SURE – The ifp Software for Dense Image Matching

KONRAD WENZEL, MATHIAS ROTHERMEL, NORBERT HAALA, DIETER FRITSCH, Stuttgart

ABSTRACT

Dense image matching methods enable the extraction of 3D surface geometry from images acquired from multiple views. Tpyical applications vary from aerial imaging, where such methods can be used to retrieve digital surface models with high density, up to cultural heritage data recording, where the acquisition of images using digital cameras represents an efficient method to retrieve 3D data for documentation purposes. For every application, the desired density and precision of the 3D surface information can be selected flexibly by choosing an appropriate image sensor and acquisition configuration. Within this paper, the dense image matching software SURE is presented, which has been developed by the Institute for Photogrammetry at the University of Stuttgart. It uses a multi-view stereo (MVS) approach, where first stereo pairs are matched against each other. This stereo matching step is based on the library libTSGM, which implements a modified version of the *Semi Global Matching* (SGM) algorithm – enabling the determination of 3D information for almost each pixel. The modification uses a hierarchical approach, which enables the processing of complex scenes with large depth variations with short processing time and low memory consumption. Within a second step, the results of image matching are fused by triangulating rays for multiple stereo models at once. This improves the precision of object points, but also enables the rejection of outliers as well as the determination of quality values for each 3D point. The whole implementation is parallelized and optimized for scalability. Thus, large datasets regarding image size and image count can be processed on common desktop PCs. SURE is available online for free for non-commercial use at http://www.ifp.uni-stuttgart.de/publications/software/.

1. INTRODUCTION

3D reconstruction methods are applied in an increasing range of applications. For example, mobile phone images can be used to acquire objects in 3D, DSLR images can be used for surveying applications, UAV imagery can be used for flexible mapping purposes, up to large frame aerial applications, where thousands of images are acquired to reconstruct surface models of whole countries. Thus, one key challenge for image based 3D reconstruction algorithms is to be able to deal with large datasets of scenes with arbitrary geometry.

Currently, the algorithms for the retrieval of 3D information can be separated into two categories. The first category covers the retrieval of image orientations using manual or automatically determined distinct features in the images, followed by a bundle adjustment. The other category represents surface reconstruction methods, where dense image matching algorithms exploit the previously derived orientation of the images to derive complete surfaces.

This paper focuses on the latter – dense image matching. By finding corresponding pixels between multiple images, viewing rays of the corresponding object point can be intersected in space, which leads to the 3D position. In order to retrieve dense surface information, the correspondences are ideally determined for each pixel, leading to high density 3D surface information.

Even though many surface reconstruction algorithms have been published, only few solutions are publically available. Popular solutions are for example PMVS [Furukawa and Ponce, 2010] and its extension CMVS, as well as MicMac [Pierrot-Deseilligny and Paparoditis, 2006] or CMPMVS [Jancosek & Pajdla, 2011]. SURE – the solution presented in this paper is publically available as well for free for non-commercial and research use. It has been developed for the use within very

different applications and offers high flexibility regarding scene complexity and dataset size, as well as moderate processing times and low memory consumption.

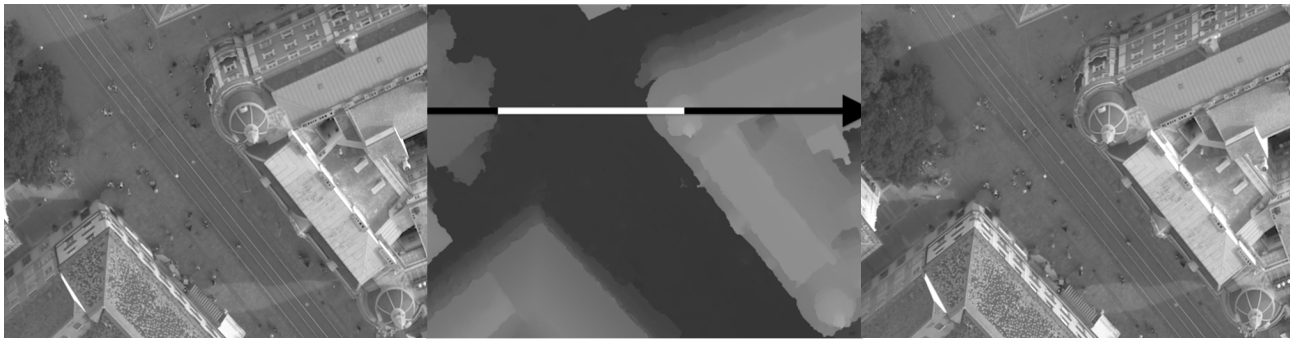

Figure 1: Left image, disparity image, right image. The disparity image or parallax image stores the correspondence information as disparity, which is correlated to the depth.

2. SURE

2.1. Semi Global Matching

At the Institute for Photogrammetry, dense image matching methods have been investigated and applied for several projects. In 2010 the *Semi Global Matching* method [Hirschmüller, 2005] has been implemented firstly, to evaluate its performance in aerial applications. The advantage of *Semi Global Matching* is its semi-global optimization, where neighboring pixels are taken into account to extract smooth surfaces. In contrast to local matching methods, such as correlation where pixel windows are compared separately, the resolution of matching ambiguities is much more reliable. This is in particular important for repetitive or low textured images. In such areas, *Semi Global Matching* is still able to retrieve reliable results. Beside higher matching stability, matching can be performed pixelwise with global matching methods, which does not lead to smearing effects like for local methods. While other global matching methods suffer from high computational efforts, *Semi Global Matching* uses a recursive method, approximating the smoothness constraint as 1D paths through the image, which enables efficient implementations.

Figure 1 shows a typical result of *Semi Global Matching.* The right hand and the left hand image represent an epipolar image pair. By using epipolar rectification, corresponding pixels only need to be searched along the column direction of the image, which reduces the complexity. The middle image shows a disparity image, where a disparity (or parallax) is stored for each pixel. As visible, a disparity can be determined for each pixel – leading to consistent surfaces. Within subsequent steps, occluded and inconsistent pixels can be filtered from this disparity image, e.g. by enforcing consistency within a left-right and right-left match.

The result of the stereo matching method is the disparity image – containing the correspondence information. In order to retrieve 3D coordinates from the image, a triangulation has to be performed. For this purpose, the viewing rays corresponding to the pixel measurements are constructed using the orientation, in order to intersect them in space. For the case of matching within a stereo pair, each pixel has one corresponding measurement. Thus, two rays are intersected, leading to one 3D point for each successfully matched pixel. Figure 2 shows a result for such a single stereo pair.

Figure 2: Results for the matching and triangulation for a single stereo model (Microsoft Ultracam, 8cm GSD).

2.2. Hierarchical strategy

Within the original algorithm of *Semi Global Matching*, the disparity range to be evaluated for each pixel is fixed over the whole image. For each pixel, exactly the same range of parallaxes is evaluated within a cube shaped cost structure as shown in figure 3. Subsequently, the surface is extracted from this cube, where the matching cost (the similarity measure) is optimal (low). This is suitable for aerial imagery and images with low resolution. In contrast, close range images or images with high ground resolution as well as convergent image acquisition can result in very large parallax ranges over the image. This leads to a high memory consumption and computation time if a fixed range is used.

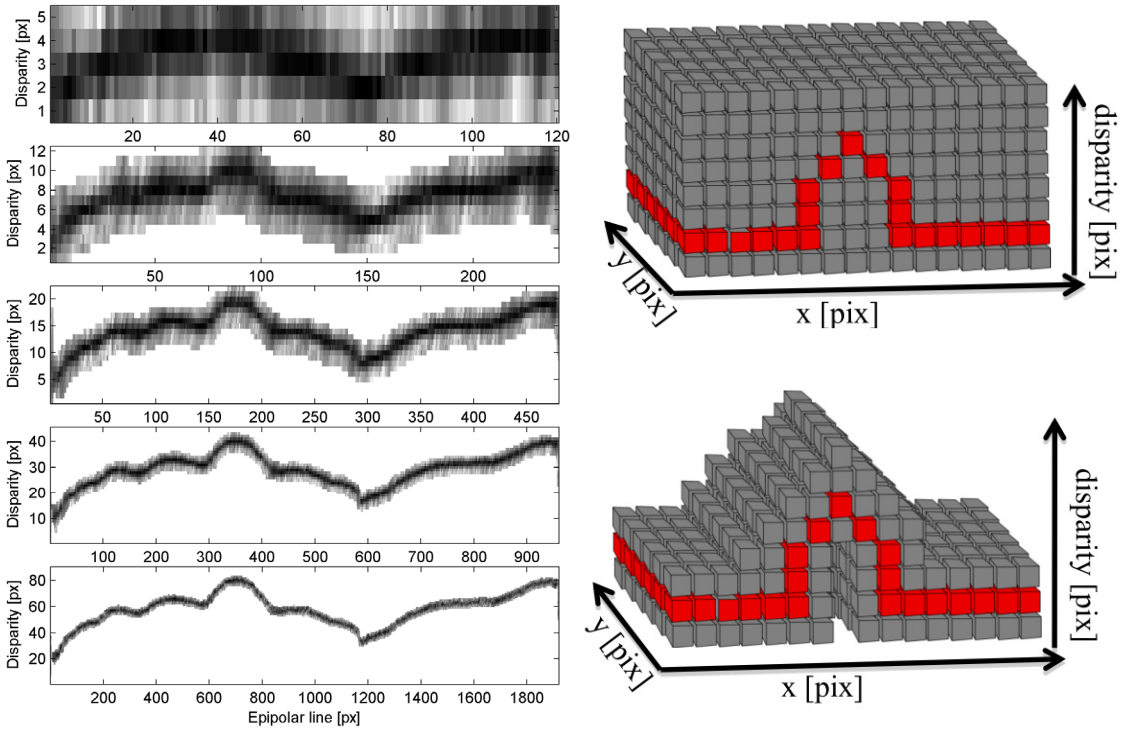

Figure 3: Hierarchical strategy. The disparity search range for each pixel is narrowed down using different resolution levels. Left: from low resolution to high resolution, the range is narrowed down for each pixel individually, enabling the processing of large depth variations at low memory consumption and processing efforts. Upper right: The original cost cube as used in SGM and the dynamic concept as used within SURE (lower right).

Thus, we developed a hierarchical approach, where the search range is narrowed down using different resolution levels [Wenzel et al., 2011, Rothermel et al., 2012]. For this purpose, an image pyramid is applied, where each level has quarter resolution (half image width and height). From the resolution of the image, an initial level is determined. Subsequently, an image matching is performed on the highest level of the image pyramid (lowest resolution) using a disparity search range over the whole image. This enables depth estimation without initial depth values, while consuming only little time on such levels. With each level the complexity is reduced by factor $2^3=8$, since it is a 3 dimensional problem, where the disparity range adapts to the resolution level.

After this quick initialization of the depth, the resulting disparity image can be passed to the next higher resolution level. Here, a new disparity search range can be determined from the previous result using the previous disparity of each pixel and adding a buffer. Within SURE, a method (*tSGM*) using a tube shaped disparity range is used. It is selected from a window surrounding the current pixel (e.g. 7x7 pixels) and added by a range of 16 disparities or 32 disparities. Consequently, a much smaller amount of disparities must be evaluated as visible in figure 3, leading to significant lower memory consumption and processing time. As shown within an example of a typical aerial imagery application in [Rothermel et al., 2012], about 70% memory can be saved while speeding up the computation by 30%. These savings increase even more for datasets with high resolution and large depth variations – enabling the processing using current desktop PCs.

2.3. Multi-view stereo extension

The direct triangulation of the result of a single stereo pair leads to reasonable results. However, even though the disparity image is filtered, mismatches can remain – for example poorly textured areas or objects moving along the epipolar direction, where the movement leads to invalid disparities and thus to invalid heights as shown in figure 4. In order to detect such outliers, multiple stereo models can be used at once within the triangulation step [Rothermel et al, 2011] [Haala et al., 2012]. Consequently, a multi-photo consistency is not used within the matching step, where more than two images would be matched at the same time, but the consistency is enforced within the triangulation step, as shown and described in figure 5 and [Rothermel et al., 2012].

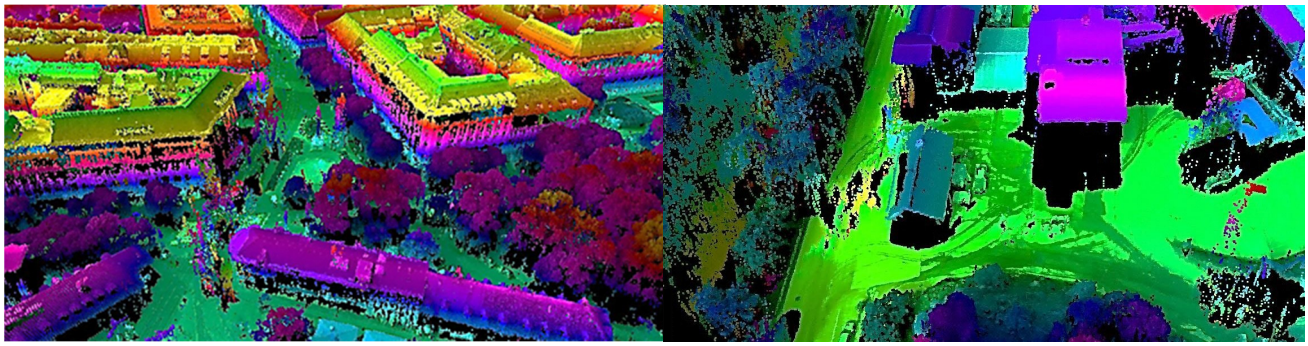

Figure 4: Typical problems for matching of single stereo models only – moving objects such as the tram in the middle of the left image or the river (left) and the car (lower right) in the right image lead to erroneous points. These problems can be overcome by a multi-view stereo extension as used within SURE.

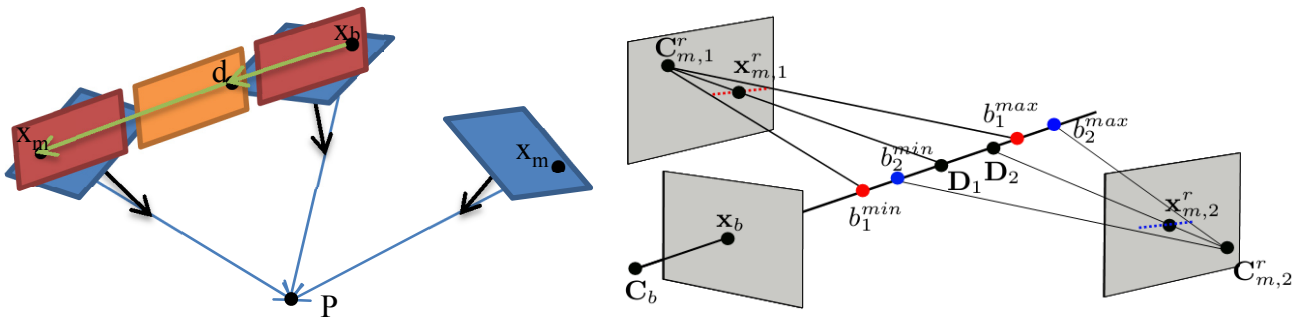

Figure 5: Consistency between multiple models enforced during the triangulation step. Left: from the disparity image (yellow), the correspondence between the pixels in the epipolar images (red) is known. By using the homographies from epipolar rectification, the correspondence between the epipolar images and the original images (blue) is known. Thus, for a pixel within a base image, the correspondence can be determined and triangulated for pixels in multiple stereo models. Right: triangulation can be simplified to a point along one ray instead of finding the point where multiple lines are closest. Confidence intervals along this ray (blue and red) are used to determine, if measurements from multiple models are consistent.

2.3.1. Automatic selection of stereo models

The selection of suitable stereo models is an important step within a multi-stereo solution like SURE. If approximate 3D information of the acquired object is available, e.g. a common terrain height or feature points from *Structure from Motion* methods, the optimal stereo models can be selected based on the evaluation of ray intersection angles between surface and cameras. However, SURE does not rely on such initial ground truth information and offers several functionalities to select suitable models only based on the orientation.

Within a first step, stereo models are filtered according to the viewing direction of the camera. For this purpose, the angles between the viewing directions of the cameras are determined for each model and stored. Therefore, divergent stereo models and models with too strong convergence can be detected and rejected (e.g. > 60 degrees). Within a second step, the models of each image are sorted according to their baseline. Only a certain amount of models per image is kept (e.g. 10 stereo models per image). Alternatively, suitable stereo models can be detected automatically using a coarse image matching on low resolution. Beside those default procedures, further functionalities for the selection of stereo models are available. Furthermore, the connectivity can be defined manually if preferred.

3. WORKFLOW FOR SURE

SURE can be used with any kind of overlapping images. Beside the images, the image orientation containing the interior (camera parameters) and the exterior orientation (camera rotation and translation) are used as input for SURE. Within SURE, dense image matching determines correspondences for each image pixel for multiple models per image, which can be triangulated simultaneously. The resulting dense point cloud representing the surface can be used to derive further products such as orthophotos or digital surface models (DSMs).

The image orientations serving as input for SURE can be determined by common orientation methods, such as camera calibration using patterns (e.g. for fixed camera setups) or automatically for unordered image datasets using *Structure from Motion* methods. For aerial applications, results from automatic aerial triangulation (AAT) can be used. Since the accuracy of the final point cloud

is directly dependent on the quality of the bundle block adjustment, the method should be selected to the specific requirements of the application.

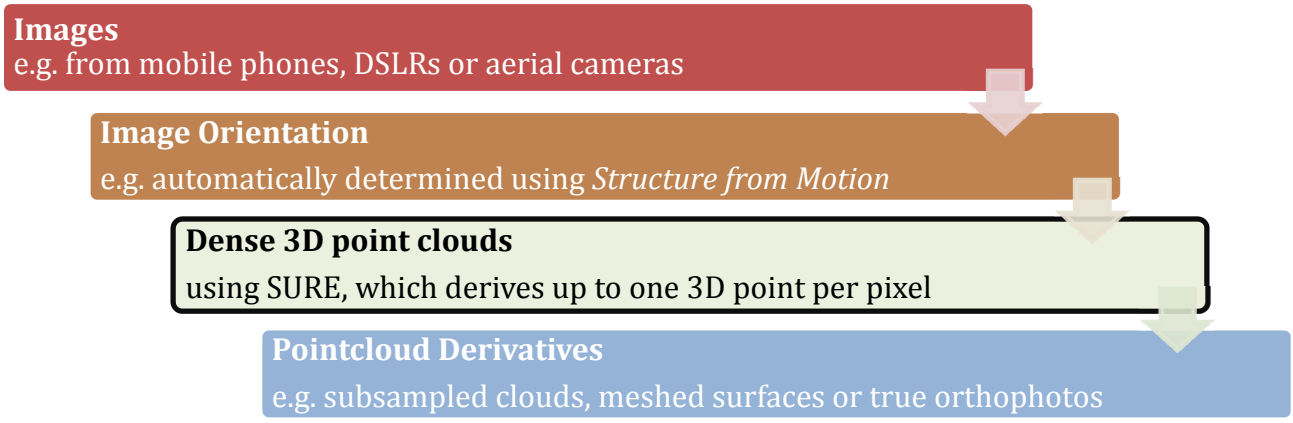

Figure 6: General data workflow for the usage of SURE within 3D reconstruction from images. The images are used to determine the orientation – for example with *Structure from Motion* methods, calibration methods using patterns, or aerial triangulation procedures. This orientation is used together with the imagery as input for the dense image matching step – in this case with SURE. The resulting dense point cloud can be used to derive further products.

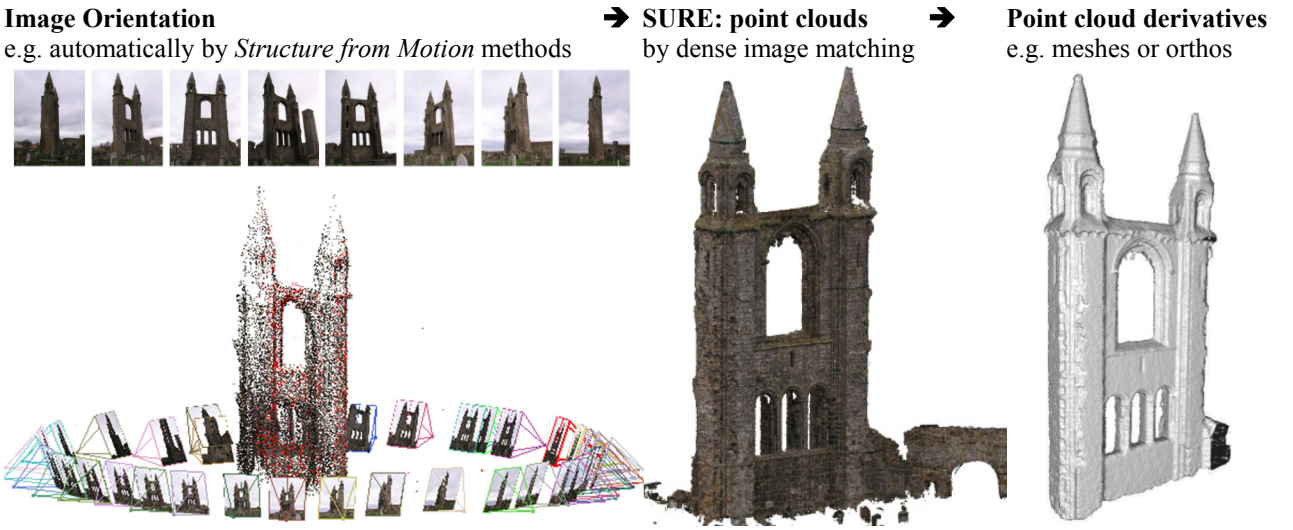

Figure 7: Typical workflow for using SURE. Firstly, the image orientations containing intrinsic camera parameters and exterior orientations are determined. In this example, the determination was performed automatically using the *Structure from Motion* software VisualSFM [Wu, 2011]. The shown dataset contains 43 images with 5 Megapixels resolution of the cathedral ruins in St. Andrews, Scotland. Subsequently, the orientations can be used to run SURE, which leads to a dense point cloud with 43 Mio 3D points. One derivative product can be meshed surfaces, like the example derived by volumetric surface integration [Zach, 2008] [Korcz, 2011].

The simplest way for the determination of orientation parameters are automatic methods such as *Structure from Motion*, where feature points are used to reconstruct structure from unordered image sets. Typical solutions are for example Bundler [Snavely et al., 2006], VisualSFM [Wu, 2011], Apero [Deseilligny and Clery, 2011] or commercial products such as Agisoft Photoscan or Pix4D. Other solutions are optimized for the use with large frame aerial imagery, such as Inpho/Trimble Match-AT. Fixed camera setups, e.g. for inspection purposes within industrial applications or stereo vision for cars, can be calibrated using software like OpenCV, Halcon (MVTec), Australis (Photometrix) or Photomodeller (Eos Systems). Independent of the used method, the resulting

orientations can be used to run SURE. No initial depth information is required, which is in particular beneficial for applications with fixed camera setups.

Several interfaces are available for the transfer of orientation parameters from arbitrary sources. Furthermore, several image distortions models are implemented, so that images are either passed to SURE undistorted, or are undistorted automatically. The final point cloud resulting from SURE can be used for the derivation of further products. One example is a true orthophoto, which can be used for mapping purposes. Due to the high point density of SURE, the point cloud can be projected onto an ortho plane without the need of interpolation. Other products can be meshed surfaces, e.g. derived by volumetric range image integration. Furthermore, 3D models can be extracted – e.g. for 3D city models.

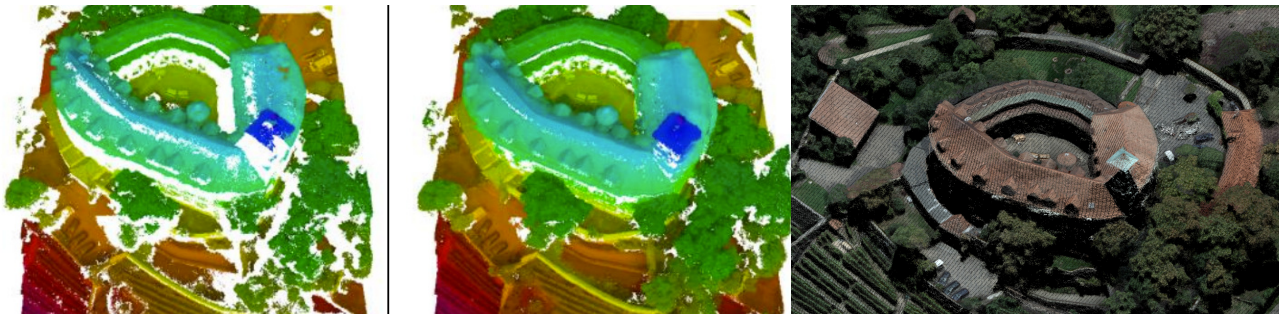

Figure 8: Point cloud derived with SURE for an 8cm GSD dataset. Left: single flight strip, Middle/Right: 3 flight strips.

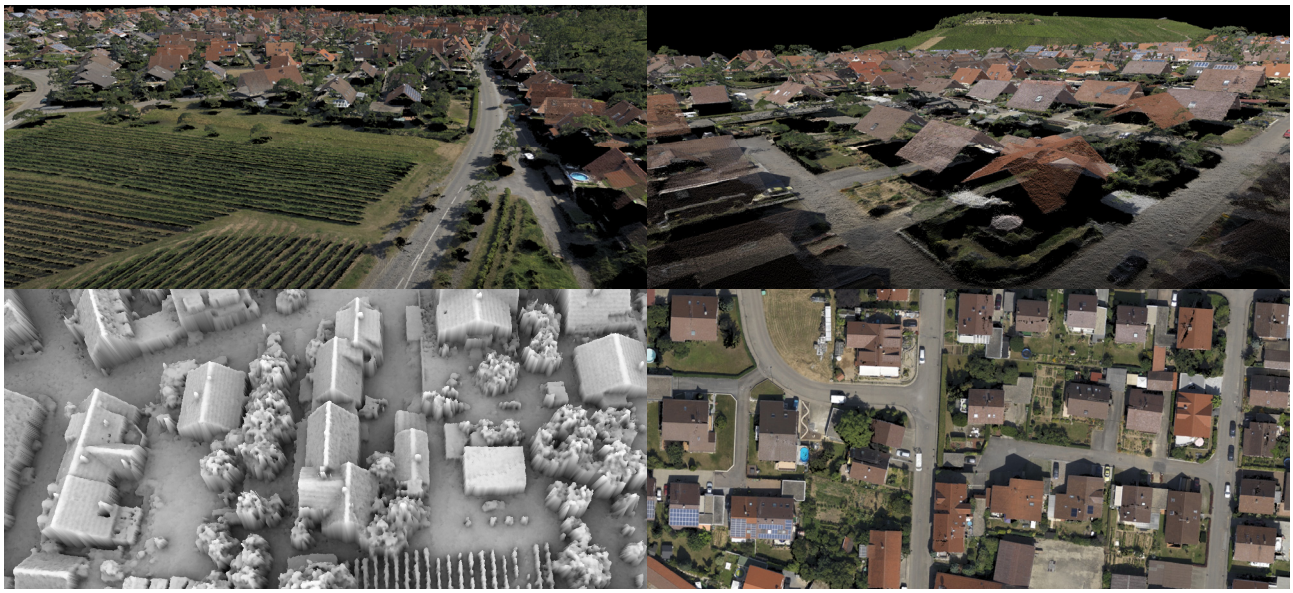

Figure 9: Aerial image dataset with 5cm GSD acquired from a gyrocopter. Upper images: point cloud, lower left: shaded point cloud, lower right: true orthophoto automatically derived from the point cloud.

4. APPLICATIONS

SURE can be used for a variety of applications, since it scales well to datasets with many and large images. This scalability is in particular possible due to the matching on single stereo models, where beside the hierarchical approach (*tSGM,* chapter 2.2) several techniques like image tiling are used to be able to process large images at low memory consumption. Thus, applications vary from large frame aerial imagery to mobile phone imagery.

4.1. Large frame aerial photogrammetry

For the typical airborne image acquisition of huge areas, large frame sensors are currently used in a nadir configuration with image overlaps of 60 to 85% in strip and 20-85% across strip. SURE can be used for the processing of such datasets on common desktop PCs. For example, 36 Microsoft Ultracam-X images of undulating terrain with 135 Megapixels each and overlap of 60% in both directions can be processed within 4 hours and 37 minutes using an Intel i7 processor with 4 cores [Haala, 2013]. Figure 8 shows another dataset of the Vaihingen Enz test site acquired using a Microsoft Ultracam-X at 8cm GSD and 80% overlap in and 60% overlap across strip. Another example is shown in figure 9, where a gyrocopter and a medium format IGI DigiCAM with 50 Megapixels was used to acquire 131 images with 5cm GSD [Fritsch et al., 2013]. The overlap was 80% in flight and 60% across flight direction. Figure 10 shows an example, where oblique images were acquired using an IGI Quattro DigiCAM Oblique [Fritsch et al., 2012].

Figure 10: Point clouds from a IGI Quattro DigiCAM Oblique with four cameras looking in a 45° angle.

Since precision and resolution are proportional to the image scale, image blocks can be acquired to the specific needs of the application. Consequently, high ground resolution can be used for datasets with need for high point density to higher depth precision, while lower resolutions might be suitable for large scale mapping applications. Nevertheless, high image overlap is beneficial for all datasets, since the resulting high redundancy improves the point precision while enabling the rejection of outliers. Also, models with higher image similarity are available, which is beneficial for image matching and consequently for the completeness. Furthermore, the amount of occlusions is low, which is in particular beneficial orthophoto projects.

4.2. UAV photogrammetry

Unmanned aerial vehicles enable cost efficient mapping by images. Different systems are currently available – in particular multicopters and fixed wing solutions. Fixed wing solutions are currently rather used for the acquisition of large areas at short time. Resulting blocks are similar to large frame aerial image datasets (e.g. [Cramer, 2013]). In contrast, multicopter systems can be used more flexibly – in particular in applications with different flying heights.

Figure 11 shows an example application, where an octocopter was used for the acquisition of a church façade [Cefalu et al., 2013]. While flying along the façade from up to down, one image per second was acquired. The resulting image dataset was automatically oriented using *Structure from Motion* methods [Abdel Wahab et al., 2012]. Subsequently, dense image matching was used to retrieve a dense point cloud. By defining the point cloud onto an ortho plane, an image for mapping purposes can be extracted, which can be used for drawings like stone cadasters as within this application. The high image and the resulting high point cloud and ortho image resolution are beneficial for the identification of small details.

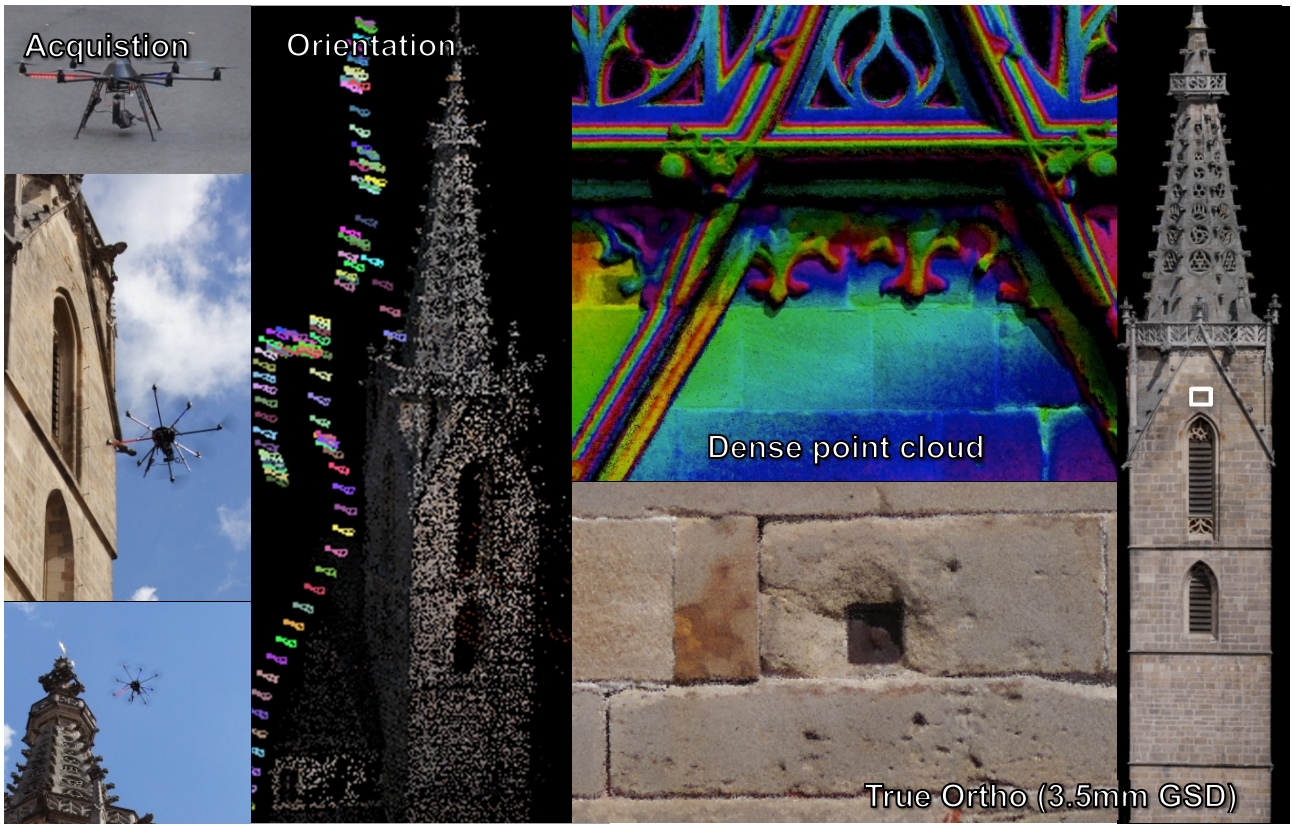

Figure 11: Façade acquisition in Rottenburg am Neckar, Germany. A church tower façade was mapped using an Octocopter. From left to right: image acquisition, automatic image orientation, dense point cloud, true orthophoto.

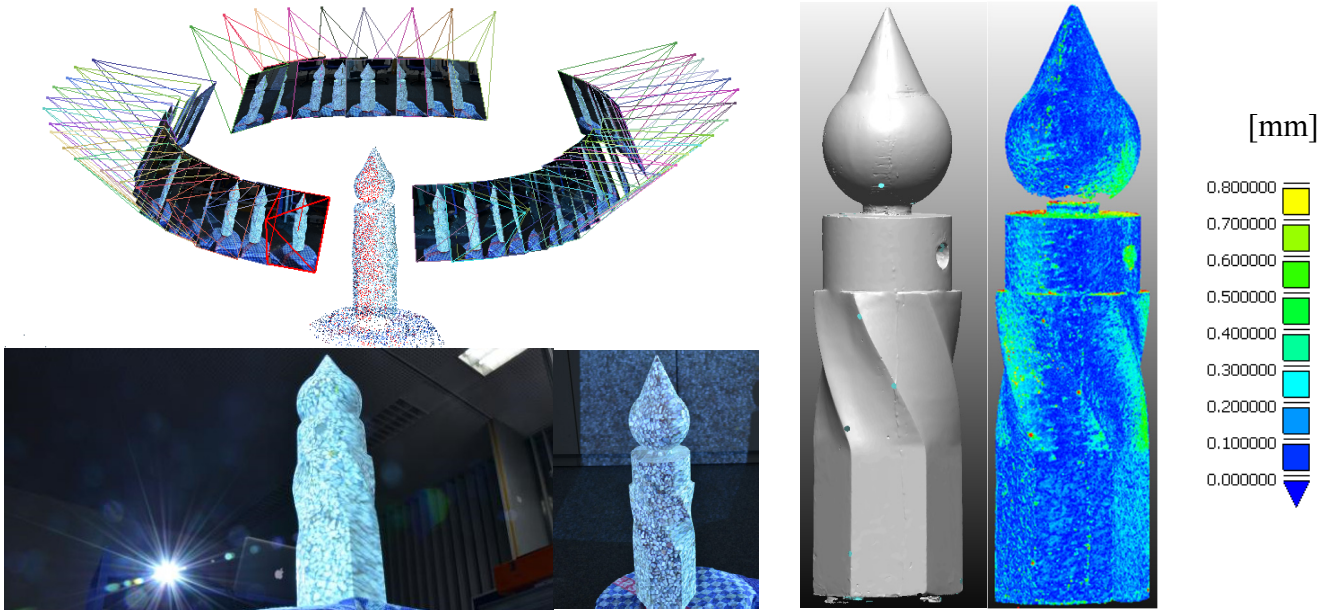

Figure 12: acquisition of the test object “Testy”. 45 images acquired with a Nikon D7000, oriented using *VisualSFM,* dense image matching with SURE – leading to a point cloud with 80 Mio. points. Right: reference surface acquired with GOM Atos 1 and difference to reference (standard deviation 0.18mm).

4.3. Close range imagery

Image datasets for close range applications can be acquired by mobile phones up to professional DSLR cameras. Also, industrial cameras can be used within a calibrated rig of multiple cameras to

be able to retrieve a dense point cloud from each shot, e.g. as shown in [Wenzel et al., 2011]. The ability of SURE to be not dependent on initial depth information is beneficial here.

Figure 12 shows an example, where the 35cm large test object "Testy" was acquired, which was developed by [Reulke et al., 2012] for the evaluation 3D measurement methods. Since it does not provide sufficient texture for dense image matching, artificial texture was projected using 3 video projectors. The point cloud resulting from SURE contains 80 Mio. points with a standard deviation to a reference surface of 0.18mm. Figure 13 shows an example of cultural heritage documentation, where 9 images were acquired using a Nikon D7000 DSLR camera.

Figure 13: Dense image matching for cultural heritage applications. From left to right: images, image orientation by *VisualSFM* (cameras and sparse point cloud), dense point cloud from SURE, shaded point cloud from SURE.

Imagery at close distance can have largely varying acquisition configurations. For a surface reconstruction by dense image matching, highly overlapping imagery with short baselines is beneficial. Here, image similarity is higher, leading to easier matching and thus, higher completeness of the resulting point cloud. Furthermore, the redundancy is beneficial. However, often it is difficult to acquire the imagery without gaps. Thus, taking a high amount of imagery is recommended – for example with the strategy "One panorama each step" [Wenzel et al., 2013].

5. CONCLUSION

The dense image matching implementation SURE is a publically available solution for surface reconstruction from imagery. It implements a modified version of *Semi Global Matching* for dense stereo matching within a multi-stereo frame. The modified stereo method *tSGM* narrows down disparity search ranges hierarchically, to enable the processing of imagery with large depth variations at short processing time and low memory consumption. The resulting disparities are used within a triangulation step, to intersect rays while using correspondence information from multiple stereo models at once. This enables an improvement on precision, the estimation of quality measures and the rejection of outliers. Thus, SURE represents a flexible solution for surface reconstruction from imagery.

SURE is available for free for non-commercial use at:
http://www.ifp.uni-stuttgart.de/publications/software/sure/index.html

6. REFERENCES

Abdel-Wahab, M., Wenzel, K., & Fritsch, D. (2012): Automated and Accurate Orientation of Large Unordered Image Datasets for Close-Range Cultural Heritage Data Recording. Photogrammetrie-Fernerkundung-Geoinformation, 2012(6), pp. 679-689.

Cefalu, A., Abdel-Wahab, M., Wenzel, K., Peter, M., & Fritsch, D. (2013): Image-based 3D Reconstruction in Cultural Heritage Preservation. ICINCO 2013 (to be published).

Cramer, M. (2013): The UAV@LGL BW Project – A NMCA Case Study. In: Photogrammetric Week 2013 (Ed. D. Fritsch). Stuttgart, Germany (to be published).

Deseilligny, M. P., & Clery, I. (2011, March): Apero, an open source bundle adjusment software for automatic calibration and orientation of set of images. In: Proceedings of the ISPRS Symposium, 3DARCH11.

Fritsch, D., Kremer, J., & Grimm, A. (2012): A Case Study of Dense Image Matching Using Oblique Imagery – Towards All-in-One Photogrammetry. GIM International. April 2012, Volume 26, No 4.

Fritsch, D., Grimm, A., Kremer, J., Rothermel, M., & Wenzel, K. (2013): Bilddatenerfassung mit einem Gyrocopter – Erste Erfahrungen zur "Photogrammetrie nach Bedarf". DGPF Tagungsband 22/2013, Dreiländertagung, Freiburg.

Furukawa, Y., & Ponce, J. (2010): Accurate, dense, and robust multiview stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(8), pp. 1362-1376.

Haala, N., & Rothermel, M. (2012): Dense Multi-Stereo Matching for High Quality Digital Elevation Models. Photogrammetrie-Fernerkundung-Geoinformation, 2012(4), pp. 331-343.

Haala, N. (2013): The Landscape of Dense Image Matching Algorithms. In: Photogrammetric Week 2013 (Ed. D. Fritsch). Stuttgart, Germany (to be published).

Hirschmüller, H. (2005): Accurate and efficient stereo processing by Semi Global Matching and Mututal Information. IEEE Conference for Computer Vision and Pattern Recognition, 2, pp. 807-814. San Diego, CA, USA.

Hirschmüller, H. (2005, June): Accurate and efficient stereo processing by semi-global matching and mutual information. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 2, pp. 807-814. IEEE.

Hirschmüller, H. (2008): Stereo processing by semiglobal matching and mutual information. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30(2), pp. 328-341.

Jancosek, M., & Pajdla, T. (2011, June): Multi-view reconstruction preserving weakly-supported surfaces. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. pp. 3121-3128. IEEE.

Korcz, D. (2011): Volumetric Range Image Integration. Diplomathesis, ifp, University of Stuttgart.

Pierrot-Deseilligny, M., & Paparoditis, N. (2006): A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36(1/W41).

Reulke, R., & Misgaiski, M. (2012): Test body "Testy" for Laser Scanning and Optical Systems. Photogrammetrie-Fernerkundung-Geoinformation, 2012(6), Zum Titelbild.

Rothermel, M., & Haala, N. (2011): Potential of Dense Matching for the Generation of High Quality Digital Elevation models. In Proceedings of ISPRS Hannover Workshop High-Resoultion Earth Imaging for Geospatial Information.

Rothermel, M., Wenzel, K., Fritsch, D., & Haala, N. (2012): SURE: Photogrammetric Surface Reconstruction from Imagery. Proceedings LC3D Workshop, Berlin, December 2012.

Snavely, N., Seitz, S. M., & Szeliski, R. (2006, July): Photo tourism: exploring photo collections in 3D. In: ACM transactions on graphics (TOG) (Vol. 25, No. 3, pp. 835-846). ACM.

Wenzel, K., Abdel-Wahab, M., Cefalu, A., & Fritsch, D. (2011): A Multi-Camera System for Efficient Point Cloud Recording in Close Range Applications. In: LC3D workshop, pp. 37-46.

Wenzel, K., Rothermel, M., Fritsch, D. & Haala, N. (2013): Image Acquisition and Model Selection for Multi-View Stereo. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-5/W1, pp. 251-258, 2013.

Wu, C. (2011): "VisualSFM: A Visual Structure from Motion System", http://homes.cs.washington.edu/~ccwu/vsfm/, 2011.

Zach, C. (2008, June): Fast and high quality fusion of depth maps. In: Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), Vol. 1.