Single-lens Device for Stereoscopic Modelling

Radoslaw HOFMAN¹, Krzysztof KOWALAK², Lukasz KAMINSKI², Pawel GARDZINSKI² and Slawomir MACKOWIAK²

¹3D Vision sp. z o.o., ul. Piotrowo 3, 61-138 Poznan, Poland radekh@3d-vision.com.pl
²Poznan University of Technology, Poland

Keywords: 3D modeling, 3D scanning, Mobile device.

Abstract. In this paper the authors present an innovative approach which may be used for 3D-modelling of real-world objects for the purpose of virtual reality, 3D printing etc. The main idea is based on the matching characteristic points between images made by single lens device (for example a mobile). The method brings fully automated solution allowing to be used without calibration method. The experimental results show that although the modelled objects contain some distortions, the final result is suitable for non-professional usage.

Introduction

3D printing, virtual reality and a stereoscopic video content creating were became very popular in the last years. Typically, the users of 3D printers and VR use 3D models prepared by specialists or scanned with sophisticated equipment for the purpose of prototyping processes, simulation, testing of arrangements of the scene in VR etc. The preparation of 3D models is however still quite expensive—it requires skills or expensive tools (or both), therefore this technology is almost out of reach for regular users. It may be assumed, that a solution to this problem would open new possibilities for development of 3D technology and general usability of modern technologies in everyday life.

In this paper the authors present their proposal addressing the problem mentioned above: 3D reconstruction and modelling of objects with the usage of commonly used single-lens device (mobile or camera). The approach utilizes images acquired with such a device and delivers complete, fully automated algorithms which discover the 3D model from images.

In the literature there may be found some approaches to the problem—some authors propose 3D object reconstruction and modelling based on depth maps estimated on two or more images of the object [7,8]. Other authors use the approach of matching certain characteristic features of the object on images [9,10] or use a volumetric reconstruction techniques which are based on multiple images of the object [11,12].

Some of the above approaches appear to be too complex to be used on regular devices with an expectation of reasonable computation time (for example the approach with depth maps). The process may be supported by dedicated hardware (for example cameras with structured light), however this approach is against the idea of creating low cost and widely available solution. The approach based on volumetric reconstruction also appeared to give unsatisfactory results when the authors were using regular (low cost) cameras. Therefore the main approach for the solution presented in this paper is based on characteristic feature points recognition and matching between images. Furthermore, for the purpose of images correction and spatial orientation typical approaches utilize kind of marker for images calibration. Such a requirement would limit the possibility to use the solution in any situation (without carrying the marker). Therefore the authors has decided calibration method not requiring the marker. The solution is presented in details in the next section.

The approach and application on one type of devices was presented by the authors in 2015 [13], however in this paper the more general approach to the problem is presented.
Proposed Solution

The solution proposed in this paper offers 3D model reconstruction using a single-lens device. The object has to be photographed from different angles or recorded on a movie. The images captured are then used to determine the spatial information about the images themselves and the object. The step of the proposed approach is presented in Figure 1.

![Figure 1. General steps of the proposed approach [13]](image)

The first stage is understood as a technical process of making pictures of the object from different angles to obtain maximum coverage of views. The next, more complex phase is related to feature extraction as in this phase the system needs to estimate spatial information about the scene. There are many different approaches to camera calibration [1] based on marker with known geometry or without the marker. The second group of approaches appears to be more complex, as the spatial information may only be obtained from the images and camera information. However, the authors have decided to use marker less method to aim for easy to use solution. The adopted method is based on SIFT points [2], SIFT points which are expected to be stable in case of scale, orientation and change of illumination.

![Figure 2. Example of the identification of SIFT points](image)

The next stage of the algorithm is focused on matching feature points between each two views. This approach is more complex than chain matching but it provides better resistance level for case of impossibility of the matching of an adjacent pair of views. The point matching technique is based on the comparison of the descriptors of these points. To avoid incorrect matching the authors have used fundamental matrix, which is calculated for each two images in RANSAC [3] approach.

![Figure 3. SIFT points matching](image)

After the images were compared the system may finally determine the relative position of cameras between these images. The calculation utilizes the fundamental matrix of points and also technical information about camera, which has to be known accurately. The solution assumes that the camera is identical for each picture (all are made with the same device) therefore all distortions which are the result from imperfection of the device are the same on each image. Under this assumption the problem of camera characteristics determination may be re-formulated as an error minimization problem and
may be solved with the usage of Levenberg - Marquadt [4] algorithm. The algorithm proposed by the authors starts with single pair of images which should have a relatively high number of matching feature points. Then a DLT[5] and RANSAC[3] algorithms are executed which result with determined camera parameters. The process may be repeated for other pairs of images in case of symptoms, that the parameters were identified incorrectly.

The next stage of processing is called dense model reconstruction and is based on SIFT points and matrices calculated in the previous steps. The main goal of this step is to increase model coverage by expanding cloud of points to with patches. Each point identified in the previous steps is assigned a patch of \( \mu \times \mu \) size, where \( \mu \) is equal 5 or 7 and a vector directed on outside of the model which represents the spatial position of the patch. In the next step the patch is expanded to fill the space between closest patches. After this step the filtering process is started where the created dense model is compared with the initial images to verify location of the patches. When the filtering is finished the surface may be recreated.

The experiments have shown that the model may also contain some additional distortions and also the model requires surface map to be reconstructed. The first step of post processing is based on SOR [6] filtering which is based on stochastic analysis of the cloud of points and removal of those, which do not meet criteria (which is based on the calculation of distance between points and application of the corollary, that distribution of the distance should appear as Gaussian). After the outstanding points are removed the algorithm reconstructs the surface. This process is complex, as the quality of input images is assumed to be low. Surface reconstruction is based on Poisson [7] algorithm – the method is well enough resistant to noise in cloud of points as it is considering all points simultaneously.

**Experimental Results**

The algorithm presented in this paper is utilizing certain approaches to create low-cost and easy to use solution based on single-lens device. The approach was implemented and tested in several scenarios based on the images acquired on typical cameras (among other: mobile phone cameras). The sequences were prepared as two sets: series of photos and a video sequences with varying length (verifying accuracy of the method between high and low density scans).

![Figure 4. Example of reconstructed surface from point cloud](image)

One of the main assumption for the work was to prepare solution available to be used without specialist knowledge and without expensive devices. As it was already described in the previous section, the proposed solution does not require any calibration or configuration by the user.

The reconstruction effectiveness was evaluated on the several objects’ reconstructions which were assessed for quality on the basis of pictures rendered from the models. One of such results is presented in Figure 4.

The proposed system has brought promising results in terms of easiness and accuracy of 3D models recreation. The authors are convinced that the quality is sufficient for non-professional applications, however the quality of the end result is mainly dependent from the quality of acquiring process, what leads to conclusion that professional camera may increase the quality of the model. On the other hand the quality of typical devices is increasing over time, therefore it may be expected that the quality of results from the presented algorithm will increase. However even with typical mobile phone camera
the quality is good enough for 3D printing or VR without requiring any 3D modelling skills from the user.

Conclusions
This paper the authors have presented an innovative approach for easy to use 3D model reconstruction algorithm based on typical single-lens devices. All calculations of the algorithm are conducted automatically so one of the main advantages of proposed approach is that user does not need to have any technical knowledge about 3D modelling and also does not need to carry any calibration marker.

The algorithm ends with recreation of complete 3D model of scanned object while the experimental results confirm that the quality of such models is satisfactory in terms of 3D printing or VR usage. In the future the algorithm may be further developed to be optimized (create 3D models in real-time) and utilize cameras with improved parameters.

Acknowledgement
The project Innovative stereoscopic webcam with dedicated software for image processing prototype construction is the project co-founded by European Union (project number: UDA-POIG.01.04.00-30-018/10-03, realization period: 02.2012-06.2014). European funds for the development of innovative economy (Fundusze Europejskie - dla rozwoju innowacyjnej gospodarki).

References
[1] Cyganek B (2013) Object Detection and Recognition in Digital Images: Theory and Practice, Wiley.
[2] Lowe David G. (1997) Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 60, 2:91-110.
[3] Fischler M., Bolles R. (1987) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Readings in computer vision: issues, problems, principles and paradigms, 726 740.
[4] Nocedal J., Wright S. (1999) Numerical Optimization, Springer Verlag, New York.
[5] Hartley R., Zisserman A. (2003) Multiple view geometry in computer vision, Cambridge University Press.
[6] Point Cloud Library (PCL) Documentation http://dev.pointclouds.org/projects/pcl/wiki
[7] Seitz S.M., Curless B., Diebel J., Scharstein D., Szeliski R. (2006) A Comparison and Evaluation of Multi-View Stereo Re-construction Algorithms. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 519-528.
[8] Xiao J., Chen J., Yeung D.-Y. and Quan L., (2008) Learning twoview stereo matching, In Proceedings of the 10th European Conference on Computer Vision: Part III, ECCV 08, Springer-Verlag, Berlin, Heidelberg, pp. 15-27.
[9] Snavely N., Seitz S. and Szeliski R., (2006) Photo tourism: exploring photo collections in 3d, In ACM Transactions on Graphics (TOG), volume 25. ACM, pp. 835-846.
[10] Snavely N., Seitz S. and Szeliski R., (2008) Modeling the world from internet photo collections", International Journal of Computer Vision, 80(2): pp. 189-210.
[11] Goldlcke B., Magnor M., (2003) Real-time, free-viewpoint video rendering from volumetric geometry, In Proc. Visual Computation and Image Processing (VCIP), Lugano, Switzerland, 5150(2): pp. 1152-1158.
[12] Baker H., (1977) Three-dimensional modelling, In Proc. of the 5th International Joint Conference on Artificial Intelligence, Cambridge, MA, USA, pp. 649-655.

[13] Kowalak K., Kaminski L., Gardzinski P., Mackowiak S., Hofman R., (2015) 3-D Reconstruction of Real Objects Using an Android Device in Choras R (ed) Image Processing and Communications Challenges 7, Volume 389 of the series Advances in Intelligent Systems and Computing pp. 135-142, DOI: 10.1007/978-3-319-23814-2_16, Springer International Publishing.