Image-based 3D Building Reconstruction Using A-KAZE Feature Extraction Algorithm

H. Seong, H. Choi, H. Son, and C. Kim*

Department of Architectural Engineering, Chung-Ang University, Seoul 06974, South Korea
E-mail: gusdn7543@cau.ac.kr, vianj@cau.ac.kr, hjson0908@cau.ac.kr, changwan@cau.ac.kr (* Correponding author)

Abstract

The development of 3D reconstruction from 2D building images enables cost-effective and accurate acquisition of spatial data. Feature extraction is a fundamental technique for 3D reconstruction method. However, buildings mostly consist of planar surfaces whose entities are feature-less. This study presents 3D building reconstruction using A-KAZE feature extraction algorithm. Because A-KAZE algorithm does not use Gaussian blurring like SIFT and SURF, A-KAZE algorithm has potential to extract correct visual features for feature matching and 3D reconstruction. The proposed method was tested on actual building scenes acquired from a high-resolution camera. The experimental results showed that the A-KAZE algorithm can detect the sufficient number of features with low computation time. It is expected that the proposed method can be implemented in comprehensive 3D reconstruction of civil infrastructures.

Keywords –

3D reconstruction; feature extraction; A-KAZE; binary descriptor; structure from motion; photogrammetry

1 Introduction

The demand for 3-dimensioanl (3D) reconstruction of a building is increased over the last decade in the field of the Architecture, Engineering, Construction and Facilities Management (AEC&FM), with the purpose of collecting as-built spatial data. As-built spatial data can be used for a variety of applications including inspection of defects [1], building performance assessment [2], and document updates [3]. Traditional approaches of collecting spatial data like manual measurement with visual inspection and total station survey are not only labor-intensive and time-consuming, but also inaccurate because such measurements become highly dependent on the skill, experience, and expertise of the workforce. A potential alternative to these conventional approaches, therefore, is a laser scanning technology, which can generate 3D point cloud data of a structure promptly with a high level of accuracy. However, in spite of the distinctive advantage of this remote sensing technology, its use is limited by its expensive cost and requirement of highly skilled operators. To overcome this challenge, image-based 3D reconstruction has drawn attention because of its ease of use, inexpensive cost, and higher accessibility with the development of computer vision-based techniques.

Structure from Motion (SfM) is a photogrammetric method that generate 3D point cloud data using multiple images acquired from different view angles [4], which is an equivalent principle to a human's ability to perceive a 3D structure from 2D projection. To estimate the intrinsic and extrinsic parameters of camera and to reconstruct a 3D geometry, SfM method uses a process of finding corresponding features from the images. So far, numerous feature extraction algorithms have been proposed in the field of computer vision and image processing since extracting and matching such pairwise features among images are the key steps of SfM method [5]. However, using only visual features to reconstruct a 3D scene is a challenging task. Recently, many researchers have used different types of geometric entities as visual features to estimate the parameters of camera. The most commonly used visual feature is the point-based feature, also known as interest-point or keypoint. Since point-based methods have significantly high time complexity, lines and planes are also used as high-level visual features [6]. These visual features can be found on building scenes, whose objects are mostly consisting of planar surfaces. However, high-level visual features are commonly variant to changes in orientation of scenes and illumination information [7].

Recently, retrieval algorithms with localized keypoint are widely used in 3D reconstruction techniques. The keypoint is mostly detected on regions with significant variation in edges and keypoint descriptors are invariant to orientation changes of images acquired through different angle-view. In computer vision, several 3D reconstruction commercial software packages have emerged to gain robustness and enhance efficiency of extracting features from images.

During recent decades, Scale-Invariant Feature Transform (SIFT) [8] algorithm has been perceived as the most common choice for feature detector and descriptor. However, SIFT and Speeded Up Robust Feature (SURF) [9] exploit Gaussian scale space which blurs the details with noises to equivalent extent and does not preserve boundaries of an object [10]. Since buildings have elements with sharp edges like wall and roof, other alternative algorithm is needed rather than SIFT and SURF. Otherwise, manual optimization of miss-matched image pairs is required when using an unreliable feature extraction algorithm. Barazzetti et al [11] presented 3D reconstruction method to obtain accurate point clouds automatically. This study evaluated the performance of SIFT descriptor on different sizes of test images. The evaluation results showed that computation time increased considerably when the size of the image was increased.

This study presents a 3D reconstruction method using A-KAZE feature extraction algorithm for finding pairwise features from building images, to generate an accurate 3D points clouds automatically. A-KAZE is a non-linear scale space-based algorithm that has a great potential of preserving sharp boundaries of buildings' elements. In this study, the comparative analysis among SIFT, SURF, and A-KAZE has served to evaluate each algorithm as a feature detector and descriptor. To achieve a practical implementation, a typical building has set to be a target for 3D reconstruction.

2 Methodology

Accelerated-KAZE (A-KAZE) feature detector and descriptor, which is proposed by Alcantarilla et al. [12], uses nonlinear scale spaces to extract corresponding features of images. This algorithm is developed from KAZE algorithm [13] by embedding mathematical Fast Explicit Diffusion (FED) in pyramidal structure to speed up the nonlinear scale space computation. In comparison to Gaussian scale space of SIFT and SURF, nonlinear scale space can extract corresponding features of images while maintaining details and reducing noises by means of nonlinear diffusion filtering. Moreover, the computational complexity $O(n^2)$ (where n is the number of input images) of A-KAZE is lower than that of SIFT and SURF [14].

A-KAZE is mainly comprised of three steps: a nonlinear scale space building with FED, feature detection, and feature description. The first step is to define a set of evolution times for building the nonlinear scale space. Then A-KAZE detects the feature points that are the local maxima of the determinant of the Hessian matrix using nonlinear scale space. Finally, given feature points are represented by Modified-Local Difference Binary (M-LDB).

In the first step, A-KAZE builds pyramidal structure of nonlinear scale space. The scale space is discretized into a series of octaves (O) and sub-levels (S). The set of octaves and sub-levels are identified by discrete octave (o) and a sub-level (s) and they are mapped to their corresponding scale (σ) as shown in equation (1).

$$\sigma_i(o,s) = 2^{o+s/S},\ o \in [0...O-1],\ s \in [0...S-1],\ i \in [0...M] \quad (1)$$

Where M is the total number of filtered images. Then the set of discrete scale levels in pixel units is transformed into σi time units because nonlinear diffusion filtering is defined in time term as shown in equation (2).

$$t_i = \frac{1}{2}\sigma_i^2,\ i = \{0...M\} \quad (2)$$

Moreover, the input image can be convolved with a Gaussian blurring of standard deviation σ to reduce noise and possible artefacts. From the filtered input image, the contrast factor (λ) is computed in an automatic way as the 70 % of the gradient histogram. To overcome the main drawback of KAZE which is its high computational cost, FED scheme is embedded into the pyramidal structure.

In the second step, feature points which are local maxima are detected from the filtered images (Li), by computing the determinant of the Hessian matrix in the nonlinear scale space. Hessian matrix of the images is computed by a normalized scale factor as shown in equation (3),

$$L^i_{Hessian} = \sigma^2_{i,norm}(L^i_{xx}L^i_{yy} - L^i_{xy}L^i_{xy}) \quad (3)$$

Then the determinant of the given Hessian Matrix is computed to find local maxima points at each evolution level. The pre-defined threshold for potential local maxima points is set to be a window of 3×3 pixels. The computation is implemented efficiently by discarding non-local maxima points. These candidate points are recalculated to determine whether the points are maxima by comparing each point form i+1 and i-1 evolution level. Finally, the given maxima points are assigned to be feature points.

A Modified-Local Difference Binary (M-LDB) is employed as a feature descriptor in A-KAZE. The M-LBD descriptor follows the principle of a LDB descriptor [15] that computes binary strings for image patches using gradient and intensity information from non-linear scale space. In the LDB descriptor, each image patch is divided into various sizes of n×n grid cells and those partitions are efficient to compute average intensity. However, the LDB descriptor is sensitive to rotation of image patches that requires high computational cost. Thus, A-KAZE achieves a scale and rotation invariance by scale-dependent sampling and estimating the main orientation information. The M-LDB uses horizontal and vertical derivatives from detected features to construct the descriptor.

In this study, a brute-force method is employed for efficient feature matching. The brute-force approach finds corresponding features from multiple images in different angle. The process mainly consists of three steps: comparing binary descriptors and testing ratio, using estimators to compute fundamental matrixes, and using epipolar constraint for eliminating outlier matches. Then, only inlier matches are exploited for homography estimation by using random sample consensus (RANSAC) method.

3 Experimental Results

To demonstrate that A-KAZE algorithm is an optimal feature extraction algorithm for building scenes as a feature detector and descriptor, a comparative analysis has presented. OpenCV library is used for implementation of feature extraction and matching algorithms, with the default parameters. The test images of building's wall surface are captured by a Zenmuse Z3 camera equipped in a Quadcopter DJI Matrice 100 (see Figure 1). Resolution of the images are ultra-high-definition of 4000×2250 pixels (4K). Since the size of the building is incompatible to any other objects, 3D reconstruction is needed to be facilitated by using images of high resolution to provide a 3D model with valuable and practical information for building operators.

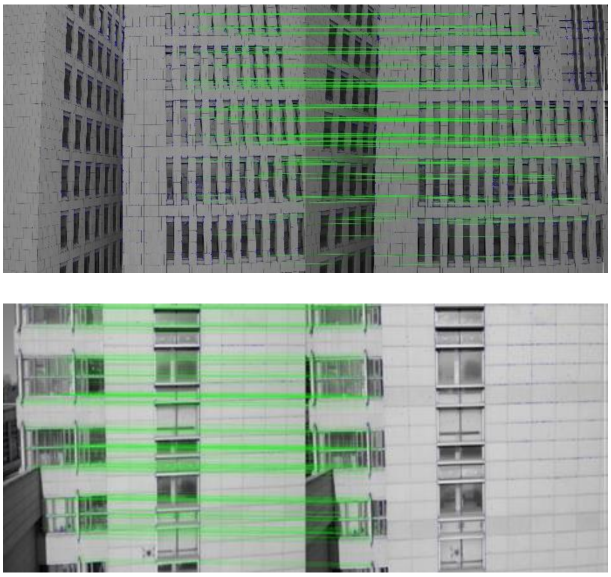

Figure 1. Exemplar images of feature extraction results of building scenes

Table 1. The results of feature extraction and matching of building scenes. The highest values are indicated in bold.

	Detected keypoints	Matching score	Computation time (sec,)
SIFT	2051	26.09	6.88
SURF	**16603**	16.26	11.06
A-KAZE	3317	**32.47**	**5.91**

As shown in Figure 1, keypoints are sufficiently detected and matched. The results also imply that A-KAZE algorithm reliably extracted corresponding keypoints from building scenes whose objects had poor-textured surface. Table 1 presents the matching results of Figure 1 using SIFT, SURF, and A-KAZE. As presented in Table 1, the matching results indicate that SURF algorithm detected the highest number of keypoints from the images and for the matched features. However, SURF algorithm showed the lowest matching score among algorithms. Although SIFT algorithm showed compatible matching score result, the number of detected keypoints and matched features are relatively lower than other algorithms. A-KAZE algorithm achieved a stable performance in detecting keypoints with the highest matching score.

Since feature extraction and matching are repetitive processes on many multiple images in 3D reconstruction procedure, computation time is a critical determinant. The time complexity for computing A-KAZE algorithm is presented as the lowest among the three algorithms, as shown in Table 1. These results potentially imply that A-KAZE is an optimal algorithm for extracting and matching correct features from building images.

Figure 2 is a result of 3D building reconstruction of which surface is poor-textured. The 3D reconstruction model is generated by 43 sequential images of 4K resolution using A-KAZE algorithm. The result presents a similar scene to its source images. This potentially imply that A-KAZE is an optimal algorithm for extracting and matching correct features from building images.

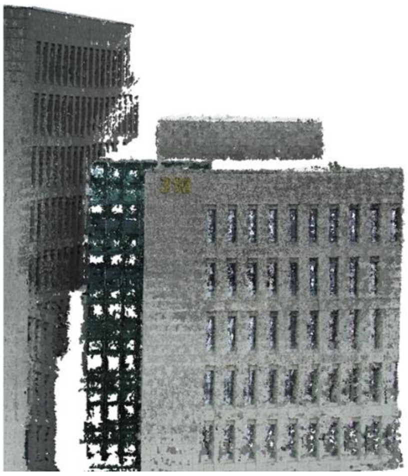

Figure 2. Exemplar images of 3D reconstruction of building

4 Conclusion

This study presented a 3D reconstruction method using A-KAZE feature extraction algorithm to detect correct corresponding keypoints and to reconstruct accurate 3D point clouds of buildings automatically. Because buildings mostly consist of texture-less elements with planar surfaces, conventional feature extraction algorithms encountered challenges. To solve this problem, A-KAZE feature extraction algorithm was employed in this study. To assess the performance of A-KAZE algorithm as feature detector and descriptor, analysis of detecting and matching results has conducted. The results presented that A-KAZE algorithm reliably and efficiently detected corresponding features from high-resolution building images which were acquired in different view angles. This study, thus far, contributes to establishing a comprehensive 3D reconstruction method for buildings.

Acknowledgements

This research was supported by a grant (18CTAP-C129716-02) from infrastructure and transportation technology promotion research Program funded by Ministry of Land, Infrastructure and Transport (MOLIT) of Korea government and Korea Agency for Infrastructure Technology Advancement (KAIA).

References

[1] Hamledari H., Rezazadeh Azar E., and McCabe B. IFC-based Development of As-built and As-is BIMs using construction and facility inspection data: Site-to-BIM data transfer automation. *Journal of Computing in Civil Engineering*, 32(2), 2017.

[2] Habibi S. The promise of BIM for improving building performance. *Energy and Buildings*, 153, 525–548, 2017.

[3] Klein L., Li N., and Becerik-Gerber, B. Imaged-based verification of as-built documentation of operational buildings. *Automation in Construction*, 21:161–171, 2012.

[4] Siebert S., and Teizer J. Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. *Automation in Construction*, 41:1–14, 2014.

[5] Moreels P. and Perona P. Evaluation of features detectors and descriptors based on 3d objects. I*nternational Journal of Computer Vision*, 73(3): 263–284, 2007.

[6] Bartoli A. and Sturm P. Structure-from-motion using lines: Representation, triangulation, and bundle adjustment. *Computer vision and image understanding*, 100(3): 416–441, 2005.

[7] Hasheminasab M., Ebadi H., and Sedaghat A. An integrated ransac and graph based mismatch elimination approach for wide-baseline image matching. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(1): 297, 2015.

[8] Lowe D. G. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on Computer Vision*, pages 1150–1157, Kerkyra, Greece, 1999.

[9] Bay H., Tuytelaars T., and Van Gool L. (2006). Surf: Speeded up robust features. *Computer vision–ECCV*, 404–417.

[10] Nabiyev V. V., Yılmaz S., Günay A., Muzaffer G., and Ulutaş G. Shredded banknotes reconstruction using AKAZE points. *Forensic science international*, 278: 280–295, 2017.

[11] Barazzetti L., Scaioni M., and Remondino, F. Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation. *The Photogrammetric Record*, 25(132): 356-381, 2010.

[12] Alcantarilla P. F., Nuevo J., and Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces, In Proceedings of the British Machine Vision, 2013.

[13] Alcantarilla P. F., Bartoli A., and Davison A. J. (2012, October). KAZE features. In *Proceedings of European Conference on Computer Vision*, pages 214-227, Berlin, Germany, 2012.

[14] Jiang G., Liu L., Zhu W., Yin S., and Wei S. A 127 fps in full HD accelerator based on optimized AKAZE with efficiency and effectiveness for image feature extraction. In *Proceedings of Design Automation Conference, pages* 1-6, 2015.

[15] Yang X., and Cheng K. T. LDB: An ultra-fast feature for scalable augmented reality on mobile devices. In *Proceedings of Mixed and Augmented Reality (ISMAR)*, pages 49-57, 2012.