IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 28, 2019, accepted February 9, 2019, date of publication February 18, 2019, date of current version March 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2899390

High-Accuracy Image Rotation and Scale Estimation Using Radon Transform and Sub-Pixel Shift Estimation

TAKANORI FUJISAWA, (Member, IEEE), AND MASAAKI IKEHARA, (Member, IEEE)
Department of Electronics and Electrical Engineering, Keio University, Yokohama 223-8522, Japan
Corresponding author: Takanori Fujisawa (fujisawa@tkhm.elec.keio.ac.jp)

ABSTRACT Rotation and scale estimation of images are fundamental tasks in image registration. The conventional estimation method uses log-polar transform and 1D shift estimation to estimate rotation and scale regardless of the shift of images. However, this transform requires interpolation of the frequency components, which causes estimation error. We propose a rotation and scale estimation algorithm based on the Radon transform and sub-pixel shift estimation. The Radon transform can estimate the rotation independent of the shift and can reduce the influence of interpolation error. In addition, sub-pixel shift estimation using a linear approximation of the phase component improves the precision of 1D shift estimation and achieves accurate rotation estimation. The proposed method was evaluated on test images, and the results demonstrate that the proposed method has higher accuracy compared with the log-polar transform.

INDEX TERMS Image registration, rotation estimation, scale estimation, phase only correlation.

I. INTRODUCTION

With the recent advancements in hardware technology, processing huge volumes of image and video data has become common. Image registration, which matches two or more images, is an important technology. These images are taken from, for example, video frames, different cameras, or different viewpoints, and image registration can be used in many applications, such as medical image analysis, object recognition, and industrial vision.

The fundamental task in image registration is to estimate the shift, scale, and rotation between two images. Conventional approaches for rotation and shift estimation are categorized by phase-only correlation (POC) approaches [1]–[5], intensity-based registration [6]–[9], and feature-based registration [10]. The intensity-based image registration is proposed by Thevenaz and Unser [6] and Thevenaz *et al.* [7]. A typical intensity-based algorithm maximizes Mattes' mutual information [8] using step gradient descent or a one-plus-one evolutionary algorithm [9].

The feature-based approach extracts the speeded-Up robust features (SURF) [11] or binary robust invariant scalable keypoints (BRISK) [12] and estimates a transform matrix using existing methods [10], [13], [14]. The intensity-based approach compares two images by plain and achieves high precision if the rotation angle is small and the scale is close to 1. However, to estimate a large angle or scale, the number of iterations required to calculate the optimization problem increases, and an incorrect estimation may be obtained because of a local solution. The feature-based approach depends on the number of features extracted from the images and the precision of the point-to-point estimation of features. Therefore, the precision of the feature-based approach is reduced when a small number of feature points are extracted or incorrect matching occurs in point-by-point estimation.

POC is a common method for image shift estimation [2] that calculates the cross-power spectrum of two images. The image shift is represented as the peak of this spectrum, and shift estimation is achieved by searching this peak. Searching the peak of the POC produces large error. Therefore, some methods propose estimating the sub-pixel shift directly from the phase component of the cross-power spectrum [1], [15]–[18]. In this case, phase wrapping ambiguity appears on the phase component to deal with. Hoge's method [15] applies rank-1 approximation using SVD to reduce errors of the unwrapping operation. Tong *et al.* [17] combine Hoge's SVD and unified random sample consensus (RANSAC) to improve the precision. Dong *et al.* [18]

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

model an additional noise component as a mixture of Gaussian and reduce an effect of the noise. However, these methods are sensitive to the spectrum aliasing.

In order to estimate the image rotation, the rotation is converted into the geometrical translation and estimate the translation by an algorithm such as POC. The combination of log-polar transform and 1D POC is one of these methods [3], [19], [20]. The method translates the frequency of the image into a polar coordinate. The image rotation is represented as a horizontal dimensional shift in this coordinate. Note that the amount of shift does not depend on the scaling or spatial shift of the image. On this basis, the rotation estimation is achieved by estimating the horizontal shift by the 1D POC. However, log-polar transform must approximate the frequency components using an interpolation algorithm, and the approximation precision deteriorates significantly in high-frequency components, which results in low precision for the estimated rotation. In order to reduce the effect of interpolation error, the new transforms are proposed such as MLFFT [21] or MPFT [22]. These methods still suffer from interpolation error.

Wan and Wei [23] proposed the rotation estimation algorithm based on Radon transform. This method converts the rotation into the geometrical shift and estimate the shift by the phase correlation method which is robust to noise. However, the precision of the phase correlation method is limited to integer precision. Therefore there is the limitation of the precision of the rotation estimation.

We improve the accuracy of the rotation estimation by introducing a new algorithm based on Radon transform and sub-pixel estimation in the phase domain. In the proposed method, the frequency component of input images are transformed by Radon transform. As well as log-polar transform, the rotation of the image is represented as a horizontal shift. On this bases, we estimate the degree of rotation by estimating the shift on the transformed images. Radon transform sums all pixels along the specified angle, which reduces the error caused by interpolating the image pixels.

Furthermore, we introduce a high-accuracy shift estimation based on the linear approximation of the phase component. Conventional algorithms search the peak of the POC by fitting a model function. In the estimation of discrete images, the precision of this search is limited to the integer. The proposed approach estimates the shift by linear approximation of the phase component. This method enables estimation of the non-integer part of the shift, which improves the precision of the sub-pixel shift.

We present an evaluation of the proposed method using standard images with different rotations and scales. The comparison of the estimation errors demonstrates that the proposed method achieves correct shift estimation for the degree of the scale and rotation. In addition, the results indicate that the proposed method is robust compared to feature-based and intensity-based approaches.

II. PREVIOUS ALGORITHM

A. ROTATION AND SCALE IN LOG-POLAR TRANSFORM

We define image $X_F(x, y)$ and its translation $X_G(x, y)$ using scale factor S, shift (δ_x, δ_y), and rotation angle θ_0 as follows:

$$X_G(x, y) = X_F(Sx\cos\theta_0 - Sy\sin\theta_0 - \delta_x, \quad Sx\sin\theta_0 + Sy\cos\theta_0 - \delta_y). \tag{1}$$

Both the images are translated to the frequency domain by Fourier transform.

$$\mathcal{F}_G(\omega_x, \omega_y) = \frac{1}{S^2}\mathcal{F}_F\Big(\frac{\omega_x}{S}\cos\theta_0 - \frac{\omega_y}{S}\sin\theta_0, \quad \frac{\omega_x}{S}\sin\theta_0 + \frac{\omega_y}{S}\cos\theta_0\Big)e^{-j(\omega_x\delta_x+\omega_y\delta_y)} \tag{2}$$

Here $\mathcal{F}_F$ and $\mathcal{F}_G$ denote the 2D Fourier transform of X_F and X_G respectively. Log-polar transform is defined as the coordinate translation of expressions $\mathcal{F}_F$ and $\mathcal{F}_G$ to log distance ρ and angle θ. In other words, ρ and θ are defined by ω_x and ω_y as follows:

$$\rho = \log\left(\sqrt{\omega_x^2 + \omega_y^2}\right), \quad \theta = \tan^{-1}(\omega_x/\omega_y). \tag{3}$$

We express $F(\rho, \theta)$ as the log-polar transform of $|\mathcal{F}_F|$ and $G(\rho, \theta)$ as the log-polar transform of $|\mathcal{F}_G|$. $|\mathcal{F}_F|$ is the magnitude of $\mathcal{F}_F$ and $|\mathcal{F}_G|$ is the magnitude of $\mathcal{F}_G$. Using these expressions, the relationship of the scaling and rotation (2) is rewritten by the shift in the ρ and θ directions respectively.

$$G(\rho, \theta) = \frac{1}{S^2}F(\rho - \log S, \theta + \theta_0) \tag{4}$$

B. SHIFT ESTIMATION USING PHASE-ONLY CORRELATION

On the basis of the relationship given by (4), the image scale and rotation are obtained by estimating the 1D shift. To estimate the rotation θ_0, S is assumed to be 1 and ρ is fixed to 0. Here, we use $F(\theta)$ and $G(\theta)$ to express $F(\rho, \theta)$ and $G(\rho, \theta)$, respectively, with fixed ρ. Note that rotation θ_0 satisfies the following relationship:

$$G(\theta) = F(\theta + \theta_0). \tag{5}$$

The major shift estimation algorithm uses POC. The POC of $F(\theta)$ and $G(\theta)$ is calculated as follows:

$$R(\theta) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(F(\theta)) \circ \mathcal{F}(G(\theta))^*}{|\mathcal{F}(F(\theta)) \circ \mathcal{F}(G(\theta))^*|}\right). \tag{6}$$

Here, $\mathcal{F}(\cdot)$ denotes the 2D Fourier transform of the matrix, $\mathcal{F}^{-1}(\cdot)$ is the inverse 2D Fourier transform, $\circ$ is element-wise matrix multiplication (division is also calculated in an element-wise manner), and $\mathcal{F}(G)^*$ is the complex conjugate of $\mathcal{F}(G)$.

POC $R(\theta)$ has a peak according to the shift of the images. Figure 1 (a) shows the POC characteristics, where the blue line represents the POC with a peak according to the shift (the yellow line). To search the best location of the peak, POC $R(\theta)$ is fitted to a shift of the sinc function as follows:

$$R(\theta) \simeq \frac{c}{\pi}\,\mathrm{sinc}\left(\frac{c(\theta - \theta_0)}{\pi}\right). \tag{7}$$

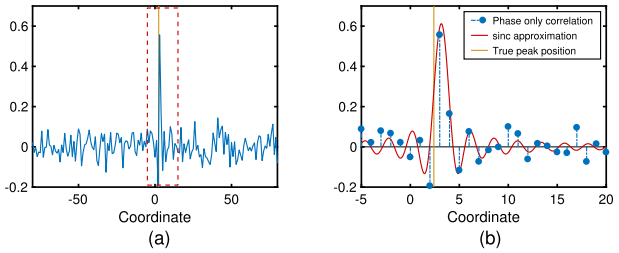

FIGURE 1. Peak expression in the POC and its approximation by sinc function. (a) Phase only correlation. (b) Fitting POC to the sinc function around the peak.

The approximation result of the sinc function is shown as the red line in Fig. 1 (b). In addition to the rotation, scale S can be estimated by fixing θ and fitting the 1D POC between $F(\rho)$ and $G(\rho)$.

In practice, log-polar transform is conducted on the discrete image; therefore, the value of angle θ is quantized into the multiply of the step θ_{step}. In the discrete expression, the precision of the non-integer shift is limited if the estimation is given by fitting the POC to the model function because the true peak of the POC may not appear in the discrete expression and the peak of the model function does not match the peak of the POC.

III. PROPOSED METHOD

Log-polar transform is the translation of image frequency $\mathcal{F}_F(\omega_x, \omega_y)$. This transform requires interpolation of the image frequency to obtain the specified frequency component $F(\rho, \theta)$; however, interpolation of the frequency component causes a large error, particularly in the high-frequency component. To suppress this interpolation error, we propose a new rotation estimation algorithm using Radon transform.

The proposed method applies Radon transform to the frequency domain and the spatial domain of images. From the Radon transform of the frequency response of the images, we can obtain the rotation estimation regardless of the shift. From the Radon transform of the spatial images, we can obtain the scale estimation. After compensating the rotation and scale of an image, we can match two images by estimating the 2D shift. The whole process of the proposed method obtaining rotation, scale and shift is described in Fig. 2. Section III-A describes the method to estimate rotation by Radon transform, section III-C describes the scale estimation method, and section III-D describes the 2D shift estimation.

A. ROTATION ESTIMATION BASED ON RADON TRANSFORM ON THE FREQUENCY DOMAIN

The Radon transform $F(\rho, \theta)$ of the Fourier transform $\left|\mathcal{F}_F(\omega_x, \omega_y)\right|$ of the image is calculated by summing all pixels along angle θ:

$$F(\rho, \theta) = \int_{-\infty}^{\infty} \left|\mathcal{F}_F(\xi\cos\theta - \rho\sin\theta,\ \xi\sin\theta + \rho\cos\theta)\right| d\xi. \tag{8}$$

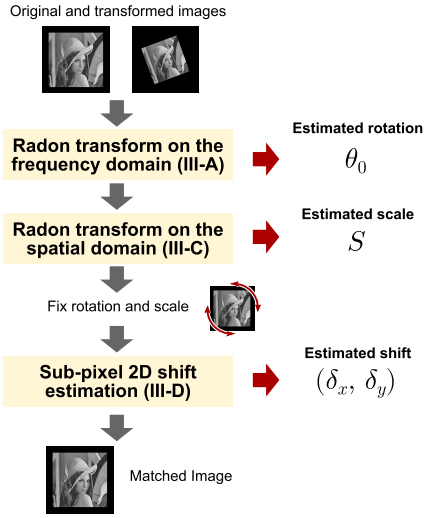

FIGURE 2. The whole process of the proposed method to estimate rotation, scale and shift.

Similarly, Radon transform $G(\rho, \theta)$ is calculated from $\left|\mathcal{F}_G(\omega_x, \omega_y)\right|$. In Radon transform F, G, scaling S, and the rotation θ_0 become

$$G(\rho, \theta) = S \cdot F\left(\frac{1}{S}\rho,\ \theta + \theta_0\right). \tag{9}$$

This equation means that the rotation of two images is expressed as the 1D shift for the θ direction. Using this feature, we estimate rotation θ_0 using the 1D shift estimation algorithm. Note that the shift (δ_x, δ_y) is cancelled because we take an magnitude of $\mathcal{F}_F$ and $\mathcal{F}_G$ before calculating Radon transform. We obtain $F(0, \theta)$ and $G(0, \theta)$, as shown in Fig. 4. If we express $F(0, \theta)$ and $G(0, \theta)$ as $F(\theta)$ and $G(\theta)$ respectively, the rotation estimation can be expressed as follows:

$$G(\theta) = F(\theta + \theta_0). \tag{10}$$

Figure 5 shows a comparison of the two transforms for rotation estimation. Radon transform adds all pixels along the specified angle θ. This operation reduces the influence of interpolation error that occurs in the points that are distant from a pole.

In the case of processing real images, (ω_x, ω_y) or (ρ, θ) have only discrete coordinates. In other words, if we define m, n as an integer value and ρ_{step}, θ_{step} as the precision of the transform, the coordinates (ρ, θ) is defined as follows. In the experiment, we set $\rho_{\text{step}} = 1$ and $\theta_{\text{step}} = 1°$ for each transform precision:

$$\rho = m\rho_{\text{step}}, \quad \theta = n\theta_{\text{step}}. \tag{11}$$

For calculating the Radon transform of $\mathcal{F}_F$, we assume ξ has the discrete value $\xi = q\xi_d$ and sum all pixels in the range of $q \in [q_{\min}, q_{\max}]$. $q_{\min}$ and $q_{\max}$ are the smallest/largest value where the coordinate $(q\xi_d\cos\theta - \rho\sin\theta, q\xi_d\sin\theta + \rho\cos\theta)$ is in the range of $\mathcal{F}_F$. Using these expressions, the Radon transform $F(\rho, \theta)$ can be calculated from the

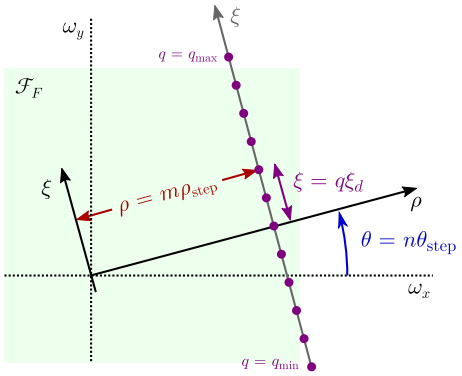

FIGURE 3. Calculating the discrete radon transform for $\mathcal{F}_F$.

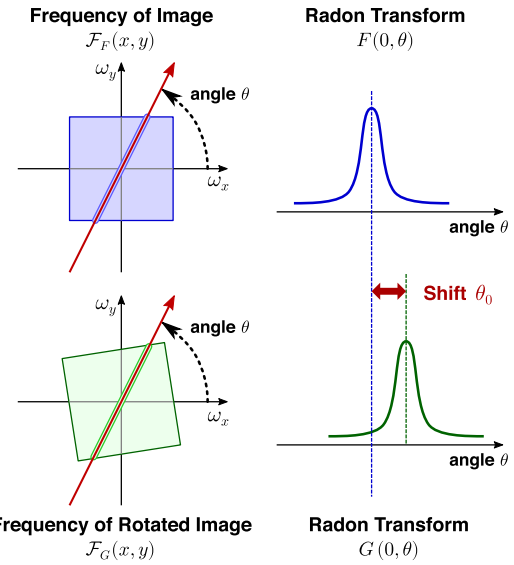

FIGURE 4. Rotation estimation using radon transform for the frequency responses.

following expression:

$$F(\rho, \theta) = \sum_{q=q_{\min}}^{q_{\max}} |\mathcal{F}_F(q\xi_d \cos\theta - \rho \sin\theta, q\xi_d \sin\theta + \rho \cos\theta) \cdot \xi_d|. \quad (12)$$

The step of ξ is set to $\xi_d = 0.25$. Figure 3 shows the methods to calculate $F(\rho, \theta)$ from $\mathcal{F}_F$. The Radon transform $F(\rho, \theta)$ is the sum of the values indicated by purple dots.

B. SUB-PIXEL SHIFT ESTIMATION BY LINEAR APPROXIMATION OF THE PHASE COMPONENT

As discussed in Section II-B, the precision of the non-integer part of shift θ_0 given by POC is uncertain. To improve the estimation of the non-integer shift, we introduce a non-integer shift estimation based on linear approximation of the phase component. We express the phase component of the correlation of $F(\theta)$ and $G(\theta)$ as $C(k)$. k is the axis of the transformed

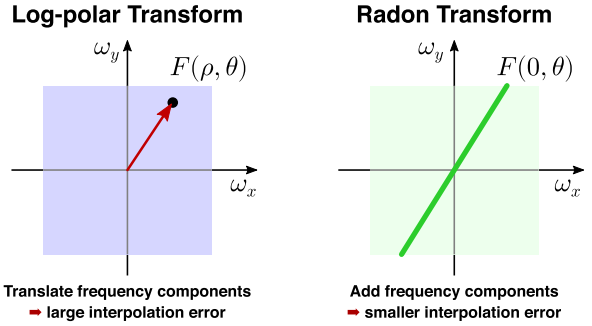

FIGURE 5. Comparison of log-polar transform and radon transform.

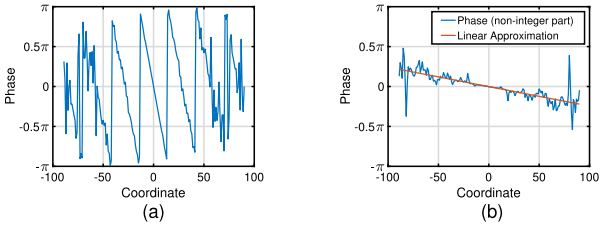

FIGURE 6. Linear approximation for the phase of non-integer shift. (a) Wrapped phase components with integer shift. (b) Unwrapped phase component and its linear approximation.

domain of θ.

$$r(k) = \frac{\mathcal{F}(F(\theta)) \circ \mathcal{F}(G(\theta))^*}{|\mathcal{F}(F(\theta)) \circ \mathcal{F}(G(\theta))^*|} \quad (13)$$

$$C(k) = \tan^{-1} \frac{\Im(r(k))}{\Re(r(k))} \quad (14)$$

$\Re(\cdot)$, $\Im(\cdot)$ denote the real and imaginary parts of the content, respectively. The correlation $C(k)$ can be approximated by the following linear expression:

$$C(k) \simeq a \cdot k \quad (15)$$

Thus, the coefficient a can be estimated by the following linear regression.

$$a = \underset{a}{\text{argmin}} \sum_k (C(k) - a \cdot k)^2 \quad (16)$$

However, the output range of $\tan^{-1}$ in (14) is limited to the range $[-\pi, \pi]$. As a result, $C(k)$ becomes the wrapped form shown in Fig. 6 a. To address this issue, we split phase component C into an integer component C' and a decimal component C''.

$$C'(k) \simeq a'k, \quad C''(k) \simeq a''k \quad (17)$$

$$C(k) = C'(k) + C''(k) \quad (18)$$

A diagram of the shift estimation splitting the integer and non-integer components is shown in Fig. 7. First, we estimate integer component a' by searching the peak of the conventional POC. Then, we subtract the phase of the integer shift from $C(k)$. We then obtain the phase of the decimal component $C''(k)$.

$$C''(k) = C(k) - a'k \quad (19)$$

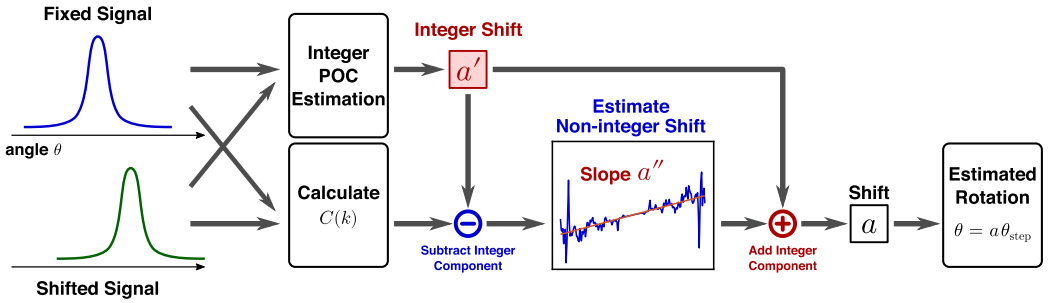

FIGURE 7. Sub-pixel shift estimation for the radon transformed axis.

Note that this expression does not contain the wrapping effect as shown in Fig. 6 b. By modifying the minimization (16) to the estimation of the non-integer part, we obtain the minimization of a'' as follows:

$$a'' = \underset{a}{\operatorname{argmin}} \sum_{k} \left(C''(k) - a \cdot k\right)^2 \tag{20}$$

By combining integer shift a' and non-integer shift a'' from (20), we obtain image rotation θ_0 as follows:

$$\theta_0 = a\theta_{\text{step}} = (a' + a'')\theta_{\text{step}} \tag{21}$$

C. RADON-TRANSFORM-BASED SCALE ESTIMATION

Using the rotation estimation described in the previous section, we obtain the scale estimation. For this purpose, we use Radon transform for the spatial images X_F and X_G. The Radon transform $F(\rho, \theta)$ of spatial image $X_F(x, y)$ is calculated by summing all pixels along angle θ.

$$F(\rho, \theta) = \int_{-\infty}^{\infty} X_F(\xi \cos\theta - \rho \sin\theta,\ \xi \sin\theta + \rho \cos\theta)\, d\xi \tag{22}$$

Similarly, Radon transform $G(\rho, \theta)$ is calculated from $X_G(x, y)$. In the real case, ρ, θ, ξ have the discrete value and Radon transform of the discrete image X_F can be calculated in the same manner with the section III-A. Using expressions (1) and (22), the relationship of Radon transform F, G, shift (δ_x, δ_y), scaling S, and the rotation θ_0 becomes the following expression:

$$G(\rho, \theta) = \frac{1}{S} F\left(S\rho - \sqrt{\delta_x^2 + \delta_y^2} \sin\left(\theta - \tan^{-1}\frac{\delta_y}{\delta_x}\right), \theta + \theta_0\right). \tag{23}$$

Then we adjust the rotation by the estimated value θ_0.

$$G'(\rho, \theta) = G(\rho, \theta - \theta_0) \tag{24}$$

$G'(\rho, \theta)$ is expressed by $F(\rho, \theta)$ as follows:

$$G'(\rho, \theta) = \frac{1}{S} F\left(S\rho - \sqrt{\delta_x^2 + \delta_y^2} \cdot \sin\left(\theta - \tan^{-1}\frac{\delta_y}{\delta_x}\right), \theta\right). \tag{25}$$

Using this expression, we can obtain scale factor S by calculating the ratio of coordinates whose values exceed a specified threshold. In other words, we express ρ_{F1} as the first coordinate whose value $F(\rho_{F1}, \theta)$ exceeds the threshold t_1 and ρ_{F2} as the last coordinate where $F(\rho_{F2}, \theta)$ exceeds the threshold t_2. Similarly, we express ρ_{G1}, ρ_{G2} for the coordinate of $G(\rho, \theta)$. The estimate of the scale is obtained as follows:

$$S \simeq \frac{\rho_{G1} - \rho_{G2}}{\rho_{F1} - \rho_{F2}} \tag{26}$$

To reduce interpolation error and calculation cost, we use the ρ-axis of the Radon transform in $\theta = 0°$ and $\theta = 90°$. In other words, the proposed method calculates $S_{\theta=0}$ from $F(\rho, 0)$ and $G(\rho, -\theta_0)$, and $S_{\theta=90}$ from $F(\rho, 90)$ and $G(\rho, 90 - \theta_0)$. Then, the proposed method obtains the estimated scale S as follows:

$$S = \frac{1}{2}(S_{\theta=0} + S_{\theta=90}) \tag{27}$$

The precision of the scale estimation is improved by taking the average of the two transform axes. Note that we selected transform angles $\theta = 0°$ and $\theta = 90°$ because Radon transform can be calculated with less interpolation error with these angles.

D. IMAGE MATCHING USING 2D SUB-PIXEL SHIFT ESTIMATION

After obtaining the rotation θ_0 and scale S, we can match two images $X_F(x, y)$ and $X_G(x, y)$ by fixing rotation and scale from X_G. If we define $X'_G(x, y)$ as a rotated and scaled image from X_G, X'_G has a 2D shift from X_F as follows:

$$X'_G = X_F(x - \delta_x, y - \delta_y). \tag{28}$$

In order to estimate the 2D shift of X_F and X'_G, we extend a shift estimation algorithm in III-B from 1D estimation to 2D estimation. We calculate the phase component of the correlation of X_F and X'_G as $C(k, l)$ by the following expression:

$$r(k, l) = \frac{\mathcal{F}(X_F(x, y)) \circ \mathcal{F}(X'_G(x, y))^*}{\left|\mathcal{F}(X_F(x, y)) \circ \mathcal{F}(X'_G(x, y))^*\right|}, \tag{29}$$

$$C(k, l) = \tan^{-1}\frac{\Im(r(k, l))}{\Re(r(k, l))}. \tag{30}$$

TABLE 1. Comparison of rotation estimation errors [degrees].

Image	Size	SURF	BRISK	Intensity	Log-polar	Proposed
Lenna	256	0.2411	0.1615	0.4668	0.2544	**0.1126**
Airplane	256	0.2132	**0.0946**	0.4152	0.3326	0.2074
Barbara	256	0.2294	0.1232	0.2306	0.3542	**0.0990**
Boats	256	0.2139	0.1394	0.4398	0.2306	**0.0513**
Brickhouse	256	0.3508	**0.1289**	0.3770	0.2522	0.1455
Bridge	256	0.1499	0.0881	0.2820	0.3964	**0.0850**
Cameraman	256	0.2321	0.1538	0.2932	0.2970	**0.0807**
Girl	256	0.2324	**0.0759**	0.2973	0.2778	0.1590
Goldhill	256	0.1580	0.1447	0.3089	0.2760	**0.0638**
Lighthouse	256	0.2008	0.0865	0.3733	0.2698	**0.0642**
Mandrill	256	0.3109	**0.0846**	0.3349	0.2806	0.1354
Milkdrop	256	0.3124	0.1348	0.3137	0.2353	**0.0713**
Parrots	256	0.3321	0.2159	0.1998	0.2578	**0.1481**
Pepper	256	0.1289	0.1575	0.4157	0.2395	**0.1133**
Sailboat	256	0.1921	**0.0432**	0.2736	0.3071	0.0839
Woman	256	0.1732	0.2129	0.2263	0.2352	**0.1573**
Airplane	512	0.0530	**0.0320**	0.3224	0.2951	0.2071
Barbara	512	0.0708	**0.0591**	0.3786	0.2483	0.1430
Boats	512	0.0994	**0.0378**	0.4226	0.3116	0.0818
Bridge	512	0.0498	**0.0227**	0.3417	0.3613	0.1176
Goldhill	512	0.0461	**0.0440**	0.4245	0.2732	0.0821
Houses	512	0.0584	**0.0257**	0.4065	0.4327	0.0717
Lenna	512	0.0819	**0.0711**	0.4166	0.2969	0.1052
Mandrill	512	0.0685	**0.0475**	0.3654	0.2712	0.1506
Milkdrop	512	**0.0727**	0.1110	0.4043	0.2866	0.0896
Pepper	512	0.0877	**0.0501**	0.4170	0.3005	0.1035
Sailboat	512	0.0502	**0.0168**	0.3276	0.3002	0.0981
Average		0.1483	**0.0875**	0.3562	0.2940	0.1124

TABLE 2. Comparison of average errors of rotation estimation for all images.

Size	Scale	SURF	BRISK	Intensity	Log-polar	Proposed
256	0.6	0.4031	0.2488	0.5566	0.3299	**0.1078**
	0.8	0.2020	**0.1071**	0.2717	0.3285	0.1122
	1.0	0.0832	**0.0565**	0.1557	0.1847	0.1134
512	0.6	0.1168	**0.0788**	0.5832	0.3385	0.1136
	0.8	0.0585	**0.0409**	0.3579	0.3013	0.1141
	1.0	0.0261	**0.0215**	0.2118	0.2814	0.1134

Here, $\mathcal{F}$ denotes 2D Fourier transform and (k, l) are the two axes of the transformed domain. This correlation can be approximated by the following linear expression.

$$C(k, l) \simeq \delta_x k + \delta_y l. \tag{31}$$

Thus, the coefficients (δ_x, δ_y) are approximated as follows:

$$\{\delta_x, \delta_y\} = \underset{\{\delta_x, \delta_y\}}{\operatorname{argmin}} \sum_{k,l} (C(k, l) - \delta_x k - \delta_y l)^2 \tag{32}$$

$C(k, l)$ has the wrapped form because of the output range of $\tan^{-1}$. To address this issue, we split phase component C into an integer component C' and a decimal component C''.

$$C'(k, l) \simeq \delta'_x k + \delta'_y l, \quad C''(k, l) \simeq \delta''_x k + \delta''_y l \tag{33}$$

$$C(k, l) = C'(k, l) + C''(k, l) \tag{34}$$

The integer shift (δ'_1, δ'_2) is estimated by the conventional 2D POC and subtracted from $C(k, l)$ to obtain decimal component $C''(k, l)$.

$$C''(k, l) = C(k, l) - \delta'_x k - \delta'_y l. \tag{35}$$

This expression does not have the wrapping effect. Therefore we obtain the non-integer part (δ'_x, δ'_y) by the following minimization:

$$\{\delta''_x, \delta''_y\} = \underset{\{\delta''_x, \delta''_y\}}{\operatorname{argmin}} \sum_{k,l} (C''(k, l) - \delta''_x k - \delta''_y l)^2 \tag{36}$$

We combine the integer shift (δ'_x, δ'_y) and non-integer shift (δ''_x, δ''_y) to obtain the shift with sub-pixel precision:

$$\delta_x = \delta'_x + \delta''_x, \quad \delta_y = \delta'_y + \delta''_y. \tag{37}$$

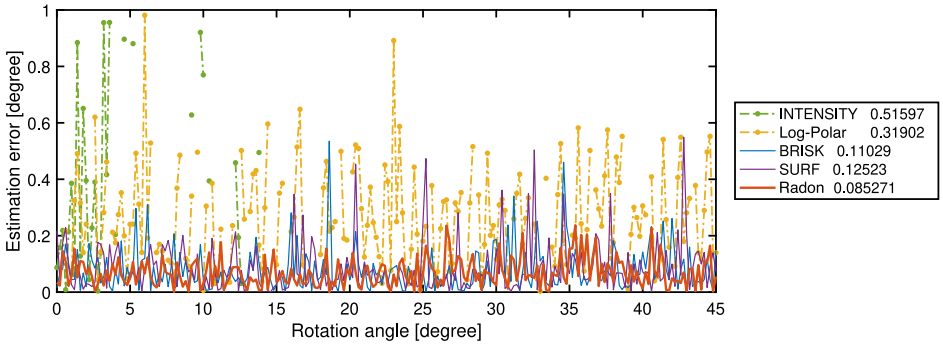

FIGURE 8. Rotation errors for the rotation estimation of lenna image (scale 0.6).

TABLE 3. Comparison of rotation estimation errors for noisy images with a standard deviation of 20 [degrees].

Image	Size	SURF	BRISK	Intensity	Log-polar	Proposed
Lenna	256	0.3373	0.2516	0.4248	0.3583	**0.1211**
Airplane	256	0.2889	**0.1315**	0.4726	0.3848	0.1942
Barbara	256	0.3643	N/A	0.2836	0.4493	**0.1103**
Boats	256	0.3115	N/A	0.4637	0.3006	**0.0744**
Brickhouse	256	0.3427	0.2419	0.3379	0.3148	**0.1588**
Bridge	256	0.2670	0.1493	0.2552	0.4381	**0.0920**
Cameraman	256	0.2910	0.1715	0.2947	0.3466	**0.0922**
Girl	256	0.3336	0.2477	0.3106	0.3962	**0.1843**
Goldhill	256	0.2784	0.2775	0.3068	0.3004	**0.0656**
Lighthouse	256	0.2585	0.1694	0.3822	0.3098	**0.0680**
Mandrill	256	0.2788	N/A	0.3223	0.3707	**0.1518**
Milkdrop	256	0.3499	0.1991	0.3090	0.3277	**0.0878**
Parrots	256	0.4039	0.2927	0.2917	0.3310	**0.1759**
Pepper	256	0.1911	0.1829	0.2997	0.3013	**0.1160**
Sailboat	256	0.2293	0.0997	0.2950	0.4002	**0.0962**
Woman	256	0.3296	0.2500	0.2166	0.3208	**0.1811**
Airplane	512	0.0852	**0.0537**	0.4457	0.2811	0.2144
Barbara	512	**0.0997**	0.1237	0.3611	0.3271	0.1609
Boats	512	0.1225	**0.0762**	0.3963	0.3041	0.0890
Bridge	512	0.0779	**0.0561**	0.3196	0.3630	0.1167
Goldhill	512	0.0956	0.1375	0.3115	0.2586	**0.0813**
Houses	512	0.0710	**0.0440**	0.4080	0.4750	0.0765
Lenna	512	0.1470	0.1441	0.4127	0.2682	**0.1162**
Mandrill	512	0.1193	**0.1173**	0.3404	0.3881	0.1588
Milkdrop	512	0.0952	0.1819	0.3535	0.2929	**0.0920**
Pepper	512	0.1090	**0.1010**	0.3188	0.2784	0.1091
Sailboat	512	0.0588	**0.0347**	0.3554	0.2925	0.0960
Average		0.1496	0.1242	0.3566	0.3289	**0.1202**

The amount of the estimated shift (δ_x, δ_y) strongly depends on the location of the centroid of rotation. Therefore we cannot compare the precision of shift estimation.

IV. EVALUATION

A. EXPERIMENTAL SETTINGS

To evaluate the proposed algorithm, we compared rotation and scale estimation for standard images. The images were rotated, scaled and shifted with a specified angle, scale value and shift value. Then, the estimation algorithm was applied, and the error between the estimated and ground-truth values was calculated. In this evaluation, we used 14 images with size 256 × 256 and 11 images with size 512 × 512. To compare the precision of the rotation estimation algorithm, images were rotated from 0° to 45° at steps of 0.2°. In addition, scales of 0.6, 0.8, and 1.0 were tested. When we compare scale estimation precision, the input rotation was fixed at 35° and the input scale was varied from 0.6 to 1.0 at a step of 0.01. To evaluate estimation robustness, white Gaussian noise was added to the fixed and shifted images. The standard deviation of the Gaussian noise was $\sigma = 0$ (no noise), $\sigma = 10$, and $\sigma = 20$. We compared the

TABLE 4. Comparison of average errors of rotation estimation for noisy images with a standard deviation of 20.

Size	Scale	SURF	BRISK	Intensity	Log-polar	Proposed
256	0.6	0.4417	0.2234	0.5100	0.3843	**0.1182**
	0.8	0.3021	0.2308	0.2930	0.3772	**0.1235**
	1.0	0.1768	0.1667	0.1844	0.2980	**0.1277**
512	0.6	0.1548	0.1376	0.5416	0.3692	**0.1164**
	0.8	**0.0885**	0.0945	0.3468	0.3282	0.1204
	1.0	**0.0517**	0.0615	0.2088	0.2650	0.1207

TABLE 5. Comparison of scale estimation errors for images rotated by 35°.

Image	Size	SURF	BRISK	Intensity	Log-polar	Proposed
Lenna	256	0.00487	0.00752	0.26553	0.00415	**0.00401**
Airplane	256	0.00660	0.00275	0.26139	0.00425	**0.00205**
Barbara	256	0.00520	N/A	0.26938	0.00522	**0.00309**
Boats	256	0.01114	0.01126	0.25324	0.00944	**0.00280**
Brickhouse	256	0.00927	0.01158	0.22652	0.00550	**0.00247**
Bridge	256	0.00459	0.00310	0.20188	**0.00287**	0.00325
Cameraman	256	0.00610	0.00308	0.24872	0.00309	**0.00264**
Girl	256	0.00555	0.00563	0.18613	**0.00246**	0.00377
Goldhill	256	0.00339	0.00460	0.23055	0.00625	**0.00300**
Lighthouse	256	0.00439	**0.00185**	0.10414	0.00530	0.00257
Mandrill	256	0.00779	N/A	0.30288	0.00375	**0.00271**
Milkdrop	256	N/A	0.02680	0.24015	**0.00255**	0.00658
Parrots	256	0.00864	0.00621	0.37079	**0.00233**	0.00330
Pepper	256	0.00265	0.00304	0.25068	0.00320	**0.00235**
Sailboat	256	0.00240	**0.00122**	0.42498	0.00519	0.00259
Woman	256	0.00305	0.00714	0.16196	0.00242	**0.00136**
Airplane	512	0.00119	**0.00051**	0.23599	0.00250	0.00148
Barbara	512	0.00113	0.00152	0.27915	0.00222	**0.00098**
Boats	512	0.00138	0.00092	0.15451	0.00162	**0.00091**
Bridge	512	0.00112	**0.00081**	0.21990	0.00148	0.00126
Goldhill	512	0.00121	0.00234	0.27085	0.00409	**0.00080**
Houses	512	0.00103	**0.00081**	0.24387	0.00301	0.00144
Lenna	512	0.00129	0.00262	0.23907	0.00189	**0.00100**
Mandrill	512	0.00124	0.00184	0.29119	0.00186	**0.00116**
Milkdrop	512	0.00180	0.00219	0.23572	0.00150	**0.00104**
Pepper	512	0.00149	0.00098	0.24259	0.00200	**0.00093**
Sailboat	512	0.00086	**0.00060**	0.44799	0.00332	0.00060
Average		0.00382	0.00444	0.25406	0.00346	**0.00223**

proposed algorithm to the conventional estimation algorithm, two feature-based algorithms, i.e., SURF [11] and BRISK features [12], the intensity-based algorithm using Mattes' mutual information [8], and the POC-based algorithm using log-polar transform [3].

B. ROTATION ESTIMATION ERROR

Table 1 lists the rotation estimation errors for all images. Here, the results for the different input scales (0.6, 0.8, and 1.0) are averaged. "N/A" means that there were few BRISKs in the image and rotation estimation was unavailable. The proposed method outperformed the log-polar transform. Table 2 shows the rotation estimation errors for each scale. Note that the errors of all images are averaged. The proposed Radon-based method yielded higher precision, particularly with small-scale images. The results demonstrate that the proposed method can estimate rotation regardless of scale.

The rotation estimation errors for the Lenna image (scale 0.6) are shown in Fig. 8. As can be seen, BRISK features achieve the higher estimation precision than SURF features. However, the number of BRISKs tend to be small, and estimation fails if the number of keypoints is less than three. The intensity-based method estimates with the lower error under 15°; however, estimating larger angles is difficult, and an inappropriate local solution produces a large estimation error.

Tables 3 and 4 compare the rotation estimation errors for the noisy images (standard deviation $\sigma = 20$). As can be seen, the proposed method achieves rotation estimation that is robust against noise, and feature-based methods return large estimation errors or "N/A" values because of the small number of feature points.

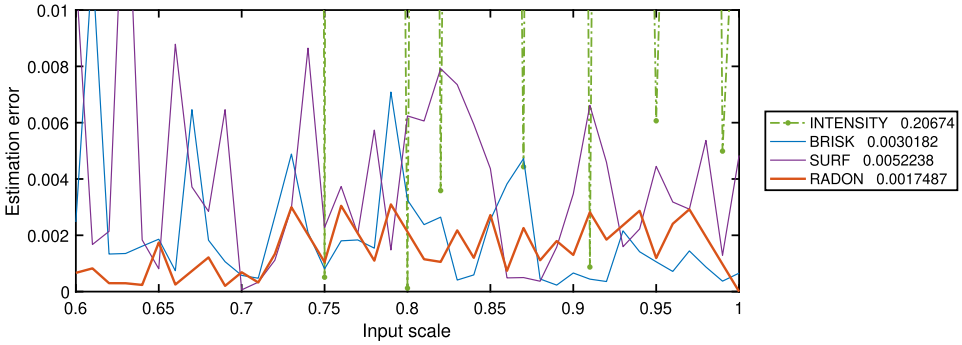

FIGURE 9. Comparison of errors of scale estimation for milkdrop image with rotation angle 35°.

TABLE 6. Comparison of execution times for rotation and scale estimation.

Size	Scale	SURF	BRISK	Intensity	Log-Polar	Proposed
256	0.6	0.0452	0.3161	0.8353	0.0214	**0.0149**
256	0.8	0.0427	0.4650	0.9745	0.0261	**0.0150**
256	1.0	0.0482	0.5050	1.0480	0.0363	**0.0182**
512	0.6	0.0874	0.4451	2.4960	0.0709	**0.0311**
512	0.8	0.0989	0.4477	2.8349	0.0871	**0.0367**
512	1.0	0.1230	0.4864	3.1434	0.1170	**0.0471**

C. SCALE ESTIMATION ERROR

The scale estimation errors for each images rotated by 35° are compared in Table 5. These results demonstrate that the proposed Radon-based transform can estimate the image scale with less error compared to feature-based and intensity-based methods, particularly when the image size is 256. The proposed method takes the average of two scale estimations using (27) to reduce the estimation error, which improves estimation precision for images of size 512. Figure 9 compares the scale estimation errors for the Milkdrop image. The proposed algorithm demonstrates higher estimation precision when the input scale is in the range of 0.6 to 0.7. In this range, it is difficult for the feature or intensity-based methods to estimate the scale value accurately.

D. EXECUTION TIME

We compared the comparison of the execution times of the rotation and scale estimation algorithms. All angles were evaluated from 0° to 45° in 0.2° steps, and the results were averaged. The rotation and scale estimation for each image size (256 and 512) and for each scale (0.6 to 1.0) are compared in Table 6. For all image sizes and scales, compared to the conventional methods, the proposed method demonstrates a faster execution time because the Radon transform calculation is limited to three axes, and each estimation is processed in 1D signals.

V. CONCLUSION

In this paper, we have proposed a rotation and scale estimation algorithm based on Radon transform and sub-pixel shift estimation. Radon transform can reduce the influence of interpolation error because it is applied to a spatial image rather than the frequency. The number of ranges of Radon transform can be limited to $\rho = 0$, $\theta = 0°$, and $\theta = 90°$, therefore, the proposed method has an advantage relative to execution time. As well as introducing the Radon transform, we have proposed a sub-pixel shift estimation method that uses linear approximation of the phase component, which enables an accurate non-integer shift. In the proposed method, we have combined the conventional POC method for the integer part of the shift and linear approximation for the non-integer part. Evaluations using test image demonstrate that the proposed method accurately estimates rotation compared to other POC-based approaches regardless of the degree of the input rotation, scaling, and shifting. In particular, the proposed method realizes robust estimation when the input scale differs from 1 or noise is added to the images.

REFERENCES

[1] M. Balci and H. Foroosh, "Subpixel estimation of shifts directly in the Fourier domain," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1965–1972, Jul. 2006.

[2] T. Dobashi and H. Kiya, "A parallel implementation method of FFT-based full-search block matching algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 2644–2648.

[3] X. Zhang, N. Homma, K. Ichiji, M. Abe, N. Sugita, and M. Yoshizawa, "A faster 1-D phase-only correlation-based method for estimations of translations, rotation and scaling in images," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 97, no. 3, pp. 809–819, Mar. 2014. [Online]. Available: http://ci.nii.ac.jp/naid/40019997382/

[4] T. Tsuboi and S. Hirai, "Detection algorithm of position and orientation of planar motion objects using Radon transform and one-dimensional phase-only matched filter," in *Proc. JSME Annu. Conf. Robot. Mechatron. (Robomec)*, 2003, p. 2P1-1F-C5. [Online]. Available: http://ci.nii.ac.jp/naid/110002506417/

[5] S. Hirai, T. Tsuboi, A. Masubuchi, and M. Zakouji, "Parallel processing of one-sided Radon transform for the detection of the position and the orientation of planar motion objects," *IEICE Tech. Report. Digit. Signal Process.*, vol. 101, no. 384, pp. 31–38, Oct. 2001. [Online]. Available: http://ci.nii.ac.jp/naid/110003280532/

[6] P. Thevenaz and M. Unser, "A pyramid approach to sub-pixel image fusion based on mutual information," in *Proc. 3rd IEEE Int. Conf. Image Process.*, vol. 1, Sep. 1996, pp. 265–268.
[7] P. Thevenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 27–41, Jan. 1998.
[8] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, "Nonrigid multimodality image registration," *Proc. SPIE*, vol. 4322, pp. 1609–1620, Jul. 2001. doi: 10.1117/12.431046.
[9] M. Styner, C. Brechbuhler, G. Szckely, and G. Gerig, "Parametric estimate of intensity inhomogeneities applied to MRI," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 153–165, Mar. 2000.
[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004. doi: 10.1023/B:VISI.0000029664.99615.94.
[11] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314207001555
[12] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput.*, Nov. 2011, pp. 2548–2555.
[13] M. Muja and D. G. Lowe, "Fast matching of binary features," in *Proc. Comput. Robot Vis.*, May 2012, pp. 404–410.
[14] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proc. VISAPP Int. Conf. Comput. Vis. Theory Appl.*, 2009, pp. 331–340.
[15] W. S. Hoge, "A subspace identification extension to the phase correlation method [MRI application]," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 277–280, Feb. 2003.
[16] S. Leprince, S. Barbot, F. Ayoub, and J. P. Avouac, "Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1529–1558, Jun. 2007.
[17] X. Tong *et al.*, "A novel subpixel phase correlation method using singular value decomposition and unified random sample consensus," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4143–4156, Aug. 2015.
[18] Y. Dong, T. Long, W. Jiao, G. He, and Z. Zhang, "A novel image registration method based on phase correlation using low-rank matrix factorization with mixture of Gaussian," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 446–460, Jan. 2018.
[19] H. Sasaki, K. Kobayashi, T. Aoki, M. Kawamata, and T. Higuchi, "Rotation measurements using rotation invariant phase only correlation," *ITE Tech. Rep.*, vol. 22, no. 45, pp. 55–60, Sep. 1998. [Online]. Available: http://ci.nii.ac.jp/naid/110003690056/
[20] S. Nagashima, K. Ito, T. Aoki, H. Ishii, and K. Kobayashi, "A high-accuracy rotation estimation algorithm based on 1D phase-only correlation," in *Proc. Int. Conf. Image Anal. Recognit. (ICIAR)*, M. Kamel and A. Campilho, Eds. Berlin, Germany: Springer, 2007, pp. 210–221. doi: 10.1007/978-3-540-74260-9_19.
[21] W. Pan, K. Qin, and Y. Chen, "An adaptable-multilayer fractional Fourier transform approach for image registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 400–414, Mar. 2009.
[22] Y. Dong, W. Jiao, T. Long, G. He, and C. Gong, "An extension of phase correlation-based image registration to estimate similarity transform using multiple polar Fourier transform," *Remote Sens.*, vol. 10, no. 11, p, 1719, 2018. [Online]. Available: http://www.mdpi.com/2072-4292/10/11/1719
[23] Y. Wan and N. Wei, "A fast algorithm for recognizing translated, rotated, reflected, and scaled objects from only their projections," *IEEE Signal Process. Lett.*, vol. 17, no. 1, pp. 71–74, Jan. 2010.

TAKANORI FUJISAWA received the B.E. and M.E. degrees in electrical engineering from Keio University, Yokohama, Japan, in 2014 and 2016, respectively, where he is currently pursuing the M.E. degree under the supervision of Prof. M. Ikehara. His research interests include applications of sparse coding to image super-resolution, image inpainting, and automatic music transcription.

MASAAKI IKEHARA received the B.E., M.E., and Dr. Eng. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1984, 1986, and 1989, respectively. He was an Appointed Lecturer with Nagasaki University, Nagasaki, Japan, from 1989 to 1992. In 1992, he joined the Faculty of Engineering, Keio University, where he is currently a Full Professor with the Department of Electronics and Electrical Engineering. From 1996 to 1998, he was a Visiting Researcher with the University of Wisconsin, Madison, and Boston University, Boston, MA, USA. His research interests include multi-rate signal processing, wavelet image coding, and filter design problems.