Can. J. Remote Sensing, Vol. 33, No. 6, pp. 541–550, 2007

A self-adaptive motion estimation algorithm for superresolution reconstruction of multiframe SPOT panchromatic images

Bing Tan, Qing Xu, Jonathan Li, and Yan Zhang

Abstract. High spatial resolution of imaging sensors onboard Earth observation satellites is often critical for land use and land cover mapping. In this paper, we propose a self-adaptive motion estimation algorithm that combines the consistency characteristics of image distortion, high accuracy feature point extraction, and image matching techniques to achieve subpixel accuracy of motion estimation. Using the wavelet superresolution and the modulation transfer function (MTF), we implemented the superresolution reconstruction from multiframe Satellite pour l'Observation de la Terre (SPOT-4) panchromatic images under a framework of interpolation–restoration. The results show that this algorithm is promising for reconstruction of a higher resolution frame.

Résumé. La haute résolution spatiale des capteurs imageurs à bord des satellites d'observation de la terre est souvent critique dans la cartographie de l'utilisation du sol et du couvert. Dans cet article, nous proposons un algorithme auto-adaptatif d'estimation du mouvement combinant les caractéristiques de cohérence de la distorsion d'images, l'extraction haute précision des points caractéristiques et les techniques d'appariement d'images afin d'assurer la précision au niveau du sous-pixel dans l'estimation du mouvement. Utilisant la superrésolution par transformée en ondelettes et la fonction de transfert par modulation (FTM), nous avons réalisé une reconstruction superrésolution à partir d'images panchromatiques multiples de Satellite pour l'Observation de la Terre (SPOT-4) dans un contexte d'interpolation-restauration. Les résultats montrent que cet algorithme offre un bon potentiel pour la reconstruction d'une image à plus haute résolution.
[Traduit par la Rédaction]

Introduction

Land use and land cover mapping of urban–suburban areas is one of many applications where high spatial resolution satellite images are useful and often critical. However, most of such tasks done at the city or regional level have been utilizing low to medium spatial resolution satellite images, such as Landsat thematic mapper (TM) or enhanced thematic mapper plus (ETM+) (15 m panchromatic, Pan and 30 m multispectral (MS)) and Satellite pour l'Observation de la Terre (SPOT-4; 10 m Pan and 20 m MS) imagery, as the major data sources in terms of their large ground coverage per image frame (e.g., 185 km × 185 km for a Landsat-4/5 TM or Landsat-7 ETM+ frame and 60 km × 60 km for a SPOT-4 frame) and low cost over the past two decades. An approach to obtain high spatial resolution is to use commercial satellite images such as 1 m resolution IKONOS and 0.6 m resolution QuickBird imagery or aerial photographs. Although this approach is feasible, covering a large area for urban and regional planning using such imagery may be very costly (e.g., US$20 per square kilometre for a Standard Ortho IKONOS image frame covering 11 km × 11 km on the Earth's surface). An alternative approach is to reconstruct a high-resolution image from a sequence of low-resolution images using digital image processing techniques. This is possible if there is a subpixel motion between the acquired frames, for example, SPOT images.

In the digital image processing community, a number of algorithms for the resolution enhancement of video images have appeared in the literature (Gunturk et al., 2002; Altunbasak et al., 2002; Mateos et al., 2000). Among these algorithms, the focus of this paper is on image-superresolution solutions. Image superresolution refers to image processing algorithms that produce high-quality, high-resolution (HR) images from a set of low-quality, low-resolution (LR) images (Ur and Gross, 1992). Lukosz (1966) first realized the possibility of image superresolution, and Tsai and Huang (1984) first presented the seminal work on image superresolution. From the viewpoint of image sampling, LR

Received 4 February 2006. Accepted 3 August 2007. Published on the *Canadian Journal of Remote Sensing* Web site at http://pubs.nrc-cnrc.gc.ca/cjrs on 15 February 2008.

B. Tan,[1] Q. Xu, and Y. Zhang. Department of Remote Sensing, Institute of Surveying and Mapping, University of Information Engineering, Zhengzhou, Henan 450052, China.

J. Li. Department of Geography, University of Waterloo, Waterloo, ON N2L 3G1, Canada

[1]Corresponding author (e-mail: tanbing@133sh.com).

image data are nonuniformly sampled when considering all the LR frames together on an HR grid, although each frame can be sampled on a rectangular grid. Thus the image reconstruction process can be performed under an interpolation–restoration framework. Sauer and Allebach (1987) were the first researchers to consider superresolution as an interpolation method for non-uniformly sampled data. They used a projection onto convex sets (POCS) algorithm to reconstruct the unknown values. Aizawa et al. (1991) also modeled superresolution as an interpolation problem with nonuniform sampling and used a formula related to the Shannon sampling theorem to estimate values on an HR grid. However, these two methods ignored the atmospheric blurring effect on the remote sensor. Tekalp et al. (1992) extended these algorithms to include blurring effects and sensor noise and proposed an additional restoration step. Ur and Gross (1992) considered the Papoulis generalized multichannel sampling theorem for interpolating values on an HR grid. The wavelet-based interpolation method for processing nonuniformly sampled LR data has only been considered in recent years. Ford and Etter (1998) performed a reconstruction of nonuniformly sampled one-dimensional (1D) signals using a wavelet basis in a multiresolutional setting. Nguyen (2000) then expanded this method to a two-dimensional (2D) case in a multiresolution framework. Lertrattanapanich (2003) contributed further to those efforts by strengthening the structural foundation to support the use of multiresolution analysis. A good overview of a variety of superresolution methods proposed in the past years can be found in Farsiu et al. (2004). However, little has been published in the literature about the superresolution solution to complex satellite images.

In this paper, we propose a self-adaptive algorithm for motion estimation with the consideration of local geometric distortion in satellite imagery. Based on the theory of multiresolution analysis, the wavelet superresolution method for image reconstruction is briefly summarized and the post-processing of wavelet-interpolated SPOT-4 Pan images is performed experimentally. Following this introduction, some issues related to motion estimation are examined, and a description of the principle of the wavelet superresolution algorithm is presented. The modulation transfer function (MTF) and its corresponding formulas are detailed, the results obtained using the proposed algorithm to process SPOT-4 Pan imagery are presented, and conclusions are drawn and challenges in future research on superresolution reconstruction are outlined.

Motion estimation

Superresolution produces high-quality HR images from a set of degraded LR images where relative frame-to-frame motions provide different looks at the scene. Image superresolution is possible only in the presence of relative frame-to-frame motions of multiple frames. For the SPOT-4 images under consideration in this study, the relative motions are unknown, or are only approximately known according to the orbital velocity and path of the satellite. Thus the relative motions must be estimated from raw data in a preprocessing step of superresolution reconstruction.

To extract subpixel information content from the LR image sequences, image registration must be performed to estimate the relative motion parameters between the reference image and other images at the subpixel accuracy level. For the complexity of distortion, the motion between two sequential images is nonglobal and cannot be modeled with one fixed motion model. Thus high accuracy motion estimation for satellite imagery has been difficult and remains a considerable challenge. A comprehensive review on motion estimation can be found in Brown (1992).

In this study, we use the least-squares matching (LSM) technique at an accuracy level of 0.10–0.01 pixels (Ackermann, 1984; Baltsavias, 1991). Since motion estimation can be performed on each pixel and the accuracy of motion estimation depends on the accuracy of image matching, it is ideal to use the LSM technique for processing the images with different gray scales, positions, orientations, and scales. However, it is also too tedious to perform LSM on a pixel by pixel basis; despite its low efficiency, mismatching may occur due to lack of texture information in some regions, and the accuracy of matching cannot be guaranteed. It is well known that the motion between two sequential images cannot be simulated as one global motion model for the whole image, but under the assumption of small translation and rotation, no occlusions, and sufficient smoothness, an affine motion model can serve as a good approximation model for local small regions. Hence, if a large image region is divided into several small regions with appropriate sizes, motion estimation can be performed with the local piecewise affine transform model in these small regions. To solve the problem of region separation and local motion estimation, we adopt the high-accuracy feature point extraction operator in photogrammetry, such as the Förstner operator (Förstner and Gülch, 1987) to extract the feature points from the reference image, and then the LSM is performed to get its tie points in the registered image. With these pairs of points, we can use Delaunay triangulation (Davoine et al., 1996) to divide a large image region into smaller regions, and then local motion estimation can be performed with the piecewise affine transform model.

As the motion estimation is based on the affine transform model for a small region, there are six motion parameters that should be estimated. Each pair of points can be written in a generic form

$$\begin{bmatrix} x_i & y_i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_i & y_i & 1 \end{bmatrix} \mathbf{m} = \begin{bmatrix} x_i' \\ y_i' \end{bmatrix} \tag{1}$$

where $\mathbf{m} = (m_1\ m_2\ m_3\ m_4\ m_5\ m_6)^{\mathrm{T}}$. For a triangular region, let (x_i, y_i) be the feature points in the reference image and (x_i', y_i') be their corresponding tie points in the registered image, where $i = 1, 2, 3$. After substituting all pairs of points, Equation (1) can be represented in matrix form as

$$\mathbf{Am} = \mathbf{b} \tag{2}$$

where **A** is the coefficient matrix and **b** is the constant matrix. The minimum norm least-squares solution can then be expressed as

$$\mathbf{m} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b} \tag{3}$$

With these motion parameters estimated, we can register one small triangular region to the reference region. Motion estimation will be completed after all the triangular regions are registered in the same way.

Theoretically speaking, this may serve as an ideal method for the motion estimation between two sequential images, but the problem is how to implement it. It is well known that the geometric distortion of satellite imagery is very complicated and its complexity corresponds to the characteristics of the imaging area. The feature point extraction algorithm is based on gray scale, and no geometric distortion is considered, so it is difficult to determine how many feature points should be extracted and how they are distributed. Although the number of feature points extracted may be controlled manually with different thresholds during the extraction process, it is the most demanding and tedious work to properly perform the process of feature point extraction from all kinds of satellite images. Moreover, the process may lead to some unexpected results. For example, more feature points can be extracted in flat areas with less distortion but more texture information, and fewer feature points can be extracted in hilly areas with large distortion but less texture information.

To solve this problem, we consider the fact that the distortion of the satellite imagery is continuous if the image is smooth enough. Hence we can divide the whole image into some initial regions, calculate the affine transform parameters of each region, and then compare the parameters with those of neighbor regions. According to the continuous distortion assumption, the parameters of one region and those of its neighbor regions should be similar or the same if the triangular regions are divided well enough. If there is a large difference between these parameters (i.e., larger than a given threshold *T*), the distortion in this triangular region cannot be described precisely by the local affine transform model. Then more feature points should be extracted and added to this region to split them into more subregions. If the parameters of one region and its neighbor regions are the same, we can merge it with its neighbor regions. It is a simple iterative process and the iteration will not stop until the difference between parameters of all the regions and its neighbor regions is less than a given threshold *T*. To this end, when conducting feature point extraction and matching in a certain region, local geometric distortion of the region should be considered and the extracted feature points can be automatically adapted to a reasonable distribution.

The procedures of self-adaptive motion estimation are illustrated in **Figure 1**. As shown in **Figure 1**, the reference image and registering images are chosen and a set of initial feature points in the reference image are extracted with a high-accuracy positioning operator, and the process of self-adaptive triangulation is performed. With these initial points extracted and those tie points matched by the LSM method, the whole image is divided into some initial triangular regions by Delaunay triangulation. The motion parameters are then estimated through the initial triangular regions. If there are large differences between the motion parameters estimated from one region and those of its neighbors, more feature points and tie points are added to split the region of interest. If the differences between them are small enough, they can be merged into one region. As shown in **Figure 2**, during the merging process, the two vertices v_t and v_s are merged together to form a new vertex v'_s. The derived region now contains one vertex and two fewer triangles.

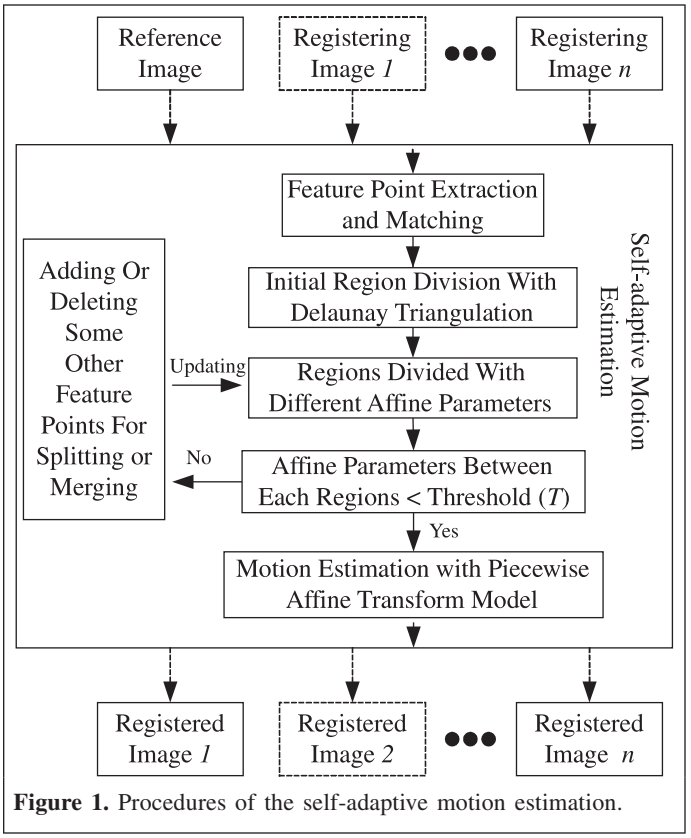


Figure 1. Procedures of the self-adaptive motion estimation.

This is an iterative process; the difference between the regions of interest and their neighbors needs more points, and

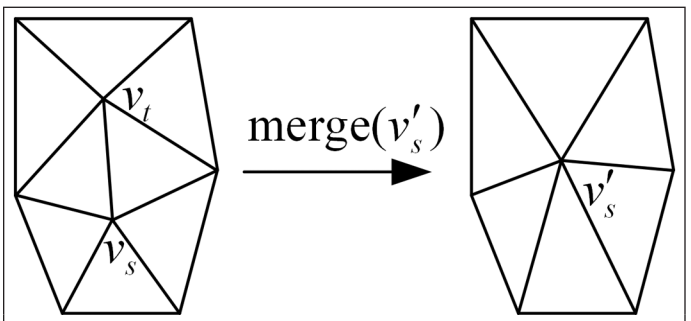


Figure 2. The process of triangle merging.

the added points will decrease the difference between them, which will not converge until the motion parameters of all the regions divided and their neighbor regions are just or almost the same. Lastly, motion estimation based on a piecewise affine transform model is performed to acquire the final motion estimation information.

Wavelet superresolution

The wavelet superresolution is a wavelet-based superresolution algorithm in a multiresolution framework. The wavelet coefficients and coarse-scale approximation coefficients are estimated at different scales through the irregularly sampled data, and then the values of the HR grid are calculated by the estimated decomposition coefficients. Since several studies on multiresolution analysis with orthonormal wavelets can be found in the literature, here we only summarize the wavelet superresolution method proposed by Lertrattanapanich (2003) without derivation that can be used with both the global and local motion models. A more detailed description of the wavelet superresolution method can be found in Nguyen (2000) and Lertrattanapanich (2003).

Interpolation for 1D nonuniformly sampled data

In this section, we first consider the wavelet-based interpolation technique for 1D nonuniformly sampled data. Suppose that there are P nonuniformly sampled data of 1D signal $f(t)$, which are defined as $\{[x_t, f(x_t)] | t = t_0, t_1, \ldots, t_{P-1}\}$; we want to obtain the values of M uniformly sampled data points of $f(t)$ at $t = 0, 1, \ldots, M-1$, where typically $P < M$, $0 \le t_i \le M$. Substituting each coordinate $\{[x_t, f(x_t)] | t = t_0, t_1, \ldots, t_{P-1}\}$ into the wavelet decomposition equation of the 1D signal (Nguyen, 2000), we have P linear equations

$$f(t_i) = \sum_k a_{J,k}\phi_{J,k}(t_i) + \sum_{j=J}^{-1}\sum_k b_{j,k}\psi_{j,k}(t_i) \quad i = 0, \ldots, P-1 \qquad (4)$$

where J is a given decomposition scale or subspace of Hilbert space, which is illustrated in **Figure 3**. The first term on the right-hand side of Equation (4) is the projection of $f(x)$ onto the subspace V_J, and this projection represents a coarse approximation of $f(x)$; and the second term is the projection of $f(x)$ onto the subspaces W_j, $j \ge J$, which provides increased details of $f(x)$. A more detailed description of this interpolation can be found in Nguyen (2000) and Lertrattanapanich (2003).

It is obvious that if we can estimate the scaling coefficients $\{a_{J,k}\}$, $k \in Z$ and $\{b_{j,k}\}$, $j \ge J$, $k \in Z$ (where Z is the data space) in Equation (4) from the given nonuniformly sampled data, we can substitute the uniformly sampled coordinates into Equation (4) to get their corresponding values. In this way, superresolution reconstruction can easily be achieved when the scaling function $\phi(t)$ and wavelet function $\psi(t)$ are finitely supported.

Suppose that $[0, N]$ is the finite support interval for $\phi(t)$ and let $t_{\max} = \max\{t_i\}$ and $t_{\min} = \min\{t_i\}$, then $S_J = \{-N + [2^J t_{\min}], \ldots, [2^J t_{\max}]\}$ should be the set of shifts with nonzero contribution in the right-hand side of Equation (4), and therefore Equation (4) can be rewritten in vector form as

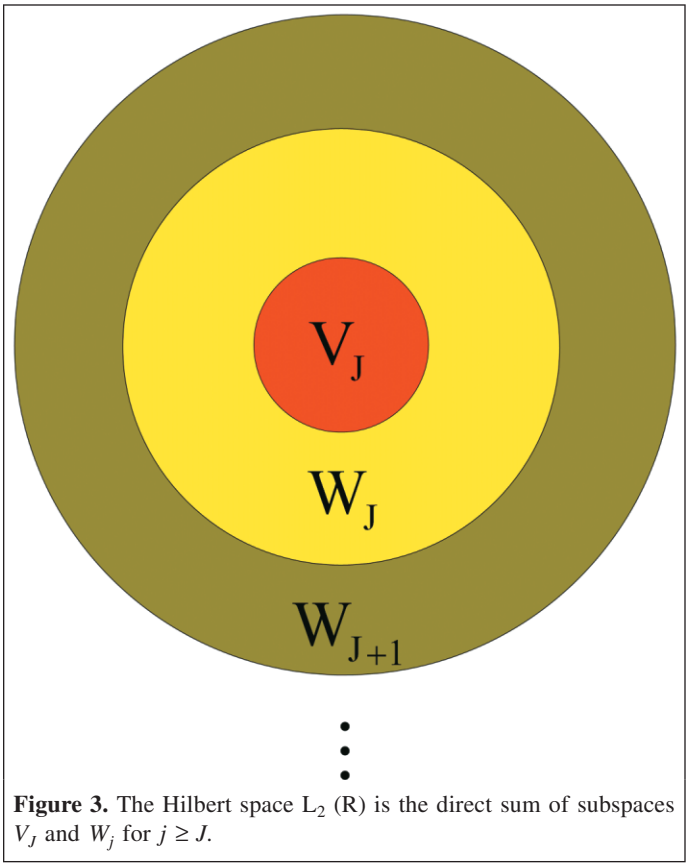


Figure 3. The Hilbert space L_2 (R) is the direct sum of subspaces V_J and W_j for $j \ge J$.

$$\boldsymbol{f} = \boldsymbol{G}_J \boldsymbol{a}_J + \sum_{j \ge J} \boldsymbol{H}_j \boldsymbol{b}_j \qquad (5)$$

where $\boldsymbol{f} = [f(t_i)]_{i=0,\ \ldots,\ P-1}$, $\boldsymbol{a}_J = (a_{J,k})_{k \in S_J}$, $\boldsymbol{b}_J = (b_{J,k})_{k \in S_J}$, $\boldsymbol{G}_J = [\phi_{J,k}(t_i)]_{i=0,\ldots,P-1}^{k \in S_J}$, and $\boldsymbol{H}_J = [\psi_{J,k}(t_i)]_{i=0,\ldots,P-1}^{k \in S_j}$.

To construct $\boldsymbol{G}_J$ and $\boldsymbol{H}_J$, we need to know the basic function values at sampling points $\{t_i\}$, which can be calculated by recursion. The key problem is to calculate the scaling coefficients and the wavelet coefficients. To calculate the scaling coefficients in Equation (5), we can ignore the detail components and just consider the low-frequency component:

$$\boldsymbol{f} \approx \boldsymbol{G}_J \boldsymbol{a}_J \qquad (6)$$

We can then solve Equation (6) for a regularized least-squares estimation in wavelet domain for some regularization parameter $\lambda > 0$, which is

$$\hat{\boldsymbol{a}}_J = (\boldsymbol{G}_J^{\mathrm{T}} \boldsymbol{G}_J + \lambda \mathbf{I})^{-1} \boldsymbol{G}_J^{\mathrm{T}} \boldsymbol{f} \qquad (7)$$

where $\mathbf{I}$ is the unit matrix. With the least-squares estimation value $\hat{\boldsymbol{a}}_J$, a coarse-scale estimation of $\boldsymbol{f}$ is yielded, which is denoted by

$$\hat{\boldsymbol{f}}_J = \boldsymbol{G}_J \hat{\boldsymbol{a}}_J \tag{8}$$

Then, $\hat{\boldsymbol{b}}_J$, the estimation of wavelet coefficient $\boldsymbol{b}_J$, can be calculated by computing the difference between $\boldsymbol{f}$ and $\hat{\boldsymbol{f}}_J$ with the least-squares estimation method:

$$\boldsymbol{H}_J \boldsymbol{b}_J \approx \boldsymbol{f} - \boldsymbol{G}_J \hat{\boldsymbol{a}}_J \tag{9}$$

The additional wavelet coefficient vectors $\boldsymbol{b}_j$, $j > J$, can be obtained in the same way:

$$\boldsymbol{H}_J \boldsymbol{b}_j \approx \boldsymbol{f} - \hat{\boldsymbol{f}}_j \qquad j > J \tag{10}$$

Thus the values of $f(t)$ at the HR grid points $t = 0, 1, \ldots, M-1$ can be calculated with the scaling coefficients and wavelet coefficients estimated as follows:

$$f(t) \approx \sum_{k \in S_J} \hat{a}_{J,k} \phi_{J,k}(t) + \sum_{j \geq J} \sum_{k \in S_j} \hat{b}_{j,k} \psi_{J,k}(t) \qquad t = 0, \ldots, M-1 \tag{11}$$

Interpolation for 2D nonuniformly sampled data

The wavelet superresolution in the previous section can easily be extended to the 2D case. Suppose that a set of P nonuniformly sampled data points are obtained after image registration $\{[(t_i, s_i), f(t_i, s_i)] | i = 0, 1, \ldots, P-1\}$, all these sampled coordinates can be substituted into the wavelet decomposition equation of 2D signal in the following form:

$$\begin{aligned} f(t_i, s_i) = & \sum_{k,l \in Z} a_{J,k,l} \Phi_{J,k,l}(t_i, s_i) \\ & + \sum_{j \geq J} \sum_{k,l \in Z} b^{\mathrm{h}}_{j,k,l} \Psi^{\mathrm{h}}_{j,k,l}(t_i, s_i) \\ & + \sum_{j \geq J} \sum_{k,l \in Z} b^{\mathrm{v}}_{j,k,l} \Psi^{\mathrm{v}}_{j,k,l}(t_i, s_i) \\ & + \sum_{j \geq J} \sum_{k,l \in Z} b^{\mathrm{d}}_{j,k,l} \Psi^{\mathrm{d}}_{j,k,l}(t_i, s_i) \qquad i = 0, 1, 2, \ldots, P-1 \end{aligned} \tag{12}$$

Equation (12) can also be expressed in matrix form as

$$\boldsymbol{f} = \boldsymbol{G}_J \boldsymbol{a}_J + \sum_{j \geq J} \boldsymbol{H}^{\mathrm{h}}_j \boldsymbol{b}^{\mathrm{h}}_j + \sum_{j \geq J} \boldsymbol{H}^{\mathrm{v}}_j \boldsymbol{b}^{\mathrm{v}}_j + \sum_{j \geq J} \boldsymbol{H}^{\mathrm{d}}_j \boldsymbol{b}^{\mathrm{d}}_j \tag{13}$$

Similarly, the detailed components (last three terms $\boldsymbol{b}^{\mathrm{h}}_j$, $\boldsymbol{b}^{\mathrm{v}}_j$, and $\boldsymbol{b}^{\mathrm{d}}_j$) are ignored to estimate $\mathbf{a}_J$; b^{v}_j and b^{d}_j are ignored to estimate b^{h}_j through $r_J = f - \hat{f}_J$, where $\hat{f}_J = G_J \hat{a}_J$; and then b^{d}_j are ignored to estimate b^{v}_j through $r_J = f - \hat{f}_J$, where

$$\hat{f}_J = G_J \hat{a}_J + \sum_{j \geq J} H^{\mathrm{h}}_j b^{\mathrm{h}}_j,$$

and by following the same method to estimate b^{d}_j. A more detailed description of wavelet superresolution can be found in Nguyen (2000) and Lertrattanapanich (2003). This demonstrates that $\boldsymbol{b}^{\mathrm{h}}_j$, $\boldsymbol{b}^{\mathrm{v}}_j$, and $\boldsymbol{b}^{\mathrm{d}}_j$ can be estimated in the same way as dealing with the 1D case described in the previous section.

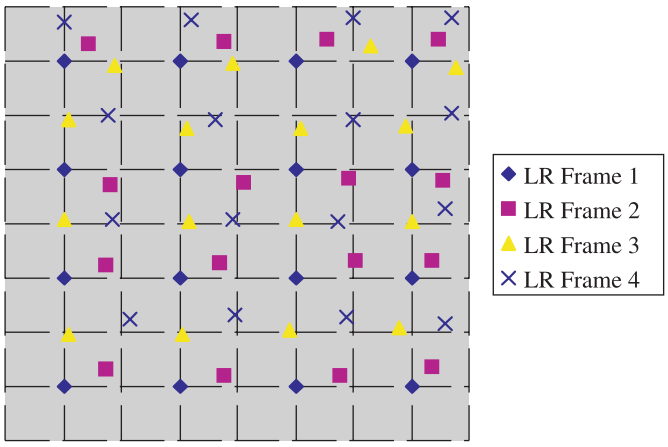


Figure 4. The relationship between motion estimation and superresolution interpolation.

Superresolution is implemented in this paper at the framework of interpolation and restoration. **Figure 4** illustrates the relationship between motion estimation and superresolution interpolation. When the motion estimation information from all the LR frames is obtained, the sampling points of LR frames can be considered as irregular samples on an HR grid (see **Figure 4**). Thus the HR image reconstruction becomes the issue of interpolation from nonuniformly sampled data.

HR image denoising and deblurring

Because the wavelet interpolation process does not consider the effect of sensor blurring and noise degradation, the reconstructed HR image usually is the blurred, noisy, and undersampled version of the original image. Hence HR image denoising and deblurring can further improve the signal-to-noise ratio. For the special case of SPOT-4 Pan image superresolution, the methods of image denoising and deblurring are analyzed in this section as the postprocessing for the interpolated HR image.

To analyze the problem from the viewpoint of signal processing, the acquisition of digital images can be modeled as

$$g(x) = h(x) f(x) + n(x) \tag{14}$$

where $f(x)$ is the original signal, $g(x)$ is the observed signal, $h(x)$ identifies the property of the imaging system, and $n(x)$ is the effect of noise. As shown in Equation (14), the quality of the observed image is mainly affected by the point spread function (PSF), which is the Fourier transform of $h(x)$, and noise. The original signal can be well restored if we know the PSF and noise intensity exactly. However, this is impossible for the case of the general image acquisition process. The noise effect can be modeled as a Poisson distribution in the general case, although it is very complex. Under this assumption, a Weiner filter is employed to reduce the noise effect to an acceptable intensity level. According to Latry and Rougé (1998), the

SPOT pushbroom imaging system is the product of convolution of three main contributing elements: the PSF of the optical sensor, the PSF corresponding to the integrated effect on the surface of the elementary detector, and the PSF corresponding to the effect of satellite motion along its orbit, which is caused by the displacement of the detector during the integration time. For convenience, here we express the equation in the Fourier transform of the PSF, which is called the modulation transfer function (MTF):

$$\mathrm{MTF}_{\mathrm{global}} = \mathrm{MTF}_{\mathrm{optical}}\mathrm{MTF}_{\mathrm{detector}}\mathrm{MTF}_{\mathrm{motion}} \tag{15}$$

Let f_x and f_y be the spatial frequencies expressed in rows and columns, respectively, let ΔX and ΔY be the dimensions of the detector projected to the ground corresponding to the row and column sides, respectively, and with V the velocity of the subsatellite point and t_i the integration time, the MTF of a SPOT image can be described as

$$\begin{aligned}
\mathrm{MTF}_{\mathrm{optical}}(f_x, f_y) &= \exp\left(-\alpha\sqrt{f_x^2 + f_y^2}\right) \\
\mathrm{MTF}_{\mathrm{detector}}(f_x, f_y) &= \exp(-\beta f_x)\sin c(\pi f_x \Delta X)\sin c(f_y \Delta Y) \\
\mathrm{MTF}_{\mathrm{motion}}(f_x, f_y) &= \sin c(\pi f_y V t_i)
\end{aligned} \tag{16}$$

where $\sin c(x) = \sin x/x$ and the values of parameters α and β are determined by the values of the MTF for the frequencies$(1/2\Delta X, 0)$, $(0, 1/2\Delta Y)$, which can be calculated by recursion. **Figure 5** illustrates the simulated MTF for a SPOT-4 imaging system; because the MTF is finite support, only the nonzero values are plotted as a grid size of 11 × 11.

Experimental results

Test datasets

In this study, three datasets are used to test the proposed algorithm, two of which are simulated datasets and one is the real dataset. The first simulated dataset consists of two subsections of an orthorectified SPOT-4 Pan image and its corresponding digital elevation model (DEM). The two images follow the strict distortion geometry of linear pushbroom imagery, and each scan line has an unique set of perspective centre and rotation angles. The collinearity equation between a pixel on the ith scan line $(x_i, 0)$ of the simulated image and its corresponding object point (X, Y, Z) in object space is written as

$$\left.\begin{aligned}
x_i &= -f\frac{a_1(X - X_{si}) + b_1(Y - Y_{si}) + c_1(Z - Z_{si})}{a_3(X - X_{si}) + b_3(Y - Y_{si}) + c_3(Z - Z_{si})} \\
0 &= -f\frac{a_2(X - X_{si}) + b_2(Y - Y_{si}) + c_2(Z - Z_{si})}{a_3(X - X_{si}) + b_3(Y - Y_{si}) + c_3(Z - Z_{si})}
\end{aligned}\right\} \tag{17}$$

The y coordinate along the direction of flight is implied in the position and attitude of the satellite at a given time, which can be linearly related to the location and attitude of the central linear array as follows:

$$\left.\begin{aligned}
\varphi_i &= \varphi_0 + \dot{\varphi}y \\
\omega_i &= \omega_0 + \dot{\omega}y \\
\kappa_i &= \kappa_0 + \dot{\kappa}y \\
X_{si} &= X_{s0} + \dot{X}_s y \\
Y_{si} &= Y_{s0} + \dot{Y}y \\
Z_{si} &= Z_{s0} + \dot{Z}_s y
\end{aligned}\right\} \tag{18}$$

where f is the focal length; φ_0, ω_0, κ_0, X_{s0}, Y_{s0}, and Z_{s0} are the exterior orientation elements of the central scan line; φ_i, ω_i, κ_i, X_{si}, Y_{si}, and Z_{si} are the exterior orientation elements of the ith scan line; a_i, b_i, and c_i are the elements of the rotation matrix; and $\dot{\varphi}$, $\dot{\omega}$, $\dot{\kappa}$, $\dot{X}_s$, $\dot{Y}_s$, and $\dot{Z}_s$ are the variation rates of the exterior orientation elements.

With 12 given elements (φ_0, ω_0, κ_0, X_{s0}, Y_{s0}, Z_{s0}, $\dot{\varphi}$, $\dot{\omega}$, $\dot{\kappa}$, $\dot{X}_s$, $\dot{Y}_s$, $\dot{Z}_s$), the orthorectified image, and the DEM, the distortion can be modeled precisely by Equation (18). The given elements that are used to generate simulated images are shown in **Table 1**. The ground elevation of the DEM range varies from 73 to 455 m.

The second simulated dataset was generated from an original HR image with dimensions of 300 × 300 pixels. The HR image was first blurred by a Gaussian PSF with a window of 3 and the variance of 1 and then was down-sampled by a factor of 2 with a random shift value to generate five LR image frames with dimensions of 150 × 150 pixels that include the motion information.

The third dataset consists of seven SPOT-4 Pan images with dimensions of 512 × 512 pixels each. Their acquisition time and viewing angles are listed in **Table 2**.

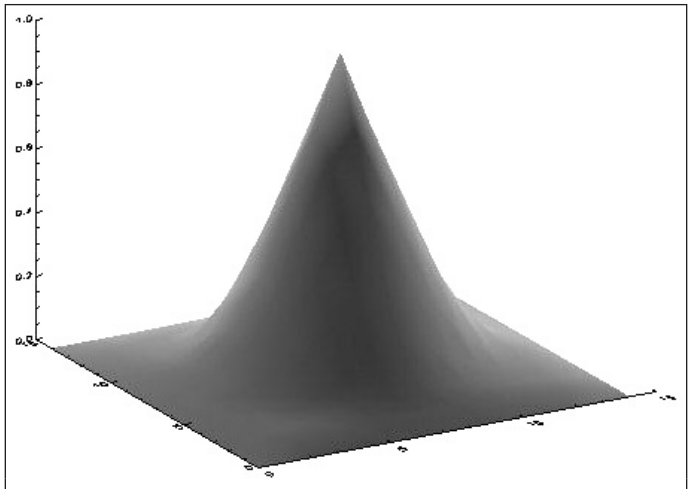


Figure 5. The simulated global MTF of SPOT.

Results and discussion

In the case of dataset 1, we first use a given position (X, Y, Z) to calculate its pixel coordinate (x, y) in each simulated image with the different coefficients given in **Table 1**. Thus the

Table 1. The elements of two simulated images.

Element	Simulated image 1	Simulated image 2
X_{s0} (m)	-4.232490×10^{5}	-4.451286×10^{4}
Y_{s0} (m)	1.694972×10^{5}	-3.375800×10^{4}
Z_{s0} (m)	8.127166×10^{5}	8.291104×10^{5}
φ_0 (rad)	4.789554×10^{-1}	6.243126×10^{-2}
ω_0 (rad)	-1.820755×10^{-1}	3.881773×10^{-2}
κ_0 (rad)	-2.547037×10^{-1}	-2.570903×10^{-1}
$\dot{X}_s$(m/pixel)	1.392486×10^{2}	2.644680×10^{1}
$\dot{X}_s$(m/pixel)	1.392486×10^{2}	2.644680×10^{1}
$\dot{Y}_s$ (m/pixel)	7.494897×10^{2}	5.535497×10^{2}
$\dot{Z}_s$ (m/pixel)	-1.853987×10^{2}	-2.769930×10^{1}
$\dot{\varphi}$ (rad/pixel)	1.849889×10^{-4}	2.039998×10^{-4}
$\dot{\omega}$ (rad/pixel)	-3.613227×10^{-5}	2.324194×10^{-4}
$\dot{\kappa}$ (rad/pixel)	2.814000×10^{-5}	-8.659719×10^{-6}

Table 2. Seven SPOT-4 Pan images.

Frame No.	Acquisition date (month–day–year)	Viewing angle (°)
1	05–08–2000	10.8
2	06–08–2000	26.4
3	02–21–2001	0.0
4	08–16–2001	4.8
5	10–10–2001	–20.4
6	01–04–2002	23.4
7	01–05–2002	–11.4

motion information between the two simulated images can be predetermined using the precisely calculated pixel coordinates. The proposed self-adaptive motion estimation algorithm is then used to estimate the motion information of these two simulated images. A total of 100 points evenly distributed in these two simulated images are randomly chosen to compare the estimated motion information with the predetermined motion information. For comparison purposes, the motion information is also estimated using the gradient-based motion estimation algorithm (Tekalp et al., 1992). The error between the estimated motion information and the predetermined motion information is shown in **Figure 6**. The estimated motion error by the proposed algorithm is plotted as a solid line in **Figure 6** and has a maximum error of 0.6 pixels and an average error of 0.2 pixels. The estimated motion error by the gradient-based algorithm is plotted as the broken line in **Figure 6** and has a maximum error of 4.7 pixels and an average error of 3.2 pixels. This result demonstrates that the proposed algorithm can achieve subpixel registration accuracy, which is much better than that of the gradient-based motion estimation algorithm. The lower accuracy of motion estimation obtained using the gradient-based algorithm may be due to the complex distortion of satellite images.

In the case of dataset 2, we randomly selected an LR frame from the five LR frame images as the reference frame and then estimated the motion information of the other LR frames based on the reference frame. **Figure 7a** shows one of the LR frames,

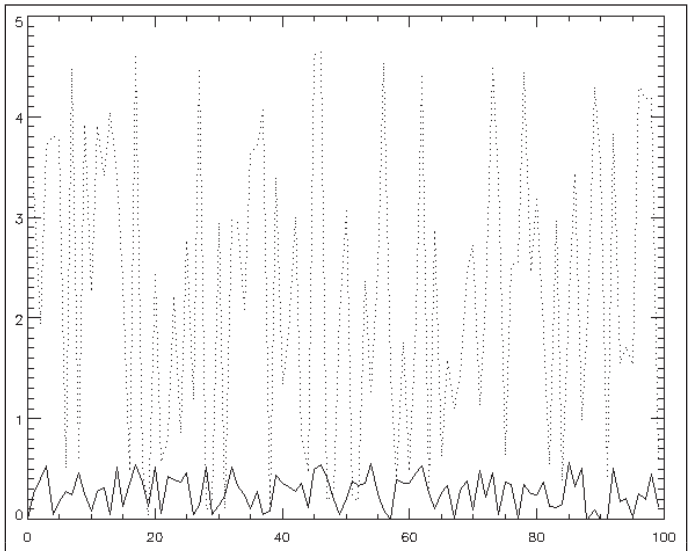


Figure 6. Motion error distribution with different motion estimation algorithms. The solid line is the estimated motion error by the proposed algorithm, and the broken line is the estimated motion error by the gradient-based algorithm.

Figure 7b the simple averaging of the registered frames, and **Figure 7c** the reconstructed superresolution image with a superresolution factor of 2, which is double that of the original shown in **Figure 7a**. We then use the mean square error (MSE) to evaluate the similarity between the reconstructed superresolution image and the original HR image. MSE is defined as follows:

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i,j}(x_{ij} - \hat{x}_{ij})^2$$

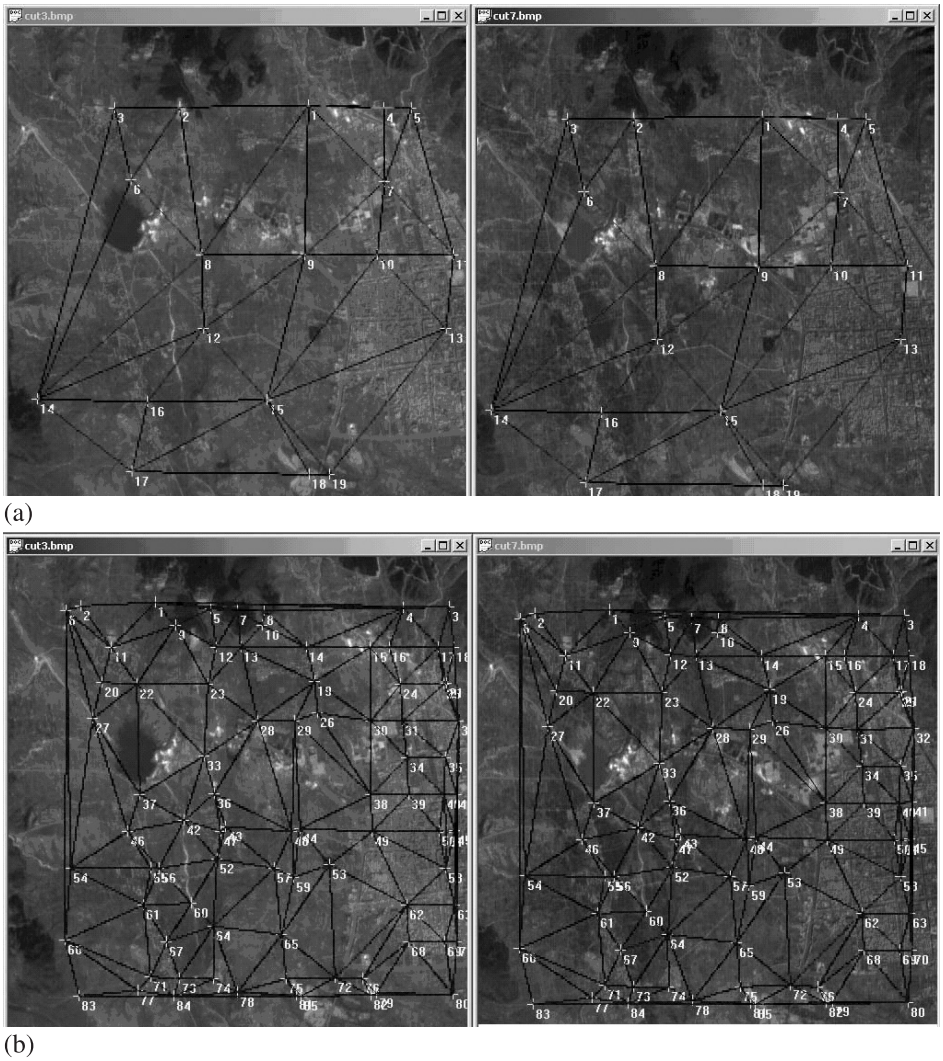
where x_{ij} and $\hat{x}_{ij}$ are the values of the original HR image and the reconstructed image, respectively. The results obtained demonstrate that the reconstructed superresolution image has an MSE value of 9.87, which is much higher than that of the simulated LR image with an MSE value of 32.99 and that of the averaging of the registered frames with an MSE value of 23.97 (see **Figure 7**).

In the case of dataset 3, we chose frame 3 as the reference frame because it is the only nadir viewing image and has a lower noise level. The motion information between the reference frame and the other frame images is then estimated using the proposed motion estimation algorithm. **Figure 8** illustrates the results of self-adaptive region division, with **Figure 8a** presenting initial triangulation results based on the feature points extracted, and **Figure 8b** the final triangulation results based on the feature points extracted after the self-adaptive process is completed.

The superresolution reconstruction of SPOT-4 Pan images is implemented using the wavelet superresolution algorithm described in the previous section, and the resolution enhancement factor employed is 2. The Daubechies DB4 filter (Ford and Etter, 1998; Nguyen, 2000) is used in this study

(a) (b) (c)

Figure 7. Superresolution experiment with dataset 2. (a) One of the LR frames. (b) Simple averaging of the registered frames. (c) Reconstructed superresolution image with a superresolution factor of 2.



(a)

(b)

Figure 8. Self-adaptive region division. (a) Initial triangulation with feature points extracted. (b) Final triangulation with self-adaptive feature point extracted.

(a) (b)

Figure 9. Superresolution image reconstructed from the original SPOT-4 Pan image. (a) Subsection of the original LR frame. (b) Reconstructed HR frame.

because of its good behavior in superresolution and simplicity. **Figure 9a** shows the original 10 m resolution SPOT-4 Pan image (as the LR frame), and **Figure 9b** is the superresolution image reconstructed from the original LR image, the resolution of which is double that of the original shown in **Figure 9a**.

Visually, the reconstructed image has a higher level of detail than that of the original SPOT-4 Pan image. To further prove the higher resolution and better quality of the reconstructed image, some object positioning experiments were performed (using Equation (15)) with the original and reconstructed images, respectively. We assume that object positioning is carried out under the same conditions, and the better positioning accuracy results in the better image resolution. Under this assumption, we first choose 11 ground control points (GCPs) and 13 checkpoints from both the original image and the reconstructed image. The exterior orientation elements of Equation (15) are calculated with these GCPs. Then the corresponding ground coordinates of these checkpoints can be calculated with the exterior elements, the DEM, and their image coordinates. To this end, we get two sets of ground coordinates, one of which is calculated and the other measured. Since all these calculations are conducted under the same condition, the difference between the calculated and measured ground coordinates can be used to identify the positioning accuracy. As shown in **Table 3**, the positioning accuracy of the reconstructed image is better than that of the original image, which implies that the reconstructed image has a higher resolution than that of the original SPOT-4 Pan image.

Table 3. Statistical analysis of positioning accuracy (in m) of the reconstructed HR image and the original LR image.

	Reconstructed HR image			Original LR image		
Object	$d_X = X' - X$	$d_Y = Y' - Y$	$d_{XY} = \sqrt{d_X^2 + d_Y^2}$	$d_X = X' - X$	$d_Y = Y' - Y$	$d_{XY} = \sqrt{d_X^2 + d_Y^2}$
1	25.048	15.847	29.640	45.149	3.706	45.301
2	–1.354	–30.571	30.601	–15.188	–15.106	21.422
3	40.642	–10.072	41.871	72.559	–44.896	85.325
4	–7.707	0.1203	7.708	–30.483	–5.0837	30.904
5	–22.515	4.976	23.059	–3.773	–28.647	28.894
6	13.823	17.718	22.473	25.213	1.173	25.240
7	35.821	–41.842	55.080	20.081	–31.058	36.985
8	20.669	–11.697	23.749	15.867	22.554	27.577
9	31.140	–40.047	50.729	–30.440	15.575	34.193
10	7.730	–25.594	26.736	60.585	–15.924	62.643
11	20.274	8.592	22.019	60.524	–50.501	78.827
12	9.629	6.362	11.542	–45.149	–43.731	62.856
13	6.025	–12.422	13.806	15.845	28.849	7.000
$\sqrt{\frac{\sum_{i=1}^{N}\Delta_i^2}{N}}$	22.046	21.557	6.467	39.530	28.244	32.914

Note: X' and Y' are calculated coordinates, and X and Y are measured coordinates.

Conclusions and outlook

A self-adaptive motion estimation algorithm for superresolution reconstruction of the multiframe SPOT-4 Pan images has been presented. The proposed method integrates feature point extraction with image matching, which can easily adapt to any motion estimation model and better solve the motion estimation for complex satellite images. The experiments have demonstrated that the proposed self-adaptive motion estimation algorithm is very promising when applying the wavelet superresolution algorithm to reconstruct the multiframe SPOT-4 Pan images. The results obtained show higher spatial resolution of the reconstructed images than the original images. This method can also be used in other areas where high accuracy motion estimation or image registration is required. However, the wavelet superresolution algorithm works well for the LR images with small dimensions. As the image dimensions increase, the increase in geometric series may slow down the process owing to its computation complexity. Therefore, improving the computation efficiency of the wavelet superresolution algorithm when processing a large satellite image (e.g., a typical SPOT-4 frame image is 6000×6000 pixels) is still a very challenging task. In addition, the considerable differences introduced by the images acquired at different dates during a long interval may also cause problems when using the proposed algorithm.

Acknowledgments

This work was sponsored in part by a grant from the Pre-Research and Development Program, Chinese Surveying and Mapping Agency. The authors wish to thank Prof. Zexun Gen and Dr. S. Lertrattanapanich for providing their publications on wavelet superresolution. The authors would also like to thank the two anonymous reviewers for their valuable comments.

References

Ackermann, F. 1984. High precision digital image correlation. In *Proceedings of the 39th Photogrammetric Week*. Institute for Photogrammetry, University of Stuttgart, Stuutgart, Germany. Vol. 9, pp. 231–243.

Aizawa, K., Komatsu, T., and Saito, T. 1991. Acquisition of very high resolution images using stereo cameras. In *Visual Communications and Image Processing '91: Visual Communication*. Edited by K.-H. Tzou and T. Koga. The International Society for Optical Engineering (SPIE), Bellingham, Wash. Proceedings of SPIE, Vol. 1605, pp. 318–328.

Altunbasak, Y., Patti, A.J., and Mersereau, R.M. 2002. Super-resolution still and video reconstruction from MPEG-coded video. IEEE *Transactions on Circuits and Systems*, Vol. 12, pp. 217–226.

Baltsavias, E. 1991. *Multiphoto geometrically constrained matching*. Ph.D. thesis, Institute of Geodesy and Photogrammetry, ETH Zürich. Mitteilungen No. 49.

Brown, L. 1992. A survey of image registration techniques. *ACM Computing Survey*, Vol. 24, No. 4, pp. 325–376.

Davoine, F., Antonini, M., Chassery, J.-M., and Barlaud, M. 1996. Fractal image compression based on Delaunay triangulation and vector quantization. *IEEE Transactions on Image Processing*, Vol. 5, No. 2, pp. 338–346.

Farsiu, S., Robinson, D., Elad, M., and Milanfar, P. 2004. Advances and challenges in superresolution. *International Journal of Imaging Systems and Technology*, Vol. 14, No. 2, pp. 47–57.

Ford, C., and Etter, D. 1998.Wavelet basis reconstruction of nonuniformly sampled data. *IEEE Transactions on Circuits and Systems*, Vol. 45, No. 8, pp. 1165–1168.

Förstner, W., and Gülch, E. 1987. A fast operator for detection and precise location of distinct point, corners and centres of circular features. In *Proceedings of the ISPRS Intercommission Conference on the Fast Processing of Photogrammetric Data*, 2–4 June 1987, Interlaken, Switzerland. pp. 281–305.

Gunturk, B.K., Altunbasak, Y., and Mersereau, R.M. 2002. Multiframe resolution-enhancement methods for compressed video. *IEEE Signal Processing Letters*, Vol. 9, No. 6, pp. 170–174.

Latry, C., and Rougé, B. 1998. SPOT5 THR mode. In *Earth Observing Systems III*, San Diego, Calif. Edited by W.L. Barnes. The International Society for Optical Engineering (SPIE), Bellingham, Wash. Proceedings of SPIE, Vol. 3439, pp. 480–492.

Lertrattanapanich, S. 2003. *Superresolution from degraded image sequence using spatial tessellations and wavelets.* Ph.D. thesis, Department of Electrical Engineering, Pennsylvania State University, University Park, Pa.

Lukosz, W. 1966. Optical systems with resolving power exceeding the classical limit. *Journal of the Optical Society of America (A: Optics, Image Science and Vision)*, Vol. 56, No. 11, pp. 1463–1472.

Mateos, J., Katsaggelos, A.K., and Molina, R. 2000. Resolution enhancement of compressed low resolution video. In *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2000),* 5–9 June 2000, Istanbul, Turkey. IEEE, New York. Vol. 4, pp. 1919–1922.

Nguyen, N.X. 2000. *Numerical algorithms for image superresolution*. Ph.D. thesis, Program in Scientific Computing and Computational Mathematics, Stanford University, Stanford, Calif.

Sauer, K., and Allebach, J. 1987. Iterative reconstruction of band-limited images from non-uniformly spaced samples. *IEEE Transactions on Circuits and Systems*, Vol. CAS-34, pp. 1497–1505.

Tekalp, A.M., Ozkan, M.K., and Sezan, M.I. 1992. High resolution image reconstruction from low-resolution image sequences and space-varying image restoration. In *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-92)*, 23–26 March 1992, San Francisco, Calif. Vol. III. Multidimensional Signal Processing. IEEE, New York. pp. 169–172.

Tsai, R.Y., and Huang, T.S. 1984. Multi-frame image restoration and registration. In *Advances in computer vision and image processing*. Edited by T.S. Huang. JAI Press, Greenwich, Conn. pp. 317–339.

Tsai Ur, H., and Gross, D. 1992. Improved resolution from subpixel shifted pictures. *CVGIP: Graphical Models and Image Processing*, Vol. 54, No. 2, pp. 181–186.

