3D Modeling of Building Sides from the Stereo Images for the Realistic Virtual City in 3D GIS

Yun Koo Chung*, Kyung-Ok Kim* and Joon Hee Han**

***ETRI / Image Processing Department**
Gajeong-dong 161, Yusong-gu, Taejon,Korea
Phone: +8242-860-5032, Fax:+8242-860-4844
E-mail: ykchung@etri.re.kr

**** Pohang University of Science and Technology /**
Computer Science & Engineering Department
Phone: +82562-279-2247
E-mail: joonhan@postech.ac.kr

Abstract

Remote sensing (RS) data show the surfaces of the earth only but cannot provide the shape data of building sides. The proposed method recovers a 3D shape of building sides from stereo images. Its result shows a higher possibility for recovering a large shaped object by overcoming the difficulties of traditional stereo matching techniques. The urban area will be visualized more realistically than the current model based on graphic and vector data.

Keywords: stereo matching, reconstruction of 3D shape

I. Introduction

There have been active research activities in stereo matching during the past twenty years. The results have been improved in correct matching ratio and processing time for practical usage. Many stereo techniques [1-3] use feature based technology since unique matching for all the pixels is not available since some pixels of the local area on the images are similar. In this paper, we suggest a technique to recover the three dimensional shape of a large object from the image sequence using three images. We focused to recover an approximated large shape and to represent its shape even though the extracted information is not complete enough. The texture image is put on the reconstructed shape using texture mapping. This method enables a large shaped objects in the real world to be represented realistic. The recovering procedures are as following: feature extracting, feature matching, dense matching, and recovering 3D shape.

II. Stereo Matching

The total scheme consists of stereo matching and reconstruction of 3D shape. The first phase is a stereo matching. The target task is to find the corresponding features between two images. We propose the following steps to get the highly correct matching.

1) Feature extraction

Feature is distinguished geometrically and by brightness difference to its neighborhood points.

This characteristic is utilized to find correspondent feature points on the other image. The corner points are used as a feature. In this paper the corner detector proposed by Smith and Brady[4] is modified. The corner is detected by the degree of similarity to its neighborhood points. The similarity degree is defined as S(p,q) as following:

$$S(p,q)=\begin{cases}1 & if \quad thr_{low} < |I(p)-I(q)| < thr_{up} \\ 0 & otherwise\end{cases} \qquad (1)$$

$$N(p) = \sum_{W} S(p,q) \qquad (2)$$

$$C(p) = \begin{cases}1 & if \quad N(p) < 0.5 mask\_size \\ 0 & otherwise\end{cases} \qquad (3)$$

The Unique Segment Assimilating Nucleus (USAN) at the point p is computed as N(p). The degree of corner is depending on the value of N(p). The detector determines a point as a corner point if its mask window size is less than the half of the mask window. Also, if the value is small, the corner degree becomes bigger. We call this corner detector SUSAN (Smallest USAN) corner detector.

2) Initial matching

The feature points are obtained from the three images. Then, matching processes are performed to find corresponding points between features of two images. The NCC(Normalized Correlation Coefficient) method is used, where the matching is processed using normalized correlation scores to the corner features. Normalized correlation score C(p1,p2) is defined as following formula:

$$C(p_1,p_2)=\frac{1}{K}\sum_u\sum_v\left[I_1(p_1.x+u,p_1.y+v)-\bar{I}(p_1)\right]$$

$$x\left[I_2(p_2.x+u,p_2.y+v)-\bar{I}(p_2)\right] \qquad (4)$$

,where $K=(2m+1)(2n+1)\sigma(p_1)\sigma(p_2)$.

The C(p1,p2) is scored from −1 to +1. The possibility of feature matching becomes higher when the value is higher. Normalized correlation simply lessons the affects of distortion by scaling and viewing angles. The processing time can be saved by the limited size of searching area.

3) Feature rematching by the epipolar geometry

After initial matching results, the false matching could exist. There exists epipolar geometry between the stereo images and it is recovered using the fundamental matrix **F**. The corresponding feature points from initial matching points are selected for computing the fundamental matrix **F**. The fundamental matrix **F** is a matrix which has all of the geometrical information between two images. To the three dimensional point M, let's call the projective point on one image plane as m, and the point on the second image plane as **m'**. There exists a following relationship between two images:

$$\mathbf{m'}^T\mathbf{F}\mathbf{m}=\mathbf{0} \qquad (5)$$

This equation (5) is called the epipolar equation. If we use this **F**, we can find the corresponding feature points on the other image to any feature points on one image using the following line equations:

$$l' = \mathbf{F}m, \quad l = \mathbf{F}^T m' \qquad (6)$$

After this process, the false matching points are removed as the following "least median of squares" method. The number of initial matching points is n. We can make arbitrary set of eight feature points. Let's call this group of eight points J set. The fundamental matrix $\mathbf{F_J}$ is determined as following equation.

$$M_j = \mathrm{med}_{i=1..n}[\,d^2(m_{2i}, F_j m_{1i}) + d^2(m_{1i}, F_j^T m_{2i})] \qquad (7)$$

The M_j, which is the least value among the all of the M_j's, is called SM_J. The robust standard deviation is computed using the following formula:

$$\hat{\sigma} = 1.4826[1+5/(n-p)]\sqrt{SM_J} \qquad (8)$$

The false matching features are determined if

$$r_i^2 \le (2.5\hat{\sigma})^2 \quad (9)$$

where, $r_i^2 = d^2(m_{2i}, Fm_{1i}) + d^2(m_{1i}, F^T m2)$.

III. Reconstruction of 3D Building Sides

The shape of 3D building sides is recovered using projective reconstruction and Euclidean reconstruction using three images. There are two ways to recover the shape of 3D object. The first way shows the feeling of 3D shape for representing the 3D shape in the cyber space. If an object is of an arbitrary shape and needs a dense 3D information, the next step is to do the dense matching. And if it needs not real world coordinates and only needs a 3D shape, the next process is to compute a disparity map with scaling and texture mapping.

The second way provides a 3D geometrical shape with known knowledge. For example, the geometrical shapes of streets and buildings mostly consist of planes and polygons. Therefore, it needs not pixel by pixel information to represent buildings and streets. It needs a coarse geometrical information with digital image. Texture mapping makes the recovered shape more realistic representation. Here, we use the second way to represent a building in the cyber space.

1) Projective reconstruction

When the 3D coordinates of an image point is reprojected through the camera matrix, the position of the projected point should be the same as the position of the image point. Using this constraint, matching features from the given image sequence are used to compute the camera matrix and their 3D coordinates. We call this process "projective reconstruction". The projective reconstruction method is based on the factorization recovering method by Bill Triggs.[7] The projective depth λ_{ik} from more than two images is defined as following: Let's call X_k homogeneous coordinates of 3D coordinates and call P_i camera projective matrix. λ_{ik} is a projective depth.

$$\lambda_{ik} x_{ik} = P_i X_k \quad (10)$$

Using the equation 10, all the points on an image can be defined as following matrix:

$$W = \begin{bmatrix} \lambda_{11}x_{11} & \lambda_{12}x_{12} & \cdots & \lambda_{1n}x_{1n} \\ \lambda_{21}x_{21} & \lambda_{22}x_{22} & \cdots & \lambda_{mn}x_{mn} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{m1}x_{m1} & \lambda_{m2}x_{m2} & \cdots & \lambda_{mn}x_{mn} \end{bmatrix} = \begin{bmatrix} P_1 \\ P_1 \\ \cdots \\ P_1 \end{bmatrix} \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix} \quad (11)$$

$$\lambda_{ik} = \frac{(e_{ij} \wedge x_{ik}) \cdot (F_{ij} x_{jk})}{\left\| e_{ij} \wedge x_{ik} \right\|^2} \lambda_{jk} \quad (12)$$

If the value of λ_{1k} is assigned by 1, all the other λ_{jk} can be computed. The matrix W is computed from x_{ik} and λ_{ik}. W is transformed to UWV^T using SVD method. The rank of W should be 4. Hence, the values from the fifth elements are reset to zero.

$$W = U\hat{W}V^T = (U\hat{W}_1)(U\hat{W}_2 V^T) = \hat{U}\hat{V} \quad (13)$$

$\hat{U}\hat{V}$ is computed. The computed matrix is called P_i. In order to compute the P_i, let's set P_1 to [I|0]. From this, we get P_i, i = 1,....,m. P_iT and $T^{-1}X$ are projective reconstruction. T is multiplied to P_i, i = 1, ... ,m.

$$T = \begin{bmatrix} \tilde{p}_1^{-1} & -\tilde{p}_1^{-1} p_0 \\ 0 & 1 \end{bmatrix} \quad (14)$$

2) Euclidean reconstruction

If there is no skew and if the aspect ratio and the position of the principal point can be known, assign the principal point (u,v) to (0,0), aspect ratio α_u and α_v as 1. Then, the linear method can be applied to solve the parameters of Euclidean reconstruction matrix[7]. Let $\alpha_x = \alpha_y = \alpha$, s=0, $u_0 = v_0 = 0$as initial values.

$$H = \begin{bmatrix} K & 0 \\ d^T & 1 \end{bmatrix}, \qquad d^T = [a,b,c]^T \quad (15)$$

$$\lambda_i P_i H H^T P_i^T = KK^T \quad (16)$$

$$\lambda_i P_i \begin{bmatrix} \alpha^2 & 0 & 0 & a\alpha \\ 0 & \alpha^2 & 0 & b\alpha \\ 0 & 0 & 1 & c \\ a\alpha & b\alpha & c & a^2+b^2+c^2 \end{bmatrix} P_i^T = \begin{bmatrix} \alpha^2 & 0 & 0 \\ 0 & \alpha^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (17)$$

$$w = KK^{T.} \quad q=[\ \alpha^2, a\alpha\ , b\alpha\ , c, \|d\|^2, 1\ ]^T. \quad (18)$$

From the above equations, q can be solved using SVD linear method. The values of K and H are solved from the value of q using the nonlinear minimization method "Levenberg-Marquartdt" algorithm.

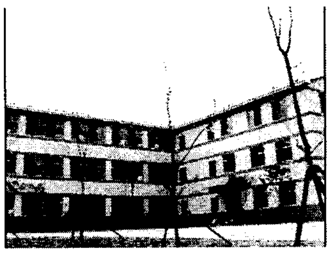
At the next step, the P^e와 X^e are recovered using the P and H such that $P^e = PH$ and $X^e = H^{-1}X$.

3) **Texture mapping**

The texture image is put on the recovered 3D shape of building sides. The mapping processes consist of eliminating a background, selecting the reference points, and computing the mapping function between the reference points of the shape and the texture image.

IV. Experimental Results

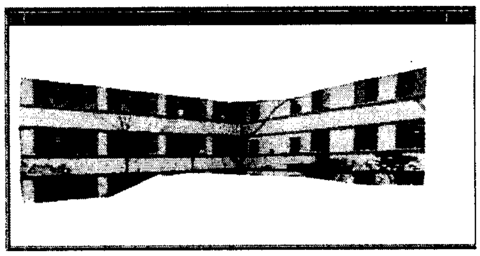
We applied the whole processes described at the above to recover a shape of building sides. We did not complete the whole processing for reconstruction of an entire building, but we have shown the high possibility of reconstructing the shape of building sides through implementing the algorithms. <Figure 1> is an original image of a building. <Figure 2> shows results of feature extraction and matching, and <figure 3> shows views of the recovered shapes of the building sides.



<Figure 1> Original image

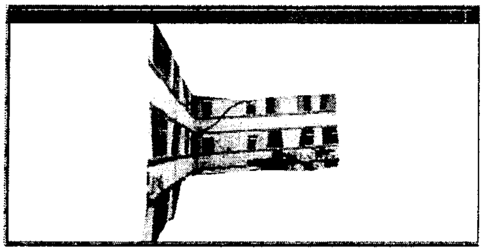
a) Feature extraction b) Feature matching

<Figure 2> Feature based matching result

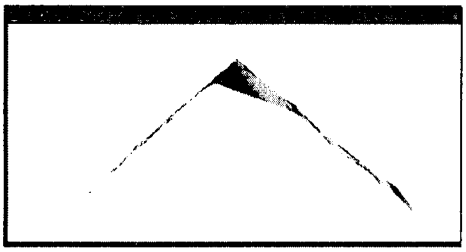


a) viewing 1

<Figure 3> Recovered building sides (continues)



b) viewing 2



c) viewing 3

<Figure 3> Recovered building sides

V. Conclusion

We presented a recovering process of the 3D building sides and implemented the scheme. This method can be utilized to represent the shape of urban buildings for the virtual city in 3D GIS. The traditional method has limitations in the size of an object and illumination environment. But this scheme has advantage over these limitations and gives more convenience to use. Our scheme has shown the high possibility of recovering building shapes. In future, we are to improve correct matching ratio and position errors of features for higher precision of the recovered 3D shape. The work will be connected to the GIS data for representation of realistic 3D buildings in 3D GIS.

Reference

[1] Geiger, D., Ladendorf, B. and Yuille, A.L., "Binocular stereo with occlusions", *Second European Conference on Computer Vision*, 1992, pp.425-433

[2] Medioni, G. and Nevatia, R., "Segmented-based Stereo Matching," *Computer Vision Graphics and Image Processing*, vol. 31, pp. 2-18, 1985.

[3] Marapane, S. B. and Trivedi, M. M., "Region-based Stereo Analysis for Robotic Application," *IEEE Transaction on System Man and Cybernetics*, vol. 19, no. 6, pp. 1447-1464, Nov./Dec. 1989.

[4] Smith, S.M. and Brady, J.M., "New approach to low level image processing", *Int. J. Computer Vision*, vol.23(1), pp.45-78,1997

[5] Harris, C.G., "Determination of ego-motion from matched points," in proc. *Alvey Vision Conf.*,1987

[6] Deriche, R., Zhang, Z., Luong, Q.T. and Faugeras, O., "Robust recovery of the epipolar geometry from an uncalibrated stereo rigidity," *ECCV*,vol.800,pp.567-575,1994

[7] Sturm, P. and Triggs, B., "A factorization based algorithm for multi-image projective structure and motion", *ECCV'96*, pp.II:709-720,1996