Mathematics for Computer Vision (MCV 2018)
February 15-16th 2018, Trento, ITALY
Demo Session 2: 3D Reconstruction

3D reconstruction in your pocket

Fabio Poiesi and Paul Chippendale

February 16th, 2018

Abstract We present a pipeline to create digital 3D replicas of real-world objects using off-the-shelf smartphones. Our methodology uses a hybrid approach where keyframes from a smartphone's camera are automatically selected and sent to a server for 3D reconstruction. The smartphone periodically receives back 3D model updates for progress visualisation, thus users can concurrently and collaboratively perform acquisitions and speed up the reconstruction of large objects, e.g. buildings. We will show examples of 3D reconstructions performed by both individuals and groups.
Keywords: Image filtering, Structure from Motion, Epipolar Geometry, Optimisation.

1 Introduction

The minutuarisation of imaging devices and the increasing power of embedded computational hardware have led to an explosion of new applications (apps) that use computer vision on smartphones [6]. Smart photography [3], biometrics [1], Augmented Reality [4], Virtual Reality [8] and 3D reconstruction [7] are just a few examples of these applications.

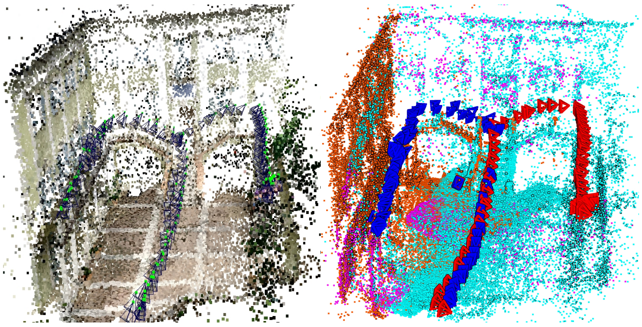
In this demo, we will show a client-server 3D reconstruction pipeline in action. Visitors will be able to try the pipeline for themselves and use *Replicorder* to digitise real-world objects.

Fabio Poiesi and Paul Chippendale
Technologies of Vision, Fondazione Bruno Kessler, Povo (TN), Italy. E-mail: {poiesi, chippendale}@fbk.eu.

2 Overview

Replicorder is a hybrid approach where image acquisition, reconstruction and visualisation are split across a smartphone and remote server. Such a hybrid approach enables users to collaboratively reconstruct real-world objects and receive feedback about global reconstructions. Replicorder acquires images from a smartphone's camera and automatically selects a subset of them based on both their quality and on their novelty [7]. These selected images are uploaded to a reconstruction server that progressively creates 3D models. By transmitting only some of the images, we can reduce bandwidth demands between smartphone and server. Replicorder generates haptic feedback each time a new image is chosen. When a user is scanning an object, the Replicorder app receives near real-time feedback about the status of the reconstruction from the server, enabling users to focus on parts of the scanned object that deserve more attention. Such a client-server architecture produces high-quality 3D models without placing high resource demands on the smartphone.

3D model creation on the server starts with Structure from Motion (SfM) [7]. SfM uses multiple threads to process the independent and asynchronous uploading of images from different users. When SfM is performed on images from a single device, camera intrinsic parameters are typically assumed to be the same for all images with Bundle Adjustment [9] being used to estimate these parameters (fixing them globally whilst estimating extrinsic parameters and minimising the reprojection error). In the case of images acquired from different devices, Bundle Adjustment uses instead the concept of 'intrinsic groups', where parameters from images from a specific camera are assumed to be fixed. After the images have been oriented, multi-view stereo



(a) (b)

Fig. 1 Sparse point clouds resulting from the concurrent acquisition of Villa Tambosi in Trento, Italy. (a) Sparse point cloud shown together with the oriented images and (b) colour-coded sparse point cloud showing the contribution of the two users. The 3D points triangulated by the blue smartphone (Sony Z5) are coloured cyan, those triangulated by the red smartphone (LG Nexus 5X) in orange and those triangulated by both in magenta.

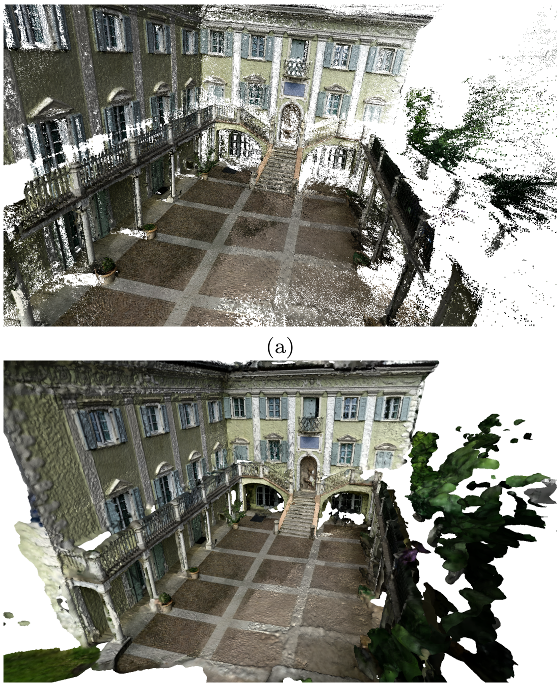
is performed to generate a dense point cloud [5]. Multi-view uses the concept of photo-consistency between sets of input images and verifies agreements in terms of illumination, materials and 3D geometry. The final mesh is obtained using Poisson reconstruction computed on the dense point cloud [2]. Created 3D models are stored on a reconstruction server and can be downloaded via a RESTful API[1].

3 Collaborative 3D reconstruction

Fig. 1 shows the sparse point cloud of Villa Tambosi in Trento (Italy) collaboratively created by two users. Fig. 1a shows the global point cloud of the reconstructed Villa Tambosi in addition to the positions of the cameras (green points with projected images). Fig. 1b shows the 3D point cloud with the colour-coded cameras representing each smartphone's pose during acquisition. The 3D points triangulated from the blue smartphone's images are coloured cyan, those triangulated from the red smartphone's are coloured orange, and those triangulated by both are coloured magenta. The total number of 3D points in this example is 71816, of which 39768 (55%) were triangulated from solely the blue smartphone, 25398 (35%) were triangulated from solely the red smartphone and 6650 (9%) were triangulated from images from both devices.

Acknowledgements: This research has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 687757.

[1] `http://www.restapitutorial.com/`, last accessed: Dec 2017.



(a)

(b)

Fig. 2 3D reconstruction performed with two smartphones that concurrently acquired images of Villa Tambosi in Trento, Italy: (a) Dense point cloud and (b) Mesh.

References

1. M.O. Derawi, D. Yang, and C. Busch. Fingerprint recognition with embedded cameras on mobile phones. In *Proc. of Security and Privacy in Mobile Information and Communication Systems*, Aalborg, DK, May 2011.
2. M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proc. of Eurographics Symposium on Geometry Processing*, Cagliari, IT, Jun. 2006.
3. J. Lin and C.-K. Yang. Collaborative panoramic image generation from multiple mobile phones. In *Proc. of Multimedia & Expo Workshops*, Hong Kong, CN, Jul. 2017.
4. H. Liu, G. Zhang, and H. Bao. Robust Keyframe-Based Monocular SLAM for Augmented Reality. In *IEEE International Symposium on Mixed and Augmented Reality*, Merida, MX, Sep. 2016.
5. A. Locher, M. Perdoch, and L.V. Gool. Progressive prioritized multi-view stereo. In *Proc. of Computer Vision and Pattern Recognition*, Las Vegas, US, Jun. 2016.
6. O. Muratov, Y. Slynko, V. Chernov, M. Lyubimtseva, A. Shamsuarov, and Victor Bucha. 3dcapture: 3d reconstruction for a smartphone. In *Computer Vision and Pattern Recognition Workshops*, Las Vegas, US, Jun. 2016.
7. F. Poiesi, A. Locher, P. Chippendale, E. Nocerino, F. Remondino, and L. Van Gool. Cloud-based collaborative 3d reconstruction using smartphones. In *Proc. of European Conference on Visual Media Production*, London, UK, Dec. 2017.
8. N. Pretto and F. Poiesi. Towards gesture-based multi-user interactions in collaborative virtual environments. In *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, pages 203–208, Hamburg, DE, Nov. 2017.
9. B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. *Bundle adjustment - a modern synthesis.* Vision Algorithms: Theory and Practice, Springer-Verlag, Berlin, GE, 2000.