Study of stereo matching algorithms used in Satellite Stereo Pipeline for the 3D reconstruction of digital elevation models.

BEDDIAF Amir, BERDEAUX Alexandre, BOULAHBAL Houssem, BOURGOIS Astrid and SADOU Loïc

Abstract—In this paper, we study the different stereo matching algorithms used in the software S2P and the possibility of expanding it to build a DEM as a 3D mesh. We offer a comparison of the stereo-matching algorithm following criterias such as robustness to artifacts. Then we explore the state of art of the methods reconstructing a 3D mesh from a point cloud generated by S2P.

Keywords—S2P, stereo-vision, stereo vision, stereo-matching, stereo matching, satellite image processing, mesh, point cloud

I. Introduction

With many satellite observation missions in orbit, there is a great abundance of material allowing accurate three dimensional reconstruction of earth's surface. From Lidar telemetry to optic stereoscopic imaging, Reconstruction data can be obtained from different methods, resolutions and qualities. Such data may not be easily available to the potential users but commercial distribution of image from optic satellite like Pleïades or Ikonos allow the use of stereo-vision methods for the construction of a digital surface model (DSM).

The satellite stereo pipeline (S2P) is a software that allow the extraction of a DSM from satellites images and their corresponding rational polynomial coefficient (RPC) in separate xml files. Satellite images come with additional challenge in the context of stereo-vision that need to be overcome before applying standard stereo-matching algorithm. It need divides the images in tiles and then makes a stereo-rectification on said tiles using affine camera approximation [1]. It must also apply a correction to the images to ensure that the epipolar line associated with a view cross its corresponding pixel in the other one [2]. Once those problem solved, it uses a stereo-matching algorithm chosen by the user to extract a DEM of each tiles as a 3D point cloud [3].

This pre-processsing step allows to abstract the specificities of satellite imaging and focus the choice on efficient stereo-matching techniques. The cloud point represents a sampling of the surface, in order to reconstruct a 3D surface we need to process the cloud point into a 3D mesh.

In this paper we will give direction on the way to use S2P in order to build a DSM in the form of a 3D surface. Seeing as the specificities of satellite imaging are handled by the software we will set or focus on two major development choice, the stereo-matching algorithm used and the technique of reconstruction from the given cloud point [3].

First we will present the software S2P which natively offer stereo matching of images, presenting different categories of algorithm and putting some focus on the point cloud artefact that can appear and influence the transformation into a mesh. Then we will present a selection of pertinent 3D reconstruction method that can be used to transform the point cloud into a 3D surface as a mesh.

II. Stereo Matching

A. General concepts

Stereo can in general be identified as belonging to two broad categories : local or global methods. They are also comprised of a subset of component namely the matching cost computation, cost aggregation, disparity optimization and disparity refinement [4].

Matching cost associate a pixel coordinate and a disparity, this create a space that we will try to separate on each pixel coordinate with a disparity of minimal cost. There is a great number of potential matching cost from ones using intensity differences such as SSD (Sum of Squared method) to matching specialized in handling specifics imaging problem like sampling [5]

Computing disparities with a local methods involve to sweep the secondary image with a patch centered on each pixel x' like:

$$x' = x + d \tag{1}$$

Where d is the disparity value for each pixel x' and x is the pixel in the reference image which we are looking for a correspondence. The disparity $d(x)$ is the value of d that provide the minimum value of a cost function for each pixel x' in the secondary image [6].

There are two difficulties during the disparity computation with a local algorithm:

- Matching ambiguities: it happens when there is a repetitive pattern in both reference and secondary image. The patch used to compute the disparity return similar results for identical contexts, thus the repetitive patterns are identical contexts but at different places. This is why repetitive pattern (or textureless area) generate matching errors, they will make ambiguities on the choice of the corresponding pixels in the secondary image.
- Depth discontinuities: This type of discontinuity is challenging. It occurs when an object have a big depth difference with its background. Computing the disparity use patch correspondences. However, high depth discontinuity generally produce occlusion and dis-occlusion (Fig 1). Occlusion or dis-occlusion are parts of the

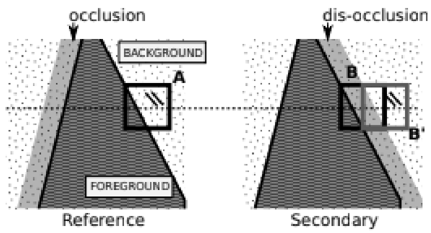

Fig. 1: Sketch showing occlusion and dis-occlusion situations Source: [6]

reference image that respectively disappear or appear in the secondary image due to the modified point of view between the two images. These differences are problematic when computing the disparity of a point on an edge of depth discontinuity, a part of the patch can compute an occlusion or dis-occlusion area and the resulting disparity will be low even if the point is theoretically the correct corresponding point.

A reliable local algorithm have to deal with this two main difficulties for decent results.

On the other hand global method do not use a window of neighbors to match pixels but formulate the problem as an energy minimization problem. We then try to find a disparity function that minimize global energy presented by the sum of the costs of pixels in the image and a smoothness term that usually measure the disparity difference of neighbor in a function ρ [4].

Another class of global method is the use of dynamic programming, the previous optimization being in the two dimension it can be NP-hard. The minimal cost is computed by using dynamic programming to match scan-lines, there is problem for enforcing inter-scan-line consistency that result in streaking artifact [4].

B. Point Cloud Artifacts

The point cloud generated by the matching algorithms present various artifacts. understanding the nature and scale of these artifacts will allow us to choose the right algorithm of mesh reconstructing. we provide below the artifacts that have the most impact on the reconstruction of satellite images: [7]

- Sampling error : The distribution of point sampling is not uniform. The points generated by stereo matching algorithms can produce a non-uniform sampling patterns that may be due to the orientation of the scanner as well as to the geometry of the observed shape.
- Noise : Points that are randomly distributed near the surface.distribution of the noise is a function of sensor noise, depth quantification and distance or orientation of the surface from the scanner.
- Outliers : Points that are away from the original surface are classified as outliers. Outliers are due to structural artifacts in the acquisition process. In some cases, outliers are distributed randomly in the volume, where their density is lower than the density of the points sampling the surface.

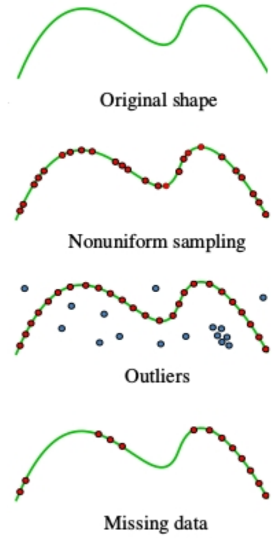

Fig. 2: Different forms of point cloud artifacts Source: [7]

Procedure TV-L^1_optical_flow($I_0, I_1, u^0, \tau, \lambda, \theta, \varepsilon, N_{maxiter}, N_{warps}$)

$\mathbf{p}_1 \leftarrow (0,0)$
$\mathbf{p}_2 \leftarrow (0,0)$
for $w \leftarrow 1$ **to** N_{warps} **do**
  Compute $I_1(\mathbf{x}+\mathbf{u}^0(\mathbf{x})), \nabla I_1(\mathbf{x}+\mathbf{u}^0(\mathbf{x}))$ using bicubic interpolation
  $n \leftarrow 0$
  while $n < N_{maxiter}$ **and** *stopping_criterion* $> \varepsilon$ **do**
    $\mathbf{v} \leftarrow TH(\mathbf{u}, \mathbf{u}^0)$
    $\mathbf{u} \leftarrow \mathbf{v} + \theta \operatorname{div}(\mathbf{p})$
    $\mathbf{p} \leftarrow \frac{\mathbf{p}+\tau/\theta \nabla \mathbf{u}}{1+\tau/\theta|\nabla \mathbf{u}|}$
    $n \leftarrow n+1$
  end
end

Fig. 3: TV-L1 algorithm by Janvier Sánchez and co Source: [8]

- Missing Data : Missing data are the details that were lost during the matching part. Many factors can lead to missing data such as limited sensor range, high light absorption and occlusions in the scanning process where large parts of the shape are not sampled.
- Occlusion : Points that are visible from one point of view and invisible from the other. As a result, some pixels will not have a match.

C. Stereo-matching algorithm used in S2P

1) Total Variation minimization with L^1 Norm (TV-L1): TV-L1 (Total Variation minimization with L^1 norm) A method based on the minimization of a function which contains a data term using the L1 norm and a regularization term using the total variation of the flow. This formulation allows discontinuities in the flow field, while being more robust to noise than the classical approach by Horn and Schunck. [8]

This algorithm (Fig 3) takes 9 parameters. I_0 and I_1 are the images; u^0 is given by the enclosing multi scale procedure and it is zero at the coarsest level; τ is the time step with default value 0.15; λ is the data attachment weight with default

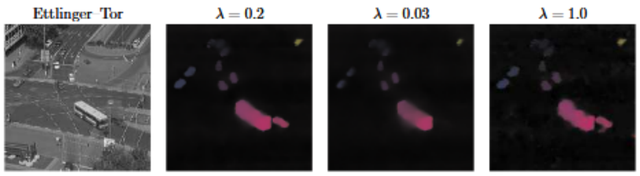

Fig. 4: EttlingerTor example with different value of λ
Source: [8]

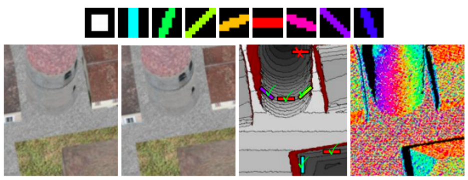

Fig. 5: Images showing the window shape used to compute the best disparity for some points
Source: [6]

value 0.15; θ is the tightness with default value 0.3; ε is the stopping criterion threshold with default value 0.01; N_{scales} and N_{warps} are respectively the number of scales and the number of warps with default value 5. [8]

This algorithm allows to detect the displacement of the objects in a scene like in (Fig 4). A small values of return a smoother solution with underestimate optical flow for the moving objects. However, big values of return a more important attachment term. This doubled with increased sensitivity to noise, resulting in unstable flow fields [8]. This algorithm can detect geometric shapes in a scene as well as their displacements. We do not have result for satellite images with this algorithm.

2) Multi-Scale Multi-Window (MSMW): This method aims to solve the two problems of local methods (previously explained). MSMW can be split in two parts: the multi-scale and the multi-window. Therefore, we will introduce those two methods behaviors beginning with the multi-window aspect and finishing with the multi-scale aspect. The main goal of the Multi-Window concept is to solve the depth discontinuities problems by adapting the patch shape to the discontinuity shape. Basically, instead of sweeping the secondary image with a 5x5 (for exemple) patch which compute all its 25 pixels, we use several patch shapes. The (Fig 5) shows, for some points, which window shape give the best disparity, it demonstrates a real adaptation to the discontinuities borders. Analyzing the context of pixels using only the pixel in the same depth shows very good results for depth discontinuities and also for slanted surface.

While the Multi-Window concept has a good answer to the discontinuity problem, Multi-Scale concept aims to solve the problem of repetitive patterns or lack of textures. This method is based on a scale pyramid, each level is determined by subsampling and blur (convolution with a Gaussian kernel) the previous level. Thus, the final level shows only low frequency of the originals images giving coarse information and then eliminate noise. Executing stereo matching on a specific level restrict the search area at the previous level then it reduce the computation time and the probability of analyzing repetitive patterns.

By combining the concepts of Multi-Window and Multi-Scale, MSMW offer a good answer to the local methods difficulties. Moreover there is plenty of match criteria to apply at each scale [6], so we can say MSMW is more focused on reliable disparity than a complete disparity.

3) Semiglobal Matching (SGM): SGM is a global method using Dynamic Programming in order to optimize matching following multiple directions around the pixel including [9]. However in a rectified image the presence of the best match inside the search space is guaranteed [1].

It is based on compensating radiometric differences of input images by using the Mutual Information as a matching cost function for pixel-wise matching. This method performs four principal steps of stereo matching in addition to another step which is multibaseline adaptation. This step is optional for satellite images. These same four steps (pixelwise matching cost calculation,cost aggregation, disparity computation, disparity refinement) define the Scharstein and Szeliski's taxonomy [4] used on Middleburry's stereo pages to compare the different algorithms.

- Pixelwise Matching Cost Calculation : In state of the art two methods for computing the matching cost used by SGM, the first concept used is the sampling insensitive measure of Birchfield and Tomasi [5], the cost is defined as the absolute minimum difference of intensities at the base image pixel p and match(target) image pixel q. The second method consists on the Mutual information (MI) which is insensitive to recording and illumination changes [9]. The SGM algorithm implemants the HMI method (hierarchical MI computation), the implementation of all the method is detailed on Hirschmuller's works [9]
- Cost Aggregation : noise, outliers and other artifacts may affect the computing of the Pixelwise cost, So in order to minimize the effect of these artifacts a second constraint was introduced to support the smoothness constraint, this constraint consists on penalizing changes of neighboring disparities. a function that relates these two constraints has been introduced :

$$E(D) = \sum_{p} \Big(C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \Big) \quad (2)$$

after introducing this energy function the problem of stereo matching is reformulated in Hirschmuller's works as finding the disparity image that minimizes the energy E(D). this reformulation is a global minimization in 2D, which is NP-complete, the solutions suffers easily from streaking.
The idea of Hirschmuller is aggregating matching costs in 1D from all directions equally (Fig 6),the principle is presented by equations :

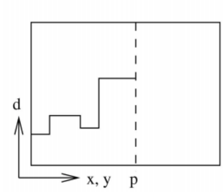

Fig. 6: Cost Aggregation in disparity space Source: [9]

$$\begin{aligned} L_r(p,d) =& C(p,d) + \min(L_r(p-r,d), \\ & L_r(p-r,d-1) + P_1, \\ & L_r(p-r,d+1) + P_1, \\ & \min_i L_r(p-r,i) + P_2) - \min_k L_r(p-r,k). \end{aligned} \tag{3}$$

- Disparity Computation : the computing of disparity images D_b et D_m for the pair of base and match(represents the match image of I_b) images is based on the same method used in local stereo methods, which consists on defining the disparity d that corresponds to the minimum cost.
  for the artifacts present on the disparity images D_b and D_m are filtered using median filter with a dimension of 3*3 .
  Also the disparities can help us to determine other artifacts like occlusions and mismatches by performing a consistency check.
  this calcul is presented in these equations:
  $$D_p = \begin{cases} D_{bp} \quad if \quad |D_{bp} - D_{mq}| <= 1 \\ D_{inv} \quad\quad otherwise \end{cases}$$
- Disparity refinement Disparity images resulted may always presents some artifacts or invalid disparities, so the need to deal with these errors such as discontinuities due to textures or noise ,holes or outliers. by filtering the peaks, interpolating the occlusions and mismatches which called Discontinuity Preserving Interpolation.

4) Semiglobal Block Matching (SGBM): This algorithm is a modification of SGM replacing pixel by pixel matching with block matching. It also change the cost function to the sub-pixel metric detailed in [5]. The studied implementation only do a single pass of dynamic programming and also use pre- and post-filtering of the image (enforce unique match, filter speckling, interpolation) [10].

5) More Global Matching (MGM): MGM algorithm is an apporixamation of global methods for stereo matching driven from SGM algorithm in order to treat streaking artifacts present in SGM. this new version uses the same pixelwise cost calculation and disparity computation with different aggregation of costs. The energy minimization of this algorithm is reduced by the factor of 5 in comparison with SGM. SGM and streaking artifacts: two adjacent in the structure used in SGM will share only the horizontal line plus 4 points of intersection (Fig 7). When the data term on the horizontal line is weak, i.e. all disparity hypothesis are equally plausible [11], the vertical

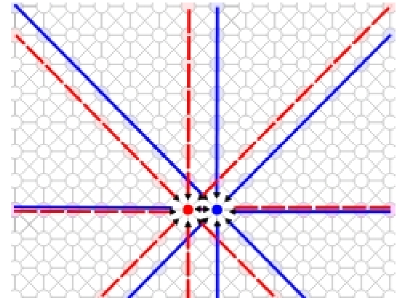

Fig. 7: Star-shaped graphs associated in SGM to two adjacent pixel Source: [11]

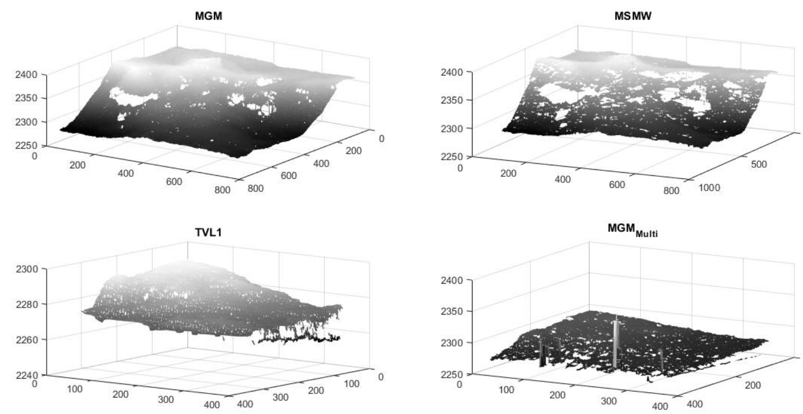

Fig. 8: Cloud comparison with Matlab

lines that are unrelated will have the greatest impact and, as a result, the regularity of the matching is poorly enforced by SGM.

More Global Matching method : The contribution can be summarized as a change in the recursive update formula: this algorithm proposes to update using information from more than one direction. The strategy is to inject information from the 2D problem in the processing of signs 1D path, this done by incorporating messages from the node visited in the previous scan line.

$$L_r(p,d) = C_p(d) + \sum_{x \in [r,r]} \frac{1}{2} \min_{d' \in D} \left(L_r(p-x,d') + V(d,d') \right) \tag{4}$$

MGM uses dependencies on scan lines, these cannot be processed in parallel in the same way as it is done with SGM. Therefore the parallelization in MGM is achieved diagonal-by-diagonal.

D. Result's comparison

In order to compare the different methods previously introduced, we display point clouds of the same tile returned by different methods with the software MatLab (Fig 8).

(Fig 8) shows that the most efficient results are for the MGM method, the density of the point cloud is higher with MGM, there are fewer missing data and sampling error artifacts, however there are high toolmaker artifacts . For the mesh part we assume that we use MGM method, the method is chosen by empirical strategy, MGM has always shown the best results for satellite image.

III. 3D reconstruction

IV. Conclusion

The conclusion goes here.

References

[1] C. d. Franchis, E. Meinhardt-Llopis, J. Michel, J. Morel, and G. Facciolo, "On stereo-rectification of pushbroom images," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct. 2014, pp. 5447–5451.

[2] ——, "Automatic sensor orientation refinement of Pliades stereo images," in *2014 IEEE Geoscience and Remote Sensing Symposium*, Jul. 2014, pp. 1639–1642.

[3] C. de Franchis, E. Meinhardt-Llopis, J. Michel, J.-M. Morel, and G. Facciolo, "An automatic and modular stereo pipeline for pushbroom images," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Zrich, Switzerland, Sep. 2014. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01109526

[4] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, vol. 47, pp. 5–12, 2001.

[5] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 4, pp. 401–406, Apr. 1998.

[6] A. Buades and G. Facciolo, "Reliable Multiscale and Multiwindow Stereo Matching," *SIAM Journal on Imaging Sciences*, vol. 8, no. 2, pp. 888–915, Jan. 2015. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/140984269

[7] M. Berger, A. Tagliasacchi, L. Seversky, P. Alliez, J. Levine, A. Sharf, and C. Silva, "State of the Art in Surface Reconstruction from Point Clouds," *EUROGRAPHICS star reports*, vol. 1, no. 1, pp. 161–185, Apr. 2014. [Online]. Available: https://hal.inria.fr/hal-01017700/document

[8] J. S. Prez, E. Meinhardt-Llopis, and G. Facciolo, "TV-L1 Optical Flow Estimation," *Image Processing On Line*, vol. 3, pp. 137–150, Jul. 2013. [Online]. Available: http://www.ipol.im/pub/art/2013/26/

[9] H. Hirschmuller, "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.

[10] "OpenCV: cv::stereo::StereoBinarySGBM Class Reference." [Online]. Available: https://docs.opencv.org/3.4.0/d1/d9f/classcv_1_1stereo_1_1StereoBinarySGBM.html#details

[11] G. Facciolo, C. d. Franchis, and E. Meinhardt, "MGM: A Significantly More Global Matching for Stereovision," in *BMVC*, 2015.