A STUDY ON AUTOMATIC IMAGE ARRANGEMENT IN URBAN AREA - TOWARD THE NEW CONCEPT OF URBAN VISUALIZATION -

T. Fuse [a, *], T. Wayama [b], E. Shimizu [c]

[a] National Institute for Land and Infrastructure Management, Asahi 1, Tsukuba, Ibaraki, 305-0804, Japan -
fuse-t92ta@nilim.go.jp
[b] Mizuho Securities Co., Ltd., Otemachi 1-5-1, Chiyoda-ku, Tokyo, 100-0004, Japan -
takayuki.wayama@mizuho-sc.com
[c] Dept. of Civil Engineering, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-8656, Japan, -
shimizu@civil.t.u-tokyo.ac.jp

Commission IV, WG IV/4

KEY WORDS: GIS, Matching, Multiresolution, Visualization, Urban, Navigation

ABSTRACT:

Recently, visualization of urban scenes attracts attention from the perspective of tourism and cultural city activation. Considering the popularity of the digital still cameras, a more impressive form of urban scenes, which allows the user to view the scenes from connected various points, is required. The still images can be connected in three dimensional space relatively. By compensating some lacks based on human spatial cognitive ability, a lively urban visualization is expected. In order to implement the urban visualization technique by using many still images, multiple resolution matching of images and spatial arrangement of the images are required. This study proposes a technique of urban visualization consisted of multiple resolution image matching, spatial image arrangement, and effective image projection for impressive visualization, and then explores the possibility of the new concept of urban representation based on the technique. The matching strategy for multiple resolution matching is composed of multiple resolution image creation by wavelet analysis, resolution ratio seeking, coarse to fine strategy matching, estimation of templates deformation, voted block matching. According to the matching results, relative orientation is applied, and then images are arranged in a three dimensional space. To keep the stability of the solution, calculation method for relative orientation is modified. The proposed technique was applied to images taken around a building, and formed a closed network. The matching result was enough to visualize the urban scene compared with manual matching. Through the application, the significance of the technique and limit of the technique was confirmed.

1. INTRODUCTION

Recently, visualization of urban scenes attracts attention from the perspective of tourism and cultural city activation. For the transmission of urban scenes, impressive representation is still challenging. Though walk-thorough and fly-through animation by using three dimensional city models is noteworthy (Debevec *et.al.*, 1996; Grzeszczuk, 2002; Pollefeys *et.al.*, 2004), the manual dependent modeling takes a lot of time and effort.

On the other hand, digital still cameras are now widespread, and urban scenes can be easily extracted as images. Since the images are actual images taken from same view points of humans, and have effectiveness to convey atmosphere of the city (Tanaka, 2002; Snavely, 2006). Such images are expected to contribute to build attractive urban scene representation. So far, such kind of representation can be monotonous, due to the fixed point used. Considering the popularity of the digital still cameras, a more impressive form of urban scenes, which allows the user to view the scenes from connected various points, is required.

Generally, the still images lack spatial continuity and stereoscopic effect. In fact, human recognize urban scenes discretely and can feel three dimension from a image. The still images can be associated with each other, namely the images can be connected in three dimensional space relatively. By compensating the lack based on human spatial cognitive ability, a lively urban visualization is expected. Accordingly, attractive urban visualization can be accomplished by the combination of many still images. In order to implement the urban visualization by using many still images, multiple resolution matching of images taken in urban areas and spatial arrangement of the images based on the matching results become key techniques.

This study proposes a technique of urban visualization consisted of multiple resolution image matching, spatial image arrangement based on the image matching result, and effective image projection for impressive visualization, and then explores the possibility of the new concept of urban representation based on the technique.

2. MULTIPLE RESOLUTION IMAGE MATCHING

2.1 Matching Strategy

In this study, area-based image matching is adopted as basic method. Some difficulties with the matching between images taken in urban area exist. First, as geometric aspects, difference of resolution and deformation has to be mainly considered. In particular, resolution of same objects shot while walking in urban area is subject to change significantly. Second, occlusion caused by moving objects, such as vehicles, human, or so on, often occur. When the area of occlusion is large compared with total area, the matching solution becomes unstable, and then the possibility of correct solution will be lower. Third, poor feature area, which has small intensity variance in template windows, make the matching solution unstable.

To deal with above mentioned difficulties, the following technique is set as a matching strategy. In the matching strategy for the images with different resolution and geometric distortion, stable solution and fast seeking of the solution is necessary. The matching strategy in the realization of the multiple resolution matching is composed of multiple resolution image creation (image pyramids creation) by wavelet analysis, resolution ratio seeking, coarse to fine strategy matching, estimation of templates deformation, voted block matching (Figure 1).

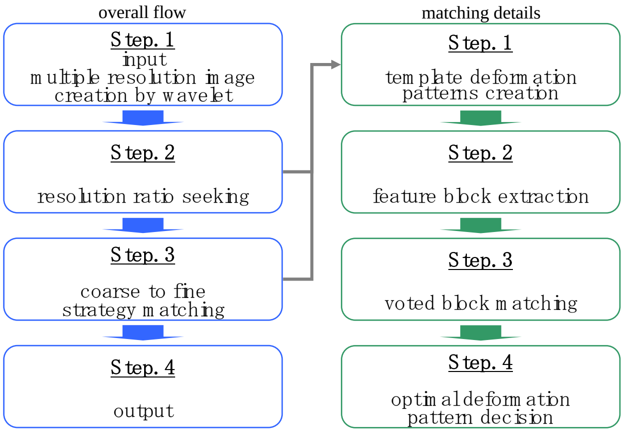


Figure 1. Framework of multiple resolution matching

2.2 Coarse to Fine Approach Based on Multiresolution Analysis

A possible approach of stable and fast image matching with different resolution is coarse to fine approach (Goshtasby *et al.*, 1984). In this approach, corresponding resolution level is given in advance. On the other hand, two images taken in the urban area, the level (resolution ratio) has to be estimated. This study proposes the resolution ratio seeking process, which is included in the coarse to fine strategy.

In the coarse to fine strategy, multiple resolution images (image pyramids) are created. Let T_i and I_i denote master and slave image respectively, and *DS*() down sampling operator. The image pyramids are represented as follows.

$$\begin{aligned} &\{T_0, T_1, \cdots, T_n\}, \quad T_{i+1} = DS(T_i) \\ &\{I_0, I_1, \cdots, I_n\}, \quad I_{i+1} = DS(I_i) \end{aligned} \tag{1}$$

The down sampling method will be explained in the next section.

The T_n and $\{I_n, I_{n-1}, ..., I_{n-k}\}$ are utilized to seek resolution ratio ($k = d$) between master and slave images based on cross correlation maximization. The corresponding levels in the image pyramids are determined by the ratio, and then coarse to fine strategy matching is performed by using the corresponding levels. Namely, the coarse to fine approach is applied to the following image pyramids.

$$\begin{aligned} &\{T_n, T_{n-1}, \cdots, T_d\} \\ &\{I_{n-d}, I_{n-d-1}, \cdots, I_0\} \end{aligned} \tag{2}$$

In the case that sub pixel matching is necessary, up sampling is applied to I_0 by a super resolution technique.

2.3 Image Pyramids Based on Wavelet Analysis

Methods for multiresolution images are called as multiresolution analysis (MRA) (Tanimoto, 1981; Tanimoto and Pavlids, 1975). One of the most popular method is MRA based on discrete wavelet transformation (DWT). In the MRA, signal is orthogonally decomposed into high- and low-frequency component sequentially (Mallat, 1989; Chui, 1992).

To simplify, DWT is applied to one dimensional signal as an example. DWT can be obtained by applying low- and high-pass filter, and then down sampling (Walker, 1999).

$$\begin{aligned} y_{low}[n] &= \sum_{k=-\infty}^{\infty} x[k] \cdot g[2 \cdot n - k] \\ y_{high}[n] &= \sum_{k=-\infty}^{\infty} x[k] \cdot h[2 \cdot n - k] \end{aligned} \tag{3}$$

$$\begin{aligned} y_{low} &= DS(y_{low}[n]) \\ y_{high} &= DS(y_{high}[n]) \end{aligned} \tag{4}$$

where x = one dimensional signal
y_{low} = low-frequency component (approximation coefficients)
y_{high} = high-frequency component (detail coefficients)
g = impulse response of low-pass filter
h = impulse response of high-pass filter

A diagram of filter bank by DWT and frequency domain of each DWT component is shown in Figure 2 and Figure 3, respectively.

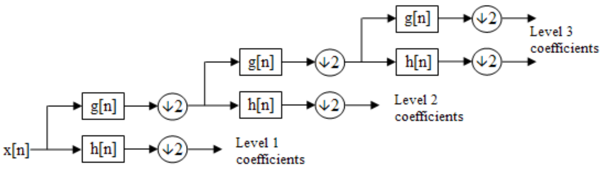


Figure 2. Filter bank by DWT

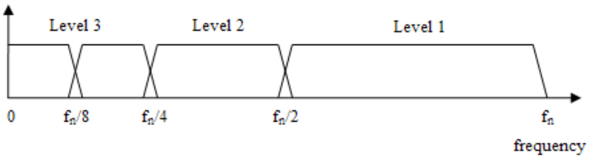


Figure 3. Frequency domain of each component

In the case of two dimensions, signal is decomposed into horizontal, vertical, and diagonal direction, simultaneously. As a result, the following four components are obtained:
O^{LL}: low-frequency component;
O^{LH}: high-frequency component in *y* direction;
O^{HL}: high-frequency component in *x* direction;
O^{HH}: high-frequency component

in 45- and 135-degree direction.

By applying DWT to the low-frequency component repeatedly, any resolution level is obtained (Figure 4). Figure 5 shows an example of DWT application.

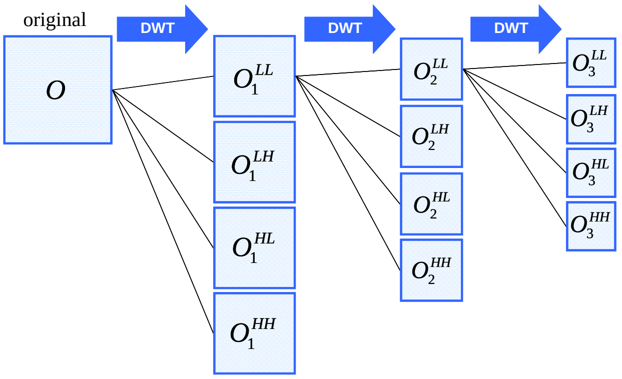


Figure 4. Two dimensional wavelet analysis

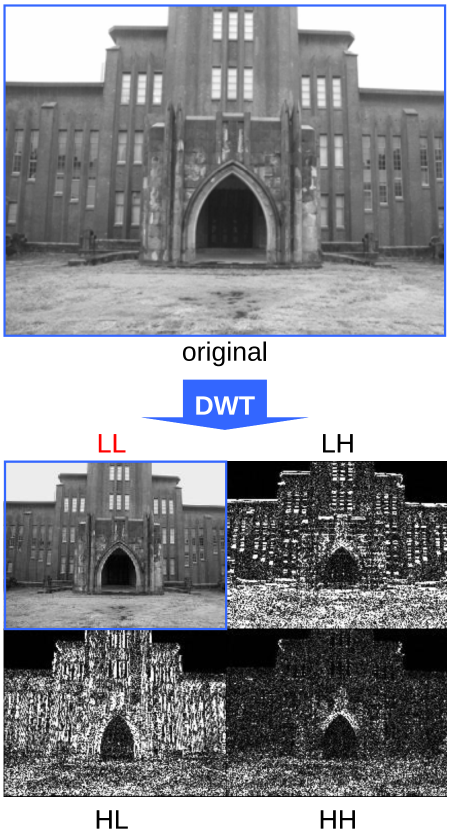


Figure 5. Example of DWT application

2.4 Geometric Transformation Patterns of Template

To deal with deformation, geometric transformation has to be introduced. Additionally, when the above mentioned MRA is applied, the resolution varies in proportion to the power of 2 times. Accordingly, small difference of resolution still remains. By applying projective transformation as an approximate transformation, the deformation and the small resolution difference is solved (Szeliski, 1996).

$$x' = \frac{a_1 x + a_2 y + a_3}{a_7 x + a_8 y + 1} \qquad (5)$$

$$y' = \frac{a_4 x + a_5 y + a_6}{a_7 x + a_8 y + 1}$$

where x' y' = transformed coordinates
$[a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8]$ = parameters

The master images are set as template, which are deformed compared with slave images. When the parameters of transformation are estimated by the least-squares method normally, the calculation requires much time. To reduce the calculation time and avoid instability depending on initial value, various deformed template patterns (parameters sets) are provided in advance (Figure 6), and then the matching is processed in each deformed template patterns.

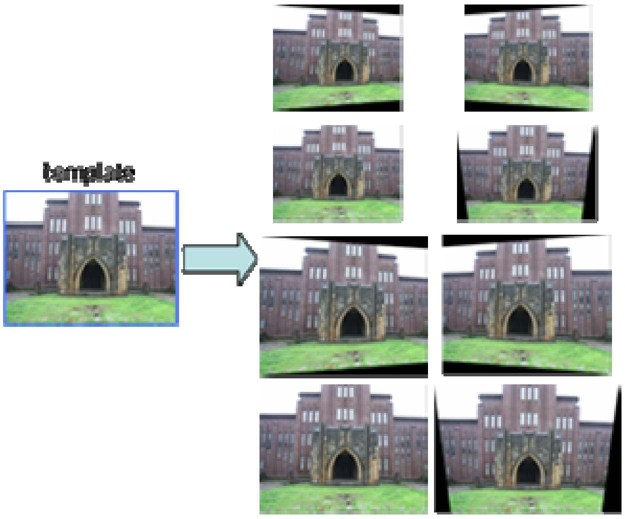


Figure 6. Example of deformed template patterns

Among the patterns, the most appropriate pattern will be selected by maximizing correlation between template and slave images after the following voted block matching. The matching method can acquire stable solution, when the images have only small deformation. According to the characteristic of the matching method, the geometric transformation deals with only large deformation here.

2.5 Voted Block Matching

In the matching, a block matching technique is adapted, since images taken in urban areas have often occlusion. The block matching technique is knows as a robust technique against the occlusion (Saitoh, 2001). Additionally, the block matching technique can deal with small deformation of images. Here, voted block matching (VBM) is utilized as the block matching technique.

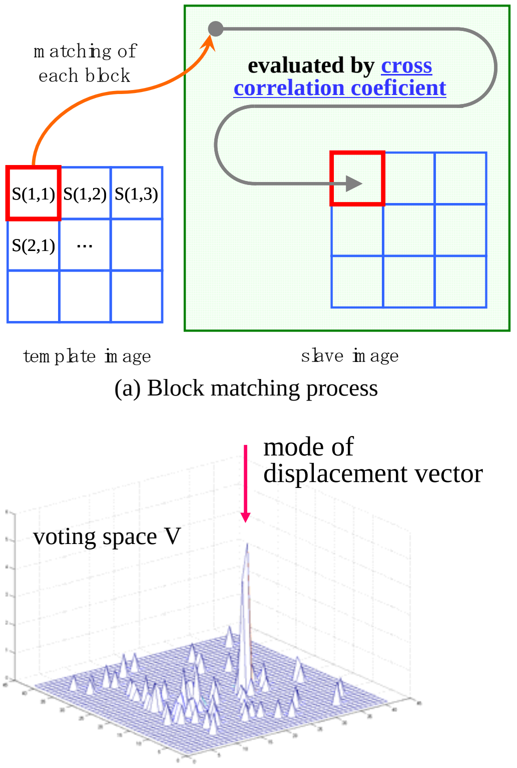
In the block matching, the original template is divided into multiple blocks, and then the displacement vector of each block is searched by maximizing cross correlation coefficient. The displacement vectors of all blocks are voted, and mode of the displacement vectors is calculated. The vote is based on Hough transform. The voting process makes matching results stable and robust. When the displacement vector with maximum cross correlation coefficient is ($a_{\max}$, $b_{\max}$), voting process to the voting space V in the following.

$$V\left(a_{\max}(u,v), b_{\max}(u,v)\right) \leftarrow V\left(a_{\max}(u,v), b_{\max}(u,v)\right) + 1 \quad (6)$$

The final matching ratio is equal to cumulative number of votes divided by the number of blocks (*MN*).

$$C_V(a,b) = \frac{V(a,b)}{MN} \quad (7)$$

Figure 7 depicted the concept of voted block matching.



(a) Block matching process

(b) Voting process

Figure 7. Concept of voted block matching

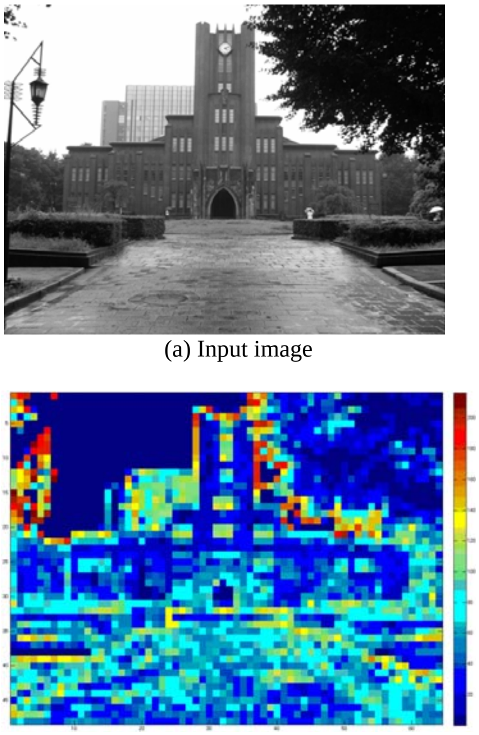
10 types of images, which had different occlusion area ratio, were prepared to validate performances of matching techniques. Compared with accumulated block matching (ABM), normalized correlation matching (NCM), incremental sign correlation (ISC), VBM provided better performance. The maximum occlusion area ratio was 87.5 % (Saitoh, 2001).

Because some blocks do not have enough texture (e.g. sky area), it is desirable to exclude such blocks, called as non-feature blocks, from the voting. The non-feature blocks are specified by variance histogram analysis and cut out. Figure 8 shows an example of variance calculation in matching blocks, whose size is 10 x 10 pixels. Blocks in sky area are recognized as small variance blocks.

3. IMAGE ARRANGEMENT

3.1 Relative and Successive Orientation

Using the matching results, relative and successive orientation is applied, and then images are arranged in a three dimensional space.



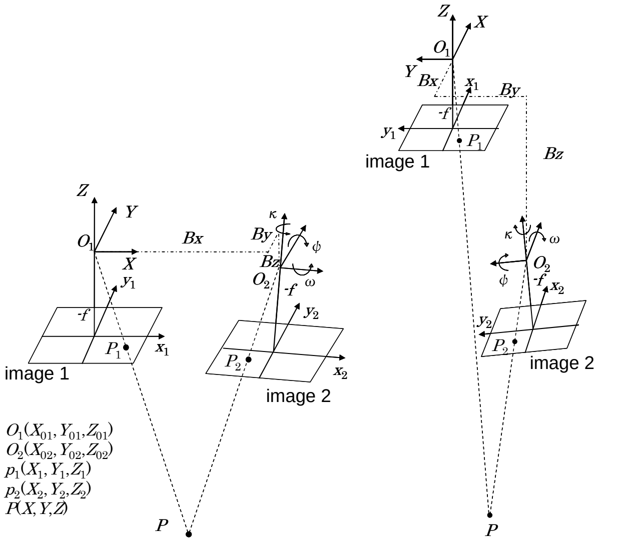
(a) Input image

(b) Variance of matching blocks

Figure 8. Variance calculation in matching blocks

In case of urban visualization, images are often taken along with street. Accordingly, the base length or parallax in relative orientation is not enough large (Figure 9), and so the solution is unstable. Normally, B_x is set as 1, namely the coplanarity condition is derived as follows.

$$\begin{vmatrix} 1 & b_y & b_z \\ P_1 & Q_1 & R_1 \\ P_2 & Q_2 & R_2 \end{vmatrix} = 0 \quad (8)$$



(a) normal case (b) in case of urban visualization

Figure 9. Camera configuration

where $b_y = B_y/B_x$, $b_z = B_z/B_x$
$P_1=X_1-X_{01}$, $Q_1=Y_1-Y_{01}$, $R_1=Z_1-Z_{01}$
$P_2=X_2-X_{02}$, $Q_2=Y_2-Y_{02}$, $R_2=Z_2-Z_{02}$

To keep the stability of the solution, the following coplanarity condition is adopted in this study.

$$F\left(b_x, b_y, \omega, \varphi, \kappa\right) = \begin{vmatrix} b_x & b_y & 1 \\ P_1 & Q_1 & R_1 \\ P_2 & Q_2 & R_2 \end{vmatrix} = 0 \tag{9}$$

where $b_x = B_x/B_z$, $b_y = B_y/B_z$
A coplanarity equation can be obtained in each block, and the initial value of the solution is set by using geometric transformation parameters (Mikhail *et al*., 2001).

$$\begin{aligned}
\omega_0 &= \tan^{-1}\left(f \cdot a_8 \right) \\
\varphi_0 &= \tan^{-1}\left(- f \cdot a_7 \cdot \cos\omega_0 \right) \\
\kappa_0 &= \tan^{-1}\left(-a_4/a_1 \right) \quad if\ \varphi = 0 \\
\kappa_0 &= \tan^{-1}\left(-a_2/a_5 \right) \quad if\ \varphi \neq 0, \omega = 0 \\
\kappa_0 &= \tan^{-1}\left\{ -\left(A_1A_3 - A_2A_4 \right)\left(A_1A_2 + A_3A_4 \right) \right\} \\
&\qquad\qquad if\ \varphi \neq 0, \omega \neq 0 \\
Z_0 &= f\cos\omega_0 \sqrt{\left(A_2^2 + A_3^2 \right) / \left(A_1^2 + A_4^2 \right)} + Z_m \\
X_0 &= a_3 - \left(\tan\omega_0 \sin\kappa_0 / \cos\varphi_0 - \tan\varphi_0 \cos\kappa_0 \right)\left(Z_m - Z_0 \right) \\
Y_0 &= a_6 - \left(\tan\omega_0 \cos\kappa_0 / \cos\varphi_0 + \tan\varphi_0 \sin\kappa_0 \right)\left(Z_m - Z_0 \right)
\end{aligned} \tag{10}$$

where Z_m = average depth
$A_1 = 1 + \tan^2\varphi_0$
$A_2 = a_1 + a_2 \tan\varphi_0 / \sin\omega_0$
$A_3 = a_4 + a_5 \tan\varphi_0 / \sin\omega_0$
$A_4 = \tan\varphi_0 / \left(\cos\varphi_0 \tan\omega_0 \right)$

Through successive orientation after relative orientation, spatial image arrangement is conducted.

3.2 Error Adjustment by Analogy with Traverse Survey

Since height of camera (*x*-coordinate in this study) is almost uniform, position of horizontal plan (*y* and *z* coordinates) should be more important.

To improve the accuracy of the image positioning (*y* and *z* coordinates), an adjustment method in analogy with traverse survey is applied. In particular, when route of image shot is closure, accumulated residuals (closure difference) only in *y* and *z* coordinates is adjusted (Figure 10).

4. EXPERIMENT

The proposed technique was applied to images taken in urban area. The images were taken around a building, and formed a closed network. The number of images 66 (600 x 800 pixels).

Haar wavelet, which performs as smoothing, is utilized to create image pyramid. The mother wavelet is following.

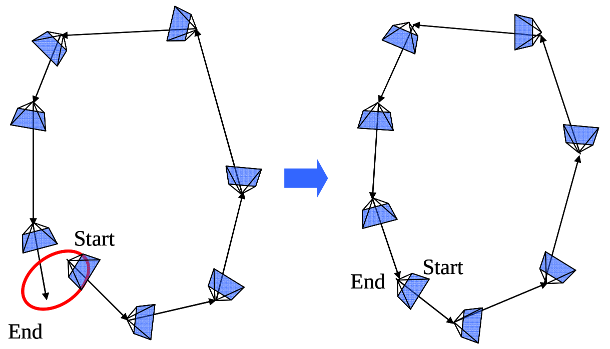


Figure 10. Improved positioning in analogy with traverse survey

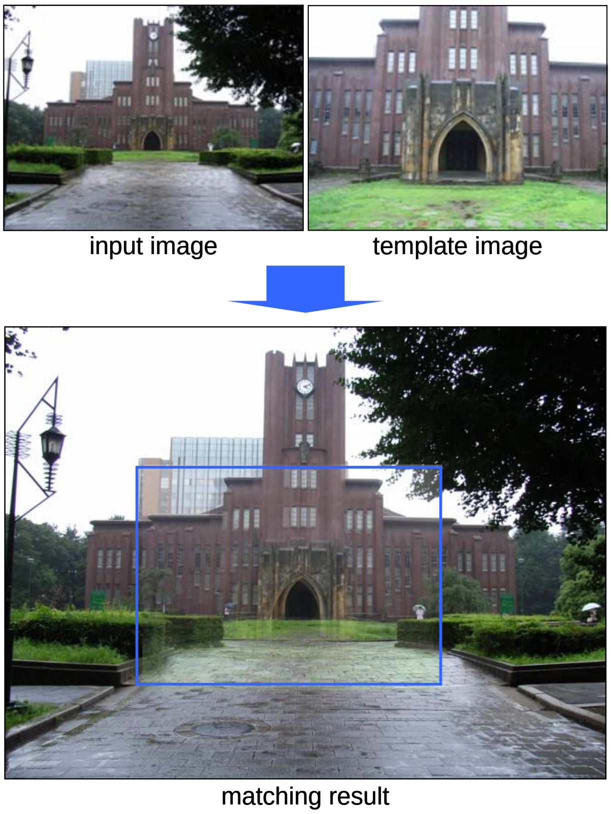


Figure 11. Experimental result of image matching

$$\psi\left(x\right) = \begin{cases} 1 & 0 < x < 1 \\ -1 & -1 < x < 0 \\ 0 & others \end{cases} \tag{11}$$

In the voted block matching, block size and threshold to specify non-feature blocks is needed. Through trial and error, results were not depending on the block size very much. The threshold was determined as 30 (variance) by mode method of histogram analysis.

The matching result was enough to visualize the urban scene compared with manual matching. The success rate was 97 %. Figure 11 shows an experimental result of multiresolution matching. However, when the camera pose was suddenly changed (in case of turning points), the matching sometimes failed. This implies that the images should be taken in more high-density.

Figure 12 shows an experimental result of image arrangement. In the part that the matching failed, the trajectories of camera position looks unnatural. Additionally, final matching results was determined by using all blocks, which contain outlier blocks. By eliminate such outlier blocks, the accuracy will be expected improved.

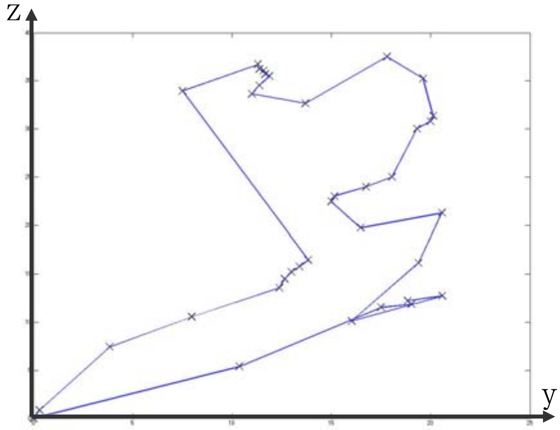


Figure 12. Experimental result of image arrangement

5. CONCLUSIONS

This study proposes a technique to deal with some difficulties inherent in urban visualization. The technique consists of multiresolution image matching and spatial image arrangement. The multiresolution image matching method is combination of coarse to fine strategy, wavelet-based image pyramid creation, deformed template preparation, and voted block matching. The spatial image arrangement method introduces stabilization of coplanarity condition equation, and error adjustment of by analogy with traverse survey. The significance of the technique and limit of the technique is confirmed. Through the application, possibility of the new concept of urban representation is indicated.

As a further work, one of the most important issue is investigation of effective presentation method. Transition speeds, image blending ratios, distances between neighbouring images are important points for the effective visualization. Those effects will be verified via psychological tests. In this study, a filter for MRA is simple Haar wavelet. As filters for MRA, not only linear filters also nonlinear filters can be applied. For example, median filter can keep edge information, the characteristic may have influence on matching improvement. Furthermore, combination of feature-based matching will be considered.

REFERENCES

Chui, C.K., 1992. *An Introduction to Wavelets*. Academic Press, Boston.

Debevec, P.E., Taylor, C.J., Malik, J., 1996. Modeling and rendering architecture from photographs: A hybrid geometry and image-based approach. *SIGGRAPH Conf. Proc.*, pp.11-20.

Goshtasby, A., Gage, S.H., and Bartholic, J.F., 1984. A two-stage cross correlation approach to template matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(3), pp.374-378.

Grzeszczuk, R., 2002. Course 44: Image-based modelling. *SIGGRAPH 2002*.

Mallat, S.G., 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), pp.674-693.

Mikhail, E.M., Bethel, J.S., McGlone, J.C., 2001. *Introduction to Modern Photogrammetry*. John Wiley & Sons, Inc, New York.

Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Kock, R., 2004. Visual modeling with a hand held camera. *International Journal of Computer Visiion*, 59(3), pp.207-232.

Saitoh, F., 2001. Robust image matching for occlusion using vote by block matching. *IEICE Transactions on Information and Systems*, J84-D-II(10), pp.2270-2279 (in Japaneses).

Snavely, N., Seitz, S.M., Szeliski, R., 2006. Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics*, 25(3), pp.835-846.

Szeliski R., 1996. Video mosaics for virtual environments. *IEEE Compter Graphics and Applications*, 16(2), pp.22-30.

Tanaka, H., Arikawa, M., Shibasaki, R., 2002. A 3-D photo collage system for spatial navigations. *Digital Cities II, Computational and Sociological Approaches*. Springer, New York, pp.305-316.

Tanimoto, S., 1981. Template matching and pyramids. *Computer Graphics and Image Processing*, 16, pp.356-369.

Tanimoto, S. and Pavlids, T., 1975. A hierachical data structure for picture processing, *Computer Graphics and Image Processing*, 4, pp.104-119.

Walker, J.S., 1999. *A Primer on Wavelets and Their Scientific Applications*. Chapman & Hall, London.