

UAV AND LIDAR IMAGE REGISTRATION: A SURF-BASED APPROACH FOR GROUND CONTROL POINTS SELECTION

B. Kalantar¹, N. Ueda¹, H. A. H. Al-Najjar², H. Moayedi^{3,4,*}, A. A. Halin⁵, S. Mansor⁶

¹RIKEN Center for Advanced Intelligence Project, Goal-Oriented Technology Research Group, Disaster Resilience Science Team, Tokyo 103-0027, Japan - ([bahareh.kalantar](mailto:bahareh.kalantar@riken.jp), [naonori.ueda](mailto:naonori.ueda@riken.jp))@riken.jp

²Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, 2007 NSW, Australia - HusamAbdulrasoolHammadeh.AL-NAJJAR@student.uts.edu.au

³Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁴Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam - [hossein.moayedi@tdtu.edu.vn](mailto:hosseini.moayedi@tdtu.edu.vn)

⁵Dept. of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia - alfian@ieee.org

⁶Dept. of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia - shattri@upm.edu.my

KEY WORDS: Image registration, UAV images, LiDAR, SURF, Feature based.

ABSTRACT:

Multisource remote sensing image data provides synthesized information to support many applications including land cover mapping, urban planning, water resource management, and GIS modelling. Effectively utilizing such images however requires proper image registration, which in turn highly relies on accurate ground control points (GCP) selection. This study evaluates the performance of the interest point descriptor SURF (Speeded-Up Robust Features) for GCPs selection from UAV and LiDAR images. The main motivation for using SURF is due to it being invariant to scaling, blur and illumination, and partially invariant to rotation and view point changes. We also consider features generated by the Sobel and Canny edge detectors as complements to potentially increase the accuracy of feature matching between the UAV and LiDAR images. From our experiments, the red channel (Band-3) produces the most accurate and practical results in terms of registration, while adding the edge features seems to produce lacklustre results.

1. INTRODUCTION

Multisource remote sensing utilizes data from various image sources in order to improve upon single data remote sensing. However, multiple sources of images are only useful after proper image registration, which can be achieved through accurate selection of ground control points (GCPs). Generally, image registration techniques are divided into two categories, (i) parametric (i.e. using all required parameters of the remote sensing platform), and (ii) non-parametric (i.e. only a set of GCPs are considered) (Bouchiha & Besbes., 2013). This work focuses on the latter category, which is further divided into manual and automatic registration methods (Wong & Clausi., 2007).

In manual registration, the selection of control points (CP) is performed by a human operator, which can be prone to inaccuracies due to human error (Eastman et al., 2007). Moreover, manual registration is impractical for complex images where the human eye might not be able to decide on suitable CPs. This is where automatic techniques come in as it potentially avoids the pitfalls of human limitations. Automatic techniques can fall into two categories, namely area-based methods (ABM) and feature-based methods (FBM) (Zitova & Flusser., 2003; Kalantar et al., 2017a; Kalantar et al., 2018). Also referred to as correlation-like or template matching methods (Fonseca & Manjunath., 1996), ABM calculates the statistics within a small pixel window, both in the sensed image and the reference image (Hong & Zhang., 2007).

On the other hand, while FBM can also use similar windows, it can also take the entire image for correspondence estimation (Althof et al., 1997; Barnea & Silverman., 1972; Pratt., 1974). Compared to ABM, FBM is more robust and reliable (Hong & Zhang., 2007). A feature descriptor such as Scale Invariant Feature Transform (SIFT) (Lowe, 1999; Lowe, 2001; Lowe, 2004), Harris corner (Harris & Stephens., 1988) speed up robust feature (SURF) (Bay et al., 2006) or shape context (SC) (Belongie et al. 2002) can be used in FBM.

The Harris corner detector (Harris & Stephens., 1988) is a popular feature based on the eigenvalues of the second moment matrix. However, Harris corners are not scale invariant making it less suitable for dynamically varying remote sensing footage. Lowe (2004) proposed using SIFT, owing to its effectiveness in domain-specific image registration, image mosaicking and image retrieval (Lowe, 2004; Ledwich & Williams., 2004; Harandi et al., 2015; Yang & Guo., 2008). SIFT's main limitation however is its high computational cost (Harandi et al., 2015). Attempts have been made to speed up the algorithm such as the work by (Mikolajczyk & Schmid., 2005). Here, the authors attempted to make SIFT lighter, more robust and discriminative by proposing a variant based on the Gradient Location-Orientati0on Histogram (GLOH) (Mikolajczyk & Schmid., 2005). GLOH requires a lot of priori samples to generate a projection matrix in the feature space. Other than that, shape context (Belongie et al., 2002) was also used to compute the correspondences between points.

* Corresponding author at: Ton Duc Thang University, Ho Chi Minh City, Vietnam ([hossein.moayedi@tdtu.edu.vn](mailto:hosseini.moayedi@tdtu.edu.vn))

Another common interest-point detector is SURF (Bay et al., 2008) as it is invariant to rotation, scale and illumination and

also the computational cost is low so it is suitable for any real time application. Conceptually, it is very similar to SIFT but with faster computational time. The work in Bouchiha and Besbes (2013) proposed a SURF-based automatic registration approach for remote-sensing. They used three different image datasets where one reference image has more than four sensed images. Two sensors were utilized namely Landsat TM and Landsat ETM+. The reported results were robust to rotation, scale and illumination changes and can be used to register remotely sensed images obtained under varying conditions. Brook & Ben-Dor (2011) proposed a novel method for automatic image registration based on topology (AIRTop) for change detection and multi-sensor (airborne and spaceborne) fusion. SURF was used to extract landmark structures (roads and buildings). A remote sensing image registration method was also proposed in Panchal et al. (2013) using an optimized variant of SURF. Two test sets were used: (i) the first being Landsat 7 ETM+ as the reference image and the Landsat 4-7 sets combined as the sensed images, and (ii) the second includes SIR-C Radar as the reference image and the Landsat 4-5 MSS as the sensed image. Results showed that this variant of SURF was robust and accurate in performing registration.

To the best of our knowledge, no work has attempted to utilize interest point detection algorithms that involve UAV and LiDAR images. Therefore, this paper will attempt to investigate and evaluate the use of the SURF algorithm for UAV and LiDAR images registration. We foresee that an interest point detector would be able to accurately identify GCPs, with minimal time (hence SURF over SIFT). In the following section, the datasets and the study area are presented, including a brief explanation of the SURF algorithm as well as the overall workflow of our proposed approach. Section 4 presents the experimental results followed by discussions in Section 5. This paper is concluded in Section 6 with remarks regarding future research.

2. STUDY AREA AND DATA USED

The sensed image used in this study is obtained on February 16th 2016, from the fixed-wing UAV platform over the study area located between $102^{\circ} 19' 55''$ E to $103^{\circ} 27' 08''$ E and $02^{\circ} 50' 36''$ N to $02^{\circ} 39' 22''$ N, over Universiti Putra Malaysia (UPM), Malaysia (Fig. 1). The UAV image has three bands [green (G), red (R), blue (B)] with a ground resolution of 0.068 m/pix. These images show several geometric transformations/changes such as scale (scale factor between 1 and 5), rotation (varies between 25° and 175° with a step of 25°) and photometric (e.g. changing illumination). The other dataset we used is from an airborne LiDAR (Light Detection and Ranging) system on March 8, 2013 over the same study area. Along with LiDAR point clouds, orthophotos were captured with a camera with a spatial resolution of 13 cm. The laser scanner has a scanning angle of 60° with a camera angle of $\pm 30^{\circ}$. The posting density of the LiDAR data was 3–4 pts/m² (average point spacing = 0.41 m).

3. METHODOLOGY

Figure 2 shows the overall workflow of the proposed work. First, the UAV and LiDAR data are preprocessed using three main steps: (i) orthorectification, (ii) mosaicking, and (iii) grayscale conversion (Kalantar et al., 2017b). Then, two edge

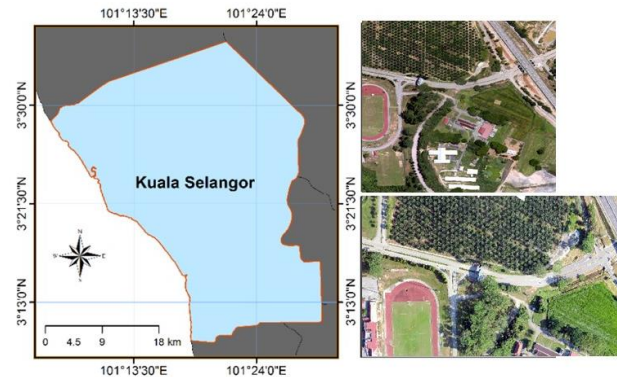


Figure 1. The study area located over UPM.

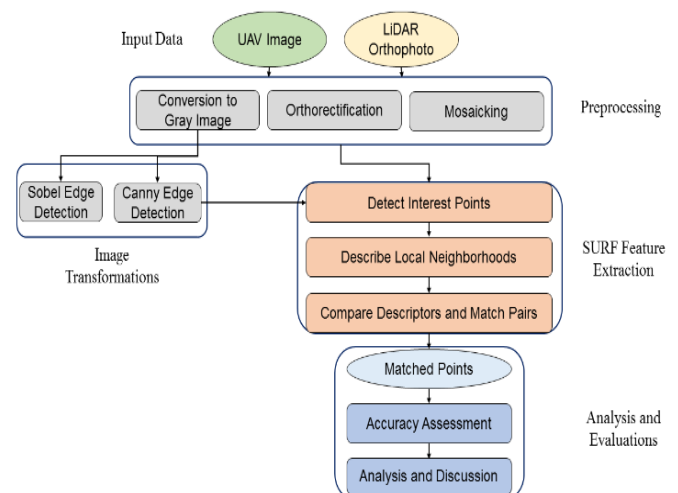


Figure 2. Overall workflow of this study.

images are generated using the Sobel and Canny edge detectors, respectively. The SURF algorithm is then used to detect the pertinent features, followed by the computation of their descriptors. The matching pairs are then identified between the UAV and LiDAR datasets. The performance of SURF is scrutinized for each input dataset based on the matching results and visualizations. This generally entails evaluation of the pre-processed images and the edge filtered images.

3.1 Pre-Processing

UAV images were preprocessed using Agisoft PhotoScan software which allows generating georeferenced dense point clouds, textured polygonal models, digital elevation models (DEMs), and orthomosaics from a set of overlapping images with the corresponding referencing information (Kalantar et al. 2017b). On the other hand, LiDAR data were preprocessed using ArcGIS 10.3 software. The point-cloud LiDAR data was interpolated into a digital terrain model (DTM) and aerial orthophoto images were rectified and mosaicked based on the LiDAR point cloud.

3.2 The principles of SURF

As explained in Teke & Temizel (2010), SURF is based on the approximated Hessian Matrix:

$$H(x, \sigma) = \begin{bmatrix} L_{XX}(x, \sigma) & L_{XY}(x, \sigma) \\ L_{YX}(x, \sigma) & L_{YY}(x, \sigma) \end{bmatrix}$$

where $L_{XX}(x, \sigma)$ is the convolution result of the second order derivative of Gaussian filter $\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image in point x , and similarly for $L_{XY}(x, \sigma)$ and $L_{YY}(x, \sigma)$.

The first step of SURF includes fixing a reproducible orientation according to the information from a circular region around the interest point. Secondly, a square region aligned to the selected orientation is constructed where the SURF descriptor are extracted from. The Haar-wavelet responses in x and y direction is calculated in order to be invariant to rotation.

3.3 Implementation

We implemented the SURF algorithm in Matlab (version R2015) on a personal computer with a Core i5 CPU (2.4 GHz) and 4-gigabytes of RAM. UAV and LiDAR orthophoto images were used as inputs. Both images were then converted to grayscale followed by Sobel and Canny edge detection. The SURF features were then extracted from the images followed by computation of the descriptors at the interest points. These descriptors are subsequently used to match the UAV and LiDAR features. A geometric transformation is utilized to locate the matched features in the scene. Finally, after removing the outliers, the matched features are displayed. The average time required for executing the code was approximately 6-seconds.

4. RESULTS

In all, ten (10) different experiments were conducted with respect to three different geometric transformations namely projective, similarity, and affine (Table 1). The initial observation from the table reveals that best registration is achieved by using band 3 (red channel) from both the UAV and LiDAR orthophoto images, with an accuracy of 100%. Note however, that this accuracy is not the registration accuracy, but only an indication about the accuracy of feature matching. To obtain the accurate registration accuracy, the number of detected features without outliers in each dataset was carefully analysed. This is because the geometric transformation can only be executed (and its accuracy estimated) if at least four accurate matching features are found.

The results show that the highest number of features was identified from the Canny filtered images. This was achieved using band 2 channels from the UAV and LiDAR orthophoto images. In contrast, in the Canny images, very low number of features were matched between the UAV and LiDAR orthophoto images. The highest accuracy rate achieved is 40 by using band 2, but with the similarity transformation method applied. In addition, the experiments indicate that the Sobel filtered images are not suitable for registration of UAV and LiDAR orthophoto images. This is because very low number of features were identified. When comparing the three geometric transformations, the Similarity method was most efficient based on all the experiments conducted in the current study. However, the best feature matching is achieved by the Projective method when using band 3 for both datasets. In this experiment, five (5) features could be correctly matched between the UAV and LiDAR orthophoto images. Based on these results, it can be postulated that the edge filters are not recommended for UAV-LiDAR data registration. The use of a single band instead of taking the average of three available bands is more suitable as

feature matching accuracy is highest. Specifically, the use of band 3 (red channel) was found to most efficient and practical.

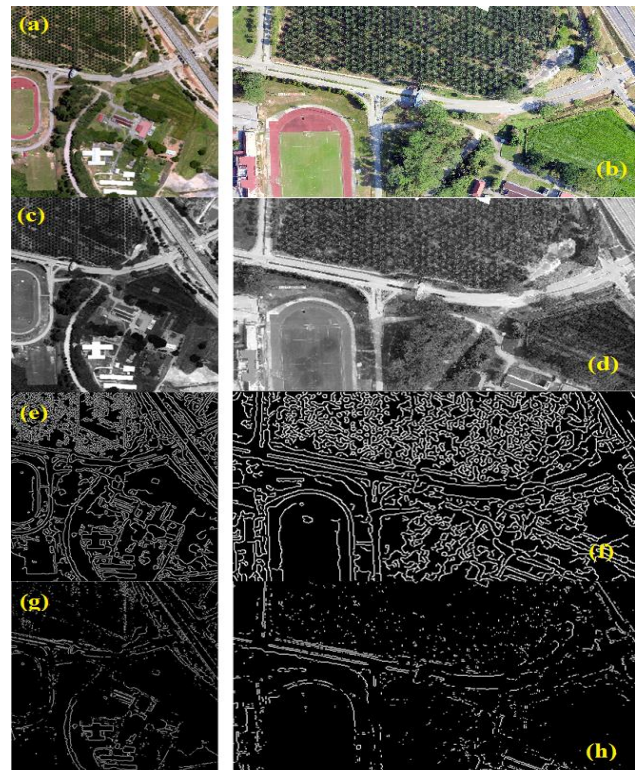


Figure 3. (a) LiDAR, (b) UAV, (c) Band 3 LiDAR, (d) Band 3 UAV, (e) Canny transformed LiDAR Band 3, (f) Canny transformed UAV Band 3, (g) Sobel transformed LiDAR Band 3, (h) Sobel transformed UAV Band 3.



Figure 4. (a) UAV Band1, (b) UAV Band2, (c) UAV Band3, (d) LiDAR Band1, (e) LiDAR Band2, (f) LiDAR Band 3 (The green circles show the strongest feature points selected by SURF).

5. DISCUSSION

Registration of two different data sources such as UAV and LiDAR is a challenging task using automatic procedures because of the difference of spatial resolution and the amount of information contained in the two images are different. Based on the results, band 3 seems to provide the highest accuracy for matching features. This further validates results from a previous study by Teke and Temizel (2010) who also found that single bands (i.e. red and green) performs better than the average of three RGB bands.

Type of Transformation	Exp.	Source data	Sense data	Number of Identified features including outliers	Number of identified features with outliers removed	Accuracy Rate (i.e. number of correct matches/ total number of matched features)
Projective	1	UAV	LiDAR	4	1	0%
	2	UAV-Band1	LiDAR-Band1	5	4	75%
	3	UAV-Band2	LiDAR-Band2	9	5	60%
	4	UAV-Band3	LiDAR-Band3	11	5	100%
	5	Canny (UAV-Band1)	Canny (LiDAR-Band1)	16	4	25%
	6	Canny (UAV-Band2)	Canny (LiDAR-Band2)	22	5	20%
	7	Canny (UAV-Band3)	Canny (LiDAR-Band3)	19	5	0%
	8	Sobel (UAV-Band1)	Sobel (LiDAR-Band1)	3	2	0%
	9	Sobel (UAV-Band2)	Sobel (LiDAR-Band2)	2	1	0%
	10	Sobel (UAV-Band3)	Sobel (LiDAR-Band3)	5	4	0%
Similarity	1	UAV	LiDAR	4	3	100%
	2	UAV-Band1	LiDAR-Band1	5	3	100%
	3	UAV-Band2	LiDAR-Band2	9	3	100%
	4	UAV-Band3	LiDAR-Band3	10	4	100%
	5	Canny (UAV-Band1)	Canny (LiDAR-Band1)	15	4	25%
	6	Canny (UAV-Band2)	Canny (LiDAR-Band2)	22	5	0%
	7	Canny (UAV-Band3)	Canny (LiDAR-Band3)	20	5	40%
	8	Sobel (UAV-Band1)	Sobel (LiDAR-Band1)	3	2	0%
	9	Sobel (UAV-Band2)	Sobel (LiDAR-Band2)	2	1	0%
	10	Sobel (UAV-Band3)	Sobel (LiDAR-Band3)	6	4	0%
Affine	1	UAV	LiDAR	3	2	100%
	2	UAV-Band1	LiDAR-Band1	5	3	66%
	3	UAV-Band2	LiDAR-Band2	8	3	66%
	4	UAV-Band3	LiDAR-Band3	10	4	100%
	5	Canny (UAV-Band1)	Canny (LiDAR-Band1)	14	3	33%
	6	Canny (UAV-Band2)	Canny (LiDAR-Band2)	21	3	0%
	7	Canny (UAV-Band3)	Canny (LiDAR-Band3)	19	3	0%
	8	Sobel (UAV-Band1)	Sobel (LiDAR-Band1)	3	2	0%
	9	Sobel (UAV-Band2)	Sobel (LiDAR-Band2)	2	1	0%
	10	Sobel (UAV-Band3)	Sobel (LiDAR-Band3)	5	3	0%

Table 1. Results of the experiments analysed in the current study for selecting optimal case for registration UAV and LiDAR datasets.



Figure 5. Best matching features by using band 3 of UAV and LiDAR orthophoto and projective transformation method.

Although sensitive to image features, the red band seems to highlight important scene features (e.g. Figure 3.), increasing the number of matched points. Conversely, as projective transformation uses a more complex equation than similarity and affine methods and requires a minimum of four identified features, this projective transformation method performs better than other methods in terms of identifying more number of features without outliers (Table 1). On the other hand, similarity transformation scales, rotates, and translates the data. It also maintains the aspect ratio of the features transformed which was seen to be more effective than affine methods that can skew the

data. Furthermore, the Canny edge filter highlights the edge in the images so that more features could be determined. However, albeit having more features, the result of matching these features in both UAV and LiDAR orthophoto images did not achieve high accuracy rates. This might be due to the complex features presented in the study area. In contrast, a study by Pandya et al. (2013) showed that the use of Sobel increased the accuracy rate of matching features. In their study, a simple photo captured by normal handheld camera was used. In the current study, we suggest the use of single band such as red channel for detecting features and image registration for remote sensing datasets.

LiDAR can be a significant complementary data for UAV images that can serve many applications such as crop mapping, object detection and recognition, and urban planning. However, proper registration of UAV and LiDAR data is necessary for accurate information extraction. Several methods are available in the literature to register UAV and Lidar data sets such as SIFT, SURF, and GIS-based registration. As each method has its own advantages and disadvantages, the evaluation of the methods is important to determine the best method for UAV and Lidar data registration. This study evaluates SURF algorithm with the aim of providing guidelines for the practical projects.

6. CONCLUSION

This study evaluated the SURF algorithm for UAV and LiDAR orthophoto images registration. In addition, two edge filters (Sobel and Canny) were also utilized in the preprocessing stage, with the hope of identifying more useful features. Results show that best registration can be achieved by merely considering band 3 (the red band) from the input data. Although filtering the RGB images into edge images increased the number of features extracted by SURF, the features seemed to just add more noise and overall ineffective. On the other hand, grayscale images could be matched better than using edge images. Overall, this study suggests using SURF method for UAV and LiDAR image registration is effective. Our future plan is to consider transforming the input images using high pass image filtering techniques.

References

- Althof, R. J., Wind, M. G., & Dobbins, J. T. (1997). A rapid and automatic image registration algorithm with subpixel accuracy. *IEEE transactions on medical imaging*, 16(3), 308-316.
- Barnea, D. I., & Silverman, H. F. (1972). A class of algorithms for fast digital image registration. *IEEE transactions on Computers*, 100(2), 179-186.
- Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3), 346-359.
- Bay, H., Tuytelaars, T., & Van Gool, L. (2006, May). Surf: Speeded up robust features. In *European conference on computer vision* (pp. 404-417). Springer, Berlin, Heidelberg.
- Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (4), 509-522.
- Bouchiha, R., & Besbes, K., 2013. Automatic remote-sensing image registration using SURF. *International Journal of Computer Theory and Engineering*, 5(1), 88-92.
- Brook, A., & Ben-Dor, E. (2011). Automatic registration of airborne and spaceborne images by topology map matching with SURF processor algorithm. *Remote Sensing*, 3(1), 65-82.
- Eastman, R. D., Le Moigne, J., & Netanyahu, N. S. (2007, June). Research issues in image registration for remote sensing. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE.
- Fonseca, L. M., & Manjunath, B. S. (1996). Registration techniques for multisensor remotely sensed imagery. *PE & RS-Photogrammetric Engineering & Remote Sensing*, 62(9), 1049-1056.
- Harandi, B. K. G., Mansor, S. B., & Shafri, H. Z. M. Image mosaic methods for UAV sequence images. in *36 th conference on ACRS*, 2015.
- Harris, C., & Stephens, M. (1988, August). A combined corner and edge detector. In *Alvey vision conference* (Vol. 15, No. 50, pp. 10-5244).
- Hong, G., & Zhang, Y. (2007, July). Combination of feature-based and area-based image registration technique for high resolution remote sensing image. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International* (pp. 377-380). IEEE.
- Kalantar, B., Mansor, S. B., Halin, A. A., Shafri, H. Z. M., & Zand, M. (2017a). Multiple moving object detection from UAV videos using trajectories of matched regional adjacency graphs. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9), 5198-5213.
- Kalantar, B., Mansor, S. B., Sameen, M. I., Pradhan, B., & Shafri, H. Z. (2017b). Drone-based land-cover mapping using a fuzzy unordered rule induction algorithm integrated into object-based image analysis. *International journal of remote sensing*, 38(8-10), 2535-2556.
- Kalantar, B., Mansor, S., Halin, A. A., Ueda, N., Shafri, H. Z. M., & Zand, M. (2018, October). A graph-based approach for moving objects detection from UAV videos. In *Image and Signal Processing for Remote Sensing XXIV* (Vol. 10789, p. 107891Y). International Society for Optics and Photonics.
- Ledwich, L., & Williams, S. (2004, December). Reduced SIFT features for image retrieval and indoor localisation. In *Australian conference on robotics and automation* (Vol. 322, p. 3).
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on* (Vol. 2, pp. 1150-1157). Ieee.
- Lowe, D. G. (2001). Local feature view clustering for 3D object recognition. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. 1-1). IEEE.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.
- Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, 27(10), 1615-1630.
- Panchal, P. M., Panchal, S. R., & Shah, S. K. (2013). A comparison of SIFT and SURF. *International Journal of Innovative Research in Computer and Communication Engineering*, 1(2), 323-327.
- Pandya, M. M., Chitaliya, N. G., & Panchal, S. R. (2013). Accurate image registration using SURF algorithm by increasing the matching points of images. *International Journal of Computer Science and Communication Engineering*, 2(1), 14-19.
- Pratt, W. K. (1974). Correlation techniques of image registration. *IEEE transactions on Aerospace and Electronic Systems*, (3), 353-358.
- Teke, M., & Temizel, A. (2010, August). Multi-spectral satellite image registration using scale-restricted SURF. In *Pattern*

Recognition (ICPR), 2010 20th International Conference on (pp. 2310-2313). IEEE.

Wong, A., & Clausi, D. A., 2007. ARRSI: Automatic registration of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5), 1483-1493.

Yang, Z. L., & Guo, B. L. (2008, August). Image mosaic based on SIFT. In *Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP'08 International Conference on* (pp. 1422-1425). IEEE.

Zitova, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21(11), 977-1000.