0$ is searched in the second image $J_0(n_1, n_2)$.

If $\boldsymbol{p}_0$ and $\boldsymbol{q}_0$ are feature points, it may be possible to match these two points on the basis of their feature attributes. However, if $\boldsymbol{p}_0$ and $\boldsymbol{q}_0$ are not feature points, it is difficult to match them. The number of feature points to be used as candidates for corresponding points is often not sufficient in many applications. The area-based matching methods solves this problem by introducing neighboring areas of these two points ($\boldsymbol{p}_0$ and $\boldsymbol{q}_0$) for block matching. The POC-based technique described in Sect. 2 is useful for block matching.

Our algorithm consists of two stages: a stage for pixel-level correspondence estimation and a stage for sub-pixel-level correspondence estimation. In pixel-level estimation, we detect the corresponding point $\boldsymbol{q}_0$ with pixel-level accuracy using coarse-to-fine POC, so that the estimation error becomes less than 1 pixel. In the sub-pixel estimation stage, we recursively improve the sub-pixel accuracy of POC block matching by adjusting the location of the window function (Eq. (6)) with sub-pixel accuracy. We call this technique *sub-pixel window alignment*.

3.2 Correspondence Estimation with Pixel-Level Accuracy

The pixel-level estimation stage mainly employs the coarse-to-fine correspondence search technique, where some coarser versions of the original input images are created. The POC-based block matching starts at the coarsest image layer and the operation gradually moves to finer layers. Figure 2 shows an overview of the technique. Let $\boldsymbol{p}_l$ and $\boldsymbol{q}_l$
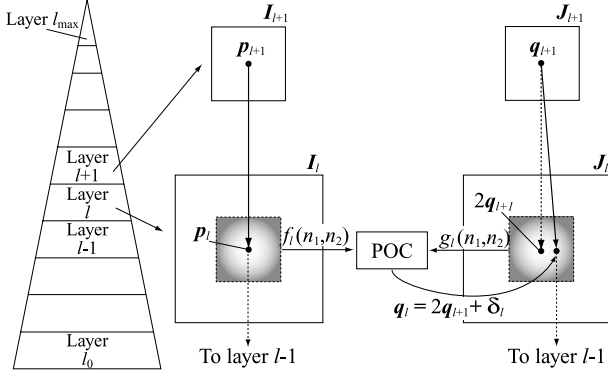
Fig. 2 Pixel-level correspondence search using a coarse-to-fine approach.

be coordinate vectors of matching image points at the l–th layer. The goal in the pixel-level estimation is to find $\boldsymbol{q}_0$ that corresponds to the given reference point $\boldsymbol{p}_0$ at the layer 0 (the original image layer) with pixel accuracy.

Procedure for pixel-level correspondence estimation
Input:
   images $I_0(n_1, n_2)$ and $J_0(n_1, n_2)$,
   reference point $\boldsymbol{p}_0$ in $I_0(n_1, n_2)$
Output:
   corresponding point $\boldsymbol{q}_0$ in $J_0(n_1, n_2)$
Step 1: For $l = 1, 2, \cdots, l_{max}$, create the l–th layer images $I_l(n_1, n_2)$ and $J_l(n_1, n_2)$, i.e., coarser versions of $I_0(n_1, n_2)$ and $J_0(n_1, n_2)$, recursively as follows:

$$I_l(n_1, n_2) = \frac{1}{4} \sum_{i_1=0}^{1} \sum_{i_2=0}^{1} I_{l-1}(2n_1 + i_1, 2n_2 + i_2),$$

$$J_l(n_1, n_2) = \frac{1}{4} \sum_{i_1=0}^{1} \sum_{i_2=0}^{1} J_{l-1}(2n_1 + i_1, 2n_2 + i_2).$$

In our experiments, we choose $l_{max} = 4$.

When the maximum disparity (the distance between corresponding points) in the original image layer $l = 0$ is given by D, the maximum disparity in the coarsest layer $l = l_{max}$ is given by $D/2^{l_{max}}$. Hence, the greater the disparity D, the more layers are required. In our experiments, we choose $l_{max} = 4$ such that the disparity at the coarsest layer is amply small compared to the size of input images for the calculation of the POC function.
Step 2: For every layer $l = 1, 2, \cdots, l_{max}$, calculate the coordinate $\boldsymbol{p}_l = (p_{l1}, p_{l2})$ corresponding to the original reference point $\boldsymbol{p}_0$ recursively as follows:

$$\boldsymbol{p}_l = \left\lfloor \frac{1}{2} \boldsymbol{p}_{l-1} \right\rfloor = \left(\left\lfloor \frac{1}{2} p_{l-1\,1} \right\rfloor, \left\lfloor \frac{1}{2} p_{l-1\,2} \right\rfloor \right), \tag{9}$$

where $\lfloor x \rfloor$ denotes the operation to round the element of x to the nearest integer towards minus infinity.
Step 3: We assume that $\boldsymbol{q}_{l_{max}} = \boldsymbol{p}_{l_{max}}$ in the coarsest layer. Let $l = l_{max} - 1$. (Actually, we can omit the creation of the l_{max}–th layer images.)

Step 4: From the l–th layer images $I_l(n_1, n_2)$ and $J_l(n_1, n_2)$, extract two sub-images (or image blocks) $f_l(n_1, n_2)$ and $g_l(n_1, n_2)$ with their centers on $\boldsymbol{p}_l$ and $2\boldsymbol{q}_{l+1}$, respectively. For accurate matching, the size of image blocks should be reasonably large. In our experiments, we use 31×31 image blocks.
Step 5: Estimate the displacement between $f_l(n_1, n_2)$ and $g_l(n_1, n_2)$ with pixel accuracy using POC-based image matching, which is the simplified version of the matching algorithm described in Sect. 2. We eliminate the technique (i) function fitting and (iii) spectral weighting for simplicity. Only the windowing technique (ii) using Eq. (6) is employed. The displacement between the two image blocks is estimated by detecting the position of the maximum value of the POC function $\hat{r}(n_1, n_2)$ with pixel accuracy. Let the estimated displacement vector be $\boldsymbol{\delta}_l$. The l–th layer correspondence $\boldsymbol{q}_l$ is determined as follows:

$$\boldsymbol{q}_l = 2\boldsymbol{q}_{l+1} + \boldsymbol{\delta}_l. \tag{10}$$

Step 6: Decrement the counter by 1 as $l = l - 1$ and repeat from **Step 4** to **Step 6** while $l \geq 0$. □

3.3   Correspondence Estimation with Sub-Pixel Accuracy

The sub-pixel estimation stage, on the other hand, employs the sub-pixel window alignment technique illustrated in Fig. 3.
Procedure for sub-pixel correspondence estimation
Input:
   images $I_0(n_1, n_2)$ and $J_0(n_1, n_2)$,
   reference point $\boldsymbol{p}_0$ in $I_0(n_1, n_2)$ given with pixel-level accuracy,
   corresponding point $\boldsymbol{q}_0$ in $J_0(n_1, n_2)$ given with pixel-level accuracy
Output:
   reference point $\boldsymbol{p}_{-1}$ in $I_0(n_1, n_2)$ given with sub-pixel accuracy,
   corresponding point $\boldsymbol{q}_{-1}$ in $J_0(n_1, n_2)$ given with sub-pixel accuracy
Step 1: From the original images $I_0(n_1, n_2)$ and $J_0(n_1, n_2)$, extract $N_1 \times N_2$ sub-images $f(n_1, n_2)$ and $g(n_1, n_2)$ with their centers on $\boldsymbol{p}_0$ and $\boldsymbol{q}_0$, respectively. For mathematical simplicity, we assume that the index ranges of the two sub-images are given by $n_1 = -M_1, \cdots, M_1$ and $n_2 = -M_2, \cdots, M_2$, and hence $N_1 = 2M_1 + 1$ and $N_2 = 2M_2 + 1$.
Step 2: Set the initial window functions for $f(n_1, n_2)$ and $g(n_1, n_2)$, denoted by $w_f(n_1, n_2)$ and $w_g(n_1, n_2)$, as $w_f(n_1, n_2) = w_g(n_1, n_2) = w(x_1, x_2)|_{x_1=n_1, x_2=n_2}$, where for real variables x_1 and x_2, the window function $w(x_1, x_2)$ is defined as

$$w(x_1, x_2) = \frac{1 + \cos\left(\dfrac{\pi x_1}{M_1}\right)}{2} \frac{1 + \cos\left(\dfrac{\pi x_2}{M_2}\right)}{2}.$$

Instead of the Hanning window function in the above equation, other window functions such as Hamming, Gaussian, and Kaiser can also be used.
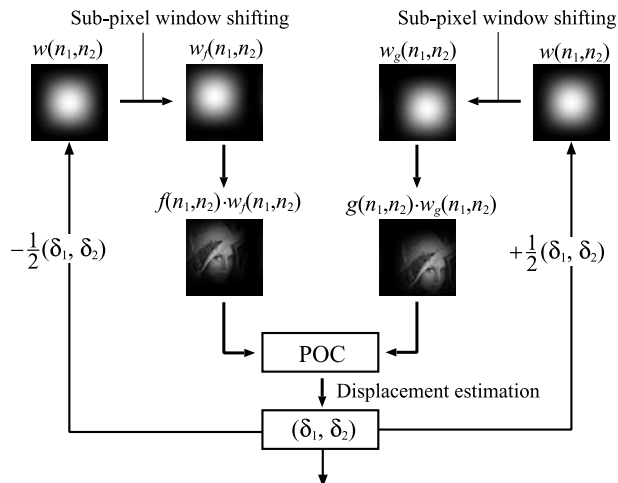
Sub-pixel window shifting   Sub-pixel window shifting

$w(n_1,n_2)$ $w_f(n_1,n_2)$ $w_g(n_1,n_2)$ $w(n_1,n_2)$

$f(n_1,n_2){\cdot}w_f(n_1,n_2)$ $g(n_1,n_2){\cdot}w_g(n_1,n_2)$

$-\frac{1}{2}(\delta_1,\delta_2)$ $+\frac{1}{2}(\delta_1,\delta_2)$

POC

Displacement estimation

(δ_1,δ_2)

Fig. 3 Sub-pixel correspondence search using a sub-pixel window alignment technique.

Step 3: Estimate the displacement between $f(n_1,n_2)$ and $g(n_1,n_2)$ with sub-pixel accuracy using the POC-based image matching described in Sect. 2. Let the estimated sub-pixel displacement vector be represented as $\boldsymbol{\delta} = (\delta_1, \delta_2)$, where $\delta_1, \delta_2 < 1$. In our experiments, 5×5 data points around the maximum peak of a correlation array are used in function fitting ((i) in Sect. 2). As for the spectral weighting technique ((iii) in Sect. 2), the parameters U_1 and U_2 of the low-pass-type weighting function $H(n_1, n_2)$ defined in Eq. (7) are set as $U_1 = \lceil M_1/2 \rceil$ and $U_2 = \lceil M_2/2 \rceil$, where $\lceil x \rceil$ denotes the operation to round the element of x to the nearest integer towards infinity. These parameters are optimized empirically.

Step 4: Update the window functions $w_f(n_1,n_2)$ and $w_g(n_1,n_2)$ by shifting their centers as

$$w_f(n_1,n_2) = w\left(x_1 + \frac{\delta_1}{2}, x_2 + \frac{\delta_2}{2}\right)\Big|_{x_1=n_1,x_2=n_2},$$

$$w_g(n_1,n_2) = w\left(x_1 - \frac{\delta_1}{2}, x_2 - \frac{\delta_2}{2}\right)\Big|_{x_1=n_1,x_2=n_2}.$$

Step 5: Repeat from **Step 3** to **Step 4** for specific times. This is called the "sub-pixel window alignment" technique, which gradually reduces the sub-pixel error in the estimated displacement vector. In our experiments, the number of iterations is 5.

Even though we did not experience any convergence problems in our experiments, we cannot ensure that the calculated displacement (δ_1, δ_2) converges. In our experiments, the displacement (δ_1, δ_2) converges after about 2 iterations. We have experimented with other window functions such as Hamming, Gaussian, and Kaiser, and found that the convergence property is almost the same. In this paper, we use ample iteration time (5 iterations) for reliable estimation. Hence, the computation cost of the sub-pixel window alignment technique becomes 5 times as high as that of simple POC matching without using the technique. However, we can reduce the number of iterations from 5 to 2 when we need to reduce computation time.

Step 6: As a result, the final matching points are estimated with subpixel accuracy as follows:

$$\boldsymbol{p}_{-1} = \boldsymbol{p}_0 - \frac{\boldsymbol{\delta}}{2},$$

$$\boldsymbol{q}_{-1} = \boldsymbol{q}_0 + \frac{\boldsymbol{\delta}}{2}.$$

□

Note that in some applications we must keep the reference point unchanged, i.e., $\boldsymbol{p}_{-1} = \boldsymbol{p}_0$. In this case, we use the following update equations for window functions:

$$w_f(n_1, n_2) = w(x_1, x_2)|_{x_1=n_1,x_2=n_2},$$

$$w_g(n_1, n_2) = w(x_1 - \delta_1, x_2 - \delta_2)|_{x_1=n_1,x_2=n_2}.$$

Accordingly, the final matching points are given by

$$\boldsymbol{p}_{-1} = \boldsymbol{p}_0,$$

$$\boldsymbol{q}_{-1} = \boldsymbol{q}_0 + \boldsymbol{\delta}.$$

The key point of our proposed technique is to align the image areas for calculation of the POC function so as to improve the sub-pixel accuracy of POC matching. Note that we can consider two possible ways of performing sub-pixel alignment of image areas: image shift and window shift. Sub-pixel image shift requires interpolation techniques, and its accuracy is limited by the accuracy of interpolation. It is much easier to perform sub-pixel shift of the window function, since the window function is defined on the continuous variables (x_1, x_2).

4. Experimental Evaluation of Correspondence Accuracy

In this section, we describe a set of experiments for evaluating the accuracy of the proposed correspondence search algorithm. Figure 4 shows the experimental system. Images are captured with a CCD camera (JAI CVM10 with Nikon Rayfact TM2514B1 lens), which is mounted on a micro stage that allows precise alignment of the camera position. The camera is fixed with an angle of approximately 30 degrees against the direction of micro stage movement so that the images taken with the camera have displacements in both horizontal and vertical directions. The target we use in the experiments is a wooden board with rich texture. The board is fixed 45 cm away from the camera in such a way that it is almost parallel to the focal plane of the CCD camera. The micro stage is moved horizontally 60 times, and the length of each movement is 15 μm. Images of the board are taken with the camera at each position. Each image is the size of 640×480 pixels, and captures a $10.5\,\text{cm} \times 8\,\text{cm}$ area of the wooden board.

In general, for high-accuracy sub-pixel correspondence of small areas, a high S/N ratio is desirable, because the information contained in an image decreases with the image size. In our experiments, the S/N ratio of the images taken with the CCD camera is not so high, and hence we take 30 sequential images of the board at each position, where
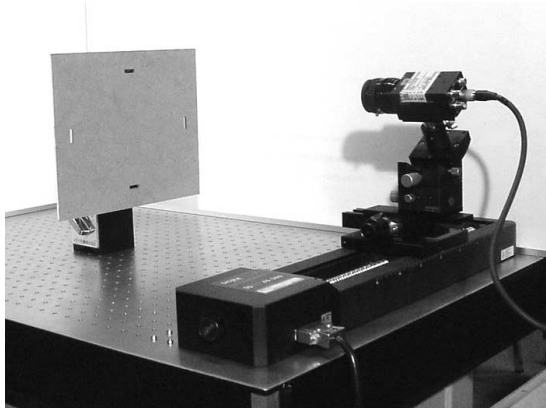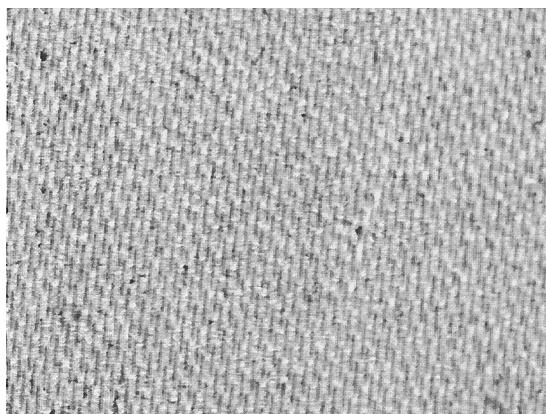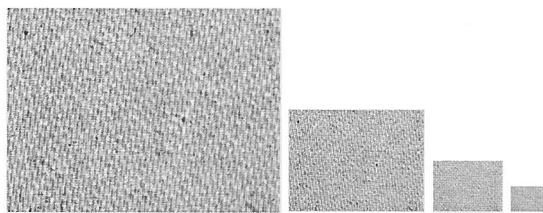
**Fig. 4**   Experimental system.

(a)


(b)   (c)   (d)   (e)

Fig. 5   Original image and the coarse images: (a) $I_0(n_1, n_2)$, (b) $I_1(n_1, n_2)$, (c) $I_2(n_1, n_2)$, (d) $I_3(n_1, n_2)$, and (e) $I_4(n_1, n_2)$.

the 30 static frames are averaged to improve image quality. This step does not necessarily need to be performed. However, it is effective for improving the sub-pixel matching accuracy when images taken with the cameras have low S/N ratios. Of course, it is possible to perform high-quality image matching without averaging by using low-noise CCD cameras, when applications require higher accuracy.
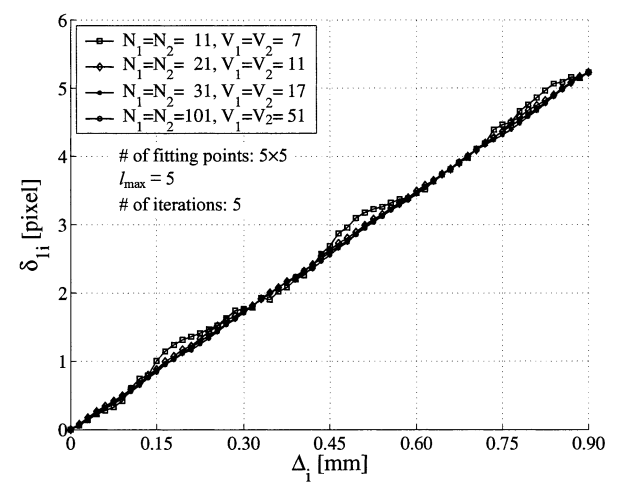
The reference image $I_0(n_1, n_2)$ is taken before moving the camera stage, where we choose a reference point $\boldsymbol{p}_0$ in advance. Figure 5 shows an original image $I_0(n_1, n_2)$ taken with the camera and four coarse images $I_l(n_1, n_2)$, $l = 1, \cdots, 4$, created from $I_0(n_1, n_2)$ in the pixel-level correspondence estimation. At every camera position i (=
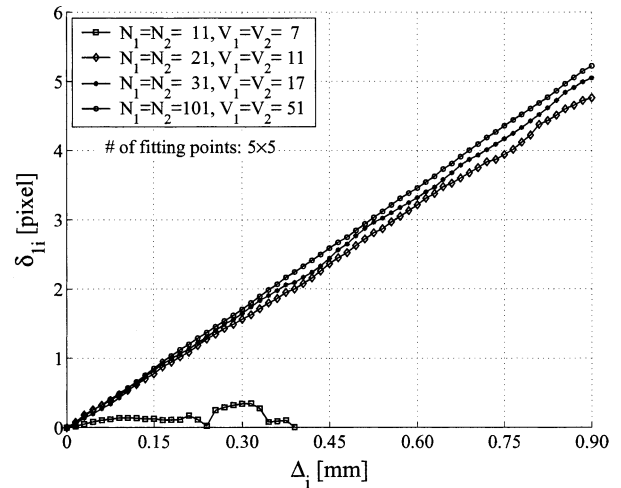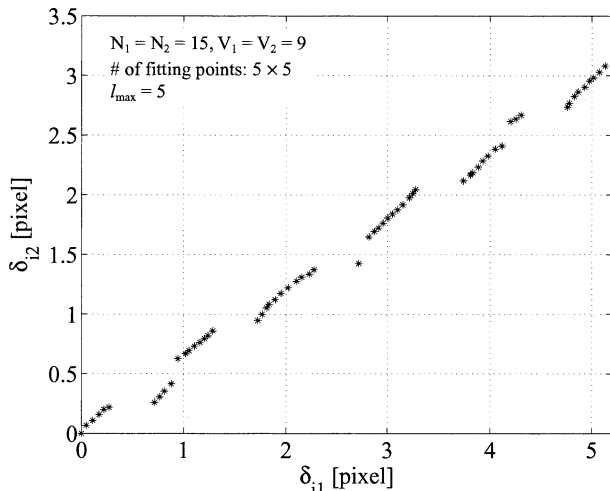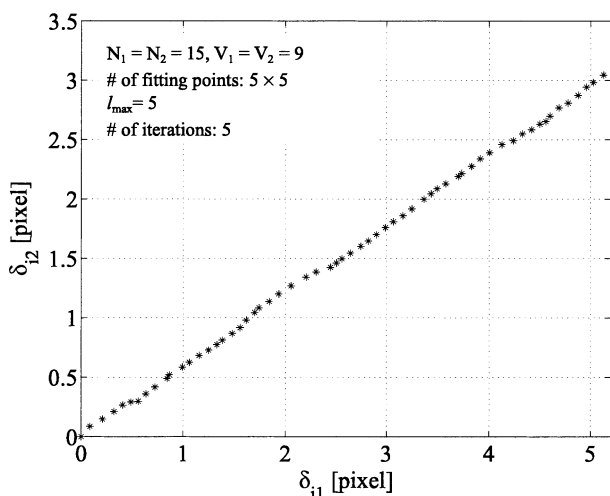

(a)
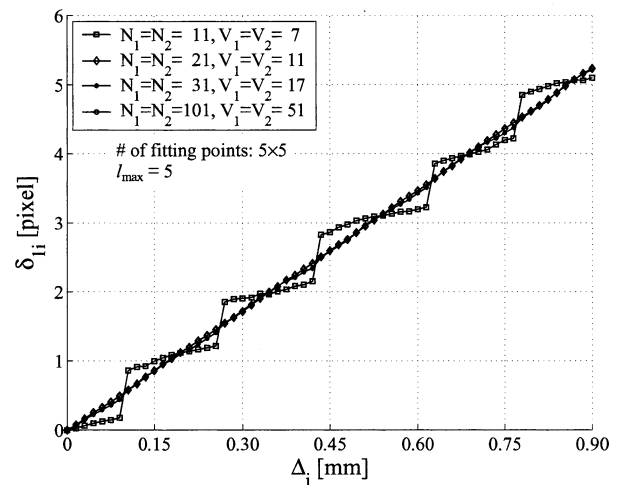
(b)

(c)

Fig. 6   Estimated displacements δ_{i1} in horizontal direction: (a) simple POC-based image matching, (b) algorithm A, and (c) algorithm B.

(a)

(b)

Fig. 7 Estimated displacements $(\delta_{i1}, \delta_{i2})$: (a) algorithm A and (b) algorithm B.

$1, 2, \cdots, 60$), we search the corresponding point with sub-pixel accuracy. Let the displacement vector between the reference point in the initial image and the corresponding point in the i-th image be $\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2})$. We evaluate the performance of the image matching technique in the following manner. There are 60 sets of data containing the actual displacement Δ_i [mm] (i.e., the displacement of the micro stage) and the estimated displacement $(\delta_{i1}, \delta_{i2})$ [pixel]. Ideally, the estimated displacement δ_{i1} (or δ_{i2}) must be proportional to Δ_i. To evaluate the correspondence accuracy, we calculate a linear approximation $\delta_{im} = a_m \times \Delta_i$ ($m = 1, 2$) in a least-square sense, where a_m [pixel/mm] is a constant coefficient. In this experiment, the parameters a_1 and a_2 are evaluated as $a_1 \cong 5.80$ [pixel/mm] and $a_2 \cong 3.46$ [pixel/mm], respectively. The error ϵ_i in displacement estimation at the i–th position is calculated by
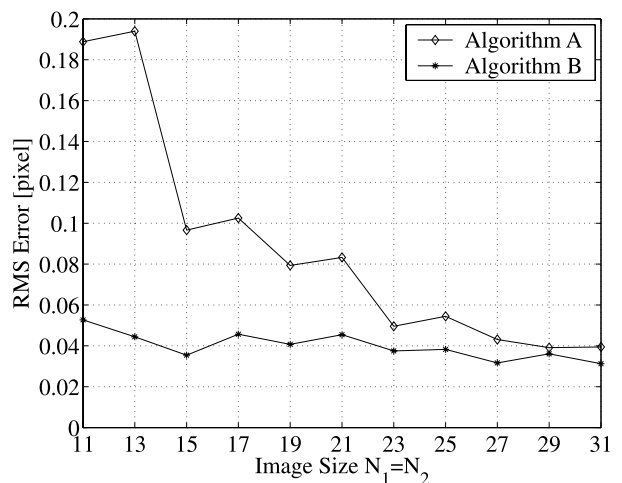

Fig. 8 RMS error versus image size.

$$\epsilon_i = \sqrt{(\delta_{i1} - a_1 \times \Delta_i)^2 + (\delta_{i2} - a_2 \times \Delta_i)^2}. \qquad (11)$$

We evaluate the following two new algorithms in comparison with simple POC-based image matching described in Sect. 2.

[Algorithm A] The sub-pixel correspondence search algorithm that employs the coarse-to-fine search technique (in Sect. 3.2) followed by simple sub-pixel POC matching (in Sect. 2).

[Algorithm B] The sub-pixel correspondence search algorithm that employs the coarse-to-fine search technique (in Sect. 3.2) followed by the sub-pixel window alignment technique (in Sect. 3.3).

Figure 6(a) shows the result of displacement estimation using simple POC-based block matching (in Sect. 2) for various sizes of image blocks, such as 11×11, 21×21, 31×31 and 101×101. The horizontal axis represents the actual displacement Δ_i [mm] of the camera, and the vertical axis represents the estimated horizontal displacement δ_{i1} [pixel]. Figures 6(b) and (c) show the results when using the new algorithms A and B, respectively. From Fig. 6(a), we can see that the simple algorithm tends to show some inconsistency as the block size decreases, and that the algorithm fails completely when the image size is 11×11. In Fig. 6(b) (algorithm A), the plots lie on the same line for $N_1 = N_2 \geqq 21$, while systematic stepwise error is observed when the block size is 11×11. Using the sub-pixel window alignment technique, this error is reduced significantly as shown in Fig. 6(c) (algorithm B).

Figures 7(a) and (b) show the effect of the sub-pixel window alignment technique more clearly, where the displacements δ_{i1} and δ_{i2} are plotted for the cases of (a) algorithm A and (b) algorithm B. Ideally, the displacements $(\delta_{i1}, \delta_{i2})$ ($i = 1, 2, \cdots, 60$) lie on a straight line. In Fig. 7(a), however, some discontinuity is observed. The error can be significantly reduced using the sub-pixel window alignment technique as shown in Fig. 7(b). Figure 8 shows the impact of the sub-pixel window alignment technique on the reduction of the RMS (Root Mean Square) error in displacement

estimation. The sub-pixel window alignment technique is highly effective especially for smaller image sizes.

5. Application to 3D Measurement

In this section, we implement a high-accuracy 3D measurement system using stereo vision by applying our proposed correspondence search algorithm. In stereo vision, area-based matching methods are frequently used to determine correspondences between stereo images (usually with pixel-level accuracy). The overall accuracy of 3D measurement system is mainly determined by (i) the baseline length between two cameras and (ii) the accuracy of estimated disparity between corresponding points. A conventional approach to high-accuracy 3D measurement is to employ a wide-baseline camera pair combined with edge-based correspondence matching. In this approach, only limited number of corresponding points located on edges can be used for 3D reconstruction. On the other hand, the area-based correspondence matching technique (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 the area-based approach is severely restricted by the narrow baseline between cameras. Our initial observation [10] shows that POC-based sub-pixel correspondence matching makes possible high-accuracy 3D surface reconstruction of a target object, even if the baseline between two cameras is relatively short. Also, the proposed technique is robust to intensity variation between stereo images. This section analyzes the impact of the proposed correspondence search technique on the accuracy of 3D measurement system.

In stereo vision, at least two images from different viewpoints are required for disparity estimation as illustrated in Fig. 9(a), where we assume parallel cameras for simplicity. The 3D coordinate $P_0 = (x, y, z)$ is a real world point, which is projected to p_0 and q_0 on the left and right camera images, $I_0(n_1, n_2)$ and $J_0(n_1, n_2)$, respectively. To determine the 3D coordinate P_0, we need to calibrate the cameras in advance and to know the corresponding points p_0 and q_0. The proposed correspondence search technique is used to find p_0 and q_0 in stereo images.

Figure 9(b) shows the experimental system for evaluating the accuracy of 3D measurement. We measure a wooden board using the same camera system as described in the previous section. Two images are taken at two camera positions with a distance of only 1 cm by moving the camera on the micro stage. The wooden board is placed 30 cm away from the camera and is set slightly panned with respect to the focal plane of the camera. The baseline is so narrow that almost no projective distortion is observed. This enables our stereo system to measure objects with various degrees of pan. The size of images taken with the camera is 640×480 and approximately 5000 points are chosen as reference points for correspondence search and disparity estimation. The block size used for matching is 15×15.

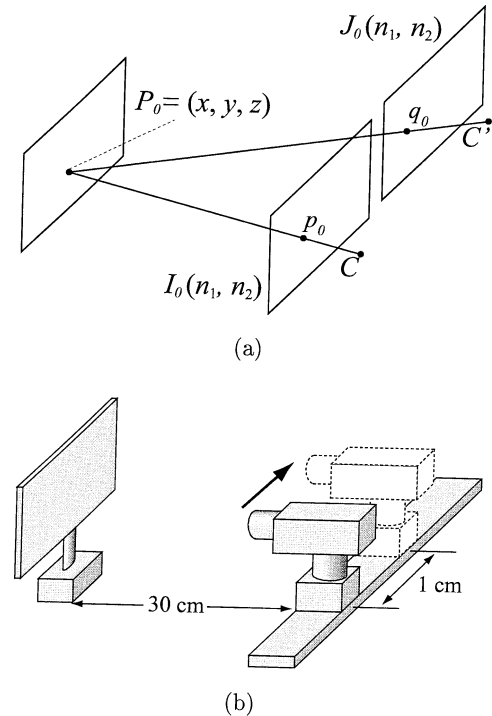In general, for the case of using area-based method, it is

(a)

(b)

Fig. 9 3D measurement based on stereo vision: (a) the principle and (b) the experimental system.
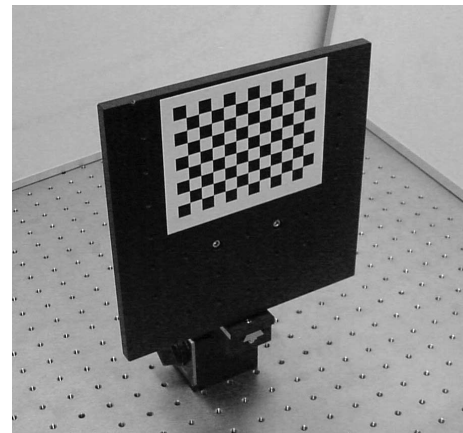

Fig. 10 Calibration object.

important to adjust the baseline length of the stereo cameras to reduce projective image distortion. We are now studying the design of a high-accuracy multi-camera 3D measurement system, where 3 pairs of narrow-baseline stereo cameras are employed with the POC-based correspondence search technique (see [10] for earlier discussion on this topic). The proposed sub-pixel correspondence search technique is useful in such applications.

We calibrate the camera at each position with a conventional method [8], where the object with the checker pattern shown in Fig. 10 is used. We detect the corners on the checker pattern using the Harris corner detector [11], and determine the corresponding corners in stereo im-
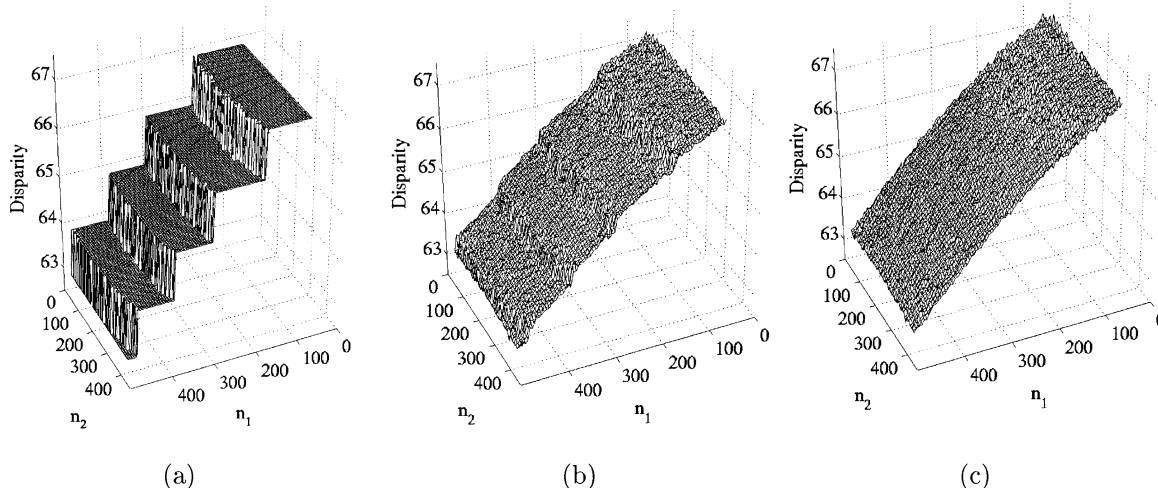
Fig. 11 Disparity maps of the wooden board: (a) conventional pixel-level correspondence search algorithm, (b) algorithm A, and (c) algorithm B.
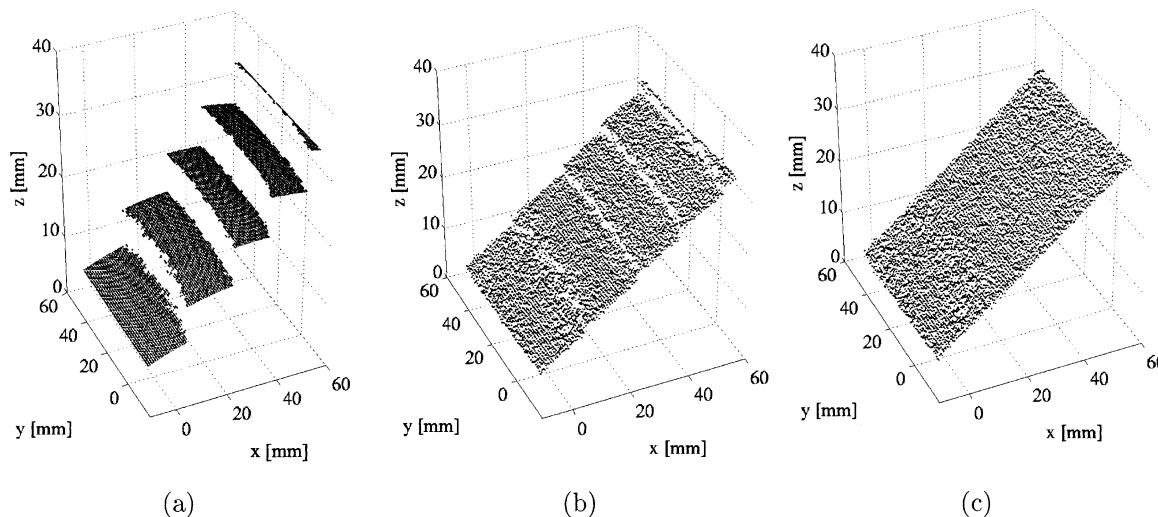

Fig. 12 Reconstructed 3D surfaces of the wooden board: (a) conventional pixel-level correspondence search algorithm, (b) algorithm A, and (c) algorithm B.

ages. Using these information, we determine the basic camera parameters, such as the relative rotation/translation between the cameras and lens distortion parameters, which are needed to reconstruct 3D coordinates of the target object. After the camera calibration, we take two images at two camera positions, and determine 5000 correspondences between the two images. We calculate disparity vectors for the 5000 pairs of corresponding points.

Figures 11(a), (b), and (c) show the measured disparity (i.e., length of a disparity vector), where (a) is the case of a conventional pixel-level correspondence search algorithm, and (b) and (c) are the cases of algorithms A and B (described in Sect. 4), respectively. In Fig. 11(a), we can see stepwise disparity error clearly. This error is significantly reduced in the algorithm A as shown in Fig. 11(b). The algorithm B (shown in Fig. 11(c)) produces a smooth disparity change, which directly reflects the flat surface of

the wooden board. Figures 12(a), (b), and (c) reconstruct the 3D coordinates of surface of the board using the disparity maps of Figs. 11(a), (b) and (c), respectively. Figure 12(c) clearly shows that the sub-pixel window alignment technique achieves higher accuracy in 3D reconstruction.

The reconstruction accuracy is evaluated by fitting an ideal plane to the 3D surface data in Figs. 12(a), (b), and (c) by least squares. The error is evaluated by calculating the distances of the measured points from the ideal plane. Table 1 shows the evaluated RMS errors. While the RMS error is 1.54 mm when using the conventional pixel-level correspondence search algorithm, the RMS error decreases to 0.39 mm when using algorithm A and to 0.20 mm when using algorithm B. This result clearly demonstrates the significant impact of the proposed method on the accuracy of 3D measurement.

Table 1 Comparison of RMS errors in 3D measurement.

	RMS error [mm]
Conventional pixel-level correspondence search algorithm	1.54
Algorithm A	0.39
Algorithm B	0.20

6. Conclusion

In this paper, we have proposed a high-accuracy correspondence search algorithm, which employs (i) a POC-based image matching technique, (ii) a coarse-to-fine search technique for pixel-level correspondence estimation, and (iii) a sub-pixel window alignment technique for sub-pixel correspondence estimation. Experimental evaluation of matching accuracy shows that the proposed method makes possible to estimate image displacements with approximately 0.05-pixel accuracy when using 11×11 image blocks. We have also implemented a high accuracy 3D measurement system using the proposed technique. Our system can achieve sub-mm accuracy in 3D measurement, even when the baseline length between stereo cameras is 1 cm and the camera-object distance is about 30 cm. Also, we are planning to apply the proposed technique to improve the performance of POC-based industrial vision sensors developed by our business-academia collaboration (see [12] for example).

References

[1] L.G. Brown, "A survey of image registration techniques," ACM Computing Surveys, vol.24, no.4, pp.325–376, Dec. 1992.

[2] Q. Tian and M.N. Huhns, "Algorithms for subpixel registration," Computer Vision, Graphics, and Image Processing, vol.35, no.2, pp.220–233, Aug. 1986.

[3] G.A. Thomas, "Television motion measurement for DATV and other applications," BBC Research Department Report, Sept. 1987.

[4] H. Foroosh, J. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," IEEE Trans. Image Process., vol.11, no.3, pp.188–200, 2002.

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

[6] T. Aoki, K. Takita, T. Higuchi, and K. Kobayashi, "Phase-based image matching and its application to intelligent vision systems," Proc. Int. Symp. New Paradigm VLSI Computing, pp.95–100, Dec. 2002.

[7] 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.

[8] O.D. Faugeras, Three-Dimensional Computer Vision, MIT Press, 1993.

[9] C.P. Sung, K.P. Min, and G.K. Moon, "Super-resolution image reconstruction: A technical overview," IEEE Signal Process. Mag., vol.20, no.3, pp.21–36, 2003.

[10] M.A. Muquit, K. Takita, T. Aoki, and T. Higuchi, "High-accuracy passive 3D measurement using multi-camera system based on phase-only correlation," Proc. IEEE Int. Symp. Intelligent Signal Processing and Communication Systems, pp.91–95, Nov. 2002.

[11] C. Harris and M.J. Stephens, "A combined corner and edge detector," Proc. 4th Alvey Vision Conference, pp.147–152, 1988.

[12] http://www.aoki.ecei.tohoku.ac.jp/poc/

Kenji Takita   received the B.E. degree in information engineering, and the M.S. and the Ph.D. degrees in information sciences from Tohoku University, Sendai, Japan, in 1999, 2001, and 2004, respectively. He is currently working at Matsushita Electric Industrial Co., Ltd. His research interests include computer vision and digital signal processing.

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.

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, and the Best Paper Award at the 1999 IEEE International Symposium on Intelligent Signal Processing and Communication Systems.

Tatsuo Higuchi received the B.E., M.E., and D.E. degrees in electronic engineering from Tohoku University, Sendai, Japan, in 1962, 1964, and 1969, respectively. He is currently a Professor at Tohoku Institute of Technology. From 1980 to 1993, he was a Professor in the Department of Electronic Engineering at Tohoku University. He was a Professor from 1994 to 2003, and was Dean from 1994 to 1998 in the Graduate School of Information Sciences at Tohoku University. His general research interests include the design of 1-D and multi-D digital filters, linear time-varying system theory, fractals and chaos in digital signal processing, VLSI computing structures for signal and image processing, multiple-valued ICs, multiwave opto-electronic ICs, and biomolecular computing. Dr. Higuchi received the Outstanding Paper Awards at the 1985, 1986, 1988, 1990, 2000 and 2001 IEEE International Symposiums on Multiple-Valued Logic, the Outstanding Transactions Paper Award from the Society of Instrument and Control Engineers (SICE) of Japan in 1984, the Technically Excellent Award from SICE in 1986, and the Outstanding Book Award from SICE in 1996, the Outstanding Transactions Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan in 1990 and 1997, the Inose Award from IEICE in 1997, the Technically Excellent Award from the Robotics Society of Japan in 1990, the IEE Ambrose Fleming Premium Award in 1994, the Outstanding Book Award from the Japanese Society for Engineering Education in 1997, the Award for Persons of scientific and technological merits (Commendation by the minister of state for Science and Technology), the IEE Mountbatten Premium Award in 1999 and the Best Paper Award at the 1999 IEEE International Symposium on Intelligent Signal Processing and Communication Systems. He also received the IEEE Third Millennium Medal in 2000. He received the fellow grade from IEEE and SICE.