PRELIMINARY INVESTIGATION OF THE ROBUSTNESS OF MAXIMALLY STABLE EXTREMAL REGIONS (MSER) MODEL FOR THE AUTOMATIC REGISTRATION OF OVERLAPPING IMAGES

O. G. Ajayi [a], I. J. Nwadialor [b], I. C. Onuigbo [b], O. A. Kemiki [b]

[a] *Department of Surveying and Geoinformatics, Federal University of Technology, Minna, PMB 65, Minna, Nigeria, Nigeria*
[b] *Federal University of Technology, Minna, Nigeria*

Abstract: *Various researchers in Digital Image Processing have developedkeen interest in the automation of object detection, description and extraction process used for various applications, and this has led to the development of series of feature detection and extraction models, one of which is the Maximally Stable Extremal Regions Features Algorithm (MSER). This paper investigates the robustness of MSER algorithm (a blob-like and affine-invariant feature detector) for the detection and extraction of corresponding features used for the automatic registration of series of overlapping images. The robustness investigation was carried out in three different registration campaigns using overlapping images extracted from google earth online image data repository and image pairs acquired from an Unmanned Aerial Vehicle (UAV) flight mission. Sum of Square Difference (SSD) and Bilinear interpolation models were used to establish the similarity measure between the registered images, resampling of the pixel-values and computation of non-integer coordinates respectively while Random Sampling Consensus (RANSAC) algorithm was used to exclude the outliers and to compute the transformation matrix using affine transformation function. The results obtained from this preliminary investigation shows that the processing speed of MSER is quite high for Automatic Image Registration with a relatively high accuracy. While an accuracy of 61.54% was obtained from the first campaign with a processing time of 11.92 seconds, the second campaign gave an accuracy of 52.02% with a processing time of 11.20 seconds and the third campaign produced an accuracy of 55.62% with a processing time of 3.27 seconds. The obtained speed and accuracy shows that MSER is a very robust model and as such, can be deployed as a feature detection and extraction model in the development of an automatic image registration scheme.*

Ajayi, O.G et al. (2018). Preliminary investigation of the robustness of maximally stable extremal regions (mser) model for the automatic registration of overlapping images. *Geoplanning: Journal of Geomatics and Planning, 5*(1), 63-74. doi: 10.14710/geoplanning.5.1.63-74

1. INTRODUCTION

Recently, obtaining descriptors of features by analyzing the linear scale space of an image have proved to be robust and reliable in the recognition, extraction and matching of objects (Sivic & Zisserman, 2003) which is a very crucial stage in mosaic generation from overlapping images. One of the essential image processing operations in remote sensing is image registration. Diverse applications such as change detection, image fusion, etc. are made possible by mosaicking overlapping image pairs differently acquired under different imaging conditions and circumstances, at different time epochs, covering the same imaging area (Kumar, Manjunath, & Rao, 2003). The basic aim of image registration is to ensure that two overlapping image pairs are aligned and matched spatially such that analogous pixels in the overlapping image pairs will correspond to the exact imaged scene of interest (physical region). This alignment and matching is achieved by estimating scale, rotation and translation using a defined or selected appropriate transformation function. It is an unavoidable issue in various fields of application of digital image

processing, especially, those that involves pixel by pixel comparison of two or more images of the same scene (Dai & Khorram, 1999).

The determination of a transformation function which best aligns features in the base image with the conjugate features in the observed image is the major task of image registration. According Zitová & Flusser (2003), automatic image registration can be classified into both area based method which directly aligns the intensity or pixels of the image pair holistically, and feature based method which extracts higher-level structures from the image pair, and find similar features to execute the registration task which makes this method more useful when the detection and extraction of conjugate features are reliably possible. Nevertheless, low spatial resolution and the presence of blur and noise in the base and moving image pairs will negatively affect the accurate detection and extraction of corresponding or conjugate feature points (Zitová & Flusser, 2003). Brown (1992) presented the four major components of Image Registration as Feature Space, Search Space, Search Strategy and Similarity Measure. The selection of each of these model components is determined by the nature and properties of the images (Rao, Rao, Manjunath, & Srinivas, 2004).

Various feature detection and extraction models have been implemented for accurate generation of mosaic from overlapping image pairs. Some of these include Speeded Up Robust Feature (SURF) algorithm, Scale Invariant Feature Transform (SIFT) algorithm, Harris and Stephen Corner Detector, etc (Olaleye et al., 2015) (Ajayi et al., 2014). Though these models have proved to be relatively robust, current research focus still seeks to investigate the possibility of obtaining a more robust feature detection and extraction model, which necessitated the need for this study. The Maximally Stable Extremal Regions (MSER) model is a robust feature extraction algorithm. MSER extracts a number of co-variant regions from an image, and it is a blob-like, local and affine invariant feature detector which is invariant to illumination changes and image resolution. It scales well for both small and large objects (Varah & Grujić, 2013). An MSER is a stable connected component of some gray-level sets of the image and it is based on the idea of extracting regions which stay nearly the same through a wide range of thresholds. While MSER has been widely and successfully applied in different image processing applications (Mikolajczyk et al., 2005) (Fraundorfer & Bischof, 2005) some of which include tracking and 3D segmentation (Donoser & Bischof, 2006), retrieval or restoration of images (Nister & Stewenius, 2006), matching of wide baselines (Matas et al., 2004) and curvilinear structures (Lemaitre et al., 2011), object recognition (Obdrzalek & Matas, 2002), real-time visual surveillance (Salahat et al., 2015), Field Programmable Gate Array- FPGA (Kristensen & MacLean, 2007), cell detection and analysis (Kaakinen et al., 2013), etc, research efforts aimed at implementing it for mosaic generation or automatic image registration is relatively unknown. This paper presents some preliminary findings of the investigation of the robustness of MSER for the automatic registration of overlapping image pairs using images acquired from Trimble Ux-5 Unmanned Aerial Vehicles (UAV) and google earth online image data repository.

2. DATA AND METHODS

2.1. Mathematical Formulation of The MSER Model

Obdrzalek & Matas (2002) presented the formal definition of MSER and mathematical annotation of its terms such as image region, region boundary, extremal region and Maximally Stable Extremal Regions. According to (Kimmel, Zhang, Bronstein, & Bronstein, 2011), the following is the mathematical formulation of MSER model:
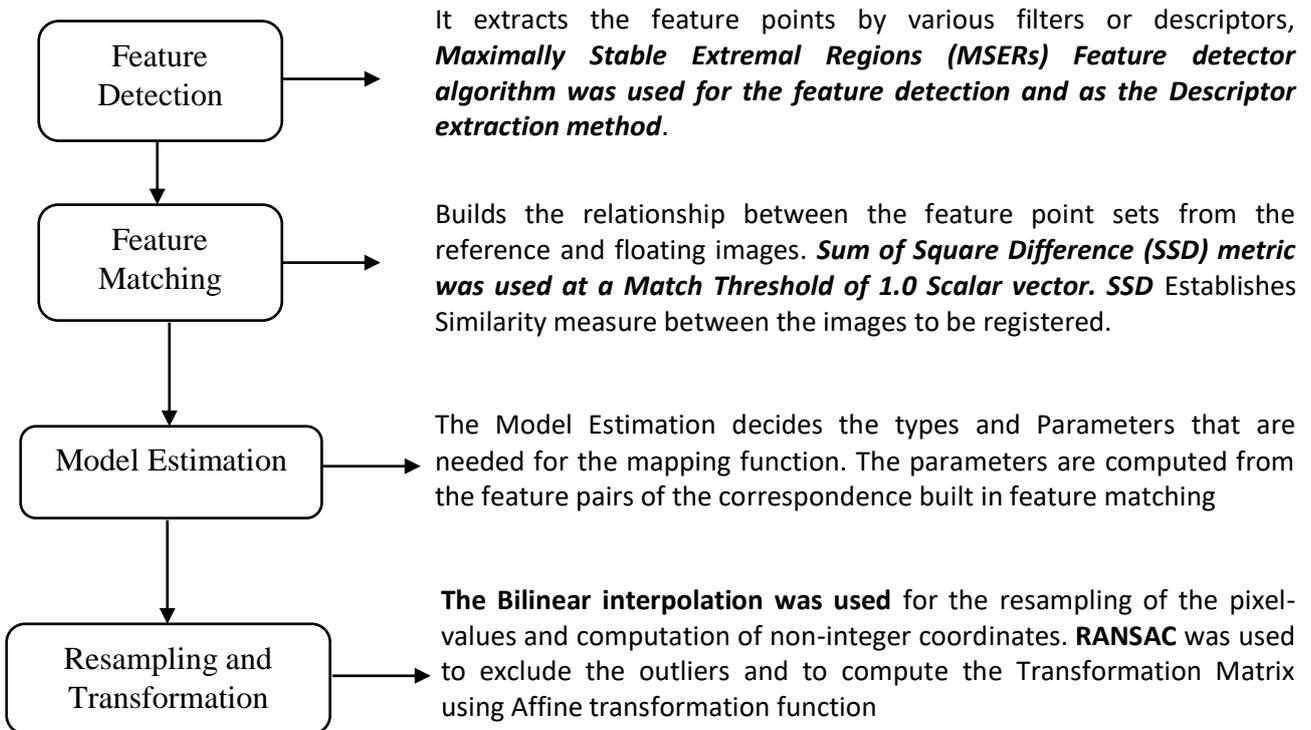
Let R_t be the family of connected components representing an edge in the component tree. Such regions are referred to as extremal by Obdrzalek & Matas (2002) since either: $I|_{int}(R_t) \prec I|_{\partial} R_t$ or $I|_{int}(R_t) \succ I|_{\partial} R_t$ which implies that the value of all the pixels within the regions are either completely darker or completely brighter than the pixel values along the boundary where the intensity is exactly equal to t.

Equation (1) gives the stability of a Region R_t:

$$\psi_1(R_t) = \frac{A(R_t)}{\frac{d}{dt}A(R_t)}$$, where $A(R_t)$ = Area R_t .... (1)

A region whose area remains slightly thesame with the change of the threshold t is considered stable while a region R_t is termed maximally stable if $\psi_1(R_t)$ (an affine – invariant property) has a local maximum at t since area ratio is retained under affine transformation. This suggests that for an affine transformation T of the domain X, the corresponding regions R and R' detected in images I d and $I(T^{-1})$, respectively will be related to $TR = R'$. These are the kind of regions that the MSER model detects.

Briefly highlighted in Figure 1 are the stages involved, techniques and the mathematical models used in the automatic registration of overlapping images. Figures 1 and 2 present a flowchart of the methodology.

Feature Detection	It extracts the feature points by various filters or descriptors, ***Maximally Stable Extremal Regions (MSERs) Feature detector algorithm was used for the feature detection and as the Descriptor extraction method***.
Feature Matching	Builds the relationship between the feature point sets from the reference and floating images. ***Sum of Square Difference (SSD) metric was used at a Match Threshold of 1.0 Scalar vector. SSD*** Establishes Similarity measure between the images to be registered.
Model Estimation	The Model Estimation decides the types and Parameters that are needed for the mapping function. The parameters are computed from the feature pairs of the correspondence built in feature matching
Resampling and Transformation	**The Bilinear interpolation was used** for the resampling of the pixel-values and computation of non-integer coordinates. **RANSAC** was used to exclude the outliers and to compute the Transformation Matrix using Affine transformation function

Figure 1: Stages and techniques used for the Automatic Image Registration

Detailed discussion of the Mathematical models of SSD, RANSAC, Bilinear Interpolation and Affine Transformation functions used can be found in (Ajayi et al., 2014). All the necessary computations and image registration were carried out using code scripts written in MATLAB R2014a environment.

2.2. Data Used for Experimentation

The model experimentation is subdivided into three (3) different image registration campaigns with overlapping image pairs showing different geographical areas used for the experimentation. Pairs of overlapping images with image size of 700x1028 pixels each, showing part of the University of Lagos, Akoka-Campus, Lagos Nigeria, extracted from google earth online image data repository was used for the first image registration campaign while overlapping image pairs showing part of the Federal University of Technology, Main Campus, Minna Nigeria, also extracted from google earth was used for the second image registration campaign. The images are of the size 2745x4800 each. Overlapping image pairs of part of The Federal Polytechnic Ado-Ekiti, acquired with the aid of Trimble UX5 UAV, was used for the model experimentation in the third image registration campaign. The image pairs for the first image registration campaign is presented as Figures 3a and 3b while the image pairs for the second image registration campaign is presented as Figures 4a and 4b. The image pairs used for the third registration campaign is presented as Figures 5a and 5b and its size is 800x532x3 pixels each.

Figure 3a: An Image of part of UNILAG Campus (The Base Image).

Figure 3b: An Image of part of UNILAG Campus (The Moving Image).

Figure 4a: Image part of FUTMinna, Main Campus (Base image).

Figure 4b: Image part of FUTMinna, Main Campus (Moving image).

Figure 5a: Image part of Federal Polytechnic Ado, Ekiti (Base image)

Figure 5b: Image part Federal Polytechnic, Ado Ekiti (Moving Image)

Affine transformation function was adopted for the computation of the model's transformation matrix (equation 2). It has six unknown parameters and requires a minimum of three reference points.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \dots\dots\dots (2)$$

Where: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ are scalar Quantities, $\begin{bmatrix} e \\ f \end{bmatrix}$ are Translation Parameters.

α = rotation angle, T_x = Translation in x axis and T_y = Transformation on y axis

Where: $a = \cos \alpha$, $b = \sin \alpha$ (α is the rotation angle), $c = T_x$ and $d = T_y$ (Translation in x and y axis respectively)

```
                              ┌─────────────────────┐
                              │   INPUT THE IMAGES  │
                              └─────────────────────┘
```
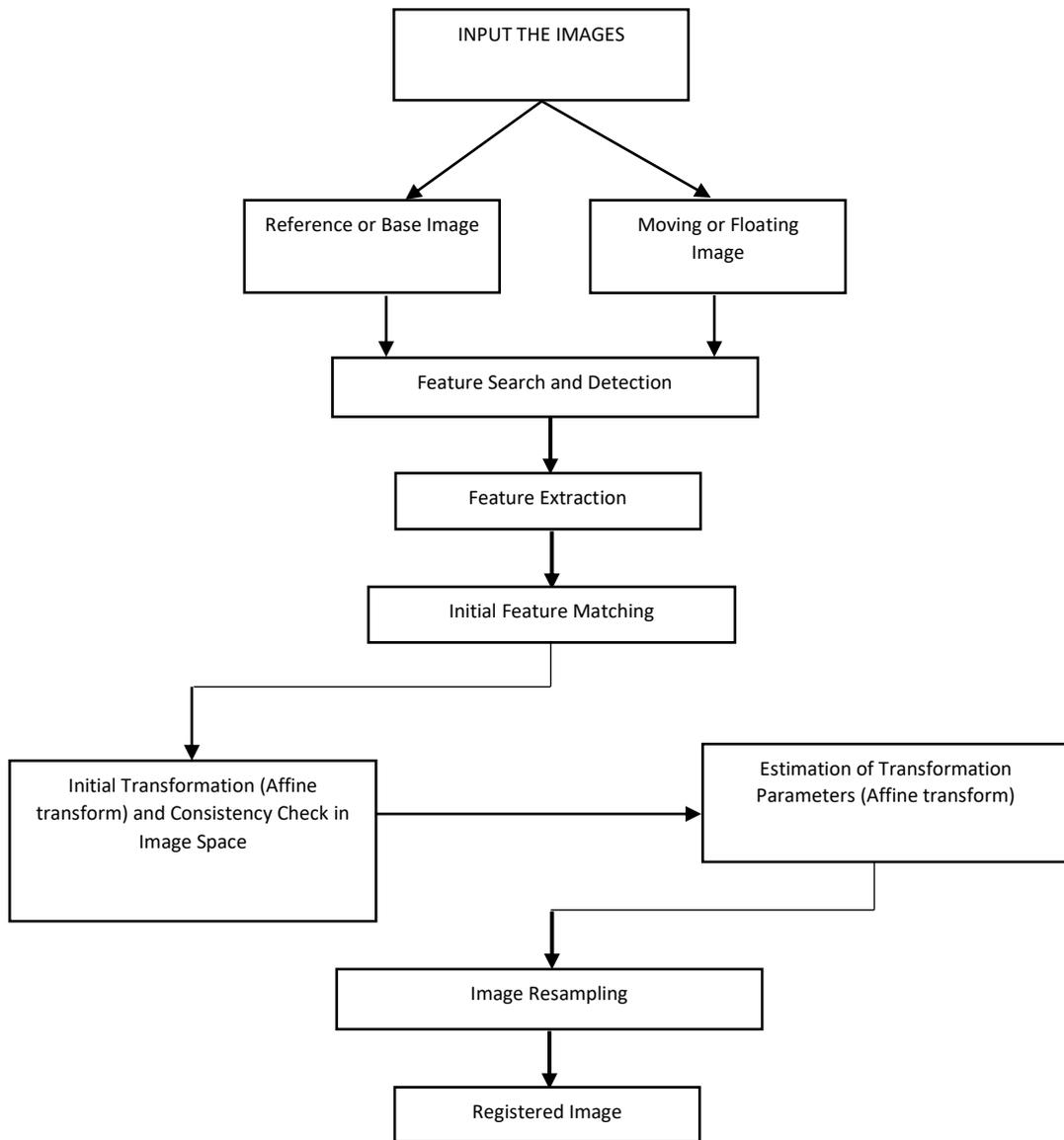

Figure 2: Flowchart of the step by step procedure of executing
feature based auto-registration (Adapted from Ajayi et al., 2014)

2.3. Accuracy Assessment

In order to ascertain the robustness of the registration model, an accuracy assessment method developed by (Olaleye et al., 2015) was adopted. This method (Equation 3) makes use of the total percentage of matched inliers, out of the total matched conjugate points (inliers and outliers) to determine how robust the feature extraction process is and in turn, define the degree of accuracy or accuracy measure of the entire image registration process since the success of the process depends largely on the robustness of the feature detection and extraction model and the model used for the exclusion of outliers.

$$x = \frac{a}{b}(100\%) \dots\dots\dots\dots (3)$$

Where x = the overall accuracy (%), a = total matched inliers (excluding outliers), and b = total matched points (both inliers and outliers).

3. RESULTS AND DISCUSSION

The results obtained from the first automatic image registration campaign using MSER as the feature detection and extraction model are as given in Figures 6a, 6b and 6c. Figure 6a shows the identified and extracted corresponding features, figure 6b presents the extracted corresponding points/features having excluded the outliers using RANSAC while Figure 6c presents the Mosaic generated by registering the two overlapping images. All units are measured in pixels.
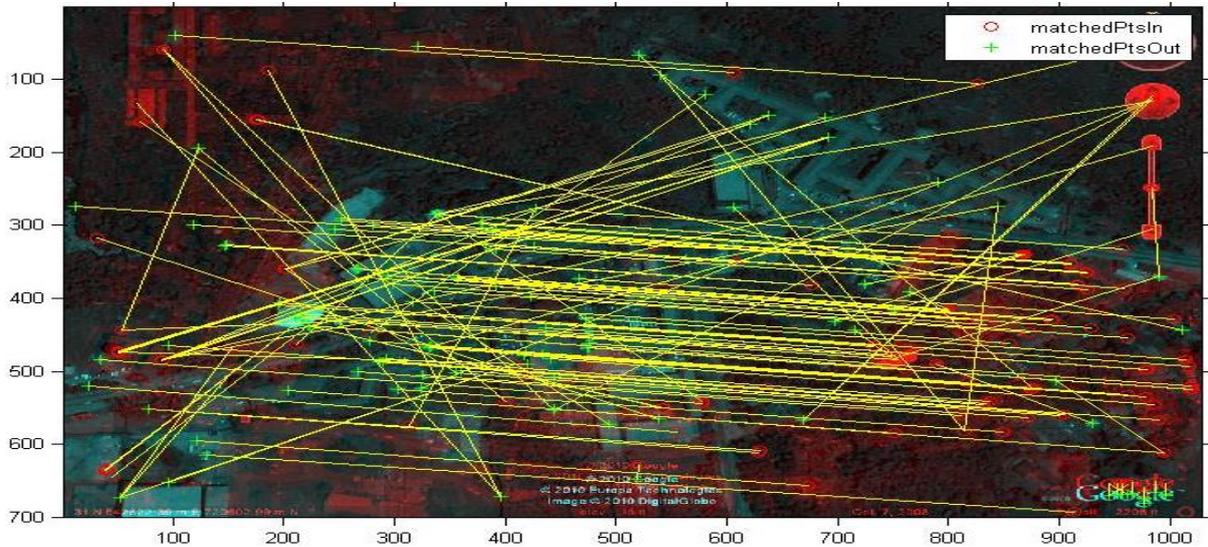
Figure 6a: All Matched and extracted corresponding points (Inliers and Outliers) for the first campaign (In pixels).
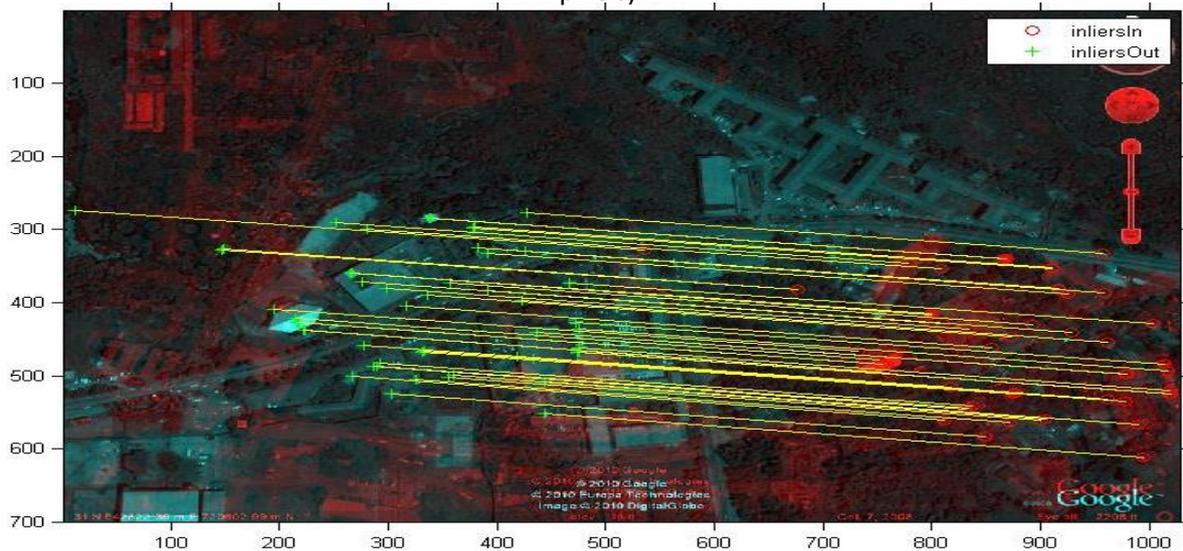
Figure 6b: Extracted Corresponding points used for the computation of Transformation Matrix (outliers excluded) for the first campaign (in pixels).

Figure 6c: The registered Image or Mosaic generated from the overlapping image pairs for Campaign 1 (in pixels) using the developed model.
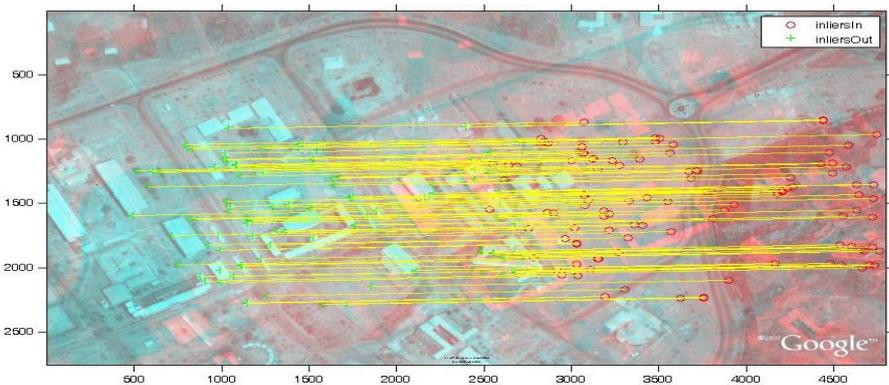
With a registration accuracy of 61.44%, the computed transformation matrix of the first registration campaign is:

$$T_1 = \begin{bmatrix} 1.0171 & 0.0756 & 502.0217 \\ 0.0009 & 1.0169 & 49.8708 \\ 0 & 0 & 1.0000 \end{bmatrix}$$

The results obtained from the second image registration campaign are presented in Figures 7a-7c. Figure 7a presents all the matched corresponding features (inliers and outliers inclusive) while figure 7b presents the extracted corresponding points/features, having excluded the outliers using RANSAC. Figure 7c presents the Mosaic generated by registering the two overlapping images. All units are measured in pixels.
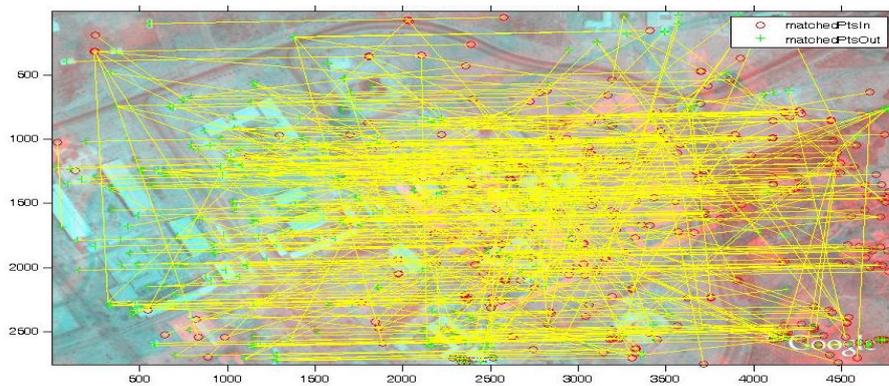
Figure 7a: All matched corresponding points (inliers and outliers inclusive) for campaign 2 (in pixels).

**Figure 7b**: Matched inliers used for the image registration for campaign 2 (in pixels).

Figure 7c: Mosaic generated for campaign 2 (in pixels) using the developed model.

With a registration accuracy of 52.02%, the computed transformation matrix of the second registration campaign is:

$$T_2 = \begin{bmatrix} 0.0010 & 0.0000 & 2.0287 \\ -0.0000 & 0.0010 & -0.0504 \\ 0 & 0 & 0.0010 \end{bmatrix}$$

The results obtained from the third registration campaign are presented in Figures 8a-8c. While figure 8a presents all the matched corresponding features (inliers and outliers inclusive), figure 8b presents the extracted corresponding points/features, having excluded the outliers using RANSAC and the Mosaic generated by registering the two overlapping images is presented in figure 8c. All units are measured in pixels. The accuracy obtained from the third registration campaign is 55.62% while computed transformation matrix of the registration for the third registration campaign is:

$$T_3 = \begin{bmatrix} 1.0051 & 0.0041 & 115.8039 \\ -0.0014 & 1.0197 & 15.3436 \\ 0 & 0 & 1.0000 \end{bmatrix}$$
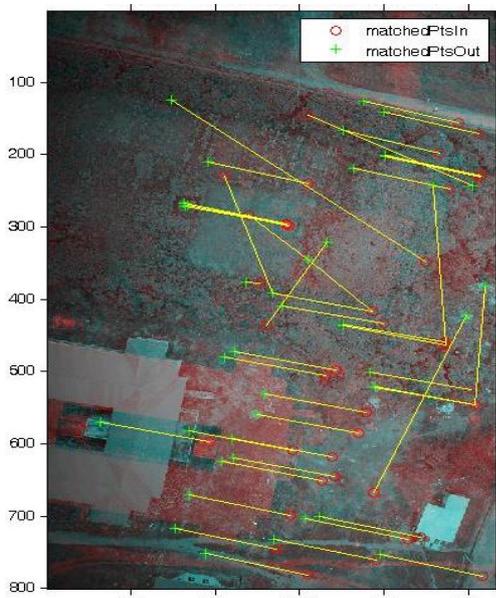
Figure 8a: All matched corresponding points (Camp. 3)
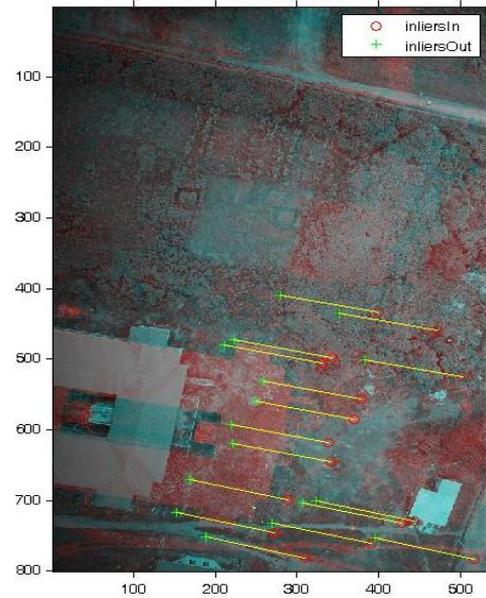
Figure 8b: Matched Inliers (Campaign 3) in pixels

Figure 8c: Mosaic generated from the UAV acquired image pair (Campaign 3) in pixels

A total of 191, 662 and 27 corresponding points were automatically extracted from the overlapping images of the first, second and third image registration campaigns respectively. The number of matched corresponding points of the second image registration campaign was more than the matched corresponding features of the first and third campaigns because the image pairs used for the second registration campaign have a larger total area coverage or image size (approximately 2745x4800 pixels) compared to the image pairs used for the first and third registration campaigns which have image sizes of 700x1028 and 800x532 pixels respectively. Also, the image pairs used for the second registration campaign have a higher overlapping percentage, and as such, the model has sufficient latitudes for the extraction of corresponding features. It was observed that the number of matched points using MSER are very few which makes it an excellent model for applications where only small matches are needed such as the computation of epipolar geometry (Mikolajczyk et al., 2005). This suggest that integrating MSER with epipolar correlation in the development of an automatic image registration scheme will result into a highly robust model with respect to time and accuracy.

Approximately 61.44%, 52.02% and 55.62% of the matched corresponding points for the first, second and third registration campaigns respectively were used for the automatic image registration. These percentages were obtained after the exclusion of outliers (mis-matches) from all the matched points using Random Sampling Consensus (RANSAC). This shows that 38.56%, 47.98% and 44.38% of all the matched points were outliers for the first, second and third image registration campaigns respectively. This imply that MSER automatically extracted more outliers during the second image registration campaign which can be attributed to the quality of the spatial resolution of the used overlapping image pairs. They also depict the accuracy level or measure of reliability of the automatic registration model (Olaleye et al., 2015).

The time taken for the complete execution of the automatic image registration was approximately 11.92 seconds (speed of the developed registration scheme) for the first registration campaign, about 11.20 seconds was expended on the second campaign and 3.27 seconds was used for the third campaign. The processing time of each of the registration campaigns was determined automatically using the system run time generated by the MATLAB report. This is in tandem with the results obtained by (Mikolajczyk et al., 2005) which attests to the speed of MSER when compared to 5 other detectors. An improved computational time was also recorded by Nistér & Stewénius (2008) and (Kaakinen et al., 2013) and this is as a results of MSER's negligible need for pre-optimization procedures which dramatically reduces its computation time (Kaakinen et al., 2013).

Though MSER is regarded as one of the best region detectors due to its robustness against view point, scale and lightning changes, this fair results obtained especially in registration campaign 2 can be attributed to

the sensitivity of MSER algorithm to image blurs (Śluzek, 2016) and as such, the resolution of the image may be influential to the process of determining the efficiency of the successfully matched inliers and consequently, the registration result. This is evident in the fact that better accuracies were recorded in the first and third registration campaigns compared to the accuracy obtained from the second registration campaign because the image pairs used for the second registration campaign are quite blurry.

4. CONCLUSION

The result of the preliminary investigation of the robustness of MSERs feature identification, detection and extraction algorithm for the extraction of corresponding features used for the automatic registration of overlapping images is herein presented. All computation was done with code scripts written in MATLAB R2014a environment. Though the percentage matched inliers used for the registration of the images were quite good (61.44%, 52.02% and 55.62% for the first, second and third registration campaigns respectively), registration campaign 2 produced the least satisfactory results of the three campaigns which is an evidence that MSER is highly sensitive to image blurs (Lemaitre et al., 2011), since the spatial resolution of the image pairs used for the second campaign is not as high as the resolution of the image pairs used for the first and third campaigns. It is however noted that MSER is highly robust and can be a choice model for feature detection and extraction in an automatic image registration scheme.

For further studies, attempt shall be made to investigate the possibility of pre-defining a minimum threshold (radius) as the acceptable region for smaller MSERs so as to avoid unnecessarily small regions. The robustness of edge-enhanced Maximally Stable Extremal Regions which should help the sensitivity of this algorithm to image blurs and noise will also be investigated.

5. REFERENCES

Ajayi, O. G., Odumosu, J. O., Okorocha, C. V, Nzelibe, I. C., Ahmadu, H. A., & Bawa, S. (2014). Semi-Automatic Generation of Mosaic from Overlapping Images Using MATLAB. *Int J Engg Advance Tech Studies*, 2(4), 1–15.

Brown, L. G. (1992). A Survey of Image Registration Techniques. *ACM Computing Surveys*, 24(4), 325–376. [Crossref]

Dai, X., & Khorram, S. (1999). A Feature-Based Image Registration Algorithm Using Improved Chain-Code Representation Combined with Invariant Moments. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5), 2351–2362. [Crossref]

Donoser, M., & Bischof, H. (n.d.). Efficient Maximally Stable Extremal Region (MSER) Tracking. in 2006 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* - Volume 1 (CVPR\textquotesingle06). IEEE. [Crossref]

Fraundorfer, F., & Bischof, H. (n.d.). A novel performance evaluation method of local detectors on non-planar scenes. In 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR\textquotesingle05) - Workshops. IEEE*. [Crossref]

KAAKINEN, M., HUTTUNEN, S., PAAVOLAINEN, L., MARJOMÄKI, V., HEIKKILÄ, J., & EKLUND, L. (2013). Automatic detection and analysis of cell motility in phase-contrast time-lapse images using a combination of maximally stable extremal regions and Kalman filter approaches. *Journal of Microscopy*, 253(1), 65–78. [Crossref]

Kimmel, R., Zhang, C., Bronstein, A. M., & Bronstein, M. M. (2011). Are MSER Features Really Interesting? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11), 2316–2320. [Crossref]

Kristensen, F., & MacLean, W. J. (2007). Real-Time Extraction of Maximally Stable Extremal Regions on an FPGA. In 2007 *IEEE International Symposium on Circuits and Systems. IEEE*. [Crossref]

Kumar, A. S., Manjunath, A. S., & Rao, K. M. M. (2003). Merging IRS multispectral and PAN images by A-Trous wavelets. *International Journal of Remote Sensing*.

Lemaitre, C., Perdoch, M., Rahmoune, A., Matas, J., & Miteran, J. (2011). Detection and matching of curvilinear structures. *Pattern Recognition*, 44(7), 1514–1527. [Crossref]

Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10), 761–767. [Crossref]

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., … Gool, L. Van. (2005). A Comparison of Affine Region Detectors. *International Journal of Computer Vision*, 65(1–2), 43–72. [Crossref]

Nister, D., & Stewenius, H. (n.d.). Scalable Recognition with a Vocabulary Tree. In 2006 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* - Volume 2 (CVPR\textquotesingle06). IEEE. [Crossref]

Nistér, D., & Stewénius, H. (2008). Linear Time Maximally Stable Extremal Regions. *In Lecture Notes in Computer Science* (pp. 183–196). Springer Berlin Heidelberg. [Crossref]

Obdrzalek, S., & Matas, J. (2002). Object Recognition using Local Affine Frames on Distinguished Regions. In Procedings of the British Machine Vision Conference 2002. *British Machine Vision Association*. [Crossref]

Olaleye, J. B., Ajayi, O. G., Omogunloye, O. G., Odumosu, J. O., & Okorocha, C. V. (2015). AUTOMATIC REGISTRATION OF SIMULTANEOUSLY OVERLAPPING IMAGES. *NED University Journal of Research*, 12(4).

Rao, C. V., Rao, K. M. M., Manjunath, A. S., & Srinivas, R. V. N. (2004). Optimization of automatic image registration algorithms and characterization. *In Proceedings of the ISPRS Congress* (pp. 698–702).

Salahat, E., Saleh, H., Sluzek, A., Al-Qutayri, M., Mohammad, B., & Ismail, M. (2015). A maximally stable extremal regions system-on-chip for real-time visual surveillance. *In IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society. IEEE*. [Crossref]

Sivic, & Zisserman. (2003). Video Google: a text retrieval approach to object matching in videos*. In Proceedings Ninth IEEE International Conference on Computer Vision. IEEE*. [Crossref]

Śluzek, A. (2016). Improving performances of MSER features in matching and retrieval tasks. *In European Conference on Computer Vision* (pp. 759–770).

Varah, S., & Grujić, N. (2013). Target Detection and Tracking Using a Parallel Implementation of Maximally Stable Extremal Region. *In GPU Technology Conference, Canada*.

Zitová, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11), 977–1000. [Crossref]