Using Stereo-photogrammetry for Interiors Reconstruction in 3D Game Development

Mariusz Szwoch[1], Adam L. Kaczmarek[2], and Dariusz Bartoszewski[3]

[1] Gdansk University of Technology, Gdansk, Poland szwoch@eti.pg.edu.pl
[2] Gdansk University of Technology, Gdansk, Poland adam.l.kaczmarek@eti.pg.edu.pl
[3] Forever Entertainment s.a., Gdynia, Poland dariusz.bartoszewski@forever-entertainment.com

Summary. This paper proposes a new approach to the reconstruction of building interiors based on stereo-photogrammetry. The proposed technology can be used, among others, for modeling rooms for video games whose action takes place in locations reflecting real interiors. The original, large-scale experiment showed that the proposed approach is economically justified in the case of larger and more complex spaces which are more difficult to model in a traditional way. Also novel approach is proposed to deal with the problem of scanning flat uniform surfaces.

1 Introduction

Modern 3D video games require hard use of high quality 3D assets for both game characters as well as level scenery. In the traditional approach, all such assets are created by 3D artists or bought from a store. The first approach is a very labor-intensive task, while the second one does not guarantee originality of the assets, what is easily recognized by players. Situation becomes more complicated in case when game concept requires that the action takes place in a location that represents a real world place. In such case, traditional, handmade modeling seems to be non-creative, yet time and labor consuming task that could be replaced by automatic or at least semi-automatic machine work. Automation of room reconstruction not only reduces time and the cost of modeling the 3D stage, but also supports rapid prototyping, which is a very important element of agile development process and often determines the further direction of the game development. That is why, it is so important to accelerate the process of modeling 3D scenes that have the appearance as close as possible to the target. As long as this target appearance is a figment of the artist's imagination, the only solution is to improve the 3D modeling tools and hire skilled 3D artist. If, however, objects appear on the scene to reflect real objects, then there is the possibility of using additional tools allowing for automatic or semi-automatic reconstruction of these objects for the needs of the game.

There are several possible solutions of reconstruction of 3D inanimate objects, including photogrammetry [1],[2], methods based on lasers and augmented depth sensors, e.g. structured light RGB-D ones [3],[4],[5]. Unfortunately, all of the mentioned solutions have some drawbacks that limit their usage in video game industry, especially by small and medium developer teams. Laser based scanners are very precise as to geometry scanning but are very expensive and do not support textured models. Time-of-Flight (ToF) sensors are cheaper but have very limited resolution and still do not provide texture information. RGB-D scanners, based mostly on the structured infrared light patterns, like Microsoft Kinect or Intel RealSense, can provide complete 3D models of usually enough quality but are limited to small distances of about several meters. Finally, photogrammetry methods can provide high quality textured models but requires numerous input data and high computational costs. Moreover, it often fails to reconstruct flat surfaces with no salient points.

In this paper, the main concept of the STERIO project is presented, which aims to overcome problems of above approaches by using stereo-photogrammetry methods which, under some assumptions, can offer reliable scanning of textured objects at reasonable financial and computational costs. Performed experiments indicate that such approach is faster than traditional one. Additionally, a special way of treating flat surfaces is proposed, which uses graphic patterns displayed on a problematic surface.

2 Background

Three-dimensional (3D) video games are the most important and the most spectacular part of the video games market. They dominate primarily on the platforms of personal computers and video consoles, but they are also an essential and still growing part of games for mobile and browser platforms.

Because independent modeling of three-dimensional objects is very time-consuming, for some time the methods, algorithms, tools and equipment supporting the 3D scanning have been developed. Initially, they focused on scanning small and medium 3D objects, including motion capture systems. To scan such smaller objects, relatively cheap camera sets and various types of 3D scanners and depth sensors could be used, like Kinect or RealSense. Modeling of larger objects and 3D locations could be carried out mainly using expensive laser scanners which are available mainly to larger development studios. With a smaller game budget, the modeling of larger assets remained the domain of 3D graphic designers. It is only recently that methods allowing the use of cameras for scanning buildings, their interiors and even vegetation have attracted a lot of interest [2],[4],[6],[7],[8].

3D scanning technology has been developed for the last decades and used in different fields, such as preserving objects significant for a cultural heritage

[6]. The best solution to acquire relatively small objects are 3D scanners using different scanning technologies, e.g. 3D structural-light scanners [7]. Unfortunately, due to their limitations, such 3D scanners cannot be at reasonably costs used for scanning bigger objects, e.g. buildings and their interiors.

This problem is less significant when scanning devices based on lasers are used. This class of devices includes *time-of-flight* (TOF) cameras and devices based on *light detection and ranging* (LIDAR) technologies [8],[9]. LIDAR has a movable laser which is pointed in different directions during scanning. A 3D scan results from a series of measurements performed around a single point. TOF makes a scan with the use of light laser beam distracted in different directions. The disadvantage of both of these devices is such that the resulting 3D scan is sparse and lacks color information. LIDARs and TOF cameras are also relatively expensive. However, these scanners are more immune to light conditions than 3D structural-light scanners.

3D scans can be also obtained with the use of standard, optical cameras which are very accessible and relatively cheap. In general, there are two kinds of methods for acquiring 3D scans with cameras, which are *photogrammetry* and *stereo-photogrammetry*. The photogrammetric approach bases on a series of usually unrelated photos taken from around a scanned object. Special algorithms, such as *scale-invariant feature transform* (SIFT) and *speeded up robust features* (SURF) are used to detect the characteristic points, or *landmarks*, in those images [10]. Matching algorithms, such as *random sample consensus* (RANSAC) are then used to find the correspondence between images pair. On this base, algorithms, such as *structure from motion* (SfM), allow to reconstruct a 3D point cloud representing the scanned object. This approach is the most popular nowadays as it allows for the reconstruction of buildings, plants, and other objects based on uncorrelated images possible taken using different cameras at different time, e.g. as is the case in social media services. This approach is also used in a professional tools, such as Agisoft Photoscan [11]. The greatest disadvantage of this approach is the computational power needed to process big sets of images.

Another technique, which allows to retrieve a 3D view of objects, is stereo-photogrammetry that belongs to the same class of technologies as stereo vision [2],[12]. In this method, a rigid set of two cameras is used, called a *stereo-camera*. As the relative position of cameras is fixed, it is possible to reconstruct partial point cloud of the scanned object directly from a stereo-image, simplifying the calculations time. This can be an essential advantage over SfM in some application area such as rapid modeling and 3D scenes prototyping in video games. Although, this approach requires using a tripod, it can become another advantage in some situations, such as insufficient illumination or hard-to-process surfaces, when it is possible to take several photos from exactly the same location in different lighting conditions [13].

3 The STERIO project

The goal of the STERIO project is to develop an efficient technology and programming tools for the effective development of 3D models using stereoscopic photography [13]. The results of the project will allow for faithful recreation of real world scenery in the virtual world of video games. Applying the proposed technology will also allow for more effective and less costly location generation for games which action takes place in real locations. Such approach will have a big advantage over the indirectly competitive 3D laser-scanning technology as it produces a similar effect with fewer resources and shorter production time. To facilitate operation, the technology will first be integrated with the most popular Unity 3D engine.

3.1 The STERIO data set

In the first stage of the STERIO project a comprehensive data set of images was prepared, allowing for further research in different aspects of 3D scanning of interiors for video games application. The whole STERIO dataset consists of over 10 thousand images of 25 various test locations and scenes that can be used for validation of developed algorithms. For each location, a different number of data subsets were prepared, consisting of images taken under different lighting conditions with different exposition parameters, camera location and orientation points, additional visual markers displayed, and other experiment setting. Some datasets are accompanied by additional images of a special chessboard pattern for cameras calibration. Sample images from the STERIO dataset are presented in Fig.1.

3.2 Reconstruction of smooth surfaces using displayed patterns

Reconstruction of flat, uniform and monochromatic surfaces with no landmarks, or salient points, is one of the main problems of photogrammetry. Such surfaces, which are often found in interiors (e.g. walls and ceilings), are most often not reproduced by existing applications and libraries supporting photogrammetry (e.g. [11]), resulting in the discontinuity, or a hole, in the object's grid. There are several approaches possible to solve this problem. It is possible to place some markers on such surfaces, which is rather inconvenient, time-consuming and do not guarantee the final success, as there are still no landmarks between the markers. Another approach is to manually model such improperly reconstructed surfaces. In the STERIO project, we propose a novel approach to overcome this issue and increase the completeness of automatically obtained 3D model. In this approach, a special pattern is displayed on such planes (Fig.1a), which provides details necessary for the proper scanning of the surface. This approach is not possible or at least not practical when using popular photogrammetric approach, as the displayed pattern overlaps the original texture of the surface. To avoid this undesirable effect, we propose

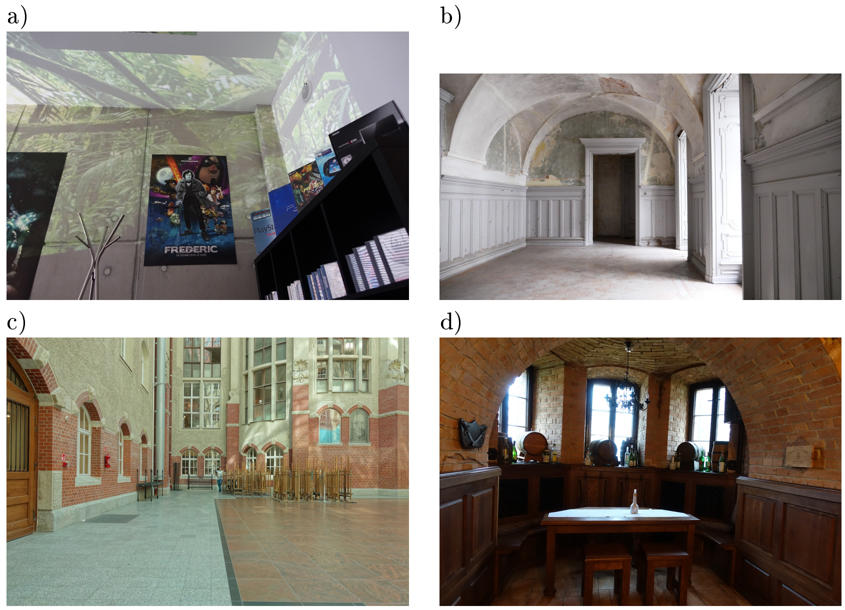

Fig. 1. Sample images from the STERIO dataset: modern interior (an office) with additional visual pattern displayed, a corridor from an abandoned manor, a university hall, a restaurant's cellar

to use the stereo-photogrammetry for such surfaces, in which two stereo-pairs are taken from the same stereo-kit position and orientation - one with the pattern displayed and one - without it. A set of images containing displayed pattern are used to reconstruct the interior's geometry while the other images are used to restore interior's original material, or texture.

Because stereo-photogrammetry with two cameras forces the use of a photographic tripod, in the applied approach there is no problem with maintaining the same position and orientation of both cameras and the fitting of photos from both sets, which is not possible when using individual photos in the photogrammetry process. There are two drawbacks of the presented approach. Firstly, it requires frequent switching on and off the LCD projector used to project a pattern, which can be easily omitted, e.g. by 'displaying' the black image. The second is the necessity of additional texture processing in order to restore their original colors. Fortunately, the proposed extension does not complicate or significantly extend the time of the whole process of interiors scanning. The initial results of performed experiments, presented in the next section, are very encouraging.

4 Experiments

Two experiments were carried out to validate the suitability of the use of stereo-photogrammetry in reconstruction of interiors for video game purposes. In the first experiment (4.1), several different interiors were reconstructed by the traditional modeling and the proposed approach. This experiment is pioneering on a global scale and no such comparison has been published yet. In the next experiment (4.2) several different patterns were tested to indicate which is the best one in this specific application.

4.1 Evaluation of the profits from using photogrammetric approach in modeling interiors

The goal of this experiment was to compare the time-complexity of modeling of the selected rooms using classic and photogrammetry-based approaches. The comparison made will allow estimating the possible benefits related to the development of the technology proposed within the project. As there are no such comparisons available, neither in the scientific nor industry literature, the results of this experiment are very important for further realization of the STERIO and similar projects.

In the case of the classic approach, models are created on the basis of a set of photos and basic distance measurements allowing for appropriate scaling of the model. In order to obtain the models based on photogrammetry, the Agisoft Photoscan software was used [11], which was considered as one of the best software available in the market at the moment. In both cases, the final stage of the work was to adapt the models to the requirements of the Unity environment and video games. In order to increase the reliability of the obtained results, the task was broken up into subtasks performed by different 3D graphic designers.

The experiment assumption was to model six varied interiors (Fig.2) for a video game purposes using both approaches and compare their total time and estimated costs. In the first approach, three computer graphic designers were supposed to traditionally model two different interiors each, using a series of photos and very simplified models as a scale reference. All geometry was modeled by hand and textures were created or adopted from available assets. Only room specific textures, such as paintings or frescos, were prepared based on photos. Final models were optimized for using within game engines and finally imported into Unity as separate interactive scenes.

In the second approach, each graphic designer was to model interiors, processed in the first stage by another one, using results of photogrammetry-based 3D scanning with Photoscan. As 3D scans received from Photoscan has numerous imperfections they cannot be directly used in video games. Graphic designers had to manually reduce complexity of the models mesh and textures, correct bad shapes, conceal defects and fill-in all mesh holes.

Fig. 2. Six interiors modeled within the experiment: Room3, Room4, Room1, SmallRoom, Parking, and the Corridor (finally not used in the comparison)

The final models should have similar complexity and quality as those created in the first approach.

Thanks to the detailed labor reporting, it was possible to compare the average effort needed to create usable 3D interior models using these two approaches. Unfortunately, Photoscan could not deal to successful recover one room, so finally five rooms could be used for a comparison. The received results indicate that the approach based on 3D photogrammetry scanning was faster by nearly 30%, in the average (Tab.1). Closer analysis shows that the advantage appears mainly for larger and more complex interiors. It is a positive fact as just such rooms consume most modeling time, what is confirmed by the weighted average profit of 41%. This rough comparison confirmed the thesis underlying the sense of the whole project. Moreover, the advantage of the proposed approach would be much higher and reach even 50%, if there existed tools aiding the 3D scan improving and post-processing. Such tools will be developed in the further stage of the STERIO project.

Table 1. The comparison of modeling time for the traditional and photogrammetry approach. Time given in hours [h]

	Room3	Room4	Room1	SmallRoom	Parking	Totally
Traditional approach [h]	180	103	95	18.5	13	409.5
Photogrammetry approach [h]	118	60	82.5	19	12	291.5
Time advantage	34.4%	44.8%	13.1%	-2.6%	7.7%	28.8%

4.2 Comparison of different patterns types for reconstruction of flat surfaces

The goal of this experiment was to compare and verify different types and classes of patterns that can be used to improve reconstruction of interiors using the proposed stereo-photogrammetric approach. In general, several classes of patterns have been identified, including photographs, semi-regular geometric motifs and pseudo-random patterns. For each class several sample images were chosen (Fig.3). For the experiment a stereo-photography kit were used equipped with a pair of high-resolution Sony A7R cameras. The experiment was carried out for two single-colored walls, one of which ended with a smooth, cylindrical column.

The experiment results were evaluated manually by a human expert. Such approach was sufficient, because the aim of the experiment was to estimate the suitability of individual classes of pattern, and not to accurately determine the errors of reconstruction of scanned surfaces. Analyzing the reconstruction results, one can draw the following conclusions:

- Regular geometric patterns (Fig.3a-b) do not allow to reconstruct the whole flat surfaces as particular landmarks are too similar one to each other and may be easily mistaken. Additionally, they add more or less regular moiré effect displacements that must be filtered out at post processing stage.
- Adding some color to pattern variations (Fig.3c-d) allows to reconstruct the whole flat surfaces in most cases but still does not allow for reconstruction of the curved surfaces. The moiré effect is still clearly visible.
- Pseudo-random motifs (Fig.3e-f) allow to reconstruct all surfaces still limiting the local displacements which has more random character that is easier to filter out.

5 Conclusion and future work

In this paper, a new approach was proposed to speed-up the reconstruction of real interiors for video game purposes. The unique experiment confirmed that using photogrammetric approach can partially automatize this important element of 3D game production pipeline and shorten the modeling time of more complex interiors. Although in many cases, a collection of individual photographs (photogrammetry) may be sufficient to reconstruct a room, serious problems with the reconstruction of flat uniform surfaces often arise. That is why, within the STERIO we propose the stereo-photogrammetric approach. The conducted experiment proved that proposed novel method is fully useful for scanning such problematic surfaces.

Future work includes further development of stereo-photogrammetry methods and a set of tools to maximize the automation level of the entire process with the options of manually editing and performing advanced post-processing algorithms. These tools will be verified and validated within the production of a commercial 3D horror video game, which action will take place in the interiors of an abandoned manor house modeled on existing historic building.

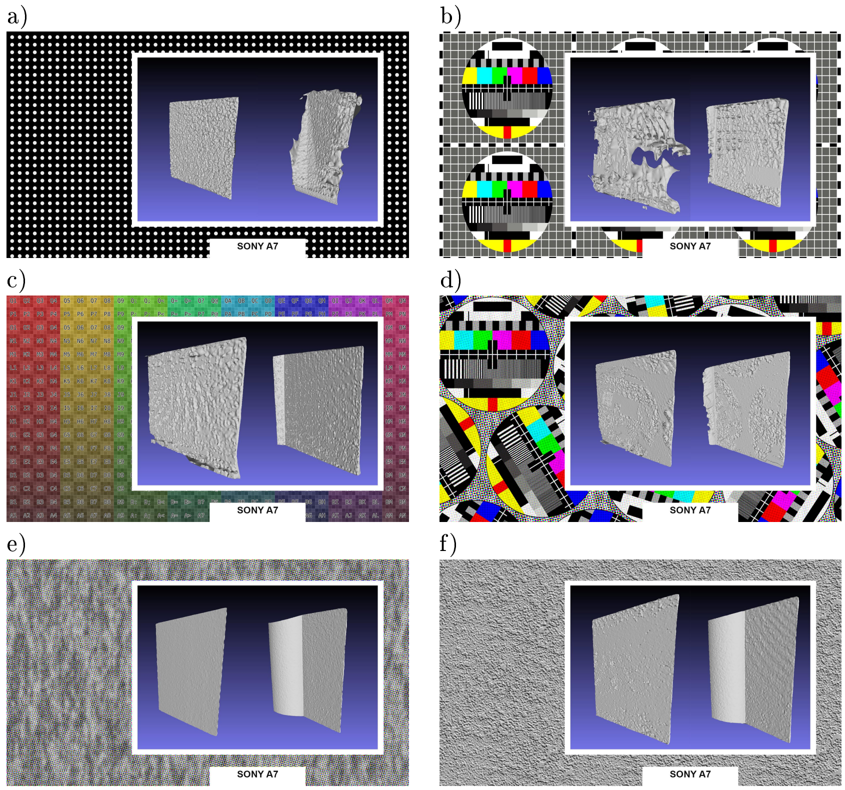

Fig. 3. Sample results of geometry reconstruction of uniform surfaces (a wall and a column) using different displayed pattern classes

6 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. Ahmadabadian AH et al. (2013) A comparison of dense matching algorithms for scaled surface reconstruction using stereo camera rigs. In: ISPRS Journal of Photogrammetry and Remote Sensing. 78:157–167
2. Kaczmarek AL (2015) Improving depth maps of plants by using a set of five cameras. Journal of Electronic Imaging 24(2):023018
3. Khoshelham K, Elberink SO (2012) Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. Sensors 12(2):1437–1454

4. Escorcia V, Dávila MA, Golparvar-Fard M, Niebles JC (2012) Automated Vision-Based Recognition of Construction Worker Actions for Building Interior Construction Operations Using RGBD Cameras. In: Construction Research Congress
5. Szwoch M, Pieniazek P (2016) Detection of Face Position and Orientation Using Depth Data. In: The 8th Intern. Conf. on Image Processing and Communications; Image Processing And Communications Challenges 7 Book Series: Advances in Intelligent Systems and Computing. Springer International Publishing 389:239–251
6. Balsa-Barreiro J, Fritsch D (2017) Generation of visually aesthetic and detailed 3D models of historical cities by using laser scanning and digital photogrammetry. Digital Applications in Archaeology and Cultural Heritage
7. Ha H, Oh TH, Kweon IS (2015) A multi-view structured-light system for highly accurate 3d modeling In: Int. Conf. on 3D Vision (3DV). IEEE 118–126
8. Kazmi W, Foix S, Alenya G (2012) Plant leaf imaging using time of flight camera under sunlight, shadow and room conditions. In: Int. Symp. on Robotic and Sensors Environments (ROSE). IEEE 192–197
9. Tse RO, Gold C, Kidner D (2008) 3D City Modelling from LIDAR Data, Advances in 3D Geoinformation Systems. Springer Berlin Heidelberg 161–175
10. Bay H, Ess A, Tuytelaars T, van Gool L. (2008) Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding 110(3):346–359
11. Agisoft LLC (2014) Agisoft PhotoScan user Manual
12. Kaczmarek AL (2017) Stereo vision with Equal Baseline Multiple Camera Set (EBMCS) for obtaining depth maps of plants. Computers and Electronics in Agriculture 135:23–37
13. Szwoch M, Kaczmarek AL, Bartoszewski D (2017) STERIO - Reconstruction of 3D Scenery for Video Games Using Stereo-photogrammetry. Computer Game Innovations in Monograph of the Lodz University of Technology In. Napieralski P, Wojciechowski A (eds) Lodz University of Technology, Lodz, Poland