Short Papers

Robust FFT-Based Scale-Invariant Image Registration with Image Gradients

Georgios Tzimiropoulos, *Member*, *IEEE*,
Vasileios Argyriou, *Member*, *IEEE*,
Stefanos Zafeiriou, *Member*, *IEEE*, and
Tania Stathaki

Abstract—We present a robust FFT-based approach to scale-invariant image registration. Our method relies on FFT-based correlation twice: once in the log-polar Fourier domain to estimate the scaling and rotation and once in the spatial domain to recover the residual translation. Previous methods based on the same principles are not robust. To equip our scheme with robustness and accuracy, we introduce modifications which tailor the method to the nature of images. First, we derive efficient log-polar Fourier representations by replacing image functions with complex gray-level edge maps. We show that this representation both captures the structure of salient image features and circumvents problems related to the low-pass nature of images, interpolation errors, border effects, and aliasing. Second, to recover the unknown parameters, we introduce the normalized gradient correlation. We show that, using image gradients to perform correlation, the errors induced by outliers are mapped to a uniform distribution for which our normalized gradient correlation features robust performance. Exhaustive experimentation with real images showed that, unlike any other Fourier-based correlation techniques, the proposed method was able to estimate translations, arbitrary rotations, and scale factors up to 6.

Index Terms—Global motion estimation, correlation methods, FFT, scale-invariant image registration, frontal view face registration.

1 Introduction

The estimation of the relative motions between two or more images is probably at the heart of any autonomous system which aims at the efficient processing of visual information. Motions in images are induced due to camera displacements or displacements of the individual objects composing the scene. Image registration methods for global motion estimation address the problem of compensating for the camera ego-motion and finally aligning the images. Applications are numerous: from global scene representation and image mosaicking to object detection/tracking and video compression.

In this work, we focus on global registration schemes which make use of all image information. In particular, we propose a robust correlation-based scheme which operates in the Fourier domain for the estimation of translations, rotations, and scalings in images. For the class of similarity transforms, a frequency-domain approach to motion estimation possesses several appealing properties. First, through the use of correlation, it enables an exhaustive search for the unknown motion parameters, and therefore, large motions can be recovered with no a priori information (good initial guess). Second, the approach is global which equips the algorithm with robustness to noise. Third, the method is computationally efficient. This comes from the *shift property* of the Fourier Transform (FT) and the use of FFT routines for the rapid computation of correlations.

The work in [1] introduces the basic principles for translation, rotation, and scale-invariant image registration in the frequency domain. Given two images related by a similarity transform, the translational displacement does not affect the magnitudes of the FTs of the two images. Resampling the Fourier magnitudes on the log-polar grid reduces the problem of estimating the rotation and scaling to one of estimating a 2D translation. Thus, the method relies on correlation twice: once in the log-polar Fourier domain to estimate the rotation and scaling and once in the spatial domain to recover the residual translation. In the usual way, the authors use phase correlation (PC) [2] instead of standard correlation while they perform conversion from Cartesian to log-polar using standard interpolation schemes (e.g., bilinear interpolation).

To enhance accuracy, the authors in [3], [4], [5] introduce new sampling schemes and algorithms which reduce the inaccuracies induced by resampling the magnitude of the FT on the log-polar grid. To recover the rotation and scaling, the method in [3] relies on the pseudopolar FFT [6], which rapidly computes a discrete FT on a nearly polar grid. The pseudopolar grid serves as an intermediate step for a log-polar Fourier representation which is obtained using nearest-neighbor interpolation. Overall, the total accumulated interpolation error is decreased; nevertheless, the pseudopolar FFT is not a true polar Fourier representation and the method estimates the rotation and scaling in an iterative fashion. In [4], the authors propose to approximate the log-polar DFT by interpolating the pseudo-log-polar FFT. The method is noniterative but the gain in registration accuracy is not significant. The main idea in [5] is to obtain more accurate log-polar DFT approximations by efficiently oversampling the lower part of the Fourier spectrum using the Fractional FFT. The presented experimental results do not explicitly show the applicability of the algorithm in real images related by large-scale factors while oversampling inevitably increases the execution time.

The work in [7] introduces a robust technique to handle arbitrary rotations and large translations and scale factors. It leverages the log-polar transform in the spatial domain to achieve these results. The main contribution of our work is to demonstrate that FFT-based scale-invariant registration in real images is also feasible even for large-scale factors (up to 6), arbitrary rotations, and large translations.

In particular, we provide reasoning and experimentation which show that robustness in FFT-based scale-invariant image registration depends on the image representation used and the type of correlation employed rather than the method used to approximate the log-polar DFT. In our scheme, we first replace image functions with complex gray-level edge maps and then compute the standard Cartesian FFT. Using simple arguments, we show that this step both captures the structure of salient image features and provides an efficient solution to problems induced by the low-pass nature of images, interpolation errors, border effects, and aliasing. Next, we simply resample the Cartesian FFT on the log-polar grid using bilinear interpolation. Neither sophisticated FFT nor oversampling is employed to enhance accuracy. To perform robust correlation, we replace phase correlation with gradient-based correlation schemes [8], [9]. We present a novel theoretical analysis which shows that under a reasonable assumption, the use of image gradients tailors correlation to the nature of real images and provides a mechanism to reject outliers induced by real-world registration problems. Following our analysis, we introduce the normalized gradient correlation (NGC), and finally, we estimate the rotation and scaling using NGC in the log-polar Fourier

- *G. Tzimiropoulos, S. Zafeiriou, and T. Stathaki are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. E-mail: {gt204, szafeiri, t.stathaki}@imperial.ac.uk.*
- *V. Argyriou is with the Faculty of Computing, Information Systems, and Mathematics, Kingston University London, Penrhyn Road, Surrey KT1 2EE, U.K. E-mail: vasileios.argyriou@kingston.ac.uk.*

Manuscript received 12 Mar. 2009; revised 23 Oct. 2009; accepted 14 Apr. 2010; published online 13 May 2010.
Recommended for acceptance by C. Stewart.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2009-03-0159.
Digital Object Identifier no. 10.1109/TPAMI.2010.107.

domain. Contrary to the common belief that FFT-based schemes are unable to handle real-world registration problems [7], exhaustive experimentation with popular image data sets demonstrates that, unlike any other Fourier-based techniques, our formulation provides a fast and robust framework for scale-invariant image registration.

The rest of the paper is organized as follows: Section 2 gives the necessary background in scale-invariant FFT-based image registration using correlation. Section 3 presents in detail the key features of the proposed scheme. We present performance evaluation experiments in Section 4, while Section 5 presents results for the application of frontal view face registration. Finally, Section 6 summarizes the contributions of this work.

2 FFT-Based Scale-Invariant Image Registration

Let $I_i(\mathbf{x}), \mathbf{x} = [x, y]^T \in \mathcal{R}^2$, $i = 1, 2$, be two image functions. We denote by $\widehat{I}_i(\mathbf{k})$, $\mathbf{k} = [k_x, k_y]^T \in \mathcal{R}^2$, the Cartesian FT of I_i and M_i the magnitude of $\widehat{I}_i$. Polar and log-polar Fourier representations refer to computing the FT as a function of $\mathbf{k}_p = [k_r, k_\theta]^T$ and $\mathbf{k}_l = [\log k_r, k_\theta]^T$, respectively, where $k_r = \sqrt{k_x^2 + k_y^2}$ and $k_\theta = \arctan(k_y/k_x)$.

2.1 Translation Estimation Using Correlation

Assume that we are given two images, I_1 and I_2, related by an unknown translation $\mathbf{t} = [t_x, t_y]^T \in \mathcal{R}^2$:

$$I_2(\mathbf{x}) = I_1(\mathbf{x} + \mathbf{t}). \tag{1}$$

We can estimate $\mathbf{t}$ from the $2D$ cross-correlation function $\mathrm{C}(\mathbf{u})$, $\mathbf{u} = [u, v]^T \in \mathcal{R}^2$ as $\widehat{\mathbf{t}} = \arg_{\mathbf{u}} \max\{\mathrm{C}(\mathbf{u})\}$, where[1]

$$\mathrm{C}(\mathbf{u}) \stackrel{\triangle}{=} I_1(\mathbf{u}) \star I_2(-\mathbf{u}) = \int_{\mathcal{R}^2} I_1(\mathbf{x}) I_2(\mathbf{x} + \mathbf{u}) d\mathbf{x}. \tag{2}$$

From the *convolution theorem* of the FT [10], C can be alternatively obtained by

$$\mathrm{C}(\mathbf{u}) = F^{-1}\left\{\widehat{I}_1(\mathbf{k})\widehat{I}_2^*(\mathbf{k})\right\}, \tag{3}$$

where F^{-1} is the inverse FT and $*$ denotes the complex conjugate operator. The *shift property* of the FT [10] states that if I_1 and I_2 are related by (1), then, in the frequency domain, it holds

$$\widehat{I}_2(\mathbf{k}) = \widehat{I}_1(\mathbf{k}) e^{j\mathbf{k}^T\mathbf{t}} \tag{4}$$

and therefore (3) becomes

$$\mathrm{C}(\mathbf{u}) = F^{-1}\left\{M_1^2(\mathbf{k}) e^{-j\mathbf{k}^T\mathbf{t}}\right\}. \tag{5}$$

The above analysis summarizes the main principles of frequency domain correlation-based translation estimation. For finite discrete images of size $N \times N$, correlation is efficiently implemented through (3) by zero padding the images to size $(2N-1) \times (2N-1)$ and using FFT routines to compute the forward and inverse FTs. If no zero padding is used, the match is cyclic and, in this case, the algorithm's complexity is $O(N^2 \log N)$.

2.2 Estimation of Translation, Rotation, and Scaling Using Correlation

Assume that we are given two images, I_1 and I_2, related by a translation $\mathbf{t}$, rotation $\theta_0 \in [0, 2\pi)$, and scaling $s > 0$, that is,

$$I_2(\mathbf{x}) = I_1(D\mathbf{x} + \mathbf{t}), \tag{6}$$

where $D = s\Theta$ and $\Theta = \begin{bmatrix} \cos\theta_0 & \sin\theta_0 \\ -\sin\theta_0 & \cos\theta_0 \end{bmatrix}$. In the Fourier domain, it holds [11] that

$$\widehat{I}_2(\mathbf{k}) = (1/|\Delta|)\widehat{I}_1(\mathbf{k}') e^{j\mathbf{k}'^T\mathbf{t}}, \tag{7}$$

where

$$\mathbf{k}' = D^{-T}\mathbf{k} \tag{8}$$

and Δ is the determinant of D. Taking the magnitude in both parts of (7) and substituting $D^{-T} = \Theta/s$, $\Delta = s^2$ gives

$$M_2(\mathbf{k}) = (1/|\Delta|)M_1(\mathbf{k}') = (1/s^2)M_1(\Theta\mathbf{k}/s). \tag{9}$$

Using the log-polar representation gives (ignoring $1/s^2$) [1]

$$M_2(\mathbf{k}_l) = M_1(\mathbf{k}_l - [\log s, \theta_0]^T). \tag{10}$$

We can observe that in the log-polar Fourier magnitude domain, the rotation and scaling reduce to a 2D translation which can be estimated using correlation. After compensating for the rotation and scaling, we can recover the remaining translation using correlation in the spatial domain. Note that if $\widetilde{\theta}_0$ is the estimated rotation, then it is easy to show that $\widetilde{\theta}_0 = \theta_0$ or $\widetilde{\theta}_0 = \theta_0 + \pi$. To resolve the ambiguity, one needs to compensate for both possible rotations, compute the correlation functions, and, finally, choose as the valid solution the one that yields the highest peak [1].

3 Robust FFT-Based Scale-Invariant Image Registration

3.1 Robust Translation Estimation Using Normalized Gradient Correlation

To estimate the translational displacement, we can replace standard correlation with gradient-based correlation schemes. Gradient correlation (GC) combines the magnitude and orientation of image gradients [9]:

$$\mathrm{GC}(\mathbf{u}) \stackrel{\triangle}{=} G_1(\mathbf{u}) \star G_2^*(-\mathbf{u}) = \int_{\mathcal{R}^2} G_1(\mathbf{x}) G_2^*(\mathbf{x} + \mathbf{u})\, d\mathbf{x}, \tag{11}$$

where

$$G_i(\mathbf{x}) = G_{i,x}(\mathbf{x}) + jG_{i,y}(\mathbf{x}) \tag{12}$$

and $G_{i,x} = \nabla_x I_i$ and $G_{i,y} = \nabla_y I_i$ are the gradients along the horizontal and vertical direction, respectively. Orientation correlation (OC) considers orientation information solely [8] by imposing

$$G_i(\mathbf{x}) \leftarrow \begin{cases} G_i(\mathbf{x})/|G_i(\mathbf{x})|, & \text{if } |G_i(\mathbf{x})| > \epsilon, \\ 0, & \text{otherwise,} \end{cases} \tag{13}$$

and ϵ is the value of a threshold. Thresholding $|G_i(\mathbf{x})|$ removes the contribution of pixels where gradient magnitude takes negligible values. In the following analysis, we focus primarily on GC.

3.1.1 Frequency-Domain Analysis

From (5), we observe that the phase difference term $e^{-j\mathbf{k}^T\mathbf{t}}$, which contains the translational information, is weighted by the magnitude M_1. In practice, in cases where (1) holds approximately only and $M_1 \neq M_2$, we estimate the translational displacement through (3). In this case, the phase difference function is weighted by the term $M_1 M_2$. Due to the low-pass nature of images, the weighting operation results in a correlation function with broad peaks of large magnitude and a dominant peak whose maximum is not always located at the correct displacement.

To tackle the problem, phase correlation (PC) [2] considers the phase difference function solely

$$\mathrm{PC}(\mathbf{u}) \stackrel{\triangle}{=} F^{-1}\left\{\frac{\widehat{I}_2(\mathbf{k})\widehat{I}_1^*(\mathbf{k})}{|\widehat{I}_2(\mathbf{k})||\widehat{I}_1^*(\mathbf{k})|}\right\} = F^{-1}\left\{e^{j\mathbf{k}^T\mathbf{t}}\right\} = \delta(\mathbf{u} - \mathbf{t}). \tag{14}$$

Thus, the resulting correlation function will be a 2D Dirac located at the unknown translation. In the presence of noise and dissimilar

1. To be more precise, we assume hereafter that the images are of finite energy such that correlation integrals such as the one in (2) converge.

parts in the two images, the value of the peak is significantly reduced and the method may become unstable [1].

GC is an approach which lies somewhere between C and PC. In the frequency domain, we have

$$\mathrm{GC}(\mathbf{u}) = F^{-1}\{\widehat{G}_1(\mathbf{k})\widehat{G}_2^*(\mathbf{k})\}. \tag{15}$$

Spatial-domain differentiation is equivalent to high-pass filtering in the Fourier domain. Taking the FT in both parts of (12) yields

$$\widehat{G}_i(\mathbf{k}) = jk_x\widehat{I}_i(\mathbf{k}) - k_y\widehat{I}_i(\mathbf{k}). \tag{16}$$

Plugging (16) into (15) and using (4), we get

$$\mathrm{GC}(\mathbf{u}) = F^{-1}\{M_{G_1}^2(\mathbf{k})e^{-j\mathbf{k}^T\mathbf{t}}\}, \tag{17}$$

where M_{G_i} denotes the magnitude of $\widehat{G}_i$:

$$M_{G_i}(\mathbf{k}) = |\widehat{G}_i| = k_r M_i(\mathbf{k}). \tag{18}$$

In this case, the weighting operation results in a peak of large magnitude in the GC surface with very good localization accuracy.

3.1.2 Spatial Domain Analysis

From the definition of GC and using (12), we can easily derive

$$\begin{aligned}\mathrm{GC}(\mathbf{u}) = {} & G_{1,x}(\mathbf{u}) \star G_{2,x}(-\mathbf{u}) + G_{1,y}(\mathbf{u}) \star G_{2,y}(-\mathbf{u}) \\ & + j\{-G_{1,x}(\mathbf{u}) \star G_{2,y}(-\mathbf{u}) + G_{1,y}(\mathbf{u}) \star G_{2,x}(-\mathbf{u})\}.\end{aligned} \tag{19}$$

The imaginary part in the above equation is equal to zero;[2] therefore

$$\mathrm{GC}(\mathbf{u}) = G_{1,x}(\mathbf{u}) \star G_{2,x}(-\mathbf{u}) + G_{1,y}(\mathbf{u}) \star G_{2,y}(-\mathbf{u}). \tag{20}$$

Using the polar representation of complex numbers, we define

$$R_i = \sqrt{G_{i,x}^2 + G_{i,y}^2}$$

and $\Phi_i = \arctan G_{i,y}/G_{i,x}$. Based on this representation, (20) takes the form

$$\mathrm{GC}(\mathbf{u}) = \int_{\mathcal{R}^2} R_1(\mathbf{x})R_2(\mathbf{x}+\mathbf{u})\cos[\Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x}+\mathbf{u})]d\mathbf{x}. \tag{21}$$

Equation (21) shows that GC is a joint metric that consists of two terms, each of which can be an error metric itself. The first term is the correlation of the gradient magnitudes R_i. The magnitudes R_i reward pixel locations with strong edge responses and suppress the contribution of areas of constant intensity level which do not provide any reference points for motion estimation. The second term is a cosine kernel applied on gradient orientations. This term is responsible for the Dirac-like shape of GC and its ability to reject outliers induced by the presence of dissimilar parts in the two images.

To show the latter point, let us first define the orientation difference function

$$\Delta\Phi_{\mathbf{u}}(\mathbf{x}) = \Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x}+\mathbf{u}). \tag{22}$$

For a fixed $\mathbf{u} \neq \mathbf{t}$, we recall that the images do not match, and therefore, it is not unreasonable to assume that, for any spatial location $\mathbf{x}_0 \in \mathcal{R}^2$, the difference in gradient orientation $\Delta\Phi_{\mathbf{u}}(\mathbf{x}_0)$ can take any value in the range $[0, 2\pi)$ with equal probability. Thus, for $\mathbf{u} \neq \mathbf{t}$, we assume that $\Delta\Phi_{\mathbf{u}}(\mathbf{x})$ is a stationary random process $y(t)$, with "time" index $t \triangleq \mathbf{x} \in \mathcal{R}^2$, which $\forall t$ follows a uniform distribution $U(0, 2\pi)$. If we define the random process $z(t) = \cos y(t)$, then it is not difficult to show that $\forall t$ the random variable $Z = z(t)$ has a density function $f_Z(z) = 1/\{\pi\sqrt{1-z^2}\}$ defined in $[-1, 1]$ with mean value $E\{Z\} = 0$.

The integral $s = \int_a^b z(t)dt$ of the stochastic process $z(t)$ is also a random variable s. By interpreting the above as a Riemannian integral and using the linearity of the expectation operator, we conclude that

$$E\{s\} = \int_a^b E\{z(t)\}dt = \int_a^b \int_{-\infty}^{+\infty} z f_Z(z)dz = 0. \tag{23}$$

The above result shows that the integral $\int_{\mathcal{R}^2} \cos\Delta\Phi_{\mathbf{u}}(\mathbf{x})d\mathbf{x}$ is equal to zero in mean value.

We can derive a stricter result by further assuming mean-ergodicity. In this case, the time average is equal to the mean. Thus, we get

$$E(Z) \propto \int z(t)dt \equiv \int_{\mathcal{R}^2} \cos\Delta\Phi_{\mathbf{u}}(\mathbf{x})d\mathbf{x} = 0, \mathbf{u} \neq \mathbf{t}, \tag{24}$$

which is essentially OC or, alternatively, GC after imposing $R_i = 1$, $i = 1, 2$.

Experimentation has shown that the above assumptions hold approximately for a wide range of image data sets. For example, Fig. 1a shows the "Pentagon" image. We circularly shift the image in two different fashions and, for each shift, we compute the difference $\Delta\Phi$ between the original and the shifted image. For each case, Figs. 1b and 1c show the histogram with the distribution of $\Delta\Phi$. In both cases, $\Delta\Phi$ is well-described by a uniform distribution, and therefore, the value of $\sum_i \cos\Delta\Phi(\mathbf{x}_i)$ will be approximately equal to zero.

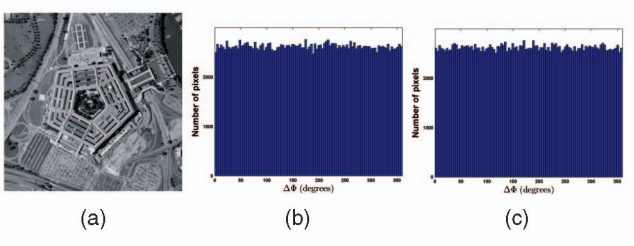


Fig. 1. (a) The 512×512 pentagon image. (b)-(c) The distribution of the difference in orientation $\Delta\Phi$ between the original image and two circularly shifted versions.

Under the above assumptions, OC will be a Dirac function even when the given images match only partially. To show this, we model dissimilar parts by relaxing (1) as follows:

$$I_1(\mathbf{x}+\mathbf{t}) = I_2(\mathbf{x}), \quad \mathbf{x} \in \Omega \subseteq \mathcal{R}^2. \tag{25}$$

That is, after shifting I_1 by $\mathbf{t}$, I_1 and I_2 match only in $\mathbf{x} \in \Omega$. From the above analysis, we may observe that

$$\mathrm{OC}(\mathbf{u})\big|_{\mathbf{u}\neq\mathbf{t}} = 0. \tag{26}$$

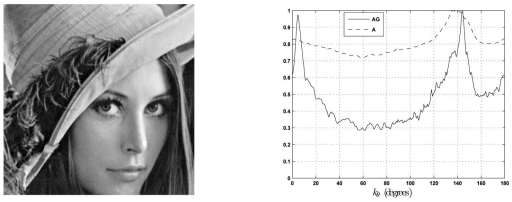
At $\mathbf{u} = \mathbf{t}$, we have

$$\mathrm{OC}(\mathbf{t}) = \int_\Omega \cos\Delta\Phi_{\mathbf{t}}(\mathbf{x})d\mathbf{x} + \int_{\mathcal{R}^2-\Omega} \cos\Delta\Phi_{\mathbf{t}}(\mathbf{x})d\mathbf{x} = \int_\Omega d\mathbf{x} \tag{27}$$

since $\Delta\Phi_{\mathbf{t}}(\mathbf{x}) = 0\ \forall\mathbf{x} \in \Omega$ and, in $\mathbf{x} \in \mathcal{R}^2 - \Omega$, the two images do not match. Overall, OC will be nonzero only for $\mathbf{u} = \mathbf{t}$, and its value at that point will be the contribution from the areas in the two images that match solely.

Our analysis does not impose any bound to the number of outliers. In fact, as their number increases, one would expect that accuracy is enhanced since $\Delta\Phi$ will better approximate the uniform distribution. In practice, we expect that deviations from our above assumptions will limit the dynamic range of the algorithm. Additional sources of performance degradation are errors in estimating the image gradients, possible image noise, and aliasing effects. To conclude, we mention that the above analysis

2. This result is not exact for real image pairs where translations induce nonoverlapping regions.



(a) (b)

Fig. 2. (a) "Lena" and (b) the 1D representations A (dashed line) and A_G (solid line).

agrees with experimental results which have shown that gradient-based correlation schemes are able to estimate translational displacements reliably even when the overlap between the given images is less than 20 percent. Note that phase correlation is able to register images when the overlap is of the order of 40 percent [12].

3.1.3 Normalized Gradient Correlation

In the above analysis, we assumed $R_i = 1$, $i = 1, 2$. To optimize the orientation difference function $\Delta\Phi$ of the image-salient structures solely, we introduce the normalized gradient correlation

$$\mathrm{NGC}(\mathbf{u}) \triangleq \frac{G_1(\mathbf{u}) \star G_2^*(-\mathbf{u})}{|G_1(\mathbf{u})| \star |G_2(-\mathbf{u})|} = \frac{\int_{\mathcal{R}^2} G_1(\mathbf{x}) G_2^*(\mathbf{x}+\mathbf{u}) d\mathbf{x}}{\int_{\mathcal{R}^2} |G_1(\mathbf{x}) G_2(\mathbf{x}+\mathbf{u})| d\mathbf{x}}. \quad (28)$$

Following the above analysis, (28) takes the form:

$$\mathrm{NGC}(\mathbf{u}) \triangleq \frac{\int_{\mathcal{R}^2} R_1(\mathbf{x}) R_2(\mathbf{x}+\mathbf{u}) \cos[\Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x}+\mathbf{u})] d\mathbf{x}}{\int_{\mathcal{R}^2} R_1(\mathbf{x}) R_2(\mathbf{x}+\mathbf{u}) d\mathbf{x}}. \quad (29)$$

NGC has two interesting properties:

1. $0 \leq |\mathrm{NGC}(\mathbf{u})| \leq 1$.
2. Invariance to affine changes in illumination.

The first property provides a measure to assess the correctness of the match. To show the second property, consider $I_2'(\mathbf{x}) = aI_2(\mathbf{x}) + b$ with $a \in \mathcal{R}^+$ and $b \in \mathcal{R}$. Then, by differentiation, $G_2' = aG_2$; therefore, the brightness change due to b is removed. Additionally, $R_2' = aR_2$ and $\Delta\Phi_2' = \Delta\Phi_2$; thus, the effect of the contrast change due to a will cancel out in (29). Note that if $a \in \mathcal{R}$, we can achieve full invariance by looking for the maximum of the absolute correlation surface.

3.1.4 Analysis in the Presence of Additive White Gaussian Noise

In general, signal differentiation exacerbates noise effects. Nevertheless, under the assumption of white noise, correlation is not affected by the degradation of the signal-to-noise ratio. Consider the case of a 1D signal s corrupted by additive white Gaussian noise n:

$$r(t) = s(t) + n(t), \quad (30)$$

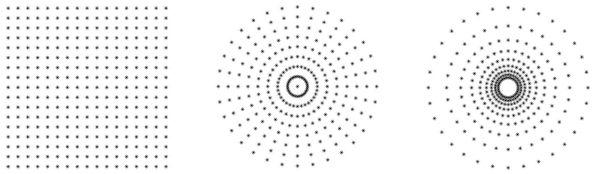
where the noise is assumed to be uncorrelated with the signal. The noise autocorrelation is given by $R_n(\tau) = \sigma_n^2 \delta(\tau)$, where σ_n^2 is the noise variance. If we perform differentiation, then

$$d(t) = s_d(t) + n_d(t), \quad (31)$$

where $d(t) \triangleq \frac{dr(t)}{dt}$, $s_d(t) \triangleq \frac{ds(t)}{dt}$, and $n_d(t) \triangleq \frac{dn(t)}{dt}$. Obviously, n_d is uncorrelated with s_d. Its autocorrelation is given by [13]

$$R_{n_d}(\tau) = -\frac{d^2 R_n(\tau)}{d\tau^2} = -\sigma_n^2 \frac{d^2\delta(\tau)}{d\tau^2}. \quad (32)$$

Assume that we are given two signals related by a shift ξ, that is, $s_2(t) = s_1(t+\xi)$, and corrupted by additive white Gaussian noise. The cross-correlation of d_1 and d_2 is



(a) (b) (c)

Fig. 3. (a) Cartesian, (b) polar, and (c) log-polar grids.

$$\begin{aligned} R_{d_1 d_2}(\tau) &= E\{d_1(t) d_2(t+\tau)\} \\ &= E\{[s_{d_1}(t) + n_d(t)][s_{d_1}(t+\xi+\tau) + n_d(t+\tau)]\} \\ &= E\{s_{d_1}(t) s_{d_1}(t+\xi+\tau)\} - \sigma_n^2 \frac{d^2\delta(\tau)}{d\tau^2} \end{aligned} \quad (33)$$

since noise and signals are assumed to be uncorrelated. The above result shows that uncorrelated white noise does not affect the estimation process. On the contrary, one can show that white noise deteriorates the performance of phase correlation [2].

3.2 Robust Estimation of Rotation and Scaling

In our scheme, to estimate the rotation and scaling, we replace I_i with G_i and then use M_{G_i} as a basis to perform correlation in the log-polar Fourier domain. This is possible since, from (8), we have $k_r = sk_r'$, and therefore,

$$\begin{aligned} M_{G_2}(\mathbf{k}) &= k_r M_2(\mathbf{k}) \\ &= (1/s) k_r' M_1(\mathbf{k}') \\ &= (1/s) M_{G_1}(\mathbf{k}') \\ &= (1/s) M_{G_1}(\Theta\mathbf{k}/s). \end{aligned} \quad (34)$$

The use of M_{G_i} is a key element of our approach. It equips the method with accuracy and robustness. We discuss the above arguments in detail as follows:

First, M_{G_i} captures the frequency response of the image-salient features solely. Areas of constant intensity level induce low-frequency components which hinder the estimation of the rotation and scaling. To illustrate this, consider the "Lena" image and the scenario where the motion is purely rotational. To estimate the rotation, we use the 1D representation $A(k_\theta) = \int M(k_r, k_\theta) dk_r$ and correlation over the angular parameter k_θ. The image contains a wide range of frequencies and, consequently, A is almost flat (Fig. 2b, dashed line). In this case, matching by correlation can be unstable. On the contrary, A_G (obtained by averaging M_G) efficiently captures possible directionality of the image-salient features: The two main orientations of the edges in the image give rise to two distinctive peaks in A_G (Fig. 2b, solid line). Using A_G to perform correlation, matching will be more accurate and robust.

Second, conversion from Cartesian to polar/log-polar induces a much larger interpolation error for low-frequency components. Fig. 3 clearly illustrates the problem. We may observe that, near the origin of the Cartesian grid, less data are available for interpolation. It is also evident that, for Cartesian-to-log-polar conversion, the situation becomes far more problematic since the log-polar representation is extremely dense near the origin. Thus, recently proposed DFT schemes [3], [4], [5] sample the FT on non-Cartesian grids, which geometrically are much closer to the polar/log-polar ones. Therefore, accuracy is enhanced, however, at the cost of additional computational complexity. On the contrary, our approach to alleviating the problem differs substantially: Eliminate the effect of low-frequency components by using the representation M_{G_i}. This comes naturally since the bottom line from the "Lena" example is that discarding low frequencies from the representation will also result in more robust and accurate registration. Our algorithm uses the Cartesian FFT and bilinear interpolation without oversampling, and hence, it is significantly faster than the schemes in [3], [4], [5].

TABLE 1
Filter Coefficients for Central Difference Estimators of the Theoretical Differentiator Up To Third Order

order	h_{-3}	h_{-2}	h_{-1}	h_0	h_1	h_2	h_3
1			-1	0	1		
2		1/12	-2/3	0	2/3	-1/12	
3	-1/60	3/20	-3/4	0	3/4	-3/20	1/60

Third, the periodic nature of the FFT induces boundary effects which result in spectral leakage in the frequency domain. Attempting to register images with no preprocessing typically returns a zero-motion estimate ($\theta_0 = 0,\ s = 0$). To reduce the boundary effect, one can use window functions [14]. Assuming that there is no prior knowledge about the motion to be estimated, the reasonable choice is to place the same window at the center of both images. In this case, windowing not only results in loss of information but also attenuates pixel values in regions shared by the two images in different ways. On the other hand, our scheme is based on image gradients, and therefore, discontinuities due to periodization will appear only if very strong edges exist close to the image boundaries. Thus, unlike previously proposed schemes, our method does not rely on image windowing.

Fourth, the estimation of the Fourier spectrum using FFT routines is largely affected by aliasing effects. Rotations and scalings in images induce additional sources of aliasing artifacts which are aggravated by the presence of high frequencies. For example, the commutativity of the FT and image rotation does not hold in the discrete case: The DFT of a rotated image differs from the rotated DFT of the same image resulting in rotationally dependent aliasing [15]. Using filters with bandpass spectral selection properties to compute G_i reduces the effect of high-frequency noise and aliasing in the estimation process. Elementary filter design suggests that we can obtain filters with such properties by approximating the ideal differentiator with central differences of various orders. Table 1 gives the filter coefficients for central difference estimators up to third order.

3.3 The Algorithm

Based on our analysis in the previous sections, we propose a robust gradient-based approach to scale-invariant image registration in Algorithm 1.

Algorithm 1. ROBUST FFT-BASED SCALE-INVARIANT IMAGE REGISTRATION ALGORITHM

Inputs: Two images $I_i, i = 1, 2$ of size $X_i \times Y_i$ related by a translation $\mathbf{t}$, rotation θ_0, and scaling s.

Step 1. Estimate G_i using central differences of second order and zero-pad the images to size $N \times N$, where $N = 2^n$ and n is the smallest integer such that $N \geq \max\{X_1, Y_1, X_2, Y_2\}$.

Step 2. Compute the $N \times N$ Cartesian FFT of G_i, and then, its magnitude M_{G_i}.

Step 3. Resample M_{G_i} on an $N/2 \times N/2$ log-polar grid. Use $\text{base} = \exp\{N/2 \log N/2\}$ as the logarithmic base for the log conversion along the radius axis and bilinear interpolation. Denote by L_i the corresponding log-polar Fourier magnitude representations.

Step 4. From L_i, extract the corresponding complex gradients G_{L_i} using central differences of second order. Using the FFT with no zero-padding and G_{L_i}, implement (28). Let (m, k) be the location of the maximum in the NGC surface. Estimate θ_0 (in degrees) and s as $\hat{\theta}_0 = 180m/(N/2)$ and $\hat{s} = \text{base}^k$.

Step 5. Scale down and rotate the zoomed image by $\hat{\theta}_0$ and $\hat{\theta}_0 + \pi$ using bilinear interpolation. Use NGC in the spatial domain to resolve the π ambiguity and estimate the residual translation $\hat{\mathbf{t}}$.

Fig. 4. An example of image pairs considered in our experiments and registration accuracy achieved by the proposed scheme. The reference image is scaled down, rotated, and translated according to the estimated motion parameters, and then superimposed on the target image.

From our experiments, we observed that the choice of the filter order is not critical; we suggest the use of the second-order central difference estimator since it gave slightly better results. Additionally, notice that, after **Step 2**, the distance of the available Cartesian data (to be interpolated) from the origin is in the range $[0, N/2]$. Thus, our choice of an $N/2 \times N/2$ log-polar grid implies that no oversampling takes place during the Cartesian-to-log-polar conversion. The choice of the base ensures that the extrapolated data will also span the range $[0, N/2]$. As for the type of interpolation used in the log-polar Fourier domain, we found that bilinear interpolation, compared to nearest neighbor, enhanced the performance of our scheme substantially, without adding significant computational overhead.

4 Performance Evaluation

To evaluate the performance of our scheme, we used a popular database with real images [16]. We examined two registration problems: **P.1**: Translations and scalings. **P.2**: Translations, rotations, and scalings. The database provides a set of 6 and 10 data sets for P.1 and P.2, respectively. Each data set consists of a collection of images capturing a particular scene. Depending on the data set, the image resolution varies from 348×512 to 650×850. We used approximately 1,000 image pairs, covering a wide range of rotations and scale factors up to 6. Fig. 4 shows a representative example of image pairs used in our experiments.

4.1 Comparison with State-of-the-Art

The target of this section is twofold. First, we present a comparison between OC, PC, and the proposed NGC. For this purpose, we also implemented the proposed scheme (**Algorithm 1**) using OC and PC in the log-polar Fourier domain (**Step 4**). For all variants, we preserved the original image resolution and used FFT length equal to 1,024. Second, we assess the performance of the state-of-the-art in FFT-based image registration. In particular, we implemented an improved version of method given in [3] as follows: We replaced the pseudopolar FFT with an accurate polar FFT recently proposed in [17]. Next, to approximate the log-polar FT, we resampled the polar FFT on the log-polar grid using bilinear interpolation. Finally, to estimate the rotation and scaling, we used PC. Since the method failed badly for most data sets without windowing, we preprocessed all images using a Tukey window prior to applying the algorithm.

To compare all schemes, we examined the maximum scale factors that each method recovered successfully for each data set. We obtained these factors by attempting to register the first image in each data set (reference image) with all the other images in the particular data set (target images). Table 2 gives an overview of the results. For each data set, we present the maximum scale factor $\widehat{s}$ and the corresponding rotation $\widehat{\theta_0}$ estimated by all schemes along with the ground truth s and θ_0 as given in [16], [18].

The proposed scheme (using NGC) gave excellent results. For most data sets ("Asterix," "Belledonee," "Bip," "Laptop1," "Bark," "Boat," "East Park," "East South," "Laptop2," "Resid," and "UBC"), the algorithm correctly estimated the maximum scale change considered. For "Van Gogh," the maximum scale factor detected was only 3.4. The reason is explained in detail in the next

TABLE 2
The Maximum Scale Factors and the Corresponding Rotations Recovered by the Proposed Scheme Using NGC, OC, PC, and the State-of-the-Art, Respectively

	Proposed scheme - NGC		Proposed scheme - OC		Proposed scheme - PC		Polar FFT - PC	
P.1	$(s, \theta_0 = 0)$	$(\hat{s}, \hat{\theta}_0)$	$(s, \theta_0 = 0)$	$(\hat{s}, \hat{\theta}_0)$	$(s, \theta_0 = 0)$	$(\hat{s}, \hat{\theta}_0)$	$(s, \theta_0 = 0)$	$(\hat{s}, \hat{\theta}_0)$
"Asterix"	$(6.0, 0.0°)$	$(5.78, 0.0°)$	$(2.71, 0.0°)$	$(2.65, 0.0°)$	$(4.21, 0.0°)$	$(4.11, 0.0°)$	$(1.98, 0.0°)$	$(1.91, 0.0°)$
"Belledonee"	$(5.34, 0.0°)$	$(5.57, 0.35°)$	$(-)$	$(-)$	$(3.22, 0.0°)$	$(3.22, 0.35°)$	$(1.77, 0.0°)$	$(1.78, 0.44°)$
"Bip"	$(3.75, 0.0°)$	$(3.73, 0.0°)$	$(1.50, 0.0°)$	$(1.51, 0.0°)$	$(2.69, 0.0°)$	$(2.68, 0.0°)$	$(1.5, 0.0°)$	$(1.51, 0.09°)$
"Crolle"	$(4.01, 0.0°)$	$(3.97, 0.7°)$	$(2.15, 0.0°)$	$(2.15, 0.7°)$	$(3.23, 0.0°)$	$(3.26, 0.0°)$	$(1.7, 0.0°)$	$(1.78, 0.09°)$
"Laptop1"	$(6.25, 0.0°)$	$(6.22, 0.35°)$	$(3.55, 0.0°)$	$(3.55, 0.35°)$	$(4.87, 0.0°)$	$(4.87, 0.0°)$	$(2.96, 0.0°)$	$(2.97, 0.26°)$
"Van Gogh"	$(3.4, 0.0°)$	$(3.38, 0.0°)$	$(2.82, 0.0°)$	$(2.82, 0.0°)$	$(2.47, 0.0°)$	$(2.49, 0.0°)$	$(1.7, 0.0°)$	$(1.71, 0.0°)$
P.2	(s, θ_0)	$(\hat{s}, \hat{\theta}_0)$	(s, θ_0)	$(\hat{s}, \hat{\theta}_0)$	(s, θ_0)	$(\hat{s}, \hat{\theta}_0)$	(s, θ_0)	$(\hat{s}, \hat{\theta}_0)$
"Bark"	$(4.09, 153.4°)$	$(4.01, 150.1°)$	$(2.49, 119.9°)$	$(2.48, 119.9°)$	$(3.01, 22.77°)$	$(3.03, 22.85°)$	$(1.22, 31.5°)$	$(1.23, 31.2°)$
"Boat"	$(4.36, 46.0°)$	$(4.26, 45.7°)$	$(2.35, 8.0°)$	$(2.38, 7.7°)$	$(3.34, 12.78°)$	$(3.38, 13.01°)$	$(2.35, 8.02°)$	$(2.36, 7.82°)$
"East Park"	$(5.77, 0.6°)$	$(5.78, 0.4°)$	$(2.38, 9.8°)$	$(2.38, 9.8°)$	$(3.97, 65.36°)$	$(3.96, 66.09°)$	$(2.38, 9.75°)$	$(2.36, 9.76°)$
"East South"	$(5.09, 60.0°)$	$(5.18, 59.4°)$	$(1.61, 61.9°)$	$(1.61, 61.9°)$	$(3.80, 1.68°)$	$(3.77, 2.46°)$	$(2.07, 35.74°)$	$(2.06, 35.86°)$
"Ensimag"	$(4.92, 40.7°)$	$(4.76, 41.5°)$	$(1.84, 0.8°)$	$(1.82, 0.7°)$	$(2.38, 36.14°)$	$(2.43, 36.21°)$	$(1.82, 0.81°)$	$(1.83, 0.88°)$
"Inria"	$(4.03, 0.8°)$	$(3.91, 0.7°)$	$(2.87, 31.6°)$	$(2.89, 31.3°)$	$(2.87, 31.61°)$	$(2.89, 31.29°)$	$(2.87, 31.61°)$	$(2.87, 31.38°)$
"Inria Model"	$(4.79, 50.82°)$	$(4.82, 51.0°)$	$(1.58, 20.3°)$	$(1.57, 20.0°)$	$(2.78, 29.71°)$	$(2.78, 29.88°)$	$(1.58, 20.25°)$	$(1.57, 20.21°)$
"Laptop2"	$(1.51, 45.4°)$	$(1.51, 45.0°)$	$(1.51, 45.4°)$	$(1.51, 45.3°)$	$(1.51, 45.4°)$	$(1.51, 45.0°)$	$(1.51, 45.4°)$	$(1.51, 45.09°)$
"Resid"	$(5.89, 33.2°)$	$(5.85, 31.6°)$	$(2.37, 76.0°)$	$(2.38, 76.3°)$	$(3.31, 34.33°)$	$(3.34, 33.75°)$	$(1.87, 6.52°)$	$(1.88, 6.42°)$
"UBC"	$(2.89, 9.6°)$	$(2.89, 9.5°)$	$(2.09, 71.8°)$	$(2.1, 71.7°)$	$(2.89, 9.6°)$	$(2.89, 9.49°)$	$(2.89, 9.6°)$	$(2.87, 9.58°)$

section. Moreover, the method estimated translations and rotations to nearly one pixel and degree accuracy, respectively.

Replacing NGC with OC gave considerably worse results. In particular, the maximum scale factors recovered were approximately reduced by half compared to those detected using NGC. OC is robust only if the orientation difference function $\Delta\Phi$ follows a uniform distribution for displacements other than the correct. This is not the case for images resampled on a log-polar grid. More specifically, near the origin, resampling induces artifacts since very few data are available for interpolation in the original Cartesian representation. The structure (and therefore, the orientation) of the artifacts is more related to the Cartesian-to-log-polar conversion rather than the image to be interpolated. The result is a bias in the detection process. We conclude that to achieve robust performance, both magnitude and orientation information must be considered.

Additionally, a simple visual inspection of Table 2 reveals the performance improvement obtained using NGC instead of PC. Using our NGC, we were able to detect successfully maximum scale changes in the range $[4, 6]$. Replacing NGC with PC, the maximum scale factors recovered were limited in the range $[2.5, 4]$.

Finally, the gain in performance compared to the state-of-the-art is evident. Interestingly, we can observe that the implementation of our scheme using PC in the log-polar Fourier domain gave significantly better results. We conclude that the choice of sophisticated methods to approximate the log-polar DFT is not a critical element of robustness in FFT-based scale-invariant image registration.

TABLE 3
The Detection Ratio for Each Scale Range and Data Set

	P.1 Scalings and translations			
Dataset	Small	Moderate	Large	Very Large
"Asterix"	46/46	44/44	26/26	20/20
"Belledonee"	0/0	3/3	2/2	1/1
"Bip"	15/15	15/15	4/4	2/2
"Crolle"	9/9	12/12	4/4	2/3
"Laptop1"	68/75	56/81	26/33	20/21
"Van Gogh"	45/45	46/46	22/22	0/23
Overall	183/190	176/201	84/91	45/70
	P.2 Scalings, rotations and translations			
Dataset	Small	Moderate	Large	Very Large
"Bark"	4/4	8/8	2/2	1/1
"Boat"	16/16	17/17	8/8	4/4
"East Park"	18/18	20/20	9/9	8/8
"East South"	16/16	15/15	8/8	6/6
"Ensimag"	19/19	20/20	8/8	7/8
"Inria"	17/17	20/20	9/9	6/9
"Inria Model"	18/18	20/20	9/9	6/8
"Laptop2"	77/77	1/1	0/0	0/0
"Resid"	17/17	21/21	9/9	8/8
"UBC"	41/41	33/33	4/4	0/0
Overall	243/243	175/175	66/66	46/52

4.2 Performance Evaluation for All-Possible Image Pairs

In this section, we present an exhaustive evaluation of our scheme by attempting to register all possible image pairs for problems P.1 and P.2. Since the method recovered rotations and translations within very good accuracy, we examined the ability of the method to detect scale changes solely. In particular, we grouped together all possible scale factors into four groups as follows: Small: $s \leq 1.5$, Moderate: $1.5 < s \leq 2.5$, Large: $2.5 < s \leq 3.5$, and Very Large: $3.5 < s \leq 6$. For each group and data set, we computed the detection ratio $\stackrel{\triangle}{=}$ (number of correct detections)/(number of image pairs).

Table 3 gives an overview of the obtained results. The robustness of the proposed scheme is evident. With the exception of "Laptop1" and "Van Gogh," only a few misdetections were observed for all data sets and scale changes considered.

For "Laptop1" and "Van Gogh," the method appeared to be unstable. Figs. 5a and 5b show an image pair taken from the "Laptop1" data set, for which the method failed. We may observe that the camera does not zoom in/out, it is the laptop that is moving toward the camera. Therefore, there is no single global rotation, scale, and translation to be recovered but two motions: the scale change induced by the laptop movement and the zero motion of the background (i.e., the background remains unchanged). This yields two peaks in the resulting correlation function, but the scheme in its current form just picks the maximum which corresponds to the zero motion of the background. Figs. 5c and 5d show an image pair taken from the "Van Gogh" data set. The image on the right has a white frame placed on a black background. This yields very strong edge responses which bias the detection process. This extreme case will rarely be encountered in most real-world applications. In both cases, to reduce the

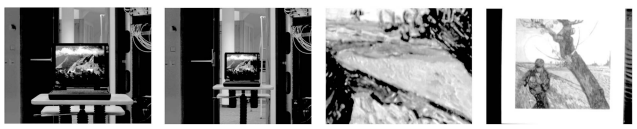


Fig. 5. Two examples of image pairs for which the method failed. (a)-(b) From the "Laptop1" data set. (c)-(d) From the "Van Gogh" data set.

TABLE 4
Detection Ratio for "Laptop1" and "Van Gogh" Using Tukey Windowing

	P.1 Scalings and translations			
Dataset	Small	Moderate	Large	Very Large
"Laptop1"	75/75	81/81	33/33	21/21
"Van Gogh"	45/45	46/46	22/22	22/23

TABLE 5
The Overall Detection Ratio for PSNR Equal to 20 and 14 dB

	P.1 Scalings and translations			
Overall	Small	Moderate	Large	Very Large
PSNR = 20 dB	190/190	198.8/201	89/91	48.05/70
PSNR = 14 dB	185.5/190	190.3/201	75.95/91	23.1/70
	P.2 Scalings, rotations and translations			
Overall	Small	Moderate	Large	Very Large
PSNR = 20 dB	233/243	175/175	64.75/66	32.1/52
PSNR = 14 dB	220.6/243	164.15/171	48.35/66	11.05/52

TABLE 6
The Overall Detection Ratio for Various Image Resolutions

	P.1 Scalings and translations			
Overall	Small	Moderate	Large	Very Large
$N_{\max} = 512$	190/190	201/201	91/91	65/70
$N_{\max} = 128$	190/190	196/201	78/91	22/70
	P.2 Scalings, rotations and translations			
Overall	Small	Moderate	Large	Very Large
$N_{\max} = 512$	243/243	175/175	66/66	35/52
$N_{\max} = 128$	233/243	171/171	41/66	12/52

$N_{\max}$ *and FFT Length were set to 512 and 128, respectively.*

background effect, we used a Tukey window [14]. Table 4 shows the obtained results. Only one misdetection occurred.[3]

4.3 Performance Evaluation in the Presence of Gaussian Noise

In this section, we assess the performance of our scheme in the presence of additive zero-mean white Gaussian noise. We considered two large noise levels: $\mathrm{PSNR} = 20$ and $\mathrm{PSNR} = 14$ dB. For each PSNR and image pair, to assure the validity of the classification results, we repeated the experiment using a total of 20 noisy image pairs. Table 5 outlines the overall results for each PSNR value. For each scale range, we present the mean value of the detection ratio.

For $\mathrm{PSNR} = 20$ dB, the method appeared to be very robust. Compared to the noise-free case, we observed a degradation in performance, only for $s > 3.5$. In the same scale range and for $\mathrm{PSNR} = 14$ dB, the method failed. Nevertheless, the method recovered scale factors up to 3.5 consistently for most data sets.

4.4 Performance Evaluation for Various Image Resolutions

In this experiment, we assess the performance of the method for various image resolutions. To simulate low resolution, we low-pass filtered the original image and then decreased its dimensions using nearest-neighbor interpolation [15]. We considered two cases such that the maximum dimension $N_{\max}$ of the low-resolution image was 512 and 128, while the FFT length was also set to 512 and 128 respectively. Table 6 gives the overall results for each case.

For the typical case $N_{\max} = 512$, no misdetections occurred for $s \leq 3.5$. For the same resolution and $s > 3.5$, a slight degradation in performance was observed. For very low-resolution images ($N_{\max} = 128$), the method appeared to be robust for scale changes up to 2.5. In the range $2.5 < s \leq 3.5$, performance was still satisfactory.

3. For the remaining performance evaluation results, we applied the same Tukey window to "Laptop1" and "Van Gogh" only.

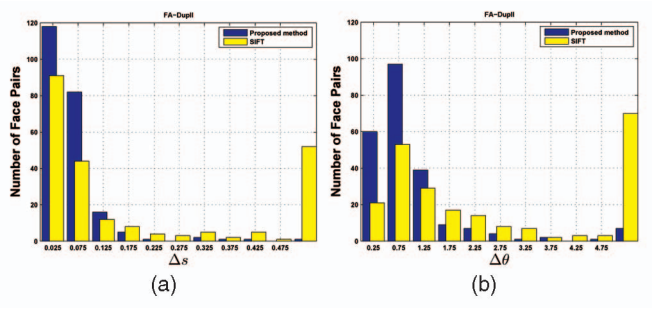


Fig. 6. (a) The distribution of the scale difference Δs. The width of each histogram bin is 0.05 units, while its center is indicated by the numbering of the x-axis. (b) The distribution of the rotation difference $\Delta\theta$. The width of each histogram bin is 0.5 units, while its center is indicated by the numbering of the x-axis. Blue color: Proposed scheme. Yellow color: SIFT.

Fig. 7. Registration accuracy achieved by the proposed scheme for the application of face registration. The zoomed-in image is scaled down, rotated, and translated according to the estimated motion parameters, and then superimposed on the zoomed-out image.

5 Application to Frontal View Face Registration

Accurate registration of face data is a typical prerequisite for most face recognition and verification algorithms. Even small alignment errors may result in significant degradation performance; thus, researchers usually report results after manual alignment, which is often performed using eyes annotation.

Assuming frontal view faces, it is not unreasonable to model global motion with a similarity transform which can be estimated using the proposed scheme. Thus, we chose to perform a representative set of registration attempts using the FA and DupII data sets of the FERET database [19], [20]. Matching FA with DupII results in very challenging registration cases since the assumption for a similarity transform is often violated by other sources of misalignment such as $3D$ rotations, nonuniform illumination changes, occlusions, and facial expression variations.

The data sets share a total of 75 subjects. FA contains one face image per subject while DupII at least two faces per subject and a total of more than 200 face images. For each subject, we attempted to register each face from FA with the corresponding faces from DupII. Since no ground truth data are available, we present a qualitative performance analysis of our scheme by calculating, for each pair of faces, the absolute differences between the estimated scale and rotation parameters and the ones obtained by performing manual eye-based registration. In a similar spirit, we assess the performance of SIFT-based registration [21] using RANSAC. For each method, Figs. 6a and 6b show the histograms (obtained using all pairs) with the distribution of the scale Δs and rotation $\Delta\theta$ absolute differences, respectively.

Our scheme outperforms SIFT-based matching in two aspects. First, it appears to be more accurate. This is illustrated by the total number of face pairs for which Δs and $\Delta\theta$ are relatively small (for example, $\Delta s \leq 0.1$ and $\Delta\theta \leq 1°$). Second, it is significantly more robust. This is illustrated by the total number of face pairs for which Δs and $\Delta\theta$ are relatively large (for example, $\Delta s \geq 0.15$ and $\Delta\theta \geq 3°$). Fig. 7 shows an example illustrating the registration accuracy achieved by the proposed scheme.

In general, assuming that the given images share a sufficient number of image features, spatial domain registration schemes [21], [22] are able to handle more challenging registration problems than the proposed method does such as affine distortions and severe partial matching scenarios. However, this is not the case for face registration problems, where face data may substantially vary not only due to different capturing conditions but also due to significant appearance changes of the individual subjects. Our scheme measures global similarity, and therefore appears to be more suitable for handling cases where robust matching of local features is not feasible.

Other advantages of our algorithm over SIFT-based registration methods are:

- Computational efficiency. Our approach naturally draws advantages from very recent advances in parallel implementations of the FFT [23], [24], [25]. We expect that, using such optimized architectures, near real-time performance can be achieved. Even with a conventional 3 GHz Pentium IV computer, to register a pair of 512×512 images, a Matlab implementation of our algorithm requires about 1 second, while Lowe's precompiled code typically requires about 4-12 seconds. For larger image sizes, the gain in performance is consistently more than $10\times$.
- Constant time of processing. Our approach makes use of all image information and has a fixed time of processing which depends on the FFT resolution. On the contrary, SIFT's execution time depends on the image content and, more specifically, on the number of detected keypoints.
- Ease of implementation. Contrary to spatial domain methods, our scheme requires fine-tuning of very few parameters with the most important being the FFT length.

6 Conclusions

We presented a gradient-based approach which operates in the frequency domain for the estimation of scalings, rotations, and translations in images. We attribute the robustness of the proposed scheme to both the image representation used and the type of correlation employed. We provided reasoning and experimentation which verify the validity of our arguments. There is no other FFT-based technique which is able to recover large motions in real images. A key feature of Fourier-based registration methods is the speed offered by the use of FFT routines. The proposed scheme estimates large motions accurately and robustly without the need of excessive zero-padding and oversampling, thus without sacrificing part of the computational efficiency which typifies the frequency-domain formulation.

The supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputer society.org/10.1109/TPAMI.2010.107, of our paper includes:

1. Additional experiments on a very challenging data set.
2. Comparison with the method in [5].
3. Comparison with the method in [7].
4. Additional registration examples from the Inria database.
5. Detailed results for each data set for Sections 4.3 and 4.4, respectively.

Acknowledgments

The authors would like to thank Professor Wolberg and Dr. Zokai for graciously running their algorithm [7] on the image pairs of Section 1 of the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputer society.org/10.1109/TPAMI.2010.107, as well as for providing the image pairs of Section 3 of the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2010.107. They are grateful to Professor Michael Elad for providing the Matlab implementation of the polar FFT. They would also like to thank the associate editor and the anonymous reviewers whose suggestions and comments greatly improved the quality of this work. Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office. This work was supported by the Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre established by the UK Ministry of Defence.

References

[1] B.S. Reddy and B.N. Chatterji, "An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration," *IEEE Trans. Image Processing*, vol. 5, no. 8, pp. 1266-1271, Aug. 1996.
[2] C.D. Kuglin and D.C. Hines, "The Phase Correlation Image Alignment Method," *Proc. IEEE Conf. Cybernetics and Soc.*, pp. 163-165, 1975.
[3] Y. Keller, A. Averbuch, and M. Israeli, "Pseudopolar-Based Estimation of Large Translations, Rotations and Scalings in Images," *IEEE Trans. Image Processing*, vol. 14, no. 1, pp. 12-22, Jan. 2005.
[4] H. Liu, B. Guo, and Z. Feng, "Pseudo-Log-Polar Fourier Transform for Image Registration," *IEEE Signal Processing Letters*, vol. 13, no. 1, pp. 17-21, Jan. 2006.
[5] W. Pan, K. Qin, and Y. Chen, "An Adaptable-Multilayer Fractional Fourier Transform Approach for Image Registration," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 400-413, Mar. 2009.
[6] A. Averbuch, D.L. Donoho, R.R. Coifman, and M. Israeli, "Fast Slant Stack: A Notion of Radon Transform for Data in Cartesian Grid Which Is Rapidly Computable, Algebrically Exact, Geometrically Faithful and Invertible," *SIAM J. Scientific Computing*, to appear.
[7] S. Zokai and G. Wolberg, "Image Registration Using Log-Polar Mappings for Recovery of Large-Scale Similarity and Projective Transformations," *IEEE Trans. Image Processing*, vol. 14, no. 10, pp. 1422-1434, Oct. 2005.
[8] A.J. Fitch, A. Kadyrov, W.J. Christmas, and J. Kittler, "Orientation Correlation," *Proc. British Machine Vision Conf.*, pp. 133-142, 2002.
[9] V. Argyriou and T. Vlachos, "Estimation of Sub-Pixel Motion Using Gradient Cross-Correlation," *Electronic Letters*, vol. 39, no. 13, pp. 980-982, 2003.
[10] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, second ed. Pearson Education, 2002.
[11] R.N. Bracewell, K.-Y. Chang, A.K. Jha, and Y.-H. Wang, "Affine Theorem for Two-Dimensional Fourier Transform," *Electronics Letters*, vol. 29, no. 3, pp. 304-309, 1993.
[12] Y. Keller and A. Averbuch, "A Projection-Based Extension to Phase Correlation Image Alignment," *Signal Processing*, vol. 87, pp. 124-133, 2007.
[13] A.L. Garcia, *Probability and Random Processes for Electrical Engineering*, second ed. Pearson Education, 2004.
[14] F.J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51-83, Jan. 1978.
[15] H.S. Stone, B. Tao, and M. MacGuire, "Analysis of Image Registration Noise Due to Rotationally Dependent Aliasing," *J. Visual Comm. and Image Representation*, vol. R.14, pp. 114-135, 2003.
[16] Image Database, http://lear.inrialpes.fr/people/mikolajczyk/, 2010.
[17] A. Averbuch, R.R. Coifman, D.L. Donoho, M. Elad, and M. Israeli, "Fast and Accurate Polar Fourier Transform," *Applied and Computational Harmonic Analysis*, vol. 21, pp. 145-167, 2006.
[18] Image Database, http://www.robots.ox.ac.uk/vgg/research/affine/, 2010.
[19] P.J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The Feret Database and Evaluation Procedure for Face Recognition Algorithms," *Image and Vision Computing J.*, vol. 16, no. 5, pp. 295-306, 1998.
[20] P.J. Phillips, H. Moon, P.J. Rauss, and S. Rizvi, "The Feret Evaluation Methodology for Face Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.
[21] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
[22] C.V. Stewart, C.-L. Tsai, and B. Roysam, "The Dual-Bootstrap Iterative Closest Point Algorithm with Application to Retinal Image Registration," *IEEE Trans. Medical Imaging*, vol. 22, no. 11, pp. 1379-1394, Nov. 2003.
[23] Intel Math Kernel Library, http://www.intel.com/software/products/mkl, 2010.
[24] M. Frigo and S.G. Johnson, "FFTW on the Cell Processor," http://www.fftw.org/cell/index.html, 2007.
[25] N.K. Govindaraju, B. Lloyd, Y. Dotsenko, B. Smith, and J. Manferdelli, "High Performance Discrete Fourier Transforms on Graphics Processors," *Proc. ACM/IEEE Conf. Supercomputing*, 2008.