# UNIFORM FULL-INFORMATION IMAGE MATCHING USING COMPLEX CONJUGATE WAVELET PYRAMIDS

He-Ping Pan
Cooperative Research Centre for Sensor Signal and Information Processing
SPRI Building, Technology Park, Adelaide
The Levels, SA 5095, Australia
Email: heping@cssip.edu.au

Commision III, Working Group 2

**KEY WORDS:** Image Matching, Surface Reconstruction, Algorithms, Automation

## ABSTRACT

Stereo image matching is reconsidered from the viewpoint of full-information exploitation via a uniform transformation of information through scale space. We consider the general stereo situation where both interior and relative orientation of two images are unknown. It is shown that wavelet multiresolution analysis provides an adequate transformation and representation of image signal information with desired properties such as good space-frequency locality and information preservation. In particular, complex conjugate wavelets are used for phase-based matching. Technically, this paper presents a basic procedure for top-down matching two stereo images using complex conjuage wavelet pyramids for the standard case where two images may have a lower bound of stereo overlapping of 60% and relative rotation around principal axis is small. A strategy of spiral parallax propagation is developed for tackling the unknown partial correspondence on the top level. A complete example on matching two real aerial images is shown.

## 1 INTRODUCTION

Image matching may be considered as the central and most difficult problem in photogrammetry and stereo vision for surface reconstruction from multiple images. It has received great attention from many photogrammetrists and computer vision specialists, as well as researchers from pattern recognition and artificial intelligence over last three or more decades. The problem is extremely hard to solve perfectly, partly because the problem domain of image matching in general is not a closed one, partly because of the lack of adequate fundamental mathematical and informatic theories and tools for a thorough understanding of the information-processing mechanism throughout the image matching process.

Due to the length limit, this paper does not give a comprehensive overview on the related literature of general image matching and wavelets. Briefly, existing approaches for stereo image matching may be classified into several clusters according to the choice of matching primitives, matching criterion and strategies as follows.

### Signal Correlation

The most obvious approach to stereo image matching is to correlate two image functions over each pair of local areas. It is thus often called image correlation, or area-based matching, etc. This is perhaps the earliest approach, and obviously an engineering solution. (Heleva, 1976; Ackerman, 1984)

### Feature Matching

Feature matching was introduced naturally to overcome the inabilities of area-based signal correlation by attempting matching only on information-rich points or more complicated primitives such as edges, regions, etc. It was inspired by the studies on biological vision (Grimson, 1981; Förstner, 1986)

### Global Matching

Instead of matching local areas or features separately, the approach of global matching attempts to match all pairs of homologous image points or features within a simultaneous framework, typically via least-squares adjustment or other relaxation procedures (Grün, 1985; Poggio et al, 1985; Rosenholm, 1987; Rauhala, 1987; Barnard 1989; Zhang et al 1992).

### Object-Space Image Matching

Object-space image matching, so-called typically by photogrammetrists, assumes a coherent facet model of the scene surfaces *a priori*. This is largely true for the terrain viewed from a relatively high altitude in aerial photography (Ebner et al, 1987; Wrobel, 1987; Helava, 1988, Heipke, 1992).

### Image-Domain Approach Revisited

The approach that we are proposing here, briefly called *uniform full-information image matching*, may be considered as a natural development of the three image-domain approaches (signal correlation, feature matching, and global matching). We rely on exploiting the full information beared in the image signals. We require the representation of image signal information to be uniform through scale space. We do not distinguish explicit features such as points, edges/lines, regions, textures, shading, etc; instead, we use full-information representation which may be considered as implicit feature vectors. In particular, we use wavelet multiresolution analysis (wavelet pyramid) as information representation of image signals for image matching. We also use general effective matching strategies inspired by biological vision. Large continuity and minor discontinuity of parallax field is also considered in practical algorithms.

## 2 UNIFORM FULL-INFORMATION IMAGE MATCHING

The notion of *uniform full-information image matching* may be best described briefly as follows. A digital image is a function $f(x,y)$ on a 2-dimensional support. For image matching or in general, pattern recognition, a representation of $f(x,y)$ is to be chosen in such a way so that the constructs in the new representation may be related to salient information of the original signal function $f(x,y)$. In general, let us assume $f(x,y)$ is to be represented by a vector of projections of $f(x,y)$ onto $n$ basis functions $\psi_j(x,y)$

$$f(x,y) \longrightarrow (a_1, a_2, \ldots, a_n) \quad (1)$$

$$a_j = < f(x,y), \psi_j(x,y) >, \; j = 1,2,\ldots,n \quad (2)$$

where $< \cdot, \cdot >$ denotes the inner product of two functions, $a_j$'s are called the representation coefficients.

By appropriately choosing the basis function $\psi_j$'s, we intend to extract salient information of $f(x,y)$ in the form of the coefficients $a_j$'s. A representation of the form (1) is said to be a *full-information* representation if with $a_j$'s computed via (2), the equation

$$f(x,y) = F(a_1, a_2, \ldots, a_n; \psi_1, \psi_2, \ldots, \psi_n) \qquad (3)$$

holds, where $F()$ is a computable function. An example of (3) is that when $\psi_j$'s constitute an orthonormal basis, we have a simple reconstruction procedure

$$f(x,y) = \sum_{j=1}^{n} a_j \, \psi_j(x,y) \qquad (4)$$

A representation of the form (1) and (3) is said to be *uniform* because each representation coefficient $a_j$ is defined and computed with exactly the same simple mathematical form of (2). With some contrast to previous image-domain approaches, we do not scrutinize on the explicit interpretation of the representation coefficients, which may highly nonlinearly relate to intensity differentials, textures, shading, surface reflectance variations, etc.

For image matching purpose, it is desirable if the representation of the form (3) has properties of good dimensional orthogonality, discriminative uniqueness, space-frequency locality, multiresolution adaptivity, and computational efficiency and robustness. For the particular problem of stereo matching, we may also require the information representation and matching strategies to be invariant to translation, rotation, scale and partial correspondence between two stereo images.

Fourier analysis is a classical example of the uniform and full-information representation, which, however, is known to be very poor in spatial locality. Wavelet analysis is a new approach in this sense, as fundamental as Fourier analysis, but with inherently good locality in both spatial and frequency (scale) domain and a number of other desired properties.

## 3 WAVELETS AND COMPLEX WAVELETS

In this section, we briefly draw the essentials of wavelet theory and lay the mathematical foundation of this image matching approach.

### 3.1 Wavelet Transform and Wavelet Pyramid

The wavelet transform is a relatively recent development in mathematics and signal processing (Grossmann and Morlet, 1984; Mallat, 1989), as a signal decomposition approach to overcome the shortcomings of the window Fourier transform. This decomposition is to project the signal $f(x)$ onto a family of functions which are the dilations and translations of a unique function $\psi(x)$. The function $\psi(x)$ is called a wavelet and the corresponding wavelet family is given by

$$\psi_{s,t}(x) = \sqrt{s}\psi(s(x-t)), (s,t) \in \mathbf{R}^2 \qquad (5)$$

where $\mathbf{R}$ denote the set of real numbers, $s$ and $t$ are called the scale and translation respectively. Let $L^2(\mathbf{R})$ denote the vector space of measurable, square-integrable one-dimensional functions $f(x)$. The wavelet transform of a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ with a given wavelet $\psi$ is defined by

$$Wf(s,t) = \int_{-\infty}^{+\infty} f(x)\psi_{s,t}(x)dx =< f(x), \psi_{s,t}(x) > \qquad (6)$$

A wavelet transform as defined in (6) can be interpreted as a decomposition of a signal $f(x)$ into a set of frequency channels having the same bandwidth on a logarithmic scale. Under a certain condition, $f(x)$ can be reconstructed from its wavelet transforms $Wf(s,t)$. The adaptivity to scale $s$ and translation $t$ leads to its good locality in both frequency and spatial domain, a property desired by image matching algorithms.

For a special class of functions, the redundancy of the continuous wavelet transform (6) can be cleared by discretizing both the scale factor $s$ and the translation $t$,

$$s = \frac{1}{2^j} \text{ and } t = k, \text{ with } (j,k) \in \mathbf{Z}^2 \qquad (7)$$

where $\mathbf{Z}$ denotes the set of integers, the wavelet family

$$\psi_{j,k}(x) = \frac{1}{\sqrt{2^j}}\psi(\frac{x-k}{2^j}), \qquad (j,k) \in \mathbf{Z}^2 \qquad (8)$$

is called dyadic discrete wavelets.

In the dyadic scale space of the form (8), let $A_j f$ denote the approaximation of a given function $f(x)$ at a scale $s = \frac{1}{2^j}$. In practice, we only consider a limited number of levels $j = 0, 1, 2, \ldots, n$, for some $n$ chosen to be the coarsest level, corresponding to the smallest scale $s = \frac{1}{2^n}$. Let $D_j f$ denote the difference between two approximations $A_{j-1}f$ and $A_j f$, i.e.

$$D_j f = A_{j-1}f - A_j f, \qquad j = 1, 2, \ldots, n \qquad (9)$$

where $A_0$ is an identity operator. The function $f(x)$ can be decomposed as

$$\begin{aligned} f(x) &= A_1 f + D_1 f \\ &= A_2 f + D_2 f + D_1 f \\ &\vdots \\ &= A_n f + \sum_{k=1}^{n} D_k f \end{aligned} \qquad (10)$$

It was proved that a multiresolution analysis can be realized by a scaling function $\phi(x)$ and its associated wavelet function $\psi(x)$,

$$A_j f(x) = \sum_{k=-\infty}^{+\infty} < f(u), \phi_{j,k}(u) > \phi_{j,k}(x) \qquad (11)$$

$$D_j f(x) = \sum_{k=-\infty}^{+\infty} < f(u), \psi_{j,k}(u) > \psi_{j,k}(x) \qquad (12)$$

where

$$\phi_{j,k}(x) = \frac{1}{\sqrt{2^j}}\phi(\frac{x-k}{2^j}), \qquad (j,k) \in \mathbf{Z}^2 \qquad (13)$$

and $\psi_{j,k}(x)$ has the form of (8), which relates to $\phi_{j,k}(x)$ by.

$$\hat{\psi}(2\omega) = e^{-i\omega}\overline{H(e^{-i\omega})}\hat{\phi}(\omega) \qquad (14)$$

where $\hat{\phi}$ denotes the Fourier transform of the function $\phi$, $H$ is the transfer function of $\phi$, $\overline{H}$ denotes the complex conjugate of $H$.

For a 2-D image function $f(x,y)$, a multiresolution analysis can be written as

$$\begin{aligned} f(x,y) &= A_1 f + D_{1,1} f + D_{1,2} f + D_{1,3} f \\ &= A_2 f + D_{2,1} f + D_{2,2} f + D_{2,3} f \\ &\quad + D_{1,1} f + D_{1,2} f + D_{1,3} f \\ &\vdots \\ &= A_n f + \sum_{j=1}^{n} [D_{j,1} f + D_{j,2} f + D_{j,3} f] \end{aligned} \quad (15)$$

Each approximation $A_j f(x,y)$ and difference component $D_{j,p} f(x,y)$ can be fully characterized with a 2-D scaling function $\Phi(x,y)$ and its associated wavelet functions $\Psi_p(x,y)$, $p = 1,2,3$,

$$A_j f(x,y) = \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} a_{j,k,l} \Phi_{j,k,l}(x,y) \quad (16)$$

$$D_{j,p} f(x,y) = \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} d_{j,p,k,l} \Psi_{j,p,k,l}(x,y) \quad (17)$$

where

$$\Phi_{j,k,l}(x,y) = \frac{1}{2^j} \Phi(\frac{x-k}{2^j}, \frac{y-l}{2^j}),\ (j,k,l) \in \mathbf{Z}^3 \quad (18)$$

$$\Psi_{j,p,k,l}(x,y) = \frac{1}{2^j} \Psi_p(\frac{x-k}{2^j}, \frac{y-l}{2^j}) \quad (19)$$

$$a_{j,k,l} = < f(x,y), \Phi_{j,k,l}(x,y) > \quad (20)$$

$$d_{j,p,k,l} = < f(x,y), \Psi_{j,p,k,l}(x,y) > \quad (21)$$

For a separable multiresolution analysis, the scaling function $\Phi(x,y)$ and wavelet functions $\Psi_p(x,y), p = 1,2,3$ can be written as

$$\Phi(x,y) = \phi(x)\phi(y) \quad (22)$$

$$\Psi_1(x,y) = \phi(x)\psi(y) \quad (23)$$

$$\Psi_2(x,y) = \psi(x)\phi(y) \quad (24)$$

$$\Psi_3(x,y) = \psi(x)\psi(y) \quad (25)$$

where $\phi$ is a one-dimensional scaling function, $\psi$ is the 1-D wavelet function associated with $\phi$. Apparently $\Psi_1$, $\Psi_2$, $\Psi_3$ extract the details of the 2-D image function $f(x,y)$ in the y-axis, x-axis and diagonal directions respectively.

The representation (15) may be vividly called the wavelet pyramid of an image $f(x,y)$. Given a discrete image $f(x,y)$ with a limited support $x = 1,2,\ldots,m_x$, $y = 1,2,\ldots,m_y$, the actual procedure for constructing this pyramid involves computing the coefficients $a_{j,k,l}$, $d_{j,p,k,l}$, which can be grouped into four matrices $A_j$, $D_{j,p}$, $p = 1,2,3$, on each level $j$

$$A_j = (a_{j,k,l})_{(k=1,2,\ldots,\frac{m_x}{2^j}; l=1,2,\ldots,\frac{m_y}{2^j})} \quad (26)$$

$$D_{j,p} = (d_{j,p,k,l})_{(k=1,2,\ldots,\frac{m_x}{2^j}; l=1,2,\ldots,\frac{m_y}{2^j})} \quad (27)$$

Let $h$ and $g$ be the impulse response of the filter $\phi$ and $\psi$, the coefficients $a_{j,k,l}$, and $d_{j,p,k,l}$, $p = 1,2,3$, can be computed via an iterative procedure. The wavelet pyramid of image $f(x,y)$ and its constructing process are illustrated in Fig.1. ($x|2$ means dyadic subsampling).

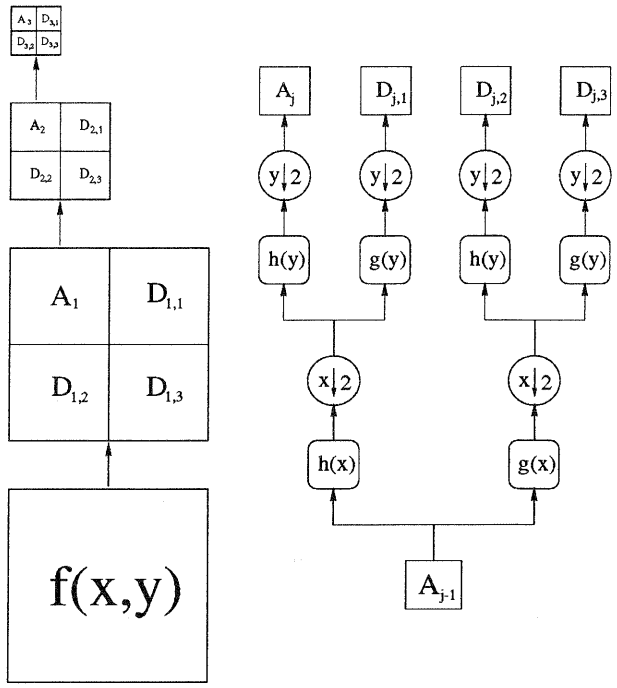


Figure 1: Wavelet pyramid of an image $f(x,y)$ (left) and the flowchart for the analysis from level $j-1$ to level $j$ (right)

### 3.2 Complex Wavelets for Phase-Based Matching

Wavelet pyramid is ideal for scale adaptive image matching as it has the advantage at the good locality in both spatial and frequency domain. However, wavelet pyramid of an image is neither translation-invariant, nor rotation-invariant. At this stage, let us concentrate on the translation-invariance problem, while assuming that either the rotation angle $\gamma$ of the matched image about its principal axis relative to the reference image is small enough, or two stereo images have been resampled along the epipolar lines.

The wavelet pyramid of real-valued wavelets is not translation-invariant implies that the phase information is not readily represented. In order to explore the phase information in the image signals and still on multiscales, complex-valued wavelets are a suitable representation as the translation in the spatial domain is represented as a rotation in the complex phase domain. This gives rise to the interpolability of the wavelet transform, yielding the possibility of subpixel matching through the multilevels of the wavelet pyramids.

The complex wavelets used in this work were designed by Margarey and Kingsbury (1995), first used for motion estimation of video frames. Bergeaud and Mallat (1984) proposed similar complex wavelets. It should be pointed out that the similarity distance measures and various matching strategies to be described in the following sections are not limited to those particular wavelets used in this work, they may rather be generally applicable with other well-designed wavelets with good properties.

For general image matching purpose, we require the wavelet filter pair $(h,g)$ (impulse response of the scaling and wavelet function $\phi$ and $\psi$) to be compactly supported in spatial domain, regular (differentiable up to a high order), symmetric

(leading to linear phase). Orthogonality mostly cannot be strictly emphasized in practice.

For the particular complex wavelets used in this work, the impluse reponses $g$ of the wavelet function and $h$ of the scaling function are a complex pair of even-length modulated windows.

$$g(k) = b_1 w_1(k+0.5) e^{i\omega_1(k+0.5)} \quad (28)$$

$$h(k) = \tilde{b}_1 \tilde{w}_1(k+0.5) e^{i\tilde{\omega}_1(k+0.5)} \quad (29)$$

$$(k = -n_w, -n_w+1, \ldots, n_w-1)$$

where $b_1$ and $\tilde{b}_1$ are complex constants, $w_1$ and $\tilde{w}_1$ are a pair of real-valued low-pass windows of width $2n_w$, symmetric about $k=0$ and decaying to zero at each end. A commonly used one of this type is Gaussian

$$w_1(k) = \exp\left(-\frac{k^2}{2\sigma_1^2}\right), \quad \tilde{w}_1(k) = \exp\left(-\frac{k^2}{2\tilde{\sigma}_1^2}\right) \quad (30)$$

Due to the compromise between the good locality of matching and information sufficiency, the minimum width of window functions should be 4, thus $n_w = 2$. The modulation frequencies $\omega_1$ and $\tilde{\omega}_1$ should be complementary

$$\omega_1 + \tilde{\omega}_1 = \pi \quad (31)$$

in order to cover the frequency range $[0, \pi]$. Because $\phi$ and $\psi$ are a pair of low- and high-pass filters, we have $\omega_1 > \tilde{\omega}_1$. With the Gaussian window functions defined in (30), the Fourier transforms of $g$ and $h$ have conjugate symmetry about the modulation frequencies $\omega_1$ and $\tilde{\omega}_1$. Since real 1-D signals have conjugate symmetric spectra, the neglect of the negative half spectrum $[-\pi, 0]$ does not exclude any significant information about a real 1-D input. Ideally, a maximum coverage on the frequency range $[0, \pi]$ without significant gaps and with minimal overlap can be effective achieved if on each level $j$,

$$\omega_j = 3\,\tilde{\omega}_j \quad (32)$$

thus, by (31), we have the modulation frequencies on the first level (bottom-up)

$$\omega_1 = \frac{5\pi}{6}, \qquad \tilde{\omega}_1 = \frac{\pi}{6} \quad (33)$$

Ideally, the modulation frequencies are to be decomposed through levels

$$\omega_j = \frac{\omega_{j-1}}{2} \quad (34)$$

In practice, if Gaussian windows of (30) are used, the frequency decomposition through levels approximates asymptotically to (34).

Using the 1-D complex wavelet and scaling filters defined by formulas (28)-(29), we can implement the 2-D complex wavelet analysis in the same separable way as described in section (3.1). The 2-D wavelet filters so formed will be predominantly first quadrant filters in the frequency domain. As real discrete images contain significant information in the first and second quadrant of the unit frequency cell, we need to use the complex conjugate filters $\bar{g}$ and $\bar{h}$ in addition to $g$ and $h$ in order to produce a mirror set of difference coefficient matrices $\tilde{D}_{j,p}$, $p = 1, 2, 3$, for each $j$-th level, containing the second quadrant information.

The algorithm for the complex conjugate wavelet analysis of an image is similar to the one shown in Fig.1. The wavelet analysis from level $j-1$ to level $j$ correspond to transforming two complex approximation submatrices to eight complex approximation and difference submatrices

$$\{A_{j-1}, \tilde{A}_{j-1}\} \longmapsto \{A_j, \tilde{A}_j, D_{j,p}, \tilde{D}_{j,p}, p = 1, 2, 3\} \quad (35)$$

where $\tilde{A}_j$ is the mirror of $A_j$, and $\tilde{D}_{j,p}$ is the mirror of $D_{j,p}$. The algorithm is illustrated in Fig.2 - 3.

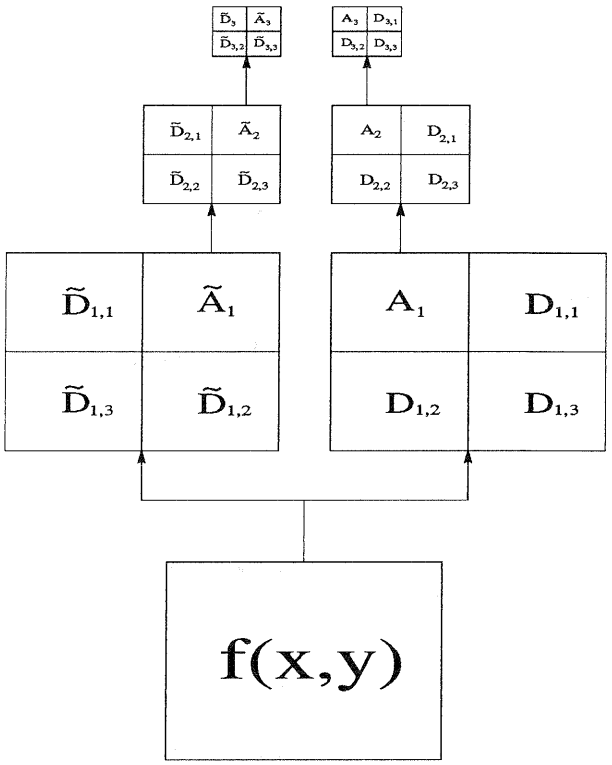


Figure 2: Complex conjugate wavelet pyramid of an image $f(x, y)$

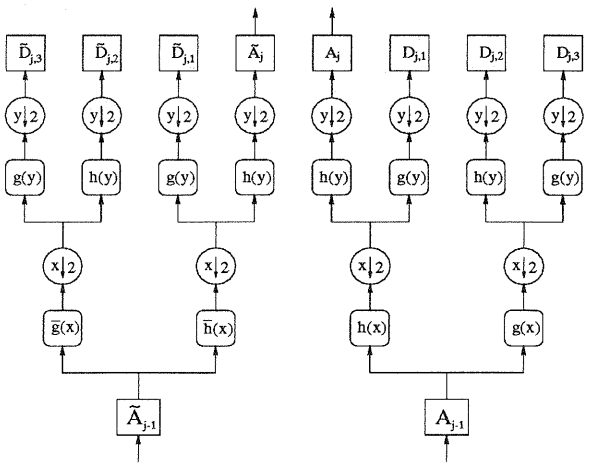


Figure 3: Flowchart of complex conjugate wavelet analysis

## 4 SIMILARITY DISTANCE AND CONTINUOUS MATCHING

Using the complex conjugate wavelet analysis, we can define the similarity distance $S_j((x,y),(x',y'))$ for any pair of image points on the reference image (e.g. left) $f(x,y)$ and the matched image (e.g. right) $f'(x',y')$. We shall use $z'$ to denote any thing on the matched image corresponding to $z$ on the reference image.

### 4.1 Implicit Feature Vectors

On a given $j$-th level of the wavelet pyramid, for a given position $(x,y)$, we have 8 coefficients of complex conjugate wavelet analysis $(A_j(x,y), \tilde{A}_j(x,y), D_{j,p}(x,y), \tilde{D}_{j,p}(x,y), \quad p = 1,2,3)$. Note that $A_j$ and $\tilde{A}_j$ are approximation components whose information are further decomposed onto the next higher level $j+1$. In order to match two images on the $j$-th level, the similarity distance can be defined using only 6 differential components $D_{j,p}(x,y), \tilde{D}_{j,p}(x,y), p = 1,2,3$. For image matching to be invariant to local image intensity, a normalization proved, with real image data, to be adequate, which leads to the implicit feature vector $B_j(x,y)$for each position $(x,y)$

$$\mathbf{B}_j(x,y) = (\frac{D_{j,1}(x,y)}{|A_j(x,y)|}, \frac{D_{j,2}(x,y)}{|A_j(x,y)|}, \frac{D_{j,3}(x,y)}{|A_j(x,y)|}, \frac{\tilde{D}_{j,1}(x,y)}{|\tilde{A}_j(x,y)|}, \frac{\tilde{D}_{j,2}(x,y)}{|\tilde{A}_j(x,y)|}, \frac{\tilde{D}_{j,3}(x,y)}{|\tilde{A}_j(x,y)|}) \quad (36)$$

where $|.|$ denotes the module of a complex number. The components $B_{j,p}, p = 1,2,\ldots,6$ are also called subbands of wavelet analysis.

### 4.2 Standard Similarity Distance

In the standard case where the relative rotation angle $\gamma$ is small enough, a similarity distance $S_j((x,y),(x',y'))$ can be defined as

$$S_j((x,y),(x',y')) = \sum_{p=1}^{6} S_{j,p}((x,y),(x',y')) \quad (37)$$

where $S_{j,p}$'s are the subband similarity distances, defined by

$$S_{j,p}((x,y),(x',y')) = |B_{j,p}(x,y) - B'_{j,p}(x',y')|^2 \quad (38)$$

In a top-down hierarchical matching scheme, for an image point $(x = k, y = l)$, $(k,l) \in Z^2$, on the reference image, we may know its approximate correspondence $(x' \approx k', y' \approx l')$ on the matched image. The precise correspondence may be somewhere $(x',y')$ around $(k',l')$:

$$f(k,l) \mapsto f'(k'+u, l'+v) \quad (39)$$

where $(u,v) \in R^2$, denoting the differences

$$u = x' - k', \qquad v = y' - l' \quad (40)$$

Note $(k,l)$ and $(k',l')$ take integer coordinates in the down-sampled image on the level $j$. With this approximation, the subband similarity distance is reformed to

$$S_{j,p}((k,l),(k'+u,l'+v)) = |B_{j,p}(k,l) - B'_{j,p}(k'+u,l'+v)|^2 \quad (41)$$

The best matching point in the simplest sense corresponds to

$$\min_{u,v} S_j((k,l),(k'+u,l'+v)) \quad (42)$$

### 4.3 Continuous Interpolation for Fine Matching

In order to minimize the similarity distance $S_j((k,l),(k'+u,l'+v))$ in (42), we need to interpolate $B'_{j,p}(k'+u,l'+v)$. Ideally, we should use four integer-indexed positions surrounding $(k'+u,l'+v)$, which leads to complicated interpolation and minimization procedures. If we limit $|u|,|v| \leq 0.5$, we can use single nearest neighbor $(k',l')$ with a phase shift relative to the spatial shift $(u,v)$,

$$B'_{j,p}(k'+u,l'+v) \approx B'_{j,p}(k',l')e^{i2^j \mathbf{\Omega}_{j,p}\cdot(u,v)} \quad (43)$$

where $\mathbf{\Omega}_{j,p}$ denotes the pair of the modulation frequencies (in x- and y-directions) for $p$-th subband $D'_{j,p}$ on the $j$-th level, given by (for $p = 1,2,\ldots,6$)

$$(\tilde{\omega}_j,\omega_j),(\omega_j,\tilde{\omega}_j),(\omega_j,\omega_j),(-\tilde{\omega}_j,\omega_j),(-\omega_j,\tilde{\omega}_j),(-\omega_j,\omega_j)$$

where $\omega_j$, $\tilde{\omega}_j$ are given inductively by (32) - (34).

It can be shown that with the continuous interpolation of (43), the similarity distance can be written as

$$S_j((k,l),(k'+u,l'+v)) = s_1(u-u_0)^2 + s_2(v-v_0)^2 + s_3(u-u_0)(v-v_0) + s_4 \quad (44)$$

where the coefficients $s_1, s_2, s_3, s_4, u_0, v_0$ can be computed directly from given data $B_{j,p}(k,l)$, $B'_{j,p}(k',l')$, $p = 1,2,\ldots,6$, and $(\omega_j,\tilde{\omega}_j)$. $(u_0,v_0)$ is the minimum point of the similarity distance surface $S_j$. $s_1$, $s_2$, $s_3$ are the curvature (second derivatives) along $x$-, $y$-, and diagonal directions respectively, characterizing the uncertainty of the estimate $(u_0,v_0)$. $s_4$ is the minimum value of the similarity distance $S_j$.

### 4.4 Local Parallax Continuity and Generic Pattern Matching

For the robustness of image matching, local parallax continuity should be taken into account when matching two given positions $(k,l)$ and $(k',l')$. To maintain a compromise between fine locality of matching and robustness of matching, we could define a new similarity distance $P_j((k,l),(k',l'))$ in the sense of generic pattern matching

$$P_j((k,l),(k',l')) = \sum_{(r,c)} \min_{u,v} S_j((k+r,l+c),(k'+r+u,l'+c+v)) \quad (45)$$

$$(r,c) \in [(0,0),(-0.5,-0.5),(-0.5,.5),(0.5,-0.5),(0.5,0.5)]$$

where $(u,v)$ is a fine-tuning shift vector, which is variational for each $(r,c)$ pair. Note that $(k+r,l+c)$ with $r,c = 0.5$ or $-0.5$ correspond to diagonal positions which can also be computed with rigorous bottom-up wavelet transform. Similarly, a normal position $(k,l)$ may be matched to a normal $(k',l')$ or a diagonal position $(k'+r,l'+c)$.

## 5 SPIRAL AND HIERARCHICAL PARALLAX PROPAGATION

### 5.1 Spiral Parallax Propagation

Without loss of generality, let us consider a stereo pair of square images with size $2^n \times 2^n$. With minimum overlapping of 60%, the central area (e.g. $2 \times 2$ on the level of $8 \times 8$) on a chosen highest level of each image should have correspondence on the other image. Image matching thus can start with an exhaustive search for the best matching of

the central area of the reference image to the matched image (or vice versa) using the similarity distance defined by (45), which yields an approximate parallax vector for this central area. Parallax vector for each integer-indexed position of this central area can then be fine-tuned also by using (45). Known parallax vectors can then be propagated from the central area to the outer rings, ring by ring, until the boundary of partial correspondence is reached, and this can be done fully automatically. Gross errors of the resultant parallax field can be detected and corrected automatically by using the local continuity constraint.

### 5.2 Hierarchical Parallax Propagation

After image matching on a higher $(j+1)$-th level, the parallax field should then be propagated to the next lower (finer) $j$-th level. The initial parallax field on the current $j$-th level can be obtained by interpolating the parallax field at the $(j+1)$-th level. The inverse of the similarity distance of (45) for each position on the higher level may be taken as the weighting factor for linear or nonlinear interpolation.

After matching through intermediate levels, a number of homologous matched point pairs can then be selected automatically, the focal length of each image and five relative orientation parameters can then be solved via a direct closed-form solution [Pan et al, 1995] from these pure image coordinates. Note that the standard aerial stereo pairs with closely parallel principal axes correspond to a degeneracy of that direct solution. The solution of two focal lengths is sensitive, though indeed solvable. For robotic stereo images with an essential vergence angle, the solution is robust enough.

### 5.3 An Example of Real Aerial Images

The complete procedure consists of complex wavelet transform, spiral matching on the top level, and hierarchical matching through intermediate levels, solving the two focal lengths and relative orientation, up to surface reconstruction and visualization. This procedure has been implemented and tested with real images. Fig.4-6 show an example of matching a pair of real aerial images. Through visual checking of each matched position pairs, no gross error is found. As we only match regular points to normal and diagonal regular points, the matching errors bound to 0.5 pixel on each level. This wavelet-based approach can in principle reach a resolution of $2 \times 2$. We shall leave the final pixel-level or subpixel-level matching to least-square global matching, to which, wavelet features may still be useful.

## 6 CONCLUSIONS

This paper presents a basic theory of uniform full-informatin image matching using complex conjugate wavelet pyramids. The basic procedure including the bottom-up wavelet multiresolution analysis and top-down image matching has been implemented and tested with real images. The result is promising. The feasibility of this approach is confirmed. Rotation-invariant image matching is not discussed here due to the length limit, which is the main focus of our current research.

## REFERENCES


- Ackermann F. (1984): Digital image correlation: performance and potential application in photogrammetry. Photogrammetric Record, II(64): 429-439.
- Barnard S.T. (1989): Stochastic stereo matching over scale. IJCV 3:17-32.
- Ebner H., Fritsch D., Gillessen W., and Heipke C. (1987): Integration von Bildzuordnung und Objectrekonstruktion innerhalb der digitalen Photogrammetrie. Bildmessung und Luftbildwesen 55(5):194-203.
- Förstner W. (1986): A feature based correspondence algorithm for image matching. IAPRS 26, Part 3/3, pp. 150-166.
- Grimson W.E.L. (1981): From Images to Surfaces: A computational study of the human early visual system. MIT Press, Cambridge, MA.
- Grossmann A. and Morlet J. (1984): Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM J. Math., 15:723-736.
- Grün, A. (1985): Adaptive least squares correlation: A powerful image matching technique. South African Journal of Photogrammetry, Remote Sensing and Cartography, 14(3):175-219.
- Heipke, C. (1992): A global approach for least-squares image matching and surface reconstruction in object space. PE&RS 58(3):317-323.
- Helava, U.V. (1976): Digital correlation in photogrammetric instruments. IAPRS, Congress Helsinki, Com. II, Vol. 21.
- Helava, U.V. (1999): Object-space least-squares correlation. PE&RS, 54(6): 711-714.
- Magarey J. and Kingsbury N. (1995): Motion estimatin using complex wavelets. Technical Report, Department of Engineering, Cambridge University, U.K.
- Mallat S.G. (1989): A theory for multiresolution signal decomposition: the wavelet representation. IEEE-PAMI 11(7):674-693.
- Pan H.P., Brooks M.J., Newsam G.N. (1995): Image resituation: initial theory. SPIE Proceedings 2598, pp.162-173.
- Pan H.P., Huynh D.Q., Hamlyn G.K. (1995): Two-image resituation: practical algorithms. SPIE Proceedings 2598, pp.174-190.
- Poggio, T., Torre V., and Koch C. (1985): Computational vision and regularization theory. Nature 317(26):314-319.
- Rauhala (1987): Fast compiler positioning algorithms and techniques of array algebra in analytical and digital photogrammetry. ISPRS Proc. Fast Processing of Photogrammetric Data, Interlaken.
- Rosenholm (1987): Multi-point matching using the least squares technique for evaluation of the three-dimensional models. PE&RS 53(6):621-626.
- Wrobel B. (1987): Facets Stereo Vision (FAST Vision) - A new approach to comptuer stereo vision and to digital photogrammetry. ISPRS Proc. Fast Processing of Photogrammetric Data, Interlaken.
- Wrobel B. (1991): Least-squares methods for surface reconstruction from images. IJPRS 46:67-84.
- Zhang Z.X, Zhang J.Q., Wu X.L. and Zhang H. (1992): Global image matching with relaxation method. Proc. International Colloquium on Photogrammetry, Remote Sensing and Geographic Information System, Wuhan, China.

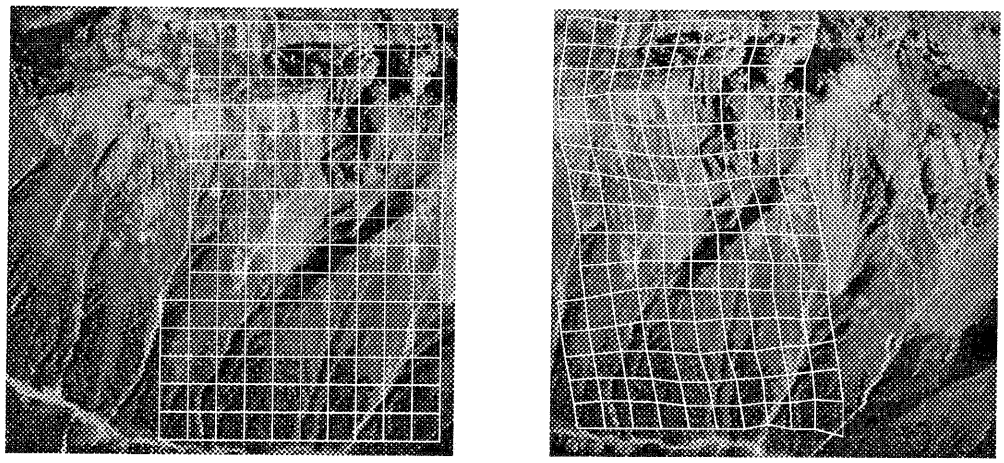

Figure 4: Matched positions on left and right images of level $16 \times 16$

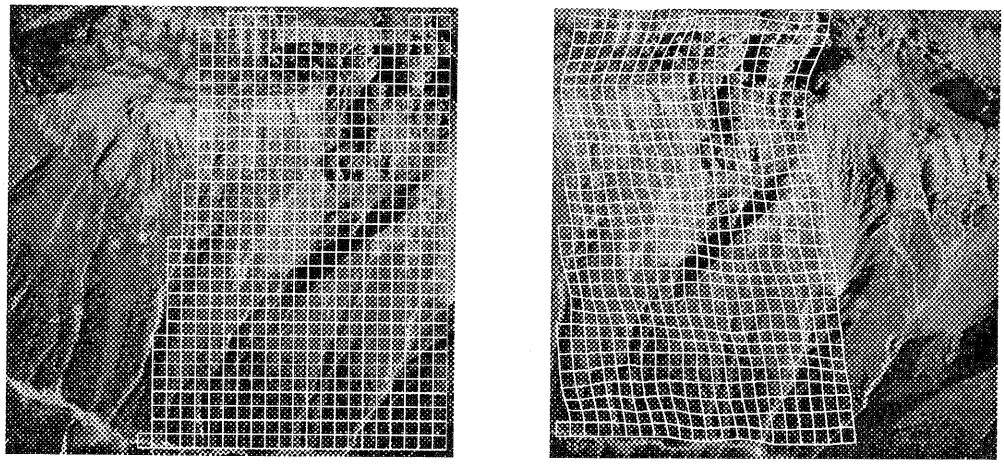

Figure 5: Matched positions on left and right images of level $32 \times 32$

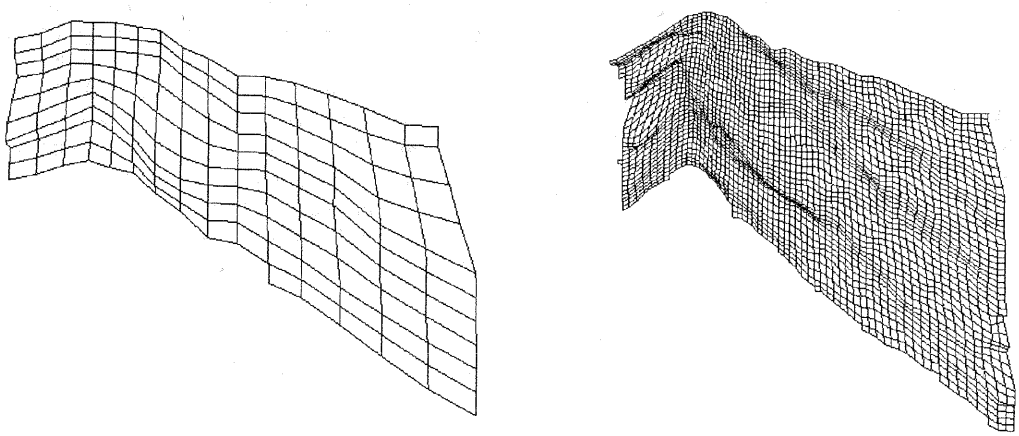

Figure 6: Reconstructed surfaces on level $16 \times 16$ (left), and $64 \times 64$ (right)

Surfaces are viewed from left to right relative to the original images.

The author would like to thank G.N. Newsam, N. Kingsbury, and J. Magarey for helpful discussions, and Prof. M.J. Brooks for the support to this work.