

DENSE MULTI-VIEW IMAGE MATCHING FOR DSM GENERATION FROM SATELLITE IMAGES

A. Mahphood¹, H. Arefi^{1,*}, A. Hosseinaveh², A. A. Naeini³

¹ School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran - (ahmadmahphood, hossein.arefi)@ut.ac.ir

² Department of Photogrammetry and Remote Sensing Faculty of Geodesy and Geomatics Engineering K. N. Toosi University of Technology - hosseinaveh@kntu.ac.ir

³ Department of Geomatics Engineering, Faculty of Civil Engineering and Transportation, University of Isfahan, Isfahan 8174673441, Iran - a.alizadeh@eng.ui.ac.ir

Commission VI, WG VI/4

KEY WORDS: DSM, Multi-View, Semi-Global Matching, Stereo Images, Dense Image Matching.

ABSTRACT:

Digital Surface Model (DSM) can be generated from stereo pairs of satellite or aerial images. Among the most state-of-the-art matching algorithms, Semi-Global Matching (SGM) has widely been used for generating DSM from both satellite and aerial images. This paper presents an approach to improve the accuracy of DSM generated by SGM from multi-view satellite images using a novel technique including several filters. The filters are used for deleting mismatches between very tall buildings in urban areas and removing the sea regions. The technique, in contrast to the recent multi-view matching approaches, considers some of the points generated with only a pair of images in the final DSM. The approach is implemented on five sequential high resolution images acquired by the Worldview-2 satellite. The results are locally evaluated in shape and quantitative terms in comparison with commercial software to reveal the capability of the approach to generate a reliable and dense point cloud. Experiments show that the proposed method can achieve below half-pixel accuracy.

1. INTRODUCTION

DSM is a digital model or 3D representation of a terrain's surface derived from various sources such as photogrammetry, LiDAR and IFSAR (Li et al., 2005). DSM gives some information about the shape of the earth's surface which can be used for various applications such as urban planning, flood control, resource management, telecommunication planning, military mapping, and etc. (Li et al., 2005).

Recently, researchers have used the multi-image techniques to generate a digital model from close range images (Ahmadabadian et al., 2013, Innmann et al., 2019), aerial (Ameri et al., 2002, Haala et al., 2012) and satellite images (Zhang et al., 2006, Krishna et al., 2008, d'Angelo et al., 2012, Giribabu et al., 2013, Qin, 2017, Gong et al., 2018). The accuracy of the final point clouds directly related to the accuracy of image matching algorithms. Most of the current approaches (d'Angelo et al., 2012, Haala et al., 2012) used SGM (Hirschmüller, 2008) for both pushbroom satellite and aerial images. SGM is performed through pixel-wise matching of Mutual Information (MI) and approximating a global, two dimensional (2D) smoothness constraint by combining eight or sixteen one dimensional (1D) constraints.

These researches have used some filters such as consistency check (Haala et al., 2012), sparse-area removal (d'Angelo et al., 2012) and removing points appeared in only a pair of images (d'Angelo et al., 2012, Haala et al., 2012) to remove mismatches. Although the proposed filters can improve the

accuracy, there are some other issues that aren't considered. For instance, some specific regions such as sea areas or very tall buildings which are leading to the occluded areas, cannot be recognized in the matching process.

Since the object coordinates of points computed from only two corresponding image points are less accurate than the other points which are supported from more than two corresponding images, the points visible in only a pair of images have been removed by researchers (Haala et al., 2012) in order to increase the accuracy. However, if the number of images is low and the images provided from an urban area with towers and very tall buildings, removing such points that are visible only in a pair of images may lead to removing important points.

This paper aims to improve both the accuracy and competences of the final DSM based on a novel approach in the multi-view matching algorithm. In this approach, the accuracy is improved through filtering sea area (water masking) as well as considering some points visible only in a pair of images.

2. METHODOLOGY

The proposed method consists of several major stages, as can be seen in Figure 1. All stages comprise some smaller steps which are described in detail in following sub-sections:

* Corresponding author

2.1 Generating Epipolar Images

Epipolar resampling aims at generating normalized images where conjugate points are located along the same row. The process provides normalized images that are important for many applications such as automatic image matching, aerial triangulation, DEM and orthophoto generation, and stereo-viewing (Morgan et al., 2006). More importantly, using epipolar images increases the speed of the correlation process and reduces the possibility of incorrect matches. This process can now be performed automatically in commercial software packages such as PCI Geomatica.

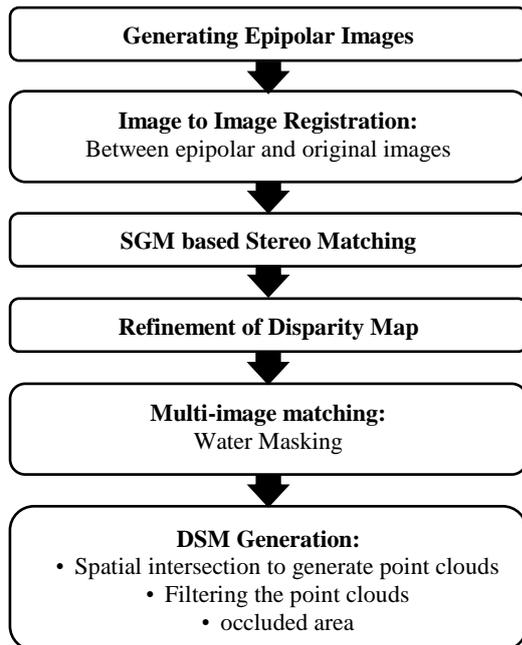


Figure 1. The Workflow of the DSM generation approach

2.2 Image to Image Registration

One of the final stages of DSM generation is the spatial intersection using Rational Polynomial Coefficients (RPCs). However, instead of original images, epipolar images are used in the process of image matching. As a result, image coordinates must be transformed from epipolar images to the original ones. In other words, a 2D geometric transformation should be computed between the original and epipolar images. Some researchers have tried to solve this problem by rotating stereo images to the base direction of flight and producing quasi-epipolar images (Alobeid et al., 2010, Dini, 2014). However, it was proved that in this method Y parallax was not corrected properly, unless the images were high resolution with a small inclination angles.

The first step of transformation between the original and epipolar images is to find corresponding points. For this purpose, SURF method (Bay et al., 2006) is utilized to extract key points from both original images and their epipolars. Then, corresponding points are matched using normalized cross correlation method (NCC) (Zitova et al., 2003), and the outliers are removed using the RANSAC method (Fischler et al., 1981). Finally, the conformal parameters are estimated using corresponding points and least squares technique.

2.3 SGM based Stereo Matching

To generate DSM from multiple images, the central image is selected as the base image and the rest of the images are considered as search images. For each pair, left to right image matching is performed using SGM and accordingly a disparity map is generated (Hirschmüller, 2008). It should be noted that utilizing rectified images in SGM reduces the search area and mismatches in one hand, and increases the speed of the matching process in another hand.

2.4 Refinement of Disparity Map

Disparity maps should be refined to have more reliable corresponding points. For this purpose, two filters, namely, consistency check and removing isolated areas are applied.

2.4.1 Consistency check: Consistency check (d'Angelo et al., 2012, Haala et al., 2012) is applied by changing the roles between the base and search images. Only disparity estimations consistent to this forward-backward matching are considered as valid matches.

2.4.2 Removing isolated areas: As second filter, isolated areas whose pixel numbers are smaller than 200 pixels in the DSM are removed as outliers (d'Angelo et al., 2012).

2.5 Multi-Image Matching

In this section, corresponding points between all original images have been recognized using both transformations (Section 2.2) and disparity maps (Section 2.3). Disparity maps contain parallax values which can be used for finding corresponding points in the epipolar images and not in original ones. However, in order to do a spatial intersection for point cloud generation, the corresponding points must be projected back to the original images where RPCs are known. To tackle this problem, 2D geometric transformations (Table 2) are employed.

Nonetheless, the corresponding points among all images must be recognized using a multi-image matching process. For this purpose, transformations between original and epipolar images are performed. This process is started from one image which is the original base image. The pixels of the original base image are transformed to their epipolar images by the transformations. After that, considering the fact that the central image is left for all epipolar image pairs, the corresponding pixels in the right epipolar images (image 1, image 2, and image 3 in Figure 2) are found by using the disparity values of epipolar image pairs (equation (1)). Thus, corresponding points between left epipolar images (base epipolar) and right epipolar images (search images) are found.

$$\begin{aligned} i_r &= i_i \\ j_r &= j_i - \text{disparitymap}(i_i, j_i) \end{aligned} \quad (1)$$

where i_r and j_r = pixel coordinates of the right epipolar image.
 i_i and j_i = pixel coordinates of the left epipolar image.

Next, corresponding pixels in the right epipolar image pairs are transformed to their original images by the Inverse of geometric transformation. Consequently, corresponding pixels in all the original images have been found (Figure 2). For reducing computational time and increasing the accuracy, pixels related

to the water bodies are masked in the same time as they are projected from original base image to the epipolar ones.

2.5.1 Water Masking: Since the water bodies, e.g. seas, in images cause mismatches after matching, therefore they should be ignored in matching process. There are many investigations to extract water body from satellite imagery such as (Nath et al., 2010) suggesting the IR spectral band or (Wohlfeil et al., 2012) who have employed SRTM DEM. A new pixel-based method is proposed here to recognize water areas. This filter consists of following three processing steps:

1. Image segmentation is implemented and sea region is determined by using the value of pixels of sea region. However, the pixels of other regions might have the same gray value and will be determined as well (Figure 5(a)). Moreover, due to different reflections from water surface, there may exist region in sea area with different intensities. Although morphological tools can be used for solving these problems, they lead to changing the maritime border. Therefore, the second and third stages are proposed for removing wrongly recognized regions and filling the gaps within the sea region.
2. The wrongly recognized regions identified in the previous stage are removed by defining a threshold for regions smaller than a predefined value specified by considering the size of the sea region as shown in Figure 5(b). The used threshold is 50000 pixels, because there are many buildings that are connected together.
3. The gaps within sea region are filled and joined to sea area, so that the image resulted from previous stage contains two sections, determined region (sea region) and undetermined regions (gaps within sea region and ground region). Now the gaps in undetermined regions are determined through defining a threshold for region smaller than a certain value specified with considering the size of the undetermined regions (Figure 5(c)). The used threshold is 1000 pixels.

2.6 DSM Generation

Given refined corresponding points from previous stage, point clouds are generated using spatial intersection (Grodecki et al., 2003). Next, the points are classified based on the number of rays used for computing their coordinates. Points visible in only an image pair are chosen in one group named non-reliable point cloud (single), and other points are named reliable point clouds. For DSM generation, (Kuschik, 2013, Qin, 2017, Gong et al., 2018) generate DSM for each stereo pair. In the end, all these DSMs are fused by simple median filter (Kuschik, 2013, Gong et al., 2018) or adaptive depth fusion method that considers the spatial consistency (Qin, 2017). In other word, the fusion is implemented in 2.5D space. In our approach, the fusion is implemented in 3D space (point cloud). So, there are at least two points for each pixel in the base image. These points are fused by median filter. The resulted point cloud is the reliable point cloud.

As mentioned in the introduction, previous researchers removed points visible in only an image pair (non-reliable point cloud). However, due to occlusion and self-occlusion areas on the one hand and on the other the limitation in the number of images, removing non-reliable point cloud leads to loss of important

information. Therefore, this group of points is preserved and processed in this paper.

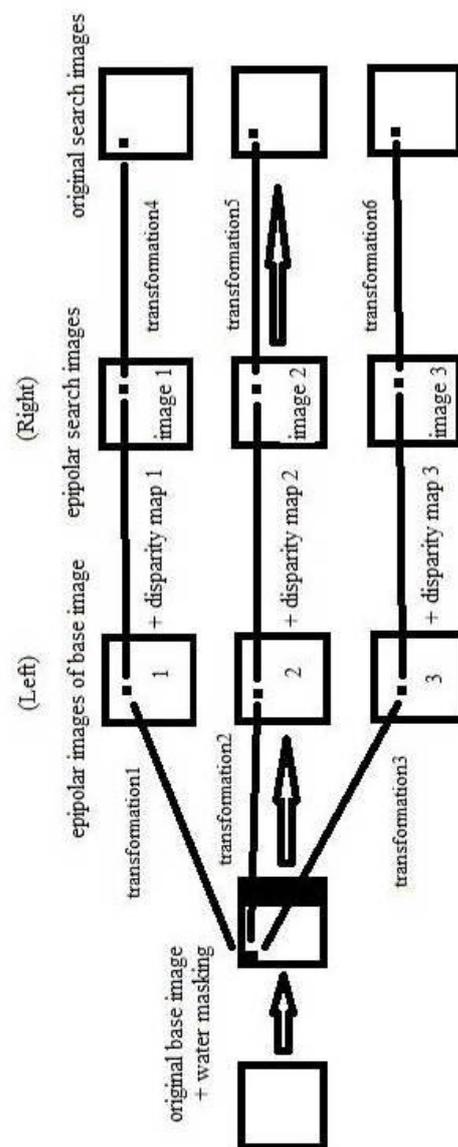


Figure 2. Multi-image matching process for one pixel.

2.6.1 Handling Occlusions: For occluded areas, similar to multi-image matching Methodology (from Section 2.3 to Section 2.6), the procedure is repeated by considering other images as base image. As outcome, for each image a new point cloud with a lot of common points with the original point cloud, resulting from Section 2.5 are generated. Points which are common between two point clouds must be removed because they are generated before. However, the rest of points can be either mismatches or located in occluded areas. If they are located in occluded area, they must be preserved; otherwise they must be removed and this work is done in Section 2.6.2. In this way, a point cloud is generated by deleting the largest possible number of mismatches and at the same time the important points are preserved.

2.6.2 Filtering Non-Reliable Point Cloud: Here, a filter is proposed to find points that must be preserved and points that must be removed. This filter has two steps. In the first step, the points are segmented by using region growing technique. A single point (seed) is randomly chosen and the neighbouring points are added to the region if they are within a specific distance (e.g. one meter). This step is repeated for every new point and if no more points can be added, the process is terminated.

In second step, these regions are checked if they complete the reliable point cloud. So that, if one point of the reliable cloud points is near to the previous resulted region, then this region must be preserved. Otherwise, the region should be deleted. For this step, we used a threshold equal to the two times of the pixel size (1 m).

3. EXPERIMENTS AND RESULTS

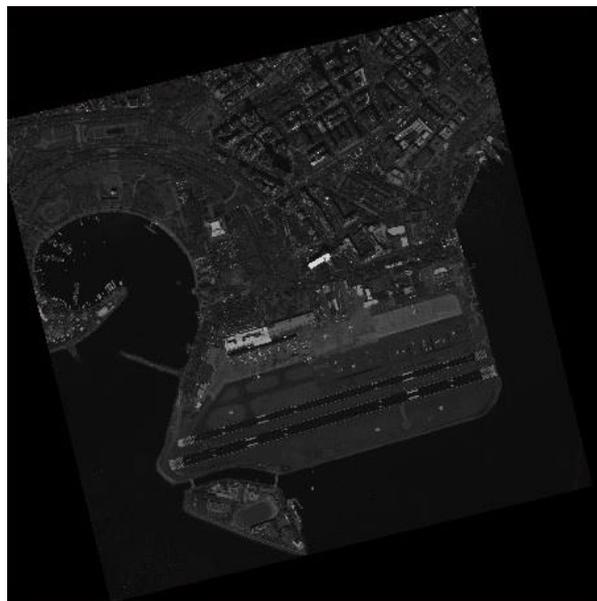
The approach was applied on a dataset containing five sequential images of World-View 2 with 4000 by 4000 pixels from the city of Rio de Janeiro in Brazil (Figure 3(a)). The pixel size is 0.5 m with more than 90% overlap. The type of images used for generating DSM is PAN images. For all images, the RPCs were provided from the metadata files. The details of images used to generate DSM are shown in the Table 1.

Output format	GeoTIFF
Date	31-03-2010
Scan Direction	Reverse
Number of Rows	4000 pixel
Number of Columns	4000 pixel
Bits Per Pixel	16
GSD	0.5 m
Band Id	PAN

Table 1. Details of images used to generate DSM.



(a)



(b)

Figure 3. a) WorldView-2 image for test site, b) epipolar base image.

3.1 Generating Epipolar Images

The trial version of PCI Geomatica was used to generate epipolar images. Then the mathematical model for each image pair was computed. RMS for this process was for each model about 0.23 pixel. As is obvious, the accuracy is very good for transforming pixels from epipolar images to original ones. The epipolar base image is illustrated in Figure 3(b).

3.2 Stereo Matching using SGM and Filtering

Running SGM on each image pairs resulted in a disparity map which shows the parallaxes along the X axis with respect to the right epipolar image. Next, the consistency check was applied for each image pair. The Figure 4 illustrates some of the removed points after this filter.



(a)



(b)

Figure 4. a) Disparity map without consistency check, b) disparity map with consistency check.

Among different 2D geometric transformations, two low-order polynomials (conformal and affine) and a second-order polynomial (bilinear) were tested and results are listed in Table 2. As shown, the results of all three polynomials are almost identical. Accordingly, 2D conformal with low computational time has been selected.

Image	Number of extracted points	RMS for types of transformations		
		Second-order polynomial (pixel)	Affine (pixel)	Conformal (pixel)
image 1	626	0.22	0.22	0.23
image 2	821	0.23	0.23	0.23
image 3	1099	0.23	0.23	0.24
image 4	811	0.22	0.23	0.24
image 5	648	0.22	0.22	0.23

Table 2. RMS for types of transformations.

While pixels were being transformed from the original base image to their epipolar images, the proposed approach of water masking was applied to pixels. It increases the speed of the proposed method by transforming fewer pixels to the epipolar images and more importantly filtering non-reliable points in the spatial intersection (Figure 5).

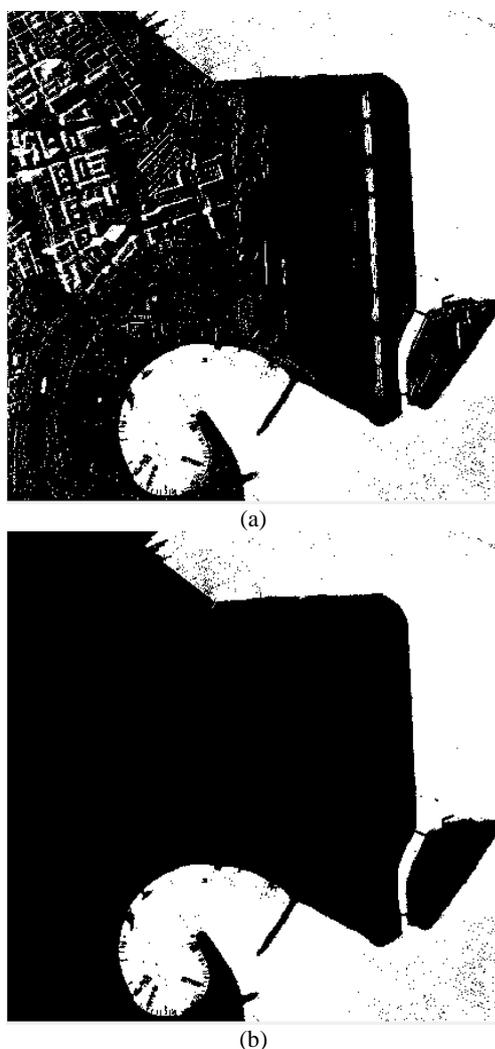


Figure 5. base image filtered. a) first stage, b) second stage and c) third stage.

After that, the point clouds were generated and divided into non-reliable and reliable points. The filter of non-reliable point cloud was applied and finally interpolation was employed using Nearest Neighbor technique (Li et al., 2005). Two DSM were generated, one was generated from only reliable points, and the other was generated from both the filtered non-reliable and reliable points.

As can be seen in Figure 6, the streets were removed in the DSM generated from only reliable points. However, they were preserved in the second DSM (Figure 6(b)). The streets are shown by rectangles. These objects were preserved and appeared in the second DSM because the filtered non-reliable points with different elevations were involved in the interpolation process.

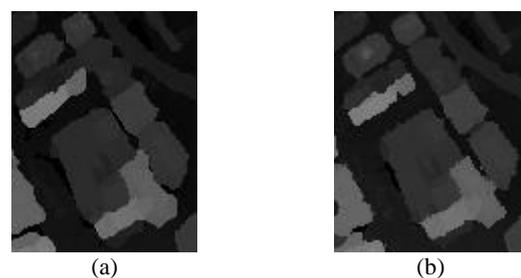


Figure 6. a) DSM generated from reliable points, b) the DSM after adding filtered non-reliable points

4. EVALUATION

In order to evaluate, our new approach for DSM generation from multiple satellite images, the result was compared to the DSM generated by using the trial version of PCI Geomatica. In PCI, the DSM was generated by computing the mean of the models generated from multi-pair of images (Figure 7).

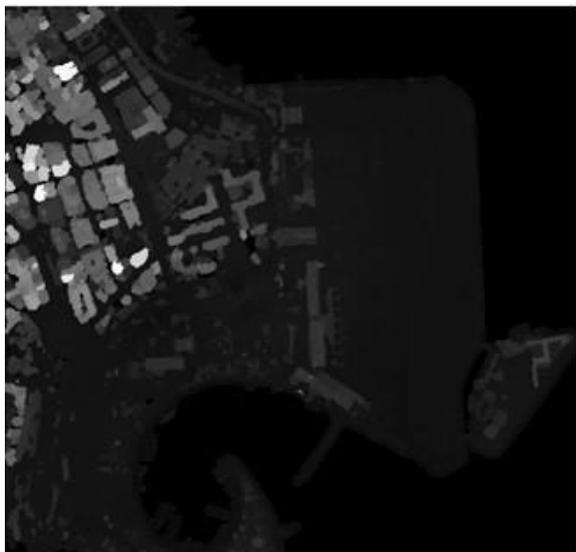
As shown in Figure 7, there are several blocks of buildings which are not separately recognizable in the DSM generated by PCI Geomatica. This is due to the algorithm implemented in PCI Geomatica which cannot identify streets between very high buildings. This problem has been solved in the DSM generated

by the proposed method. As it is illustrated in Figure 7(a), employing water-masking filter led to a better preserving coastal regions in the DSM.

Type	number of points	Standard deviation (m)
Building1	3051	0.236
Building2	5555	0.396
Building3	2831	0.122
Ground	21180	0.256

Table 3. standard deviation of tested sites points.

In order to quantitatively evaluate the accuracy of the generated DSM, the accuracy was tested based on several sample points were collected manually from flat surfaces (e.g. some building roofs). Next, the standard deviation of the sample points from a fitted plane on the surface was computed. As reported in Table 3, the accuracy is very reasonable with respect to image resolution, i.e. 0.5 m. It implies that we have achieved half-pixel level of accuracy or even better.



(a)



(b)

Figure 7. a) Generated DSM using the proposed method, b) DSM of PCI Geomatica.

5. CONCLUSION

DSM can be generated from high resolution multi-view satellite images using SGM. A new approach was proposed here to generate a complete, reliable and accurate DSM using SGM on multi view image and several filters two of which are new. The Multi-view imaging enable removing outliers and mismatches considerably. However, two new improvements, i.e. water masking and preserving some of non-reliable point clouds, are introduced in this study which not only reduce the cost burden of computational time but also preserve important points which were ignored in the previous research. In the water masking filtering, the sea area is completely removed resulting in reducing computational time and less mismatches. In the second filter, contrary to the previous researchers (d'Angelo et al., 2012, Haala et al., 2012), the points visible in only an image pair were not removed completely for the sake of completeness. And it can be concluded that the use of one geometric transformation between the original images and their epipolar images for high-resolution images and using the SGM method led into acceptable results, at least in the direction of the Z-axis. Accordingly, a reliable and accurate DSM is generated after applying these two steps.

It should be stated that, some of the buildings in the final generated DSM were still connected to each other due to the shadows of very tall building that have covered parts of streets. Therefore, in future, we will try to apply more filters on the urban images as pre-processing (like wallis filter) to remove the largest sources of mismatches, decrease the shadow effect and increase the performance of image matching process.

REFERENCES

- Ahmadabadian, A. H., S. Robson, J. Boehm, M. Shortis, K. Wenzel and D. Fritsch 2013. A comparison of dense matching algorithms for scaled surface reconstruction using stereo camera rigs. *ISPRS Journal of Photogrammetry and Remote Sensing* **78**: 157-167.
- Alobeid, A., K. Jacobsen and C. Heipke 2010. Comparison of matching algorithms for DSM generation in urban areas from Ikonos imagery. *Photogrammetric Engineering & Remote Sensing* **76**(9): 1041-1050.
- Ameri, B., N. Goldstein, H. Wehn, A. Moshkovitz and H. Zwick 2002. High Resolution Digital Surface Model (DSM) Generation using Multi-View Multi-Frame Digital Airborne Images. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci* **34**: 419-424.
- Bay, H., T. Tuytelaars and L. Van Gool 2006. Surf: Speeded up robust features. *Computer vision–ECCV 2006*: 404-417.
- d'Angelo, P. and G. Kusch 2012. Dense multi-view stereo from satellite imagery. *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*: 6944-6947.
- Dini, G. R. 2014. *Toward an Automatic Solution for Updating Building Databases Using Space-borne Stereo Imaging*,

Technische Informationsbibliothek und Universitätsbibliothek
Hannover (TIB).

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**(6): 381-395.

Giribabu, D., S. S. Rao and Y. K. Murthy 2013. Improving Cartosat-1 DEM accuracy using synthetic stereo pair and triplet. *ISPRS Journal of Photogrammetry and Remote Sensing* **77**: 31-43.

Gong, K. and D. Fritsch 2018. POINT CLOUD AND DIGITAL SURFACE MODEL GENERATION FROM HIGH RESOLUTION MULTIPLE VIEW STEREO SATELLITE IMAGERY. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci* **42**(2).

Grodecki, J. and G. Dial 2003. Block adjustment of high-resolution satellite images described by rational polynomials. *Photogrammetric Engineering & Remote Sensing* **69**(1): 59-68.

Haala, N. and M. Rothermel 2012. Dense multiple stereo matching of highly overlapping UAV imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci* **39**: B1.

Hirschmüller, H. 2008. Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **30**(2): 328-341.

Innmann, M., K. Kim, J. Gu, M. Niessner, C. Loop, M. Stamminger and J. Kautz 2019. NRMVS: Non-Rigid Multi-View Stereo. *arXiv preprint arXiv:1901.03910*.

Krishna, B. G., T. Amitabh, P. Srinivasan and K. Srivastava 2008. Dem generation from high resolution multi-view data product. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci* **37**: 1099-1102.

Kuschik, G. 2013. Large scale urban reconstruction from remote sensing imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci* **5**: W1.

Li, Z., Q. Zhu and C. Gold 2005. *Digital terrain modeling : principles and methodology*. New York, CRC Press.

Morgan, M., K.-O. Kim, S. Jeong and A. Habib 2006. Epipolar resampling of space-borne linear array scanner scenes using parallel projection. *Photogrammetric Engineering & Remote Sensing* **72**(11): 1255-1263.

Nath, R. K. and S. K. Deb 2010. Water-body area extraction from high resolution satellite images-an introduction, review, and comparison. *International Journal of Image Processing (IJIP)* **3**(6): 353-372.

Qin, R. 2017. Automated 3d recovery from very high resolution multi-view satellite images. *ASPRS (IGTF) annual Conference*: 10.

Wohlfeil, J., H. Hirschmüller, B. Piltz, A. Börner and M. Suppa 2012. Fully automated generation of accurate digital surface models with sub-meter resolution from satellite imagery. *Int. Arch. Photogramm. Rem. Sens. Spatial Inf. Sci*: 34-B33.

Zhang, L. and A. Gruen 2006. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* **60**(3): 195-211.

Zitova, B. and J. Flusser 2003. Image registration methods: a survey. *Image and vision computing* **21**(11): 977-1000.