Stereo Vision Based Reconstruction of Huge Urban Areas from an Airborne Pushbroom Camera (HRSC)

Heiko Hirschmüller[1], Frank Scholten[2], and Gerd Hirzinger[1]

[1] Institute of Robotics and Mechatronics, German Aerospace Center (DLR) Oberpfaffenhofen
[2] Institute of Planetary Research, German Aerospace Center (DLR), Berlin, Germany
heiko.hirschmueller@dlr.de

Abstract. *This paper considers the application of capturing urban terrain by an airborne pushbroom camera (e.g. High Resolution Stereo Camera). The resulting images as well as disparity ranges are expected to be huge. A slightly non-linear flight path and small orientation changes are anticipated, which results in curved epipolar lines. These images cannot be geometrically corrected for matching purposes such that epipolar lines are exactly straight and parallel to each other. The proposed novel processing solution explicitely calculates epipolar lines for reducing the disparity search range to a minimum. This is a necessary prerequisite for using an accurate, but memory intensive semi global stereo matching method that is based on pixelwise matching. It is shown that the proposed approach performs accurate matching of urban terrain and is efficient on huge images.*

1 Introduction

The High Resolution Stereo Camera (HRSC) has been developed by the Institute of Planetary Research (DLR)[1] for the exploration of the Marsian surface from orbit. The airborne version HRSC-AX is currently used for capturing earths landscape and cities from flight altitudes between 1500m to 5000m. The camera contains nine sensor arrays, which are arranged orthogonally to the flight direction in different angles. All arrays have a resolution of 12000 pixels. Five arrays are panchromatic. The other four capture red, green, blue and infrared light. The position and orientation of the camera is continuously measured by a sophisticated GPS/IMU system. Current post-processing [1–3] includes radiometric corrections as well as refinements of all camera positions, orientations and time offsets by means of photogrammetric methods based on HRSCs multi-stereo image information. A geometric correction step projects the pixels of each array at all camera positions onto an artificial plane, resulting in nine 2D images, in which effects caused by high and low frequent orientation variations are eliminated, while disparities caused by terrain and buildings still remain (Fig. 1).

The current stereo method [1] performs hierarchical, correlation based stereo matching of these 2D images. Epipolar lines are assumed to be parallel to the overall flight direction. Violations of this assumption are handled by searching correspondences in a 2D area around suspected epipolar lines. The matching result is used for calculating Digital Elevation Models (DEM) and subsequently for the generation of true ortho images based on the DEM. This paper proposes the explicit calculation of epipolar lines in general, non-linear pushbroom images combined with a semi-global stereo matching method. The aim is to increase the accuracy in scenes of urban areas.

DAGM 2005, Vienna, Austria, 31 Aug. - 2 Sept. 2005.

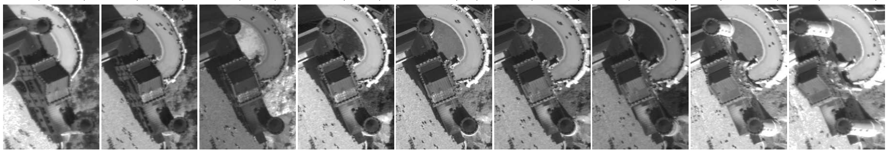

Fig. 1. Parts of corrected 2D pushbroom images with a resolution of 15cm/pixel.

2 Related Literature

There are different possibilities for modeling the movement of pushbroom cameras. The *Linear Pushbroom Camera* model [4–6] assumes a movement with constant velocity and orientation in a straight line. This movement is not realistic for an airborne camera, which is exposed to wind. General movements may be approximated by polynomial functions [7] or explicitely represented by discrete positions and orientations [1, 8, 9].

Geometric corrections of the images as a pre-requisite for matching and subsequent 3D surface modeling are commonly done by projecting all pixels onto an artifical plane [1, 8, 9]. Straight lines and epipolar lines are *almost* straight in the corrected images, if the flight path is nearly linear. Thus, these corrected images can be treated as *almost* rectified [9]. However, a rectification that results in exactly straight epipolar lines is not possible, since epipolar lines are generally hyperbolas, even if the camera movement is exactly linear [4]. The complex shape of epipolar lines must be considered during stereo matching, either by searching in 2D around linearly approximated epipolar lines [1, 8] or by explicitely calculating them [6] for reducing the search space to 1D. 2D search areas can also be reduced by ortho rectification implying rough terrain instead of simple geometric correction [2].

Stereo matching must be efficient, due to typically huge images and disparity ranges. Local, correlation based approaches are often applied, either by hierarchical matching [1] or by region growing [8, 6], which starts at high confidence correspondences. However, correlation based methods are known to blur sharp object boundaries [10], which makes them less suitable for urban areas. Global cost minimization methods have been shown [11] to perform much better at object boundaries. However, global methods are typically slow and memory intensive, which makes them unsuitable for huge images.

3 Reconstruction from Pushbroom Images

The following sections introduce the camera model (Section 3.1) as base for calculating epipolar lines (Section 3.2). Epipolar lines are required for limiting the correspondence search during stereo matching (Section 3.3). The matching result is finally used for calculating an orthographic projection of height and image information (Section 3.4).

3.1 Modeling Non-Linear Pushbroom Cameras

Pushbroom cameras use an array of sensor elements that is arranged in a straight line on the image plane (Fig. 2a). Lens distortion is modeled by treating the sensor array

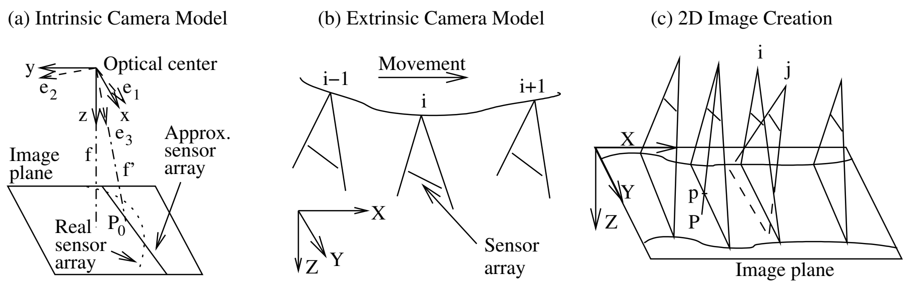

Fig. 2. Camera model and the creation of a corrected 2D image.

as being curved. The intrinsic parameters are the focal length f (i.e. distance between optical center and image plane) and the positions x_k, y_k on the image plane for all pixels k. Thus, the 3D location of a pixel k is $S_k = \left(x_k\ y_k\ f\right)^T$ in the camera coordinate system.

A 2D image is captured line by line, while the camera moves. A linear movement cannot always be guaranteed (e.g. while flying with an airplane). Therefore, it is assumed that the path is only roughly a straight line, the speed is not constant and the orientation is changing slightly (Fig. 2b). The lack of constraints requires to measure the orientations R_i and locations T_i at all capturing positions i with high accuracy. Since original GPS/IMU measurements describe the orientation of the IMU-axes, a photogrammetric reconstruction of the orientation of the camera axes is performed by means of photogrammetric methods [3]. This leads to the relationship $P = sR_iS_k + T_i$ between a world point P and the kth pixel of the ith capturing position. The reconstruction of the ray of light that ends in pixel k is, based on the previously described photogrammetric reconstruction, straight forward. However, the inverse (i.e. finding pixel k as projection of P) is difficult, because the list of pixel positions S is unsorted. The calculation is simplified by approximating the sensor array by the best fitting line, i.e.

$$P = sR_iS_k + T_i, \qquad S_k = kU + V + \varepsilon_k. \tag{1}$$

This allows the definition of a new camera coordinate system e_1, e_2, e_3 (Fig. 2a) in which e_1 is parallel to the sensor line and $f'e_3$ intersects the line in the point P_0.

$$e_1 = \frac{U}{|U|} \qquad f'e_3 = P_0 = -\frac{UV}{UU}U + V \qquad e_2 = e_3 \times e_1 \tag{2}$$

The resulting closed form definition of the relationship becomes,

$$P = sR_iR_c\left(k'\ 0\ 1\right)^T + T_i, \qquad R_c = \left(e_1\ e_2\ e_3\right). \tag{3}$$

The relationship between the pixel k' and k can be derived easily as $k' = \frac{k|U| + e_1V}{f'}$. Corrected 2D images are obtained by projecting all pixel values onto a common image plane at $z = 0$ (Fig. 2c) using (1). The projection is generally irregular. The values at regular grid positions are calculated as linear interpolations of nearby pixels. Orientation changes that destroy the order of projected pixels (e.g. camera position j in Fig. 2c) are treated by removing disturbing camera positions and their projected pixels.

For 3D computer vision purposes it is important to have a projection model of the corrected 2D pushbroom images. The intrinsic parameters are the path of optical center positions T_i and corresponding viewing directions R_iR_c. The movement between discrete positions is assumed to be linear. Equation (3) can then rewritten as,

$$P = sR_iR_c \left(k' \; 0 \; 1\right)^T + T_i + s_i \frac{T_{i+1} - T_{i-1}}{2}. \tag{4}$$

The factor s_i controls the linear movement of the ith optical center for projecting P exactly on the sensor array (i.e. $y = 0$). Solving for s_i in dependence of P results in,

$$s_i(P) = 2\frac{r_2(P - T_i)}{r_2(T_{i+1} - T_{i-1})}, \text{ with } R_iR_c = \left(r_1 \; r_2 \; r_3\right). \tag{5}$$

The position $C(P)$ of the optical center is calculated from the closest base i, i.e.

$$C(P) = T_i + s_i(P)\frac{T_{i+1} - T_{i-1}}{2}, \text{ for } i \text{ such that } s_i(P) \text{ is a minium} \tag{6}$$

The closest base i can be found by a binary search in $O(\log_2 n)$ steps. Removing overlapping projections (e.g. j in Fig. 2c) ensures a sorted list. The sign of s_i determines the search direction. The determination of the optical center permits the projection of a world point P onto an image point p by calculating the intersection of P, $C(P)$ at $z = 0$.

$$p = f_{proj}(P) = \frac{C_z}{C_z - P_z}\begin{pmatrix} P_x - C_x \\ P_y - C_y \end{pmatrix} + \begin{pmatrix} C_x \\ C_y \end{pmatrix}, \text{ with } C = C(P). \tag{7}$$

Similarly, the world point P is reconstructed from a given pixel p at the distance z.

$$P = f_{rec}(p,z) = \frac{C_z - z}{C_z}\begin{pmatrix} p_x - C_x \\ p_y - C_y \\ -C_z \end{pmatrix} + \begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix}, \text{ with } C = C(\left(p_x \; p_y \; 0\right)^T) \tag{8}$$

3.2 Calculation of Epipolar Lines

A pixel p_1 and the corresponding optical center $C_1(p_1)$ define a line, which contains the world point P that is projected on p_1. The projection of this line into a second image is called epipolar line. The projection of P in the second image must be on the epipolar line (Fig. 3a). This is formally defined as,

$$p_2 = e_{12}(p_1, d) = f_{proj,2}(f_{rec,1}(p_1, -d\Delta z)). \tag{9}$$

(a) Calculation of Corresponding Points

(b) Epipolar Lines (1) approx. by same pixel column, (2) exact, (3) approx. as straight line

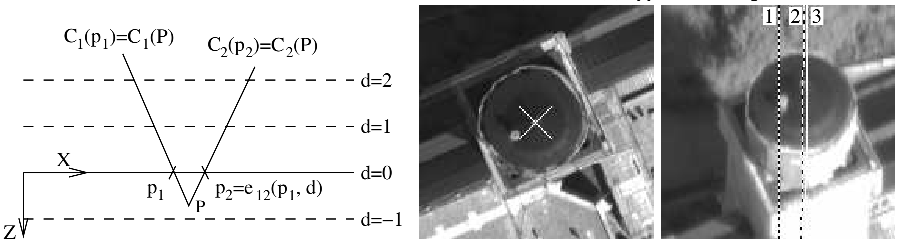

Fig. 3. Calculation of epipolar lines.

The disparity d controls the position on the epipolar line. The constant Δz is set such that a disparity step of 1 causes a mean translation of 1 pixel on the epipolar line. Figure 3b shows an example of an exactly calculated epipolar line and approximations by the same pixel column and as straight line. The overall flight path is vertical. The example shows that the approximated epipolar lines miss the correct correspondence by several pixel, which enforce a 2D search for finding correspondences. The exact point by point calculation of epipolar lines reduces the search range to a minimum. The efficiency of this approach is optimized by calculating only a few points on the line and assuming piecewise linearity in between.

3.3 Stereo Matching

Stereo matching is done with the Semi-Global Matching (SGM) method [12], which aims to determine the disparity image D, such that the cost $E(D)$ is a minimum.

$$E(D) = \sum_p C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \qquad (10)$$

The cost function evaluates pixelwise matching costs $C(p, D_p)$ at the pixel p with the disparity D_p. Piecewise smoothness of the disparity image is supported by adding a small cost P_1 for all small disparity changes and a higher cost P_2 for all higher disparity changes. Adding a *constant* cost for all higher disparity changes preserves discontinuities. Finding the minimum of equation (10) is an NP-complete problem. The SGM algorithm approximates the global minimization by pathwise minimizations from all directions. The complexity is only $O(ND)$, but the memory consumption is also proportional to ND (i.e. number of pixels times the disparity range). The pixelwise matching cost $C(p, D_p)$ is based on hierarchically computing Mutual Information (MI) [12] instead of intensity differences. This makes it robust against recording differences and illumination changes, which can easily happen since pushbroom cameras capture corresponding points at different times on the flight path. Finally, the SGM method provides a multi-baseline extension [12] for reducing mismatches and increasing accuracy.

Applying the SGM method to HRSC images includes three adaptations. Firstly, multi-baseline matching uses the five panchromatic images of the HRSC (Sect. 1), weighted by their recording angle. Optionally, the red and green images are also matched against the panchromatic nadir image, which is possible with MI matching. Secondly, the huge images are split into manageable pieces (i.e. tiles) for matching. The tiles are defined slightly overlapping and pixels near image borders are rejected, because they receive support only from one side by the global cost function. Thirdly, the disparity range is determined automatically, by first processing downscaled images (e.g. factor 16) with a very large disparity range. A reduced range is determined from the result and used for higher resolutions. The disparity range determination is done during the hierarchical computation of MI.

3.4 Orthographic Projection

The resulting disparity image corresponds to the nadir image of the HRSC. The model of this image (Sect. 3.1) is complex due to low constraints on the flight path and a mixture of perspective and parallel projection models. The disparity image permits the conversion of the data into a simple orthographic model. Each pixel p of the disparity image D is reconstructed by $P = f_{rec,D}(p, -D_p \Delta z)$. An orthographic projection is used, which stores each height value P_z at the pixel position P_x, P_y. Double mappings are resolved by using the height that is closest to the camera. This orthographic model also supports fusing results of different recordings, by taking the mean or median of heights, which decreases outliers. Finally, gaps are filled by interpolation. The result is a Digital Elevation Model (DEM) of the scene. The corresponding intensity or color value at P_x, P_y and the stored height value P_z is determined by bilinear interpolation in the image I at the position $q = f_{proj,I}(P)$. The results is a true ortho-image.

4 Experimental Results

The first row of Fig. 4 shows small parts of three scenes. The DEMs of Neuschwanstein castle are the combination of 4 flights in cross directions over the castle. Similarly, the DEMs of Garmisch-Partenkirchen are the combination of 3 overlapping parallel flights. The DEMs of Rosenheim are the result of one flight only.

The second row shows DEMs that are produced by a hierarchical, correlation based method (HC) [1]. The method avoids the calculation of epipolar lines by searching correspondences in a 2D area. Matching is done in images that are sampled with 25cm/pixel. However, only every second pixel is calculated, which results in 50cm/pixel. It can be seen that the boundaries of houses are severely blurred and towers are unrecognized.

The third and fourth row of Fig. 4 present DEMs that have been produced by SGM with calculated epipolar lines with images at different resolutions. The boundaries of houses are much sharper and all towers are properly detected. The average processing time is one hour on a 2.8GHz Xeon computer for producing 11MPixel of the DEM by matching 5 images with an average disparity range of 400 pixel. The fifth row of Fig. 4 shows reconstructions, using the ortho image as well as the S1 and S2 images (e.g. Fig. 1) for top and side textures. The last row presents reconstructions of a $110 km^2$ area of Berlin in a resolution of 20cm/pixel. The area has been captured by 6 parallel, partly overlapping flights. Each flight contributed approximately 1 billion height values to the DEM. The total processing time was 18 days on a 2.8GHz Xeon computer.

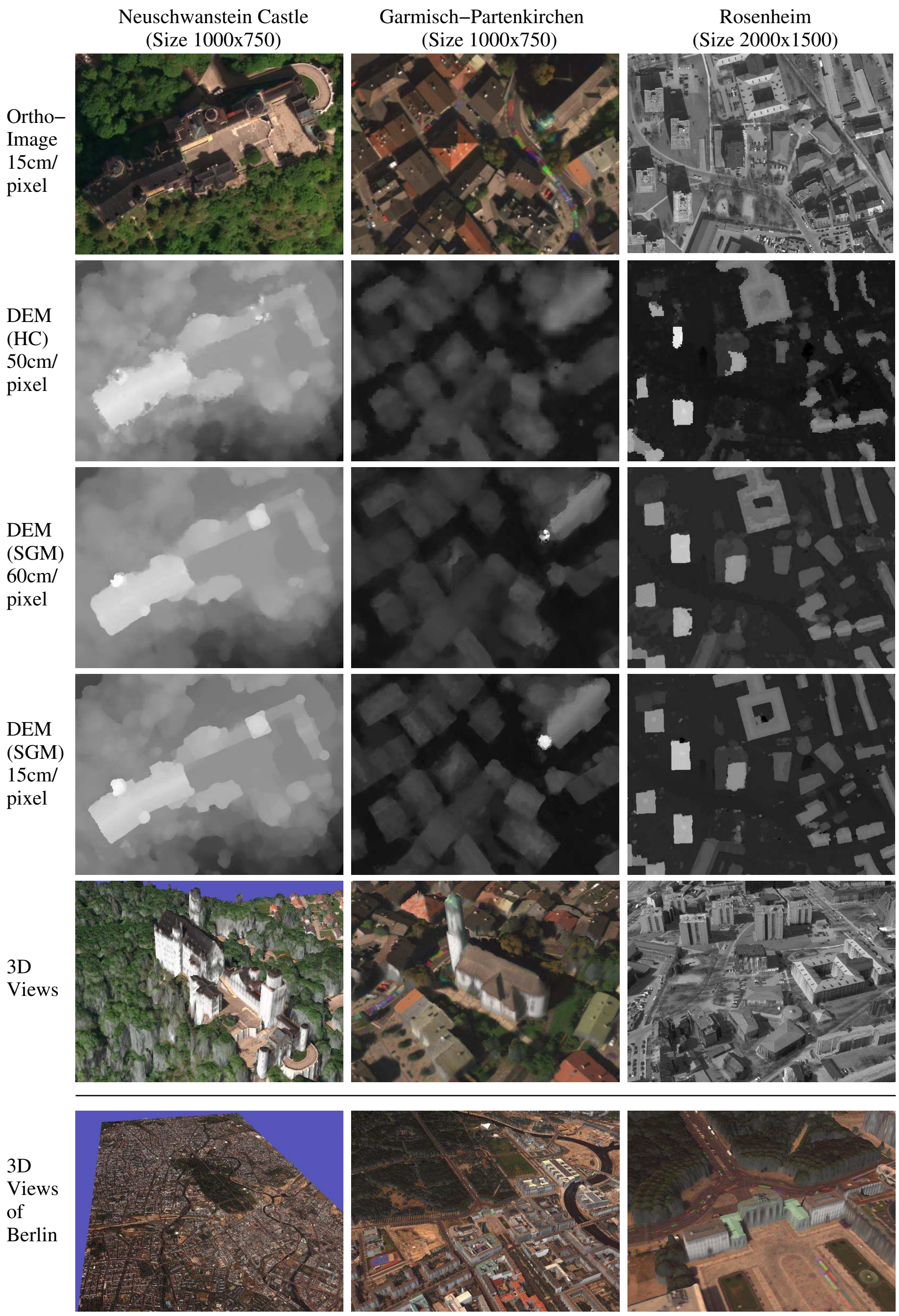

Fig. 4. Results of a hierarchical, correlation base method (HC) that finds correspondences within a 2D area and the SGM method that finds correspondences along calculated epipolar lines.

5 Conclusion

It has been shown that epipolar lines can be efficiently calculated in general non-linear pushbroom images. This permits the minimization of the correspondence search range for using the accurate, but memory intensive Semi-Global Matching method. Experiments confirmed that the proposed stereo processing solution performs accurate matching of urban terrain with sharp boundaries of buildings. The approach permits the efficient and fully automatic 3D reconstruction of whole cities. Future plans include the implementation of the approach on a processing cluster.

Acknowledgments

We would like to thank Klaus Gwinner, Johann Heindl, Frank Lehmann, Martin Oczipka, Sebastian Pless and Frank Trauthan for inspiring discussions.

References

1. Wewel, F., Scholten, F., Gwinner, K.: High resolution stereo camera (hrsc) - multispectral 3d-data acquisition and photogrammetric data processing. Canadian Journal of Remote Sensing **26** (2000) 466–474
2. Scholten, F., Gwinner, K., Wewel, F.: Angewandte digitale photogrammetrie mit der hrsc-a. In: Photogrammetrie, Fernerkundung, Geoinformation. Volume 5., Stuttgart, Germany, E. Schweizerbartsche Verlagsbuchhandlung (2002) 317–332
3. Scholten, F., Wewel, F., Sujew, S.: High resolution stereo camera airborne (hrsc-a): 4 years of experience in direct sensor orientation of a multi-line pushbroom scanner. In: ISPRS Proceedings Sensors and Mapping from Space, Hannover, Germany (2001)
4. Gupta, R., Hartley, R.I.: Linear pushbroom cameras. IEEE PAMI **19** (1997) 963–975
5. Morgan, M., Kim, K., Jeong, S., Habib, A.: Epipolar geometry of linear array scanners moving with constant velocity and constant attitude. In: ISPRS, Istanbul, Turkey (2004) 1024
6. Lee, H.Y., Park, W.K., Kim, T., Lee, H.K.: Accurate demlk extraction from sport stereo pairs: A stereo matching algorithm based on the geometry of the satellite. In: Asian Conference on Remote Sensing. (1999)
7. Lee, C., Theiss, H., Bethel, J., Mikhail, E.: Rigerous mathematical modeling of airborne pushbroom imaging systems. Journal of Photogrammetric Engineering and Remote Sensing **66** (2000) 385–392
8. Boerner, A.: Entwicklung und Test von Onboard-Algorithmen für die Landfernerkundung. PhD thesis, Technical University of Berlin, Germany (1999)
9. Haala, N., Stallmann, D., Cramer, M.: Calibration of directly measured position and attitude by aerotriangulation of three-line airborne imagery. In: ISPRS, Object Recognition and Scene Classification from Multispectral and Multisensor Pixels, Columbus/Ohio, USA (1998) 23–30
10. Hirschmüller, H., Innocent, P.R., Garibaldi, J.M.: Real-time correlation-based stereo vision with reduced border errors. IJCV **47** (2002) 229–246
11. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV **47** (2002) 7–42
12. Hirschmüller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In: IEEE CVPR, San Diego, USA (2005)