SUB-PIXEL REGISTRATION AND ESTIMATION OF LOCAL SHIFTS DIRECTLY IN THE FOURIER DOMAIN

Hassan Foroosh Murat Balci

School of Computer Science
University of Central Florida, FL, USA
{foroosh, balci}@cs.ucf.edu

ABSTRACT

In this paper, we propose a new approach to sub-pixel registration, and the estimation of local shifts between a pair of images, directly in the Fourier domain. For this purpose, we establish the exact relationship between the continuous and the discrete Fourier phase of two shifted images or their subregions. In particular, we show that the discrete phase difference of two shifted images is a 2-dimensional sawtooth signal, with the exact shifts determined to sub-pixel accuracy by the number of periods of the phase difference along each frequency axis. The sub-pixel portions of the shifts are determined by the non-integer fraction of a period of the phase difference.

1. INTRODUCTION

Registration is a crucial step in the analysis and fusion of information between multiple images. Examples can be found in remote sensing, robotics, and bio-medical imaging, and many other important areas [1-3]. In some applications, such as coding and compression registration needs to be established locally, while other applications such as stereovision require registering locally but along some specific directions defined by the epipolar lines.

Registration techniques typically assume that motion can be modeled using a given family of transformations such as rigid, affine, or Euclidean. Registration is performed by searching for a particular transformation within the family that optimizes some similarity or redundancy criterion, e.g. correlation coefficients or mutual entropy.

Herein, we are interested in investigating the information contained in the phase domain. We are particularly motivated by applications that require registration at sub-pixel accuracy. Examples of such applications include super-resolution from multiple views [4-7] or examination of same-patient MRI data in a clinical setting [8].

This work was partially supported by ONR grant #N000140410512 and Sun Microsystem's grant #EDUD-7824-030482-US

2. THEORETICAL BACKGROUND

Let $f_1(x,y)$ and $f_2(x,y)=f_1(x-x_o,y-y_o)$ be two bandlimited functions with their relative shifts given by (x_o,y_o). Their cross power spectrum is then given by

$$\hat{c}(u,v)=\frac{\hat{f}_1\hat{f}_2^*}{\left|\hat{f}_1\hat{f}_2^*\right|} \tag{2.1}$$

where the hat sign as usual indicates the Fourier transform and the asterisk stands for the complex conjugate.

As is well known due to the whitening (i.e. the denominator) the magnitude is normalized to unity for all frequencies, and due to the Fourier shift property the spatial translations lead to linear phase shifts between the two phase functions along each frequency axis, i.e.

$$\angle\hat{c}(u,v)=x_o u+y_o y \tag{2.2}$$

which is a planar surface through the origin.

Using discrete Fourier transform (2,1) is shown to be equally applicable in the discrete case, and yield remarkably good results for images. The shift parameters can then be determined by inverse transforming the discrete $\hat{c}(u,v)$, which would clearly lead to a unit impulse centered at (x_o,y_o). This of course is true only for integer displacements. For non-integer (i.e. shifts with sub-pixel components) additional care needs to be taken, which is discussed in [9].

0-7803-8554-3/04/$20.00 ©2004 IEEE.

When applying this approach locally, within a small window size, the estimation of local motion (disparities) becomes inaccurate and dominated by noise. The main cause for this is due to the fact that the Fourier transform is an operator that can not localize signals both in space and frequency. As a result the noise process and the aliasing errors, which often are localized at the high frequency components of the Fourier spectrum, become dispersed in the spatial domain.

A possible approach to overcome this problem is therefore to avoid inverse transforming the cross power spectrum, and try to estimate the disparities directly in the Fourier domain. The often overlooked difficulty with this approach is that due to discretization the components of the phase matrix do not form a planar surface unless the 2-dimensional phase matrix is first unwrapped. It is worth noting that 2-dimensional phase unwrapping is a notoriously ill-posed problem. We will show that the problem is not truly that difficult in this case.

Let us denote the discrete phase matrix of the cross power spectrum by

$$\mathbf{P} = [p_{mn}] \tag{2.3}$$

where $m=0,\ldots,M-1$, and $n=0,\ldots,N-1$.

The phase of the underlying continuous signal has the following representation in the spatial domain

$$\begin{aligned}\varphi(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_o u + y_o v)\exp(iux + ivy)\,du\,dv \\ &= -x_o \frac{d\delta(x)}{dx} y_o \frac{d\delta(y)}{dy}\end{aligned} \tag{2.4}$$

Although inverse transforming the phase function in this way may not be customary, it streamlines the understanding of the relationship between the discrete phase matrix **P** and the phase of the underlying continuous signal, as shown below. From bandlimited sampling theory and (2.4) it follows that the spatial domain representation of a phase component in **P** is given by

$$\begin{aligned}\varphi(\mathrm{k},\mathrm{l}) = -\frac{x_o}{\pi k}&\left(2\operatorname{sinc}\frac{\pi}{x_o}k - 2\cos\frac{\pi}{x_o}k\right) \\ \frac{y_o}{\pi l}&\left(2\operatorname{sinc}\frac{\pi}{y_o}l - 2\cos\frac{\pi}{y_o}l\right)\end{aligned} \tag{2.5}$$

where the scaling by the inverse of x_o and y_o in the sinc and cos functions reflect the fact that the absolute value of the discrete phase can not exceed π.

This latter equation can also be written as

$$\varphi(m,n) = x_0\left(\frac{2\operatorname{sinc}\frac{\pi}{x_o}k - 2\cos\frac{\pi}{x_o}k}{i\pi k}\right) \; y_0\left(\frac{2\operatorname{sinc}\frac{\pi}{y_o}l - 2\cos\frac{\pi}{y_o}l}{i\pi l}\right) \tag{2.6}$$

$$\begin{aligned}&= x_0\left(\frac{1}{i\pi k} + \pi\delta(k)\right)\left(2\operatorname{sinc}\frac{\pi}{x_o}k - 2\cos\frac{\pi}{x_o}k\right) \\ &\quad y_0\left(\frac{1}{i\pi l} + \pi\delta(l)\right)\left(2\operatorname{sinc}\frac{\pi}{y_o}l - 2\cos\frac{\pi}{y_o}l\right)\end{aligned} \tag{2.7}$$

which shows that $\frac{d\mathbf{P}}{dm}$ and $\frac{d\mathbf{P}}{dn}$ consist of periodic sequences of superposition of a pulse corresponding to the sinc function and a pair of impulses corresponding to the cos function. It therefore follows that

$$p_{mn} = \begin{cases} 2\pi\left(x_0\frac{m}{M} + y_0\frac{n}{N}\right) & \text{if } \left|x_0\frac{m}{M} + y_0\frac{n}{N}\right| < \frac{1}{2} \\ \text{otherwise repeated periodically} & \end{cases} \tag{2.8}$$

Comparing this with equation (2.2) we notice that the discrete phase matrix is not really a plane through the origin, but rather a 2D sawtooth function whose slopes along the two frequency axes are given by x_0 and y_0.

This implies that the subspace approximation of the phase matrix using singular value decomposition as suggested recently in [12] (see also [10]) can be avoided. This follows immediately from the fact that the discrete phase in matrix **P** contains the samples of a sawtooth signal, and hence is not really of rank 2, but rather becomes of rank 2 after unwrapping. Equivalently, $\exp(i\mathbf{P})$ becomes of rank 1 only after unwrapping. Since the unwrapping of **P** should yield a rank 2 matrix, it immediately follows that the unwrapping process in the case of 2-dimensional shifts is a separable process, i.e. it can be performed independently along each frequency axis directly on the phase matrix. Note that unlike 2D unwrapping, 1D unwrapping is a well-posed problem, and hence easy to solve.

Once the matrix **P** is unwrapped, we obtain a new phase matrix **U**. In theory, any row of this matrix provides a solution for x_0, and equivalently any column of **U** provides a solution for y_0. For instance, for a given row $\mathbf{r}_m$ and a column $\mathbf{c}_n$ of the unwrapped phase matrix **U**, we have

$$x_0 = \frac{M}{2\pi}\frac{d\mathbf{r}_m}{dx}, \quad y_0 = \frac{N}{2\pi}\frac{d\mathbf{c}_n}{dy} \qquad (2.9)$$

These are by definition the instantaneous frequencies of the correlation matrix, if we use a short-length Fourier transform. What these analyses show is that for translational motion - since the unwrapped phase matrix is only rank-2 - the instantaneous frequencies defined by short-length Fourier transform and the formal definition in terms of the derivative of phase coincide.

However, due to noise and other sources of error (see for instance [9-11] for a thorough description of sources of error), the derivatives in equations (2.9) may lead to inaccurate results. Fortunately, one can readily overcome this problem by using the fact that N such rows (respectively M such columns) are available for regression. A robust solution can thus be obtained by using the least median square criterion

$$(\hat{x}_0, \hat{y}_0) = \arg \min_{\substack{m=1,\ldots,M \\ n=1,\ldots,N}} \left\{ \operatorname{med}\left(\frac{2\pi}{M}x_0 - \frac{d\mathbf{r}_m}{dx}\right)^2 + \operatorname{med}\left(\frac{2\pi}{N}y_0 - \frac{d\mathbf{c}_m}{dy}\right)^2 \right\} \qquad (2.10)$$

Note that this is a simple fitting problem and not a complex non-linear optimization issue. Therefore the complexity of the algorithm is essentially determined by the Fast Fourier Transform (FFT) (i.e. N log(N), for N points).

3. EXPERIMENTAL RESULTS

We applied the technique to both global sub-pixel registration problem and local motion (disparity) estimation. In both cases excellent results were obtained.

For global registration, we used the approach described in [9], to generate images with sub-pixel shifts, i.e. starting from a real high resolution image, we lowpass filtered and downsampled shifted versions of the image. Using appropriate downsampling rates, shifts with sub-pixel contents were produced. Figure 1 shows some of the images to which the technique was applied. Results are shown in table 1 and are compared to those reported in [9]. On average the accuracy was equal or higher than [9], but of course with less required computational time since inverse transforming is avoided.

We also applied the technique in a framework using short-length Fourier transform. This would allow us to build a space-frequency representation of the data directly in the Fourier domain, using the instantaneous frequencies characterized by a windowed Fourier transform. We particularly applied the results to some rectified stereo pairs. This implies that the motion is restricted locally to be along the epipolar lines that are typically warped and mapped to the image scan lines. Results are shown in Figures 2 and 3 for the Pentagon image and the baseball image. Textured map models built from these space-frequency representations are also shown for comparison.

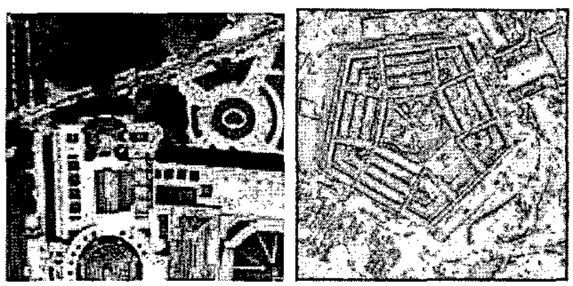


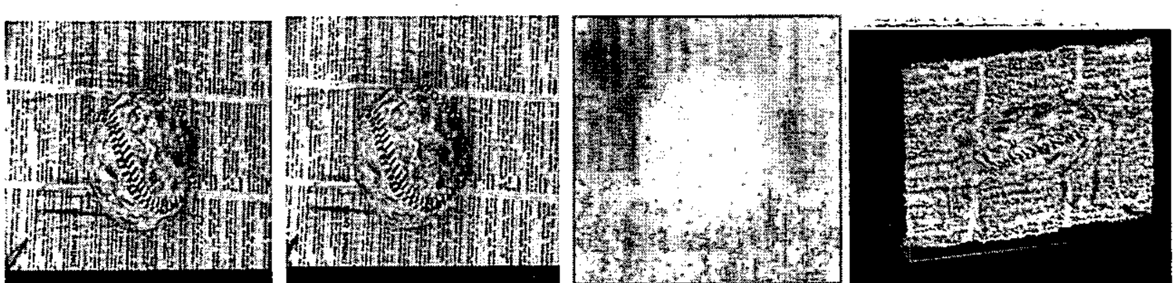
Figure 1: Some of the images used for simulation.

Image	True Shifts	Foroosh et al. (2001)	Proposed Method
Pentagon	**(0.50,-0.50)**	**(0.48,0.51)**	**(0.49,0.51)**
	(0.25,0.50)	**(0.28,0.49)**	**(0.26,0.51)**
	(-0.25,-0.50)	**(-0.25,-0.52)**	**(-0.26,-0.49)**
	(0.0,0.75)	**(0.0,0.80)**	**(0.0,0.77)**
Paris	**(0.167,-0.5)**	**(0.152,-0.49)**	**(0.161,-0.49)**
	(0.67,0.25)	**(0.69,0.33)**	**(0.67,0.27)**
	(-0.33,-0.167)	**(-0.32,-0.15)**	**(-0.31,-016)**
	(0.33,0.33)	**(0.325,0.32)**	**(0.33,0.34)**

TABLE 1: Results for simulated global registration

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

Figure 2: (a) & (b) Stereo pair, (c) space-frequency representation, (d) snapshot of a textured 3D Model.



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

Figure 3: (a) and (b): Baseball stereo pair, (c) the recovered space-frequency representation,(d) texture mapped 3D model.

4. CONCLUSION

The analyses and the experimentations presented in this paper show that accurate results can be obtained for sub-pixel registration directly in the Fourier domain, even when applied to small image regions. We have shown that in the discrete case the phase matrix of the cross-power spectrum of two image signals has a 2-dimensional sawtooth form. This implies that the sub-pixel portions of the shifts are determined by the non-integer fraction of a period of the phase difference. Furthermore, the instantaneous frequencies can be readily determined by differentiating the phase matrix using a robust estimator, such as the least median square. This provides information about the 3D structure of the scene.

5. REFERENCES

[1] B. Zitova and J. Flusser, Image Registration Methods: A Survey, Image and Vision Computing, vol. 21, 977-1000, 2003.

[2] L.G. Brown, A survey of image registration techniques, ACM Computing Surveys, vol. 24 , 326–376, 1992.

[3] J.B.A. Maintz, M.A. Viergever, A survey of medical image registration, Med. Image Anal vol. 2, 1–36, 1998.

[4] H. Shekarforoush (Foroosh), M. Berthod, M. Werman and J. Zerubia, *Reconstruction of High Resolution 3D Visual Information*, International Journal of Computer Vision, vol. 19, no.3, pp. 289-300, 1996.

[5] H. Shekarforoush (Foroosh) and R. Chellappa, *"Data-Driven Multi-channel Super-resolution with Application to Video Sequences"*, Journal of Optical Society of America-A, vol. 16, no. 3, pp. 481-492, 1999.

[6] M. Irani and S. Peleg, Improving Resolution by Image Registration, Graphical Models and Image Processing, vol. 53, no. 3, pp. 231-239, 1991.

[7] S. Peleg, D. Keren and L. Schweitzer, Improving Image Resolution using Subpixel Motion, Pattern Recognition Letters, vol. 5, pp. 223-226, 1987.

[8] J.V. Hajnal, D.L.G. Hill, D.J. Hawkes, Medical Image Registration, CRC Press, Baton Rouge, Florida, 2001.

[9] H. Foroosh, J. Zerubia, and M. Berthod, Extension of Phase Correlation to Sub-pixel Registration, IEEE Trans. Image Proc., vol. 11, Issue 3, pp. 188-200, 2002.

[10] H. Foroosh and S. Hoge, *Motion Information in the Phase Domain*, Chapter 3, Video registration", Kluwer Academic Publishers, 2003.

[11] H. S. Stone, M. Orchard, E.-C. Chang, and S. Martucci, A fast direct Fourier-based algorithm for subpixel registration of images, *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 10, pp. 2235–2243, 2001.

[12] W.S. Hoge, Subspace Identification Extension to the Phase Correlation Method *IEEE Trans. Medical Imaging*, 22(2):277-280, 2003