

Hybrid Photogrammetry Structure-from-Motion Systems for Scene Measurement and Analysis

Shawn Recker^{1,2}, Mikhail M. Shashkov¹, Mauricio Hess-Flores¹, Christiaan Gribble², Rob Baltrusch²,
Mark A. Butkiewicz², Kenneth I. Joy¹

¹Email: idav.ucdavis.edu/people
Institute of Data Analysis and Visualization
Department of Computer Science
University of California Davis
1 Shields Ave, Davis, CA 95616
530-752-1077

²Email: {shawn.recker, christiaan.gribble,
rob.baltrusch, mark.butkiewicz}@survice.com
SURVICE Engineering Company
Applied Technology Operation
1362 Brass Mill Road, Suite 5 Belcamp, MD 21017
410-297-2378

Given the recent advances in both photogrammetry and structure-from-motion, a pipeline that capitalizes on the strengths of both fields is now possible. This paper presents a hybrid system that uses photogrammetric information to improve the accuracy of structure-from-motion, which in turn provides a more dense reconstruction. The procedure maintains the required metrological accuracy and permits measurements between points with no corresponding targets. The paper provides an analysis of the effects of various camera parameters to determine optimal scene configurations. Results generated by the hybrid system for real and synthetic data demonstrate that both more accurate and more dense reconstructions are obtained than with structure-from-motion alone.

Word count: 5427, Image count: 30

1 Introduction

The reconstruction of scenes from multiple images or video streams has become an essential component in many modern applications. Many fields, e.g., robotics, surveillance, and virtual reality, employ systems to reconstruct a three-dimensional (3D) representation of an environment in order to accomplish specific goals. There are a number of systems capable of reconstructing scenes from thousands of images^{1,2,3}. Increasing the accuracy of these systems is an area of ongoing research.

In metrology, photogrammetry is often used to obtain measurements from a 3D representation of an object or scene derived from images³. Typically, targets (individual markers) are placed on the object to identify points of interest. These targets are usually retro-reflective and can be coded to easily distinguish one from another. Non-coded targets can be employed as well, and are identified in relation to a coded target. The choice and placement of targets depend on the desired measurement. Photographs of the object are taken from various camera positions while ensuring that the same target appears in at least in two images. The center of each target is found in the images where it appears, and triangulation is used to determine the 3D coordinate of each target. The resulting point cloud can later be processed to obtain a 3D polygonal mesh or other representation³.

In computer vision, a common technique used to obtain a 3D model of an object is structure-from-motion (SfM)⁴. While SfM has developed into a subfield of computer vision, this field began in photogrammetry. A number of fundamental mathematics used in current SfM techniques, e.g. bundle adjustment⁴, were adapted from the photogrammetry literature. As such, many algorithms span the two fields. For example, SfM involves feature tracking (target tracking), camera pose estimation (camera calibration), triangulation, and bundle adjustment (block adjustment)⁴.

Despite the similarities between these fields, their goals now diverge. Generally, photogrammetry focuses on obtaining surface points on an object that have very low covariances with the goal of measuring specific targets. SfM typically produces denser reconstructions so that more knowledge of the overall scene can be obtained for specialized purposes. Though visually accurate, results obtained with SfM are typically not sufficient for metrology applications.

Given advances in photogrammetry and SfM, a pipeline harnessing the symbiotic relationship between these fields is now possible. The proposed system uses photogrammetric information to enhance the accuracy of SfM: SfM provides a dense reconstruction, while photogrammetry is used to correct camera parameters or scene points that have high uncertainty. This additional SfM information allows measurements between points that do not have corresponding targets. This procedure is less expensive, maintains the required metrological accuracy, and provides a more complete reconstruction than is possible with either method alone. In addition, this paper provides an analysis of the effects of various camera parameters, such as focal length and depth, on reconstruction to determine the correct setup for optimal results.

The remainder of this work is organized as follows: An overview of related technologies is provided in Section 2. The components of the system are presented in Section 3. Experimental results are discussed in Section 4. The paper offers some conclusions and outlines areas of future work in Section 5.

2 Background

Multi-view reconstruction attempts to obtain a 3D representation of an underlying scene from a collection of images taken from various camera viewpoints. Rays from each camera center through its image plane define the 3D intersection point for corresponding pixels between images. The output of scene reconstruction is characteristically a point cloud, though post-processing can be applied to obtain a polygonal mesh or other representations. A set of corresponding pixels, known as a feature track, can be computed using sparse or dense tracking algorithms. Feature tracking is the most important step in scene reconstruction, as errors in this stage affect all subsequent stages.

There exist many algorithms for multi-view reconstruction, but the following sequential stages are common to most systems. Typically, feature matches (between consecutive pairwise views) and tracks (concatenating across multiple views) are generated. Such tracking can be sparse^{6,7} or dense⁸ and consists of computing and linking the pixel coordinates in all images for each scene point, whenever it is visible. Frame decimation⁹ is often applied at this point, particularly for sequential image sets, to remove images with very small or very large baselines. A baseline is the relative separation of images in world space. Small baselines lead to bad numerical conditioning in pose and structure estimation, whereas large baselines introduce problems in feature tracking. Next, camera intrinsics are estimated through a process called self-calibration⁴. In most cases, much of this information is already known. Also, epipolar geometry can be estimated from pairs or triplets of views. Epipolar geometry encapsulates the intrinsic projective geometry between groups of views and is used in the process of determining camera pose (position and orientation). Only relative positions and orientations can be obtained between (or among) views. Once camera parameters are estimated, computation of scene structure is achieved through triangulation methods, such as linear triangulation⁴. In this method, the 3D position of a scene point, given a set of cameras and pixel feature track positions corresponding to the point, is computed as the best-fit intersection position in space for the set of rays from each camera center and through the feature track positions. Finally, because errors in all of the above steps influence the accuracy of the computed structure, bundle adjustment is performed to optimize some, or even all, of the camera and structure parameters^{10,11}. Bundle adjustment usually employs the Levenburg-Marquardt algorithm^{10,11} to minimize reprojection error of all computed structure, though the algorithm may converge to a local minimum and not achieve the globally optimal result.

Many scene reconstruction algorithms are based on some combination of the steps outlined above. Current examples include: Akbarzadeh et al.¹², who introduce a method for dense reconstruction of urban areas from a video stream, and Snavely et al.¹, who present a system for interactively browsing and exploring large unstructured collections of photographs of a scene. The latter system uses an image-based modeling front end that automatically computes the viewpoint of each photograph, as well as a sparse 3D model of the scene. The most recent system incorporating many modern reconstruction algorithms is VisualSfM².

The reconstruction procedure in photogrammetry is very similar, but has several key differences. The main goal is accurate measurement, so the setup conditions are often known and more important than in SfM. Camera intrinsics are known and baselines are carefully designed. Therefore, self-calibration and frame decimation are not relevant. Furthermore, there is more consideration to covariance analysis for input parameters and its propagation to output scene points³.

There is also recent work focusing on the effect of various parameters in the reconstruction process. Beder and Steffen¹³ introduce an uncertainty metric for a single scene point computed using linear triangulation. They prove that the optimal answer for an N -view triangulation lies within this 3D uncertainty ellipse. Knoblauch et al.¹⁴ introduce a metric for determining the source of error for a given feature match. This method does not rely on any *a priori* scene knowledge. More recently, Recker et al.¹⁵ introduce a scalar field visualization system based on an angular error metric for understanding the sensitivity of a 3D reconstructed point to various parameters in multi-view reconstruction. The present work investigates the advantages possible with a hybrid photogrammetry/SfM system, and explores the impact of various parameters on reconstruction quality.

3 Methodology

The proposed system utilizes photogrammetry to enhance the accuracy of reconstructions obtained with SfM, and capitalizes on SfM to provide a dense reconstruction, thereby improving the ability to analyze scenes' underlying photogrammetric results. The hybrid system diagram is presented in figure 1.

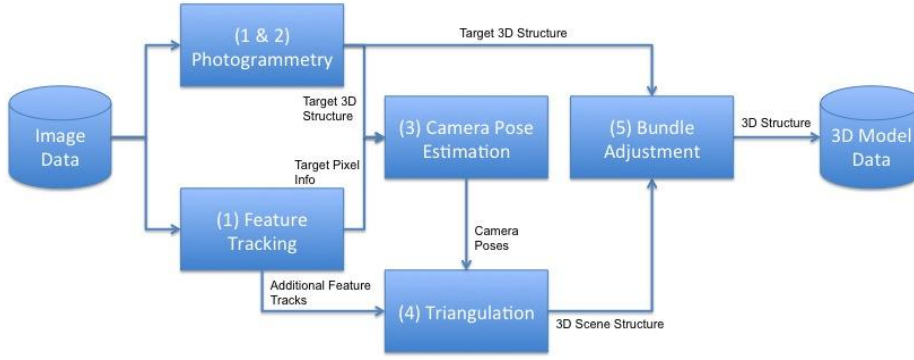


Figure 1: *Hybrid photogrammetry/SfM system*. (1) The process begins by identifying targets in the input image sequences. In addition, images are processed using an automatic feature tracking algorithm. (2) Camera calibration and triangulation of targets is performed in the photogrammetry stage. (3) In the event that camera data cannot be extracted from the photogrammetry system, camera pose estimation is used to obtain position and orientation. (4) Camera parameters are passed to the scene triangulation stage and are used to triangulate the remaining scene points obtained by feature tracking. (5) Bundle adjustment is performed to optimize the resulting structure, and the point cloud is stored.

First, images of a targeted asset are provided to a photogrammetry system and 3D target positions are computed. Images are simultaneously processed for additional feature tracks using SIFT⁶. Some photogrammetry systems also provide camera intrinsics and pose. If this information is not available, camera pose for all images is estimated using 3D target positions and 2D projections by solving the perspective N -point problem (PnP), using the Efficient Perspective N -Point (EPnP) algorithm¹⁶. Finally, any remaining feature tracks are triangulated using statistical error-based angular triangulation⁵, and bundle adjustment¹⁷ is performed to optimize the structure and computed camera pose. Finally, the point cloud is stored and can be post-processed to obtain another representation, if necessary.

4 Results

The proposed system is tested extensively for accuracy and general behavior using both real and synthetic datasets. The implementation is written in C++, and results are obtained on a MacBook Pro with an 2.66 GHz Intel Core i7 processor and 4 GB of RAM, running Mac OS X Mavericks 10.9.1.

4.1 Synthetic Tests

A variety of synthetic scenes are used to verify elements of the hybrid technique. Computation of 3D structure is accomplished using linear triangulation, the accuracy of which depends on the previously computed camera projection matrix and feature track position. The following tests examine the impact of these parameters on the resulting 3D structure.

Three error metrics are recorded for each experiment: reprojection error, rotational error, and positional error. The formulas are shown in equations 1, 2, and 3, respectively:

$$f_{reprojectionerror}(X, P) = \|PX - x\| \quad (1)$$

$$f_{rotationalerror}(R_i, R_j) = \frac{\|R_i - R_j\|}{\|R_j\|} \quad (2)$$

$$f_{positionalerror}(X_i, X_j) = \|X_i - X_j\| \quad (3)$$

Here, X is a 3D scene point, P is a camera projection matrix, and x is a 2D feature track location corresponding to X . R_i and R_j are quaternions that encapsulate a 3D rotation. X_i and X_j are also 3D positions.

Automated feature tracking techniques often do not exactly match corresponding points between an image set. The slight mismatch introduces error into the final reconstruction when feature tracks are assumed to be correct. In computer vision, feature tracking inaccuracy is defined as image noise and is modeled according to a Gaussian distribution. These inaccuracies are tested in the following experiments. In addition, numerical conditioning can affect the 3D computed structure and the final reconstruction. Therefore, the effects of 3D positional noise on camera pose estimation are also tested.

4.1.1 Feature Tracking Noise and Camera Depth

The first experiment in this set examines the effect of camera depth on accuracy of 3D reconstruction. Camera depth is defined as the distance from the camera center to the scene points being viewed. A ground truth camera was positioned at increasing distances, in units of two along the Z-axis, from a $1 \times 1 \times 1$ box with 100 ground truth 3D positions on its surface. Ground truth positions are projected into each image plane. At each distance, image noise was introduced using a zero mean Gaussian distribution with increasing standard deviation from zero to five pixels. The original distance from the camera center and each ground truth point is computed. Then, using the noisy 2D projection, a ray with unit direction from the camera center is computed. A 3D point along that ray is computed using the original distance. The positional error between the original point and new position is computed according to equation 3 for all 100 scene points. Results are averaged and standard deviation is computed, as shown in figure 2.

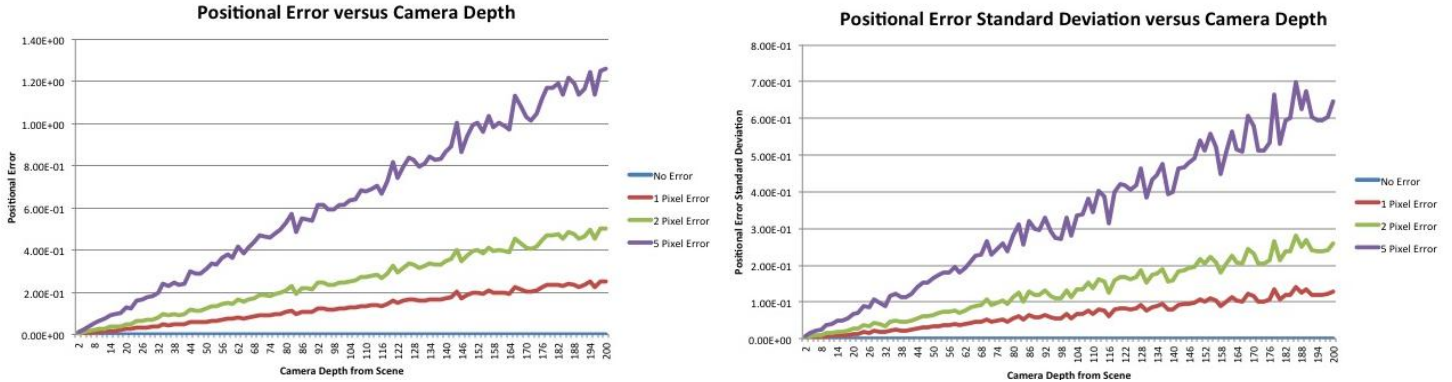


Figure 2: *Impact of camera depth on resulting 3D structure in the presence of image noise.* A ground truth camera was positioned at increasing distances from a $1 \times 1 \times 1$ box with 100 ground truth 3D positions. At each distance, image noise is introduced using a zero mean Gaussian distribution with increasing standard deviation from zero to five pixels. The original distance from the camera center and each ground truth point is computed. Using the noisy 2D projection, a ray with unit direction from the camera center is computed. A 3D point along that ray is computed using the original distance. The positional error between the two points is computed. The results are averaged (left) and standard deviation is computed (right). It is easily observed that, as the camera moves away from the scene, inaccuracies in feature tracking manifest in larger distances from the ground truth point.

The effects of depth are easily seen in figure 2. As the camera moves away from the scene, displacing a feature track location on the image plane, even a small amount, manifests in computed 3D positions that are large distances from the original data. These results demonstrate two important principles to consider when reconstructing objects from images. First, objects should be reasonably close to the camera, which allows the tracking to be slightly inaccurate without a significant impact on computed 3D positions. Second, there is an inverse relationship between an object’s distance from the camera and the accuracy required of the feature tracking mechanism: if the object is far from the cameras, feature tracking must be as accurate as possible to obtain reasonable reconstructions. Recognizing—and, to the extent possible, mitigating—the impact of this relationship will help to compute more accurate 3D positions.

4.1.2 Camera Pose Estimation with Positional Noise and Camera Depth

The second experiment examines the effect of camera depth on pose estimation in the presence of positional noise using the same setup as above; however, ground truth 2D projections are fixed. 3D positional noise is introduced to the box structure, again sampling from a Gaussian distribution with zero mean, and standard deviation increasing from zero to one mm. The experiment ensures that all 3D points are positioned in front of the cameras. For each depth, camera pose is solved and reprojection error, rotational error, and positional error are recorded. For rotational and positional error, the ground truth camera is used as reference. The results of this test are shown in figure 3.

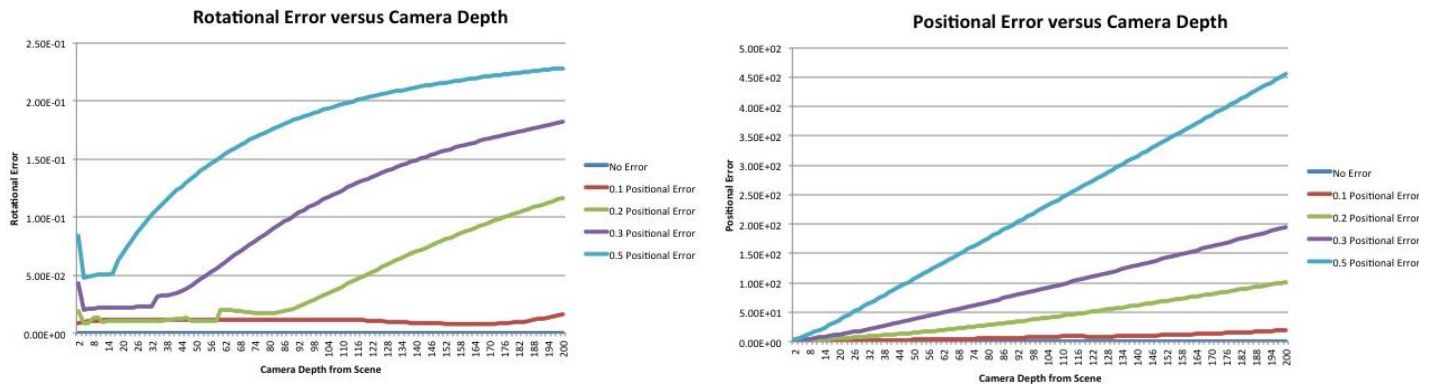


Figure 3: *Impact of camera depth on pose estimation in the presence of positional noise.* The experiment setup is the same as for results depicted in figure 2. Here, 3D positions are varied according to a Gaussian distribution with zero mean and increasing standard deviation from zero to one mm in steps of 0.1 units. Rotational error (left) and positional error (right) are recorded with respect to the original ground truth camera data. It is easily observed that, as the distance between the scene and the camera increases, accuracy of the computed pose decreases.

The data in figure 3 demonstrates that, as the camera moves away from the scene and pose is estimated with slightly noisy 3D positions, parameter estimates are less accurate compared to the actual camera parameters. From rotational error (figure 3, left), for certain distances away from the scene, error stabilizes for each noise level tested, suggesting there exists an ideal range of distances between camera and target. In this scenario, the object was $1 \times 1 \times 1$ mm in dimension, and the ideal distance is between four and 18 mm for higher positional noise. In lower noise cases, there is no discernible pattern for ideal camera distance.

4.2 3D Model Tests

Two synthetic datasets, *bradley* and *maxxpro*, are used to further analyze camera pose estimation. The objects are rendered using known camera parameters, and ground truth feature tracks are generated from the 3D object data. Views of these datasets are presented in figure 4.



Figure 4: *Synthetic datasets used to analyze camera pose estimation.* These 3D models correspond to real assets, but can be rendered using known parameters, offering more control over the corresponding experimental results. Ground truth feature tracks are generated from the 3D object data. Two images from the *bradley* sequence are shown at left, while two images from the *maxxpro* sequence are shown at right. Both datasets are provided by SURVICE Engineering Company.

4.2.1 Camera Pose Estimation in the Presence of Image Noise

The first experiment in this set determines accuracy of computed camera parameters when introducing image noise. Image noise is sampled from a Gaussian distribution with zero mean and standard deviation increasing from zero to ten pixels. Camera parameters are estimated and reprojection error, rotational error, and positional error are recorded for each camera. The *bradley* sequence contains 1234 cameras and *maxxpro* contains 278 cameras. Error metrics are averaged and standard deviation is computed. Additional trials that vary the number of ground truth 3D points from five to 1000 are also conducted. The results of this experiment are presented in figure 5.

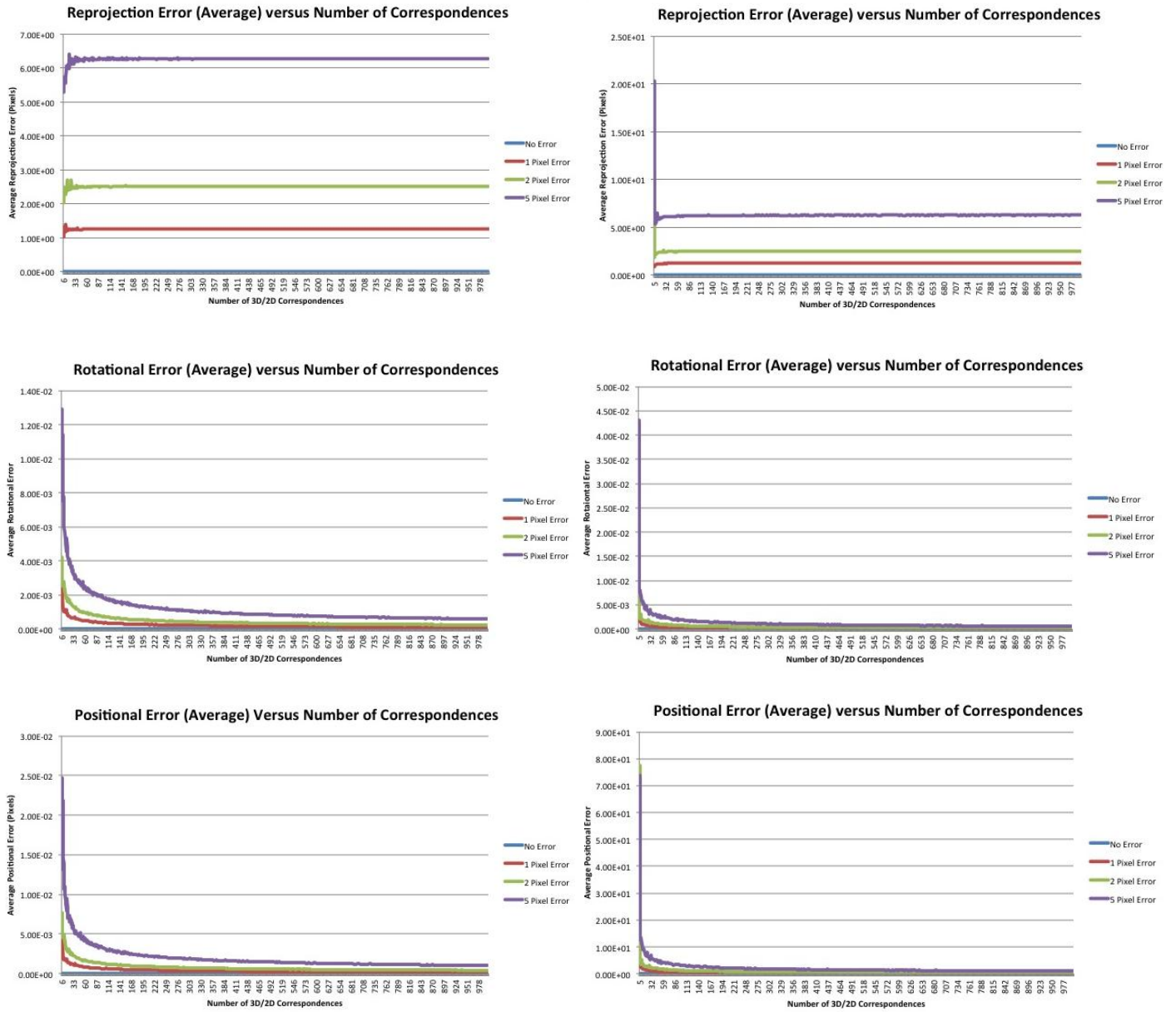


Figure 5: Impact of the number of 3D/2D correspondences on camera pose in the presence of image noise. Results from the *bradley* (left) and *maxxpro* (right) sequences are shown. Ground truth models are projected into each image and image noise is introduced using a Gaussian distribution with zero mean and standard deviation increasing from zero to ten pixels. Camera parameters are computed and error metrics recorded. For each sequence, average reprojection error is shown in the top row, average rotational error in the middle row, and average positional error in the bottom row.

The data in figure 5 demonstrates that, for low numbers of 3D/2D correspondences, there is larger uncertainty in computed camera pose. However, error tends to stabilize for all metrics across noise levels as the number of correspondences increases. Based upon these results, having approximately 50 targets on the object is sufficient to obtain accurate camera pose estimates in the presence of higher noise. For lower noise levels, approximately 20 targets produce accurate results.

4.2.2 Camera Pose Estimation in the Presence of 3D Point Error

The second experiment determines accuracy of computed camera parameters when introducing 3D positional noise. Positional noise is sampled from a Gaussian distribution with zero mean and standard deviation increasing from zero to five world space units. Camera parameters are estimated and reprojection error, rotational error, and positional error are recorded for each camera. Error metrics are averaged and standard deviation is computed. Additional trials that vary the number of ground truth 3D points from five to 1000 are also conducted. The results of this experiment are presented in figure 6.

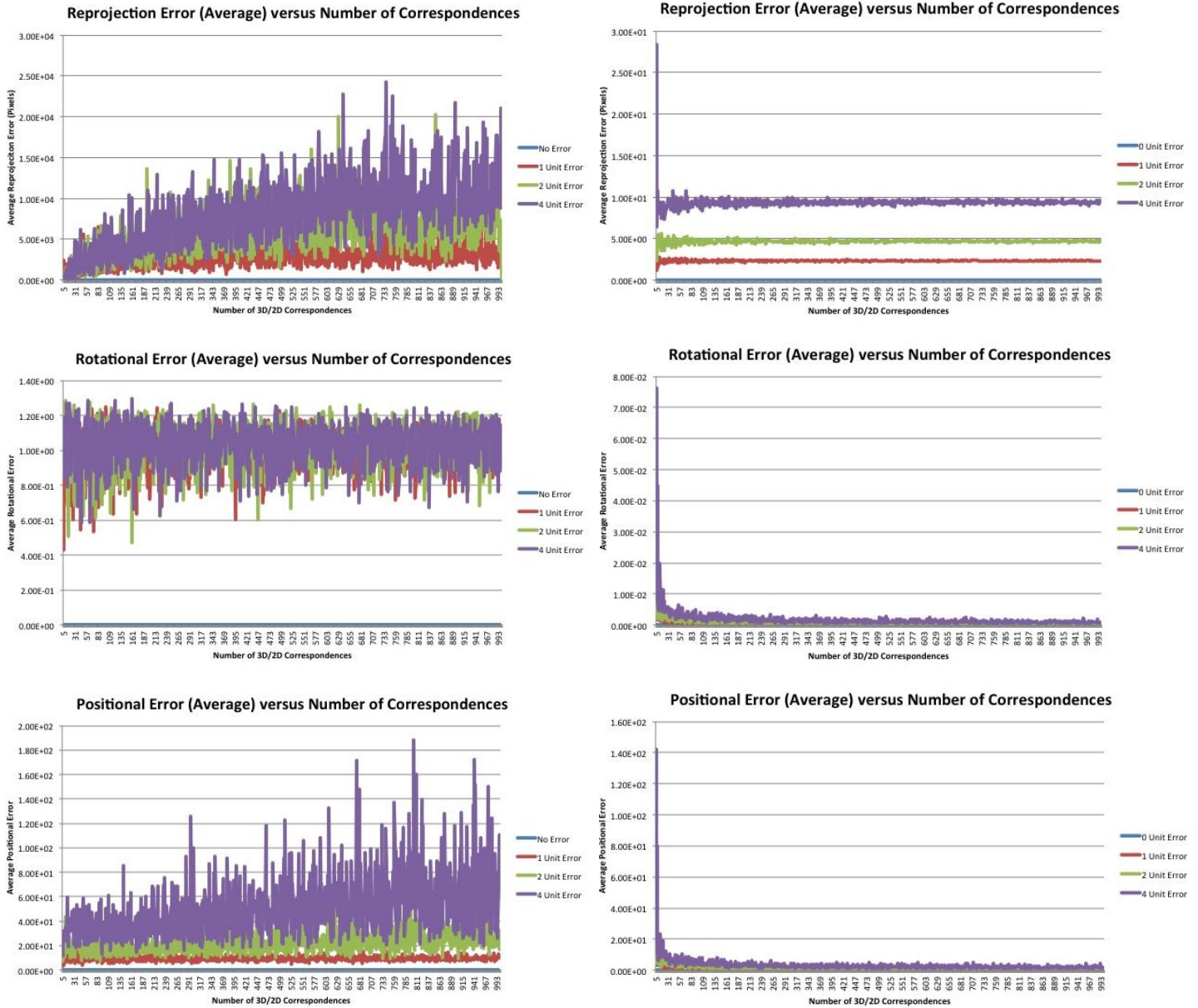


Figure 6: Impact of the number of 3D/2D correspondences on camera pose in the presence of 3D position noise. Results from the *bradley* (top) and *maxxpro* (bottom) sequences are shown. Positional noise, sampled from a Gaussian distribution with zero mean and standard deviation increasing from zero to five world space units, is introduced to the computed 3D structure. Camera pose is computed and error metrics recorded. For each sequence, average reprojection error is shown in the left column, average rotational error in the middle column, and average positional error in the right column.

The results in figure 6 show that the overall sensitivity of pose estimation to noise in 3D structure is much higher compared to the impact of 2D image noise. For large movement in the 3D point, camera pose estimates becomes less accurate. When compared to the original camera data that was used to generate the 2D projections, reprojection error is much improved. However, rotational and positional error from the original camera indicate a large change. The effect of displacing the 3D structure is higher for the *bradley* dataset compared to *maxxpro*; however, the relative scale of these models is different: four world-space units in *bradley* is a much larger relative movement than for *maxxpro*. This difference leads to greater reconstruction error. Based on these results, camera pose estimation is particularly sensitive 3D structure computation, so higher accuracy triangulation schemes should be employed.

4.2.3 Camera Pose with Different Spatial Distributions of Points

The final experiment determines accuracy of computed camera pose when using different spatial distributions of 50 points that are representative of photogrammetry targets placed on an object. The first distribution clusters points on the object, while the second distributes points randomly over the object. Camera parameters are estimated and the error metrics used in the previous tests are computed. The results of this experiment are presented in figure 7.



Figure 7: *Impact of the spatial distribution of 3D scene points on pose estimation.* For this experiment, 50 ground truth points are used to solve for camera pose. These points are arranged in two spatial distributions: a small cluster on the object (left) and a random distribution across the object (right). Accuracy of the camera pose is the same for each point distribution. The results for the *bradley* dataset are shown here: camera positions are displayed in green, the points selected for pose estimation are shown in red, and remaining points from the model are shown in dark grey.

The data in figure 7 indicate that the spatial distribution of 3D scene points has no significant impact on pose estimation. Given accurate tracking and 3D position data, EPnP solves for camera pose accurately, independent of the point distribution. Photogrammetry targets, therefore, need not be placed uniformly on an object.

4.3 Real Test

To verify the hybrid technique, a real dataset is run on the system and results are compared to those generated by an SfM-only system. The dataset contains 14 images of an armoire with 37 photogrammetry targets randomly distributed throughout the scene. The AICON DPA Pro system solves for the targets' 3D positions. All images have the same camera intrinsics, which are provided by the AICON system. Camera intrinsics and target positions are input to the SURVICE HawkEye SfM system. HawkEye computes feature tracks using SIFT⁶ and triangulates the results.

Results of the hybrid technique are compared to those generated by VisualSFM². Note that these systems have different submodules, which could manifest in the difference in reprojection error, but VisualSFM (freeware) is regarded as the state-of-the-art reconstruction system and a comparison must be made. The hybrid system generates 6659 total computed scene points with an average reprojection error of 0.4376 pixels, while VisualSFM generates 3073 computed points with an average reprojection error of 6.1426 pixels. These results indicate that 3D target locations, as provided by the photogrammetry system, allow SfM to provide more accurate results when compared to standard SfM alone.

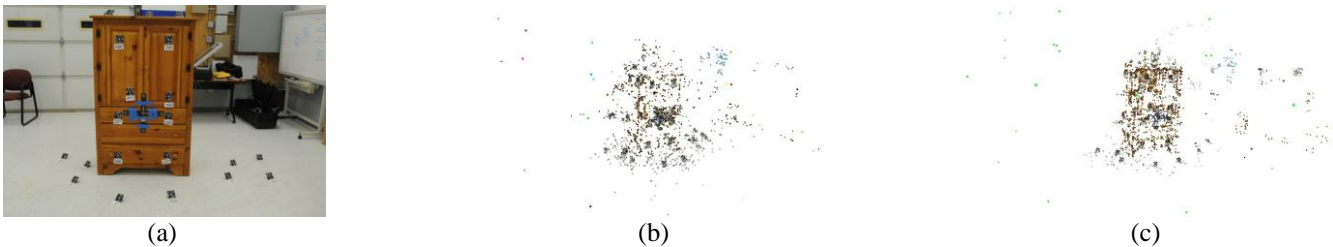


Figure 8: *Comparing hybrid photogrammetry/SfM results to those produced by SfM alone.* This figure contains the results of the *armoire* dataset, which contains 14 images of 37 photogrammetry targets distributed randomly across the armoire. The AICON DPA Pro system solves for target positions, which are input to HawkEye SfM. HawkEye generates the final reconstruction. An image from the *armoire* sequence is in panel (a). Panel (b) shows the point cloud generated by the hybrid system (6659 points and average reprojection error of 0.4376 pixels), while panel (c) shows the results generated by VisualSFM (3073 points and average reprojection error of 6.1426 pixels).

4.3.1 Additional Reconstruction Parameters

The previous experiments explore the impact of various reconstruction parameters on the accuracy of the results. However, other factors also affect the final results—lighting, for example. For automated feature tracking techniques, changes in lighting can lead to mismatches or cause a feature point in one image to be missed in another. Ideally, lighting

should approximate constant global illumination from all angles (i.e., uniform hemispheric lighting). If uniform lighting is not possible, then an overhead light source can work, but care must be taken to avoid shadows and glare. The affect of lighting can be seen in figure 9, which shows one image from a sequence of eight depicting a desk under normal lighting conditions and then again under “darker” conditions. The sequences are run through VisualSfM² followed by dense reconstruction with PMVS¹⁸. VisualSfM was unable to acquire good feature tracks from the darker image sequence due to a lack of contrast in the images. Poor feature tracks lead to inaccurate camera parameters, which in turn lead to a sparse and imprecise point cloud. In fact, the system computed two different models for the darker sequence.

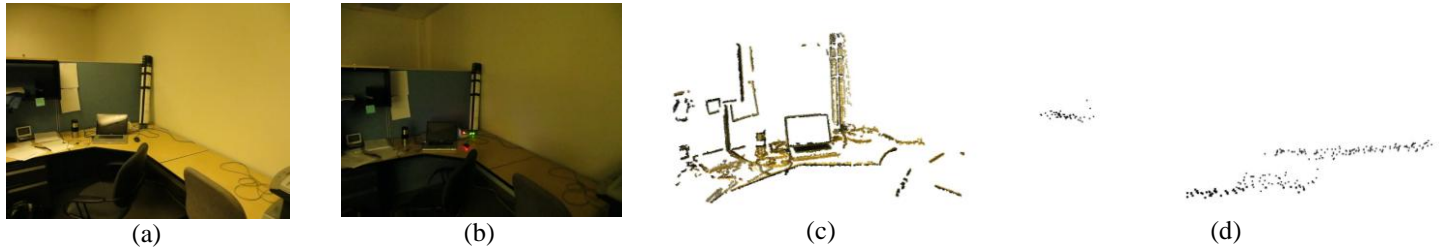


Figure 9: *Comparing the effects of lighting on reconstruction.* Two image sequences of a desk under normal and “darker” lighting conditions are input to VisualSfM^{2,18} and reconstruction is attempted. Panels (a) and (b) show reference images from each sequence, while panels (c) and (d) show the resulting reconstructions. The sparse point cloud from the darker sequence indicates that feature tracking is unable to make good matches, which affects all subsequent stages of reconstruction.

Focal length and image resolution also affect accuracy of reconstruction. Unfortunately, these factors are not independent of one another. Focal length specifies the distance between the image plane and the camera center of projection. Increasing focal length decreases the area encapsulated by a single pixel. Large focal lengths result in a narrow field-of-view, so less of the scene is captured by a single image. Recker et al.¹⁵ show that focal length has a limited effect on the accuracy of a reconstructed 3D point, mainly affecting the scale of a reconstruction. To capture more of the scene while minimizing the area a single pixel covers, image resolution can be increased. However, increasing focal length too much may nullify the benefit of increased image resolution. On a practical level, a pixel covering too much area can lead to inaccuracies in target location. If the pixel covers a significant portion of the target, then there will only be a few pixels in which the target is present: limited target coverage leads to poor target tracking and, ultimately, to low quality reconstruction³.

5 Conclusion

The experiments in Section 4 analyze the effect that certain parameters have on accuracy of 3D reconstruction. Distance between the camera and scene (camera depth) is an important factor to consider. It has been shown that the closer the camera is to the scene, the lower the impact of a tracking error on final reconstruction. When the camera is far from the scene, the effects of tracking errors are amplified. However, it is also shown that having a camera that is too close to the scene can be problematic for rotation estimation and that there is a range of distance values resulting in accurate camera rotation estimations. In the presence of accurate target tracking, approximately 20 targets are sufficient to accurately compute camera pose. In the presence of higher noise, approximately 50 targets are required. Finally, changes in lighting are shown to present difficulties in feature tracking, which results in poor quality reconstructions.

In summary, this paper presents a hybrid 3D reconstruction system that combines photogrammetry with SfM techniques. The system uses photogrammetric information to enhance accuracy of SfM results. SfM provides a dense reconstruction, while photogrammetry is used to correct camera parameters or scene points that have high uncertainty. This procedure permits measurements between points that do not have corresponding targets, maintains the required metrological accuracy, and provides a more complete reconstruction than with either method alone. Results generated by the hybrid system for real and synthetic data demonstrate that both more accurate and more dense reconstructions are obtained than with SfM alone.

The development of hybrid systems for 3D reconstruction helps improve both photogrammetry and structure-from-motion. The continued exploration of combining techniques in these fields may lead to both improvement of current algorithms and the development of new ones. Investigation of photogrammetry-assisted volume-based reconstruction is an interesting and important topic for future applications. Moreover, additional visualization techniques are needed to show reconstruction accuracy, and to highlight key locations at which to place photogrammetry targets so that overall reconstruction accuracy is improved.

Acknowledgment

This work was supported in part by Lawrence Livermore National Laboratory, the National Nuclear Security Agency through Contract No. DE-FG52-09NA29355, the US Marine Corps SBIR program, and SURVICE Engineering's Internal R&D program. The authors thank their colleagues in the Institute for Data Analysis and Visualization (IDAV) at UC Davis and in the Applied Technology Operation at SURVICE Engineering for their support.

End Notes

- ¹ N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3D," in *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*. plus 0.5em minus 0.4emNew York, NY, USA: ACM, 2006, pp. 835–846.
- ² Changchang Wu, "VisualSfM: A visual structure from motion system," 2011. [Online]. Available: <http://homes.cs.washington.edu/ccwu/vsfm/> =0pt
- ³ T. Dodson, R. Ellis, C. Priniski, S. Raftopoulos, D. Stevens, and M. Viola, "Advantages of high tolerance measurements in fusion environments applying photogrammetry," in *Fusion Engineering, 2009. SOFE 2009. 23rd IEEE/NPSS Symposium on*, June 2009, pp. 1–4.
- ⁴ R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. plus 0.5em minus 0.4emCambridge University Press, 2004.
- ⁵ S. Recker, M. Hess-Flores, and K. I. Joy, "Statistical angular error-based triangulation for efficient and accurate multi-view scene reconstruction," in *Workshop on the Applications of Computer Vision (WACV)*, 2013.
- ⁶ D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal On Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- ⁷ H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision – ECCV 2006*, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds. plus 0.5em minus 0.4emSpringer Berlin / Heidelberg, 2006, vol. 3951, pp. 404–417, 10.1007/11744023.32. [Online]. Available: <http://dx.doi.org/10.1007/11744023.32>
- ⁸ E. Tola, V. Lepetit, and P. Fua, "Daisy: an efficient dense descriptor applied to wide baseline stereo," in *PAMI*, vol. 32, no. 5, May 2010, pp. 815–830.
- ⁹ D. Nistér, "Frame decimation for structure and motion," in *SMILE '00: Revised Papers from Second European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*. plus 0.5em minus 0.4emLondon, UK: Springer-Verlag, 2001, pp. 17–34.
- ¹⁰ M. Lourakis and A. Argyros, "The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm," Institute of Computer Science - FORTH, Heraklion, Crete, Greece, Tech. Rep. 340, August 2000.
- ¹¹ B. Triggs, P. McLauchlan, R. I. Hartley, and A. Fitzgibbon, "Bundle Adjustment - A Modern Synthesis," in *ICCV '99: Proceedings of the International Workshop on Vision Algorithms*. plus 0.5em minus 0.4emLondon, UK: Springer-Verlag, 2000, pp. 298–372.
- ¹² A. Akbarzadeh, J.-M. Frahm, P. Mordohai, B. Clipp, C. Engels, D. Gallup, P. Merrell, M. Phelps, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch, H. Towles, D. Nister, and M. Pollefeys, "Towards urban 3d reconstruction from video," in *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, June 2006, pp. 1–8.
- ¹³ C. Beder and R. Steffen, "Determining an Initial Image Pair for Fixing the Scale of a 3D Reconstruction from an Image Sequence." in *DAGM-Symposium '06*, 2006, pp. 657–666.
- ¹⁴ D. Knoblauch, M. Hess-Flores, M. Duchaineau, and F. Kuester, "Factorization of Correspondence and Camera Error for Unconstrained Dense Correspondence Applications," *5th International Symposium on Visual Computing, Las Vegas, Nevada*, pp. 720–729, 2009.
- ¹⁵ S. Recker, M. Hess-Flores, M. A. Duchaineau, and K. I. Joy, "Visualization of scene structure uncertainty in a multi-view reconstruction pipeline," in *Vision, Modeling and Visualization Workshop*, 2012, pp. 183–190.
- ¹⁶ V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnnp: An accurate o(n) solution to the pnp problem," *International Journal Computer Vision*, vol. 81, no. 2, 2009.
- ¹⁷ M. I. A. Lourakis and A. A. Argyros, "The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm," Institute of Computer Science – FORTH, Heraklion, Crete, Greece, Tech. Rep. 340, Aug. 2000.
- ¹⁸ Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–8.