High-Accuracy Subpixel Image Registration Based on Phase-Only Correlation

Kenji TAKITA[†a)], *Student Member*, Takafumi AOKI[†], Yoshifumi SASAKI[††], *Regular Members*, Tatsuo HIGUCHI[†††], *Fellow, and* Koji KOBAYASHI[††††], *Regular Member*

SUMMARY This paper presents a high-accuracy image registration technique using a Phase-Only Correlation (POC) function. Conventional techniques of phase-based image registration employ heuristic methods in estimating the location of the correlation peak, which corresponds to image displacement. This paper proposes a technique to improve registration performance by fitting the closed-form analytical model of the correlation peak to actual two-dimensional numerical data. This method can also be extended to a spectrum weighting POC technique, where we modify cross-phase spectrum with some weighting functions to enhance registration accuracy. The proposed method makes possible to estimate image displacements with 1/100-pixel accuracy.
key words: *image registration, subpixel registration, image matching, phase-only correlation, phase correlation*

1. Introduction

High-accuracy image registration is an important fundamental task in many fields, such as computer vision, remote sensing, medical imaging, etc. Especially for such applications as stereo-vision 3D measurement [1] and super-resolution imaging (that reconstructs a high-resolution image from multiple low-resolution images), the estimation of image displacements with subpixel accuracy is essential.

Over the years, various techniques for image registration have been developed. Typical examples include methods using image correlation functions as well as those using image features. Recently, a high-accuracy image registration technique using a *Phase-Only Correlation* (POC) function (or simply a "phase correlation function") has been developed [2]–[6]. The POC-based image registration enables to estimate the displacement between images with subpixel accuracy from the location of the correlation peak.

The image matching using POC can be extended

to the registration of translated, rotated and scaled images [7]. In this registration technique, the angle of image rotation and the scale factor of image reduction/magnification are detected by converting them into image translation. Therefore, estimating translational displacements of images with high accuracy is important. However, conventional techniques of phase-based image registration employ ad hoc methods for estimating the location of the correlation peak (usually using some interpolation techniques for the POC function), which limits the accuracy of displacement estimation to 1/10-pixel level. Another problem is that there have been no systematic reports on the experimental evaluation of subpixel registration techniques.

In this paper, we propose a method to significantly enhance the registration performance of the POC technique by fitting the analytical model of the correlation peak to two-dimensional numerical data array. We also extend this technique to a spectrum weighting POC technique, where we modify cross-phase spectrum with some weighting functions to reduce the effect of noise and to control the shape of correlation peak for better function fitting. Another original contribution of this paper is that we give systematic experimental analysis of the registration performance of the proposed technique by evaluating the errors in estimating the translational displacements (δ_1, δ_2), the rotation angle θ and the scale factor λ. A set of experiments demonstrate that the proposed technique can estimate image translation with 1/100-pixel accuracy, which leads to 1/40-degree accuracy in rotation estimation and to 1/10000-scale accuracy in scale factor estimation (when image size is 251×251).

Original contribution of this paper are summarized as follows: (i) derivation of a closed-form peak model for the POC function in terms of Discrete Fourier Transform (DFT), (ii) proposal of a high-accuracy image registration technique based on the function fitting using the derived peak model, (iii) proposal of a spectrum weighting technique that can control the shape of peak models so as to improve registration accuracy, and (iv) systematic experimental evaluation of registration performance ($\sim 1/100$-pixel accuracy) using actual images captured by a commercial-off-the-shelf CCD camera.

2. Phase-Only Correlation Function

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

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

where $k_1 = -M_1, \cdots, M_1 \quad k_2 = -M_2, \cdots, M_2$, $W_{N_1} = e^{-j\frac{2\pi}{N_1}}$ $W_{N_2} = e^{-j\frac{2\pi}{N_2}}$, 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 spectrum $R(k_1, k_2)$ between $F(k_1, k_2)$ and $G(k_1, k_2)$ is given by

$$
\begin{aligned}
R(k_1, k_2) &= F(k_1, k_2) \overline{G(k_1, k_2)} \\
&= A_F(k_1, k_2) A_G(k_1, k_2) e^{j\theta(k_1, k_2)}, \qquad (3)
\end{aligned}
$$

where $\overline{G(k_1, k_2)}$ denotes the complex conjugate of $G(k_1, k_2)$ and $\theta(k_1, k_2) = \theta_F(k_1, k_2) - \theta_G(k_1, k_2)$. On the other hand, the cross-phase spectrum (or normalized cross spectrum) $\hat{R}(k_1, k_2)$ is defined as

$$
\begin{aligned}
\hat{R}(k_1, k_2) &= \frac{F(k_1, k_2) \overline{G(k_1, k_2)}}{|F(k_1, k_2) \overline{G(k_1, k_2)}|} \\
&= e^{j\theta(k_1, k_2)}. \qquad (4)
\end{aligned}
$$

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

$$
\hat{r}(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1 k_2} \hat{R}(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2}, \quad (5)
$$

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

3. Subpixel Image Registration

In this section, we propose high-accuracy displacement estimation technique using the POC function. 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 subpixel 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 $f_c(x_1 - \delta_1, x_2 - \delta_2)$, 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 $n_2 = -M_2, \cdots, M_2$. Let $F(k_1, k_2)$ and $G(k_1, k_2)$ be the 2D DFTs of $f(n_1, n_2)$ and $g(n_1, n_2)$, respectively. Considering the difference of properties between the Fourier transform defined in continuous space and that defined in discrete space carefully, we can now say that

$$
G(k_1, k_2) \cong F(k_1, k_2) \cdot e^{-j\frac{2\pi}{N_1} k_1 \delta_1} e^{-j\frac{2\pi}{N_2} k_2 \delta_2}.
$$

Thus, $\hat{R}(k_1, k_2)$ is given by

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

The POC function $\hat{r}(n_1, n_2)$ will be the 2D IDFT of $\hat{R}(k_1, k_2)$, and is given by

$$
\begin{aligned}
&\hat{r}(n_1, n_2) \\
&= \frac{1}{N_1 N_2} \sum_{k_1 k_2} \hat{R}(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2} \\
&\cong \frac{\alpha}{N_1 N_2} \frac{\sin\{\pi(n_1 + \delta_1)\}}{\sin\{\frac{\pi}{N_1}(n_1 + \delta_1)\}} \frac{\sin\{\pi(n_2 + \delta_2)\}}{\sin\{\frac{\pi}{N_2}(n_2 + \delta_2)\}}, \quad (6)
\end{aligned}
$$

where $\alpha = 1$. The above equation represents the shape of the peak for the POC function for common images that are minutely displaced with each other. The peak position of the POC function corresponds to the displacement between the two images. We can prove that the peak value α decreases (without changing the function shape itself), when small noise components are added to the original images. Hence, we assume $\alpha \leqq 1$ in practice.

Figure 1 shows the POC function when $(\delta_1, \delta_2) = (0, 0)$ and $(\delta_1, \delta_2) = (0.5, 0)$. Figures 1(a) and (c) show the 3D plots of the POC function $\hat{r}(n_1, n_2)$, and (b) and (d) show the enlarged 2D plots of the POC function around the correlation peak. In Figs. 1(b) and (d), the black dots indicate the discrete data points of $\hat{r}(n_1, n_2)$ calculated by 2D DFT/IDFT, and the solid lines represent the estimated shape of the POC function given by (6). As shown in Figs. 1(c) and (d), the maximum value in the discrete 2D data array of the POC function $\hat{r}(n_1, n_2)$ could be less than 1, even if the two images $f(n_1, n_2)$ and $g(n_1, n_2)$ are captured from an identical real-world image $f_c(x_1, x_2)$.

Our approach is to use Eq. (6)—the closed-form peak model of POC function for images with minutely displaced with each other—directly for estimating the peak position by function fitting. This is possible only when the displacement (δ_1, δ_2) is small enough in comparison with the total image size (N_1, N_2). When the
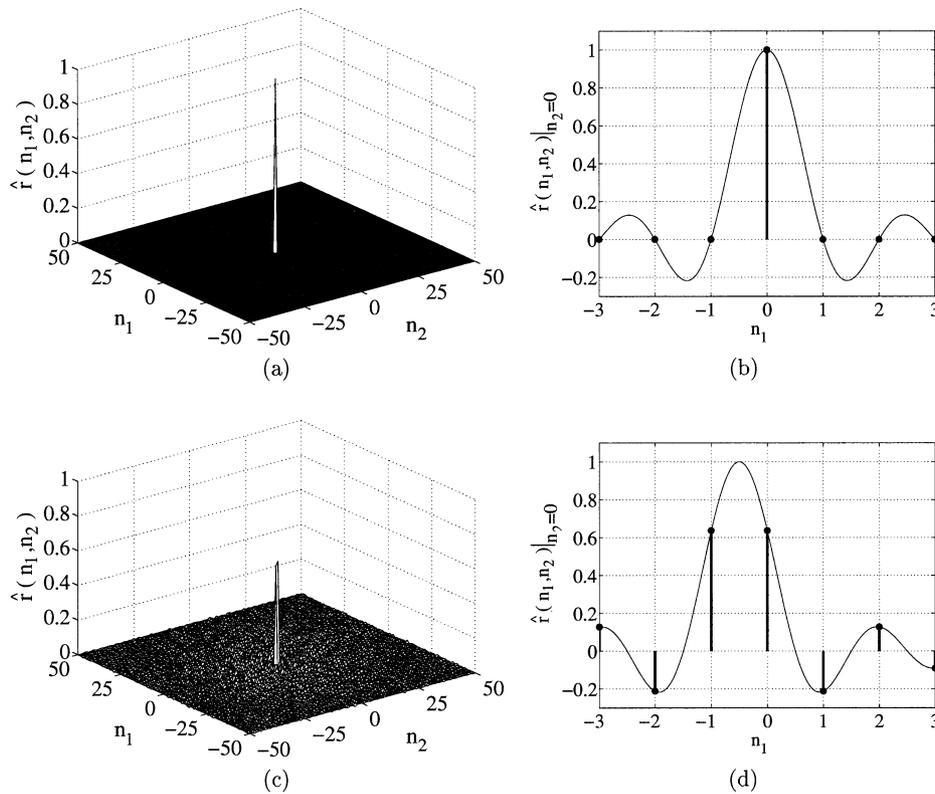
Fig. 1 3D and 2D plots of the POC function $\hat{r}(n_1, n_2)$: (a) for the case $(\delta_1, \delta_2) = (0, 0)$ and (b) its enlarged view $(n_2 = 0)$; and (c) for the case $(\delta_1, \delta_2) = (0.5, 0)$ and (d) its enlarged view $(n_2 = 0)$.

displacement is relatively large, we must employ coarse-to-fine search technique, where pixel-level image shift is detected first and then subpixel displacement is estimated by extracting sub-images from the original images so that the mutual translation of the two sub-images is small enough to assume the peak model (6) to be valid. For example, in stereo vision, we use multi-resolution images to estimate the disparities between stereo images, because we need to estimate relatively large displacements using small block images [1]. The coarse-to-fine search technique is to begin with images of the lowest resolution to determine coarse displacements, which gives an approximate solution for finer displacement estimation in the next higher resolution layer. This process is repeated to achieve subpixel registration accuracy. In this paper, we focus on the case where displacements between images are very small, so the coarse-to-fine search technique is not employed.

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 (6) to the calculated data array around the correlation peak, where α, δ_1 and δ_2 are fitting parameters. Figure 2(b) shows an example where Eq. (6) is fitted to the data array of $\hat{r}(n_1, n_2)$.

Listed below are important considerations for achieving high accuracy.

(i) Optimizing the number of points used in function fitting
We employ a least-square fitting technique for estimating image displacements (δ_1, δ_2) and correlation peak value α. Careful optimization of the number of fitting data points (around the maximum peak) is important in order to improve registration accuracy. Since the POC function (6) has a very sharp peak, limited number of data points $(3 \times 3 \sim 9 \times 9)$ are enough to achieve high-accuracy function fitting.

(ii) Windowing to eliminate the effect of periodicity in DFT
Due to the DFT 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 discontinuity 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 by

$$w(n_1, n_2) = \frac{1 + \cos(\frac{\pi n_1}{M_1})}{2} \frac{1 + \cos(\frac{\pi n_2}{M_2})}{2}. \quad (7)$$

(a) $H_0(k_1, k_2)$

(b) $\hat{r}_0(n_1, n_2)$

(c) $H_1(k_1, k_2)$

(d) $\hat{r}_1(n_1, n_2)$

(e) $H_2(k_1, k_2)$

(f) $\hat{r}_2(n_1, n_2)$

(g) $H_3(k_1, k_2)$

(h) $\hat{r}_3(n_1, n_2)$
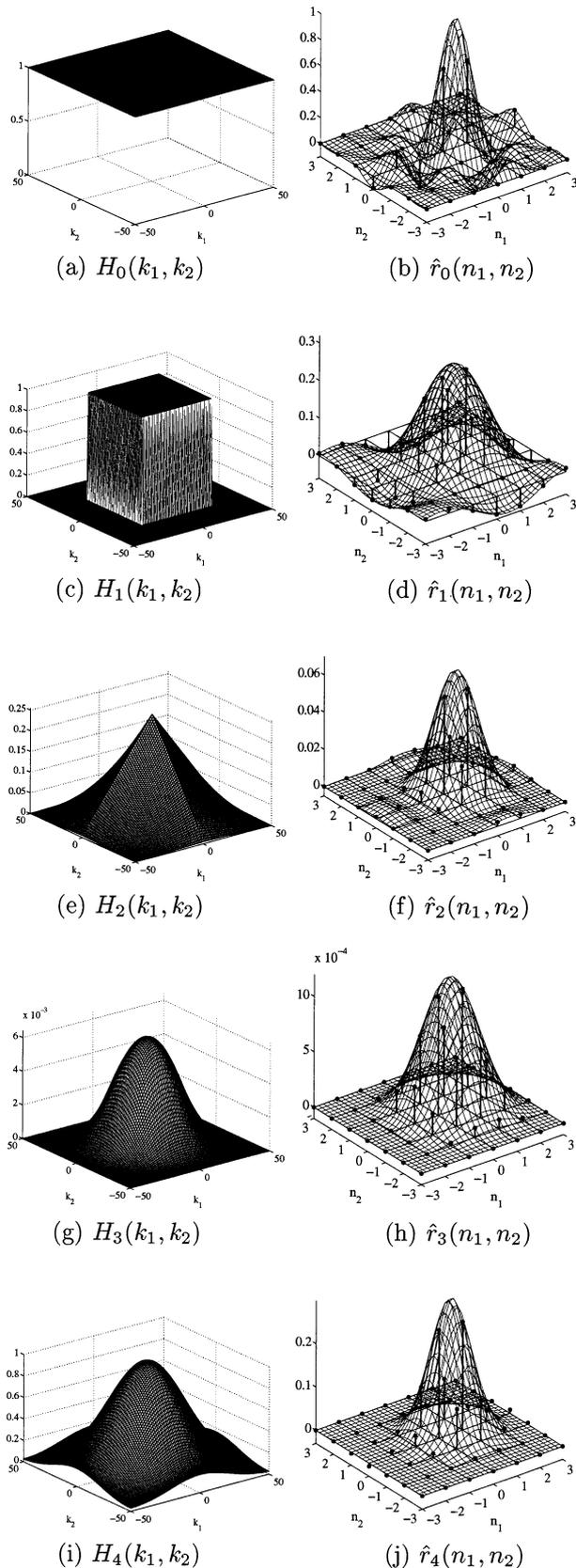
(i) $H_4(k_1, k_2)$

(j) $\hat{r}_4(n_1, n_2)$

Fig. 2 Spectrum weighting functions and corresponding correlation peak models: (a) $H_0(k_1, k_2)$, (b) $\hat{r}_0(n_1, n_2)$, (c) $H_1(k_1, k_2)$, (d) $\hat{r}_1(n_1, n_2)$, (e) $H_2(k_1, k_2)$, (f) $\hat{r}_2(n_1, n_2)$, (g) $H_3(k_1, k_2)$, (h) $\hat{r}_3(n_1, n_2)$, (i) $H_4(k_1, k_2)$, and (j) $\hat{r}_4(n_1, n_2)$.

(iii) Modifying cross-phase spectrum by weighting functions

For natural images, typically most of the energy is concentrated in the low spatial frequency components. Equation (4) of cross-phase spectrum $\hat{R}(k_1, k_2)$ implies that the calculation of POC emphasizes the high frequency components, which 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 to $\hat{R}(k_1, k_2)$ in frequency domain and eliminating the high frequency components having low reliability. A typical example of the weighting function is a rectangular low-pass function $H_1(k_1, k_2)$ defined as

$$H_1(k_1, k_2) = \begin{cases} 1 & |k_1| \leqq U_1, |k_2| \leqq U_2 \\ 0 & \text{otherwise} \end{cases}, \qquad (8)$$

where U_1 and U_2 are integers satisfying $0 \leqq U_1 \leqq M_1$ and $0 \leqq U_2 \leqq M_2$. The cross-phase spectrum $\hat{R}(k_1, k_2)$ is multiplied by the weighting function $H_1(n_1, n_2)$ when calculating the 2D IDFT. Let $\hat{r}_1(n_1, n_2)$ be the POC function when using the weighting function $H_1(k_1, k_2)$. Then $\hat{r}_1(n_1, n_2)$ will be given by

$$
\begin{aligned}
& \hat{r}_1(n_1, n_2) \\
&= \frac{1}{N_1 N_2} \sum_{k_1 k_2} \hat{R}(k_1, k_2) H_1(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2} \\
&\cong \frac{\alpha}{N_1 N_2} \frac{\sin\{\frac{V_1}{N_1}\pi(n_1 + \delta_1)\}}{\sin\{\frac{\pi}{N_1}(n_1 + \delta_1)\}} \frac{\sin\{\frac{V_2}{N_2}\pi(n_2 + \delta_2)\}}{\sin\{\frac{\pi}{N_2}(n_2 + \delta_2)\}},
\end{aligned}
$$
$$(9)$$

where $V_1 = 2U_1 + 1$ and $V_2 = 2U_2 + 1$. Figure 2(d) shows the situation when (9) is used in function fitting. Note that the main lobe of the POC function extends more than the case of no filtering (shown in Fig. 2(b)).

In general, if we use a weighting function, we need to change the peak model for function fitting correspondingly. In other words, we can change the peak model for function fitting by changing the weighting function to be multiplied with cross-phase spectrum. Assume that we are to use the following peak model for function fitting:

$$\hat{r}(n_1, n_2) = h(n_1 + \delta_1, n_2 + \delta_2), \qquad (10)$$

where $h(x_1, x_2)$ denotes an arbitrary closed-form function defined on real variables x_1 and x_2, and δ_1 and δ_2 are minute image displacements. Note that we need to select the function model $h(x_1, x_2)$ so that it can be reconstructed completely using $N_1 \times N_2$ discrete sampling points based on sampling theorem. For example, the peak model (9) is a typical example of such functions. Then, the corresponding weighting function $H(k_1, k_2)$ is obtained by computing the 2D DFT of $h(n_1, n_2)$ numerically. As a result, we have the following relationship:

$$\hat{r}(n_1, n_2)$$

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

$$\cong \frac{1}{N_1 N_2} \sum_{k_1 k_2} e^{j \frac{2\pi}{N_1} k_1 \delta_1} e^{j \frac{2\pi}{N_2} k_2 \delta_2} H(k_1, k_2)$$

$$\cdot W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2}$$

$$= h(n_1 + \delta_1, n_2 + \delta_2). \tag{11}$$

For selecting $h(n_1, n_2)$ and $H(k_1, k_2)$, we must carefully consider the following: (i) we had better use the function $h(n_1, n_2)$ that makes possible high-accuracy function fitting, and (ii) we must select $H(k_1, k_2)$ for realizing adequate weighting of cross-phase spectrum according to the reliability of frequency components. Also, we need to optimize both the number of data points used for function fitting and pass-band width of weighting function $H(k_1, k_2)$.

In addition to the above mentioned $H_1(k_1, k_2)$, we consider the weighting functions $H_2(k_1, k_2)$ and $H_3(k_1, k_2)$, which are derived from $H_1(k_1, k_2)$ as:

$$H_2(k_1, k_2) = \frac{1}{N_1 N_2} H_1(k_1, k_2) \otimes H_1(k_1, k_2), \quad (12)$$

$$H_3(k_1, k_2) = \frac{1}{N_1^2 N_2^2} H_2(k_1, k_2) \otimes H_1(k_1, k_2), \quad (13)$$

where $\otimes$ denotes the convolution operator. The corresponding closed-form peak models are given below

$$\hat{r}_2(n_1, n_2)$$

$$= \frac{1}{N_1 N_2} \sum_{k_1 k_2} \hat{R}(k_1, k_2) H_2(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2}$$

$$\cong \left\{ \frac{\alpha}{N_1 N_2} \frac{\sin\{\frac{V_1}{N_1} \pi (n_1 + \delta_1)\}}{\sin\{\frac{\pi}{N_1} (n_1 + \delta_1)\}} \frac{\sin\{\frac{V_2}{N_2} \pi (n_2 + \delta_2)\}}{\sin\{\frac{\pi}{N_2} (n_2 + \delta_2)\}} \right\}^2, \tag{14}$$

$$\hat{r}_3(n_1, n_2)$$

$$= \frac{1}{N_1 N_2} \sum_{k_1 k_2} \hat{R}(k_1, k_2) H_3(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2}$$

$$\cong \left\{ \frac{\alpha}{N_1 N_2} \frac{\sin\{\frac{V_1}{N_1} \pi (n_1 + \delta_1)\}}{\sin\{\frac{\pi}{N_1} (n_1 + \delta_1)\}} \frac{\sin\{\frac{V_2}{N_2} \pi (n_2 + \delta_2)\}}{\sin\{\frac{\pi}{N_2} (n_2 + \delta_2)\}} \right\}^3. \tag{15}$$

Another useful example of weighting function is Gaussian function $H_4(k_1, k_2)$ defined as

$$H_4(k_1, k_2) \cong e^{-2\pi^2 \sigma^2 (k_1^2 + k_2^2)}, \tag{16}$$

where σ is a parameter that controls the function width. In this case, we need to use the following peak model:

$$\hat{r}_4(n_1, n_2) \cong \frac{1}{2\pi\sigma^2} e^{-(n_1^2 + n_2^2)/2\sigma^2}. \tag{17}$$

Figures 2(a), (c), (e), (g) and (i) show the weighting functions $H_0(k_1, k_2)$, $H_1(k_1, k_2)$, $H_2(k_1, k_2)$, $H_3(k_1, k_2)$ and $H_4(k_1, k_2)$, respectively, where $H_0(k_1, k_2)$ is the case of no filtering. The corresponding peak models $\hat{r}_0(n_1, n_2)$, $\hat{r}_1(n_1, n_2)$, $\hat{r}_2(n_1, n_2)$, $\hat{r}_3(n_1, n_2)$ and $\hat{r}_4(n_1, n_2)$ are shown in Figs. 2(b), (d), (f), (h) and (j), respectively.

4. Experiments of Image Displacement Estimation

This section describes a set of experiments for estimating translational image displacements using the proposed technique. We have estimated the displacements between two images taken by a CCD camera (JAI CVM10 with Sony VCL–16WM lens). The target object is a wood cube with the size of $10\,\text{cm} \times 10\,\text{cm} \times 10\,\text{cm}$, which is mounted on a micro stage that allows precise alignment of the cube position (Fig. 3). The cube is placed so that one side of the cube is parallel with the focal plane of CCD. The distance between the camera and the cube is 70 cm, and the size of the cube in the captured images is about 200×200. We have moved the micro stage 53 times horizontally with each step of 0.05 mm displacement and took 30 sequential images (1 second) at each position. These 30 static frames are averaged to improve image quality.

From the captured images we have extracted 101×101 sub-images having a wooden texture, which are used for registration experiments (i.e., $N_1 = N_2 = 101$ and $M_1 = M_2 = 50$). These sub-images contain only the texture of the wood cube and do not contain background textures. We have calculated the POC function between a reference image before moving the cube and each image after moving it. We have estimated the horizontal displacement δ_1 [pixel] by fitting Eqs. (6), (9), (14), (15), or (17) depending on the weighting function used in each experimental trial. The $p \times p$ data points around the maximum peak of a correlation array are used in function fitting with the least-square method, where p is optimized in every experiment. The parameters: V_1 and V_2 (i.e., U_1 and U_2) for $H_1(k_1, k_2)$–$H_3(k_1, k_2)$ and σ for $H_4(k_1, k_2)$ control the pass-band
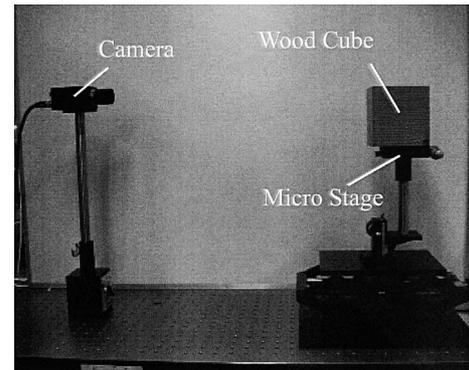
Fig. 3 Experimental system.

Table 1　Error [pixel] in displacement estimation.

	Maximum Error	RMS Error
"Original"	0.5045	0.2837
"Fitting" # of fitting points: 3×3 Fitting function: $\hat{r}_0(n_1, n_2)$	0.1232	0.0421
"Fitting+Window" # of fitting points: 3×3 Fitting function: $\hat{r}_0(n_1, n_2)$	0.0227	0.0101
"Fitting+Window+$H_1(k_1, k_2)$" # of fitting points: 7×7 Fitting function: $\hat{r}_1(n_1, n_2)$ $U_1/M_1 = U_2/M_2 = 0.5$	0.0165	0.0059
"Fitting+Window+$H_2(k_1, k_2)$" # of fitting points: 7×7 Fitting function: $\hat{r}_2(n_1, n_2)$ $U_1/M_1 = U_2/M_2 = 0.34$	0.0138	0.0055
"Fitting+Window+$H_3(k_1, k_2)$" # of fitting points: 7×7 Fitting function: $\hat{r}_3(n_1, n_2)$ $U_1/M_1 = U_2/M_2 = 0.48$	0.0083	0.0038
"Fitting+Window+$H_4(k_1, k_2)$" # of fitting points: 7×7 Fitting function: $\hat{r}_4(n_1, n_2)$ $\sigma = 0.71$	0.0080	0.0037

Table 2　RMS error [pixel] when changing the cutoff frequencies and the number of fitting points (for the case of $H_1(k_1, k_2)$).

$\frac{U_1}{M_1} = \frac{U_2}{M_2}$	Number of Fitting Points			
	3×3	5×5	7×7	9×9
0.1	0.0303	0.0321	0.0313	0.0289
0.2	0.0118	0.0112	0.0103	0.0099
0.3	0.0075	0.0071	0.0070	0.0070
0.4	0.0072	0.0067	0.0063	0.0065
0.5	0.0077	0.0062	0.0059	0.0065
0.6	0.0119	0.0072	0.0090	0.0082
0.7	0.0157	0.0103	0.0106	0.0114
0.8	0.0200	0.0200	0.0166	0.0169
0.9	0.0178	0.0279	0.0300	0.0298
1.0	0.0105	0.0210	0.0282	0.0327

width of weighting functions. These parameters are also optimized in every experiment.

We evaluate the measurement errors in the following manner. We have 53 sets of data of the actual displacements δ_1 [pixel] estimated from images and the displacements Δ [mm] of the micro stage. Using these data, we calculate an approximate line $\delta_1 = a \times \Delta$ in a least-square sense, where a [pixel/mm] is a constant. In this experiment, the parameter a is evaluated as $a \cong 2.04$ [pixel/mm] by least-squares fitting. This means that 1 mm displacement of the micro stage corresponds to the displacement of about 2.04 pixels in the captured images. Now, let Δ_i [mm][†] be the actual displacement of the micro stage after the i-th minute movement, and let δ_{1i} [pixel] be the displacement estimated from images taken at the corresponding i-th position. We evaluate the estimation error ϵ_{Ti} [pixel] by the following equation:

$$\epsilon_{Ti} = \delta_{1i} - a \times \Delta_i. \qquad (18)$$

Table 1 summarizes the error [pixel] in displacement estimation, where RMS error represents *Root Mean Square* error. The "Original" corresponds to pixel-level displacement estimation, "Fitting" corresponds to displacement estimation by function fitting using (6), "Fitting+Window" means combination of the function fitting technique and the windowing technique using (7), and "Fitting+Window+$H_j(k_1, k_2)$" means combination of the function fitting technique, the windowing technique and the spectrum weighting technique using the weighting function $H_j(k_1, k_2)$ ($j = 1, 2, 3, 4$).

The parameters shown in Table 1 are selected by

experimental optimization. For example, Table 2 shows parameter optimization when using spectrum weighting function $H_1(k_1, k_2)$. By changing the parameters of the weighting function and the number of fitting points, we can reduce the estimation error. From this table, the best result is obtained when $U_1/M_1 = U_2/M_2 = 0.5$ and the number of fitting points is 7×7. This result shows that eliminating unreliable high frequency components is effective to achieve higher accuracy.

Figure 4 shows the registration error ϵ_{Ti} [pixel] (in each estimated δ_1) versus a actual displacement Δ_i [mm] of the cube. As for the spectrum weighting function, only the case of $H_4(k_1, k_2)$ is shown. Our initial observation shows that the estimation accuracy could not be improved when using only weighting functions. However, combination of windowing and spectrum weighting has a significant impact on registration error reduction; RMS error decreases below 1/100-pixel level.

Figure 5 shows the registration error when changing image size, where we employ the combination: "Fitting+Window+$H_4(k_1, k_2)$." As shown in this plot, the estimation accuracy significantly depends on the image size. However, the estimation error is still 1/100-pixel level even if the image size is 41×41.

5. Registration of Translated, Rotated and Scaled Images

The proposed high-accuracy image registration technique can be extended to the registration for images including translation, rotation and scaling simultaneously. This section is to evaluate the accuracy of rotation angle estimation and scale factor estimation using the proposed technique. In this experiment, we employ the algorithm described in [7] to transform the image rotation and scaling into image translation. The rotation angle and the scale factor are estimated by detecting the corresponding translational displacements by the proposed technique.

[†]$\Delta_i = 0.05 \times i$ [mm], since the cube is moved by 0.05 mm at every minute movement.

(a)

(b)

(c)

(d)
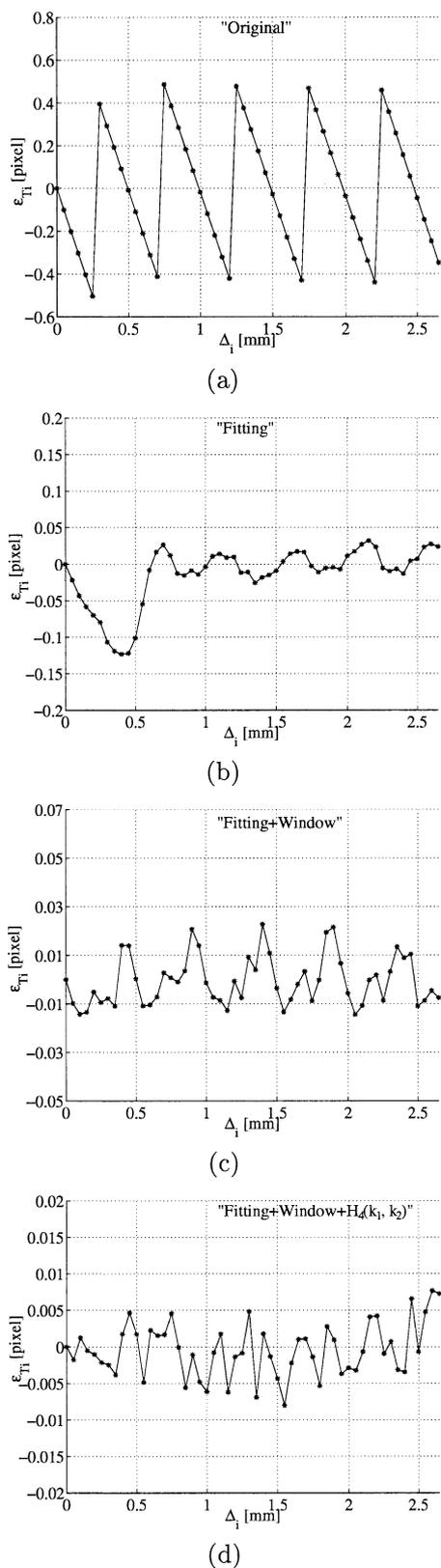
Fig. 4 Error in displacement estimation for the case: (a) "Original" (b) "Fitting," (c) "Fitting+Window," and (d) "Fitting+Window+$H_4(k_1, k_2)$."
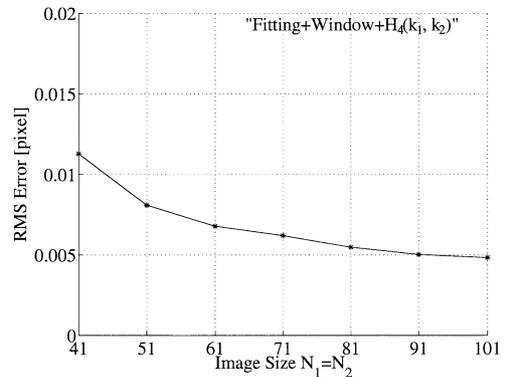
Fig. 5 RMS error when changing the image size.

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 $g_c(x_1, x_2)$ be the image obtained by translating, rotating and scaling the image $f_c(x_1, x_2)$ by the displacements (δ_1, δ_2), the angle θ and the scale factor λ, respectively. 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 $g_c(x_1, x_2)$ as

$$
\begin{aligned}
f(n_1, n_2) &= f_c(x_1, x_2), \\
g(n_1, n_2) &= g_c(x_1, x_2) \\
&= f_c(\ \lambda(x_1 - \delta_1)\cos\theta - \lambda(x_2 - \delta_2)\sin\theta, \\
&\quad \lambda(x_1 - \delta_1)\sin\theta + \lambda(x_2 - \delta_2)\cos\theta\ ),
\end{aligned}
$$

where $n_1 = -M_1, \cdots, M_1$, $n_2 = -M_2, \cdots, M_2$, $x_1 = n_1 T_1$ and $x_2 = n_2 T_2$. For simplicity, we assume that $M_1 = M_2 = M$, $N_1 = N_2 = N$ and $T_1 = T_2 = 1$ in the following discussion   Let $F_c(\Omega_1, \Omega_2)$ be the 2D Fourier Transform (2D FT) of $f_c(x_1, x_2)$, and $F(k_1, k_2)$ and $G(k_1, k_2)$ be the 2D DFTs of $f(n_1, n_2)$ and $g(n_1, n_2)$, respectively. Then, we have the approximation:

$$|F(k_1, k_2)| \cong |F_c(\Omega_1, \Omega_2)|, \qquad (19)$$

$$
\begin{aligned}
|G(k_1, k_2)| \cong \frac{1}{\lambda^2} \bigg| F_c &\bigg(\frac{1}{\lambda}(\Omega_1 \cos\theta - \Omega_2 \sin\theta), \\
&\frac{1}{\lambda}(\Omega_1 \sin\theta + \Omega_2 \cos\theta)\bigg) \bigg|, \qquad (20)
\end{aligned}
$$

where $k_1 = -M, \cdots, M$ $k_2 = -M, \cdots, M$, $\Omega_1 = 2\pi k_1/N$ and $\Omega_2 = 2\pi k_2/N$. Thus, we could omit the translational displacements (δ_1, δ_2) by considering only amplitude spectra $|F(k_1, k_2)|$ and $|G(k_1, k_2)|$. These amplitude spectra are treated as real-valued images.

We estimate (δ_1, δ_2), θ and λ as follows: (i) estimate the rotation angle θ and the scale factor λ using the amplitude spectra $|F(k_1, k_2)|$ and $|G(k_1, k_2)|$, (ii) normalize the second image $g(n_1, n_2)$ with the scale factor $1/\lambda$ and the rotation angle $-\theta$ to obtain a new image $g'(n_1, n_2)$, and (iii) estimate the translation (δ_1, δ_2) between $f(n_1, n_2)$ and $g'(n_1, n_2)$.

The main problem here is the step (i) of the above procedure. Let $F_{LP}(l_1, l_2)$ and $G_{LP}(l_1, l_2)$ be the *log-polar* mapping of $F(k_1, k_2)$ and $G(k_1, k_2)$, respectively,

where $l_1 = -M \cdots M$ and $l_2 = -M \cdots M$. We can write $|F_{LP}(l_1, l_2)|$ and $|G_{LP}(l_1, l_2)|$ as

$$|F_{LP}(l_1, l_2)| \cong |F_c(\ (\pi \log_N r) \cos \phi, (\pi \log_N r) \sin \phi \ )|, \tag{21}$$

$$|G_{LP}(l_1, l_2)| \cong \frac{1}{\lambda^2} \left| F_c \left(\left(\pi \log_N \frac{r}{\lambda} \right) \cos(\phi + \theta), \right. \right.$$
$$\left. \left. \left(\pi \log_N \frac{r}{\lambda} \right) \sin(\phi + \theta) \right) \right|, \tag{22}$$

where $\phi = l_1 \pi / N$ and $r = N^{(2l_2 + 2M + 1)/2N}$. From the above equations, we can now say that

$$|G_{LP}(l_1, l_2)| \cong \frac{1}{\lambda^2} \left| F_{LP} \left(l_1 + \frac{N}{\pi} \theta, l_2 - N \log_N \lambda \right) \right|. \tag{23}$$

From the above equation, we can see that the rotation angle θ and the scale factor λ is transformed to the translational displacements $(N\theta/\pi, -N \log_N \lambda)$ between two images $F_{LP}(l_1, l_2)$ and $G_{LP}(l_1, l_2)$. Hence, we can estimate θ and λ using the proposed image registration technique described in the last section.

We can summarize the procedure of translation (δ_1, δ_2), rotation θ and scaling λ estimation as follows:
[Step 1]   Calculate 2D DFTs of the discrete images $f(n_1, n_2)$ and $g(n_1, n_2)$ to obtain $F(k_1, k_2)$ and $G(k_1, k_2)$.

[Step 2]   Calculate the amplitude spectra $|F(k_1, k_2)|$ and $|G(k_1, k_2)|$. For natural images, most energy is concentrated in low-frequency domain. Hence, we had better use $log|F(k_1, k_2)|$ and $log|G(k_1, k_2)|$ in stead of $|F(k_1, k_2)|$ and $|G(k_1, k_2)|$.

[Step 3]   Calculate the log-polar mapping $|F_{LP}(l_1, l_2)|$ and $|G_{LP}(l_1, l_2)|$.

[Step 4]   Estimate the image displacement between $|F_{LP}(l_1, l_2)|$ and $|G_{LP}(l_1, l_2)|$ using the POC technique to obtain θ and λ.

[Step 5] Normalize the second image $g(n_1, n_2)$ with the scale factor $1/\lambda$ and the rotation angle $-\theta$ to obtain a new image $g'(n_1, n_2)$, and estimate the translation (δ_1, δ_2) between $f(n_1, n_2)$ and $g'(n_1, n_2)$.   □

Our careful experimental observation shows that the accuracy of log-polar mapping in the above procedure (Step 3) is particularly important. In log-polar mapping, we need to interpolate pixel intensity between discrete sampling points. The accuracy of this interpolation has a significant impact on the total accuracy of image registration. We employ a high-accuracy interpolation technique that combines zero-padding image extension and bilinear image interpolation. (We must omit the detailed discussion on this interpolation issue due to limited space in this paper.)

We have carried out a set of experiments for evaluating the accuracy of rotation angle θ and scale factor
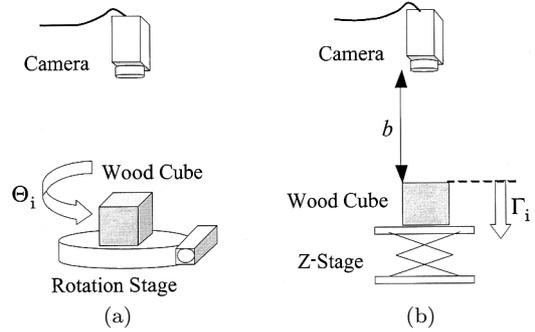
Fig. 6   Experimental system setup: (a) rotation estimation and (b) scale estimation.

λ estimated by using the proposed method. The target object used in the following experiments is the same wood cube shown in Fig. 3. We have mounted the cube on a micro stage that allows precise rotation and z-axis alignment as shown in Fig. 6. For rotation estimation (Fig. 6(a)), the distance from the camera to the cube is 70 cm, and the size of the cube in the captured images is about 350×350. We use 251×251 sub-images with a wooden texture extracted from the camera images. These sub-images contain only the texture of the wood cube and do not contain background textures. We have rotated the micro stage from 0 to 90 degrees with each step of 1 degree, and took time-averaged images (30 frames at each position). For scale factor estimation (Fig. 6(b)), the initial distance from the camera to the cube is about 50 cm. We have moved the z-stage 12 times with each micro step of 5 mm.

The measurement error for angle estimation is evaluated by

$$\epsilon_{Ri} = \theta_i - \Theta_i,$$

where θ_i [degree] and Θ_i [degree] are the estimated angle and the actual angle of the rotation stage for the i-th micro step. Figure 7 shows the error in rotation estimation, where "Original" corresponds to pixel-level angle estimation, "Fitting" corresponds to rotation estimation by function fitting, and "Optimized" means the combination of the function fitting technique, the windowing technique and the spectrum weighting technique using the weighting function $H_4(k_1, k_2)$ (with $\sigma = 0.74$). The number of fitting points used for function fitting is 9×9. The RMS errors for "Original," "Fitting" and "Optimized" are 0.209 [degree], 0.0473 [degree] and 0.0277 [degree], respectively.

On the other hand, the measurement error for scale estimation is evaluated by

$$\epsilon_{Si} = \lambda_i - \frac{b}{b + \Gamma_i}$$

where λ_i and Γ_i are the estimated scale factor and the z-axis shift [mm] of a micro stage for the i-th micro movement. The parameter b, the initial distance from
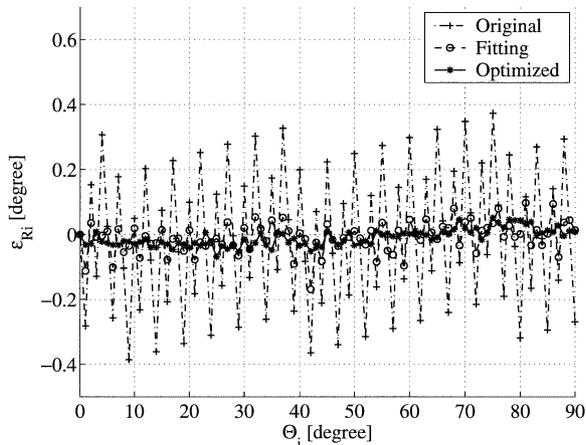
Fig. 7   Error in rotation estimation.
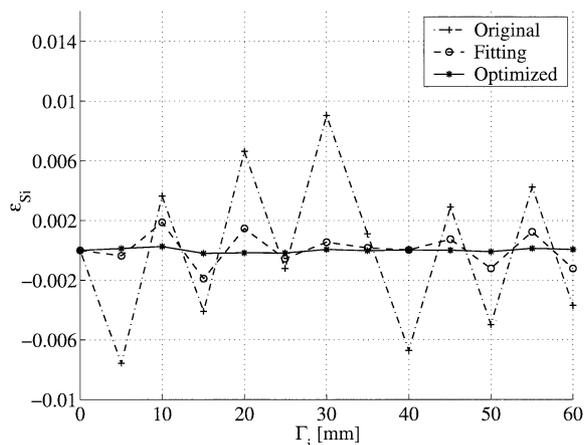
Fig. 8   Error in scale estimation.

the camera to the object, is optimized by the least-square method. Figure 8 shows the error in scale estimation, where "Original" corresponds to pixel-level angle estimation, "Fitting" corresponds to rotation estimation by function fitting, and "Optimized" means the combination of the function fitting technique, the windowing technique and the spectrum weighting technique using the weighting function $H_4(k_1, k_2)$ (with $\sigma = 0.56$). The number of fitting points used for function fitting is 9×9. The RMS errors for "Original," "Fitting" and "Optimized" are 5.00×10^{-3}, 1.08×10^{-3} and 1.29×10^{-4}, respectively.

As is observed in the above experiments, the proposed technique for high-accuracy translation estimation is effective also for rotation and scale estimation; the proposed method achieves 1/40-degree accuracy in rotation estimation and 1/10000-scale accuracy in scale estimation when the image size is 251×251.

6.   Conclusion

Original contribution of this paper are summarized as follows: (i) derivation of closed-form model for the

POC function in terms of Discrete Fourier Transform (DFT), (ii) proposal of a high-accuracy image registration technique based on the function fitting using the derived peak model, (iii) proposal of a spectrum weighting technique that can control the shape of peak models so as to improve registration accuracy, and (iv) systematic experimental evaluation of registration performance ($\sim$ 1/100-pixel accuracy) using actual images captured by a commercial-off-the-shelf CCD camera. We are planning to apply the proposed technique to POC-based commercial products developed by our group (see [8] for example). Also, a part of the methods described in this paper has been successfully applied to the implementation of a high-accuracy passive 3D measurement system based on stereo image matching [1]. Another interesting application is super-resolution imaging, which reconstructs a high-resolution image using multiple low-resolution images.

References

[1] 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.

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

[3] G.A. Thomas, "HDTV bandwidth reduction by adaptive subsampling and motion-compensation DATV techniques," Society of Motion Picture and Television Engineers Journal, pp.460–465, May 1987.

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

[5] H. Shekarforoush, M. Berthod, and J. Zerubia, "Subpixel image registration by estimating the polyphase decomposition of the cross power spectrum," INRIA Technical Report, no.2707, Nov. 1995.

[6] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy image registration based on phase-only correlation and its experimental evaluation," Proc. IEEE Int. Symp. Intelligent Signal Processing and Communication Systems, pp.86–90, Nov. 2002.

[7] Q. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.16, no.12, pp.1156–1168, Dec. 1994.

[8] http://www.higuchi.ecei.tohoku.ac.jp/poc/

Kenji Takita received the B.E. degree in information engineering, and the M.S. degree in information sciences from Tohoku University, Sendai, Japan, in 1999 and 2001, respectively. He is currently working toward the Ph.D. degree. His research interests include computer vision and digital signal 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, and the Best Paper Award at the 1999 IEEE International Symposium on Intelligent Signal Processing and Communication Systems.

Yoshifumi Sasaki received the B.E. and M.E. degrees in electronic engineering, and the D.E. degree in information sciences from Tohoku University, Sendai, Japan, in 1991, 1993 and 1996, respectively. He is currently a Lecturer at Ishinomaki Senshu University. His research interests include VLSI computing structures for robot, signal processing and image processing.

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, IEICE, and SICE.

Koji Kobayashi received the B.E. and M.E. degrees in electronic engineering from Tohoku University, Sendai, Japan, in 1976, and 1978, respectively. He is currently General Manager, Vision Sensing Department, Yamatake Corporation, Isehara, Japan. His research interests include image processing, biometric authentication, VLSI design, and CMOS image sensors.