# A Method for Removing Shadows from Photos Taken with a Drone and Stitching the Photos Together

YuJie Wu, Megumi Wakao, Naoki Morita
School of Information and Telecommunication Engineering
Tokai University
Tokyo, Japan
e-mail: 2cjnm024@mail.u-tokai.ac.jp, 9bjt2103@cc.u-tokai.ac.jp, morita@tokai.ac.jp

Kenta Morita
Faculty of Medical Engineering
Suzuka University of Medical Science
Mie, Japan
e-mail: morita@suzuka-u.ac.jp

***Abstract*—We are using a drone to obtain aerial views of grape farms to check the growth of grapevines. The shadows cast by the branches (cordons) in captured images differ depending on the location of the drone when the images were captured. The problem with stitching images of cordons with shadows together is that the cordons often appear disjointed at the edges of individual images. To solve this problem, we propose a method for removing elements other than cordons in the images captured from above, so that cordons only can be extracted from the images. We tested the effectiveness of this method using images in which the cordons were disjointed at the edges of the images after stitching by conventional methods.**

***Keywords- Aerial image stitching; feature point; segment***

## I. INTRODUCTION

Viticulture in Japan is practiced by cultivating grapevines along trellises above the vines. Figure 1 shows a photograph of a vineyard, with branches(cordons) extending outward from the main trunk of the grapevines along a trellis positioned about 2 m above the ground. In order to harvest tastier grapes and manage nutrients effectively, it is necessary to monitor the number of shoots growing from the cordons [1]. However, since it is difficult to check cordon growth from the ground, we sought to use drones to check the state of the cordons from the air.

The characteristics of a grapevine are described here using aerial photographs obtained using a drone. Figures 2 and 3 show photographs of a grapevine taken from a height of approximately 10 m above the vines in a vineyard. The area of the image shown in Figure 3 was offset from the center of the area shown in Figure 2 by approximately 2 m. As indicated by the orange boxes in Figures 2 and 3, the lower half of Figure 2 and the upper half of Figure 3 show the same area. The feature indicated by the yellow arrow in Figure 2 is a cordon, the gray-white lines that are connected to a cordon are referred to as arms, and the brown line is also a branch. The black lines on the ground are the shadows cast by the vine. However, due to differences in the photographic angles, the appearance of cordon shadows is different. For example, the shadows of the same feature indicated by the blue arrows in Figures 2 and 3 are contained by the red circles in the two figures. Examination of the red circles shows that the distance between the cordon and the shadow is shorter in Figure 2 than in Figure 3. As a result of these differences, conventional image stitching methods could not be used to combine the images.

The aim of this study was to combine two images so that the cordons at the edges of adjacent images are correctly stitched together. Here we describe the stitching process that we developed to overcome the problems of disjointed cordons between images using the aerial photographs that we took.

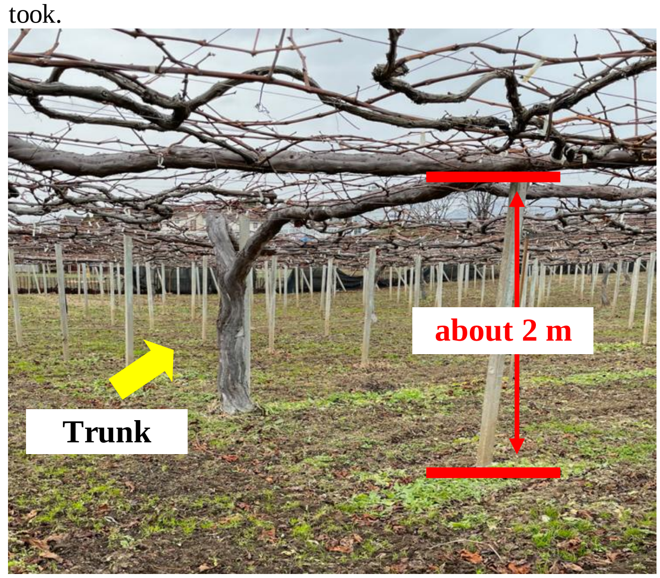


Figure 1. Photograph of a grapevine in a vineyard in Japan.

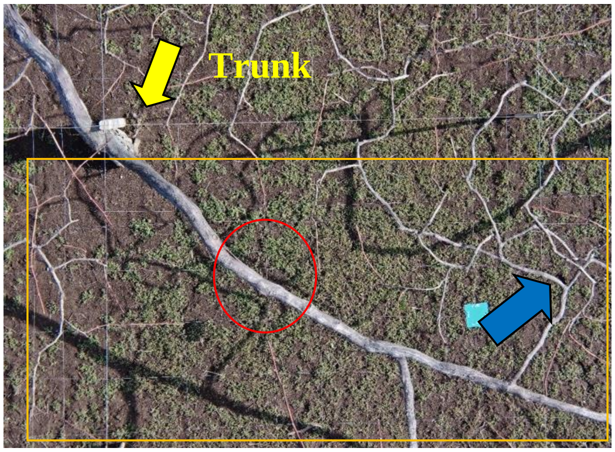


Figure 2. Aerial photograph of part of a grapevine to show the distribution of cordons and their shadows.

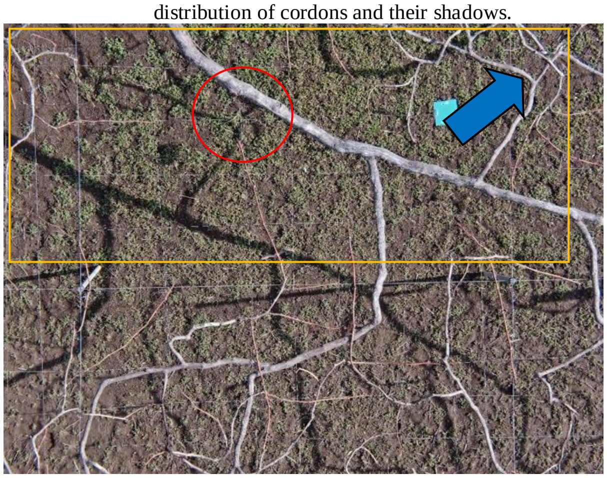

Figure 3. As in Figure 1, except that the image was taken so that the center of the images is offset by approximately 2 m.

In Section II we will explains the difficulties encountered with combining aerial photos and describes previous studies on the topic.

## II. DIFFICULTIES IN COMBINING IMAGES

Images that were combined using photographs taken from slightly different locations showed that cordons in the image appeared cut off or disjointed. For example, Figure 4 shows the result of combining Figures 2 and 3 with the projective transformation using a Scale-Invariant Feature Transform (SIFT) descriptor. As shown by the red box in Figure 4, stitching of the two images resulted in the cordons and other features appearing disjointed.

In order to investigate the cause of these disjointed areas, we generated key points and descriptors using SIFT algorithm [2]. There are many methods for extracting feature values [3]-[5]. SIFT features are local, based on the appearance of the object at particular points of interest, and they are invariant to image scale and rotation. Moreover, the SIFT algorithm has a high recognition rate for feature points in regions with significant color shifts in the peripheral area. It is therefore considered to be particularly well suited for stitching together aerial images of orchards obtained in this study.

Figure 5 shows an image obtained by applying the SIFT algorithm to the central part of Figure 3. The circles in Figure 5 indicate SIFT key points and descriptors. The radius of the circle represents the feature scale. After checking the SIFT key points in Figure 4, the circles representing the feature points were generated for the entire image. Many of the key points identified were associated with features above the cordons. In addition, several circles were also associated with shadows, some of which were even larger than those associated with cordons and arms. This means that shadows are also being treated as features to be used for combining the image, i.e., just like cordons. Consequently, corrections need to be made to negate the effect of shadows when combining images of cordons.

Numerous studies have been conducted to correct distortions in terrain in aerial images [6]-[11], especially for buildings standing on the ground. However, the cordons are not attached to the ground. Orthoimages created from aerial photographs using GPS require reference points on the ground for correction. However, the “floating” nature of cordons makes it difficult to set reference points without including shadows, leading to significant image distortion. Conventional methods are thus not suitable for analyzing “floating” cordons.

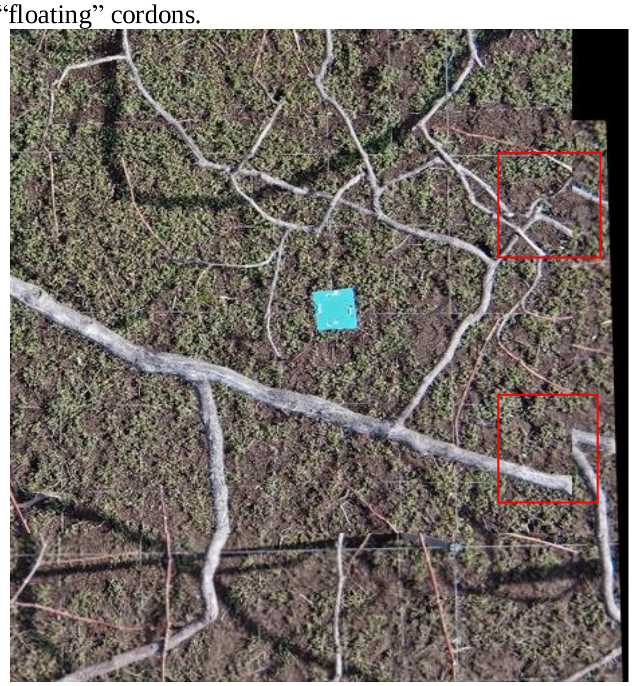

Figure 4. Image showing the photographs shown in Figures 1 and 2 stitched together using a traditional method.

 ISBN: 978-1-68558-079-7 

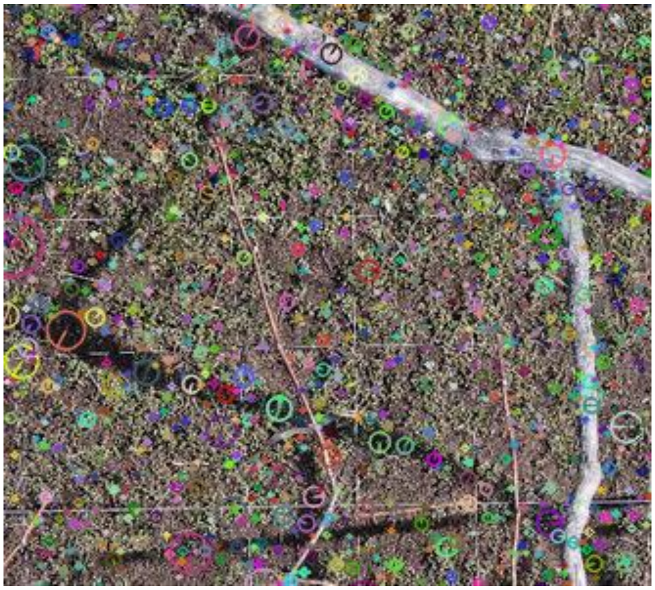

Figure 5. Transformed image showing key points and descriptors.

In Section III we will discusses the proposed solution.

## III. PROPOSED METHOD

If the cordons and their shadows overlap each other when images are stitched, then the distortion increases. Therefore, we developed a method that ignores the effect of cordon shadows and stitches the photos so that only the cordons overlap each other.

Figure 6 shows a general overview of an image stitching procedure that uses feature points extracted from images to combine them. Figure 7 shows our proposed method. Our method first segments out the cordons from img1 and img2 to produce the seg1 and seg2 images that contain only cordons. Then, we apply the feature points for cordons extracted from seg1 and seg2 to img1 and img2 for image stitching. This method allows us to eliminate the effect of cordon shadows on image stitching and achieves seamless stitching of cordons at the edges of the images.

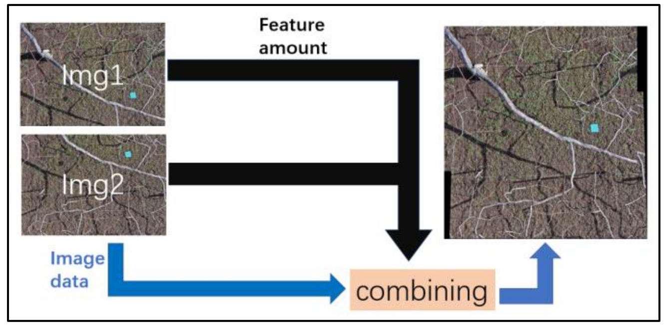


Figure 6. Conventional method for stitching images.

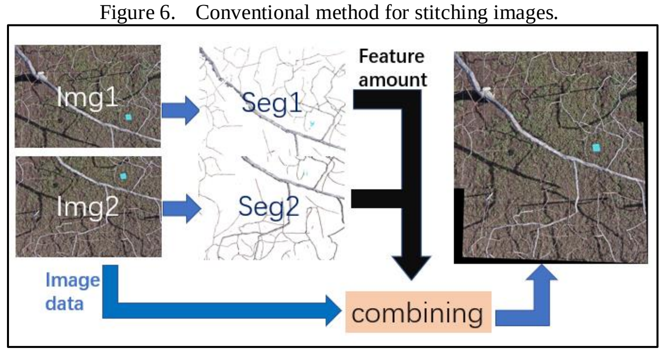


Figure 7. Proposed method for stitching images.

In Section IV we will describes the developed image-stitching system.

## IV. SYSTEM DEVELOPMENT

To correct areas where the cordons are misaligned at the edges of aerial photographs, we developed an aerial photograph correction system. This alignment procedure can be divided into three steps.

Step1: Remove non-vine elements from aerial photos and create images that only show the silhouettes of cordons.

Step2: Extract feature points from the silhouettes.

Step3: Apply projective transformation to aerial photos based on the extracted feature points and stitch them together.

Using these three steps we created images that show only the silhouettes of the cordons. To remove the non-vine elements, we performed semantic segmentation using SegNet [12], a deep convolutional encoder-decoder architecture for image segmentation. We trained the network with 4000 images of cordons and segmented them with a network trained using 100,000 iterations. The segmentation results are shown in Figure 8. The reddish-brown regions shown in the figure represent the silhouettes of cordons.

Feature points and descriptors were extracted from the silhouette images using the SIFT algorithm. Figure 8 shows the SIFT feature points and descriptors obtained within the same region in Figure 5.

A projection transformation matrix was then computed based on the feature points extracted from the silhouette image, and the aerial photographs were merged. We employed the projection transformation function in the OpenCV library and stitched the images together based on the extent of matching among feature points.

 ISBN: 978-1-68558-079-7 

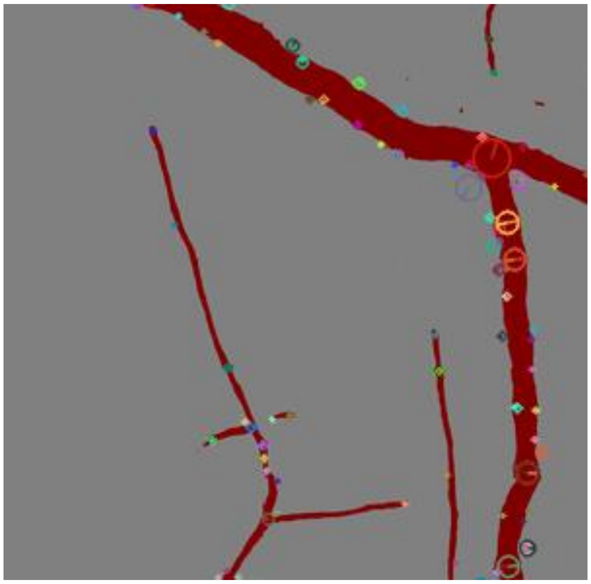

Figure 8. Results of image segmentation

In Section V we will describes the experiments conducted and presents the results and considerations, and the conclusions are given in Section VI.

## V. Experiments And Considerations

In this section, we clarify the effectiveness of the proposed method. We compared the synthetic precision between the traditional method of stitching after extracting feature points from aerial images, and the method involving the extraction of feature points and stitching using images of cordon silhouettes.

Figure 9 shows one of the experimental results. It can be seen that the region enclosed in Figure 4 appears to be fully connected. As shown in Figure 4, the upper red box indicates a misalignment of four or more cordon widths, while the lower red box indicates a misalignment of three cordon widths. However, these defects are fully corrected in Figure 7.

Table I shows the measurement results. Using the conventional method, 53 cordons out of 60 were disjointed at the edges of images, whereas using the proposed method, only one cordon was disjointed. In the conventional method, one cordon was disjointed by a distance of twice the thickness of the original cordon, but using the proposed method, the cordon was disjointed by a distance of only one cordon thickness.

TABLE I. Measurement results

| **Disjointed distance** | Traditional method | Proposed method |
|---|---|---|
| ***Nothing*** | 7 | 59 |
| ***One cordon*** | 17 | 1 |
| ***Two cordons*** | 16 | 0 |
| ***Three cordons*** | 10 | 0 |
| ***More*** | 10 | 0 |

Compared with the conventional method, the proposed method improved the synthetic precision. Using the proposed method, only one cordon was not perfectly stitched. We considered that the reason for this discontinuity along the cordon was because, in terms of area, too little of the cordon was shared between the two original images being stitched.

From the experimental results, the method developed for recognizing cordons from aerial images and extracting feature points based on shared features was considered to be effective for connecting cordons seamlessly.

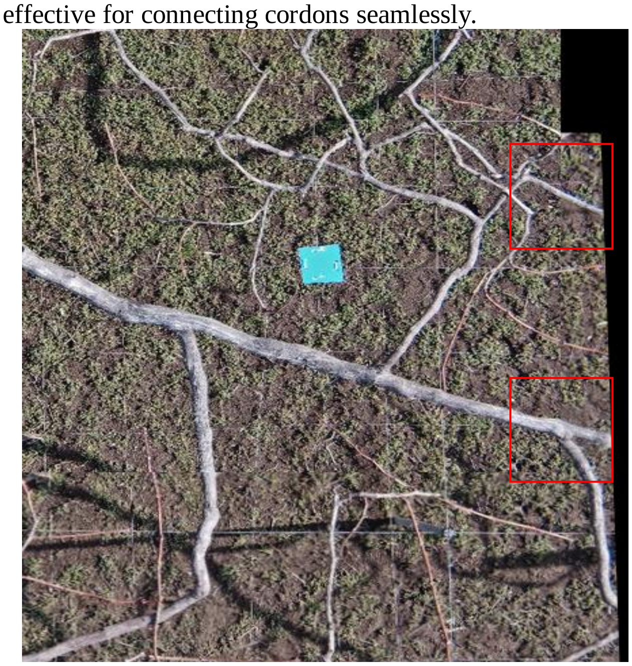

Figure 9. Image produced using the photographs shown in Figures 1 and 2 stitched together using the proposed method.

## VI. Conclusion

The aim of this research was to accurately overlap cordons at the edges of aerial images being stitched. When combining photographs taken at two locations, cordons are often disjointed where the images are stitched. We therefore developed a method to remove elements other than cordons from aerial images and then extract feature points from images to show only the silhouettes of cordons. The findings showed that, compared with previous methods, the proposed method has a higher rate of synthetic precision for the stitching of cordons at the edges of aerial images.

In the future, we would like to stitch all of the aerial photographs so that the cordons overlap.

## References

[1] Nirasaki-shi, Yamanashi, CUPOD FARM [Online]. Available from: https://www.cupidfarm.co.jp/ [accessed: 2023-03-12]

[2] D. G. Lowe, "Object recognition from local scale-invariant features," Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999, pp. 1150-1157 vol.2, doi: 10.1109/ICCV.1999.790410.

[3] S. Leutenegger, M. Chli and R. Y. Siegwart, "BRISK: Binary Robust invariant scalable keypoints," 2011 International Conference on Computer Vision, Barcelona, Spain, 2011, pp. 2548-2555, doi: 10.1109/ICCV.2011.6126542.

[4] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," 2011 International Conference on Computer Vision, Barcelona, Spain, 2011, pp. 2564-2571, doi: 10.1109/ICCV.2011.6126544.

[5] Alcantarilla, P.F., Bartoli, A., Davison, A.J. (2012). KAZE Features. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds) Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol 7577. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33783-3_16

[6] Guoqing Zhou, "Urban large-scale orthoimage standard for national orthophoto program," Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05., Seoul, Korea (South), 2005, pp. 1214-1217, doi: 10.1109/IGARSS.2005.1525336.

[7] CAO Jian, LI Kan, GAO Chun-xiao, LIU Qiong-xin. Application of Local Features in Aerial Image Mosaic[J]. Journal of University of Electronic Science and Technology of China, 2013, 42(1): 125-129. doi: 10.3969/j.issn.1001-0548.2013.01.026

[8] Yamazaki Y., Itaya T. (2010) An attempt to create a wide-area ortho-mosaic image from analog aerial photographs. Journal of the Japan Society of Forest Survey, 32(4), 197-200.

[9] Kim, M., Kim, J., & Lee, D. (2018). A study on orthoimage generation method using unmanned aerial vehicle photogrammetry. Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, 36(6), 535-544.

[10] Guo, W., Wu, J., & Xie, X. (2019). Study on Ortho-image Mosaicing of Super-long-span Cable-stayed Bridge Based on Multi-sensors. Journal of Sensors, 2019.

[11] Ayhan, E., & Uysal, M. (2017). Estimation of stand height and aboveground biomass of Mediterranean pine forests using QuickBird-2 orthoimages. International Journal of Remote Sensing, 38(12), 3655-3677.

[12] Badrinarayanan, V., Kendall, A., Cipolla, R., “SegNet:A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.” IEEE Trans. Pattern Anal. Mach. Intell. Vol 39, No.12, pp.2481–2495, Dec.2017

 ISBN: 978-1-68558-079-7 