INVESTIGATION OF IMAGE STITCHING REFINEMENT WITH ENHANCED CORRELATION COEFFICIENT

Stephen Khor Wen Hwooi[1*], Aznul Qalid Md Sabri[2*]

[1,2]Faculty of Computer Science and Information Technology, University of Malaya
Kuala Lumpur, Malaysia

E-mail: stphnkhrwnhw@gmail.com[1*] (corresponding author), aznulqalid@um.edu.my[2*] (corresponding author)

DOI: https://doi.org/10.22452/mjcs.vol33no1.2

ABSTRACT

Image stitching combines multiple overlapping scenes into a panorama. It can be applied in video stitching, super-resolution, and medical imaging. Post-processing routines such as exposure compensation, seam line adjustment, and blending are often performed to increase its visual appeal. Our aim is to increase the registration accuracy between adjacent image pairs in a stitching algorithm. This is done by adding an area-based registration step, namely Enhanced Correlation Coefficient (ECC), which refines the original feature-based image registration. The performance of registrations in the stitching algorithm is evaluated using root-mean-squared error of control points, structural similarity index, and universal image quality index. The Boat and Graffiti datasets from the Oxford Robotics Database are used for experiments. It is found that ECC largely improves registrations in stitching except for a slight increase in root-mean-squared error of control points in the Boat dataset. In addition, ECC's enhancement makes the registration outperform the ground truth in one image of both datasets.

Keywords: Image registration, Image stitching

1.0 INTRODUCTION

Image stitching is the compositing of a set of images into a panorama [1]. It is used in generation of photo spheres, photo mosaics, and image stabilisation. The first step of a generic algorithm is to detect and describe features in each image. These features are matched between overlapping images. Then, the homography between adjacent images is estimated through Random Sample Consensus (RANSAC) [2] or Maximum Likelihood Estimation Sample Consensus (MLESAC) [3].

Post-processing routines are applied to the composited images to make the panorama more visually appealing. For example, bundle adjustment is applied to refine the camera parameters which may have accumulated errors from the concatenation of pairwise homographies. Exposure differences between image pairs are compensated by applying colour correction. Besides that, the seam line between images is adjusted to minimise its visibility. Lastly, the images are blended and projected into a composite surface once all post-processing adjustments have been made [4].

Our paper's objective is to increase the accuracy of image registrations in stitching. This increase in accuracy leads to less ghosting issues for post-processing routines. In addition, it contributes to more structural similarity, quality, and registration accuracy between overlapping regions of images.

Our stitching algorithm takes a set of images as input. Feature detection, description and matching is performed between each image and the next. After that, a geometric transformation and inliers are computed from the matched points. These transformations are concatenated to compute transformations relative to the first image. The output limits of each transform is used to determine the centre image of the panorama.

The surface of the panorama into which all images are mapped is created. The size of the panorama is computed by finding the output limits over all transformations. Finally, the transformations map and overlay images together on the surface of the panorama.

Our experimental protocol involves datasets from the Oxford Robotics Database, namely the Graffiti and Boat datasets which test for affine and scale change respectively [5]. Traditional evaluation metrics such as structural similarity [6] and image quality index [7] are used to quantify the quality of registrations in the stitched panorama. Additionally, the root-mean-squared-error of control points is computed between registered images to assess its accuracy [8]. The Maximally Stable Extremal Regions ($mser$) [9], Speeded-Up Robust Features ($surf$) [10], and

Harris ($harris$) [11] detectors and $surf$ descriptor are chosen as the feature-based image registration methods for the datasets.

The novelty of our approach lies mainly in the use of ECC [12] as the area-based image registration refinement method. The number of iterations for ECC is set to the default setting of 15 iterations. It is predicted that ECC's refinement on registrations increases the structural similarity, quality, and registration accuracy between adjacent images. This is amplified by the cumulative nature of alignments as more images are added for stitching.

The following sections are structured as follows: Section 2 provides a literature review of related works with explanations on the research gap and similarities with other works. Section 3 details each step in the process of our proposed stitching algorithm and the experimental protocol employed. Section 4 shows the results of our approach compared with benchmark algorithms in two widely-used image registration datasets. Section 5 gives justifications for experimental results and possible improvements in experimental design. Section 6 summarises the findings of our investigation, possible improvements in the experimental protocol, and applications of this study.

2.0 LITERATURE REVIEW

With the ever increasing hardware processing power of recent times, the relevance of our accuracy refinement approach cannot be overstated as it could fill a gap in research by increasing registration accuracy in stitching for applications in medicine, remote sensing, and motion estimation.

Several recent approaches have used ECC for registration in the medical field. A paper used Arterial Spin Labelling Magnetic Resonance Imaging (ASL MRI) to measure perfusion in kidneys. In this work, the affine version of ECC was used to coregister ASL, equilibrium magnetisation (M0), and relaxation time (T1) maps [13, 14].

Aside from that, ECC was used as the similarity measure between adversarial and original images in a recent work on data augmentation where Perceptive Adversarial Similarity Score (PASS), hard positive generation, and a diverse set of adversarial perturbations were introduced [15].

Besides that, the feature-based methods which we use as benchmarks in experiments are well known and widely used in current literature. For example, improved $surf$ was proposed as a way to stitch varied scenes quickly. They used Support Vector Machines (SVM) to simplify matching [16]. The robustness of this approach to handle scene variability can be improved by our proposed approach of ECC refinement. Multimodal image stitching in the agricultural sector was investigated with the aim of weed control [17]. The location of the weeds can be triangulated more accurately if the registration is enhanced with ECC.

A high performance stitcher was created using the Oriented Fast and Rotated Brief (ORB) feature detector and Exposure Compensation blending [18]. ECC refinement can potentially increase ORB's accuracy seeing that it has high efficiency but lacks high accuracy. A real-time stitching algorithm was proposed using ORB and Progressive Sampling Consensus (PROSAC) to reduce non-randomness in the sampling procedure [19]. Their selected feature-based approach (ORB) can be enhanced with our proposed ECC refinement algorithm.

A study with the aim of increasing the field of view of fundus images was carried out using a combination of the weber descriptor and seamless blending [20]. Our proposed ECC refinement approach can provide the high accuracy usually required for detailed analysis in the medical field. Image stitching was performed in an automotive vehicle using multiple monitors to solve the problem of blind spots in rear-view mirrors [21]. They used $surf$ and Haar wavelet transform to stitch the images. Our enhancement approach may increase the accuracy of stitching, thereby reducing the possibility of ghosting issues. This in turn would reduce the chances of a car accident.

We chose to compare the registration performance of our proposed stitching approach with the $surf$ detector, $surf$ descriptor, $mser$ detector, and $harris$ detector because they are still widely-used either as benchmark algorithms or part of the solution in recent papers. $surf$ was used to detect global control points before local matching using diffeomorphic Demons. This was done with the aim of handling general and nonlinear deformation during image registration [22]. Brain tissue slice registration was investigated by first performing Wavelet-based Edge Correlation (WEC) for detection before using the $surf$ descriptor. After that, matching was performed considering local constraints [23]. In order to solve the problem of low feature matches, Random Resample Consensus (RANRE-SAC) was proposed to regenerate matches. The texture component of the $surf$ descriptor was used to make their approach robust to noise [24].

An approach tackling nonlinear changes in remote sensing images evaluated three state-of-the-art descriptors. They also proposed OSURF for colour images, which is a combination of the $mser$ detector and the $surf$ descriptor [25].

In an investigation comparing the performance of region and interest point detectors, the $mser$ detector was compared with the $surf$ detector. Their suitability in the presence of global and local distortion was studied [26]. Multispectral registration was improved when the $mser$ detector was used on an edge-enhanced image. They used the original $mser$ detector as one of their benchmark algorithms [27].

In a study on image change detection, images from the same region but different times were registered using the $harris$ detector and the sift descriptor. After that, the difference image was extracted and its dimension was reduced using Principal Component Analysis (PCA). The feature vector space was classified using hybrid genetic fuzzy C-means (FCM) algorithm before being compared with the reference image [28]. The registration of cadastral images was investigated by first using $harris$ to perform feature detection. Then, RANSAC was used to estimate the geometric transformation before stitching [29]. In crack image matching, the uneven lighting in images were first reduced using Mask Dodging. After that, the noise was removed and the binary image was extracted. $harris$ was used for feature detection before image matching using Hausdorff distance [30].

In order to justify the usage of selected feature-based registration algorithms as benchmarks, we present the state of the art feature-based image registration algorithms being used for various state of the art image stitching research works in Table 1.

A direct comparison with the image stitching problems in Table 1 is non-valid since the goal of the paper is to propose the usage of an additional area-based registration refinement step, namely ECC, in a standard image stitching framework. This is done to improve the image stitching performance by way of improving the registration accuracy between adjacent images of a dataset. Thus, comparison with state of the art feature-based image registration approaches is sufficient to demonstrate the effectiveness of our proposed approach.

Table 1. Feature-based image registration methods and their usage in image stitching papers.

Feature-based method	Image stitching paper
$surf$	Multimodal Image Stitching Algorithm for Weed Control Applications in Organic Farming by Holtorf et. al. 2016 [17]
	Seamless Image Stitching Using Structure Deformation with HoG Matching by Lee et. al. 2015 [41]
	An Adaptive Image-stitching Algorithm for an Underwater Monitoring System by Li et. al. 2014 [37]
$mser$	Real Time Image Mosaicing System Based on Feature Extraction Techniques by Adel et. al. 2014 [42]
	Image Stitching System Based on ORB Feature-Based Technique and Compensation Blending by Adel et. al. 2015 [18]
	Research on Image Matching Technology for the Spherical Stereo Vision by Zhang et. al. 2015 [43]
$harris$	Image Registration and Mosaicking Based on the Criterion of Four Collinear Points by Jinwei et. al. 2016 [35]
	An Image Registration Method For Engineering Images by Jiang and Lin 2016 [29]
	Image Mosaic Algorithm Based on the Combination of Harris and Edge Extraction by Zhihong and Songming 2016 [34]

On another note, there are many approaches which are conceptually similar to our approach. We aim to identify the similarities and differences with these approaches in the following paragraphs. Shunzhi proposes a hessian algorithm which combines grayscale image matching with feature-based detection [31]. This results in more matches and a quicker search for matches. It is similar to our approach in that it aims to increase registration accuracy by finding more accurate feature locations. However, it seeks to refine feature point locations whereas our algorithm aims to refine the geometric transformation.

Bae suggests a piecewise 3D registration approach. Each image can be treated as a 2D segment defined by features [32]. Then, the segments are stitched together based on their inclination in 3D space. It seeks to find a geometric relationship between images. It differs in that it uses feature segments in 3D instead of 2D features used in our algorithm.

Improvement in stitching of digital radiography X-ray images is proposed by Yang [33]. First, the images are downsampled to reduce computation time. The overlap region is identified through improved phase correlation. This region is evaluated using the correlation coefficient which is the basis for ECC in our algorithm. Our approach utilises feature-based image registration to recognise overlap regions instead of phase correlation.

A mosaicking algorithm which combines Canny edge detection and $harris$ corner extraction is proposed by Zhi-hong [34]. The corners are locally optimised before NCC and RANSAC. Lastly, image fusion's weighted average method is used. It uses feature-based and area-based methods for registration. Our algorithm utilises a combination of a feature-based image registration method and ECC. In contrast, they use Canny, $harris$ and NCC.

Jinwei proposes to refine feature points for robust feature matching in image stitching [35]. Firstly, $harris$ corners are detected. Collinear point sets which are larger than four points are extracted using the Cross-Ratio criterion. Four groups of points from this set is used to compute the homography. Results show that this approach makes it robust to a high number of mismatches. It seeks to refine the registration method. Our approach refines the homography instead of feature points.

A fast image registration algorithm based on block processing and iterative improvement is proposed by Han [36]. In the same way, ECC is iterative in its refinement and relies on a good initialisation. We use feature-based image registration rather than block processing.

Hengyu suggests using $surf$ to stitch images if the number of features are larger than a threshold [37]. Otherwise, the phase correlation method is used. Our approach uses the 64-dimension $surf$ as the feature descriptor. They use different modes of stitching based on the situation. In contrast, we integrate both feature-based and area-based methods.

A mosaicking algorithm for visual sensor networks is proposed by Yi-qin [38]. The overlap regions are first identified with the strip search algorithm to reduce its complexity. Image registration is performed using block search. The registration is refined with improved absolute difference value sum algorithm. It is similar to our approach in that it suggests a refinement of image registration. However, our algorithm uses ECC for refinement.

Xingxing proposes using Location Sensitive Hashing to accelerate feature matching [39]. The algorithm uses Euclidean distance ratio and median filter to produce a stable homography. This is like our algorithm which aims to enhance the homography. They seek to speed up the stitching process while we aim to improve the registration accuracy.

A multiresolution approach integrating both feature-based and area-based registration is proposed by Xianyong [40]. It utilises block matching, a low contrast filter, and RANSAC to refine feature matches. This is similar to our approach of combining feature-based and area-based methods. However, we use ECC for refinement and MLESAC for transformation estimation.

3.0 METHODOLOGY

3.1 Proposed Stitching Algorithm

The image stitching algorithm is based on MATLAB's automatic feature-based panoramic image stitching frame-work from the computer vision system toolbox as shown in Fig. 1 where I and T are the image and transformation matrix respectively. Subscript n is the image order and starts from the value of two. It can handle uncallibrated cameras. Hence, it is assumed that there is no lens distortion in the images.

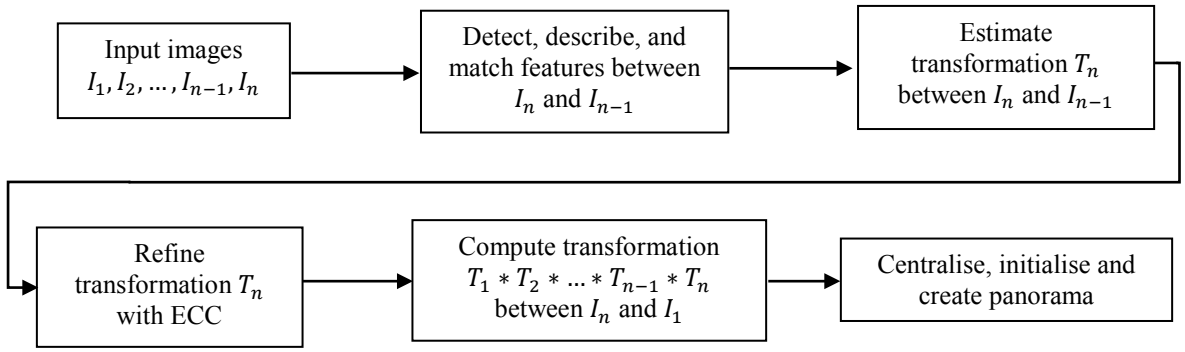

Fig. 1. Flowchart of image stitching algorithm with ECC refinement.

After the images are loaded, feature-based image registration is applied sequentially on the images starting from the first image. Detectors that are considered include $mser$, $surf$, and $harris$. The 64-dimension $surf$ is used to describe the detected interest points.

$mser$ is a blob detector. It detects covariant regions which are stable connected components on some level sets of an image. In other words, the detected regions stay similar even when different intensity thresholds are applied. $mser$ can be either brighter or darker than their outer boundary. Rejection criteria include the allowable range and variation of its area.

Wavelet responses in a $20s \times 20s$ neighbourhood is computed around a detected keypoint, where s is a predefined pixel size. After that, $surf$ divides the region into 4×4 subregions where each subregion is a 4-element vector as shown in Eq. 1. It uses the trace of Hessian to only match descriptors with the same contrast type i.e. bright blobs on dark backgrounds or vice versa.

$$v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|\right) \tag{1}$$

Feature matches are exhaustively searched by computing the sum of squared differences, SSD, between both feature vectors. To determine the uniqueness of a match, a distance ratio, d, as defined in Eq. 2 is utilised, where f_1, f_2, and f_2' are the reference image feature, best target image feature match, and second best target image feature match respectively. If d is close to one, a match is ambiguous. If d is close to zero, it is considered unique. In addition, if the SSD is larger than a predefined matching threshold, it is considered a false positive.

$$d = \frac{SSD(f_1, f_2)}{SSD(f_1, f_2')} \tag{2}$$

MLESAC uses feature matches to compute the geometric transformation and inliers. It estimates putative solutions by maximising likelihood. Then, it performs parameterisation through constrained optimisation.

Transformation matrices spatially transform coordinates of a target image from (x, y) to (u, v) based on a motion model. The projective model, which preserves straight lines, is used for this algorithm. Eq. 3 shows components of a homography matrix in homogeneous form. a and e control scale. g and h are responsible for translation.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

The formula for rotation of θ degrees is defined in Eq. 4.

$$\begin{bmatrix} cos\ \theta & sin\ \theta & 0 \\ -sin\ \theta & cos\ \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

Then, the pairwise transformation is refined by forward-additive ECC. Implementation for ECC is taken from the Image Alignment Toolbox (IAT). Firstly, the normalised reference image vector, $\widehat{\iota_r}$, is computed using its zero-mean vector, $\overline{\iota_r}$, as shown in Eq. 5.

$$\widehat{\iota_r} = \frac{\overline{\iota_r}}{\|\overline{\iota_r}\|} \tag{5}$$

The mapping function is defined as $\phi(x: p_{j-1})$ where x and p_0 are the reference image coordinates and initial warp respectively. j is set to one for the first iteration. The Jacobian, $\bar{G}(p_{j-1})$, is computed from this mapping function. After that, the zero-mean of the warped image vector, $\overline{\iota_w}(p_{j-1})$, is computed.

$\widehat{\iota_r}^t \overline{\iota_w}$ is compared with $\widehat{\iota_r}^t P_G \overline{\iota_w}$ where $P_G = \bar{G}(\overline{G}^t \bar{G})^{-1} \overline{G}^t$. The result of this comparison determines the computation method for Δp_j. Finally, the warp is updated as shown in Eq. 6. The algorithm is terminated by setting the maximum number of iterations, j_{max}, and a predefined Mean-Square-Distance threshold, T_{MSD}, for $\|\Delta p_j\|$where $j_{max} = 15$ and $T_{MSD} = 1$ pixel^2 as suggested in [12].

$$p_j = p_{j-1} + \Delta p_j \tag{6}$$

Refined pairwise transformations between adjacent images are concatenated to compute the transformation relative to the first image.

Transformations are modified by making the centre image the least distorted one. This is done by inverting the centre image transformation and applying to the other transformations. The centre image can be found through the output limits of each transformation. An empty panorama is initialised. Its size is determined by computing the maximum and minimum output limits over all transformations. Lastly, the transformations are used to map and overlay images together into the panorama.

3.2 Experimental Protocol

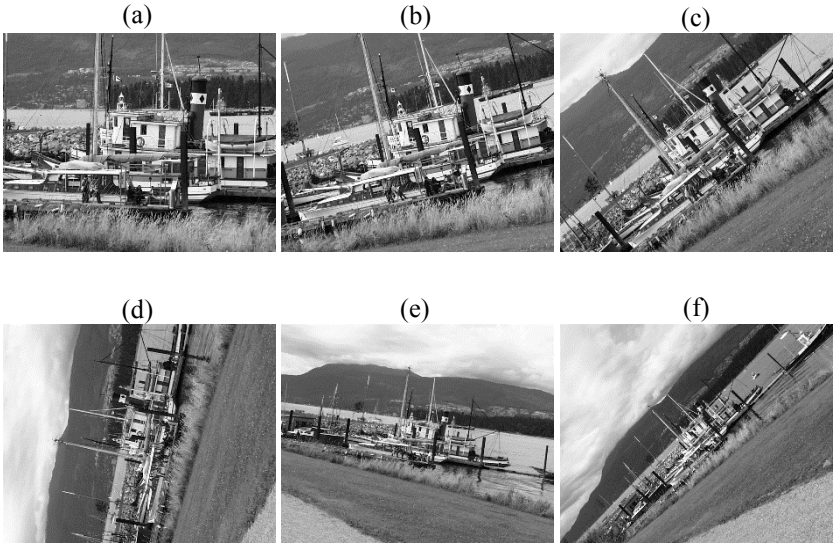

Fig. 2. Images from the Boat dataset. (a) is the reference image. (b) - (f) are target images to be registered to the reference image.

We selected state of the art feature-based image registration approaches, namely the $mser$, $surf$, and $harris$ detector paired with the 64-dimension surf descriptor, as benchmark algorithms in our experiments.

The performance of the image stitching algorithms are evaluated using three metrics, namely root-mean-squared error of control points, $rmse$, structural similarity index, $ssim$, and universal image quality index, Q. The overlapping region between the reference and the target image is evaluated for all metrics. The Boat and Graffiti dataset as shown in Fig. 2 and Fig. 3 are used.

Root-mean-squared error of control points, $rmse$, evaluates the registration accuracy between two images. It is computed by first selecting n ground truth control points in the reference image and target image using the ground truth homography provided along with the datasets. Then, the distance from ground truth produced by the estimated transformation is computed according to Eq. 7 where e is the Euclidean distance error of corresponding points. n is set to 20 as is employed in [44] and [8].

$$rmse = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n}} \tag{7}$$

Structural similarity index, $ssim$, considers luminance, contrast, and structural information between images x and y. It is defined as shown in Eq. 8 where μ, σ, and σ_{xy} are the mean, standard deviation, and cross-covariance respectively. C_1 and C_2 are regularisation constants for when the local mean or standard deviation of an image region is close to zero.

$$ssim(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{8}$$

Universal image quality index, Q, factors in correlation, luminance, and contrast between images $\bar{x}$ and $\bar{y}$. Its formula is shown in Eq. 9 where σ_x and σ_y are considered contrast estimates.

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x})^2 + (\bar{y})^2]} \tag{9}$$

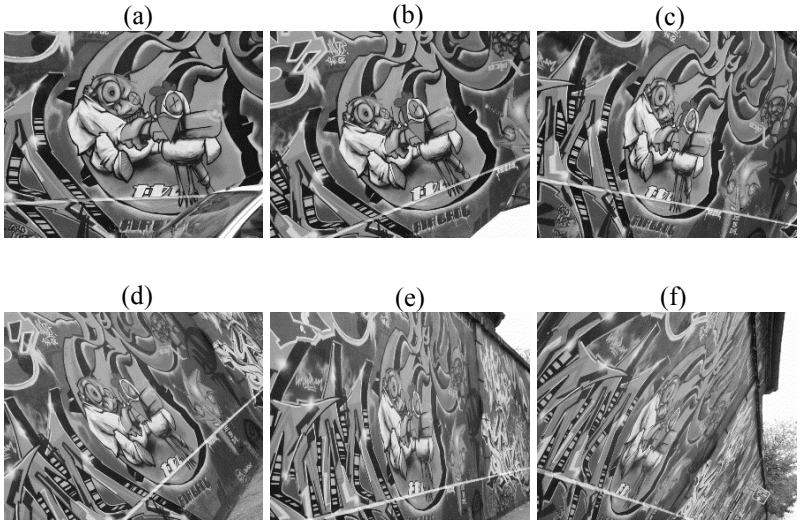

Fig. 3. Images from the Graffiti dataset. (a) is the reference image. (b) - (f) are target images to be registered to the reference image.

4.0 RESULTS

The following sections describe results obtained from the Boat and Graffiti datasets. For each dataset, the root-mean-squared error of control points, $rmse$, structural similarity index, $ssim$, and universal image quality index, Q between img1 and target images are computed. The subscript $-ecc$ indicates that the original algorithm is used. Subscript $+ecc$ means that the algorithm refined with ECC is used. Tru is the error generated by the ground truth homography. The average error across all images is computed to represent the overall performance of an algorithm.

4.1 Boat

Table 2 indicates that the $rmse$ increases by 0:0875% overall with ECC for $mser$. The error is reduced by at least half in img4 and img5. However, $rmse$ in img2 increases by half. The error in img6 increases twofold. $rmse$ in img3 is increased by 0:0412%. In $mser$, ECC increases the structural similarity by 1:7812% overall. This performance improvement occurs in all target images. $ssim_{mser+ecc}$ surpasses $ssim_{mserTru}$ by 0.0171 points in img2. The overall ssim produced by the ground truth homography is 15.3533% higher than the original algorithm.

With ECC refinement, Q increases by 1.6373% overall while it decreases by 0.0297% in img3 for $mser$. $Q_{mserTru}$ is 17.2872% higher than $Q_{mser-ecc}$ across all images, which indicates high potential for improvement. $Q_{mser+ecc}$ outperforms $Q_{mserTru}$ by 0.0161 points in img2.

$surf$'s $rmse$ is increased marginally by 0.0205% overall after ECC (Table 2). However, more target images experience positive effects, namely img2, img4, and img5. In contrast, only img3 and img6 are adversely affected. The overall $ssim$ of $surf$ is increased by 1.3333% with ECC. This is largely due to the positive effect on img2 which experiences an increase of almost 7%. img4 also reaps benefits from ECC. The images which experience a decrease in $ssim$ are img3, img5, and img6. With ECC, the Q of $surf$ is increased by 1.2415% overall. img2 and img4 experience increases in Q while the opposite occurs in img3, img5, and img6. The overall positive effect on Q is mostly due to the large increase in Q for img2 by 6:2037%. The increases in $surf$'s $ssim$ and Q after ECC makes it surpass the ground truth performance in img2 by more than two points. The ground truth performance presents a potential improvement of about 15% in $surf$'s $ssim$ and Q compared to the original algorithm.

According to Table 2, the $rmse$ of $harris$ is increased by 0:0468% overall after ECC. $rmse$ decreases occur in img4 and img5 while the opposite occurs in img2, img3, and img6. With ECC, $harris$'s $ssim$ increases by 0.1933% overall. Positive effects occur in img2, img5, and img6. Adverse effects occur in img3 and img4. $harris$'s Q improves overall by 0.1664% after ECC. img2, img5, and img6 experience benefits in Q while the opposite happens to img3 and img4. ECC positive effect on $ssim$ and Q allows it to outperform the ground truth by more than two points in $harris$. The ground truth performance is about 14% higher than that of the original algorithm.

Table 2. The $rmse$, $ssim$, and Q of original ($-ecc$), refined ($+ecc$), and ground truth homography (Tru) using the $mser$, $surf$, and $harris$ feature-based image registration methods in the Boat dataset.

Feature \ Target image		img2	img3	img4	img5	img6	Average
$mser$	$rmse_{mser-ecc}$	0.0004	17.8894	0.0007	0.0002	0.0039	3.5789
	$rmse_{mser+ecc}$	0.0006	17.8968	0.0003	0.0001	0.0125	3.5821
	$rmse_{mserTru}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$ssim_{mser-ecc}$	0.6837	0.4766	0.7175	0.8277	0.8788	0.7169
	$ssim_{mser+ecc}$	0.7415	0.4768	0.7189	0.8309	0.8800	0.7296
	$ssim_{mserTru}$	0.7244	0.7841	0.8375	0.9012	0.8875	0.8269
	$Q_{mser-ecc}$	0.7236	0.4680	0.7124	0.8273	0.8771	0.7217
	$Q_{mser+ecc}$	0.7763	0.4679	0.7144	0.8301	0.8787	0.7335
	$Q_{mserTru}$	0.7602	0.8145	0.8535	0.9113	0.8927	0.8464
$surf$	$rmse_{surf-ecc}$	0.0028	17.8922	0.0006	0.0002	0.0014	3.5794
	$rmse_{surf+ecc}$	0.0006	17.8968	0.0003	0.0002	0.0029	3.5802
	$rmse_{surfTru}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$ssim_{surf-ecc}$	0.6931	0.4775	0.7174	0.8311	0.8807	0.7200
	$ssim_{surf+ecc}$	0.7415	0.4768	0.7190	0.8306	0.8798	0.7296
	$ssim_{surfTru}$	0.7244	0.7841	0.8375	0.9012	0.8875	0.8269
	$Q_{surf-ecc}$	0.7310	0.4685	0.7125	0.8309	0.8789	0.7244
	$Q_{surf+ecc}$	0.7763	0.4679	0.7145	0.8300	0.8782	0.7334
	$Q_{surfTru}$	0.7602	0.8145	0.8535	0.9113	0.8927	0.8464
$harris$	$rmse_{harris-ecc}$	0.0001	17.8896	0.0004	0.0002	0.0021	3.5785
	$rmse_{harris+ecc}$	0.0006	17.8968	0.0003	0.0002	0.0028	3.5801
	$rmse_{harrisTru}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$ssim_{harris-ecc}$	0.7323	0.4782	0.7196	0.8300	0.8794	0.7279
	$ssim_{harris+ecc}$	0.7415	0.4768	0.7182	0.8303	0.8798	0.7293
	$ssim_{harrisTru}$	0.7244	0.7841	0.8375	0.9012	0.8875	0.8269
	$Q_{harris-ecc}$	0.7678	0.4693	0.7156	0.8290	0.8779	0.7319
	$Q_{harris+ecc}$	0.7763	0.4679	0.7138	0.8296	0.8781	0.7331
	$Q_{harrisTru}$	0.7602	0.8145	0.8535	0.9113	0.8927	0.8464

4.2 Graffiti

In $mser$, $rmse$ is reduced by 40.8403% overall based on Table 3. ECC decreases $rmse$ for all images except img4 where it increases by 6.4364%. ECC increases $ssim$ by 1.7270% for $mser$. In img2, $ssim_{mser+ecc}$ exceeds $ssim_{mserTru}$ by 0.0259 points. Across all images, $ssim_{mserTru}$ presents a possible improvement of 25.1710% from the original algorithm. The universal image quality index, Q, of $mser$ across all images increases by 1.6801% with ECC. $Q_{mser+ecc}$ outperforms $Q_{mserTru}$ by 0.0229 points in img2. $Q_{mserTru}$ is a 28.1748% improvement on $Q_{mser-ecc}$ overall.

With ECC, $surf$'s $rmse$ decreases by 5.1924% overall. Only img5 experiences adverse effects while the remaining images reap benefits from ECC. The $ssim$ of $surf$ improves by 2.5125%. All images experience improvements in $ssim$. Similarly, improvements in Q occur in all images with an overall positive effect of 2.3666% for $surf$. The $ssim$ and Q of $surf$ after ECC surpasses that of the ground truth by around three points. The gap in $ssim$ and Q performance between the ground truth and original algorithm is about 28%.

$harris$'s $rmse$ is increased by 0.0170 pixel overall after ECC. Positive effects occur in img3 and img5 while the opposite occurs in img2, img4, and img6. With ECC, the $ssim$ of $harris$ increases by 1.5368%. All images experience this positive effect in $ssim$. In the same way, $harris$'s Q for all images increases with an overall increase of 1.5212%. ECC's benefits allow $harris$'s $ssim$ and Q to leapfrog the ground truth performance by about up to 3.4456 points. The ground truth performance in $ssim$ and Q presents a potential improvement of around 27% over the original algorithm.

Table 3. The $rmse$, $ssim$, and Q of original ($-ecc$), refined ($+ecc$), and ground truth homography (Tru) using the $mser$, $surf$, and $harris$ feature-based image registration methods in the Graffiti dataset.

Feature	Target image	img2	img3	img4	img5	img6	Average
$mser$	$rmse_{mser-ecc}$	0.5441	0.1000	0.0613	0.3388	0.1416	0.2372
	$rmse_{mser+ecc}$	0.1176	0.0877	0.0652	0.3388	0.0923	0.1403
	$rmse_{mserTru}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$ssim_{mser-ecc}$	0.7465	0.4795	0.5348	0.6649	0.7214	0.6294
	$ssim_{mser+ecc}$	0.7857	0.4841	0.5394	0.6683	0.7239	0.6403
	$ssim_{mserTru}$	0.7598	0.7575	0.7875	0.8293	0.8052	0.7878
	$Q_{mser-ecc}$	0.7778	0.4733	0.5297	0.6615	0.7190	0.6323
	$Q_{mser+ecc}$	0.8157	0.4783	0.5344	0.6644	0.7218	0.6429
	$Q_{mserTru}$	0.7927	0.7813	0.8106	0.8450	0.8224	0.8104
$surf$	$rmse_{surf-ecc}$	0.0668	0.1275	0.0679	0.3388	0.1433	0.1489
	$rmse_{surf+ecc}$	0.1193	0.0877	0.0652	0.3388	0.0946	0.1411
	$rmse_{surfTru}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$ssim_{surf-ecc}$	0.7137	0.4808	0.5374	0.6673	0.7237	0.6246
	$ssim_{surf+ecc}$	0.7857	0.4841	0.5394	0.6683	0.7239	0.6403
	$ssim_{surfTru}$	0.7598	0.7575	0.7875	0.8293	0.8052	0.7878
	$Q_{surf-ecc}$	0.7479	0.4748	0.5324	0.6635	0.7215	0.6280
	$Q_{surf+ecc}$	0.8157	0.4783	0.5344	0.6644	0.7218	0.6429
	$Q_{surfTru}$	0.7927	0.7813	0.8106	0.8450	0.8224	0.8104
$harris$	$rmse_{harris-ecc}$	0.0624	0.1210	0.0582	0.3389	0.0431	0.1247
	$rmse_{harris+ecc}$	0.1193	0.0877	0.0652	0.3388	0.0977	0.1417
	$rmse_{harrisTru}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$ssim_{harris-ecc}$	0.7524	0.4795	0.5362	0.6638	0.7211	0.6306
	$ssim_{harris+ecc}$	0.7857	0.4841	0.5394	0.6683	0.7239	0.6403
	$ssim_{harrisTru}$	0.7598	0.7575	0.7875	0.8293	0.8052	0.7878
	$Q_{harris-ecc}$	0.7826	0.4734	0.5311	0.6604	0.7189	0.6333
	$Q_{harris+ecc}$	0.8157	0.4783	0.5344	0.6643	0.7218	0.6429
	$Q_{harrisTru}$	0.7927	0.7813	0.8106	0.8450	0.8224	0.8104

5.0 DISCUSSION

Results show that ECC largely improves registration of images in stitching based on $rmse$, $ssim$ and Q. Only 5 out of 30 images (5 target images × 3 feature-based methods × 2 datasets) showed minor decreases in *ssim* of between 0.0005 and 0.0014 across both datasets whereas the others showed improvements of between 0.0002 and 0.0720.

In terms of Q, 6 out of 30 images showed decreases of between 0.0001 and 0.0018 while the rest had improvements of between 0.0002 and 0.0677. In the same way, 43% of $rmse$ results across both datasets show minor negative results (i.e. an increase in $rmse$ of between 0.0002 and 0.0569 pixel) while the rest show possible improvements of up to 0.4265 pixel.

The sparse decreases in $rmse$, $ssim$ and Q performance occur mostly in the Boat dataset as the large scale change between target images prove to be challenging. Only 21% of negative results occur in the Graffiti dataset, all of which are under the $rmse$ metric.

The large scale variation in the Boat dataset causes adverse results for both the $ssim$ and Q metrics in this dataset because both of them assume that there are strong interdependencies between spatially close pixels. Both evaluation metrics perform adversely on the same images (except img3 by $mser$) in this dataset because $ssim$ and Q are very similar in the way that $ssim$ is an improved version of Q. Besides that, mostly negative results in img3 of the Boat dataset on all evaluation metrics are complications arising from poor initialisations by the feature-based methods.

The largest improvement occurs at the $rmse$ in the Graffiti dataset with a decrease of around 40% overall. The only exception occurs at the $rmse$ in the Boat dataset which may be because the ECC implementation is not optimised. ECC is fixed to 15 iterations as opposed to a convergent approach which is more optimal.

ECC's refinement allows the algorithms to outperform the ground truth $ssim$ and Q in img2 of both datasets. $rmse$ does not follow this trend because it uses the ground truth homography as the benchmark for its computation. The large performance gap (15% - 29%) between the ground truth and original algorithm suggests that there is room for improvement in the feature-based registration step of stitching.

Fig. 4. Panorama produced using the proposed stitching algorithm on the Boat and Graffiti dataset.

The evaluation metrics use grayscale images for computation. A metric which uses the RGB values of an image in its computation can better simulate the perception of a human visual system by weighting each colour channel accordingly. Other than that, the visual appeal of the stitching product can be enhanced with post-processing routines that perform exposure compensation, seam line adjustment, and blending.

The outcome of experiments agree with previous literature on the suitability of feature-based registration for images which undergo rotation, scale, and viewpoint changes. Close values between the original and refined algorithm show that the area-based registration performance relies on a good initialisation by feature-based registration. ECC's largely positive effect on registration supports the previous assertion that area-based registration has higher accuracy than feature-based-registration.

6.0 CONCLUSION

An image stitching algorithm with registration of image pairs refined by ECC is carried out. Results show that ECC is mostly effective in increasing registration accuracy in stitching. The only exception is a slight increase in average $rmse$ in the Boat dataset. The large gap between the ground truth and the original algorithm suggests that much improvement can be done in this area. Even so, the refined algorithm manages to outperform the ground truth at $ssim$ and Q in img2. The algorithm encourages an alternative approach to stitching as opposed to the more common way of using rotation and camera matrices. This approach offers a potentially more efficient solution if correction of lens distortion is not required.

ECC's implementation can be improved by using a convergent approach to increase its effectiveness. Post-processing routines can be applied to the panorama to make it more visually appealing. Consequently, metrics can be applied on the blended panorama to assess the image quality closer to its final product. Other than that, they can be modified to weight each colour channel of an RGB image to better replicate the human visual system. The robustness of this approach can be evaluated more thoroughly by using different datasets. Its applications include reduction of artefacts in stitching and increase in accuracy of motion estimation between images.

ACKNOWLEDGEMENT

This work is funded under the University of Malaya's Research Grant (UMRG), grant number RP030A-14AET and the Fundamental Research Grant (FRGS), grant number FP061-2014A.

REFERENCES

[1] R. Szeliski, "Image alignment and stitching: A tutorial," *Found. Trends. Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1–104, Jan. 2006. [Online]. Available: http://dx.doi.org/10.1561/0600000009

[2] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981. [Online]. Available: http://doi.acm.org/10.1145/358669.358692

[3] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, pp. 138–156, 2000.

[4] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vision*, vol. 74, no. 1, pp. 59–73, Aug. 2007. [Online]. Available: http://dx.doi.org/10.1007/s11263-006-0002-3

[5] "Oxford robotics database," http://www.robots.ox.ac.uk/ vgg/data/data-aff.html, 2012.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE TRANSACTIONS ON IMAGE PROCESSING, vol. 13, no. 4, pp. 600–612, 2004.

[7] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, Mar. 2002. [Online]. Available: http://dx.doi.org/10.1109/97.995823

[8] A. Wong and J. Orchard, "Robust multimodal registration using local phase-coherence representations," *Journal of Signal Processing Systems for Signal Image and Video Technology*, pp. 89–100, 2008.

[9] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. BMVC*, 2002, pp. 36.1–36.10, doi:10.5244/C.16.36.

[10] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, Jun. 2008. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2007.09.014

[11] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[12] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1858–1865, 2008.

[13] K. Gillis, C. McComb, R. K. Patel, K. Stevens, M. P. Schneider, A. Radjenovic, S. T.W. Morris, G. Roditi, C. Delles, and P. Mark, "Non-contrast renal magnetic resonance imaging to assess perfusion and corticomedullary differentiation in health and chronic kidney disease," vol. 133, 07 2016.

[14] K. A. Gillis, C. McComb, J. E. Foster, A. H. Taylor, R. K. Patel, S. T. Morris, A. G. Jardine, M. P. Schneider, G. H. Roditi, C. Delles, and P. B. Mark, "Inter-study reproducibility of arterial spin labelling magnetic resonance imaging for measurement of renal perfusion in healthy volunteers at 3 tesla," *BMC Nephrology*, vol. 15, no. 1, p. 23, Jan 2014. [Online]. Available: https://doi.org/10.1186/1471-2369-15-23

[15] A. Rozsa, E. M. Rudd, and T. E. Boult, "Adversarial diversity and hard positive generation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 25–32.

[16] C. Cai, P. Wang, and Y.-h. Liang, "Fast image stitching based on improved surf," in *International Conference on Computer Supported Cooperative Work in Design*, 2016, pp. 411–416.

[17] T. Holtorf, F. J. Knoll, and S. Hussmann, "Multimodal image stitching algorithm for weed control applications in organic farming," in *Proceedings of the 2016 SAI Computing Conference*, 2016, pp. 336–342.

[18] E. Adel, M. Elmogy, and H. Elbakry, "Image stitching system based on orb feature-based technique and compensation blending," *International Journal of Advanced Computer Science and Applications*, pp. 55–62, 2015.

[19] H.-k. Jeon, J.-m. Jeong, and K.-y. Lee, "An implementation of the real-time panoramic image stitching using orb and prosac," in *International SoC Design Conference*, 2015, pp. 91–92.

[20] M. U. Akram, D. Usmani, and T. Ahmad, "Seamless fundus image stitching using wld to improve field of view," in *International Conference on Digital Information and Communication Technology and its Applications*, 2015, pp. 106–110.

[21] K.-C. Hu, F.-Y. Lin, C.-C. Chien, T.-S. Tsai, C.-H. Hsia, and J.-S. Chiang, "Panoramic image stitching system for automotive applications," in *1st IEEE International Conference on Consumer Electronics - Taiwan*, 2014.

[22] M. A. Ullah and S. M. M. Rahman, "Low-complexity nonrigid image registration using feature-based diffeomorphic log-demons," in *Advances in Intelligent Systems and Computing*, 2016, pp. 357–366.

[23] W.-Y. Hsu, "A hybrid approach for brain image registration with local constraints," *Integrated Computer-Aided Engineering*, pp. 73–85, 2017.

[24] A. Nakazawa, "Noise stable image registration using random resample consensus," in *International Conference on Pattern Recognition*, 2016, pp. 853–858.

[25] R. Bouchiha and K. Besbes, "Comparison of local descriptors for automatic remote sensing image registration," *Signal Image and Video Processing*, pp. 463–469, 2015.

[26] M. Kashyap and M. Bhattacharya, "On global transform preservation by region based interest points for image registration," in *IEEE International Conference on Research in Computational Intelligence and Communication Networks*, 2015, pp. 61–65.

[27] L. Liu, H. Y. Tuo, T. Xu, and Z. L. Jing, "Multi-spectral image registration and evaluation based on edge-enhanced mser," *Imaging Science Journal*, pp. 228–235, 2014.

[28] L. Yufeng and H. Wei, "Research on sar image change detection algorithm based on hybrid genetic fcm and image registration," *Multimedia Tools and Applications*, pp. 15 137–15 153, 2017.

[29] J.-D. Jiang and G.-S. Lin, "An image registration method for engineering images," in *International Computer Symposium*, 2016, pp. 425–428.

[30] W. Chunwei, Z. Liqiang, W. Yaodong, Y. Zujun, and Y. Lingzhi, "Tunnel crack image registration algorithm using hausdorff-distance," in *6th International Symposium on Test Automation and Instrumentation*, 2016, pp. 44–48.

[31] S. Zhu, L. Liu, and S. Chen, "Image feature detection algorithm based on the spread of hessian source," *Multimedia Syst.*, vol. 23, no. 1, pp. 105–117, 2017. [Online]. Available: https://doi.org/10.1007/s00530-015-0453-x

[32] H. Bae, W. Kang, S. Lee, and Y. Kim, "Development of a piecewise linear omnidirectional 3d image registration method," *Review of Scientific Instruments*, 2016.

[33] F. Yang, Y. He, Z. S. Deng, and A. Yan, "Improvement of automated image stitching system for dr x-ray images," *Computers in Biology and Medicine*, pp. 108–114, 2016.

[34] X. Zhihong and J. Songming, "Image mosaic algorithm based on the combination of harris and edge extraction," *6th International Symposium on Test Automation and Instrumentation (ISTAI 2016)*, pp. 1–7, 2016.

[35] C. Jinwei, G. Bin, and G. Gangxiang, "Image registration and mosaicking based on the criterion of four collinear points," *2nd International Conference on Mechanical, Electronic and Information Technology Engineering (ICMITE)*, pp. 78–84, 2016.

[36] W. Han and Y. Wang, "A fast algorithm of image registration," *International Conference on Mechatronics, Electronic, Industrial and Control Engineering (MEIC)*, pp. 820–823, 2015.

[37] H. Li, J. Luo, C. Huang, Y. Yang, and S. Xie, "An adaptive image-stitching algorithm for an underwater monitoring system," *International Journal of Advanced Robotic Systems*, 2014.

[38] Y.-q. Cao, X.-c. Zhou, and X.-s. Huang, "An improved image mosaicking algorithm for visual sensor networks," *International Conference on Energy, Environment and Materials Engineering (EEME)*, pp. 892–900, 2014.

[39] X. Li and Z. Yuan, "Automatic panoramic stitching of omni-directional remote sensing images," *6th International Congress on Image and Signal Processing (CISP)*, pp. 352–357, 2013.

[40] X. Fang, J. Zhu, and B. Luo, "Image mosaic with relaxed motion," *Signal, Image and Video Processing*, vol. 6, no. 4, pp. 647–667, 2012. [Online]. Available: https://doi.org/10.1007/s11760-010-0194-4

[41] S. Lee, Y. Park, and D. Lee, "Seamless image stitching using structure deformation with hog matching," in *International Conference on Information and Communication Technology Convergence*, 2015, pp. 933–935.

[42] E. Adel, M. Elmogy, and H. Elbakry, "Real time image mosaicing system based on feature extraction techniques," in *9th International Conference on Computer Engineering & Systems*, 2014, pp. 339–345.

[43] B. Zhang, N. Liu, Y. Jiao, Y. Li, and J. Zhu, "Research on image matching technology for the spherical stereo vision," in *IEEE International Conference on Mechatronics & Automation*, 2015, pp. 2402–2407.

[44] Q. Zhang, Y. Wang, and L. Wang, "Registration of images with affine geometric distortion based on maximally stable extremal regions and phase congruency," *Image Vision Comput.*, vol. 36, no. C, pp. 23–39, Apr. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.imavis.2015.01.008