Highly Dense 3D Surface Generation Using Multi-image Matching

Myoung-Jong Noh, Woosug Cho, and Ki In Bang

This study presents an automatic matching method for generating a dense, accurate, and discontinuity-preserved digital surface model (DSM) using multiple images acquired by an aerial digital frame camera. The proposed method consists of two main procedures: area-based multi-image matching (AMIM) and stereo-pair epipolar line matching (SELM). AMIM evaluates the sum of the normalized cross correlation of corresponding image points from multiple images to determine the optimal height of an object point. A novel method is introduced for determining the search height range and incremental height, which are necessary for the vertical line locus used in the AMIM. This procedure also includes the means to select the best reference and target images for each strip so that multi-image matching can resolve the common problem over occlusion areas. The SELM extracts densely positioned distinct points along epipolar lines from the multiple images and generates a discontinuity-preserved DSM using geometric and radiometric constraints. The matched points derived by the AMIM are used as anchor points between overlapped images to find conjugate distinct points using epipolar geometry. The performance of the proposed method was evaluated for several different test areas, including urban areas.

Keywords: Matching, multiple images, epipolar, vertical line locus, occlusion area.

Manuscript received Jan. 15, 2011; revised Oct. 6, 2011; accepted Nov. 25, 2011.

This research was supported by a grant from Cutting-Edge Urban Development-Korean Land Spatialization Research Project funded by Ministry of Land, Transport and Maritime Affairs (06KLSGB01).

Myoung-Jong Noh (phone: +82 32 860 7571, nmj@inhaian.net), Woosug Cho (corresponding author, wcho@inha.ac.kr), and Ki In Bang (biske@naver.com) are with the Department of Civil Engineering, Inha University, Incheon, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.12.1611.0011

I. Introduction

A wide variety of matching methods for automatic 3D surface generation has been developed and presented in the literature, but most of these have not provided satisfactory results, especially for complex urban areas. The problematic areas in urban aerial images are discontinuous regions, hidden and occluded areas, homogeneous areas, and repetitive pattern areas, which result in very difficult and ambiguous matching. As recent airborne digital images, however, have higher radiometric resolution and higher overlap, the problems faced by automatic matching procedures could be reduced or overcome with the development of a novel matching approach. The primary goal of the matching method proposed in this paper is to generate an accurate and dense regular-grid 3D surface model, especially for urban areas, using aerial multi-images. Before presenting our matching approach, we review some of the work published in this field, especially that related to multi-image matching.

Conventionally, to generate a discontinuity-preserved 3D surface model, distinct points or edges are first extracted and matched to determine the 3D position. Reference [1] performs epipolar line matching to find conjugating edge points in a stereo model based on feature-based and area-based approaches, and dynamic programming (DP). Since this method computes a simple similarity measure with left and right pixel information of each edge point on an epipolar line without geometric constraints, the result of matching can strongly be affected by local intensity differences. This method also has disadvantages of DP, such as a streaking effect in the epipolar direction and a smearing effect [2]. For multi-image matching, [3] uses distinct points and matched these using the weighted sum of squared-differences (SSDs). Even though this

method tries to avoid falsely matched points in occlusion areas using the SSD profile, it is not certain that this approach can rigorously detect the occlusion areas. Instead of SSDs, [4] uses the intersection of epipolar lines from multiple images to find the conjugate points among interest points. Since this method does not define a search range along the epipolar lines, it can produce false conjugate points. Similar to the epipolar line intersection, [5] finds a voxel where 3D rays from multiple images intersect in 3D space and uses the space-sweep approach to match edge points. The matching quality of this method can be strongly affected by the size of a voxel in terms of accuracy of 3D points. References [6], [7] suggest a feature-based matching method using linear features. However, this approach does not suggest how to deal with the linear features in occlusion areas to avoid false matching. References [8], [9] develop a method to match corresponding points along the edges using epipolar geometry and least median of squares for rays. Instead of matching only points or edges, [10] proposes a multiple primitive multi-image matching method that matches interest points, edges, and grid points using multiple images. The sum of normalized cross correlation (SNCC) using geometrically constrained cross correlation is applied for measuring the similarity of feature points and grid points. For application of the SNCC, the search height range for a selected point is determined by an approximate digital surface model (DSM), and each normalized cross correlation (NCC) is computed stepwise within the search height range. Similar to [10], [11] classifies matching entities as closed edges, straight edges, curved edges, edge points, and grid points. The multi-patch size matching using NCC and least squares matching are applied. References [10], [11] do not take into consideration the occluded areas in measuring the similarity, and hence they can cause falsely matched points, especially in urban areas. For implementing these methods, pre-processing to extract interest points or edges should be performed, and thus the computational load is increased.

In the following studies, attempts are made to match all image pixels in order to generate a DSM with maximally dense image resolution without pre-processing to extract edges. Reference [12] presents a hierarchical correlation-based matching technique for generating a disparity map using the epipolar constraints. This technique uses a support function and relaxation algorithm to resolve ambiguities. However, this technique does not take into consideration the occluded areas, and the search height range along the epipolar line and matching candidate pixels may therefore be detected incorrectly. For detecting and removing occluded points along epipolar lines, [13] proposes semiglobal matching and mutual information using a joint probability distribution and smoothed cost for generating a disparity map. Since constraints of the smoothed cost are empirically determined by the image information, the resulting disparity map can strongly be affected by local intensity differences. For a video sequence, [14] presents methods for recovering a consistent depth map using a bundle optimization framework to improve the depth estimation around the occlusion boundaries. The depth maps generated in [12]-[14] perform well for areas with strong texture but have difficulties with weakly textured areas [15]. Based on the vertical line locus [16], which is able to partially overcome repetitive pattern problems, [17] proposes a multi-image correlation with radiometric attenuation to measure the similarity without taking occlusion into consideration. Reference [18] proposes an algorithm for multi-view stereopsis that outputs a dense set of small rectangular patches covering the surfaces visible in the images. Because of the lack of regularization, the patch generation step reconstructs 3D points only where there is reliable texture information, and post-processing is necessary to fill in possible holes and obtain a complete mesh model.

As discussed in previous research, dense and accurate DSM generation typically involves three common problems. The first issue is how to constrain the search range of an object point in order to accelerate the matching process and also reduce the number of ill-matched points. The second issue is how to exactly match discontinuous regions such as building boundaries and homogeneous areas such as ground. The occlusion area is the third problem, especially when applying area-based multi-image matching (AMIM) over a complex urban area. Since multiple images have different views, they have different occluded and visible areas. Without a reliable method to determine occlusion areas, final products can be seriously deteriorated.

II. Overview of Matching Approach

The matching approach proposed in this paper takes advantage of both area-based and feature-based matching. The area-based matching plays a key role in the generation of a dense 3D surface model, while feature-based matching copes with the inadequate performance in area-based matching over problem areas such as homogeneous and discontinuous areas. The proposed approach also takes occlusions into account through an intermediate DSM. The salient feature of the proposed approach is in the use of object-space-guided multi-image matching. The matching approach consists of successive and iterative steps, where each step is appropriate for dense and accurate 3D surface generation. The overall procedure of the approach is shown in Fig. 1. An image pyramid is first prepared using a Gaussian filter with a sub-sampling factor of 2. It is well known that a coarse-to-fine approach drastically

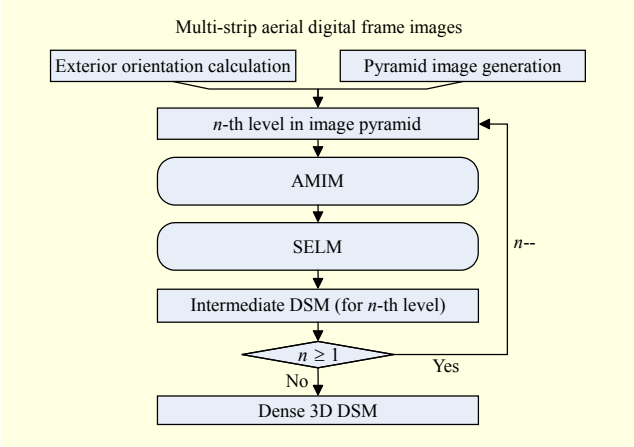


Fig. 1. Overall procedure of proposed approach.

reduces the search space as well as the number of ill-matched points. Also, the image pyramid solves the problem of determining the approximate heights in the region, starting with a horizontal plane as an approximation of the terrain [4]. In the proposed approach, the intermediate DSM generated from the lower level of the image pyramid is used for matching upper-level images. More precisely, the intermediate DSM constrains the search space of object points to match. Thus, we are not concerned with the case where matched points are missing through a consecutive image pyramidal level.

The first main step in the approach is AMIM based on the vertical line locus. This process creates a regular-grid DSM on every level of the image pyramid through an analysis of the correlation profile constructed from multi-images. Since the AMIM entails area-based matching, it runs on the assumption that the terrain is a smooth surface, and thus may not perform well over breaklines such as building boundaries. To improve the performance over discontinuities as much as possible, we incorporate the concept of an adaptive window into the AMIM in computing correlation coefficients [19]. In addition, the intermediate DSM allows detection of the occluded areas among multi-images in the upper level of the image pyramid, which ultimately improves the quality of the final DSM.

The second step aims to successfully cope with the limitations of the AMIM in matching problematic areas such as homogeneous areas and building boundaries. Stereo-pair epipolar line matching (SELM), a type of feature-based matching, attempts to match distinct points along epipolar lines using geometric and radiometric constraints and an intermediate DSM. In the SELM, the matched points from the AMIM, which lie on epipolar lines, are selected and used as anchor points. Only those distinct points detected between anchor points on epipolar lines are rigorously matched through a predefined cost function. In doing so, the search space for distinct points is reduced drastically; therefore, the number of false matches and computation time decrease. The heights of the remaining unmatched pixels between the anchor point and matched distinct points are then linearly interpolated if the condition meets the pre-defined requirements. The major role of the SELM is to improve the DSM from the AMIM, which involves preserving breakline information and filling heights on homogeneous areas.

This hierarchical process is repeated at every level of the image pyramid to refine the intermediate DSM. Finally, the final DSM is generated from the highest resolution image. Thus, the grid size of the final DSM corresponds to the ground pixel size of the original image. The following sections describe the detailed procedure of the proposed approach.

III. Area-Based Multi-image Matching

The goal of AMIM is to determine an optimal height for every single grid cell for a raster DSM at each level of the image pyramid. The proposed AMIM is an area-based multi-image matching method with a vertical line locus. Initially, the concept of the vertical line locus was introduced to automatically generate a DSM using a stereo image pairing. It is well known that the vertical line locus is a strong geometric constraint reducing the search space to a large extent. However, it is not appropriate to use a pre-defined and fixed search range for each vertical line locus. In this paper, we propose a novel approach to constrain the search range adaptively by using an intermediate DSM generated at each level of the image pyramid.

As shown in Fig. 2, the object space is defined as regularly spaced grids. On each grid, approximate heights within the search range are assigned along the vertical line within the object space, and the plausible points are then projected back to multiple images. For every assigned height, the similarity

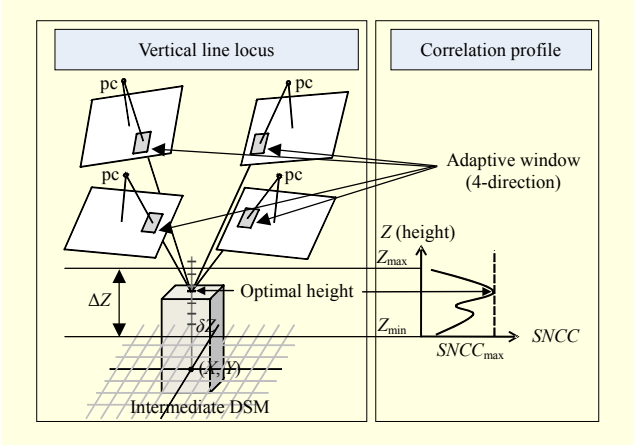


Fig. 2. Proposed method estimating optimal height using vertical line locus and SNCC (adapted from [17]).

measure is computed with multiple images. The height yielding the highest similarity is then determined as the optimal height for that grid point. In contrast to the case of stereo image matching, multi-image matching uses the SNCC as a similarity measure. However, the difference between our method and other methods is that we take occlusion into account in computing the SNCC, which will be explained in subsection 2. The cross correlation coefficient is quite sensitive to the size of the window, and the AMIM uses a 4-directional adaptive window in the process to reflect the context of the image.

1. Height Search Range

In order to reduce computational complexity and erroneous image matching, the height search range (ΔZ in Fig. 2) for each DSM cell needs to be determined adaptively along the vertical line in an object space.

The proposed method defines the search range by using an intermediate DSM, or, more precisely, the triangulated irregular network (TIN) model generated from the matched points in the lower level of the image pyramid. As the matching process moves forward through the image pyramid, the intermediate DSM becomes refined, and thus the height search range is also refined.

In Fig. 3, small dots denote matched cells in the lower-level DSM generated by the AMIM. T_U represents grid cells in the upper-level DSM to be constrained by the search range. T_L is a corresponding grid cell in the lower level, which is not a matched cell. To determine the height search range for T_U, the triangular patch containing T_L must first be selected. All triangular patches (N_1 to N_{11}) sharing any one of the 3 vertices (H_1, H_2, H_3) of the triangular patch containing T_L are then searched and selected. The height search range [Z_{min}, Z_{max}] is

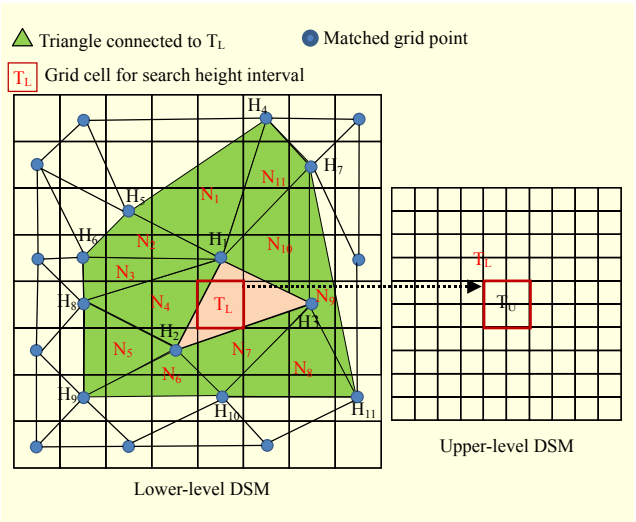


Fig. 3. Determination of height search interval by TIN model.

then determined by the lowest and highest heights that are drawn from all triangular patches involved. Where T_L is a matched cell such as H_1, all triangular patches sharing H_1 as a vertex are searched and selected. The remainder of the process is the same as that explained for T_L. The probability that the search height range can include the correct height for T_U increases using this method. Furthermore, false-matching errors will decrease, and this will also decrease the computational complexity. It should be noted that, in the lowest level of the image pyramid, the search height range is set from a priori knowledge of the terrain.

In addition, the increment in height (δZ) within the search range also plays a role in improving the quality of the DSM as well as the performance of the matching process. In practice, δZ represents the height resolution for an SNCC profile within the search range (see δZ in Fig. 2). We develop a simple and effective method that makes use of ground pixel resolution (or expected height accuracy) for determining δZ adaptively through the image pyramid. The expected height accuracy σ_Z for a stereo model is derived by the law of error propagation as

$$\sigma_Z = \frac{\sqrt{2}\sigma}{(B/H)f} H, \tag{1}$$

where σ denotes the standard errors of an image measurement, B is the length of the air base, H is the average flying height above the ground level, and f is the principal distance of the camera.

The relief displacement Δr corresponding to the height accuracy σ_Z in the stereo model can be determined by

$$\Delta r = r\frac{\sigma_Z}{H} = r\frac{1}{B/H}\frac{\sqrt{2}\sigma}{f}, \tag{2}$$

where r refers to the distance from the principal point to an image point where relief displacement occurs.

For example, if aerial images having 13,824 pixels by 7,680 pixels are captured with 80% endlap and 60% sidelap, the maximum Δr is about 1.8σ in the case of a 120-mm principal distance and 1,000-m flying height. σ can be used as the standard error of the image matching point. If σ is about 1.0 pixel for the AMIM, Δr_{max} corresponding to σ_Z is about 2.0 pixels. If half of σ_Z is used as δZ, Δr_{max} corresponding to σ_Z is about 1.0 pixel, and thus all pixels corresponding to relief displacement in an image can be computed for SNCC. In this way, the image matching process can achieve a height accuracy of at least one pixel. Therefore, based on the error propagation, δZ using multiple models can be used by

$$\delta Z = \frac{1}{2\sqrt{m}}\left(\sum_{i=1}^{m}\sigma_{Z_i}\right), \tag{3}$$

where m refers to the number of models.

2. Selection of Reference Image

The method of selecting reference images proposed in the paper differs significantly from that of previous research [3], [10], [17]. As the AMIM proceeds, the reference image can be changed for each grid cell of the DSM. First, the planimetric distances between each grid cell and nadir positions of the multiple images involved are computed. The image that has the shortest distance among them is then selected as the reference image for the grid cell to be matched. In this way, it is assured that we will have a greater likelihood of selecting the image directly down-viewing the grid cell, which eventually improves the performance of multi-image matching with the vertical line locus. It is obvious that this approach involves significant computing time. However, it is implemented for better performance in urban areas.

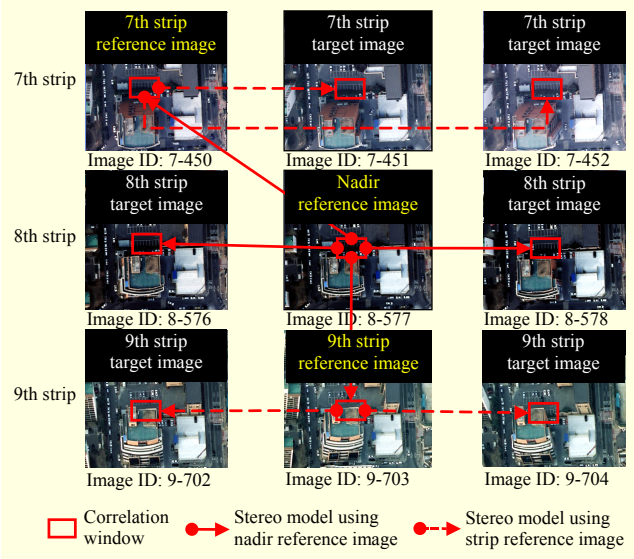


Fig. 4. Selection of nadir and strip reference images for computing SNCC (in image ID, first number refers to strip ID and second number is photo ID).

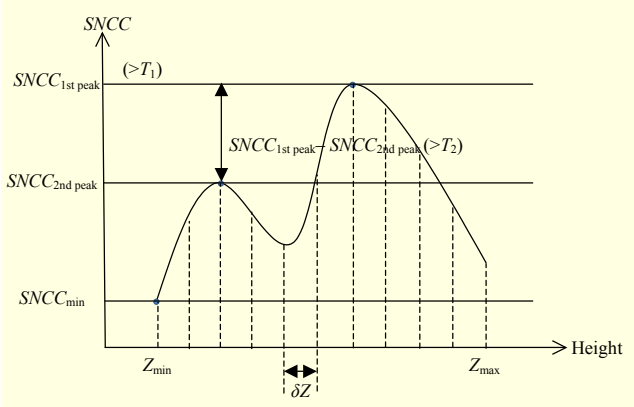


Fig. 5. Two decision measures for determining optimal height.

When the reference image, called the nadir reference image, is chosen for each grid cell, the similarity measures are computed for all overlapped images on which the grid cell of interest is shown. One image of each across-strip having the highest similarity measure is then selected as the strip reference image. In contrast to previous studies, we choose one nadir reference image and one strip reference image for each across-strip in computing the SNCCs for each grid along the vertical line. As shown in Fig. 4, the image ID 8-577 is selected as the nadir reference image and the images ID 7-450 and ID 9-703 are chosen as the strip reference images for the 7th and 8th strip, respectively.

In previous research [3], [10], [17], one reference image is selected and the NCCs are computed with each target image involved, while the proposed approach chooses both one reference image and one strip reference image for each strip, and the NCCs are computed with each target image strip by strip. It should be noted that the nadir reference image is also a strip reference image. In addition, the proposed approach takes the occlusion into consideration in computing the NCCs. This means that target images that are occluded do not contribute to the computation of the NCCs. The intermediate DSM determines whether or not the target images are occluded. The previous approach and our approach are implemented, and it is found that our approach outperforms the previous approach in computing the NCCs as well as the SNCCs.

3. Optimal Height Determination

The AMIM determines the optimal height by evaluating the SNCC profile. According to [19], the size of a window should be adaptively changed due to local variations of gray values. Thus, the 4-directional adaptive window is applied in computing the similarity measure among multiple images. In addition, the adaptive window has an advantage in improving the area-based image matching quality around man-made objects.

The SNCC profile is constructed for determining the optimal height of each grid cell of the DSM. Each SNCC for every height increment δZ within the height search range $[Z_{\min}, Z_{\max}]$ is calculated by

$$SNCC = \frac{1}{n}\sum_{i=1}^{n} NCC, \tag{4}$$

where n refers to the number of NCCs computed.

The proposed approach uses two decision rules that determine the optimal height for each grid (see Fig. 5).

- Decision rule 1: the first highest peak ($SNCC_{\text{1st peak}}$) > T_1,
- Decision rule 2: the difference between the first and second highest peaks ($SNCC_{\text{1st peak}} - SNCC_{\text{2nd peak}}$) > T_2,

where $SNCC_{1st\ peak}$ and $SNCC_{2nd\ peak}$ are the first and second highest peaks, respectively, and $SNCC_{min}$ is the minimum value in the profile. T_1 is a threshold value for the first highest peak, and T_2 is a threshold value computed by

$$T_2 = \frac{SNCC_{1st\ peak} - SNCC_{min}}{K}, \quad (5)$$

where K is an empirically determined constant.

In order to determine the optimal height, the two decision rules should be satisfied simultaneously. If there is one peak in the profile, only the first rule is evaluated in the process. In most previous research using the SNCC, the first rule is applied as presented above; however, the second rule is applied differently. To implement the second rule, in the previous research, a simple and fixed threshold value is used, such as 0.3 for T_2, while the proposed approach uses a method in which T_2 is computed adaptively reflecting the SNCC profile.

IV. Stereo-Pair Epipolar Line Matching

The concept of the AMIM involves determining the correct height by evaluating the SNCC profile and enforcing the search range along the vertical line locus. Although multiple images are used and the search space is reduced, the AMIM suffers from similar problems, to other area-based image matching techniques such as discontinuous and homogeneous areas. Thus, the SELM is proposed to cope with the problems that the AMIM cannot completely avoid, especially in urban areas.

The SELM extracts distinct points along epipolar lines and runs on each stereo pair configured from multiple images involved. The distinct points are defined as inflection points of image gray-level profiles, which is similar to the approach of a gradient operator detecting edges. The matched image points from the AMIM are used as anchor points that define the search range along the epipolar lines of each stereo pair. Taking such matched image points as anchor points can reduce the search space, thereby reducing the computation load and increasing the possibility of correct matches for the distinct points extracted. In measuring the similarity between distinct points in a stereo pair, a new cost function is proposed using geometric and radiometric constraints. After matching distinct points, unmatched image points between matched distinct points can be linearly interpolated for heights.

1. Search Range for Distinct Points along Epipolar Line Pair

The SELM selects a nadir image as the reference image with respect to the target area of interest since there is less chance of an occluded area in the reference image. The epipolar geometry between the nadir image and all of the images overlapped with the nadir image is established by the coplanarity condition. The anchor points that lie on the epipolar line pair are searched and selected. Due to occlusion, those anchor points that do not belong to the epipolar line pair are excluded from each stereo pair. With a radiometric transformation such as the min-max linear stretch, the gray-value difference between the two epipolar lines of the stereo pair is balanced for the similarity measure. The distinct points extracted along the epipolar line pair can represent the parts of discontinuities, which will be regarded as the matching candidate points for the SELM. Since the distinct points are the inflection points on a gray-level profile, the characteristics of inflections such as peaks and valleys will be used as a cue for matching.

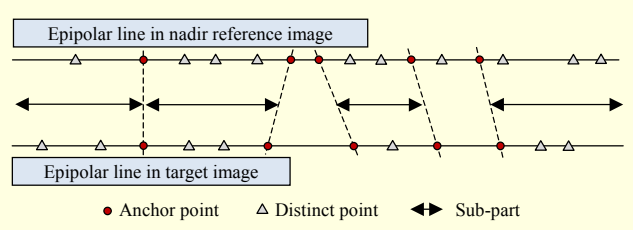
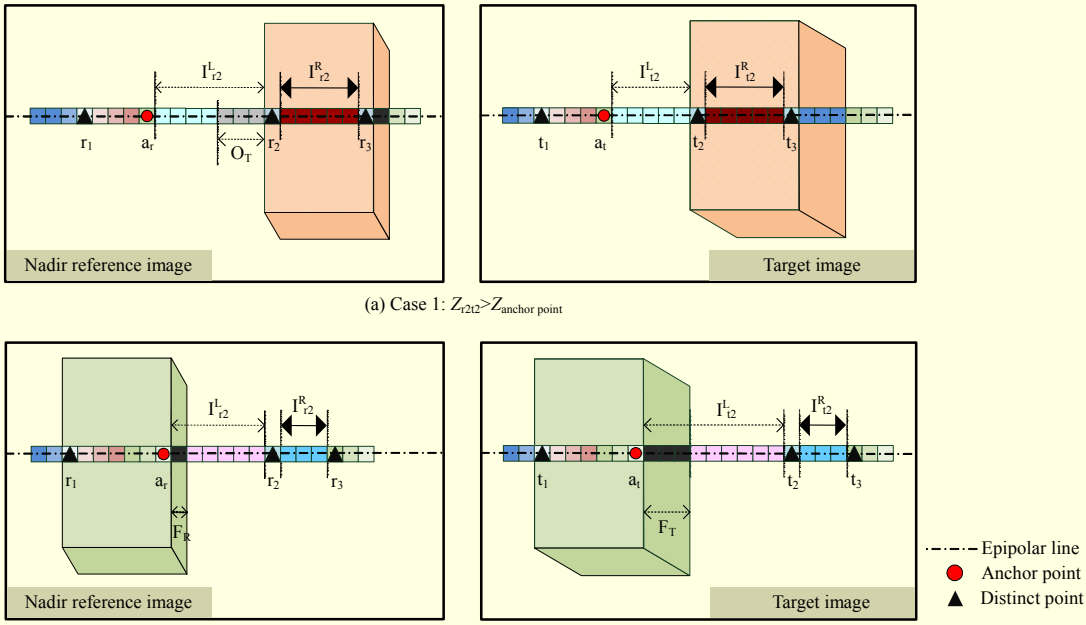


Fig. 6. Anchor and distinct points along epipolar line pair.

Once the anchor points and distinct points are arranged on the epipolar line pair, the SELM divides the epipolar lines into sub-parts using anchor points, as shown in Fig. 6. Each sub-part subtended by two anchor points defines the search range where the SELM searches and matches rigorously distinct points. The novel approach proposed in this paper can avoid the ambiguity problem, especially in epipolar line matching.

2. Geometric Constraint

In order to match the distinct points along an epipolar line pair, the gray-value distribution on both the left and right sides of the distinct points are evaluated for a similarity measure in a stereo pair [1]. In the case where distinct points lie on smooth terrain, both sides of the distinct points can be evaluated for the similarity measure. However, if distinct points lie on breaklines such as building boundaries, either the left or right side of these points may not represent a similar gray-value distribution in an epipolar line pair due to occlusion or foreshortening. For this reason, the SELM incorporates a geometric constraint in order to cope with problematic areas such as occluded and foreshortened areas. More precisely, the SELM makes a decision about which side can be occluded with respect to the distinct point by comparing the height of the anchor point with the approximate height of the distinct point. Once the occluded or foreshortened side is determined, that side is excluded from



(a) Case 1: $Z_{r2t2} > Z_{\text{anchor point}}$

(b) Case 2: $Z_{r2t2} < Z_{\text{anchor point}}$

Fig. 7. Geometric constraint for similarity measure for (a) case 1 and (b) case 2.

computation of the similarity measure for matching the distinct point. In this way, we can match the distinct points lying on breaklines more reliably.

In Fig. 7, case 1 shows that the height of the anchor point is lower than that of the distinct points r_2 and t_2 in the reference and target images, respectively. In this case, one can see that the distinct points r_2 and t_2 sit on breaklines. Therefore, the epipolar line pair, I^L_{r2} and I^L_{t2}, on the left side of r_2 and t_2 should be excluded for computing the similarity measure because of the possibility that the area O_T occluded in the target image is high. However, the other side of the epipolar line pair, I^R_{r2} and I^R_{t2}, has less possibility of occlusion; hence, this side contributes only to computing the similarity measure. In case 2, the height of an anchor point is higher than that of the distinct points r_2 and t_2, which shows that the anchor points a_r and a_t lie on the breaklines. Therefore, the epipolar line pair, I^L_{r2} and I^L_{t2}, in the left side of r_2 and t_2 should not be considered in computing the similarity measure due to the foreshortening. The other side of the epipolar line pair, I^L_{r2} and I^L_{t2}, for r_2 and t_2 is selected only for a similarity measure. In summary, if the difference between the height of the anchor point and the approximate height of the distinct points exceeds the predefined threshold, the SELM selects the opposite side to the anchor point for computing the similarity measure of the distinct point. Otherwise, the SELM includes the part between the anchor and the distinct point when computing the similarity measure.

The similarity measure based on the geometric constraint is the normalized correlation coefficient, p_{array}. In most cases, the number of corresponding pixels along the epipolar line pair will differ due to the image scale, terrain undulation, and perspective geometry. Simply, the corresponding pixels are linearly matched in computing p_{array}. Equation (6) represents the cost function, GC_{diff}, drawn from the similarity measure based on the geometric constraint:

$$GC_{\text{diff}} = \frac{(1 - p_{\text{array}})}{2} \quad (0 \le GC_{\text{diff}} \le 1). \tag{6}$$

3. Radiometric Constraint

The radiometric constraint together with the geometric constraint is implemented into the SELM. The major role of the geometric constraint proposed in this paper is to determine the possible occluded side with respect to the distinct point to be matched and to evaluate the similarity of gray-value distributions over the selected line segments on the epipolar lines. In addition to the geometric constraint, the radiometric constraint developed in the paper is to evaluate the similarity of relative differences between the gray value of the anchor points and the gray value of the distinct points in the reference image and the target image, respectively.

The SELM makes use of two measures derived by the radiometric constraint in matching the distinct points. First, the gray-value differences (Δp_{ae}^R and Δp_{ae}^T) between the anchor points and the distinct points (matching candidate points) in the reference image and the target image, respectively, are computed. Second, the gray-value differences (Δp_{se}^R and

Δp_{se}^{T}) between the two end points of the line segment selected by the geometric constraint are computed. These four measures are normalized into two cost measures and then converted into a single cost measure. The two normalized cost measures (p_{ae}^{diff} and p_{se}^{diff}) derived by the radiometric constraint are shown in (7a) and (7b), respectively. The two normalized cost measures are combined into one cost measure (RC_{diff}) as shown in (8). The cost, RC_{diff}, ranges from 0.0 to 1.0. In the case where Δp_{ae}^{R} and Δp_{ae}^{T} have opposite signs, p_{ae}^{diff} is less than 0.0 or greater than 1.0. The two distinct points will therefore not be matched since the radiometric constraints are not satisfied.

$$\begin{cases} p_{ae}^{diff} = \dfrac{\Delta p_{ae}^{R} - \Delta p_{ae}^{T}}{\Delta p_{ae}^{R}}, & \text{if } \left|\Delta p_{ae}^{R}\right| \geq \left|\Delta p_{ae}^{T}\right|, \\ p_{ae}^{diff} = \dfrac{\Delta p_{ae}^{T} - \Delta p_{ae}^{R}}{\Delta p_{ae}^{T}}, & \text{if } \left|\Delta p_{ae}^{T}\right| \geq \left|\Delta p_{ae}^{R}\right|, \end{cases} \tag{7a}$$

$$\begin{cases} p_{dd}^{diff} = \dfrac{\Delta p_{se}^{R} - \Delta p_{se}^{T}}{\Delta p_{se}^{R}}, & \text{if } \left|\Delta p_{se}^{R}\right| \geq \left|\Delta p_{se}^{T}\right|, \\ p_{dd}^{diff} = \dfrac{\Delta p_{se}^{T} - \Delta p_{se}^{R}}{\Delta p_{se}^{T}}, & \text{if } \left|\Delta p_{se}^{T}\right| \geq \left|\Delta p_{se}^{R}\right|, \end{cases} \tag{7b}$$

$$RC_{diff} = \frac{p_{ae}^{diff} + p_{dd}^{diff}}{2} \quad (0 \leq RC_{diff} \leq 1). \tag{8}$$

4. Matching Strategies

After extracting distinct points in a stereo image, the SELM approximately selects matching candidate distinct points using the valleys and peaks. In order to find a correct corresponding distinct point in a target image, the similarity measure is computed for the matched valleys or peaks. Once a distinct point pair is selected, the SELM also assesses whether the height calculated by the point pair is within the height range pre-determined by the intermediate DSM from the AMIM. The SELM computes the similarity measure only for the distinct point pairs that satisfy the above conditions: distinct point type (valley or peak) and height range. The matching cost Φ of the distinct point pair is calculated by (9), which is a cost function used in the SELM. The matching cost consists of the geometric constraints (GC_{diff}) and radiometric constraints (RC_{diff}). In (9), w_G and w_R represent the weights for the geometric and radiometric constraints, respectively, and e is the threshold value for the matching cost.

$$\begin{aligned} &F = w_G \cdot GC_{diff} + w_R \cdot RC_{diff},\ F \leq e, \\ &w_G + w_R = 1,\ 0 \leq w_G \leq 1,\ 0 \leq w_R \leq 1,\ w_G + w_R = 1. \end{aligned} \tag{9}$$

After completion of distinct point matching, the SELM estimates unmatched image points between matched points along the epipolar line. To estimate the heights of unmatched image points, the following two conditions are checked between adjacent matched points. First, the height difference between these points along the epipolar line should be less than the expected height accuracy, which can be estimated by the law of error propagation. Second, the grey-level profiles of the ranges defined by adjacent matched points should be homogenous, or change smoothly. In addition, two ranges in stereo images should be radiometrically similar to be two ranges in the same area. If the above two conditions are satisfied, the heights of the unmatched points are linearly interpolated under the assumption that the surface between the matched points is smooth or homogenous.

The final height of matched points is determined by all height values from all considered stereo pairs. Each stereo pair provides height values with statistical values such as posteriori variances of a space intersection procedure. The final height value for a point in question is then determined by a weighted average method using the posteriori variances of the estimated height as weight values.

V. Experimental Results

The test data used in this research were 3-strip digital aerial images taken by a DMC aerial frame camera over the city of Daejeon in February 2007. The average flying height is about 1,010 m above ground level and the ground sampling distance is about 10 cm, and thus the geometric resolution of the final DSM is 10 cm. The endlap and sidelap are about 80% and 60%, respectively. For evaluating the quality of the DSM generated by the proposed method, lidar data acquired by Optech ALTM3070 in February 2005 were used as reference surface data. The density of lidar data is about 2.5 points per square meter, and thus the geometric resolution is about 60 cm. To investigate the impact of the developed procedure, three experimental cases were configured. In case 1, the AMIM was tested without the use of strip reference images and occlusion detection. In case 2, strip reference images were also utilized as well as nadir reference images. In addition, occluded image points were detected in each image to avoid false similarity measures. Case 3 added the SELM to the conditions of case 2. Table 1 presents a summary of the three experimental cases.

Figure 8 shows the test areas captured in nine overlapped images, which are selected from the test dataset consisting of three strips. As can be seen, the test area contains many occluded areas caused by high-rise apartments. In addition, the tested multiple images have significant radiometric differences with respect to contrast and brightness. Furthermore, the right image in the 9th strip is radiometrically contaminated, as shown in this figure. Therefore, the test images are an ideal test-bed to

Table 1. Three experimental cases for evaluating performance of proposed method.

Case	Strip reference image	Occlusion detection	AMIM	SELM
1			●	
2	●	●	●	
3	●	●	●	●

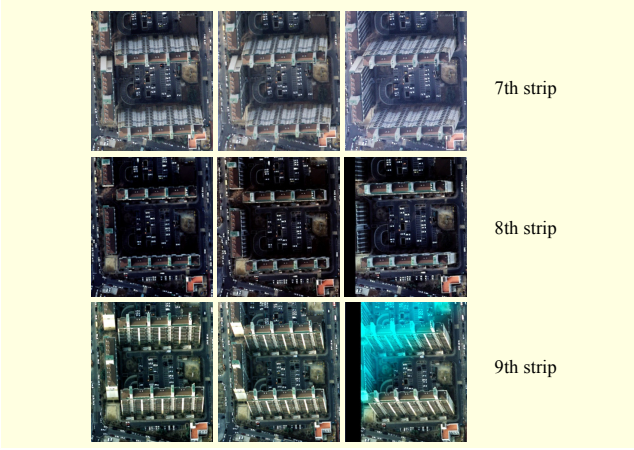


Fig. 8. Nine multiple images taken from three overlapped strips.

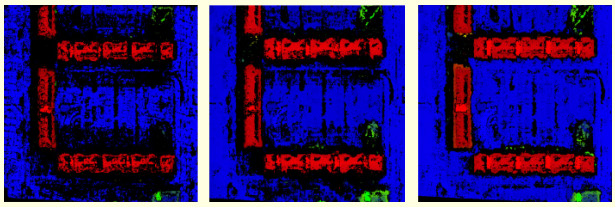


Fig. 9. Result DSMs generated in (a) case 1, (b) case 2, and (c) case 3, described in Table 1.

verify the performance and robustness of the proposed method regardless of image quality. Figure 9 presents the results of the DSMs generated in the three cases. As shown in Figs. 9(a) and 9(b), the use of strip reference images and the geometric occlusion detection procedure significantly improves the quality of the DSM, especially around building areas.

In addition, case 2 has relatively lower chance of unmatched DSM cells compared to the result of case 1. The contribution of the SELM can be verified by comparing Figs. 9(b) and 9(c). It can be seen that the building boundaries in Fig. 9(c) are much clearer than those in Fig. 9(b). In addition to the clear breaklines, it can be confirmed that the SELM increases the number of matched DSM cells even on a homogeneous surface.

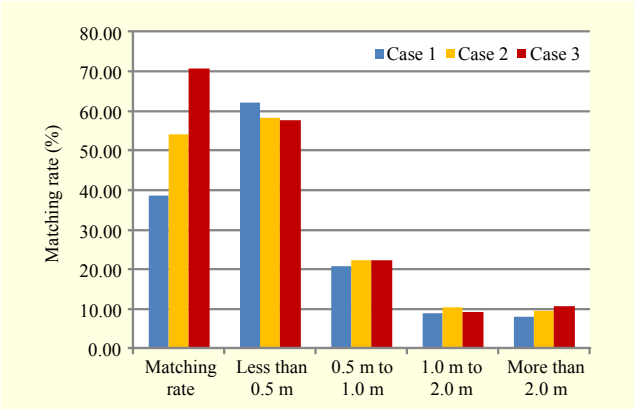


Fig. 10. Matching rate and height accuracy for each case.

Figure 10 presents a quantitative evaluation of the experimental results. This figure shows the matching rate and height accuracy analysis of the generated DSM for each case. The matching rate of case 1 is less than 40%, which is much lower than the matching rates of cases 2 and 3. The highest matching rate is about 70% in case 3. Case 3 has a significantly higher matching rate than that (54%) of case 2 because the SELM has functionality to estimate the height of smooth surfaces including homogenous areas. Even though the matching rates differ significantly, the accuracies of the generated DSMs are compatible. The height accuracy is calculated by the height differences between the generated DSM and the lidar reference surface. The overall height accuracy of about 60% of the matched points is less than 50 cm; about 80% of the estimated heights satisfy 1.0 m accuracy in all three cases. This means that the similarity measure used in the proposed method efficiently identifies matched image points, and consequently the height values of DSM are correctly calculated. From these experimental results, it can be demonstrated that the use of the strip reference images and occlusion detection procedure can considerably improve the matching rate. In addition, it can be confirmed that the SELM contributes to generating clear breaklines and preserving discontinuities. This stems from matching distinct points representing edges along the epipolar line. Furthermore, the matching rate is increased even on smooth surfaces because interpolation of the range between adjacent matched points satisfies the smooth surface. Another feasibility test is presented below. The performance of the proposed method (same as in case 3) will be discussed in relation to three different test areas.

Figure 11 shows aerial images and lidar data captured in three different areas, as well as the final DSMs generated by the proposed method over the same areas using multiple images. U1 and U2 are areas containing high-rise buildings, while relatively low and small buildings are densely distributed in U3. As can be seen in Fig. 11, the generated DSMs are quite

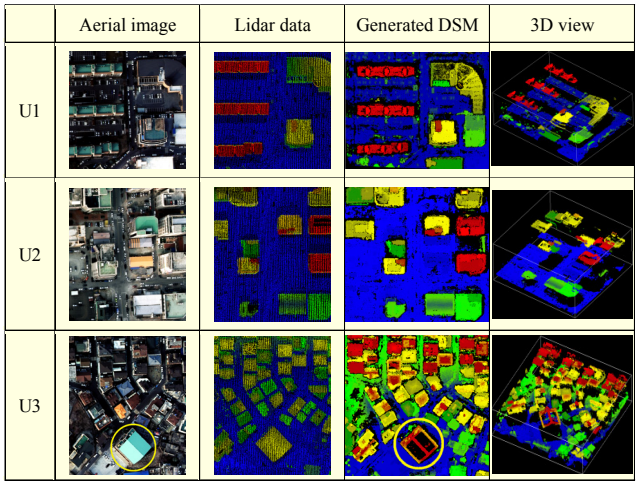


Fig. 11. Aerial images and lidar data for three test areas, and generated DSM using proposed method.

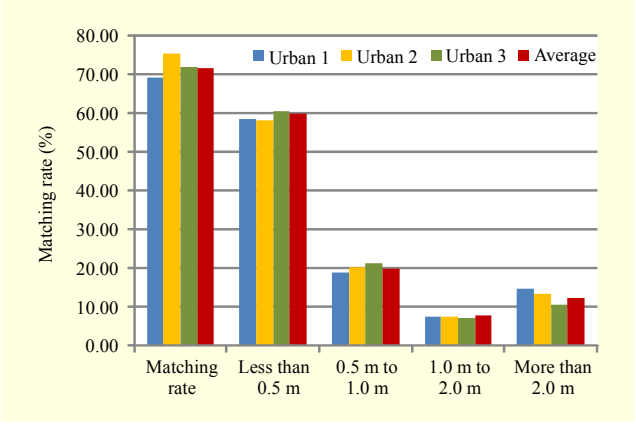


Fig. 12. Matching rate and accuracy for three urban areas.

visually compatible with the reference lidar data. Since the ground resolution of the used multiple images is about 10 cm, the generated DSMs naturally look denser than the lidar data. It should be noted that the proposed method has a weak point in radiometrically homogeneous large areas where area-based matching could fail, and distinct points and breaklines are not observed. The circled area in the U3 shows a failed area on a gable roof. As can be confirmed in the aerial image, the gable roof is quite large and appears to be radiometrically homogeneous. For this reason, the proposed method failed to generate matched points along the roof. In the case of a flat roof, the SELM can interpolate unmatched points along the epipolar line. However, if the heights of both ends of unmatched areas differ, such as in the case of a gable roof, the proposed method does not run the interpolation because the area is defined as a discrete surface. The authors recommend using a common DSM interpolation filter to solve this problem because this type of failure occurs in local inclined surfaces.

The accuracy and matching quality of the generated DSM in each area are quantitatively analyzed in Fig. 12. As shown in this figure, the matching rates for the three urban areas are quite compatible, and their average matching rate is about 70%. In terms of height accuracy, about 60% of the matched points are less than 50 cm, and about 80% of the matched points have less than 1.0-m vertical errors in all three urban areas. This is a noticeable feature of the proposed method since it indicates that the performance of the proposed method is fairly robust. The fully automatic method has been implemented on a personal computer having 2.4-GHz dual core CPUs and 3.0-Gbyte RAM. The test matching area, as shown in Fig. 8, is about 0.15 km^2 and composed of 9 images in 3 strips. This matching procedure took around 5 hours.

VI. Conclusion

This paper introduced new methods to generate an accurate, dense, and discontinuity-preserved DSM using multiple images. The proposed method consists of two procedures: the AMIM and SELM. The AMIM is an object-space-guided area-based matching using the vertical line locus, while the SELM is a type of feature-based matching with geometric and radiometric constraints using the epipolar geometry. The AMIM generates a dense DSM using a novel method to define search ranges adaptively while taking occlusion into consideration. The major role of the SELM is to cope with common problems that the area-based matching cannot avoid such as discontinuities and homogeneous areas. The experimental results demonstrate the advantages of the proposed method in terms of matching quality and accuracy of the final DSM. In particular, the DSM generated by the proposed method is as dense as the image ground resolution, and discontinuities such as building boundaries are preserved. In addition, the proposed method is not expected to be seriously affected by the radiometric quality of the images, and the height accuracy of the generated DSM is also quite reliable. The vertical error of 60% of the generated DSM points is less than 50 cm, constituting fairly reasonable accuracy, because the derived height values based on the error propagation for test images can be estimated with about 60-cm standard deviation for 80% and 60% overlap ratio. Therefore, the matching algorithm presented in this paper can be useful in applications for generating a large-scale DSM.

References

[1] C. Baillard and O. Dissard, "A Stereo Matching Algorithm for Urban Digital Elevation Models," *Photogrammetric Eng. Remote Sens.*, vol. 66, no. 9, Sept. 2000, pp. 1119-1128.

[2] A. Alobeid, K. Jacobsen, and C. Heipke, "Comparison of Matching Algorithms for DSM Generation in Urban Areas from Ikonos Imagery," *Photogrammetric Eng. Remote Sens.*, vol. 76, no. 9, Sept. 2010, pp. 1041-1050.

[3] J.S. Ku, K.M. Lee, and S.U. Lee, "Multi-mage Matching for a General Motion Stereo Camera Model," *Proc. IEEE Int. Conf. Image Process.*, Chicago, IL, USA, 1998, pp. 608-612.

[4] H.G. Mass, "Automatic DEM Generation by Multi-image Image Feature Base Matching," *Proc. ISPRS Congress Vienna*, Vienna, Austria, 1996, pp. 484-489.

[5] R.T. Collins, "A Space-Sweep Approach to True Multi-image Matching," *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition*, San Francisco, CA, USA, 1996, pp. 358-363.

[6] J. Shao, R. Mohr, and C. Fraser, "Multi-image Matching Using Segment Features," *Proc. ISPRS Congress Amsterdam*, Amsterdam, Netherlands, 2000, pp. 837-844.

[7] A.F. Elaksher, "Multi Image Matching of Straight Lines with Geometric Constraints," *Proc. ISPRS GEOgraphic Object Based Image Analysis*, Calgary, Alberta, Canada, 2008 (electronic form).

[8] F. Jung, V. Tollu, and N. Paparoditis, "Extracting 3D Edgel Hypotheses from Multiple Calibrated Images: A Step towards The Reconstruction of Curved and Straight Object Boundary Lines," *Proc. ISPRS Photogrammetric Comput. Vision*, Graz, Austria, 2002, pp. 100-104.

[9] F. Jung and N. Paparoditis, "Extracting 3D Free-Form Surface Boundaries of Man-Made Objects from Multiple Calibrated Images: A Robust, Accurate and High Resolving Power Edgel Matching and Chaining Approach," *Proc. ISPRS Photogrammetric Image Analysis*, Munich, Germany, 2003, pp. 39-44.

[10] L. Zhang, *Automatic Digital Surface Model(DEM) Generation from Linear Array Images*, doctoral dissertation, Swiss Federal Institute of Technology, Zurich, 2005.

[11] M.N. Pateraki, *Adaptive Multi-image Matching for DEM Generation from Airborne Linear Array CCD Data*, PhD dissertation, Swiss Federal Institute of Technology, Zurich, 2005.

[12] U.M. Leloglu, M. Roux, and H. Maitre, "Urban DEM with Three or More High-Resolution Aerial Images," *Proc. ISPRS GIS between Visions Appl.*, Stuttgart, Germany, 1998, pp. 347-352.

[13] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, Feb. 2008, pp. 328-341.

[14] G. Zhang et al., "Consistent Depth Maps Recovery from a Video Sequence," *IEEE Trans. Pattern Anal. Mach.*, vol. 31, no. 6, June 2009, pp. 974-988.

[15] D. Bulatov and J.E. Lavery, "Reconstruction and Texturing of 3D Urban Terrain from Uncalibrated Monocular Images Using L1 Splines," *Photogrammetric Eng. Remote Sens.*, vol. 76, no. 4, Apr. 2010, pp. 439-449.

[16] T. Schenk, *Digital Photogrammetry*, vol. 1, Ohio: TerraScience, 1999.

[17] N. Paparoditis, C. Thom, and H. Jibrini, "Surface Reconstruction in Urban Areas from Multiple Views with Aerial Digital Frame Cameras," *Proc. ISPRS Congress Amsterdam*, Amsterdam, Netherlands, 2000 (CD ROM).

[18] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust Multiview Stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, Aug. 2010, pp. 1362-1376.

[19] T. Kanade, "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 9, Sept. 1994, pp. 920-932.

Myoung-Jong Noh received the BSc in civil engineering, the MSc and the PhD in geoinfomatic engineering from Inha University, Incheon, Rep. of Korea, in 2003, 2005 and 2011, respectively. He is currently a postdoctoral researcher with the Byrd Polar Research Center, Ohio State University, Columbus, Ohio, USA. His research interests include the fields of digital photogrammetry, geospatial data processing, and earth science.

Woosug Cho received the BSc from the Department of Civil Engineering, Hanyang University, Rep. of Korea, the MSc in Department of Civil Engineering from Iowa State University, USA, and the PhD from the Department of Geodetic Science & Mapping, Ohio State University, USA, in 1985, 1988, and 1995, respectively. He worked for Korea Research Institute for Human Settlement (KRIHS) in 1996. Since 1997, he has been a faculty member of the Department of Civil Engineering, Inha University, Rep. of Korea. His research interests include digital photogrammetry, airborne laser and online mapping, automatic 3D data generation, land monitoring, and forest application with lidar.

Ki In Bang received the BSc in civil engineering from Inha University, Incheon, Rep. of Korea, in 1998; the MSc in geoinformatics engineering from Inha University in 2001; and the PhD in geomatics engineering from the University of Calgary, Calgary, Canada, in 2010. He worked for Daewoo Construction Ltd. as a civil engineer in 1998 and was a research scientist of ETRI from 2001 to 2004. He is currently an adjunct professor with the Department of Civil Engineering, Inha University, and a CTO of GeonTechKorea, Seoul, Rep. of Korea. His research interests include the fields of digital photogrammetry, laser scanning systems, and geospatial data processing.