Research Article

Automatic 3D City Modeling Using a Digital Map and Panoramic Images from a Mobile Mapping System

Hyungki Kim,[1] Yuna Kang,[2] and Soonhung Han[1]

[1] *Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea*
[2] *1st R&D Institute, Agency for Defense Development, 488 Bugyuseong-daero, Yuseong-gu, Daejeon 305-152, Republic of Korea*

Correspondence should be addressed to Soonhung Han; shhan@kaist.ac.kr

Three-dimensional city models are becoming a valuable resource because of their close geospatial, geometrical, and visual relationship with the physical world. However, ground-oriented applications in virtual reality, 3D navigation, and civil engineering require a novel modeling approach, because the existing large-scale 3D city modeling methods do not provide rich visual information at ground level. This paper proposes a new framework for generating 3D city models that satisfy both the visual and the physical requirements for ground-oriented virtual reality applications. To ensure its usability, the framework must be cost-effective and allow for automated creation. To achieve these goals, we leverage a mobile mapping system that automatically gathers high-resolution images and supplements sensor information such as the position and direction of the captured images. To resolve problems stemming from sensor noise and occlusions, we develop a fusion technique to incorporate digital map data. This paper describes the major processes of the overall framework and the proposed techniques for each step and presents experimental results from a comparison with an existing 3D city model.

1. Introduction

Three-dimensional city models are widely used in applications in various fields. Such models represent either real or virtual cities. Virtual 3D city models are frequently used in movies or video games, where a geospatial context is not necessary. Real 3D city models can be used in virtual reality, navigation systems, or civil engineering, as they are closely related to our physical world. Google Earth [1] is a well-known 3D representation of real cities. It illustrates the entire earth using satellite/aerial images and maps superimposed on an ellipsoid, providing high-resolution 3D city models.

In the process of 3D city modeling, both the cost and the quality requirements must be considered. The cost can be estimated as the time and resource consumption of modeling the target area. The quality factor considers both visual quality and physical reliability. The visual quality is proportional to the degree of visual satisfaction, which affects the level of presence and reality. The physical reliability is the geospatial and geometrical similarity between the objects—in our case, mainly buildings—in the modeled and physical worlds.

Generally, accomplishing a satisfactory level for both requirements is difficult.

Numerous techniques can be used for 3D city modeling. For instance, color and geometry data from LiDAR are mainly used if the application requires detailed building models for a small area. If the city model covers a large area and does not need detailed features, reconstruction from satellite/aerial images is more efficient [2]. This means that the effective approach can differ according to the target application using the 3D city model. The goal of our research is to propose a 3D city modeling method that can be applied in ground-oriented and interactive virtual reality applications, including driving simulators and 3D navigation systems, which require effective 3D city modeling methods for diverse areas.

2. Related Work

Existing 3D city modeling methods can be divided into manual modeling methods, BIM (Building Information Model) data-based methods, LiDAR data-based methods, and image-based methods.

The manual modeling method is highly dependent on the modeling experts. Older versions of Google Earth and Terra Vista employed this method in their modeling systems. Although current applications employ manual modeling because of its high quality, the method is also a high-cost, labor-intensive process. Hence, it is not efficient for urban environments that include numerous buildings. The BIM data-based method facilitates the use of building design data from the construction stage and is applied in city planning [3] and fire and rescue scenario [4]. However, this method is only efficient for applications in which the activity areas are strictly constrained, as gathering BIM data is problematic or even impossible given the large size of urban environments. Moreover, the BIM data should be postprocessed for the use of virtual reality applications. This is because the information in BIM does not contain the as-built 3D model; so the properties for the visual variables should be mapped.

To address these problems, remote sensing techniques are being aggressively adopted and studies of LiDAR data-based methods and image-based methods are increasingly common [5]. LiDAR is a device that samples precise data from the surrounding environment using laser scanning technology. In several studies (e.g., [6, 7]), high-quality 3D city models have been reconstructed using LiDAR data. However, as noted in other work [6], ground-level scanning has a limited data gathering range, meaning that redundant data collection is unavoidable in the modeling of diverse areas, whereas airborne LiDAR [8] is limited in terms of its cost and color data collection methods.

Image-based methods include those based on stereo matching and inverse procedural modeling approaches. In earlier research [7, 9], a method based on stereo matching was used to recover 3D data from the feature point matching between a series of images. This approach usually requires numerous images to satisfy the accuracy and robustness requirements of feature point matching. Several recent inverse procedural modeling approaches [10–12] have modeled buildings using relatively few (mainly one) images. This can overcome the difficulties of data-collection in stereo matching. This approach employs a plausible assumption; that is, that the shape of a building consists of a set of planes in three dimensions, to reconstruct individual 3D buildings without pixel-wise, 3D information. However, because image-based methods are not robust against instances of occlusion, user input or strong constraints are frequently necessary. This reduces their cost effectiveness and/or physical reliability.

In our research, the approach which preserves the cost-efficiency by using the existing image database while increasing physical reliability will be proposed. The image database is relatively easy to access than LiDAR database so cost on the data collection can be decreased. On the other hand the method based on the stereo matching requires numerous images on the large-scale modeling that decreases the universal applicability of method. Therefore inverse procedural modeling approach is preferred on our objective, while the physical reliability can be increased by combining accurate reference data [13].

3. Proposed Method

3.1. Mobile Mapping System and Digital Map. In this study, we propose a framework that uses a massive number of images gathered from a mobile mapping system (MMS). This addresses many problems in existing methods, which cannot simultaneously provide feasible levels of cost effectiveness, visual quality, or physical reliability. An MMS collects and processes data from sensor devices mounted on a vehicle. Services such as Google Street View [14], Naver Maps [15], and Baidu Maps [16] present information in the form of high-resolution panoramic images that include the geospatial position and direction of each image taken. The main focus of these services is to offer visual information about the surrounding environment at a given location. The advantages of data collected from MMS are as follows.

(1) Nationwide or even worldwide coverage following the development of remote sensing technologies and map services.

(2) Rich, visual, and omnidirectional information.

(3) Sensor information that allows geospatial coordination with the physical world.

Using these advantages, we can model a diverse city area for ground-oriented interactive systems in a cost-effective way with the existing image database. Moreover, high visual quality at ground level can be provided by high-resolution panoramic images [17]. However, there are currently several disadvantages in the data collected from MMS.

(1) Sensor data includes noise, which lowers its physical reliability.

(2) The number of images in a given area is limited and is insufficient for stereo matching-based reconstruction.

(3) Inclusion of an enormous amount of unnecessary visual information, including occlusions, cars, and pedestrians.

Noise is unavoidable in the sensing process. The amount of error this introduces differs according to the surrounding environment; a ± 5 m positional error and $\pm 6°$ directional error have been reported in Google Street View data [18]. Such error levels can be problematic in the analysis required for 3D modeling. Moreover, the current service has an interval of ~10 m between images, which lowers the possibility of successful reconstruction using stereo matching. Additionally, the uncontrolled collection environment results in a severe disadvantage for inverse procedural modeling. MMS data also requires an additional process to classify individual buildings, unlike the inverse procedural modeling approaches.

To address these problems with MMS data, we propose a method that incorporates 2D digital map data. Digital maps have accurate geospatial information about various features in the physical world. For instance, the 1 : 5000 digital maps applied in our framework have a horizontal accuracy of 1 m, which is five times better than that of the MMS position data.
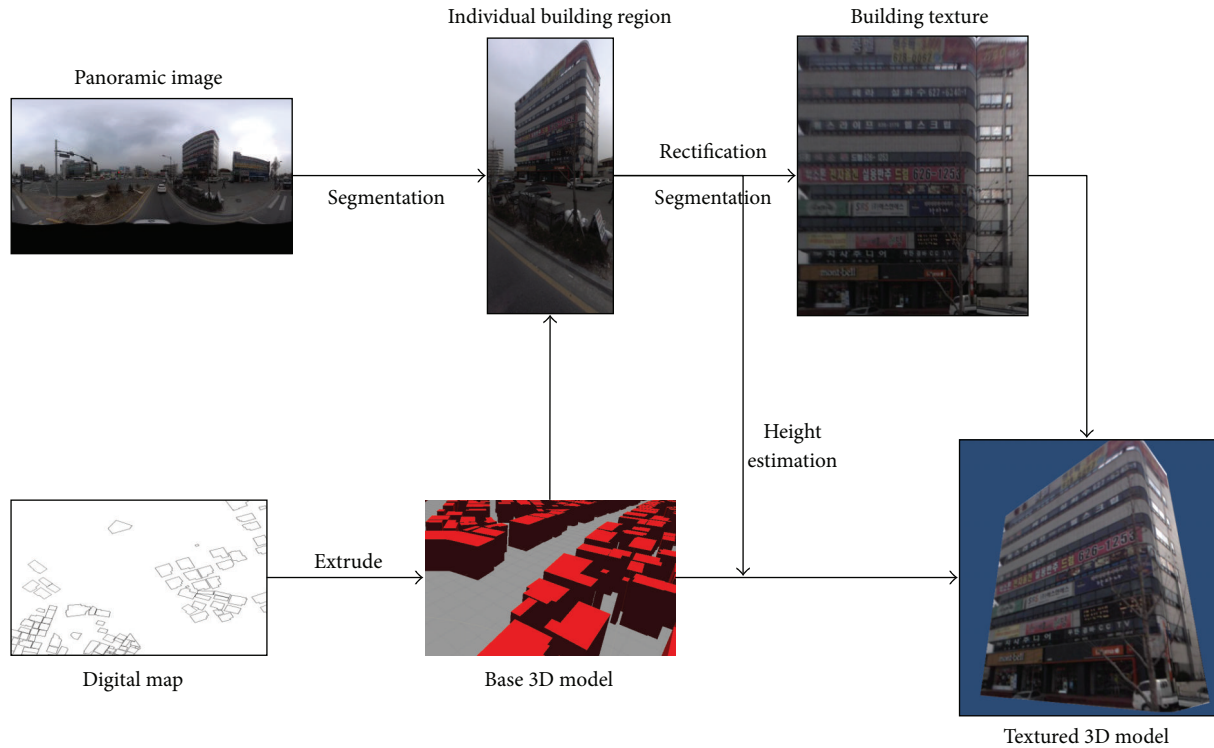
FIGURE 1: Overview of the proposed 3D city modeling framework.

Therefore, by combining data, the problems of sensor errors can be overcome and the selective use of visual information is possible. On the other hand, the geometrical characteristic of the building is restricted to a quasi-Manhattan world model. The quasi-Manhattan world model is the assumption that structures consist of vertical and horizontal planes, and is an extension of the Manhattan world model that assumes structures consist of vertical and horizontal planes orthogonal to each other.

3.2. Process Overview. The proposed framework is illustrated in Figure 1. The input data are the aforementioned digital maps, which contain building footprint information and panoramic images from the MMS system with sensor data. The base 3D model is generated from the footprint information of the buildings; the individual building regions are segmented and reprojected according to the combined GPS/INS (Inertial Navigation system) information. The reprojected region is further segmented and rectified to produce the texture image. Height estimation is possible by combining the building contour information from the texture image and the reprojected image. We can then obtain the textured 3D model by applying the height information to modify the height of the base 3D models.

The detailed process is illustrated in Figure 2. The entire modeling procedure can be divided into the following four stages.

(1) Image/building analysis.

(2) Error correction/compensation.

(3) Segmentation and validation.

(4) Texture mapping.

The error correction/compensation and segmentation and validation processes include a feedback loop to sequentially obtain the texture of individual buildings.

3.3. Image/Building Analysis. This stage analyzes the correlation between each image and the digital map. To do this, a base 3D model is generated from the footprint of the buildings by extending the model in the vertical direction. The footprint data consists of the geospatial coordinates of the building contour projected to the ground surface. This data should retain a certain level of accuracy and therefore contains precise information about the buildings.

The input image can then be positioned in the 3D environment according to the GPS/INS sensor information. Next, buildings are classified according to the three criteria listed below. The objective of this classification is to separate the buildings into texture-acquirable examples and others at a high resolution. The proposed criteria are as follows.

(1) The distance between the location from the GPS sensor corresponding to the image and the building.

(2) The occlusion between the buildings.

(3) The region occupied by the building in the image.

The distance criterion is quite straightforward: more distant buildings are less likely to appear in the image. The occlusion criterion is also reasonable, as an occluded building
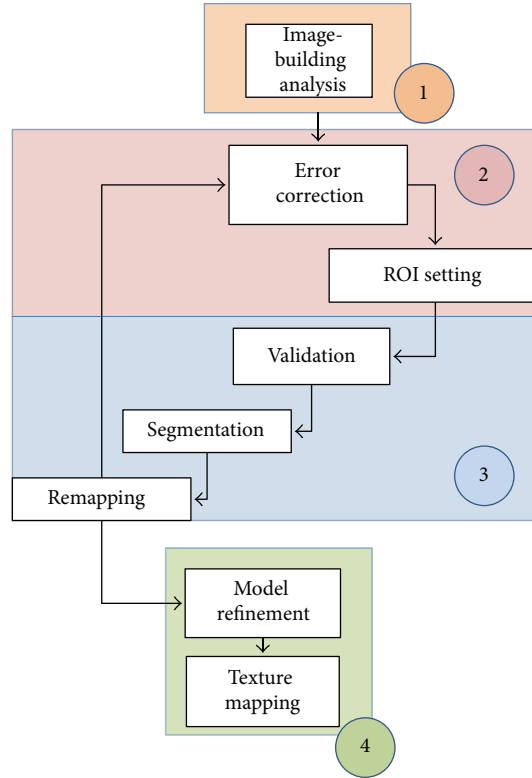
FIGURE 2: Detailed process of the proposed approach.

cannot appear in the image. The third criterion considers information about the brief texture resolution of each image by calculating the façade angle and width of the image, and then estimating the resolution of the texture. Assuming W is the width of the panoramic image that has a 360-degree field of view (FOV), and then the pixel per radian in the horizontal direction can be calculated as W/π. The location of the captured image can be expressed as $C \in \mathbb{R}^2$ in the 2D digital map and both ends of the façade footprints can be expressed as $P_l, P_r \in \mathbb{R}^2$. Then the pixel per meter, which is the texture resolution for each façade, can be calculated using the law of cosines

$$\text{Texture Resolution} = \frac{W \cdot \theta}{\pi},$$

where

$$\theta = \arccos\left(\frac{\left(\|C - P_l\|\right)^2 + \left(\|C - P_r\|\right)^2 - \left(\|P_r - P_l\|\right)^2}{2\left(\|C - P_l\|\right)\left(\|C - P_r\|\right)}\right).$$

(1)

These criteria are illustrated in Figure 3.

3.4. Error Correction/Compensation. The correction/compensation process turns the omnidirectional building detection problem into the simple problem of segmenting a single image. Noise in the GPS/INS sensor is the primary source of the mismatch between the image/building analysis results and the ground truth.

First, correction of the GPS/INS sensor error is performed using image-based localization methods. Image-based localization is a major research issue in robot vision and location-based services. The localization of a panoramic image can be conducted by applying semantic segmentation [18]. However, the authors assumed the existence of a detailed 3D cadastral model, but these are not often available in city areas. Other research [19] has proposed a localization method that uses images and digital maps, but this is strongly dependent on user input. Hence, their method is not appropriate for our framework, which deals with a large number of images. Instead, our framework utilizes a method that localizes the panoramic image based on the orientation descriptor [20]. The footprint orientation (FPO) descriptor encodes the relative angle between the lines emitted radially from a certain location on the map and the footprints of the buildings. In the same way, the FPO descriptor can be calculated from the panoramic image because the panoramic image has omnidirectional information so that by vanishing point estimation we can calculate the angle between the location of the image and the footprints of the visible buildings. By finding the minimum distance between the FPO descriptor calculated from the image and the sampled locations on the map, we can estimate where the panoramic image has been taken. Experiments showed that the error after estimation is less than 2 m, which is sufficient for our framework, and to proceed to the processing stage. Meanwhile, the 360-degree FOV panoramic image is preferred because of this error correction. As can be seen in earlier research, a single image from a normal lens contains a limited amount of visual information about the
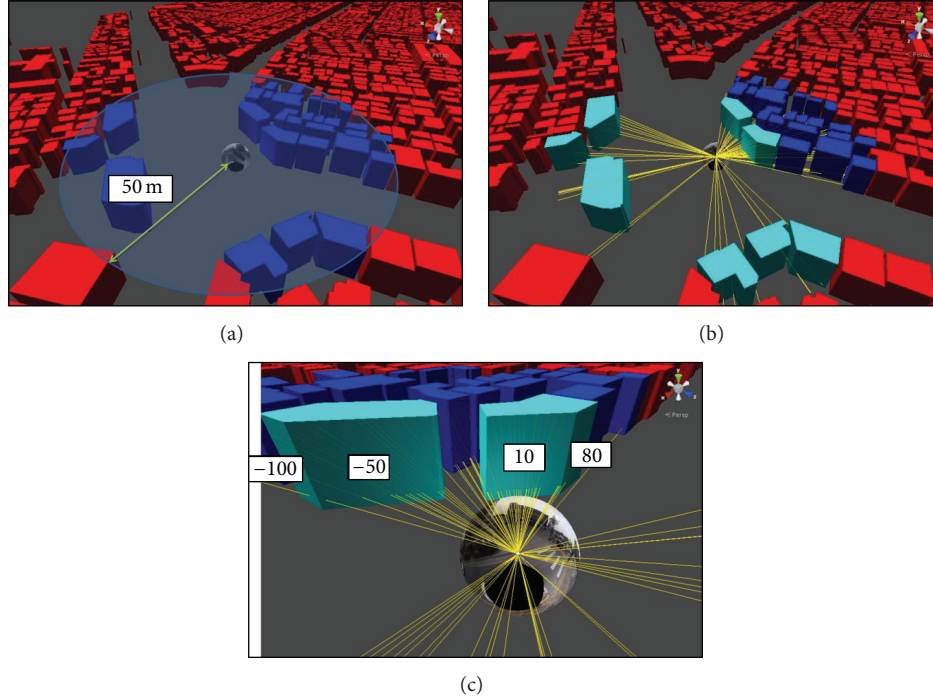
FIGURE 3: Criteria of the image/building analysis. (a) Distance between the image and buildings; (b) occlusion between buildings; and (c) facing viewing angle of the building façade, that is used to calculate the texture resolution.

surrounding environment [21], so user-input should be considered in order to more accurately estimate the location where the image was taken [22].

The error compensation utilizes this error bound to set the region of interest (ROI). Our objective in error compensation is to process the single 360-degree panoramic image into several normal-lens images to build each target, by partitioning and reprojecting. As we mentioned before, the error correction reduces the position and orientation error but there still exist mismatches, up to 2 m in position, between the base 3D model from the digital map and the panoramic image. So we calculate the FOV for each targeted building, which is ROI, with a 2-m margin. After that, the panoramic image is reprojected using rectilinear projection to generate an image which preserves the straight lines in 3D space in the projected image. Then the ROI contains the complete image of the target building and the complexity of the image-segmentation process is reduced.

3.5. Segmentation and Validation. This step involves the segmentation and validation of individual buildings. As noted in earlier studies [23, 24], the most important information is the outer boundary in the image. Previous research has attempted to obtain the outer boundary of a building by estimating the vanishing point and corresponding line segments. Usually, the method proposed in [23] gives robust results, because horizontal line segments are less affected by occlusions, whereas [24], which relies on the vertical vanishing point, suffers from occlusions caused by pedestrians, trees, and cars.

The outer boundary can be obtained by minimizing the 1D Markov random field energy, which is defined as

$$E(\mathscr{L})$$
$$= \sum_{i=1}^{n} D_i(l_i) + \sum_{i=1}^{n-1} V_{i,i+1}(l_i, l_{i+1}) \quad \text{(see [18])} \tag{2}$$

$$D_i(l_i) = \begin{cases} [T(x_i) > \beta] \cdot \lambda & \text{if } l_i = 0 \\ 100 - C_{l_i}(x_i) & \text{otherwise} \end{cases} \tag{3}$$

$$V_{i,j}(l_i, l_j) = [l_i \neq l_j] \cdot \mu \left(C_{l_i}(x_i) + C_{l_j}(x_j) \right), \tag{4}$$

where $\mathscr{L} = (l_1, l_2, \ldots, l_n)$ is the labeling of the entire column in the image so that n is the pixel width of the image and $l_i \in \{1, 2, \ldots, m\}$ is the label for each column x_i where m is the number of detected horizontal line orientations (e.g., $m = 2$ in Figure 4(a)). $C_{l_i}(x_i)$ is the number of line segments with the specific horizontal line orientation in the column x_i, and the total number of line segments of any orientation in x_i is $T(x_i) = \sum_j C_j(x_i)$. Therefore $D_i(l_i)$ is the unary potential, which has a lower energy when there are more horizontal line segments crossing x_i where β is the threshold and λ controls the cost for no-façade region. $V_{i,i+1}(l_i, l_{i+1})$ is the pairwise potential for the line segments corresponding to the vanishing points of the x_i and x_{i+1}, where μ is the weight factor. This describes the smoothness factor when labeling the different pixel values by providing higher energy when the labels differ between those pixels. The result is illustrated in Figure 4.
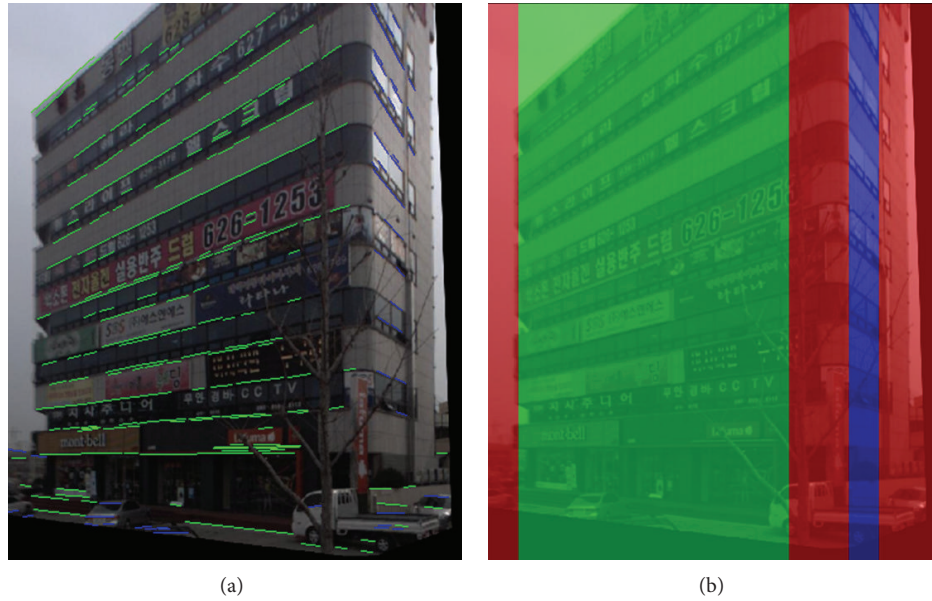
FIGURE 4: (a) Detected horizontal line segments (green and blue); (b) segmentation results using [23].

In Figure 4(a), line segments corresponding to the two vanishing points from two major facades of the building are illustrated on top of the image with green and blue lines. The result labeling is illustrated in Figure 4(b), which minimizes the energy defined in (2). Each green and blue label indicates the same façade area in the image and the red label indicates no façade detected. To generate the rectified image according to the vanishing point, we applied metric rectification [25] and bilinear interpolation in image transformation, to fill in lost pixels. This is not the finished image, because the horizontal line segment method relies on the detected line segment. Hence, a novel image segmentation method that adopts the additional characteristics in the energy term is needed.

For validation, the segmented building boundary is evaluated for the presence of three possible defects in the segmented image.

(1) The existence of the building in the segmented region.

(2) The equality of the segmented building with a building in the 2D map.

(3) The completeness of the segmentation, which checks for pixel loss in the segmentation stage.

If the building consists of a set of planes, its existence can be confirmed by means of vanishing point estimation. The relationship between the normal directions of the planes can be calculated, and this can be compared to the relationship between the normal directions from the digital map to check the degree of sameness. This data can also be fed back to the second step to correct the position and orientation information.

Finally, pixel loss can occur even when the error correction/compensation process is complete. We expect that the continuity of the image segments used in the vanishing point estimation process can be used with the color information

to guarantee subpixel accuracy in the segmented result. The resulting image is expected to resemble that shown in Figure 5. The leftmost figure shows the reprojected ROI image from the error correction/compensation process; through the segmentation, we obtain the outer boundary image of the target building illustrated in the middle. Note that the image is vertically rectified here. The rightmost figure illustrates an individually segmented image of major façades, which is classified from the vanishing point estimation and metric rectification.

3.6. Model Refinement and Texture Mapping. After segmentation, a process of model refinement and texture mapping is required. Model refinement involves modifying the geometrical shape of the buildings, with height being the most significant factor. The building height can be estimated from the vertical field of view in the original panoramic image, which was mapped inversely from the segmented vertical edge in the previous step. In addition, because we are referring to digital map data, building heights can be estimated using simple trigonometry by combining the distance from the camera to the building edge. Additional refinements can be done by adopting the methods in [26, 27], which extracts detailed shapes from images. The base 3D model is a quasi-Manhattan world in which buildings are composed of a set of planes. Thus, further refinement can improve the quality of the resulting 3D model both visually and physically.

The texture mapping stage is quite straightforward once the previous steps are complete. The resulting model is rendered using color data from the texture image and the UV-coordinate values of the vertices in the 3D model. To simplify the UV-coordinate modification, we consider the width of a single texture image to give the total perimeter of the entire building and assign to each façade image the ratio between this perimeter and the façade length. The resulting texture
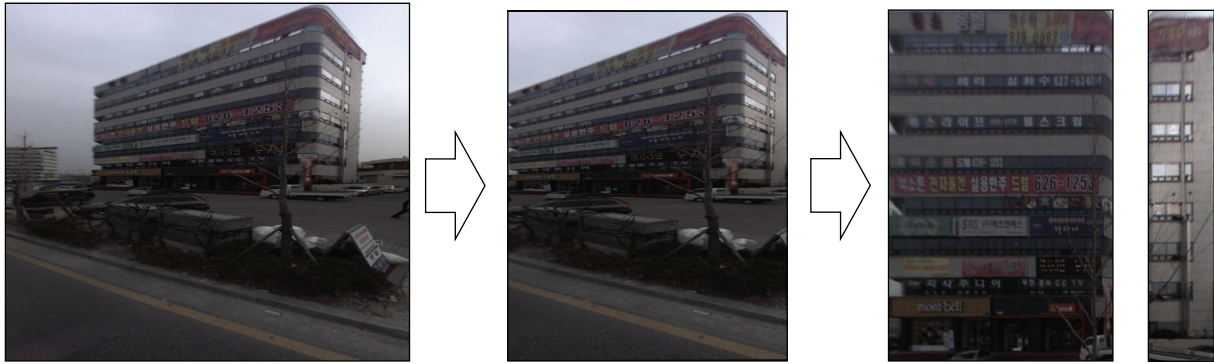
FIGURE 5: Image segmentation and validation. From the regional image of the building to the individual façade.
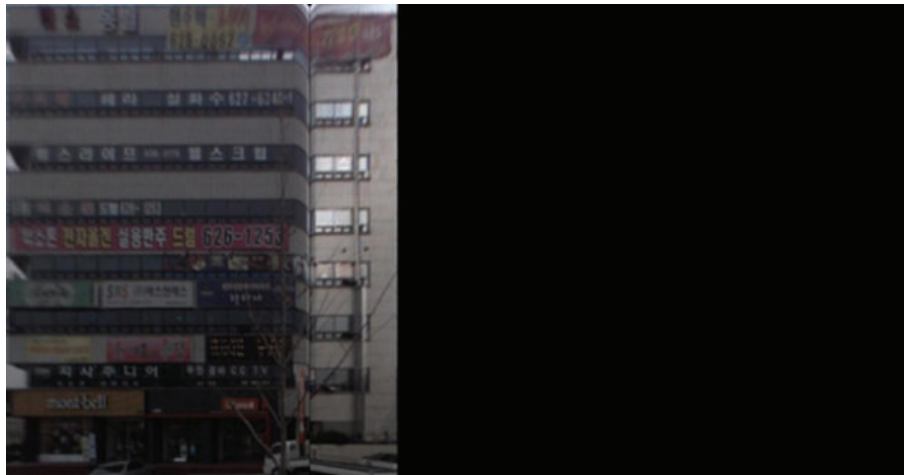

FIGURE 6: Resulting texture image.

is illustrated in Figure 6. Notice that the right portion of the texture image is empty since the panoramic image from the MMS only catches the façade facing the road. Therefore, only the image of the façade collected from panoramic image appears in the left portion of the texture image.

4. Experiments

4.1. Experimental Environment. For the experiments, we obtained panorama images from around Daejeon, South Korea, as well as coupled position and orientation data from a GPS/INS sensor. The source panorama images have a resolution of 5400 × 2700 pixels and a 360° horizontal/180° vertical field of view. Each panorama image was generated by stitching six perspective images using equidistant cylindrical mapping. A total of 233 images were gathered in the area bounded by [36.358933°, 36.339820°] latitude and [127.432165°, 127.436540°] longitude. The applied digital map was drawn at a scale of 1 : 5000, giving a horizontal accuracy of 1 m. By selecting digital map data through its classification code, only information related to the buildings was utilized.

For the 3D image/building analysis environment, we adopted Unity3D version 4.3 with ray casting and vector calculation functionality. Before the image/building analysis,

the digital map data was parsed using the shapefile C Library in the Unity3D environment to generate the base 3D model. Error correction/compensation, image segmentation, and validation processes were implemented using MATLAB R2013a. A standard desktop PC (Intel Core i5 CPU 3.4 GHz, 4 GB RAM, Windows 7) was used as hardware.

4.2. Experimental Results. The resulting model, illustrated in Figure 7, has a photo-realistic appearance. This is because of the high-resolution texture obtained from the panoramic image. The following characteristics of the resulting model can be observed.

(1) Rich visual information from the photo-realistic appearance.

(2) Identical height values as for buildings in the physical world.

(3) Accurate geospatial information, identical to that of the digital map.

(4) Quasi-Manhattan world model that has a complete prismatic mesh structure.

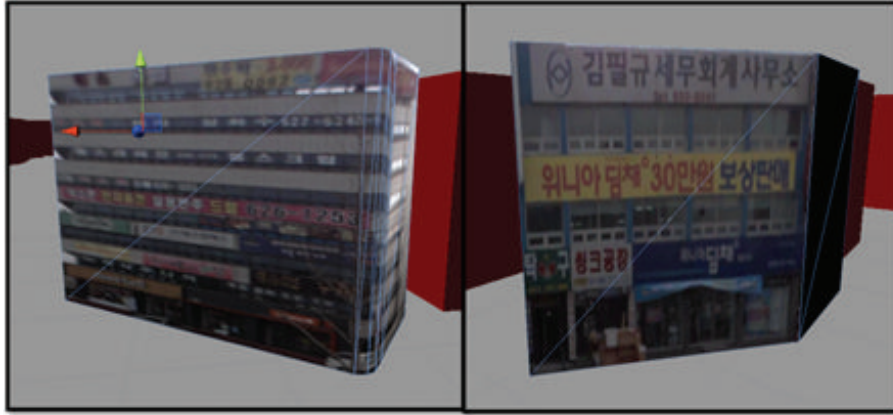The photo-realistic appearance provides an identical visual experience to the user, which is an important factor

FIGURE 7: Resulting textured building models.

(a)   (b)
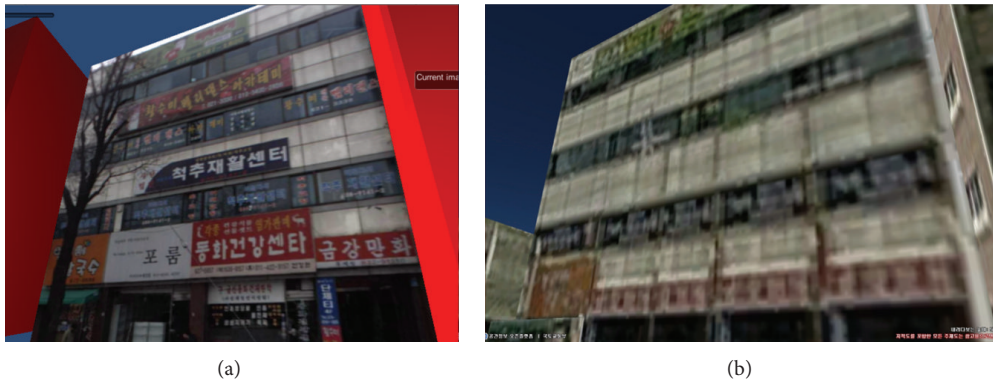
FIGURE 8: Comparison between our method (a) and aerial image-based method (b).

in immersive virtual environments. Additionally, the remaining text on buildings and signs provides additional visual information. The difference in visual quality is illustrated in Figure 8, which compares a 3D building model from an aerial image in V-World [28], a virtual earth serviced by the National Geographic Information Institute of Korea, and our result. The height values in our result are identical to those of the actual buildings; this information was not provided in the digital map. The height information directly influences the physical reliability. Moreover, because the position and direction of the resulting building models are coordinated with the information from the digital map, physical reliability is maintained for most applications. For instance, Figure 9 provides a comparison between the view from the center of the panoramic images mapped into the unit sphere and the resulting 3D world. The difference is less noticeable in terms of the buildings. Finally, as the geometry of the building is based on the footprint extruded from the digital map, the resulting 3D city model can be easily utilized in interactive virtual reality systems. This is because the complete prismatic mesh structure produces credible results for the algorithms that operate interactive content, for instance, in collision detection. In contrast to the results obtained using the stereo matching-based reconstruction, our approach does not require postprocessing.

Since the objective our research is offline generation of the model from the source data, the computation time is not the important target. Nevertheless, the image/building analysis takes 3.2 seconds for about 120 image locations with 7711 buildings in the digital map. The computation time of this process depends on the number of images and buildings in the targeted region. On the other hand, the computation time in the vanishing point estimation and reprojection in the error correction/compensation process as well as in segmentation depends on various factors, including the number of line segments detected, the resolution of the panoramic image, and the region occupied by the target building. For instance, the reprojection takes 36 seconds for a 30-degree horizontal and 120-degree vertical FOV which produces a 647×2977 resolution image from a 5400×2700 panoramic image. The processing time increases to 71 seconds for a 60-degree horizontal image with the same vertical FOV which produces a 1311×2977 resolution image. Vanishing point estimation takes 10 seconds on average by applying the line segments detector algorithm [29] and standard RANSAC technique. Then, the segmentation takes 2 seconds to solve the 1D minimization problem described in (2) using a dynamic programming method. The reprojection, vanishing point estimation, and segmentation are computed independently per building. Therefore the overall computation linearly increases

(a)   (b)
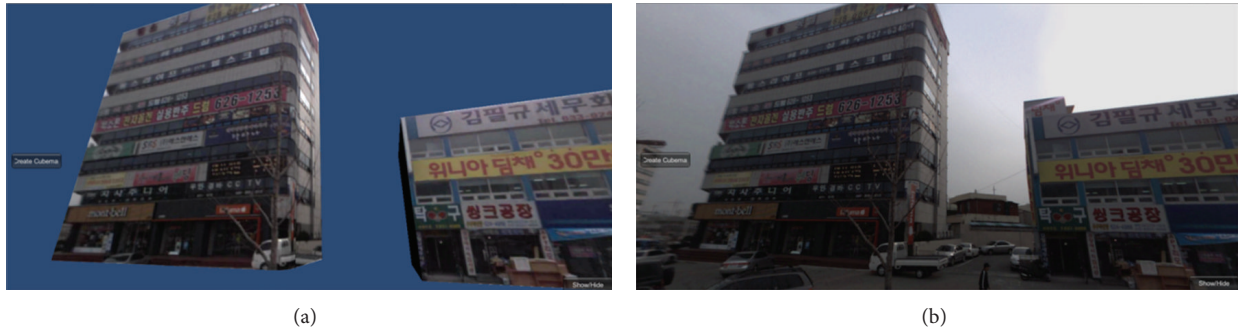
Figure 9: Scene from the same geospatial position in resulting city model (a) and panoramic image in unit sphere (b).

according to the number of buildings. Although the overall results are positive, the proposed framework has several limitations. First, it is possible for small features to be truncated in the segmentation process, and in some cases, the segmentation was not successful. As we mentioned in Section 3.3, the current segmentation method is based purely on previous vanishing point estimation and vertical/horizontal labeling research [23]. Hence, the segmentation results are dependent on the existence of edge information, which can be disturbed by occlusions. As we can guarantee that the building exists in the ROI after error correction, color, and texture features [30] may be needed to complement the vanishing-point method and a novel segmentation algorithm [31] could be adopted. At this point, the error correction process is not complete, so in some cases buildings did not exist in the expected ROI. The error correction process can be improved using edge-based wide-baseline stereo matching. Edges would be identified completely after segmentation, because the validation process guarantees the completeness of the segmentation.

5. Conclusion

In this paper, we proposed a framework for the automatic generation of a textured high-resolution 3D city model for ground-level applications. Our approach was employed to produce a complete prismatic mesh based on a quasi-Manhattan world model. The cost-effectiveness of the framework is ensured by its use of MMS data, which allows access to a massive number of images and a large coverage area. To address the problems of existing MMS-based methods, our framework uses digital map data in four major steps. The proposed approach combines existing techniques with novel processes in each of these steps.

In future work, we will consider combining data from different sources. This is because data from the MMS system are restricted, as images of the façade facing away from the road cannot be gathered. The missing data can be supplemented by processing aerial images [32] or pedestrian-collected data. Aerial images can only provide low-resolution textures at ground level but can also be source data for the updated 2D digital map [33]. Pedestrian-collected data can generate high-resolution images that are similar to the panoramic images, although the data collection process cannot be made cost-effective. Thus, a compromise approach should be proposed

according to the intended application. Moreover, the overall appearance of the building can be changed by reconstruction or alteration of its exterior over time, which could cause a mismatch between the 2D digital map and the panoramic image. Therefore, research should continue on detection of mismatches using automated, ground-level photography. The removal of occluding objects should focus on redundant images of the same building. Occluding objects can be severely disadvantageous to the visual experience, so occlusion removal will increase the visual quality of the resulting 3D city model.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publishing of this paper.

Acknowledgments

References

[1] Google Earth, http://www.google.com/earth/.

[2] P. Musialski, P. Wonka, D. G. Aliaga, M. Wimmer, L. Gool, and W. Purgathofer, "A survey of urban reconstruction," *Computer Graphics Forum*, vol. 32, no. 6, pp. 146–177, 2013.

[3] J. Yao, H. Tawfik, and T. Fernando, "A GIS based virtual urban simulation environment," in *Proceedings of the 6th International Conference on Computational Science—Volume Part III*, pp. 60–68, 2006.

[4] B. Hagedorn and J. Döllner, "High-level web service for 3D building information visualization and analysis," in *Proceedings of the 15th ACM International Symposium on Advances in Geographic Information Systems (GIS '07)*, pp. 51–58, November 2007.

[5] R. B. Rusu, *Semantic 3D Object Maps for Everyday Robot Manipulation*, Springer, New York, NY, USA, 2013.

[6] N. J. Shih, C. Y. Lee, and T. Y. Chan, "3D scans of as-built street scenes for virtual environments," in *Proceedings of the Symposium on Simulation for Architecture and Urban Design*, pp. 46–51, 2011.

[7] D. Visintini, A. Spangher, and B. Fico, "The VRML model of Victoria Square in Gorizia (Italy) from laser scanning and photogrammetric 3D surveys," in *Proceedings of the 12th International Conference on 3D Web Technology (Web3D '07)*, pp. 165–168, Umbria, Italy, April 2007.

[8] G. Vosselman and H. G. Mass, *Airborne and Terrestrial Laser Scanning*, CRC Press, New York, NY, USA, 2010.

[9] X. Li, C. Wu, C. Zach, S. Lazebnik, and J. Frahm, "Modeling and recognition of landmark image collections using iconic scene graphs," in *Proceedings of the 10th European Conference on Computer Vision*, pp. 427–440, 2008.

[10] P. E. Debevec, C. J. Taylor, and J. Malik, "Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, pp. 11–20, New York, NY, USA, August 1996.

[11] X. Zheng, X. Zhang, and P. Guo, "Building modeling from a single image applied in urban reconstruction," in *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry (VRCAI '11)*, pp. 225–234, December 2011.

[12] O. Barinova, V. Konushin, A. Yakubenko, K. Lee, H. Lim, and A. Konushin, "Fast automatic single-view 3-d reconstruction of urban scenes," in *Proceedings of the 10th European Conference on Computer Vision*, pp. 100–113, 2008.

[13] H. Kim, Y. Kang, and S. Han, "A study on the photo-realistic 3D city modeling using the omnidirectional image and digital maps," Transactions of the Society of CAD/CAM Engineers, Accepted.

[14] Naver Maps, http://map.naver.com/.

[15] Google Street View, https://www.google.com/maps/.

[16] Baidu, http://www.map.baidu.com/.

[17] N. Haala and M. Kada, "An update on automatic 3D building reconstruction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 6, pp. 570–580, 2010.

[18] A. Taneja, L. Ballan, and M. Pollefeys, "Registration of spherical panoramic images with cadastral 3D models," in *Proceedings of the 2nd Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT '12)*, pp. 479–486, October 2012.

[19] R. Cipolla, D. Robertson, and B. Tordoff, "Image-based localization," in *Proceedings of the International Conference on Virtual Systems and Multimedia*, pp. 22–29, 2004.

[20] P. David and S. Ho, "Orientation descriptors for localization in urban environments," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '11)*, pp. 494–501, San Francisco, Calif, USA, September 2011.

[21] Z. Kang, L. Zhang, S. Zlatanova, and J. Li, "An automatic mosaicking method for building facade texture mapping using a monocular close-range image sequence," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 3, pp. 282–293, 2010.

[22] R. Cipolla, D. Robertson, and B. Tordoff, "Image-based Localisation," in *Proceeding of International Conference on Virtual System and Multimedia (VSMM '04)*, pp. 22–29, 2004.

[23] G. Wan and S. Li, "Automatic facades segmentation using detected lines and vanishing points," in *Proceedings of the 4th International Congress on Image and Signal Processing (CISP '11)*, pp. 1214–1217, Shanghai, China, October 2011.

[24] D. Kim, H. Trinh, and K. Jo, "Object recognition of outdoor environment by segmented regions for robot navigation," in *Proceedings of the Intelligent Computing 3rd International Conference on Advanced Intelligent Computing Theories and Applications*, pp. 1192–1201, 2007.

[25] D. Liebowitz and A. Zisserman, "Metric rectification for perspective images of planes," in *Proceedings of the 11th IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '98)*, pp. 482–488, June 1998.

[26] N. Haala and C. Brenner, "Extraction of buildings and trees in urban environments," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 54, no. 2-3, pp. 130–137, 1999.

[27] K. Schindler and J. Bauer, "A model-based method for building reconstruction," in *Proceedings of the International Conference on Computer Vision Workshop on Higher-Level Knowledge in 3D Modeling and Motion (HLK '03)*, pp. 74–82, 2003.

[28] Vworld, http://map.vworld.kr/.

[29] R. Grompone von Gioi, J. Jakubowicz, J. Morel, and G. Randall, "LSD: a fast line segment detector with a false detection control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 722–732, 2010.

[30] S. K. Vipparthi and S. K. Nagar, "Color directional local quinary patterns for content based indexing and retrieval," *Human-Centric Computing and Information Sciences*, vol. 4, no. 1, pp. 1–13, 2014.

[31] H. T. Manh and G. Lee, "Small object segmentation based on visual saliency in natural images," *Journal of Information Processing Systems*, vol. 9, no. 4, pp. 592–601, 2013.

[32] L. H. Liew, B. Y. Lee, Y. C. Wang, and W. Cheah, "Aerial images rectification using non-parametric approach," *Journal of Convergence*, vol. 4, no. 2, pp. 15–22, 2013.

[33] A. P. Nyaruhuma, M. Gerke, G. Vosselman, and E. G. Mtalo, "Verification of 2D building outlines using oblique airborne images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 71, pp. 62–75, 2012.