

RESEARCH ARTICLE

Open Access

Spatial and visual data fusion for capturing, retrieval, and modeling of as-built building geometry and features

Zhenhua Zhu* and Sara Donia

Abstract

Background: As-built building information, including building geometry and features, is useful in multiple building assessment and management tasks. However, the current process for capturing, retrieving, and modeling such information is labor-intensive and time-consuming.

Methods: In order to address these issues, this paper investigates the potentials of fusing visual and spatial data for automatically capturing, retrieving, and modeling as-built building geometry and features. An overall fusion-based framework has been proposed. Under the framework, pairs of 3D point clouds are progressively registered through the RGB-D (Red, Green, Blue plus Depth) mapping. Meanwhile, building elements are recognized based on their visual patterns. The recognition results can be used to label the 3D points, which could facilitate the modeling of building elements.

Results: So far, two pilot studies have been performed. The results show that a high degree of automation could be achieved for the registration of building scenes captured from different scans and the recognition of building elements with the proposed framework.

Conclusions: The fusion of spatial and visual data could significantly facilitate the current process of retrieving, modeling, and visualizing as-built information.

Keywords: As-built information; Data fusion; RGB-D mapping; Visual recognition; Automation

Introduction

Three dimensional (3D) as-built building information records existing conditions of buildings, including their geometry and actual details of architectural, structural, and mechanical, electrical, and plumbing (MEP) elements (Institute for BIM in Canada IBC 2011). Therefore, such information is useful in multiple building assessment and management tasks. For example, as-built building information could be used to identify and quantify the deviations between design and construction, which could significantly reduce the amount of rework during the construction phase of a project (Liu et al. 2012). Also, the use of as-built building information could facilitate the coordination of the MEP designs, when renovating and retrofitting existing old buildings. This facilitation was expected

to reduce almost 60% of the MEP field-to-finish workflow (ClearEdge3D 2012).

Although as-built building information is useful, the current process for capturing, retrieving, and modeling such information requires a lot of manual, time-consuming work. This labor-intensive and time-consuming nature increases the cost of using as-built building information in practice. As a result, it was not a value adding task for most general contractors and has not been widely used in the vast majority of construction and renovation/retrofit projects (Brilakis et al. 2011), unless the time and cost of the current process for capturing, retrieving, and modeling as-built building information could be significantly reduced, and the as-built building information could be constantly updated and closely reviewed (Pettee, 2005).

In order to automate the current manual process for capturing, retrieving, and modeling as-built building information, several research studies have been initiated.

* Correspondence: zhenhua.zhu@concordia.ca
Department of Building, Civil and Environmental Engineering, Concordia University, Montreal H3G 1 M8, Canada

Most of them were built upon the remote sensing data captured by laser scanners or digital cameras. For example, (Okorn et al. 2010) presented an idea of creating as-built floor plans in a building by projecting the 3D point clouds captured by a laser scanner. (Fathi and Brilakis 2012) relied on the color images captured by digital cameras to model as-built conditions of metal roof panels. (Geiger et al. 2011) proposed an approach to build 3D as-built maps from a series of high-resolution stereo video sequences.

The sole reliance on one type of sensing data makes existing studies have inherent limitations. For example, when capturing, retrieving, and modeling 3D as-built building information in the building indoor environments, (Furukawa et al. 2009) noted that the prevalence of texture-poor walls, floors, and ceilings in the environments may fail those studies built upon the images or video frames captured by digital cameras. Also, (Adan et al. 2011) reported that a 3D laser scanner had to be set up at hundreds of locations in order to complete the scan of 40 rooms.

Compared with existing research studies, this paper investigates the potential of fusing two different types of sensing data (i.e. color images and depth images) to capture, retrieve, and model 3D as-built building information. The focus has been placed on building geometry and features. An overall fusion-based framework has been proposed in the paper. Under the framework, 3D point clouds are progressively registered through the RGB-D (Red, Green, Blue plus Depth) mapping, and building elements in the point clouds are recognized based on their visual patterns for 3D as-built modeling. So far, two pilot studies have been performed to show that a high degree of automation could be achieved with the proposed framework for capturing, retrieving, and modeling as-built building information.

Background

The benefits of using 3D as-built building information have been well acknowledged by researchers and professionals in the architecture, engineering, and construction (AEC) industry. Meanwhile, the process for capturing, retrieving, and modeling such information has been identified as labor-intensive (Tang et al. 2010). Typically, the process starts with the collection of as-built building conditions using remote sensing devices, such as laser scanners or digital cameras. Then, the sensing data collected from multiple locations are registered (i.e. building scenes registration), and building elements in the sensing data are recognized (i.e. building elements recognition). This way, the semantic or high-level building geometric information in the sensing data could be used for domain related problem solving.

So far, a lot of manual work has been involved in the registration of building scenes and the recognition of

building elements. In order to address this issue, multiple research methods have been initiated. Based on the type of the sensing data they work on, the methods can be classified into three main categories. The methods in the first category were built upon the 3D point clouds directly captured by the devices, such as terrestrial laser scanners. The methods in the second category relied on the color images or videos taken by digital cameras or camcorders. Recently, the methods with the use of RGB-D data (i.e. pairs of color and depth images) have been proposed with the development of RGB-D cameras (e.g. Microsoft® *Kinect*). Below are the details of these methods.

Point cloud based methods

In the point cloud based methods, terrestrial laser scanners are commonly used to capture the detailed as-built building conditions. One laser scan may collect millions of 3D points in minutes. Using this rich information, (Okorn et al. 2010) proposed an idea of creating as-built building floor plans by projecting the collected 3D points onto a vertical (z-) axis and a ground (x-y) plane. The projection results could indicate which points can be grouped. This way, the floor plans were created (Okorn et al. 2010). In addition to the generation of floor plans, (Xiong and Huber 2010) employed the conditional random fields to model building elements from 3D points, but their work was limited to those building elements with planar patches, such as walls and ceilings. The openings on the walls or ceilings could be further located by checking the distribution density of the points around the openings (Ripperda and Brenner, 2009) or using a support vector machine (SVM) classifier (Adan and Huber, 2011).

Although the detailed as-built building information can be captured with laser scanners, the scanners are heavy and not portable. Typically, at least two crews are needed to use one laser scanner to collect as-built conditions (Foltz, 2000). This non-portable nature makes laser scanners inconvenient to use, especially when capturing, retrieving, and modeling the as-built information in the building indoor environments. (Adan et al. 2011) mentioned that a laser scanner had to be set up at 225 different locations in order to scan 40 rooms, which was approximately 5.6 locations per room.

Also, the 3D points collected by laser scanners only record the spatial as-built building conditions, which limits the capability to recognize building elements. Existing recognition methods mainly relied on the points spatial features, which can be described globally or locally (Stiene et al. 2006). Global descriptors (Wang et al. 2007; Kazhdan et al. 2003) captured all the geometrical characteristics of a desired object. Therefore, they are discriminative but not robust to cluttered scenes (Patterson et al. 2008). Local descriptors (Shan et al. 2006; Huber

et al. 2004) improved the recognition robustness, but they were computationally complex (Bosche and Haas 2008). Both global and local descriptors cannot recognize building elements with different materials. For example, it is difficult for the methods to differentiate between a concrete column and a wooden column just based on their 3D points, if both of them have the same shape and size.

Color based methods

The methods in the second category target on the visual data (i.e. color images or videos) from digital cameras or camcorders as an affordable and portable alternative for as-built modeling. So far, several critical techniques have been created. (Snavely et al. 2006) explored the possibility to create a 3D sparse point cloud from a collection of digital color images (i.e. structure-from-motion). (Furukawa and Ponce 2010) developed a multi-view stereo matching framework, which can generate an accurate, dense, and robust point clouds from stereo color images. In the work of (Durrant-Whyte and Bailey 2006), a robot was used to build a 3D map within an unknown environment while at the same time keeping track of its current locations (i.e. simultaneous localization and mapping). Based on these techniques, many image/videos based methods for capturing, retrieving, and modeling 3D as-built building information have been proposed, and the 3D reconstruction of the built environments has become possible. For example, the 3D models of built environments with high visual quality can be automatically derived even from single facade images of arbitrary resolutions (Müller et al. 2007). Also, (Pollefeys et al. 2008) used the videos from a moving vehicle to model the urban environments.

Compared with 3D point clouds, there are multiple recognition cues that can be extracted from images or videos, including color, texture, shape, and local invariant visual features. (Neto et al. 2002) used color values to recognize structural elements in digital images, while (Brilakis et al. 2006) presented the concept of “material signatures” to retrieve construction materials, such as concrete, steel, wood, etc. (Zhu and Brilakis 2010) combined two recognition cues (shape and material) to locate concrete column surfaces. Although many recognition methods have been developed, the robust recognition of buildings elements has not been achieved yet. One main limitation lies in the fact that the recognition results from images or videos are limited to two-dimensional (2D) space. Therefore, it is difficult to directly use the recognition results for capturing, retrieving, and modeling 3D as-built building information.

In addition, digital cameras or camcorders are easy and convenient to use in building environments with multiple interior partitions. However, most existing image/videos based methods for capturing, retrieving, and modeling 3D

as-built building information heavily relied on the extraction of visual features from images or video frames. Building environments are commonly prevalent with texture-poor walls, floors, and ceilings. This may fail the procedures of these methods, unless the scene-specific constraints for the environments are pre-created manually (Furukawa et al. 2009).

RGB-D based methods

Point cloud based methods and color based methods relied on one type of sensing data for capturing, retrieving, and modeling as-built building conditions. As a result, they have inherent limitations associated with the type of the sensing data they work on. Specifically, the point cloud based methods can accurately retrieve the detailed as-built building conditions, but the data capturing process is time-consuming. Most laser scanners are not portable and the setup of a laser scanner also takes time (Foltz, 2000). In addition, the recognition of building elements from the point clouds solely relied on their spatial features. Compared with the point cloud based methods, the recognition of building elements from color images or videos can be performed using multiple recognition cues, which could increase the recognition accuracy. The main limitation of the color based methods lies in their modeling robustness. Most color based methods may fail when building scenes have texture-poor elements. Table 1 summarizes the comparison results between the methods in these two categories.

The emergence of RGB-D cameras provides another idea to capture, retrieve, and model 3D as-built building information. The RGB-D cameras are novel sensing systems, which can capture pairs of mid-resolution color and depth images almost in real time. One example of the RGB-D cameras is Microsoft® *Kinect*. The resolution of the *Kinect* is up to 640×480 , which is equivalent to the capture of 307,200 3D points per frame. Also, its sensing rate is up to 30 frames-per-second (FPS) sensing rate, which means almost 1 million points could be captured by the *Kinect* in one second. Considering these characteristics, (Rafibakhsh et al. 2012) once concluded that the RGB-D cameras, such as *Kinect*, have great potential for spatial sensing and modeling applications at construction sites. So far, several research studies (Weerasinghe et al. 2012; Escorcía et al. 2012) have been proposed. However, most of them focused on the recognition and tracking of construction workers, when the positions of the RGB-D cameras are fixed. To the authors' knowledge, none of existing research studies has been focused on modeling as-built building information.

In addition to the research studies in the AEC domain, researchers in computer vision also investigated the use of RGB-D cameras. For example, (Henry et al. 2010) and (Fioraio and Konolige 2011) examined how a dense 3D

Table 1 Comparison of point cloud based methods and color based methods

Criteria	Point cloud based methods	Color based methods
Sensing Data Properties	Accurate and detailed	Not accurate and detailed
Scene Registration	Point clouds distance	Visual features matching
Building Elements Recognition	One recognition cue only (spatial feature)	Multiple recognition cues (color, texture, shape, etc.)

map of building environments can be generated with RGB-D cameras (i.e. RGB-D mapping). More impressively, (Izadi et al. 2011) showed that the rapid and detailed reconstruction of building indoor scenes can be achieved; however, their work is currently limited to the reconstruction of a relatively small building scene (3 m cube) (Roth and Vona 2012).

Objective and scope

Although the preliminary results from existing RGB-D based methods are promising, the full potentials of the fusion of the spatial and visual data have not been well investigated for capturing, retrieving, and modeling as-built building information. The objective of this paper is to fill this gap, and the focus is placed on how to automate the current process for capturing, retrieving, and modeling as-built building geometry and features. Specifically, an overall fusion-based framework for capturing, retrieving, and modeling as-built building geometry and features has been proposed. The framework mainly includes two parts: 1) 3D scenes registration and 2) 3D building elements recognition. First, the 3D point clouds captured from different scans are progressively registered through the RGB-D mapping. Then, the building elements in the 3D point clouds are recognized based on their visual patterns. The potential and effectiveness of the framework have been tested with two pilot studies. In the studies, the issues such as the accuracy and automation for 3D scene registration and elements recognition have been evaluated. Although the pilot studies use the RGB-D camera, Microsoft® *Kinect*, as the sensing device, it is worth noting that the fusion-based idea behind the proposed framework is also applicable for other stand-alone or combined sensing devices as long as the sensing devices could provide pairs of color and depth images with the one-on-one correspondence.

Methods

As illustrated in Figure 1, the proposed fusion-based framework mainly includes two parts, 3D scene registration and 3D building elements recognition. The scene registration refers to the merge and alignment of the point clouds from multiple scans into one single point cloud under a pre-defined global coordinate system. The registration is always necessary in the process of capturing, retrieving, and modeling as-built building information, since one scan typically cannot capture all as-built

conditions in a building environment. Here, the global coordinate system is defined as the coordinate system used by the point cloud in the first scan. The point clouds in other scans are aligned to the cloud in the first scan. The idea behind the RGB-D mapping (Henry et al. 2010) has been adopted to achieve the registration purpose. First, the visual feature detectors are used to determine the positions of the features in the color images. The features are distinctively described with feature descriptors. Then, the common features in consecutive color images. Based on the matching results, the pairs of 3D matching points can be determined by referring to the locations of the matched feature positions in the color images and their corresponding depth values. This way, the point clouds in the scans could be progressively registered.

A point cloud includes a set of 3D points with known coordinates. The coordinates indicate the positions of the points in the building environments without any semantic or high-level geometric as-built building information. In order to retrieve this information, building elements in the environments need to be recognized from the point cloud. This way, the geometries and dimensions of the elements can be estimated and modeled correspondingly. However, the direct recognition of the building elements from the point cloud has proven difficult, especially when the detailed prior information is not available (Bosche and Haas 2008). Therefore, the visual features of building elements has been exploited here. Under the workflow of building elements recognition, the elements are first recognized in the color images based on their unique visual patterns. The patterns include the topological configurations of the elements' contour and texture features. For example, concrete columns in buildings are dominated by long near-vertical lines (contour features) at sides and concrete surfaces (texture feature) in the middle, no matter they are rectangular or circular. Therefore, they can be located by searching such cues in the color images. When building elements are recognized from the color images, the recognition results could be used to classify the 3D points, so that the points belonging to the same building elements could be grouped and modeled separately.

Results

So far, two pilot studies have been performed to show the effectiveness of the proposed framework. In the studies, the sensing data collected by the Microsoft® *Kinect* have

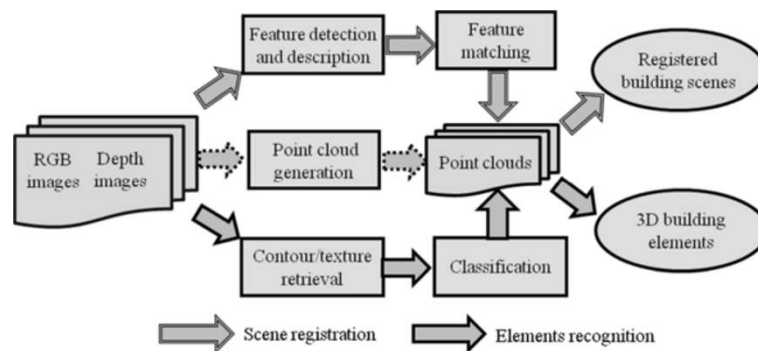


Figure 1 Fusion-based framework.

been used to evaluate the framework. The *Kinect* is a novel hybrid sensing system, which can provide a stream of color and depth images with a resolution of 640×480 pixels and 2048 levels of depth sensitivity (Biswal 2011). The typical sensing range of the *Kinect* is limited from 1.2 meters to 3.5 meters with an angular field of view of 57° horizontally and 43° vertically (Biswal 2011). According to the report from (Rafibakhsh et al. 2012), the average sensing error of the *Kinect* was around 3.49 cm. The authors also compared the sensing data from the *Kinect* with the data from a Leica total station, and found that the average absolute and percentage errors were 3.60 cm and 2.898% (Zhu and Donia, 2013).

Although the *Kinect* has a limited sensing range and considerably lower accuracy than a high-end terrestrial laser scanner or total station, the main focus of the studies is not to evaluate the potential of the *Kinect* for construction applications. Instead, the emphasis is placed on evaluating the automation of the proposed framework for registering building scenes and recognizing building elements with the idea behind the fusion of visual and spatial data. Therefore, although the *Kinect* has been selected in the studies, other stand-alone or combined

devices could be used, as long as they provide pairs of color and depth images. For example, the sensing data captured by the combination of a digital camera and a terrestrial laser scanner are also supported by the proposed framework, if the color images captured by the digital camera are well calibrated with the data captured by the laser scanner.

Scenes registration

The first pilot study is to test the 3D building scenes registration. Figure 2 shows the detection of the visual features in the color image with the proposed framework, where the ORB (Oriented FAST and Rotated BIEFF) detector and descriptor (Rublee et al. 2011) have been used. In the figure, the green points indicated the visual features detected in the color image, and their corresponding positions in the 3D point cloud could also be located, considering one-on-one correspondence between the pixels in the pair of the color and depth images. Figure 3 illustrates an example of matching common visual features in consecutive color images. The matching of visual features in consecutive color images was built upon the fast approximate nearest neighbor (FANN) searches.

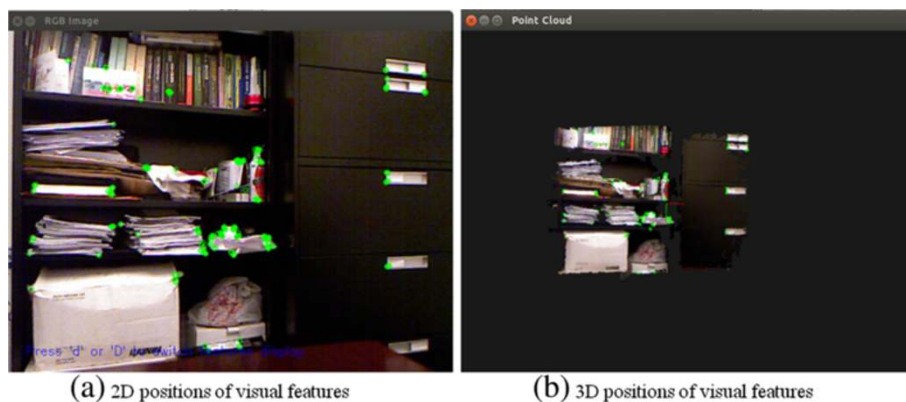


Figure 2 2D and 3D positions of visual features.

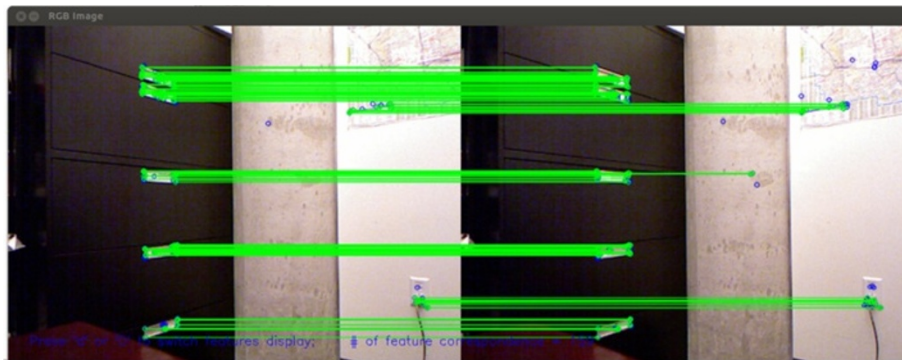


Figure 3 Matching of 2D visual features.

The detailed information about the FANN searches could be found in the work of (Muja and Lowe 2012) and (Muja and Lowe 2009), and the specific steps for feature detection, and description could be referred to the work of (Rublee et al. 2011). The matching results could indicate the common 3D points in the consecutive point clouds. This way, the point clouds from different scans could be progressively and automatically registered (Figure 4). The final 3D point cloud after the registration is displayed in Figure 5.

The accuracy of the scenes registration in the proposed framework has been tested with the common datasets prepared by (Sturm et al. 2012) in the Computer Vision Groups at the Technische Universität München. Here, the accuracy is measured by comparing the difference between the Kinect scan positions (i.e. estimated trajectory)

with its real scan positions (i.e. ground truth). According to the test results, it was found that the average error of using the ORB detector and descriptor for the scenes registration was around 0.24 meter. The error may come from the sensing error from the sensing device. For example, the average sensing error of the *Kinect* in one scan was around 3.49 cm (Rafibakhsh et al. 2012). In addition to the sensing error, the automatic pair-wise registration may also produce the registration error, since no explicit control reference points have been used during the registration process. Such registration errors could not be completely eliminated but accumulated along the registration progress. A recent study indicated that the selection of appropriate visual feature detector and descriptor may significantly reduce the pair-wise registration error (Zhu, 2013). Therefore, more studies are needed to investigate

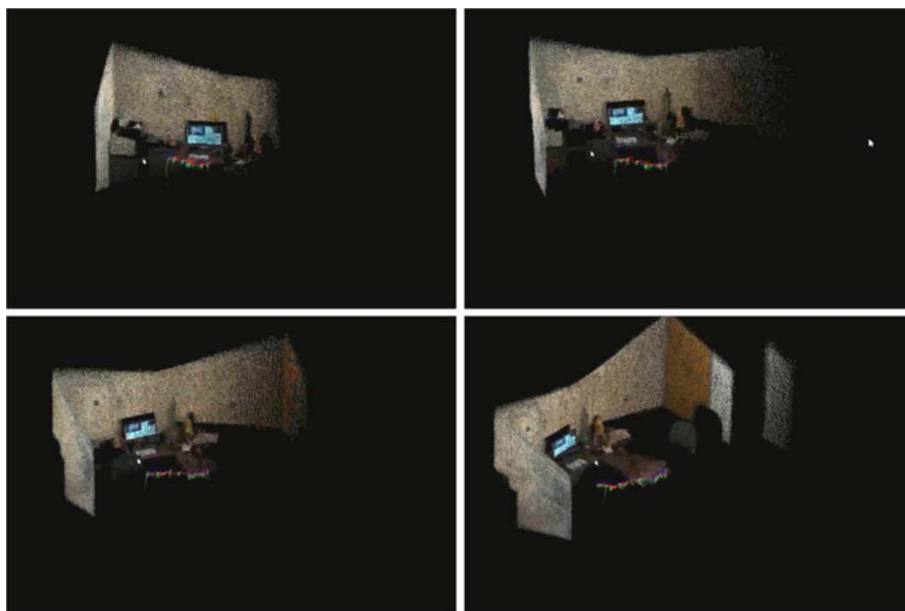


Figure 4 Progressive 3D scenes registration.

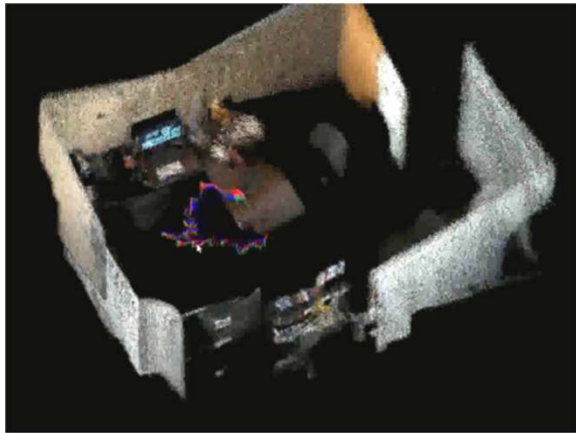


Figure 5 Registered 3D point cloud of an office.

the effectiveness of the color and texture information available in the building environments on the accuracy of the scenes registration.

Building elements recognition

The second pilot study is to test the 3D building elements recognition. Figures 6 and 7 show an example of 3D points segmentation and clustering with their visual cues. In this example, the color image is first converted from the red, green, and blue (RGB) space to the $L^*a^*b^*$ space, which includes a luminosity layer (L^*) and two chromaticity layers (a^* and b^*). Then, the image is segmented and clustered with the k-means algorithm (Arthur and Vassilvitskii, 2007) based on the information in the two chromaticity layers. The luminosity layer is not considered here in order to remove the effect of the brightness on the segmentation and clustering results. This way, building elements, such as walls, doors, etc., in the color image could be identified (Figure 6). Moreover, the corresponding 3D points belonging to the same types of building elements could be extracted (Figure 7). Compared with the direct segmentation and clustering

of 3D point cloud, it can be seen that the use of visual and spatial data could provide a convenient and fully automatic way to segment and cluster 3D point clouds.

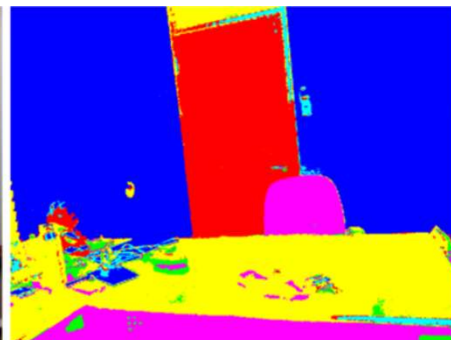
Moreover, specific 3D building elements in the point cloud could be further recognized, when considering more visual cues. For example, concrete columns (rectangular or circular) in a color image are dominated by long near-vertical lines (contour cues) and concrete surfaces (color and texture cues). Therefore, they could be located by searching these cues in the color image. This idea of recognizing concrete columns has been implemented by the writer. More details can be found in the writer's pervious work (Zhu and Brilakis, 2010). When such cues are found in the image, the corresponding 2D column surface pixels could be mapped to the 3D points in the cloud. This way, the points belonging to the column surface can be classified and grouped to indicate the concrete column in 3D, as illustrated in Figure 8.

Discussion

The main purpose of the two pilot studies is to show the degree of automation that the proposed framework could reach with the idea of fusing the spatial and visual data. The studies are not used to validate that more dense and/or accurate 3D points could be captured with the fusion of spatial and visual data. Although the sensing density and accuracy play an important role in the process of capturing, retrieving, and modeling as-built building information, they are not in the scope of this paper. Instead, the studies are used to indicate that a high degree of automation could be achieved in the current process of capturing, retrieving, and modeling as-built building information with the fusion of spatial and visual data. The automation has been highly valued to promote the use of 3D as-built building information, especially considering the current process of capturing, retrieving, and modeling as-built building information is labor-intensive and time-consuming.



(a) Original image



(b) Segmentation result

Figure 6 Image segmentation.

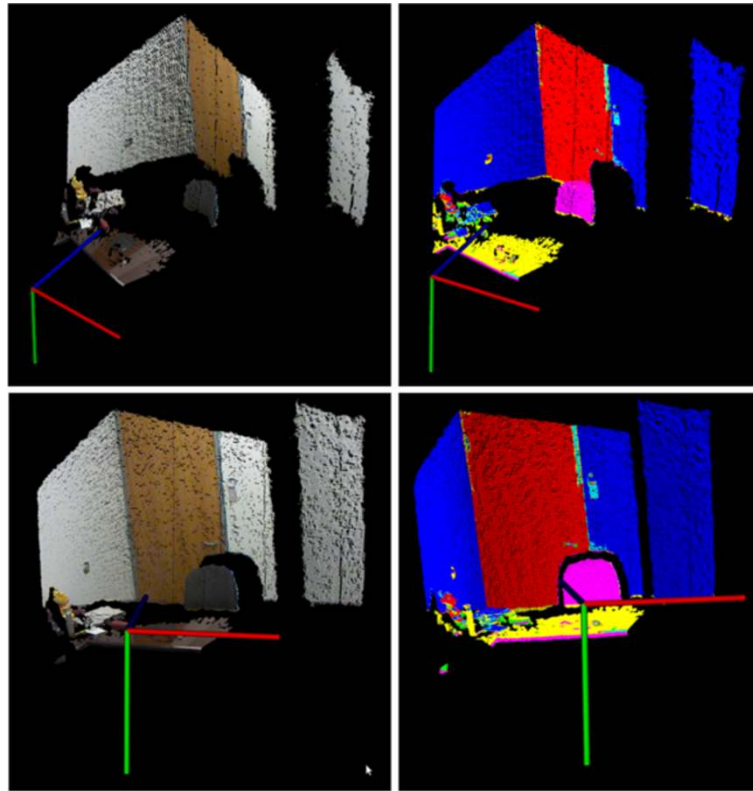


Figure 7 Elements segmentation in 3D point cloud.

The preliminary results from the two pilot studies have shown that building scenes could be progressively and automatically registered by detecting, describing, and matching visual features. Also, building elements in the environments could be automatically recognized based on their visual cues. In both studies, no specific user interventions are required for setting up the reference/control points in the building environments and editing the sensing data for building scenes registration and building elements recognition. A user just needs to

hold the sensing device and keep scanning the building environments. The 3D information of the building environments could be automatically captured and retrieved. Meanwhile, the results could be fed back to the user almost simultaneously for his/her timely review. Compared with existing work in capturing and retrieving as-built building information, to the writers' knowledge, none of existing point cloud based methods or color based methods could reach such a high degree of automation. This is a significant advantage of fusing the spatial and

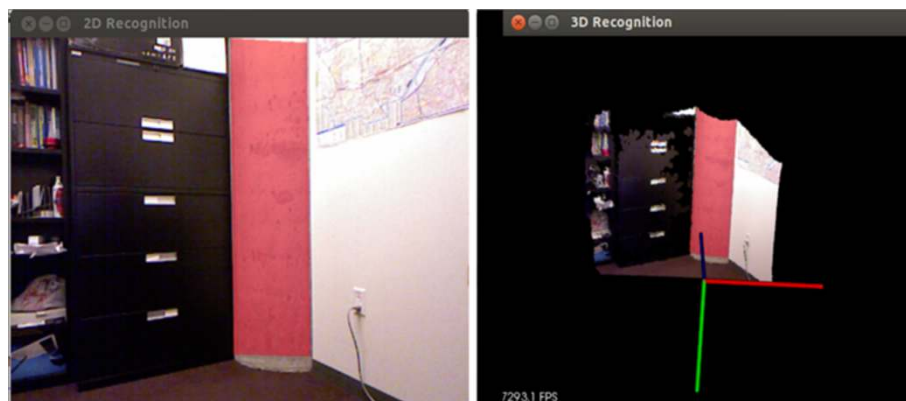


Figure 8 Example of concrete column recognition.

visual data learned from the two pilot studies, and also the contribution of this paper to the existing body of knowledge in the areas of capturing, retrieving, and modeling as-built building information.

Conclusions

3D as-built building information, including building geometry and features, is useful in multiple building assessment and management tasks, but the current process for capturing, retrieving, and modeling as-built building information is labor-intensive and time-consuming. This labor-intensive and time-consuming nature limited the wide use of as-built building information in the AEC industry. In order to address this issue, several research studies have been proposed, but they were built upon one single type of sensing data. The sole reliance on one type of sensing data has several limitations.

This paper investigates the fusion of visual and spatial data to facilitate the process of capturing, retrieving, and modeling as-built building information. An overall fusion-based framework has been proposed. Two pilot studies have been performed. The first study tested the automation of building scenes registration with visual feature detection, description, and matching. It was found that the full automation of 3D building scenes registration could be achieved with the fusion of visual and spatial data, but the registration accuracy needs to be further improved. The second study tested the recognition of building elements in the 3D point cloud with their visual recognition cues. It was noted that the use of visual recognition cues could facilitate the recognition of building elements in 3D.

So far, the proposed fusion-based framework rely on the visual features for 3D building scenes registration and building elements recognition. The precise extraction and description of visual features in the environments play an important role on the success of the proposed framework. That is why the selection of different visual feature extractors and descriptors may affect the scene registration accuracy as indicated in the pilot study 1. However, no matter what visual feature extractor and descriptor are selected, the visual features extraction and description could still be affected by the environmental conditions (e.g. lighting, lack of texture and color information, shadows, etc.), since the visual feature detectors and descriptors currently available are not ideal (Mikolajczyk and Schmid, 2005). Also, these environmental conditions might affect the segmentation and clustering of building scenes for building elements recognition. In the pilot study 2, although the RGB color space was converted to $L^*a^*b^*$ space and the luminosity layer (L^*) was not considered in the segmentation and clustering process, the illumination effects on the segmentation and clustering results could still not be completely removed. Therefore, future work will focus on 1) how to increase the robustness

of the proposed framework against common environmental conditions (e.g. lighting, lack of texture and color information, shadows, etc.), and 2) how to recognize other types of 3D building elements with different materials.

Competing interests

The authors declare that they have no competing interests.

Authors' contribution

ZZ carried out the spatial and visual data fusion study, participated the sequence alignment and drafted the manuscript, and directed the entire processes of this study. SD undertook the experiments to collect the spatial and visual data using a Kinect camera, and provided suggestions on the data collection and analysis. Both authors read and approved the final manuscript.

Acknowledgement

This paper is based in part upon the work supported by the National Science and Engineering Research Council (NSERC) of Canada. Any opinions, findings, C and conclusions or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the NSERC.

Received: 3 March 2013 Accepted: 23 August 2013

Published: 3 September 2013

References

- Arthur, D., & Vassilvitskii, S. (2007). "k-means++: the advantages of careful seeding". In *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms*.
- Adan, A., & Huber, D. (2011). "3D reconstruction of interior wall surfaces under occlusion and clutter". Hangzhou, China: Proc. of 3DPVT 2011.
- Adan, A., Xiong, X., Akinci, B., & Huber, D. (2011). "Automatic creation of semantically rich 3D building models from laser scanner data". Seoul, Korea: Proc. of ISARC 2011.
- Biswal, S. (2011). "Xbox 360 Kinect review, specification and price". <http://www.techibuzz.com/xbox-360-kinect-review/> Accessed January 10, 2013.
- Bosche, F., & Haas, C. (2008). Automated retrieval of 3D CAD model objects in construction range images. *Auto in Construction*, 17(4), 499–512.
- Brilakis, I., German, S., & Zhu, Z. (2011). Visual pattern recognition models for remote sensing of civil infrastructure. *Journal of Computing in Civil Engineering*, 25(5), 388–393.
- Brilakis, I., Soibelman, L., & Shinagawa, Y. (2006). Construction site image retrieval based on material cluster recognition. *Journal of Advanced Engineering Informatics*, 20(4), 443–452.
- ClearEdge3D. (2012). "3D imaging services reduces MEP workflow on Pharma project by 60% using edgewise plant 3.0". <http://www.clearedge3d.com/> Accessed June 10, 2012.
- Durrant-Whyte, H., & Bailey, T. (2006). Simultaneous localization and mapping. *IEEE Robotics and Automation Magazine*, 13(2), 99–110.
- Escorcia, V., Golparvar-Fard, M., & Niebles, J. (2012). "Automated vision-based recognition of construction worker actions for building interior construction operations using RGBD cameras" (pp. 879–888). West Lafayette, Indiana: ASCE Construction Research Congress.
- Fathi, H., & Brilakis, I. (2012). "A videogrammetric as-built data collection framework for digital fabrication of sheet metal roof panels". Hersching, Germany: The 19th EG-ICE Workshop on Intelligent Computing in Engineering. July 4–6, 2012.
- Fioraio, N., & Konolige, K. (2011). "Realtime visual and point cloud slam". In *Proc. of the RGB-D workshop on advanced reasoning with depth cameras at robotics: Science and Systems Conf. (RSS)*.
- Foltz, B. (2000). Application: 3D laser scanner provides benefits for PennDOT bridge and rockface surveys. *Professor Survey*, 20(5), 22–28.
- Furukawa, Y., Curless, B., Seitz, S., & Szeliski, R. (2009). "Reconstructing building interiors from images". Kyoto, Japan: Proc. of ICCV.
- Furukawa, Y., & Ponce, J. (2010). Accurate, dense, and robust multi-view stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), 1362–1376.
- Geiger, A., Zeigler, J., & Stiller, C. (2011). *Stereoscan: dense 3D reconstruction in real-time*. BadenBaden, Germany: Proc. of IEEE Intelligent Vehicles Symposium.

- Henry, P, Krainin, M, Herbst, E, Ren, X, & Fox, D. (2010). *RGB-D Mapping: using depth cameras for dense 3D modeling of indoor environments*. Delhi, India: Proc. of International Symposium on Experimental Robotics.
- Huber, D, Kapuria, A, Donamukkala, R, & Hebert, M. (2004). *"Parts-based 3D object recognition"*. Washington DC: Proc. of CVPR 2004.
- Institute for BIM in Canada (IBC). (2011). *"Environmental scan of BIM tools and standards"*. Canadian Construction Association. <http://www.ibt-bim.ca/documents/Environmental%20Scan%20of%20BIM%20Tools%20and%20Standards%20FINAL.pdf> Accessed Oct. 25 2012.
- Izadi, S, et al. (2011). *"KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera"*. Santa Barbara, CA: ACM Symposium on User Interface Software & Technology (UIST).
- Kazhdan, M, Funkhouser, T, & Rusinkiewicz, S. (2003). *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing (SGP '03)*. pp. 156–164. Aire-la-Ville, Switzerland: Eurographics Association.
- Liu, X, Eybpoosh, M, & Akinci, B. (2012). *"Developing as-built building information model using construction process history captured by a laser scanner and a camera"* (pp. 1232–1241). West Lafayette: Construction Research Congress (CRC). IN., May, 21–23, 2012.
- Mikolajczyk, K, & Schmid, C. (2005). A Performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10), 1615–1630.
- Muja, M, & Lowe, D. (2012). *"Fast matching of binary features"*. In *Proceedings of the 2012 Ninth Conference on Computer and Robot Vision (CRV '12)* (pp. 404–410). Washington, DC, USA: IEEE Computer Society.
- Muja, M, & Lowe, D. (2009). *Fast approximate nearest neighbors with automatic algorithm configuration* (International Conference on Computer Vision Theory and Applications, pp. 331–340). Lisboa, Portugal.
- Müller, P, Zeng, G, Wonka, P, & Van Gool, L. (2007). "Image-based procedural modeling of facades". *ACM Transactions on Graphics*, 26, 1–9. No. 3, Article 85.
- Neto, JA, Arditi, D, & Evens, MW. (2002). Using colors to detect structural components in digital pictures. *Computer Aided Civil Infrastructure English*, 17, 61–67.
- Okorn, B, Xiong, X, Akinci, B, & Huber, D. (2010). *"Toward automated modeling of floor plans"*. Paris, France: Proc. of 3DPVT 2010.
- Patterson, A, IV, Mordohai, P, & Daniilidis, K. (2008). *"Object detection from large-scale 3D datasets using bottom-up and top-down descriptors"*. Marseille, France: Proc. of ECCV 2008.
- Pettee, S. (2005). *"As-builts - problems and proposed solutions"* (CM eJournal, First Quarter: 1-19). <http://cmaanet.org/files/as-built.pdf> (June 10, 2012).
- Pollefeys, M, Nister, D, Frahm, JM, Akbarzadeh, A, Mordohai, P, Clipp, B, Engels, C, Gallup, D, Kim, SJ, Merrell, P, Salmi, C, Sinha, S, Talton, B, Wang, L, Yang, Q, Stewenius, H, Yang, R, Welch, G, & Towles, H. (2008). Detailed real-time urban 3D reconstruction from video. *International Journal of Computer Vision*, 78(2), 143–167.
- Rafibakhsh, N, Gong, J, Siddiqui, M, Gordon, C, & Lee, H. (2012). *"Analysis of XBOX Kinect sensor data for use on construction sites: depth accuracy and sensor interference assessment"* (pp. 848–857). West Lafayette: Constitution Research Congress. IN. May 21–23.
- Ripperda, N, & Brenner, C. (2009). *Application of a formal grammar to facade reconstruction in semiautomatic and automatic environments*. Hannover, Germany: Proc. of AGILE conference on geographic information science. 2009.
- Roth, H, & Vona, M. (2012). *"Moving volume KinectFusion"*. Surrey, UK: British Machine Vision Conf. (BMVC). September 2012.
- Rublee, E, Rabaud, V, Konolige, K, & Bradski, G. (2011). *"ORB: an efficient alternative to SIFT or SURF"* (pp. 2564–2571). Barcelona, Spain: Proc. of ICCV 2011.
- Shan, Y, Sawhney, H, Matei, B, & Kumar, R. (2006). Shapeme histogram projection and matching for partial object recognition. *IEEE Transactions on Pattern Anal Mach Intell*, 28(4), 568–577.
- Snavely, N, Seitz, S, & Szeliski, R. (2006). *"Photo tourism: exploring image collections in 3D"*. MA, Boston: Proc. of SIGGRAPH 2006.
- Stiene, S, Lingemann, K, Nuchter, A, & Hertzberg, J. (2006). *"Contour-based object detection in range images"*. Chapel-Hill, NC: Proc. of 3DPVT 2006.
- Sturm, J, Engelhard, N, Endres, F, Burgard, W, & Cremers, D. (2012). *"A benchmark for the evaluation of RGB-D SLAM systems"*. In *Proc. of the International Conference on Intelligent Robot Systems (IROS)*. Oct. 2012.
- Tang, P, Huber, D, Akinci, B, Lipman, R, & Lytle, A. (2010). Automatic reconstruction of as-built building information models from laser-scanned point clouds: a review of related techniques. *Auto in Construction*, 19(7), 829–843.
- Wang, D, Zhang, J, Wong, H, & Li, Y. (2007). *"3D model retrieval based on multi-shell extended gaussian image"* (Lecture notes in computer science: advances in visual information systems, pp. 426–437).
- Weerasinghe, I, Ruwanpura, J, Boyd, J, & Habib, A. (2012). *"Application of microsoft kinect sensor for tracking construction workers"* (pp. 858–867). West Lafayette, Indiana: ASCE Construction Research Congress (CRC).
- Xiong, X, & Huber, D. (2010). *"Using context to create semantic 3D models of indoor environments"*. Aberystwyth, UK: Proc. of BMVC 2010.
- Zhu, Z, & Brilakis, I. (2010). "Concrete column recognition in images and videos". *Journal of Computing in Civil Engineering, American Society of Civil Engineers*, 24(6), 478–487. November/December 2010.
- Zhu, Z, & Donia, S. (2013). *"Potentials of RGB-D cameras in as-built indoor environments modeling"*. Los Angeles, CA: 2013 ASCE International Workshop on Computing in Civil Engineering. June 23–25, 2013.
- Zhu, Z. (2013). *"Effectiveness of visual features in as-built indoor environments modeling"*. In *13th International Conference on Construction Applications of Virtual Reality*. London, UK. October, 30-31 2013 (accepted).

doi:10.1186/2213-7459-1-10

Cite this article as: Zhu and Donia: Spatial and visual data fusion for capturing, retrieval, and modeling of as-built building geometry and features. *Visualization in Engineering* 2013 1:10.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com