

FAST STEREO PAIR GENERATION USING SRTM REGISTRATION AND OPTICAL FLOW

Yoldaş Ataseven^{a,b*}, A. Aydın Alatan^a

^a METU, Electrical and Electronics Engineering, 06531 Balgat Ankara, Turkey - (ataseven, alatan)@metu.edu.tr

^b ASELSAN Inc., P.K. 1 Yenimahalle, Ankara, Turkey – yataseven@aselsan.com.tr

Commission WG I/4

KEY WORDS: Registration, SRTM, Stereo, interpolation, RPC, IKONOS

ABSTRACT:

In Digital Surface Model generation applications, obtaining stereo correspondences is a crucial step. The geometry of stereo image generation causes significant disparities between match points. Unlike the pinhole camera model, in pushbroom sensor images, search is less structured, since there exists no epipolar geometry. This situation brings significant computational load over the stereo correspondence generation step. In this study, a fast sparse stereo pair generation scheme is presented. This approach uses SRTM data, RPC coefficients and quadratic polynomial interpolation to generate an estimate for the geodetic coordinates of each pixel in the reference image. Harris corner detector is used to generate many feature points in the reference image. The geodetic coordinate estimates at these feature points are projected onto the second image using RPC coefficients and this projection is taken as an initial estimate for a pyramidal Kanade-Lucas-Tomasi (KLT) optical flow estimation method. Since the initial estimates for optical flow are quite reliable, the resulting match points are obtained using more precise spatio-temporal derivatives during optical flow estimation step. The pyramidal nature of the optical flow estimation method allows quite fast and promising results, even for cluttered and perspective-distorted regions, which are usually difficult to mark manually. On satellite images, the accuracy of the geodetic coordinate estimates and the success of the proposed pair generation scheme are presented through experiments. The differences for urban and rural areas are also investigated. The limits of the KLT algorithm for satellite stereo match generation and the potential of the method are discussed. Based on the experiments, it is shown that the proposed approach gives quite reliable sparse stereo correspondences for urban and rural areas in relatively small execution time.

1. INTRODUCTION

Modern Earth Observation Satellites (EOS) provide sub-meter resolution and images are in hundred mega-pixels range. The utilization of along-track stereo imaging provides massive amounts of data, which result in very expensive computations, especially for DTM/DSM generation.

Sub-meter geo-location and elevation accuracies require rigorous satellite models (which are not made available by data providers), or non-linear projection information, namely *rational polynomial coefficients* (RPCs), that are provided by the images, and ground control points (GCPs) collected from the field of interest.

Terrain reconstruction is often based on a convex optimization scheme, which optimizes the geodetic coordinates by minimizing the projection errors with respect to the stereo correspondences. For accurate DSM generation, many accurate stereo correspondences are required.

In projective stereo imaging, the stereo correspondences obey the epipolar constraint, which states that the cameras' focal points and the actual point lie on a plane. This plane intersects the image plane of the searched image on a line, which is defined as the epipolar line. It is possible to use linear operators for rectifying one image (image-2) w.r.t. the reference image (image-1) such that for every point in image-1, the corresponding epipolar line in image-2 is horizontal. This way, fast stereo pair generation can be performed.

On the other hand, unlike the projective camera images, satellite images do not have epipolar geometry, hence it is not possible to use epipolar lines in the search for stereo correspondences.

There are various studies for determining the stereo correspondences. Some of the developed methods aim at stereo matching for specific structures such as buildings, while others are more generic. Zhang and Gruen [Zhang 2006] use IKONOS triples and take the middle image as the reference. Then they assume a height (Z) value for the point and project it onto the other two images. They use a hierarchical search with different Z values and use the normalized cross-correlation. Zahng *et al.* [Zhang 2000] use correlation coefficients on image pyramids to register satellite images. Lee *et al.* [Lee 2003] use the linear pushbroom model of Gupta & Hartley [Gupta 1997] as a starting point and then use their epipolar curve as the guide for the search path and the match window orientation. Park *et al.* [Park 2006] use edge tracing and cross-correlation for rectangular building roofs.

To avoid the computational expense of the full-search, a commonly used approach is to use hierarchy. Although it is a fast and useful approach, it brings potential mismatches since the disparities in satellite images can be large. To handle large disparities, taller pyramids are required and as the number of the pyramid levels increase, it is possible to direct the search in the wrong direction due to errors in the upper levels, where

* Corresponding author.

the details are lost. Thus, it is beneficial to constrain the search region.

Optical flow is widely used in video processing for point matching. It is based on linearizing the displacement of the objects in two successive images to fasten the search. Although optical flow is not capable of handling large disparities due to the invalidity of the linearization for such displacement, if accurate initial estimates are provided, it can be used as a very fast alternative in stereo pair generation.

Generation of accurate initial estimates for stereo pairs is generally done using coarse DSMs obtained from downsampled images [Zhang 2004] or surface fitting to sparse reconstruction [Goncalvez 2007].

One alternative is to use the maps or static DTMs of various accuracies. Shuttle Radar Topography Mission (SRTM) data are among them [Farr 2007].

This study presents a pair generation scheme which utilizes Kanade-Lucas-Tomasi (KLT) tracker on coarsely geo-registered stereo satellite image pairs. For geo-registration, SRTM data and RPCs are used.

2. METHODOLOGY

2.1. Object-to-Image SRTM Registration

The proposed algorithm uses the data vendor provided RPCs to perform the registration *from the object domain to the image domain*. RPCs define the rational polynomial functions that map a real world coordinate to the image domain. Therefore, once we have the RPC coefficients, we have the projection functions. Thus, for each point in the SRTM data, we can find its projection pixel in the image.

It is known that SRTM accuracy is unchanged when 90m resolution SRTM data are upsampled by 3 to obtain 30m resolution, via bi-cubic interpolation [Keeratikasikorn 2008]. Therefore, in this study, the SRTM data are upsampled by 3 with bi-cubic interpolation and registration is performed for the upsampled version.

Once the SRTM points are projected onto the image, an interpolation is required to fill the empty pixels. The SRTM data is regularly-sampled (30m or 90m Ground Sample Distance; GSD). On the other hand, satellite images have higher resolution and do not sample the Earth surface on a regular latitude-longitude grid. 1 degree x 1 degree SRTM patches and satellite images are never aligned. In other words, the SRTM grid does not project to another regular grid in the image domain and some SRTM points will fall outside the image, especially for narrow FOV satellite images, such as IKONOS. Thus, an interpolation scheme is required that provides acceptable accuracy, and leaves no empty pixels.

For that purpose, quadratic polynomial surface fitting is applied as follows: For each SRTM point p_s , eight neighbors of that SRTM point are projected to the image domain to define a surface patch sampled at 9 points, together with the center point p_s (Figure 1). Then the projection points are taken as examples from quadratic polynomials (on image coordinates) that define a height function, a latitude function and a longitude function;

$$F_i = a_i u^2 + b_i v^2 + c_i u.v + d_i u + e_i v + f_i \quad (1)$$

where

$$\begin{aligned} u, v &= \text{image row and column indexes} \\ i &= 1, 2, 3 \text{ for height, latitude and longitude.} \\ a_i, b_i, c_i, d_i, e_i, f_i &= \text{coefficients of the polynomial} \end{aligned}$$

In other words, interpolation functions are defined separately for latitude, longitude and height.

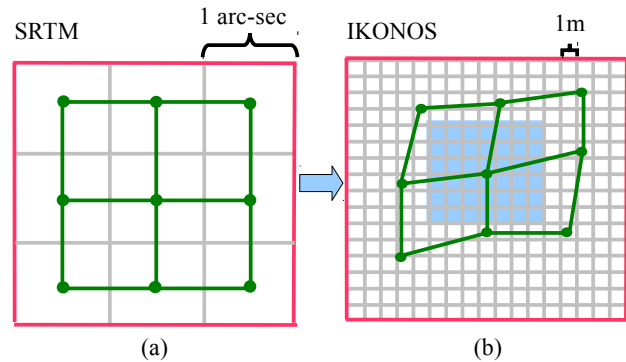


Figure 1. (a) The nine-point SRTM grid and (b) their projections on the satellite image.

The resulting registration is observed to have bias, as presented in the results section. To reduce the bias, a few accurate stereo pairs are used. The bias is reduced by performing full search for additive correction terms in the object domain (latitude, longitude and height) for every pair point in the first image and minimizing the projection error in the second image. The details are explained in [Ataseven 2010]. The method used in this study is a faster alternative, where the search along each geodetic coordinate is isolated. A search for a correction term for height bias is also performed.

2.2. Optical Flow for Stereo Pair Generation

Optical flow is a well known topic for video processing and machine vision applications. The underlying theory of optical flow is based on the assumption that the motion field is unchanged for a small neighborhood. In this neighborhood, performing a Taylor expansion for the second image using the first image and omitting the higher order terms of the expansion result in a differential equation [Fleet 2006]:

$$I_x V_x + I_y V_y + I_t = 0 \quad (2)$$

where, I_x and I_y correspond to directional derivatives in image 1, V_x and V_y are the velocity components and I_t is the time derivative which corresponds to the local difference of the images.

KLT tracker uses equation 2 with the assumption that the motion field is constant for a restricted neighborhood [Lucas 1981]. The result is an overdetermined system of equations obtained from the points in the neighborhood, which can be solved either by the least squares solution or iterative methods. KLT uses a Newton-Raphson type iteration.

Obviously, since the time difference between the stereo satellite image pair is not infinitesimal, the time derivative which is approximated by the difference of the images is invalid. Thus, it is needed to provide initial estimates for the search. These initial estimates are obtained from the SRTM registration output. Still, the derivative term remains inaccurate for significant height discontinuities (occlusions),

which is the case for the edges of the tall building roofs. This is a typical problem in stereo match generation with large base distance and is generally addressed by dedicated methods for urban areas [Fischer 1998, Park 2006].

In this study, Intel's OpenCV implementation of pyramidal KLT method [Bouquet 2000], which is a fast implementation optimized for Intel Processors, is used.

The feature points are selected in the reference image, with the Harris corner detector, again using OpenCV implementation. Harris corner detector is based on the local derivatives and seeks for points where both x and y gradients are significant. It is preferred in this application because of its suitability to KLT tracker and the flexibility in the distribution and quality of feature points.

The used KLT implementation outputs the match error in MSE for the neighborhood of the feature point. The size of the neighborhood is user-defined (11x11 in the experiments). Thus, one can select the acceptable MSE for a feature to be named successful.

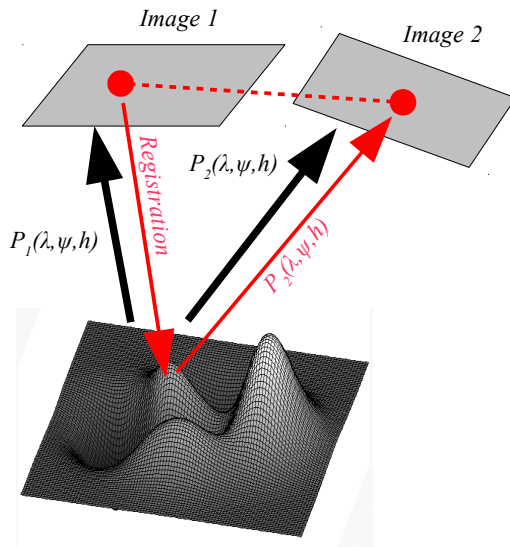


Figure 2. Generation of initial estimates for the KLT algorithm.

When the Harris corners are determined in image-1, the initial estimate for geodetic coordinates are immediately available from the SRTM registration output. This geodetic coordinates are then projected onto image-2 with its RPC-based projection function, P_2 . The projection on image-2 is given to the KLT algorithm to perform matching in a small region.

The accuracy of SRTM-IKONOS registration determines the required range for the KLT algorithm. We have two parameters for the KLT algorithm for determination of the search range: i) the size of neighborhood around the feature point, ii) the number of levels in the image pyramid. Increasing the number of levels in the image pyramid brings the possibility of false matching results due to the reduced detail (local information) in the lower resolution pyramid levels. On the other hand, increasing the size of the neighborhood brings the possibility of an increase in the number of rejected pairs. However, it is also observed that, larger neighborhoods increase the performance for the regions where the texture is alike for small patches, but

distinguishable for large patches. Forests are examples to such regions.

3. EXPERIMENTS AND RESULTS

3.1. SRTM-Single IKONOS Image Registration

For single-image case, the registration accuracy is evaluated at 121 Ground Control Points (GCPs). The results are presented in Table 1.

	Mean error	σ_{error}	RMS
Latitude	$2.43 \times 10^{-5} \text{ }^\circ$	$1.53 \times 10^{-5} \text{ }^\circ$	$2.873 \times 10^{-5} \text{ }^\circ$
Longitude	$-1.76 \times 10^{-5} \text{ }^\circ$	$1.46 \times 10^{-5} \text{ }^\circ$	$2.288 \times 10^{-5} \text{ }^\circ$
Height	-5.30 m	4.27m	6.8m
Planimetric error	3.28 m	1.71 m	3.69 m
Along UTM X	-1.43m	1.15 m	1.83 m
Along UTM Y	2.75m	1.65 m	3.21 m

Table 1. Error figures for SRTM-image registration, after bias compensation using stereo pairs, at 121 GCPs. IKONOS, Hobart region, Australia

As it can be observed in Table 1, the registration results for the IKONOS stereo pair correspond to a few pixels of error. Note that all of the GCPs are taken from the ground level even for the urban regions and thus, errors for tall buildings are not seen in this table. Therefore, to handle such urban structures, the search region must be kept larger than the table implies.

3.2. Stereo Pair Generation

For the feature points, Harris features that are separated at least by 30 pixels are used to allow uniformly distributed feature points.

A pyramidal KLT with three-level hierarchy is used with a block size of 11x11. The corresponding search range is 35 pixels. For a base-height ratio of 1, this corresponds to handling structures of approximately 50m high, when the reference image is nadir.

For the experiments, IKONOS images of Hobart, Australia are used. The site of the image has a terrain range of 0-1250m. The images contain the sea, urban regions, rural areas, forests and cloud cover.

Two experiments are conducted:

- i) Harris features are determined on the entire reference image. KLT match score is kept strict to avoid inaccurate matches.
- ii) Reference image is divided into 186 sub-images and the Harris features are determined on each sub-image. KLT match error threshold is relaxed for allowing more pairs.

The resultant feature point distributions for the successful matches are given in Figure 3. Only the reference image is displayed.

In the experiments (i) and (ii), totally 22330 and 93977 pairs are generated, respectively.

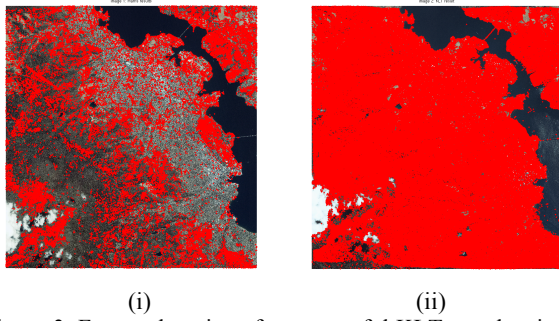


Figure 3. Feature locations for successful KLT matches in the two experiments.



Figure 4. Pairs with the best MSE scores for experiment (i)

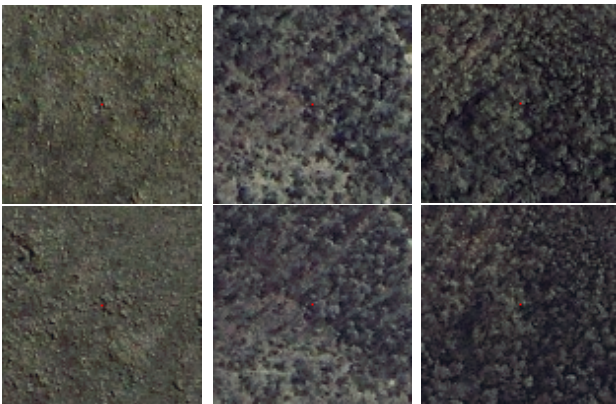


Figure 5. Pairs with the worst MSE scores for experiment (i)

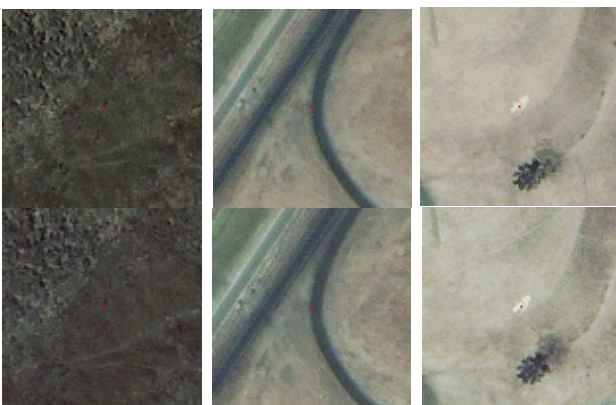


Figure 6. Pairs with the best MSE scores for experiment (ii)



Figure 7. Pairs with the worst MSE scores for experiment (ii)

The quality of the generated pairs are difficult to evaluate automatically. Due to the non-perspective distortions in the pushbroom stereo imaging geometry, the measures that are focused on pixel-wise agreement in a neighborhood may generate misleading results. A better measure of matching accuracy should determine displacement between the ground points to which the match points are assigned. Visual inspection can be used for that purpose. However, visual inspection is a time consuming process, since one needs to consider the perspective-like effects while deciding on the success of the match. Therefore, for a high number of feature points, the visual inspection is unfeasible. Thus the pairs with the worst and best MSE scores are observed.

The results of visual inspection are as follows:

- All of the 100 best-score matches are correct for both experiments.
- 10% of the 100 worst-score matches are wrong for experiment (i), excluding the 2 pairs in cloud cover
- 17% of the 200 worst-score matches are wrong for experiment (ii), excluding the 10 pairs in cloud cover.

It is observed in the results that, although the number of wrong matches increase with increasing match error, a significant amount of additional correct matches can be obtained with relaxed match score threshold.

The algorithm performs well for the regions where occlusions are small or non-existent. For the points at building edges, the algorithm sets the pairs at the ground points near the buildings.

3.3. Performance in Urban Regions



Figure 8. Pairs in an urban area.

In Figure 8, it is possible to observe the uniform distribution of the feature points, except for the tall building, where there are both sunlight reflection changes, and strong parallax.

3.4. Computational Time

The experiments are performed on a PC that has an Intel Pentium Q9450 quad-core processor and 8GB RAM. Only a single core of the processor is used. The execution times for different phases of the algorithm are as follows:

The SRTM registration is completed in 59 minutes (MATLAB). Note that the registration is performed for each pixel in the image, not only for the feature points. Bias removal in SRTM registration is completed (MATLAB) within 5 seconds. The generation of 96113 feature points with the Harris feature detector (C++) is completed within 30 seconds. The matching of 93977 pairs with KLT (C++) is completed within 20 seconds. The matching of 22330 pairs with KLT (C++) is completed within 5 seconds.

4. CONCLUSION

The proposed algorithm is a fast and accurate way of generating many stereo pairs in a relatively short time. The density of the feature points is user-controlled. Since the satellite images are usually rich in terms of Harris features, it is possible to generate high number of stereo pairs.

There exist a significant amount of erroneous pairs, if the Harris criteria or KLT match error thresholds are relaxed. These pairs can be eliminated by performing the KLT algorithm in the inverse direction and forcing backward consistency.

When the quality of Harris features and the KLT error threshold are kept strict, the resultant pairs are very reliable. Such pairs might be used for accurate tie points.

The algorithm is in general unable to place pairs at the edges of the buildings. Instead, it places the pairs on the ground points that are very near to the roof corners. This result is caused mainly due to the strong contrast between the building roofs and the ground, and nature of Harris feature detector. Thus, the algorithm is required to be used together with some pair generators that are specialized on buildings.

The execution time of the algorithm allows the generation of high number of pairs. It should also be noted that the proposed algorithm is fully automatic and does not contain any heuristics.

As the future work, the consistency check will be applied and the performance of the registration algorithm on dense reconstruction will be experimented. There are other feature selectors other than the Harris feature detector; and optical flow methods other than KLT. Once the registration output is at hand, other matching methods may also perform well.

REFERENCES

Ataseven, Y., Alatan, A. A., 2010, "SRTM Registration for Electro-Optic Satellite Images without GCP", *Proceedings of ISPRS PCV 2010*, Paris, France, September 1-3 (in press).

Bouguet, J., 2000, "Pyramidal Implementation of the Lucas-Kanade Feature Tracker: Description of the Algorithm", *Technical Report, OpenCV Document*, Intel Microprocessor Research Labs

Farr, T. G., et al., 2007, The Shuttle Radar Topography Mission, *Rev. Geophys.*, 45, RG2004, doi:10.1029/2005RG000183

Fischer, A., Kolbe, T.H., Lang, F., Cremens, A.B., Förstner, W., Plumer, L., Steinhage, V., 1998, "Extracting Buildings from Aerial Images using Hierarchical Aggregation in 2D and 3D", *Computer Vision and Image Understanding*, Vol. 72, Issue 2, November 1998, pp. 185-203

Fleet, D.J., Weiss, Y., 2006, "Optical Flow Estimation", in Paragios *et al.* *Handbook of Mathematical Models in Computer Vision*. Springer. ISBN 0387263713.

Gonçalves, J.A., Marçal, A.R.S., 2007, "Automatic Orthorectification of ASTER Images by Matching Digital Elevation Models", *ICIAR 2007*, LNCS 4633, pp. 1265–1275

Park, C.H., Sohn, H.G. Song, Y.S., 2006, "Stereo Matching Strategy for 3-D Urban Modeling", *ICCSA 2006*, LNCS 3981, pp. 1043 – 1050,

Rodriguez, E., Morris, C.S., Belz, J. E., Chapin, E.C., Martin, J.M., Daffer, W., Hensley, S., 2005, "An assessment of the SRTM topographic products", *Technical Report JPL D-31639*, Jet Propulsion Laboratory, Pasadena, CA, USA, 143 pp.

Lee, H.Y., Kim, T., Park, W., Lee, H.K., 2003, "Extraction of digital elevation models from satellite stereo images through stereo matching based on epipolarity and scene geometry", *Image and Vision Computing* 21, pp. 789–796

Zhang, Z., Zhang J., Uao, M., Zhang, U., 2000, "Automatic Registration of Multi-Source Imagery Based on Global Image Matching", *Photogrammetric Engineering & Remote Sensing* Vol. 66, No. 5, May 2000, pp. 625-629.

Park, C.H., Sohn, H.G., Song, Y.S., 2006, "Stereo Matching Strategy for 3-D Urban Modeling", M. Gavrilova et al. (Eds.): *ICCSA 2006*, LNCS 3981, pp. 1043 – 1050

Gupta, R., Hartley, R., 1997, "Linear Pushbroom Cameras", *IEEE Transactions on PAMI* 19 (9), pp. 963–975.

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

Lucas, B.D., and Kanade, T., 1981, "An Iterative Image Registration Technique with an Application to Stereo Vision. *Proceedings of Imaging Understanding Workshop*, pp. 121-130

Zhang, L., Gruen, A., 2004, "Automatic DSM Generation from Linear Array Imagery Data", *20th ISPRS Congress*, 12-23 July, Istanbul, Turkey. *The International Archives of Photogrammetry, Sensing and Spatial Information Sciences*, Volume XXXV, Part B3, Commission III, pp. 128-133

ACKNOWLEDGEMENT

The IKONOS images and GCPs are taken from the ISPRS data set collection (<http://www.isprs.org/data/default.aspx>).

This study is supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK), with National Scholarship Programme for PhD Students.