Using Open Source Libraries for Obtaining 3D Scans of Building Interiors

Adam L. Kaczmarek[1(✉)], Mariusz Szwoch[1], and Dariusz Bartoszewski[2]

[1] Gdansk University of Technology, ul. G. Narutowicza 11/12, 80-233 Gdansk, Poland
{adam.l.kaczmarek,szwoch}@eti.pg.edu.pl
[2] Forever Entertainment S.A., Gdynia, Poland
dariusz.bartoszewski@forever-entertainment.com

Abstract. This paper describes methods for making 3D scans of building interiors. The main application of these methods is the development of first-person view (FPV) perspective games or creating virtual museums. 3D scans can be made with the use of different equipment such as Light Detection and Ranging (LIDAR), time of flight (TOF) cameras and structural light 3D scanners. However, the paper focuses on using stereo cameras for obtaining 3D scans, because of its low cost. It is a significant factor for small and medium size game development studios. The paper considers both the method based on photogrammetry and stereophotogrammetry. In photogrammetry the Structure from Motion technology is used for making 3D scan on the basis of images of an object taken from different locations. Photogrammetry uses stereo vision algorithms for acquiring depth maps and point clouds representing distances between a stereo camera and objects. The paper analyses implementations of these technologies available in programing libraries OpenCV, openMVG and openMVS. The paper shows that the algorithm for Structure from Motion provided in openMVG can be successfully applied for obtaining point clouds from pair of images despite this algorithm is intended for use with a greater number of input images.

Keywords: 3D scanning · Structure from Motion · Stereo vision

1 Introduction

The paper is concerned with acquiring 3D models of buildings interiors, mainly for the purpose of their use in video games. In general, 3D scanning is usually used for making 3D models of individual objects of any size, which usually means reconstruction of their outer surface. However, scanning the inner view of building has also wide range of applications. Obtained 3D scans of interiors can be used for generating a virtual 3D map of rooms or even the entire building. Such models can be used for archiving purposes as the input to CAD (computer aided design) or BIM (building information modeling) systems [1]. Next possible application is to prepare a virtual walk in a real location, e.g. to present the inaccessible or hard to reach interiors to the interested people. This approach can be used to create virtual museums presenting historical interiors or chambers with original equipment [2]. Another possible application is the interior design.

L. Borzemski et al. (Eds.): ISAT 2018, AISC 852, pp. 201–210, 2019.
https://doi.org/10.1007/978-3-319-99981-4_19

Using the techniques of virtual reality (VR) and mixed reality (MR) it is possible to present new architectural ideas, models or solutions that partially or totally replace the existing ones. These new arrangements can be realistically presented to the audience using VR goggles or in a cave automatic virtual environment (CAVE) [3]. Finally, in the last few years, another field of application is gaining popularity and attracting the attention of many researchers. This new application is the use of automatically or semi-automatically scanned interiors in video games. Although, these solutions are usually limited to 3D games with the first-person view (FPV) perspective, which action takes place in real locations, such games are still a large part of market worth billions of dollars [4]. This paper focuses on this application area.

There is a variety of equipment that enables obtaining 3D scans, which can be divided generally into active and passive devices. The most important active devices include Light Detection and Ranging (LIDAR), time of flight (TOF) cameras and structural light 3D scanners. Although, active scanners provide high quality 3D geo-metric models (meshes) in most situations, they usually do not provide information about the objects' texture (material) that has to be manually added later. Moreover, TOF cameras and structural light scanners usually have very limited resolution that is insufficient in many applications. Finally, very important disadvantage of the active scanners is their cost. The equipment dedicated for making 3D scans is usually very expensive, which is particularly significant flaw in case of small game development studios which would like to obtain 3D scans of building interiors. However, 3D scans can be also made with the use of cameras, which are relatively cheaper and widely accessible. A studio making a video game can already be in possession of cameras that can be used for making scans. Therefore, this paper analysis the subject of using cameras for making 3D scans of building interiors.

There are many algorithms that make it possible to obtain a 3D scan from images [5, 6]. Open source implementations of these algorithms are also available. Such an implementation is included in one of the most important programming library in the field of computer vision and image processing, which is OpenCV [7].

Among many algorithms included in OpenCV there are algorithms for processing images in order to obtain 3D images and 3D scans. However, experiments presented in this paper shows that the results of these algorithms are not suitable for scanning interiors of building. The paper also presents experiments with other programming libraries called openMVS and openMVG. These libraries are much less popular than OpenCV however results of using them are much more suitable for the purpose of scanning interiors.

Original contributions of this paper include: (1) The comparison of the results of acquiring 3D scans of building interiors with the use of libraries OpenCV, openMVG and openMVS. (2) The description of the results of applying the technology of Structure from Motion to a pair images from a stereo camera. (3) The proposition of the 3D scanning method for building interiors.

2 Related Work

There are two main methods of acquiring 3D scans and 3D images with the use of cameras. The first one is stereophotogrammetry that uses the technology of stereo vision [6]. The other one is photogrammetry based usually on the Structure from Motion (SfM) algorithm [8].

2.1 Stereo Vision

In the technology of stereo vision 3D images are obtained on the basis of a pair of images taken with a stereo camera that consists of two cameras fixed to each other on a special rig.

Any object, that is visible by both cameras, is located at different coordinates in images taken by these cameras. Let us suppose that a certain object is visible at coordinates x_1, y_1 on the left image and at coordinates x_2, y_1 in right image. The difference between values x_1 and x_2 is called the *disparity* (d). If the object is closer to the camera system, then the disparity is greater. Disparities determined for the entire image form a *disparity map*. A disparity map can be easily transformed to either a *depth map* or a set of points in 3D space (*point cloud*). A depth map contains values of distances between a stereo camera and each point of the objects (so called "pixel with depth"), whereas a point cloud contains 3D coordinates of each point in a certain local 3D coordinate system.

In order to transform a disparity map to a depth map it is necessary to have data related to physical features of a stereo camera set such as distances between cameras b (baseline) and focal length f of lens. These data can be recovered from technical specification of a stereo camera or on the basis of its calibration.

The calibration is a process that is required when a stereo camera is used. The problem with stereo cameras is such that their optical axes are not parallel to each other. In real application there will always be imperfections. Moreover, cameras are also slightly rotated relatively to each other. However, there are also distortions which occur in every single camera. Lens of cameras cause that straight lines visible in real objects became bent in images. This effect occurs in different extent for different kinds of lens. It is the most evident in case of fish-eye lens that cause a barrel kind of a distortion [7]. In images made with this kind of lens, straight parallel lines would look like they were drawn on a sphere.

The problem with distortions is such that they cause deterioration in the quality of disparity maps. Because of distortions, errors occur in these maps. Distortions which occur in images can be neutralized by transforming these images. Different kind of transformations can be used including shifts and rotations. In order to appropriately transform images in order to neutralized distortions it is necessary to resolve which distortions occur in images and to what extent.

These data are obtained in the process of image calibration. The calibration of a stereo camera is based on making images of a specific, regular image pattern. These images are analyzed by an algorithm for calibrating images. The algorithm recognizes characteristic points and on their basis it calculates transformation parameters. A typical kind of an image patters used for such calibration is a chessboard.

The calibration leads to the improvement in the quality of images, but it does not resolve all the problems with obtaining disparity maps from stereo pairs. The main problem is that the pair of images needs to be processed with the use of a stereo matching algorithm in order to determine the location of the same objects in the pair of images. Different kinds of algorithms were developed, which differ in their speed, quality of results and the level of required computer resources to operate.

There are many rankings of stereo matching algorithms. One of the most popular is the ranking provided in Middlebury stereo vision page [9]. Another popular ranking of stereo matching algorithms is KITTY (http://www.cvlibs.net/datasets/kitti/) [10]. These ranking include both well-known algorithms such as those provided in the OpenCV library and many novel algorithms tested only by their authors.

2.2 Structure from Motion

Another approach to obtaining 3D data on the basis of images is the technology of Structure from Motion. In this technology, images of an object are made from many points of view located around the object. The set of images is then analyzed by the SfM algorithm that searches for characteristic (*salient*) points in all images. There are different methods of searching for these landmark points, however one of the most important algorithms designed for this purpose are scale-invariant feature transform (SIFT) and Speeded up robust features (SURF) [11].

On the basis of location of these characteristic points in images the algorithm estimates positions of a camera in the moment, when images were taken. This information is used to determine the location of characteristic points in 3D space. The result of a SfM algorithm is a points cloud reflecting the shape of real objects. This is a significant difference in comparison to algorithms for stereo vision, which produce a depth map acquired from a single point of view. In order to obtain a whole shape of a real object it is necessary to merge disparity maps acquired from different points of view. In case of SfM technology the whole point cloud representing the object is the direct result of the algorithm. The result can also have a form of a triangle mesh used in computer graphics.

There are different rankings of SfM algorithms similarly as there are rankings of algorithms for stereo vision. The Middlebury Stereo Vision Page, apart from the ranking of stereo matching algorithms, contains also a ranking of algorithms for obtaining Structure from Motion (http://vision.middlebury.edu/mview/eval/) [8].

2.3 OpenCV Library

The OpenCV library is open source and it contains implementations of a large number of algorithms used in computer vision and image processing including algorithms for image filtering, recognition and machine learning. The library contains also algorithms for stereo vision and Structure from Motion techniques.

The OpenCV library contains four stereo matching algorithms for obtaining disparity maps from images taken by a stereo camera, which are: Stereo Block Matching (Ste-reoBM), Stereo Semi-Global Block Matching (StereoSGBM), Heiko Hirschmüller algorithm (StereoHH) and Stereo Variational methods (StereoVar)

[12–14]. OpenCV contains also algorithms for stereo camera calibration. The calibration can be performed on the basis of image patterns containing black and white chess-board or circles. The OpenCV calibration makes it possible not only to improve the quality of disparity maps but also to acquire point clouds from these maps. OpenCV library also contains SfM algorithms. They are available in this library starting from version 3.0.

2.4 openMVG and openMVS Libraries

An alternative method to using OpenCV for obtaining 3D scans is taking advantage of Open Multiple View Geometry (openMVG) and Open Multiple View Stereovision (openMVS) libraries [15, 16]. Our experiments showed that these libraries make it possible to acquire a 3D scan that has much higher quality than OpenCV.

OpenMVG library contains computer vision algorithms including those for estimating positions of camera using the SfM technology. The input data to the algorithm are the images taken from different points of view. OpenMVG not only determines camera positions but also produces a sparse point cloud corresponding to real objects visible in images.

Unfortunately, point clouds obtained from openMVG are sparse and not sufficient for using as models in 3D video games. However, this sparse point cloud can be expanded to a dense one with the use of the openMVS library. The output data from openMVG which are the location of cameras and initial point cloud can be provided as input data to openMVS. Taking into account the set of images processed by openMVG the openMVS library can generate a dense point cloud of the scene.

3 Technological Requirements

The research presented in this paper is dedicated for making 3D scans of building interiors. This assumption imposes requirements on the technology, which needs to be used. Foremost, as described in the introduction, only passive scanning technology, based on making images, is concerned. There are also some additional requirements that have to be fulfilled.

In case of building interiors it is not possible to take images from points of view located around objects as is expected by SfM technology. Inside buildings it is only possible to take photos of walls from one, visible side.

Moreover, it would be advisable to make it possible to exclude some images from the set of input images. That is why, a set of input images needs to be prepared manually. It is necessary to select images that will be used. In case of SfM technology, the entire image set is analyzed by the algorithm for obtaining 3D scans, what is time consuming. When a scan of a building interior is acquired, some fragments of the space can be unnecessary. The possibility of adding and deleting some parts of a 3D scan without the necessity to recalculate the entire model would enhance the efficiency of preparing scans.

Another significant technological limitation is such that inside the building space there may appear cavities with very limited access. Such cavities occur in narrow spaces

blocked from many sides by building structure or furniture. In these cases it is only possible to make photos from only one point of view.

Because of the requirements, restrictions and constraints listed above, the research presented in this paper is based on using stereo cameras that enables to obtain a point cloud representing objects in 3D space visible from one point of view. These point clouds can be included and excluded from the resulting, complete 3D scan of an interior consisting of many point clouds obtained from different points of view.

Using stereo cameras do not cause that only the technology of stereo vision can be used. The application designed for making 3D scans of interiors can also take advantage of the SfM. In case of scanning building interiors this technology is particularly advantageous when it is applied to pairs of images from a stereo camera.

4 Experiments

The purpose of the experiments was to select a mature technology that would make it possible to obtain 3D scans of building interiors on the basis of images from a stereo camera. We did not concern novel algorithms which were not verified in different circumstances and environment. These algorithms have a potential to become commonly used methods. We have focused on optimized and well-tested implementations of algorithms for processing images. Therefore, for our tests, we have selected algorithms available in open source libraries OpenCV, openMVG and openMVS.

The input data set, that we used in the experiments, consisted of images obtained by ourselves and images provided on Middlebury Stereo Vision Page. We used the Sony RX100 camera for making images. The camera has a 1.0” sensor size.

4.1 Stereo Vision in OpenCV

We have performed experiments with stereo vision using the OpenCV library. We have encountered such a problem that algorithms for calibrating stereo cameras used in OpenCV are intended for use for a specific range of distances between a stereo camera and real object. Image transformations performed during calibration include shifting images, so that they can be matched by stereo matching algorithm for a specific disparity range that corresponds to a range of distances.

The problem with this method is such that it is inconvenient to use this kind of calibration for scanning building interiors. It would be necessary to calibrate cameras with the use of an image pattern which has a size of a building wall. It is problematic to prepare such samples. Cameras calibrated with OpenCV for a certain distance can be used for other distances however it causes significant deterioration in the quality of the results. Moreover, there are also problems with determining point clouds from such obtained disparity maps. Sample calibrated images showing furniture and a part of a wall are presented in Fig. 1. Cameras were calibrated with the use of a chessboard which had a size 78 cm $\times$ 62.5 cm.

Fig. 1. A sample calibrated images from a stereo camera.

Calibrated images are input data to a stereo matching algorithm which obtains a disparity map. Figure 2 present disparity map obtained for the input images presented in Fig. 1. The Stereo Semi-Global Block Matching algorithm was used. Results show that there are many points for which disparities were not obtained appropriately. Although, we did not have ground truth with real values of distances between cameras and objects, we could determine, that the point cloud generated on the basis of the disparity map also did not reflect the shape of objects visible in images.

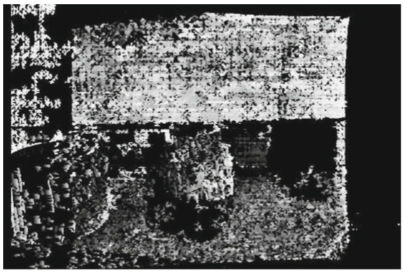


Fig. 2. A disparity map obtained for images presented in Fig. 1.

Stereo matching algorithms that are available in the OpenCV library, require providing a disparity range which is analyzed by the algorithm. If the range is wide, then the algorithm considers identifying the location of objects that are within a long range of distances from the stereo camera. However, the increase of this range causes the increase in the number of errors occurring in the results. In case of making 3D scans of building interiors, it would be advantageous to make scans of all objects located both close and far from cameras. Using the technology described in this subsection it is not possible, if a high quality of the results need to be achieved.

4.2 Structure from Motion in OpenCV

The authors of this paper have also verified results of using algorithms for SfM implemented in the OpenCV library. The algorithm did not manage to produce readable results for the image set of building interiors prepared by the authors. Therefore, the authors conducted an experiment that involved processing a sample series of images for testing

the technology of SfM provided in Middlebury Stereo Vision Page. It was the Temple data set. In the experiment the authors obtained a 3D scan on the basis of 5 images acquired from this set. These images are presented in Fig. 3.

Fig. 3. Images from the temple data set provided in Middlebury Stereo Vision.

The results are presented in Fig. 4. The image shows location of cameras retrieved by the algorithms included in the library (marked with yellow color) and the 3D scan obtained from these images.

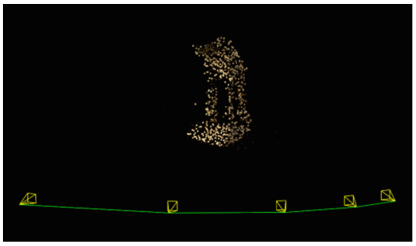


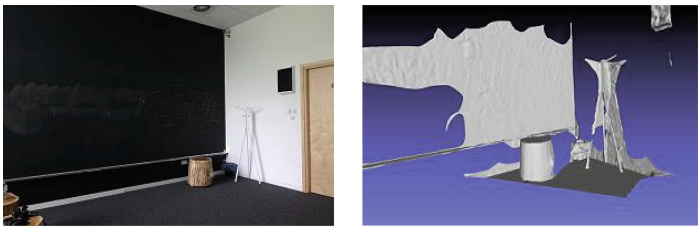
Fig. 4. The location of cameras retrieved by the SfM algorithms and the obtained 3D scan.

The scan contains a small amount of points representing the object. Ground truth for this data set is available in Middlebury Stereo Vision Page and results of using OpenCV are included in the ranking presented in this page. However, the algorithm did not manage to obtain realistic results when it was executed with only two input images. Even without comparing them with ground truth the authors could determine that they did not met our expectations. As described in Sect. 3 obtaining results for two images was a desired technology for scanning building interiors.

4.3 Structure from Motion in openMVG and openMVS

We have conducted experiments in order to verify the performance of these libraries for the purpose of obtaining 3D scans of building interiors. Experiments were focused on obtaining data from a pair of images taken with the use of a stereo camera. Experiments included making images of a parking lot, a conference room and a room resembling medieval chambers.

Figure 5(a) shows a sample input image from a pair of images. The image shows a conference room with a blackboard, modern chairs and a hanger. Figure 5(b) presents the result of using algorithms available in openMVG and openMVS.



(a) (b)

Fig. 5. An image from the input set (a) and a 3D scan obtained on the basis of openMVG and openMVS libraries.

Our experiments showed that openMVG leads to correct results, even if there are only two input images. In case of this data set the authors did not have ground truth, however in contrast to the results obtained from OpenCV, the results acquired from openMVG were not blurred and it was possible to recognize shapes of real objects in the generated point cloud. The general shape was reconstructed, though, there are some areas for which a 3D model was not obtained. This observation applies to all kinds of areas considered in experiments. However, in case of reconstructing building interiors there are other stereo images taken from other locations and different angles. A complete scan will consists of many point clouds obtained with the use of a stereo camera from different points of view. Thus, blank areas can be filled in, when stereo images from other camera locations are taken into account. The greatest advantage of this technology is such that it provides reliable results for the area of a scene for which a 3D scan is obtained even if there are areas which are not included in the scan.

This kind of areas occurs in particular in case of monochrome flat surfaces. These elements of the scene can be relatively easily added to the 3D scan manually in order to obtain a complete scan. There are also other kinds of surface which are problematic for 3D scanning like reflective or transparent materials. The scan of a building interior containing such materials will have to be edited manually in order to be correct, however the technology of 3D scanning will provide the base of the reconstruction of buildings interiors.

5 Summary

One of the most significant conclusions from the research presented in this paper is such that the technology of Structure from Motion can be used for pairs of images obtained with the use of stereo camera. In general, SfM is not dedicated for stereo cameras, but for a large set of input images from different points of view. However, algorithms implemented in openMVG and openMVS libraries make it possible to acquire 3D data

from image pairs. This feature is crucial for developing an application for making 3D scans of building interiors.

Acknowledgment. This work was supported by the Sectoral Programme GAMEINN within the Operational Programme Smart Growth 2014–2020 under the contract no POIR. 01.02.00-00-0140/16.

References

1. Jung, J., Hong, S., Jeong, S., Kim, S., Cho, H., Hong, S., Heo, J.: Productive modeling for development of as-built BIM of existing indoor structures. Autom. Constr. **42**, 68–77 (2014)
2. Gomes, L., Regina, O., Bellon, P., Silva, L.: 3D reconstruction methods for digital preservation of cultural heritage: a survey. Pattern Recogn. Lett. **50**, 3–14 (2014)
3. Lebiedz, J., Szwoch, M.: Virtual sightseeing in immersive 3D visualization lab. ACSIS-Ann. Comput. Sci. Inf. Syst. **8**, 1641–1645 (2016)
4. Szwoch, M., Kaczmarek, A., Bartoszewski, D.: STERIO - reconstruction of 3D scenery for video games using stereo-photogrammetry. In: Proceedings of the Conference on Game Innovations (CGI), Lodz (2017)
5. Kaczmarek, A.L.: Improving depth maps of plants by using a set of five cameras. J. Electr. Imaging **24**(2), 023018 (2015)
6. Kaczmarek, A.L.: Stereo vision with equal baseline multiple camera set (EBMCS) for obtaining depth maps of plants. Comput. Electr. Agric. **135**, 23–37 (2017)
7. Bradski, D.G.R., Kaehler, A.: Learning OpenCV, 1st edn. O'Reilly Media Inc., Sebastopol (2008)
8. Seitz, S.M., Curless, B., Diebel, J., Scharstein D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), vol. 1, pp. 519–528. IEEE (2006)
9. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vis. **47**(1/2/3), 7–42 (2002)
10. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The KITTI vision benchmark suite. In: Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3354–3361. IEEE (2012)
11. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
12. Konolige, K.: Small vision systems: hardware and implementation. In: Shirai, Y., Hirose, S. (eds.) Robotics Research, pp. 203–212. Springer, London (1998)
13. Kosov, S., Thormhlen, T., Seidel, H.P.: Accurate real-time disparity estimation with variational methods. In: Advances in Visual Computing. LNCS, vol. 5875, pp. 796–807, Springer, Heidelberg (2009)
14. Hirschmuller, H.: Stereo processing by semi-global matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. **30**, 328–341 (2008)
15. openMVG Homapage. http://openmvg.readthedocs.io/en/latest/. Accessed 16 May 2018
16. openMVS Homapage. http://openmvg.readthedocs.io/en/latest/software/MVS/OpenMVS/. Accessed 16 May 2018