MOBILE MAPPING FROM STEREO OMNI-DIRECTIONAL VISION

S. Ozawa[a], I. Miyagawa[a], K. Wakabayashi[a], T. Arikawa[a]

[a] NTT Cyber Space Laboratory, NTT Corporation, 1-1 Hikarinooka Yokosuka-Shi Kanagawa, 239-0847, Japan
ozawa.shiro@lab.ntt.co.jp

Commission V, WG V/2

KEY WORDS: Mapping, Vision, Urban, Building, GPS, Virtual Reality, Mobile, Image

ABSTRACT:

We have developed a mobile mapping system for the purpose of acquiring building models having three-dimensional information such as building shape and position in urban space. The system mounts two omnidirectional cameras and position-sensing equipment on a vehicle and obtains images and positional information in a synchronous manner. It features three major functions. The first is a GPS interpolation function that enables positioning even at places in an urban environment where signals cannot be received from GPS satellites. The second is a synchronization function that synchronizes positional information and omnidirectional images and enables reconstructed models to be accurately converted to global coordinates. The third is a shutter control function that enables the shooting intervals of the two omnidirectional cameras to be either synchronized or non-synchronized. These functions make it possible to select images for all sorts of baselines from multiples images captured by the two omnidirectional cameras and to make stereo measurements for an optimal baseline according to the height of the target building and its distance from the camera. This system can perform high-speed data acquisition at a rate of four square kilometers per hour.

1. INTRODUCTION

Three-dimensional (3D) digital maps that contain building-related information such as position and shape in urban space are eagerly awaited in a wide range of fields (Teller, 1999; Virtual Helsinki, 1999).

In particular, 3D digital maps can be used as an important information platform in fields like Intelligent Transportation Systems (ITS) and Geographic Information Systems (GIS), and they can be applied to the simulation of radio-wave propagation, disaster countermeasures, etc. They are also expected to find use in the entertainment field such as in video-games.

In short, there is an extremely broad range of needs for 3D digital maps. In this regard, not only will the demand for accurate shape measurement and high-quality texture intensify, but the acquisition of such information will also have to be automated and the cost of acquisition lowered.

Against this background, various techniques have been proposed for acquiring urban models by the airborne approach using aerial photographs, laser rangefinders, and the like. The airborne approach has a major advantage in that a wide area can be efficiently reconstructed. For this reason, even our research group has been active in this area, researching a technique for reconstructing building shape from aerial images and a technique for reconstructing building shape from laser range data (Miyagawa, 2000; Horiguchi, 1999).

The airborne approach, however, lacks information on the sides of buildings and cannot support the use of building models as seen from the ground such as for "walkthroughs." On the other hand, there has been much activity in the research of "mobile mapping" that mounts cameras and range sensors on a vehicle to reconstruct building shape (Ellum, 2002; GIS Development, 2000; Tamura, 1998). Not only can mobile mapping reconstruct the sides of buildings, it can also acquire data in a relatively short time by utilizing key properties of vehicles.

With the above in mind, we set out to develop a mobile mapping system for the purpose of automatically reconstructing urban building models by installing cameras on a vehicle and making mobile measurements.

The high-accuracy acquisition of positional information is essential to mobile mapping, and to this end, the Global Positioning System (GPS) is often used. In urban areas, however, the quality of position data may drop due to shielding by buildings, to multi-path effects, etc., and many mobile mapping systems have come to incorporate a mechanism for interpolating GPS position data. For our mobile mapping system, we have developed and incorporated a function that interpolates GPS position data in real time while monitoring the quality of GPS positioning.

Various techniques have been proposed for reconstructing shape including a sensor-based technique using range sensors (Zhao, 2000) and an image-based technique using captured images (Uehara, 2000). The sensor-based technique can directly acquire the shape of a building but cannot obtain building texture. The image-based technique, on the other hand, has the advantage of being able to acquire shape and texture simultaneously. A conventional camera, however, has a limited angle of view, which means that the range of measurement will likewise be limited. Such a camera is therefore not conducive to an urban area that wills more than likely feature buildings of various heights. For this reason, we have adopted in our system an omnidirectional camera that can capture a 360 degree image at one time.

An omnidirectional camera may be of the mirror-projection type, which captures an image projected on a mirror with a camera, or of the lens type that uses a fish-eye lens. (Kawasaki, 2001). The mirror-projection type enables the mirror shape to be

TIB

TIB Hannover

designed as desired, but its drawback is that the camera itself will also be captured in the picture. The lens type, in contrast, can capture a completely omnidirectional image without capturing the camera, but its drawback is that design is quite difficult. In our system, we have adopted a fish-eye lens that can support the capturing of tall buildings without capturing the camera itself. It must also be pointed out, however, that spatial resolution in an omnidirectional camera is low since a wide-area image must be projected onto a finite pickup device at one time. There are two approaches to solving this problem: a hardware approach such as using a high-definition camera, and a software approach such as processing multiple images to raise resolution. We have adopted the former approach, using high-definition cameras in our mobile mapping system.

2. MOBILE MAPPING SYSTEM

2.1 Configuration of the mobile mapping system

The configuration of this system is shown in Fig. 1 and an external view of our mobile mapping system is shown in Fig. 2.

The system can be broadly divided into an "Imaging Section" that captures and stores omnidirectional images and a "Positioning Section" that collects positional information when capturing images.

The Imaging Section consists of omnidirectional cameras, a computer for recording captured images, and a shutter controller for controlling each camera's shutter. The cameras are placed on the roof of the vehicle to prevent the vehicle from being captured in the images.

The Positioning Section consists of a GPS antenna, a Differential Global Positioning System (D-GPS) device, a vehicle-speed sensor, a gyro sensor, a GPS interpolation device, and a computer for recording positional information. The Imaging Section and Positioning Section are synchronized by a GPS-based time server.

This system can accurately acquire the data required for reconstructing building models while moving.

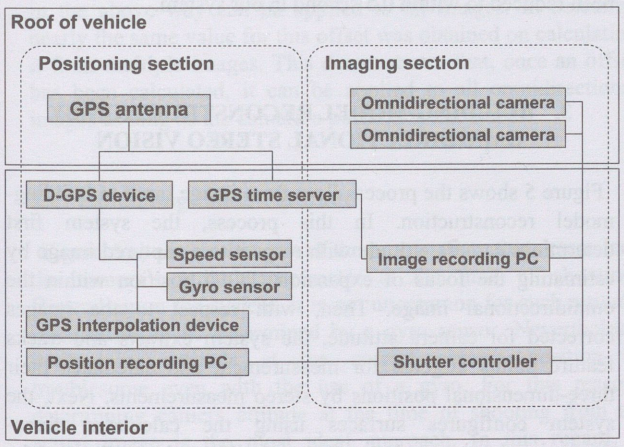

Figure 1. Mobile mapping system composition

Figure 2. External view of our mobile mapping system

2.2 Omnidirectional cameras adopted for the system

In an urban area, a mobile mapping system must be able to capture tall buildings. This condition can be met by an omnidirectional camera that can capture a 360° scene at one time. As mentioned above, there are two major types of omnidirectional cameras: a mirror-projection type and a lens type. Because the camera is captured in the image and forms an obstruction in the former type, we have adopted an omnidirectional camera that employs a fish-eye lens for which the camera will not be picked up. This type of camera, however, while being able to capture an expansive 360° area, suffers from low spatial resolution. As a countermeasure to this problem, we decided to use a digital camera having a 6-million-pixel CCD. Using a camera with high spatial resolution makes it possible to reconstruct high-quality texture and high-accuracy building models.

On the other hand, it takes about two seconds to capture one image with the type of camera adopted in our system, making for a low shooting rate. For this reason, the interval between the shooting positions for each image is large, and variation occurs in texture quality and model accuracy according to the distance from the shooting position and target building. To avoid this problem, we mount two omnidirectional cameras in our system and incorporate a shutter controller so as to control the shutter interval for each camera and facilitate flexible control of the shooting interval. This, in turn, makes it possible to adjust the baseline distance in omnidirectional stereo according to the distance to the target building and to therefore eliminate the variation in accuracy and texture quality among the buildings to be reconstructed.

Table 1 lists main specifications of our omnidirectional camera and Fig. 3 shows an actual example of an omnidirectional image taken with our mobile mapping system. The omnidirectional images captured by this system are of high quality and can accommodate tall buildings.

On the basis of these features, the omnidirectional cameras adopted for our system are considered to be an effective means of acquiring data in an urban area.

TIB
TIB Hannover

Table 1. Specification of omnidirectional Camera

Focal Length	5.9 [mm]
Image Circle	ϕ 18[mm]
Image Size	3032×2008 [pix]
Shutter Interval	2 [sec/frame]
Color Depth	24/32[bit/pix]

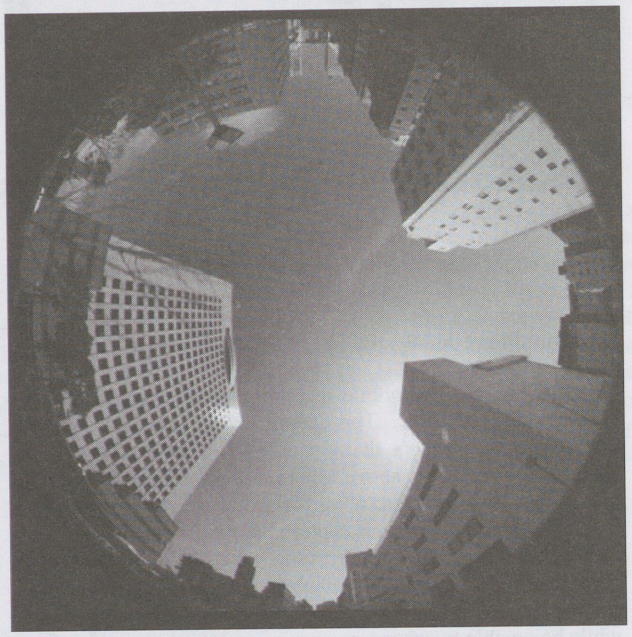

Figure 3. Example of omnidirectional image

2.3 Real-time positioning by D-GPS and GPS interpolation device

When using GPS equipment in an urban area, the quality of positioning can drop due to shielding and multi-path effects caused by buildings and other objects. In recognition of such problems, we have adopted a D-GPS device having a multi-path-elimination function. In addition, to deal with shielding effects that not even GPS equipment with the above function can eliminate, we have also developed for our system a GPS interpolation device that employs a gyro sensor and speed sensor. A key feature of this device is its ability to perform positioning by interpolating in real time while directly inputting data from GPS equipment and monitoring the quality of GPS data. Figure 4 shows an example of actual position data for a certain area. For comparison purposes, the figure also shows typical position data from GPS equipment. Specifically, the solid circles indicate typical GPS position data with no interpolation while the empty circles indicate position data obtained by our system. These circles draw a curve initially moving downward from the left and then changing direction and moving upward to the right. Examining this data, we see that typical GPS position data is missing in the "lacking section" while position data by our system features interpolated data in the "interpolated section." In addition, typical GPS position data includes disorder as a result of poor quality data in the "noisy area." The results demonstrate that our mobile mapping system can obtain high-accuracy positional information in a stable manner even under conditions that prevent positioning by GPS.

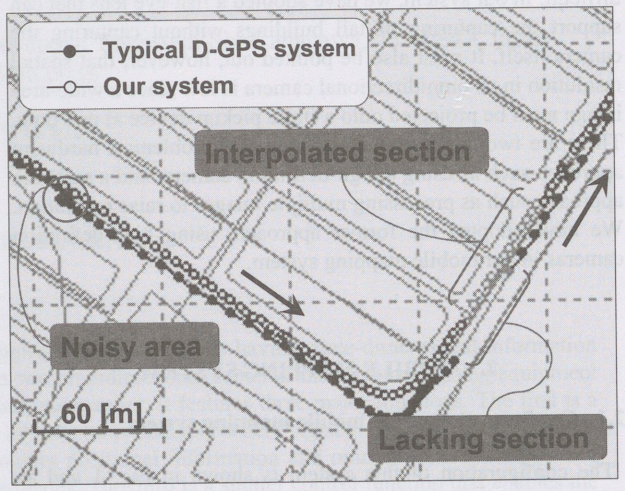

Figure 4. Comparison our system's position data with typical D-GPS data

2.4 Synchronization of positional information and omnidirectional images

If mistakes are made when assigning correspondence between omnidirectional images and positional information, it will not be possible to correctly convert reconstructed building models to global coordinates. It is therefore important to establish strict synchronization between GPS-based positional information and captured omnidirectional images. In the former system that we created, the lag between shooting time and positioning time was a maximum of 8 seconds due to Windows Operating System delays as a result of image-storage processing. In our current system, however, the positioning computer records the shutter opening time of an omnidirectional camera in GPS time. As a result, the lag between shooting time and positioning time has been reduced to within 0.5 second in our system.

3. BUILDING-MODEL RECONSTRUCTION BY OMNIDIRECTIONAL STEREO VISION

Figure 5 shows the process flow from image input to building-model reconstruction. In this process, the system first determines camera attitude with respect to a captured image by estimating the focus of expansion (FOE) position within the omnidirectional image. Then, with respect to the images corrected for camera attitude, the system extracts and tracks feature points targeted for measurement and calculates their three-dimensional positions by stereo measurements. Next, the system configures surfaces using the calculated three-dimensional points and reconstructs building shape. Finally, the system re-projects the reconstructed shape onto the building's lateral omnidirectional image and extracts texture to reconstruct the building model.

The following describes the above process in more detail.

TIB
TIB Hannover

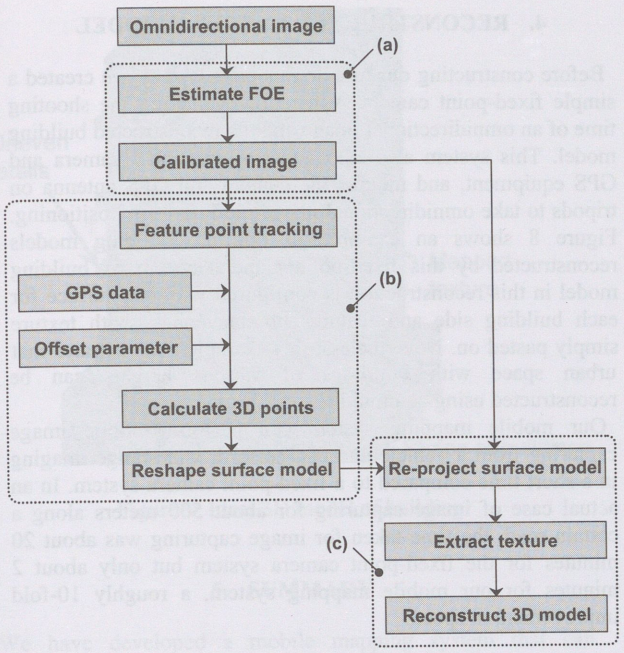

Figure 5. Process flow of reconstruct 3D model from omnidirectional Stereo

3.1 Calculating offset between camera position and GPS system position

To obtain the shooting position of an omnidirectional camera, the offset between the camera's position and GPS system position as determined by the GPS antenna must be determined beforehand. This offset is calculated by using points whose positions in space are known (SCP[1]) as reference points. Here, the position of a SCP projected onto a captured image is manually plotted, and the system position is then optimized so that the difference between it and the projected position is minimum. The difference between this optimal GPS system position and its initial value is taken to be the offset. If the installation positions of the camera and GPS antenna on the roof of the vehicle are not changed, an offset parameter determined in the above way can be applied to all images. In actuality, nearly the same value for this offset was obtained on calculating it from multiple images. This demonstrates that, once an offset has been calculated, it can be applied to all omnidirectional images as long as the camera's position is not altered.

3.2 Estimating camera attitude

Figure 5(a) shows that part of the process flow that determines the camera's attitude parameter at the time of shooting. Depending on the system, it is not uncommon for such attitude information to be determined by a gyro sensor. Nevertheless, determining attitude at the correct camera position is troublesome even with the use of a gyro. For this reason, determining camera attitude at the time of shooting from an actual image is the most ideal approach. In this regard, a property of an omnidirectional image is that camera attitude can be approximated by the difference between FOE in the image and the center of the image. Accordingly, FOE can be estimated by using, for example, a vertical segment of a building in the image. Estimating FOE in this way makes it easy to calculate camera attitude at the time of shooting from the captured image itself. This, in turn, makes it possible to carry out high-accuracy 3D measurements that take camera attitude into account.

[1] SCP (Spatial Control Point). An extension of the Ground Control Point (GCP) concept, a SCP means a point whose 3D position in space has been surveyed. Specifically, it refers to 3D values of points on a building's corner, side, etc. (Ishikawa, 2001)

3.3 3D measurements by omnidirectional stereo vision

Figure 5(b) shows that part of the process flow that performs 3D measurements by omnidirectional stereo vision. We here describe the optical projection formula of an omnidirectional camera and the equations for calculating depth and height through omnidirectional stereo. Figure 6 and 7 shows the relationships between the omnidirectional cameras and points on target buildings. The plane which consists of an x-axis-y-axis is made into the ground of height 0 among a figure, and z-axis is taken in the height direction.

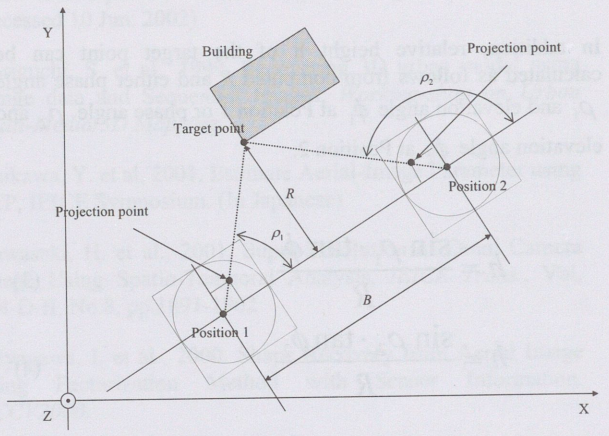

Figure 6. Horizontal position of a building and omnidirectional cameras

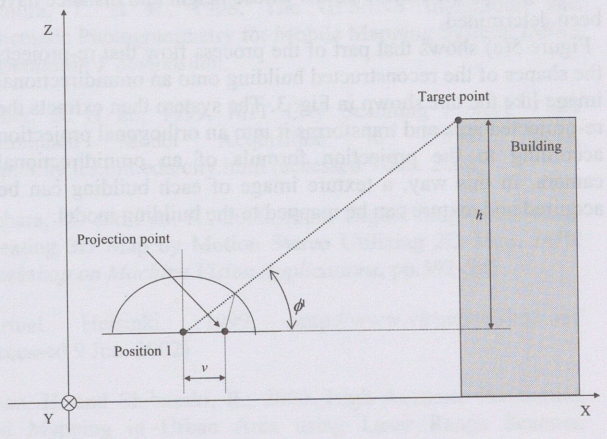

Figure 7. Vertical position of a building and an omnidirectional camera

Now, from the projection formula of a fish eye lens, the distance v from image center to the position where the target point is projected on the image can be given by Equation (1).

$$v = \frac{f}{D}\left(\frac{\pi}{2} - \phi_1\right) \quad (1)$$

Here, f is focal length and D is the physical size of one pixel of the image sensor at the focal surface.

Equations (2) to (4) are used to calculate the 3D position of a target point when performing stereo measurements by two omnidirectional images. In Fig. 6, B is the length of the stereo baseline while R is the distance from the line connecting the shooting positions of each image to the target point. Here, depth R can be calculated as follows from baseline B and phase angles ρ_1 and ρ_2 in each image.

$$R = \frac{B}{\cot \rho_1 - \cot \rho_2} \quad (2)$$

In addition, relative height h of the target point can be calculated as follows from computed R and either phase angle ρ_1 and elevation angle ϕ_1 at Position 1 or phase angle ρ_2 and elevation angle ϕ_2 at Position 2.

$$h = \frac{\sin \rho_1 \cdot \tan \phi_1}{R} \quad (3)$$

$$h = \frac{\sin \rho_2 \cdot \tan \phi_2}{R} \quad (4)$$

The above equations can be used to calculate the height and depth of all feature points of a target building. The shape of the building can be reconstructed by configuring the surfaces of the building from the feature points whose height and distance have been determined.

Figure 5(c) shows that part of the process flow that re-projects the shapes of the reconstructed building onto an omnidirectional image like the one shown in Fig. 3. The system then extracts the re-projected area and transforms it into an orthogonal projection according to the projection formula of an omnidirectional camera. In this way, a texture image of each building can be acquired and texture can be mapped to the building model.

4. RECONSTRUCTED BUILDING MODEL

Before constructing our mobile mapping system, we created a simple fixed-point camera system for evaluating the shooting time of an omnidirectional image and the reconstructed building model. This system also uses an omnidirectional camera and GPS equipment, and mounts the camera and GPS antenna on tripods to take omnidirectional images and perform positioning. Figure 8 shows an example of multiple building models reconstructed by this fixed-point camera system. A building model in this reconstruction is configured with one surface for each building side and features no unevenness with texture simply pasted on. Nevertheless, this example demonstrates that urban space with buildings of various heights can be reconstructed using an omnidirectional camera.

Our mobile mapping system can perform mobile image capturing from a vehicle and can perform wide-range imaging in a short time compared to a fixed-point camera system. In an actual case of image capturing for about 500 meters along a certain road, the time taken for image capturing was about 20 minutes for the fixed-point camera system but only about 2 minutes for our mobile mapping system, a roughly 10-fold improvement in speed.

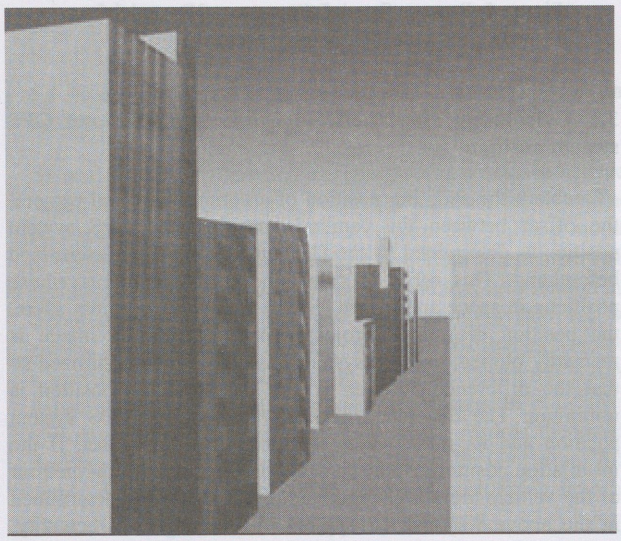

Figure 8. Example of 3D city space

Figure 9 shows the kind of building-model reconstruction that we are targeting using our mobile mapping system. This ideal model reconstructs not only building shape but also uneven building features such as balconies. Texture mapping will also be performed for each building. Reconstructing urban space in this way using building models having detailed shapes and high-quality texture will enable this mobile mapping system to support background-model applications such as walkthroughs and various kinds of urban simulations.

TIB
TIB Hannover

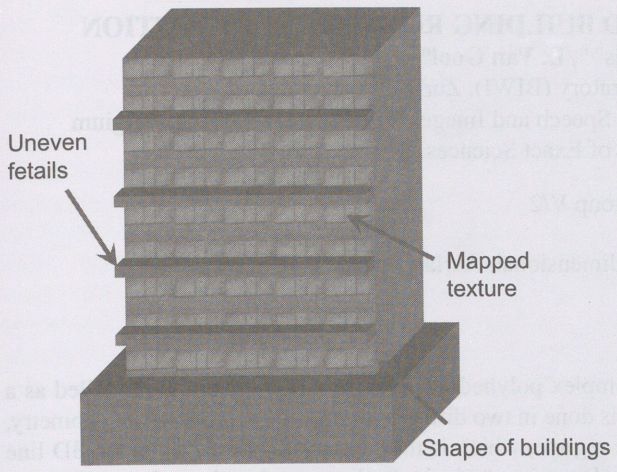

Figure 9. Ideal 3D model of building

5. SUMMARY

We have developed a mobile mapping system that can synchronously acquire images and positional information using position-sensing equipment and two omnidirectional cameras spaced 5 meters apart. This system has the following features.

An urban environment will have areas in which positioning by GPS will not be possible due to shielding effects caused by tall buildings. This system has therefore been designed to acquire positional information with apparently high GPS accuracy even in places where GPS positioning cannot be performed. This positional information is synchronized with omnidirectional images and enables reconstructed building models to be accurately converted to global coordinates.

A particular advantage of this system is that once offset between GPS system position and a camera has been acquired from an image, it can be applied to all images. In addition, camera attitude information at the time of shooting can be easily obtained since this information can be approximately substituted by the FOE of the omnidirectional image in question.

A shutter control function enables omnidirectional images to be acquired for any baseline distance and for models to be obtained by selecting an optimal baseline distance according to height of the target building and its distance from the camera.

The omnidirectional cameras of the system can capture high-resolution omnidirectional images enabling high-quality texture to be obtained for mapping to building models.

Constructing a mobile mapping system that can synchronously acquire omnidirectional images and positional information in the above way has improved mobility in terms of model acquisition and has enabled image capturing at a rate of four square kilometers per hour. These achievements mean that the time from data acquisition to model reconstruction can be shortened by about 2/3 and that personnel expenses and other costs can be held down.

A future issue is finding ways of making this mobile mapping system even more accurate with even higher levels of quality. This might be accomplished, for example, by raising the resolution of omnidirectional cameras through the use of even better high-definition digital cameras, or by introducing a position acquisition system of even greater accuracy such as Virtual Reference Stations (VRS). With these improvements, our eventual aim is to raise the accuracy of acquired models to several centimeters.

As for the future of this system, we can envision the installation of cameras at positions lower than building roofs to enable the reconstruction of the lower parts of buildings. Also, by making the entire system compact and mounting it on general automobiles, it might be possible to model a wide area in a short time and to also reconstruct urban models corresponding to changes over time such as seasonal changes.

REFERENCE

Ellum, C. and El-Sheimy, N., 2002. LAND-BASED MOBILE MAPPING SYSTEMS, *Photogrammetric Engineering & Remote Sensing*, pp. 13-28.

GIS Development, 2000. http://www.gisdevelopment.net/ (accessed 10 Jun. 2002)

Horiguchi, S. et al., 1999, Recovering 3D urban model using Range data and Sequential Images, *Working-shop on Urban Multi-Media/3D Mapping*.

Ishikawa, Y. et al, 2001, Estimate Aerial-Image Parameter using SCP, IEICE Symposium. (In Japanese)

Kawasaki, H. et al., 2001. Super-Resolution of Omni Camera Image Using Spatio-Temporal Analysis, *IEICE Trans.*, Vol. J84-D-II, No.8, pp.1891-1902

Miyagawa, I. et al., 2000. Shape Recovery from Aerial Image using Factorization Method with Sensor Information. *ACCV2000*.

Okutomi, M. and Kanade, T., 1993. A Multiple-baseline Stereo, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.15, No.4, pp.353-363

Tamura, T. et al., 1998. The GPS/INS Integration and Kinematic Photogrammetry for Mobile Mapping System, *ISPRS Commission V Symposium*.

Teller, S. et al., 1999. MIT City Scanning Project: Fully Automated Model Acquisition in Urban Areas, http://city.lcs.mit.edu/city.html (accessed 9 Jun. 2002)

Uehara, M. and Zen, H., 2000. From Digital Map to 3D Map: Creating 3D Map by Motion Stereo Utilizing 2D Map, *IAPR Workshop on Machine Vision Applications*, pp.592-595

Virtual Helsinki, 1999. http://www.virtualhelsinki.net/ (accessed 9 Jun. 2002)

Zhao, H. and Shibasaki, R., 2000. High Accurate Positioning and Mapping in Urban Area using Laser Range Scanner, Working-*shop on Urban Multi-Media/3D Mapping*.

TIB

TIB Hannover