INFERRING MOTION FROM THE RANK CONSTRAINT OF THE PHASE MATRIX

Murat Balci *Hassan Foroosh*

School of Computer Science, University of Central Florida

ABSTRACT

In this paper, we investigate the rank constraint of the discrete phase difference, and derive its exact parametric model. We show that the discrete phase difference of two shifted images, or their subregions, is a 2-dimensional sawtooth signal. This allows us to determine the motion parameters to subpixel accuracy by simply counting the number of cycles of the phase difference along each frequency axis. The subpixel portion is given by the non-integer fraction of the last cycle along each axis. The problem is formulated as a homogeneous cost function under rank constraint for the phase matrix, and the shape constraint for the filter that computes the group delay, and is solved using a robust technique.

1. RELATION BETWEEN DISCRETE AND CONTINUOUS PHASE DIFFERENCES

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

$$\hat{c}(u,v) = \frac{\hat{f}_1 \hat{f}_2^*}{|\hat{f}_1 \hat{f}_2^*|} \tag{1}$$

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 Fourier shift property the spatial translations lead to linear phase differences between the two functions along each frequency axis, i.e.

$$\angle \hat{c}(u,v) = x_o u + y_o v \tag{2}$$

which is a planar surface through the origin.

Eq. (1) is shown to be equally applicable in the discrete case, and yield remarkably good results for image registration. The motion parameters are determined by inverse transforming the discrete $\hat{c}(u,v)$, which yields a discrete Dirichlet function [1]. The solution can then be found by a least-squares fitting. However, if we apply this approach locally, the estimation of local motion parameters becomes inaccurate and dominated by errors. The main cause for this is the noise process and the aliasing errors, which often are localized at the high frequency components of the Fourier spectrum, but become dispersed in the spatial domain upon inversion of $\hat{c}(u,v)$. To overcome this problem, we can estimate the displacements directly in the Fourier domain.

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

A practical solution for this problem was first proposed by Hoge [2], who suggested that the two frequency axes can be decoupled by a subspace approximation of the phase correlation matrix $\exp(i\mathbf{P}(m,n))$. The advantage is that the unwrapping step can then be performed on the 1-dimensional dominant left and right eigenvectors of the phase correlation matrix, rather than directly on the phase matrix itself i.e. $\mathbf{P}$ - recall that 2D phase unwrapping is known to be notoriously ill-posed. We will show below that due to the rank constraint of the unwrapped phase matrix, the unwrapping process becomes separable along the two frequency axes. In other words, it reduces to two 1-dimensional unwrapping steps. As a result very good results can also be found without subspace approximation. Furthermore, We will show that even phase unwrapping is an unnecessary step, since we will determine the exact parametric shape of the phase difference matrix.

Proposition *Let $\mathbf{P} = [p_{mn}]$ be the discrete phase difference matrix of two images shifted by (x_o, y_o), where $m = 0, ..., M-1$, and $n = 0, ..., N-1$. We maintain that $\mathbf{P}$ is a 2D sawtooth signal, with periods $\frac{2\pi}{x_o}$ and $\frac{2\pi}{y_o}$.*

Proof: The phase difference of the underlying continuous signals is given in the spatial domain by

$$\varphi(x,y) = \int_{\infty}^{\infty}\int_{\infty}^{\infty} (x_0 u + y_o v)\exp(iux + ivy)dudv \tag{3}$$

$$= -ix_o \frac{d\delta(x)}{dx} - iy_o \frac{d\delta(y)}{dy} \tag{4}$$

where the derivatives are in distributional sense [3]. From bandlimited sampling theory and (4), the spatial domain representation of the discrete phase difference is given by

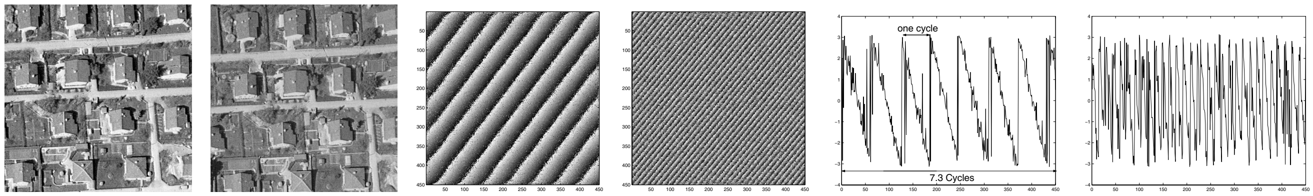
$$\varphi_{kl} = -i\frac{x_o}{\pi k}\left(2\text{sinc}\frac{\pi k}{x_o} - 2\cos\frac{\pi k}{x_o}\right) - i\frac{y_o}{\pi l}\left(2\text{sinc}\frac{\pi l}{y_o} - 2\cos\frac{\pi l}{y_o}\right) \tag{5}$$

Therefore

$$\mathcal{R}e\{\varphi_{kl}\}$$
$$= \frac{x_o^2}{\pi k^2}\left(2\frac{\pi k}{x_o}\cos\frac{\pi k}{x_o} - 2\sin\frac{\pi k}{x_o}\right) + \pi\frac{y_o^2}{\pi l^2}\left(2\cos\frac{\pi l}{y_o} - 2\sin\frac{\pi l}{y_o}\right)$$
$$= x_o \frac{2}{2\pi/x_o}\int_{-\frac{\pi}{x_o}}^{\frac{\pi}{x_o}} u\sin ku \,\mathrm{d}u + y_o\frac{2}{2\pi/y_o}\int_{-\frac{\pi}{y_o}}^{\frac{\pi}{y_o}} v\sin lv\,\mathrm{d}v \tag{6}$$

On the other hand, it can be verified that

$$x_o\frac{2}{2\pi/x_o}\int_{-\frac{\pi}{x_o}}^{\frac{\pi}{x_o}} u\cos ku\mathrm{d}u + y_o\frac{2}{2\pi/y_o}\int_{-\frac{\pi}{y_o}}^{\frac{\pi}{y_o}} v\cos lv\mathrm{d}v = 0 \tag{7}$$



(a) (b) (c) (d) (e) (f)

Fig. 1. (a) & (b) Aerial images with some shifts, (c) & (d) noisy sawtooth phase matrices corresponding to shifts of (7.3,5.6) and (30.5,25.4) pixels, (e) & (f) one row of the phase matrices shown in (c) & (d), respectively.

and similarly

$$x_o \frac{2}{2\pi/x_o} \int_{-\frac{\pi}{x_o}}^{\frac{\pi}{x_o}} u \mathrm{d}u + y_o \frac{2}{2\pi/y_o} \int_{-\frac{\pi}{y_o}}^{\frac{\pi}{y_o}} v \mathrm{d}v = 0 \tag{8}$$

From (6), (7), and (8), and using the definition of the Discrete Fourier Transform (DFT) based on Fourier series [4], it follows immediately upon substituting $u = n\frac{2\pi}{N}$ and $v = m\frac{2\pi}{M}$ that φ_{kl} is a DFT coefficient of the following discrete periodic signal

$$p_{mn} = \begin{cases} 2\pi\left(x_o \frac{n}{N} + y_o \frac{m}{M}\right) & \\ p_{m'n} & \text{if } \ m'\frac{2\pi}{M} = m\frac{2\pi}{M} + J\frac{2\pi}{x_o} \\ p_{mn'} & \text{if } \ n'\frac{2\pi}{N} = n\frac{2\pi}{N} + J\frac{2\pi}{y_o} \end{cases} \tag{9}$$

where J is an arbitrary integer. □

Therefore $\mathbf{P}$ is a 2D sawtooth signal as opposed to the continuous phase difference in (2), which is a plane through the origin. Figure 1 illustrates this result. The first illuminating observation that can be made from this result is that the unwrapping of a 2-dimensional sawtooth signal is separable, since its unwrapped matrix has to be rank-2. This implies that a subspace approximation is not really required. The second observation is that unwrapping is also an unnecessary step, since the motion parameters can be determined simply by the number of cycles along each frequency axis. For instance, since the period of the sawtooth signal along the u-axis is $\frac{2\pi}{x_o}$, there are x_o repeated cycles along each row of $\mathbf{P}$, where x_o may or may not be an integer. This process of counting the number of cycles along the rows and the columns of the phase matrix is essentially all that is required to determine the local or the global motions.

2. SOLVING FOR LOCAL OR GLOBAL MOTIONS

As we showed above, the key to solve the problem is to find how many cycles of the sawtooth phase difference fit in the range $[0, 2\pi]$ along each frequency axis. The number of cycles i.e. x_o and y_o may or may not be integer values, and are given by

$$x_o = \frac{\text{cycles}}{2\pi} = \frac{N}{2\pi}\frac{d\mathbf{P}(m,n)}{dn} \quad \text{and} \quad y_o = \frac{\text{cycles}}{2\pi} = \frac{M}{2\pi}\frac{d\mathbf{P}(m,n)}{dm} \tag{10}$$

However, due to noise and the discontinuities of the sawtooth signal, counting the number of cycles per 2π using the equations in (10) would lead to inaccurate results. To overcome this problem, we need to use the fact that in a window of width N and height M there are $M \times N$ data points available for regression. Furthermore, the gradient at the discontinuities along a row $\mathbf{r}_m$ or a column $\mathbf{c}_n$ may be treated as outliers. Therefore, a robust estimator can be designed to eliminate the influence of outliers and noise. We will show the derivations for the columns of $\mathbf{P}$. However, the approach is equally applicable to the rows.

Essentially, we need to design an optimal filter $\mathbf{h}$ that can compute the slope of the noisy sawtooth signal along the columns of $\mathbf{P}$. We will model the filter as a finite impulse response (FIR) filter. Since a gradient filter is expected to be anti-symmetric, we will assume that $\mathbf{h}$ is a type III FIR filter, of length $2L+1$, i.e. $\mathbf{h} = [h_1, ..., h_{2L+1}]^T$. The gradient of the n^{th} column is then given by

$$\mathbf{H}\mathbf{c}_n = \mathbf{c}'_n \tag{11}$$

where $\mathbf{c}'_n = [c'_{L+1}, ..., c'_{M-L}]^T_n$ is the gradient vector truncated at both ends to avoid border artifacts, and $\mathbf{H}$ is a $(M-2L) \times M$ matrix given by

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}^T & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{h}^T \end{bmatrix} \tag{12}$$

For a general phase matrix, the problem would be a blind one, since we would need to determine both $\mathbf{H}$ and $\mathbf{c}'_n$. However, in our case, since $\mathbf{P}$ is a sawtooth signal, except for a small number of discontinuities, we have $\mathbf{c}'_n = \frac{2\pi}{M} y_o \mathbf{l}$, where $\mathbf{l} = [1, ..., 1]^T$ is a vector of length $M-2L$. As a result, after some algebraic manipulation, the equation in (11) can be written in the homogeneous form as

$$\tilde{\mathbf{C}}\tilde{\mathbf{h}} = \mathbf{0} \tag{13}$$

where

$$\tilde{\mathbf{C}} = \begin{bmatrix} \mathbf{C} & \frac{-2\pi}{M}\mathbf{l} \end{bmatrix} = \begin{bmatrix} c_1 & \cdots & c_{2L+1} & -\frac{2\pi}{M} \\ c_2 & \cdots & c_{2L+2} & -\frac{2\pi}{M} \\ \vdots & \dots & \vdots & \vdots \\ c_{n-2L} & \cdots & c_n & -\frac{2\pi}{M} \end{bmatrix} \tag{14}$$

and $\tilde{\mathbf{h}} = [h_1 \ \dots \ h_{2L+1} \ y_o]^T$.

There are two important constraints that apply to $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{h}}$. By inspection, we can verify that $\mathbf{C}$ is a rank-2 matrix (assuming discontinuities are due to outlying data), and hence so is $\tilde{\mathbf{C}}$. On the other hand, we know that for a noise-free phase matrix, the filter $\mathbf{h}$ should be anti-symmetric, i.e. $h_{L+1+i} = h_{L+1-i}$, for $i = 1, ..., L$ and $h_{L+1} = 0$. Therefore, we can formulate the problem as follows:

$$\tilde{\mathbf{h}}_{\text{opt}} = \arg\min \|\tilde{\mathbf{C}}\tilde{\mathbf{h}}\| + \lambda\|\mathbf{A}\mathbf{h}\| \tag{15}$$

where λ is the regularization parameter, and $\mathbf{A} = [a_{ij}]$ is a $(L+1) \times (2L+1)$ matrix defined by

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \text{ or } i + j = 2L + 2 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

This formulation is basically a semi-norm Tikhonov-Arsenin regularization of the problem in (13). The first term imposes the rank constraint on $\tilde{\mathbf{C}}$ by assuming that the last column of $\tilde{\mathbf{C}}$ is a constant, and the second term imposes the anti-symmetry constraint on $\mathbf{h}$. The solution is given by $\tilde{\mathbf{h}}_{\text{opt}} \sim [\mathbf{h}_{\text{opt}}^T \;\; 1]^T$, where

$$\mathbf{h}_{\text{opt}} = (\mathbf{C}^T\mathbf{C} + \lambda\mathbf{A}^T\mathbf{A})^{-1}\mathbf{C}\mathbf{l} \tag{17}$$

Note that, since our formulation in (15) is homogeneous, the solution is found only up to a scale factor as indicated by the $\sim$ notation. This scale ambiguity, however, can be readily resolved by assuming that $\mathbf{h}$ is the discrete gradient of a smoothing kernel that preserves the first moment. This implies that the components of $\mathbf{h}_{\text{opt}}$ should satisfy

$$\sum_{j=1}^{2L+1}\sum_{i=1}^{j} h_i = 1 \tag{18}$$

We now have the solution given by (17) and (18) up to an unknown regularization parameter. The optimal value of this parameter is given by the method of Generalized Cross Validation (GCV), which amounts to minimizing

$$\text{GCV}(\lambda) = \frac{\|(\mathbf{I} - \mathbf{C}(\mathbf{C}^T\mathbf{C} + \lambda\mathbf{A}^T\mathbf{A})^{-1}\mathbf{C}^T)\mathbf{l}\|^2}{(\text{tr}(\mathbf{I} - \mathbf{C}(\mathbf{C}^T\mathbf{C} + \lambda\mathbf{A}^T\mathbf{A})^{-1}\mathbf{C}^T))^2} \tag{19}$$

with respect to λ. In the existing literature, the minimizer of (19) is usually obtained by using numerical techniques, e.g. the quadrature rules and the Lanczos algorithm [5]. However, in our case, due to the rank constraint of $\mathbf{C}$, we can find a simplified closed-form solution.

Lemma *Let $\mathbf{K} = \mathbf{C}\mathbf{A}^\dagger$, where $\mathbf{A}^\dagger$ is the Moore-Penrose pseudo-inverse of $\mathbf{A}$. Let also $\mathbf{V}\Sigma\mathbf{V}^T$ be the spectral decomposition of $\mathbf{K}\mathbf{K}^T$. A first order approximation of the optimal minimizer of the GCV function in* (19) *is given by*

$$\lambda^* = \frac{\sigma_1 \sum_{j=2}^{N-2L} s_j^2}{(N - 2L - 1)s_1^2 - \sum_{j=2}^{N-2L} s_j^2} \tag{20}$$

where σ_1 is the dominant eigenvalue in the diagonal matrix Σ, and s_j's are the components of the vector $\mathbf{V}^T\mathbf{l}$.

The proof is left out due to lack of space. □

Using (17), (18), and (20), we can compute a total number of $T = (M - 2L) \times N$ estimated values for y_o (or equivalently $(N - 2L) \times M$ values for x_o). The question now is how to use this highly redundant amount of information to estimate y_o and x_o, robustly.

3. ROBUSTIFYING THE SOLUTION

The computed values for y_o, i.e. $y_o^j, j = 1, \ldots, T$ are noisy and with outliers. The probability density function of the contaminated y_o can therefore be modeled as a mixture of the inlying and the outlying densities. Assuming that these two densities are normally distributed $\mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$, our model can be written as

$$p(y_o|\Theta) = \sum_{i=1}^{2} p_i p_i(y_o|\theta_i) \tag{21}$$

where p_i is the prior probability that y_o is drawn from the distribution i, such that $\sum_{i=1}^{2} p_i = 1$, $\theta_i = [p_i, \mu_i, \sigma_i^2]$, and

$$p_i(y_o|\theta_i) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(y_o - \mu_i)^2}{2\sigma_i^2}\right) \tag{22}$$

This clearly is a parametric model, and the parameter vector to be estimated is given by $\Theta = [p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2]^T$, where $p = p_1$ and $p_2 = 1 - p$. A maximum likelihood estimation of the parameter vector, and hence separation of the inlying density can be done using the Expectation Maximization (EM) algorithm [6]. Separating the inlying values in this manner is essentially equivalent to truncating the quadratic error in (15), which is known to yield a robust technique. The following EM steps can be readily derived for our mixture model in (21) using the general theory of the EM algorithm. The $t + 1$ iteration is given by

- **E-step:** Estimate $p(i|y_o^j, \Theta^t)$, given the current estimate of the parameter set Θ^t. From Bayes' law this is given by

$$p(i|y_o^j, \Theta^t) = \frac{p_i^t p_i(y_o^j|\theta_i^t)}{p(y_o^j|\Theta^t)} = \frac{p_i^t p_i(y_o^j|\theta_i^t)}{\sum_{i=1}^{2} p_i(y_o^j|\theta_i^t)} \tag{23}$$

- **M-step:** Update the parameters of the model to maximize the likelihood of the data [6].

$$p_i^{t+1} = \frac{1}{N}\sum_{j=1}^{T} p(i|y_o^j, \Theta^t) \tag{24}$$

$$\mu_i^{t+1} = \frac{\sum_{j=1}^{T} y_o^j p(i|y_o^j, \Theta^t)}{\sum_{j=1}^{T} p(i|y_o^j, \Theta^t)} \tag{25}$$

$$\sigma_i^{t+1} = \left(\frac{\sum_{j=1}^{T} p(i|y_o^j, \Theta^t)(y_o^j - \mu_i^{t+1})^2}{\sum_{j=1}^{T} p(i|y_o^j, \Theta^t)}\right)^{\frac{1}{2}} \tag{26}$$

Applying the above two steps iteratively to both y_o^j and x_o^j, allows us to separate the inlying values from the outlying ones. The mean of the inlying values can then be used as an unbiased and robust estimation of the motion parameters.

4. RESULTS AND DISCUSSION

We applied the technique to an extensive set of images, some of which are shown below. Both global sub-pixel registration problem and local motion (disparity) estimation were evaluated. In both cases excellent results were obtained.

For global registration, we used the approach described in [1], 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 2 shows some of the images to which the technique was applied. Results are shown in table 1 and are compared to those reported in [1]. The accuracy was predominantly higher than [1], but also with less required computational time since inverse transforming is not required.

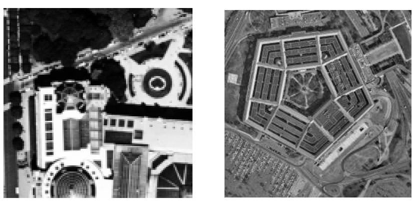


Fig. 2. Some of the images used for simulation.

Image	True Shifts	Foroosh et al. 2001	Proposed Method
Paris	(0.50, -0.50)	(0.48,-0.51)	(0.495,-0.496)
	(0.25, 0.50)	(0.28,0.49)	(0.256,0.499)
	(-0.25, -0.50)	(-0.25,-0.52)	(-0.25,-0.51)
	(0.0,0.75)	(0.0,0.80)	(0.0,0.745)
Pentagon	(0.167, -0.5)	(0.152,-0.49)	(0.16,-0.5)
	(0.67, 0.25)	(0.69,0.33)	(0.68,0.24)
	(-0.33, -0.167)	(-0.32,-0.15)	(-0.34,-0.16)
	(0.33, 0.33)	(0.325,0.32)	(0.333,0.328)

Table 1. Results for global shifts of the images in Figure 2

We also applied the technique to real data 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. We particularly applied the technique to some rectified stereo pairs of aerial and indoor images. For rectified stereo pairs the local motion is along the epipolar lines that are typically warped and mapped to image scan lines. Results are shown in Figures 3 and 4 for the Pentagon image and the baseball image.

In conclusion, this paper shows that accurate results can be obtained for sub-pixel registration directly in the Fourier domain, even when applied to small image regions. Examples of Fourier imaging modalities that can benefit of such approach are magnetic resonance imaging (MRI) [7, 8], and synthetic aperture radar (SAR) [9].

(a) (b) (c)

Fig. 3. (a) & (b) Pentagon's stereo pair, (c) space-frequency representation.

(a) (b) (c)

Fig. 4. (a) & (b) Stereo pair of a baseball, (c) space-frequency representation.

5. REFERENCES

[1] H. Foroosh, J. Zerubia, and M. Berthod, "Extension of phase correlation to sub-pixel registration," *IEEE Trans. Image Processing*, vol. vol. 11, no. 3, pp. pp. 188–200, 2002.

[2] W.S. Hoge, "Subspace identification extension to the phase correlation method," *IEEE Trans. Medical Imaging*, vol. 22, no. 2, pp. 277–280, 2003.

[3] A.H. Zemanian, *Distribution Theory and Transform Analysis*, Dover Publications Inc., 1965.

[4] A. V. Oppenheim, R. W. Schafer, and with J.R. Buck, *Discrete-Time Signal Processing*, Prentice-Hall International Inc., 2nd edition, 1989.

[5] G.H. Golub and M. Heath, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. pp. 195–197, 1979.

[6] G.J. McLachlan, *The EM Algorithm and Extensions*, Wiley, 1997.

[7] E.M. Haacke, R.W. Brown, M.R. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design*, John Wiley and Sons, 1999.

[8] G.A. Wright, "Magnetic resonance imaging," *IEEE Sig. Processing Magazine*, vol. 14, no. 1, pp. 56–66, 1997.

[9] M.I. Skolnik, *Radar Handbook*, McGraw-Hill, 1990.