

# **AUTOMATIC GEOREFERENCING OF AERIAL IMAGES USING HIGH-RESOLUTION STEREO SATELLITE IMAGES**

**Jaehong Oh**, PhD candidate  
**Charles K. Toth**, Senior Research Scientist  
**Dorota A. Grejner-Brzezinska**, Professor  
Satellite Positioning and Inertial Navigation (SPIN) Laboratory  
The Ohio State University  
470 Hitchcock Hall, 2070 Neil Ave, Columbus, OH 43212  
[oh.174@osu.edu](mailto:oh.174@osu.edu)  
[toth@cfm.ohio-state.edu](mailto:toth@cfm.ohio-state.edu)  
[dbrzezinska@osu.edu](mailto:dbrzezinska@osu.edu)

## **ABSTRACT**

For airborne surveys, direct georeferencing has become the primary source for EOPs (Exterior Orientation Parameters) determination since integrated GPS/INS (Global Positioning System/Inertial Navigation System) systems were introduced. However, there is still need for alternative indirect georeferencing since there are remote, inaccessible areas that lack a geodetic infrastructure and thus GPS/INS-based georeferencing is not feasible. In addition, terrain-referenced navigation is gaining momentum, where the assumption is that no GPS is available. High-resolution satellite images have been globally available and newer high-resolution satellite images offer not only better spatial resolution with decreasing revisit time, but high positional accuracy up to subpixel if ground control is available. Therefore, high-resolution satellite imagery has high potential as a ground control source for aerial image georeferencing and terrain-referenced navigation. Indirect georeferencing of aerial images usually requires accurate 3D ground control points. Unfortunately, ortho-rectified imagery, which is conventionally used as a reference for image-to-image georegistration, contains relief displacement due objects on the ground, resulting in horizontal errors; note that accurate DSM (Digital Surface Model), including terrain and all features on the ground, is usually not available globally. In this study, a high-resolution stereo satellite image-based automatic georeferencing approach is proposed. The use of stereo images can avoid the impact of relief displacement and requires no DSM to obtain ground heights. The matching between aerial and satellite stereo images is based on the SIFT (Scale-Invariant Feature Transform) features, and outliers are pruned utilizing RANSAC (RANdom SAMple Consensus). Finally, to recover 3D ground coordinates, cross correlation matching is performed for epipolar resampled satellite images using FFT (Fast Fourier Transform). An experiment was carried out for a strip of aerial images, including various terrain textures, and showed good potential for the approach.

## **INTRODUCTION**

Digital imaging systems with direct georeferencing have been widely used for mapping. With the advance in semiconductor technology to produce large CCD arrays, aerial digital cameras have been rapidly spreading all over the world. The main advantages of aerial digital cameras over film cameras are a completely digital data flow and improved radiometric image quality with the capability to simultaneously acquire panchromatic, color and near-infrared imagery. High-performance airborne digital sensor systems can generally acquire large amounts of data and provide frequent updates of geospatial image information with the capability of fast and automated digital processing. Highly automated georeferencing is based on integrated GPS/INS systems (Grejner-Brzezinska, 1999). However, GPS signals could be vulnerable to interference such as broadcast television, ultrawide-band communications, over-the-horizon radar and cellular telephones (Carroll, 2001), in which cases indirect georeferencing is needed. Note that navigation-grade INS can maintain position accuracy within about 1–4 km during GPS outages of one hour (Jekeli, 2001), which is certainly inadequate for image georeferencing in mapping. In addition, there are remote, inaccessible areas that lack a geodetic infrastructure and thus GPS/INS-based georeferencing is not feasible: therefore, there is a need for alternative georeferencing.

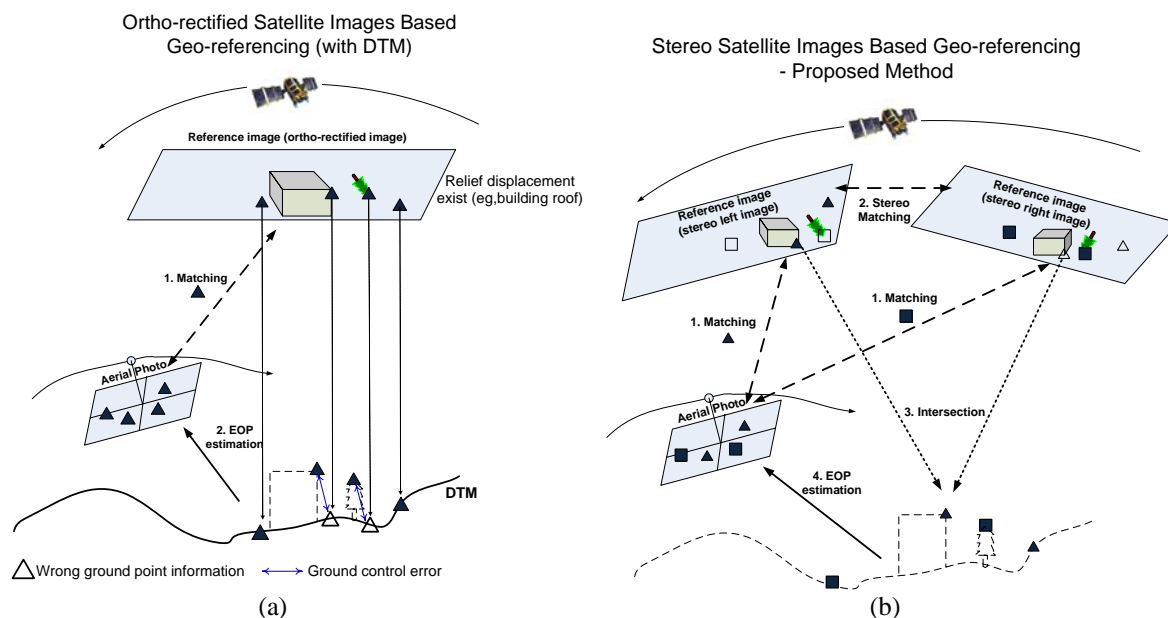
Image-to-image matching can be one of the alternatives for indirect georeferencing that can serve mapping in remote areas as well as terrain-referenced navigation. By image matching, reference images provide newly acquired images with ground control information to obtain georeferencing. Therefore, reference data quality is one of the key

components to achieve accurate georeferencing. Ideal reference data would be images acquired with the same sensor at similar geometry in the similar time and season; obviously it is rarely the case. The general requirements of reference image are high spatial and temporal resolution and high positional accuracy.

High-resolution satellite images have good potential as a ground control source for aerial image georeferencing. Since the first 1m resolution satellite, IKONOS-2, was launched in September 1999, many high-resolution satellite images have been available such as Quickbird (DigitalGlobe, 2001, 60cm), SPOT-5 (SPOT, 2002, 2.5m), OrbView-3 (GeoEye, 2003, 1m), KOMPSAT-2 (KARI, 2006, 1m), EROS-B (ImageSat, 2006, 70cm), WorldView-1 (DigitalGlobe, 2007, 50cm), CARTOSAT-2A (ISRO, 2008, 80cm), GeoEye-1 (GeoEye, 2008, 41cm), WorldView-2 (DigitalGlobe, 2009, 46cm). Moreover, several high performance satellites are scheduled to be launched in the near future such as EROS-C, and GeoEye-2. High-resolution imagery is acquired worldwide with good temporal resolution and is available at good-positional accuracy, up to subpixel level accuracy when bundle adjusted using Ground Control Points (GCP) (Fraser and Ravanbakhsh, 2009). These images have a relatively large swath width, usually more than 10 km, and are used to create worldwide image maps, such as Google Earth. In addition, many images are acquired in multispectral bands, which have advantages in feature extraction.

The image-to-image approach for automatic image co-registration has already been studied (Wong and Clausi, 2007; Le Moigne *et al.*, 2006; Ali and Clausi, 2002); most of the methods are focused on two-dimensional co-registration between low-resolution satellite images based on simple geometric models, such as affine and polynomial transformations. In contrast, accurate aerial image georeferencing requires 3D ground coordinates, which, for example, can be obtained from a combination of ortho-rectified imagery and DSM. However, there are practical limitations to generate and update “true” ortho-rectified imagery for large areas, including the high production cost of DSM with accurate breaklines. Standard ortho-rectified images and DTM (Digital Terrain Model) can be adopted instead, though ground control points from this combination can be highly contaminated by relief displacement caused by objects on the ground, such as buildings and trees.

This study proposes an automatic georeferencing method that is based on using high-resolution stereo satellite images to overcome the requirement of “true” ortho-rectified images and accurate DSM. Figure 1 compares the proposed method to the technique of using a combination of standard ortho-rectified imagery and DTM. Figure 1(a) shows a standard ortho-rectified image and a DTM, as a reference for georeferencing. The image points (solid triangles on both images) are obtained by image matching between the aerial and reference images. Note that wrong ground point information (hollow triangles) is obtained due to relief displacement of the building and tree in the reference image; the DTM does not contain height information of these objects. In contrast, in Figure 1(b), correct ground point coordinates are computed by stereo matching of reference images, and no external ground height information, such as DSM, is required.

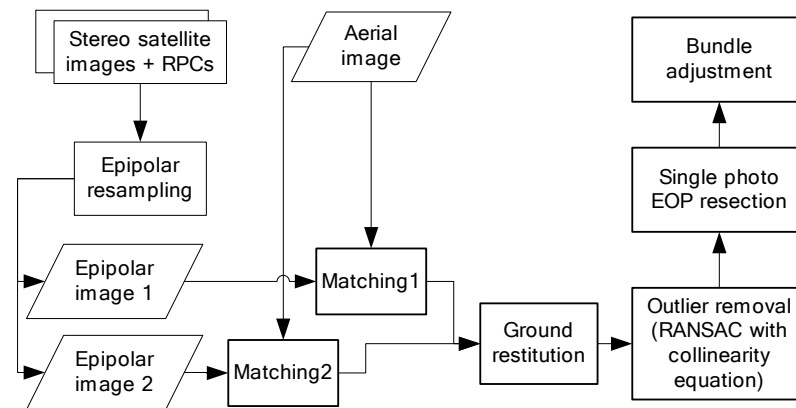


**Figure 1.** Image-to-image matching georeferencing methods using reference data: (a) a standard ortho-rectified image and DTM (b) stereo satellite images.

Another component for successful georeferencing is robust image matching because aerial images and satellite images tend to have large image differences due to differences in sensors used, acquisition time and season, camera angles and so on. Therefore, highly invariant feature matching methods are required. For our work, the popular point feature extraction and matching method, SIFT (Lowe, 1999) was selected because it has been recognized to be very reliable and invariant to imaging condition changes. Note that there are also some modifications to SIFT to make it more effective: PCA-SIFT (Ke and Sukthankar, 2004), GLOH (Gradient Location-Orientation Histogram) (Mikolajczyk and Schmid, 2005), CSIFT (Abdel-Hakim and Farag, 2006), SR-SIFT (Yi *et al.*, 2008), SURF (Speeded-Up Robust Features) (Bay *et al.*, 2008) and Robust SIFT (Li *et al.*, 2009), though, they are conceptually similar. One concern in image matching is the existence of matching outliers. Therefore, in this paper, the combination of SIFT and RANSAC (Fischler and Bolles, 1981) is tested.

Figure 2 depicts the flowchart of the proposed method. Following epipolar image resampling of stereo satellite images based on RPC, aerial images are SIFT-matched to each epipolar resampled image. By stereo matching between the epipolar resampled images, 3D ground coordinates for each matching point can be computed. Since there could be outliers in the matched aerial image points and the corresponding ground coordinates, the outliers are removed using RANSAC with the collinearity equation. Finally, EOPs are estimated using single photo resection. When there are overlaps between aerial images, bundle adjustment is carried out to improve georeferencing performance.

The paper is structured as follows: first, a new algorithm for epipolar resampling of satellite images to support efficient stereo satellite image processing will be briefly introduced, and followed by correlation matching using FFT between stereo satellite images to reconstitute 3D ground coordinates. Second, SIFT based multi-scale image matching between aerial images and satellite images is discussed, including an analysis on invariance and accuracy of SIFT matching. Finally, experimental results of the proposed method are presented, followed by a brief summary.



**Figure 2.** Flowchart of the proposed method for automatic georeferencing of aerial images using stereo satellite images.

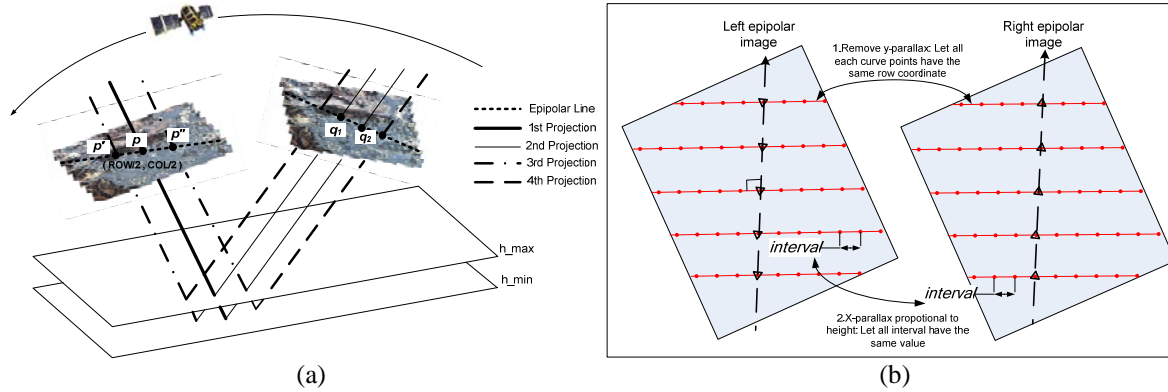
## PROCESSING OF STEREO SATELLITE IMAGES

### Epipolar Curve Pair Determination and Epipolar Image Resampling

Accurate epipolar resampling is pivotal for efficient processing of a stereo satellite image pair that is used as reference data in the study. Unlike frame cameras that have well-known epipolar geometry, the pushbroom camera model, which is adopted for most high-resolution satellite imaging systems, does not produce straight epipolar lines, and no epipolar pair exists for the entire scene (Gupta and Hartley, 1997; Kim, 2000). These properties make it difficult to establish epipolar geometry of pushbroom cameras for accurate epipolar image resampling, which, in general, is essential to confine the search space for efficient and accurate matching between stereo satellite images. In the study, a new epipolar resampling method is briefly introduced exploiting the RPC (Rational Polynomial Coefficients), which is provided by most high-resolution satellite image vendors.

Based on the finding that epipolar curve pairs approximately exist for local areas determined by ground height ranges (Kim, 2000), global epipolar curve pairs are determined piecewise as shown in Figure 3(a). Starting from a point in the left image,  $p$ , for example the center point, an inverse RPC projection to the maximum and minimum ground heights is performed to obtain two ground points. Then, the two ground points are projected to the right image to determine corresponding image points,  $q_1$  and  $q_2$ , by forward RPC projection. The epipolar line between

the two image points,  $q_1$  and  $q_2$ , is approximated by a straight line. Now starting from  $q_1$  and  $q_2$  in the right image, the corresponding two left image points  $p'$  and  $p''$  can be obtained in the same way. The epipolar curve between  $p'$  and  $p''$  and the epipolar curve between  $q_1$  and  $q_2$  can be paired. By continuing the projections, a linear array of image points in an epipolar curve are obtained in the left and right image. They constitute an epipolar curve pair which shows the sensor trajectory direction. In the same way, epipolar curve pairs for the entire image can be generated by selecting start points along the orthogonal direction to the trajectory direction.



**Figure 3.** Epipolar curve pair determination and epipolar image resampling (a) iteratively epipolar curve point determination (b) arrangement of the epipolar curve points to satisfy epipolar image conditions.

The epipolar resampling consists of three steps: (1) piecewise generation of the epipolar curve points, (2) rearrangement of the curve points to satisfy epipolar image conditions, and (3) transformation for image resampling. The x-axis of the resampled image should be aligned along the trajectory and the y-axis should be orthogonal to the trajectory. In addition, there should be no y-parallax, and the x-parallax should be linearly proportional to ground height (Morgan, 2004). Therefore, all piecewise generated points are relocated to meet the epipolar resampled image conditions. As can be seen in Figure 3(b), images are firstly aligned by aligning the start points in the left image (downward triangles) and the corresponding image points in the right image (upward triangles) along the y-axis in the epipolar resampled domain. Removal of y-parallax can be achieved by assigning a constant row coordinate value to each epipolar curve pair in both images. The linear relationship between x-parallax and ground height can be attained if the interval between the epipolar curve points is fixed in both images. It is because the epipolar curve points are computed from the fixed height interval in the proposed piecewise method. The constant interval can be obtained from the mean interpoint distance (between neighboring points along the epipolar curve) in both images.

### Stereo Matching Between Satellite Images to Restitute 3D Ground Coordinates

Since the geometric and spectral difference between the along-track stereo satellite images are rather small, the well known normalized cross-correlation can be used for the epipolar resampled images as shown in equation (1).

$$c = \frac{\sum_{i=1}^{r1} \sum_{j=1}^{c1} [(A_{ij} - \bar{A})(B_{ij} - \bar{B})]}{\sqrt{[\sum_{i=1}^{r1} \sum_{j=1}^{c1} (A_{ij} - \bar{A})^2] [\sum_{i=1}^{r1} \sum_{j=1}^{c1} (B_{ij} - \bar{B})^2]}} \quad (1)$$

where,  $\mathbf{A}$  is the template target image, size of  $r1 \times c1$ ,  $\mathbf{B}$  is a subarray of template source image,  $A_{ij}$  is the digital number from image  $\mathbf{A}$  at row  $i$  and column  $j$ ,  $B_{ij}$  is the digital number from image  $\mathbf{B}$  at row  $i$  and column  $j$ ,  $\bar{A}$  is the average of all digital numbers in  $\mathbf{A}$ , and  $\bar{B}$  is the average of all digital numbers in  $\mathbf{B}$ .

Even if the accurate epipolar image resampling significantly reduces the search space, there is still computational load. Therefore, FFT is utilized for fast matching performance. It can be seen that  $\bar{B}$  can be computed using FFT. By

decomposing the numerator of equation (1), we can obtain equation (2). Note that  $\sum_{i=1}^{r1} \sum_{j=1}^{c1} (A_{ij} - \bar{A})$  is computed once for each template target image, and  $\sum_{i=1}^{r1} \sum_{j=1}^{c1} B_{ij} (A_{ij} - \bar{A})$  can be computed using FFT.

$$\sum_{i=1}^{r1} \sum_{j=1}^{c1} [(A_{ij} - \bar{A})(B_{ij} - \bar{B})] = \sum_{i=1}^{r1} \sum_{j=1}^{c1} B_{ij} (A_{ij} - \bar{A}) - \bar{B} \sum_{i=1}^{r1} \sum_{j=1}^{c1} (A_{ij} - \bar{A}) \quad (2)$$

The decomposition of the denominator yields equation (3). Note that  $\sum_{i=1}^{r1} \sum_{j=1}^{c1} (A_{ij} - \bar{A})^2$  is computed once for each template target image, and the two terms,  $\sum_{i=1}^{r1} \sum_{j=1}^{c1} B_{ij}$ ,  $\sum_{i=1}^{r1} \sum_{j=1}^{c1} B_{ij}^2$ , can be computed using FFT.

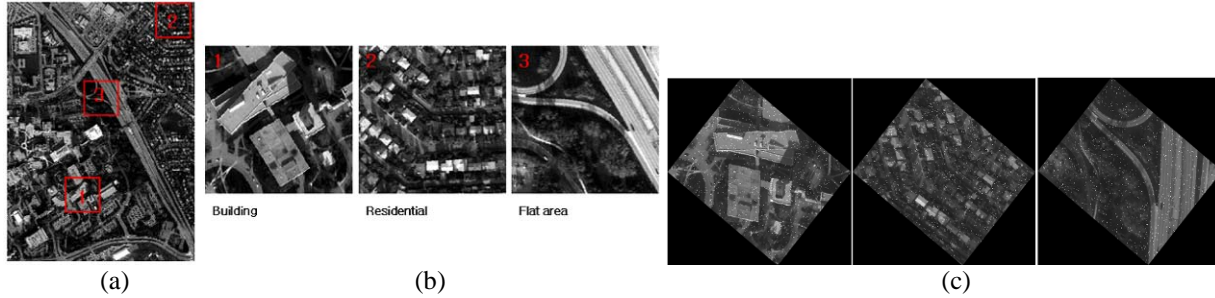
$$\left[ \sum_{i=1}^{r1} \sum_{j=1}^{c1} (A_{ij} - \bar{A})^2 \right] \left[ \sum_{i=1}^{r1} \sum_{j=1}^{c1} (B_{ij} - \bar{B})^2 \right] = \left[ \sum_{i=1}^{r1} \sum_{j=1}^{c1} (A_{ij} - \bar{A})^2 \right] \left[ \sum_{i=1}^{r1} \sum_{j=1}^{c1} B_{ij}^2 - 2\bar{B} \sum_{i=1}^{r1} \sum_{j=1}^{c1} B_{ij} - \sum_{i=1}^{r1} \sum_{j=1}^{c1} \bar{B}^2 \right] \quad (3)$$

## SIFT-BASED IMAGE MATCHING BETWEEN AERIAL AND SATELLITE IMAGES

### SIFT Matching Performance Analysis

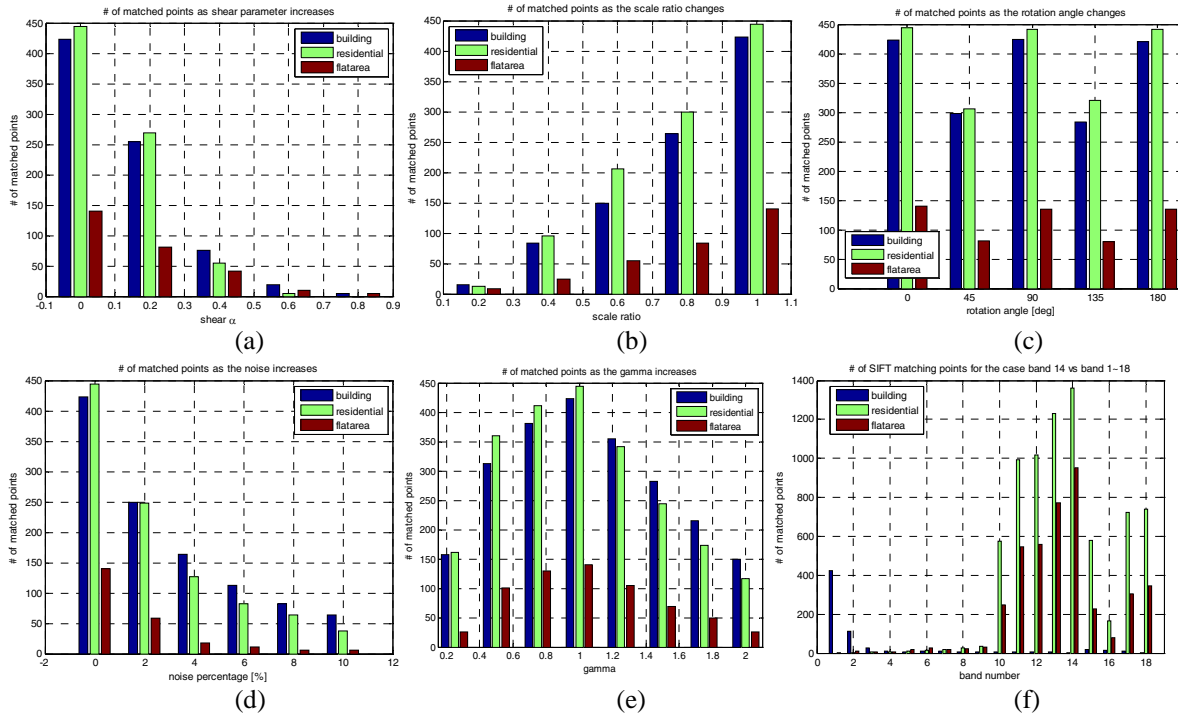
Robust and accurate image matching is critical due to significant differences between aerial and satellite images. Before using SIFT in this study, matching performance is analyzed for simulated image differences such as shear distortion, scale, rotation, noise, intensity, and spectra difference. Three test images (Figure 4(b)), building, residential, and flat area, are a subset from a 1m resolution aerial hyperspectral aerial image that has 18 spectral bands (Figure 4(a)), and image differences are simulated (Figure 4(c)). The reason for testing a hyperspectral image is because its abundant spectral bands enable the analysis of the impact of spectral differences on matching performance.

Matching between the simulated images (Figure 4(c)) and the original image (Figure 4(a)) is performed. Figure 5 shows the number of matching points as a function of changing simulation variables. The value 0.6 is used as the SIFT matching criterion, which means that the smallest vector angle in radians between normalized SIFT feature descriptor vectors is 0.6 times less than the second smallest vector angle. Figure 5(a) shows that as the shear distortion ( $\alpha$ ) increases, the number of matching points decreases. Note that SIFT matching seems not to overcome large distortions. Figure 5(b) shows that a large image resolution difference significantly affects the matching result since smaller scale parameters (larger scale differences) tend to yield fewer matching points. In Figure 5(c), it is observed that the same number of matching points is acquired when rotating the image by 90° and 180°, though not 45° and 135°. The probable reason is because 45° and 135° rotations require image resampling while 90° and 180° do not. From Figure 5(d) it can be seen that the SIFT matching tends to not produce enough matching points for the image with much noise. Figure 5(e) shows that the number of matching points decreases as the intensity difference increases (gamma 1 represents no intensity difference). Figure 5(f) is when the matching is performed between spectral band 14 and the other bands of the hyperspectral image. Considering bands from 1 to 10 are visible spectrums, matching between the visible spectrum and near-infrared spectrum seems not to produce many matching points, and the number of matching points decreases significantly as the spectral gap increases, i.e., the SIFT seems not to overcome a large spectral gap.

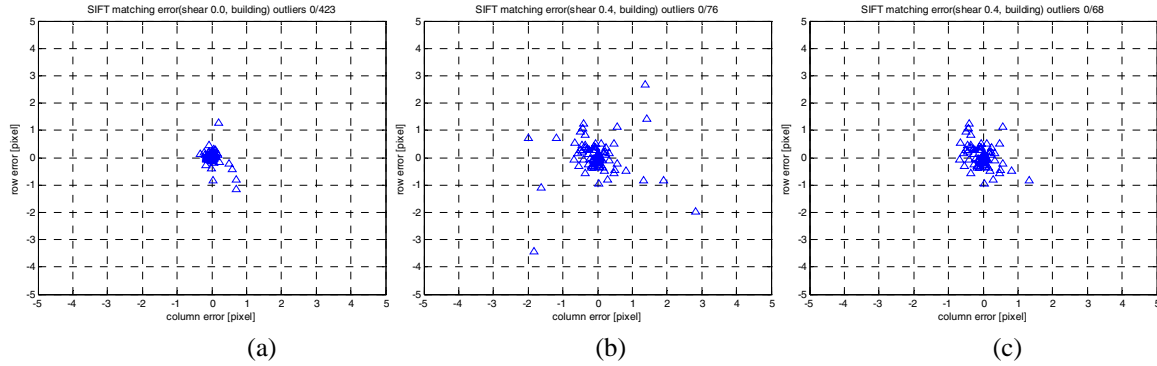


**Figure 4.** Test images by CASI-1500 (ITRES Research) (a) full scene (a) subset scenes (b) simulated images using scale difference: 0.8, shear: 0.2, rotation  $45^\circ$  and noise 1%.

Following the invariance test, matching accuracy is investigated. Figure 6 presents the matching accuracy of the building region as a function of the shear parameter. Figure 6(a) shows the accuracy when there is no image distortion. Ideally, the matching error should be zero for the no-distortion case, but the test showed that the errors for some points are not zero. The likely reason is the multi-scale approach of the SIFT. Similar results can be shown when an image is matched to a subset of the image (when an image is matched to itself, there should be no error). A number of low-accuracy matching points are observed in Figure 6(b). Therefore, RANSAC is tested for outlier removal. In RANSAC, model parameters are estimated from a randomly selected observation set and then every observation is tested if it fits the model and is added to the consensus set if it does. Through iterations, a new consensus set is obtained and a better model is estimated. Table 6(c) shows the outlier removal test results using RANSAC with the affine model. Note that low accuracy matching points were pruned.



**Figure 5.** SIFT invariance test results: number of matching points as changing (a) shear (b) scale difference (c) rotation (d) noise (salt and pepper) (e) image stretch (f) spectral band 14 vs. the others.

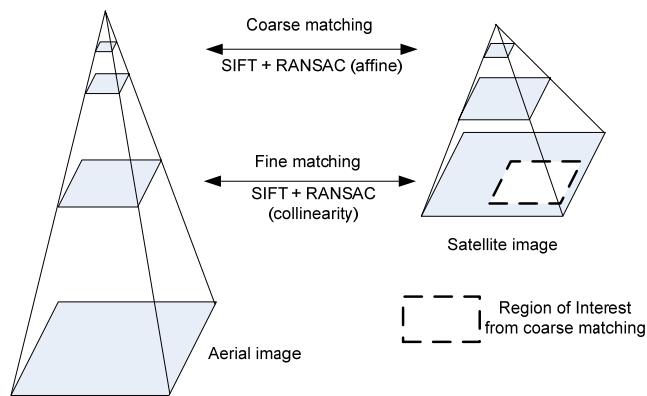


**Figure 6.** SIFT accuracy test results: column and row errors (a) shear 0.0 (b) shear 0.4 (c) shear 0.4 with outlier removal using RANSAC.

### SIFT Based Multi-scale Image Matching

In addition to robustness and accuracy of image matching, efficiency is also an important criterion for practical application since images acquired for mapping purposes or terrain-referenced navigation are typically large in size. Therefore, a multi-scale approach consisting of coarse and fine image matching between aerial and satellite images is used as shown in Figure 7.

By utilizing coarse matching, an aerial image is approximately located and coregistered to the satellite images, and the region of interests (ROI) is obtained in satellite images to which the aerial image will be fine-matched. Coarse matching utilizes down-sampled images; a Gaussian image pyramid is generated from each aerial and satellite image to attenuate aliasing effects (Fosyth and Ponce, 2000). Gaussian down-sampled images of similar spatial resolution are matched to each other, and outliers in the matching results are removed using RANSAC. Since the down-sampled images are often low-resolution, not significantly affected by relief displacement, a 2D affine model should be enough to model the transformation. Fine matching is performed between the aerial image and subset satellite image (each stereo image) to obtain control points from the satellite images. Note that aerial images, which usually have higher spatial resolution, need to be down-sampled close to the resolution of satellite images before matching, since resolution difference is one of the major factors affecting matching performance. For example, if the aerial image has 25cm resolution and satellite images are of 1m resolution, pyramid level 3 of the aerial image is selected. This pyramid image generation has the advantage of significantly reducing computational load. In contrast to coarse matching, the outliers in the matching results should be suppressed using a rigorous model, such as the collinearity equation, exploiting well-calibrated camera information because relief displacement needs to be handled.



**Figure 7.** Multi-scale image matching approach between an aerial image and satellite images.

## EXPERIMENT

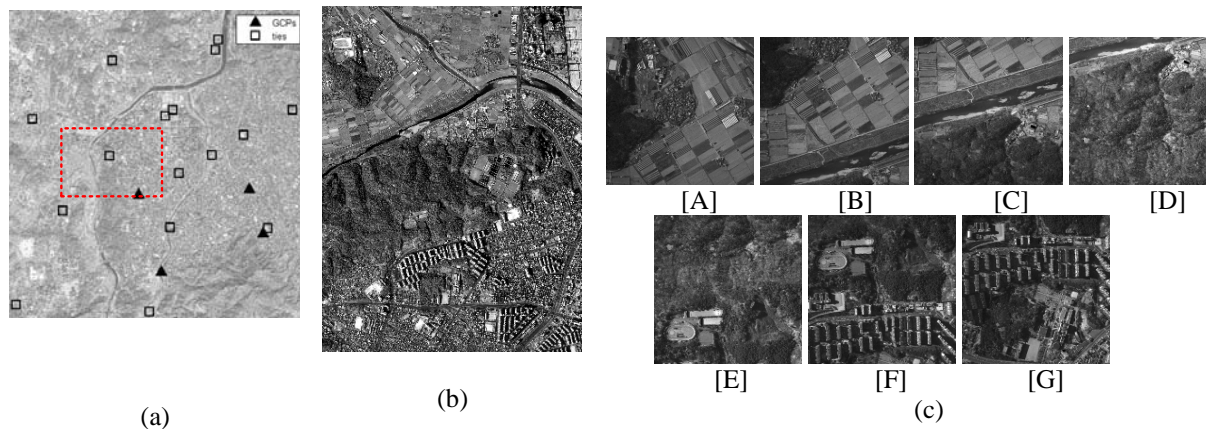
### Data Specification and Reference Image Preparation

Table 1 presents the test data specification of aerial and IKONOS stereo images used in our tests, and Figure 8 shows the test area satellite and aerial images. Note that the aerial images were acquired in June 2003 and the satellite images in November 2001. Due to time and seasonal gap, there are large seasonal and terrain differences between the images.

Since IKONOS RPCs have relatively poor accuracy, the IKONOS RPCs were refined by estimating only shift terms in the RPCs error adjustment model (Dial and Grodecki, 2002; Fraser and Hanley, 2005) using four GCPs shown in Figure 8(a), note that Toutin (2006) showed that IKONOS RPCs have similar accuracy with Toutin's 3D physical model when refined with a shift, while Quickbird RPCs need to be refined with linear functions due to its relief dependency. Image coordinate residuals after adjustment were at one-pixel level. Following the RPC adjustment, epipolar image resampling was performed based on the proposed piecewise method, and y-parallax in the epipolar resampled images was at one-pixel level. Ground restitution residuals of the epipolar resampled images for four GCPs were horizontal and vertical 1 meter. Figure 8(b) presents a subset image of the left epipolar resampled image as indicated in Figure 8(a). A strip of aerial images containing various ground textures including open field, forest and buildings, was selected as shown in Figure 8(c). Since the matching algorithm is based on a single image band, it is important to select image bands having similar spectral ranges as the image matching seems unlikely to overcome high spectral gaps shown in Figure 5(f). Therefore, the red band in the aerial image was selected because the IKONOS relative spectral response (GeoEye, 2008) shows that the red band has the best overlap with panchromatic spectra among the color bands.

**Table 1.** Test data specification.

Data	Site	Date	Focal length	Resolution	Spectral
IKONOS, Level2 Stereo	Daejeon, Korea	Nov 2001	10m	1m	Pan
Aerial images	Daejeon, Korea	Jun 2003	55mm	25cm	Color



**Figure 8.** Test data (a) IKONOS image and GCP distribution (b) subset epipolar resampled IKONOS image (c) aerial images [A][B] shows fields, [D][E] are forests, [G] contain buildings, [C][F] are mixed area.

### SIFT Based Multi-scale Image Matching

The initial localization for images [A], [D], and [G] by coarse matching using Gaussian down-scaled images is presented in Table 2; Gaussian images of spatial resolution of 2m were used for image matching. Coarse matching is based on SIFT + RANSAC using an affine model. Table 2 shows that regions of interest could be obtained successfully; note that the IKONOS images are epipolar resampled.

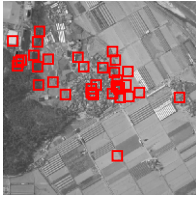

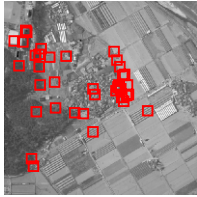
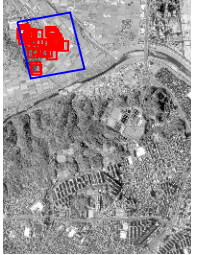
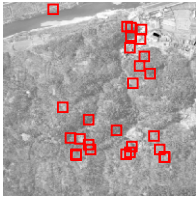

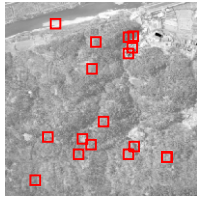

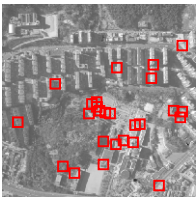
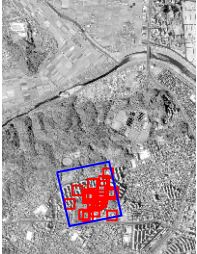

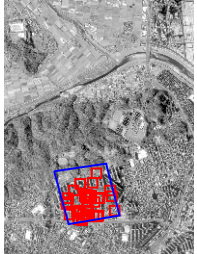
Based on the region of interest from the coarse matching, subsets of the IKONOS images with a margin were used for fine matching. Since the spatial resolution of IKONOS and aerial images are 1m and 25cm respectively, the aerial images were downscaled to 1m for efficient matching.



Table 3 shows the matching results where rectangles are the SIFT matching results. A 3D ground coordinate for each rectangular point was computed by IKONOS stereo matching (cross-correlation matching with FFT). Then, the ground coordinates and corresponding aerial image points were refined from RANSAC with the collinearity equation, shown in

Table 3 by triangles. The images [A] and [G] show that a decent number of points could be obtained over the image. However, matching points could hardly be obtained over the forest in image [D] due to nonexistence of unique features and large seasonal difference.

**Table 2.** Coarse matching results for images [A], [D], and [G].

	Aerial image	Left IKONOS image	Aerial image	Right IKONOS image
[A]	aerial photo  row [pixel] column [pixel]	 row [pixel] column [pixel]	aerial photo  row [pixel] column [pixel]	 row [pixel] column [pixel]
[D]	aerial photo  row [pixel] column [pixel]	 row [pixel] column [pixel]	aerial photo  row [pixel] column [pixel]	 row [pixel] column [pixel]
[G]	aerial photo  row [pixel] column [pixel]	 row [pixel] column [pixel]	aerial photo  row [pixel] column [pixel]	 row [pixel] column [pixel]

**Table 3.** Fine matching results for images [A], [D], and [G].

	Aerial image	Left IKONOS image	Right IKONOS image
[A]	<p>row [pixel]</p> <p>column [pixel]</p> <p>□ all matched points △ refined with RANSAC</p>	<p>row [pixel]</p>	<p>row [pixel]</p>
[D]	<p>row [pixel]</p> <p>column [pixel]</p>	<p>row [pixel]</p>	<p>row [pixel]</p>
[G]	<p>row [pixel]</p> <p>column [pixel]</p>	<p>row [pixel]</p>	<p>row [pixel]</p>

### Georeferencing and Accuracy Assessment

First, single photo resection was tested for EOPs estimation. Before real experiments, a simulation was performed for the same camera specification at a similar altitude (1250m) to check how accurate EOPs can be estimated based on 1m-high-resolution satellite images. Nine GCPs were evenly distributed over the entire aerial image, and 1m and 1 pixel of random measurement errors were added to 3D ground coordinates and image coordinates. Several single photo resection tests indicated that the trajectory position and attitude errors are bounded to less than 10 meters and sub-degree level, respectively. Height and heading were better estimated than the others, as the X or Y components seemed to be affected by ground/image coordinate errors.

Following the simulation, the experiment continued by estimating EOPs from the ground control points obtained by image matching. The EOPs were compared to reference EOPs known from bundle adjustment using accurate ground control points, and the differences are presented in

Table 4. Compared to the forest area ([C], [D] and [E]), the open field and building areas show better estimation performance. Estimation for image [D] is the worst because, as shown in

Table 3, there are not enough invariant features in the forest area due to the large seasonal gap between the aerial and IKONOS images. Most aerial images, except the forest areas, show EOPs positional accuracy of less than 20 meters; note that the height was relatively well estimated. The reason that positional accuracy is worse than the simulation is because there are more error sources in the real data set, such as inaccurate matching. In terms of attitude estimation performance, sub-degree accuracy could be achieved in most cases, with yaw (heading) angles showing the best accuracy.

**Table 4.** Difference between Single Photo Resection and reference EOPs.

EOPs difference	X [m]	Y [m]	Z [m]	Roll [degrees]	Pitch [degrees]	Yaw [degrees]
[A] Field	-11.53	-1.81	-4.68	0.0665	-0.5333	0.1089
[B] Field	1.20	-0.75	-4.06	0.0188	0.0632	0.0818
[C] Field + Forest	66.61	-13.58	-38.00	0.5255	3.4157	-0.5450
[D] Forest	534.11	62.75	-603.52	-4.7506	40.0899	-2.7743
[E] Forest	-44.21	18.48	-4.46	-1.0249	-2.2213	0.3907
[F] Forest + building	8.26	4.68	2.70	-0.2552	0.4523	0.0361
[G] Building	4.55	-13.87	-2.02	0.6786	0.3217	-0.0854

Second, a bundle adjustment was performed by generating tie points between adjacent aerial images taking advantage of image overlap. Table 5 lists EOP estimation errors. Note that EOP accuracy of the forest area images ([C], [D], and [E]) significantly improved. However, the EOPs for some images are a little deteriorated, most likely due to the fact that the EOP errors of the forest area images are distributed to the adjacent images.

**Table 5.** Difference between Bundle Adjustment (without outlier removal) and reference EOPs.

EOPs difference	X [m]	Y [m]	Z [m]	Roll [degrees]	Pitch [degrees]	Yaw [degrees]
[A] Field	-2.03	-6.87	-6.81	0.3361	-0.0765	0.1444
[B] Field	-3.72	-6.38	-6.33	0.3520	-0.1225	0.1250
[C] Field + Forest	-5.96	-8.79	-5.62	0.4921	-0.1929	0.0777
[D] Forest	-5.47	-6.91	-4.30	0.4151	-0.1537	0.0063
[E] Forest	-9.30	-5.64	-2.54	0.3402	-0.3279	-0.0632
[F] Forest + building	-9.25	-7.10	-0.64	0.3851	-0.3321	-0.0940
[G] Building	-8.44	-7.56	0.97	0.3745	-0.2943	-0.1198

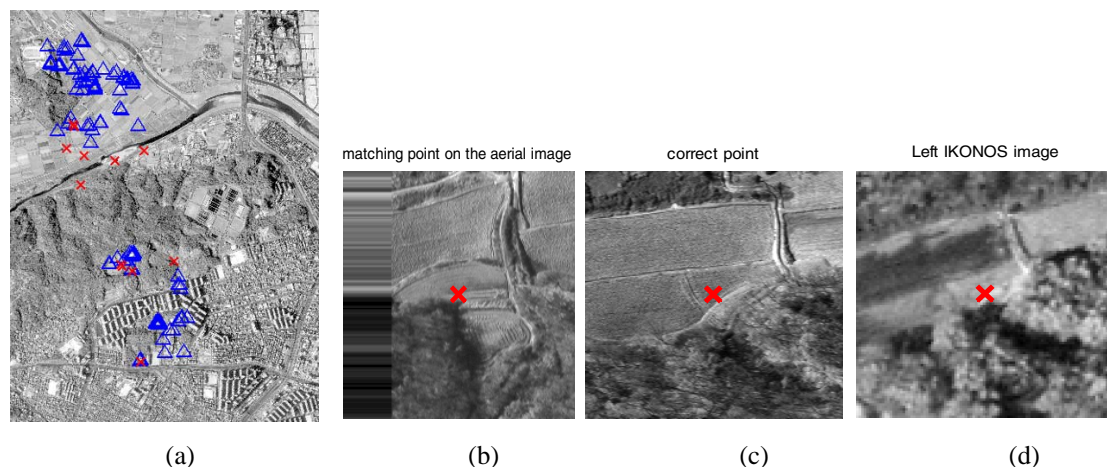
Next, a bundle adjustment with outlier removal was carried out using Baarda's data snooping method (Baarda, 1968). Figure 9(a) presents which ground controls were removed as outliers. Note that most of them are from the aerial images [C], [D], and [E] that are forest areas. Since a small number of matching points was obtained over the forest area, no redundant ground controls were available to successfully refine control points in the matching process. Figure 9(b) depicts a removed matching point as outlier in aerial image [D], while Figure 9(c) and Figure 9(d) show the correct position on the aerial image [D] and the left IKONOS image, respectively. Table 6 shows the EOPs accuracy. Note that the flight direction is the X-direction. Even though there is a strong correlation between X and pitch, and between Y and roll, EOPs accuracy improved and is less than 7 meters. The building area images ([F] and [G]) show fairly accurate Z values. High ground height variation in the building area seems to lead to relatively accurate Z value estimation. Among the attitude angles, again yaw angle shows the best accuracy.

**Table 6.** Difference between Bundle Adjustment (with outlier removal) and reference EOPs.

EOPs difference	X [m]	Y [m]	Z [m]	Roll [degrees]	Pitch [degrees]	Yaw [degrees]
[A] Field	-6.10	-4.70	-5.05	0.1964	-0.2799	0.0471
[B] Field	-5.22	-3.65	-3.45	0.1590	-0.2243	0.0397
[C] Field + Forest	-5.16	-4.68	-2.39	0.2171	-0.2115	0.0343
[D] Forest	-3.39	-4.51	-1.54	0.2169	-0.1243	0.0217
[E] Forest	-5.72	-3.21	-0.79	0.1561	-0.2096	-0.0063
[F] Forest + building	-4.42	-4.23	0.11	0.2023	-0.1309	-0.0198
[G] Building	-3.76	-4.28	0.53	0.1974	-0.0832	-0.0286

The ground restitution accuracy for tie points was computed and presented in Table 7. Without outlier removal, the bundle adjustment shows horizontal accuracy up to 3.45 meters and vertical accuracy up to 5.77 meters. Considering the existence of outliers, this accuracy is quite good. The effect of the outliers seems to be attenuated by the combined impact of many good ground controls and tie points. In contrast, the bundle adjustment with outlier removal improves horizontal and vertical accuracy to 2.18 and 4.53 meters, respectively. Since 1m resolution stereo images are used as reference, these results are relevant.

The error sources affecting the estimation accuracy can be summarized as: satellite image positional accuracy, epipolar image resampling accuracy, image matching between aerial and satellite images, stereo matching, GCP distribution on the aerial image, and image overlap.



**Figure 9.** Outlier removal of ground controls in the bundle adjustment (blue triangles: refined ground controls, red crosses: removed as outliers); (a) distribution on the IKONOS image (b) a removed outlier point located at the edge of aerial image [D], (c) and (d) show the correct position on the aerial image and IKONOS image, respectively.

**Table 7.** Ground restitution accuracy.

Ground restitution accuracy		X [m]	Y [m]	Z [m]
Bundle adjustment (without outlier removal)	Mean	2.15	0.98	2.77
	Max	3.45	2.69	5.77
Bundle adjustment (with outlier removal)	Mean	1.15	0.30	1.56
	Max	2.18	0.76	4.53

## CONCLUSION

A new concept of automatic aerial image georeferencing using high resolution stereo satellite imagery is proposed. Significant improvements in high-resolution satellite image specification, including high spatial and temporal resolution, good positional accuracy and large swath width, motivated the idea of using satellite imagery as a reference for image-to-image based indirect georeferencing. For aerial image georeferencing that requires accurate 3D ground coordinates, stereo satellite images are used as reference to provide 3D ground coordinates. A multi-scale image matching scheme, including coarse and fine matching, is adopted for efficient matching between aerial and satellite images. For both matching stages, a combination of SIFT and RANSAC is utilized to obtain accurate matching points. Cross-correlation matching, implemented by FFT, is utilized to compute ground coordinates from satellite stereo pairs.

Following the SIFT matching analysis on invariance and accuracy, an experiment is carried out for a strip of aerial images and IKONOS stereo images. Test results showed that good matching results could be obtained over open and built-up areas; however, the matching was poor over forest areas, especially when there is a large seasonal difference between the aerial image and the reference. EOPs of aerial images are estimated from ground control information acquired from the matching points. Both single photo resection and bundle adjustments were tested. Using the bundle adjustment approach with blunder removal, accurate EOPs could be obtained and the ground restitution accuracy could be improved up to the level that can be expected from the positional accuracy of stereo satellite images.

Even though georeferencing performance is highly subjected to many error sources, including reference data positional accuracy and matching quality, the high potential of stereo satellite images as reference data has been confirmed. The approach could be used not only for aerial image georeferencing in mapping at small scale, but also provides a good basis for robust navigation under GPS signal loss (in post-processing). One important issue is that the outlier removal by RANSAC might yield slightly different results at each run due to the nature how RANSAC draws samples randomly from the entire dataset.

## REFERENCES

- Abdel-Hakim, A.E., and A.A. Farag, 2006. CSIFT: A SIFT Descriptor with Color Invariant Characteristics, In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.1978-1983.
- Ali, M., and D. Clausi. 2002. Automatic registration of SAR and visible band remote sensing images, In: *Proceedings of IEEE International Geoscience and Remote Sensing*, Symp. Vol.3, pp.1331-1333.
- Baarda, W., 1968. *A testing procedure for use in geodetic networks*, Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 2, No. 5, Delft, 1968.
- Bay, H., A. Ess, T. Tuytelaars, L.V. Gool, 2008. SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding (CVIU)*, 110(3): 346-359.
- Carroll, J., 2001. Vulnerability assessment of the transportation infrastructure relying on the global positioning system, Technical report, Volpe National Transportation Systems Center, August 2001, Report for US Department of Transportation.
- Dial, G., and J. Grodecki, 2002. Block adjustment with rational polynomial camera models, In: *Proceedings of the ACSM-ASPRS 2002 Annual Conference*, 22-26 April, Washington DC, unpaginated CD-ROM.
- Fischler, M. A. and R. C. Bolles, 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of the ACM*. 24: 381-395
- Fosyth, D., and J. Ponce, 2000. *Computer Vision: A Modern Approach*, Prentice Hall, p.154.
- Fraser, C.S., and H.B. Hanley, 2005. Bias-compensated RPCs for sensor orientation of high-resolution satellite imagery, *Photogrammetric Engineering & Remote Sensing*, 71(8):909-915.
- Fraser, C.S., and M. Ravanbakhsh, 2009. Georeferencing accuracy of GeoEye-1 imagery, *Photogrammetric Engineering & Remote Sensing*, 75(6):634-638.
- GeoEye, Inc. 2008. IKONOS Relative Spectral Response, <http://www.geoeye.com/CorpSite/resource/white-papers.aspx>.
- Grejner-Brzezinska, D.A., 1999. Direct exterior orientation of airborne imagery with GPS/INS system: Performance analysis, *Navigation*, 46(4):261-270.
- Gupta, R., and R. Hartley, 1997. Linear Pushbroom Cameras, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9), pp. 963-975.
- Jekeli, C., 2001. *Inertial Navigation Systems with Geodetic Applications*, Walter de Gruyter, p.102.
- Ke, Y and R. Sukthankar, 2004. PCA-SIFT: A more distinctive representation for local image descriptors, In: *Proceedings International Conferences on Computer Vision*, Washington DC, 2004, pp.506-513.
- Kim, T., 2000. A Study on the epipolarity of linear pushbroom images, *Photogrammetric Engineering & Remote Sensing*, 66(8), pp.961-966.
- Le Moigne, J., A. Cole-Rhodes, R. Eastman, P. Jain, A. Joshua, N. Memarsadeghi, D. Mount, N. Netanyahu, J. Morissette, and E. Uko-Ozoro, 2006, Image Registration and Fusion Studies for the Integration of Multiple Remote Sensing Data, In: *Proceedings of 2006 IEEE International Conference on Acoustics, Speech and Signal Processing*, 14-19 May 2006. Vol.5, pp.1189-1192.
- Li, Qiaoliang, Guoyou Wang, Jianguo Liu, and Shaobo Chen, 2009. Robust Scale-Invariant Feature Matching for Remote Sensing Image Registration, *IEEE Geoscience and Remote Sensing Letters*, 6(2):287-291.
- Lowe, D.G., 1999. Object recognition from local scale-invariant features, In: *Proceedings International Conferences on Computer Vision*, Corfu, Greece, pp.1150-1157.
- Mikolajczyk, K., and C. Schmid, 2005. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 27, No. 10, pp.1615-1630, Oct. 2005.
- Morgan, M., 2004. *Epipolar Resampling of Linear Array Scanner Scenes*, Ph.D. dissertation, University of Calgary, Alberta, Canada, 187 p.
- Toutin, T., 2006. Comparison of 3D Physical And Empirical Models for Generating DSMs From Stereo HR Images, *Photogrammetric Engineering & Remote Sensing*, 72(5), pp.597-604.
- Wong, A., and D.A. Clausi, 2007, ARRSI: Automatic Registration of Remote-Sensing Images, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 45, No. 5, pp.1483-1493.
- Yi, Z., C. Zhiguo, and X. Yang, 2008. Multi-spectral remote image registration based on SIFT, *Electron. Lett.*, Vol. 44, No.2, pp.107-108.