

Semi-Global Matching – Motivation, Developments and Applications

HEIKO HIRSCHMÜLLER, Oberpfaffenhofen

ABSTRACT

Since its original publication, the Semi-Global Matching (SGM) technique has been re-implemented by many researchers and companies. The method offers a very good trade off between runtime and accuracy, especially at object borders and fine structures. It is also robust against radiometric differences and not sensitive to the choice of parameters. Therefore, it is well suited for solving practical problems. The applications reach from remote sensing, like deriving digital surface models from aerial and satellite images, to robotics and driver assistance systems. This paper motivates and explains the method, shows current developments as well as examples from various applications.

INTRODUCTION

Stereo matching is used for finding corresponding pixels in a pair of images, which allows 3D reconstruction by triangulation, using the known intrinsic and extrinsic orientation of the cameras. Unfortunately, the problem is ill-posed, since images are locally very ambiguous. Many techniques have been proposed in photogrammetry and, later on, in computer vision. However, all of them are far away from being optimal. This paper motivates and explains the Semi-Global Matching (SGM) technique, which offers a good trade off between accuracy and runtime and is therefore well suited for many practical applications, as shown in the end of the paper.

MOTIVATION

Correlation based stereo matching has long been the preferred method for dense reconstruction, especially if runtime is an issue, due to real-time constraints or huge amounts of data. Correlation methods compute the similarity between pixels by *comparing* windows around pixels of interest. This is done for all possible correspondences for finding the pixel with the highest similarity. Normalized Cross Correlation (NCC) is often used in photogrammetry for comparing windows. NCC handles radiometric differences like bias and gain changes and assumes Gaussian noise in the values of corresponding pixels.

Correlation methods implicitly assume that all pixels within the window have the same distance

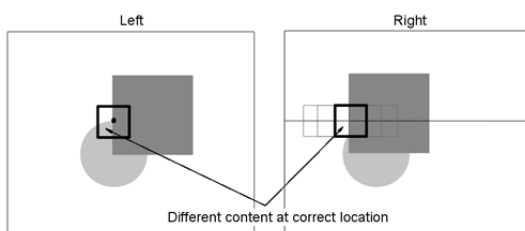


Figure 1: Correlation at object border.

from the camera, i.e. lie on a fronto-parallel surface. Slanted surfaces and abrupt changes, as caused by depth discontinuities, result in wrongly including non-corresponding image parts in the calculation as shown in Figure 1.

Slanted surfaces can be handled by permitting affine transformations of corresponding windows, i.e. Least-Squares Matching (Grün, 1985). However, this is computationally expensive and does not work at fine

structures or depth discontinuities. Matching partly wrong content near depth discontinuities (Figure 1), leads to severe correlation errors, especially for NCC, since matching wrong pixels is not well modeled as Gaussian noise. Therefore, the computer vision community often uses the sum of absolute or squared differences and optionally limits differences (i.e. truncated costs as discussed by Scharstein and Szeliski (2002)) for reducing errors due to matching non-corresponding pixels. Non-

parametric costs like Rank and Census further reduce the problem (Zabih and Woodfill, 1994) since outliers have a much lower weight. The problem can also be reduced by adapting the size and shape of windows (Kanade and Okutomi, 1994; Fusiello et al., 1997; Hirschmüller et al., 2002).

However, a real solution of this problem is only possible by matching pixels individually instead of matching windows. Of course, individual pixels do not contain enough information for unique matching. Therefore, global methods additionally use a smoothness constraint that penalizes discontinuities. This is typically formulated in a cost function,

$$E(D) = \sum_p (C(p, D_p)) + \sum_{q \in N_p} PT[|D_p - D_q| \geq 1]. \quad (1)$$

The first term of the function sums all pixel-wise matching costs over the whole image, while the second term adds a penalty for all pixels with neighbors that have a different disparity. In this way, discontinuities are permitted if pixel-wise matching is *stronger* than the penalty, i.e. if the texture indicates a discontinuity. It is noteworthy that the second term indirectly connects all pixels with each other in the image and makes the function global.

In this formulation, the disparity image D is sought that minimizes equation (1). Unfortunately, this is an NP complete problem (Boykov et al., 2001). Famous approximate solutions to this problem are known as Graph Cuts (Kolmogorov and Zabih, 2001) and Belief Propagation (Felzenszwalb and Huttenlocher, 2004). The drawback of these and many other global methods are the speed and memory consumption, which often does not scale well with image size.

SEMI-GLOBAL MATCHING (SGM)

Semi-Global Matching (Hirschmüller, 2005 and 2008) successfully combines concepts of global and local stereo methods for accurate, pixel-wise matching at low runtime.

Algorithm

The core algorithm considers pairs of images with known intrinsic and extrinsic orientation. The method has been implemented for rectified and unrectified images. In the latter case, epipolar lines are efficiently computed and followed explicitly while matching (Hirschmüller et al., 2005).

Matching Costs

Radiometric differences often occur due to different imaging characteristics of the camera due to the vignetting effect, different exposure times, etc, and properties of the scene like non-lambertian reflection, which is viewpoint dependent, changing of the light source (e.g. movement of the sun), etc. The imaging characteristics of the camera can be calibrated and images corrected accordingly. However, the properties of the scene and lighting are typically unknown. Thus, the matching cost, which measures the dissimilarity of corresponding pixels, has to handle radiometric differences.

The original SGM implementation (Hirschmüller, 2005) used Mutual Information (MI) as matching cost. MI has been introduced in computer vision by Viola and Wells (1997) for the registration of images from different sensors and is extensively used in the medical image community. With MI, the global radiometric difference is modeled in a joint histogram of corresponding intensities. Kim et al. (2003) applied MI to pixel-wise matching with Graph Cuts by Taylor expansion. MI has been successfully adapted for SGM (Hirschmüller, 2005 and 2008). The pixel-wise MI matching cost is very suitable for matching unrectified images, as corresponding image parts may be rotated or scaled against each other, which is irrelevant if only individual pixels are considered.

An extensive study of different matching costs (Hirschmüller and Scharstein, 2009) showed that MI can model all global radiometric differences as well as image noise very well, but it degrades with increasing local radiometric differences as caused by vignetting, different shadows, etc. Furthermore, MI does not scale well with increasing radiometric depth (i.e. 12 or 16 bit instead of 8 bit quantization per pixel), since the joint histogram becomes too sparse in this case.

The same study identified Census as the most robust matching cost for stereo vision. Census has been introduced by Zabih and Woodfill (1994). It encodes the local neighborhood (e.g. window of size 9x7 pixel) around each pixel into a bit vector that only stores if the compared, neighboring pixel has a lower value than the center pixel or not. Pixel-wise matching is done by computing the Hamming distance of bit vectors of corresponding pixels. In this way, all radiometric changes that maintain the local ordering of intensities become completely irrelevant. Census is also well suited for local radiometric changes since it considers the local neighborhood only. However, for matching unrectified images, the affine transformation of the local neighborhood has to be considered.

Since the Census transformation is window based, it suffers from the same problem as correlation based stereo methods. However, the weight of outliers in the window is very low (i.e. 1 bit per outlier) and SGM uses the Hamming distances pixel-wise and not the sum of Hamming distances over another window as proposed in the original publication (Zabih and Woodfill, 1994). In general, Census is slightly inferior to MI, if there are only global radiometric differences, but it is much better for local radiometric changes, which are present in many real-world applications.

Pathwise Aggregation

SGM uses a slightly different global cost function as shown in equation (1) for penalizing small disparity steps, that are often part of slanted surfaces, less than real discontinuities,

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

The cost function is optimized similarly to scanline optimization (Scharstein and Szeliski, 2002), which can be seen as a subset of dynamic programming. It efficiently optimizes equation (2) in a time that is linear to the number of pixels and disparities. However, the optimization is performed in 1D only (i.e. typically along scanlines) and not in 2D, which is NP complete. Due to individual optimizations along scan lines, streaking artefacts are inherent in scanline optimization and dynamic programming methods.

The novel idea of SGM is the computation along several paths, symmetrically from all directions through the image. As shown in Figure 2, eight paths from all directions meet at every pixel.

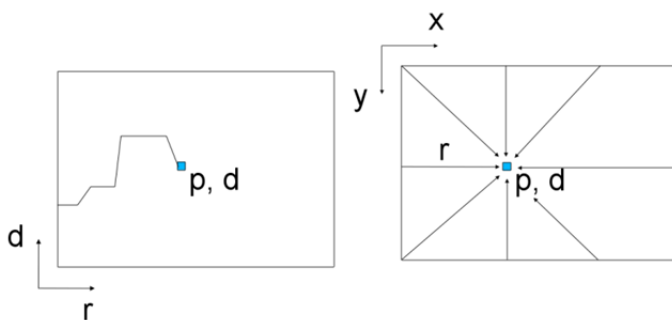


Figure 2: Eight optimization paths from different directions meet at every pixel.

Each path carries the information about the cost for reaching a pixel with a certain disparity. For each pixel and each disparity, the costs are summed over the eight paths. Thereafter, at each pixel, the disparity with the lowest cost is chosen. This formulation ignores occlusions. Thus, arbitrary results occur in occluded areas. However, occlusions can be identified by computing the disparities separately for the left and right image and comparing the results by a left-right consistency check. Further post-processing steps are possible for cleaning up the

disparity image. More details about SGM aggregation, matching of multiple and huge images is given in the literature (Hirschmüller, 2008).

CPU AND CLUSTER IMPLEMENTATION

The algorithm has been implemented and optimized by coding some core loops manually with vector commands (i.e. SSE2 instruction set), that can operate on multiple values in parallel. Nevertheless, matching many huge images, as in case of aerial image processing, requires a lot of processing time, but each image can be matched independently against its neighboring, overlapping images. Therefore, SGM based image processing has been implemented on a computer cluster, which consists of two blade systems with a total of 256 Intel X5570 CPU cores and another blade system with older and slower CPUs. Each CPU core has 3 GB of RAM available for running one matching process. A GPFS¹ based central file system with a total capacity of 55 TB is distributed over three RAID systems and accessed through four server computers for high performance access to image data. Processing is controlled by the Grid Engine² software.

GPU AND FPGA IMPLEMENTATION

For real-time robotics applications, SGM has been implemented on the GPU (Graphics Processing Unit) using OpenGL/Cg as programming language (Ernst and Hirschmüller, 2008). Mutual Information as well as Census have been implemented as matching cost. The implementation runs on a NVidia GeForce GTX 275 on rectified images with a size of 640×480 pixel and 128 pixel disparity range with about 4.5 Hz. Other examples of SGM implementations on the GPU are in the literature (Gibson and Marques, 2008; Rosenberg et al., 2006).

In a parallel development effort, FPGA (Field Programmable Gate Array) implementations of SGM with Census as matching cost have been investigated. Since the SGM algorithm has a very regular structure with operations that can be reduced to comparing and adding integer values, it is well suited for massive parallel implementations on FPGAs. Gehrig et al. (2009) presented a SGM implementation on a FPGA that processes images at 27 Hz and requires less than 3 W of power. Banz et al. (2010) published an idea for reducing the external memory bandwidth, which is the bottleneck for an SGM implementation on a FPGA.

These GPU and FPGA implementations target at real-time applications on rather small images. For processing huge images with large disparity ranges, the DLR Institute of Robotics and

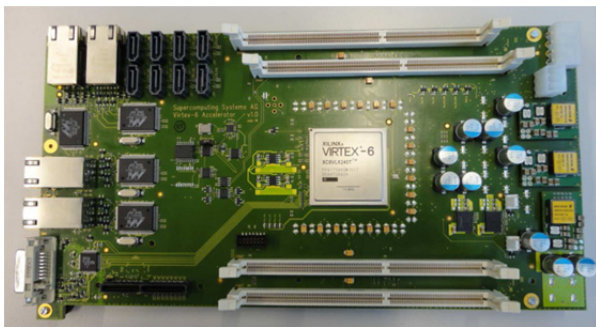


Figure 3: SGM accelerator board for huge images with large disparity ranges.

Mechatronics delegated an implementation to the Supercomputing Systems AG³. Within the project, a new Xilinx Virtex 6 (LX240) board has been developed with four parallel memory banks, equipped with 2 GB DDR3 memory modules for avoiding a memory bus bottleneck. The communication with a host PC is implemented by Gigabit Ethernet (Figure 3).

The SGM implementation expects rectified images with 16 bit radiometric depth and uses Census as matching cost. Images can have a width and height between 512 and 2048 pixels in steps of 8 pixels.

The disparity range can be chosen between 32 and 4096 pixels for an image width up to 1024 pixels. It is limited to 1024 disparities for images with a larger width. The width of the right image is extended by the disparity range, such that each pixel of the left image has a correspondence in the right image, even if the disparity range is larger than the image width. This feature is used for tilewise processing of huge images. Left-right consistency

¹ <http://www.ibm.com/systems/software/gpfs>

² <http://www.gridengine.org>

³ <http://www.scs.ch>

checking is done on the host PC, since it is computationally not demanding. The implementation is thought as an external SGM acceleration box. The communication via Gigabit Ethernet permits creating a network of host PCs and processing boards.

Currently, the FPGA implementation is running at only half of the planned core frequency of 160 MHz. The implementation requires 4.6 s for matching twice (i.e. for left-right consistency checking) an image of size 2048×2048 pixels with 1024 pixel disparity range. In comparison, the optimized CPU implementation, running on a very fast Intel X5570 CPU with 3.0 GHz is 41 times slower than the FPGA implementation. The goal of the current optimization phase of the project is to double the speed of the FPGA. Despite this high computational performance, the FPGA box has a total power consumption of just 54 W, including power supply. This speedup and the advantage of significantly less hardware costs, infrastructure and air-conditioning in comparison to CPU or GPU solutions, makes the FPGA a very attractive alternative for processing huge amounts of data in short time.

EVALUATIONS

SGM has participated in several tests and evaluations. The Middlebury stereo pages (Scharstein and Szeliski, 2011) currently list 108 stereo methods. The C-SGM variant (Hirschmüller, 2006) that is modified for structured indoor scenes has a Rank of 30 and an average error of 5.8%. In comparison, the currently best method has an average error of 4%. However, runtime and scalability, with respect to image size, is not considered in the evaluation. Furthermore, the test image set is quite limited as it includes only four pairs of small images of indoor scenes. Practical experience has shown that SGM is not only much faster and more scalable than most other methods, but also robust in very different applications and does not require parameter tuning.

Another test compares SGM and other methods for reconstructing the Martian surface from HRSC images that have been taken by the ESA probe MarsExpress (Heipke et al, 2006; Hirschmüller et al, 2006). In this application, SGM performed as well as other methods that were specially tuned to this application in contrast to SGM.

The main power of SGM can be demonstrated on aerial images of urban scenes. A study (Hirschmüller and Bucher, 2010) compared SGM based digital surface models from images of different commercial aerial cameras with each other, with a laser model and ground control points. Very precise DSMs in the ground sampling distance of the images were computed, especially on datasets with sufficient image overlap (Figure 4). The competitiveness of SGM as compared to aerial laser scanning has also been confirmed by others (Gehrke et al., 2010).



Figure 4: Automatic untextured and textured reconstruction from DGPF UltraCam-X images with 8 cm/pixel ground sampling distance (from Hirschmüller and Bucher, 2010).

APPLICATIONS

The applications for SGM range from aerial and satellite image processing and robotics to driver assistance systems. It is noteworthy that algorithmic parameters never need to be changed within an application and many parameters are also fixed between very different applications like aerial image matching and robotics, which shows the generality and maturity of the approach.

AERIAL AND SATELLITE IMAGE PROCESSING

The SGM cluster implementation (Section 3.2) has been used in various projects for creating digital surface models (DSM) and ortho images from aerial pinhole, pushbroom as well as satellite images. More than 100 TB of registered images have been processed fully automatically in recent years.



Figure 5: Automatic untextured and textured reconstruction from DGPF UltraCam-X images with 8 cm/pixel ground sampling distance (from Hirschmüller and Bucher, 2010).

Figure 5 shows a typical result from high resolution aerial images. The DSM is always created in the ground sampling distance of the images, with sharp object boundaries and fine details. Side textures for vertical structures like building walls are generated fully automatically from aerial images as well, which results in photo realistic, high resolution city reconstruction.



Figure 6: Reconstruction of Zugspitze in 20 cm/pixel from HRSC pushbroom images.

Figure 6 shows a reconstruction of Germany's highest mountain "Zugspitze" in 20 cm/pixel from pushbroom images, taken by the High-Resolution Stereo Camera (HRSC) that has been developed at the DLR Institute of Planetary Research. Pushbroom images are matched unrectified, by explicitly calculating and following epipolar lines (Hirschmüller et al., 2005), since pushbroom images cannot be rectified in general.

A reconstruction of Berlin from satellite images is shown in Figure 7. Despite the low spatial and radiometric resolution as compared to aerial images, all houses and main structures are precisely represented in the DSM. In this case, only an ortho-image has been created from panchromatic input images for texturing the model.

SGM based reconstruction of the worlds highest mountain (Mt. Everest) is shown in Figure 8. As in all other examples, the DSM has been created in the ground sampling distance of the images, which

is in this case 50 cm/pixel. The data set contains a height range of almost 5000 m from the lowest point to the mountain top, which results in disparity ranges of several thousand pixels. The maturity of the approach has convinced many researchers and companies to re-implement SGM for remote sensing applications (e.g. Gehrke et al., 2010).

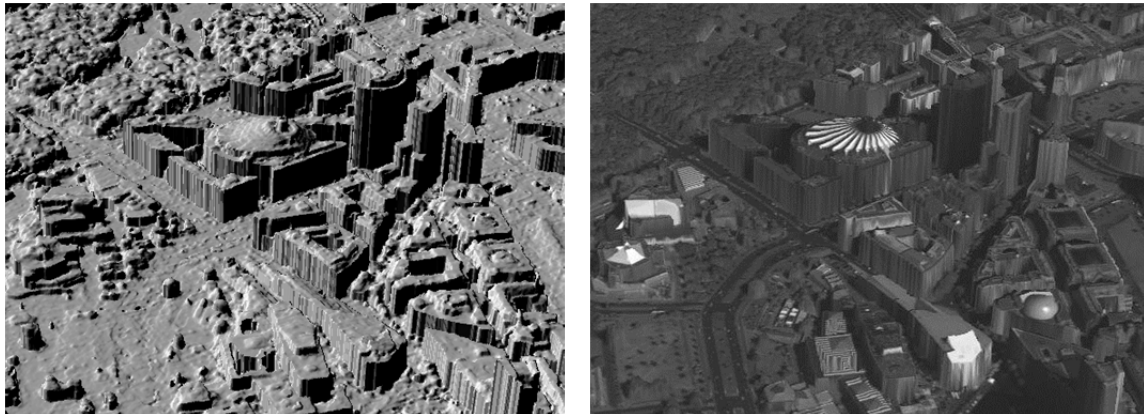


Figure 7: Reconstruction of Berlin in 50 cm/pixel from World View satellite images, provided by Digital Globe.

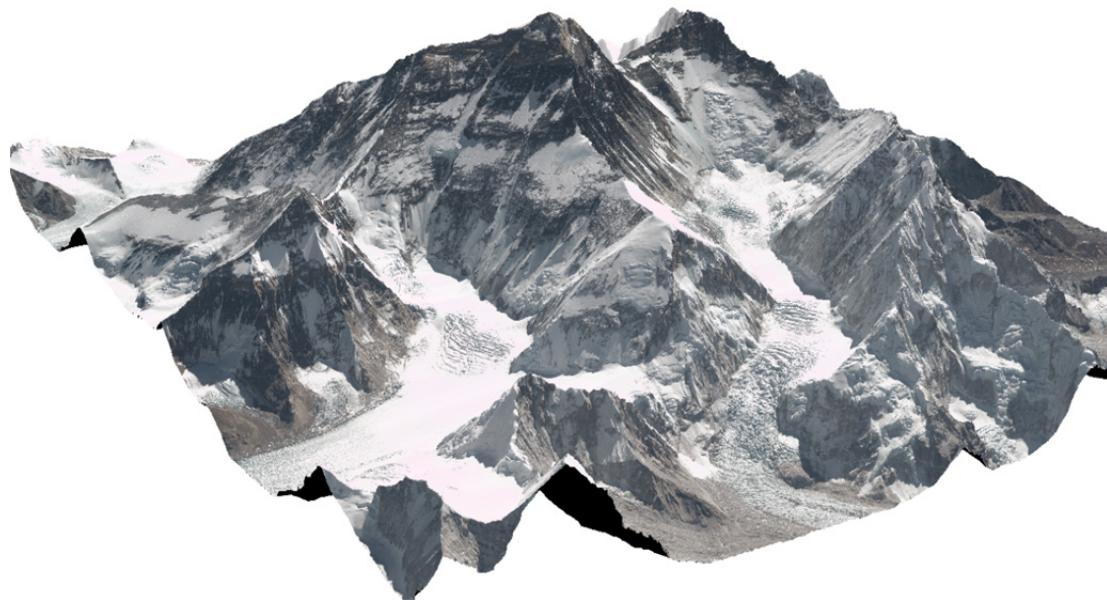


Figure 8: Reconstruction of Mt. Everest in 50 cm/pixel from World View satellite images, provided by Digital Globe.

MODELLING FROM LIGHT WEIGHT, UNMANNED MULTICOPTERS

The creation of DSMs and fully textured reconstructions from consumer grade cameras is another application for SGM. In this case, an off-line 3D viewpoint planning algorithm was used on a coarse DSM (Schmid et al., 2011). An optimal path was planned through all viewpoints and flown autonomously by a light weight multicopter. Images were taken by a 10 MPixel Panasonic Lumix DMC-LX3. Image registration was performed by the bundler software (Snavely et al., 2008) and matched by SGM. The results in Figure 9 appears very accurate with a high level of details.

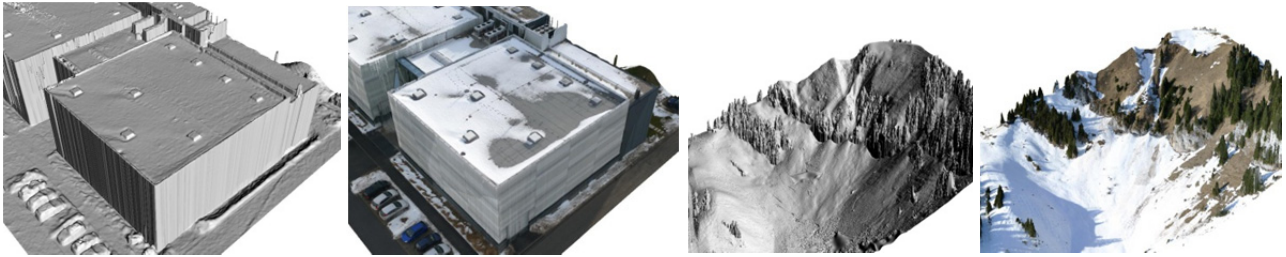


Figure 9: Building “DLR TechLab” and hillside, reconstructed by SGM in 5 cm/pixel from images of a 10 MPixel camera mounted on a flying multicopter (from Schmid et al., 2011).

MOBILE ROBOTICS

Real-time implementations of SGM on the GPU are used in robotics. Figure 10 shows the DLR crawler (Görner et al., 2010), which is equipped with a synchronized stereo camera. Stereo matching and visual odometry with six degrees of freedom are performed in 5 Hz. The depth images and visual odometry information are used for mapping and navigation in unknown, rough terrain (Chilian and Hirschmüller, 2009) for search and rescue as well as for planetary missions. Current work includes the on-board implementation of SGM on FPGAs for mobile robotics.

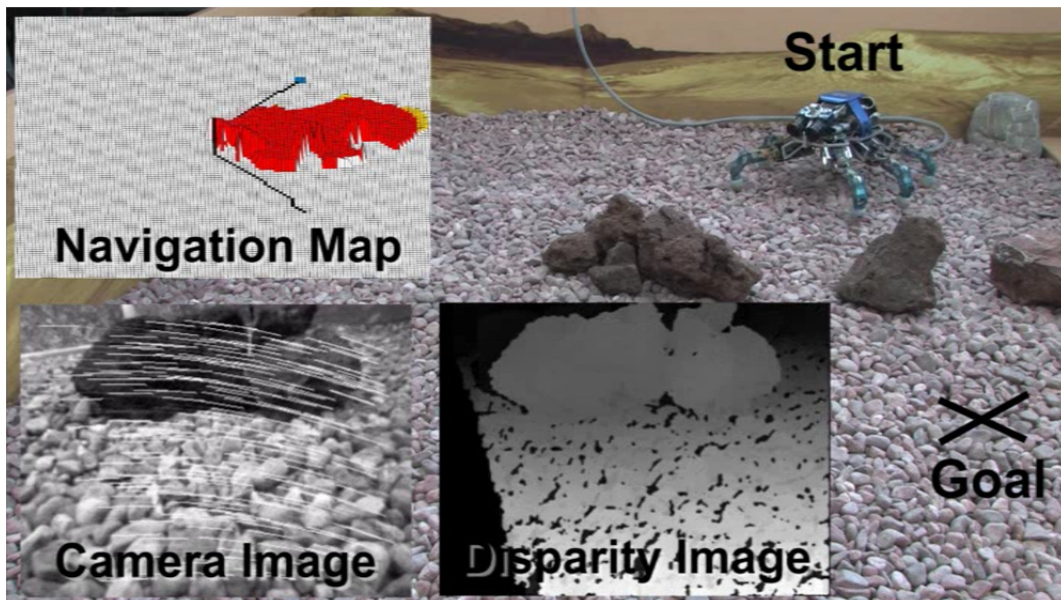


Figure 10: DLR Crawler with stereo camera and real-time GPU implementation of SGM.

DRIVER ASSISTANCE SYSTEMS

SGM has also been discovered by Daimler AG for supporting driver assistance systems. In this application, the car is equipped with a stereo camera that observes the scene in front of the car as shown in Figure 11. A low-power FPGA implementation of SGM (Gehrig et al., 2009) is used for computing depth images in real-time. The depth images are used as base for various driver assistance tasks. For this purpose, SGM gives much better results as correlation based stereo matching (Steingrube et al., 2009). Figure 11 shows that SGM delivers much more depth values. Moreover, it also produces significantly less errors in low contrast areas and shows robustness against repetitive structures like at the rail of the bridge.

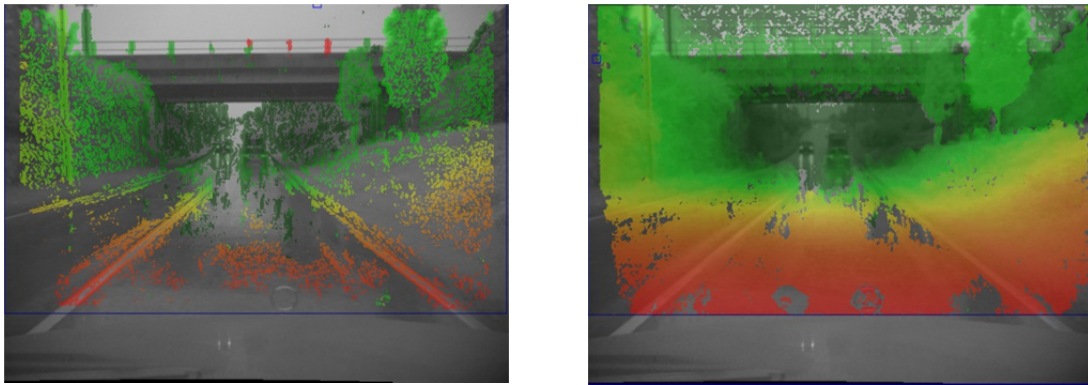


Figure 11: Camera image with depth image overlay. Red is near and green is far away.
Left: Stereo correlation. Right: SGM. Courtesy of Stefan Gehrig (Daimler AG).

FUTURE WORK

Current work includes the GPU implementation of SGM in OpenCL for getting rid of some OpenGL size and data type restrictions. Furthermore, FPGA based matching of huge images will be optimized for speed and FPGA implementations for mobile systems realized. Future work also includes the creation of real 3D models from SGM disparity images, additionally to the currently produced 2.5D surface models.

ACKNOWLEDGEMENTS

While the idea to SGM, its continuous algorithmic refinement and CPU implementation has been done by myself, several colleagues at the DLR Institute of Robotics and Mechatronics in Oberpfaffenhofen and Berlin were involved in its implementation on the GPU and FPGA. The FPGA hardware and implementation for huge images has been realized by Felix Eberli, Reto Stalder, Robert Hunger, David Stadelmann and Fabian Schenkel (Supercomputing Systems in Zürich, Switzerland) on our request.

For acquiring aerial images and performing image registration, I would like to thank many colleagues at the DLR Institute of Robotics and Mechatronics in Oberpfaffenhofen and Berlin, as well as colleagues at the DLR Institute of Planetary Research for HRSC image acquisition and preprocessing.

I would like to thank Michael Cramer (IFP, University of Stuttgart), the DGPF Projektkonsortium and Vexcel Imaging Graz for providing the registered UltraCam-X test images of Vaihingen/Enz within the project “DGPF Kameraevaluierung”. The presented reconstructions have been taken from an earlier publication (Hirschmüller and Bucher, 2010). Furthermore, I would like to thank Digital Globe for providing World View images of Berlin and Mt. Everest, which we have registered and matched by SGM for creating the presented reconstructions.

Last but not least, I would like to thank Stefan Gehrig and Uwe Franke (Daimler AG) for the fruitful cooperation, valuable feedback and the images shown in Figure 11.

REFERENCES

- Banz, C., Hesselbarth, S., Flatt, H., Blume, H. and Pirsch, P. (2010): "Real-Time Stereo Vision System using Semi-Global Matching Disparity Estimation: Architecture and FPGA-Implementation", IEEE Conference on Embedded Computer Systems (SAMOS X), July 2010.
- Boykov, Y., Veksler, O. and Zabih, R. (2001): "Fast Approximate Energy Minimization via Graph Cuts", IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (11), pp. 1222-1239.
- Chilian, A. and Hirschmüller, H. (2009): "Stereo Camera Based Navigation of Mobile Robots on Rough Terrain", IEEE International Conference on Intelligent Robots and Systems (IROS), October 2009, St. Louis, MO, USA.
- Ernst, I. and Hirschmüller, H. (2008): "Mutual Information based Semi-Global Stereo Matching on the GPU", International Symposium on Visual Computing (ISVC08), December 2008, Las Vegas, Nevada, USA.
- Felzenszwalb, P. F. and Huttenlocher, D. P. (2004): "Efficient Belief Propagation and Early Vision", IEEE Conference on Computer Vision and Pattern Recognition.
- Fusiello, A., Roberto, V. and Trucco, E. (1997): "Efficient stereo with multiple windowing", IEEE Conference on Computer Vision and Pattern Recognition, June 1997, Puerto Rico, pp. 858-863.
- Gehrig, S., Eberli, F. and Meyer, T. (2009): "A Real-Time Low-Power Stereo Vision Engine using Semi-Global Matching", International Conference on Computer Vision Systems (ICVS), LNCS 5815, Liege, Belgium.
- Gehrke, S., Morin, K., Downey, M., Boehrer, N. and Fuchs, T. (2010): "Semi-Global Matching: An Alternative to LIDAR for DSM Generation?", Canadian Geomatics Conference and Symposium of Commission I, ISPRS, June 2010, Calgary, Canada.
- Gibson, J. and Marques, O. (2008): "Stereo Depth with a Unified Architecture GPU", IEEE Conference on Computer Vision and Pattern Recognition.
- Görner, M., Chilian, A. and Hirschmüller, H. (2010): "Towards an Autonomous Walking Robot for Planetary Surfaces", International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS), August/September 2010, Sapporo, Japan.
- Grün, A. (1985): "Adaptive Least Squares Correlation – A Powerful Image Matching Technique", South African Journal of Photogrammetry, Remote Sensing and Cartography, Volume 14 (3).
- Heipke, C., Oberst, J., Attwenger, M., Dorninger, P., Ewe, M., Gehrke, S., Gwinner, K., Hirschmüller, H., Kim, J. R., Kirk, R. L., Lohse, V., Mayer, H., Muller, J.-P., Rengarajan, R., Schmidt, R., Scholten, F., Shan, J., Spiegel, M., Wählich, M., Yoon, J.-S. and Neukum, G. (2006): "The HRSC DTM Test", International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI, Part 4.
- Hirschmüller, H., Innocent, P. R. and Garibaldi, J. M. (2002): "Real-Time Correlation-Based Stereo Vision with Reduced Border Errors", International Journal of Computer Vision, 47 (1/2/3), pp. 229-246.

- Hirschmüller, H. (2005): “Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information”, IEEE Conference on Computer Vision and Pattern Recognition, June 2005, San Diego, CA, USA, Volume 2, pp. 807-814.
- Hirschmüller, H., Scholten, F. and Hirzinger, G. (2005): “Stereo Vision Based Reconstruction of Huge Urban Areas from an Airborne Pushbroom Camera (HRSC)”, Lecture Notes in Computer Science: Pattern Recognition, Proceedings of the 27th DAGM Symposium, Vienna, Austria, Volume 3663, pp. 58-66.
- Hirschmüller, H. (2006): “Stereo Vision in Structured Environments by Consistent Semi-Global Matching”, IEEE Conference on Computer Vision and Pattern Recognition, June 2006, New York, NY, USA.
- Hirschmüller, H., Mayer, H. and Neukum, G. (2006): “Stereo Processing of HRSC Mars Express Images by Semi-Global Matching”, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVI, Part 4, September 2006 in Goa, India.
- Hirschmüller, H. and Scharstein, D. (2007): “Evaluation of Cost Functions for Stereo Matching”, IEEE Conference on Computer Vision and Pattern Recognition, 18-23 June 2007, Minneapolis, Minnesota, USA.
- Hirschmüller, H. (2008): “Stereo Processing by Semi-Global Matching and Mutual Information”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 30(2), February 2008, pp. 328-341.
- Hirschmüller, H. and Gehrig, S. (2009): “Stereo Matching in the Presence of Sub-Pixel Calibration Errors”, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA.
- Hirschmüller, H. and Scharstein, D. (2009): “Evaluation of Stereo Matching Costs on Images with Radiometric Differences”, in IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 31(9), September 2009, pp. 1582-1599.
- Hirschmüller, H. and Bucher, T. (2010): “Evaluation of Digital Surface Models by Semi-Global Matching”, DGPF, Juli 2010, Vienna, Austria.
- Kanade, T. and Okutomi, M. (1994): “A stereo matching algorithm with an adaptive window: Theory and experiment”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16 (9), pp. 920-932.
- Kolmogorov, V. and Zabih, R. (2001): “Computing visual correspondence with occlusions using graph cuts”, International Conference for Computer Vision.
- Rosenberg, I. D., Davidson, P. L., Muller, C. M. R. and Han, J. Y. (2006): “Real-Time Stereo Vision using Semi-Global Matching on Programmable Graphics Hardware”, International Conference on Computer Graphics and Interactive Techniques – SIGGRAPH.
- Scharstein, D. and Szeliski, R. (2002): “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms”, International Journal of Computer Vision, 47 (1/2/3), pp. 7-42.

- Scharstein, D. and Szeliski, R. (2011): Middlebury Stereo Evaluation, Web Page, <http://vision.middlebury.edu/stereo>.
- Schmid, K., Hirschmüller, H., Dömel, A., Grixia, I., Suppa, M. and Hirzinger, G. (2011): “View planning for multi-view stereo 3D reconstruction using an autonomous multicopter”, ICUAS, May 2011, Denver, CO, USA.
- Snavely, N., Seitz, S. and Szeliski, R. (2008): “Modelling the World from Internet Photo Collections”, International Journal of Computer Vision, Volume 80, pp. 189-210.
- Steingrube, P., Gehrig, S. and Franke, U. (2009): “Performance Evaluation of Stereo Algorithms for Automotive Applications”, International Conference on Computer Vision Systems (ICVS), LNCS 5815, Liege, Belgium.
- Viola, P. and Wells, W. M. (1997): “Alignment by Maximization of Mutual Information”, International Journal of Computer Vision, 24 (2), pp. 137-157.
- Zabih, R. and Woodfill, J. (1994): “Non-parametric local transforms for computing visual correspondance”, Proceedings of the European Conference of Computer Vision, Stockholm, Sweden, pp. 151-158.