Check for
updates

Lorentz-modulated multiscale nonlinear diffusion for stitching in near-uniform scenes

Tze Kian Jong[1,2] · David B. L. Bong[1]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Image stitching finds diverse applications in multimedia contexts, and creating panoramic images with this technique can be particularly challenging in near-uniform scenes. Traditional feature detectors often struggle to identify distinctive features concealed within these scenes. The problem arises from the presence of featureless or homogeneous content lacking the necessary distinctiveness required to provide abundant and widely dispersed corresponding interest points. Such problems can result in unsatisfactory visual outcomes during the stitching process, manifesting as conspicuous artifacts like seams, ghosting, and geometric distortion due to insufficient matchable inliers between overlapping images. This paper presents a novel approach to feature detection, employing a nonlinear diffusion method that involves modifying the conductivity function of the partial differential equation. Inspired by the time dilation phenomenon in Einstein's theory of special relativity, we incorporate the Lorentz factor into the conductivity function, enabling the construction of novel multiscale nonlinear scale spaces that can effectively detect features in homogeneous regions and accurately stitch multiple images. Our experimental findings reveal that the proposed method consistently surpasses other state-of-the-art techniques in detecting extensive features and enhancing image stitching quality in near-uniform scenes.

Keywords Feature detection · Image stitching · Lorentz factor · Near-uniform scene · Nonlinear diffusion · Partial differential equation

1 Introduction

Image stitching is commonly used across diverse multimedia applications to enhance visual experiences and create immersive content. Examples include 3D modeling, virtual and augmented reality, educational multimedia, large-scale event coverage, aerial mapping,

✉ David B. L. Bong
davidblbong@yahoo.com; bbldavid@unimas.my

Tze Kian Jong
johntkian@gmail.com; tkjong@planet.gov.my

[1] Faculty of Engineering, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

[2] National Planetarium, Ministry of Science, Technology and Innovation, Lot 53, Jalan Perdana, 50480 Kuala Lumpur, Malaysia

Springer

e-commerce product presentations, architectural visualization, virtual tours, and more. The core algorithm of image stitching involves aligning and merging information from two or more images with overlapping field-of-view (FOV) to create a single composite image with a wider FOV and higher resolution. The image stitching process involves three key steps: firstly, extracting and accurately matching interest points between input images, then the overlapping images are subjected to deformation or warping, aligning them precisely using estimated geometric or homography transformation models (e.g., affine, similarity and projective transformation) [1]. Following the alignment process, images are blended and seam-cut to create a seamless, wider FOV image [2–4]. Hence, the quality of alignment and stitching results greatly depends on finding enough and correct matchable interest points (also known as keypoints) between the input images [5].

Image stitching has undergone significant advancements, benefiting various aspects of daily life by overcoming FOV limitations in images or videos [6, 7]. However, stitching images with near-uniform scenes presents specific challenges compared to scenes featuring distinctive textures. In near-uniform or low-texture scenes like the skies, oceans, deserts, planets' surface etc., traditional feature-based image stitching algorithms struggle to find enough distinctive features or keypoints to match and align the images accurately. This is due to the lack of distinctive content that can provide reliable corresponding interest points. Insufficient matchable interest points often lead to difficulties in estimating precise transformation models for alignment, resulting in visual artifacts (e.g., seams, ghosting and blurring) and geometric distortion in the stitched output.

Moreover, substantial image distortion problems can arise due to "clustering" or "concentration" of corresponding interest points in specific overlapping regions. This issue typically arises when many corresponding interest points are detected only in small or narrow feature-rich areas within the overlapping region, leaving fewer or no points in most near-uniform regions. Consequently, the stitching result is further degraded because the "clustering" corresponding interest points provide a poor fit for accurately estimating the transformation model. Despite ongoing efforts by researchers to develop improved image alignment and advanced image composition methods, addressing severe misalignment and distortion in stitching near-uniform scene images remains challenging [8].

Image stitching methods rely heavily on robust feature detectors to obtain matching keypoints from overlapping images. Feature detection involves identifying interest structures and primitives (e.g., points, lines, curves, and regions) that highlight the salient content of images (e.g., corners, edges, blobs, and ridges). While these methods perform well on feature-rich images, they encounter challenges when dealing with near-uniform scene images. A comprehensive survey on recent developments in visual feature detection, categorizing methods into edge, corner, and blob detections, is elaborated by Li et al. [9]. The earliest feature detector used for image stitching algorithms dates back to the work of Harris corner features in 1988 [10]. Although it is a corner and edge features detector, it lacks scale invariance, affecting its accuracy in providing matches for images of different sizes. Later, Scale-Invariant Feature Transform (SIFT) [11] emerged and gained widespread adoption in computer vision and image stitching. SIFT exhibits remarkable distinctiveness and invariance to image scale, rotation and translation, as well as robustness against illumination and viewpoint changes. Thereafter, researchers have expanded the feature detection method to enhance its robustness and efficiency, while making it more suitable for real-time systems by optimizing its computational complexity. Concerning filtering techniques, SIFT [11, 12] and Speeded Up Robust Features (SURF) [13] excel in detecting blob-like features by utilizing a pyramid of Gaussian scale spaces. In contrast, CenSurE features [14] are estimated by two

variants of bi-level Gaussian approximation filters, allowing for rapid computation with integral images in real-time.

Gaussian smoothing does not preserve object boundaries. Both image details and noise are blurred to some degree in Gaussian scale spaces, resulting in a diminution in localization accuracy and distinctiveness of the interest points [15]. This problem can be addressed using nonlinear diffusion filtering, which could generate smooth scale spaces and simultaneously preserve the natural boundaries of regions and objects [16]. Thus, instead of Gaussian smoothing, several methods, such as BFSIFT [17], KAZE [15], A-KAZE [18] and SRP-AKAZE [19], have adopted nonlinear diffusion filtering to improve the searching performance of local extrema at different scale levels. Other techniques like MSER [20] use extremal regions, whereas FAST [21] and AGAST [22] use accelerated segment tests. The most recent methods, such as RIFT [23] and MSFD [24], utilize phase congruency to tackle nonlinear radiation distortion in multi-model images. Multiple features are also employed to address multi-modal image registration [25]. Some researchers place greater emphasis on refining feature descriptors like WLD [26], BRIEF [27], M-LDB [18], BRISK [28], ORB [29], FREAK [30] and DFOB [31]. Their goal is to expedite computation speed while minimizing storage demands.

Although deep learning (DL) has gained prominence in tackling intricate computer vision (CV) challenges lately, it may not always be the one-size-fits-all solution for every application. Some scenarios may benefit more from using traditional algorithms [32]. For example, when it comes to general image stitching, classical techniques like SIFT [11] and other classical feature detection methods [9] excel in their performance. On the other hand, DL relies on specific training datasets, leading to performance degradation when dealing with images outside its training set. In autonomous robotics, the limitations of robotic hardware and the lack of real-time annotated data often make classic computer vision methods a practical choice for robot applications [33]. Hybrid approaches that merge classical algorithms with DL have shown potential in addressing CV challenges that are not readily solvable by DL alone in the modern context [32, 34]. Conventional CV techniques can enhance DL performance in various applications, including panoramic stitching [35], simultaneous localization and mapping (SLAM), 3D vision, etc. [36]. As such, classical CV techniques remain significant in the present landscape. The main focus of this paper is on the conventional image stitching method, utilizing a novel feature-based detection algorithm. This paper does not delve into DL, as it is beyond the purview of this paper.

In this paper, we propose a novel feature detection method to improve image stitching performance and reduce severe misalignment or projective distortion, especially in the presence of near-uniform or low-texture images. The contributions of our work can be summarized as follows:

a. Introducing a novel conductivity function in partial differential equation (PDE) based on the Lorentz factor to create an alternative nonlinear scale-space.
b. Presenting a robust feature detection approach that relies on Lorentz-modulated non-linear diffusion scale-space. This technique substantially enhances the number of reliable corresponding or matching interest points in overlapping images, offering notable advantages, particularly for images with near-uniform or low-texture characteristics.
c. Broadening the evaluation criteria to assess the performance of corresponding or matching interest points across images. We accomplish this by studying their spatial distribution and investigating the connection between their recall (RC) and spread-overlap (S_o) metrics, represented as the RC/S_o score.

The rest of this paper is organized as follows: Section II explains the related work. Section III details the proposed method and evaluation metrics. Section IV reports the experimental results and analyses. Finally, the paper concludes in Section V.

2 Related work

In this section, we begin with a review of nonlinear diffusion filtering, followed by a concise definition of the Lorentz factor within the context of time dilation.

2.1 Nonlinear diffusion filtering

Scale-space filtering is a powerful image processing technique that decomposes an image into a series of gradually smoother images across increasing scales or time units. The derived image representations can extract potential interest features and be applied in many applications such as denoising, segmentation, and multiscale analysis [37]. Over the years, several approaches related to scale-space filtering have been developed, notably the linear, nonlinear isotropic, and nonlinear anisotropic diffusion models [38]. While these methods are designed to simplify images at multiple scales, the diffusivity of nonlinear diffusion models excels at improving edges.

According to [39], the earliest theory of linear scale-space has already been axiomatically derived by Taizo Iijima. However, the ideas of linear scale-space introduced by Witkin [40] and Koenderink [41] are more popular among researchers. In brief, Witkin introduced the Gaussian scale-space representation by convolving the original image with a Gaussian kernel. With reasonable assumptions, Koenderink [41] and Lindeberg [37] showed that the Gaussian function and its derivatives are the only sensible linear scale-space kernels. The Gaussian kernel is generally defined as follows:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{1}$$

where x and y are the Cartesian coordinates of the image plane, and σ is the scale level. The Gaussian scale-space of an image, $L(x, y, \sigma)$ can be easily constructed by convolving a variable-scale Gaussian kernel with an input image, $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{2}$$

where $*$ indicates the convolution operation in x and y. Gaussian kernel with larger scale level tends to produce simpler or smoother image representation. Similarly, Duits et al. consider Poisson scale-space as a feasible alternative to Gaussian [42]. A recent technique finds that multiscale Poisson kernel produces stable features in scale space [43].

Gaussian scale-space is useful for noise reduction and emphasizes prominent structures at selecting coarser scales. The major downside of Gaussian smoothing is that it does not preserve object boundaries and the loss of localized structure details increases at coarser scales. This limitation can be addressed by the nonlinear diffusion approach proposed by Perona and Malik [16] for edge detection and image restoration. Nonlinear diffusion is described as a partial differential equation (PDE) that regulates the prior information of image features through the diffusion coefficient in the filtering processing. Nonlinear scale-space appears relatively stable in the presence of noise

while keeping details or edges well localized. For an input digital image I, nonlinear diffusion can be formulated mathematically as:

$$\frac{\partial I}{\partial t} = div(g(x, y, t) \bullet \nabla I) \tag{3}$$

where div is the divergence operator, $g(x, y, t)$ is the conductivity function that defines the diffusion weight, and ∇ is the spatial gradient operator. The variable t in the function $g(x, y, t)$ represents the 'time' scale parameter. It is used to enumerate iteration 'time' steps to lead the preceding image to simpler image representations in discrete implementation. Thus, the function $g(x, y, t)$ controls the diffusion process, adapting it to each pixel's local image differential structure.

Generally, there are three different formulations of conductivity functions. Perona and Malik [16] proposed the following conductivity functions in their work.

$$g_1 = exp\left(-\frac{|\nabla I(x, y, t)|^2}{k^2}\right) \tag{4}$$

$$g_2 = \frac{1}{1 + \frac{|\nabla I(x,y,t)|^2}{k^2}} \tag{5}$$

The parameter k is the contrast factor that controls the diffusion weight magnitudes concerning the image spatial gradient, thereby regulating boundaries sharpness. According to [15] and [16], the parameter k can either be fixed manually at a constant value or automatically computed from the image gradient histogram. On the other hand, Weickert used a different form of the conductivity function g_3 to rapidly decrease diffusivity where smoothing on both sides of the edge is stronger than smoothing across it [44, 45].

$$g_3 = \begin{cases} 1 & , |\nabla I|^2 = 0 \\ 1 - exp\left(-\frac{3.315}{(|\nabla I(x,y,t)|/k)^8}\right) & , |\nabla I|^2 > 0 \end{cases} \tag{6}$$

The nonlinear scale spaces generated by these three forms of conductivity function are somewhat dissimilar: g_1 favors high-contrast edges, g_2 favors wide regions over smaller ones, and g_3 favors intraregional smoothing over interregional blurring. According to Alcantarilla et al. [15], g_1 and g_3 are more suitable for corner detection, whereas g_2 is better suited for detecting blob-like features.

2.2 Time dilation phenomenon: lorentz factor

According to Einstein's theory of special relativity, time dilation is fundamentally described as a phenomenon in which there is a difference in elapsed time between two events, measured by two clocks that are either moving relatively to each other or due to a gravitational potential difference between their proximity locations [46]. Generally, time dilation can be expressed as:

$$\Delta t = \gamma \triangle t' \tag{7}$$

where Δt is the elapsed time for the clock observed in motion, $\Delta t'$ is the elapsed time for the clock observed at rest, and γ is a scaling factor determining how much time is relatively stretched and contracted. γ is also known as the Lorentz factor and is defined from the Lorentz transformations [46] as:

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{8}$$

where v is the velocity of the moving object and c is the speed of light. Since $\gamma > 1$, the Δt measured in the clock in motion is longer than the $\Delta t'$ measured in the clock at the resting reference frame. This phenomenon is known as time dilation. In simple terms, the faster the object moves through space, the slower the object moves through time.

3 Our method

This section introduces a newly devised conductivity function inspired by the concept of time dilation, which provides mathematical principles for the proposed method. Subsequently, we explain the feature detection algorithms and the steps for image stitching. Finally, we elaborate on the performance evaluation method.

3.1 Conductivity function formulation

As previously stated, the idea of the proposed method draws inspiration from the phenomenon of time dilation and involves modifying the Lorentz function. Analogously, the Lorentz function expressed in (8) can be extrapolated to approximate an improved conductivity function that defines the diffusion weight in the diffusion equation.

Let's consider an input image in which its image spatial gradient in each pixel is represented as $\nabla I(x, y, t)$. According to (3), filtering the image to simpler scale space representation requires a diffusion weight, g, to preserve object boundaries (i.e., high image gradient) and to smooth non-boundaries or homogenous regions (i.e., low image gradient). In other words, the greater the image gradient in image space, the slower the image gradient degrades over the time scale. This characteristic indeed performs analogously to the Lorentz function. Thus, by replacing the variable v and the constant c in (8) with the image spatial gradient $|\nabla I(x, y, t)|$ and the contrast parameter k, respectively, we obtain a newly found conductivity function expressed as:

$$g_4 = \frac{1}{\sqrt{\left| 1 - \frac{|\nabla I(x,y,t)|^2}{k^2} \right|}} = \frac{1}{\sqrt{\alpha}} \tag{9}$$

Since the magnitude of image spatial gradient $|\nabla I(x, y, t)|$ for a digital image is typically 0 to 255, we take the absolute value of α to avoid any complex numbers when computing the square root term of g_4. In our experiments, we manually select a value between 0.1 and 0.9 for the parameter k because it generally yields considerably stable diffusivity output.

Figure 1 demonstrates how image spatial gradients $|\nabla I(x, y, t)|$ are affected under different conductivity coefficients for a fixed parameter k. As shown in Fig. 1, we can see that the

conductivity coefficient of g_4 tends to have a stronger impact on smaller image gradients (i.e., homogenous regions) when compared to the standard g_2. However, the g_4 coefficient for $|\nabla I(x, y, t)| \geq 2$ is notably high, leading to potentially blurry scale space at coarser scales. To reduce this blurry effect, we revise the g_4 function by raising it to the power of 4, resulting in a new function, g_5.

$$g_5 = \frac{1}{\left(1 - \frac{|\nabla I(x,y,t)|^2}{k^2}\right)^2} \tag{10}$$

For $|\nabla I(x, y, t)| = 1$, the modified g_5 also demonstrates a stronger coefficient, as shown in Fig. 1. A strong coefficient for g_5 is expected to rapidly smooth the homogenous regions. On the other hand, the g_5 coefficient approaches zero when $|\nabla I(x, y, t)| \geq 2$, which explains that a high image gradient (i.e., object boundaries) is expected to degrade at a much slower rate over the time scale, or in other words, the object boundaries remain well-preserved at coarser scales. The improved g_4 and g_5 functions generate distinct scale space image structures, convincingly improve near-uniform scene feature detection and image stitching, as further explained in Section IV.
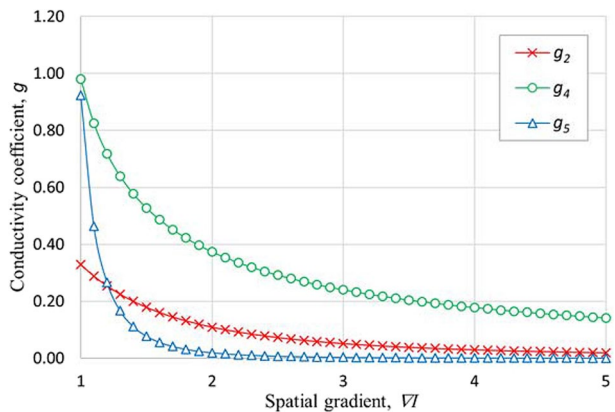
3.2 Building nonlinear scale spaces

To build nonlinear diffusion scale spaces from a digital image, we use Weickert's modified semi-implicit scheme, namely the Additive Operator Splitting (AOS) [44, 45] scheme, to numerically approximate the nonlinear partial differential equation (PDE) in discretized form. In the AOS scheme, discretization of (3) can be expressed in a vector–matrix notation as:

$$L^{t+1} = \frac{1}{m} \sum_{l=1}^{m} \left(Id - m\tau A_l\right)^{-1} L^t \tag{11}$$

where L^t represents the nonlinear scale spaces at evolution time t, A_l is the block of tridiagonal square matrices, τ is the step size, m is the number of dimensions ($m = 2$ in our method), and Id is the identity matrix. Under consecutive pixel numbering along the direction l, the operators $\left(Id - m\tau A_l\right)$ interpret one-dimensional diffusive interaction along the axes are diagonally dominant tridiagonal matrices of a linear system of equations. Such a linear system of equations in the AOS scheme can be efficiently solved by the Thomas

Fig. 1 The variation of conductivity coefficients g_2, g_4 and g_5 with fixed parameter k against image spatial gradients ∇I

algorithm or the Tri-Diagonal Matrix Algorithm (TDMA) [45, 47]. For every step size τ in AOS scheme, all coordinate axes are treated the same way to create discrete nonlinear scale spaces.

Prior to building the scale space image structures, the first step is to compute a set of evolution times t_i from which we find the step size $\tau \left(= t_{i+1} - t_i \right)$ to apply in (11). The scale spaces are arranged in a sequential discrete octaves o and sub-levels s and analyzed via up-scaling SURF's box filter to approximate second-order Gaussian derivatives [13, 48]. Each octave-sublevel pair value is then mapped to a corresponding filter size as:

$$f_i = 3((2^o \times s) + 1), i = \{0 \ldots N\} \tag{12}$$

with an initial filter size f_0 ($=9 \times 9$) corresponding to the Gaussian derivatives of initial sigma σ_0 ($=1.6$ in our method). When filter size increases, the associated Gaussian scale also increases and can be easily calculated because the filter layout ratio remains constant. Since nonlinear diffusion works in time units, the set of discrete scale sigma σ_i can be matched to their corresponding time units t_i by using:

$$t_i = \frac{1}{2}\sigma_i^2, i = \{0 \ldots N\} \tag{13}$$

where N is the total number of 2-dimensional arrays of scale space image structures. For our method, we chose to create an array of 12 scale space image structures, which will be divided into 5 octaves, each comprises 4 sublevels. The first octave consists of 4 sequential scale spaces, and the remaining octaves comprise the last 2 scale spaces from the previous octave and the following sequence of 2 scale spaces.

For an input image of a near-uniform scene, Fig. 2 shows the difference between nonlinear scale-space structures computed using conductivity functions from (5), (9) and (10) for several contrast factors k. Each scale space image in Fig. 2 is cropped to square shape from its original dimensions, and only the 12th layer of the scale space image is presented for every conductivity function and contrast parameter k. As shown in Fig. 2, each conductivity function performs at a different diffusivity rate, generating scale space images with distinctive degree of smoothness. When compared to the standard g_2, the proposed g_4 smooths the input images at a much faster rate and results in expediting blurry structures in the output scale space, whereas g_5 smooths the input image at a much slower rate and maintains most of the image's prominent structures. At increasing contrast factors, g_2 and g_4 generated scale spaces develop blurry effects and rapidly lose the most prominent structures. In contrast, almost all the structural information of g_5 generated scale spaces are well preserved, and the strong image edges remain unaffected even at higher evolution time units.

3.3 Feature detection and description

In search of scale-invariant interest points, we employ Hessian matrix by converting all nonlinear scale spaces into integral images, enabling fast computation to be later implemented using box filters [13, 48]. For every integral image, the scale-normalized Hessian determinant matrix is determined through the application of box filters to approximate the Laplacian of Gaussian [13]. The Hessian determinant essentially acts as a measure of blob responses, which will be further examined in the non-maxima suppression process. Only Hessian responses above a predetermined threshold will be retained to regulate detection capability. After thresholding, non-maximum suppression is performed in a $3 \times 3 \times 3$
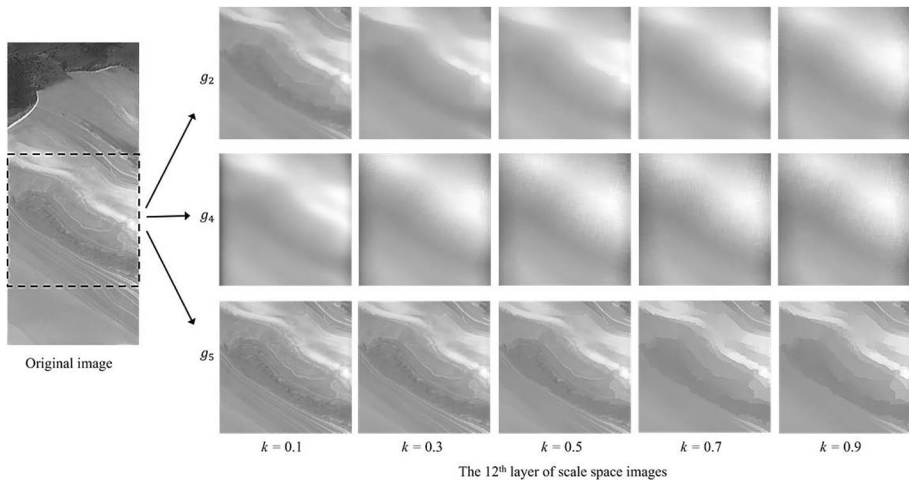
Fig. 2 Example of nonlinear scale-space images computed using conductivity functions g_2, g_4 and g_5 for several contrast factors k. Each processed image is obtained by cropping the 12th layer of the scale-space image into a square shape

neighborhood of sequential scale spaces to find candidate points [11]. A point is classified as an interest point only if it is greater than its 8 neighbors in the current scale space and 9 neighbors in the scale spaces above and below. Lastly, a 3D quadratic function will interpolate adjacent data points for sub-pixel accuracy and stable localization [11], eliminating unstable candidates with low contrast or inadequate edge localization. The interpolated extrema location is capable of providing substantial improvement in matching and stability.

Every identified interest point must define its distinctive feature description for image matching purposes. The feature description is extracted from a square region on the original input image, aligned to a dominant orientation at each interest point. The dominant orientation can be determined by calculating the Gaussian weighted Haar-wavelet responses within a circular neighborhood of radius $6\,\sigma_i$ around the interest point [13]. Wavelet responses are then summed within a rotating circular segment of $\pi/3$ around the interest point, with the longest vector signifying the dominant orientation of the interest point. The final step in defining the feature description is to build the descriptor vector for each interest point. We apply the M-SURF descriptor by computing Haar-wavelet responses in the horizontal and vertical directions relative to the dominant direction (i.e., represented as d_x and d_y) over a larger square region grid of size $24\,\sigma_i \times 24\,\sigma_i$, which is split into smaller 4×4 square subregions with $2\,\sigma_i$ overlapping zone from the original input image [15]. In each subregion, wavelet responses are weighted using a subregion-centered Gaussian and aggregated into a 4-dimensional descriptor vector, denoted as $d_v = \left\{ \sum d_x, \sum d_y, \sum |d_x|, \sum |d_y| \right\}$. This contributes to an overall length of 64 ($=4 \times 4 \times 4$) element feature vector for each interest point. Each descriptor vector is normalized to a unit vector to achieve contrast invariance.

3.4 Image stitching

To stitch a pair of overlapping scene images successfully, it is essential to have a substantial number of accurately matched interest points between the images. In this study, we

use the matching algorithm in the VLFeat open-source library [49] to obtain a collection of indexed corresponding interest points and their squared Euclidean distance. To exclude outliers, we use the M-estimator SAmple Consensus (MSAC) algorithm [50], to provide better matching of the correct matches or inliers between the images. Given the inherent randomness in the MSAC algorithm, the count of inliers may exhibit slight variation between each execution, though the differences are generally inconsequential. To get the highest possible number of inliers, we employ an iterative approach, execute the MSAC algorithm in 20 trials. Finally, we estimate the global geometric transformation model that aligns, warps, and blends the overlapping images, resulting in a decent stitched representation. Table 1 recaps the proposed method and all the related algorithms used in our image stitching procedure.

3.5 Evaluation method

Assessing the performance of feature detectors and descriptors is a fundamental aspect of computer vision. The evaluation metrics introduced by Mikolajczyk et al. [51, 52], are widely adopted in various studies involving local features. However, it is worth noting that these metrics, including repeatability and recall measures, may not directly reflect the performance of image stitching in the context of the extracted inliers. For example, an effective feature detector will deliver high repeatability. Still, it does not guarantee that the stable inliers extracted from such feature detector are sufficient for yielding decent image stitching. The quality of image stitching relies on the presence of sufficient quantity of stable inliers and their well-distributed spatial placement within the images' overlapping

Table 1. Summary of algorithms used in image stitching procedure (Algorithm 1) and the proposed method (Algorithm 2)

Algorithm 1 Feature detection, description, matching and image stitching
Input Images I_n, Gaussian sigma σ, octave o, sub-level s, initial sigma σ_o, and contrast factor k **Output** Final stitched image I_o

1. Feature detection and description:
     for $n = 1 \rightarrow 2$
       a. Denoise I_n with Gaussian kernel $\sigma \rightarrow L_n$
       b. Use **PDEaos** function to compute N nonlinear scale spaces L_i
       c. Extract Hessian determinant and Laplacian sign response map from each L_i
       d. Apply non-maximum suppression
       e. Extract 64 elements feature vector
     end for
 2. Apply matching algorithm provided in VLFeat library [40]
 3. Exclude outliers using MSAC algorithm [41]
 4. Estimate global geometric transformation model
 5. Warp and blend images $I_n \rightarrow I_o$

Algorithm 2 PDE-based nonlinear diffusion scale space filtering
Input Gaussian denoised images L_n, octave o, sub-level s, initial sigma σ_o and contrast factor k **Output** Set of integral nonlinear scale space images L_i, $i = 1 \dots N$

function PDEaos
 for $i = 1 \rightarrow N$
     a. Compute a set of filter size f_i
     b. Compute corresponding discrete scale sigma σ_i
     c. Compute evolution times t_i
     d. Compute step size $\tau_i = t_{i+1} - t_i$
     e. Compute diffusivity matrix $A_l(L_n)$ for the desired k
     f. Solve vector-matrix (11) using AOS and TDMA algorithms
     g. Convert nonlinear filtered images to integral images L_i
 end for

regions. Therefore, we employ the latest evaluation metrics, such as Spread-overlap (S_o) metric and the RC/S_o score, which we believe are well-suited for evaluating the image stitching performance [53].

Recall (RC) Measure To match a pair of planar scene images, a detected interest point x_i in image I_i will typically be repeated in image I_j as the corresponding point x_j. Recall (RC) is defined as the number of inliers divided by the number of corresponding points visible within the overlapping scene [51]. In mathematical notation, the recall (RC) measure can be expressed as:

$$RC = \frac{\left|N_m(\epsilon_s)\right|}{\left|N_c(\epsilon)\right|} > 0 \tag{14}$$

where N_m is the number of inliers or correct matches, and N_c is the number of corresponding points. Generally, a repeated point x_i will not be precisely detected at the position x_j, but rather in the neighborhood of x_j, denoted by ϵ. Hence, $N_c(\epsilon)$ can only be satisfied if the location uncertainty of x_i does not exceed ϵ in size within the neighborhood of x_j [54]. Instead of ϵ, we employ the classical approach to determine the number of correspondences based on the smallest Euclidean distance (multiplied by a threshold value of 1.5) between each interest point's feature vectors in both images I_i and I_j. According to [51], the number of inliers is defined based on a maximum overlap error ($\epsilon_S = 0.5$) measuring the accuracy of matching corresponding regions under a homography transformation. As an alternative way to exclude outliers, we use M-estimator SAmple Consensus (MSAC) algorithm [50] to determine the number of inliers based on the maximum distance error (i.e., 1.5 pixels) from an interest point in image I_i to its projective corresponding point in image I_j.

Spread-overlap (S_o) Measure Inspired by Marmol et al. [55], we develop the spread-overlap (S_o) metric to compute the spatial distribution of interest points across the overlapping region between images. Marmol et al. constructed a uniform 10×10 grid cell mask on the image inside the view of arthroscope's eyepiece. They calculated the ratio of number of grid cells containing at least one interest point. Instead of using a 10×10 grid cell mask, we partitioned the entire image area into square grid cells, each occupying 0.25% of the total image area. This approach yielded a consistent grid of 400 square covering the entire image area, regardless of the image's size. To ensure each grid cell is big enough to hold a few interest points, we use sample images with at least 100×100 pixels or higher in our study. The spread-overlap (S_o) measure is thus defined as the number of grid cells within the overlapping region that contain no less than one inlier, according to the following expression:

$$S_o = \frac{n_o}{N_o} = \frac{n_o}{400} \tag{15}$$

where n_o refers to the number of valid grid cells contain at least one correct match or inlier, and N_o is the total number of grid cells ($N_o = 400$) overlaid only the overlapping region.

RC/S_o Score The score is computed as the ratio of the recall (RC) to the spread-overlap (S_o) measures. In principle, the RC/S_o score measures how well the inliers' spatial distribution estimates the global homography transformation, which is used to stitch the overlapping images precisely with minimal misalignment or distortion. A higher value of the

RC/S_o score can lead to obvious misalignment, distortion, or even failure in image stitching, primarily due to lack of inliers and their narrowed or concentrated distribution within a small area of the overlapping region. To illustrate this, consider a set of 30 inliers (depicted as red dots in Fig. 3) with an RC measure of 0.750 (equivalent to 75%). These inliers are scattered in three distinct distribution patterns across the overlapping region (see Fig. 3(a)-(c)). The overlapping area is divided into a 7×5 uniform square grid cell for simplicity. When inliers are widely scattered in the overlapping region, the RC/S_o score is likely to be closer to a value of one (see Fig. 3(a)-(b)). This suggests that in Fig. 3(a), the inliers are more reliable for achieving decent alignment in image stitching. Conversely, when inliers are intensely concentrated within a smaller area, as shown in Fig. 3(c), a higher RC/S_o score would be reached, indicating a greater probability of misalignment, distortion, or even failure in image stitching. This happens because the narrow spread of inliers provides insufficient information for accurate estimation of the global homography transformation needed for proper image stitching.

4 Result and discussion

In our experiments, we validate the effectiveness and robustness of the proposed feature detection method by stitching various pairs of images and assessing the detected interest points using evaluation metrics discussed in the previous section. To implement the proposed method and the related algorithms for image stitching and evaluate their performance, we use the MATLAB computer vision system toolbox along with the mexOpenCV interface in our research. Additionally, we adapt and make use of certain algorithms provided in Ralli diffusion code [47], OpenSURF library [48] and VLFeat open-source library [49] to accomplish the proposed feature detection and description algorithms. The effectiveness of our feature detection method is subsequently validated against various state-of-the-art methods, including MSER, SIFT, SURF, BRISK, KAZE, A-KAZE, AGAST, ORB, and the recent upright variant of RIFT (denoted as U-RIFT). Table 2 provides a summary

(a) $S_o = 0.771$; $RC/S_o = 0.972$ (b) $S_o = 0.571$; $RC/S_o = 1.313$ (c) $S_o = 0.200$; $RC/S_o = 3.750$
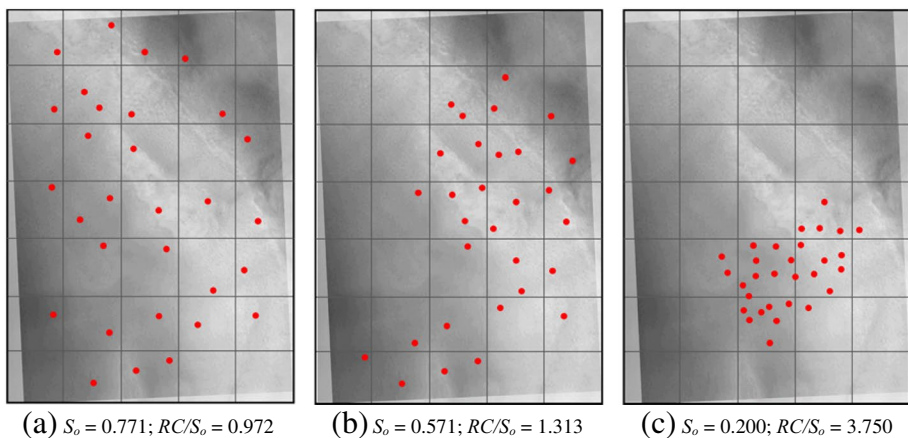
Fig. 3 Explanation of the RC/S_o scores in the context of 30 sample inliers within 7×5 square grid cells in the overlapping region. The recall (RC) measure is set to 0.750 with red dots indicate inlier locations

Table 2 Summary of feature detection and description methods used for validation

Detector	Descriptor	Feature Type				Data Type	
		Corner	Edge	Blob	Region	nonbinary	binary
MSER	SURF	-	-	-	√	√	-
SIFT	SIFT	-	-	√	-	√	-
SURF	SURF	-	-	√	-	√	-
BRISK	BRISK	-	-	√	-	-	√
KAZE	M-SURF	-	-	√	-	√	-
A-KAZE	M-LDB	-	-	√	-	-	√
AGAST	FREAK	√	-	-	-	√	-
ORB	BRIEF	√	-	-	-	-	√
U-RIFT	U-RIFT	√	√	-	-	√	-

(a) Bikes   (b) Trees   (c) Leuven   (d) Iguazu   (e) Ubc
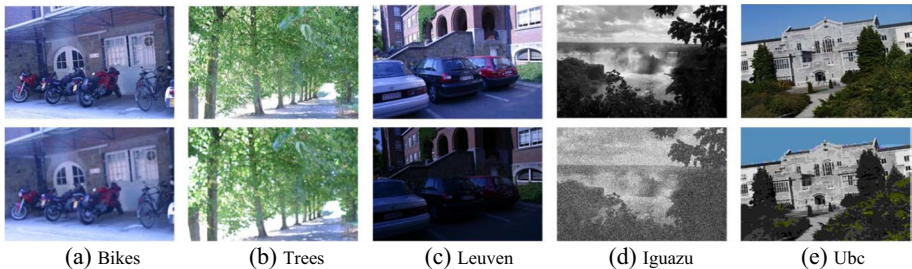
Fig. 4 Examples of benchmark image pairs with (**a**)-(**b**) image blur, (**c**) light change, (**d**) Gaussian noise, and (**e**) JPEG compression, used for feature detector performance comparison

of these methods based on their feature detector, associated descriptor, the targeted features, and data types. The default settings for each method are basically retained in our experiments. This study is carried out on a Windows 10 64-bit computer equipped with an Intel core i5-6300U CPU operating at 2.40 GHz and 8.00 GB of RAM.

Concerning the experimental datasets, we use 25 benchmark (available for download at [15, 52, 56]) and 75 real-world image pairs to evaluate and compare our method with state-of-the-art feature detectors, both quantitatively and qualitatively. Figure 4 shows 5 examples of benchmark image pairs, including 'bikes' and 'trees' for image blur, 'leuven' for illumination change, 'iguazu' for Gaussian noise, and 'ubc' for JPEG compression. Each set of benchmark images consists of 6 same-scene images that gradually change their photometric transformation, respectively. The real-world images contain scenes where certain regions within the image content exhibit homogeneity or low-textured characteristics. As discussed in Section I, near-uniform scene images tend to lead to less accuracy and sensitivity for state-of-the-art feature detection methods due to near-homogeneous or low-texture content. These sample images are sourced from publicly available datasets compiled by previous researchers (refer to [57–60] for examples), online resources from the NASA Photojournal [61], and images captured by the authors in real-world scenarios (available upon request). All color images are converted to grayscale before processing in feature detection and image stitching algorithms.

4.1 Benchmark image analysis

For ease of reference, the proposed feature detectors that utilize partial differential equations (PDE) with the new conductivity g_4 and g_5 will be referred to as ePDE-g_4 and ePDE-g_5 in the following discussion. Knowing that the diffusion weight is regulated by the contrast factor k (as expressed in (9) and (10)), Fig. 5 illustrate how the performance of our proposed feature detectors, ePDE-g_4 and ePDE-g_5, varies across various value of the contrast factors k for the 25 benchmark image pairs (see Fig. 4). For each contrast factor k, we compute and average the evaluation response of each feature detector across all benchmark images. As shown in Fig. 5(a)-(b) for ePDE-g_4, both the average number of inliers and spread-overlap measure experience a declining trend at higher values of the contrast factor k. The result is not unexpected because the nonlinear scale spaces generated by ePDE-g_4 generally contain poorer local structure information at higher contrast factor k (see Fig. 2). On the other hand, the performance of ePDE-g_5 is reasonably stable, with only a slight fall-off observed across contrast factors k. Its detected inliers are greater in quantity and spread wider within the overlapping region when compared to ePDE-g_4. This is because the dominant structure of nonlinear scale spaces generated by ePDE-g_5 is prominently well-preserved (see Fig. 2), due to its stable diffusivity regulated by (10). Figure5 (c) shows that ePDE-g_4 can hold marginally more inliers than ePDE-g_5 for their detected corresponding feature points. In Figure 5(d), the average RC/S_o result for ePDE-g_5 is comparatively more consistent than ePDE-g_4, which implies that ePDE-g_5 is likely to create less distortion to the final stitched image. Based on the results in Fig. 5 and considering computation complexity concern, we choose a value of k (=0.5) for the proposed method in the subsequent experiments.

Figure 6 shows the performance results for various feature detectors by averaging their evaluation responses across benchmark images. Given that these benchmark images are
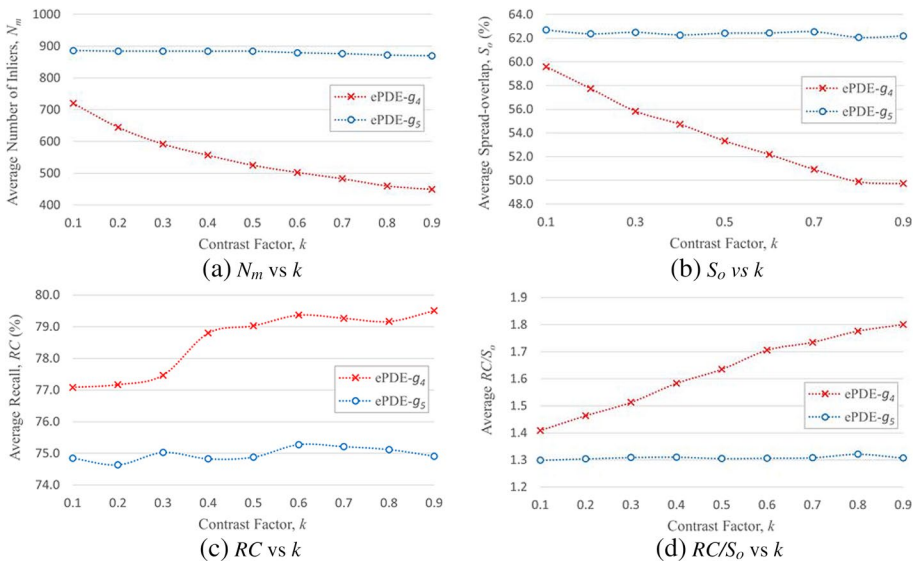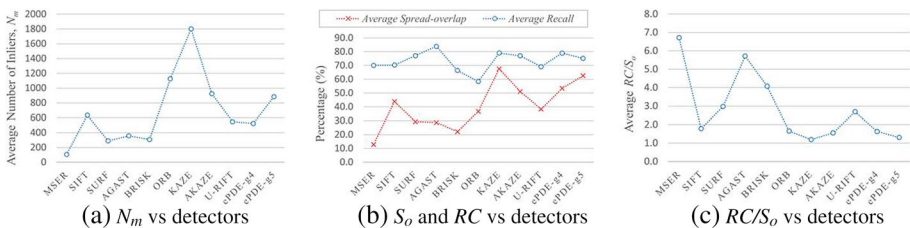

Fig. 5 Performance analysis of the proposed feature detectors ePDE-g_4 and ePDE-g_5 against contract factor k. For each evaluation metric, the results are obtained by averaging the evaluation responses for each contract factor k across 25 pairs of benchmark datasets (see Fig. 4)
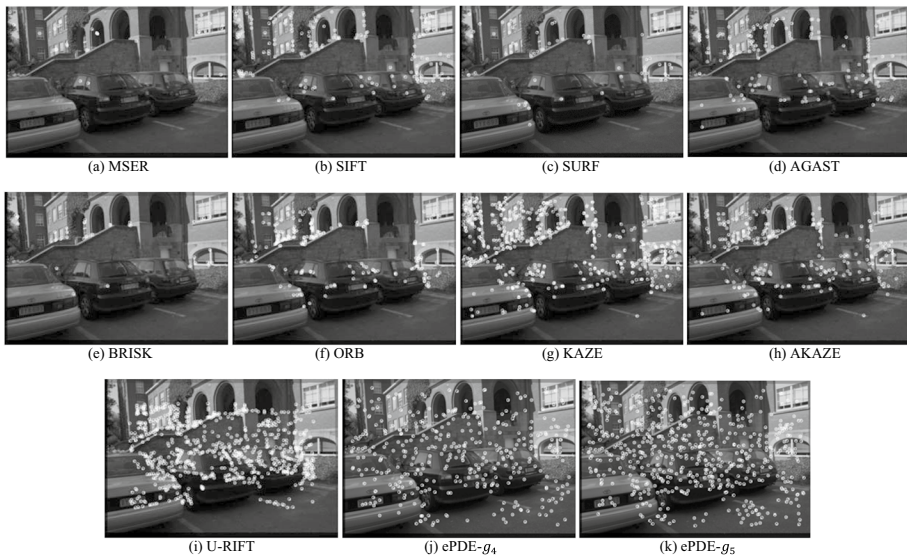
generally feature-rich but gradually become blurry, dimmer, lossy or noisy (see Fig. 4), ePDE-g_5 offers a more significant number of inliers than other state-of-the-art feature detectors except ORB, KAZE and AKAZE as shown in Fig. 6(a). As expected, ePDE-g_4 generates lesser number of inliers when compared to ePDE-g_5 due to its vague structure of nonlinear scale spaces (see Fig. 2). Despite having a lesser number of inliers, ePDE-g_4 still outperformed MSER, SURF, AGAST and BRISK in detecting stable inliers. As shown in Fig. 6(b), ePDE-g_4 and ePDE-g_5 produce broader spread of inliers across the overlapping region than other feature detectors. Their spread-overlap performance is relatively equivalent to KAZE and AKAZE over feature-rich images. Figure 6(b) also shows that both ePDE-g_4 and ePDE-g_5 could hold at least 70% of correct matches from their detected interest points, which imply that they can generate enough reliable and matchable interest points within the overlapping region. In Fig. 6(c), both ePDE-g_4 and ePDE-g_5 have RC/S_o scores closer to the value of one (similar to SIFT, ORB, KAZE and AKAZE), suggesting that their detected inliers are more reliable in estimating proper alignment for feature-rich image stitching. An example in Fig. 7 demonstrates that both ePDE-g_4 and ePDE-g_5 can detect more inliers, not only widely spread across the overlapping region but also increase its detection capability at smoother, dimmer, lossy and near-uniform areas when compared to other state-of-the-art feature detectors.

4.2 General images analysis

In this section, we further examine the performance of the proposed method by combining the benchmark images and an additional 75 pairs of real-world and near-uniform scene images. These images are arbitrarily gathered from various datasets (captured by authors, researchers' datasets, and online resources), where certain areas of these images are either near-uniform or featureless. To determine the image stitching success rate for each feature detector, as shown in Fig. 8, we visually inspect the end result of every image stitching process based on how well the images pairs are aligned to each other, devoid of any visible severe distortion effects on the final stitched image. The image stitching success rate is calculated based on the percentages of the number of correctly stitched images out of the 100 pairs of sample images. As shown in Fig. 8, the performance of both ePDE-g_4 and ePDE-g_5 stands out when it comes to stitching near-uniform scene images, achieving a success rate of no less than 98%, which surpasses other feature detectors that achieved success rate below 95%. MSER shows the worst performance in our study, demonstrating only a 62% success rate.


Fig. 6 Performance analysis of the proposed method compared to other state-of-the-art feature detection methods. Each datum is obtained by averaging the evaluation responses across benchmark images (see Fig. 4)

(a) MSER   (b) SIFT   (c) SURF   (d) AGAST

(e) BRISK   (f) ORB   (g) KAZE   (h) AKAZE

(i) U-RIFT   (j) ePDE-g_4 (k) ePDE-g_5

Fig. 7 Both ePDE-g_4 and ePDE-g_5 demonstrate a significantly wider distribution of inliers when compared to other feature detection techniques in their resulting 'leuven' stitched images. White circles and cross markers indicate the approximate inlier locations

Fig. 8 Success rate of image stitching and the average number of inliers for various feature detectors across 100 pairs of sample images. Both ePDE-g_4 and ePDE-g_5 achieve a success rate of 98% or higher when compared to other feature detectors
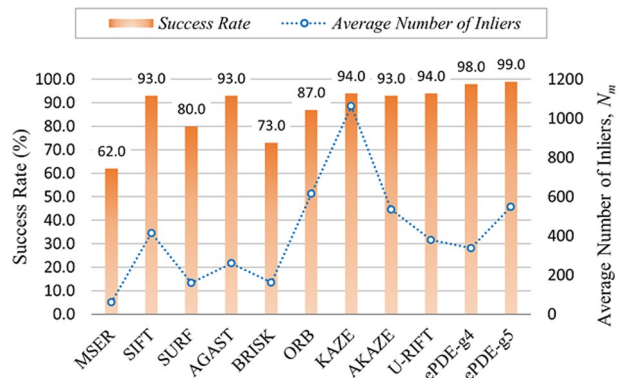

Most of the unpleasantly stitched images generally occur when the detected inliers are insufficient and intensely scattered within a smaller area of the overlapping region. Figure 8 also clarifies that even some popular feature detectors, like KAZE that holds the greatest number of detected inliers, do not necessarily reflect its capability to achieve a better success rate for near-uniform scene image stitching. KAZE scores 5% lower in its image stitching success rate than the proposed ePDE-g_4 and ePDE-g_5. This implies that both ePDE-g_4 and ePDE-g_5 are comparatively more robust and sensitive than other feature detectors in near-uniform scene image stitching, as they create sufficient inliers within the overlapping region. The advantage of both ePDE-g_4 and ePDE-g_5 is further supported by their inliers' spread-overlap results, as shown in Fig. 9(a), which reveals their potential to produce a larger number of widely distributed inliers within the overlapping region when compared

(a) N_m and S_o vs detectors (b) N_m and RC vs detectors (c) N_m and RC/S_o vs detectors
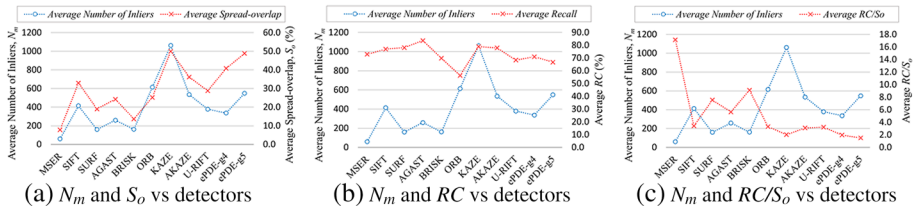
Fig. 9 Performance analysis for various feature detectors over 100 pairs of sample images that consist of 25 pairs of benchmark images and 75 pairs of real-world near-uniform scene images
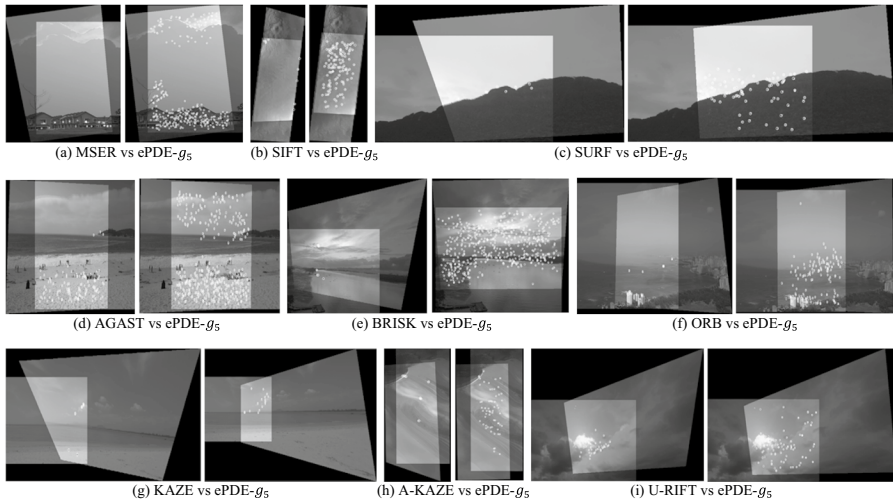

(a) Stitched images 1 (b) Stitched images 2 (c) Stitched images 3 (d) Stitched images 4
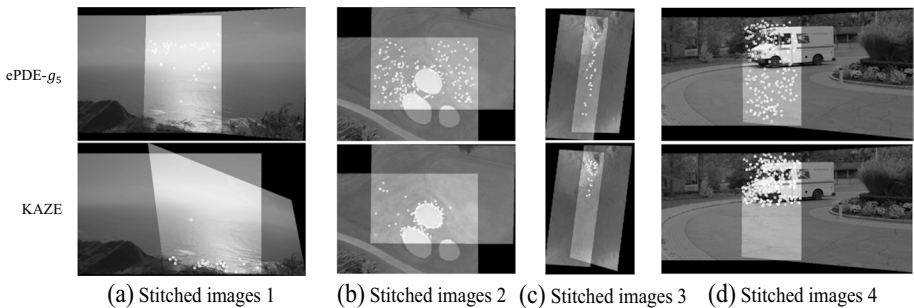
Fig. 10 The ePDE-g_5 exhibits a notably wider spread of inliers compared to KAZE, leading to less distortion in stitched images (specifically Fig. 10(**a**) and (**b**)). White circles and cross markers represent the approximate locations of inliers

to other feature detectors. Although KAZE demonstrates an impressive result in terms of the average inliers count and spread-overlap (as shown in Fig. 9(a)), it does not appear to excel in stitching near-uniform images, as indicated in Fig. 7, Fig. 10, and Fig. 11. By visually inspecting over 100 pairs of images, we notice that KAZE and A-KAZE generally perform well over feature-rich regions but not in the near-uniform areas. To support this claim, we provide a comparison of final stitched images for the detected inliers between KAZE and ePDE-g_5, as illustrated in Fig. 10.

For each feature detector, the evaluation metrics in Fig. 9 are expressed as average values across the 100 pairs of sample images. Figure 9(b)-(c) illustrates additional performance comparisons together with the number of inliers for various feature detectors, i.e., the recall (RC) measure and the RC/S_o score, respectively. Figure 9(b) shows that all feature detectors can retain at least 65% of their corresponding interest points on average as inliers, except for ORB, which scores only 56.4% in the recall measure. The recall outcomes for both ePDE-g_4 and ePDE-g_5 are regarded as reasonably satisfying. This is because the success of image stitching typically depends more on the quantity and distribution of inliers than on the recall percentage of correct matches. As depicted in Fig. 9(c), both ePDE-g_4 and ePDE-g_5 outperform other feature detectors by expressing the RC/S_o ratio closer to one, which implicitly suggests that the detected inliers are expected to be more precise in approximating decent image alignment. Hence, we can anticipate precise image stitching from the proposed ePDE-g_4 and ePDE-g_5, given their capability to identify sufficient inliers with more extensive spatial distribution. This is, in fact, important for the precise estimation of alignment between near-uniform scene images. On the contrary, a

Table 3 Comparison of feature detection runtimes in ascending order

Method	ORB	A-KAZE	SURF	AGAST	SIFT	MSER	BRISK	KAZE	U-RIFT	ePDE-g_5	ePDE-g_4
Averaged Runtime (s)	0.022	0.047	0.053	0.077	0.087	0.226	0.229	0.404	2.375	3.710	3.979

Fig. 11 Examples demonstrating how the ePDE-g_5 method surpasses other feature detection techniques in terms of the spatial distribution of inliers within the overlapping region of the resulting stitched images (in grayscale). White circles and cross markers represent the approximate locations of inliers

higher RC/S_o ratio tends to create noticeable image distortion, often leading to failure in image stitching.

Fig. 11 presents the resulting stitched near-uniform scene images, each highlighting the estimated position of detected inlier within the overlapping region. These images are provided for qualitative comparisons among various feature detectors. For each comparison in Fig. 11, the state-of-the-art method produces the first image, while the second image is created using the proposed ePDE-g_5 method. Stitching these images (as shown in Fig. 11) is undoubtedly challenging because of their relatively homogeneous and low-texture image content. When limited inliers are tightly clustered in a small region, it generally offers less information to an accurate estimation of the global geometric transformation between overlapping images. This often results in noticeable misalignment, ghost effects, and image distortion. Such problems can be seen in Fig. 11. For example, the stitched images produced by BRISK and KAZE, illustrated in Fig. 11(e) and Fig. 11(g), exhibit distortion, while MSER, AGAST and U-RIFT appear to produce severe misalignment, as seen in Fig. 11(a), Fig. 11(d), and Fig. 11(i), when compared to the end products of ePDE-g_5. Notice also that the majority of the state-of-the-art feature detectors face challenges in detecting a sufficient inlier, as illustrated in examples in Fig. 11(b), Fig. 11(c), Fig. 11(f), and Fig. 11(h)), particularly when working with near-uniform scene images. Considering these comparisons and analyses, the proposed methods demonstrate superior performance in image stitching when compared to other state-of-the-art methods, particularly for the near-uniform scene images. This is achieved by generating a more extensive and widely distributed set of inliers.

However, this enhancement comes at the expense of increased computation complexity compared to other state-of-the-art methods (see Table 3). As shown in Table 3, the runtimes are averaged across 20 images, highlighting the fastest execution times observed for each investigated feature detection method. The longer runtime of our method is primarily due to the iterative process involved in generating comprehensive nonlinear diffusion scale-space

representations and detecting extensive inliers in our feature detection method. While we acknowledge that the longer execution runtime is a significant concern for real-time applications, we believe that the trade-off in computational cost is well justified by the substantial improvement in image stitching quality and overall performance of our proposed method. In future work, we aim to explore optimization techniques, including patch-based methods that operate on localized image patches instead of the entire image, as well as the potential application of deep learning strategies to mitigate computational complexity challenges without compromising the effectiveness of our approach.

5 Conclusion

Inspired by Einstein's theory of special relativity, we have developed a new feature detection method based on Lorentz-modulated nonlinear scale spaces. This approach aims to enhance the performance of image stitching, particularly in challenging near-uniform scenes that often lack distinctive features due to their featureless or low texture nature. Our method addresses this challenge by incorporating the Lorentz factor into the formulation of the conductivity function of partial differential equation (PDE). This results in novel nonlinear scale spaces that offer richer multiscale structural information, making feature detection more robust. Our experimental results show that our method significantly outperforms many state-of-the-art methods, such as MSER, SIFT, SURF, BRISK, KAZE, A-KAZE, AGAST, ORB, and U-RIFT. Indeed, our method significantly enhances feature detection efficiency and the spatial distribution of inliers across the overlapping region of near-uniform scene images. Although KAZE and A-KAZE excel in feature-rich regions, its performance tends to decline in near-uniform areas. This paper primarily focuses on the conventional image stitching approach, utilizing a novel feature-based detection algorithm. We do not delve into deep learning (DL) as it is not within the scope of this study.

Furthermore, we have extended the evaluation method for image stitching performance by employing the latest criteria measures: the spread-overlap S_o measure and the RC/S_o score. These criteria offer several advantages over conventional evaluation metrics for assessing the performance of feature detectors and image stitching. The spread-overlap S_o measure provides valuable information about the inlier's spatial distribution within the overlapping region. On the other hand, the RC/S_o score is a reliable indicator to determine the success rate of image stitching. They are both of utmost importance in accurately evaluating the effectiveness of feature detectors and image stitching.

Our proposed feature detection method can be applied to enhance a wide range of multimedia applications, including panoramic stitching, virtual tours, surveillance, satellite imaging, automobile vision, virtual reality, immersive full dome visualization, and more. In line with our future interest, we intend to apply this method to enhance the fusion of astronomical images, thereby improving their matching accuracy for precise image stitching. Astronomical images often capture scenes that are near-uniform, featuring dim and blurry objects against a primarily uniform and noisy background. Detecting sufficiently correct and matchable feature points in these overlapping images for precise alignment is a challenging task. The difficulty often results in misalignment, severe distortion, and visible artifacts (such as ghosting and blurring effects), leading to misinterpretations of astronomical studies. We firmly believe that our proposed Lorentz-based nonlinear diffusion feature detection holds the potential in addressing and improving the challenges associated with astronomical image stitching.

Declarations

Conflict of Interests The authors declare that they have no conflict of interests.

References

1. Szeliski R (2007) Image Alignment and Stitching: A Tutorial. Found Trends® Comput Graph Vis 2:1–104. https://doi.org/10.1561/0600000009
2. Hossein-Nejad Z, Nasri M (2022) Clustered redundant keypoint elimination method for image mosaicing using a new Gaussian-weighted blending algorithm. Vis Comput 38:1991–2007. https://doi.org/10.1007/S00371-021-02261-9/METRICS
3. Miao X, Qu T, Chen X, He C (2023) Superpixel-based foreground-preserving image stitching. Mach Vis Appl 34:1–13. https://doi.org/10.1007/S00138-022-01363-1/METRICS
4. Qin Y, Li J, Jiang P, Jiang F (2021) Image stitching by feature positioning and seam elimination. Multimed Tools Appl 80:20869–20881. https://doi.org/10.1007/S11042-021-10694-6/TABLES/2
5. Brown M, Lowe DG (2007) Automatic panoramic image stitching using invariant features. Int J Comput Vis 74:59–73. https://doi.org/10.1007/S11263-006-0002-3/METRICS
6. Lyu W, Zhou Z, ChenZhou LY (2019) A survey on image and video stitching. Virtual Real Intell Hardw 1:55–83. https://doi.org/10.3724/SP.J.2096-5796.2018.0008
7. Pandey A, Pati UC (2019) Image mosaicing: A deeper insight. Image Vis Comput 89:236–257. https://doi.org/10.1016/j.imavis.2019.07.002
8. Xiang T-Z, Xia G-S, Bai X, Zhang L (2018) Image stitching by line-guided local warping with global similarity constraint. Pattern Recognit 83:481–497. https://doi.org/10.1016/j.patcog.2018.06.013
9. Li Y, Wang S, Tian Q, Ding X (2015) A survey of recent advances in visual feature detection. Neurocomputing 149:736–751. https://doi.org/10.1016/j.neucom.2014.08.003
10. Harris C, Stephens M (1988) A combined corner and edge detector. In: Taylor CJ (ed) Proceedings of the Alvey Vision Conference 1988. Alvey Vision Club, pp 23.1–23.6. https://doi.org/10.5244/C.2.23
11. Lowe DG (2004) Distinctive Image Features from Scale-Invariant Keypoints. Int J Comput Vis 60:91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94
12. Şekeroğlu K, Soysal ÖM (2017) Comparison of SIFT, Bi-SIFT, and Tri-SIFT and their frequency spectrum analysis. Mach Vis Appl 28:875–902. https://doi.org/10.1007/S00138-017-0868-9/METRICS
13. Bay H, Tuytelaars T, Van Gool L (2006) SURF: speeded up robust features. In: Leonardis A, Bischof H, Pinz A (eds) Computer Vision – ECCV 2006. ECCV 2006. Lecture notes in computer science, vol 3951. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11744023_32
14. Agrawal M, Konolige K, Blas MR (2008) CenSurE: center surround extremas for realtime feature detection and matching. In: Forsyth D, Torr P, Zisserman A (eds) Computer Vision – ECCV 2008. ECCV 2008. Lecture notes in computer science, vol 5305. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-88693-8_8
15. Alcantarilla PF, Bartoli A, Davison AJ (2012) KAZE features. In: Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C (eds) Computer Vision – ECCV 2012. ECCV 2012. Lecture notes in computer science, vol 7577. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33783-3_16
16. Perona P, Malik J (1990) Scale-Space and Edge Detection Using Anisotropic Diffusion. IEEE Trans Pattern Anal Mach Intell 12:629–639. https://doi.org/10.1109/34.56205
17. Wang S, You H, Kun Fu (2012) BFSIFT: A Novel Method to Find Feature Matches for SAR Image Registration. IEEE Geosci Remote Sens Lett 9:649–653. https://doi.org/10.1109/LGRS.2011.2177437
18. Alcantarilla PF, Nuevo J, Bartoli A (2013) Fast explicit diffusion for accelerated features in nonlinear scale spaces. In: Proceedings British Machine Vision Conference 2013 – BMVC 2013, pp 13.1–13.11. https://bmva-archive.org.uk/bmvc/2013/Papers/paper0013/
19. Li D, Xu Q, Yu W, Wang B (2020) SRP-AKAZE: an improved accelerated KAZE algorithm based on sparse random projection. IET Comput Vis 14:131–137. https://doi.org/10.1049/iet-cvi.2019.0622

20. Zhang Q, Wang Y, Wang L (2015) Registration of images with affine geometric distortion based on Maximally Stable Extremal Regions and phase congruency. Image Vis Comput 36:23–39. https://doi.org/10.1016/j.imavis.2015.01.008

21. Rosten E, Porter R, Drummond T (2010) Faster and Better: A Machine Learning Approach to Corner Detection. IEEE Trans Pattern Anal Mach Intell 32:105–119. https://doi.org/10.1109/TPAMI.2008.275

22. Mair E, Hager GD, Burschka D, Suppa M, Hirzinger G (2010) Adaptive and generic corner detection based on the accelerated segment Test. In: Daniilidis K, Maragos P, Paragios N (eds) Computer Vision – ECCV 2010. ECCV 2010. Lecture notes in computer science, vol 6312. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-15552-9_14

23. Li J, Hu Q, Ai M (2020) RIFT: Multi-Modal Image Matching Based on Radiation-Variation Insensitive Feature Transform. IEEE Trans Image Process 29:3296–3310. https://doi.org/10.1109/TIP.2019.2959244

24. Hu M, Sun B, Kang X, Li S (2023) Multiscale structural feature transform for multi-modal image matching. Inf Fusion 95:341–354. https://doi.org/10.1016/J.INFFUS.2023.02.026

25. Chen Y, Zhang X, Zhang Y, Maybank SJ, Fu Z (2018) Visible and infrared image registration based on region features and edginess. Mach Vis Appl 29:113–123. https://doi.org/10.1007/S00138-017-0879-6/METRICS

26. Banerjee A, Das N, Santosh KC (2022) Weber local descriptor for image analysis and recognition: a survey. Vis Comput 38:321–343. https://doi.org/10.1007/S00371-020-02017-X/METRICS

27. Calonder M, Lepetit V, Strecha C, Fua P (2010) BRIEF: binary robust independent elementary features. In: Daniilidis K, Maragos P, Paragios N (eds) Computer Vision – ECCV 2010. ECCV 2010. Lecture notes in computer science, vol 6314. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-15561-1_56

28. Leutenegger S, Chli M, Siegwart RY (2011) BRISK: Binary Robust invariant scalable keypoints. In: Proceeding IEEE International Conference on Computer Vision, 2548–2555. Barcelona, Spain. https://doi.org/10.1109/ICCV.2011.6126542

29. Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: an efficient alternative to SIFT or SURF. In: Proceeding IEEE International Conference on Computer Vision, 2564–2571. Barcelona, Spain. https://doi.org/10.1109/ICCV.2011.6126544

30. Alahi A, Ortiz R, Vandergheynst P (2012) FREAK: fast retina keypoint. In: Proceeding IEEE Conference on Computer Vision and Pattern Recognition, 510–517. Providence, RI, USA. https://doi.org/10.1109/CVPR.2012.6247715

31. Xu Z, Liu Y, Du S, Wu P, Li J (2016) DFOB: Detecting and describing features by octagon filter bank for fast image matching. Signal Process Image Commun 41:61–71. https://doi.org/10.1016/j.image.2015.12.001

32. O'Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez GV, Krpalkova L, Riordan D, Walsh J (2020) Deep Learning vs. Traditional Computer Vision. Adv Intell Syst Comput 943:128–144. https://doi.org/10.1007/978-3-030-17795-9_10

33. Romero-González C, García-Varea I, Martínez-Gómez J (2022) Shape binary patterns: an efficient local descriptor and keypoint detector for point clouds. Multimed Tools Appl 81:3577–3601. https://doi.org/10.1007/S11042-021-11586-5/FIGURES/18

34. Sun S, Park U, Sun S et al (2023) Fusion representation learning for keypoint detection and description. Vis Comput 39:5683–5692. https://doi.org/10.1007/s00371-022-02689-7

35. Youssef S, el Shehaby M, Fayed S (2020) A Smart multi-view panoramic imaging integrating stitching with geometric matrix relations among surveillance cameras (SMPI). Multimed Tools Appl 79:30917–30981. https://doi.org/10.1007/S11042-020-09432-1/TABLES/17

36 Ma J, Jiang X, Fan A, Jiang J, Yan J (2021) Image Matching from Handcrafted to Deep Features: A Survey. Int J Comput Vis 129(23):79. https://doi.org/10.1007/s11263-020-01359-2

37. Lindeberg T (1994) Scale-Space Theory : A Basic Tool for Analysing Structures at Different Scales. J Appl Stat 21:225–270. https://doi.org/10.1080/757582976

38. Weickert J (1996) Theoretical foundations of anisotropic diffusion in image processing. In: Kropatsch W, Klette R, Solina F, Albrecht R (eds) Theoretical foundations of computer vision. Computing supplement, vol 11. Springer, Vienna. https://doi.org/10.1007/978-3-7091-6586-7_13

39. Weickert J, Ishikawa S, Imiya A (1999) Linear scale-space has first been proposed in Japan. J Math Imaging Vis 10:237–252. https://doi.org/10.1023/A:1008344623873/METRICS

40. Witkin AP (1984) Scale-space filtering: A new approach to multi-scale description. In: ICASSP '84. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp 150–153. San Diego, CA, USA. https://doi.org/10.1109/ICASSP.1984.1172729

41. Koenderink JJ (1984) The structure of images. Biol Cybern 50:363–370. https://doi.org/10.1007/BF00336961/METRICS

42. Duits R, Florack L, De Graaf J, Ter Haar RB (2004) On the axioms of scale space theory. J Math Imaging Vis 20:267–298. https://doi.org/10.1023/B:JMIV.0000024043.96722.AA/METRICS

43. Bhatia H, Kirby RM, Pascucci V, Bremer PT (2021) Vector Field Decompositions Using Multiscale Poisson Kernel. IEEE Trans Vis Comput Graph 27:3781–3793. https://doi.org/10.1109/TVCG.2020.2984413

44. Weickert J (2001) Efficient image segmentation using partial differential equations and morphology. Pattern Recognit 34:1813–1824. https://doi.org/10.1016/S0031-3203(00)00109-6

45. Weickert J, Ter Haar Romeny BM, Viergever MA (1998) Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Trans Image Process 7:398–410. https://doi.org/10.1109/83.661190

46. Bambi C (2018). Introduction to General Relativity. https://doi.org/10.1007/978-981-13-1090-4

47. Ralli J (2011) Fusion and regularisation of image information in variational correspondence methods. Dissertation, Universidad de Granada. https://produccioncientifica.ugr.es/documentos/618f4f639f f8c939aaca3fd6?lang=en

48. Evans C (2009) Notes on the OpenSURF library. University of Bristol, Tech. Rep. CSTR-09-001. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6c7cf406a47048730c1a08d46 cb0166b16566524

49. Vedaldi A, Fulkerson B (2010) Vlfeat - An open and portable library of computer vision algorithms. In: MM'10 - Proceedings of the 18th ACM international conference on Multimedia, 1469–1472. https://doi.org/10.1145/1873951.1874249

50. Torr PHS, Zisserman A (2000) MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. Comput Vis Image Underst 78:138–156. https://doi.org/10.1006/cviu.1999.0832

51. Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE Trans Pattern Anal Mach Intell 27:1615–1630. https://doi.org/10.1109/TPAMI.2005.188

52. Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T, Van GL (2005) A Comparison of Affine Region Detectors. Int J Comput Vis 65:43–72. https://doi.org/10.1007/ s11263-005-3848-x

53. Jong TK, Bong DBL (2019) Evaluation of local features for near-uniform scene images. In: Proceedings 2019 IEEE international conference signal image processing application ICSIPA 2019, 95–99. Kuala Lumpur, Malaysia. https://doi.org/10.1109/ICSIPA45851.2019.8977755

54. Schmid C, Mohr R, Bauckhage C (2000) Evaluation of interest point detectors. Int J Comput Vis 37:151–172. https://doi.org/10.1023/A:1008199403446/METRICS

55. Marmol A, Peynot T, Eriksson A, Jaiprakash A, Roberts J, Crawford R (2017) Evaluation of Keypoint Detectors and Descriptors in Arthroscopic Images for Feature-Based Matching Applications. IEEE Robot Autom Lett 2:2135–2142. https://doi.org/10.1109/LRA.2017.2714150

56. Affine Covariant Features (2007). Visual Geometry Group, Department of Engineering Science, University of Oxford. https://www.robots.ox.ac.uk/~vgg/research/affine/. Accessed 19 Nov 2023

57. Zaragoza J, Chin TJ, Tran QH, Brown MS, Suter D (2014) As-Projective-As-Possible Image Stitching with Moving DLT. IEEE Trans Pattern Anal Mach Intell 36:1285–1298. https://doi.org/10.1109/ TPAMI.2013.247

58. OpenCV: Open Source Computer Vision Library, OpenCV extra data. GitHub. https://github.com/ opencv/opencv_extra/tree/master/testdata/stitching. Accessed 19 Nov 2023

59. Li J, Deng B, Tang R, Wang Z, Yan Y (2020) Local-Adaptive Image Alignment Based on Triangular Facet Approximation. IEEE Trans Image Process 29:2356–2369. https://doi.org/10.1109/TIP.2019.2949424

60. Adobe Open Source Data Sets. Adobe Systems, Inc. https://sourceforge.net/adobe/adobedatasets/home/ Home/. Accessed 19 Nov 2023

61. Photojournal: NASA's Image Access Home Page. Jet Propulsion Laboratory, California Institute of Technology. https://photojournal.jpl.nasa.gov/. Accessed 11 Feb 2023

Tze Kian Jong received his B.Sc. degree in Applied Physics from University of Malaya in 2001 and the M.Eng. degree in Telecommunication Engineering from Universiti Malaysia Sarawak in 2007. He is currently pursuing Ph.D. degree in Electronic Engineering at Universiti Malaysia Sarawak. He was involved in academia and university R&D for more than 12 years. He is currently a government official of the Ministry of Science, Technology, and Innovation Malaysia located at the National Planetarium, Kuala Lumpur, Malaysia. His research interests include image processing, astronomy, and planetarium. He is a member of the Malaysia Board of Technologists.

David B. L. Bong received his Bachelor of Electrical Engineering degree from Universiti Teknologi Malaysia, Master of Science degree in Computer Control & Automation from Nanyang Technological University, Singapore, and Ph.D. degree in Digital Signal Processing from Universiti Sains Malaysia. He has 7 years of industrial working experience in the R&D and manufacturing sectors in Singapore prior to his academia involvement for more than 19 years. He is a Senior Member of IEEE, Fellow of the Institution of Engineers Malaysia, and Professional Engineer of the Board of Engineers Malaysia. He is currently an associate professor with the Faculty of Engineering, Universiti Malaysia Sarawak.