ON STEREO-RECTIFICATION OF PUSHBROOM IMAGES

C. de Franchis⋆ *E. Meinhardt-Llopis*⋆ *J. Michel*† *J.-M. Morel*⋆ *G. Facciolo*⋆

⋆ CMLA, Ecole normale supérieure de Cachan, France
† CNES - DCT/SI/AP, France

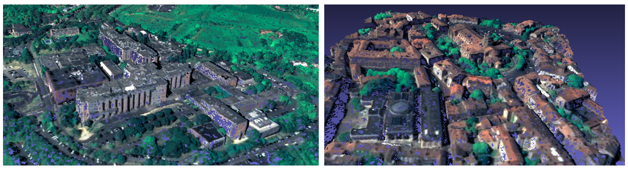


Fig. 1: 3D point clouds automatically generated from Pléiades tri-stereo datasets, without any manual intervention, with the *s2p stereo pipeline* which is available online [1] on the IPOL website.

ABSTRACT

Image stereo pairs obtained from pinhole cameras can be stereo-rectified, thus permitting to test and use the many standard stereo matching algorithms of the literature. Yet, it is well-known that pushbroom Earth observation satellites produce image pairs that are not stereo-rectifiable. Nevertheless, we show that by a new and adequate use of the satellite calibration data, one can perform a precise local stereo-rectification of large Earth images. Based on this we built a fully automatic 3D reconstruction chain for the new Pléiades Earth observation satellite. It produces 1/10 pixel accurate Earth image stereo pairs at a high resolution. Examples will be made available online to the computer vision community.

***Index Terms*—** pushbroom, stereo-rectification, epipolar, remote sensing, Pléiades satellite

1. INTRODUCTION

Stereo-rectification is a technique that permits to simplify the computation of image point correspondences between the views of a stereo pair. It restricts the search for corresponding image points from the entire image plane to a single line. For any point $\mathbf{x}$ in a view of the pair, the corresponding point $\mathbf{x}'$ in the other view, if it exists, lies on the *epipolar line* of $\mathbf{x}$ denoted by $\mathrm{epi}^{\mathbf{x}}$. Conversely $\mathbf{x}$ lies on $\mathrm{epi}^{\mathbf{x}'}$. The rectification aims to resample the images in such a way that corresponding points are located on the same row, thus simplifying the matching task and permitting to use all classic stereo matching algorithms. This however is not always possible.

Aknowledgements: work partially supported by Centre National d'Etudes Spatiales (MISS Project), European Research Council (Advanced Grant Twelve Labours), Office of Naval Research (under Grant N00014-97-1-0839), Direction Générale de l'Armement, Fondation Mathématique Jacques Hadamard and Agence Nationale de la Recherche (Stereo project).

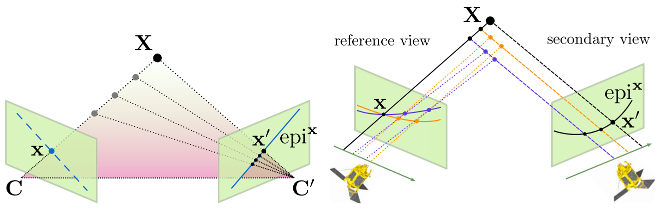


Fig. 2: In the pinhole case the epipolar plane defines a correspondence between epipolar lines. In the pushbroom case the projection of a 3-space ray on the secondary view generates a ruled quadric. The projection of this quadric on the reference view contains many epipolar curves: epipolar curves are not conjugate.

For images taken with pinhole cameras there is a correspondence between the epipolar lines of the two views. All the points $\mathbf{x}'$ of the second view lying on the epipolar line $\mathrm{epi}^{\mathbf{x}}$ share the same epipolar line in the first view. Epipolar lines $\mathrm{epi}^{\mathbf{x}}$ and $\mathrm{epi}^{\mathbf{x}'}$ are said to be *conjugate*. Figure 2 illustrates the conjugacy of epipolar lines. It is well-known [2, 3] that images can be resampled in order to produce a rectified pair in which the epipolar lines are horizontal and match up between views. Matching rectified images is much simpler than matching the original images, because the search of correspondences is performed along horizontal lines only [4].

Satellite images however can't be rectified because they are taken with pushbroom sensors, for which the pinhole model is invalid. Using various pushbroom camera models [5, 6], it can be shown that pushbroom image pairs have non-straight epipolar curves and that these curves are not conjugate, making stereo-rectification impossible [7, 8]. In this work we study to what extent it is possible to stereo-rectify pushbroom images anyway, in order to use standard matching algorithms for processing satellite stereo pairs. This study is motivated by the availability of high resolution images from new satellites with stereo capabilities such as Pléiades.

The Pléiades constellation is composed of two Earth observation satellites able to deliver images with a resolution of 70 cm and a swath width of 20 km. The unique agility of the Pléiades satellites allows them to capture multiple views of the same target in a single pass. This permits nearly simultaneous stereo or tri-stereo acquisitions with a small base to height ratio, ranging from 0.15 to 0.8.

The non rectifiability of pushbroom images didn't block the development of 3D reconstruction solutions for satellite images. Many solutions have been proposed to circumvent this issue. We may group them into three categories:

1. No rectification [9, 10, 11, 12]: many authors propose to keep the original images unchanged and to perform stereo matching by following the non-straight epipolar curves. This approach eliminates the need for stereo-rectification while keeping the benefits of one-dimensional exploration. However, non-straight epipolar curves may prevent from applying stereo matching optimizations and from using off-the-shelf correlators.

2. Affine camera approximation [13, 14, 15, 16]: other authors propose to approximate the pushbroom sensor with an affine camera model. This approach often uses Ground Control Points (GCP) to estimate the affine model for each image, and the overall achieved precision is on the order of one pixel on images from Spot and Ikonos satellites.

3. Polynomial epipolar resampling [17, 18]: Oh *et al.* show that even if pairs of epipolar curves don't exist in the pushbroom case, for small altitude ranges of the scene one may assume that curve pairs exist with small error. Thus they build whole epipolar curve pairs on Ikonos stereo images by putting together small pieces of corresponding curves. Then they resample the images to transform these curves into straight horizontal lines. They report a maximal error of one pixel. Since their resampling procedure is non-linear, it can't guarantee that straight lines are preserved.

It is important to note that errors in the rectification are critical as they may result in a vertical disparity between corresponding points in the rectified images, which may hurt the performance of the stereo matching. We refer to this vertical disparity as *epipolar error*. The epipolar error is the ultimate performance measure for the different methods. Current state of the art methods attain errors on the order of one pixel. Our method lowers this error by one order of magnitude.

Our contribution. Our goal is to build an automatic 3D reconstruction pipeline for satellite images. We started by observing that a large-scale stereo-rectified pair is not needed for applying a stereo matching algorithm. Thus we propose, like Morgan *et al.* [15], to approximate the sensor by an affine camera model. But, unlike Morgan [15], this approximation is made only on small image tiles. This limits the discrepancy between epipolar curves, as was studied by Oh *et al.* [17]. It leads in practice to an almost perfect stereo-rectification, with a very small epipolar error.

In our approach the rectification is seen as an intermediary step to efficiently solve the stereo correspondence problem, not as a product per se. For each locally rectified tile a standard off-the-shelf stereo algorithm can be applied to estimate a horizontal disparity map, with high chances of success thanks to the high precision of the stereo-rectification. The computed correspondences are then transferred to the coordinate system of the original images. This eliminates the need for stereo-rectifying the full images all at once.

The proposed local rectification hinges on the external and internal camera calibration information, provided for satellite images as co-localization functions [19, 20]. This information permits to quantify a priori the epipolar error due to the affine approximation. Conducting our experimentation on several satellite stereo pairs of urban, flat and mountainous regions, we obtained errors on the order of tenth of pixel.

The proposed method is currently being used by CNES on Pléiades images, as part of the *s2p 3D reconstruction stereo pipeline*, which is an automatic pipeline for computing digital elevation models and 3D points clouds from stereo and tri-stereo datasets. Figure 1 shows some of its results. This tool can be tested online [1] and provides rectified stereo pairs to be used by the image processing community.

We detail the proposed solution in the next section and in section 3 we validate our approach with extensive experimentation carried out using images from the Pléiades satellites.

2. LOCAL RECTIFICATION WITHOUT GCP

While Morgan *et al.* [15] use GCPs to estimate the affine camera models, this work uses the standard computer vision approach for stereo-rectification [2]: first estimate the affine fundamental matrix between the two views, then compute a pair of affine transformations to rectify the images. The fundamental matrix estimation requires only image matches, eliminating the need for GCPs and manual intervention.

2.1. In defense of the affine approximation

The suitability of the affine camera model in approximating a satellite pushbroom sensor can be attributed to Okamoto *et al.* [21]. Their arguments are all applicable to Pléiades images:

- Altitude differences in the photographed terrain are small in comparison with the flying altitude of the satellite, whose mean is 694 km for Pléiades.
- The angular field of view of the sensor is narrow. For a full Pléiades image it is less than 2°, and it is much less if one considers only a small tile.
- The acquisition time of such a tile is less than one second, thus the sensor may be assumed to have the same attitude and speed while capturing the scene.

Our locally affine stereo-rectification is summarized as follows: given at least 4 correspondences $(\mathbf{x}_i, \mathbf{x}'_i)_{i=1,\ldots,N}$ between the two views, the affine fundamental matrix $\mathtt{F}$ is estimated using the Gold Standard algorithm. It is worth noting that in the *affine* case the Gold Standard algorithm is reduced to a very simple linear algorithm [2]. Two rectifying affine transformations are then extracted from $\mathtt{F}$ [3].

Algorithm 1: Locally affine stereo-rectification of pushbroom images.
Data: $\text{RPC}_1, \text{RPC}_2$: RPC's of input images;
$x, y, w, h \in \mathbf{R}$: coordinates of ROI in image 1
Result: $\mathtt{H}_1, \mathtt{H}_2$: rectifying homographies
1 estimate altitude range ; `// from RPCs or SRTM`
2 compute N virtual matches $(\mathbf{x}_i, \mathbf{x}'_i)$; `// section 2.2`
3 estimate $\mathtt{F}$ from $(\mathbf{x}_i, \mathbf{x}'_i)$; `// Gold Standard algorithm`
4 estimate $\mathtt{H}_1$ and $\mathtt{H}_2$; `// Loop Zhang algorithm`

2.2. The problem of finding correspondences

A natural way to compute correspondences between the two views is to extract feature points, compute descriptors and match them, as done by SIFT [22]. But this may lead to a set of keypoints all lying on the same plane, *i.e.* on the ground. This configuration is degenerate and $\mathtt{F}$ cannot be computed from it. Even if the keypoints do not exactly lie on the same plane, as relief reduces to zero, the covariance of the estimated $\mathtt{F}$ increases [2]. We found that a safer way to estimate $\mathtt{F}$ is to use the calibration data [17, 19] to generate virtual correspondences between the two views.

For each Pléiades image the internal and external parameters of the pushbroom sensor are known [20] and are described using the rational polynomial camera model. Each image is accompanied by a pair of functions, called RPCs (Rational Polynomial Coefficients) [19], that allow to convert from image coordinates to coordinates on the globe and back. The projection from object space to image plane is denoted by $\text{RPC} : \mathbf{R}^3 \to \mathbf{R}^2$, $(\varphi, \lambda, h) \mapsto (x, y)$, where 3-space points are represented by their spheroidal coordinates. In that system a point of 3-space is identified by its latitude $\varphi \in (-90, 90)$, longitude $\lambda \in (-180, 180]$ and altitude h, in meters, above the Earth surface. Its inverse, with respect to the first two components, is denoted by $\text{RPC}^{-1} : \mathbf{R}^3 \to \mathbf{R}^3$, $(x, y, h) \mapsto (\varphi, \lambda, h)$. It takes a point $\mathbf{x} = (x, y)$ in the image together with an altitude h, and returns the coordinates of the unique 3-space point $\mathbf{X} = (\varphi, \lambda, h)$ whose altitude is h and whose image is $\mathbf{x}$.

Virtual correspondences generation. Given a region Ω in the reference image and an estimated altitude range $[h_m, h_M]$ for the associated 3-space points (*i.e.* points that were imaged into Ω) Ω is back-projected on the Earth surface thanks to RPC^{-1}. Let denote by $\Gamma = \text{RPC}^{-1}(\Omega \times [h_m, h_M]) \subset \mathbf{R}^3$ the back-projected domain, and by $(\mathbf{X}_i)_{i=1,\dots,N}$ a regular sampling of Γ. Each 3-space point $\mathbf{X}_i$ is projected on the two images using the associated RPCs, leading to a virtual correspondence $(\mathbf{x}_i, \mathbf{x}'_i)$. The images contents at locations $\mathbf{x}_i$ and $\mathbf{x}'_i$ may not correspond, but $\mathbf{x}'_i$ is located on the epipolar curve of $\mathbf{x}_i$, and that is enough to estimate a fundamental matrix.

The locally affine stereo-rectification algorithm is summarized in Algorithm 1.

Dataset	Scene dim. (km)	RPC altitude validity (m)
calanques	25×24	40 – 1090
cannes	21×20	50 – 830
giza	26×23	10 – 290
mera	25×42	-10 – 8610
mont_blanc	21×15	850 – 4730
montevideo	22×20	-10 – 150
new_york	48×37	-120 – 190
ossoue	22×22	-10 – 3320
spitsberg	21×20	-10 – 640
toulouse	25×21	150 – 340
tregor	26×24	50 – 160
ubaye	22×15	1100 – 3050
mercedes	25×23	10 – 90
fray_bentos	22×20	0 – 80

Table 1: Pléiades datasets used for the experiments.

3. EXPERIMENTAL VALIDATION

The proposed stereo-rectification method is evaluated by measuring the epipolar error, which is completely determined by the fundamental matrix $\mathtt{F}$. This error is measured by

$$\max_{i \in \{1,\dots,N\}} \max\{d(\mathbf{z}'_i, \mathtt{F}\mathbf{z}_i), d(\mathbf{z}_i, \mathtt{F}^\top \mathbf{z}'_i)\},$$

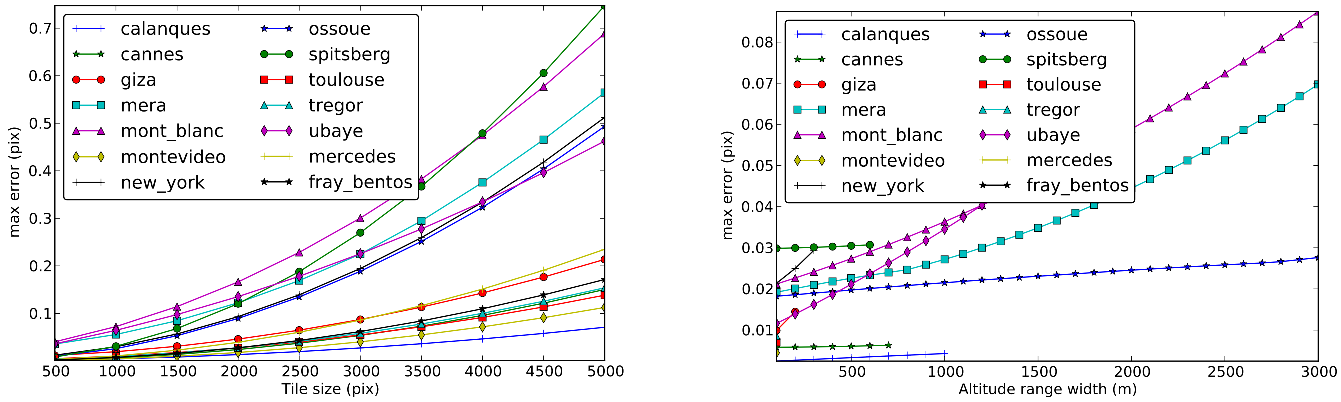
where $d(\mathbf{z}, \mathbf{l})$ is the distance, in pixels, between a point $\mathbf{z}$ and a line $\mathbf{l}$. The matches $(\mathbf{z}_i, \mathbf{z}'_i)_{i=1\dots,N}$ are new virtual correspondences obtained from a sampling of volume Γ. This error is the maximal distance between a point's epipolar line and the matching point in the other image (computed for both points of the match). The distance $d(\mathbf{z}'_i, \mathtt{F}\mathbf{z}_i)$ between a point $\mathbf{z}'_i$ and the epipolar line it is supposed to lie on $(\mathtt{F}\mathbf{z}_i)$ is computed as

$$d(\mathbf{z}'_i, \mathtt{F}\mathbf{z}_i) = \frac{|\mathbf{z}'^\top_i \mathtt{F}\mathbf{z}_i|}{\sqrt{(\mathtt{F}_1^\top \mathbf{z}_i)^2 + (\mathtt{F}_2^\top \mathbf{z}_i)^2}},$$

where $\mathtt{F}_1^\top, \mathtt{F}_2^\top$ and $\mathtt{F}_3^\top$ denote the three rows of matrix $\mathtt{F}$.

Numerical results. From a geometric viewpoint, the locally affine stereo-rectification method described here amounts to approximate the two pushbroom sensors with affine camera models. The validity of such an approximation relies on the dimensions of the 3-space domain on which it is applied. These dimensions depend on two parameters: the tile size and the width of the altitude range. To understand the influence of these two parameters on the epipolar error, experiments were performed on numerous Pléiades stereo datasets, listed in Table 1. Each dataset contains a stereo pair of images together with their calibration data given by the RPCs. The validity domain of the RPC functions in the altitude variable gives a rough estimation of the altitude range of the scene.

Figure 3 shows the error measured on each dataset by varying the tile size and the width of the altitude range separately. Figure 3a shows the errors obtained when varying the tile size up to 5000×5000 pixels, while keeping the width of the altitude range fixed to 2500 m (or less if needed by the



(a) Influence of the tile size. A tile of size ranging from 500×500 to 5000×5000 pixels was selected in the middle of the reference image. Virtual matches between this tile and the secondary image were computed using an altitude range of width fixed to 2500 m.

(b) Influence of the width of the altitude range. A tile of size fixed to 1000×1000 pixels was selected in the reference image. Virtual matches between this tile and the secondary image were computed using an altitude range of width ranging from 100 m to 3000 m.

Fig. 3: Dependence of the epipolar error with the tile size and the width of the altitude range, on several datasets. The altitude range was centered in the RPCs validity domain, and shrinked, if needed, to fit the validity domain. The error reported by each of the plots is the maximal distance between a point and the epipolar line of its match. The epipolar line was computed thanks to the estimated affine fundamental matrix.

RPCs validity range). This is enough because at Pléiades resolution (50 cm per pixel), 5000 pixels correspond to 2.5 km, and it is reasonable to assume that no place on Earth is steeper than $45°$ when averaged on a 2.5×2.5 km area (even in the Himalaya). It is clear from plot 3a that for 1000×1000 pixels tiles the error is always under 0.1 pixel. The error increases with the tile size, and it increases faster for mountainous regions like Mont Blanc (Alps), Mera peak (Himalaya), Ossoue glacier (Pyrenees), Ubaye valley (Alps) and Spitsberg island. This may be due to the fact that the RPCs of the other datasets have an altitude validity range narrower than 2500 m.

These results are confirmed by plot 3b, where the tile size is fixed to 1000×1000 pixels while the width of the altitude range is varied up to 3000 m. This figure confirms that on any Pléiades dataset, for any altitude variations in the scene, if one considers a tile of 1000×1000 pixels the epipolar error due to the affine stereo-rectification is smaller than 0.1 pixel.Whenever the validity domain of RPCs for the altitude is narrower than 3000 m, the width of the altitude range is varied only up to that limit.

In practice, this result can be improved by using SRTM data [23]. The Shuttle Radar Topography Mission (SRTM) is an international research effort that obtained digital elevation models on a near-global scale at a resolution of three arcseconds, *i.e.* 90 m. The SRTM data may be used to estimate more precisely the altitude range for a given tile. The precision gained on the altitude range allows to use bigger tiles, while keeping the epipolar error lower than 0.1 pixel. Figure 4 shows the influence of the tile size on the error when the altitude range is estimated from SRTM data. These results show that it is always possible to stereo-rectify tiles of size 1000×1000 with an epipolar error lower than 0.05 pixels.

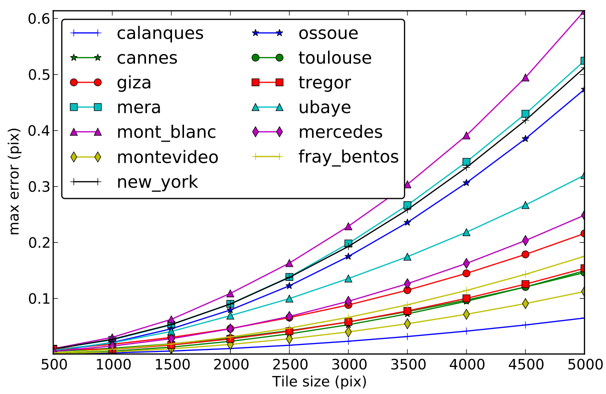


Fig. 4: Dependence of the epipolar error with the tile size. A tile of size ranging from 500×500 to 5000×5000 pixels was selected in the middle of a Pléiades reference image. Virtual matches were computed using the altitude range given by SRTM data. Spitsberg is located outside SRTM domain, thus omitted here.

4. CONCLUSION

In this work we showed that, for the purpose of stereo matching, pushbroom satellite images can be stereo-rectified and therefore be reduced to standard stereo pairs. Thorough experimentation on numerous Pléiades datasets has shown that using tiles of size 1000×1000 pixels ensures a precision of 0.1 pixel, regardless of the altitude range of the scene. The stereo-rectification described here is implemented in the Satellite Stereo Pipeline (s2p), which can be tested online [1]. This method could be used on images from other Earth observation satellites such as WorldView-2 and GeoEye-1. However, since for these images only the direct RPC projection function is provided, the inverse RPC^{-1} must be estimated.

5. REFERENCES

[1] C. de Franchis, G. Facciolo, and E. Meinhardt-Llopis, "s2p IPOL demo," http://dev.ipol.im/~carlo/s2p_icip/, 2014.

[2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, second edition, 2004.

[3] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999, vol. 1, pp. 125–131.

[4] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 139–154, Mar. 1985.

[5] A. B. Orun and K. Natarajan, "A modified bundle adjustment software for spot imagery and photography: tradeoff," *Photogrammetric Engineering and Remote Sensing*, vol. 60, no. 12, pp. 1431–1437, 1994.

[6] R. Hartley and R. Gupta, "Linear Pushbroom Cameras," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Sept. 1997, pp. 963–975.

[7] T. Kim, "A study on the epipolarity of linear pushbroom images," *Photogrammetric Engineering and Remote Sensing, vol 66*, , no. 8, pp. 961–966, 2000.

[8] A. F. Habib, M. Morgan, S. Jeong, and K.-O. Kim, "Analysis of Epipolar Geometry in Linear Array Scanner Scenes," *The Photogrammetric Record*, vol. 20, no. 109, pp. 27–47, Mar. 2005.

[9] H.-Y. Lee, T. Kim, W. Park, and H. K. Lee, "Extraction of digital elevation models from satellite stereo images through stereo matching based on epipolarity and scene geometry," *Image and Vision Computing*, vol. 21, no. 9, pp. 789–796, Sept. 2003.

[10] H. Hirschmüller, F. Scholten, and G. Hirzinger, "Stereo Vision Based Reconstruction of Huge Urban Areas from an Airborne Pushbroom Camera (HRSC)," in *Pattern Recognition*, W. Kropatsch, R. Sablatnig, and A. Hanbury, Eds., vol. 3663 of *Lecture Notes in Computer Science*, pp. 58–66. Springer Berlin Heidelberg, 2005.

[11] H. Hirschmüller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[12] R. T. Collins, "A space-sweep approach to true multi-image matching," in *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on*, 1996, pp. 358–363.

[13] T. Ono, "Epipolar resampling of high resolution satellite imagery," *International Archives of Photogrammetry and Remote Sensing*, Sept. 1999.

[14] S. C. Fraser, P. M. Dare, and T. Yamakawa, "Digital surface modelling from spot 5 hrs imagery using the affine projective model," in *XXth ISPRS Congress*, July 2004, vol. 35, pp. 385–388.

[15] M. Morgan, K.-O. Kim, S. Jeong, and A. Habib, "Epipolar Resampling of Space-borne Linear Array Scanner Scenes Using Parallel Projection," *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 11, pp. 1255–1263, Nov. 2006.

[16] M. Wang, F. Hu, and J. Li, "Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 347–355, May 2011.

[17] J. Oh, W. Hee Lee, C. K. Toth, D. A. Grejner-Brzezinska, and C. Lee, "A Piecewise Approach to Epipolar Resampling of Pushbroom Satellite Images Based on RPC," *Photogrammetric Engineering and Remote Sensing*, vol. 76, no. 12, pp. 1353–1363, Dec. 2010.

[18] E. Christophe, J. Inglada, and A. Giros, "Orfeo toolbox: a complete solution for mapping from high resolution satellite images," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, pp. 1263–1268, 2008.

[19] C. V. Tao and Y. Hu, "A comprehensive study of the rational function model for photogrammetric processing," *Photogrammetric Engineering and Remote Sensing*, vol. 67, no. 12, pp. 1347–1358, 2001.

[20] Astrium EADS, *Pléiades Imagery User Guide V 2.0*, October 2012.

[21] A. Okamoto, S.-I. Akamatu, and H. Hasegawa, "Orientation theory for satellite ccd line-scanner imageries of hilly terrains," *International Archives of Photogrammetry and Remote Sensing*, vol. 29, pp. 217–222, 1993.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[23] T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, et al., "The shuttle radar topography mission," *Reviews of Geophysics*, vol. 45, no. 2, 2007.