

Image Based 3D Reconstruction of Texture-less Objects for VR Contents

Jahanzeb Hafeez¹, Seunghyun Lee¹, Soonchul Kwon¹, Alaric Hamacher^{1†}

¹Graduate School of Information Contents, Kwangwoon University, Seoul, Korea
{jahanzeb, shlee, ksc0226, [†]stereo3d}@kw.ac.kr

Abstract

Recent development in virtual and augmented reality increases the demand for content in many different fields. One of the fast ways to create content for VR is 3D modeling of real objects. In this paper we propose a system to reconstruct three-dimensional models of real objects from the set of two-dimensional images under the assumption that the subject does not have distinct features. We explicitly consider an object that is made of one or more surfaces and radiates constant energy isotropically. We design a low cost portable multi camera rig system that is capable of capturing images simultaneously from all cameras. In order to evaluate the performance of the proposed system, comparison is made between 3D model and a CAD model. A simple algorithm is also proposed to acquire original texture or color of the subject. Using best pattern found after the experiments, 3D model of the Pyeongchang Olympic Mascot “Soohorang” is created to use as VR content

Keywords: virtual reality, multi camera rig, 3D reconstruction, texture replacement, surface comparison

1. Introduction

With the emerging trend in virtual and augmented reality, a new wave has started in many different fields. This trend has influenced the entertainment industry with new applications and games based in virtual environment [1-2], medical and healthcare for the training of surgeons [3], cultural heritage for preserving the digital archive of museum objects [4], education [5], fashion [6] and engineering [7].

Where this emerging trend in VR and AR opened new door to many fields, however the demand for the content is also increasing. VR content from the real world can be categorized into three categories of 360 degree video capture, volumetric 3D capture and photogrammetry. In this work we mainly focus on photogrammetry and the term refers to the geometry acquisition of real objects.

Geometric acquisition approaches can be categorized into image-based and range-based techniques. Image based techniques can generate a precise 3D model in which 3D coordinates of objects are identified from a sequence of 2D overlapped images captured at arbitrary viewpoints. In computer vision, Structure from Motion (SfM) is most common technique that has been implemented in different software. In our previous work, we showed the 3D reconstruction of objects full of textures and unstructured surfaces [8]. This technique has been

improving rapidly in diverse applications such as in medicines for accurate investigation of scoliosis torso [9], in the industry for reverse engineering, and 3D printing of manmade and natural objects [10]. This technique usually drives to high quality results but it reaches its limitations when subjects under consideration are identical or textureless [11].

In this paper, we address the problem of reconstructing 3D model of the objects of unrecognizable texture and made of smooth surfaces, from a collection of cameras. This is one of the classical problems in computer vision and broad literature exists [11]. When the surface of the interesting object is smooth and weakly textured, the SfM step in the implemented algorithm designed to find correspondences between pairs of images, faces difficulties. It can be mentioned that if the surface of the object is smooth and texture-less, all of the algorithms based on feature point detection such as SIFT, will face difficulties to find key feature points [12]. In order to overcome this problem, several methods have been implemented such as structured light in which a predefined pattern is projected on an object and reflected patterns is compared with the predefined pattern to extract the depth information [13]. Similarly, a texture-full pattern can be projected on the object before capturing the images. Even though this technique improves the quality of model, but the original texture of the object hides under the patterns. We aim to introduce a new, simple and applicable system for 3D reconstruction of smooth and texture-less objects as well as recover the original texture or color of the object that can be further use in virtual world.

2. Related Work

Systems used for 3D model reconstruction can be listed into three classes comprises of those which work with active sensor (range-based systems) [14], passive sensors (image-based systems) [15] and those which work with both active and passive sensors [16]. However, range-based systems are more costly than image-based systems. Image-based systems are cheaper because they use normal cameras for image acquisition which are further used to produce 3D model [17].

In university of Calgary, a system was designed for the reconstruction of a torso surface. This system was made of several amateur cameras fixed on four metal arms. Four projectors were also used in order to provide artificial markers onto the surface of torso. Although all the cameras were operate simultaneously, however, they produce four different scans by each arm which further combined to build up entire surface through utilizing the overlap between them [9]. Even though the system was useful as compare to X-ray scanning but it is expensive due to use of several amateur cameras.

One of the low-cost and portable systems was designed by A. Hosseininaveh et al. for the surface reconstruction of small and texture-less objects. This system comprises of one camera and eight dot laser pointers to project patterns over the object in order to provide artificial markers. A stepper motor was used on which object is placed while capturing images [18]. Although, the system is low cost and portable but it is not expandable could be harmful for users due to use of strong lasers.

In current work, we develop a low-cost and automatic imaging system for the surface reconstruction of small and texture-less objects. Using multiple low-cost cameras and small mobile projectors in the system make it convenient, with minimum or no user intervention and hence lesser time of modeling process.

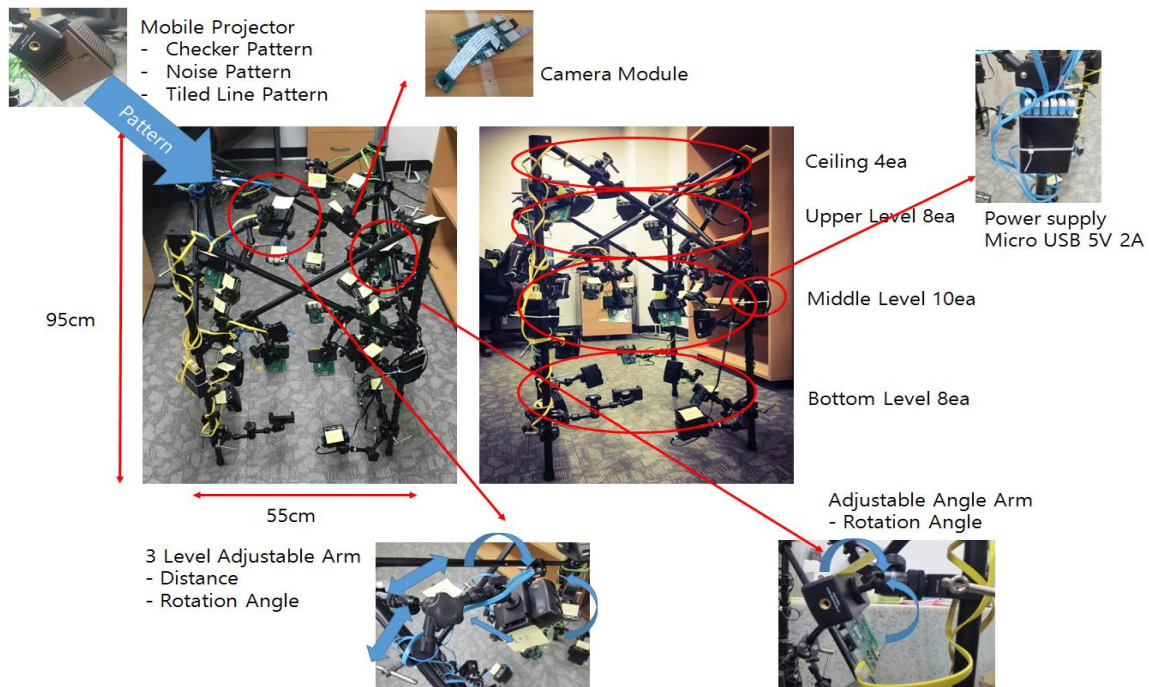


Figure 1. Multi camera rig setup and components used

3. Proposed System

In the proposed system a multi camera rig is designed for 3D modeling of texture-less and smooth objects. This rig has 30 Raspberry Pi 2 camera (5 megapixels) modules that are variable with respect to camera angle and distance. They are able to capture still images of size 2592 X 1944 and 1080p at 30 fps or 720p at 60 fps videos. Also, they have 53 degrees horizontal and 40 degrees vertical field of view. These cameras are connected with a central PC via Hub. For synchronization protocol, each camera is at signal waiting mode while central PC generates trigger signals. The Figure 1 shows the multi-camera network configuration diagram. A unique IP address is given to each camera from 192.168.0.101~130 and connects with central PC 192.168.0.100 via hub. Every time when computer generates trigger signals, each camera capture an image and store it into camera module memory which further transferred to central PC via FTP.

Cameras distribution over rig is composed of 4 ceiling part, 8 upper, 10 middle and 8 lower levels as in the Figure 1. Among these, 18 cameras are attached with 3-level variable and 12 ball-type free angle which enables the angle and distance control to face the object. Five electronic power supplies of 5V/2A are providing power to each camera by Micro USB port.

Two small mobile projectors are used, that also support wireless connection to project pre-computed patterns (again see Figure 1), over texture-less and smooth object. These projectors aim to provide artificial markers through projection of patterns onto the surface of the object. This is necessary because the surface of the object is homogeneous and without artificial markers it would be impossible to find distinct feature points in the captured pairs of images.

4. Methodology

By utilizing this system, the obtained data is processed to generate 3D model of the object. This is done by first detecting feature points from every image and matching conjugate features in the corresponding image

pairs. After that, the matched features are traced through all the overlapping images and used as tie points which generate a point cloud relative to the position and orientation of the points. Then, the obtained point cloud is used for triangulation and mesh is generated as a result. We compared the generated mesh with Computer Aided Design (CAD) model to evaluate the performance of the proposed system. At the end, we replace the texture of obtained mesh that is generated by artificial markers with the original texture of the object. The Figure 2 shows the stepwise breakdown of the implemented method.

4.1 Feature detection using SIFT

The detection of distinct features of the surface is a prerequisite for image matching. In the proposed method, these features are extracted using SIFT algorithm, which is a feature point detector [12]. The detected features describe the image rotation, image scale, noise, exposure and contrast changes. Features extracted from each image are saved into a feature descriptors that are further used for feature matching.

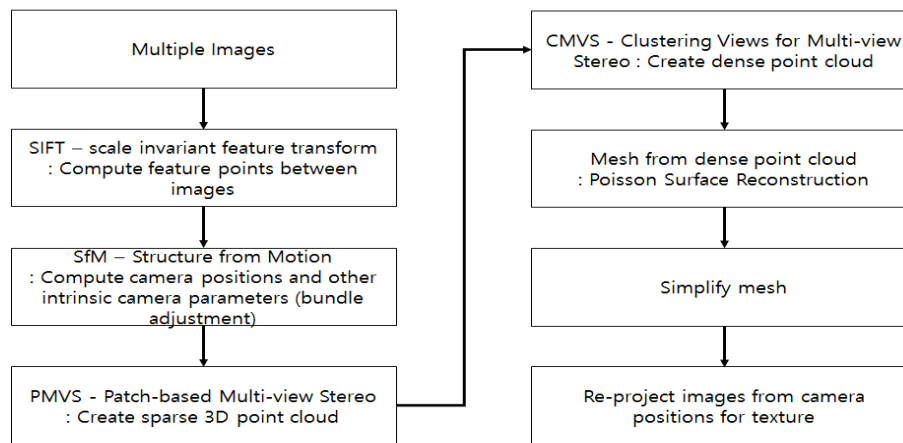


Figure 2. The proposed procedure for 3D surface reconstruction of a texture-less object

4.2 SfM - Feature matching and Bundle Adjustment

Feature matching and Bundle Adjustment are crucial elements in SfM (Structure from Motion). The feature matching algorithm implemented in our methodology used the preemptive matching technique that accelerates the matching process and reduces the reasonable amount of time [19-20]. It arranges all the features of each image into decreasing scale order and enlists all the pairs to be matched.

4.3 Dense point cloud reconstruction

Dense reconstruction of the point cloud is performed using either Patch-based Multi-view Stereo (PMVS) [21] or Clustering Views from Multi-view Stereo which are able to reconstruct global point clouds.

4.4 Mesh reconstruction

Once a dense point cloud is generated, mesh reconstruction is performed using Poisson mesh [22] that reconstructs surfaces of the object from oriented points. This algorithm creates a triangular mesh that interpolates all or most of the points.

4.5 Mesh simplification

The purpose of mesh simplification is to provide better 3D model of the object by removing the unwanted vertices and faces of the mesh.

4.6 Texture replacement

Sometimes we may want to acquire the original texture of an object of interest but when artificial markers are used through projection of patterns, original texture of the object vanishes under it. In our proposed system, we design an algorithm that replaces the images before texture mapping of the meshed model reconstructed with patterns projection to acquire the original texture of the object. It is done by capturing the images of the surfaces of the object without patterns of the same perspective, scale of the images and distance at which images were captured with patterns. It become possible with proposed multi camera rig system that is stable enough because of no user intervention during image capturing process. In order to ensure the precision with respect to image scale, distance and field of view, difference between image sets, with and without patterns is calculated. After that the images in data sets with and without patterns are interchanged. Once all the images interchange, texture mapping is performed with image set captured without artificial markers on the mesh model reconstructed with patterns projection using [23].

5. Comparison and Evaluation

We reconstruct 3D models from the sequence of images from different data sets. Four different patterns are used to acquire different data sets and their corresponding models (Figure 4) reconstructed with the images captured under the projection of patterns are showed in Figure 3.

Defining an evaluation scheme is a difficult part focusing on the accuracy of the reconstructed models. We made comparison between each created mesh model to the point cloud of the ground truth data (CAD model).

The 3D model comparison pipeline consists of the open source software Meshlab [24] and CloudCompare [25] for all 3D operations. A rough direction and size alignment of ground truth data is performed based on the point clouds of each created model in the Meshlab. At current stage, each created model along with ground truth data is loaded into CloudCompare. Then, each model is registered with the ground truth using the iterative point algorithm (IPC) by Besel and McKay [26] with a target error difference of 1×10^{-8} . For each registered model the minimum distance between every point to any triangular face of the meshed model is calculated. Note that, the created model is set as reference because the comparison is made between each created model to the point cloud of the ground truth. We show the results of comparison and distance computed alongside standard deviation in Table 1. The 3D models of smooth and texture-less object generated from multiple images are compared with ground truth to evaluate the performance of the proposed system. In Table 1, it is clear that the model reconstructed with images captured under the projection of tiled lines pattern exhibits exceedingly few deviations to the ground truth followed by square pattern and noise pattern. The model generated under the projection of checkerboard pattern shows maximum deviations.

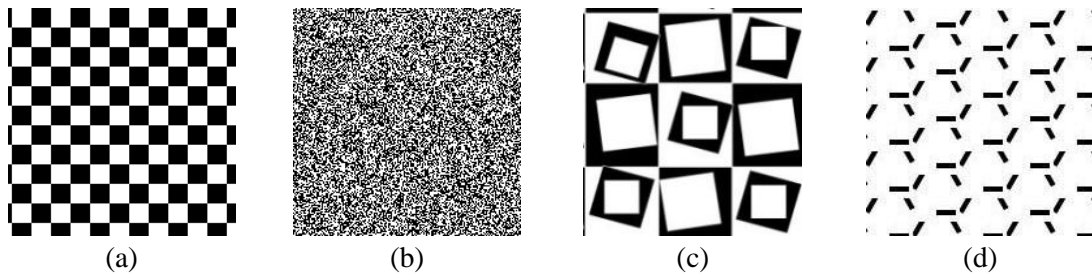


Figure 3. Test patterns namely (a) checkerboard, (b) noise, (c) squares and (d) tiled lines used in capturing process

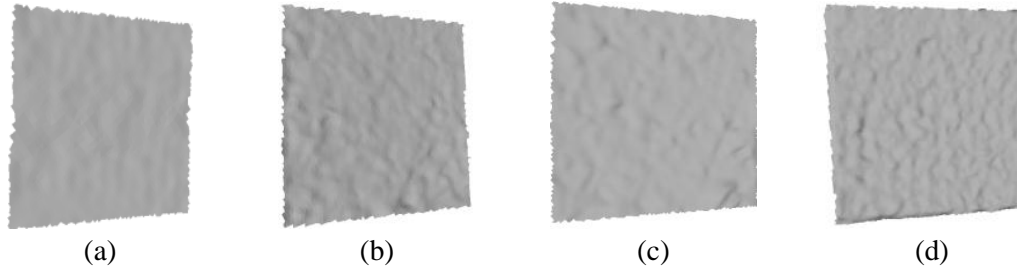


Figure 4. 3D models of a smooth and texture-less object generated with (a) checkerboard pattern, (b) noise pattern, (c) square pattern and (d) tiled lines pattern

6. Result and Discussion

The 3D models of smooth and texture-less object generated from multiple images are compared with ground truth to evaluate the performance of the proposed system. Figure 5 shows the histogram of each model reconstructed to represent the distance distribution between each generated model and ground truth. Distance distribution for each created model is colorized in the scheme blue-white-red. In this color scheme, blue color represents the points of created models which are under the surface of ground truth while red color represents the points which are above the surface of ground truth. White color represents the points which are perfectly aligned to the surface of ground truth. The histogram plots also make clear that the model reconstructed with images captured under the projection of tiled lines patterns exhibits exceedingly few deviations to the ground truth followed by square pattern and noise pattern. The model generated under the projection of checkerboard pattern shows maximum deviations. In Table 1, we also calculated the mean value and standard deviation of distance distribution for each created model.

Table 1. Mean value and standard deviation computed for each created model to the ground truth data.

<i>Pattern Type</i>	<i>Mean Distance (μ)</i>	<i>Standard Deviation (σ)</i>
Checkerboard	3.8×10^{-5}	3.01×10^{-3}
Noise	1.1×10^{-5}	1.91×10^{-3}
Squares	2.3×10^{-5}	0.81×10^{-3}
Tiled Lines	0.4×10^{-5}	0.77×10^{-3}

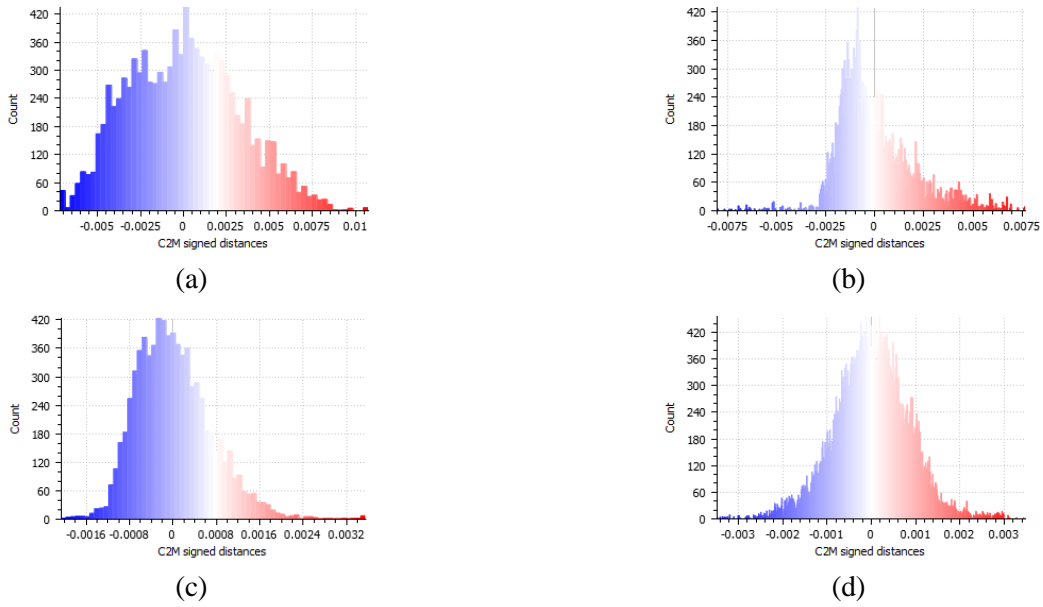


Figure 5. Histogram plots of the distance distribution between the created models and the ground truth.

7. Making of VR Contents

Previously, from comparison and evaluation we find that the model reconstructed with dataset acquired under the projection of tiled lines pattern provides minimum deviation towards ground truth. Further, we consider this pattern to reconstruct 3D model of the Pyeongchang Olympic Mascot “Soohorang” because the surface of the Olympic Mascot is smooth and weakly textured. We also obtain another dataset with no projection to compare the resultant model with first model that is reconstructed with the dataset acquired under the projection of tiled lines pattern. Moreover, this dataset is also used to obtain the original texture of the Olympic Mascot. In Figure 6, difference is clear between two models. In Figure 6(b) model reconstructed with images acquired under the projection of pattern, is more accurate and complete. However, model in Figure 6(a) is inaccurate and useless which is created with the images without projection. We texture the second model that is more accurate and complete with the second dataset using the proposed algorithm (Figure 6(c)). Further, textured model is imported into Unity game engine to create VR content as shown in Figure 7.

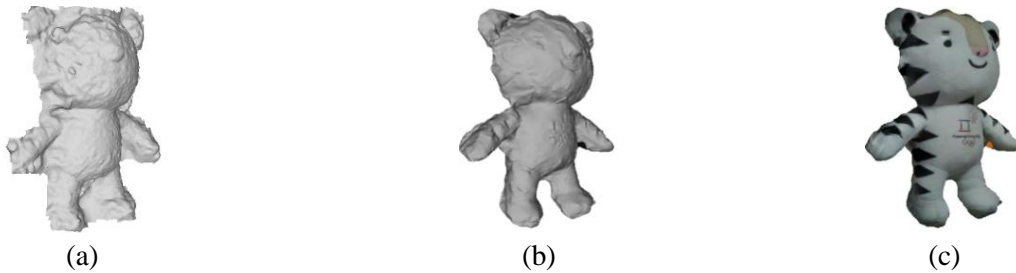


Figure 6. 3D models of Olymic Mascot reconstructed using proposed system. (a) without projection, (b) with projection and (c) textured model using images captured without projection.

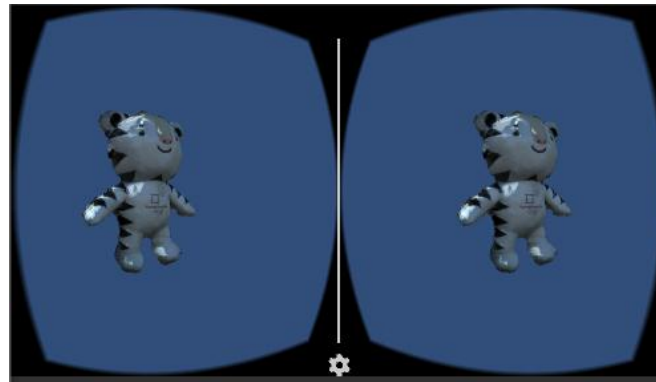


Figure 7. 3D model of Olympic Mascot in the virtual environment.

8. Conclusion

VR content is rapidly improving to attract the consumer market. In present study, we proposed a multi camera rig system that is capable of capturing images simultaneously from all cameras. Especially, the system is designed for 3D modeling of smooth and texture-less surfaces from a number of images. The projection of patterns makes possible the reconstruction of 3D model of selected subject. We show the accuracy of system by comparing the reconstructed model with the ground truth CAD model. The 3D model reconstructed by data set acquired under the projection of tiled lines pattern exhibits minimal mean distance deviation to the ground truth. We show a simple algorithm by means of which original texture of the object can be acquired using proposed system.

We also found that the model of Pyeongchang Olympic Mascot “Soohorang” created with tiled patterns is complete and accurate as compare to the model created with images having no projection. Further, this model is used to create VR content using Unity game engine.

Acknowledgement

“The present Research has been conducted by the Research Grant of Kwangwoon University in 2016”

References

- [1] A.S.R. Junior, E. C. W. Gonzales, L. Valente, and B. Feijó, “First steps towards Live-action Virtual Reality Games,” *SBC J. Interact. Syst.*, Vol. 7, No. 1, pp. 3–16, November 2016.
- [2] J. Torrente, B.Á. Del, Á. Serrano-Laguna, J.Á. Vallejo-Pinto, P. Moreno-Ger, and B. Fernández-Manjón, “Towards a low cost adaptation of educational games for people with disabilities,” *Comput. Sci. Inf. Syst.*, Vol. 11, No. 1, pp. 369–391, January 2014.
- [3] A. Hamacher, S.J. Kim, S.T. Cho, S. Pardesahi, S.H. Lee, S.J. Eun, T.K. Whangbo, “Application of Virtual, Augmented, and Mixed Reality to Urology,” *Int. Neurourol. J.*, Vol. 20, No. 3, pp. 172–181, September 2016.
- [4] F. Bruno, S. Bruno, G. De Sensi, M.-L. Luchi, S. Mancuso, and M. Muzzupappa, “From 3D reconstruction to virtual reality: A complete methodology for digital archaeological exhibition,” *J. Cult. Herit.*, Vol. 11, No. 1, pp. 42–49, March 2010.
- [5] E. Madrigal, S. Prajapati, and J. C. Hernandez-Prera, “Introducing a Virtual Reality Experience in Anatomic Pathology Education,” *Am. J. Clin. Pathol.*, Vol. 146, No. 4, pp. 462–468, October 2016.

- [6] F. Carroll, "Designing (for) experiences in photorealistic VR environments," *New Rev. Hypermedia Multimed.*, Vol. 16, No. 1–2, pp. 181–194, April 2010.
- [7] S. Aromaa and K. Väänänen, "Suitability of virtual prototypes to support human factors/ergonomics evaluation during the design," *Appl. Ergon.*, Vol. 56, pp. 11–18, September 2016.
- [8] J. Hafeez, A. Hamacher, H. Son, S. Pardeshi, and S.H. Lee, "Workflow Evaluation for Optimized Image - Based 3D Model Reconstruction," in the *Proc. of International Conference on Electronics, Electrical Engineering, Computer Science (EEECS) : Innovation and Convergence*, vol. 2, pp. 62–65, 2016.
- [9] I. Datchev, A. Habib, and Y.-C. Chang, "Image Matching and Surface Registration for 3D Reconstruction of a Scoliotic Torso," *GEOMATICA*, Vol. 65, No. 2, pp. 175–187, June 2011.
- [10] J. Schöning and G. Heidemann, "Evaluation of Multi-view 3D Reconstruction Software," in *Proc. Of 16th International Conf. on Computer Analysis of Images and Patterns*, pp. 450–461 2015.
- [11] C. Santagati, L. Inzerillo, and F. Di Paola, "Image-Based Modeling Techniques For Architectural Heritage 3d Digitalization: Limits And Potentialities," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, Vol. XL-5, No. W2, pp. 555–560, September 2013.
- [12] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, Vol. 60, No. 2, pp. 91–110, November 2004.
- [13] M. Maruyama and S. Abe, "Range sensing by projecting multiple slits with random cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 15, No. 6, pp. 647–651, June 1993.
- [14] S. Mahadevan, H. Pandzo, M. Bennamoun, and L. Williams, "A 3D acquisition and modelling system," in *Proc. Of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, pp. 1941–1944, 2001.
- [15] S.Y. Chen and Y.F. Li, "Automatic sensor placement for model-based robot vision," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, Vol. 34, No. 1, pp. 393–408, February 2004.
- [16] S. Larsson and J.A.P. Kjellander, "Path Planning for Laser Scanning with an Industrial Robot," *Robot Auton Syst.*, Vol. 56, No. 7, pp. 615–624, July 2008.
- [17] S. Paulus, J. Behmann, A.-K. Mahlein, L. Plümer, and H. Kuhlmann, "Low-Cost 3D Systems: Suitable Tools for Plant Phenotyping," *Sensors*, Vol. 14, No. 2, pp. 3001–3018, February 2014.
- [18] A. Hosseininev, R. Yazdan, A. Karami, M. Moradi, and F. Ghorbani, "A Low-Cost and Portable System for 3D Reconstruction of Texture-Less Objects," *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, Vol. 15, pp. 327–332, December 2015.
- [19] C. Wu, "Towards Linear-Time Incremental Structure from Motion," in *Proceedings of the 2013 International Conference on 3D Vision*, pp. 127–134, 2013.
- [20] C. Wu, S. Agarwal, B. Curless, and S. M. Seitz, "Multicore Bundle Adjustment," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3057–3064, 2011.
- [21] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust Multiview Stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 32, No. 8, pp. 1362–1376, August 2010.
- [22] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," in *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, pp. 61–70, 2006.
- [23] Y. Wang, W. Wang, and E. Wu, "Optimizing the Parameters for Patch-based Texture Synthesis," in *Proceedings of the 2006 ACM International Conference on Virtual Reality Continuum and Its Applications*, pp. 75–82, 2006.
- [24] MeshLab, (version 1.3.4), <http://meshlab.sourceforge.net/>.
- [25] CloudCompare, (version 2.7.0), <http://www.cloudcompare.org/>.
- [26] P.J. Besl and H.D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, February 1992.