applied sciences

Article

Deep Learning-Based Automated Background Removal for Structural Exterior Image Stitching

Myung Soo Kang and Yun-Kyu An *

Department of Architectural Engineering, Sejong University, Seoul 05006, Korea; kms35954@sju.ac.kr
* Correspondence: yunkyuan@sejong.ac.kr; Tel.: +82-2-6935-2426

Abstract: This paper presents a deep learning-based automated background removal technique for structural exterior image stitching. In order to establish an exterior damage map of a structure using an unmanned aerial vehicle (UAV), a close-up vision scanning is typically required. However, unwanted background objects are often captured within the scanned digital images. Since the unnecessary background objects often cause serious distortion on the image stitching process, they should be removed. In this paper, the automated background removal technique using deep learning-based depth estimation is proposed. Based on the fact that the region of interest has closer working distance than the background ones from the camera, the background region within the digital images can be automatically removed using a deep learning-based depth estimation network. In addition, an optimal digital image selection based on feature matching-based overlap ratio is proposed. The proposed technique is experimentally validated using UAV-scanned digital images acquired from an in-situ high-rise building structure. The validation test results show that the optimal digital images obtained from the proposed technique produce the precise structural exterior map with computational cost reduction of 85.7%, while raw scanned digital images fail to construct the structural exterior map and cause serious stitching distortion.

Keywords: digital image stitching; automated background removal; region of interest extraction; deep learning-based depth estimation; structure exterior map

check for updates

Citation: Kang, M.S.; An, Y.-K. Deep Learning-Based Automated Background Removal for Structural Exterior Image Stitching. *Appl. Sci.* **2021**, *11*, 3339. https://doi.org/10.3390/app11083339

Academic Editor: Antonio Fernández

Received: 19 February 2021
Accepted: 5 April 2021
Published: 8 April 2021

1. Introduction

Monitoring the integrity of aging structures has become increasingly important in terms of extending structures' service life and saving maintenance costs. For effective monitoring of large-scale structures, unmanned aerial vehicles (UAVs) have recently played a key role, in that faster and safer inspection is possible than expert-dependent visual inspection, even for inaccessible areas by human beings [1–3]. One of the most popular UAV-based inspection strategies is that structural exterior damage can be effectively assessed by using UAV-captured digital images. However, damage assessment and making decisions from a number of digital images often be labor-intensive and unreliable, especially as the target structure gets larger. In particular, damage quantification as well as localization are challenging works without structural exterior map establishment. To tackle the technical issues, digital image stitching techniques have been widely accepted for entire structural exterior mapping [4–7]. As for precise structural damage quantification and localization including micro-scale damage, the close-up and high-resolution spatial scanning of a digital camera-mounted UAV along the entire structural region of interest (ROI) is often required [8–11]. To construct structural exterior maps using the digital images scanned along a large-scale structure, optimal digital images should be selected by considering the overlap ratio between adjacent digital images to be stitched. The use of all raw digital images for structural exterior map establishment is not effective in terms of high computational cost as well as image stitching accuracy. To address the optimal digital image selection issue, several techniques have been investigated. Yang et al. [12]

used a constant time interval technique, which extracts video frames every two seconds. This technique can reduce the spatial redundancy of the acquired video frames, but it cannot meet the constant overlap ratio between the selected digital images. Then, to ensure the constant overlap ratio, Bang et al. [13] proposed a key frame selection technique with the known operational condition of UAV. Similarly, Bu et al. [14] employed monocular simultaneous localization and mapping (SLAM) to stitch UAV-scanned images in real-time. They calculated the relative distance among adjacent images through the weighted combination of translation and rotation in large-scale direct SLAM, and the key frames were then selected by using a certain threshold.

However, in the UAV's close-up vision scanning, digital images often include the target ROI and unnecessary background together especially in the edge of a target structure. The background objects such as sky, mountain, river, tree, etc. disturb stitching as well as selecting optimal digital images, because the background objects have extremely different feature variations from the target structural ROI ones on the sequentially scanned images. In addition, there are more distinguishable image features in the background objects than the repetitive and local target ROI ones, which pose serious distortion and ghosting effects on the image stitching process. To solve this problem, a number of trials have been conducted. For example, Xin et al. [15] proposed a self-adaptive optical flow technique to detect target object regions on the sequential image data. They tried to enhance the object outlines from a rough optical flow field using local mean algorithm, and the target object regions were then extracted. In addition, Supreeth and Patil [16] studied a multiple moving object tracking technique, enabling them to achieve robustness against objects' occlusion, shadows and camera jitter by combining background subtraction and k-means clustering. More recently, Fang et al. [17] proposed a deep learning network, called Tiramisu trained with common objects in context (COCO) dataset, which segments target objects for background removal. Although the aforementioned background removal techniques can be effective tools when it comes to digital images obtained under constant camera pose, in scanning speed and path conditions, the UAV's close-up scanning condition especially for outdoor buildings can be sensitively altered by surrounding environmental conditions as well as operator's skill. Moreover, the conventional background removal techniques highly depend on the image blurs and noises, but the image blur and noise phenomena on the sequential images captured under continuous spatial scanning unfortunately are inevitable in reality.

In this paper, a deep learning-based automated background removal technique, which is suitable for the UAV's close-up scanning condition, is newly proposed. The proposed technique has the following superior characteristics over the existing techniques: (1) the digital image acquisition conditions are not strictly restricted; (2) it is robust against the image blur and noise phenomena; (3) the computational cost can be minimized through optimal image selection using image feature matching-based overlap ratio calculation. The proposed technique is experimentally validated using UAV-scanned digital images acquired from an in-situ high-rise building structure.

This paper is organized as follows. First, the deep learning-based automated background removal technique including an optimal digital image selection algorithm is developed in Section 2. Then, the experimental validation results are shown in Section 3. Finally, this paper is concluded with a brief discussion.

2. Structural Exterior Image Stitching through Automated Background Removal

Figure 1 shows the overview of the structural exterior map establishment through deep learning-based automated background removal and optimal image selection. Once the vision camera mounted-UAV scans the target structure with a short working distance along a predefined scanning path, the spatially continuous digital images can be acquired for high-resolution structural exterior map establishment. To properly stitch the scanned digital images, the ROI images including only a target structure need to be extracted from the field of view (FOV) images. Since the FOV images, which are especially obtained from the edge of the target structure as shown in Figure 1, inevitably contain the background objects as well as ROI, the background regions are removed using a deep learning-based depth estimation network. Subsequently, the optimal images for minimizing stitching errors as well as computational costs are selected from the entire video frames based on overlap ratio calculation. Finally, the structural exterior map is constructed using a mesh-based image stitching method as shown in Figure 1. The details of each procedure are as follows.

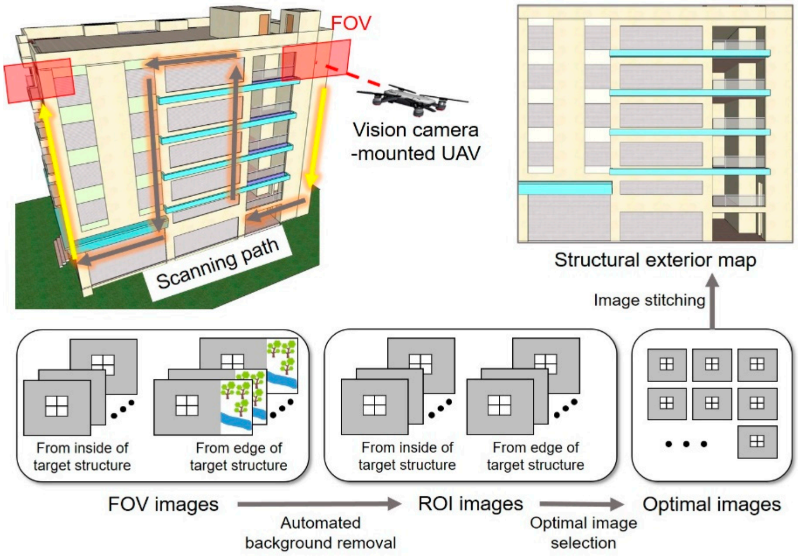

FOV
Vision camera -mounted UAV
Scanning path
Structural exterior map
Image stitching
From inside of target structure
From edge of target structure
From inside of target structure
From edge of target structure
FOV images
Automated background removal
ROI images
Optimal image selection
Optimal images

Figure 1. Overview of structural exterior map establishment through deep learning-based automated background removal and optimal image selection: FOV and ROI denote the field of view and region of interest, respectively.

2.1. Automated Background Removal Using Deep Learning-Based Depth Estimation

Figure 2 shows the deep learning-based automated background removal process. The key idea of this step is under assumption that structural ROI, which is obtained through the UAV's close-up scanning, is much closer than the background objects from the UAV-embedded digital camera within FOV. Thus, a deep learning-based depth estimation network, called Monodepth2, which is based on a U-net architecture, is employed in this study [18]. Monodepth2 is trained in a self-supervised manner by exploiting spatial geometry constraints. Monodepth2 utilizes a full-resolution multi-scale sampling method for reducing visual artifacts and an auto masking loss to ignore training pixels that violate camera motion assumption. This network can rapidly estimate depth value using only monocular RGB images, thus it is suitable for high-resolution image processing. The effectiveness of Monodepth2 was validated by comparing with 28 other depth estimation models using measurement metrics [18]. The employed model was implemented in PyTorch and trained for 20 epochs using an Adam optimizer with a batch size of 12. The learning rate of 10^{-4} is used for the first 15 epochs which is then dropped to 10^{-5} for the remainder. The smoothness term is set to 0.001. The KITTI dataset is used for pre-training, and 10% of the dataset is used as a validation set.

Once the depth values were estimated with respect to each pixel on the background, including FOV images acquired from the edge of the target structure, the depth image was obtained as shown in Figure 2. Subsequently, the ROI boundary was extracted by using depth difference. However, precise ROI boundary extraction is often difficult due to undesired noise components on the depth images. Thus, post image processing was necessary. On the depth images in Figure 2, brighter pixels indicate the closer working distance from the digital camera. The ROI boundaries within the depth image can be extracted by using the magnitude of depth gradients (G), which is given by:

$$G = \sqrt{\left(\frac{\partial I'}{\partial x}\right)^2 + \left(\frac{\partial I'}{\partial y}\right)^2} \tag{1}$$

where I' is the depth image corresponding to the FOV image, and $\frac{\partial I'}{\partial x}$ and $\frac{\partial I'}{\partial y}$ are the depth gradients of I' along the x and y directions, respectively. To precisely extract the ROI boundaries, the G image was binarized by an Otsu's method, so that each pixel had a value 1 or 0 [19]. Here, dotted pepper noises were removed on the boundary image, resulting in clear a ROI boundary consisting of consecutive pixel sets as depicted in Figure 2. Then, ROI was extracted by retaining the region, which has smaller mean depth values on the depth image across the ROI boundary. Finally, the background region was automatically removed by overlapping the extracted ROI region on the FOV image.

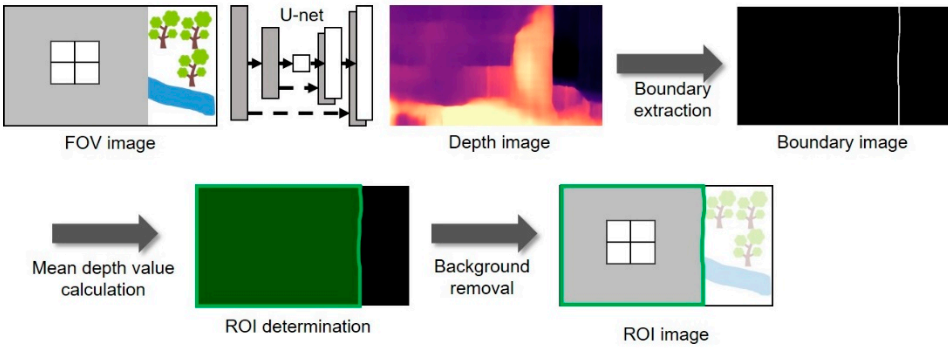


Figure 2. Automated background removal using deep learning network based on a U-Net architecture.

2.2. *Optimal Image Selection for Cost-Effective Digital Image Stitching*

To construct the structural exterior map, a number of spatially continuous digital images, which is often expressed by video frames, should be acquired, because FOV is often much smaller than the entire ROI of a large-scale structure. Moreover, image resolution should be large enough to inspect micro-scale damage on the target structure. For these reasons, structural exterior map construction using entire scanned digital images typically require tremendous computational costs. Furthermore, image stitching errors are often inversely increased when the entire scanned digital images are excessively used. In order to address these technical issues, optimal image or frame selection is necessary. In this step, the optimal image selection algorithm using image feature matching-based overlap ratio calculation was proposed.

Figure 3 describes the flow chart of the overlap ratio-based optimal image selection algorithm. First, image features such as point, corner or edges were extracted from every ROI image (I_1, I_2, I_3 … I_n) using a scale-invariant feature transform (SIFT) [20]. Since SIFT is invariant to image translation, scaling, rotation and partially invariant to illumination changes, it is advantageous for UAV's close-up scanning data processing. Next, the image features were initially matched between the adjacent images, which are defined as the matching vectors. Then, the false matching vectors were removed by using a random sample consensus (RANSAC) because similar image features on the repeated target structure's texture are often mismatched. After RANSAC, the correct matching vectors, called the inlier vectors, were obtained which physically imply how much I_q was translated from I_p along the x and y directions. Based on the assumption that there is no working distance change between the target structure and the digital camera mounted on UAV, the overlap ratio R between I_p and I_q was calculated using the inlier vectors' averaged magnitude along the x and y directions, which is given by:

$$R = \frac{100 * (Height - \mu) * (Width - \nu)}{Height * Width} \tag{2}$$

where $Height$ and $Width$ are the height and width of I_p. μ and ν are the inlier vectors' averaged magnitude along the x and y directions as shown in Figure 3. Once R was calculated, the optimal image that satisfies the predefined threshold of R (R_{th}) was determined. The above procedure was repeated by the iteration as shown in Figure 3, until the entire optimal images were obtained.

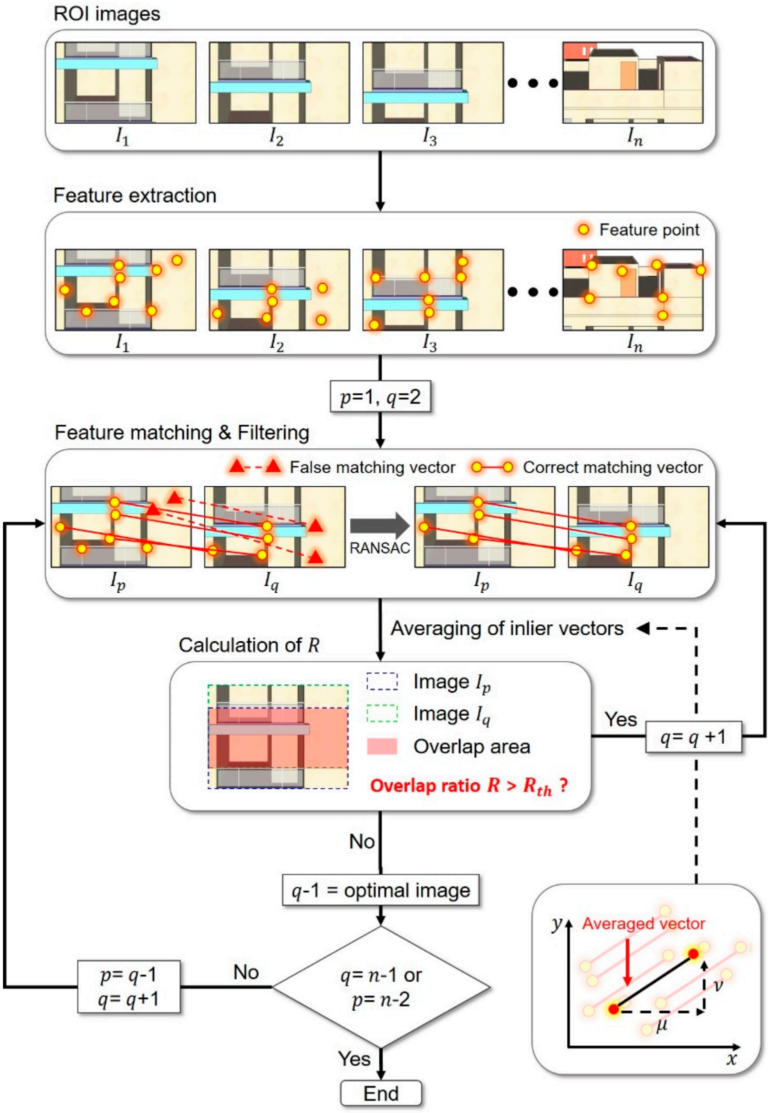


Figure 3. Optimal image selection algorithm: μ and ν are the averaged magnitude of the inlier vectors along the x and y directions. R is the overlap ratio. p and q are the iteration variables. n is the total number of the FOV images. I_p and I_q are the background-removed ROI images.

2.3. Mesh-Based Digital Image Stitching for Structural Exterior Map Establishment

In order to establish the precise structural exterior map, the local warp with a grid mesh is often used. In this study, a mesh-based digital image stitching method, called natural image stitching with the global similarity prior (NISwGSP), was employed [21]. Once the optimal images (I_j and I_{j+1}) were selected in Step 2, the homography matrix $\mathbf{H}$

($\mathbf{H} \in \mathbb{R}^{3\times3}$), which is reshaped from $\hat{h}$, was estimated between each optimal image using the following equation:

$$\hat{h} = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{N} ||a_i h||^2, \tag{3}$$

$$h = [h_1\ h_2\ h_3]^{\mathrm{T}},\ a_i = \begin{bmatrix} -x_i & -y_i & -1 & 0 & 0 & 0 & x_i x_i' & x_i' y_i & x_i' \\ 0 & 0 & 0 & -x_i & -y_i & -1 & x_i y_i' & y_i y_i' & y_i' \end{bmatrix}$$

where $[x_i\ y_i]$ and $[x_i'\ y_i']$ are the matched feature points of I_j and I_{j+1}, respectively. N is the number of the matched feature points, and h_m is the m_{th} row components of $\mathbf{H}$. The elements of $\mathbf{H}$ was be obtained using direct linear transformation (DLT), and the solution became the least significant right singular vector of $\{a_i\}_{i=1}^{N}$. Given $\mathbf{H}$, an arbitrary pixel position X_* on I_j was warped to the pixel position X_*' on I_{j+1}, using location dependent homography matrix $\mathbf{H}_*$ [22].

$$\widetilde{\mathrm{X}}_*' = \mathbf{H}_* \widetilde{\mathrm{X}}_*,\ \mathbf{H}_* = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{N} ||\omega_*^i a_i h||^2, \tag{4}$$

$$\omega_*^i = exp\left(-\frac{||\mathrm{X}_* - \mathrm{X}_i||^2}{\rho^2} \right)$$

where $\widetilde{\mathrm{X}}_*$ and $\widetilde{\mathrm{X}}_*'$ are the homogeneous coordinates of X_* and X_*', and ω_*^i is the scalar weight. ρ is the scale parameter, and X_i is the position of extracted feature point in I_j. Similarly, Equation (4) was solved by using DLT. However, solving Equation (4) with respect to all pixel position X_* on I_j, is not effective in terms of computational cost, because neighboring pixel positions often produce the same $\mathbf{H}_*$. Thus, I_j and I_{j+1} were divided into the mesh composed of $C_x \times C_y$ cells [23]. For each cell, the center point was chosen as X_*, and all pixels within the cell were warped by the same $\mathbf{H}_*$ as shown in Figure 4. Finally, the structural exterior map was constructed through mesh optimization and image mapping [21].

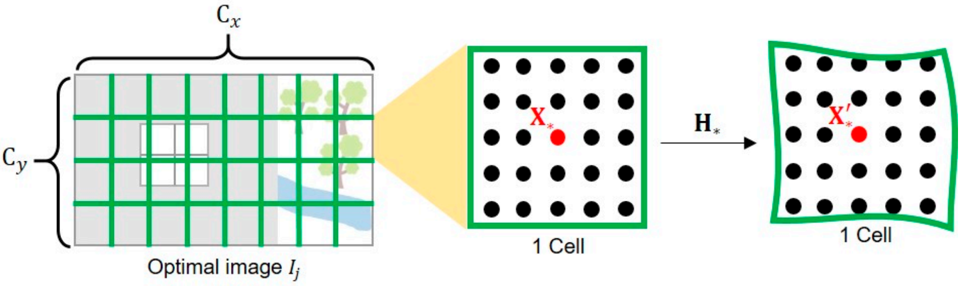


Figure 4. Mesh-based local warp: C_x and C_y are the number of cells divided from I_j along the x and y directions, X_* and X_*' are the arbitrary pixel positions on I_j and I_{j+1}.

3. Experimental Validation

The proposed technique was experimentally validated using a vision camera mounted-UAV at an in-situ 18 story building. The overall test procedures were as follows. First, the vision camera mounted-UAV scanned the target structure along a predefined scanning path to acquire the spatially continuous FOV images. Then, the automated background removal and the optimal image selection procedures were sequentially conducted. Finally, the structural exterior map was constructed using the mesh-based image stitching method. To show the superiority of the proposed technique, the test results were compared with the raw digital image ones.

Figure 5a,b show the target building and digital camera (L1D-20c of Hasselblad) mounted-UAV, Mavic 2 of DJI, used in this study. To evaluate the feasibility of the proposed technique, the UAV scanned along right-hand-side edge of the target structure, while the working distance of 4 m between the target ROI and UAV was kept. The digital images were obtained with the resolution of 3840 × 2160 pixels and 30 frames per second in video format.

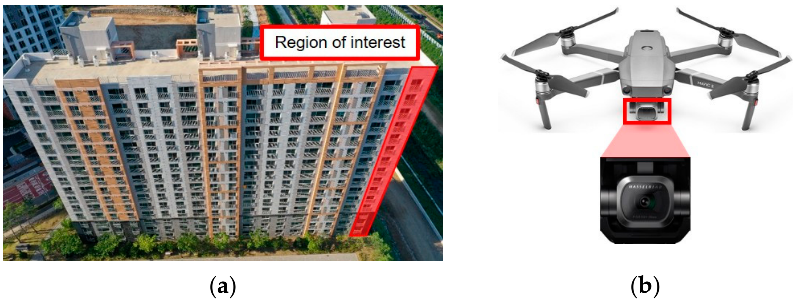


Figure 5. Experimental setup: (**a**) Front view of target structure with region of interest and (**b**) digital camera mounted-UAV.

Figure 6 shows the representative background removal results. The depth image of Figure 6b was obtained by using Monodepth2 from the raw FOV image of Figure 6a. Although the depth image contained the undesired noise components caused by scenery complexity such as window, target structure's texture and background objects, the ROI boundary between the target structure and the background objects was successfully extracted using Equation (1), as shown in Figure 6c. Then, the background region displayed in the right-side across the edge boundary on Figure 6c was removed, and the only ROI successfully remained as shown in Figure 6d. In order to quantitatively evaluate the accuracy of the depth image-based background removal results, a pixel-level error ratio between the resultant ROI image and its ground truth shown in Figure 6e was calculated.

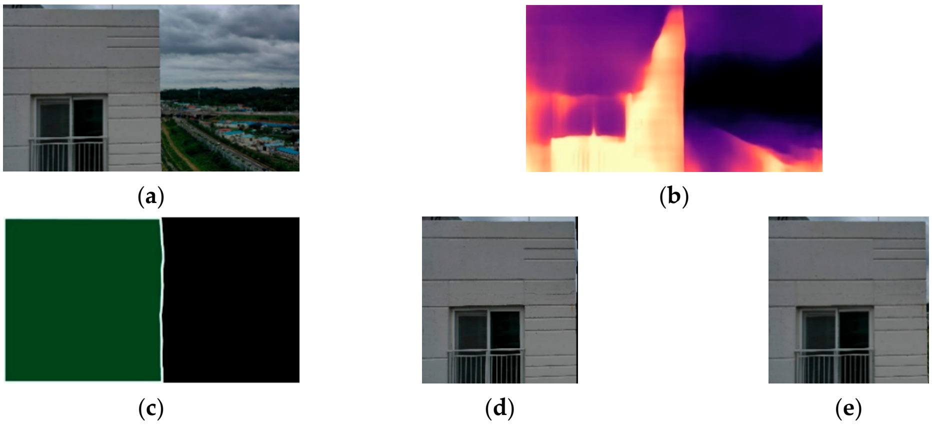


Figure 6. Automated background removal results: (**a**) FOV image, (**b**) depth image, (**c**) ROI determination by boundary extraction, (**d**) resultant ROI image and (**e**) ground truth image of ROI.

Figure 7 shows the error ratio results obtained from all the background removed-ROI images. The averaged error ratio along the entire ROI images turns out around 0.75%, which reveals that the proposed algorithm had over 99% accuracy for background removal and is acceptable for the subsequent structural exterior map construction.

Next, the representative overlap ratio calculation results are shown in Figure 8. In Figure 8a, 14,421 and 13,850 numbers of image features were extracted from I_1 and I_7, respectively. The extracted features were initially matched as displayed in Figure 8a, and only inlier vectors were then remained as shown in Figure 8b. Subsequently, 66.49% of R was calculated using Equation (2) as shown in Figure 8b.

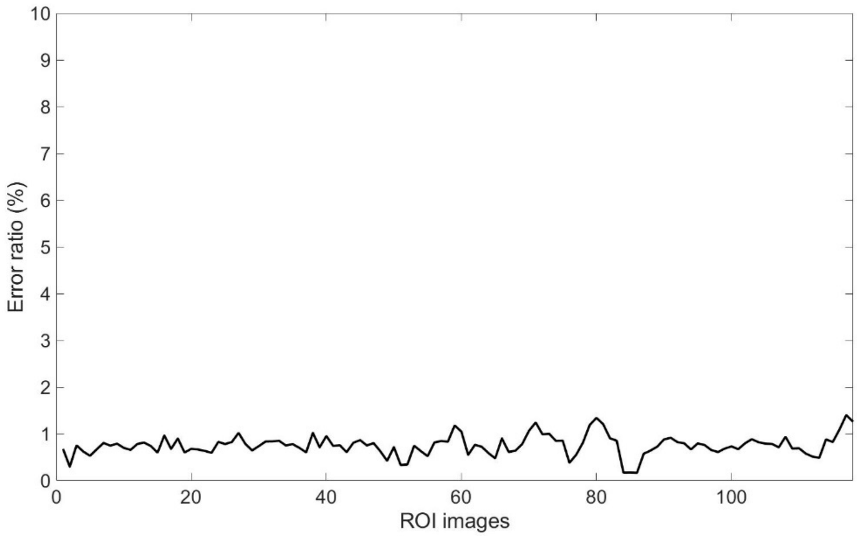


Figure 7. Error ratio between the resultant ROI images and their ground truth images.

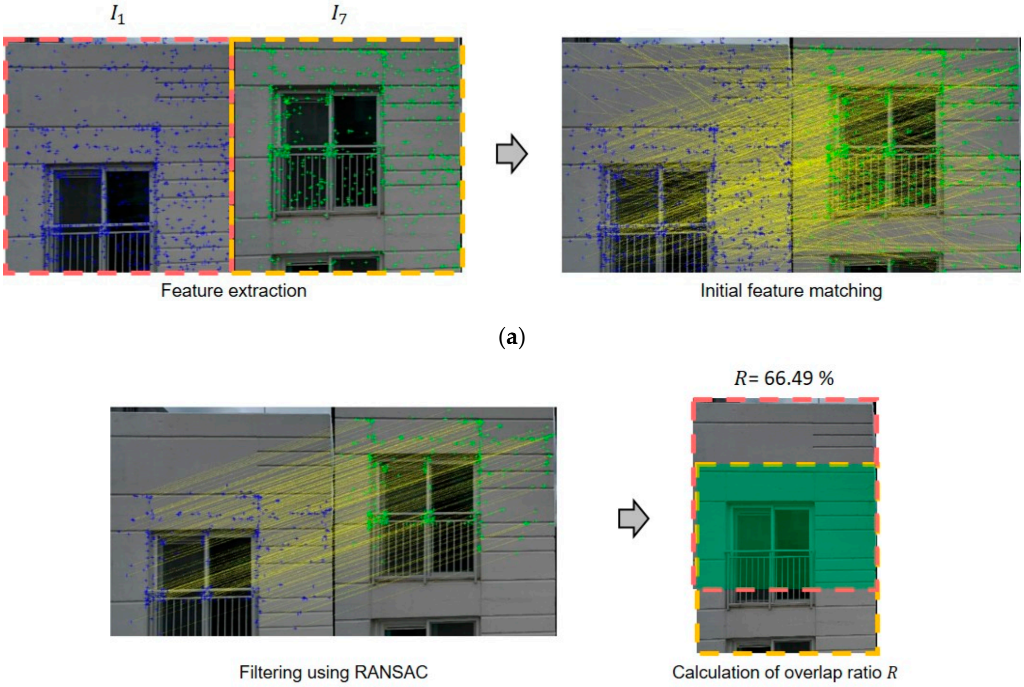


Figure 8. The representative overlap ratio calculation results when p = 1 and q = 7: (**a**) Feature extraction and initial feature matching and (**b**) inlier vectors filtered by using random sample consensus (RANSAC) and calculation of R.

In order to properly stitch the digital images, it was generally recommended that R has greater than 50% [24]. In this study, R_{th} of 80% was used due to local image features such as window and target structure's texture, which were extremely repeated in the sequential image data and may increase the false feature matching. As shown in Figure 9, R straightforwardly decreased as q increased when p = 1, R becomes 82.9% and 77.2% corresponding to q = 4 and q = 5, respectively, as shown in Figure 9. Therefore, I_4 was selected as the optimal image with respect to p = 1. Once I_4 was selected, the next optimal image can be similarly selected from I_4. This procedure was repeated until the entire optimal images were determined for structural exterior map establishment.

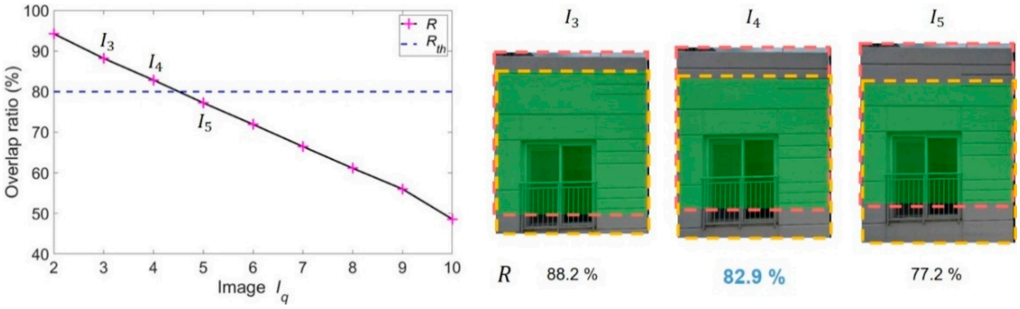


Figure 9. Optimal image selection results when p = 1.

Figure 10 shows the raw and selected optimal images. In total, 118 raw FOV images, which were acquired from UAV's close-up scanning, contained various background objects as shown in Figure 10a. On the other hand, only 35 background removed-ROI images were extracted through automated background removal and optimal image selection algorithms as shown in Figure 10b. To validate the compatibility of these two images' data, the structural exterior maps were constructed and compared.

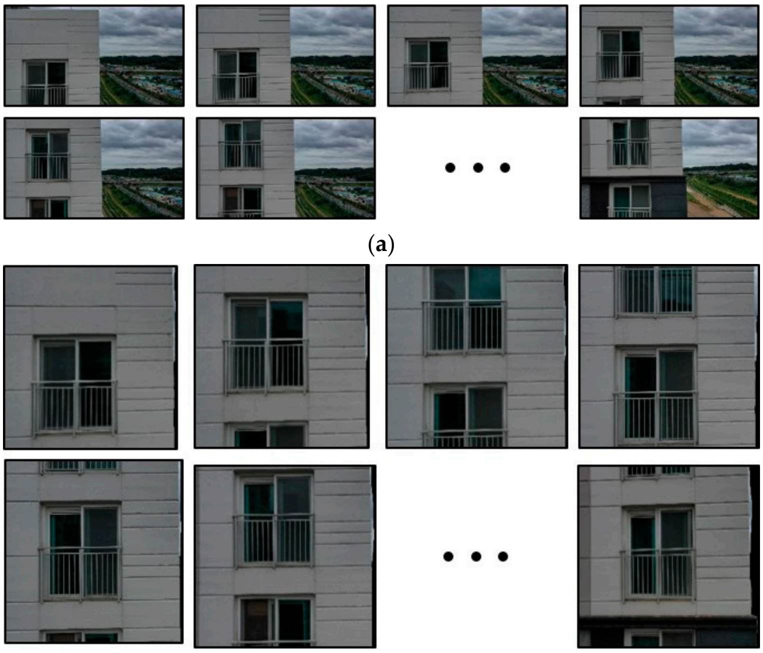
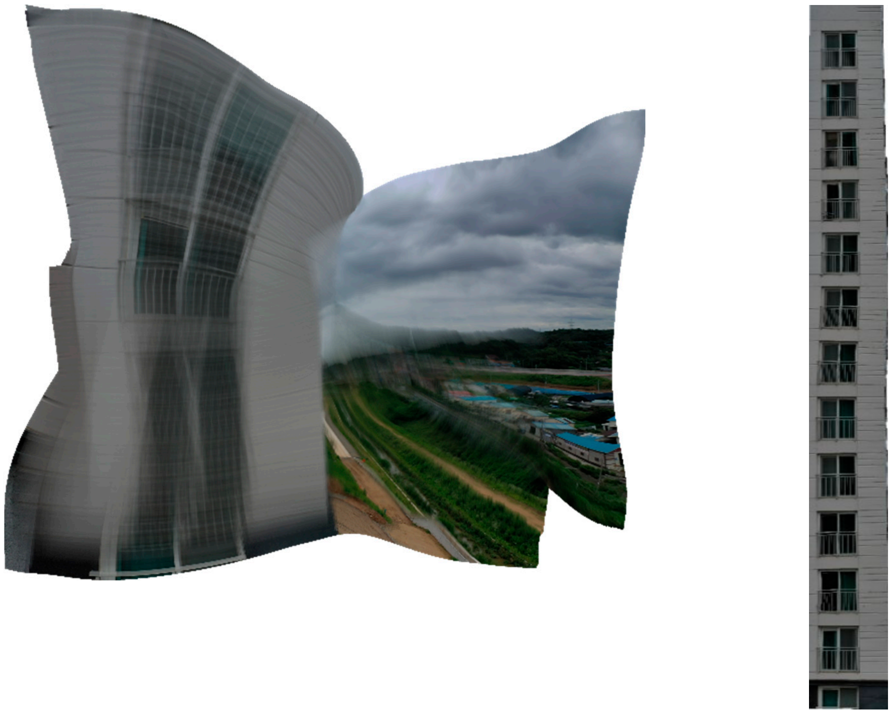


Figure 10. The representative raw and optimal images: (**a**) 118 raw images and (**b**) the corresponding 35 optimal images.

Figure 11 compares structural exterior map establishment results between using the raw and optimal images. When the raw images were used, it failed to construct the structural exterior map due to serious distortion and ghosting effects, as shown in Figure 11a. It turned out that the images were stitched according to the background features, because the background objects had more distinguishable image features than the ROI ones. On the other hand, the structural exterior map using the optimal images was properly constructed without distortion and ghosting effects on the ROI, as shown in Figure 11b.



(**a**) (**b**)

Figure 11. Comparison of structural exterior map construction results using (**a**) raw images and (**b**) optimal images.

4. Discussion

Since this work was not to newly train a deep learning network, but to employ a suitable pre-trained and pre-validated deep learning model for cost-effective and automated background removal, the suitability of the employed Monodepth2 was additionally tested. To show the effectiveness of depth estimation results, the relative mean depth values between the ROI and background regions within FOV images were calculated as shown in Figure 12. The higher depth values mean the closer region from the digital camera mounted on the UAV. It can be easily observed that the ROIs' relative mean depth values were consistently higher than the background ones without any overlap between them. This means that the relative depth values were successfully estimated using Monodepth2, and the corresponding ROIs were properly extracted. Here, Monodepth2 works well because this work is under assumption that the structural ROI is much closer than the background objects from the UAV. However, the depth estimation errors may increase when the working distance between the ROI and background objects is similar.

In addition, since one of the critical obstacles to construct the large-scale structural exterior map is the computational cost, the computational time was compared between the raw image and optimal image cases in Table 1. The structural exterior map establishment took 9 h 13 min 47 s using 118 raw images, while it just conducted within 1 h 18 min 57 s using 35 optimal images. Note that the computational times were estimated when it comes to CPU of Intel$^{®}$ Xeon E5-2630 v4 with 64 gigabytes of memory. The optimal image case includes the optimal image selection time of 40 min 49 s, which is about 51.69% out of total 1 h 18 min 57 s. These results indicate that the proposed technique is critical for structural exterior map construction performance, and also can extremely reduce the computational cost of almost 85.7%.

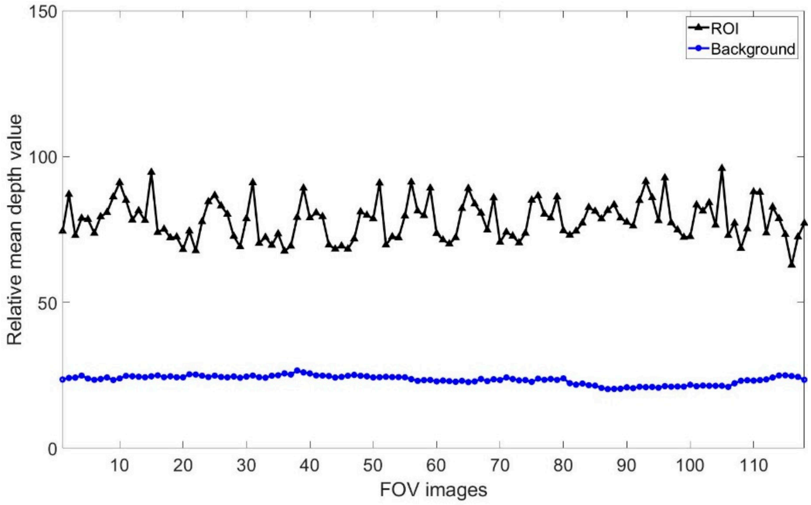


Figure 12. Relative mean depth values between the ROI and background regions.

Table 1. Comparison of the computational costs for structural exterior map establishment.

	Raw Images	Optimal Images
Number of images	118	35
Computational time	9 h 13 min 47 s	1 h 18 min 57 s

5. Conclusions

This paper proposed a deep learning-based automated background removal technique for structural exterior image stitching. The effectiveness of the proposed technique was experimentally demonstrated through in-situ high-rise building structure tests with a vision camera mounted-unmanned aerial vehicle (UAV). Then, the test results were compared with the structural exterior map constructed using non-treated raw images. The validation test results obtained using the proposed technique revealed that the structural exterior map was properly constructed without distortion and ghosting effects. On the other hand, the structural exterior map using raw images without any image processing showed serious distortion and ghosting effects on region of interest. Furthermore, the proposed technique constructed the precise structural exterior map with a computational cost reduction of 85.7% versus the raw image case. Although the proposed technique can highly depend on the accuracy of depth estimation, it can be one of the promising tools for automatically establishing structural exterior maps using UAV's close-up scanned images with low computational cost. As a follow-up study, an advanced image stitching algorithm that is robust against test environmental variation is now being developed. Furthermore, a deep learning-based automated structural damage detection algorithm incorporated with the precise structural exterior map will be developed to extend the applicability of the proposed technique.

Author Contributions: M.S.K. and Y.-K.A. conceived and designed the experiments; M.S.K. performed the validation and visualization of results; M.S.K. and Y.-K.A. wrote the paper; Y.-K.A. supervised the research. Both authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by a Grant (No. 21CTAP-C152120-03) from Technology Advancement Research Program (TARP) funded by the Ministry of Land, Infrastructure and Transport of the Korea Government.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors give special thanks to Batangsak ENG for providing the testbed.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kang, D.; Cha, Y.J. Autonomous UAVs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging. *Comput. Aided Civ. Inf.* **2018**, *33*, 885–902. [CrossRef]
2. Bae, H.; Jang, K.; An, Y.K. Deep super resolution crack network (SrcNet) for improving computer vision-based automated crack detectability in in situ bridges. *Struct. Health Monit.* **2020**, 1–15. [CrossRef]
3. Kerle, N.; Nex, F.; Gerke, M.; Duarte, D.; Vetrivel, A. UAV-based structural damage mapping: A review. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 14. [CrossRef]
4. Zhu, Z.; German, S.; Brilakis, I. Detection of large-scale concrete columns for automated bridge inspection. *Autom. Constr.* **2010**, *19*, 1047–1055. [CrossRef]
5. Jahanshahi, M.R.; Masri, S.F.; Sukhatme, G.S. Multi-image stitching and scene reconstruction for evaluating defect evolution in structures. *Struct. Health Monit.* **2011**, *10*, 643–657. [CrossRef]
6. Zhu, Z.H.; Fu, J.Y.; Yang, J.S.; Zhang, X.M. Panoramic image stitching for arbitrarily shaped tunnel lining inspection. *Comput. Aided Civ. Inf.* **2016**, *31*, 936–953. [CrossRef]
7. Yoon, S.; Gwon, G.H.; Lee, J.H.; Jung, H.J. Three-dimensional image coordinate-based missing region of interest area detection and damage localization for bridge visual inspection using unmanned aerial vehicles. *Struct. Health Monit.* **2020**, 1–14. [CrossRef]
8. Morgenthal, G.; Hallermann, N. Quality assessment of unmanned aerial vehicle based visual inspection of structures. *Adv. Struct. Eng.* **2014**, *17*, 289–302. [CrossRef]
9. Xie, R.; Yao, J.; Liu, K.; Lu, X.; Liu, Y.; Xia, M.; Zeng, Q. Automatic multi-image stitching for concrete bridge inspection by combining point and line feature. *Autom. Constr.* **2018**, *90*, 265–280. [CrossRef]
10. Won, J.; Park, J.W.; Shim, C.; Park, M.W. Bridge-surface panoramic-image generation for automated bridge-inspection using deepmatching. *Struct. Health Monit.* **2020**, *1*, 15. [CrossRef]
11. Jang, K.; An, Y.K.; Kim, B.; Cho, S. Automated crack evaluation of a high-rise bridge pier using a ring-type climbing robot. *Comput. Aided Civ. Inf.* **2021**, *36*, 14–29. [CrossRef]
12. Yang, T.; Li, J.; Yu, J.; Wang, S.; Zhang, Y. Diverse scene stitching from a large-scale aerial video dataset. *Remote Sens.* **2015**, *7*, 6932–6949. [CrossRef]
13. Bang, S.; Kim, H.; Kim, H. UAV-based automatic generation of high-resolution panorama at a construction site with a focus on preprocessing for image stitching. *Autom. Constr.* **2017**, *84*, 70–80. [CrossRef]
14. Bu, S.; Zhao, Y.; Wan, G.; Liu, Z. Map2DFusion: Real-time incremental UAV image mosaicking based on monocular SLAM. In Proceedings of the IEEE/RSJ International Conference of Intelligent Robots and Systems (IROS 2016), Daejeon, Korea, 9–14 October 2016; pp. 4564–4571.
15. Xin, Y.; Hou, J.; Dong, L.; Ding, L. A self-adaptive optical flow method for the moving object detection in the video sequences. *Optik* **2014**, *125*, 5690–5694. [CrossRef]
16. Supreeth, H.S.G.; Patil, C.M. Efficient multiple moving object detection and tracking using combined background subtraction and clustering. *Signal Image Video Process.* **2018**, *12*, 1097–1105. [CrossRef]
17. Fang, W.; Ding, Y.; Zhang, F.; Sheng, V.S. DOG: A new background removal for object recognition from images. *Neurocomputing.* **2019**, *361*, 85–91. [CrossRef]
18. Godard, C.; Mac Aodha, O.; Firman, M.; Brostow, G.J. Digging into self-supervised monocular depth estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2019), Seoul, Korea, 27 October–2 November 2019; pp. 3828–3838.
19. Otsu, N. A threshold selection method from gray level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]
20. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the International Conference on Computer Vision (ICCV 1999), Corfu, Greece, 20–25 September 1999; Volume 2, pp. 1150–1157.
21. Chen, Y.S.; Chuang, Y.Y. Natural image stitching with global similarity prior. In Proceedings of the European Conference on Computer Vision (ECCV 2016), Amsterdam, The Netherlands, 8–16 October 2016; pp. 186–201.
22. Zaragoza, J.; Chin, T.J.; Brown, M.S.; Suter, D. As-projective-as-possible image stitching with moving DLT. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013), Portland, OR, USA, 25–27 June 2013; pp. 2339–2349.
23. Schaefer, S.; McPhail, T.; Warren, J. Image deformation using moving least squares. In Proceedings of the International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2006), Boston, MA, USA, 30 July–4 August 2006; pp. 533–540.
24. Chen, Q. VR: An image-based approach to virtual environment navigation. In Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH1995), Los Angeles, CA, USA, 29–38 August 1995.