Tri-Stereo Image Matching using Watershed Segmentation on Disparity Space Image

Raghavendra H. Bhalereo
IITRAM, Ahmedabad

Abstract

Local stereo matching algorithms use winner-take-all approach to get the disparity. Many times they end up at a erroneous result. To solve this, in the present work, a novel stereo image matching technique has been developed that identifies the most likely local minimum from the several possible local minima. The technique uses properties of the physical continuity of the land surface characteristic by the watershed segmentation applied to the disparity space volume. This helps to minimize the search for the local minima for matching. The matching is further improved by combining the watersheds of two stereo pairs from the tri-stereo. In the present study, experiments have been carried out using the standard Middlebury stereo datasets and remotely sensed tri-stereo images. Based on this approach the experiments are successfully carried out using the test dataset. The experimental results are compared with the results from the currently contemporary techniques of Dynamic Programming and Semi-Global Matching which resulted in 2-10 % improvement in density of matched points for different dataset.

Keywords- Disparity Space Image, Winner Takes All, Cost Function.

Introduction

Extraction of 3D information from a stereo pair has its importance in many fields such as digital photogrammetry, computer vision, robotics, intelligent vehicle navigation, etc. Based on the legacy of the optical imaging, co-linearity condition is established between object and image space. It relates the image point in the image space, the optical centre common to both images and the object space and the object point in the object space. The relation developed is then used to convert the disparity in the position of the image feature on the ground in a stereo pair (conjugate points) to the 3^{rd} dimension of the object space using the parallax equation. The stereo matching process thus needs to establish the correspondence in the stereo pair. It is thus imperative that the accuracy and the quality of the 3D measurements depend on the density and accuracy of these points that result from the processes of matching strategies used.

The problem of stereo correspondence is solved using two broad strategies, Area-Based (Remondino et al., 2014) and Feature-Based (Jazayeri and Fraser, 2010). In the case of the photogrammetric applications, the feature-based matching is used for the image transformation and the area-based methods are used for dense matching. Based on the work by Alobeid (2011), the most popular algorithms used in the photogrammetric applications for dense matching are Least Square Matching (Gruen, 1985), Dynamic Programming (Veksler, 2005) and Semi-global Matching (SGM) (Hirschmüller and Scharstein, 2008). There are many free and open source, and commercial software that are based on these strategies (Remondino et al., 2014). SGM is the most recent and robust algorithm of all (Hirschmüller, 2011). After the advent of tri-linear scanners on satellite platforms these algorithms are seen to be modified for tri-stereo matching (Zhang and Gruen, 2006) (Mozervo et al., 2009). There are not many methods that are proposed for tri-stereo matching and even if proposed, are the extension of the binocular stereo matching methods (Tack and Gossens, 2012). Also, these methods are more constrained by the multi-resolution and the geometrical constrains enforced by imaging geometry (Baltsavias et al., 2006) . In contrast to the above approaches, there is an altogether different class of algorithms that uses segmentation-based stereo matching (Tan et al., 2014). In the proposed work the same strategy has been used on the *disparity space image* (Intille and Bobick, 1994) for the tri-stereo image matching.

The local stereo algorithms employ the block matching using correlation-based methods, such as Normalized Cross Correlation (NCC), Sum of Squared Difference (SSD), Mutual Information (MI), etc. In these methods, the disparity is computed using the "winner-takes-all" optimization, which sometimes gives incorrect results due to steep slope, high contrast object in the neighborhood, or a large search range. By applying global stereo matching algorithms, the correspondence of the conjugate points is solved, using the global optimization by applying dynamic

Table 1: Details of the datasets.

Mission	Sensor	Image ID	Image Name
CH1	TMC	TMC_NRA_20090418T093900690.IMG	
	TMC	TMC_NRA_20090418T093900690.IMG	Gassendi G
	TMC	TMC_NRA_20090418T093900690.IMG	
CH1	TMC	TMC_NRA_20090418T095023392.IMG	
	TMC	TMC_NRA_20090418T095023392.IMG	Marius D
	TMC	TMC_NRA_20090418T095023392.IMG	
ALOS	PRISM	ALPSMB201033270	
	PRISM	ALPSMF201033160	Mumbai 1
	PRISM	ALPSMN201033215	Mumbai 2
Cartosat-1	PAN	075259000201	Ground truth for
	PAN	075259000202	Mumbai 1, Mumbai 2
LORC	LOLA	LDEM 1024 00N_15N_300_330	Ground truth for Gassendi G
	LOLA	LDEM 1024 30S_15S_300_330	Ground truth for Marius D
Middlebury		Lampshade	Lamp
Stereo		Rocks	Rocks
ISPRS-EuroSDR	Aerial	40_0313_PAN	München1
project	Sensor	40_0314_PAN	München2
	DMC II 230	40_0315_PAN	

programming, belief propagation, graph-cuts, etc. where the smoothness and penalty terms take care of these problems. For local methods, this can be solved efficiently if more constraints are used. One of the widely used constraints is the epipolar constraint. For the local image matching, this epipolar constraint has to be augmented with additional steps due to the resulting streaking effect because the disparity is determined for a scan line independently. An additional constraint can be utilized if instead of two-view stereo a triplet of the image is used. Some of the approaches are found in the literature on the use of triplets for 3D modelling as by Raggam (2006); Zhang and Gruen (2006), which according to Tack and Gossens (2012) are only the extensions of the standard stereo pair matching based approaches. They suggested a composite approach based simultaneous use of the tri-stereo images for the best fit of three convergent image rays instead of the trivial intersection of two lines and incorporation of more images with different viewing angles to increase the probability of a successful match, being a unique and a correct solution.

Overall, from the literature, it is observed that there is a need to investigate the issues of (a) Limitation of the WTA to get a disparity at the global minimum for an ill-posed stereo matching problem. (b) A proper approach to make adequate use of the epipolar based local stereo matching using *Disparity Space Image*, and (c) Integration of the redundant information from the triplet of image to arrive at the optimized solution for the disparity estimation. In the present work, an attempt is made to develop a novel approach based on local minima of the stereo matching cost to determine the disparity where the global minimum deviates from the actual disparity as it exists in the real world setting. The hypothesis has been investigated using the controlled experiments on the stereo triplets with reliable ground truth. From the experiments, it has also been observed that application of the watershed segmentation to the disparity space image (DSI) can effectively extract the local minima and therefore, can be utilized for extracting the local minima. Tan et al. (2014) used watershed transformation for the stereo matching, where they applied this transformation on stereo pairs whereas, in the present studies the idea is extended to DSI.

Test Datasets

The experiments are performed for the remotely sensed images from the aerial and the satellite platforms, and the standard stereo images from Middlebury stereo dataset. In all, eight datasets are used for experimentations from four different sources. Two dataset namely Gassendi G and Marius D are from CH-1 mission's TMC sensor that covers lunar craters. Next two dataset namely Mumbai1 and Mumbai2 are from ALOS mission's PRISM sensor that covers urban and hilly terrain near Mumbai. The other two dataset München1 and München2 are very high-resolution aerial images that contain urban area of München (Munich) city, obtained from EuroSDR project (Haala, 2013). The final two dataset are from Middlebury stereo dataset that cover indoor scene. For the lunar data, DSM from the same source is used as ground truth. For ALOS PRISM, DSM generated from Cartosat-2 is used as the reference. For München dataset ground truth is available from the same source. Middlebury dataset also provides ground truth. Table 1 gives the details of the dataset and Figure 1 shows Nadir (centre) image of the triplet.

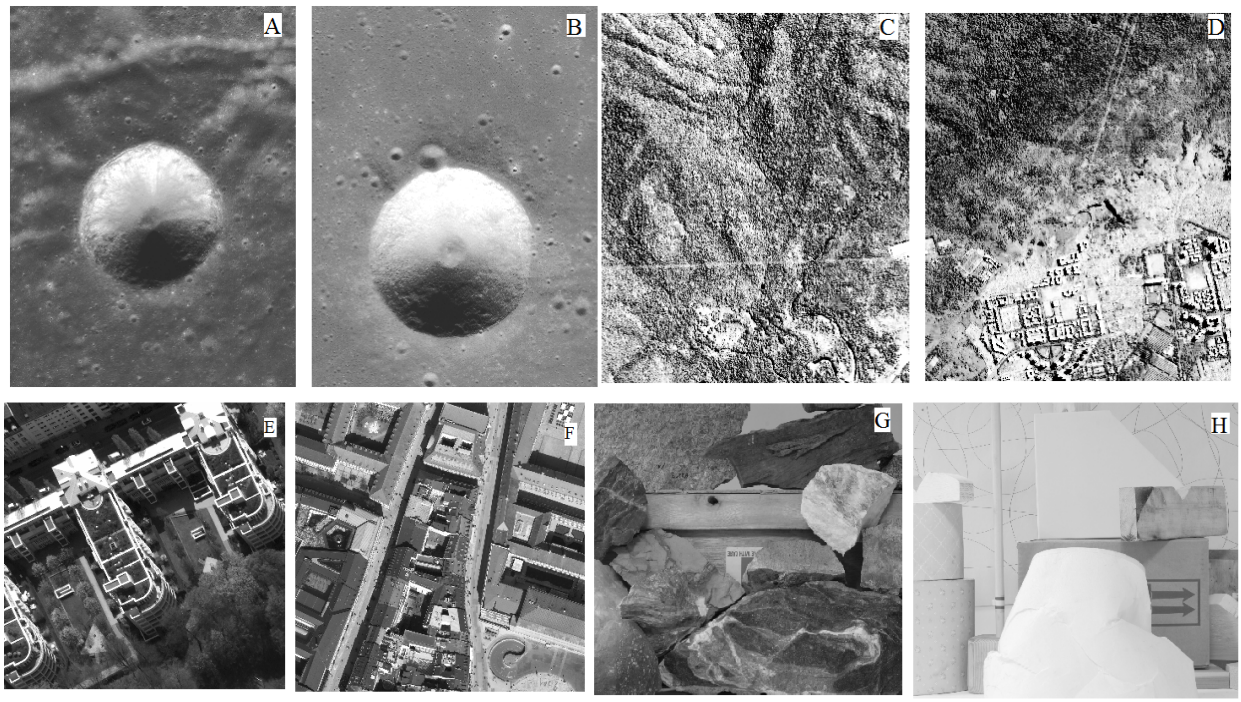

Figure 1: Nadir (Centre) view of Dataset Used (A) Gassendi G Crater and (B) Marius D Crater from CH-1 TMC (C) Mumbai1 and (D) Mumbai2 from ALOS PRISM (E) München1 and (F) München2 from EuroSDR dataset (G) Rock and (F) Lamp dataset from Middlebury stereo dataset.

Methodology for Triplet Matching

The methodology is divided into two sub-problems. The first one is the initial matching and then the disparity refinement. In the first step the Initial Correct Points (ICPs) are extracted and then density is increased by finding the best solution from various Probable Matches (PM). The work flow for determining ICP and PM is shown in Figure

Once the images are resampled the next stage is triplet matching, the following steps discuss the process. The work-flow for determining ICP and PM is shown in Figure 2

1. Disparity Space Image (DSI) Generation: Using Equation (1) the stereo matching cost for a pixel in the left image to all the pixels on its conjugate epipolar line in the right image is determined. This is carried out for all the pixels in the left scan line (i_0) and the cost is stored in DSI. In Figure 3 (A-B) left input image, and ground truth. Figure 3 (C) shows the DSI for the scan line highlighted in Figure 3 (A).

$$C(i_0, j, d) = \sum_{y \in w} [I_1(i_0, j) - I_2(i_0, j + y)]^2 \tag{1}$$

   where, $y \in \{1, 2, 3 \ldots d_{max}\}$
2. Cost volume Generation: The process is carried out for all the scan lines in the image and thus a cost volume is generated for left-centre and the another one for the centre-right stereo pair.
3. Initial Disparity: Due to virtue the in which the cost is stored in the DSI, the forward (left to right) stereo matching cost is stored in the columns and the reverse (right to left) stereo matching cost is stored in the rows. Figure 3 (D) shows initial disparity for complete image and Figure 3(E) red dots shows the initial disparity obtained in the forward direction for DSI. The alternative way to determine the two initial disparities in the forward and the reverse direction for each pixel are given as in the following equation.

$$D_F(i_0, j) = \underset{d}{\operatorname{argmin}}\, DSI(i_0, j, d) \tag{2}$$

$$D_R(i_0, j) = \underset{j}{\operatorname{argmin}}\, DSI(i_0, j, d) \tag{3}$$

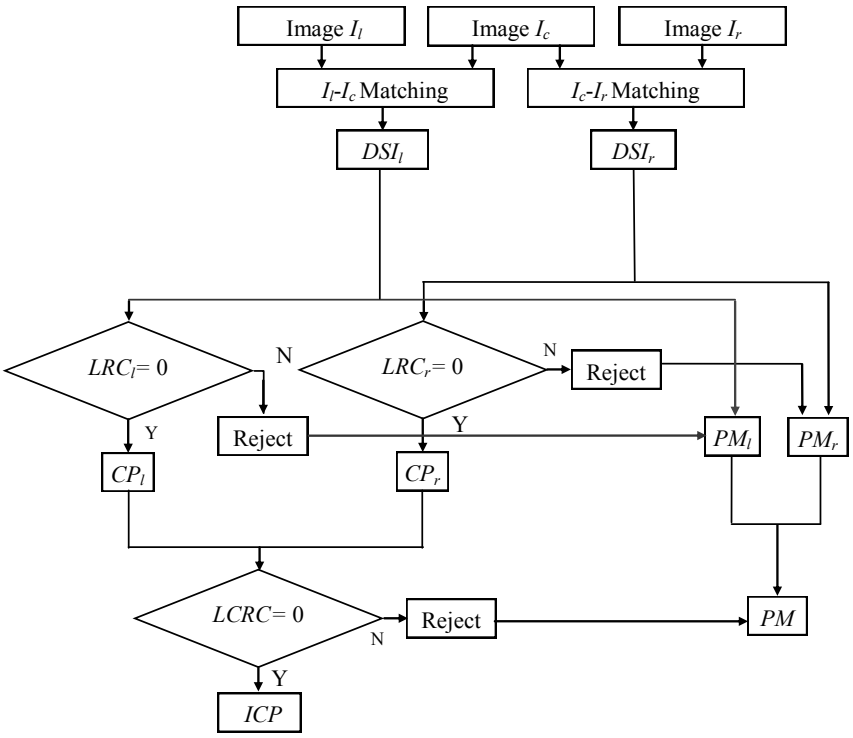

Figure 2: Work-flow for determining PM and ICP.

4. Initial Correct points (ICP): The disparities that are Left-Centre-Right Consistent(LCRC) as given in the following equation are termed as ICP, these are represented as red points in Figure 3 (F).

$$LCRC = - \mid LRC_{lc} - LRC_{cr} \mid \tag{4}$$

where, the left-right consistent disparities (LRC) are given in Equation (5) and LRC_{lc} and LRC_{cr} represents disparities obtained from left-centre and centre-right pair.

$$LRC = - \mid D_F - D_R \mid \tag{5}$$

In above Equations (4) and (5) the value of LCRC and LRC is set to '0'.

5. Probable Matches (PM) : The PM are local minima obtained from watershed transformation (WST) using Meyer's algorithm (Meyer, 1994) applied to *inverted DSI*. (Watershed lines connect local maximum of the terrain and as we require local minimum we inverted the DSI). Figure 3 (F) black line shows the watershed lines for DSI in Figure 3 (E).
6. Secondary disparity using watershed : The final step is to find the disparity at the locations where ICP are absent. The problem is similar to find a *shortest path* between ICP which is constrained by the watershed line. The path between the two ICP is the *shortest geodesic path*. To fill these holes, the discontinuity in the profile and its immediate left and right neighboring ICP are selected. The geodesic distance is determined between these two ICP that is constrained by the existing possible paths given by PM. The path with the least geodesic distance is selected as the disparity. The shortest path selection using geodesic distance is shown in Figure **??**. This process is followed for each scan line and the disparity map is generated for the complete set of the image as shown in Figure4(A)
7. Final disparity using watershed : The disparities are determined for each scan line without taking in consideration to adjacent scan lines. Therefore, a streaking effect is observed. To avoid the error and enforce smoothness in the obtained disparities in a scan line to the adjacent scan lines, disparities are determined in four different directions Figure 4 (A-D) and the mode of these four disparities is considered as the final solution Figure 4 (E).

The limitation with the best path is that it is not always a unique one, and therefore, at many places an error is observed. This is more prominent at the depth disconnectedness and the smooth regions. To remove these errors, the disparity obtained is refined. This final step of disparity refinement is detailed in the following section.

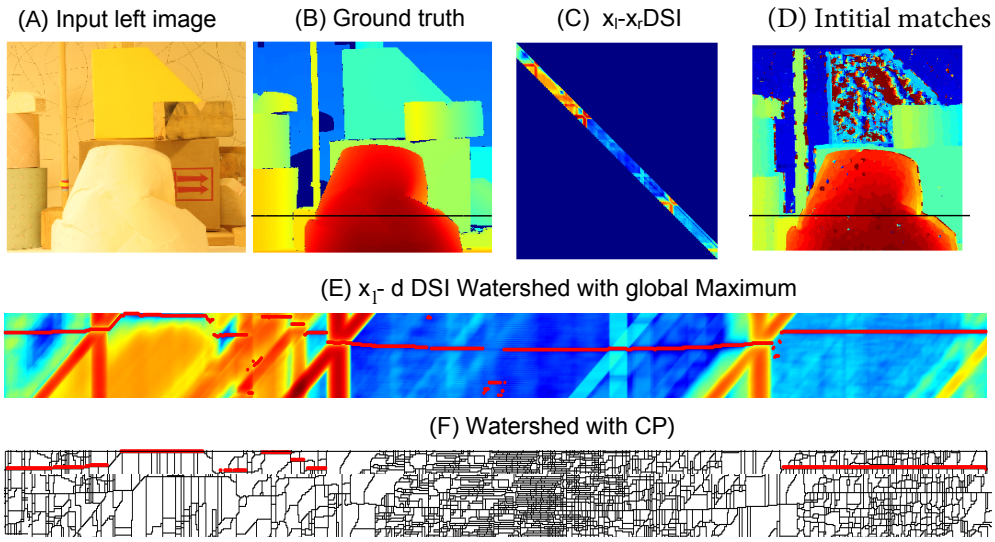

Figure 3: (A) Input images from Middlebury stereo dataset (B) Ground truth (C) Initial disparity (D)$x_r - x_l$ DSI for the scan line highlighted in (E) $x_l - x_d$ DSI for the scan line highlighted in (c) with the global minimum of each column as initial disparity for that column (F) Watershed transformation applied to above DSI showing the watershed lines (local minima) in black and LRC consistent matches in red.

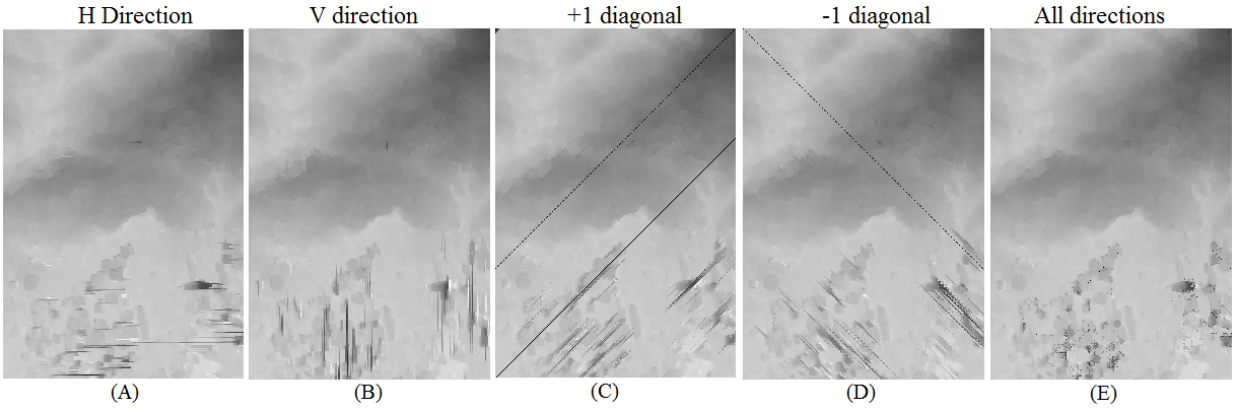

Figure 4: Results on Mumbai1 and Lamp dataset.

Results and Analysis

The proposed method determines the disparity per scanline hence, the methods that are suitable for comparison are scanline based stereo matching methods. To evaluate the proposed method, it is compared with the three scanline based methods, namely Dynamic Programming (DP), Semig-lobal Matching (SGM) and basic block matching (BM) technique. Eight different datasets cover a variety of features with a wide range of disparity and complexities. From the results obtained by applying for the proposed work, it has been observed that after considering ICP as correct match a significant number of correct points are added using PM match. The relative error from two observations has been reduced, as the points to be considered has already been filtered using Left-Right Consistency (LRC) check. The results for datasets with different resolutions are obtained non-consistent. The obtained results can't be generalized and therefore these are discussed in according to the dataset used.

Here remotely sensed datasets of three different resolutions are used. The ground resolution of ALOS and CH-1 are in meters, in these cases the results from proposed method outperform the results from the other methods. The accuracy is increased by 4-8 % for different datasets with respect to SGM and by 20 % with respect to DP. As observed from the results obtained from München2 dataset of 10 cm resolution the density of the points by the proposed method is of concern. For München1 the accuracy of the proposed method is 2 % greater to SGM and for München2 SGM is better to the proposed method by 6%. This can be improved by considering more directions. It can be thus concluded that the proposed method performs best at coarse resolution. For Middlebury dataset, SGM performs better at smooth regions with respect to the proposed method as observed from Lamp dataset by 6%. For

Table 2: Quantitative summary of correct matches obtained by various techniques for tri-stereo matching Matches.

Dataset	Ground Resolution	Size	% Correct Matches				
			BM	DP	SGM	Proposed Stereo Matching Method	
						1 Direction	4 Directions
Lamp	–	1031 × 1031	65.2	71.2	**95.1**	75.8	84.3
Rock	–	1031 × 1031	92.0	93.5	**97.4**	94.2	96.1
München1	0.1 m.	1164 × 800	81.2	88.0	**89.6**	84.3	85.1
München2	0.1 m.	1164 × 800	74.7	81.0	**89.1**	71.2	75.0
Mumbai 1	2.5 m.	2200 × 1851	74.2	87.7	91.2	92.3	**95.0**
Mumbai 2	2.5 m.	1700 × 1851	68.1	91.6	92.0	91.2	**91.8**
Marius D	5.0 m.	820 × 820	81.2	77.2	87.0	87.6	**94.7**
Gassendi G	5.0 m.	820 × 820	83.2	75.0	88.2	90.0	**93.3**

textured regions as observed from rock dataset both the methods, SGM and the proposed give the accuracy of 97 %. The probable reason of comparatively poorer results at smooth regions in higher resolution images is due to the fact that the watersheds from DSI obtained these images are too close, and they deviate the shortest path which eventually led to the error. For standard Middlebury images, SGM gives the best results.

In Table 2 the quantitative comparison of the proposed method has been presented with BM, DP and SGM. As there in no direct method to integrate the tri-stereo matching for these methods, the disparities are determined for both the views and the best disparity is considered for comparison to the proposed method. The results suggest for all the datasets that the proposed method outperforms traditional BM and DP and gives similar results as compared to the SGM. This is evident as both the algorithms consider different directions for determining the correct disparity.

This was a quantitative analysis. The results obtained are shown in Figure 5. To evaluate the results obtained qualitatively different regions of the images are chosen, and the results are compared. Table 3 gives the summary of the results for various regions in various images. Various features that are common to remotely sensed images are identified and the performance of all the algorithms are evaluated. In most of the cases the proposed algorithm gives similar results to the SGM as expected as it is similar to SGM except that in SGM the cost is determined for each pixel in the cost volume and its minimum is the solution whereas in the proposed method, only the local minima that are corresponded by the watershed line is the solution. The quality of matches at various topological features are investigated, and it has been observed that for all types of complexities the SGM and the proposed method gives best results. They can handle errors due to non-planer features, and gives correct matches to building façades as well. The object boundaries are also well preserved as presumed. As the continuity and smoothness in the four different directions is preserved it easily handles the mismatch that generally occurs due to repetitive patterns. The method is capable enough to deal with the complex structures such as trees similar to SGM. The method has all the merits of the SGM and outperforms it in terms of speed. As in the case of SGM, every pixel in the cost volume is considered for the cost the aggregation whereas, in the proposed method there is no aggregation and the search is limited to the local minima.

Conclusion

In the stereo matching using winner-take-all approach, the global minimum of stereo matching cost is considered as the correct disparity. There are instances when this criterion may end up in an error, such as when the image has a large search range or has similar artifacts with very high contrast as compared to neighboring pixels. To solve this problem, in the present thesis, a method is formulated that considers the local minima of the cost function to give the correct disparity based on delineation of the local minimum that corresponds to the correct match.

1. It has been observed that the watershed lines, when extracted from the disparity space image, are co-located with the local minima, and hence, the watershed segmentation on the disparity space image can be used to delineate local minima. The advantage of using watershed segmentation is that the watershed lines are connected, whereas the local minima are disjoint. The correct watershed line is determined using information from the initial correct matches and to traverse the shortest path for locations where such matches are absent. This also enforces the continuity constraint while determining the disparity.

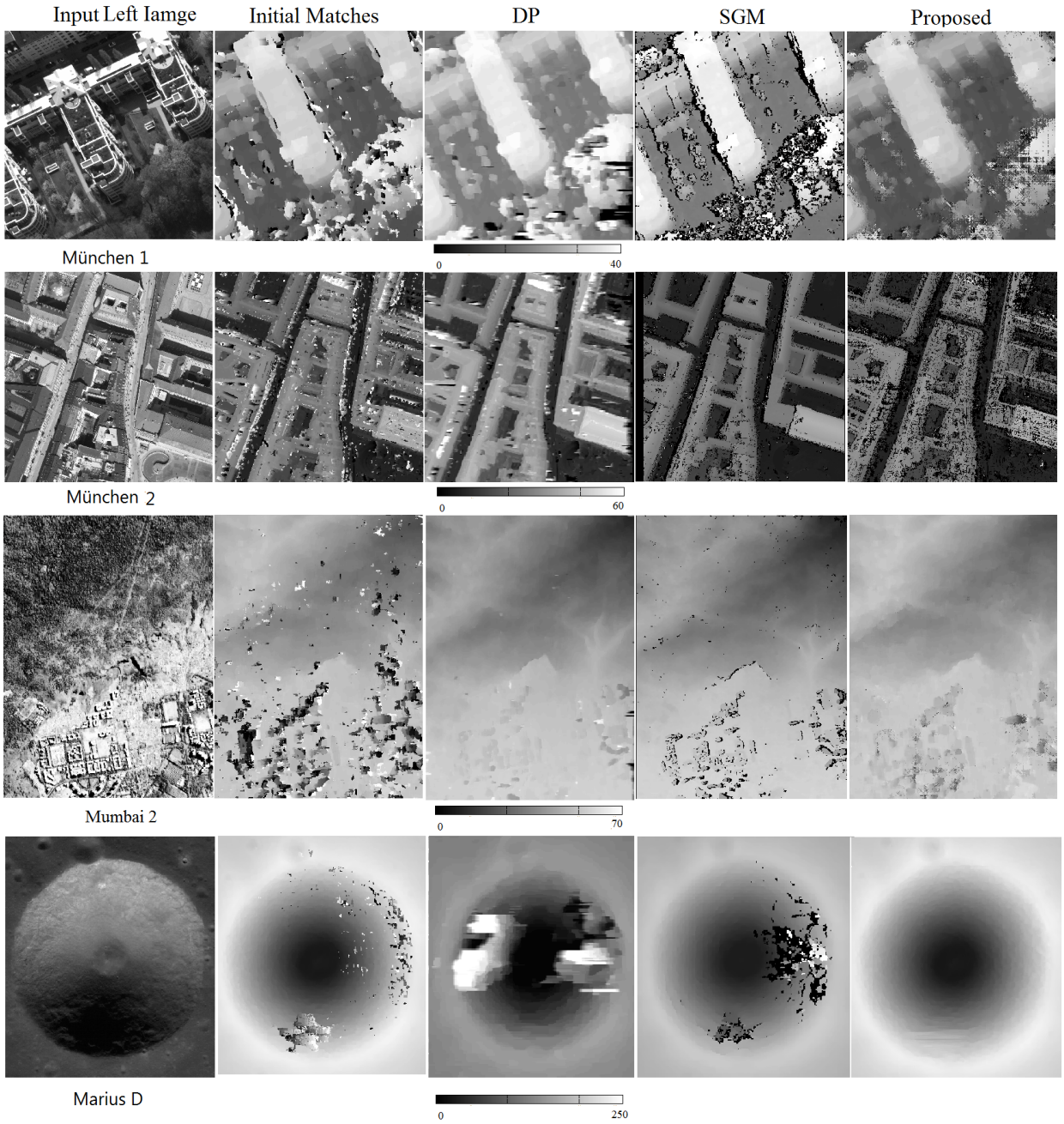

Figure 5: Results obtained for München1 München2 Mumbai2 and Marius D crater. (with disparity range). Results shows disparity maps obtained from BM, DP, SGM and Proposed method.

Table 3: Summary of qualitative analysis of the matching algorithms

	Regions	BM	DP	SGM	Proposed
1	Planer features (ground,rooftop)	Correctly Mapped	Correctly Mapped	Correctly Mapped	Correctly Mapped
2	Non-planer features	Incorrectly Mapped	Correctly Mapped	Correctly Mapped	Correctly Mapped
3	Buildings Façade (side walls)	Error	Sharp, streaking effects	Recovered	Recovered
4	Horizontal Discontinuities	Error	Sharp streaking effects	Recovered	Recovered
5	Texture less regions	Errors	Errors	Recovered	Recovered
6	Continuous textured regions	Errors	Errors	Recovered	Recovered
7	Region boundaries	Errors	Recovered in the horizontal direction	Recovered in all directions	Recovered in all directions
8	Performance speed ranking	1	2	4	3
9	Trees	Errors	Errors	Extracted	Extracted
10	Repititive patterns	Extracted within limit	streaking effects	Extracted	Extracted

2. To consider the smoothness of the surface in different directions and to get rid of the streaking effect (as in the case of dynamic programming approach), the disparity is determined in four directions similar to the semi-global matching approach. Therefore, at a given pixel location we have four disparities that correspond to the shortest path based approach in that layer. To find the correct disparity, out of these four 'mode' operation is used.

3. It has been observed that in the smooth regions, the shortest path may not be unique and hence, leads to incorrect disparity. In such cases, the disparity can be refined using connectivity information of the local minima and the global minimum from the triplet of images using the two disparity space images as evident from the present studies wherein it considers the connectivity of these probable matches with the initial correct points. It has been observed that using this measure there is a significant increase in the number of correct matches at the smooth regions.

4. As the future scope of developments on these lines, the distribution of the matches can be considered and quantitative analysis can be carried out using high-resolution ground truth from other sensors. In the present work, a possible stereo pair of the left and right image has not been used since they need special treatment due to large baseline, however a methodology can be developed for this pair.

References

Alobeid, A. (2011), Assessment of Matching Algorithms for Urban DSM Generation from Very High Resolution Satellite Stereo Images, PhD thesis, Geodäsie und Geoinformatik der Leibniz Universitä, Hannover, Germany.

Baltsavias, E., Li, Z. and Eisenbeiss, H. (2006), "DSM generation and interior orientation determination of IKONOS images using a testfield in Switzerland", *Photogrammetrie, Fernerkundung, Geoinformation* , Vol. 1, pp. 41–54.

Gruen, A. (1985), "Adaptive least squares correlation: a powerful image matching technique", *South African Journal of Photogrammetry Remote Sensing and Cartography* , Vol. 14, pp. 175–187.

Haala, N. (2013), The landscape of dense image matching algorithms, *in* 'Photogrammetric Week', pp. 271–278.

Hirschmüller, H. (2011), Semi-Global Matching Motivation, Developments and Applications, *in* 'Photogrammetric Week', Stuttgart, Germany, pp. 173–184.

Hirschmüller, H. and Scharstein, D. (2008), "Stereo processing by Semi Global Matching and Mutual Information", *IEEE Transactions on Pattern Analysis and Machine Intelligence* , Vol. 30, pp. 328–341.

Intille, S. S. and Bobick, A. F. (1994), Disparity-space images and large occlusion stereo, *in* 'European Conf. on Comput. Vision', Vol. 801 of *Lecture Notes in Comput. Series*, Springer Berlin Heidelberg, pp. 179–186.

Jazayeri, I. and Fraser, C. (2010), "Interest operators for reature-based matching in close range photogrammetry", *The Photogrammetric Record* , Vol. 24, pp. 24–41.

Meyer, F. (1994), "Topographic distance and watershed lines", *Signal Processing* , Vol. 38, pp. 113 – 125.

Mozervo, M., González, J., Roca, X. and Villanueva, J. J. (2009), Trinocular stereo matching with copmposite disparity space image, *in* 'IEEE International Conference on Image Processing', Cairo, Egypt, pp. 2089–2092.

Raggam, H. (2006), "Surface mapping using image triplets: Case studies and benefit assessment in comparison to stereo image processing", *Photogrammetric Engineering and Remote Sensing* , Vol. 72, pp. 551–563.

Remondino, F., Spera, M., Nocerino, E., Menna, F. and Nnex, F. (2014), "State of the art in high density image matching", *The Photogrammetric Record* , Vol. 29, pp. 144–166.

Tack, F. and Gossens (2012), "Assessment of photogrammetric approach for urban DSM extraction from Tri-Stereoscopic satellite imagery", *The Photogrammetric Record* , Vol. 27, pp. 293–310.

Tan, X., Sun, C., Sirault, X., Furbank, R. and Pham, T. D. (2014), "Stereo matching using cost volume watershed and region merging", *Signal Processing: Image Communication* , Vol. 29, pp. 1232 – 1244.

Veksler, O. (2005), Stereo correspondence by dynamic programming on a tree, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', Vol. 2, San Diego, CA, USA, pp. 384–390.

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