

PARTICIPATORY IMAGE-BASED MODELS' ALIGNMENT FOR RECONSTRUCTING A LARGE-SCALE INDOOR MAPPING

Ali Baligh Jahromi, Gunho Sohn*, Jaewook Jung, Kunwoo Park, David Recchia

GeoICT Laboratory, Department of Earth, Space Science and Engineering, York University,
4700 Keele Street, Toronto, Ontario, Canada M3J 1P3 - (baligh, gsohn, jwjung, kwparkv, drecchia) @yorku.ca

Commission II, WG IV

KEY WORDS: 3D Model Reconstruction, Crowd Source Modeling, Model Transformation, Model Alignment, Topology Graph

ABSTRACT:

In this paper, we introduced a recently developed image-based model alignment technique for 3D reconstruction of large-scale indoor corridors. The proposed participatory model alignment technique enables crowd source single image-based modeling since it allows various participants to incorporate their images taken from different cameras for large-scale indoor mapping. This technique is robust against changes of camera orientation and prevents miss-association of a newly generated 3D model to the previously integrated models. To investigate the possibility of aligning two individual 3D models, their respective corridor topological graphs must match, and they need to geometrically transform into the same object space. Here 3D affine transformation is applied, and the transformation parameters are estimated through corresponding vertices of both 3D models. Having integrated two models in the same 3D space, they will be back projected into the image space for evaluation using Direct Linear Transformation. Note that the proposed method performs layout model matching in image space and considers information including layout topology and geometry as well as image information to address model alignment. The advantages of using layout information in the proposed alignment technique are twofold. First, a metric constraint is imposed to insure topological model consistency and balance 3D models scale issues. Second, it will reduce alignment ambiguity related to indoor corridor scenes, where the scene is enriched with multiple structural elements including various corridors junctions. To evaluate the performance of the proposed method, we have performed the experiments on a data set collected from Ross building corridors at York University. This dataset includes single images captured by a handheld wide-angle camera. The obtained results present the ability of the proposed method in alignment of single image-based 3D models while producing limited geometric errors.

1. INTRODUCTION

The rise of world's population and rapid changes in human's lifestyle increased the rate of urbanization (Lutz et al., 2017). This imminent development drastically impacts both our lives and environment and necessitates more constructions to be accomplished. Thus, cities will grow and consequently the urge to plan, manage and monitor for analysing and updating urban infrastructure would be indisputable. For example, geometric representation of city entities even in primitive formats are foremost to the urban structure management. Thus, primitive based geometric representation of indoor models could escalate the level of understanding in drawing up building information. Consequently, 3D indoor space models' reconstruction is indispensable for updating and analysing building information. Here, the main objective of this paper is to propose a method for alignment of single 3D indoor space models to achieve large scale indoor mapping.

Sensors used to accomplish indoor mapping tasks are multiple. The most notable ones are: a) 2D/3D laser scanners; b) perspective cameras in form of monocular, stereo, omnidirectional vision; c) sonar and radio frequency beacons; and d) depth (RGBD) cameras (Baligh Jahromi et al., 2018). Laser scanners are providing accurate dense 3D point clouds and ameliorating the automation level in reconstruction of geometric models (Jung and Sohn, 2019). Yet, image data can provide valuable semantic information to indoor mapping task. For instance, layout boundaries of indoor spaces could precisely be identified in an image. Images have been utilized as an

important data source for indoor modelling. Early efforts in this regard include manually digitizing images to detect indoor space layout. Thus, in this paper indoor mapping through monocular vision is presented. Monocular cameras can gather denser visual information from the environment compare to range sensors which are not usually cheap and light.

Some researchers in photogrammetry and computer vision fields dedicated their efforts and time to come up with accurate representations of various building entities (Lee et al., 2010; Schwing and Urtasun, 2012). Yet, some studies have been presented on detection, recognition and reconstruction of building indoor models (Hedau and Hoiem, 2010; Schwing et al., 2013; Zhang et al., 2014; Liu et al., 2015; Tang et al., 2016; Zhu et al., 2016; Huang et al., 2017 and Wang et al., 2018). Recently, new approaches in computer vision established the base for automatically reconstruct indoor space models. Structure from Motion (SfM) and Simultaneous Localization and Mapping (SLAM) are well-known techniques for generation of large-scale indoor maps from a group of images (Baligh Jahromi et al., 2018). Note that 3D modeling of an indoor space is highly related to indoor mapping, navigation tasks and autonomous systems. In recent years, various applications for indoor mapping and navigation services have been introduced by companies including Microsoft, Google and Apple (Tóth et al., 2015). It should be noted that geometrically accurate 3D indoor space models are indispensable for various spatial information-based applications such as indoor positioning, navigation and security (Ochmann et al., 2016 and Lehtola et al., 2017). Moreover, novel technologies like Mobile

* Corresponding author

Augmented Reality (MAR) bestow upon a platform to utilize 3D indoor space models while interacting with surroundings via a computer or mobile device.

In the past few years, many indoor modelling and 3D reconstruction techniques have been introduced that vary in terms of data sources (single data source vs. multi-data source), adopted data processing strategies (parametric, generic or hybrid) and levels of automation (semi-automatic vs. full-automatic). Note that coming about a novel technique to create geometrically accurate indoor maps in a fully automated manner is still a big challenge. Sohn and Dowman, (2007) mentioned some critical factors that must be appraised while developing a new modelling method. These factors could be a) scene complexity, b) sensor dependency and c) incomplete cues. Indoor mapping encounters with different information of non-layout objects (e.g., tables, paintings, chairs and other clutters) in addition to layout sections (e.g., floor, walls and ceiling). Furthermore, indoor scenes have different structures and formats which should not be described by a single standard type. Thus, we deliberately simplify complex indoor scenes to obtain a suitable interpretation for indoor mapping (Yang et al., 2018). So far, various techniques have been introduced for reconstructing 3D indoor space maps using different data sources. However, the proposed techniques have limitations due to modelling accuracies, levels of automation, inherent sensor dependency and ability to solve missing data problems. A promising approach to degrade these issues would be taking advantage of both data-driven and model-driven strategies while using images as the main data source. This approach is applied in 3D modelling of indoor spaces in our previous work (Baligh Jahromi and Sohn, 2015). Thus, in this paper, we propose a new method for large-scale indoor mapping through alignment of reconstructed single image-based 3D models. We spatially generate indoor corridor topology graph to make models alignment robust against miss association of newly generated models to unrelated aligned models. This large-scale indoor corridor model alignment method allows using topological, geometric and radiometric information of both indoor layout and image to globally align single indoor corridor models.

2. DATA SET AND CAMERA MODEL

Preparing a new data set was an essential part in evaluation of the proposed image-based model alignment algorithm. Although various types of data sets were available to assess the quality of a reconstructed 3D model from a single image, these data sets were mostly covering single rooms and they were not suitable for evaluation of an image-based 3D corridor model alignment algorithm. Thus, to assess the quality of the proposed algorithm in this paper, a new data set is prepared. This dataset is specifically designed to serve our research purposes and satisfy the needs for necessary algorithm assessments. The pinhole camera model is the adopted model in this data set. This model is simply describing the imaging process. In this model the camera is recognized by a flat surface (image plane) and a light-barrier hole that represents the camera perspective centre. Each image point can be represented by a ray of light (reversible optical path). To reconstruct this ray, camera interior orientation parameters (IOPs) are needed. Also, image observations must be undistorted for lens distortions to satisfy the collinearity condition. Camera parameters, in the prepared data set, were calibrated by MATLAB calibration toolbox (Bouquet, 2004). Here, we concisely describe the prepared data sets.

The first data set includes single images acquired by handheld cameras (GoPro Hero5 and Apple iPhone 4s) covering indoor corridor environments. Table 1 reveals the specifications of two cameras used for preparing this data set. Images have been taken while crawling through different indoor corridors at York University campus buildings in Toronto, Canada. Various places such as Petrie Science, Osgoode Hall, Behavioural Science, Chemistry and Ross buildings were included.

Camera	Image Format		Field of View (degree)	Focal Length (mm)
	Row (pixel)	Col (pixel)		
Apple iPhone 4s	2,448	3,264	56.423°	4.28
GoPro Hero5	3,000	4,000	149.20°	16.80

Table 1. Digital cameras used for data set preparation.

The main test site considered in this study encompasses connected corridors located at the first floor of Ross Building. This building was selected due to having Manhattan structure aligned indoor corridors and free accessibility over time. Figure 1 shows some images taken at Ross building interiors. The actual camera positions can be inferred from this figure along with the type of images that were used in our experiments. For this test site, reference 3D indoor corridor models and their respective orientation maps (in image space) are prepared by manually identifying corridor layouts in image space (positional errors of structural corner points were less than 3 image pixels). It should be noted that indoor corridors of the selected test site have simple and rectangular outlines. Since identifying the indoor corridor layout in a single image is sometimes a very challenging task even for human eyes. The second data set includes reference laser point clouds acquired from the same test site. To prepare this laser benchmark data set, Trimble Indoor Mobile Mapping Solution (TIMMS) was used (Stott, 2016). To geo-reference the incoming laser point clouds and improve TIMMS positional accuracy, several indoor survey control points (planar accuracies under 5mm) were delicately identified inside Ross building interiors by conducting precise indoor surveying. Note that TIMMS collected laser point clouds accuracies were close to 1cm relative to TIMMS positional accuracy. Finally, the incoming 3D laser point clouds were used to generate individual ground truth 3D models that further boost the evaluation of various aspects of the proposed technique.



Figure 1. Sample images acquired at Ross building first floor.

3. SINGLE IMAGE-BASED INDOOR CORRIDOR MODELING OVERVIEW

Previously, we presented a top-down approach for reconstruction of 3D indoor corridor models from single images (Baligh Jahromi and Sohn, 2015). The proposed method permits the reconstructed indoor corridor model to represent multiple indoor corridors. Since the proposed method is specifically designed to handle the presence of side corridors and occlusions, we used the modified version of this method (Baligh Jahromi et al., 2017) for estimation of single image-based indoor corridor layout in this paper. Here, the problem of 3D indoor corridor model reconstruction from a single image is tackled through middle-level perceptual organization (Baligh Jahromi and Sohn, 2016). The method searches for indoor corridor layouts that can be converted into a physically plausible 3D model. Based on the Manhattan Rule Assumption, the proposed method sequentially creates various physically valid layout hypotheses using image line segments. To find a generated hypothesis that best matches the scene, each hypothesis will be evaluated. Finally, the best fitting hypothesis will be translated into a 3D model using the estimated orthogonal vanishing points in the image space. Note that scene layout orthogonal directions will remain intact even if the camera orientation is changed while capturing photos by crawling through an indoor corridor. In a typical corridor scene with low textures, identified orthogonal directions can provide strong clues for aligning consecutive image-based 3D models.

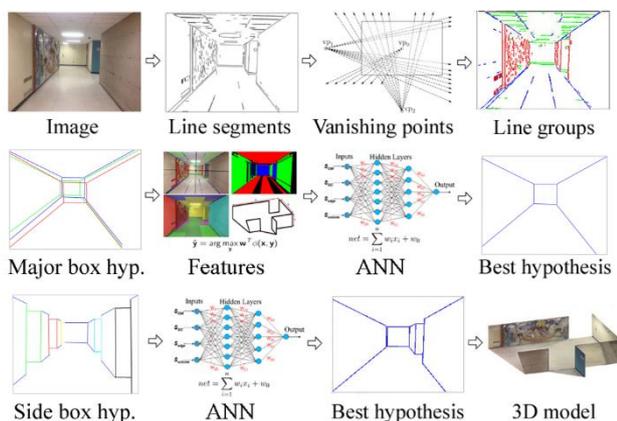


Figure 2. The proposed method extracts straight line segments and estimates orthogonal vanishing points. This method creates many layout hypotheses and uses a scoring function (parameters optimized by ANN) to evaluate them. Finally, the best fitting hypothesis will be converted into a 3D model.

It must be noted that the presented technique is introducing an approach to generate indoor corridors layouts in a hybrid way using both virtually generated rays from vanishing points and detected line segments. This technique is beneficial for two main reasons. First, the hybrid way of creating a scene layout provides a realistic solution while encountering with occlusions or objects in the scene and it is well-suited to describe many corridor spaces. It should be noted that using virtual rays solely for indoor layout creation may cause structural displacements from the true layout boundaries in lengthy corridors due to common inaccuracy of the estimated vanishing points.

Moreover, solely applying physical line segments for layout creation would be inefficient due to their ineptitude to handle occlusions. Second, the created indoor corridor model can be represented as a set of integrated individual corridors. Note that each corridor consists of various numbers of faces (maximum five) representing facades of a Manhattan form corridor. Thus, a corridor topological graph for each model can be created. Figure 2 shows the overall workflow of the introduced method. Not to repeat our previously published papers, details of the proposed method are not expressed in this paper, and readers are referred to our previous publication (Baligh Jahromi et al., 2017).

4. IMAGE-BASED MODEL ALIGNMENT

Participatory image-based model alignment is a new technique which aims to continuously update 3D models of indoor corridor environments abiding Manhattan World constraint. The proposed method considers principles of reconstructed 3D indoor corridor models from single images (Baligh Jahromi and Sohn, 2016) and performs layout model matching among the pool of generated 3D models. This method uses both 3D model information (topology and geometry) and image information (radiometric) to address model alignment. Thus, we made additional modifications to the initial indoor corridor modeling algorithm including the integration of vanishing point refinement technique and layout matching schemes to its structure. The following paragraphs summarize the overall procedure of the proposed model alignment technique in this paper. Figure 3 depicts the overall workflow of the proposed method.

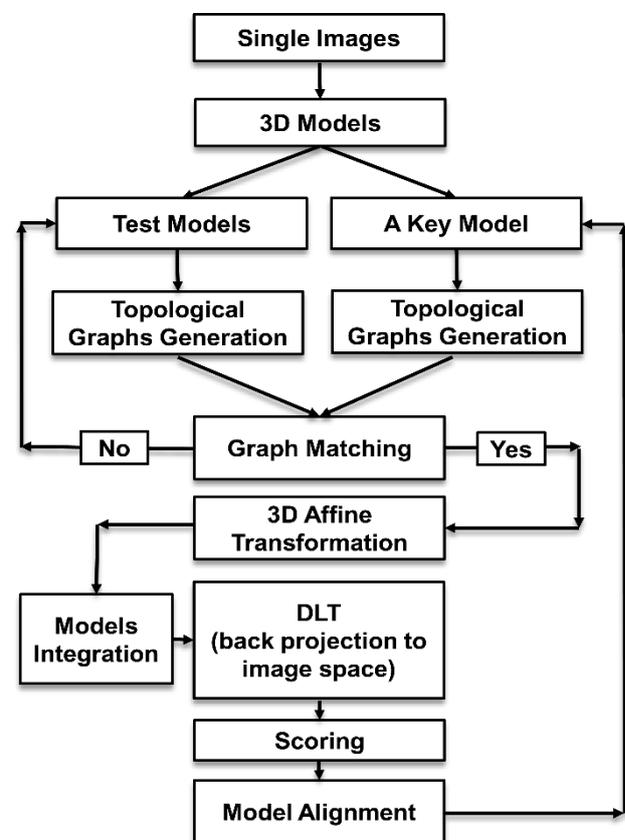


Figure 3. The workflow of the proposed image-based model alignment method.

Having various 3D indoor corridor models created from individual single images, the possibility of their alignments in both 3D and 2D spaces must be examined. Here, the procedure starts with considering a randomly chosen 3D model as the *key model* and the rest of the generated 3D models as *test models*. To identify the possible alignment between a test model and a key model, their respective corridor topological graphs must be compared. As mentioned, these corridor graphs are derived from previously generated individual models.

First, the complete structure of a key model is considered and for each side corridor in that structure a new topological graph will be generated. Therefore, the number of topological graphs for each key 3D model would be the same as the number of individual corridors in that model. Note that in the generation of new topological graphs for side corridors, the camera is hypothesised to stand inside a side corridor while facing major corridor of the original generated 3D model. More details in this regard will be presented in the following sections and readers are referred to our previous publication on integration of side corridors to the generated main corridor of a single image (Baligh Jahromi et al., 2018). Second, the corridor graph of a test model will be compared to all generated corridor graphs (including main and side corridor graphs) of the key model. If faces of a corridor topological graph which belong to a test model exactly match all the ones for any of the key model graphs, then vertices of those faces are considered as corresponding vertices.

Third, to test the possibility of having an alignment between a test model and the randomly selected key model, the test model must be geometrically transformed into the key model in 3D space. Thus, the corresponding vertices of both models are used to estimate the parameters of a 3D affine transformation using least square method. Forth, the newly transformed test model will be back projected into the original key model image (2D space) using Direct Linear Transformation (DLT). After back projection of the test model into the original key model image, a cost function is used to identify the optimal alignment. Here, the applied cost function consists of three terms considering both models topology, geometry, and radiometric similarities. The optimal alignment between a key model and a test model is determined by selecting a test model that minimizes the applied cost function. As mentioned previously, in this paper an image-based model alignment technique is proposed, and more details of this technique are presented in the following sections.

4.1 Corridor Topological Graph Generation

The main reason to adopt our single image-based indoor corridor model reconstruction method is its ability to identify multiple corridor models in a single image. Thus, the generated indoor corridor model not only represents the main corridor, but also visualize side corridors. Side corridors directly intersect with the main corridor's structure. Hence, their layout structure becomes part of the overall layout model. Normally, the presence of a side corridor in an image is identified by comparing the estimated main corridor layout geometric features to geometric features of an image. Conspicuous differences between these two geometric features would initiate the side corridor layout generation process (Baligh Jahromi et al., 2018). As mentioned, the presence of side corridors will increase the geometric complexity of the estimated indoor corridor layout structure. Thus, they can enrich the formation of the generated corridor's topological graph.

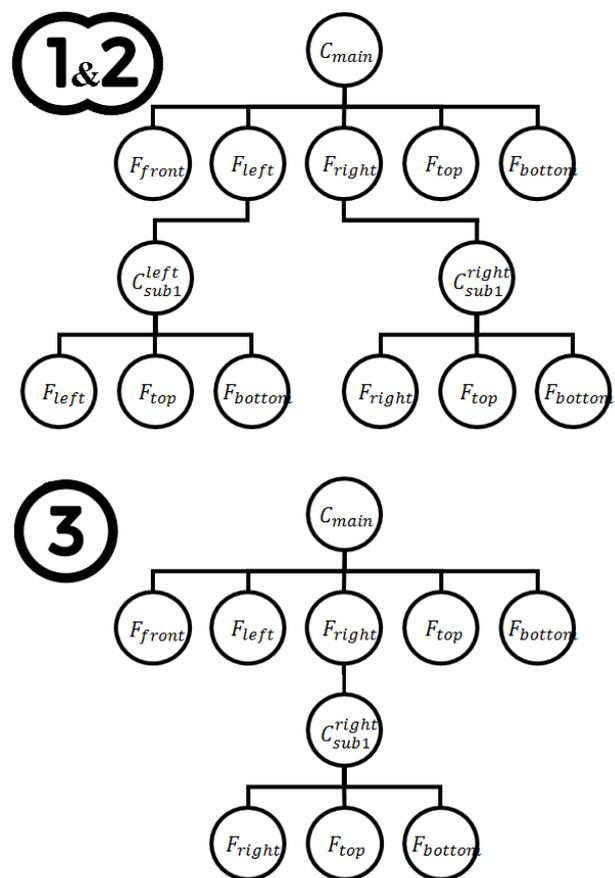


Figure 4. Top: A single image with specified corridor layout; Middle: Reconstructed 3D model; Camera is hypothesized to stand at various positions (1, 2 and 3); Bottom: The generated corridor topological graphs.

In order to explain how the corridor topological graph is created, following specific definitive terms should be presented first. Here, a 3D reconstructed model of an indoor corridor environment M can be denoted as a set of corridors $M = \{C_i | i = 1, 2, \dots, n\}$ which can contain n number of corridors. Each corridor may consist of m numbers of faces $C = \{F_j | j = 1, 2, \dots, m\}$ denoting front, bottom, top, left and right sides of a Manhattan form reconstructed 3D corridor. Note that always main corridors should be represented by five faces $C_{main} = \{F_{front}, F_{left}, F_{right}, F_{top}, F_{bottom}\}$ while side corridors may be represented by three faces $C_{sub}^{left} = \{F_{left}, F_{top}, F_{bottom}\}$ or $C_{sub}^{right} = \{F_{right}, F_{top}, F_{bottom}\}$. Here, left side and right side of a 3D model are determined based on the position and heading direction of the camera at the time of exposure. Hence, the attaching position of the side corridor structural vertices to the main corridor will be tagged respectively. Figure 4 depicts an image with its identified corridor layout in image space along with the respective 3D model, and the reconstructed corridor topological graphs for this 3D model.

It should be noted that based on the camera position and heading, the generated corridor topological graph may vary for an individual model. Figure 4 shows the camera is placed at various positions (1, 2 and 3). Thus, the respective corridor topological graphs are formed as shown in this figure. Here, side corridor numbering starts with a corridor which has the furthest distance to the camera. Thus, the corresponding side corridors should have similar numbering in their respective graph structure. Note that to investigate the possibility of having an alignment between two different 3D models, the first step is to find whether their respective corridor topological graphs match or not.

As mentioned previously, a 3D model will be randomly selected among the pool of reconstructed 3D models and considered as key model. For each corridor layout in the selected key model, a corridor topological graph must be generated. To do so, we assume the camera is standing at each corridor and facing an interior part of the reconstructed model. Figure 4 shows different positions for the camera and their respective corridor topological graphs. To have an alignment between a key model and other reconstructed models in the pool (test models), at least part of their respective corridor topological graphs must match. The matched faces in both models will have common layout vertices that can be used for estimating the appropriate transformation.

4.2 Model Transformation

To investigate the possibility of having an alignment between a test model and the selected key model which their respective corridor topological graphs match, we need to geometrically transform the test model into the key model space. To specify the corresponding vertices between a test model and the selected key model, their respective corridor topological graphs will be compared. If faces in a corridor topological graph match some faces of the other graph, the vertices belonging to those faces are declared as corresponding vertices. For instance, C_{sub1}^{left} of a test model should corresponds to C_{sub1}^{left} of the selected key model and for sure not to $C_{sub2, \dots, n}^{left}$. Thus, these corresponding vertices can be used to estimate the 3D transformation parameters. Generally, the transformation of an object in 3D space consists of three displacements, three scale differences, three axes rotations, and three shear parameters. This is the base

for the so called 12-parameter affine transformation. Here, the 12-parameter 3D affine transformation is applied as following:

$$X = a_0 + a_1x + a_2y + a_3z \quad (1)$$

$$Y = b_0 + b_1x + b_2y + b_3z \quad (2)$$

$$Z = c_0 + c_1x + c_2y + c_3z \quad (3)$$

In the above equations, X, Y and Z represent the object space coordinates of the key model matched vertices while x, y and z represent their corresponding layout vertices coordinates on the test model. Also, $a_0, a_1, a_2, a_3, b_0, b_1, b_2, b_3, c_0, c_1, c_2$ and c_3 are the 3D affine transformation parameters. Note that these parameters are calculated using the least square method. Here we select the minimum number of corresponding vertices to perform this transformation and reserve the rest of points for alignment evaluation. The evaluation process is expressed in the next section.

To be able to use the radiometric information of images in this evaluation process, we decided to back-project the integrated test model to the key model image space (figure 5). Thus, we can compare the corresponding faces radiometric information for model alignment evaluation. To perform this back-projection we used Direct Linear Transformation (DLT) equations for a single image as following:

$$x = \frac{a_1X + a_2Y + a_3Z + a_4}{a_9X + a_{10}Y + a_{11}Z + 1} \quad (4)$$

$$y = \frac{a_5X + a_6Y + a_7Z + a_8}{a_9X + a_{10}Y + a_{11}Z + 1} \quad (5)$$

In these equations, X, Y and Z represent the object space coordinates of the test model vertices while x and y represent the image coordinates of those vertices in the original key model's image space. Here, $a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}$, and a_{11} are DLT parameters. Note that these parameters are estimated using the relation between the key model vertices in 3D space and their corresponding points in the original key model's image.

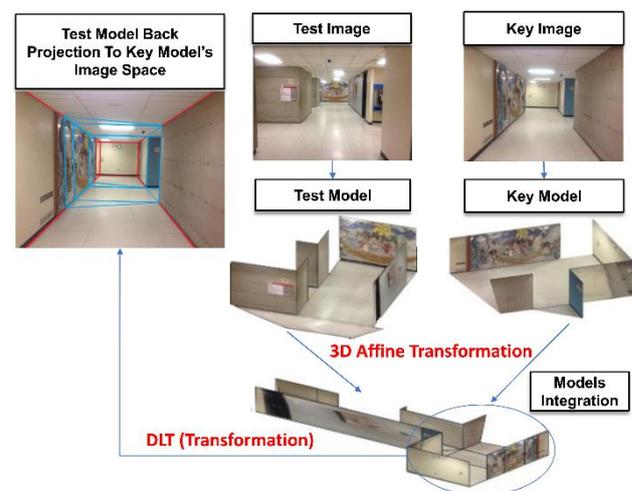


Figure 5. A schematic view of model transformation procedure.

4.3 Evaluation of Integrated Models

As mentioned before, having back-projected the integrated models into the selected key model's image space, this integration must be evaluated. Here, a cost function is used (Baligh Jahromi et al., 2018) to assess this integration and identify the optimal alignment. The proposed cost function includes three terms. These terms together assess the resemblance of test model's integrated parts to the key model's topological, geometric and radiometric information. Thus, the proposed cost function will consider both models topology, geometry, and radiometric similarities. The proposed cost function is as following:

$$Cost = 1 - [(w_R \times C_R) + (w_T \times C_T) + (w_G \times C_G)] \quad (6)$$

Where C_R , C_T , and C_G represent radiometric, topological and geometric similarities between the integrated models, respectively. Here, weight parameters are w_R , w_T , and w_G for C_R , C_T and C_G respectively. In our experiments, all cost function's weight parameters are considered as equal ($w_R = w_T = w_G = 1/3$). The radiometric similarity is defined by $C_R(t, k)$ which can be calculated by comparison of average image pixels values for corresponding layout faces $F_t \cap F_k$. Here, subscripts t and k represent test model and key model respectively. For each layout face, the average pixels values in Red, Green and Blue bands are measured. These calculated values will be assigned to each layout face. For corresponding faces in both test model and key model (F_t and F_k), their colour differences sum in three (R, G, B) bands r_{tk} is calculated. If the calculated value is less than a predefined threshold ($T_1=30$ in this paper), the indicator function δ_R would be 1 and 0 otherwise as following:

$$C_R(t, k) = \frac{\sum_{F_t \cap F_k} \delta_R}{N}, \quad \delta_R = \begin{cases} 1 & \text{if } r_{tk} \leq T_1 \\ 0 & \text{if } r_{tk} > T_1 \end{cases} \quad (7)$$

In the above equation, N represents the total number of corresponding faces between two models. As mentioned before, the corridor topological graph is generated for each individual model. Considering the generated topological graphs of a test model and a key model, the comparison of the total number of faces ($F_t \cup F_k$) and the common faces ($F_t \cap F_k$) between two models can define the topological similarity $C_T(t, k)$ as following:

$$C_T(t, k) = \frac{num(F_t \cap F_k)}{num(F_t \cup F_k)} \quad (8)$$

Finally, the geometric similarity $C_G(t, k)$ of two models is measured by calculating distances d_{tk} between the corresponding faces vertices. If the calculated distance for a pair of corresponding vertices $V_t \cap V_k$ is less than a predefined threshold ($T_2=50$ pixel in this paper), then the indicator function δ_G which measures geometric similarity is 1, and 0 otherwise:

$$C_G(t, k) = \frac{\sum_{V_t \cap V_k} \delta_G}{N}, \quad \delta_G = \begin{cases} 1 & \text{if } d_{tk} \leq T_2 \\ 0 & \text{if } d_{tk} > T_2 \end{cases} \quad (9)$$

In the above equation, N represents the total number of corresponding vertices for models under question. Having calculated all the costs for integrated test models to the selected key model, the optimal model alignment is determined by selecting a test model that minimizes the cost function as following:

$$M^* = \arg \min_{M_t} Cost(M_t) \quad (10)$$

If the minimum cost is more than a user-defined threshold ($T_3=0.15$ in this paper), then the test model is considered not to be aligned with the chosen key model. It should be noted that control parameters, weight values and thresholds used in this paper are empirically adjusted. Thus, the proposed method's performance in different conditions would be examined in future works.

5. EXPERIMENTS

To evaluate the performance of the proposed model alignment method, the new data set is used. As mentioned before, two different cameras are used for collecting images in this data set. Thus, the generated models from these two image categories would have different qualities which makes models alignment cumbersome. Since the geometric quality of the generated models in this paper is highly dependent on the accuracy of the estimated vanishing points in image space, we decided to improve model's quality prior to their alignments. In this paper, LSD method is adopted to accurately detect straight line segments in an image and the proposed method by Lee et al., (2010) is used for vanishing point estimation. To improve the vanishing point estimation results, the number of participant line segments is reduced, and a line refinement technique is applied. A line support region is created by grouping fragmented local straight-line segments which share the same level-line angle up to a specific threshold ($T_3=3^\circ$ in this paper). Thus, a rectangle can be associated with these local group of line segments which covers the whole line support region. This rectangle could represent those line segments if its angle corresponds to the level-line angle of the inside fragmented straight-line segments. This rectangle's angle is introduced as the angle of a line that represents these local straight-line segments in vanishing point estimation process. This refinement technique is tested on York Urban dataset images, where the average of 537-line segments per image were detected. Table 2 reveals the vanishing point estimation results before and after applying this line refinement technique while using the lengthier line segments.

York Urban dataset images	# of line segments (LSD)	Vanishing point estimation errors (average)		
		Frame	Vertical	Horizontal
No Refinement	100	1.6861°	2.3643°	1.1346°
Refinement applied	100	1.3969°	2.0011°	1.0093°

Table 2. Vanishing point estimation on York Urban dataset.

Having improved the accuracy of the estimated vanishing points and consequently the geometric quality of the reconstructed 3D models, the model alignment experiments are performed. As

mentioned, the main contribution of this paper is the introduction of a new method for image-based model alignment. Here, the prepared data set is used which contains images from 7 integrated corridors at Ross building. Collectively 27 images were selected which covered a closed corridors loop at Ross building first floor. Consequently, 27 image-based 3D models were reconstructed. The proposed model alignment technique was applied on the generated models.

In this paper a new method for aligning a key model to a collection of test models in 3D space is introduced. The proposed model alignment method was applied by transforming and back projecting a test model into the selected key model's image space and calculating alignment cost. The alignment cost is calculated by an established cost function presented in previous section. Figure 6 depicts two back projected sample test layout models (blue lines) into two other key layout models in image space (red lines).

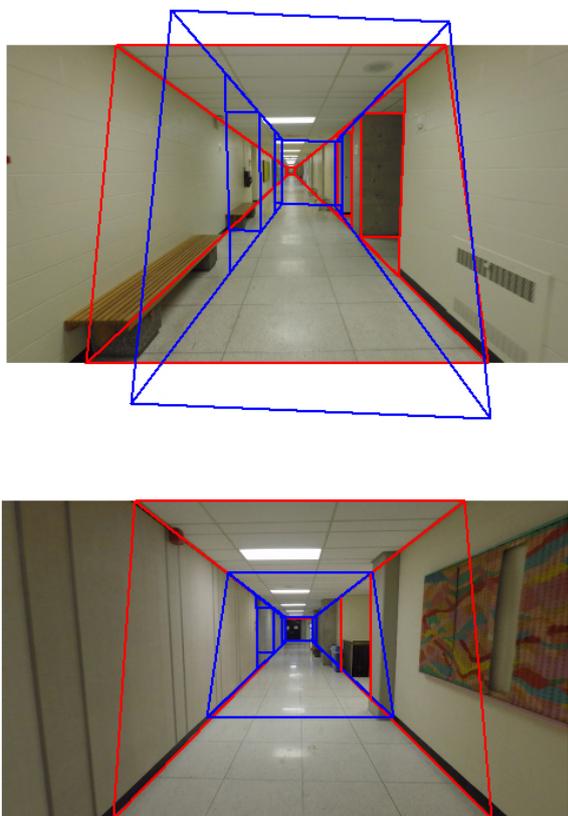


Figure 6. Test layout models (blue lines) and key layout models (red lines) are back projected into the key model's images.

Table 3 shows the calculated costs for aligning a random key model (#07) to the collection of test models. In this table only the lowest costs are presented. As shown in this table, key model #07 is well aligned with the test model #19 showing 0.14 matching cost. The threshold T_3 ($Cost \leq T_3 = 0.15$) is applied for accepting an alignment between two models. The cost function considers radiometric, topological and geometric similarities between two models in question.

Key Model #07		C_R	C_T	C_G	$Cost$
Test Model	#19	1.000	1.000	0.580	0.140
	#20	1.000	0.700	0.510	0.263
	#21	0.500	0.500	0.280	0.573
	#14	0.250	1.000	0.430	0.440
	#15	0.250	0.800	0.240	0.570

Table 3. Quantitative assessment of aligning a key model (#07) to the selected test models.

Since the reconstructed models are directly derived from single images, they may not integrate with each other perfectly in the object space. Thus, minor adjustments must be implemented to preserve the orthogonality and consistency of the generated indoor map. Here, always the key model is considered as true and the necessary changes were applied to the corresponding test models for alignment in object space. Figure 7 shows the incoming photo-textured indoor map generated by aligning 27 image-based 3D models of Ross building data set.



Figure 7. The generated photo-textured indoor map by aligning individual reconstructed image-based 3D models.

As stated previously, geometrically accurate reference 3D layout models (about 2cm accuracy) were associated with the prepared dataset. Accordingly, geometric comparison of the generated indoor map and the laser-based ground truth data (point cloud data) is performed. Note that accurate 3D indoor maps were manually extracted from TIMMS point cloud. For instance, the corridor section shown at the bottom-left of figure 7, has 2.92m width, 2.63m height and 24.14m length. The comparison showed that average ratio differences in widths, heights and lengths of individual corridors in the generated indoor map (approximately 112m path) were 2.3%, 1.9% and 7.62% respectively. Due to the high possibility of geometric errors accumulation during models' alignment, these comparisons were obtained after loop closing technique is applied on the aligned 3D models. For more information on the adopted loop closing technique readers are referred to (Baligh Jahromi et al., 2018).

6. CONCLUSIONS

In this paper, we presented a new participatory image-based models' alignment to reconstruct a large-scale indoor mapping. The focus of the proposed method is to take advantage of both 2D and 3D information of the reconstructed models. Therefore, different transformations including 3D affine transformation and Direct Linear Transformation are used to appropriately incorporate models' information. Geometrically accurate models' integration and image data association are playing a great role in the proposed model alignment technique. Here, radiometric, topological and geometric information of reconstructed image-based corridor models are considered for accurate data association. The new approach of corridor model topological graph reconstruction distinct the newly proposed technique from previous works. The proposed technique is examined on a specific data set prepared at York University. The generated results depict the ability of the proposed method to successfully identify models' alignment instances. Results comparison to the ground truth data showed average ratio differences in widths, heights and lengths of individual corridors of generated indoor map were 2.3%, 1.9% and 7.62% respectively. Thus, the proposed models' alignment technique enables individually reconstructed image-based 3D models to align with each other while producing limited mapping errors.

ACKNOWLEDGEMENTS

This research was supported by Government of Ontario, NSERC Discovery and York University. In addition, the authors wish to extend gratitude to Mr. Kivanc Babacan who consulted on this research.

REFERENCES

- Baligh Jahromi, A. and Sohn, G., 2015. Edge Based 3D Indoor Corridor Modeling Using a Single Image. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume II-3/W5, pp. 417-424.
- Baligh Jahromi, A. and Sohn, G., 2016. Geometric Context and Orientation Map Combination for Indoor Corridor Modeling Using a Single Image. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41. Volume XLI-B4, pp. 295-302.
- Baligh Jahromi, A., Sohn, G., Shahbazi, M. and Kang, J. 2017. A Preliminary Work on Layout SLAM for Reconstruction of Indoor Corridor Environments. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4.
- Baligh Jahromi, A., Sohn, G., Jung, J., Shahbazi, M. and Kang, J., 2018. Layout SLAM with Model-based Loop Closure for 3D Indoor Corridor Reconstruction. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4(2).
- Bouguet, J. Y., 2004. Camera calibration toolbox for Matlab.
- Hedau, V., and Hoiem, D., 2010. Thinking inside the box: using appearance models and context based on room geometry. In: *Proceedings of the European Conference on Computer Vision*, pp. 1-14.
- Huang, A.S., Bachrach, A., Henry, P., Krainin, M., Maturana, D., Fox, D. and Roy, N., 2017. Visual odometry and mapping for autonomous flight using an RGB-D camera. In *Robotics Research* (pp. 235-252).
- Jung, J. and Sohn, G., 2019. A line-based progressive refinement of 3D rooftop models using airborne LiDAR data with single view imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149, pp.157-175.
- Lee, D.C., Gupta, A., Hebert, M., Kanade, T., 2010. Estimating spatial layout of rooms using volumetric reasoning about objects and surfaces. In: *Advances in Neural Information Processing Systems*, pp. 1288-1296.
- Lehtola, V.V., Kaartinen, H., Nüchter, A., Kaijaluoto, R., Kukko, A., Litkey, P., Honkavaara, E., Rosnell, T., Vaaja, M.T., Virtanen, J.P. and Kurkela, M., 2017. Comparison of the selected state-of-the-art 3D indoor scanning and point cloud generation methods. *Remote Sensing*, 9(8), p.796.
- Lutz, W., Butz, W.P. and Samir, K.E. eds., 2017. *World population & human capital in the twenty-first century: an overview*. Oxford University Press.
- Ochmann, S., Vock, R., Wessel, R. and Klein, R., 2016. Automatic reconstruction of parametric building models from indoor point clouds. *Computers & Graphics*, 54, pp.94-103.
- Schwing, A. G., Urtasun, R., 2012. Efficient Exact Inference for 3D Indoor Scene Understanding. In: *Proceedings of the European Conference on Computer Vision*, pp. 299-313.
- Schwing, A. G., Fidler, S., Pollefeys, M., Urtasun, R., 2013. Box in the box: Joint 3d layout and object reasoning from single images. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 353-360.
- Sohn, G., and Dowman, I., 2007. Data fusion of high-resolution satellite imagery and lidar data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1), pp. 43-63.
- Stott, A., 2016. High-Speed Indoor Mobile Mapping and Modeling. *GeoInformatics*, 19(2), p.8.
- Tóth, Z., Magnucz, P., Németh, R. and Tamás, J., 2015. Data model for hybrid indoor positioning systems. *Production Systems and Information Engineering*, 7(1), pp.67-80.
- Wang, C., Hou, S., Wen, C., Gong, Z., Li, Q., Sun, X. and Li, J., 2018. Semantic line framework-based indoor building modeling using backpacked laser scanning point cloud. *ISPRS journal of photogrammetry and remote sensing*, 143, pp.150-166.
- Yang, Y., Jin, S., Liu, R., Bing Kang, S. and Yu, J., 2018. Automatic 3D Indoor Scene Modeling from Single Panorama. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3926-3934).
- Zhang, Y., Song, S., Tan, P., & Xiao, J., 2014. PanoContext: A whole-room 3D context model for panoramic scene understanding. In: *Proceedings of the European Conference on Computer Vision*, pp. 668-686.
- Zhu, H., Weibel, J.B. and Lu, S., 2016. Discriminative multi-modal feature fusion for RGBD indoor scene recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2969-2976.