

# Literature Review on Image Registration and Image mosaicing

Leo Chen  
Computer Science Department  
University of Cape Town

Western Cape, South Africa  
magnolia1047999863@gmail.com

## ABSTRACT

Image registration and image mosaicing have been two important topics in computer vision. Image registration is the process of aligning two or more images of the same scene taken at different time, from different viewpoints, and/or by different sensors. It has four steps: feature detection, feature matching, transformation model estimation and image transformation and resampling. Techniques are categorized into two categories which are area-based and feature-based methods. Image mosaicing is the alignment of multiple images with overlapping regions onto one reference image. It contains four steps: registration, reprojection, stitching and blending. Its technique can be classified as registration based and blending based.

## KEYWORDS

Image registration, Image mosaicing

## 1 Introduction

Image registration and image mosaicing have been two important topics in computer vision. Image registration is the process of aligning two or more images of the same scene taken at different time, from different viewpoints, and/or by different sensors [1].

Image registration has four stages: feature detection, feature matching, transformation model estimation and image resampling. Salient and distinctive features are detected in each image (feature detection). Correspondences between features in each pair of images will be established (feature matching). These will be used as a mean to estimate types and parameters of mapping functions (transformation model estimation). At the last stage, these mapping functions will be used to transform and resample the sensed image to the reference image. In some literatures, an extra stage "Outlier rejection" has been included between feature matching and transformation model estimation [2-4]. It uses statistical models such as RANSAC (Random Sample Consensus) to minimize matching errors.

Image mosaicing is the alignment of multiple overlapping images into one large composition which represents a part of a 3D scene [5]. It contains four stages: registration, reprojection, stitching and blending. First, geometric correspondences between images are established. These will be used to estimate a transformation model (usually an 8-degree planar homography). Following registration, reprojection is done by aligning images onto a same

coordinate system using the computed transformation model. After reprojection, stitching is performed where pixels of overlapping portions are mixed and pixels of the rest of the images are preserved. The final stage is blending, it can be done during or after the stitching stage. It tries to minimize the discontinuities in the global appearance of the mosaic image.

This literature review aims to review current image registration and image mosaicing techniques. Understanding their advantages and disadvantages, and from that to gain insights on how these can be applied to solve the problem of TREEREG or how these can innovate new ideas.

The rest of the review will have 2 sections: In section 2, techniques for image registration and image mosaicing will be presented, discussed and compared. Each will have its own presentation section and discussions and critical comparison section. Section 3 will be conclusion.

## 2. Presentation, Discussion and Comparison

### 2.1 Image Registration

#### 2.1.1 Presentation of Image Registration

Zitova and Flusser have categorized image registration techniques into two categories: area-based and feature-based.

Area-based methods merge feature detection stage into feature matching stage. The idea is to have windows of predefined size on both images (sensed and reference image), similarity measure is then computed between windows. Window with the highest similarity will be used to estimate a global transformation model. Common approaches are correlation-like methods, Fourier methods and mutual information methods [1].

Feature-based methods on the other hand, detect salient features such as points, lines, corners or contours. These will be represented as descriptors which are then used for finding correspondences between images. Local transformation models for each correspondence will be estimated and the sensed image will be transformed and resampled to the reference image. Common approaches are SIFT (Scale Invariant Feature Transform) [6], SURF (Speeded-Up Robust Features) [7] and ORB (Oriente FAST and Rotated BRIEF) [10].

SIFT is one of the most influential method which was presented in 2004 by D.Lowe [6]. It involves 4 steps. First,

potential key points (features) are found by searching scale-space peak. This is done by first apply Gaussian blur operator, then the different of Gaussian kernel (DoG) is calculated, which is used to calculate the Laplacian of Gaussian approximation. Pixels will be compared and local extrema is considered as potential key point. Next is key point localization where potential key points are refined and filtered by getting rid of the low contrast key points. Thirdly, orientation is assigned to each legitimate key points based on local pixel gradient. Lastly descriptors are generated for each key point based on gradient magnitude and orientation. For feature matching, SIFT usually uses nearest neighbor matching or best-bin-first algorithm [6].

SURF [7] is a fast and robust algorithm. It also relies on Gaussian scale space analysis of the images. It is composed of two parts: feature extraction and feature description. For feature extraction, it is based on determinant of the Hessian matrix to detect key points. It also exploits integral images to improve performance. For feature description, it first fixes a reproducible orientation based on information from a circular region around the key point [7]. Then a square region is constructed and aligned to the selected orientation and the SURF descriptor is extracted from it [7].

ORB was proposed in 2011 as an efficient and viable alternative to SIFT and SURF [10]. It combines FAST (feature detection technique) and a variation of BRIEF (descriptor) which is rBRIEF to achieve fast and robust registration. First FAST is applied, which compares the brightness of a pixel  $p$  to surrounding 16 pixels that are in a small circle around  $p$ . If more than 8 pixels are brighter or darker than  $p$ , then  $p$  is considered as a key point [8]. Since FAST features do not have orientation component and multiscale features, a multiscale image pyramid is applied. Once a pyramid has been created, FAST is applied to each level for feature detection which achieves partial scale invariant. Then orientation is assigned to each key point by checking the levels of intensity changes around that key point. Once features have been detected, BRIEF is applied to represent features into binary feature vectors. Since BRIEF performance falls off sharply under rotation, ORB proposed a modified version of BRIEF to achieve rotation invariant.

Machine learning techniques can also be applied to image registration. In [14] a supervised learning approach has been proposed. It self-learns the mapping function through paired patches between images. From that a transformation model will be estimated and used for registration. Unsupervised techniques can also be applied such as in [15]. In this paper authors have proposed an unsupervised learning approach. The idea is for the program to calculate and check the similarity metric between images. Window pair with highest similarities will be learned and then a transformation model is estimated.

Feature description also play an important role for the registration accuracy and robustness. [11] has categorized feature descriptors into 6 categories: gradient-based, intensity-based, spatial frequency-based, moment and probability-based, learning-based and convolutional neural

network-based. [13] has shown to represent a contour by its centroid.

There are different types of registration such as rigid, non-rigid and elastic registration. Errors in registration are usually three types: localization error, matching error and alignment error [1].

## 2.1.2 Discussion and Comparison

Area-based methods have the advantage to apply registration while there is not enough prominent details and distinctive information in images. Therefore, it is frequently applied in medical image registration. However, area-based methods have two main limitations which are: one is originated from its basic idea, since it uses rectangular windows to compare, if images are deformed by more complex transformation, this type of window cannot cover the same part of the scene in both sensed and reference image [1]. Another limitation is that a smoothed area where there are not enough prominent details, it is very likely to be matched to a similar window in another image, even though they do not represent the same scene [1].

Feature-based methods typically focus on local structural information of images. It is more robust to scaling, rotation and deformation. The difficulty lies in feature detection and feature descriptors. Features need to be robust in the sense that, under scaling, rotation, some levels of deformation, noise and illumination, the algorithm should still be able to identify those features. These features also need to be represented in a distinctive way so that each descriptor uniquely identify a feature.

Therefore, area-based methods are preferred when there is not enough prominent details and distinctive information in images. However, when the number of images being registered are large such as in image mosaicing, area-based methods are preferred even when there are sufficient details and local structural information. The reason for that is with local image registration (feature-based methods), more errors are accumulated, and it will eventually lead to low quality of registration. It is possible to combine area-based and feature-based techniques for registration. As shown in [16], authors have proposed an approach with combines wavelet-based feature extraction technique and relaxation-based image matching technique.

Common feature-based detection-description methods are compared and analyzed in [2, 12]. Both papers have concluded that SIFT is the most accurate method while ORB is the most efficient method. As shown in [2], SIFT has higher repeatability than SURF under all situations, and higher repeatability than ORB under scaling. It is also shown that ORB is more robust to rotation and affine transformation than SIFT and the repeatability remains stable and high as long as the scale remains in the range of 60% - 150%. Therefore, when the scaling is within that range, ORB is a better option than SIFT such as the case in TREEREG project since it is free to use, more robust and efficient. [12]

concluded that ORB is less robust than SIFT under rotation by comparing the matching rate. The conclusion is questionable since ORB detects much more features than SIFT (as shown in [2]), a lower matching rate does not necessarily mean it is less robust. One important difference between SIFT, SURF and ORB is that SIFT and SURF detect features across the entire image whereas ORB detects features mostly at the center of the image. Thus, if image is deformed or blurred in the middle, SIFT and SURF would be a better choice. If the number of features being detected is restricted such as in [2] where ORB and BRISK limited to 1000 features have also been tested, the accuracy and robustness of the registration will decrease accordingly. This type of registration techniques often faces a tradeoff between efficiency and accuracy.

Feature descriptors play an important role in both robustness and accuracy of registration process. There is no single best descriptors, so choosing an appropriate descriptor for features under different situations and applications is somewhat a skill that needs to be mastered. Gradient-based descriptors are more suitable for big geometric distortion and rotation [11]. Intensity-based methods are more tolerance to illumination changes, and it is computationally simple [11]. Spatial frequency-based descriptors are more accurate and faster. It is more suitable for point sets matching such as shapes [11]. Moment and probability-based methods are suitable for remote sensing images since they are efficient and robust, and can handle outliers [11]. Learning-based methods can be classified into two categories: learning low-level features and deep learning neural network for learning multiple levels of features (from low to high). These have the advantage that these do not need ground-truth data which are more flexible than conventional features. It can also accurately capture complex features thus improves the registration performance [11]. Convolutional neural network-based descriptors show improved discriminative ability for image registration, but have the drawback that it is computationally expensive [11].

## 2.2 Image Mosaicing

### 2.2.1 Presentation of Image Mosaicing

Image mosaicing aligns multiple images with overlapping area onto one reference image. The process contains four steps: (1) Registration between images is done using general image registration techniques. More typically, a global transformation model is estimated instead of local transformation model since local model is more likely to accumulate errors. This global model is usually a planar homography. (2) Images are aligned to a same coordinate system using the global transformation model estimated in step 1. (3) Stitching is performed to overlay aligned images by merging overlapping pixels while retaining other pixels. (4) Blending is performed to minimize the discontinuities in the global appearance of images [5].

Debabrata and Naima have categorized image mosaicing techniques into two categories: image mosaicing based on registration and image mosaicing based on blending [5].

Image mosaicing based on registration can be split into two groups: spatial domain-based methods and frequency domain-based methods, where spatial domain-based can be divided into area-based and feature-based [5]. Common techniques for area-based are normalized cross correlation-based methods and mutual information-based methods [5]. Common techniques for feature-based methods are low level feature-based such as Harris corner detector-based, FAST corner detector-based, SIFT feature detector-based and SURF detector-based methods [5]. These are common image registration techniques which the accuracy of it will directly affect the result of image mosaicing. For frequency domain-based methods, these methods require computation in the frequency domain to find optimal transformation parameters. It is very sensitive to noise and accurate registration often requires significant overlapping between images [5].

Image mosaicing based on blending has two categories: transition smoothing-based and optimal seam-based methods, where transition smoothing-based methods consist of feathering-based, pyramid-based and gradient-based methods [5]. Quite often inconsistencies can occur during image mosaicing, blending is important to minimize such inconsistencies of visibility in the appearance of the final mosaic.

Image mosaicing algorithms using transition smoothing-based blending attempt to minimize the visibility of seams by smoothing overlapping regions of images. It has the advantage to make boundaries become imperceptible, however the content and coherency of the overlapping region is not guaranteed since fusion does not take the content of the scene into consideration. As a result, mosaic generated using these methods often comes with blur transitions [5].

Feathering-based blending algorithms blend by taking an average value in overlapping pixels [5]. This approach may fail due to exposure difference, misalignment and moving objects in the scene [5].

Pyramid-based blending is based on low band-pass pyramid. First mask image for each source image is generated. Then these mask images will be converted into a low-pass pyramid. Per-level feathering is performed by using the resultant mask images as weights. Lastly interpolation and summing are performed for these per-level feathering [5].

Gradient-based blending blends images by suitably mixing the gradient of images. This can make the resultant image more visually appealing than previous two methods. However, it is important to have best or almost perfect registration to achieve high quality blending [5].

Optimal seam-based blending attempt to minimize the visibility of seams by looking for optimal seams in the joining boundaries between the images [5]. The goal is to find optimal location to allocate seam lines of overlapping regions [5]. It has the advantage when dealing with moving

objects and parallax but has the disadvantage that its transition is obvious when there are exposure differences [5].

Lin and Medioni [3] have proposed a method for generating mosaicing using multiple image sequences. The idea is to have a reference map (usually a satellite image) of the region and register a partial local mosaic onto it. First selected images (from image sequences) are used to form a partial local mosaic, this is done by estimating homographies between images which are then used to register these images together. Finally, the generated partial local mosaic is then registered to the reference map.

Li and Isler [4] have proposed a method which uses a combination of general Structure from Motion (SfM) techniques and standard image mosaicing techniques. Images are selected to form a sparse ground reconstruction by using standard SfM algorithm and camera pose [4]. Then a ground plane is estimated using a robust plane fitting algorithm. This algorithm will rectify each previously selected image. Homography for each image will be calculated using ground plane and camera pose. Finally, for each ground patch in the final mosaic will consider all images and find the one with the best view. Transformation and resampling are performed using previously calculated homography.

### 2.2.2 Discussion and Comparison

The performance of image mosaicing in terms of accuracy and quality of the generated mosaic mainly depend on two aspects of the process: registration and blending. Inaccurate registration will be accumulated and will affect the final quality of mosaicing. The quality of blending heavily relies on small or even perfect registration. A registration with medium to big error will cause ghost effect such as in feathering-based methods and pyramid-based methods [5]. It can also cause unappealing appearance of the mosaic in gradient-based methods [5].

One approach to avoid error accumulation is proposed in [3] as mentioned above. They first generate a partial local mosaic by using image sequences. Since consecutive images in image sequences will have large overlapping areas for the same scene at same time, high quality registration can be easily achieved. Then they register the partial local mosaic to the reference map. Using a global transformation model, the structural information in the image can be preserved, which avoids deformation. The result that was shown in the paper is promising, however, this approach has two limitations. First is that the approach assumes the scene is planar when compared to the camera. Thus, if the altitude of UAV is low, parallax between UAV image and the reference map is strong, this will cause mutual information to produce ambiguous affine registration. Secondly, inaccurate registration in the previous frames can be fixed or adjusted, which will affect the registration of future frames.

Another approach [4] avoids limitations in the above approach. They do not assume the scene is planar to the camera by constructing a ground geometry using SfM

techniques. This allows UAV to fly even at low altitude. It also avoids error accumulation by selecting the best view for a specific patch. The result is promising, as shown in [4], this approach outperforms major commercial image mosaicing software.

It is very difficult to choose an appropriate mosaicing algorithm for a specific purpose due to continuous new techniques being discovered [5].

## 3. Conclusions

Image registration and image mosaicing are two important topics in computer vision.

For image registration, area-based methods are usually preferred when there are not enough prominent details and local structural information in the images, and if the number of images being registered onto one image are large. Feature-based methods are usually more robust and accuracy against area-based methods. Among different feature-based methods, SIFT is the most accurate algorithm (in general) whereas ORB is the most efficient algorithm. ORB would be a better choice when the scaling of images is between 60%-150%. Feature descriptors play an important role in registration. It directly affects the efficiency and robustness of registration. There is no single best descriptor, so an appropriate descriptor needs to be chosen under different situations and applications.

For image mosaicing, registration and blending are two most important factors for the quality of final mosaic. Low quality registration or blending directly leads to low quality mosaic. It is very difficult to choose an appropriate mosaicing algorithm due to new techniques being continuously discovered. Two approaches have been discussed, which one uses partial local mosaic and homography, whereas the other one uses a combination of SfM techniques and standard image mosaicing techniques.

## REFERENCES

- [1] Barbara Zitová, Jan Flusser. 2003. Image registration methods: a survey. *Image and Vision Computing*, (Oct. 2003), Volume 21, Issue 11, 977-1000. DOI: [https://doi.org/10.1016/S0262-8856\(03\)00137-9](https://doi.org/10.1016/S0262-8856(03)00137-9).
- [2] Shaharyar Ahmed Khan Tareen and Zahra Saleem. 2018. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCOMET), 1-10. DOI: <https://doi.org/10.1109/ICOMET.2018.8346440>.
- [3] Yiping Lin and Gerard Medioni. 2007. Map-Enhanced UAV Image Sequence Registration and Synchronization of Multiple Image Sequences. 2007 IEEE Conference on Computer Vision and Pattern Recognition, 1-7. DOI: <https://doi.org/10.1109/CVPR.2007.383428>.
- [4] Zhengqi Li and Volkan Isler. 2016. Large Scale Image Mosaic Construction for Agricultural Applications. In *IEEE Robotics and Automation Letters*, (Jan. 2016), Vol. 1, no. 1, 295-302. DOI: <https://doi.org/10.1109/LRA.2016.2519946>.
- [5] Debabrata Ghosh and Naima Kaabouch. 2016. A survey on image mosaicing techniques. *Journal of Visual Communication and Image Representation*, (Jan. 2016), Volume 34, 1-11. <https://doi.org/10.1016/j.jvcir.2015.10.014>.
- [6] Tony Lindeberg. 2012. Scale invariant feature transform. Vol. 7, no. 5, p. 10491. DOI: <https://doi.org/10.4249/scholarpedia.10491>
- [7] Herbert Bay, Andreas Ess, Tinne Tuytelaars and Luc Van Gool. 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, (Jun.

- 2008), Volume 110, Issue 3, 346-359.  
DOI:<https://doi.org/10.1016/j.cviu.2007.09.014>.
- [8] Edward Rosten and Tom Drummond. 2006. Machine Learning for High-Speed Corner Detection. *Computer Vision – ECCV 2006*, 430–443. DOI: [https://doi.org/10.1007/11744023\\_34](https://doi.org/10.1007/11744023_34)
- [9] Michael Calonder, Vincent Lepetit, Christoph Strecha and Pascal Fua. 2010. BRIEF: Binary Robust Independent Elementary Features. *Computer Vision – ECCV 2010. ECCV 2010*, 778-792. DOI: [https://doi.org/10.1007/978-3-642-15561-1\\_56](https://doi.org/10.1007/978-3-642-15561-1_56)
- [10] Ethan Rublee, Vincent Rabaud, Kurt Konolige and Gary Bradski. 2011. ORB: An efficient alternative to SIFT or SURF. 2011 International Conference on Computer Vision, 2564-2571, DOI: <https://doi.org/10.1109/ICCV.2011.6126544>.
- [11] Chengcai Leng, Hai Zhang, Bo Li, Guorong Cai, Zhao Pei and Li He. 2019. Local Feature Descriptor for Image Matching: A Survey. In *IEEE Access*, vol. 7, 6424-6434. DOI: <https://doi.org/10.1109/ACCESS.2018.2888856>.
- [12] Ebrahim Karami, Siva Prasad and Mohamed Shehata. 2015. Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images. In *Proceedings of the 2015 Newfoundland Electrical and Computer Engineering Conference*, November 2015, St. Johns, Canada.
- [13] Manjusha Deshmukh and Udhav Bhosle. 2011. A Survey of Image Registration. *International Journal of Image Processing (IJIP)*, Volume 5, Issue 3, 245-269.
- [14] Shuang Wang, Dou Quan, Xuefeng Liang, Mengdan Ning, Yanhe Guo and Licheng Jiao. 2018. A deep learning framework for remote sensing image registration. *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 145, Part A, 148-164, DOI: <https://doi.org/10.1016/j.isprsjprs.2017.12.012>.
- [15] Bob D. de Vos, Floris F. Berendsen, Max A. VIERGVER, Marius Staring and Ivana Isgum. 2017. End-to-End Unsupervised Deformable Image Registration with a Convolutional Neural Network. In *Cardoso M. et al. (eds) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. DOI: [https://doi.org/10.1007/978-3-319-67558-9\\_24](https://doi.org/10.1007/978-3-319-67558-9_24)
- [16] Gang Hong and Yun Zhang. 2007. Combination of feature-based and area-based image registration technique for high resolution remote sensing image. *IEEE International Geoscience and Remote Sensing Symposium*, 377-380. DOI: <https://doi.org/10.1016/10.1109/IGARSS.2007.4422809>.