URBAN VISUALIZATION THROUGH VIDEO MOSAICS BASED ON 3-D MULTI-BASELINES

Jeachoon CHON, Tkashi FUSE, Eihan SHIMIZU

Dept. of Civil Engineering, TheUniversity of Tokyo, JAPAN
jjc7151@trip.t.u-tokyo.ac.jp, (fuse, shimizu)@civil.t.u-tokyo.ac.jp

KEY WORDS: Video Mosaics, 3-D space, Image sequence, Virtual realization, GIS

ABSTRACT:

In case of using an image sequence taken from a video camera mounted on a moving vehicle ban, general image mosaicing techniques based on a single baseline cannot create image mosaics. To solve the drawback, we proposed a new image mosaicing technique that can create an image mosaic in 3-D space from the image sequence utilizing the concept of 3-D multi-baselines. The key point of the proposed method is that each image frame has a dependent baseline calculated by using camera pose and average depth between a camera and 3-D objects. The proposed algorithm consists of 3 steps: feature point extraction and tracking, calculation of camera exterior orientation, and determination of multi-baselines. This paper realized and showed the proposed algorithm that can create efficient image mosaics in 3-D space from a real image sequence.

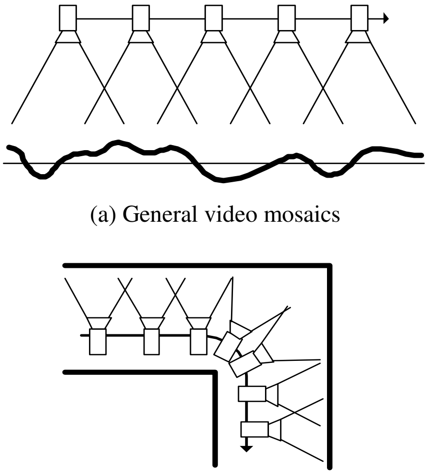
1. INTRODUCTION

Image mosaicing technique, which builds an image covering large areas through registering small 2-D images, can be used in many difference applications like satellite imagery mosaics (USGS Hurricane Mitch Program Projects), the creation of virtual reality environment (Szeliski, R., 1996), medical image mosaics (Chou *et al.,* 1997), and video compression (*Standard MPEG4).* Especially in GIS field, video mosaics are becoming more and more common in civil engineering that is representing urban environments, and managements of construction sites and road facilities.

The image mosaicing techniques are fall into two fields. In the first field, images and orthosatellite imagery, which is obtained by using the direct linear transform based on spatial data such as the digital element model, are registered to spatial vectors. In the second field, general images of a perspective projection are conjugated without spatial information. The techniques of the second field enable us to obtain spatial information and extract textures from stereo image mosaics.

Our research pertains only to the second filed. The mosaicing techniques can be mainly divided into four categories: a 360 degree panorama based on cylinder baseline projection (Shum *et al.,* 2000), a spherical mosaics based on spherical baseline projection (Coorg *et al.,* 2000), general video mosaics based on a single baseline projection (Zhu *et al.,* 2001), and x-slit images that can create image mosaics without baseline (Assaf *et al.,* 2003). In case of extracting a single texture of a facade and spatial data from panoramas and spherical mosaics, the data must be got through combing data extracted from several panoramas and spherical mosaics (Coorg *et al.,* 1999). Moreover, since the transition among image mosaics is discrete, the walkthrough in virtual reality is not smooth. On the other hand, since general video mosaics technique is to get image data in wide range, it is very efficient to extract textures of facades and spatial data. The video mosaic technique creates an image as projecting all of image sequence to a single baseline (see Fig. 1(a)), but it can't be applied to image sequence taken from a translating and rotating camera (see Fig. 1(b)). The single baseline is generally calculated by the average depth of feature points extracted and matched at 1st and 2nd image frames in case of using perspective projection.

Even if the algorithm of the x-slit images has the merit of creating an image mosaics from a translating and rotating camera, image motion per image frame is limited to 1 pixel for creating image mosaics as high resolution. Since the distance between buildings and a moving camera in urban area is very short range, generally the image motion is over 50 pixels at least. It is difficult to apply the algorithm of the x-slit images to urban area.



(a) General video mosaics

(b) A moving camera in a turning point

Fig. 1. Concept of general video mosaics.

To solve the drawback, this paper proposed a novel method that can creates video mosaics in 3-D space based on 3-D multi-baselines proposed by this paper. The core of the novel method is that each image frame has a dependent baseline calculated by using camera pose and average distance between camera focus and objects that are projected to the image plane of camera. Since each image frame is projected to itself dependent baseline, we can create video mosaics from a moving and rotating camera. The proposed algorithm consists of 3 steps: calculation of optical flow through hierarchical strategy, calculation of camera exterior orientation using collinearity equations, and determination of multi-baselines. This paper realized and showed the proposed algorithm that can create efficient image mosaics in 3D space from an real image sequence.

2. FEATURE BASED OPTICAL FLOW DETECTION

Camera orientation is computed utilizing optical flows which are obtained from sparsely located feature points that are detected using SUSAN algorithm (Smith *et al.,* 1997). Based on such feature points, the correlation and the contour style are computed and utilized to determine the best matching pair of feature points. The false optical flows, which are significantly different from others, are removed in the procedure of the repeated conversion using a median filter.

3. FEATURE BASED OPTICAL FLOW DETECTION

We discuss the camera's exterior orientation on the assumption that the interior orientation of the camera has already been established and discuss about exterior orientation. If *P(X,Y,Z)* denote the Cartesian coordinates of a scene point with respect to the camera, and if *(x,y)* denote the corresponding coordinates in the image plane, the image plane is located at the focal length *f* from the focal point *o(* X_L, Y_L, Z_L *)* of the camera. The perspective projection of a scene point *P(X,Y,Z)* on the image plane at a point where *p=(x,y)* is expressed as follows:

$$\begin{bmatrix} x \\ y \\ f \end{bmatrix} = \lambda M \begin{bmatrix} X - X_L \\ Y - Y_L \\ Z - Z_L \end{bmatrix}, \qquad (1)$$

where λ is the scale factor, X_L, Y_L, Z_L is the camera station, and M is a 3×3 rotation matrix defined as follows:

To eliminate the scale factor λ, we divided the first and second component equations in Eq. 1 by the third, leading to the following more familiar collinearity equations:

$$\begin{aligned} x &= f\frac{m_{11}(X - X_L) + m_{12}(Y - Y_L) + m_{13}(Z - Z_L)}{m_{31}(X - X_L) + m_{32}(Y - Y_L) + m_{33}(Z - Z_L)} \\ y &= f\frac{m_{21}(X - X_L) + m_{22}(Y - Y_L) + m_{23}(Z - Z_L)}{m_{31}(X - X_L) + m_{32}(Y - Y_L) + m_{33}(Z - Z_L)} \end{aligned} \qquad (2)$$

For simplicity, the collinearity equations are shown as following:

$$F = \begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} x - fU/W \\ y - fV/W \end{bmatrix}, \qquad (3)$$

where $[U \quad V \quad W]^T = M[X - X_L \quad Y - Y_L \quad Z - Z_L]^T$.

3.1 DEPENDENT RELATIVE ORIENTATION BETWEEN THE FIRST AND SECOND IMAGE FRAMES

To solve the relative orientation with the collinearity model, we can transfer the nonlinearity of the equations Eq. 4 to a linearized form Eq. 5 using Taylor series approximations. The condition equation can then be written in linearized form as:

$$-\tilde{F} + J^e\Delta^e + J^s\Delta^s + error = F, \qquad (4)$$

where $\tilde{F} = -F(\tilde{\omega}, \tilde{\phi}, \tilde{\kappa}, \overset{manual}{X_L}, \tilde{Y}_L, \tilde{Z}_L, x, y, f, \tilde{X}, \tilde{Y}, \tilde{Z})$, $\tilde{\omega}, \tilde{\phi}, \tilde{\kappa}, \tilde{Y}_L, \tilde{Z}_L, \tilde{X}, \tilde{Y}, \tilde{Z}$ are initial values, and $\underset{4n\times1}{\tilde{F}} = [\ldots \quad \tilde{F}_{x,t-1,i} \quad \tilde{F}_{y,t-1,i} \quad \tilde{F}_{x,t,i} \quad \tilde{F}_{y,t,i} \quad \ldots]^T$ $\underset{5\times1}{\Delta^e} = [\Delta\omega_t \quad \Delta\phi_t \quad \Delta\kappa_t \quad \Delta Y_{L,t} \quad \Delta Z_{L,t}]^T$ and $\underset{3n\times1}{\Delta^s} = [\ldots \quad \Delta X_i \quad \Delta Y_i \quad \Delta Z_i \quad \Delta X_{i+1} \ldots]^T$ are the vector form to the approximations for the parameters, $\underset{4n\times5}{J^e}$ and $\underset{4n\times3n}{J^s}$ are the matrix of the partial derivatives of the two functions in Eq. 4 with respect to each of the five exterior orientation elements and the three coordinates of the GCP, *i(=3…n)* is the index of the i^{th} optical flows, and *t* is the index of the image frames.

We used the first two images of an image sequence to determine the reference image frame. The world frame was typically aligned with the first image frame. The camera orientation at the second image frame can be calculated using the dependent relative orientation, Eq. 5, with the five pairs of optical flows and the $\overset{manual}{X_{L,2}}$ of the current camera station, which is input manually. Eq. 6 involves rewriting Eq. 5 for vectors, as follows:

$$\begin{bmatrix} N_{11} & N_{12} \\ N_{12}^T & N_{22} \end{bmatrix} \begin{bmatrix} \Delta^e \\ \Delta^s \end{bmatrix} = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}, \qquad (5)$$

where $N_{11} = J^{eT}J^e$, $N_{12} = J^{eT}J^s$, $N_{22} = J^{sT}J^s$, $n_1 = J^{eT}\tilde{F}$, and $n_2 = J^{sT}\tilde{F}$.

The parameter Δ^e is found as follows:

$$(N_{11} - N_{12}N_{22}^{-1}N_{12}^T)\Delta^e = n_1 - N_{12}N_{22}^{-1}n_2. \qquad (6)$$

After a few iterations of Eq. 7, we determine the exterior orientation of the camera at the second frame as follows: $[\omega_2 \quad \phi_2 \quad \kappa_2 \quad Y_{L,2} \quad Z_{L,2}]^T = [\tilde{\omega}_2 \quad \tilde{\phi}_2 \quad \tilde{\kappa}_2 \quad \tilde{Y}_{L,2} \quad \tilde{Z}_{L,2}]^T - \Delta^e$.

3.2 ABSOLUTE ORIENTATION FOR 3rd to last image framesDEPENDENT RELATIVE ORIENTATION BETWEEN THE FIRST AND SECOND IMAGE FRAMES

From the third image frame to the last image frame, the camera orientation can be calculated by using the GCP of three optical flows for three frames.

We can calculate the camera orientation using a minimum of three GCPs of optical flows that are calculated by using two camera orientations of the previous image frames. The calculation of the GCP of the optical flows is as follows:

$$\begin{aligned} X_i - X_{L,t-2} &= (Z_i - Z_{L,t-2})\overline{U}_{t-2} / \overline{W}_{t-2} \\ Y_i - Y_{L,t-2} &= (Z_i - Z_{L,t-2})\overline{V}_{t-2} / \overline{W}_{t-2} \\ X_i - X_{L,t-1} &= (Z_i - Z_{L,t-1})\overline{U}_{t-1} / \overline{W}_{t-1} \\ Y_i - Y_{L,t-1} &= (Z_i - Z_{L,t-1})\overline{V}_{t-1} / \overline{W}_{t-1} \end{aligned}, \qquad (7)$$

where $\overline{U} = m_{11}x_i + m_{21}y_i + m_{31}f$, $\overline{V} = m_{12}x_i + m_{22}y_i + m_{32}f$, $\overline{W} = m_{13}x_i + m_{23}y_i + m_{33}f$. The GCP is found with the aid of the classical least-squares solution.

$$\begin{bmatrix} X_i & Y_i & Z_i \end{bmatrix}^T = (A^T A)^{-1} A^T B, \qquad (8)$$

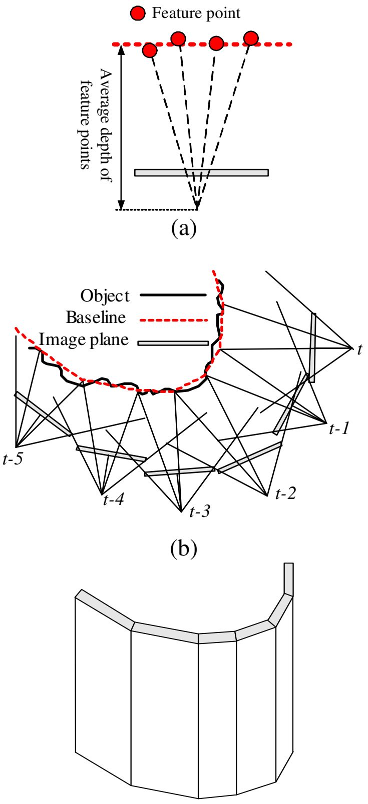
where $A = \begin{bmatrix} 1 & 0 & -\overline{U}_{t-2} / \overline{W}_{t-2} \\ 0 & 1 & -\overline{V}_{t-2} / \overline{W}_{t-2} \\ 1 & 0 & -\overline{U}_{t-1} / \overline{W}_{t-1} \\ 0 & 1 & -\overline{V}_{t-1} / \overline{W}_{t-1} \end{bmatrix}$ and $B = \begin{bmatrix} X_{L,t-2} - Z_{L,t-2}\overline{U}_{t-2} / \overline{W}_{t-2} \\ Y_{L,t-2} - Z_{L,t-2}V_{t-2} / \overline{W}_{t-2} \\ X_{L,t-1} - Z_{L,t-1}\overline{U}_{t-1} / \overline{W}_{t-1} \\ Y_{L,t-1} - Z_{L,t-1}\overline{V}_{t-1} / \overline{W}_{t-1} \end{bmatrix}$. If Eq. 6 is given at least three GCPs, the camera orientation is calculated as follows:

$$\Delta^e = (J^{eT} J^e)^{-1} J^{eT} \tilde{F}. \qquad (9)$$

After a few iterations of Eq. 10, we can determine the exterior orientation of the camera from the third image frame to the last image frame as follows: $\begin{bmatrix} \omega_t & \phi_t & \kappa_t & X_{L,t} & Y_{L,t} & Z_{L,t} \end{bmatrix}^T = \begin{bmatrix} \tilde{\omega}_t & \tilde{\phi}_t & \tilde{\kappa}_t & \tilde{X}_{L,t} & \tilde{Y}_{L,t} & \tilde{Z}_{L,t} \end{bmatrix}^T - \Delta^e$.

4. MULTI-BASELINES FOR CREATING VIDEO MOSAICS

As stated above, the 2D-image mosaic technique projects all of image frames on a single baseline to create an image with wide range. Since it can't be applied to image frames taken from a rotating camera, to solve the drawback, this paper proposes a novel method for creating image mosaics in 3-D space in case of using an image sequence taken from a translating and rotating camera.



(a)

(b)

(c)

Fig. 2. Independent baselines by using camera pose and average depth of feature points per image frame.

The core of the proposed method is determining dependent baseline to per image frame for creating multi-baselines on which all of image frames are projected. The pose of the baseline of image frame is the same pose of camera. The perpendicular distance between camera focus and baseline is evaluated as the average distance of optical flows like Fig. 2(a). the spatial transform of optical flows from the world coordinate to the coordinate of image plane is following as:

$$f(X_L, Y_L, Z_L, \omega, \phi, \kappa, x, y, z) = M^{\omega\phi\kappa}[x \quad y \quad z]^T + [X_L \quad Y_L \quad Z_L]^T \qquad (10)$$

Where (X_L, Y_L, Z_L) is camera station and (ω, ϕ, κ) is camera pose. Since the pose of baseline is the same pose of image plane, in case of Fig. 2(b), the multi-baselines are the thick dot-lines. In case that each image frames are projected on itself baseline, the result will be Fig. 2(c).

4.1 THE MODIFICATION OF FAULT BASELINE

From now, this section only describes the modification of multi-baseline in *XZ* coordination as compensating the reverse image rotation of *x* axis to itself image frame. Since the case of Fig. 2(b) based on a camera of which the image motion is the 5 DOF, $(T_X, T_Y, T_Z, \Omega_Y, \Omega_Z)$, with the exception of the image rotation of *x* axis, Ω_X, is an ideal state, it is able to create video mosaics in 3D space. In most instances, the motion of a camera has 6 DOF including the image rotation of *y* axis which is one of the difficult things to create video mosaics like Fig. 3(a) and Fig. 4. Therefore there are needed to modify the baseline like Fig. 3(b) and Fig. 5.

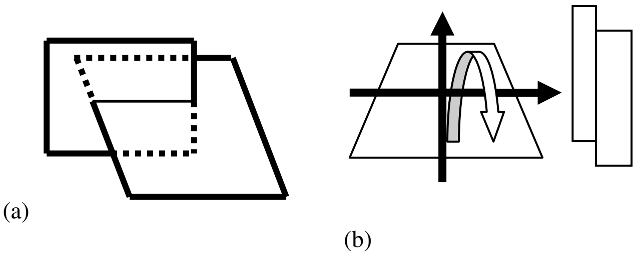


Fig. 3. The $-\Omega_X$ compensation of baseline for creating image mosaics.

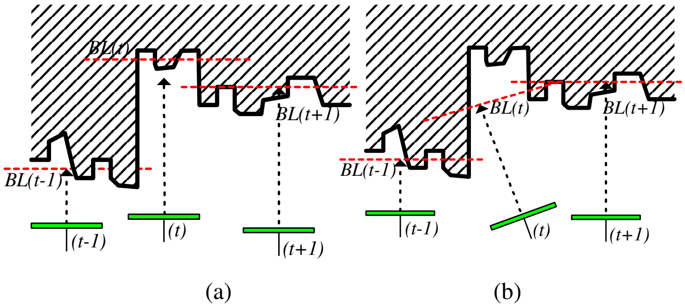


Fig. 4. Parallel baselines, fault intersection among baselines (*BL*), and infinite range in intersection area.

In case that baseline *BL(t)* parallels to neighbor baseline *BL(t-1)* and *BL(t+1)* originated that the distance among the objects projected on successive images is too long or image plane parallels neighbor image planes (see Fig. 4(a)), also there is faced with fault intersection among baselines like Fig. 4(b), it is difficult to create video mosaics. In those cases, the pose and the position of the baseline must be modified. To modify the fault intersections of baselines, this paper proposes 3 steps; at first step, evaluating all dependent baselines per image frame and then detecting the cases of Fig. 3 and Fig.4, at the last step, modifying the pose and the position of baselines like Fig. 5. At the second step, baseline must be checked whether the baseline intersects from the center of previous baseline *BL(t-1)* to $T_{X,t-1}/2$ and from $-T_{X,t+1}/2$ to the center of next baseline *BL(t+1)* or not. In case of detecting fault intersection, the baseline is modified as the line that passes $BL_{A,t-1} = f(X_{L,t-1}, Y_{L,t-1}, Z_{L,t-1}, 0, \phi_{t-1}, 0, T_{X,t-1}/2, 0, 0)$ and $BL'_{A,t} = f(X_{L,t+1}, Y_{L,t+1}, Z_{L,t+1}, 0, \phi_{t+1}, 0, -T_{X,t+1}/2, 0, 0)$. The image translation (T_X, T_Y, T_Z) is calculated as $f(0,0,0,\omega_{t-1}-\omega_t, \phi_{t-1}-\phi_t, \kappa_{t-1}-\kappa_t, X_{L,t}-X_{L,t-1}, Y_{L,t}-Y_{L,t-1}, Z_{L,t}-Z_{L,t-1})$. The equation of the baseline is following as:

$$Z = aX + b \tag{11}$$

Where $a = \dfrac{BL_{A,t-1,Z} - BL'_{A,t,Z}}{BL_{A,t-1,X} - BL'_{A,t,X}}$, $b = BL_{A,t,Z} - aBL_{A,t,X}$.

Video mosaics will be obtained through liking those baselines and general baselines like Fig. 6.

4.2 IMAGE PROJECTION ON THE MULTI-BASELINES

The equation of image projection on the multi-baselines is Eq (2) set camera station, camera pose, and the equation of baseline. In Fig. 7, the lines of *b* and *d* are calculated by using the length of baseline and the lines of *a* and *c* are calculated by using the height of image frame.

$$x = \frac{m'_{11}X + m'_{12}Y + m'_{13}}{m'_{31}X + m'_{32}Y + 1}$$
$$y = \frac{m'_{21}X + m'_{22}Y + m'_{23}}{m'_{31}X + m'_{32}Y + 1} \tag{12}$$

Where $m_{11}{}' = f\dfrac{(M_{11} + M_{13}a)}{m'_{33}}$, $m_{12}{}' = f\dfrac{M_{12}}{m'_{33}}$,
$m'_{13} = -f\dfrac{(M_{11}X_L + M_{12}Y_L + M_{13}Z_L - M_{13}b)}{m'_{33}}$,
$m_{21}{}' = f\dfrac{(M_{21} + M_{23}a)}{m'_{33}}$, $m_{22}{}' = f\dfrac{M_{22}}{m'_{33}}$,
$m'_{23} = -f\dfrac{(M_{21}X_L + M_{22}Y_L + M_{23}Z_L - M_{23}b)}{m'_{33}}$,
$m'_{31} = M_{31} + M_{33}a/m'_{33}$, $m'_{32} = M_{32}/m'_{33}$,
$m'_{33} = -(M_{31}X_L - M_{32}Y_L + M_{33}(Z_L - b))$.

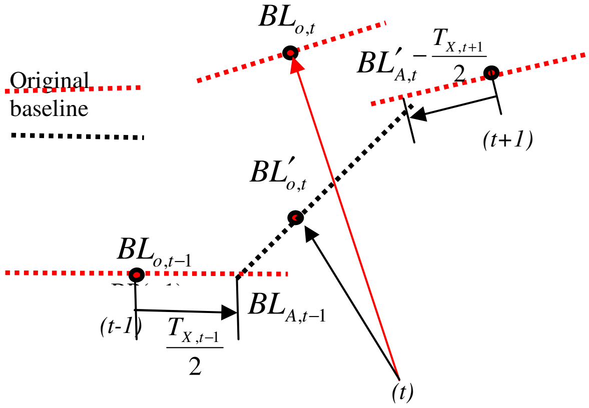


Fig. 5. New baseline in case of fault intersection among baselines (*BL*) and around intersection area.

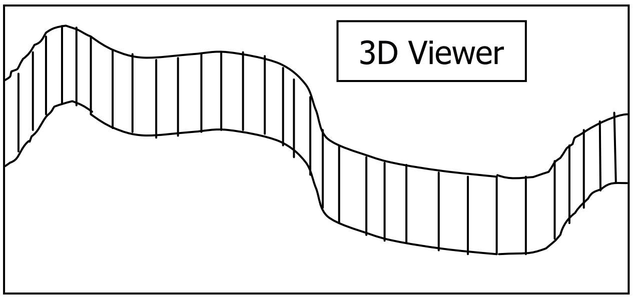


Fig. 6. Conceptual result by the proposed method.

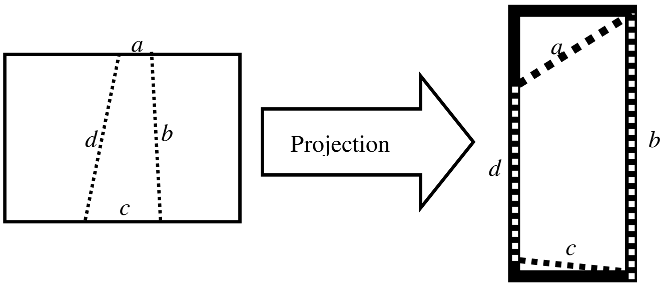


Fig. 7. Projection relationship between an image plane and a

base-plane.

5. EXPERIMENTAL RESULTS

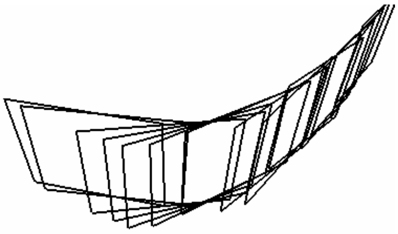
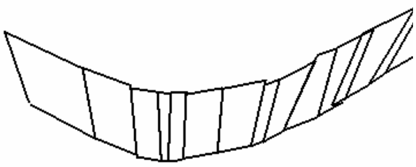
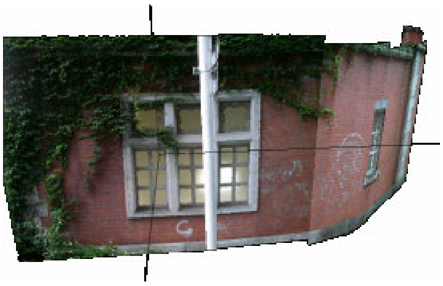


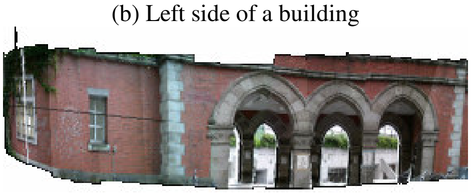
Fig. 8. Multi-baselines calculated by using an image sequence.



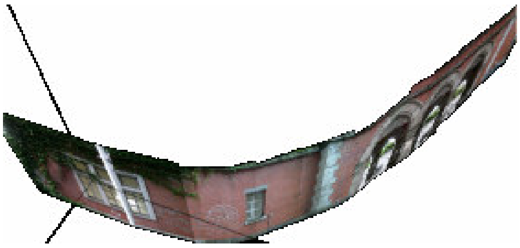
(a) Vectors of multi-baselines cut useless parts



(b) Left side of a building



(c) Front side of a building



(d) View-port with down ward angle

Fig. 9. Projection of image frames to multi-baselines.

For realizing the proposed method and displaying result in 3D space, we use ***Visual C++ 6.0*** and ***OpenGL***. The propose method has been applied to real image sequence taken from a translating and rotating video camera. Fig. 8 is about the vectors of multi-baselines in connection with the whole size of image frame. Fig. 9(a) is about cutting of useless part of the multi-baselines in Fig. 8 and Fig. 9(b) shows the left side of a house and Fig 9(c) shows the front side of the house in the result created by projecting image frames on the multi-baseline in Fig. 9(a). Figure 9(d) shows the result at view-port with downward angle.

6. CONCLUSION

The novel method for creating video mosaics in 3D space based on image sequence taken from a translating and rotating camera has been proposed in this paper. The core of the proposed algorithm is to create multi-baselines of which each baseline is determined by using the pose of camera and the average distance between camera focus and depths of optical flows per image frame. The proposed method is successfully applied to real-image sequence taken from a translating and rotating camera.

In case that the proposed method is applied to the video camera of vehicle-borne mobile mapping system which generally includes rotational and translational motion, the results by the proposed method will be applied to extract textures of facades and obtain rapid and inexpensive 3D spatial data instead of laser scanner. Also, since the results are continuous data, the results will be overcome the drawback of panoramas and spherical mosaics sequence of which the data is discrete for visual environment in large area.

References from Journals:

Assaf Zomet, Doron Feldman, Shmuel Peleg, and Daphna Weinshall, " Mosaicing New Views: The Crossed-Slits Priojection," *IEEE Transactions on PAMI*, Vol. 25, No. 6, June 2003.

Coorg, S., and S. Teller, 2000. Spherical mosaics with quarterings and dense correlation, *In IJCV,* 37(3): pp. 259-273.

Shum, H. and R. Szeliski, 2000. Systems and experiment paper construction of panoramic image mosaics with global and local alignment, *In IJCV*, 36(2): 101-130.

Smith, S.M. and J.M. Brady, 1997. SUSAN-a new approach to low level image processing, *In IJCV*, 23(1): 45-78.

Szeliski, R, 1966, Video mosaic for virtual enviroment, *Comput. Graph. Applicat.*, vol.16, no.3, pp. 22-30.

References from Other Literature:

Chou, J. S.,and J. Qian, Z. Wu, H. Schramm, 1997. Automatic mosaic and display from a sequence of peripheral angiographic images, *Proc. of SPIE*, Medical Imaging, California, 3034: 1077-1087.

Coorg, S. and S. Teller, 1999. Extracting textured vertical facades from controlled close-range imagery, *Proc. of 1999 IEEE Conference on Computer Vision and Pattern Recognition*, Colorado, pp.625-632.

Zhu, Z., A. R. Hanson, H. Schultz, E. M. Riseman, 2001. Error Characterics of Parallel-Perspective Stereo Mosaics, Proc. of *IEEE Workshop on Video Registration (with ICCV01)*, Vancouver, Canada, 13 July.

References from websites:

USGS Hurricane Mitch Program Projects, http://mitchnts1.cr.usgs.gov/projects/aerial.html