3-D Urban Environment Model Reconstruction from Aerial Data—High-definition Video and Airborne Laser Data

Kaoru Wakabayashi†, Isao Miyagawa, Yuji Ishikawa, and Kenichi Arakawa

Abstract

We are developing three-dimensional (3-D) urban environment model reconstruction techniques that use video or range images taken by aircraft. This involves reconstructing a 3-D model from a high-definition (HD) video and another model using range data called digital surface model (DSM) data. The HD video modeling consists of two processes: one for acquiring point-based spatial data and the other for reconstructing buildings from the data. Since DSM data is spatial data, the DSM data modeling consists of recreating complicated building structures from the data. The model can be reconstructed in a short time and the kind of input can be chosen according to the intended application. These new techniques will have a wide range of applications.

1. Introduction

In recent years, interest in three-dimensional (3-D) urban environmental models has been increasing in various fields such as disaster prevention, sightseeing, and electromagnetic wave propagation [1]. On the other hand, geographic information systems (GIS), computer vision (CV), and computer graphics (CG) technologies are being used to reconstruct large-scale 3-D urban models such as townscapes. However, it currently takes a long time and a lot of money to reconstruct 3-D urban models. This has been a serious obstacle to the expansion of services such as 3-D GIS. In Japan, buildings tend to have complicated structures, and a model must reconstruct large numbers of them accurately.

NTT Cyber Space Laboratories has been conducting research on 3-D urban environmental model reconstruction for eight years. This paper describes how information is acquired from aerial videos or range images and how 3-D urban environmental models are reconstructed. In particular, it describes two techniques for modeling from high-definition (HD) videos and modeling from digital surface model (DSM) data. The HD video modeling can stably generate a lot of feature points and then classify layer clusters from them. It can automatically distinguish individual buildings in the townscape. Consequently, it is a fully automatic way of modeling buildings. In other words, it can produce an accurate urban environmental model from only HD video. The DSM data modeling detects building exteriors and roofs from altitude data acquired using airborne laser scanning systems carried in a helicopter or fixed-wing aircraft. It is fully automatic and can reconstruct detailed building forms. It can reconstruct an urban environmental model from only DSM data.

These two techniques let people choose the most suitable input medium for their requirements. This is a very useful feature because the medium necessary for the 3-D model depends on the user's business field.

† NTT Cyber Space Laboratories
Yokosuka-shi, 239-0847 Japan
E-mail: wakabayashi.kaoru@lab.ntt.co.jp

2. Reconstruction modeling based on aerial images

Aerial surveys, especially ones using aerial photography, are an efficient way of acquiring data for large-scale 3-D urban environment models [2], [3]. For many years, expert operators have used photographic

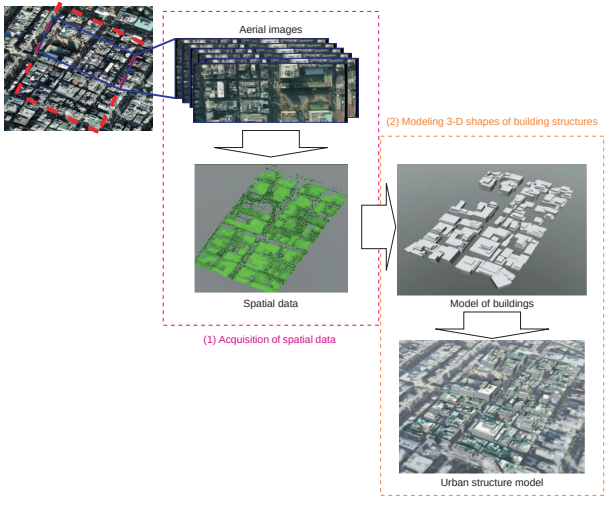


Fig. 1. Overview of 3-D model reconstruction from aerial images.

surveys to measure urban spaces, and they have manually reconstructed urban structure models. Aerial photography helicopters can now capture high-quality images using HD video. We have developed a technique that automatically reconstructs an urban structural model from aerial images. It consists of two functional steps, as shown in **Fig. 1**. First, 3-D coordinate values (spatial data) are acquired for feature points measured from aerial images. Then 3-D shapes of building structures are modeled from the spatial data.

2.1 Acquisition of global spatial data from image sequences

The factorization method [4], [5] can robustly and simultaneously recover both the camera motion and spatial data of feature points on 3-D object shapes from a sequence of images. We basically apply the perspective factorization method [6] to acquire spatial data from aerial images. We assume that a helicopter can maintain its flight vector and that aerial images can be captured stably along its flight course. The restriction of the camera motion is compensated for by acquiring spatial data more precisely using a paraperspective projection model.

The spatial data acquired from each image sequence has a local coordinate system, so it is difficult to select a coordinate system that involves the overall shape of the buildings. For example, some parts of buildings disappear from the frame as the helicopter moves. When sequentially reconstructing global spatial data, we need to determine the relationships between the local coordinate systems.

Figure 2 shows the relationships among viewpoints, aerial image sequences, and centers of gravity. Aerial images are segmented from image sequences S_1 to S_N. For example, spatial data (blue points) whose center of gravity (CoG) is O_1 is calcu-

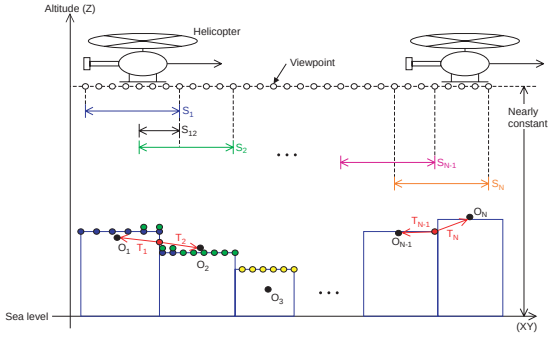


Fig. 2. Relationships among viewpoints, aerial image sequences, and centers of gravity.

lated from S_1 and spatial data (green points) whose CoG is O_2 is calculated from S_2. The CoG is the origin in the local coordinate system of each set of spatial data. To determine the geometrical relationship between the local coordinate systems, we apply the factorization method to acquire the camera motion from the overlapped sequence [7]. For example, to compute the relationship between O_1 and O_2, we compute the camera motion from image sequence S_{12}, which overlaps both S_1 and S_2. Furthermore, the 3-D coordinate values in the real world must be transformed into geographical spatial data using scaling parameters. These transformation parameters are calculated from geographical position orientation data, measured by the helicopter's GPS (global positioning system) system.

2.2 Reconstruction of building shapes from spatial data

A building's shape can be simply assumed to be a polygonal prism by adding vertical walls to the flat rooftop shapes. If the roofs consist of multiple flat rooftops, then the building is reconstructed as a set of prisms, each of which corresponds to a flat rooftop.

Figure 3 shows the reconstruction procedure, which begins with the spatial data described in section 2.1. The procedure starts by dividing the spatial data into two parts, i.e., points on the ground and

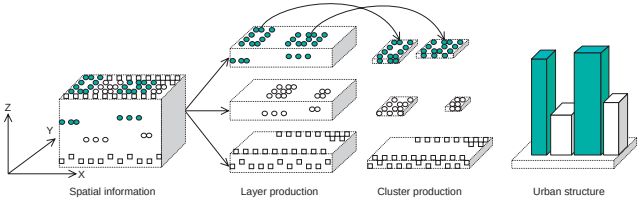


Fig. 3. Reconstruction procedure.

points on the roofs, by using a height threshold, assuming that the ground surface is a plane. A set of points located at nearly the same height is called a layer. To generate layers from the spatial data, we use hierarchical clustering, based on the minimum description length (MDL) principle [8]. Next, clusters, which are sets of points gathered together on a level plane, are generated from the layers. In this step, clusters are generated under the constraint that a polygon of any cluster on the X-Y plane may not cover any point included in the ground layer. Each cluster corresponds to a local area in which buildings exist. The clusters are divided, and the generated layers are grouped into clusters again. Finally, recursion, based on the height differences among the rooftops, generates clusters, each of which corresponds to a building feature.

By alternately generating clusters and layers, we gradually obtain the details of the urban structure. When only one cluster is generated from one layer, the procedure stops because a more detailed structure cannot be found.

2.3 Experiments

By using the method described in section 2.1, we acquired spatial data of one flight course from the aerial image sequences. **Figure 4** shows the results projected onto an aerial photograph. The helicopter flew from the bottom to the top, covering about 750 m in one course. The spatial data coincides with the road and building positions and that for urban structures is seamless for one course. The spatial data of Fig. 4 is shown from a different point of view in the upper part of **Fig. 5**, and the building shapes recovered from the data by the method described in section 2.2 are shown in the lower part. We separated roads and buildings and recovered the urban structures. Moreover, the outlines of building roofs were recov-

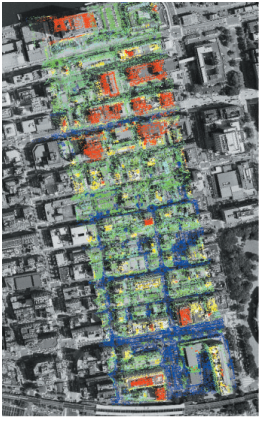


Fig. 4. Projection onto photographic image.

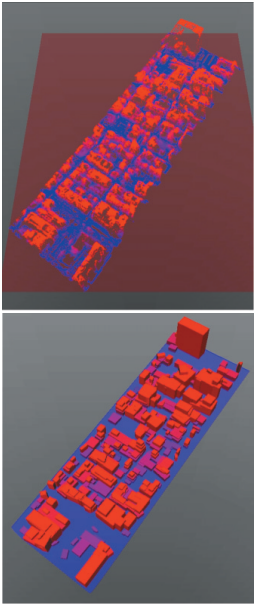


Fig. 5. Roofs modeled from spatial information.

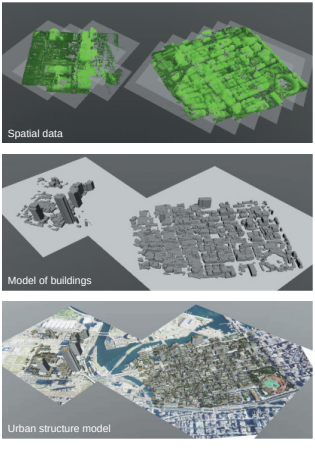


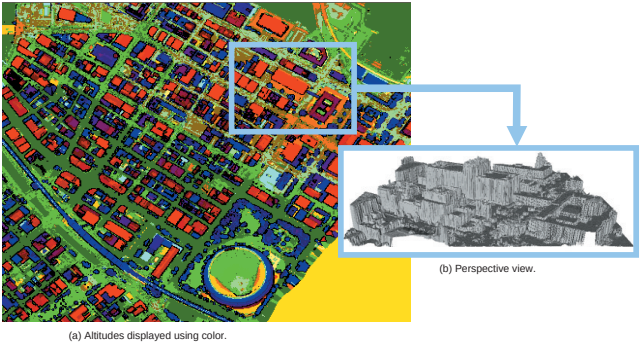
Fig. 6. Urban structural model produced from HD video.

ered using line information extracted from the aerial image.

Figure 6 shows an urban structure model of a large area of about 4 km^2 reconstructed by our system on a personal computer (CPU: 2.8-GHz Pentium 4, memory: 1 GB, OS: Linux). The processing time for reconstructing the model was 56.5 hours. Figure 6 (upper) shows spatial data acquired by the method described in section 2.1. Figure 6 (middle) shows the building shape recovered by the method described in section 2.2. It is also possible to extract texture data from the aerial photographs and map it to the ground surface and the building roof surfaces. Figure 6 (lower) shows the urban structure model with texture.

3. 3-D urban structure model reconstruction from DSM data

Here, we introduce a method for reconstructing highly precise 3-D forms using DSM data acquired with an airborne laser scanning system. DSM data is very fine-grained, occurs in regular intervals, and is horizontally accurate. The DSM data specifications used in this paper are listed in **Table 1**. **Figure 7** shows an example of DSM data. This covers part of the area treated in Fig. 6. Figure 7(a) shows the altitudes as color values. Figure 7(b) shows DSM data in a perspective view. It turns out that right-angled land-



(b) Perspective view.

(a) Altitudes displayed using color.
(green and yellow: low places, red and blue: high places)

Fig. 7. DSM data.

Table 1. Specifications of the airborne laser scanning system.

Laser emission cycle	25 kHz
Scan angle	–10° to +10°
Ground resolution	50 cm
Horizontal accuracy	–15 cm to +15 cm

scape and building features often become rounded, even in DSM data.

3.1 3-D modeling using DSM data

Techniques for modeling from DSM data include surface form analysis and division [9], [10] and semi-automatic processing [11], [12]. However, it is difficult to make complicated building forms automatically and consistently. To detect building forms consistently from homogeneous DSM data, we apply the edge analysis and classification method [13]. **Figure 8** shows the procedure for generating boundary shapes of buildings. The edge analysis and classification method computes the height changes at points that adjoin the edges from the DSM data, and based on the height changes, it classifies edges as belonging to specific patterns. The principle figures of this method are shown in **Fig. 9**. The extracted edges are classified according to these figures, which show the rises and falls when DSM data was scanned in the X and Y directions.

To acquire precise building shapes, the edges are re-extracted and re-classified. In this process, we apply the layer generation method based on the MDL criterion described in section 2.2 [8]. Even if the DSM data covers holes, steep projections, and other artifacts, an optimal building shape can still be found.

By using high-density and homogeneous DSM data, we can model uneven building roof shapes as complicated structures, consisting of two or more planes and slopes. **Figure 10** shows the algorithm for acquiring the roof shape. First, the histogram in the height direction is computed for every building, and layer generation is performed, as shown in section 2.2. Then, the layer is categorized as gables, hip roofs, planes, or patios. By examining every layer, we can extract two or more upper surface forms, so complicated building structures can be extracted. Finally,

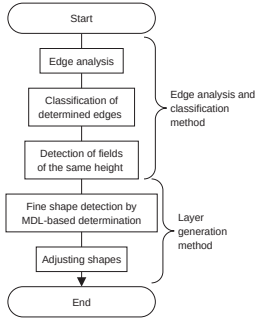


Fig. 8. Algorithm for detecting 2-D building shapes.

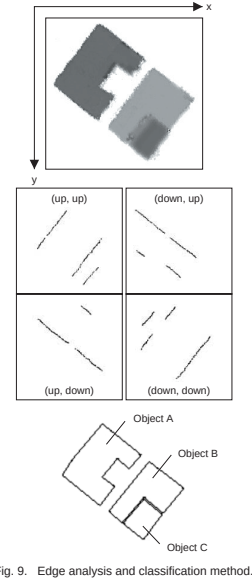


Fig. 9. Edge analysis and classification method.

the roof shapes are adjusted using histogram data.

3.2 Experiments

The model reconstructed using the DSM data shown on Fig. 7 is described below. **Figure 11** shows (a) the entire model and (b) individually modeled buildings. The model has the following features.

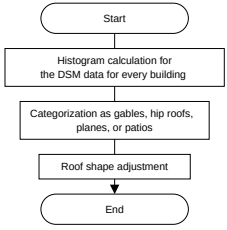


Fig. 10. Detection of a roof shape.

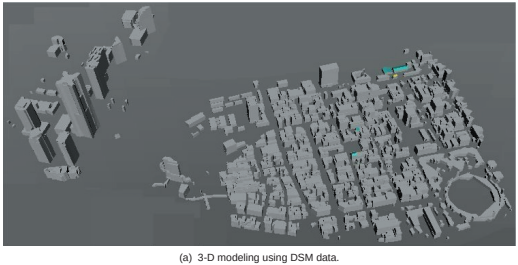
(1) It represented super-high-rise buildings as well as low-rise buildings well (the tallest building in Japan is about 300 meters high).
(2) The following buildings were represented well: buildings with two or more patios, buildings with gables or hip roofs, and building with uneven roofs.
(3) Dense building groups were represented well.

Figure 12 shows an example of the building model, with the textures taken from aerial photographs. By expressing the detailed building form correctly and giving it a texture, we can express the real building accurately.

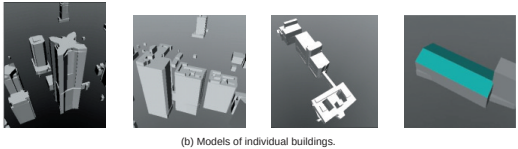
3.3 Discussion

Let us consider the precision and recall[*1] of the urban model. **Table 2** shows the precision and recall for the rooftop shapes achieved by modeling using HD video or DSM data, assuming that housing maps

*1 precision and recall are defined as follows.
precision = (area where the model buildings and the correct buildings overlap) ÷ (area of model buildings)
recall = (area where the model buildings and the correct buildings overlap) ÷ (area of correct buildings)

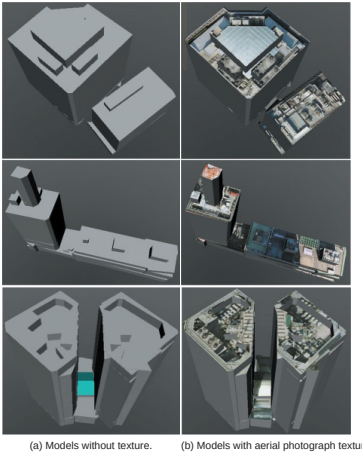


(a) 3-D modeling using DSM data.



(b) Models of individual buildings.

Fig. 11. 3-D urban modeling.



(a) Models without texture. (b) Models with aerial photograph textures.

Fig. 12. Example of adding upper surface textures to building models.

Table 2. Precision and recall of 3-D model, based on HD video or DSM data.

	HD video	DSM data
Precision	52.3%	82.3%
Recall	83.0%	72.6%

provide true data.

For modeling from HD video, the recall shows that modeling the shapes of building structures can recover building roof shapes with 80% accuracy. Modeling from HD video is an important new technique for reconstructing accurate urban structural models because they can simultaneously and rapidly acquire spatial and textural data.

On the other hand, modeling from DSM data produces a model that is very close to the actual form because its precision and recall are very high. DSM data is best for distinguishing super-high-rise buildings from surrounding low-rise buildings. It is also good for creating courtyards and building forms with gables or hip roofs and complicated building forms with uneven roofs.

The biggest difference between these techniques is in the input medium. We can choose the most suitable input medium for the application and the modeling can be done according to specific requirements.

Furthermore, no operator decisions are needed to produce consistent results and 3-D urban models of large areas can be reconstructed virtually automatically. In fact, 4-km^2 areas of Yokohama and Tokyo have been modeled with these techniques without decisions by the operators. The level and height accuracy of the reconstructed models are 2 m or better. Consequently, the accuracy of the whole urban model is very high. Compared with the technique of computing building heights by multiplying the layer height by the number of stories, as is done in existing car navigation applications, our techniques have much higher accuracy. This high accuracy will have clear benefits for human navigation applications.

4. Conclusion

This paper described two techniques for recon-

structing an urban model from aerial data. Both have a high level of reconstruction accuracy and reproduce city scenes well. The expense and time required for model reconstruction are also remarkably small. Moreover, these techniques have various uses, such as disaster prevention, sightseeing, and electromagnetic wave propagation analysis.

Aerial images have the drawback of lacking textures for the building sides. For this reason, walkthroughs in rows of houses reconstructed using our techniques will not be natural. The high expense of flying helicopters or light planes is also a drawback.

We plan to investigate ways of acquiring and applying side textures to buildings using vehicle-mounted omni-directional cameras that photograph all the buildings in their field of view. Such a technique should enable us to reconstruct an urban environment model that looks true to life. Ultimately, a lively virtual city space that seems just like the actual place will become a new communication environment.

References

[1] http://www.3dgis.jp
[2] http://www.aeroasahi.co.jp/spatial/geographic/index.html
[3] http://www.starlabo.co.jp/tls/index.html
[4] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: A Factorization Method," International J. Computer Vision, Vol. 9, No. 2, pp. 137-154, 1992.
[5] C. J. Poelman and T. Kanade, "A Paraperspective Factorization Method for Shape and Motion Recovery," IEEE Trans. Pattern Analysis & Machine Intelligence, Vol. 19, No. 3, pp. 206-218, 1997.
[6] S. Christy and R. Horaud, "Euclidean Shape and Motion from Multiple Perspective Views by Affine Iterations," IEEE Trans. Pattern Analysis & Machine Intelligence, Vol. 18, No. 11, pp. 1098-1104, 1996.
[7] I. Miyagawa, Y. Ishikawa, K. Wakabayashi, and T. Arikawa, "Spatial Chain Reconstruction from Aerial Images by the Perspective Factorization Method," IEICE, Vol. J87-DII, No. 4, pp. 942-957, Apr. 2004 (in Japanese).
[8] Y. Ishikawa, I. Miyagawa, K. Wakabayashi, and T. Arikawa, "Building Reconstruction Based on MDL Principle from 3-D Feature Points," Proc. ISPRS Commission III Sympo., Photogrammetric Computer Vision, Graz, Austria, Vol. XXXIV, No. 3B, pp. 90-95, 2002.
[9] S. Horiguchi, S. Nagai, and K. Sugiyama, "Recovering 3D Urban Model Using Range Data and Sequential Aerial Images," Urban Multi-Media/3-D Mapping (UM3-99), pp. 79-84, 1999.
[10] M. Shinya, I. Miyagawa, S. Horiguchi, K. Minamida, and N. Uemoto, "Automatic Construction of Three-dimensional Virtual Urban Models: Towards Three-dimensional Maps," NTT R&D, Vol. 49, No. 1, pp. 11-18, 2000 (in Japanese).
[11] A. Gruen and X. Wang, "CC-Modeler: a topology generator for 3-D city models," ISPRS Journal of Photogrammetry, pp. 286-295, 1998.
[12] A. Gruen and X. Wang, "Urban 3-D Mapping for a Hybrid GIS," Urban 3D, pp. 69-78, 1999.
[13] K. Fujii and T. Arikawa, "Urban Object Reconstruction using Airborne Laser Elevation Image and Aerial Image," IEEE Trans. Geoscience and Remote Sensing, Vol. 40, No. 10, pp. 2234-2240, Oct. 2002.

Kaoru Wakabayashi
Senior Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the B.E. degree in electro-communications from the University of Electro-Communications, Tokyo in 1982 and the Ph.D. degree in electronic engineering from the University of Tokyo, Tokyo in 1999. In 1982, he joined Nippon Telegraph and Telephone Public Corporation (now NTT) and has since been engaged in research on facsimile communications networks, binary image processing, map information processing, and cognitive mapping and understanding. He is a member of the Institute of Electronics, Information and Communication Engineers of Japan (IEICE) and the Information Processing Society of Japan (IPSJ).

Isao Miyagawa
Research Engineer, Visual Media Communication Project, NTT Cyber Space Laboratories.
He received the B.E. degree in electronics engineering from Fukui University, Fukui in 1991. He joined NTT Human Interface Laboratories in 1991. From 1991 to 1996, he researched communication technology for color documents and color reproduction and developed facsimile devices. Since 1997, he has studied computer vision. He is currently in the doctoral program in the Graduate School of Engineering at the University of Fukui. He is a member of IEICE and IPSJ.

Yuji Ishikawa
Researcher, Research and Development Headquarters, NTT DATA Corporation.
He received the B.S. and M.S. degrees in information science from Tokyo Institute of Technology, Tokyo in 1994 and 1996, respectively. Since 1996, he has been with NTT DATA Corporation and has been engaged in broadcast communication research. He transferred to NTT Cyber Space Laboratories in 2000 to study computer vision. In February 2004, he returned to NTT DATA Corporation and is researching sensor networks. He is a member of IPSJ.

Kenichi Arakawa
Group Leader, Senior Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the B.E. and M.E. degrees in information science from Kyoto University, Kyoto in 1984 and 1986, respectively. Since joining NTT Laboratories in 1986, he has been engaged in research on computer vision and robot vision. From 1990, he spent two years as a visiting scientist at the School of Computer Science, Carnegie Mellon University, U.S.A. He is a member of IEEE, IEICE, IPSJ, the Robotics Society of Japan, and the Institute of Image Electronics Engineers of Japan.