A Coarse Elevation Map-based Registration Method for Super-resolution of Three-line Scanner Images

Rongjun Qin, Jianya Gong, Hongli Li, and Xianfeng Huang

Abstract

Three-line scanner imagery provides three overlapped images in an along-base direction, and creates a possible avenue to obtain higher quality images through the application of the super-resolution. Accurate co-registration of the three images is a key step for super-resolution. However, discontinuities and occlusions resulting from the 3D-to-2D projection cause mis-registration in traditional 2D-image-level co-registration methods. In this paper, we address this problem by introducing 3D information extracted from image triplets by using GPS/IMU data as an approximation. The core of the proposed method is to use a number of height layers derived from feature points and image partitions, in the form of a coarse elevation map (CEM), as a 3D constraint to restrict registration on the corresponding height. In terms of super-resolution, we also propose a tree-based fast adaptive template matching method for Knife-edge detection to fully automate the SRKE super-resolution algorithm. Experimental results show that the proposed method produces improved registered images and accordingly yields significant resolution enhancement as compared to other methods.

Introduction

High Resolution (HR) images provide reliable data for many applications such as image segmentation, pattern recognition, and medical diagnostics. It is easy to distinguish similar objects in high-resolution images. In remote sensing applications however, especially regarding satellite images, image resolution is the critical factor indicating the volume of information. The more information obtained from these images, the better decision support these images can provide. Therefore, increasing the resolution of remote sensing (RS) images is an active area of research. In particular, image Super-Resolution (SR), was developed to enhance the resolution of images. It renders images with a higher resolution than that of the original image. Due to the number of original images available, super-resolution methods can be divided into two scenarios: Multi-frame super-resolution and Single frame super-resolution (Filip, 1999). Single-frame SR requires only one original image for a scene. Techniques for Single-frame SR are almost regularization methods based on prior information such as edges, smoothness, etc. (Almeida and Almeida, 2010; Li *et al.*, 2010; Zhang *et al.*, 2010). Multi-frame images usually are a set of images obtained from the imaging system that shoots the same scene at a different time instants or from a different position at almost the same time. Single-frame SR is more like image enhancement, whereas Multi-frame SR provides more information about the details of a scene. Multi-frame SR however, encounters an inevitable problem in that there must be a pixel-to-pixel match among the low-resolution (LR) images, or image registration.

Image registration is a process used to create point correspondences for all pixels between the LR images. Common registration methods such as Vandewalle *et al.* (2006) and Keren *et al.* (1988) only account for shifts and rotation. Du *et al.* (2008) proposed a control point based image registration method applied between the multi-spectral bands, and fails to process stereo images. Leila M. Fedorov developed an expert system to automatically select registration algorithms for different cases (Fedorov *et al.*, 2003; Fonseca and Manjunath, 1996). Applications in the field however usually encounter another complication: The Three Dimension (3D) to Two Dimension (2D) and (3D-to-2D) projection in stereo images causes significant projective distortion and is too coarse to be represented by shifts and rotation alone. Researchers have tried many methods to deal with this situation: Lucas and Kanade (1981) proposed a method to create an optical flow field to find correspondences between stereo images. Ogale and Aloimonos (2007) improved this method to deal with images with sharp contrast and slanted surfaces. Hamzah *et al.* (2010) proposed a Sum of Absolute Difference (SAD) method to solve stereo image correspondence, taking the absolute difference between each pixel in a block, summing them up to create a metric of similarity. Scharstein and Szeliski (2002) developed a taxonomy method, using other stereo matching techniques as building blocks for their algorithm, integrating these algorithms for better performance. However, these methods can only handle stereo images with relatively short baselines. These state-of-art matching algorithms need accurate georeferencing to find pixel-wise correspondences. Remote sensing cameras provide not only panchromatic or multi-spectral images, but also relatively accurate positioning data unavailable for ordinary camera

Rongjun Qin is with the Singapore ETH Center, Future Cities Laboratory, Singapore, and formerly with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, China (rqin@student.ethz.ch).

Jianya Gong, Hongli Li, and Xianfeng Huang are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), #129, Luoyu Road, Wuchang District, LIESMARS, Wuhan University, Wuhan, P. R. China, 430079.

Photogrammetric Engineering & Remote Sensing
Vol. 79, No. 8, August 2013, pp. 000–000.

0099-1112/13/7908–0000/$3.00/0

pictures. Satellites have a star sensor to record the transformation parameters from the Orbit Coordinate System (OCS) to the Body Coordinate System (BCS). Airplanes carry a GPS (Global Positioning System) and an IMU (Inertial Measurement Unit) to measure the position and orientation of the camera at different times. This data can be used as additional information in the 3D world to improve registration reliability.

In this paper, we focus on registration for super-resolution of Three-Line Scanner (TLS) images, since these are widely used for aerial and satellite mapping. As a typical Multi-frame image sensor, TLS has its unique characteristics: it has three almost 100 percent overlapped stripes of images (Nadir, Backward, and Forward), and each image is captured from a different perspective of view. The three images provide more texture-detail information. Along with the images, the orientation data for each scan line can be downloaded from the GPS, IMU or the star sensor affixed to the camera. These overlapped images are good sources for super-resolution. For registration methods designed for stereo images, the parallax among TLS images is too large for ideal registration results. At the same time, these stereo methods also require highly accurate geometry to generate epipolar images, which creates a heavy data preprocessing load. Based on the unique characteristics of TLS imagery, we propose a Coarse Elevation Map based (CEM) registration method that efficiently renders acceptable and robust registration results for super-resolution. This proposed method first creates a CEM of the ground space, and then corrects each stripe onto the same nadir view by incorporating the CEM as additional information. After registration, slight pixel-level differences between these registered images exist as redundant information for super-resolution. We use Super-Resolution based on knife-edges (SRKE) (Qin *et al.*, 2010) to perform super-resolution. In addition, we extend the SRKE method with a tree-based hierarchical template matching method to detect knife edges for full automation.

In summary, our contributions in this paper are threefold: First, the proposed registration method adopts the implicit geometry as height layers, which theoretically and practically outperforms similar methods; second, it robustly excludes outliers in the final super-resolution results; and third, this paper proposes an adaptive template matching algorithm to search knife-edges for automation of super-resolution, and summaries this completely automatic working-flow for generating higher-resolution image from raw TLS data.

This paper is organized as follows: in the next Section, we present general considerations about the registration method and workflow, as well as pre-processing followed by the proposed registration is introduced in detail. In the subsequent section, SRKE super-resolution is re-examined and a proposed adaptive template matching approach is presented. Finally, we draw some conclusions and discuss the drawbacks of our method.

General Methodological Considerations

Image Super-resolution has gained a great attention in both the computer vision and remote sensing fields. A variety of approaches have been developed, with several algorithms commercially available. Although they use different super-resolution methods, these registration methods cannot perform satisfactorily. An ideal solution is to firstly acquire an accurate DSM from the TLS imagery, then use different views to generate a true ortho-image, and finally perform the super-resolution on those ortho-photos. This requires precisely georeferenced images and a perfect dense matching algorithm to generate a pixel-wise, accurate DSM, which is in practice, hard to obtain without manual work or additional data such as airborne lidar data. These considerations motivated us to start the registration in similar but coarse way: first we generated very coarse 3D information as height layers, namely as CEM, and then performed image registration using this information.

The general methodology is illustrated in Figure 1: the original data downloaded from the TLS camera includes raw images, orientation data (or quaternion representation from star sensors), camera files, and support files (Tempelmann *et al.*, 2000). A fast rectifier is needed to generate the pre-rectified image, one that excludes distortion to be used to perform matching. The Geometrically Constrained Cross-Correlation (GC^3) algorithm (Zhang and Gruen, 2004) serves as the key matcher. Since "Automatic DSM Generation" is still in the research phase, mismatches often cause false digital surfaces, and moreover require highly georeferenced data, which may involve more preprocessing work. Therefore, to simplify the registration problem as well as focus on a fast processing approach, it is assumed that projective distortion is reduced to affine distortion on geometrical planes; therefore, 3D models are estimated as a set of height layers. Since a super-resolution algorithm is not our main interest in this research, we use the SRKE (Super-Resolution based on Knife-Edge) method, whose performance is recognized as one of the state-of-art algorithms for well-registered images (Qin *et al.*, 2010).

The SRKE method introduces information from knife-edges as a regularization term. The knife-edge method has unique characteristics: it varies in both size and orientation. As an adaptable template matching method, it is adopted for searching for these edges. The workflow for

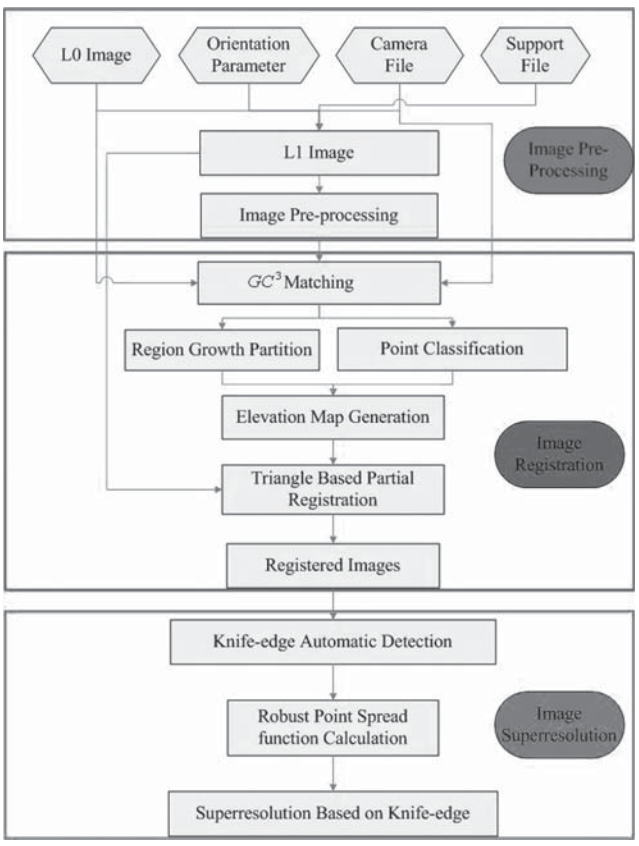


Figure 1. The Super-resolution Flow Chart.

super-resolution based on our registration method has the following characteristics:

- It has the ability to register stereo images with large projective distortion.
- It has the ability to detect mismatches, which increases robustness.
- It has the ability to automatically generate high-resolution images for each stripe in the TLS imagery from the original downloaded data, thus avoiding complicated human interaction during the process.
- It is designed specifically for imaging systems with orientation data, a key factor for solving the registration problem.

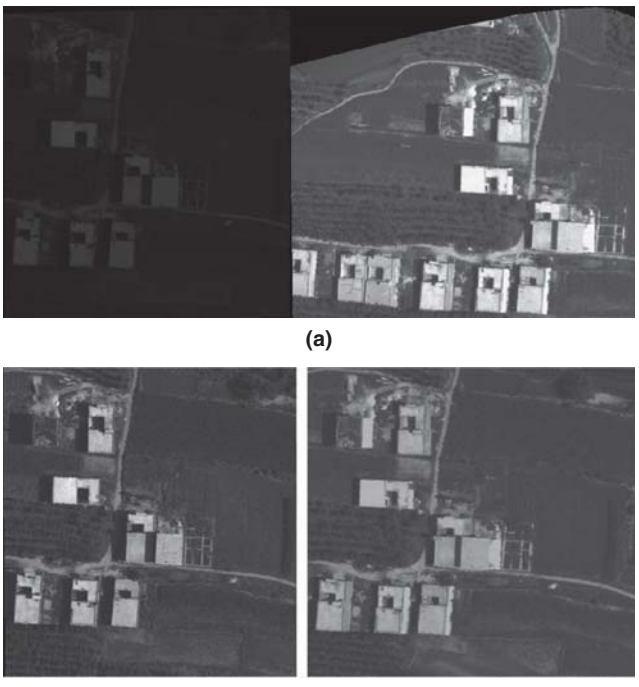
The details will be introduced in the next sections. For convenience of expression, the following sections will discuss registering forward and backward views to the nadir view, other situations are dealt with in a similar way.

Image Registration

Image Preprocessing

The raw TLS image data is rectified by projecting each view to the average height of the terrain through co-linearity equations (Tempelmann *et al.*, 2000) to produce level 1 (L1) images. The registration in turn, will be performed on these L1 images.

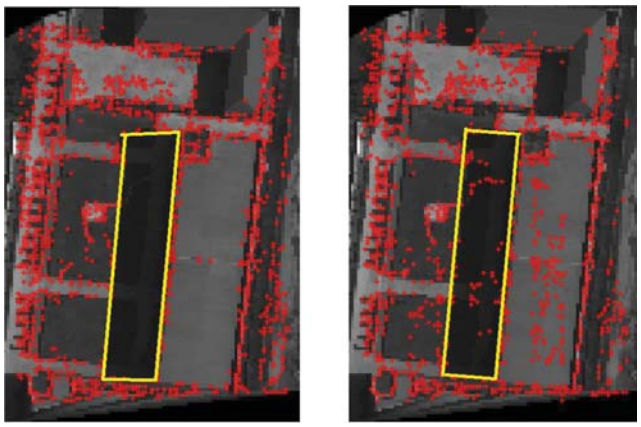
After Rectification, we adopted a "De-noising + Radiation Correction" strategy as proposed in Zhang and Gruen (2004). The Adaptive Smooth Filtering algorithm (Saint-Marc *et al.*, 1991) is used for de-noising and the Wallis filter (Zhang, 2005) is for correcting radiation difference. We use a two-step Wallis filtering for both in-between strips and inside stripes; in the in-between step, radiation difference is corrected to match more features (Figure 2a and 2b), and the inside-stripe step generates more interest points especially in the shadow areas (Plate 1, shown in yellow).



(a)

(b)

Figure 2. Wallis Filter: (a) before Wallis filter, and (b) after Wallis filter.



(a) (b)

Plate 1. Extracted interest points: (a)before, and (b) after Wallis filter.

Point Matching

The interest points are generated by the Forstner operator (Forstner and Gulch, 1987), and then we apply the advanced Geometrically Constrained Cross-Correlations (GC^3) method from Zhang (2005). Unlike the fixed widow correlation method, the advanced GC^3 method uses an adaptive window (Kanade and Okutomi, 1994) that changes according to the texture patterns and discontinuities in the images for the highest peak in cross correlation. For flat areas, the window size increases to avoid repetitive patterns and similar features, whereas in areas with significant changes of elevation, the window decreases in size, and shifts to avoid discontinuities where adaptive window matching has higher peak in correlation than the fixed window matching (Plate 2).

Coarse Elevation Map (CEM) Generation

Based on the assumption that affine registration performs well on planes, we estimated the 3D surface as a set of planes with different heights, and once certain that pixels in one area in the nadir image share the same elevation, the corresponding area in the backward and forward images can be rectified to the nadir view by an affine transformation. The transformation parameters are determined from the sparse identical points obtained by point matching.

Orientation parameters without accurate georeferencing might not be accurate for generating elevation. Thus, non-ignorable errors in elevation values exist. Hence, it is not wise to determine a plane by several points that have exactly the same elevation value. We can however, divide these points into several categories according to their elevation value. Points in a category may not share the same elevation value, but their values are relatively close (within a given range). Each category has its average value, and this value is allocated to be the elevation value of this category. These extracted 3D points are classified into several categories. We adopted K-Mean clustering (MacQueen, 1967) method to perform classification. The algorithm starts from a randomly chosen value, and iteratively classifies these values into several classes according to a fixed bearing distance. Points whose distances to the average value of a class are smaller than the bearing distance will be categorized in this class. As an iterative process, the process ends when the categories remain stable.

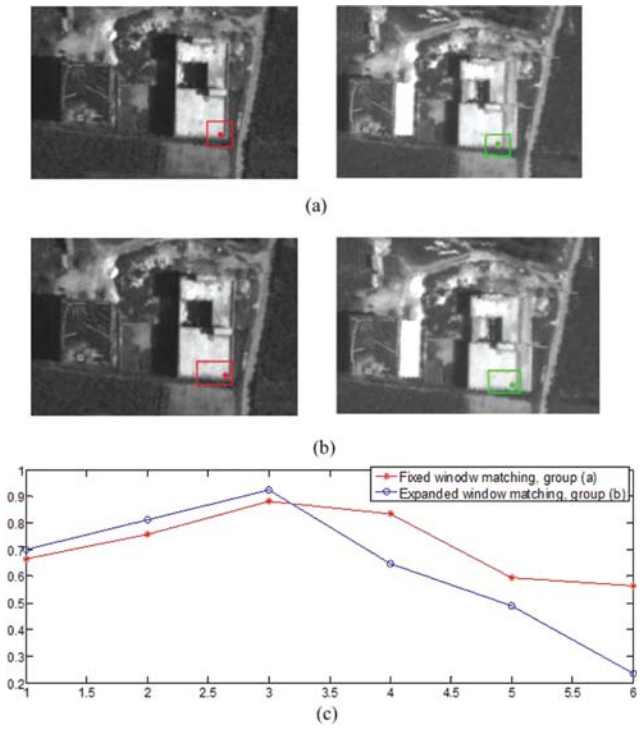


Plate 2. Adaptive window size and shift: group (a) shows the fixed window in (left side) nadir, and (right side) forward stripe, group (b) shows the adaptively corrected correlation windows, and (c) the correlation curve of (a) and (b).

In Plate 3 points with the same color belong to the same class, while the color bar indicates the elevation values. In this Plate 3, we can clearly see the elevation hierarchy: points on the roofs show higher elevations and points in grounds show lower elevations.

After 3D point classification, each matched point belongs to a specific category. Points in the same category can be seen as being in the same plane. Creating TINs (Triangle Irregular Network) among points in a category, and performing affine registration between these TINs can make the affine registration more robust against the inaccurate planes; but since points in a plane might be spread throughout the image, discontinuities can occur inside these TINs. Therefore, it is essential to coarsely estimate the category of each pixel in the image to which it belongs.

Since the plane changes very slightly in elevation, the gray level of the image pixel is a good indicator that pixels with the same gray value share the same elevation, especially in urban areas. Because the points in a same plane always reflect the same amount of sunlight, the pixels in an image plane might have similar gray values.

We adopted the Region Growing method (Adams and Bischof, 1994) to perform segmentation based on the image gray level (Plate 4). The seed for the Region Growing method is not randomly chosen; it starts from the matched identical points to ensure that each identical point is categorized in a gray-level class. Then, based on the gray level, the whole image is cut into several regions; pixels in each region have similar gray values. The threshold of the gray level for region growing is a user-defined parameter. The greater the threshold is, the less the number of regions will be. In Plate 3, each color represents a region; pixels in the black areas represent those that do not belong to any region.

Neither the gray-level based regions nor elevation categories alone are sufficient to indicate a plane, so we combine them together to estimate the elevation category of each image pixel. We adopted following criteria to estimate the elevation of each image pixel:

1. For a region A in the image, record elevation categories $\{Q_i\}$ in this region.
2. For each elevation category $Q_j a?? \{Q_i\}$, in region A, count the number of identical points, and define the elevation of all pixels in region A as the elevation of the one with most identical points. For example, if in region A, Q_2 has most points located in A, then region A will take the elevation value of Q_2 as its elevation.
3. After all available regions are labeled with elevation categories, the black areas are labeled with the same elevation category closest to labeled point in these areas.

After the above process, we obtained an estimated elevation map; the elevation map for Plate 3 is shown in Plate 5. This estimated elevation map was generated from the gray level of the image, so errors in this elevation map are likely.

Plate 3. Classified feature points in 3D, points indicate their x-y positions and the color indicate their height values.

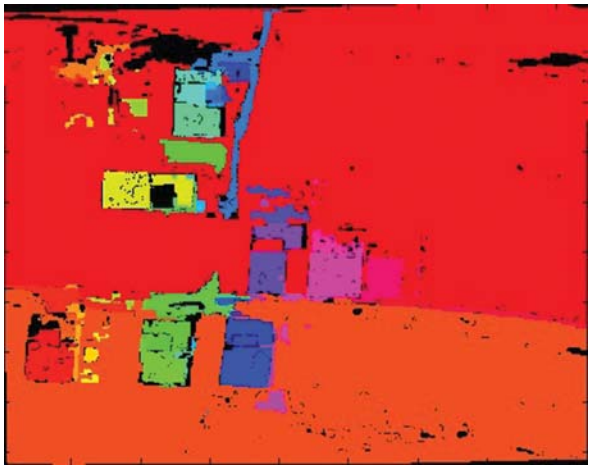
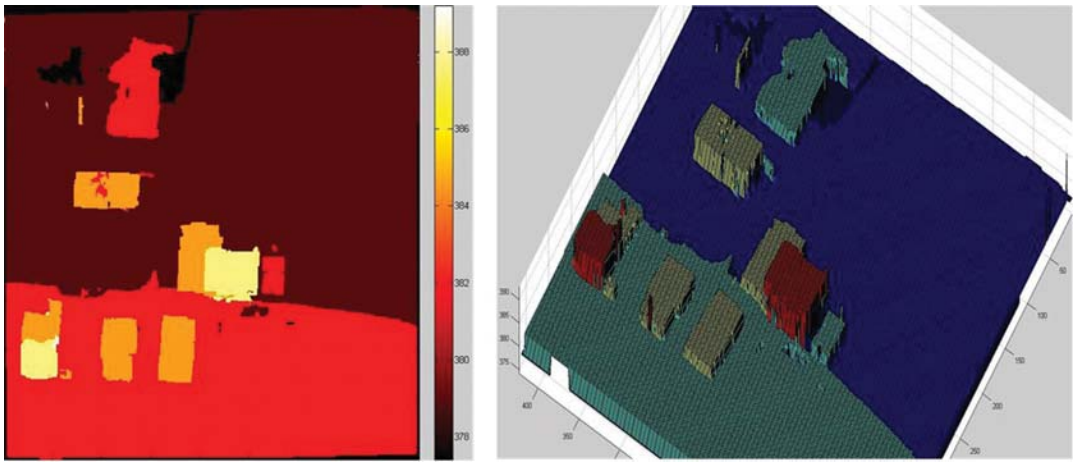


Plate 4. Region growing result.



(a) (b)

Plate 5. Estimated elevation: (a) hot map, and (b) visualization.

It only serves however, as an indicator for registration. Errors are considered in the registration process.

Affine Registration from Triangular Units with CEM Checking

The theoretical basis for our registration method is the assumption that a projection transformation equals an affine transformation in a plane. Triangles are created as small units based on the identical points; the estimated elevation map is auxiliary data to specify pixels of each triangle to planes. Delaunay TINs are first created based on the identical points of each elevation category. These triangles initially cover an area with the same elevation. We perform a six-parameter affine transformation inside each unit:

$$\begin{aligned} x_1 &= a_1 x_2 + b_1 y_2 + c_1 \\ y_1 &= a_2 x_2 + b_2 y_2 + c_2 \end{aligned} \tag{1}$$

Each triangle has one set of affine transformation parameters. However, not all the pixels in a triangle will be registered, only those that in the same plane of the vertices of this triangle will be mapped into the other view. The reason for this is that the affine parameters are calculated from vertices; therefore for each triangle, the affine registration works as follows;

$$I_r(x_1, y_1) = \begin{cases} I_2(a_1 x_2 + b_1 y_2 + c_1, a_1 x_2 + b_1 y_2 + c_1) \\ I_1(x_1, y_1) \\ if\ CEM(x_1, y_1) = CEM(v_i), i = 1,2,3 \\ otherwise \end{cases} \tag{2}$$

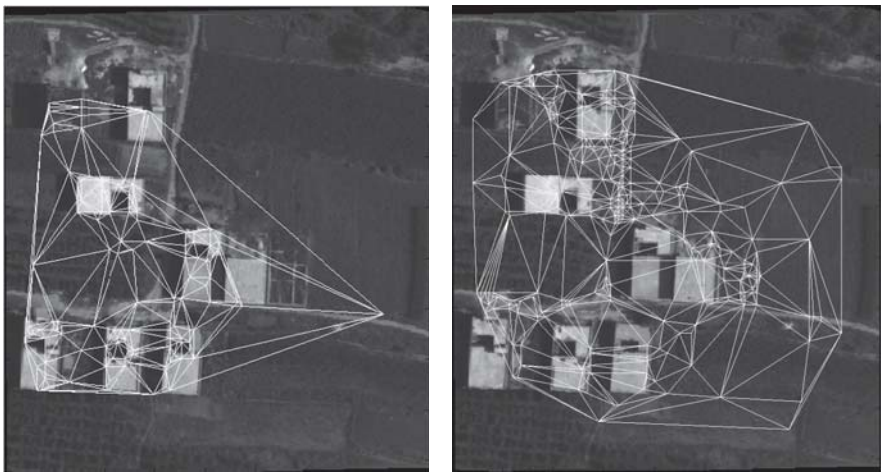
where I_1, I_2, I_r are the reference image, to-be-registered image, and registered image, respectively v_i are the vertices of a triangle; v_i is a matrix that has the same dimension of I_1, which stores the indices of planes each pixel belongs to. To avoid registration errors (which might be from either the matching or elevation map), a simple threshold test on correlation is provided to exclude wrongly registered triangles. Therefore, the proposed registration algorithm can be summarized into the following steps:

1. Generate Coarse Elevation Map.
2. Set $I_r = 0$, Size (I_r) = Size (I_1)
3. For elevation category Q_i, create TINs T_i among points in Q_i
4. For each triangle in T_i, calculate the affine transformation parameters (a_1 ,b_1, c_1 , a_2, b_2, c_2) in Equation 1) by substituting x_1, y_1 and x_2, y_2 in Equation 1 for the coordinates of the vertices.
5. For each pixel in a triangle of backward/forward image, calculate its value by Equation 2.
6. Record those pixels that have been registered to the nadir image for a correlation test. Calculate a correlation between those pixels and the corresponding pixels in nadir image. A threshold c_i is given to judge whether these pixels will be finally assigned to the right position in I_r.

Figure 3 shows the TIN created from two representative height layers. As is seen in the Figure 3a, most of the triangles are concentrated on planar faces, especially on the roofs. These roof triangles were created with the category with elevations close to the roofs; while Figure 3b, shows the triangles on the height layer close to the ground. There are some faulty triangles covering both roofs and ground. Only the part that has the same height layer as the vertices will be registered. For texture-less areas, usually there are fewer corresponding points. These triangles are larger than those found in texture-rich regions, however, because texture-less regions contribute much less to super-resolution than texture-rich regions. Those regions could even be eliminated when corresponding points are not found.

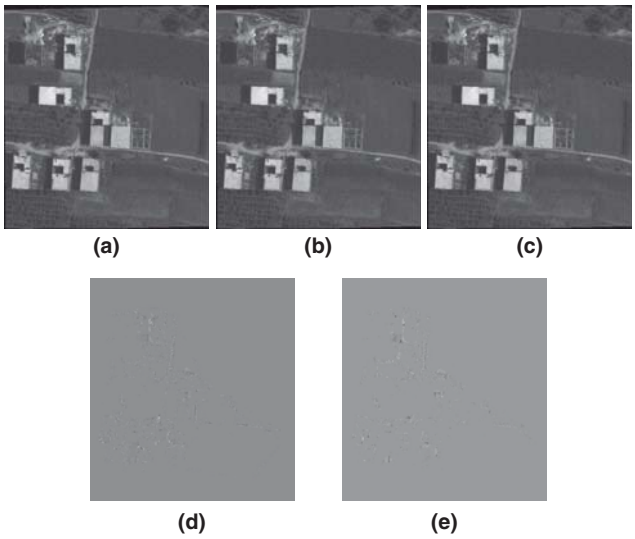
As shown in Figure 4, there are slight differences between the registered forward/backward (referred to as "registered image" hereafter). These differences are seen as detailed edges and textures, and assumed to be the complementary information for image super-resolution.

The basic assumption of the proposed method is that the 3D space is divided into several piecewise height layers, but it does not use the assumed geometry for image registration directly. It uses those height layers as auxiliary information. Therefore, each height layer contains those areas with less height variation, and affine registration is effective in those



(a) (b)

Figure 3. Triangular tins created by different height categories: (a) close to roofs, and (b) close to ground.



(a) (b) (c)

(d) (e)

Figure 4. The registered images: (a) registered backward image, (b) Registered forward image, (c) nadir image, (d) difference map between registered backward image and nadir image, and (e) difference map between registered forward image and nadir image.

regions. For cases where the surface is not piece-wise and horizontally flat, the proposed method will work differently:

1. For slant surfaces, for example, slant roofs and smooth hills if the slope is small, it will be approximate as a plane; in this case, the assumption also takes effect since the slant surfaces are also planes, and affine registration works well in those planes. While if the slope of the slant surface is large, it will be divided into several height layers (the stair-case effect), and each layer corresponds to a certain part of the slant surface; this fits the assumption as well.
2. For areas with extreme height variation within a small area, for example, towers and electrical poles, there may not be estimated planes around these objects, or faulty planes; and then the registration will be blocked by robust checking. Those unregistered regions will retain its original pixel values, since in the unregistered region there is no redundant information from other views. The result of these regions will be the same as de-convolution.

Image Super-Resolution

Our previous work "Multi-frame super-resolution based on knife-edges" (SRKE; Qin *et al.*, 2010) is adopted in the final step. It was originally derived from Sroubek and Flusser (2003) work on Multi-frame images Blind super-resolution (BSR) (Cristobal, 2007). Addressing the BSR deficiency that the parameters in BSR have to change for accommodating different datasets, SRKE can use one set of parameters to handle all images of the same type. SRKE uses a special feature widely found in remote sensing images, called knife-edge, as a constraint on a Bayesian model. The algorithm iteratively solves groups of linear equations for minimizing the following Energy function:

$$E(u,h)=\sum_{k=1}^{K}\left\| S\left(u*h_k\right)-g_k\right\|+\alpha R[u]+\beta\sum_{K=1}^{K}Q\left[h_k\right] \tag{3}$$

where K is the number of images (it is three in TLS case); u is the high-resolution image; g_k is the low resolution image; h_k is the blurring function for each lower resolution image; $\sum_{K=1}^{K}\left\| S\left(u*h_k\right)-g_k\right\|$ is the noise reduction term; and R[u]and Q [h_k] are the regularization terms of u and h_k, respectively, and take in the following form:

$$R[u]=\int_{\Omega}|\nabla u|d\Omega \tag{4}$$

$$Q[h]=\left\| S_{x,y}\left(h_h(x).h_v(y)\right)-p_h(x).p_v(y)\right\| \tag{5}$$

where S_{xy}(*a*??) denotes continuous down-sampling function, and $p_h(x)$ and $p_v(y)$ are the 1d PSF (point spread function) in horizontal function and vertical function, respectively. For knife-edges which are neither horizontal nor vertical, we adopt a robust method of extracting PSF proposed by Qin and Gong (2010).

However, the knife-edge in the previous work is chosen manually, as the knife-edge varies in both size and orientation; it fails easily if a simple template matching algorithm is adopted. Therefore, we proposed adaptive template matching, which searches knife-edges with arbitrary sizes and orientations.

Adaptive Template matching for Searching Knife-Edges

Considering the sampling effect in digital images, the orientation of knife-edges has finite possibilities: for example, a 3 × 3 kernel which contains a knife-edge has five possibilities (Figure 5a), and it increases as the size of kernel increases. Therefore, our algorithm starts to search knife-edges from a small kernel (say 3 × 3) to reduce the computation, if a knife-edge with a small size is found, the algorithm will adaptively make correlation test on the same area with a larger size of kernel. However, the algorithm does not perform an exhaustive search through all directions on the larger kernels, since the orientation of a small kernel can give a certain range of orientation in the larger one. The search will end once it finds the knife-edge.

To make the knife-edge searching more efficient, we first generate a template tree according to the successive relationship between the small kernels and large kernels: the roots are all the possible kernels of a 3 × 3 knife-edge, and then the children of each kernel are 5 × 5 kernels, the central 3 × 3 parts of which are exactly the same as their parents. (Figure 5b), the construction of the tree ends up with size 13 × 13 in our experiment; it can be even larger to estimate a more accurate PSF. However, the larger the kernels are, the more computational time it takes, and an increased possibility to fail as well; while if the size is too small, the estimated knife-edge may not be accurate enough. Therefore, as a trade-off, 13 × 13 works quite well through all the experiments as an empirical parameter. The tree structure only needs to be generated for one time and can be used in the later search. The proposed tree-based method can reduce the computation complexity dramatically. For searching a 13 × 13 kernel for all orientation (81 possible orientations) by simple fixed size template matching, it computes $13^2 \times 81 = 13{,}689$ additions. The tree-based-adaptive template matching computes $3^2 \times 5 + (5^2 + 7^2 + 9^2 + 11^2 + 13^2) \times 3 = 1{,}370$ additions for one match.

Figure 6 shows the knife-edge detection results for our test dataset. The white rectangles show the knife-edges detected by the proposed algorithm.

Experiments and Discussion

Experiments on Registration

We conducted some experiments with both the proposed algorithm and the some other methods for comparison purpose,

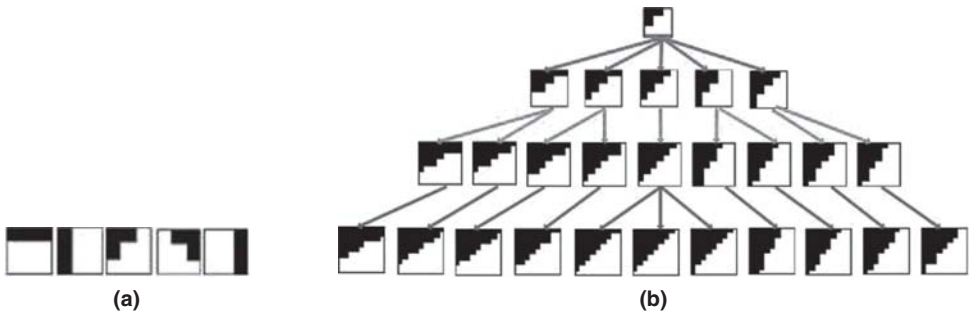


Figure 5. Example of template tree: (a) different directions of template with size 3 × 3, and (b) one branch of the template tree

Figure 6. Detected knife-edge area by adaptive template matching: (a) nadir image, (b) backward image, and (c) forward image.

and the difference maps are computed (see Figure 8). Slight differences between the registered image and the reference image come from the different information provided from each image, but differences to a large extent can be viewed as registration errors, as is shown in Figure 8.

The test dataset is from a typical TLS sensor, in this case a Leica ads40 (Figure 7); as clearly seen in the buildings that these images were captured from different views. We compared our registration method with other existing methods (Hamzah *et al.*, 2010; Kanade and Lucas, 1981; Ogale and Aloimonos, 2007). The results are shown in Figure 8.

In Figure 8, difference maps are calculated by subtracting the nadir image. The difference between the original forward image and nadir image, as well as that between the backward images, are the largest. The registration method attempts to reduce differences as much as possible. Compared to the original difference, the difference after the SAD method is reduced a little, but sharp differences rendered from projective distortion are still apparent; the optical flow method greatly reduces disparities. The differences from projective distortion are lessened, but in the backward and forward images noise is introduced which is not good for super-resolution. Backward and forward images registered by our method are visually almost the same as the nadir image; the differences are slightly different in texture pattern, and they constitute the supplementary information for super-resolution.

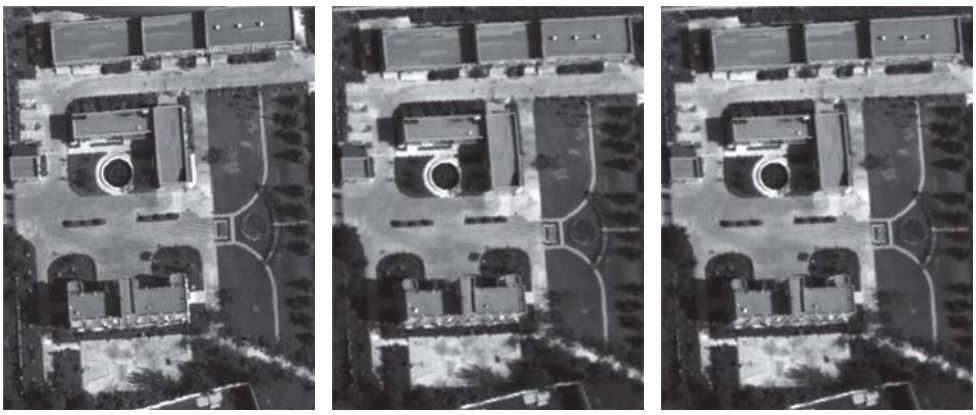
Super-resolution Experiments

The shaking and posture variations in aerial images are much larger than those in satellite images, so the registration of aerial images is much more difficult than spaceborne images. If a registration method produces a good result from airborne images, it will surely perform well for satellite imagery. Therefore, we used images from a typical airborne TLS sensor (ADS40) for experiments.

The dataset in Figure 9 has been down-sampled by a factor of 2 intentionally for evaluation; the original image could be viewed as the ground truth of the high-resolution image. In addition to the proposed method of registration, we include experiments done with other registration methods for comparison; the results are shown in Figure 9. In each row, the first three images are nadir image, registered backward image, registered forward image, and the last one is the super-resolution result, respectively. The first to the fourth row indicate the Keren registration method (Keren *et al.*, 1988), Optical flow method (Kanade and Lucas, 1981, Ogale and Aloimonos, 2007), Sum of Absolution Difference method (Hamzah *et al.*, 2010), and the proposed registration method. Keren registration considers only rotation angle and planar shifts on the X, Y axis; it is a typical registration method for images with rigid transformation. So, 3D-to-2D distortion still exists in the three images. Disparities happen on objects with relatively large elevation differences; the results are not acceptable. The result of optical flow method eliminates some disparities, but due to poor georeferencing, it introduces too much noise. Super-resolution from this registration method is subsequently flawed and unsatisfactory for practical use. Since the SAD method has smoothing effect in discontinuities and is also affected by poor georeferencing, the resulting image after super-resolution indicates no enhancement of resolution, but blurring. Because our algorithm considers the 3D planes of the ground space using orientation data when performing registration and is highly robust by checking correlations, the differences induced from projective distortion are minor. Therefore, the super-resolution results based on the proposed method shows a significant improvement (the fourth row of Figure 9) over the previous three methods in our experiments.

Since the experimental dataset is down-sampled, the original image serves as a ground truth for super-resolution. The closer the results are to the original image, the better performance the corresponding method demonstrates. We adopted the PSNR (Peak Signal to Noise Ratio) to evaluate the similarity between the super-resolution results and the original data, as shown in Table 1. It is obvious that the PSNR of our proposed method is the highest of the four, since the results obtained by other methods were affected by the registration.

After super-resolution, we can see that the proposed registration method produced better a result in both bicubic interpolation and super-resolution. We executed this experiment on the provided data for full resolution, and we interpreted small features to describe resolution enhancement. In Figure 10, there are three small buildings in the original image; each building has a break line in the area specified by a white rectangle. Counting from left to right,



(a) (b) (c)

Figure 7. Test data set from typical TLS sensor: (a) nadir (b) backward, and (c) forward images.

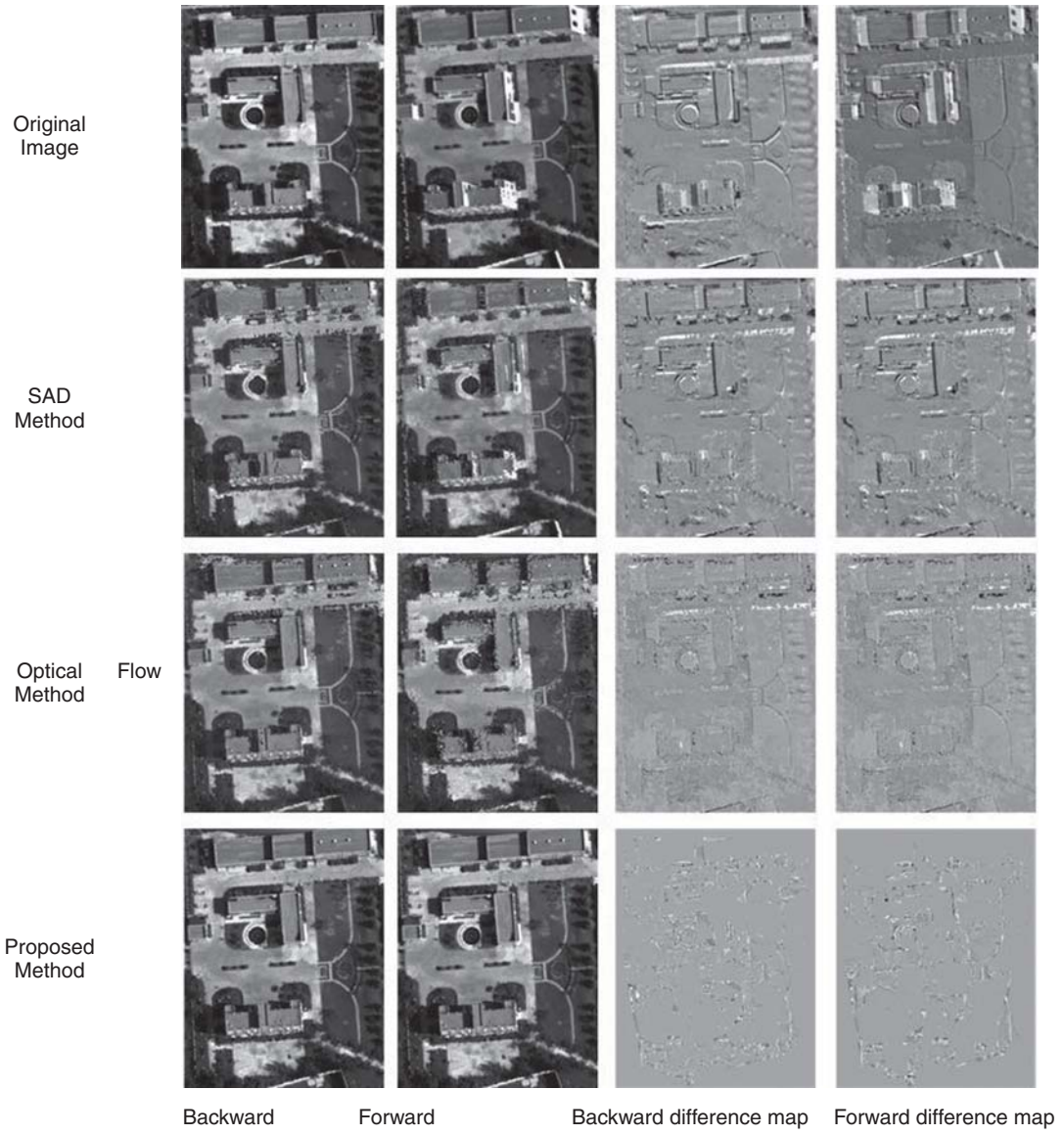


Figure 8. Registration results.

the first building's line is not clear, and is broken at the point where the thick black arrow indicates. Bi-cubic interpolation increases the pixels of the original image, but does not produce any additional information for the scene, so the breaking point still exists. Some scholars assert that the result of de-convolution is the same as super-resolution, but we can see clearly that without the information from the other views, the break point is still present in the zoomed-in image in the third row. The proposed method adopts information from the other two stripes; the resulting image shows a clear sharp edge, where the break point is eliminated.

Another set of experiments is shown in Figure 11. The proposed method enhances the texture patterns of the image by using information from other stripes, while de-convolution just sharpens the edges and enhances noise.

We also applied the proposed method to a larger dataset (with 12,000 * 21,000 pixels) covering hundred's square kilometers; 70 percent of the area was registered and no mismatch was found. Two more sets of experiments are shown in Figure 12 and Figure 13.

The SRKE result gives more information about the roof. In Figure 13 for example, in the small area marked by black circle in Figure 13d, it is flat with no texture, while in the SRKE result (Figure 13f) based on CEM registration, we can see in the same area, there is a black spot. Tracing back to the same area on the backward view and forward view, we can see slight information for the small spot. The result utilizes information from all views and renders a higher resolution image with more information.

Computational Cost

In sum, there are three parts in the whole work flow process: image preprocessing, image registration, and image

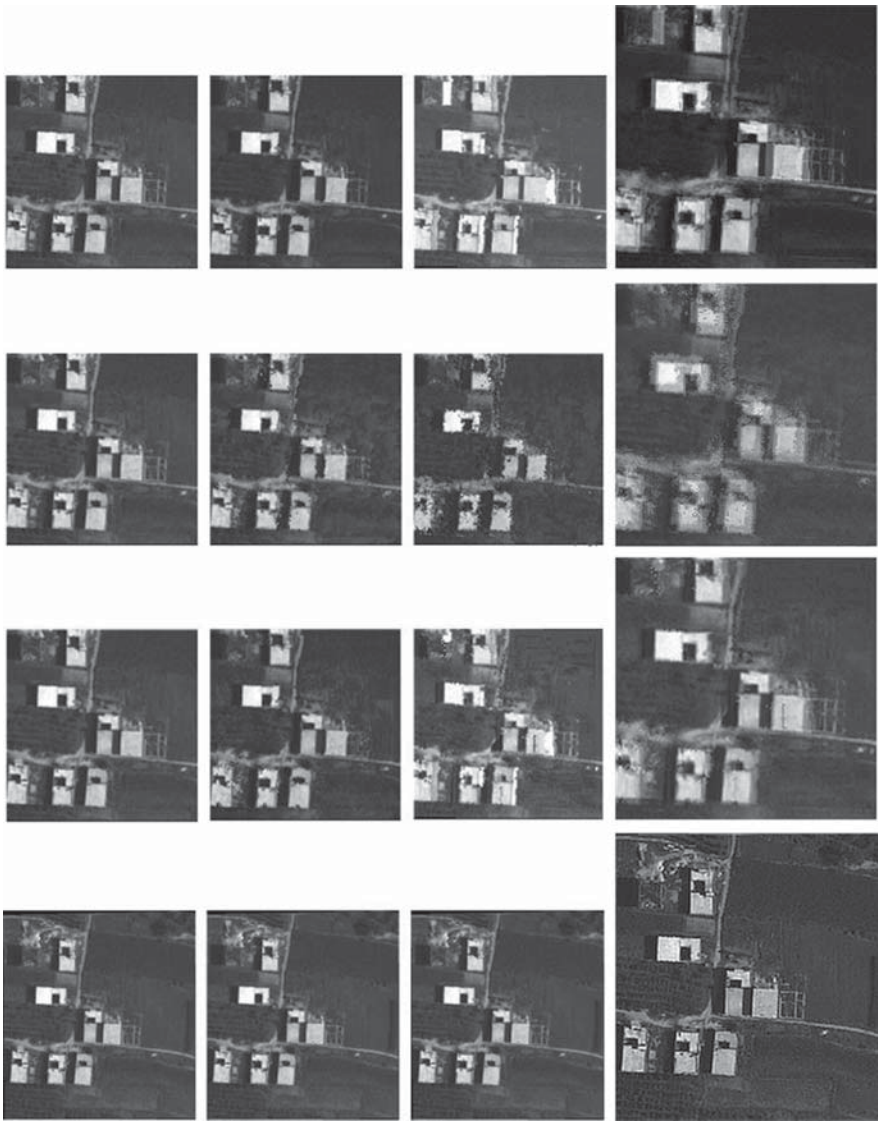


Figure 9. Experiment registration results with proposed and other methods. Each row indicates the registration method and the corresponding super-resolution results (twice as large as the original in size).

Table 1. PSNR of the Super-resolution Result based on Different Registration Method

	Keren	Optical Flow	SAD	The proposed method
Bi-cubic Interpolation	55.0545	58.6481	57.0918	62.9731
Super-resolution	56.9244	53.7460	55.8393	64.6976

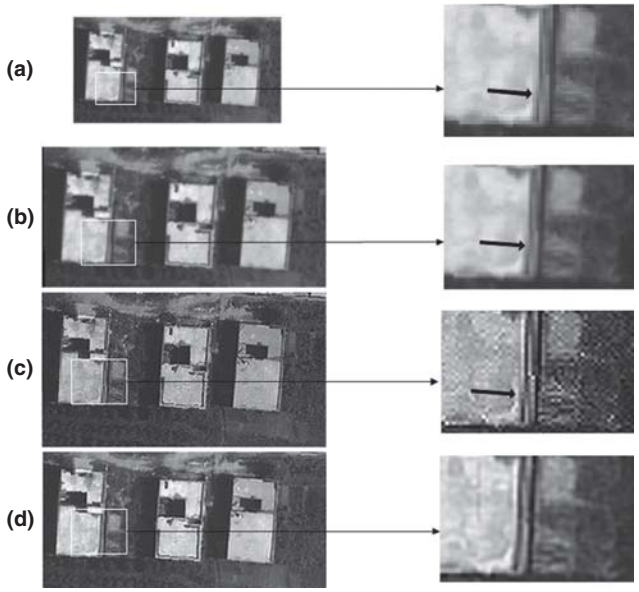


Figure 10. Result comparison in Super-resolution: (a) the original data, (b) bi-cubic interpolation, (c) Wiener deconvolution after bi-cubic interpolation, and (d) the SRKE method.

super-resolution. The raw image rectifier performs correction pixel by pixel. This step requires the most computational resources in image preprocessing. The correction can be performed on small blocks if the distortion is not very large, but for images with large stripes, it can be calculated by distributed systems, reducing the time cost considerably. Adaptive Smooth Filtering can be viewed a process of convolution, it takes the same computational time as ordinary convolution.

Image registration can be divided into three parts: image matching, the region growing process, Delaunay TIN generation, and registration. GC^3 matching takes most of the computational time during the whole process, since it adaptively changes the size and shapes. The search step along the epipolar line is a key factor determining the time the matching procedure will take. A pyramid technique is adopted to generate a coarse elevation for reducing the computation. Another procedure affecting the registration's computation is region growing for estimating the elevation map; it searches each pixel to categorize them. However, it needs only to go over all the pixels once. For machines with supercomputing ability, it is acceptable. For the generation of the TINs and the affine registration for each triangle, it is very efficient.

Image super-resolution adopts the SRKE method, with a two-step iteration to minimize the energy function. Each step needs to solve a linear equation with a number of variables equal to the number of pixels of the resulting image. For a

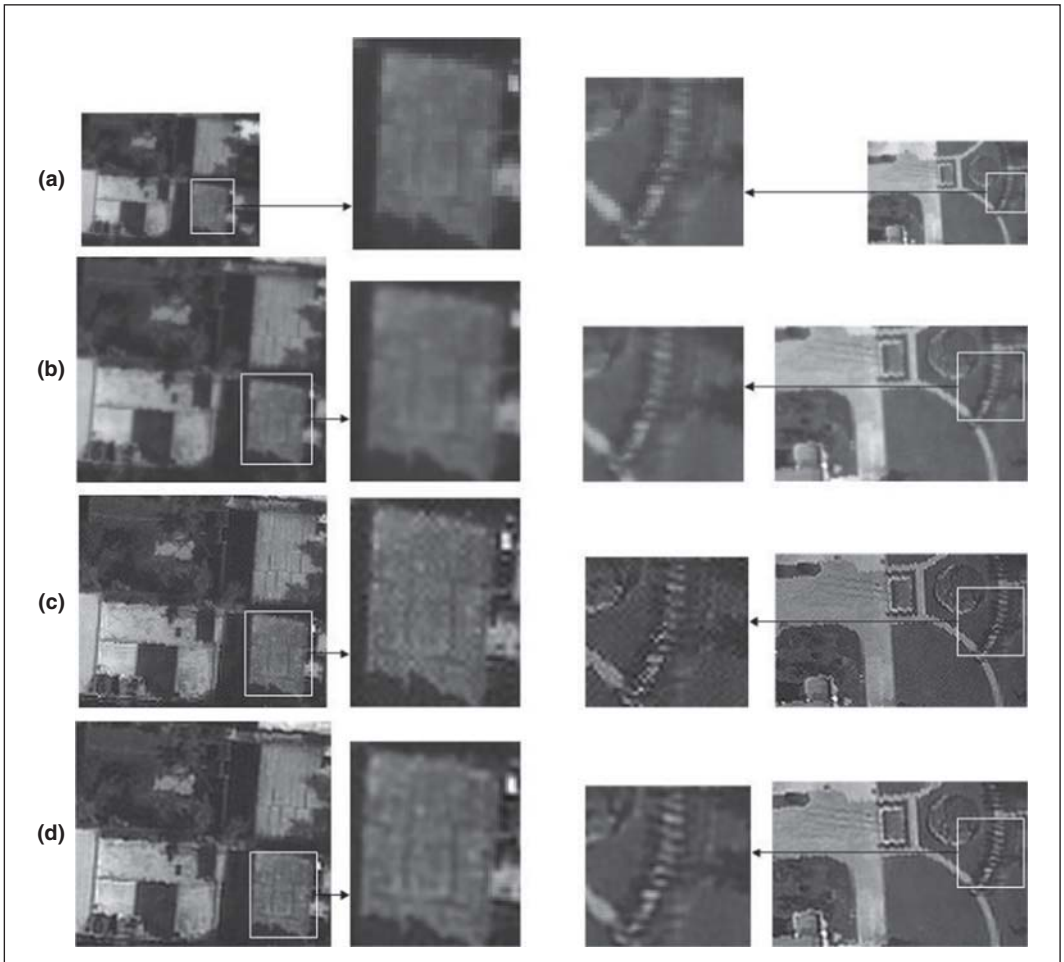


Figure 11. Result of Super-resolution: (a) the original image; (b) Bi-cubic interpolation, (c) the Wiener deconvolution after interpolation, and (d) the SRKE method

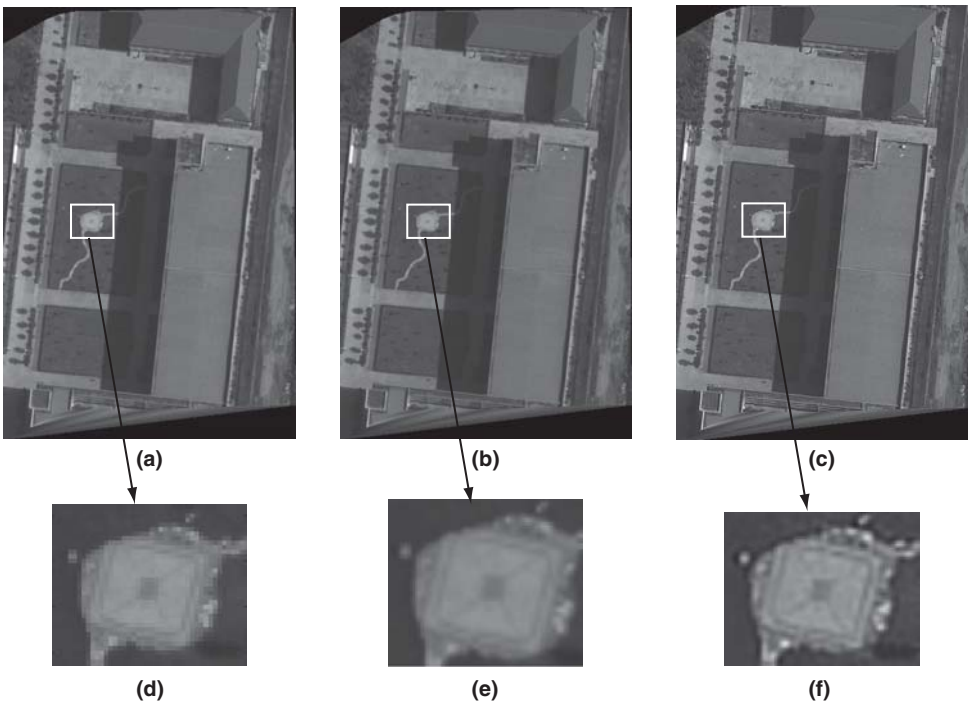


Figure 12. (a) nadir view, (b) bi-cubic interpolation, (c) SRKE result on proposed CEM registration, (d) enlargement of nadir view, (e) enlargement of bi-cubic interpolation, and (f) enlargement of SRKE result

300 × 300 image, it takes about 15 minutes on a normal PC. Fortunately, this problem can be solved by using parallel computing technology, since super-resolution can be performed block by block; hence, computing can be performed on these blocks at the same time.

Conclusions and Drawbacks

Both the CEM based registration results and the according super-resolution results show the advantages of our proposed method. As compared to other stereo registration methods, super-resolution results show visually sharp and clear results. Since there are no universally recognized methods to measure the resolution of an image except by placing ground targets, we judged the resolution by analyzing the texture patterns of the image. In the resulting super-resolution image, some features can be clearly recognized that are barely discernible in the original images. The proposed method is robust. After checking the correlation coefficients, there are few errors affecting the final super-resolution performance.

However, the proposed registration is a partial registration method, which means it may not be able to enhance the resolution in all the parts of the image. Moreover, in our algorithms the triangular TINs may not cover the whole image, especially in areas with dense variation in height (i.e., towers, electrical poles, etc.). In those areas not covered, the TIN will be replaced by the corresponding areas of the nadir image. Therefore, for those areas with no TIN coverage, the result will be equivalent to de-convolution. The more area covered by TINs, the more the resolution of the stripe will be enhanced on average. However, this problem can be solved if we can improve the 3D model of the ground space. The results of segmentation and feature point matching will also affect the final result of the method, and the errors of these results can reduce the registered areas as well.

Therefore, in the future work we will try to create 3D models more accurately, performing registration pixel by pixel based on an improved 3D model. This will allow registered images to be constructed fully from registered pixels, thus reducing computation complexity for finding correspondence.

Acknowledgments

This research work is supported by Science Fund for Creative Research Groups of the National Natural Science Foundation of China (41021061), National Natural Science Foundation of China (41023001), and National Program on Key Basic Research Project (973 Program, 2012CB719906).

References

Adams, R., and L. Bischof, 1994. Seeded region growing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):641–647.

Almeida, M.S.C., and L.B. Almeida, 2010. Blind and semi-blind deblurring of natural images, *IEEE Transactions on Image Processing*, (1):36–52.

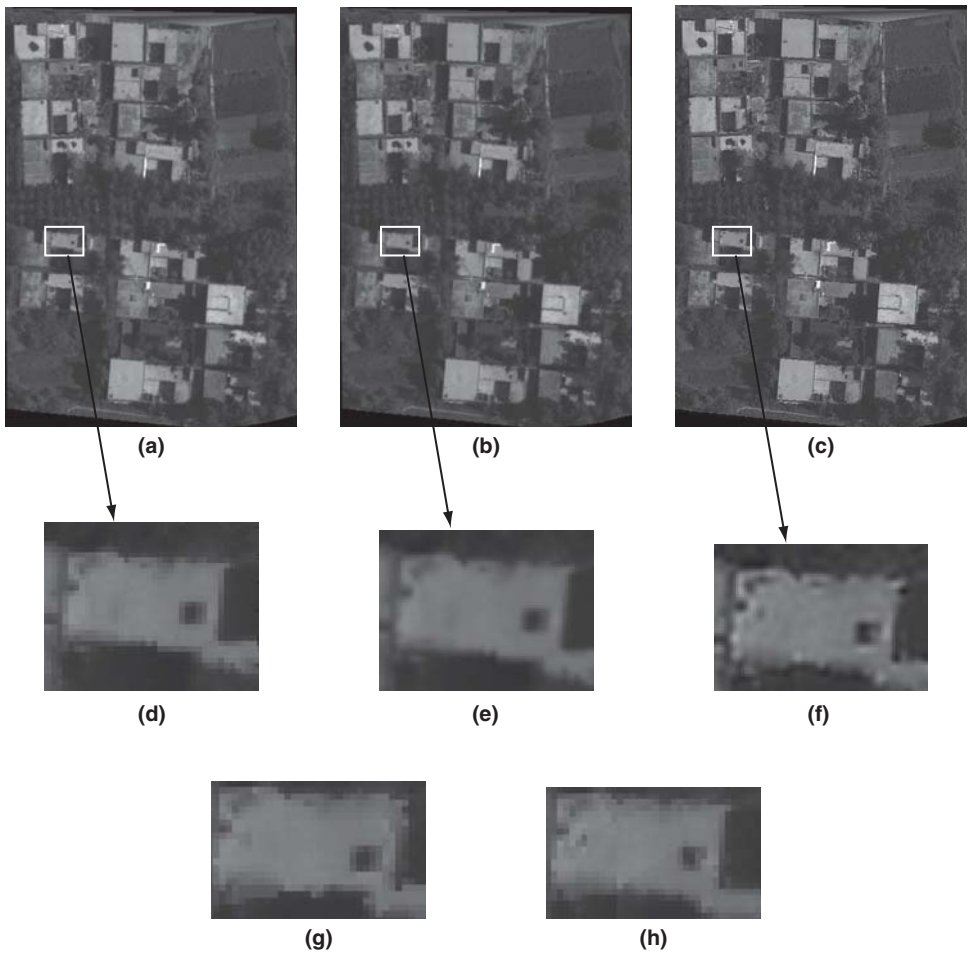


Figure 13. (a) nadir view, (b) bi-cubic interpolation, (c) SRKE result on proposed CEM registration, (d) enlargement of nadir view, (e) enlargement of bi-cubic interpolation, (f) enlargement of SRKE result, (g) the same roof on backward view, and (h) the same roof on forward view

Cristobal, G., 2007. Multiframe blind deconvolution coupled with frame registration and resolution enhancement, *Blind Image Deconvolution: Theory and Applications*, (3):317.

Du, Q., N. Raksuntorn, A. Orduyilmaz, and L.M. Bruce, 2008. Automatic registration and mosaicking for airborne multispectral image sequences, *Photogrammetric Engineering & Remote Sensing*, 74(1):169–181.

Fedorov, D., L.M.G. Fonseca, C. Kenney, and B.S. Manjunath, 2003. Automatic registration and mosaicking system for remotely sensed imagery, *Anais do XI Simpósio Brasileiro de Sensoriamento Remoto. Belo Horizonte: Instituto Nacional de Pesquisas Espaciais*, (3):317–324.

Filip-Sroubek, J.F., 1999. An overview of multichannel image restoration techniques, *Week of Doctoral Students,* Matfyzpress, pp. 580–585.

Fonseca, L.M.G., and B.S. Manjunath, 1996. Registration techniques for multisensor remotely sensed imagery, *Photogrammetric Engineering & Remote Sensing*, (62)10:1049–1056.

Forstner, W., and E. Gulch, 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features, *ISPRS*, Interlaken

Hamzah, R.A., R.A. Rahim, and Z.M. Noh, 2010. Sum of absolute differences algorithm in stereo correspondence problem for stereo matching in computer vision application, *Proceedings of Computer Science and Information Technology (ICCSIT)*, 09-11 July, Chengdu, China.

Kanade, T., and B.D. Lucas,1981. *An Iterative Image Registration Technique with an Application to Stereo Vision*, Carnegie-Mellon University of Pittsburg, Department of Computer Science, Pittsburg, Pennsylvania.

Kanade, T., and M. Okutomi, 1994. A stereo matching algorithm with an adaptive window: Theory and experiment, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (9):920–932.

Keren, D., S. Peleg, and R. Brada, 1988. Image sequence enhancement using sub-pixel displacements, *Computer Vision and Pattern Recognition*.

Li, F., X. Jia, D. Fraser, and A. Lambert, 2010. Super resolution for remote sensing images based on a universal hidden Markov Tree Model, *IEEE Transactions on Geoscience and Remote Sensing*, (11):1270–1278.

Li, Z., and A. Gruen, 2004. Automatic DSM generation from linear array imagery data, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, (1):128–133.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California, Berkeley, Califormia.

Ogale, A.S., and Y. Aloimonos, 2007. A roadmap to the integration of early visual modules, *International Journal of Computer Vision*, (1)9–25.

Qin, R., and J. Gong, 2010. A robust method of calculating point spread function from knife-edge, *Journal of Remote Sensing*, (7):895–900 (In Chinese).

Qin, R., J. Gong, and C. Fan, 2010. Multi-frame image super-resolution based on knife-edges, *Proceedings of the 2010 IEEE 10th International Conference on Signal Processing (ICSP)*, Beijing, China (C. Beijing Jiaotong University).

Saint-Marc, P., J.S. Chen, and G. Medioni, 1991. Adaptive smoothing: A general tool for early vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (4):514–529.

Scharstein, D., and R. Szeliski, 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *International Journal of Computer Vision*, (1):7–42.

Sroubek, F., and J. Flusser, 2003. Multichannel blind iterative image restoration, *IEEE Transactions on Image Processing*, (9):1094–1106.

Tempelmann, U., A.Borner, B. Chaplin, L. Hinsken, B. Mykhalevych, S. Miller, U. Recke, R. Reulke, and R. Uebbing, 2000. Photogrammetric software for the LH Systems ADS40 airborne digital sensor, *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, pp. 552–559.

Vandewalle, P., S. Ssstrunk, and M. Vetterli, 2006. A frequency domain approach to registration of aliased images with application to super-resolution, *EURASIP Journal on Applied Signal Processing*, (3):233–233.

Zhang, L., 2005. *Automatic Digital Surface Model (DSM) Generation from Linear Array Images*, Ph.D Dissertion, Swiss Federal Institute of Technology, Zurich.

Zhang, L., H. Zhang, H. Shen, and P. Li, 2010. A super-resolution reconstruction algorithm for surveillance images, *Signal Processing*, (8):848-859.

(Received 10 May 2012; accepted 22 January 2013; final version 01 March 2013)