A UNIFIED APPROACH TO FFT BASED IMAGE REGISTRATION

A. Averbuch

Tel Aviv University
School of Computer Science
Tel Aviv 69978, Israel

Y. Keller

Tel Aviv University
Dept. of Electrical Engineering Systems
Tel Aviv 69978, Israel

ABSTRACT

We present a new unified approach to FFT based image registration. Prior works divided the registration process into two stages: the first was based on phase correlation (PC) which provides pixel accurate registration [5], while the second step provides subpixel registration accuracy [3, 1]. By extending the PC method we derive a FFT based image registration algorithm which is able to estimate large translations with subpixel accuracy. The algorithm's properties resemble those of the Gradient Methods [4] while outperforming it by exhibiting superior convergence range.

1. INTRODUCTION

Image registration is an essential component in many computer vision and image processing tasks such as motion analysis, video compression, image enhancement and restoration. Many of these tasks require pixel accurate registration (motion detection) while others (video compression, image enhancement, stereo vision) require sub-pixel accuracy. Gradient based methods (GM) [4] are considered to be the state-of-the-art registration algorithm, The major drawback of GM is the need for bootstrapping due to its limited convergence range. A possible bootstrapping algorithm is the phase correlation (PC) [5] scheme which utilizes the phase shift property [**?**] of the Fourier transform to robustly estimate large image translations without prior knowledge. An extension to sub-pixel registration was presented by Shekarforoush et-al [1], which is limited to subpixel shifts and has to be bootstrapped using pixel accurate PC. We propose to use frequency analysis to derive an Unified PC (UPC) image registration algorithm being able to estimate large translations with sub-pixel accuracy by utilizing the notion that the phase shifts between continuous signals are of infinite accuracy and are invariant to the sampling process. Further analysis reveals a close relationship to GM methods, while exhibiting superior analytical properties.

The paper is organized as follows: Problem formulation and basic structure of the pixel accurate PC are given in section 2. Fourier domain analysis leading to the sub-pixel registration algorithm are given in section 3. Implementation issues and results are discussed in 4.

2. BACKGROUND

2.1. Pixel Level Phase Correlation

2.1.1. Phase correlation

The purpose of the PC algorithm is to estimate the translation $(\Delta x, \Delta y)$ between a pair of images sharing some mutual support: Let $I_1(x,y)$ and $I_2(x,y)$ be the images described in Fig. 1, their mutual support satisfies the following equation:

$$I_1(x+\Delta x, y+\Delta y) = I_2(x,y) \tag{2.1}$$

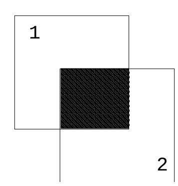

Fig. 1. Image translation

Equation 2.1 can be generalized to account for affine intensity changes :

$$I_1(x+\Delta x, y+\Delta y) = a \cdot I_2(x,y) + b, \quad a,b \in \Re \tag{2.2}$$

Equation 2.2 is solved by the PC algorithm using the *shift property* [**?**] of the FFT:

Denote:

$$\mathfrak{F}\{f(x,y)\} = \widehat{f}(\omega_x, \omega_y) \tag{2.3}$$

then

$$\mathfrak{F}\{f(x+\Delta x, y+\Delta y)\} = \widehat{f}(\omega_x, \omega_y)\, e^{j(\omega_x \Delta x + \omega_y \Delta y)} \tag{2.4}$$

We transform Eq. 2.1 to the Fourier domain using Eq. 2.4:

$$\widehat{I}_1(\omega_x, \omega_y)\, e^{j(\omega_x \Delta x + \omega_y \Delta y)} = \widehat{I}_2(\omega_x, \omega_y) \tag{2.5}$$

$$\frac{\widehat{I}_2(\omega_x, \omega_y)}{\widehat{I}_1(\omega_x, \omega_y)} = e^{j(\omega_x \Delta x + \omega_y \Delta y)} \tag{2.6}$$

The translation $(\Delta x, \Delta y)$ can be estimated by taking the inverse FFT of Eq. 2.6 :

$$Corr(x, y) = \mathfrak{F}^{-1}\left\{e^{j(\omega_x \partial x + \omega_y \Delta y)}\right\} = \delta(x - \Delta x, y - \Delta y) \tag{2.7}$$

Ideally, the peak value of the correlation surface $Corr(x, y)$ should be equal to 1.0, however, in real life scenarios due to the presence of random noise, dissimilar parts and non translational motion the peak value will be smaller than 1.0. Therefore, the translation values are estimated by:

$$(x, y) = \arg\left\{\max_{(\tilde{x}, \tilde{y})}\{Corr(\tilde{x}, \tilde{y})\}\right\} \tag{2.8}$$

In order to compensate for possible gain changes (Eq. 2.2) and since we are only interested in pure phase shifts, Eq. 2.5 can be rewritten as:

$$\widehat{Corr}(\omega_x, \omega_y) = \frac{\widehat{f_1}(\omega_x, \omega_y)\,\widehat{f_2}^{*}(\omega_x, \omega_y)}{\left|\widehat{f_1}(\omega_x, \omega_y)\right|\left|\widehat{f_2}(\omega_x, \omega_y)\right|} = e^{j(\omega_x \Delta x + \omega_y \Delta y)} \tag{2.9}$$

where * denotes the complex conjugate.

2.2. Sub-pixel registration

Shekarforoush et-al [1] presented the state-of-the-art interpolation-free, PC based, sub-pixel registration algorithm by analyzing the effects of sub-pixel shifts on the correlation function $Corr(x, y)$ in the *spatial* domain. The 1D inverse Fourier transform of a single frequency exponent, whose phase-shift relates to sub-pixel spatial shifts is:

$$\widehat{f}(i) = \left\{e^{ji\Delta\omega\Delta t}\right\}_i,\ 0 \le i < m,\ \Delta\omega = \frac{2\pi}{m}, \Delta t \notin \mathbb{Z} \tag{2.10}$$

The result $f(k) = \mathfrak{F}^{-1}\left\{\widehat{f}(i)\right\}$ is the Dirichlet Kernel:

$$D(\Delta t + k, m) = \frac{\sin(\pi(\Delta t + k))}{\sin(\pi(\Delta t + k)/m)}. \tag{2.11}$$

$D(\Delta t{+}k, N)$ coincides with $\delta(x, y)$ for integer pixel shifts while for sub-pixel shifts the center lobe's energy is split into side-lobes. Using the separability of the 2D DFT this analysis can be conducted on both axis, yielding:

$$\frac{\sin(\pi(\Delta x + x))}{\sin(\pi(\Delta x + x)/m)} \cdot \frac{\sin(\pi(\Delta y + y))}{\sin(\pi(\Delta y + y)/n)} = D(x, y), (x, y) \in \mathbb{Z} \tag{2.12}$$

where m, n are the image dimensions.

The non linear Eq. 2.12 is solved for $(\Delta x, \Delta y)$ using linearization of the denominator:

$$\frac{\sin(\pi\Delta x)}{\pi\Delta x/m} \cdot \frac{\sin(\pi\Delta y)}{\pi\Delta y/n} = D(x, y), (x, y) \in \mathbb{Z} \tag{2.13}$$

An estimate of $(\Delta x, \Delta y)$ is derived by substituting the correlation surface's main peak and neighboring peaks into Eq. 2.13. Thus, the accuracy of the estimate is linearly dependent upon the measurement accuracy of the side lobes, since the ratio of the highest side lobe to the main lobe is 21% (-13.5dB) [**?**]. Performance comparison between the results of [1] and the proposed UPC is presented in section 4.3.

3. FREQUENCY DOMAIN SUB-PIXEL REGISTRATION

The basic motivation leading to the proposed UPC sub-pixel registration is the notion that Eq. 2.5 holds for any non integer displacement values and its accuracy is only limited by errors induced by the problem setup: finite image size, non overlapping areas and illumination changes which are not modelled by Eq. 2.2. By calculating the phase shift directly, one can estimate an algebraically correct phase shift which is not limited to sub-pixel offsets or pixel accuracy.

3.1. Algorithm flow

The proposed sub-pixel evaluation process goes as follows:

1. Calculate the correlation surface $Corr(x, y)$ using Eq. 2.9 and project it into the appropriate registration uncertainty domain (section 3.2).
2. Calculate the pointwise phase shift functions in the Fourier domain, using the Fourier coefficients (section 3.3).
3. Using the robust estimator described in section 3.4, estimate the global sub-pixel accurate translation $(\Delta x, \Delta y)$.
4. Estimate and extract the mutual areas $\widetilde{I}_1$ and $\widetilde{I}_2$ created by the pixel level motion $(\lfloor\Delta x\rfloor, \lfloor\Delta y\rfloor)$ according to figure 1.
5. Move $\widetilde{I}_2$ by the sub-pixel motion $(\Delta x - \lfloor\Delta x\rfloor, \Delta y - \lfloor\Delta y\rfloor)$ (Section 3.5).
6. Go back to step # 2 until a stopping condition is met, our experience shows that a few iterations (usually 5-8 iterations) are sufficient for convergence to 10^{-3} accuracy.

3.2. Phase projection operator

Following the results of Section 2.2, the correlation surface $Corr\left(x,y\right)$ is expected to have a limited spatial support of 5÷7 pixels. Non-zeros elements beyond this support are a result of noisy measurement and should be filtered out. Hence, the phase-shift estimation can be improved by limiting the correlation surface's support to $\{[-R,R]\times[-R,R]\}$.

3.3. Phase estimation

The correlation $\widehat{Corr}(\omega_x,\omega_y)$ is calculated using Eq. 2.5, while the pointwise phase is estimated using:

$$\varphi\left(i,j\right)=\tan^{-1}\left(\widehat{Corr}(\omega_x\left(i,j\right),\omega_y\left(i,j\right))\right) \tag{3.1}$$

Where:

$$\left\{\begin{array}{l}\omega_x=j\cdot\frac{2\pi}{m}\\ \omega_y=i\cdot\frac{2\pi}{n}\end{array}\right. \quad i,j\in\sup port\left(\widehat{Corr}(\omega_x,\omega_y)\right) \tag{3.2}$$

This phase estimation is invariant to affine illumination changes since the phase is solely used. Gain differences do not affect the phase shift due to the normalization in Eq. 2.5, while offset changes affect only the DC coefficient's magnitude but not its phase.

3.4. Robust least squares estimator

Using $\varphi\left(i,j\right)$ estimated in Eq. 3.1 we derive a highly overdetermined linear equation set:

$$\left\{\varphi\left(i,j\right)=\omega_x\Delta x+\omega_y\Delta y\right\}_{i,j} \tag{3.3}$$

where $\left(\Delta x,\Delta y\right)$ are the estimated variables.

Equation 3.3 can be solved directly using least square (LS) which was found to perform well in noise-free scenarios. However, in order to register noisy images with significant dissimilarities, a robust LS estimator has to be introduced. Following [2] an iteratively re-weighted, least squares regression algorithms is used. Several weight functions were tested yielding similar results.

3.5. Sub-pixel Induced Translation

In most iterative sub-pixel registration schemes such as the GM, sub-pixel translation is induced using bilinear interpolation. We found it to be inadequate due to the non-linear shift induced at high-frequencies. In order to implement the sub-pixel translation we utilize a Fourier domain phase shift based on Eq. 2.5:

$$\widehat{I}_{shift}\left(\omega_x,\omega_y\right)=\widehat{I}_{base}\left(\omega_x,\omega_y\right)e^{j\left(\omega_x\widetilde{\Delta x}+\omega_y\widetilde{\Delta y}\right)} \tag{3.4}$$

where $\widehat{I}_{base}$ is the source image, I_{shift} is the result shifted image and $\left(\widetilde{\Delta x},\widetilde{\Delta y}\right)$ are the induced shifts.

Integral pixel shifts $\left(\lfloor\Delta x\rfloor,\lfloor\Delta y\rfloor\right)$ were implemented using mutual support extraction according to Figure 1, while the residual sub-pixel shifts $\left(\Delta x-\lfloor\Delta x\rfloor,\Delta y-\lfloor\Delta y\rfloor\right)$ were induced using Eq. 3.4.

4. EXPERIMENTAL RESULTS

4.1. Test image synthesis

In order to demonstrate the algorithm, the images shown in Fig. 2 were used: Fig. 2(a),(b) and (c) are airborne images of various types. (a) is a textured image lacking distinct features while (b) and (c) contain dominant man-made objects. Image (d) ("Barbara") was added to provide a different kind of test image. The images were shifted using a procedure similar to the one used in [1]. Let $I_1\left(x,y\right)$ and $I_2\left(x,y\right)$ be the input image and $\left(\Delta x,\Delta y\right)=\left(\frac{k_1}{m_1},\frac{k_2}{m_2}\right),\left(k_1,k_2,m_1,m_2\right)\in\mathbb{Z}$ are the induced sub-pixel shifts, then:

1. $I_1\left(x,y\right)$ and $I_2\left(x,y\right)$ are smoothed using a Gaussian kernel
2. Using Fig. 1, extract $\widetilde{I}_{base}\left(x,y\right)$ and $\widetilde{I}_{shift}\left(x,y\right)$, the mutual supports of $I_1\left(x,y\right)$ and $I_2\left(x+k_1,y+k_2\right)$ respectively.
3. $\widetilde{I}_{base}\left(x,y\right)$ and $\widetilde{I}_{shift}\left(x,y\right)$ are downsampled by the factors $\left(m_1,m_2\right)$ in $\left(x,y\right)$, respectively.

4.2. Registration accuracy estimation

Estimation of larger shifts was tested using the images in Fig. 2. They were shifted by as much as 20 pixels, each shift, having a sub-pixel component. The results which are displayed in Table 1 shows that the average UPC registration error is of $O\left(10^{-3}\right)$ pixel and convergence is achieved in 6–7 iterations. The robust LS compared well to the regular LS showing superior accuracy of $O\left(\frac{5}{1000}\right)$ pixel at a cost of a significant complexity increase, hence it seems that the regular LS is better suited for most practical scenarios. However, our results suggests that when a significant amount of noise or dissimilarity between the input images exists , the robust LS should be considered.

4.3. Comparison to prior results

We compare our results to those reported by Shekarforoush et-al by synthesizing the shifted "Pentagon" image used in [1]. Table 2 points out that our algorithm significantly outperforms the results of [1] using either the robust or regular LS. Since we use iterative refinement, Table 2 also presents

Image	$(\Delta x, \Delta y)$	UPC using robust regression	UPC using regular least square
Pentagon	$(2.50, -2.50)$	$(2.500, -2.500), 5$	$(2.501, -2.501), 6$
	$(3.25, 1.50)$	$(3.249, 1.505), 6$	$(3.248, 1.513), 8$
	$(-5.25, -4.50)$	$(-5.249, -4.509), 8$	$(-5.248, -4.510), 8$
	$(0.0, 15.75)$	$(-0.000, 15.752), 9$	$(-0.000, 15.752), 6$
Airfield	$(0.167, -0.50)$	$(0.170, -0.504), 5$	$(0.170, -0.504), 5$
	$(0.5, 3.25)$	$(0.500, 3.252), 5$	$(0.501, 3.251), 5$
	$(-7.33, -1.167)$	$(-7.338, -1.169), 8$	$(-7.334, -1.168), 6$
	$(-7.33, 2.83)$	$(-7.337, 2.838), 8$	$(-7.336, 2.839), 6$

Table 1. Result of large shift estimation

Image	$(\Delta x, \Delta y)$	Shekarforoush [1]	Unified PC 1 iteration	Unified PC 3 iterations
Pentagon [1]	$(0.50, -0.50)$	$(0.48, -0.51)$	$(0.48, -0.51)$	$(0.4997, -0.5005)$
	$(0.25, 0.50)$	$(0.28, 0.49)$	$(0.28, 0.49)$	$(0.2499, 0.5008)$
	$(-0.25, -0.50)$	$(-0.25, -0.52)$	$(-0.25, -0.52)$	$(-0.2504, -0.5002)$
	$(0.0, 0.75)$	$(0.0, 0.80)$	$(0.0, 0.80)$	$(-0.0006, 0.7510)$

Table 2. Comparison of sub-pixel shift results to Shekarforoush[1]

(a) (b)

Fig. 2. Some of the images used for verification: (a) "Pentagon" and (b) "Airfield".

the results of the UPC's first iteration indicating a significant improvement compared to [1]. Note that the algorithm in [1] is unable to directly estimate non-sub-pixel shifts, thus, no such comparisons is needed. The use of the robust LS results in a higher complexity of the UPC. However, when the regular LS is used, the complexities of the UPC and [1] are dominated by the FFT algorithm and are approximately the same.

5. SUMMERY AND CONCLUSIONS

In this paper we proposed a new approach to robust subpixel image registration. The algorithm extends the Phase Correlation method to sub-pixel accuracy and was proven to posses superior convergence range and accuracy of $\mathcal{O}\left(10^{-3}\right)$, while being invariant to illumination changes and blurring. By calculating the correlation function's phase, we derived exact expressions for least-square estimation of the phase-shift without resulting to approximations. The least-square problem was solved by either the regular or the robust least-square algorithms depending on the amount of noise present. The experimental results are backed up by analytical analysis establishing a relation to prior works.

6. REFERENCES

[1] H. Shekarforoush, M. Berthod, and J. Zerubia,"Subpixel image registration by estimating the polyphase decomposition of cross power spectrum", in Proceedings 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, USA, June 1996, pp. 532–537, IEEE.

[2] W.H. DuMouche and F.L. O'Brien, "Integrating a robust option into a multiple regression computing environment", Computer Science and Statistics: Proceedings of the 21st Symposium on the Interface, Alexandria, VA, 1989.

[3] Robert Frischholz, "Class of algorithms for realtime sub-pixel registration"

[4] L. Barron, D. J. Fleet and S. S. Beauchemin, "Performance of Optical flow Techniques", Int. J. Computational Vision, Vol. 12, pp. 43-77, 1994.

[5] C. D. Kuglin and D. C. Hines. "The phase correlation image alignment method" In IEEE 1975 Conference on Cybernetics and Society, pages 163–165, September 1975.