A Frequency Domain Approach to Roto-translation Estimation using Gradient Cross-Correlation

Georgios Tzimiropoulos*, Vasileios Argyriou and Tania Stathaki
Department of Electrical and Electronic Engineering
Imperial College London
London, SW7 2AZ , UK

Abstract

A novel frequency domain approach to roto-translation estimation is presented. The baseline gradient cross-correlation method is extended to handle rotations. A key feature of the proposed scheme is the ability to achieve good performance in the presence of both large translations and rotations as well as noise, a scenario for which other Fourier-based methods typically fail. Robustness and accuracy in conjunction with computational efficiency, offered by the frequency domain formulation, make the algorithm useful in a number of image processing tasks such as image registration.

1 Introduction

The estimation of relative motions between two images finds applications in a multitude of image processing and computer vision tasks such as image registration, video compression and object recognition. In this work, the focus is on rigid motion estimation for the class of similarity transforms and, in particular, for translational displacements and rotations.

Recently there has been an increased interest in correlation-based motion estimation techniques which operate in the Fourier domain. A frequency domain approach to motion estimation possesses two rather appealing properties: robustness to noise and computational efficiency. Robustness to noise is achieved since correlation makes use of all available image information. Computational efficiency comes from the *translational invariance property* of the magnitude of the Fourier transform (FT) and the use of fast algorithms. Given two images that are related by a roto-translation, the relative rotation affects only the magnitudes of the FTs of the two images. The magnitudes are represented in the polar Fourier domain and rotation is recovered using $1D$ correlation over the angular parameter. Then, after compensating for rotation, translation may be estimated using standard $2D$ Cartesian cross-correlation. Correlations are efficiently computed as a multiplication in the frequency domain through the use of FFT-based techniques.

*Mr G. Tzimiropoulos' funding for this work is provided by the Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre established by the UK Ministry of Defence.

In this work, to account for rotations, an extension to the baseline gradient-based cross-correlation method [1] is introduced. The main idea behind the proposed scheme is to extract the complex gray level edge maps of the given images using robust differential operators and then perform correlation. Essentially, the algorithm combines the natural advantages of a good feature selection offered by gradient-based methods with the robustness and speed provided by FFT-based correlation schemes.

An important challenge for Fourier-based motion estimation algorithms is to perform robustly in the presence of *both* large translations and rotations. In this case, large non-overlapping regions between the two images are introduced and the frequency domain motion estimation formulation holds approximately. Experimental results demonstrate that the robustness and good feature selectivity provided by the proposed approach makes the algorithm capable of estimating successfully the motion parameters for four representative images in the presence of both noise and relatively large non-overlapping parts.

2 Related work

In this section, a brief review of current state-of-the-art Fourier-based methods for global motion estimation is presented. For a representative example of frequency domain methods which are able to handle local motions, the reader is referred to [6]. In all methods, the relative translation is estimated using phase correlation [7], therefore the focus is on the rotation estimation part.

Perhaps the most popular method for planar roto-translation estimation in the frequency domain is the extended phase correlation method described in [9]. The algorithm steps on the translation invariance property of the magnitude of the FT to sequentially estimate the relative rotation and translation between two given images. Phase correlation is used instead of standard correlation since it combines both robustness to noise and good localization accuracy. A pseudopolar-based extension to [9] is proposed in [4]. To recover rotation, the method relies on the pseudopolar FT which rapidly computes a discrete FT on a nearly polar grid. The algorithm does not involve any Cartesian to polar conversion in the Fourier magnitude domain and, thus, it does not suffer from inaccuracies induced by the interpolation process; nevertheless the pseudopolar FT is not a true polar Fourier representation and the method estimates the rotation in an iterative fashion. A frequency domain approach to register images in the presence of aliasing is described in [10]. Rotation is recovered using correlation between two $1D$ functions of the angular parameter. The functions are obtained by averaging the aliasing-free part of the polar Fourier magnitude representation of the two images with respect to the radial parameter.

In addition to correlation schemes, difference functions [8],[5] provide a robust and accurate framework for Fourier-based rotation estimation. Given two images that are related by a rotation, it can be shown [8] that the difference between the magnitudes of the FTs of the first image and the flipped version of the second has a distinctive pair of orthogonal zero-crossing lines. The orientation of those lines in the Fourier plane is directly related to the unknown rotation and is recovered using a simple Hough transform. The method in [5] defines the difference function in the angular direction and proposes a pseudopolar FFT-based implementation for its accurate computation.

3 Frequency domain roto-translation estimation using gradient cross-correlation

Let $I(x,y),\ [x,y]^T \in \mathscr{R}^2$ be an image function. We denote $\widehat{I}(k_x,k_y),\ [k_x,k_y]^T \in \mathscr{R}^2$ the Fourier transform (FT) of I and $M(k_x,k_y)$ the magnitude of $\widehat{I}$, that is $M = |\widehat{I}|$. Additionally, in the polar Fourier domain, we have $\widehat{I}(k_r,k_\theta)$ and $M(k_r,k_\theta)$, where $k_r = \sqrt{k_x^2+k_y^2}$ and $k_\theta = \arctan(k_y/k_x)$.

3.1 Translation estimation

Given two images, I_1 and I_2, that are related by an unknown translation $[t_x,t_y]^T \in \mathscr{R}^2$, i.e.

$$I_1(x+t_x,y+t_y) = I_2(x,y) \tag{1}$$

the translational displacement can be recovered from the $2D$ cross-correlation function $C(u,v),\ [u,v]^T \in \mathscr{R}^2$ as $\arg_{(u,v)}\max\{C(u,v)\}$. From the *convolution theorem* of the FT [3], C is given by:

$$C(u,v) = F^{-1}\{\widehat{I_1}(k_x,k_y)\widehat{I_2}^*(k_x,k_y)\} \tag{2}$$

where F^{-1} is the inverse FT and $*$ denotes the complex conjugate operator. The *shift property* of the FT [3] states that if the relation between I_1 and I_2 is given by (1), then, in the frequency domain, it holds

$$\widehat{I_1}(k_x,k_y)e^{j(k_xt_x+k_yt_y)} = \widehat{I_2}(k_x,k_y) \tag{3}$$

and therefore (2) becomes

$$C(u,v) = F^{-1}\{{M_1}^2(k_x,k_y)e^{-j(k_xt_x+k_yt_y)}\} \tag{4}$$

The above analysis summarizes the main principles of frequency domain correlation-based translation estimation. For finite discrete images, the FT is efficiently implemented using FFT routines and the algorithm's complexity is $O(N^2\log N)$, where N is the length of the given images.

From (4), it can be seen that the phase difference term $e^{-j(k_xt_x+k_yt_y)}$, which contains the translational information, is weighted by the magnitude M_1. Then, the inverse FT is taken to yield the standard $2D$ spatial correlation function C. In practise, where (1) holds approximately and $M_1 \neq M_2$, the translational displacement is estimated through (2), and in this case, the phase difference function is weighted by the term M_1M_2. Due to the low pass nature of images, the weighting operation results in a peak of large magnitude in C, however, at the same time, good peak localization is inevitably sacrificed.

To tackle the problem, the phase correlation method performs whitening of the Fourier magnitude spectra and essentially considers the phase difference function solely. It can be easily seen that, in this case, the resulting correlation function will be a $2D$ dirac located at the unknown translation. In the presence of noise and dissimilar parts, the value of the peak is significantly reduced and the method may become unstable [9].

Gradient cross-correlation is an approach which lies somewhere between the standard and phase correlation methods. For a given image I, a gray level edge map G, which retains both *magnitude* and *orientation* information, is computed as follows

$$G = G_x + jG_y \tag{5}$$

where $G_x = \nabla_x I$ and $G_y = \nabla_y I$ are the gradients along the horizontal and vertical direction respectively. This step provides the location, magnitude and orientation of the image high-activity structures which can be used as salient features to estimate rigid motions between images. At the same time, areas of constant intensity level which do not provide any reference points for motion estimation and are very sensitive to possible uneven lighting conditions, are discarded from the representation.

Gradient cross-correlation (GC) is defined as

$$GC(u,v) = F^{-1}\{\widehat{G}_1(k_x,k_y)\widehat{G}_2^*(k_x,k_y)\} \tag{6}$$

It can be easily shown that differentiation in the image spatial domain is equivalent to high-pass filtering in the Fourier domain. Taking the FT in both parts of (5) yields

$$\widehat{G}(k_x,k_y) = jk_x\widehat{I}(k_x,k_y) - k_y\widehat{I}(k_x,k_y) \tag{7}$$

The magnitude M_G is given by

$$M_G(k_x,k_y) = \sqrt{(k_x^2+k_y^2)}M(k_x,k_y) = k_r M(k_r,k_\theta) \tag{8}$$

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

In practice, differentiation in (5) is performed using robust and efficient differentiation operators. The result is a band-pass filtered version of the original image. Therefore, in addition to low spatial frequencies, high frequency components, which are largely affected by noise and aliasing, are filtered out as well.

3.2 Roto-translation estimation

Assume now that we are given two images, I_1 and I_2, that are related by a rotation θ_0 and translation $[t_x,t_y]$, that is

$$I_1(x\cos\theta_0 + y\sin\theta_0 + t_x, -x\sin\theta_0 + y\cos\theta_0 + t_y) = I_2(x,y) \tag{9}$$

In the polar Fourier domain, we have

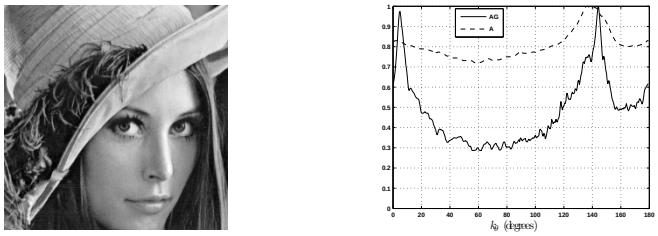
$$\widehat{I}_1(k_r,k_\theta+\theta_0)e^{j(k_x t_x + k_y t_y)} = \widehat{I}_2(k_r,k_\theta) \tag{10}$$

Taking the magnitude in both parts yields

$$M_1(k_r,k_\theta+\theta_0) = M_2(k_r,k_\theta) \tag{11}$$

which is also known as the *translation invariance property* of the magnitude of the FT [3]. It can be seen that in the polar Fourier magnitude domain, rotations reduce to translations and, therefore, one can estimate θ_0 using correlation over k_r. After compensating for rotation, the remaining unknown translation can be recovered using Cartesian correlation as described before. Note that if $\widetilde{\theta_0}$ is the estimated rotation, then it can be shown 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 Cartesian correlations and, finally, choose as valid solution the one that yields the highest peak [9].

In rotation estimation using gradient cross-correlation, M_1 and M_2 are replaced by M_{G_1} and M_{G_2}. This is a key element to the robustness of the proposed scheme, since M_G does not contain low frequencies, which do not provide any reference points for the estimation of θ_0, is little affected by noise and aliasing, and essentially emphasizes the frequency bands which reflect the orientation of the image features. To exemplify the latter point, consider the $1D$ representation $A(k_\theta) = \int k_r M(k_r, k_\theta)$ and the corresponding A_G obtained by averaging M_G computed from the "Lena" image, as sketched in Fig. (1) (b). The image contains a wide range of frequencies and, consequently, A is almost flat. In this case, matching by correlation may become unstable. In contrary, A_G contains two distinctive peaks which can be used as salient features to perform robust correlation.



(a) (b)

Figure 1: (a) "Lena" and (b) the $1D$ representations A_G (straight line) and A (dashed line).

A second advantage of the proposed scheme over other Fourier-based methods comes when implementation in a digital environment is to be considered. In particular, due to the periodic nature of the FFT, in practice, windowing should be applied to the input images to reduce the effect of boundaries whose registration corresponds to zero motion. The window is typically placed at the center of the image and the result is inevitably some loss of information. For the case of translations (even large), the problem can be tackled effectively, since, for example, phase correlation is able to recover translations when the overlap between the two images is of the order 30-40% [5]. While the same applies to the case of rotations only, the situation becomes rather problematic when (potentially large) roto-translations are to be considered. In this case, the center of rotation is unknown and windowing does not only result in loss of critical information but also attenuates pixel values non-uniformly (with respect to the rotation center). The result is a dramatic decrease in performance, as demonstrated in the following section.

On the other hand, it can be observed that the proposed scheme is based on gradient edge maps and, therefore, discontinuities due to periodization appear only if very strong edges exist close to the image boundaries. In practice, by selecting efficient differentiation operators the boundary effect is greatly reduced and the method does not apply any windowing to the input images.

From the above discussion it is evident that a key role to the robustness of the proposed scheme plays the choice of appropriate operators to perform the differentiation in (5). In our case, the operator should be robust to noise and able to reduce the border effect significantly. Not surprisingly, experimentation suggested that the operator based on the first derivative of a Gaussian [3] perfectly fits the particular application. For the $1D$ case, the filter's frequency response for standard deviation $\sigma = 0.5, 1, 2$ is shown in Fig. (2). As

σ increases, more emphasis is put on the lower part of the Fourier spectrum and, therefore, robustness to noise is enhanced; nevertheless, it is expected that smaller motions can be handled.

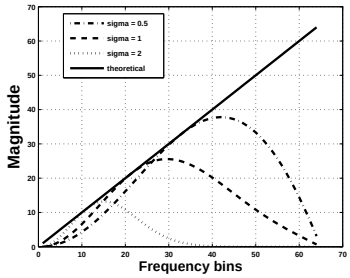
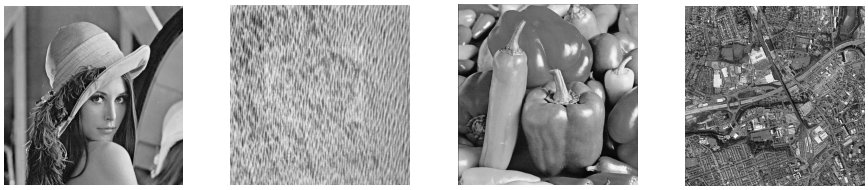


Figure 2: The response of the ideal differentiator and the first derivative of a Gaussian with standard deviation σ =0.5, 1 and 2.

4 Results

To evaluate the performance of the proposed scheme two different experiments were carried out using four representative images of size 512x512, shown in Fig. (3).



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

Figure 3: The four images used. (a) "Lena", (b) "Water", (c) "Peppers", (d) "Aerial".

The first experiment is related to rotation estimation. Given each of the four images, a set of pairs of test images of resolution 256x256 was created. Each pair I_{t_x,t_y} and I_θ is related by a roto-translation $[t_x,t_y,\theta]$. With reference to the center of the original image, I_{t_x,t_y} is obtained by cropping a 256x256 window centered at $[t_x,t_y] = 15n[\pm e_x \pm e_y]$, $n = 0,1,\ldots,7$, where e_x and e_y are the unit vectors of the $x-$ and $y-$ axis. I_θ was obtained by rotating the original image by $0^\circ \leq \theta \leq 170^\circ$ with step $\Delta\theta = 10^\circ$ and then cropping the 256x256 central window. Overall, for a fixed n, four translated images were created. Each of the translated images was registered with each rotated image. Depending on the rotation angle, the overlap between the input images is in the range 90-100% for $n = 0$, about 60% for $n = 4$, 50% for $n = 5$, 40-45% for $n = 6$ and finally about 35% for $n = 7$. For each $[t_x,t_y,\theta]$, the corresponding images were corrupted by white Gaussian noise. The signal-to-noise ratio (SNR) was set to 0 dB and each experiment was repeated 25 times. Figure 4 shows an example of a pair of misaligned noisy images considered in our experiments.

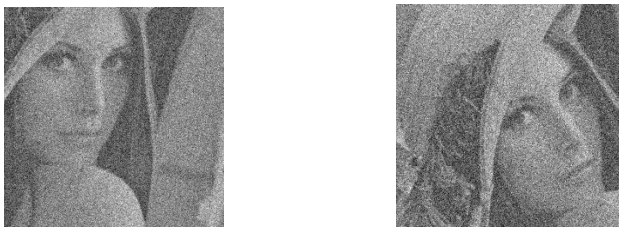


Figure 4: An example of two unregistered noisy images (SNR=0 dB, $[t_x, t_y, \theta] = [75, 75, 30]$) considered in our experiments.

Rotation estimation was attempted using three methods: gradient cross-correlation (GC), phase correlation (PC) [9] and the angular difference function (ADF) [5]. For the proposed scheme, edge maps were extracted using the first derivative of a Gaussian with standard deviation $\sigma = 1$. For PC and ADF, to reduce border effects, the input images were weighted by a $2D$ Gaussian $W_G = e^{-(x^2+y^2)/2\sigma_c^2}$ with $\sigma_c = 0.4$ which gave the best results for both methods (the image domain is supposed to be [-1/2, 1/2]x[-1/2, 1/2]). For each method, the mean value and standard deviation (std) of the registration *absolute* error in degrees with respect to n for "Lena" and "Water" are shown in Fig. (5). The registration accuracy achieved by the proposed scheme for all images is given in Table 1.

The robustness of the proposed scheme is evident. For "Lena", rotations were accurately estimated for overlapping regions up to 50% ($n = 5$) and with good precision (mean = 0.02434, std = 2.5922) when the overlap is of the order of 40-45% ($n = 6$). To this point, it should be emphasized that the size of overlap is rather indicative than representative. The main assumption for the proposed algorithm to work is that overlapping regions share a number of distinctive image features. Additionally, it cannot be expected that the algorithm features robust performance when objects with strong edges are introduced in the dissimilar parts of the two images. For example, notice that for "Peppers" (Table 1), for $n = 6$, while the mean value of the orientation absolute error is relatively small, the corresponding standard deviation is significant, which indicates that, in this case, some rotation angles were estimated inconsistently. In any case, the dynamic range of the algorithm should be defined according to the application. An application, where the proposed framework appears to fit with promising results, is the registration of various types of textured images as shown in Fig. (5) (c) and (d). In general, textured images may lack the existence of distinctive features, however, in many cases, some sort of orientation may be present which turns out to be independent of the relative translation (consider for example a brick wall) and is efficiently captured by the proposed scheme. Promising results were also obtained for "Aerial".

The second experiment aims at evaluating the performance of GC and PC for translation estimation when rotation is recovered up to a certain accuracy. In particular, with reference to the previous simulation, for a fixed rotation angle $\theta = 45^\circ$, it is assumed that rotation was estimated as $\widetilde{\theta} = 45^\circ + \theta_\varepsilon$, where $\theta_\varepsilon = \pm 1, \pm 2, \pm 3$. After compensating for $\widetilde{\theta}$, the residual translation is estimated using GC and PC. Examples which illustrate the performance of both methods, for "Lena" and $|\theta_\varepsilon| = 2$ and "Water" and $|\theta_\varepsilon| = 3$, are given in Fig. (6). For "Water", it can be seen that GC achieves relatively good performance up to $n = 2$ only. This is attributed to the large misalignment and the lack of distinctive features

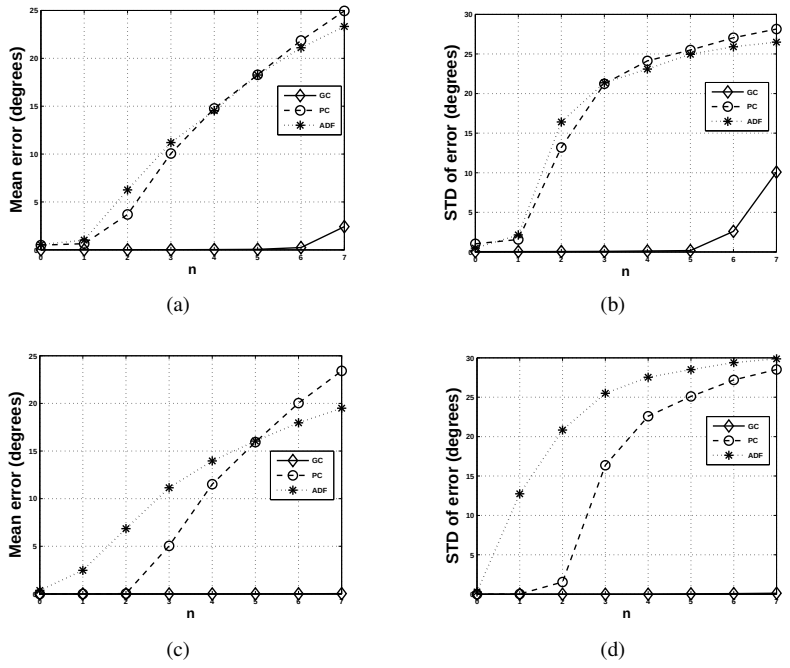


Figure 5: (a) Mean value and (b) std of the orientation error in degrees for "Lena". (c) Mean value and (d) std of the orientation error in degrees for "Water". Diamond: GC, Circle: PC, Asterisk: ADF.

n	"Lena"	"Water"	"Peppers"	"Aerial"
0	0.0025 (0.0353)	0 (0)	0.0106 (0.0719)	0 (0)
1	0.0040 (0.0447)	0 (0)	0.0187 (0.095)	0 (0)
2	0.0055 (0.052)	0 (0)	0.0297 (0.1186)	0 (0)
3	0.0117 (0.0755)	0 (0)	0.0489 (0.1501)	0 (0)
4	0.0296 (0.1191)	0 (0)	0.0769 (0.1921)	0 (0)
5	0.0549 (0.1666)	0.0033 (0.0407)	0.2444 (2.7826)	0.0006 (0.0167)
6	0.02434 (2.5922)	0.0111 (0.0737)	0.9393 (6.2438)	0.0039 (0.0441)
7	2.4194 (10.078)	0.0328 (0.1264)	2.028 (8.9526)	0.0202 (0.1052)

Table 1: Mean value (std) of the orientation error in degrees using GC.

in the particular image. Note however that the error induced by the rotation estimation step is very small and misalignments of this order are unlikely to occur. In general, for all images and $|\theta_\varepsilon|$ considered, GC outperforms PC. For $|\theta_\varepsilon| = 1$, it was found that the algorithm achieves nearly one-pixel accuracy (worst case: mean = 0.6656, std =0.7059, $n = 7$, "Aerial"). For $|\theta_\varepsilon| = 2$, the accuracy achieved by GC is given in Table 2. It can be seen that, in most cases, at least a rough estimate of the translation is provided. For $|\theta_\varepsilon| = 3$, the mean value and std were both approximately increased by one pixel compared to the case $|\theta_\varepsilon| = 2$, with the exception of "Water", where the algorithm becomes unstable.

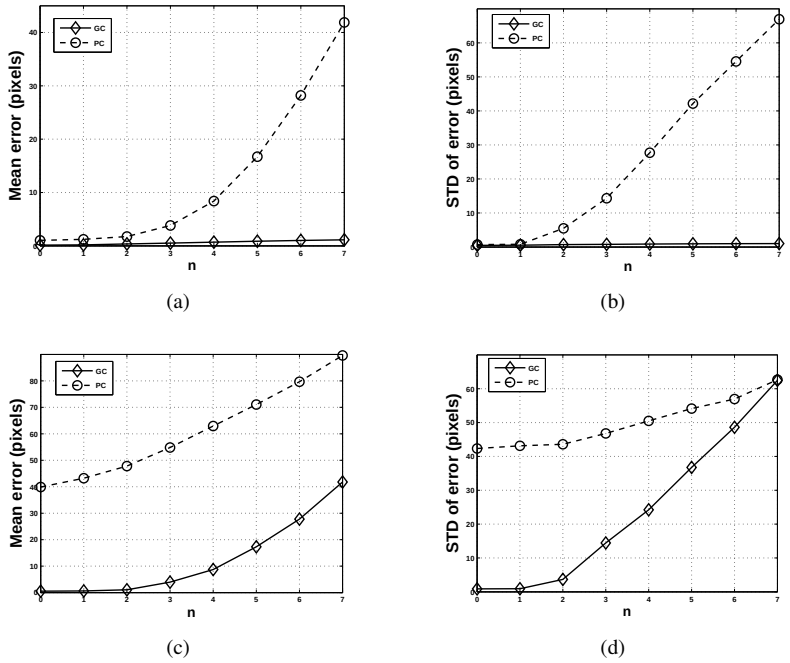


Figure 6: (a) Mean value and (b) std of the translational error in pixels for "Lena" and $|\theta_\varepsilon| = 2$. (c) Mean value and (d) std of the translational error in pixels for "Water" and $|\theta_\varepsilon| = 3$. Diamond: GC, Circle: PC

n	"Lena"	"Water"	"Peppers"	"Aerial"
0	0.1395 (0.4801)	0 (0)	1.2226 (0.5247)	1.8818 (0.4460)
1	0.1786 (0.5367)	0.02 (0.1990)	1.1804 (0.5759)	1.7845 (0.6447)
2	0.3548 (0.7077)	0.1271 (0.4836)	1.1384 (0.6277)	1.7347 (0.8109)
3	0.5123 (0.7938)	0.3905 (0.7035)	1.0957 (0.6871)	1.7951 (0.8140)
4	0.6996 (0.8687)	0.5925 (0.7556)	1.0701 (0.7231)	1.8336 (0.7912)
5	0.8702 (0.9378)	0.7397 (0.7754)	1.0572 (0.7477)	1.8808 (0.7955)
6	1.0116 (0.9781)	0.9821 (3.7081)	1.0641 (0.7572)	1.9661 (0.8089)
7	1.1280 (1.0032)	3.7377 (19.272)	1.1657 (3.018)	2.0726 (0.8204)

Table 2: Mean value (std) of the translational error in pixels for $|\theta_\varepsilon| = 2$ using GC.

5 Conclusions

Summarizing, a frequency domain algorithm for motion estimation based on gradient cross-correlation has been proposed. Experimentation suggested that, unlike other Fourier-based approaches, the scheme is able to estimation rotations when relatively large roto-translations and noise are to be considered. This is attributed to the good feature selectivity offered by the use of image gradients and the ability of the proposed scheme to perform correlations without the need of image windowing. Additionally, it was found that, de-

pending on the magnitude of the orientation error, gradient cross-correlation is able to provide at least a rough estimate of the residual translation. This could motivate an iterative approach where rotations and translations are estimated and, then, the overlapping regions between the two images are extracted and used to update the initial estimates.

The computational complexity of the proposed scheme is determined by the FFT operations and the additional cost of converting the magnitude of the FT from Cartesian to polar coordinates. If the polar FFT algorithm, recently proposed in [2], is used, then, the overall complexity of the algorithm will be $O(N^2 \log N)$.

References

[1] V. Argyriou and T. Vlachos. Performance study of gradient cross-correlation for subpixel motion estimation in the frequency domain. *IEE Vis. Image Signal Process.*, 152(1):107–114, 2005.

[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] R.C. Gonzalez and R.E. Woods. *Digital Image Processing, 2nd edition*. Pearson Education, Siganpore, 2002.

[4] 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.

[5] Y. Keller, Y. Shkolnisky, and A. Averbuch. The angular difference function and its application to image registration. *IEEE Trans. Pattern Anal. Machine Intell.*, 27(6):969–976, 2005.

[6] S.A. Kruger and A.D. Calway. A multiresolution frequency domain method for estimating affine motion parameters. In *Proc. IEEE International Conf. on Image Processing*, pages 113–116, 1996.

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

[8] L. Lucchese and G.M. Cortelazzo. A noise-robust frequency domain tecnhique for estimating planar roto-translations. *IEEE Trans. Signal Processing*, 48(3):1769–1786, 2000.

[9] 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.

[10] P. Vandewalle, S. Susstrunk, and M. Vetterli. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP Journal on Applied Signal Processing*, 2006:1–14, 2006.