IMAGE MATCHING FOR 3D PHOTOGRAMMETRIC RECONSTRUCTION

Ivan Detchev *[1], Ayman Habib [2] and Jiann-Yeou Rau [3]

[1] Graduate student, Department of Geomatics Engineering, University of Calgary,
2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada; Tel: + 1-403-220-3582;
E-mail: i.detchev@ucalgary.ca

[2] Professor and Head, Department of Geomatics Engineering, University of Calgary,
2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada; Tel: + 1-403-220-7105;
E-mail: ahabib@ucalgary.ca

[3] Assistant Professor, Department of Geomatics, National Cheng Kung University,
No. 1, University Road, Tainan 701, Taiwan; Tel: + 886-6-237-0876, Ext. 63839;
E-mail: jyrau@mail.ncku.edu.tw

KEY WORDS: Image matching, photogrammetric reconstruction, normalized cross-correlation, homogeneous texture, pattern projection

ABSTRACT: Image matching or the identification of conjugate points appearing in two or more photographs is an integral processing task in photogrammetry. More specifically, three-dimensional photogrammetric reconstruction, i.e. computing the object space coordinates of certain points of interest, would be impossible without having these points matched in the image space of at least two photographs. Observing a point of interest in only one photograph is simply not enough, because the range from the image point to the object point is inherently unknown. Thus, image matching is applied to satellite scenes and aerial photography for the generation of digital terrain models or building models, and also in close range photogrammetry for the reconstruction of objects and surfaces.

The human brain has incredible capabilities and manual image matching, i.e. one performed by a human operator, is very accurate and reliable; however, it is labour intensive, time consuming, and can become quite tedious very quickly. This is why the topic of fully automated image matching, i.e. one performed solely by a computer, has been addressed in photogrammetric research for a long time. While the problem of automatically matching signalized targets has been satisfactorily solved, the matching of natural or non-signalized features still requires some human interaction.

This paper will attempt to summarize the methodology of performing image matching of ordered photographs in a controlled environment, i.e. the interior and the exterior orientation parameters for the involved cameras and photographs are known. Example results from using a sophisticated commercial software package called CLOse RAnge MAtching (CLORAMA) and from a simple in-house matching program will be shown. The former method is based on least squares matching and requires very high resolution imagery when homogenous texture is present, while the latter method is based on the straightforward normalized cross-correlation matching and requires a pattern to be projected in order to create artificial texture.

1. INTRODUCTION TO 3D RECONSTRUCTION

Object or surface reconstruction derived from optical instruments is used in a wide range of applications. For example, surveying and mapping engineers often need digital elevation models of large areas resulted from airborne laser scanning or areal photography for the purposes of planning and design. Civil engineers make use of terrestrial laser scanning data and highly precise close range photogrammetric reconstruction of man made structures such as bridges, dams, open-pit mines, high-rises, or tunnels in order to monitor any occurring deformations. Biometric information obtained from dense scans or image processing is being used more and more in biomedical engineering, kinesiology, and the surveillance industry. Also, the creation of digital models of iconic buildings or aged monuments, for the purposes of architectural restoration or cultural heritage documentation, has already become the norm in the archaeology circles. In all these examples, the 3D reconstruction is done via range measurements using laser scanning technology or via photographic observations from digital cameras. The differences between the two optical modalities have been listed in Baltsavias (1999) for the airborne case and in Lichti *et al.* (2002) for the close range case. In the following two subsections of this paper, the advantages and disadvantages of using laser scanning or image-based reconstruction will be summarized from the close range point of view.

1.1 Terrestrial Laser Scanning

The main advantage of terrestrial laser scanning is that direct, non-contact and targetless acquisition of millions of points is possible within seconds. Laser scanning is also independent of natural light and can reconstruct points belonging to homogeneous surfaces. The shortcomings of laser scanning are that the single point positioning may have poor quality (e.g. millimetre to centimeter level precision), the scanning process is not instantaneous and the technology (both hardware and software) is very expensive. Also, breaklines and semantic information are not readily available.

1.2 Reconstruction from Photographs

One of the disadvantages of 3D reconstruction from photographs is that signalized targets are necessary for the automated computation of the image orientation. Also, external scale definition is needed, image matching between overlapping photographs must be performed, and last but not least it must be mentioned that photography depends on the light conditions of the surrounding environment. On the other hand, the benefits of image-based reconstruction are that the image acquisition can be nearly instantaneous, it only requires inexpensive off-the-shelf cameras (and/or other easily accessible electronic components), the collected image measurements are redundant, and the generated 3D points are very precise. Also, breaklines in terms of edges, and multispectral information (e.g. red, green, blue or grayscale colour) is available.

1.3 Choice of Optical Modality for 3D Reconstruction

The choice between the two depends on the user's budget and expertise, and on the application specifications. Actually, as seen from the previous two subsections, laser scanning and image-based reconstruction are not necessarily competing, but rather complementary technologies. Lately, the two technologies have been used in combination especially for cultural heritage purposes (Guidi *et al.*, 2009; Lambers *et al.*, 2007; Yastikli, 2007). This paper however assumes that a low budget is at hand and thus the focus will be only on the photographic reconstruction.

2. METHODOLOGY FOR IMAGE-BASED RECONSTRUCTION

The generation of 3D models from digital images has been under research investigation in both the computer vision and the photogrammetry fields. However, the two research parties have different priorities when it comes to the methodology and the final product. The computer vision community is primarily concerned with the automation and the speed of the processing, while the photogrammetry research community emphasizes more on the final precision of the results (Barazzetti *et al.*, 2011). Photogrammetric reconstruction achieves high precision, because the sensors used have high spatial resolution and because they undergo rigorous geometric calibration before being employed in a mapping project, where certain metric specifications must be met. The cameras usually used in computer vision, however, have lower resolution in order to handle the processing of multiple frames per second, and the reconstruction models usually follow a projective or an affine transformation which are fairly complex, but not rigorous (Barazzetti *et al.*, 2011; Patias, 2002). While the computer vision procedures are almost always fully automated and the final result may be visually pleasing, the derived 3D coordinates lack the necessary precision from an engineering point of view. In this research, priority is given to accuracy rather than automation, so the image-based reconstruction methodology will be presented from photogrammetric stand point.

The work flow of photogrammetric reconstruction generally follows the phases below:

- Project prerequisites;
- Data acquisition;
- Image pre-processing;
- Image processing;
- 3D point cloud generation;
- Point cloud quality control.

Each phase will be elaborated on in the next few subsections:

2.1 Project Prerequisites

In order for the final results to be reliable and as precise as possible, a few important prerequisites must be met before the actual data acquisition. First, the involved cameras must be calibrated in terms of their interior orientation parameters (also referred to as IOPs) so that the bundles of light rays for each particular camera can be correctly defined later on (Fraser, 1997; Habib and Morgan, 2003). Another essential task related to camera calibration is the stability analysis procedure or verifying that the camera IOPs do not significantly change over time (Habib and Morgan, 2005; Habib *et al.*, 2005). If the IOPs of the camera(s) do change significantly over short periods of time, the particular camera must be calibrated frequently or even self-calibrated on the job every time it is being used. In order to avoid the extra work and a potential loss of precision, it is advised that only cameras that are deemed stable should be employed for photogrammetric reconstruction. The last but not least prerequisite is that the exterior orientation parameters (also referred to as EOPs) must be estimated for each camera station. Knowing the EOPs is necessary to provide geometrical constraints during the image matching part of the image processing, and also to generate the final 3D point cloud. Both the camera calibration and the estimation of the EOPs are done in a bundle adjustment procedure. In cases where signalized targets are used, the bundle adjustment procedure could be fully automated and it could even be performed in near-real-time (Fraser and Edmundson, 2000). In this paper, in order to simplify matters, it is assumed that the IOPs and the EOPs are known and can be trusted.

2.2 Data Acquisition

Depending on the nature of the object or surface to be reconstructed, one camera at multiple stations or a synchronized system of cameras can be used. For example, if a static object is to be reconstructed one camera is enough, because the object will not change in between the different camera station epochs (Barazzetti *et al.*, 2011; Jazayeri and Fraser, 2010). However, if an object is in kinematic motion or it is somehow being deformed dynamically, then a minimum of two cameras (but usually three or more), which have the capability of operating simultaneously, is needed (Chong *et al.*, 2009; D'Apuzzo, 2002; Detchev *et al.*, 2011a). Also, in the cases where the object or surface to be reconstructed is very homogeneous, a pattern must be projected (Detchev *et al.*, 2011b) or some sort of grid must be applied (Jazayeri and Fraser, 2010) in order to create artificial texture. This artificial texture is helpful in terms of both completeness and reliability.

2.3 Image Pre-Processing

Typically, the image files collected with a digital camera are in either a raw format native to the particular camera or they are compressed in the more standardized JEPG format. Either way, the collected images are usually converted to the specific file format associated with the image matching software at hand. The next pre-processing step is to generate normalized images according to epipolar geometry. This guarantees that any conjugate features appearing in the overlap region of two neighbouring images will have the same y-image coordinates. At this stage, the user can optionally check for any y-parallax, and if such is present then this is an indicator that there are problems with the IOPs or the EOPs. After the epipolar normalization, the user can select the regions of interest in the collected images. The purpose for this pre-processing step is to limit the extent in the image processing to follow, and thus making the image processing both faster and more reliable. Other pre-processing steps can include building image pyramids (in the case of hierarchical image matching), selecting manual seed points (in the case of iterative dense image matching), and performing local contrast enhancement (in case there are shadows present in the input images). The local contrast enhancement can be effectively done with the Wallis filter (Jazayeri and Fraser, 2010). As seen from this subsection, most of the setup and manual interaction is done during the pre-processing phase.

2.4 Image Processing

The image processing phase is the core of the photogrammetric reconstruction work flow. First, an interest operator must be chosen. The role of the interest operator is to find features (e.g. points, edges or small regions) in the input images, which are going to be used as primitives in the image matching procedure to follow. Usually, a point or a corner detector is the preferred option. Among the many published operators, some of the most popular ones are:

- Förstner,
- Harris,
- Smallest Univalue Segment Assimilating Nucleus or SUSAN,
- Scale-Invariant Feature Transform or SIFT,
- Speeded Up Robust Feature or SURF, and
- Features from Accelerated Segment Test or FAST.

According to the investigations done by (Jazayeri and Fraser, 2010) the FAST operator outperforms the rest in both speed and accuracy.

The image matching procedure is the most important part of the image processing stage. Here, the correspondence between any existing conjugate features (e.g. the detected interest points or corners) is established (Remondino *et al.*, 2008). This could be done in pairs or in triplets, i.e. features appearing in the overlap area of any two or three neighbouring images can be identified and then – if applicable – tracked in the rest of the sequence of images. The identification of a candidate match is achieved through a similarity measure. For example, the normalized cross-correlation (NCC) coefficients between a small area around a feature of interest in one image and a number of other areas with the same size in the overlapping part of a neighbouring image can be computed. The features with the highest NCC coefficient are labeled as candidates for a matching pair as long as the computed coefficient is above a set threshold. The output from this NCC area-based matching could then be used as the initial values in a non-linear least squares matching (LSM) algorithm, which could yield candidates for matching pairs with sub-pixel precision (Remondino *et al.*, 2008). The key to having a successful matching algorithm is to limit the search space for each conjugate pair of detected features. This will decrease the computational time and at the same time increase the matching reliability. Limiting the search space can be achieved by applying geometric constraints, and the most common way of accomplishing this is through the previously described epipolar normalization, or by employing a hierarchical matching strategy via the image pyramids mentioned in the previous section.

2.5 3D Point Cloud Generation

Once all possible conjugate features are identified and tracked in all the images they appear, a multiple light ray intersection based on the collinearity equations can be run to compute the 3D coordinates of the observed points. The computation of the 3D point coordinates can be also done in a big bundle adjustment as well (Barazzetti *et al.*, 2010; Fraser and Edmundson, 2000). In the case when a certain motion is being tracked or some deformation is being monitored over numerous epochs in time, a recursive least squares algorithm or a Kalman filter can be used (Shao *et al.*, 2001).

2.6 Point Cloud Quality Control

After the final point cloud of the object or surface of interest is generated it must be checked both qualitatively and quantitatively. The qualitative check mostly involves the visual inspection of the point cloud in terms of completeness, smoothness, and point density. The quantitative check involves filtering out any points with large standard deviations, and if possible, comparing the derived 3D coordinates against any available control values or performing an iterative closest point (ICP) style surface match against any other available overlapping surface.

3. EXAMPLE RESULTS

This section of the paper will not provide a systematic review of existing image matching algorithms or photogrammetric reconstruction software packages. It will rather show some examples from literature and it will compare some preliminary results of using a commercial and an in-house software on data collected with a photogrammetric system having multiple cameras.

Barazzetti *et al.* (2011) published their work on reconstructing a damaged temple in Vietnam. The purpose of the reconstruction was to help with the temple restoration process. Some of their work involved generating a 3D point cloud from the commercial software PhotoModeler Scanner. The matching procedure in PhotoModeler is based on image pairs and it involves dense matching techniques. As a result, there were a lot of outliers in the generated point cloud and it took longer to clean it up than to generate it.

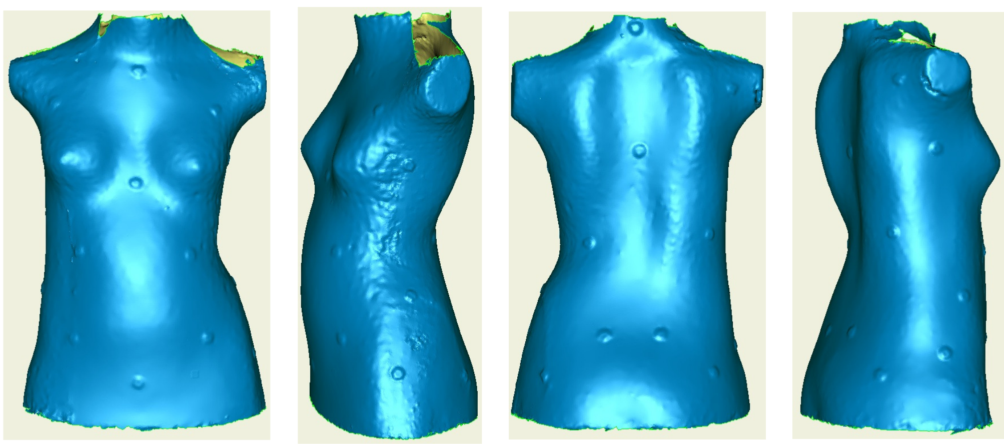
Jazayeri and Fraser (2010) published their work on the selection of an interest operator for feature-based close range photogrammetric matching. The combination of using the Wallis filter and the FAST corner detector is very impressive, because hundreds of thousands of points are detected and matched within a fraction of a second. However, the really high quality final results (e.g. 0.3 pixels for image coordinate precision, and less than 0.1 mm for object space standard deviations) comes at the cost of filtering out up to 95% of the matched features, and thus keeping only the top 5% of the observations.

Remondino *et al.* (2008) and Zhang and Gruen (2006) explain the complex algorithms behind the software package CLORAMA. The matching procedure used in this software is called multi-photo geometrically constrained (MPGC) matching. In addition to the basic processes explained in the previous section, it also employs multiple primitives, simultaneous multiple image matching, self-tuning matching parameters and other complex functions. On the other hand, the in-house software for image matching/photogrammetric reconstruction used in this research project is very simple (e.g. only image pairs and not triplets are matched at a time via NCC as opposed to LSM, the corner detector is set to only collect thousands and not hundreds of thousands of points, and the image coordinate precision is at the pixel level), but it still yields more than acceptable results (less than 0.5 mm in object space), and filtering or smoothing of the final point cloud is not necessary.

In order to compare the results from CLORAMA and the in-house software, a set of images of a torso mannequin (see Figure 1) were collected with a photogrammetric system, and processed with both packages (see results in Figure 1 and Figure 2).

(a) (b) (c) (d)

Figure 1. Photographs of the front (a), one of the sides (b), the back (c), and the other side (d) of the torso mannequin



(a) (b) (c) (d)

Figure 2. Rendered images of the front (a), one of the sides (b), the back (c), and the other side (d) of the 3D point cloud of the torso mannequin generated from CLORAMA

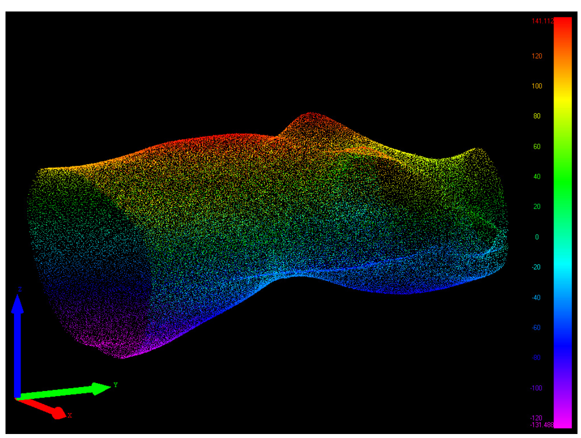


Figure 3. Raw point cloud generated from the in-house image matching/photogrammetric reconstruction software

4. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE DEVELOPMENT

In this paper, the advantages and disadvantages of using terrestrial laser scanning and image-based 3D reconstruction were reviewed. The image-based reconstruction was set as the optical modality of choice due to the much lesser cost compared to laser scanning. Next, the workflow of a photogrammetric reconstruction project was summarized with the project prerequisites, the data acquisition specifics, the image pre-processing and processing, and the point cloud generation and quality control being the main topics of discussion. Also, some general examples were given where even very well established software in the photogrammetric community needs clean-up or some sort of filtering of the final results. Even though, the in-house software used for this project yields more than satisfactory results, there are still a lot of room for development. For example, testing the functionality of the Wallis filter should be tried as well as the FAST corner detector. In addition, triplet matching could be used to further geometrically constrain the current matching procedure, and sub-pixel LSM can be experimented with as well.

LIST OF REFERENCES

Baltsavias, E.P., 1999. A comparison between photogrammetry and laser scanning. *ISPRS Journal of Photogrammetry & Remote Sensing*, 54(2-3): 83-94.

Barazzetti, L., Binda, L., Scaioni, M. and Taranto, P., 2011. Photogrammetric survey of complex geometries with low-cost software: Application to the 'G1' temple in Myson, Vietnam. *Journal of Cultural Heritage*, 12(3): 253-262.

Barazzetti, L., Scaioni, M. and Remondino, F., 2010. Orientation and 3D Modelling from Markerless Terrestrial Images: Combining Accuracy with Automation. *The Photogrammetric Record*, 25(132): 356-381.

Chong, A.K., Milburn, P. and Voert, M.t., 2009. High-accuracy Photogrammetric Technique for Human Spine Measurement. *The Photogrammetric Record*, 24(127): 264-279.

D'Apuzzo, N., 2002. Surface measurement and tracking of human body parts from multi-image video sequences. *ISPRS Journal of Photogrammetry & Remote Sensing*, 56(5-6): 360-375.

Detchev, I., Habib, A. and El-Badry, M., 2011a. Case study of beam deformation monitoring using conventional close range photogrammetry. Proceedings of the ASPRS Annual Conference, Milwaukee, Wisconsin, May 1-5.

Detchev, I., Habib, A. and El-Badry, M., 2011b. Estimation of vertical deflections in concrete beams through digital close range photogrammetry. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII(Part 5/W12).

Fraser, C.S., 1997. Digital camera self-calibration. *ISPRS Journal of Photogrammetry & Remote Sensing*, 52(4): 149-159.

Fraser, C.S. and Edmundson, K.L., 2000. Design and implementation of a computational processing system for off-line digital close-range photogrammetry. *ISPRS Journal of Photogrammetry & Remote Sensing*, 55(2): 94-104.

Guidi, G., Remondino, F., Russo, M., Menna, F., Rizzi, A. and Ercoli, S., 2009. A Multi-Resolution Methodology for the 3D Modeling of Large and Complex Archeological Areas. *International Journal of Architectural Computing*, 7(1): 39-55.

Habib, A.F. and Morgan, M.F., 2003. Automatic calibration of low-cost digital cameras. *Journal of Optical Engineering*, 42(4): 948-955.

Habib, A.F. and Morgan, M.F., 2005. Stability analysis and geometric calibration of off-the-shelf digital cameras. *Photogrammetric Engineering & Remote Sensing*, 71(6): 733-741.

Habib, A.F., Pullivelli, A.M. and Morgan, M.F., 2005. Quantitative measures for the evaluation of camera stability. *Journal of Optical Engineering*, 44(3): 033605.

Jazayeri, I. and Fraser, C.S., 2010. Interest Operators for Feature-Based Matching in Close Range Photogrammetry. *The Photogrammetric Record*, 25(129): 24-41.

Lambers, K., Eisenbeiss, H., Sauerbier, M., Kupferschmidt, D., Gaisecker, T., Sotoodeh, S. and Hanusch, T., 2007. Combining photogrammetry and laser scanning for the recording and modelling of the Late Intermediate Period site of Pinchango Alto, Palpa, Peru. *Journal of Archaeological Science*, 34(10): 1702-1712.

Lichti, D.D., Gordon, S.J., Stewart, M.P., Franke, J. and Tsakiri, M., 2002. Comparison of Digital Photogrammetry and Laser Scanning. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIV(Part 5): 39-44.

Patias, P., 2002. Medical imaging challenges photogrammetry. *ISPRS Journal of Photogrammetry & Remote Sensing*, 56(5-6): 295-310.

Remondino, F., El-Hakim, S., Gruen, A. and Zhang, L., 2008. Turning Images into 3-D Models. *IEEE Signal Processing Magazine*, 25(4): 55-65.

Shao, J., Fraser, C.S. and Wrigley, T.V., 2001. Object Point Tracking in Photogrammetric Measurement of Human Movement. *The Photogrammetric Record*, 17(97): 103-117.

Yastikli, N., 2007. Documentation of cultural heritage using digital photogrammetry and laser scanning. *Journal of Cultural Heritage*, 8(4): 423-427.

Zhang, L. and Gruen, A., 2006. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS Journal of Photogrammetry & Remote Sensing*, 60(3): 195-211.