Robust FFT-Based Scale-Invariant Image Registration

Georgios Tzimiropoulos and Tania Stathaki
Communications and Signal Processing Group, Imperial College London, Exhibition Road, London, SW7 2AZ

Abstract

We present a fast and robust gradient-based scale-invariant image registration technique which operates in the frequency domain. The algorithm combines the natural advantages of good feature selection offered by gradient-based methods with the robustness and speed provided by FFT-based correlation schemes. Experimentation with real images taken from a popular database showed that, unlike any other Fourier-based techniques, the method was able to estimate translations, arbitrary rotations and scale factors up to 6.

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 techniques for global motion estimation address the problem of compensating for the camera ego-motion and finally aligning the images. Practical applications are numerous: from global scene representation and image mosaicing to object detection / tracking and video compression.

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 [16]. Third, the method is computationally efficient. This comes from the *shift property* of the Fourier Transform (FT) and the use of Fast Fourier Transform (FFT) routines for the rapid computation of correlations.

The work in [13] 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. Re-sampling the Fourier magnitudes on the log-polar grid reduces the problem of estimating the rotation and scaling to one of estimating a 2*D* 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) [10] 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 [9,11,12] introduce new sampling schemes and algorithms which reduce the inaccuracies induced by re-sampling the magnitude of the FT on the log-polar grid. To recover the rotation and scaling, the method in [9] relies on the pseudopolar FFT [3] 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 neighbour 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 [11], the authors propose to approximate the log-polar DFT by interpolating the pseudo-log-polar FFT. The method is non-iterative but the gain in registration accuracy is not significant. The main idea in [12] is to obtain more accurate log-polar DFT approximations by efficiently over-sampling the lower part of the Fourier spectrum using the Fractional FT. The presented experimental results do not explicitly show the applicability of the algorithm in real images related by large scale factors while over-sampling inevitably increases the execution time.

In this work, we provide reasoning, intuition and experimentation which show that accuracy in FFT-based motion estimation 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 over-sampling is employed to enhance accuracy. To perform robust correlation, we replace phase correlation with gradient-based correlation schemes [5,1]. 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 domain. Exhaustive experimentation with popular image datasets [4] demonstrated that the merits of a gradient-based approach combined with the speed which typifies a frequency domain approach provide a fast and robust framework for scale-invariant image registration.

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:

$$\widehat{I}_i(\mathbf{k}), \mathbf{k} = [k_x, k_y]^T \in \mathcal{R}^2$$

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 $k_\theta = \arctan(k_y/k_x)$.

Translation Estimation using Correlation

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

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

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:

$$\widehat{\mathbf{t}} = \arg_{\mathbf{u}} \max\{\mathrm{C}(\mathbf{u})\} \quad (2)$$

where:[1]

$$C(\mathbf{u}) \triangleq I_1(\mathbf{u}) \star I_2(-\mathbf{u}) = \int_{\mathcal{R}^2} I_1(\mathbf{x}) I_2(\mathbf{x}+\mathbf{u}) d\mathbf{x} \tag{3}$$

From the *convolution theorem* of the FT [6], C can be alternatively obtained by

$$C(\mathbf{u}) = F^{-1}\left\{\widehat{I}_1(\mathbf{k})\widehat{I}_2^*(\mathbf{k})\right\} \tag{4}$$

where F^{-1} is the inverse FT and * denotes the complex conjugate operator. The *shift property* of the FT [6] states that if the relation between I_1 and I_2 is given by (1), then, in the frequency domain, it holds:

$$\widehat{I}_2(\mathbf{k}) = \widehat{I}_1(\mathbf{k}) e^{j\mathbf{k}^T\mathbf{t}} \tag{5}$$

and therefore (4) becomes:

$$C(\mathbf{u}) = F^{-1}\left\{{M_1}^2(\mathbf{k}) e^{-j\mathbf{k}^T\mathbf{t}}\right\} \tag{6}$$

The above analysis summarises the main principles of frequency domain correlation-based translation estimation. For finite discrete images of size $N \times N$, correlation is efficiently implemented through (4), 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 $O(N^2 \log N)$.

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:

$$I_2(\mathbf{x}) = I_1(D\mathbf{x} + \mathbf{t}) \tag{7}$$

where

$$D = s\Theta \text{ and } \Theta = \begin{bmatrix} \cos\theta_0 & \sin\theta_0 \\ -\sin\theta_0 & \cos\theta_0 \end{bmatrix}$$

In the Fourier domain, it holds [14]

$$\widehat{I}_2(\mathbf{k}) = (1/|\Delta|)\widehat{I}_1(\mathbf{k}') e^{j\mathbf{k}'^T\mathbf{t}} \tag{8}$$

where

$$\mathbf{k}' = D^{-T}\mathbf{k} \tag{9}$$

and Δ is the determinant of D. Taking the magnitude in both parts of (8) and substituting $D^{-T} = \Theta/s$, $\Delta = s^2$ gives

$$\begin{aligned} M_2(\mathbf{k}) &= (1/|\Delta|)M_1(\mathbf{k}') \\ &= (1/s^2)M_1(\Theta\mathbf{k}/s) \end{aligned} \tag{10}$$

Using the log-polar representation yields (ignoring the term $1/s^2$) [13]

$$(\mathbf{k}_l) = M_1(\mathbf{k}_l - [\log s, \theta_0]^T) \tag{11}$$

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 valid solution the one that yields the highest peak [13].

Robust FFT-based Scale-invariant Image Registration

Robust Translation Estimation

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 [1]

$$\begin{aligned} GC(\mathbf{u}) &\triangleq G_1(\mathbf{u}) \star G_2^*(-\mathbf{u}) \\ &= \int_{\mathcal{R}^2} G_1(\mathbf{x}) G_2^*(\mathbf{x}+\mathbf{u}) d\mathbf{x} \end{aligned} \tag{12}$$

where

$$G_i(\mathbf{x}) = G_{i,x}(\mathbf{x}) + jG_{i,y}(\mathbf{x}) \tag{13}$$

and $G_{i,x} = \nabla_x I_i$ and $G_{i,y} = \nabla_y I_i$ are the gradients along the horizontal and vertical

[1] To be more precise, we assume hereafter that the images are of finite energy such that correlation integrals such as the one in (3) converge.

direction respectively. Orientation correlation (OC) considers orientation information solely [5]

$$\mathrm{OC}(\mathbf{u}) \triangleq O_1(\mathbf{u}) \star O_2^*(-\mathbf{u}) \quad (14)$$

where

$$O_i(\mathbf{x}) = \begin{cases} G_i(\mathbf{x})/|G_i(\mathbf{x})|, \text{ if } |G_i(\mathbf{x})| > \epsilon \\ 0, \text{ otherwise} \end{cases} \quad (15)$$

and ϵ is the value of a threshold. Thresholding $|G_i(\mathbf{x})|$ aims at removing the contribution of pixels where gradient magnitude takes negligible values.

In the following analysis, the focus is primarily on GC.

Spatial Domain Analysis

From the definition of GC and using (13), we can easily derive

$$\begin{aligned} \mathrm{GC}(\mathbf{u}) &= G_{1,x}(\mathbf{u}) \star G_{2,x}(-\mathbf{u}) \\ &+ G_{1,y}(\mathbf{u}) \star G_{2,y}(-\mathbf{u}) \\ &+ j\{-G_{1,x}(\mathbf{u}) \star G_{2,y}(-\mathbf{u}) \\ &+ G_{1,y}(\mathbf{u}) \star G_{2,x}(-\mathbf{u})\} \end{aligned} \quad (16)$$

The imaginary part in the above equation is equal to zero, therefore

$$\mathrm{GC}(\mathbf{u}) = G_{1,x}(\mathbf{u}) \star G_{2,x}(-\mathbf{u}) + G_{1,y}(\mathbf{u}) \star G_{2,y}(-\mathbf{u}) \quad (17)$$

Using the polar representation of complex numbers, we define

$$R_i = \sqrt{G_{i,x}^2 + G_{i,y}^2} \text{ and}$$
$$\Phi_i = \arctan G_{i,y}/G_{i,x}$$

Based on this representation, (17) takes the form:

$$\begin{aligned} \mathrm{GC}(\mathbf{u}) = \int_{\mathcal{R}^2} &\{R_1(\mathbf{x}) R_2(\mathbf{x}+\mathbf{u}) \\ &\cos[\Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x}+\mathbf{u})] d\mathbf{x}\} \end{aligned} \quad (18)$$

Each term in (18) has its own special importance. 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. Orientation information is embedded in the cosine kernel. 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 roughly show the latter point, let us first assume that at **u**≠**t** the orientation difference function $\Delta\Phi(\mathbf{u},\mathbf{x}) = \Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x}+\mathbf{u})$ is uniformly distributed over [-π,π). This assumption appears to be reasonable, since for displacements other than the correct, the images do not match and therefore we expect that differences in gradient orientation can take any value in the range [0,2π) with equal probability.

Let us further impose $R_i = 1$, $i = 1; 2$, that is, we essentially compute a modified orientation correlation function (mOC) where, in contrary to (14), the orientation of all pixels is taken into consideration

$$\mathrm{mOC}(\mathbf{u}) = \int_{\mathcal{R}^2} \cos[\Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x}+\mathbf{u})] d\mathbf{x} \quad (19)$$

To model dissimilar parts, we modify the perfect translational model of (1) as follows:

$$I_1(\mathbf{x}+\mathbf{t}) = I_2(\mathbf{x}), \mathbf{x} \in \Omega \subseteq \mathcal{R}^2 \quad (20)$$

That is after shifting I_1 by **t**, I_1 and I_2 match only in $\mathbf{x} \in \Omega$.

At **u**≠**t**, we may observe that :

$$\mathrm{mOC}(\mathbf{u})\,|_{\mathbf{u}\neq\mathbf{t}} = 0 \quad (21)$$

since $\forall\, \mathbf{u} \neq \mathbf{t}$ we have assumed that $\Delta\Phi(\mathbf{u},\mathbf{x})$ is uniformly distributed. At **u**=**t**, we have:

$$\begin{aligned} \mathrm{mOC}(\mathbf{t}) &= \int_{\Omega} \cos \Delta\Phi(\mathbf{t},\mathbf{x}) d\mathbf{x} \\ &+ \int_{\mathcal{R}^2-\Omega} \cos \Delta\Phi(\mathbf{t},\mathbf{x}) d\mathbf{x} \\ &= \int_{\Omega} d\mathbf{x} \end{aligned} \quad (22)$$

since $\Delta\Phi(\mathbf{t},\mathbf{x}) = 0 \;\forall\, \mathbf{x} \in \Omega$ and $\Delta\Phi(\mathbf{t},\mathbf{x})$ is uniformly distributed if $\mathbf{x} \in \mathcal{R}^2 - \Omega$. Overall, mOC will be non-zero only for **u**=**t**, and its value at that point will be the contribution from the areas in the two images that match solely.

Essentially, using image gradients to perform correlation, the errors induced by outliers are mapped to a uniform distribution for which correlation is well-known to feature robust performance. 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 induced by the FFT. To conclude, we mention that the above analysis 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%. Note that phase correlation is able to register images when the overlap is of the order of 40% [8].

Normalized Gradient Correlation

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

$$\mathrm{NGC}(\mathbf{u}) \triangleq \frac{G_1(\mathbf{u}) \star G_2^*(-\mathbf{u})}{|G_1(\mathbf{u})| \star |G_2(-\mathbf{u})|} \qquad (23)$$

Following the above analysis, (23) takes the form:

$$\mathrm{NGC}(\mathbf{u}) \triangleq \frac{\int_{\mathcal{R}^2} R_1(\mathbf{x}) R_2(\mathbf{x}+\mathbf{u}) \cos\Delta\Phi(\mathbf{u},\mathbf{x}) d\mathbf{x}}{\int_{\mathcal{R}^2} R_1(\mathbf{x}) R_2(\mathbf{x}+\mathbf{u}) d\mathbf{x}} \qquad (24)$$

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:

$$I_2'(\mathbf{x}) = aI_2(\mathbf{x}) + b$$

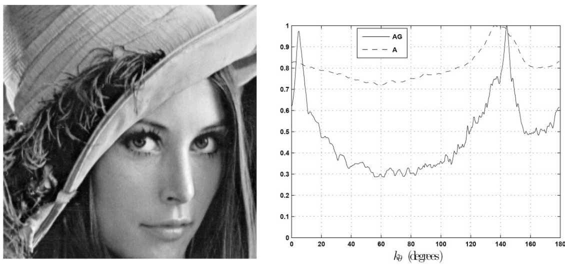
with $a \in \mathcal{R}^+$ and $b \in \mathcal{R}$

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 (24). Note that if $a \in \mathcal{R}$, we can achieve full invariance by looking for the maximum of the absolute correlation surface.

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 (9) we have $k_r = sk_r'$ and therefore:

$$\begin{aligned} M_{G_2}(\mathbf{k}) &= k_r M_2(\mathbf{k}) \\ &= (1/s) k_r' M_1(\mathbf{k}') \\ &= (1/s) M_{G_1}(\mathbf{k}') \\ &= (1/s) M_{G_1}(\Theta\mathbf{k}/s) \end{aligned} \qquad (25)$$



(a) (b)

Figure 1: (a) 'Lena' and (b) the 1*D* representations *A* (dashed line) and A_G (solid line)

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 (Figure 1 (b), dashed line). In this case, matching by correlation can be unstable. In contrary, AG (obtained by averaging MG) 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 AG (Figure 1 (b), solid line). Using AG to perform correlation, matching will be more accurate and robust.

Second, conversion from Cartesian to polar / log-polar induces much larger interpolation error for low frequency components. This is because 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 [9,11,12] 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. In 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 standard Cartesian FFT and bilinear interpolation without over-sampling, thus it is significantly faster than the schemes in [9,11,12].

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 pre-processing, typically returns a zero-motion estimate (θ_0 = 0, s = 0). To reduce the boundary effect, one can use window functions [7]. Assuming that there is no prior knowledge about the motion to be estimated, the reasonable choice is to place the same window at the centre 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. For large motions, the result can be a dramatic decrease in performance.

On the other hand, the proposed scheme is based on gray-level edge maps and, therefore, discontinuities due to periodisation will appear only if very strong edges exist close to the image boundaries. In practice, the method does not apply any windowing to the input images.

Fourth, using FFT routines to approximate the Fourier spectrum of images results in significant aliasing effects. Rotations and scalings in images induce additional sources of aliasing artefacts 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 depending aliasing [15]. Using filters with band-pass 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 [6].

The Algorithm

Based on our analysis in the previous subsections, we propose a robust gradient-based approach to estimate translations, rotations and scalings in images as follows.

> ***Algorithm 1:*** *Robust FFT-Based Scale-invariant Image Registration Algorithm*
>
> **Inputs**: Two images I_i, $i = 1;2$ related by a translation $\mathbf{t}$, rotation θ_0 and scaling s.
>
> **Step 1**: Estimate G_i and compute M_{G_i} using the standard Cartesian FFT.
>
> **Step 2**: Resample M_{G_i} on the log-polar grid using bilinear interpolation.
>
> **Step 3**: Estimate θ_0 and s using NGCorr in the log-polar domain.
>
> **Step 4**: Scale down and rotate the zoomed image. Resolve the π ambiguity and recover $\mathbf{t}$ using NGCorr in the spatial domain.

Results

To evaluate the performance of our scheme, we used a popular database with real images [4]. 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 datasets for P.1 and P.2 respectively. Each dataset consists of a collection of images capturing a particular scene. Depending on the dataset, the image resolution varies from 348×512 to 650×850. We used approximately 1000 image pairs, covering a wide range of rotations and scale factors up to 6.

The target of our experiments 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**). 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 [9] as follows. We replaced the pseudopolar FFT with an accurate polar FFT recently proposed in [2]. Next, to approximate the log-polar FT, we re-sampled the polar FFT on the log-polar grid using bilinear interpolation. Finally, to estimate the rotation and scaling, we used PC. Since we observed that the method failed badly for most datasets without windowing, we pre-processed all images using a Tukey window prior applying the algorithm.

To compare all schemes, we examined the maximum scale factors that each method recovered successfully for each dataset. We obtained these factors by attempting to register the first image in each dataset (reference image) with all the other images in the particular dataset (target images). Table 1 gives an overview of the results. For each dataset, 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 [4].

The proposed scheme (using NGC) gave excellent results. For most datasets (‘Asterix’, ‘Belledonee’, ‘Bip’, ‘Laptop1’, ‘Bark’, ‘Boat’, ‘East Park’, ‘East South’, ‘Laptop2’, ‘Resid’, ‘UBC’), the algorithm correctly estimated the maximum scale change considered. 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 others than the correct. This is not the case for images re-sampled on a log-polar grid. More specifically, near the origin, re-sampling induces artefacts since very few data are available for interpolation in the original Cartesian representation. The structure (and therefore the orientation) of the artefacts 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. Overall, we conclude that to achieve robust performance, both magnitude and orientation information must be considered.

Table 1: Experiment 1. 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"	$(2.89, 9.6°)$	$(2.89, 9.5°)$	$(2.09, 71.8°)$	$(2.1, 71.7°)$	$(2.89, 9.6°)$	$(2.89, 9.49°)$	$(2.89, 9.6°)$	$(2.87, 9.58°)$

(a) (b) (c) (d)

Figure 2: Registration accuracy achieved by the proposed scheme. (a) (s,θ_0) = (3.97,0.0°), $(\hat{s},\hat{\theta}_0)$ = (4.01,0.7°). (b) (s,θ_0) = (2.69,0:0°), $(\hat{s},\hat{\theta}_0)$ = (2.68,0.0°). (c) (s,θ_0) = (4.36,46.0°), $(\hat{s},\hat{\theta}_0)$ = (4.26,45.7°). (d) (s,θ_0) = (5.89,33.2°), $(\hat{s},\hat{\theta}_0)$ = (5.85,31.6°)

Additionally, a simple visual inspection of Table 1 reveals the performance improvement obtained using NGC instead of PC. In particular, 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.

Figure 2 illustrates the accuracy of registration 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.

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 over-sampling, thus without sacrificing part of the computational efficiency which typifies the frequency domain formulation. To register a pair of 512×512 images, a nearly optimised Matlab implementation of the algorithm on a 3 GHz Pentium IV computer requires about 1 second. It is expected that, using parallel machine-specific optimised implementations of the FFT, near real-time performance can be achieved.

Finally, a further advantage of the proposed approach is that the method is complementary to other state-of-the-art FFT-based image registration methods. Ongoing research is focused on performance evaluation of such composite schemes.

References

[1] V. Argyriou and T. Vlachos. *Estimation of sub-pixel motion using gradient cross-correlation*. Electronic Letters, 39(13):980–982, 2003.

[2] A. Averbuch, R.R. Coifman, D.L. Donoho, M. Elad, and M.Israeli. *Fast and accurate polar Fourier transform*. Appl. Comput. Harmon Anal., 21:145–167, 2006.

[3] 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, algebraically exact, geometrically faithful and invertible*. SIAM Sci. Comput, to appear.

[4] Image database. http://lear.inrialpes.fr/people/mikolajczyk/.

[5] A.J. Fitch, A. Kadyrov,W.J. Christmas and J. Kittler. *Orientation correlation*. In British Machine Vision Conference, pages 133–142, 2002.

[6] R.C. Gonzalez and R.E.Woods. Digital Image Processing. Singapore: Pearson Education, edition, 2002.

[7] F.J. Harris. *On the use of windows for harmonic analysis with the discrete Fourier transform*. Proceedings of the IEEE, 66(1):51–83, 1978.

[8] Y. Keller and A. Averbuch. *A projection based extension to phase correlation image alignment.* Signal Process., 87:124–133, 2007.

[9] Y. Keller, A. Averbuch, and M. Israeli. *Pseudopolar-based estimation of large translations, rotations and scalings in images*. IEEE Trans. Image Processing, 14(1):12–22, 2005.

[10] C.D. Kuglin and D.C. Hines. The phase correlation image alignment method. In Proc. IEEE Conf. Cybernetics and Society, pages 163–165, 1975.

[11] H. Liu, B. Guo, and Z. Feng. *Pseudo log-polar Fourier transform for image registration*. IEEE Signal Processing Letters, 13(1):17–21, 2006.

[12] W. Pan, K. Qin, and Y. Chen. *An adaptable multilayer fractional Fourier transform approach for image registration*. IEEE Trans. Pattern Anal. Machine Intell., 31(3):400–413, 2009.

[13] B.S. Reddy and B.N. Chatterji. *An FFT-based technique for translation, rotation and scale-invariant image registration*. IEEE Trans. Image Processing, 5(8):1266–1271, 1996.

[14] R.N.Bracewell, K.-Y.Chang, A.K.Jha and Y.-H.Wang. *Affine theorem for two-dimensional Fourier transform.* Electronics Letters, 29(3):304, 1993.

[15] H.S. Stone, B. Tao, and M. MacGuire. *Analysis of image registration noise due to rotationally dependent aliasing*. J. Vis. Commun. Image, R.14:114–135, 2003.

[16] S. Zokai and G.Wolberg. *Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations.* IEEE Trans. Image Process., 14(10):1422–1434, 2005.

Acknowledgements

The work reported in this paper was funded by the Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre established by the UK Ministry of Defence.