

## SHAPE RECOVERY FROM HYBRID FEATURE POINTS WITH FACTORIZATION METHOD

Isao MIYAGAWA, Shigeru NAGAI, Kazuhiro SUGIYAMA  
NTT Cyber Space Laboratories, Japan  
miyagawa@marsh.hil.ntt.co.jp

**KEY WORDS:** Factorization Method, Hybrid Feature Points, Shape from Motion, 3D Digital City

### ABSTRACT

We are developing shape recovery technology that can semi-automatically process image sequences using the idea of "Shape from Motion". In this paper, we investigate an acquisition and recovery method that processes hybrid feature points extracted from both high-resolution images and video images using the factorization method. The 3D object models are assumed to have N-point column shape, and are recovered on a plane formed from some points on the ground (delta points). The delta points correspond to orthometric height data captured by a camera parameter measurement sensor system. Acquired and recovered 3D models are mapped on Tokyo Datum. The 2D feature points measured using both hybrid feature tracking and normal feature tracking are decomposed into 3D object points using the factorization method. Comparison experiments show that the hybrid method is more effective and accurate in acquiring and recovering 3D object shape than a earlier method. Moreover, we confirm that the hybrid method yields accurately shaped top surfaces of building.

### 1 INTRODUCTION

3D cartography and geographical data are often used in various applications such as ray-tracing simulations, wind-flow mapping, sunshine occlusion, finding refuge roads after natural disasters, and so on. 3D stereo-scopic tools, e.g. (Martin Reddy, 1999), have been developed for 3D digital city projects. Aerial photogrammetry is necessary to acquire and reconstruct the many object models of buildings. Human operators who use photogrammetric equipment to acquire spatial data for a Geographical Information System (GIS) must perform the key tasks manually. It is very difficult to track feature points on aerial photogrammetry images. Automation of these tasks is difficult due to object complexity and the need for some expert knowledge.

High-resolution cameras have become recently available and used in watching system, robot vision etc. While the frame rates are lower than 30 [fps] and the captured images are 8 or 10 bit monochrome, image quality is very high, for example resolution length can correspond to 0.1 [m] per pixel in aerial images captured at altitude 300 [m]. The idea of "Shape from Motion" is important in realizing an automatic or semi-automatic tool to perform 3D cartography. As the feature points on high-resolution images are clear as 2D coordinate system values, 3D reconstruction from these images is expected to be accurate. We have investigated the effects of time-redundancy for high-resolution images on the 3D reconstruction of buildings. We have shown that adequate time-redundancy is generated by video images with transformation coefficients. In previous work, I showed that 3D feature points could be acquired from 2D feature points extracted from images, using an improved factorization method in combination with sensor information. To realize accurate shape recovery, we propose a hybrid feature tracking scheme in the next section. In it, 2D feature data is decomposed into shape matrix using a camera matrix composed from sensor information. The acquired 3D shape data are mapped onto Tokyo Datum using coordinate transformation. Moreover, each 3D object shape is reconstructed on the plane formed from delta points. Experiments compare the former method to the proposed method.

### 2 PREVIOUS WORK

The factorization method (Tomasi & Kanade, 1992)(Poelman & Kanade, 1997) was proposed as an effective and robust way of utilizing the time-redundancy of image sequences. This method recovers both camera motion and object shape. Thought random noise is present in the images, the measurement matrix is decomposed into a camera motion matrix and object shape matrix, using the rank-3 property of singular value matrix.

The camera in the helicopter commonly experiences significant disturbance by winding, which yields a form of random noise. Therefore, this current method could not acquire and recover 3D shape accurately from aerial video images. We had developed an improved factorization method (Miyagawa, 1999)(Miyagawa, 2000) that utilizes sensor information. Sensor information means camera orientation information: yaw, pitch, and roll degree based on sensor system measurements. The camera motion matrix is calculated from the rotation parameter. We have shown that this method is effective in recovering the shape of objects from aerial images. The height of each object is acquired relative to the delta plane, which

is formed by three delta points.

As most video image sequences have poor image quality, the recovered shapes are distorted case-by-case. Obviously, it is not enough to acquire and recover 3D models from time-redundancy images. In other words, There is no hardware system that can capture high-resolution images and store them on a computer disk. To require accurate 3D object shape, we develop the factorization method with high-resolution and time-redundant images.

### 3 HYBRID FEATURE TRACKING

Our idea is that time-redundant high-resolution image sequences are created by combining high-resolution images and video images. Aerial image sequences are acquired by a high-resolution camera and a video camera, both carried in the same helicopter. Time coding links the high-resolution images to the video images. Most feature points in high-resolution images are greatly displaced from one frame to the next, due to helicopter (camera) movement, for example, the variation can range from 50 to 100 pixels. The trajectories with feature points are created by combining the quality of the high-resolution images with the time-redundancy of the video images. 2D feature points extracted from high-resolution images are tracked in the video images. Different image tracking algorithms are used to extract feature points from the two types of images. To handle the occlusion in aerial images, we choose object points, for example, edge, corner, etc, and earth points lying on the ground, as 2D feature points. As the high-resolution and video images have different coordinate values, the 2D feature points from the high-resolution images are projected onto the video image plane using a transformation model. As each camera has different internal parameters (focus length, image center, and so on), the captured images must be corrected individually by a camera calibration method. As both cameras are mounted on the same helicopter, we can assume that the calibration model does not change. 2D points are first extracted from a high-resolution image and these points are then tracked in the corresponding video images. This allows us to track feature points in the two image sequences.

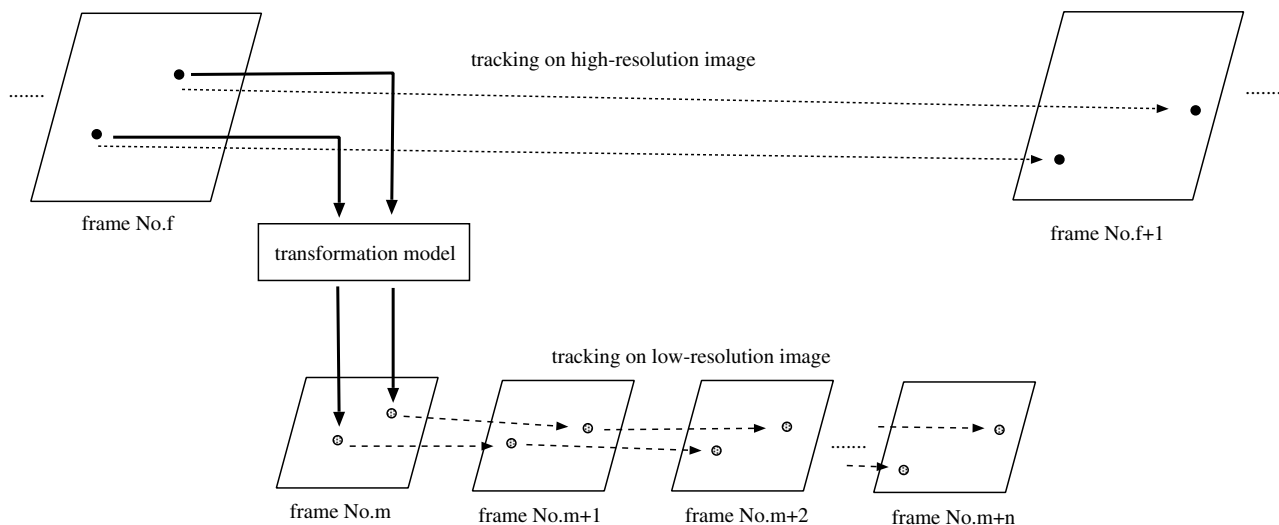


Figure 1: Hybrid Feature Tracking

#### 3.1 DELTA POINTS IN ORTHOMETRIC HEIGHT DATA

The operator must set feature points on the initial high-resolution image. Feature points are composed from some points on the top surfaces of buildings and basic points on the ground. Each object is assumed to have N-point column shape (for example, a circle column is formed by the many points). The number of delta points is at least three, and they lie on the ground plane. Here, delta points are made to correspond to orthometric height data using the camera internal parameters and camera external parameters ( $(X_p, Y_p, Z_p)$ : camera position by sensor system). When the operator set delta points on the ground, the 3D coordinate system values of these points are assessed using the corresponding orthometric height data. The delta points are not tracked on the images, because the 3D coordinate values can be projected onto 2D image planes. The orthometric height data for the ground plane is measured on a roughly 1 [m] mesh. The 3D points  $(X_o, Y_o, Z_o)$  on the ground are made to correspond to 2D points  $(x_i, y_i)$ , or  $(x_j, y_j)$  on the high-resolution images using the camera's internal and external parameters. The internal parameters, lens center:  $(C_x, C_y)$ , focus length:  $f$ , and so on, are calculated using Tsai's method (Tsai, 1986). The external parameters:  $(X_p, Y_p, Z_p)$  and  $(X_q, Y_q, Z_q)$ , are measured using a sensor system that is based on Differential-GPS.  $(X_p, Y_p, Z_p)$  is linked to time-coded, captured image in a one-to-one correspondence.  $(X_{gc}, Y_{gc}, Z_{gc})$  on the ground can be projected onto an image center  $(C_x, C_y)$ . All orthometric height data  $(X_o, Y_o, Z_o)$  can be projected into 2D coordinate values:  $(x_i, y_i)$  on the high-resolution image, assuming the pin-hole camera model.

If 3D point  $(X_o, Y_o, Z_o)$  is known, 2D points  $(x_j, y_j)$  in another camera view  $(X_q, Y_q, Z_q)$  can be established. In other words, this projection allows delta points to be matched to 2D points in each image.

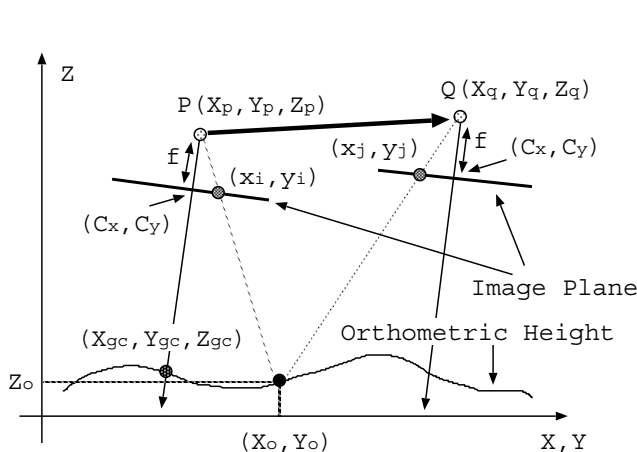


Figure 2: Delta Points on Orthometric Height

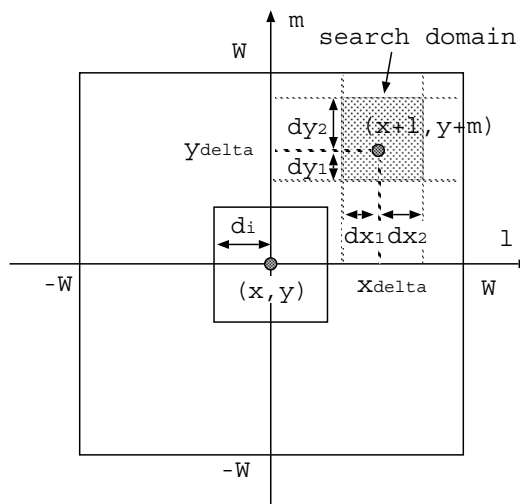


Figure 3: Search Domain

### 3.2 TRACKING ON HIGH-RESOLUTION IMAGES

In high-resolution images, it has enough resolution to achieve accurate tracking in simple template matching method. Therefore, the complex operations are not needed, we set windows to search feature points of the top surfaces of buildings. We assumed that surface size is nearly same for each frames because helicopter have maintained altitude for optical axis. In searching process, the center of balance:  $(x, y)$  (calculated by the feature points composing the top surface of building) is corresponded to the center of window to search feature points. Template matching is done between  $I(x, y)$  on this window and  $J(x + l, y + m)$  in the next window. The texture data in a template window is extracted from shape formed several points on top surface of building. Feature tracking is realized using this texture data. To put it concretely, template pattern:  $I(x, y)$  in one image is matched to texture pattern:  $J(x + l, y + m)$  in next image. When this matched value :  $D(l, m)$  is a minimum value,  $(l, m)$  is displacement between the two images. As the high-resolution image have adequate resolution, simple operation is enough to match pattern.

$$D_i(l, m) = \frac{1}{(2d_i + 1)^2} \sum_{dx=-d_i}^{d_i} \sum_{dy=-d_i}^{d_i} I(x + dx, y + dy) * J(x + l + dx, y + m + dy) \quad (1)$$

The size of search window:  $d_i$  is determined by the distance from a center of balance to the most far point in all feature points.  $(l, m)$  is parameter in search domain. If delta points can be oriented on the image, the search domain in searching feature points except delta points, becomes  $(-W \leq l \leq W), (-W \leq m \leq W)$  to  $(x_{delta} - dx_1 \leq l \leq x_{delta} + dx_2), (y_{delta} - dy_1 \leq m \leq y_{delta} + dy_2)$ .

### 3.3 TRACKING ON VIDEO IMAGES AND TRANSFORMATION COEFFICIENTS

The displacements in video images are identified using the Lucas-Tomasi-Kanade method (Tomasi & Kanade, 1991)(Jianbo Shi & Tomasi, 1994). This algorithm is robust and simple method. Feature tracking is done in video image sequence: frame No.m, m+1, ..., m+n in Figure ??.

A helicopter carries two cameras parallel. All 2D feature points:  $(x_H, y_H)$  on each calibrated high-resolution image are transformed to 2D feature points:  $(x_v, y_v)$  on the calibrated video image as follows. The transformation model is simple; we consider translation, scaling, and rotation. The transformation coefficients are calculated as follows.  $(x_{v(i)}, y_{v(i)}), i = 1 \cdot \cdot \cdot N$  are measured in the video image,  $(x_{H(i)}, y_{H(i)}), i = 1 \cdot \cdot \cdot N$  are measured in high-resolution image linked to one video image.  $p = x, y, \dagger$  means a Moore-Penrose inverse matrix.

$$\begin{bmatrix} x_H \\ y_H \end{bmatrix} = \begin{bmatrix} a_x & b_x \\ a_y & b_y \end{bmatrix} \begin{bmatrix} x_v \\ y_v \end{bmatrix} + \begin{bmatrix} D_x \\ D_y \end{bmatrix}, \quad \begin{bmatrix} a_p \\ b_p \\ D_p \end{bmatrix} = \begin{bmatrix} x_{v(1)} & y_{v(1)} & 1 \\ x_{v(2)} & y_{v(2)} & 1 \\ \vdots & \vdots & \vdots \\ x_{v(N)} & y_{v(N)} & 1 \end{bmatrix}^\dagger \begin{bmatrix} PH(1) \\ PH(2) \\ \vdots \\ PH(N) \end{bmatrix} \quad (2)$$

The high-resolution images can be captured about 2 [fps], the video images can be captured 30 [fps]. The trajectory from feature points on the high-resolution images is discrete. This trajectory can be completed by trajectory from feature points in the video images.

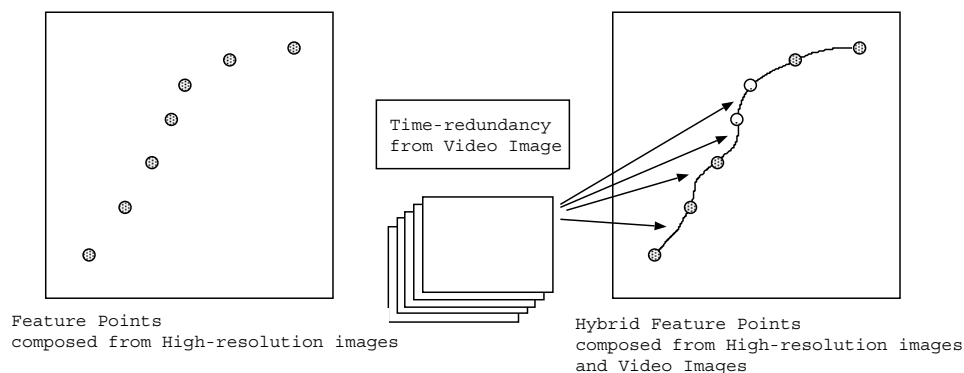


Figure 4: Trajectory of Feature Points

Moreover, this transformation is effective for texture mapping. When the operator set feature points on a initial high-resolution image, texture data corresponding to the top surface of acquired 3D object shape is cutting automatically from video image, and fitting to acquired 3D object model.

#### 4 SHAPE RECOVERY WITH FACTORIZATION METHOD

We have already developed an improved factorization method that utilize sensor information (Miyagawa, 1999)(Miyagawa, 2000). Sensor information, i.e. yaw, pitch, and roll rotation can be used to create the camera motion matrix  $[M]$ . A registered measurement matrix  $[A]^*$  composed from 2D feature points is decomposed using this camera matrix into shape matrix  $[S]$ . Here,  $[A]^* \cong [U_1][W_1][V_1]$  using the rank-3 property of singular value matrix. In this paper, we expanded this method to fit hybrid feature points.

$$[S] = (([U_1]^T[U_1])^{-1}[U_1]^T[M])^{-1}[W_1][V_1] \tag{3}$$

Moreover, 3D object shape can be placed within the unique coordinate system that we call Tokyo Datum.

##### 4.1 3D RECONSTRUCTION ON DELTA POINTS

It is important to reconstruct 3D object models on the delta planes, because the height of each object is determined as the distance between points on the building’s roof and the delta plane. However, it is difficult to recognize the height of acquired 3D delta points using the factorization method, without the orthometric height data on the ground. It is shown in Figure 5(left side) that each 3D object model is reconstructed on a flat plane. This flat plane has no relationship to sea level. If delta points can be given 3D coordinate value from the orthometric height data, we can make good use of the delta plane as a slope plane. In this case, each 3D object model can be reconstructed relative to sea level (Figure 5(right side)). We can reconstruct 3D object models on any delta plane if the delta points have orthometric height data on the ground.

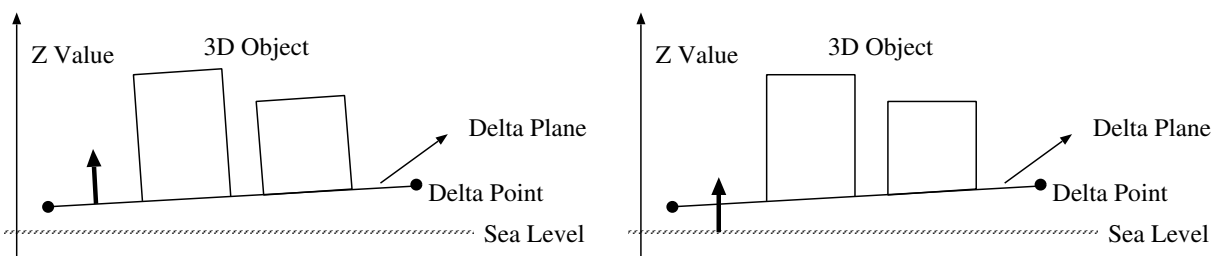


Figure 5: 3D Reconstruction on Delta Plane

## 4.2 MAPPING ON THE COORDINATE SYSTEM IN JAPAN

We introduce Tokyo Datum to map 3D object models. Tokyo Datum is the 3D-GIS coordinate system of Japan, and has been applied to many 3D-GIS applications (e.g. car navigation systems, orthometric height data on the earth, GPS). The X-axis in Tokyo Datum is along north, the Y-axis is along east, and Z value Japan is the geoid height relative to sea level. Acquired 3D values  $(X_r, Y_r, Z_r)$  are points on a relative coordinate system. These 3D values must be projected to absolute coordinate system (Tokyo Datum). Here, the 3D delta points acquired using the factorization method are  $(X_{r(i)}, Y_{r(i)}, Z_{r(i)})$ ,  $i = 1, 2, \dots, N$ , the 3D delta points measured on orthometric height data are  $(X_{jcs(i)}, Y_{jcs(i)}, Z_{jcs(i)})$ ,  $i = 1, 2, \dots, N$ .  $(X_{r(i)}, Y_{r(i)}, Z_{r(i)})$  can be transformed into  $(X_{jcs(i)}, Y_{jcs(i)}, Z_{jcs(i)})$  as follows.

$$\begin{bmatrix} X_{jcs} \\ Y_{jcs} \\ Z_{jcs} \end{bmatrix} = \begin{bmatrix} a_{X1} & a_{X2} & a_{X3} \\ a_{Y1} & a_{Y2} & a_{Y3} \\ a_{Z1} & a_{Z2} & a_{Z3} \end{bmatrix} \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} + \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix}, \quad \begin{bmatrix} a_{P1} \\ a_{P2} \\ a_{P3} \\ P_o \end{bmatrix} = \begin{bmatrix} X_{r(1)} & Y_{r(1)} & Z_{r(1)} & 1 \\ X_{r(2)} & Y_{r(2)} & Z_{r(2)} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ X_{r(N)} & Y_{r(N)} & Z_{r(N)} & 1 \end{bmatrix}^{\dagger} \begin{bmatrix} P_{jcs(1)} \\ P_{jcs(2)} \\ \vdots \\ P_{jcs(N)} \end{bmatrix} \quad (4)$$

## 5 EXPERIMENTS

In field experiments, we captured image sequences and sensor information as follows. The width of the earth plane captured by the camera ranged from 200 [m] to 250 [m]. The helicopter moved along one direction, for example, from south to north. The Differential GPS (DGPS) mounted on the helicopter captured rotation parameters as sensor information. The two kinds of images, high-resolution and video images were calibrated using Tsai’s method (Tsai, 1986).

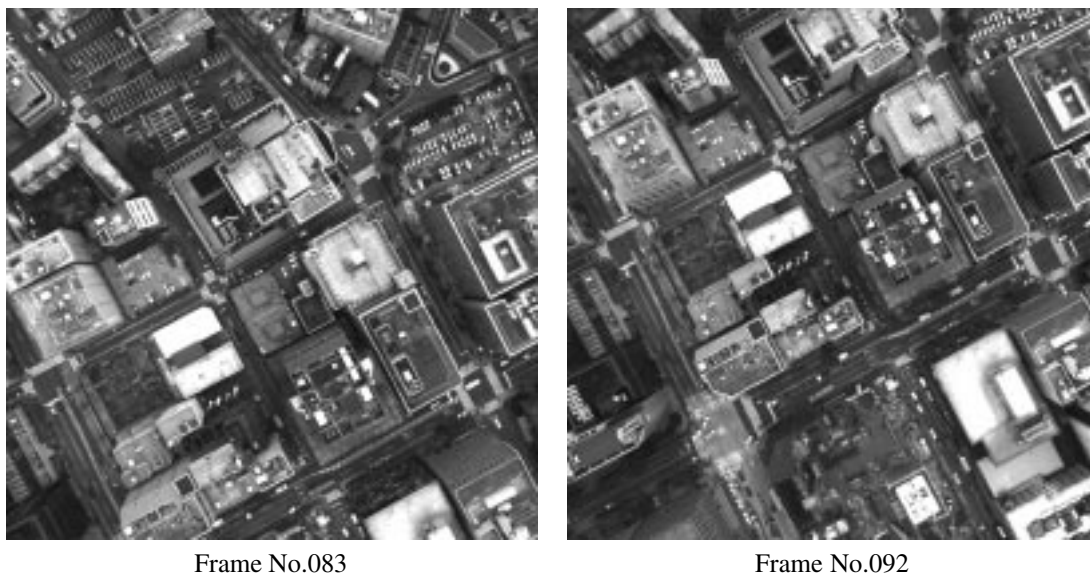


Figure 6: Aerial Images

Table 1: Field Conditions

	Specification	Remarks
Camera	video : F= 6.7 [mm], Full Color	640-by- 480 pixels, 30 [fps]
	high-resolution : F=20.0 [mm], Gray-scale	2000-by-2000 pixels, 2 [fps]
DGPS	Orientation : yaw, pitch, roll degree	measurement error : ±0.2 [deg]
Helicopter	average velocity ranged from 40 to 50 [km/h]	altitude of about 300 [m]

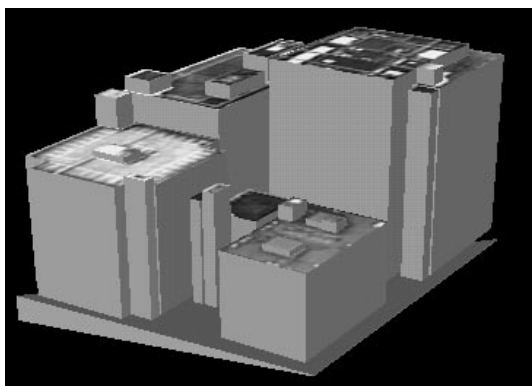
The high-resolution images are shown in Figure 6. The image labeled Frame No.083 is an initial image, on which the operator set some feature points. We used 10 images without frame-out or occlusion to acquire 3D object shape. We estimated the accuracy with 3D acquisition and recovery from both feature points and hybrid feature points. The 3D reconstruction models are shown in Figure 7. The accuracy estimation is given in Table 2. Method A means paraperspective factorization method (Poelman & Kanade, 1997), Method B means paraperspective factorization method with hybrid feature points, and Method C means paraperspective factorization method that utilize sensor information (Miyagawa, 2000).

Some DEM (Digital Elevation Map) heights in the table were measured by a laser-ranging-system. There were cases which buildings heights could not be estimated using this system. The DEM heights in this table were the average values from some points on the top surfaces of buildings.

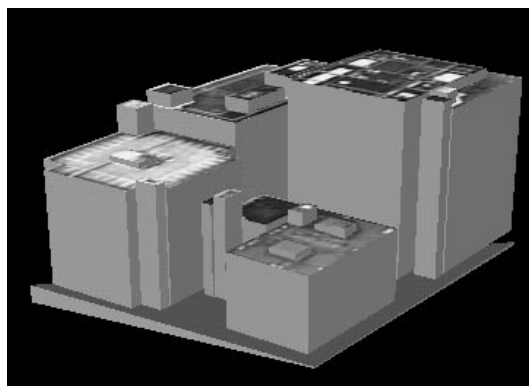
Table 2: Accuracy Estimation

DEM height	Method A	Method B	Method C
26.97	31.72	28.77	27.72
27.97	32.09	29.18	27.47
27.48	30.34	27.51	26.70
33.00	39.55	35.38	35.75
36.16	39.37	35.09	36.16
37.34	42.26	37.37	39.49
32.70	36.38	32.25	34.58
39.51	45.72	40.32	41.74
36.79	44.50	42.13	40.45
40.75	47.40	42.10	43.48

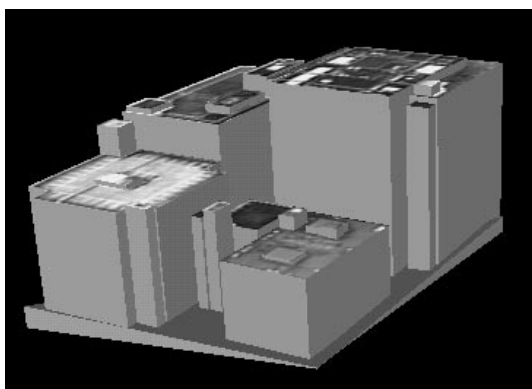
The results show that some of the 3D object shapes acquired by Method B were more accurate than those yielded by Method C. We confirmed that the factorization method with hybrid feature points was effective in achieving accurate 3D object shape.



Method A



Method B



Method C

Figure 7: 3D Reconstruction

A comparison with 3D reconstruction on a flat plane and a slope plane based on 3D delta points, is shown in Figure 8. The 3D reconstruction on flat plane fails to realize shape recovery of buildings. Though the ground is unevenness, 3D object



model stand on the flat plane. However, this assumption cause the distortion of 3D shapes. 3D model reconstruction in considering of gradient was very effective.

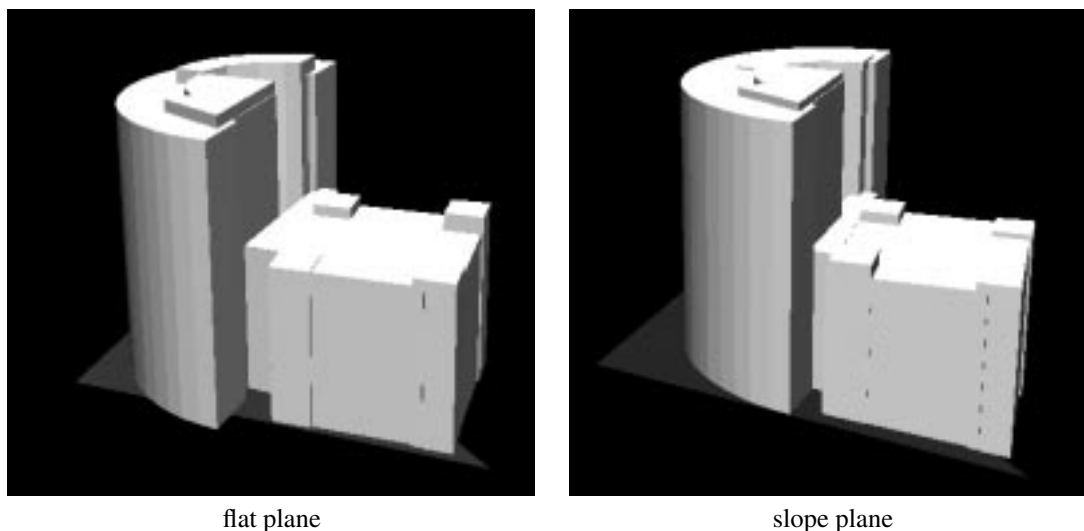


Figure 8: 3D Reconstruction on Delta Plane

We also confirmed the effect by hybrid feature points in Figure 9. The hybrid method yields accurately shaped top surfaces of buildings.

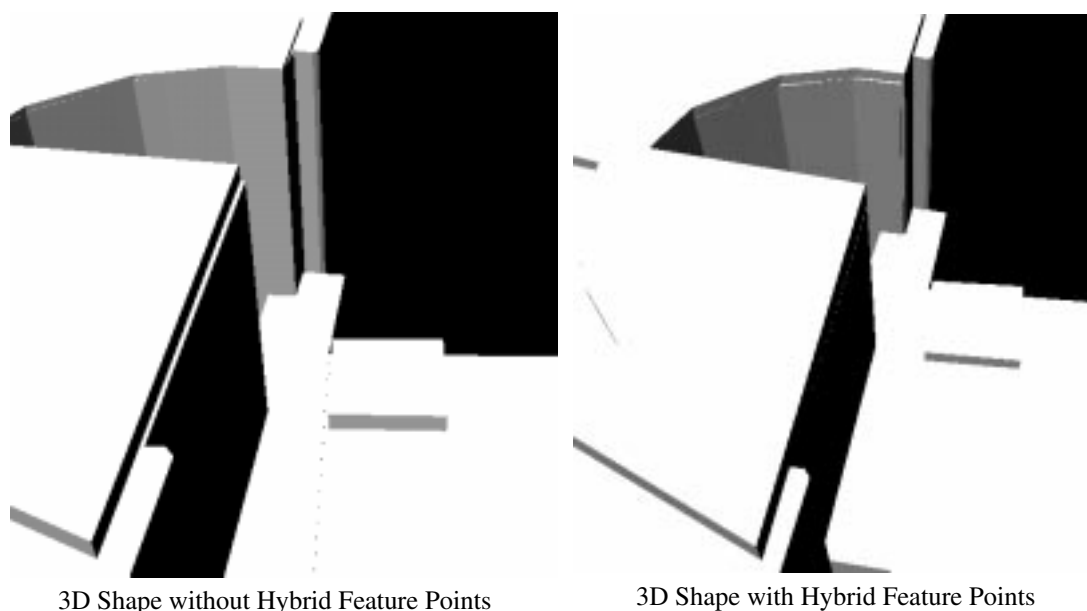


Figure 9: Effect by Hybrid Feature Points

## 6 CONCLUSION

We investigated an acquisition and recovery method that processes hybrid feature points extracted from both high-resolution and video images using the factorization method. The 3D object models were assumed to have N-point column shape, and were recovered on a plane formed from delta points corresponded to orthometric height data. Acquired and recovered 3D models were mapped on Tokyo Datum. Comparison experiments showed that the hybrid method was more effective and accurate in acquiring and recovering 3D object shape than a earlier method. Moreover, we confirmed that the hybrid method yielded accurately shaped top surfaces of building. We can expect the factorization method with hybrid feature points combining high-resolution images and Hi-Vision images to reconstruct 3D object shape more accurately.

---

**REFERENCES**

Martin Reddy, Yvan Leclerc, Lee Iverson, and Nat Beletter, 1999. TerraVision II: Visualizing Massive Terrain databases in VRML. *IEEE Computer Graphics and Applications*.

Carlo Tomasi and Takeo Kanade, 1992. Shape and Motion from Image Streams - a Factorization Method Full Report on the Orthographic Case. *CMU-92-104*.

Conrad J Poelman and Takeo Kanade, 1997. A Paraperspective Factorization Method for Shape and Motion Recovery. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.3.

Carlo Tomasi and Takeo Kanade, 1991. Detection and Tracking of Point Features. *CMU-CS-91-132*.

Jianbo Shi and Carlo Tomasi, 1994. Good features to track. *CVPR94*, p593-600.

R Y Tsai, 1986. An Effect and Accurate Camera Calibration Technique for 3D Machine Vision. *CVPR*.

I.Miyagawa, S.Nagai, K.Sugiyama, 1999. Recovery of Building Shape from Aerial Images using Factorization Method with Sensor Information. *UM3-99*.

I.Miyagawa, S.Nagai, K.Sugiyama, 2000. Shape Recovery from Aerial Images using Factorization Method with Sensor Information. *ACCV2000*.