MVA '98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

7—2

Modeling of Urban Scenes by Aerial Photographs and Simply Reconstructed Buildings

Katsuyuki Kamei* Wayne Hoy† Takashi Tamada* Kazuo Seo*
Advanced Technology R&D Center
Mitsubishi Electric Corporation

Abstract

In many fields such as city administration and facilities management, there are increasing requests for a Geographic Information System (GIS) that provides automated mapping functions to users. Especially, some mechanism for displaying 3D views of an urban scene is eagerly expected because it allows construction of an intuitive and understandable environment for managing objects in the scene. In this paper, we present a new urban modeling system utilizing both vision and image-based approaches. Our method is based on a new concept that the wide urban area should be displayed with natural photographs, and each object in the area should be identified by its figure projected on the view. The users of the system can enjoy an intuitive understanding of the area and easy identification of the target, which have been accomplished by generating natural views at any viewpoints and reconstructing shapes of objects suitably.

1 Introduction

Geographic Information System (GIS)[1] is now expected to have functions of 3D real space management. A mechanism of generating 3D views of actual scenes is one of the most anticipated, for it allows construction of an intuitive and understandable environment for managing geographical objects in the scene. One method of establishing such a mechanism is to put full 3D geometric information of the real world into a computer system. However, this requires a large amount of tedious work since there are many complex objects in an actual scene. Many methods for automatic reconstruction are challenged[2][3][4], in order to recover the detailed structures in the scene. There is another approach, which represents the urban scene by a set of natural images, such as photographs or video sequences. This image-based approach[5] will always provide good views to users, but drawbacks are that it can not show any views at those viewpoints where no photographs exist, and the view will never know the information of its contents. In mobile cases, many "Augmented Reality" systems[6][7] are developed and used for relating figures of views to their property data, though video cameras must always be just on the spot, and of course they require the shapes of real objects.

In this paper, we present a new urban modeling system utilizing both geometry-based and image-based approaches. Our method is based on a new concept that the wide urban area should be viewed by natural images, and each object in the area should be identified by indicating its figure on the view (shown in Fig.1). We use bird's-eye photographs of the urban area and generate views at arbitrary viewpoints by the method of *Image Walkthrough*[8]. In order to relate the figures to their objects, we reconstruct rough shapes of the objects using these photographs in addition to computerized maps. The users of the GIS system might want an intuitive understanding of the area and easy identification of the target, both of which are accomplished by our system.

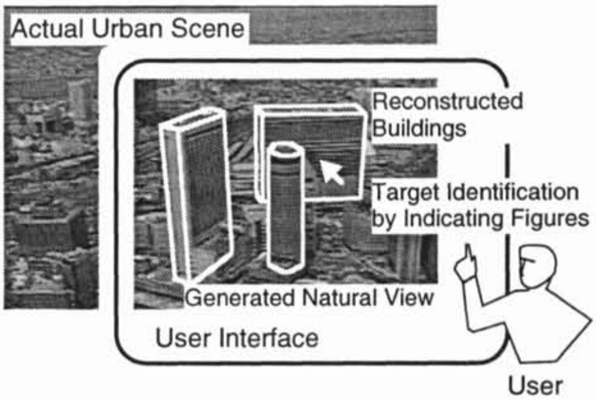
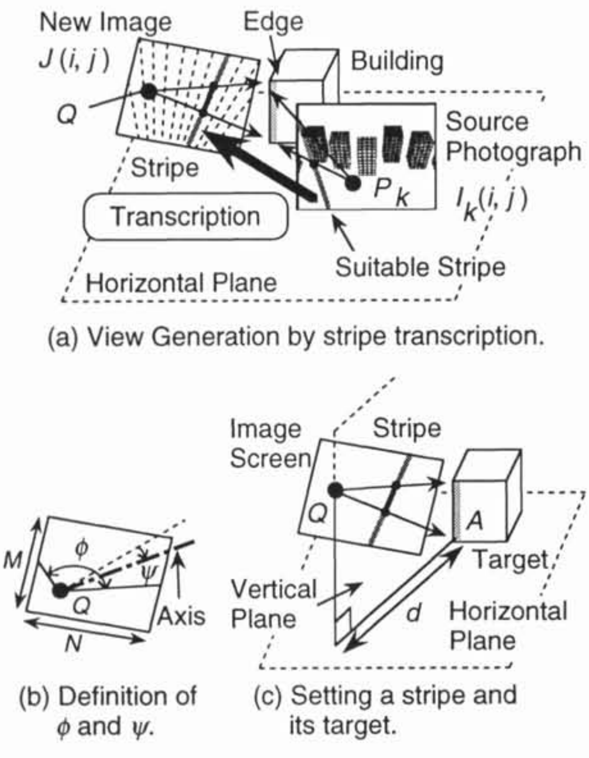


Fig.1 The concept of scene modeling.

*Address: 8-1-1 Tsukaguchi-Honmachi, Amagasaki, Hyogo 661-8661 Japan. E-mail: {kamei,tamada,seo}@sys.crl.melco.co.jp

†Current Address: University of British Columbia, 2329 West Mall, Vancouver, BC Canada V6T 1Z4.



(a) View Generation by stripe transcription.

(b) Definition of ϕ and ψ.

(c) Setting a stripe and its target.

Fig.2 View generation method.

2 Generation of Urban Views

2.1 Overview

For representing an entire urban scene, we use a set of aerial photographs and employ an image-based view-generation method called "Image Walkthrough", which we have developed. This method has several advantages for putting an actual scene into a computer system as a virtual environment, as follows:

1. It can synthesize photo-realistic views from some source photographs.
2. It is applicable to any actual scene.
3. It does not require any complicated pre-processing of the source photographs.

In our *Image Walkthrough*, source and synthesized images are considered as sets of stripes, each of which presents a narrow vertical area of the scene such as a vertical edge of a building. So these stripes should be arranged spokewise on each image. The outline of the view generation is as follows. For each stripe of the new image, the most suitable source stripe, which presents the same target on it, is detected and transcribed to it with scaled and located correctly (Fig.2). In this way, we can generate a new image of the urban area at any viewpoint, which gives the user an environment of great convenience.

2.2 Selection of Source Stripes

In this method, it is important to select the stripes of source images, because it seriously effects the quality of the resulting views. Let $I_k(i,j)$s and $J(i,j)$ be source images and the new image, and P_ks and Q be their viewpoints, respectively. As stripes of both images present a vertical area of the scene, each of them becomes to be set on the intersection of its image screen and a vertical plane. Therefore, a stripe is expressed as the following equation,

$$a\left(i-\frac{N}{2}\right)+b\left(j+\frac{N}{2\tan\psi\tan\frac{\phi}{2}}-\frac{M}{2}\right)=0 \quad (1)$$

where N, M, ϕ and ψ are width, height, angle of horizontal field of view and dip angle of the image, respectively (Fig.2(b)). And a and b are parameters which determine each stripe.

A stripe is selected and a new image is synthesized as follows:

1. Stripes are arranged on $J(i,j)$ as described above, at equal to or less than 1-pixel interval.
2. For each stripe, a main target object A is set on its ray vector (Fig.2(c)). When we have no geometrical information of the scene, A is set at the point where the horizontal distance from the viewpoint Q becomes a given value d.
3. For each stripe, the system retrieves a source stripe whose ray goes through A with the most similar angle, which may be expected to present the view of A on it (this retrieving mechanism is an extension of shown in [8]).
4. The retrieved stripe is transcribed to the correspondent of $J(i,j)$, scaled and shifted in order to draw the figure of A at the correct size and position on $J(i,j)$.

In this way, every stripe of $J(i,j)$ shows the figure of its target on it. If we can use adequate numbers of source photographs, such a source stripe may be found whose ray looks at A with the same angle as that of stripe of $J(i,j)$, which raises the quality of $J(i,j)$. Then, a fine new view can be generated at any Q and the user may freely navigate through the scene.

3 Reconstruction of Shapes

In order to access property data from the generated urban view, every figure on the view must be related to its actual object. We reconstruct the shapes of objects (buildings) and match their configuration to the view like AR systems. However, their precise reconstruction is so difficult that we approximate them by rectangular prisms, which are defined by giving the height values to their horizontal contour polygons obtained by the computerized

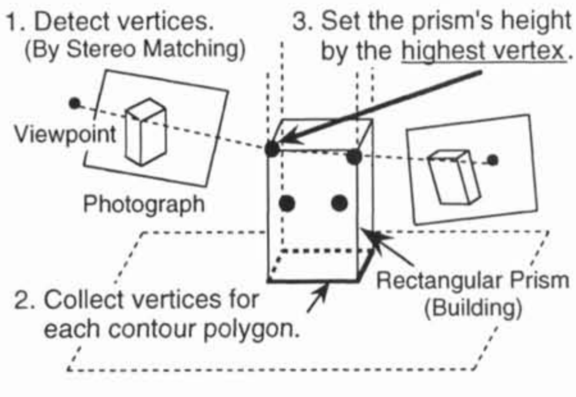


Fig.3 Decision of height of a building.

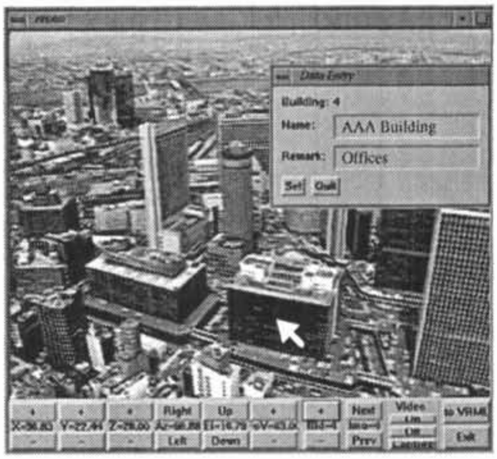


Fig.4 An example of object identification on our system.

city maps. Then, the recovering problem can be narrowed down to the decision of the correct height of every prism. We solve this problem as follows (Fig.3):

1. Detect vertices in the scene and compute their 3D coordinates by applying a stereo matching algorithm to several aerial photographs.

2. For each contour polygon, collect those vertices whose projected points to the horizontal plane are included by the polygon (these vertices are regarded as parts of the object indicated by the polygon).

3. For each polygon, select the highest vertex among the collected, and set the prism's height to the value of the highest.

In this way, we can easily set the rough height for every rectangular prism. The experimental results will show the usefulness of this method. Of course, the contour polygon may not be precise, or the height may be set incorrectly. For these cases, we prepare interactive functions of correcting them by referring to the aerial photographs. This simple prismatic model of a building is adequate for many kinds of urban simulation, such as the propagation of radio or refuge from a disaster, as well as identification of the target building.

4 Our Prototype System

We show our prototype system and its experimental results. Fig.4 shows an example of the user interface of our system identifying a target by indicating its figure on the photographic view. The information of the target, such as its name, is displayed in the pop-up window. Fig.5 shows examples of the aerial photographs and the generated view at a new viewpoint. This resultant view is made of 3 source images (for example, the right portion of Fig.5(b) is generated by transcribing Fig.5(a)). Processing time is closely related to the number of stripes. In our experiment, generating a 770 stripe, 640 x 480 pixel view took 0.2 sec. on average on an SGI Onyx workstation. If the viewpoint is moved continuously, an aerial movie can be generated.

Fig.6 shows the detected vertices and the reconstructed shapes of buildings using these vertices. Shapes have been recovered roughly, though the prisms are somewhat thicker because the given contour polygons are larger. The detection and 3D reconstruction of vertices were carried out by the method shown in [9] and took about one minute. Though we use source images captured by a VCR for consumer use, we have obtained a good result.

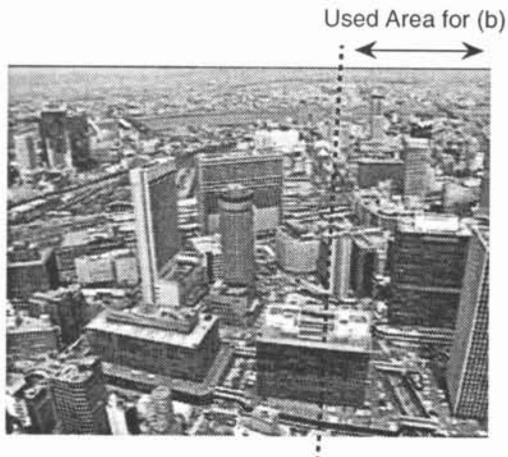
Fig.7 shows another function of our system of representing the scene by a set of texture-mapped prisms. Each texture pattern is cut from source photographs. Moreover, the system can transform this representation to VRML format, which means we can handle a rough sketch of an urban area at ease using the computer network.

These features of our method are useful for many GIS, which manage objects in actual scenes.

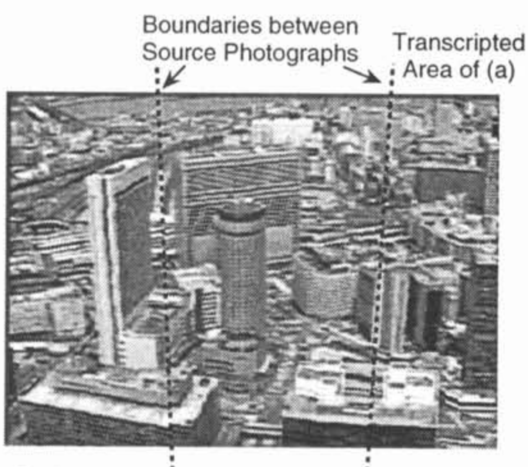
5 Conclusion

We present a new urban modeling system, which consists of the view generation and the shape reconstruction described above. We represent the scene not by either the generated view or the constructed shapes, but by relating the view to the information of its contents using the shapes. In this case, the user can specify the target clearly and obtain its information accurately by indicating its figure on the generated view, which is much more comfortable than performing the same task on the view made of the characterless polygons only, or on conventional 2D maps.

In the future, we intend to complete this system and apply it in practical systems.

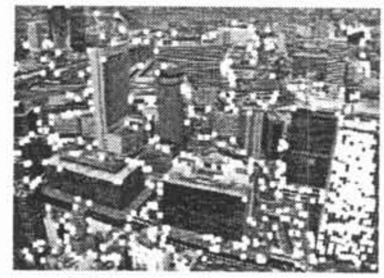


(a) A source photograph.

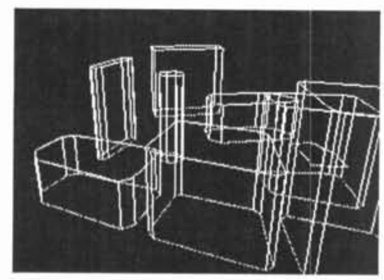


(b) A generated view at a new viewpoint.

Fig.5 Result of view generation.



(a) Detected vertices.



(b) Result of height decision.

Fig.6 Reconstraction of shapes of buildings.

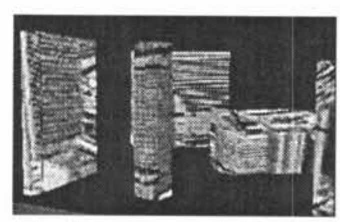


Fig.7 Scene Representation by texture-mapped prisms.

Acknowledgements

The authors acknowledge Dr. Paul Beardsley of MERL (Mitsubishi Electric Research Laboratories) for providing his 3D-recovery software and guiding us in applying it to our system.

References

[1] P.A. Burrough, "Priciples of Geographical Information Systems for Land Resources Assessment," Oxford University Press, 1986.

[2] O. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin and I. Zoghlami, "3-D Reconstruction of Urban Scenes from Image Sequences," COMPUTER VISION AND IMAGE UNDERSTANDING, Vol.69, No.3, pp.292-309, 1998.

[3] H. Zhao and R. Shibasaki, "Reconstructing Textured Urban 3D Model by Fusing Ground-Based Laser Range Image and Video Image," Proc. International Workshop on Urban Multi-Media/3D Mapping (UM3'98), pp.111-117, 1998.

[4] P. Liu, K. Ikeuchi and M. Sakauchi, "Estimating the Height of a building Using 2D Template Matching," Proc. UM3'98, pp.132-137, 1998.

[5] S. E. Chen, "QuickTime VR - An Image-Based Approach to Virtual Environment Navigation," Proc. SIGGRAPH'95, pp.29-38, 1995.

[6] R. Shibasaki, A. Takuma, H. Zhao and C. Tianen, "A Mobile User Interface for 3D Spatial Database Based on the Fusion of Live Landscape Imagery," Proc. UM3'98, pp.23-30, 1998.

[7] M. Arikawa, "Spatial Hypermedia as Augmented Reality Based on Spatial Databases," Proc. UM3'98, pp.39-30, 1998.

[8] K. Kamei, M. Maruyama and K. Seo, "Scene Synthesis by Assembling Striped Area of Source Images," Proc. ICIP'97, vol.2, pp.482-485, 1997.

[9] P. Beardsley, P. Torr and A. Zisserman, "3D Model Acquisition from Extended Image Sequences," Proc. ECCV'96, pp.II:683-695, 1996.