Mobile reconstruction and exploration of indoor structures exploiting omnidirectional images

Giovanni Pintore CRS4 *
Fabio Ganovelli ISTI-CNR †
Enrico Gobbetti CRS4
Roberto Scopigno ISTI-CNR

Figure 1: *Starting from panorama images generated with the aid of commodity mobile devices, our methods rapidly compute visual and structural 3D models of indoor environments scaled to their metric dimensions. The resulting 3D floor plans of multi-room environments can be used for a variety of planning applications, while the visual models provide location awareness within mobile application specifically developed for touch interaction on smartphones and tablets.*

Abstract

We summarize our recent advances in acquisition, reconstruction and exploration of indoor environments with the aid of mobile devices. Our methods enable casual users to quickly capture and recover multi-room structures coupled with their visual appearance, starting from panorama images generated with the built-in capabilities of modern mobile devices, as well as emerging low-cost 360° cameras. After introducing the reconstruction algorithms at the base of our approach, we show how to build applications able to generate 3D floor plans scaled to their real-world metric dimensions and capable to manage scene not necessary limited by *Manhattan World* assumptions. Then, exploiting the resulting structural and visual model, we propose a client-server interactive exploration system implementing a low-DOF navigation interface, specifically developed for touch interaction on smartphones and tablets.

Keywords: indoor scene reconstruction, omnidirectional images, mobile mapping

Concepts: •**Computing methodologies** → **Scene understanding; Reconstruction;** *Image processing; Image-based rendering;*

*Visual Computing Group, CRS4, Pula, Italy — vic.crs4.it — {pintore|gobbetti}@crs4.it
†Visual Computing Lab, ISTI-CNR, Pisa, Italy — vcg.isti.cnr.it — {ganovelli|scopigno}@isti.cnr.it

SA '16 Symp on Mobile Graphics and Interactive Applications, December 05 - 08, 2016, Macao
ISBN: 978-1-4503-4551-4/16/12
DOI: http://dx.doi.org/10.1145/2999508.2999526

1 Introduction

Acquisition and reconstruction of indoor architecture is an important and challenging task in many real world applications. Original blueprints are often hard to find, especially for older buildings, and even where they exist, they often do not represent the current layout. Moreover, since they do not provide a realistic visual representation of the scene, they can hardly be used for applications requiring visual location recognition and awareness, such as building security support [Guest et al. 2014; Pintore et al. 2016a].

Solutions to geometrically acquire the shape of indoor environments range from manual floor plans sketching to semi-automatic methods that process high-density scans (e.g., [Mura et al. 2014]). Although devices such as laser scanners often represent the most effective general solution for a dense and accurate acquisition, their use requires expensive equipment and specialized personnel, and is thus often restricted to specific application domains, such as cultural heritage, construction, architecture, or engineering. On the other hand, 3D reconstruction methods based on multiple images have become quite popular, especially with casual users. In certain situations, the accuracy of dense purely image-based methods has been shown comparable to laser sensor systems at a fraction of the cost [Seitz et al. 2006]. However, they typically require non-negligible acquisition time, and most of these approaches often fail to reconstruct surfaces with poor texture detail, which represent the majority of indoor surfaces (e.g., the perimeter walls of a room). Moreover, going from a dense acquisition to a structured model is a non-trivial task, which consumes notable memory and CPU resources [Cabral and Furukawa 2014].

Current mobile devices and emerging low-cost 360° mobile cameras offer nowadays a very attractive platform to overcome these problems in typical indoor capture applications, since they combine, in a low-cost solution, image capture with a variety of sensing hardware.

In this overview paper, we summarize our recently developed data-fusion approach for providing a fast and effective method to capture and share realistic models of buildings. The method targets end-users not necessarily skilled in virtual reality and 3D objects interaction. Our integrated system builds on our recently introduced methods for quickly capturing and recover 2.5D multi-room indoor environments, resulting in models scaled to real-world metric dimensions recovered by capturing and analyzing a single omnidirectional image per room [Pintore et al. 2016b]. This resulting structural and visual model is presented to the users using a client-server interactive system implementing a low-DOF navigation interface, specifically developed for touch interaction on smartphones and tablets [Pintore et al. 2016a].

Due to their widespread diffusion, natively integrating such a capture and explore pipeline in a mobile device would boost the development of next-generation, collaborative natural user and visual interfaces for many critical applications. Many real-world applications can benefit of this approach, including facilitating the work of interior designers or real estate operators. We imagine, in particular, a scenario where many users are able to map and share their home interior only with their mobile devices. Similarly to what happens with social networks, where people spontaneously shares much of their personal information, this mobile approach combined to a cloud system would afford to create a comprehensive database of building interiors, close to what happens for the exterior of buildings.

2 Mobile Reconstruction of Indoor Scenes

Mobile devices have become increasingly attractive due to their multi-modal acquisition capabilities and growing processing power, which enables fast digital acquisition, interactive scene understanding, and effective information extraction [Dev and Lau 2015]. Their use in terms of short-range 3D acquisition and mapping of indoor scenes has become a promising approach, as witnessed by the well known *Google Project Tango*. Mobile multi-room mapping is useful in many different real-world scenarios, such as interior and smart homes design, security management and building protection, mainly to enable non-technical people to create geometric models [MagicPlan 2014] or to add visual information to support interactive virtual tours [Sankar and Seitz 2012]. User-assisted approaches have long proven effective for floor plan reconstruction, but have the drawback of requiring extensive and repetitive user inputs. Moreover, they are prone to errors caused by imprecise device handing or manual editing. To overcome these limitations, in the last years research focused on devising data fusion and computer vision techniques that integrate data from inbuilt sensors and images [Martinelli 2012; Pintore and Gobbetti 2014; AliAkbarpour et al. 2015] to construct consistent models.

3 Omnidirectional Images and Mobiles

In recent times, there has been a renewed interest in omnidirectional images, that is images with a 360° field of view in the horizontal plane, or with a visual field that covers a hemisphere or the entire sphere. In terms of structural information, they present many advantages with respect to conventional perspective images, mainly for their wide field of view, which minimizes the possibility of fatal occlusions and improves the tracking of features. Due to the complexity of hardware and software models needed to deal with these projections [Geyer and Daniilidis 2000], their use was limited to robotic applications since few years ago, until mobile applications led to extensive utilization of automatically stitched spherical images in a variety of scenarios (e.g., *Google Photo Sphere*).

In the mobile world, we are interested on a specific category of omnidirectional projections: the *equirectangular* images – i.e., a spherical equal spacing map which has a field of view covering 360° longitude and 180° latitude (Fig. 2). When generated with the aid of sensor fusion by modern mobile devices (e.g., *Google Camera with Photo Sphere*), they have also the important property to be aligned to the gravity vector, ensuring that the vertical lines in the image of a room are also vertical lines in reality. Although equirectangular images are widespread and potentially contain a large amount of information about the architecture of the building, until recently the analysis of these images was done essentially by bringing back the problem to standard Multi-view stereo (MVS) methods for conventional perspective photos. Cabral et al. [2014], for instance, label indoor structures from omnidirectional images by exploiting extra 3D data externally calculated, making use of an additional structure-from-motion pipeline applied to the *unstitched* perspective images. Basically, these solutions inherit many of MVS approach limits, such as the need for heavy piecewise planarity and prior models, and the high computational cost make them impracticable on mobile hardware.

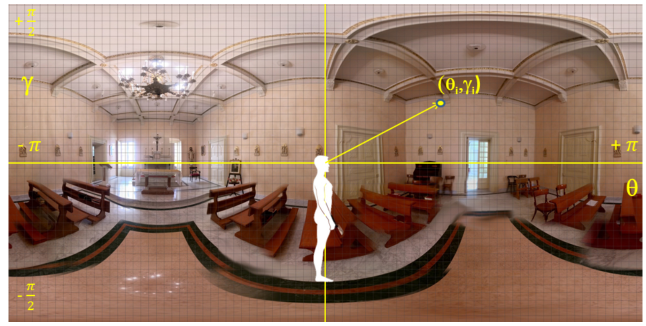

Figure 2: *The equirectangular image represent all possible views for an observer ideally located in the center of the spheremap, identified by the angles θ and γ.*

4 Mobile Mapping and Reconstruction

By exploiting theories commonly employed in catadioptric systems, we recently proposed a novel mobile solution [Pintore et al. 2016b], addressing the reconstruction problem directly in the omnidirectional space, thus avoiding the intrinsic issues of conventional multi-view approaches. This method automatically builds multi-room models from omnidirectional images, even when the walls in the scene do not form right angles. We have introduced a spatial transform for equirectangular images to recover a prior parametric model of the rooms and enable the solution of the reconstruction problem as a global optimization. Then, through the help of the device IMU, we iterate the method to map and reconstruct the entire floor plan (Fig. 3). As we recently demonstrated in Garro et al. [2016], by combining image and intertial acceleration data it is also possible to directly scale the acquired model to metric dimensions. Since these algorithms are not computationally demanding, it is possible to fully implement the entire acquisition and reconstruction pipeline on a mobile device.

This approach presents many advantages. It allows mobile device users to quickly measure and sketch an indoor environment. A single panoramic image per room can be easily obtained by off-the-shelf guided applications – a much simpler approach than multi-view methods. Instead of relying on costly offline processing, it provides immediate processing with an automatic and light-weight floor map reconstruction method. Furthermore, it returns accurate results even for scenes with surfaces lacking in texture and details, unlike MVS (Multi View Stereo) methods. In contrast to many previous approaches, neither strong *Manhattan World* constraints, nor further 3D information (e.g., original unstitched images, externally

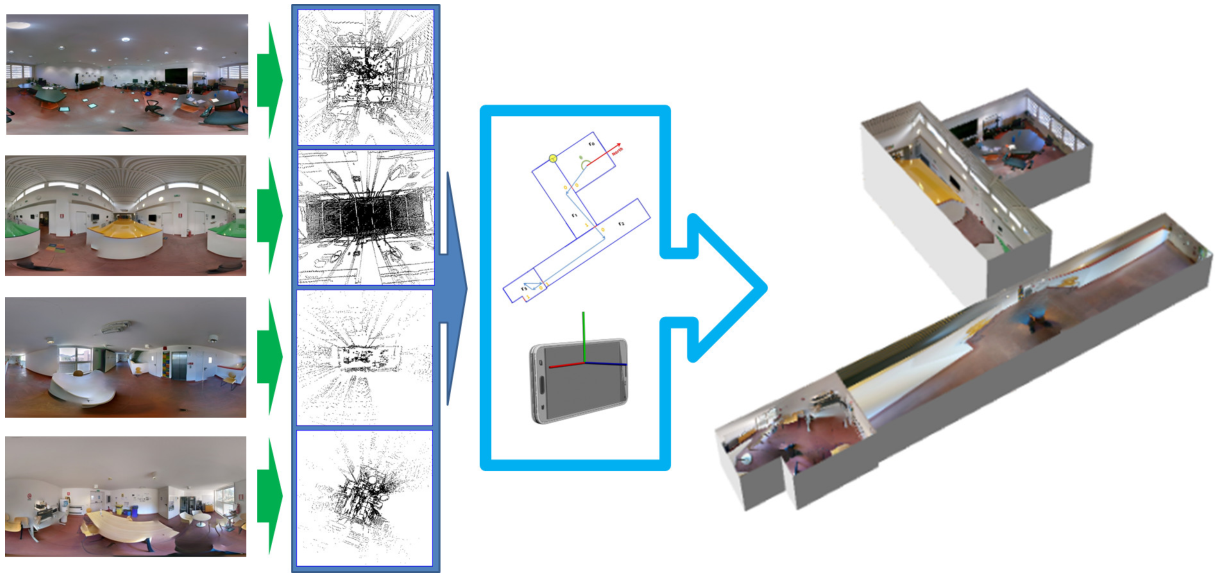

Figure 3: *We take as input one omnidirectional image for each room and project the image gradient map to a plane to highlight and extract the room shape. We iterate the method to map and reconstruct the entire floor plan, using the IMU to recover scale and relative position of rooms. As a result, we obtain a 3D representation of the captured environment, along with its visual representation through spheremaps.*

calculated 3D points, MVS data) are needed. By combining data from the images with data coming from the inertial sensors [Martinelli 2012], the whole pipeline can return rooms in real-world units, allowing the composition of multi-room models without manual interventions.

The multi-room model generated by the construction pipeline consists in a 3D floor plan, describing and indoor environment in terms of rooms bounded by walls and connected by doors. Fig 4 shows two representative examples of buildings acquired and reconstructed using our approach. The office building to the left (19 rooms) was acquired with a mobile phone in 51 minutes, and the presented model was obtained in a without any manual intervention. The building to the right (70 rooms), acquired with a 360°, was acquired in only 46 minutes, showing the interest of emerging panoramic cameras for such acquisition tasks. On the other hand, since many parts of the second building do not comply with the assumption of closed rooms connected by doors, manual interaction (2.5 hours) was required in post-processing to build the final model from disjoint pieces.

5 Interactive Mobile Exploration

The multi-room 3D scene is already simplified and structured, and can be visually enriched by associating each room to the equirectangular images taken within it, which provide original texturing information. We exploit such structuring to represent the multi-room model as a graph (Fig. 5), in which rooms are nodes and passages are edges. Such a representation offers a simple scalable way to interactively explore visually realistic indoor environment on mobile setups and/or multiplatform WebGL rendering systems [Pintore et al. 2016a]. In this approach, the viewer shows a panoramic image for each room exploiting a simple fragment shader that, for each fragment, takes the direction of the corresponding viewing ray and samples the equirectangular image. The interaction consists of simply dragging to change view orientation and pinching to zoom in/out. Thanks to the structured model recovered by our system, we are able to real-time render the transitions between rooms, creating a textured mesh by projecting the panoramas on the 3D model. This results in a considerable advantage in terms of performance with respect to similar approaches (e.g. [Di Benedetto et al. 2014; Sankar and Seitz 2012]), where the transitions between rooms are instead represented with externally rendered video sequences. The main advantage of this solution is that the network bandwidth occupancy is reduced, as well as the lag due to video buffering during the movement between rooms, resulting in real-time performance even on low-powered mobile devices or poor network coverage.

6 Conclusions and Future Trends

Starting from panorama images generated with the aid of commodity mobile devices, our methods rapidly compute visual and structural 3D models of indoor environments scaled to their metric dimensions. The resulting 3D floor plans of multi-room environments can be used for a variety of planning applications, while the visual models provide location awareness within interactive mobile interactive application specifically developed for touch interaction on smartphones and tablets. These capabilities open the door to a variety of applications in which casual users quickly capture and share structured and visually realistic indoor environments.

Our focus on using omnidirectional images as building blocks for both capture and display is reinforced by current technological trends. In particular, a new category of compact low-cost mobile cameras capable to capture spherical panoramas and omnidirectional video sequences is rapidly emerging (e.g. *Ricoh Theta S*, *Samsung Gear360* and more). These devices are often equipped with an IMU and are completely interfaced with modern smartphones, thus becoming a natural extension of them. A straightforward improvement of our approach would the exploiting multiple omnidirectional images, both to cover cases where the entire perimeter of a single room is not visible from a single point or to manage a fully automatic mapping of complex structures (such as the one in Fig. 4 right). In this context, we are exploring how to combine our method with new specialized multi-view techniques and how to improve data visualization and immersive experiences exploiting panoramic 360° videos.

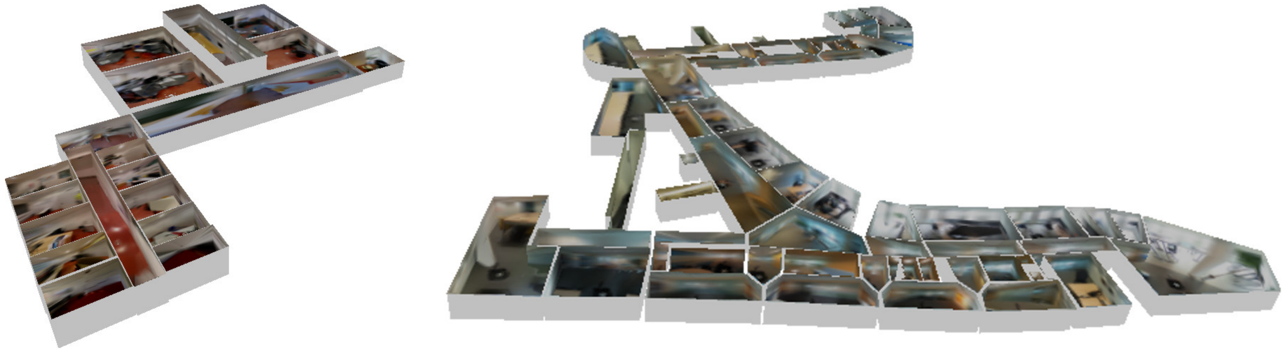

Figure 4: *Right: 3D reconstruction of a 655 mq office with 19 rooms. This environment was acquired with a mobile phone (HTC One M8) and automatically reconstructed using one spheremap stitching for each room. This is a typical multi-room environment described by closed rooms bounded by walls and connected by doors, where the entire perimeter of a single room is visible from a single point of view. Left: Reconstruction of a 70 rooms floor of the NHV ministry at Den Haag, Netherlands. The whole model was acquired with a* Ricoh Theta S *camera and represents an example of future challenge. Since many parts of the environment do not comply with the assumptions of closed rooms connected by doors, a minimal manual interaction was actually needed to help the mapping.*

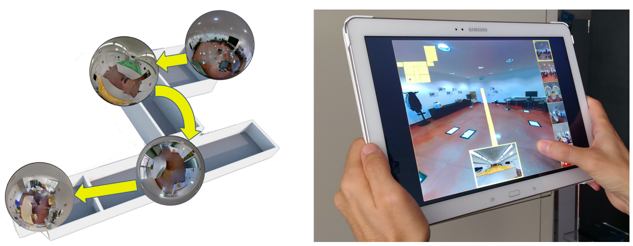

Figure 5: *Each node of the exploration graph is represented by a spheremap/room, whereas the edges (yellow) are the transition between adjacent rooms, which are generated by view interpolation using the reconstructed structural geometry. The resulting model can be quickly explored visually on a mobile platform even in presence of low-bandwidth networks.*

7 Acknowledgments

This work has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607737 (VASCO). We also acknowledge the contribution of Sardinian Regional Authorithies through project VIGEC.

References

ALIAKBARPOUR, H., PALANIAPPAN, K., AND SEETHARAMAN, G. 2015. Fast structure from motion for sequential and wide area motion imagery. In *Proc. IEEE ICCVW Video Summarization for Large-scale Analytics Workshop*.

CABRAL, R., AND FURUKAWA, Y. 2014. Piecewise planar and compact floorplan reconstruction from images. In *Proc. CVPR*, 628–635.

DEV, K., AND LAU, M. 2015. Democratizing digital content creation using mobile devices with inbuilt sensors. *IEEE CG&A 35*, 1, 84–94.

DI BENEDETTO, M., GANOVELLI, F., BALSA RODRIGUEZ, M., JASPE VILLANUEVA, A., SCOPIGNO, R., AND GOBBETTI, E. 2014. Exploremaps: Efficient construction and ubiquitous exploration of panoramic view graphs of complex 3d environments. *Computer Graphics Forum 33*, 2.

GARRO, V., PINTORE, G., GANOVELLI, F., GOBBETTI, E., AND SCOPIGNO, R. 2016. Fast metric acquisition with mobile devices. In *Proc. 21st International Workshop on Vision, Modeling and Visualization (VMV)*. To appear.

GEYER, C., AND DANIILIDIS, K. 2000. A unifying theory for central panoramic systems and practical implications. In *Proc. ECCV*, 445–461.

GUEST, J., EAGLIN, T., SUBRAMANIAN, K., AND RIBARSKY, W. 2014. Interactive analysis and visualization of situationally aware building evacuations. *Information Visualization*.

MAGICPLAN, 2014. Sensopia. www.sensopia.com.

MARTINELLI, A. 2012. Vision and imu data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination. *IEEE Transactions on Robotics 28*, 1 (Feb), 44–60.

MURA, C., MATTAUSCH, O., JASPE VILLANUEVA, A., GOBBETTI, E., AND PAJAROLA, R. 2014. Automatic room detection and reconstruction in cluttered indoor environments with complex room layouts. *Computers & Graphics 44*, 20–32.

PINTORE, G., AND GOBBETTI, E. 2014. Effective mobile mapping of multi-room indoor structures. *The Visual Computer 30*.

PINTORE, G., GANOVELLI, F., GOBBETTI, E., AND SCOPIGNO, R. 2016. Mobile mapping and visualization of indoor structures to simplify scene understanding and location awareness. In *Proc. Fourth International Workshop on Assistive Computer Vision and Robotics (ACVR)*. To appear.

PINTORE, G., GARRO, V., GANOVELLI, F., AGUS, M., AND GOBBETTI, E. 2016. Omnidirectional image capture on mobile devices for fast automatic generation of 2.5D indoor maps. In *Proc. IEEE WACV*.

SANKAR, A., AND SEITZ, S. 2012. Capturing indoor scenes with smartphones. In *Proc. UIST*, ACM, New York, NY, USA, 403–412.

SEITZ, S. M., CURLESS, B., DIEBEL, J., SCHARSTEIN, D., AND SZELISKI, R. 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. CVPR*, vol. 1, 519–528.