The Photogrammetric Journal of Finland, Vol. 24, No. 1, 2014 Received 02.05.2014, Accepted 10.11.2014
doi:10.17690/014241.1

AUTOMATED IMAGE-BASED RECONSTRUCTION OF BUILDING INTERIORS – A CASE STUDY

Ville V. Lehtola[1], Matti Kurkela[1], Hannu Hyyppä[1,2]

[1]Aalto University, Research Institute of Measuring and Modeling for the Built Environment
[2]Metropolia University of Applied Sciences

ville.lehtola@aalto.fi, matti.kurkela@aalto.fi, hannu.hyyppa@aalto.fi

ABSTRACT

In 3D reconstruction of indoor environments, automated methods offer possibilities for e.g. brokering, planning and decoration businesses. The application potential of these automated methods is, however, tied to the accuracy of these methods. In this paper, we perform a technical case study analysis on a state-of-the-art image-based method. For accuracy, we find that 60-70% of points in the reconstructed 3D point cloud are within 5 cm error range. Image-based 3D reconstruction thus offers potential for those various indoor-related applications that are satisfied with this level of accuracy. We also discuss other factors affecting to the applicability and robustness of this method.

1. INTRODUCTION

Automated 3D reconstruction of built environment is bound to open a wide playfield for professional (e.g. Building Information Management, BIM) and consumer applications, and is therefore under intense research (Bosse et al., 2012; Flint et al., 2010; Furukawa et al., 2009b; Georgantas et al., 2012; Khoshelham and Elberink, 2012; Liu et al., 2010). Automation is called for to eliminate manual work phases, as most models of today can only be manually derived – either from the hand crafted 3D designs or by post-processing laser scanning data. However, automated methods are not best suited for all tasks, especially if the aim is a professional level of detail. In contrast, they complement the professional tools with fast, rough, and easy-to-use ways to deliver value ideal for applications such as real-estate brokering and decoration planning.

Considering application scalability and the fact that consumer cameras are already very common, automated techniques founding on image sequence analysis are especially important. Also, photogrammetric techniques have long been used to assist in realizing computer-aided design (CAD) and virtual models, so their professional value is readily acknowledged. In addition, the image sequence based (ISB) reconstruction scheme does not suffer from the indoor localization problem that has to be tackled with, when automated reconstruction is attempted with mobile laser scanners.

Applications founding on image sequences taken by consumers require that the automatic reconstruction is done from an unordered and uncalibrated set of images in a robust way. The robustness is especially important, if no strong a-priori knowledge is available about the cameras. As noted by Furukawa et al. (2009a), the reconstruction is challenging in two ways. First, indoor scenes are full of textureless walls and optical surfaces (e.g. windows). Second, indoor spaces contain narrow doorways, which require complicated visibility analysis. To this background, it is important to understand how accurate and widely applicable the automated ISB reconstruction really is.

The ISB method itself is divided into six steps. First, scale invariant features are extracted from each image using SIFT (Lowe, 2004). Second, structure from motion (SfM) is gained through incremental bundle adjustment optimizations (Snavely et al., 2006). Third, the scene is clustered with CMVS (Furukawa et al., 2010). Fourth, patch-based multi-view stereo (PMVS2) creates a denser point cloud (Furukawa and Ponce, 2010). Fifth, depth maps are calculated using Manhattan-world assumption (Furukawa et al., 2009a). Sixth, volumetric optimization is performed on the whole scene (Furukawa et al., 2009b). This paper focuses on the first four steps, as the reconstructed ISB model can only be as accurate as the SfM/PMVS point cloud is.

In this paper, we set out to measure the level of accuracy of the ISB method, in order to determine its potential for applications. As our main result, we evaluate the ISB point cloud accuracy using a ground truth obtained with a tripod-mounted FARO Focus 3D laser scanner. The accuracy of the scanner is about ±2 mm (Chow et al., 2012), which suits the purpose of evaluating this method with respect to BIM-related and consumer applications. We compare the ISB method accuracy with an indoor study using APERO software (Georgantas et al., 2012) that employs rigorous photogrammetric procedures at the cost of automation e.g. preventing the use of unordered sets of images on which we concentrate.

1.1 Related work

The automated modeling of indoor environments has been attempted in roughly three ways. First, from an image sequence that is taken with one or several cameras as explained previously. Second, from depth maps that are produced with flash LIDAR or depth cameras (RGB-D) and their late consumer versions, see e.g. (Khoshelham and Elberink, 2012). And third, laser scanners have been mounted on wheels (Xiao and Furukawa, 2012), on a spring (Bosse et al., 2012) and on a backpack (Liu et al., 2010).

Mobile laser scanners (MLS) are widely used outdoors, see e.g. (Kaartinen et al., 2012; Kukko et al., 2012) and references there-in, but inside they face the so-called *localization problem* when GPS satellite signal is not available to localize the range data. This reduces MLS's greatest strength – the measurement accuracy. The state-of-the-art has coped with the reduction with assumptions about the environment. For example, the pushcart approach presented by (Xiao and Furukawa, 2012) produces only rectangular rooms. The backpack (Liu et al., 2010) is limited to modeling hallways only. The spring-mounted scanner, Zebedee, is used assuming that the laser beams do not penetrate, which mostly is true in built environment; however, this and employing the reconstructed model for localization leads into limitations with respect to some (pathological) environments (Bosse et al., 2012). Considering application scalability that these methods offer, the special equipment is expensive, scarce, and requires special training compared to consumer level cameras. Due to this fact, there is room for ISB methods that are both automated and robust.

Consumer-level depth cameras, Kinect for example, offer the benefit of performing real-time reconstruction while avoiding the problem of textureless walls. However, as these devices are not designed for the reconstruction of the built environment, they have limitations in range, resolution, and optical design (Khoshelham and Elberink, 2012). It is also an open question, how common will these depth cameras be in the future. Flash LIDARs, which basically act as high accuracy depth cameras, are an emerging technology and may offer potential for indoor 3D reconstruction in the future.

The state-of-the-art in image sequence-based (ISB) modeling is arguably still the Manhattan-world stereo by Furukawa et al. (2009b), which we will examine in the following section. However, other potential approaches exist. Flint et al. (2010) have presented an elegant dynamic programming approach, which is deterministic and can be employed on single images. However, they use assumptions that each photo contains a flat floor and/or roof, henceforth effectively invalidating the reconstruction of spaces with stairs or other non-basic structure. Their approach fails with images that contain texture lines originated from furniture (e.g. paintings or signs), which heavily impacts on algorithm robustness as the algorithm itself has no knowledge of these failures. A probabilistic 3D segmentation approach by Hernandez et al. (2007) may also prove to be a worthwhile method in indoor reconstruction later on, if pose errors can be reduced as noted by Xiao and Furukawa (2012).

2. EQUIPMENT AND CAPTURED DATA

For this paper, two scenes were captured with image set data using uncalibrated cameras. First, an unordered set of 502 images from a small town in Italy was taken in May 2012 using Canon EOS 60D with 17 mm objective. The set is from a narrow street, resembling indoor conditions except without a roof but with a lot of texture. It is used to test the self-consistency of the point cloud over a distance of 200 meters. It is also used for the initial benchmarking of the ISB method. Second, unordered indoor sets of 390 images, and 417 images were taken in May 2013 with Canon EOS 60D with 17 mm objective and Nikon D800E with 24mm objective, respectively. The latter set was then extended with 99 images totaling up to 516 images to achieve a more complete reconstruction. All images were shot with free hands to simulate non-laboratory circumstances. The lowest JPG resolution was chosen from both cameras to achieve good signal to noise ratio, 2592x1728 pixels from Canon, and 3680x2456 from Nikon. We prefer to use the images as they come from the camera, as additional resizing algorithms would add their own fingerprint to the pipeline. Our Canon represents a consumer-level system camera of below 1500 €, whereas the Nikon is a professional tool costing over 5000 €. Photos were also taken using Nikon D200 with 20mm objective, but the opening angle occurred to be too narrow for the tightest spots in the office.

In order to obtain the reference data, we employed FARO Focus 3D S 120 laser scanner (for details, see http://www.faro.com/site/resources/share/944) with scan targets. Scans were conducted from 11 different spots. The scanner's about 0.8 mm accuracy towards a scan target (Chow et al., 2012) was compensated by using from 5 to 11 scan targets to align adjacent scans. A total of 24 scan targets were in use. The estimated point cloud geometric accuracy after registration was 2.2 mm. When taking the ±2 mm (scan point) accuracy into account, the total cumulative scan error is estimated to reside below 1 cm.

3. AUTOMATED ISB RECONSTRUCTION

The foundation of the ISB method robustness is the quality of the unordered image set. Since ISB method uses jpeg-compressed images as input, possible robustness factors include not only physical properties of e.g. the objective and the CMOS chip, but also the properties of the camera's jpeg converter, and the employed color and shader scales. Some of these factors have been studied, see e.g. (Aguilar et al., 1996; Harms et al., 2014) and references therein. In order to verify if photogrammetric and other camera properties affect the robustness, we employ image sets taken with different cameras. Finally, since the image set quality is also greatly affected by the lighting of the scene, we perform benchmarking both outdoors and indoors.

Ideally, automated image sequence-based reconstruction would not require any other information than what is in the unordered photoset. In practice, camera information is transferred using the image EXIF tags, and CMOS chip (physical) widths are fed to the algorithm as extra information. Since the information in EXIF tags only gives an estimate for the focal length f and lacks distortion correction terms, e.g. radial distortion coefficients k_1 and k_2 (Fryer and Brown, 1986; Hartley and Zisserman, 2004), for the camera lens, these constants have to be put inside the bundle adjustment problem as "fitting parameters". This is bound to cause problems in the bundle adjustment outcome. Also, even if automated, the ISB method is a post-processing method, meaning that if *the chosen scene* is not fully reconstructed from *the available set of images*, the user is required to go back and take more images.

In processing the images for reconstruction, we first employ the latest freely available open source structure from motion (SfM) software, called bundler (Snavely et al., 2010; Snavely et al., 2006). The SfM output is then improved with CMVS and PMVS software (Furukawa and Ponce, 2010), which are also freely available. All reconstructions are done on Intel 8-core 3.3 GHz Xeon machine with 16 GB RAM and 64-bit Fedora 18 Linux.

4. RESULTS

4.1 Initial Benchmarking

The ISB algorithm's performance was tested with a set of 502 images from a narrow street, resembling indoor conditions except without a roof, but with a lot more texture. Images were taken with Canon EOS 60D from a small town in Italy. As the studies with tourist photos, e.g. Snavely et al. (Snavely et al., 2006), we also receive a nice 3D point cloud reconstruction using the fitting scheme for camera parameters, see Fig. 1. In contrast to challenges faced indoors, the circumstances for taking the outdoor photos were excellent: diffuse sunlight, non-reflecting surfaces, abundance of buildings, and few dynamic elements. This was also the motivation for this benchmarking. The obtained point cloud is quite dense, containing of 6.3 million points.

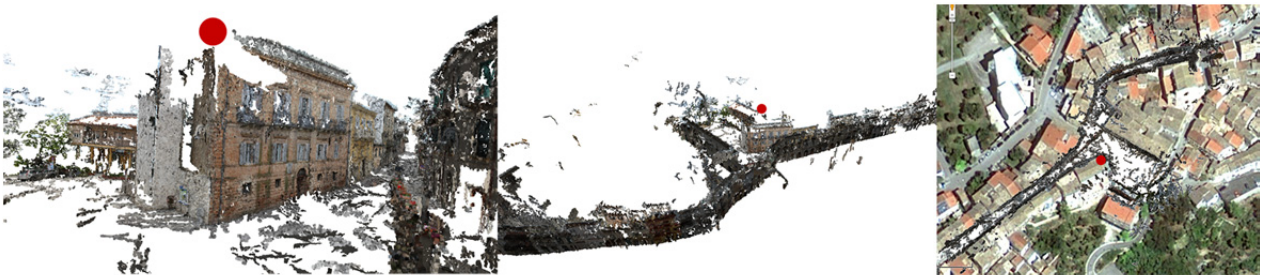

Figure 1. Close view on PMVS reconstructed point cloud (left). Unlike indoors, images contain a lot of texture. Overview of the whole point cloud extracted from 502 images taken on a narrow street in Italy (center). Visual fit of the point cloud to Google maps aerial image (right). The red dot indicates the location of the house shown in the leftmost image.

Here already, the use of CMVS clustering is a necessity to restrict the PMVS's otherwise exhaustive use of memory. With 8GB of memory, the maximum cluster size should be set somewhere between 100 and 150 images. Here, the image size is about 4.5 megapixels. PMVS reconstruction level is set to one, meaning that the image is processed with half width and half height (~1.1 megapixels).

With the default PMVS settings, outliers are quite abundant in Fig 1. To ensure the robustness of following reconstruction phases, a more conservative approach is required. There are at least two measures that can be taken to decrease the outlier abundance. First is to input more a-priori information to the pipeline by manually fixing the camera intrinsics. Second is to require a better photometric consistency in the reconstructed point cloud by reducing the amount of output information from the unreliable end.

In order to study the impact of fixed versus "fitted" camera intrinsics, the precision of the bundler estimated focal length is evaluated. This should enlighten us on how much error can be averted if the camera parameters are kept fixed. We take a set of images in a controlled environment with physically fixed camera parameters, and let bundler run bundle adjustment as if the camera parameters would be previously unknown (though EXIF tag information is still available). The distribution of bundler estimated focal lengths is plotted in Fig 2. The Gaussian-like distribution has long tails, implying that focal length estimates risk to fail in images where only few features are detected and especially in those images where most of these few features are clustered close to each other, see Fig 2. Intuitively, the Levenberg-Marquardt optimization algorithm finds these false values for *f* from local minima of the fitting scheme. Here, the bundler default weight 10^{-4} is used to constrain the focal length estimate close to its initial value from EXIF tag.

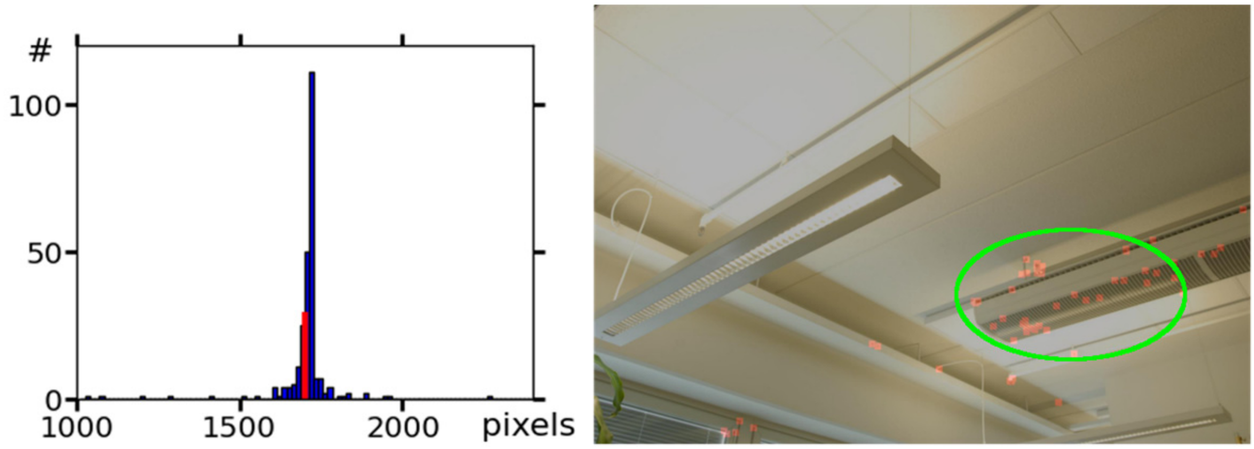

Figure 2. Left: Distribution of bundler estimated focal lengths in pixels when the physical focal length is kept fixed. Distribution is Gaussian-like, with a mean close to the physical focal length marked with a vertical red line. Farthest outliers down to 600 and up to $2*10^5$ pixels are not shown. Right: Sample image of where focal length estimate fails (f=992). Feature points are marked with red squares. Multiple points are clustered close to air conditioner, circled with green.

Now, if each single parameter has a Gaussian error behavior, it becomes important to test the overall self-consistency of the fitting scheme. For this purpose, we visually fit the point cloud to an aerial image, provided by online Google maps, see Fig 1. After choosing an appropriate scale and rotation, the fit is adequate. The point cloud appears to be self-consistently aligned around corners even at distances over 200 metres. This implies that the amount of information in feature points exceeds the inherent uncertainty caused by the fitting errors. Hence, manual camera calibration is not required, if the overall consistency of the point cloud is maintained.

To increase robustness in the point cloud obtained from PMVS output, we set the initial photometric consistency threshold from the default value of 0.7 to 0.95, as in (Furukawa et al., 2009a). PMVS relaxes the value to 0.65, when we increase the amount of patch extension iterations from the default three to seven. This effectively removes the outliers, but also reduces the point cloud density from 6.3 million to 4.3 million points and therefore results in a visually less satisfactory outcome (not shown).

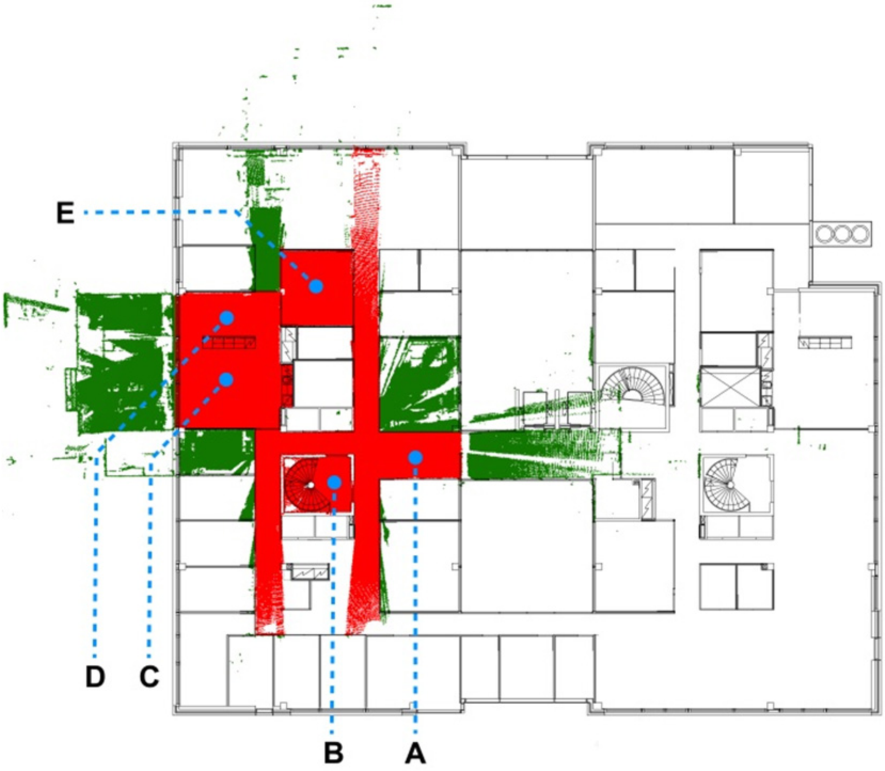

Figure 3. Laser scanned point cloud from top with the building blueprints. Red points cover the area we intended to scan. Green points are by-products that were obtained through glass. See text for map point labels A-E.

4.2 Comparison with laser scanning

In order to critically evaluate the ISB method accuracy, an office containing different kind of indoor spaces is chosen as a test site: hallways (A), a staircase (B), a coffee room (C), a separated corner with sofas (D), and a small meeting room (E), see Fig 3. The hallways and the staircase are narrow spaces having an abundance of glass surfaces and textureless walls, respectively. In other words, they provide extremely hard examples to model with the ISB method. The rooms, on the other hand, have textures but pose a different challenge due clutter (e.g. tables and chairs). A sample image from the office hallway with a spherical scan target is shown in Fig. 4, along with a snapshot from the ISB reconstructed point cloud.

The SIFT algorithm has problems in detecting all essential features in the sample image in Fig. 4. For example, white wall corners are not captured due to low contrast, which obviously reflects on the point cloud quality (not shown). In contrast, rather good reconstructions are made in places containing more contrast variation and texture, such as the coffee room, see in Fig. 4. The above mentioned problem in low-contrast feature detection sets a call for improved methodology, which is beyond the scope of this paper.

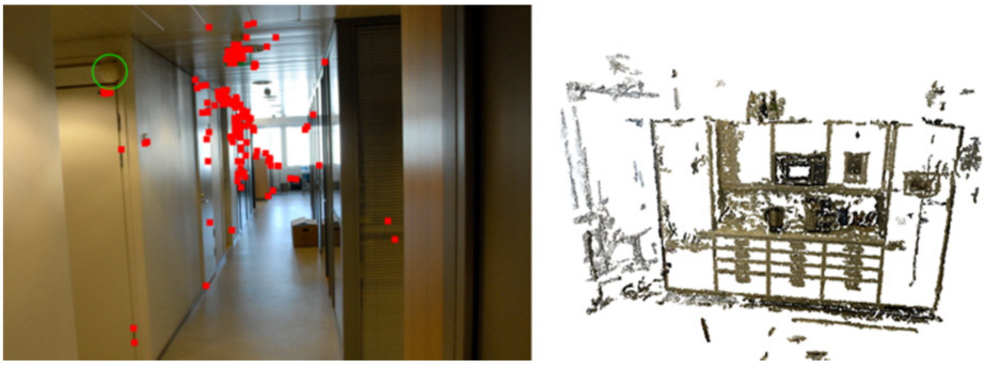

Figure 4. Left: a sample image from the office hallway. Glass walls are abundant. Sunlight from windows alters textures. A spherical scan target is circled with green. Red squares represent features remaining after bundle adjustment. The image is taken from between spots (A) and (B) shown in Fig. 3. Right: a snapshot from the PMVS reconstructed point cloud depicting the coffee room, labeled (C) in Fig. 3.

The laser scanned point cloud was coarse-grained from 115 million points to 15 million points to obtain tractability for manual processing. This *reference point cloud* is shown in Fig. 3. The loss in point cloud density was alleviated by weighing corners (i.e. curvature) over flat surfaces in coarse-graining. This was a natural choice because the ISB point cloud is more likely to contain points in corners than on plain surfaces.

An unordered set of 516 images and 390 images were taken with Nikon D800E with 24mm objective and Canon EOS 60D with 17 mm objective, respectively. Point clouds of 3.4 million and 1.5 million points obtained from these through ISB reconstruction are shown in color in Fig. 5. Capturing the images took us less than 15 minutes per set, in contrast to the laser scanning that took 1h and 30mins.

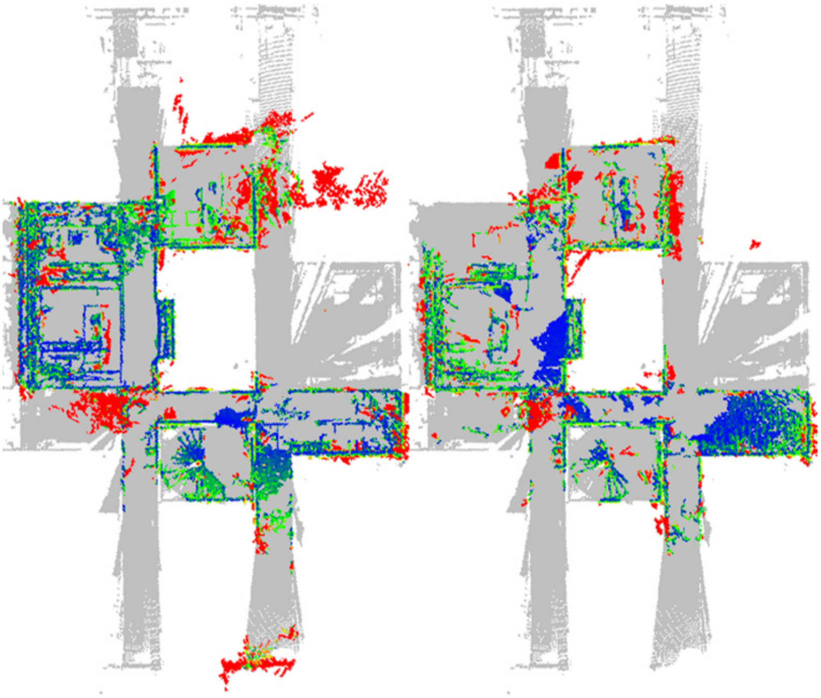

Figure 5. ISB point cloud in colors on top of the laser scanned reference point cloud in grey, for Nikon D800E (left) and Canon EOS 60D (right). See text for details.

The reference point cloud, shown in grey in Fig. 5, is kept as quantitative ground truth to evaluate ISB point cloud accuracies. In particular, the 3D distance *d* from an ISB point to the nearest reference point is measured; Octree partitioning and Chamfer distances were used for fast processing. The distribution of these point-to-point distances are shown in Fig. 6. Distances binned in Fig. 6 are matched with same colors in Fig. 5: blue points are accurate, while points with other colors are far from the ground truth. The bin width is 1 cm, below which resides the total laser scan error estimated in Section 2. From Fig. 6, 44% (27%) of ISB points reside within 2 cm error limits, and 69% (57%) of points reside within 5 cm error limits for Nikon (Canon).

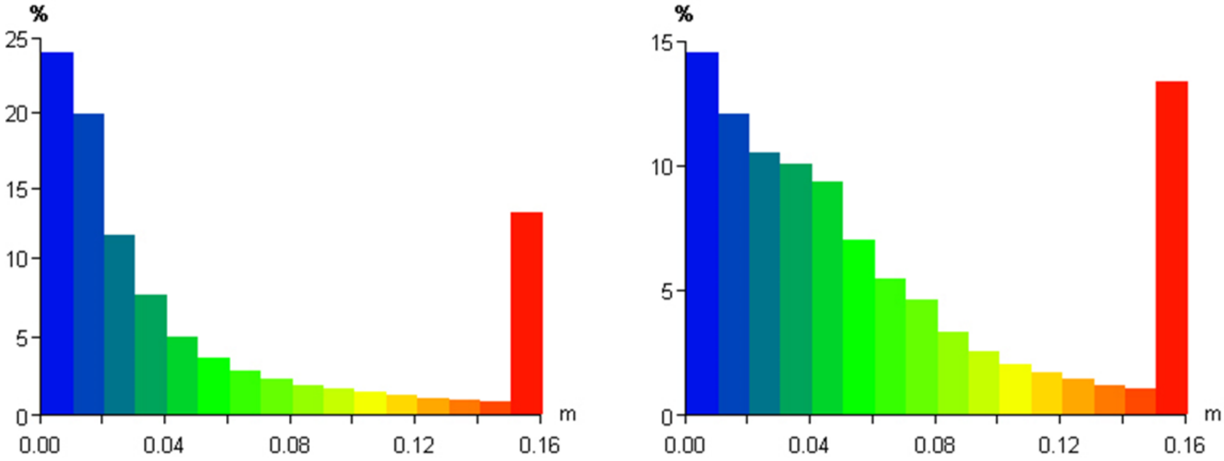

Figure 6. Distribution of the distance ***d*** *from an ISB point to the nearest reference point for Nikon D800E (left) and Canon EOS 60D (right). The bin width is 1cm. For visualization purposes, the rightmost bin contains all the probability mass from the distribution tail going over 16 cm. Note the different percentage scales.*

Our choice to use the point-to-point measure is based on conveniency: it offers a suitable way to evaluate the accuracy of ISB method that aims for BIM and consumer applications. As a consequence, our procedure results in a systematic error due to fact that the reference point cloud is not infinitely dense. On average, the nearest reference point is a distance ΔL away even if an ISB point is absolutely correct and resides on the surface. From a large sample, we estimate that the average density of reference points on a surface is around 20 000 points per m^2. This means ΔL = 0.4 cm, approximately. In order to correct the distributions of *d* in Fig. 6 with respect to this error, we can roughly interpret that they should be shifted leftwards by ΔL. Taking this correction into account, 49% (31%) of ISB points reside within 2cm error limits, and 70% (60%) of points reside within 5cm error limits for Nikon (Canon). Our results compare well with a study where a single staircase was captured using a tripod-mounted fish-eye camera and reconstructed with APERO software resulting in a mean accuracy of 6cm with a sigma of 6cm (Georgantas et al., 2012), and especially so since the reconstruction with APERO involved manual steps with pre-calibration and image pair choosing, which typically yields more accurate results than using an uncalibrated and unordered set of images on which we concentrate.

For qualitative examination of Fig. 5, we refer to the alphabetical markings shown in Fig. 3. For the Canon photo set, the sofa corner (D) completely fails to reconstruct. The Nikon set was extended with 99 extra images to provide extra coverage for these kind of places that the photographer initially captured but that the ISB method failed to reconstruct. Regardless of this extra coverage, the maintenance staircase (B) is still not well reconstructed for neither photo set due to extreme texturelessness; in contrast to the results in (Georgantas et al., 2012) where non-automatic calibration and a tripod is used. For Nikon, the end of bottom right hallway (A) is reconstructed, but shown in red since the area was not laser scanned. The large cluster of red points seen below the coffee room (C) is due to a glass wall (see Fig. 3 for laser points that went through glass). Some red points, especially those in the middle of the coffee room (C), are a result from moving furniture (e.g. cups, chairs) as we were unable to conduct the measurements at the precisely same time.

The meeting room's (E) top and right walls are distorted in both ISB point clouds shown in Fig. 5. In order to limit the possible sources for this cause, we limit the set of images to contain only images taken from inside the meeting room (E). Reconstruction obtained from this limited set of images is shown in Fig 7. Remarkably, the distortion vanishes. With closer inspection, the distortion in the original set is caused by incredibly false values for the focal length f and for the correction terms k_1 and k_2 of some images. The bundle adjustment finds these values from a local minimum as discussed in Section 4.1. In an image shot from outside the room, if the opening into the room looks narrow but still reveals a dense cloud of features while the rest of the image contains only a few features, the fitting scheme weighs the dense cloud above all. Intuitively then, because of the small diameter of the cloud, the fit can produce a wide range of values for f, k_1, and k_2. These distortions in bundle adjustment may be reduced e.g. by heavily weighing initial focal length value, but this risks to eliminate the method's primary characteristics – its automated capabilities. Cropping the meeting room from the ISB point cloud in Fig. 5 results in 54% (34%) of ISB points residing within 2cm error limits, and 75% (66%) of them residing within 5 cm error limits for Nikon (Canon) with the ΔL correction.

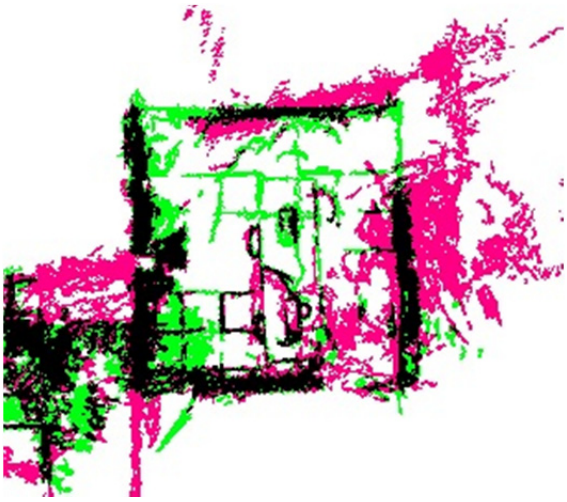

Figure 7. Close up on meeting room (E). Original point cloud from Fig. 5 for Nikon D800E is shown on magenta. Top and right side walls are distorted. Point cloud assembled from a reduced image set containing only images taken within the meeting room is shown on green. Black dots represent overlap between the two point clouds.

4.3 Robustness for further processing

The robustness of the previous step is typically vital to the success of the next step in automated reconstruction. Above, we have studied how accurate and robust the ISB point cloud is. If it is processed further, for example by computing depth maps with cubic grid Manhattan-world assumption (Furukawa et al., 2009a) or tetrahedral based discretization (Jancosek and Pajdla, 2011), the post-processing method has to maintain the accuracy of the core model while coping with outliers.

Maintaining accuracy poses a problem for very large scenes, where the ISB point cloud becomes untractable by its mere size. Now, clustering the scene is a convenient way to overcome this problem, and one known memory bottleneck set by PMVS is already circumvented by CMVS clustering. However, CMVS clustering does not always produce intuitive blocks, as points from separate sides of the scene are joined together, see Fig 8. This means that these clusters cannot be straightforwardly used for further processing, as they lack sufficient mutual overlap leading into loss of information if depth maps are created separately for each cluster. Therefore, any depth map reconstruction basically requires loading all MVS points and image projection matrices into memory, which becomes the problem. We implemented a method similar to (Furukawa et al., 2009a) to roughly test this for Manhattan world depth maps. The image textures, on the other hand, can be loaded one at a time to conserve memory and preserve computational efficiency, so the amount of scene images itself does not pose a problem.

In some circumstances when the scene contains repetitive geometry, bundler fails by locating two different "blocks" on top of each other, see Fig 9. In other words, separate SIFT feature patterns become so similar that they are matched as if they were the same pattern. This result was obtained with Nikon D200 with 20mm objective. We estimate from praxis that the probability of this pattern mixing is increased if the objective offers a narrow field of vision and if the photographer is not following three rules of thumb listed below.

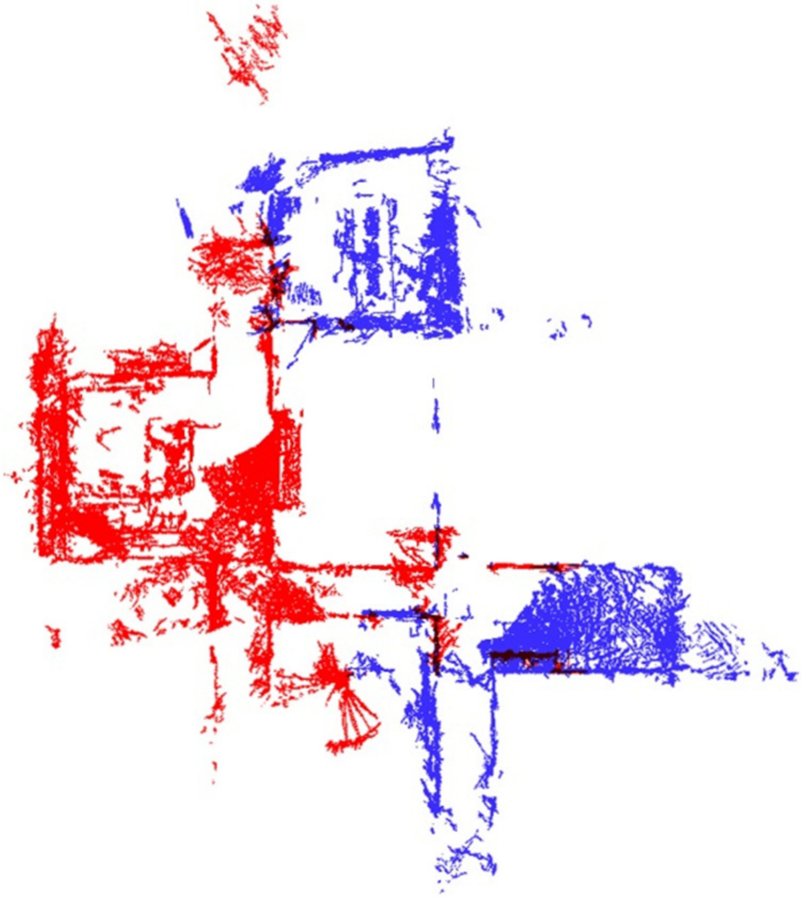

Figure 8. Top view of the point cloud obtained with ISB method (points after PMVS reconstruction). Red and blue colors indicate the two clusters in which CMVS split the scene. Blue areas top and below are not connected in any way, but still reside in the same cluster. Shots were taken with Canon EOS 60D with 17 mm objective.

4.4 Robustness with respect to photo-shooting

One vital factor affecting the quality of the image set is photo-shooting. The photographer can optimize the indoor environment by e.g. closing curtains to remove sunlight reflections and to reduce the amount of optical surfaces. In addition, three simple photo-taking rules certify that the photo album is not taken for humans but for a computer to look on:

- Indoor spaces are always tight. Keeping your back close to a wall maximizes the camera's field of vision.
- PMVS by default discards images that have an angle difference less than 10 degrees with an already accepted photo, so constant tilting and moving the camera is mandatory
- Each feature must be in at least 3 images, so after capturing a feature it still needs to be shot again and again.

These rules sound simple, but are challenging in practice. In order to capture a wanted indoor scene shown in Section 4, we had to go back to take more photos until the image set (for Nikon D800E) was complete. It would seem that the biggest risk of the method toward its robustness is the photographer.

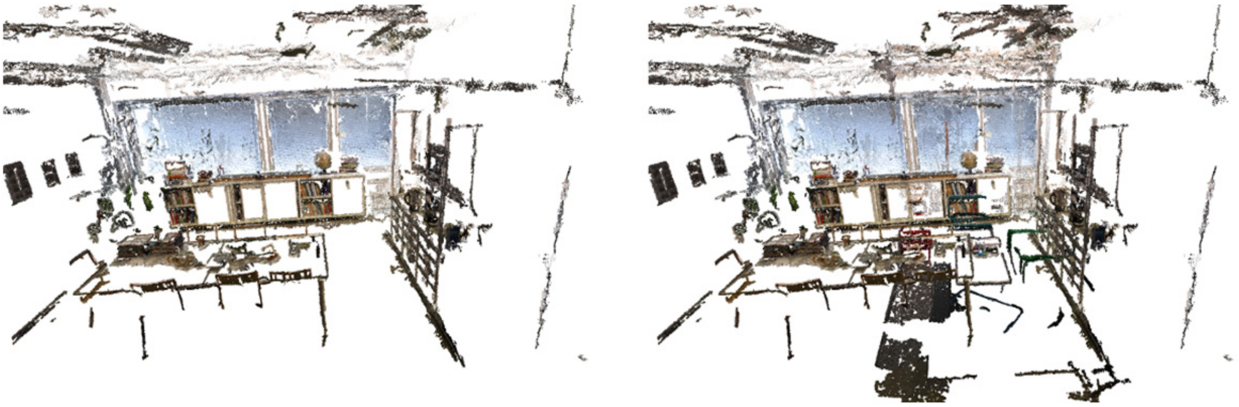

Figure 9. Left: snapshot of an ISB point cloud reconstructed from images taken only from coffee room (E). Right: same snapshot from the original reconstruction. Sofa corner (D) is wrongly reconstructed on top of the coffee room (E) due to SIFT similarities in the windows and the roof. Letters (E and D) are visible in and refer to Figure 3. The image set was taken with a Nikon D200.

5. CONCLUSIONS

In this paper, we have studied perhaps the most potent state-of-the-art method for image sequence-based (ISB) automated reconstruction of building interiors. As our main result, we quantitatively measured the ISB point cloud accuracy using a reference point cloud obtained with a tripod-mounted FARO Focus 3D laser scanner. The accuracy of our scanner is about ±2 mm (Chow et al., 2012), with the total cumulative error residing below 1 cm. We concentrated on point cloud comparison, as the reconstructed ISB model can only be as accurate as its SfM/PMVS point cloud is, and since both the laser scanning results and the ISB method attempt to represent the environment as it is at the time of measurement. From our data, the automated ISB method is able to reconstruct a 3D point cloud with 49% (31%) of points residing within 2 cm error limits, and 70% (60%) of points residing within 5 cm error limits for Nikon D800E and Canon EOS 60D, respectively. The ISB point cloud quality suffered from poor SIFT performance on low-contrast images that left important features unrecovered. We used unordered image sets taken without prior camera calibration for the purpose of maintaining wide applicability for our results.

In conclusion, image-based 3D reconstruction offers potential for those various consumer and BIM-related applications that are satisfied with the level of accuracy of 5 cm.

Future work will concentrate on improving feature detection from low-contrast images, and evaluating the applicability of post-point-cloud-reconstruction methods by studying whether they maintain the model accuracy.

6. ACKNOWLEDGEMENTS

The authors thank Academy of Finland for funding this research, Grant No. 257755 (VL), and CoE-LaSR Grant No. 272195 (VL, MK, HH).

7. REFERENCES

Aguilar, J.J., Torres, F., Lope, M.A., 1996. Stereo Vision for 3D Measurement: Accuracy Analysis, Calibration and Industrial Applications. Measurement 18 (4), 193-200.

Bosse, M., Zlot, R., Flick, P., 2012. Zebedee: Design of a Spring-Mounted 3-D Range Sensor with Application to Mobile Mapping. Robotics, IEEE Transactions on 28 (5), 1104-1119.

Chow, J., Lichti, D., Teskey, W., 2012. Accuracy Assessment of the Faro focus3D and Leica HDS6100 Panoramic Type Terrestrial Laser Scanner through Point-Based and Plane-Based User Self-Calibration. FIG Working Week 2012: Knowing to manage the territory, protect the environment, evaluate the cultural heritage. Rome, Italy. May 6-10.

Flint, A., Mei, C., Murray, D., Reid, I., 2010. A Dynamic Programming Approach to Reconstructing Building Interiors, In: European Conference on Computer Vision (ECCV) 2010, Springer Berlin Heidelberg, 01/01, pp. 394-407.

Fryer, J.G., Brown, D.C., 1986. Lens Distortion for Close-Range Photogrammetry. Photogrammetric Engineering and Remote Sensing 52 (1), 51-58.

Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R., 2010. Towards Internet-scale multi-view stereo, In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp. 1434-1441.

Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R., 2009a. Manhattan-world stereo, In: Computer Vision and Pattern Recognition (CVPR) 2009. IEEE Conference on, pp. 1422-1429.

Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R., 2009b. Reconstructing building interiors from images, In: Computer Vision, 2009 IEEE 12th International Conference on, pp. 80-87.

Furukawa, Y., Ponce, J., 2010. Accurate, Dense, and Robust Multiview Stereopsis. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32 (8), 1362-1376.

Georgantas, A., Bredif, M., Pierrot Desseilligny, M., 2012. An Accuracy Assessment of Automated Photogrammetric Techniques for 3D Modeling of Complex Interiors, In: ISPRS12, pp. -B3:23-28.

Harms, H., Beck, J., Ziegler, J., Stiller, C., 2014. Accuracy analysis of surface normal reconstruction in stereo vision, In: Intelligent Vehicles Symposium Proceedings, 2014 IEEE, pp. 730-736.

Hartley, R.I., Zisserman, A., 2004. Multiple View Geometry in Computer Vision, Second ed. Cambridge University Press, ISBN: 0521540518,.

Hernandez, C., Vogiatzis, G., Cipolla, R., 2007. Probabilistic visibility for multi-view stereo, In: Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, pp. 1-8.

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

Kaartinen, H., Hyyppä, J., Kukko, A., Jaakkola, A., Hyyppä, H., 2012. Benchmarking the Performance of Mobile Laser Scanning Systems using a Permanent Test Field. Sensors 12 (9), 12814-12835.

Khoshelham, K., Elberink, S.O., 2012. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. Sensors 12 (2), 1437-1454.

Kukko, A., Kaartinen, H., Hyyppä, J., Chen, Y., 2012. Multiplatform Mobile Laser Scanning: Usability and Performance. Sensors 12 (9), 11712-11733.

Liu, T., Carlberg, M., Chen, G., Chen, J., Kua, J., Zakhor, A., 2010. Indoor localization and visualization using a human-operated backpack system, In: Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on, pp. 1-10.

Lowe, D., 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60 (2), 91-110.

Snavely, N., Simon, I., Goesele, M., Szeliski, R., Seitz, S.M., 2010. Scene Reconstruction and Visualization from Community Photo Collections. Proceedings of the IEEE 98 (8), 1370-1390.

Snavely, N., Seitz, S., Szeliski, R., 2006. Photo Tourism: Exploring Photo Collections in 3D. ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2006 25 (3), 835-846.

Xiao, J., Furukawa, Y., 2012. Reconstructing the World’s Museums, In: Computer Vision – ECCV 2012, Springer Berlin Heidelberg, pp. 668-681.