Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography
Vol. 38, No. 2, 165-172, 2020
https://doi.org/10.7848/ksgpc.2020.38.2.165

ISSN 1598-4850(Print)
ISSN 2288-260X(Online)
Original article

A Fast Image Matching Method for Oblique Video Captured with UAV Platform

Byun, Young Gi[1)] · Kim, Dae Sung[2)]

Abstract

There is growing interest in Vision-based video image matching owing to the constantly developing technology of unmanned-based systems. The purpose of this paper is the development of a fast and effective matching technique for the UAV oblique video image. We first extracted initial matching points using NCC (Normalized Cross-Correlation) algorithm and improved the computational efficiency of NCC algorithm using integral image. Furthermore, we developed a triangulation-based outlier removal algorithm to extract more robust matching points among the initial matching points. In order to evaluate the performance of the propose method, our method was quantitatively compared with existing image matching approaches. Experimental results demonstrated that the proposed method can process 2.57 frames per second for video image matching and is up to 4 times faster than existing methods. The proposed method therefore has a good potential for the various video-based applications that requires image matching as a pre-processing.

Keywords : Video Image Processing, Feature Point Extraction, Feature Image Matching, Outlier Removal, Sensor Modeling

1. Introduction

Recently, global interest in the development of various unmanned systems such as UAV (Unmanned Aerial Vehicle), UGV (Unmanned Ground Vehicle) and USV (Unmanned Surface Vehicle) has increased owing to the development of advanced science and technology. Especially, UAV system is widely used across the world civilian, commercial as well as military application because of its low cost and ease of operation (Zhuo *et al*., 2017). Utilization technologies based on UAV vision system are constantly being developed, such as UAV navigation in GPS (Global Positioning System)-denied environments and target geo-localization (Robert *et al*., 2017).

In recent year, some research have shown that the effectiveness of UAV vision sensor in the intelligent mobility services and visual object tracking (Inder *et al*., 2019; Xue *et al*., 2018; Ke *et al*., 2019). Video Frame-matching is one of the most important issues to ensure reliable performance of video-based geo-localization because they have to track and use the same correspondence point information on video image sequences. Generally, image matching method includes a feature point extraction stage, followed by feature matching, and outlier removal stage to reduce the number of mismatching points. Corner points are generally used as feature points for image matching since they are sufficient and uniformly distributed across the image domain. Several

Received 2020. 04. 08, Revised 2020. 04. 18, Accepted 2020. 04. 29
1) Spatial Information Research Institute, Korea Land and Geospatial Informatix Corp. (E-mail: kko071@snu.ac.kr)
2) Corresponding Author, Member, Agency for Defence Development (E-mail: mutul94@add.re.kr)

methods for extracting feature points, such as Harris corner detector derived by local autocorrelation function (Harris *et al*., 1988), SUSAN (Smallest Univalue Segment Assimilating) corner detector (Smith *et al*., 1997), and KLT (Kanade-Lucas-Tomasi) corner detector (Lucas *et al*., 1981), have been proposed and shown promising results in various image matching applications (Tissainayangam and Suter, 2004).

Correlation coefficient-based method, which examine the similarity of pixel value in template, is widely used as feature matching method. To overcome the disadvantage of this method in which matching performance deteriorates around the edge, various methods have been proposed to variably adjust the template size or to integrate image feature information in a similarity calculation process (Kanade *et al*., 1994; Heo *et al*., 2011). SIFT (Scale-Invariant Feature Transform) algorithm with robust characteristics for image scale and rotation has been proposed, and it has shown superior image matching performance compared to conventional matching methods (Lowe *et al*., 2004). The SIFT-based method has a disadvantage that it takes a lot of computation time in the process of building the SIFT descriptor. To overcome this drawback, SURF (Speed Up Robust Features) algorithm has been recently proposed, which improves the computation speed dramatically while maintaining the robustness of the SIFT method (Herbert *et al*., 2013).

When extracting matching points among the feature points, inappropriate matching points may also be detected together. This mismatching points can be identified through a homography correspondence between two images. As a method for estimating a homography parameter using matching points, several methods have been proposed, such as DLT (Direct Linear Transform) technique based on least squares method (Hartley and Zisserman, 2003) and a LMedS (Least Median of Squares) technique (Zhang *et al*., 1995). However, these methods have a problem that is greatly affected by the size or ratio of errors in the homography estimation process. For overcoming this problem, a RANSAC (RANdom SAmple Consensus) algorithm has been proposed to enable robust homography parameter estimation even when the outlier ratio is very high (Fischler and Bolles, 1981). The RANSAC algorithm extracts the optimal matching points through an iterative process while randomly sampling the smallest corresponding points required to determine the homography parameter among all matching pairs extracted in the feature point matching process. A preemptive RANSAC technique, which consists in generating a fixed number of hypothesis and extracting matching points through a suitability determination process using breadth first search, has been proposed for real-time process (David, 2003).

The purpose of this paper is to develop fast matching methodology that can be used as a pre-processing of UAV video geo-localization. To do this, feature points are first extracted using the Harris corner detector, and then the matching points are extracted using the normalized NCC with mutual consistency test. In addition, the computational efficiency of NCC algorithm was improved using the integral image. Furthermore, a triangulation-based outlier removal algorithm are developed to extract more robust matching points among the initial matching points. We compared the proposed and conventional methods using oblique video image, and the experimental results demonstrated that the proposed method enable fast image matching while maintaining matching accuracy compared with existing methods.

2. Methodology

2.1 Feature Extraction

The feature point used for image matching should be easy to distinguish in both images and has a high probability of detection, even if there are geometric and radiometric differences between the images. In this study, the feature point was detected using the Harris corner detector, which is typically used among the feature point detectors. The Harris corner detector is based on computing the corner response function of each pixel in image. The corner response function is calculated by the autocorrelation matrix as follow Eq. (1):

$$M = W_G(x,y) * \begin{pmatrix} I_x^2 & I_{xy} \\ I_{xy} & I_y^2 \end{pmatrix} \tag{1}$$

where, $W_G(x,y)$ represents a Gaussian filtering function. I_x and I_y are first order partial derivatives in x and y directions with the image, respectively. I_{xy} is the mixed first order partial derivative in x and y directions.

The corner response of each pixel is calculated by using the function defined as Eq. (2) using the local structure matrix.

$$R(x,y) = \det(M) - k(trace\ M)^2 \tag{2}$$

where, k is a parameter for adjusting the number of feature points, and the smaller the value, the more feature points are extracted.

The coefficient k is an empirical value, usually lying in the interval [0.04, 0.06] (Tao *et al*., 2020). In this study, $k = 0.06$ was set for all image frames and the size of window was set to 5×5 for Gaussian filtering. In order to reduce the number of feature points and even distribution, we used bucketing techniques (Zhang *et al*., 2020). The bucketing technique consists on divide the image into a grid and, for each bucket, choose the best feature points. We selected the pixel having the maximum corner response function value of Eq. (2) as the final feature point for each bucket.

2.2 Feature Matching

2.2.1 Normalized Cross-Correlation

In this study, matching points between feature points extracted automatically are extracted using a NCC technique, which is a representative region-based image matching. The NCC method calculates the similarity between the feature point of the previous frame and the feature point of the current frame in a window area. The matching points were determined only when the NCC value is greater than a certain threshold value. For each image of the previous and the current frame, after calculating the intermediate variables A, B, and C related to the sum and squared sum in the window area as in Eq. (3), the NCC coefficient is calculated with Eq. (4) using these intermediate variables.

$$A = \sum I,\ B = \sum I^2,\ C = \frac{1}{\sqrt{nB - A^2}} \tag{3}$$

$$D = \sum I_1 I_2, \tag{4}$$
$$NCC = (nD - A_1 A_2) C_1 C_2$$

where, I_1 and I_2 represent the pixel values of the previous frame and the current frame in the window area, respectively. n represents the total number of pixels in the window area. If the pixel values of the previous frame and current frame within the window area are all the same, the NCC value is 1, and as the pixel values are different, it has a value close to 0.

In this study, the primary matching points were extracted using a 11×11 size matching window and an NCC threshold of 0.7, and a mutual consistency check was additionally performed to reduce mismatching points. That is, with respect to the primary matching points received from the previous frame to the current frame, the secondary matching is reversely performed from the current frame to the previous frame, and only the matching points identical to the primary matching result is extracted as the final matching pair.

2.2.2 Integral Images

In general, as the number of feature points extracted from the image increases and the matching window size increases, the summation processing required for NCC calculation in Eq. (3), (4) increases, which is one factor that decreases the computational efficiency in a development environment requiring real-time processing. Therefore, this study aimed to shorten the processing time by using the integral image as a part of improving the NCC calculation efficiency. The integral image is an image obtained by accumulating the pixel values of the input image as shown in Eq. (5)(Viola and Jones, 2001). When using the integral image, the sum of the pixel values in the window can be easily calculated regardless of the window size.

$$IG(x,y) = \sum_{i=0}^{y} \sum_{j=0}^{x} Org(i,j) \tag{5}$$

where, $IG(x,y)$ and $Org(x,y)$ are the values of the integral image and input image at location (x,y), respectively.

Given the integral image, it is possible to calculate the sum

of pixels within a specific window very efficiently. In order to obtain the sum of the green window areas, all values must be added from the original image in Fig. 1(a), and the value is 130. To find the sum in the same area using the integral image, simply subtract B and C from the value of point D in Fig. 1(b) and add the value of point A, and you can see that the value is the same as 130 obtained from the original image. This is because the calculation amount increases exponentially as the window size increases, whereas if the integral image is used, the same result can be achieved through only four arithmetic operations. Accordingly, if the integral image is used in an algorithm in which window-based summation processing such as Eq. (4) for all feature points such as NCC image matching frequently occurs, computational efficiency can be improved.

10	12	3	13	15
15	15	8	6	14
9	26	20	17	9
5	10	12	16	7
13	5	18	11	19

(a) input image

A 10	22	25	B 38	53
25	52	63	82	111
34	87	118	154	192
C 39	102	145	D 197	242
52	120	181	244	308

(b) integral image

Fig. 1. Illustration of block summation using the integral image

2.3 Outlier Removal based on Triangulation

Because there is a possibility that some mismatching points may still exist among matching points that have passed the NCC matching threshold condition and mutual consistency test, we developed a triangular network-based outlier removal method to extract more robust matching points among the initial matching points. Delaunay TIN as shown in Fig. 2 for the matching points is first constructed to obtain the position error of the matching points generated from the NCC-based image matching process. If the kth-triangular vertex $P_{R_i}^k$ of the Delaunay TIN in the previous frame and the corresponding node $P_{S_i}^k$ of current frame are equal to equations (6) and (7), the displacement vectors of nodes of each triangle can be obtained as in Eq. (8).

$$P_{R_i}^k = \left\{ (x_{R_i}^k, y_{R_i}^k) : i = 1,2,3, k = 1,2,\cdots,N \right\} \tag{6}$$

$$P_{S_i}^k = \left\{ (x_{S_i}^k, y_{S_i}^k) : i = 1,2,3, \ k = 1,2,\cdots,N \right\} \tag{7}$$

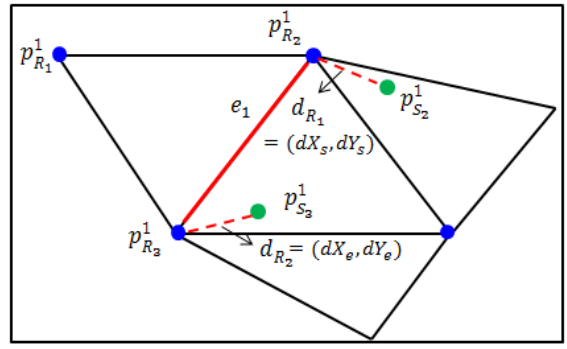


Fig. 2. Outlier removal based on Delaunay triangulation. Blue and green points represent matching points in previous and current image frame, respectively

$$d_{R_i} = \left\{ (x_{R_i}^k - x_{S_i}^k, y_{R_i}^k - y_{S_i}^k) : \ i = 1,2,3 \right\} \tag{8}$$

The outlier removal process using displacement vector information is as follows. First, a triangular edge segment such as e_1 in Fig. 2 is extracted, and if the sum of the displacement difference between the starting point and the ending point of the defined edge as in Eq. (9) is more than a certain threshold (in this study 1.2), the counting number of the node is increased.

$$D = \left| dX_s - dX_e \right| + \left| dY_s - dY_e \right| \tag{9}$$

where, (dX_s, dY_s) and (dX_e, dY_e) represent the displacement vector components of the starting point and the ending point of the edge segment , respectively. After applying the above process to all triangle node, the node having the counting number less than 50% of the number of edge segment connected to itself is classified as outlier.

3. Experiment and Evaluation Results

3.1 Test Data

The test data used in this study are real-world oblique images taken by attaching a video camera to a UAV. The experimental video is a HD color image which composed of 101 frames. The flight path of UAV includes various target areas ranging from natural forests to high-rise buildings. Fig. 3 shows the study area including the high-rise building

area extracted from some sections of the video. As you can see in the picture, it can be confirmed that it is an oblique image acquired by tilting the camera's shooting angle at low altitude. In order to evaluate the performance of the proposed technique, quantitative comparison evaluation was performed using three methods shown in Table 1, which are most commonly used as the existing feature image matching methods.

Table 1. Existing methodologies used for comparative evaluation

Method	Feature Extraction and Matching	Outlier Removal
AM	Harris + NCC	RANSAC
BM	SURF	RANSAC
CM	SURF	Preemptive RANSAC

The AM method based on Harris corner detector is frequently used in the field of image matching due to its relatively low computational complexity and ease of implementation. The BM method combining SURF and RANSAC is most frequently used because it guarantees relatively fast operation and excellent matching performance. Lastly, the CM method has been recently used in the fields that require real-time processing such as video processing. This method used a preemptive RANSAC technique that improves the RANSAC technique where the calculation speed slower as the number of feature points increases.

Fig. 3. Some image frame of study area captured from UAV video data

3.2 Results and Analysis

In this study, in the parameter setting of the AM and proposed method based on Harris corner detector, the same parameters were set for all image frame, and the initial matching was performed by setting with 0.7 as the NCC threshold.

In addition, in order to ensure the objectivity of performance verification, in the parameter setting of the BM and CM method based on SURF, the same parameters were set so that a similar number of feature points was extracted. The main SURF parameters used in this study were Hessian threshold of 0.0005, number of Octaves 4, and 4 scale spaces per octave. In the case of the RANSAC algorithm to remove mismatching points, among the matching points generated from the matching process, four matching points are randomly extracted at random to find the best matching results with the smallest position error. In the case of Preemptive RANSAC, which increased the computational efficiency of the existing RANSAC algorithm, the number of hypothesis was set to 30.

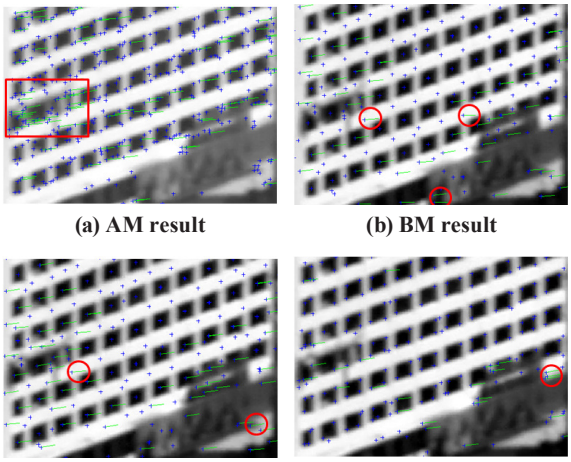
Fig. 4 shows the result images superimposed on the current frame by expressing the matching pairs extracted by applying different image matching techniques to some high-rise building areas as motion vectors. The motion vector represents the change of pixel position from the previous frame to the current frame with a green line. In the AM method, as shown in Fig. 4(a), it can be seen that a large number of feature points were extracted at the corner points of the lattice window frame structure of the building wall, and matching points also occurred at these corner points. However, it can be confirmed that there are some mismatching pairs in which the motion vector direction is different from the surroundings. This is because the NCC matching technique uses only the pixel values within a window area, so many mismatching points occur because the differences in matching correlation values between adjacent feature points are not great in images where similar texture characteristics are continuously displayed such as the study area.

On the contrary, in the BM and CM methods based on SURF that extract feature points from the blob area of the image, as shown in Fig. 4(b) and 4(c), the window frame appears as black blob in the image rather than the edge area of the wall. It can be seen that the feature points were extracted

from the region corresponding to the central region, and the matching points also occur mainly at the center point of these blobs. In addition, it can be easily visually confirmed that the consistency of the motion vector direction of matching points extracted by these methods is significantly higher than the AM method.

In the proposed method based on the Harris corner detector as shown in Fig. 4(d), the number of extracted feature points is relatively small compared to those of the AM method. In addition, it can be confirmed that the motion vectors of the final matching points extracted through the triangular network-based outlier removal process are also quite accurate.

Quantitative comparative evaluation was performed for a more rigorous comparative evaluation in this study. As a quantitative comparative evaluation method, the matching



(a) AM result (b) BM result

(c) CM result (d) result of the proposed method

Fig. 4. Magnified result images using different image matching methods. Extracted matching points according to each method are represented in the form of moving vectors with green line color.

accuracy of each method was first examined using matching success rate. Matching success rate is defined as the ratio of the number of successfully matched points to the total number of match points generated from matching process. Accuracy was evaluated through visual inspection on all video matching results of 101 frames used in the experiment. Fig. 5 shows the number of matching points extracted for each method in frame units. For convenience of analysis, the entire frame (101 frames) was divided into 10 units to calculate the section average. The AM method yielded the most matching points in all the sections compared to other methods, and it can be seen that the matching points was extracted with a relatively similar trend in the case of the proposed method and the CM method.

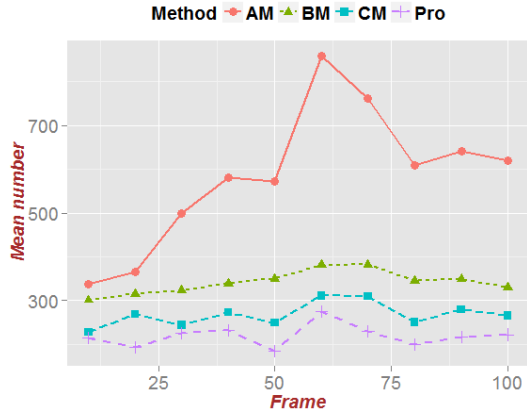


Fig. 5. Mean number of matching points extracted by using different matching methods (The mean number is calculated at interval of about ten frames)

Among the matching points extracted by each method, the average number of false matching for each section is shown in Fig. 6. The number of false matching of AM was also higher on average than the other methods, and the false matching number of the CM method was unexpectedly higher. This is because the preemptive RANSAC algorithm, which is an outlier removal technique used in the CM method, is specialized in processing speed rather than accuracy. As shown in Table 2, the average matching success rate for the entire frame also yielded better results than the existing technique.

Table 2. Comparison evaluation results in terms of running time and mean matching accuracy

Methods	AM	BM	CM	Proposed
Running Time (sec)	124.65	173.34	165.97	39.30
Mean Matching Accurary (%)	99.54	99.42	99.03	99.60

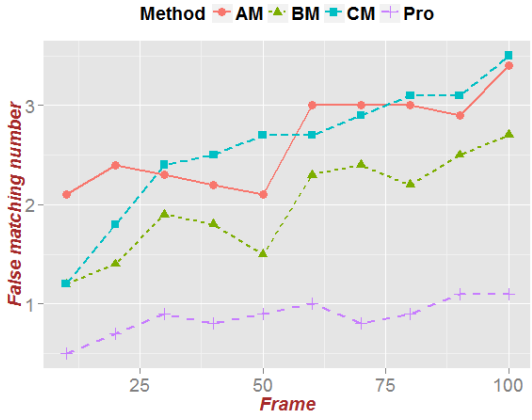


Fig. 6. Mean number of false matching points with different matching methods.

For practical use of UAV video-based matching system, not only accuracy, but also performance speed is a very important factor. Therefore, in this study, for all methods used in the experiment, the performance speed was comparatively evaluated by dividing it into a matching process including feature point extraction and subsequent outlier removal. All experiments are developed in Visual C++ 2010 and carried out on a personal computer with intel Core i7 (3.60 GHz) and 16GB of RAM. As can be seen in Fig. 7, the SURF-based BM and CM method has approximately 6 times more feature point extraction time (F-time) compared to the Harris-based AM and the proposed method. You can see that it leads to a slightly faster result. If you look at the matching part (M-time) in more detail, you can see that the CM method showed the fastest processing speed, and the matching time between the AM and BM methods differs by more than 4 times. This is because the number of feature points extracted by the AM method is much higher than the BM method, and the iteration time required for the optimization process of the RANSAC algorithm is relatively large. Finally, when looking at the feature point extraction and matching time as shown in Table 2, the proposed method showed the fastest execution speed because it took 39.30 seconds to process all 101 frames. This is considered to be a relatively satisfactory level with the ability to process 2.57 frames per second. However, given the reality that most videos are filmed at 30 frames per second, it is thought that the performance speed needs to be improved to apply the proposed method in real time. Therefore, it is expected to further improve the processing speed of the proposed method by integrating a CUDA (Compute Unified Device Architecture)-based parallel processing technique for high-speed computation in the future.

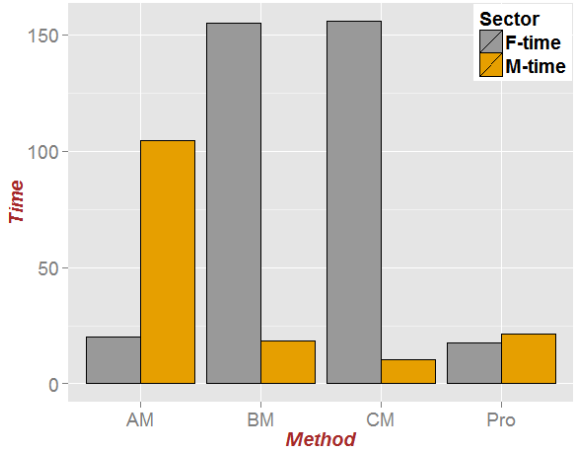


Fig. 7. Performance time comparison of four different matching methods.

4. Conclusion

In this study, matching between UAV video image frames was performed by combining NCC-based feature point matching and triangle-based outlier removal technique. As a result of applying the proposed method to a video composed of 101 frames taken from low altitude, it was confirmed that the results showed stable image matching across all frames. In order to evaluate the effectiveness of the proposed method, quantitative comparative evaluation was performed with representative feature point-based image matching methods.

As a result of the experiment, the proposed method showed a higher matching success rate than the existing method, and the execution speed also showed a processing speed up to 4 times faster. The proposed method therefore has a good potential for the various video-based applications that requires image matching as a pre-processing such as geo-localization.

In the future, we intend to improve the processing speed and performance of the proposed method by integrating the CUDA-based parallel processing method, and perform the video-based navigation in GPS-denied environments.

References

David, N. (2003), Preemptive RANSAC for live structure and motion estimation, *Proceedings Ninth IEEE International Conference on Computer Vision*, 13-16 October, Nice, France, pp. 199-206.

Fischler, M.A. and Bolles, R. C. (1981), Random sample consensus: A paradigm for model fitting with applications to image analysis, *Communications of the ACM*, Vol. 24, pp. 381-395.

Hariss, H., and Stephens, M. (1988), A combined corner and edge detector, *In Proc. of Fourth Alvey Vision Conference-1988*, 31 August-2 September, Manchester, United States, pp. 147-151.

Hartley, R. and Zisserman, A. (2003), *Multiple View Geometry in Computer Vision*, Cambridge University Press, second edition, New York, N.Y.

Heo, Y.S., Lee, K.M., and Lee, S.U. (2011), Robust stereo matching using adaptive normalized cross-correlation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, pp. 807-822.

Herbert, B., Andreas E., Tinne, T., and Luc, V. G. (2008), SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding*, Vol. 110, pp. 346-359.

Inder, K., Silver, V., and Shi, X. (2019), Learning control policies of driverless vehicles from UAV video streams in complex urban environments, *Remote Sensing*, Vol. 11, pp. 2723.

Kanade, T. and Okutomi, M. (1994), A stereo matching algorithm with an adaptive window: theory and experiment, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, pp. 920-932.

Ke, R., Li, Z., Tang, J., Pan, Z., and Wang, Y. (2019), Real-time traffic flow parameter estimation from UAV video based on Ensemble classifier and optical flow, *IEEE Transactions on Intelligent Transportation Systems*, San Francisco, United States, Vol. 20, pp. 54-64.

Lucas, B. and Kanade, T. (1981), An iterative image registration technique with an application to stereo vision, *Proceedings of the 7th international joint conference on Artificial intelligence*, 24-28 August, San Francisco United States, Vol. 2, pp. 674-679.

Lowe, D.G. (2004), Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol. 110, pp. 91-110.

Robert, C., Timonthy, W., and Randal, W. (2014), Relative navigation approach for vision-based aerial GPS-denied navigation, *Journal of intelligent and Robotic System*, Vol. 74, pp. 97-111.

Smith, S. and Brady, J. (1997), SUSAN-a new approach to low-level image processing, *International Journal of Computer Vision*, Vol. 23, pp. 45-78.

Tissainayangam, P. and Suter, M. (2004), Assessing the performance of corner detectors for point feature tracking application, *Image and Vision Computing*, Vol. 22, pp. 663-679.

Tao, L., Zaifeng, S., and Pumeng, W. (2020), Robust and efficient corner detector using non-corners exclusion, *Applied Sciences*, Vol. 10, pp. 1-14.

Viola, P. and Jones. M. (2001), Rapid object detection using a boosted cascade of simple features, *Proceeding of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8-14 December, Kauai, United State, Vol. 1, pp. 511-518.

Xue, X., Li, Y., Dong, H., and Shen, Q. (2018), Robust correlation tracking for UAV videos via feature fusion and saliency proposals, *Remote Sensing*, Vol. 10, pp. 1644.

Zhang, Z., Deriche, R., Faugeras, O., and Luong, Q. T. (1995), A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial Intelligence*, Vol. 78, pp. 87-119.

Zhang, S., Li, S., Zhang, B., and Peng M. (2020), Integration of optimal spatial distributed tie-points in RANSAC-based image registration, *European Journal of Remote Sensing*, Vol. 53, pp. 67-80.

Zhuo, X., Tobias, K., Friedrich, F., and Peter, R. (2017), Automatic UAV image geo-registration by matching UAV images to goereferenced image data, *Remote Sensing*, Vol. 9, pp. 376.