Smart Structures and Systems, *Vol. 28, No. 5 (2021) 661-673*
https://doi.org/10.12989/sss.2021.28.5.661

Feature-based image stitching for panorama construction and visual inspection of structures

Kai Cheng [1a], Jiazeng Shan*[1,2] and Yuwen Liu [1b]

[1] *Department of Disaster Mitigation for Structures, Tongji University, Shanghai 200092, China*
[2] *Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China*

(*Received November 2, 2020, Revised July 25, 2021, Accepted August 9, 2021*)

Abstract. This study presents a feature-based image stitching method with multi-level constraint criterion for panorama construction and visual inspection of building structures. The comparison of global view and local resolution over building exterior is discussed regarding practical implementation. An inspection-oriented methodology framework with optimized inlier distribution is designed for generating a feasible and reliable building panorama by using ordinary optic images. Two illustrative examples, including an earthquake-damaged masonry wall and a high-rise building with stone curtain walls, are experimentally investigated. The severely developed structural crack is fully mapped with stitched image and extracted in preparation for further quality evaluation. The curtain wall of the high-rise building is successfully constructed by using UAV-based images. The panorama quality is further compared with commercial stitching software and several improvements are illustrated in the particular case. In addition, the reliability of the proposed feature-based stitching approach is parametrically studied with different setups of input images.

Keywords: computer vision; image stitching; multi-level constraint criterion; structural health monitoring; visual inspection

1. Introduction

Structural condition assessment has been an ever-growing demand for civil infrastructures during their long-term serving life and after disastrous events such as earthquakes and strong winds. Generally, structural health monitoring (SHM) and visual inspection have been developed as two distinct but not mutually exclusive approaches (Agdas *et al*. 2016). During the past decades, significant efforts and remarkable progress have been devoted to SHM technologies, mostly related to vibration-based system identification and damage evaluation in an automated way (Iacovino *et al*. 2018, Wu and Jahanshahi 2020). Visual inspection is another important evaluation methodology mostly depending on manual observations and may be the predominant method for infrastructure inspection in a relatively long history of engineering application. Representative structural and damage indices can be visually characterized such as crack propagation (distribution, length, and width), steel corrosion, concrete spalling, and component failures.

Though the technological importance highlighted, visual inspection approaches were usually recognized as labor-consuming regarding the manual observation and the large-scale infrastructures. Correspondingly, three challenges of timing, interpretability and accessibility are summarized for visual inspection methods (Moore *et al*. 2001). Recently, such practical issues and potential limitations on visual inspection have been partially improved because of significant developments of image processing, computer vision, and autonomous equipment such as unmanned aerial vehicle (UAV) (Hoskere *et al*. 2019, Jung *et al*. 2019, Lai *et al*. 2020), and becomes practically promising in a non-contact, computer-aided and automated way (Choi *et al*. 2018, Hsu *et al*. 2020). Such visual inspection has also been characterized as image-based structural health monitoring (Ye *et al*. 2016, Akbar *et al*. 2018), and can be implemented remotely and more frequently by employing UAV-based SHM systems.

In the context of visual observation and inspection, the photogrammetry-based digital imaging provides the fundamental data and predominant source, and the image quality may determine the following assessment accuracy and reliability directly. Regarding the engineering size of high-rise buildings and long-span bridges, the constraints of camera view range, the coverage of a single photograph, and the region of interest (ROI) of target structures may inevitably generate a practical conflict in real-world implementation, as the trade-offs between the need for collecting close-up images and their relevant localization (Choi *et al*. 2018). Illustratively, the typical optical camera provides a field view of 35×50 degrees, which is much lower than that of human visual system with 135×200 degrees. Furthermore, a major challenge in visual inspection may be the extraction of small fraction of local meaningful regions from the high volumes of images

*Corresponding author, Associate Professor,
E-mail: jzshan@tongji.edu.cn
[a] Graduate Student, E-mail: 1932242@tongji.edu.cn
[b] Graduate Student, E-mail: 1932244@tongji.edu.cn

ISSN: 1738-1584 (Print), 1738-1991 (Online)

(Yeum *et al*. 2017).

To address this problem, the panorama image stitching has been an emerging solution in building inspection with a relatively acceptable balance between global vision and close-up resolution. Technically, the close-up images of the building structures are sequentially captured by state-of-the-art low-cost and high-performance vision sensors, and then a high-resolution orthophoto representing a complete view of the target structures is generated by manual or automated image stitching. Usually, there are two types of image alignment and stitching algorithms, the direct method, and the feature-based method, which are related in the computer vision field. Several commercial application software packages are available, such as Autostitch (Brown and Lowe 2007), Photoshop (Adobe 2006), and Microsoft Image Editor (Microsoft 2008), where the panorama can be created by the user inputs.

In image stitching, feature detection and matching play a crucial role and may determine the panorama quality. Correspondingly, there are several key issues requiring to be investigated, such as the reliable correspondences between two sets of image features, the high computational complexity and costs, the development of advanced outlier removal algorithms, and the capturing of the perspective changes for scenes. Recently, a new feature descriptor named L^2-SIFT has been proposed by Sun *et al*. (2014) for large images in large-scale aerial photogrammetry, a robust feature matching algorithm using spatial clustering has been developed by Jiang *et al*. (2020). A robust image stitching method with multiple registration algorithm has been presented by Herrmann *et al*. (2018). A warping residual-based image stitching for large parallax has been developed by Lee and Sim (2020), and Nie *et al*. (2021) proposed an unsupervised deep image stitching method to stitch images with few features or low resolution.

In engineering practice, the quality of the panorama images may not necessarily be guaranteed regarding engineering applications (Bang *et al*. 2017). The image stitching and registration has been primarily investigated by several researchers in different areas. Representative effort and remarkable progress include the multi-image stitching and scene reconstruction for evaluating defect evolution (Jahanshahi *et al*. 2011), the image preprocessing for automatic generation of construction site panorama (Bang *et al*. 2017), the autonomous image localization method (Yeum *et al*. 2017), and post-event visual evaluation of building facade (Choi *et al*. 2018). Obviously, the feature-based stitching may be popular due to its computational efficiency, stitching robustness, and automatic ability. Several feature descriptors have been broadly used for detection and matching, including SIFT (Scale-invariant feature transform) (Lowe 2004, Jahanshahi *et al*. 2011), SURF (Speed-up robust features) (Bay *et al*. 2008, Akbar *et al*. 2018, Yuan *et al*. 2019) and so on. Although different stitching methods have been presented, the inspection-oriented panorama construction for building structures may still be an open problem, especially for stitching images with accuracy and efficiency and for UAV-based remote sensing techniques and applications (Xiang *et al*. 2019).

In this paper, a feature-based image stitching method with multi-level constraint criterion is proposed for panorama construction and visual inspection of large-scale building structures. The necessity and framework of the image stitching method are firstly presented. The methodology of the image stitching algorithm with optimization on inlier distribution is then stated. Followingly, the potential value of the proposed image stitching method on visual inspection is presented with two detailed illustrations, while comparing the optimal stitching performance of the proposed method with two commercial software and discussing the reliable stitching performance with different setups of input images.

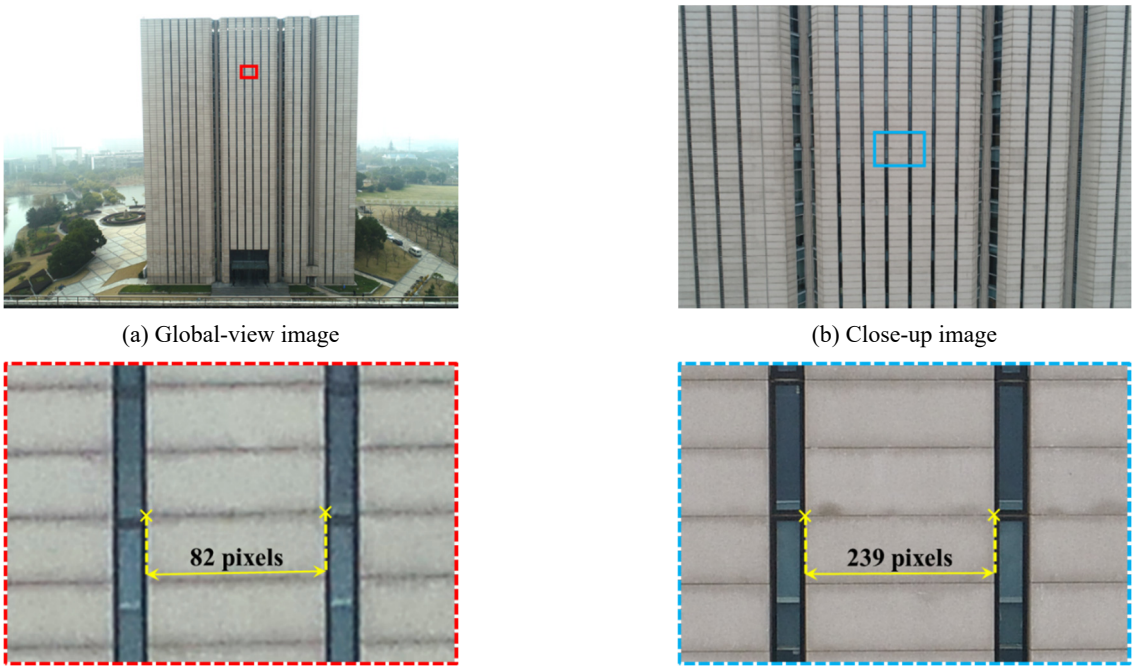
2. Motivation and framework

During the long-term serving life of civil infrastructures, the primary visual quality assessment and the following statistical quantity evaluation for the built environment is of great interest for structural maintenance. As discussed above, the time-consuming and laborious manual inspection method may be relatively difficult to meet the rapidly increasing demands of structure exterior inspection, regarding both inspection expense, efficiency and accuracy. Therefore, the automated image stitching for generating an entire presentation of building exteriors becomes popular and informative by using multiple pictures from ordinary cameras.

In the authors' opinion, there may be two essential and practice-valuable interests requiring to be addressed with trade-off: (1) the stitching quality of global vision over the mostly similar or even identical architectural texture or appearance such as stone and glass curtain walls; (2) the image resolution of close-up vision over the structural exterior areas satisfying the inspection requirements such as the pixel-level quantification of length, width, and area.

Illustratively, Fig. 1 illustrates the local ROI images from a real-world building structure with stone curtain walls under global and close-up photogrammetry scenarios. As illustrated in Fig. 1, for the same region of interest, the pixel-level width of the illustrated stone curtain wall is 82 pixels in the global-view image as shown in Fig. 1(c), while that in the close-up image is 239 pixels as shown in Fig. 1(d). Obviously, the resolution and local imaging details are much better for further usages. Meanwhile, the zoomed ROI of the red dashed block can be easily located in the global-view image of Fig. 1(a) but may not be straightforward for the blue dashed block to obtain its location relative to the whole building exteriors by using the close-up image of Fig. 1(b).

Generally, if the picture taken directly by the consumer-grade cameras covers the whole global view of the civil infrastructure, the corresponding resolution of the global image of the large-scale building exteriors may be too blurred to inspect the conditions of the region of interest (ROI). On the contrary, if the close-up images are taken with satisfactory image resolution of the ROI of the building exteriors, the overall localization of the ROI may become difficult and inaccurate. Given the above trade-off problem, without the assistance of image processing



(a) Global-view image (b) Close-up image

(c) Zoom in on global-view image (d) Zoom in on close-up image

Fig. 1 Comparison of resolution difference of the same region of interest between the global-view and the close-up exterior image of a typical high-rise building

techniques, it may be practically difficult to obtain the high-resolution image of building exteriors simultaneously with global facade view and overall localization information. Therefore, the panorama construction for the global view of large-scale building structures with the stitching over close-up images may be desired and valuable for visual inspection and structural maintenance.

To fully address this issue, a feature-based image stitching method may be an available and feasible approach for aiding the visual inspection efficiently and automatically, which can also preferably solve the problems within large-scale infrastructure maintenance (Spencer *et al*. 2019, Strauss *et al*. 2020, Sun *et al*. 2020). Typically, a feature-based image stitching procedure consists of the following steps (Szeliski 2007, Jahanshahi *et al*. 2011):

(1) Distinctive keypoints detection. Feature points such as edges, corners, vertices, line intersections are detected and used as the criterion to estimate the similarity and overlap regions between corresponding images.
(2) Feature matching and outliers exclusion. The detected keypoints are matched to determine which features come from corresponding locations in different images. Once an initial set of feature correspondences has been computed, outliers exclusion is needed to produce a high accuracy alignment and compute the homography matrix. Two widely used algorithms are called RANdom Sample Consensus (RANSAC) (Fischler and Bolles 1981) and Least Median of Squares (LMS) (Rousseeuw 1984).
(3) Image blending and composition. Blending the images to compensate for exposure differences and other misalignments. Then the homography transformation matrix is applied as geometric transformation to the images.
(4) Furthermore, bundle adjustment is needed to simultaneously adjust pose parameters for an extensive collection of overlapping images.

There are many feature detector techniques, such as SIFT, SURF, Harris (Harris and Stephens 1988), FAST (Features from Accelerated Segment Test) (Rosten *et al*. 2010), ORB (Oriented fast and Rotated Brief) (Rublee *et al*. 2011) and so on. SIFT detector is a popular choice for feature detection due to its robustness to scale and rotation changes and its invariance to changes in the three-dimensional (3D) viewpoint and illumination. SURF detector improves the computation time of SIFT by using an integral image for fast local gradient computations on images, but partially sacrificing scale and rotation invariance (Luo and Oubong 2009). Harris detector is used to detect corner keypoints but not invariant to scale changes and cross correlation. FAST detector is a corner detector and faster than SIFT or SURF detector, but has no rotation or scale invariance. ORB detector is extremely fast operation, and is scale and rotation invariant, robust to noise and affine transformations, while sacrificing very little on performance accuracy.

Following the illustration and comparison of the global-view image and close-up image in Fig. 1, the proposed framework of the feature-based image stitching methodology with inliers distribution optimization is designed and illustrated in Fig. 2. The detailed procedure of the proposed feature-based image stitching method includes

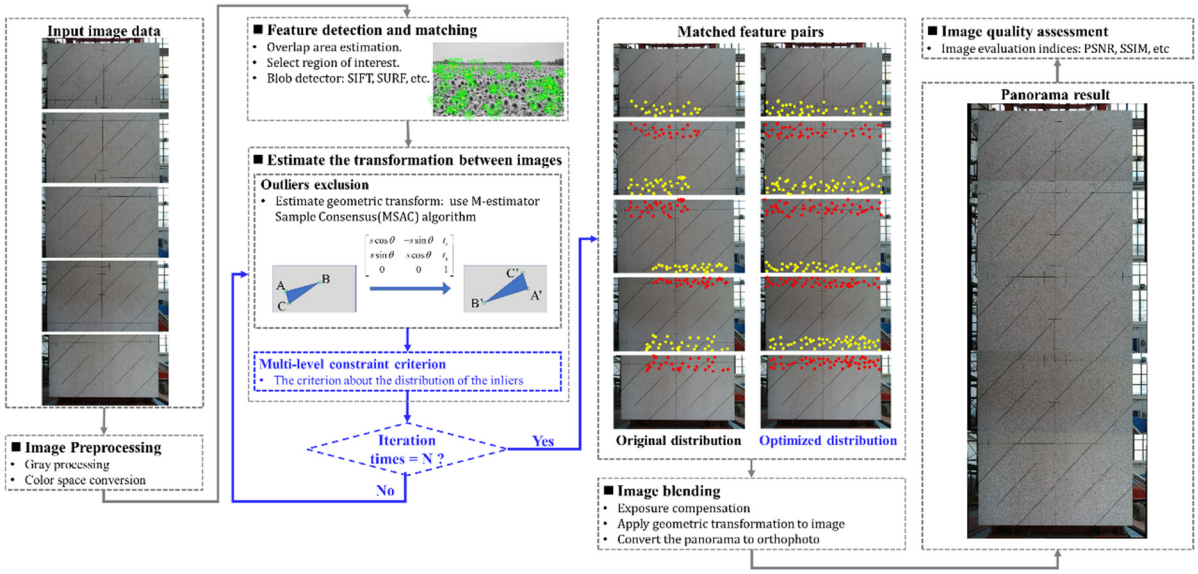


Fig. 2 Designed framework for the proposed feature-based multi-image stitching method with optimization on inlier distribution (yellow and red feature points as matched pairs)

the following essential steps:

(1) feature detection and matching, where the overlap regions between images are located and selected as ROI, and the blob detectors are adopted to be extracted initially and matched correspondingly based on the comparison of Euclidean distance between the descriptors of keypoints in corresponding images;
(2) transformation estimation, where the geometric transformation is online estimated between adjacent images by adopting the M-estimator SAmple Consensus (MSAC) algorithm (Torr and Zisserman 2000) which is a variant of the RANSAC algorithm and partially compensating the undesirable effect of threshold sensitivity, and further optimized by multi-level constraint criterion;
(3) iterative optimization, where the optimal group of geometric transform parameters is obtained to be applied to the target paired images via optimization on inliers distribution.
(4) image blending, where the contrast and exposure difference are compensated, the geometric transformation is applied to the images and the panorama result is converted to orthophoto;
(5) image quality assessment (IQA), where the full reference image evaluation indices such as Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity (SSIM), are adopted to estimate the quality of panorama objectively.

As illustrated in Figs. 1 and 2, the detected and matched features with the preliminary input of ROI may reduce the rate of feature mismatching significantly, in the particular case of panorama construction for texture-similar building exteriors. Then, the estimation of the homography transformation matrix is implemented by performing the MSAC algorithm-based outliers exclusion controlled by the proposed multi-level constraint criterion. Especially, as illustrated in the block of matched feature pairs in Fig. 2, the yellow points in each image are matched to the red points in the neighboring image right below it. It's shown that the conventional process of transform estimation directly utilizes the initial feature correspondences, while the proposed image stitching method iteratively optimizes the distribution of the matched features controlled by the multi-level constraint criterion as marked in blue blocks. Practically, such evenly distributed paired features may improve the accuracy of the homography transform matrix and the image quality of the constructed panorama.

3. Methodology

3.1 Preliminary feature extraction and matching

Input image sequences are supposed to have an overlap region between adjacent images. In order to quantify the proportion of the selected overlap region relative to the whole image, the overlap ratio is defined and further utilized to evaluate the potential influence of the image overlap on the panorama result of our proposed method. As illustrated in Fig. 3, the overlap ratio ϕ can be calculated in the following formula

$$\phi = d_o / d_v \tag{1}$$

where d_o is the pixel-level distance represents the overlapping scope, and d_v is the pixel-level distance represents the viewing scope of UAV.

After ascertaining the overlap region and overlap ratio, the ROI can be selected to frame the feature detection area. To describe the blob features within the ROI, the SIFT algorithm (Lowe 2004, VLFeat 2008) is adopted in the present study. As a classic image correlation method,

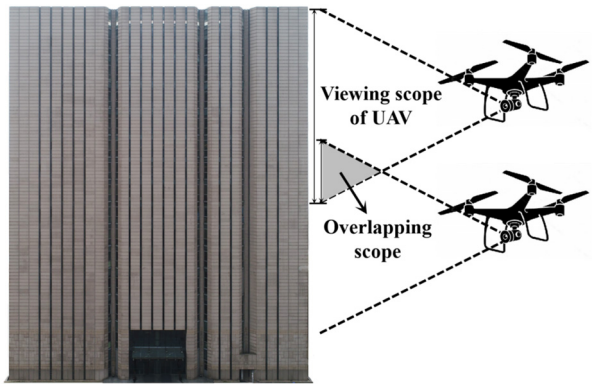


Fig. 3 Definition of the overlap ratio between two individual aerial imaging cases

a SIFT keypoint descriptor consists of a $4 \times 4 \times 8 = 128$ dimensional vector, which is reliably invariant to image translation, scaling, and rotation, and partially invariant to illumination changes and projection. Two illustrations of the SIFT descriptor detected from an earthquake-damaged masonry wall are shown in Fig. 4, and each box, with green lines and arrows in the picture, contains an eight-dimension vector which describes eight orientations in the box, respectively. Especially due to the most similar or even identical architectural texture or appearance, the difference of keypoints may be subtle, even with 128-dimensional vectors as their descriptors. The introduction of ROI may significantly reduce the rate of mismatching and reduce the computational complexity and costs.

With the extraction of SIFT features in the ROI between adjacent images, features can be then matched correspondingly according to the SIFT descriptors. By means of comparing the Euclidean distance between the feature descriptors in adjacent images, the feature pairs with the minimum Euclidean distance are supposed to be the candidate pairs of feature matching. The Euclidean distance D of descriptor d_a and descriptor d_b can be calculated as the following formula

$$D = \sqrt{\sum_{i=1}^{N=128} (d_a(i) - d_b(i))^2} \tag{2}$$

3.2 Estimate the transformation between images

The result of preliminary feature matching includes inliers and outliers. Inliers are the correct feature matching pairs we need to calculate the transformation matrix, but outliers are the incorrect pairs that mislead to work out a wrong transformation matrix, so the elimination of outliers is practically necessary. In order to detect outliers, the MSAC algorithm (Torr and Zisserman 2000) is utilized, where the feature pairs are randomly chosen to fit the homography transformation matrix. Initially, a minimal set of randomly sampled points are utilized to estimate the parameters of the homography transformation matrix by the MSAC algorithm. Then, the remaining feature points are divided into inliers and outliers in the rule of whether it can fit the model within the allowable error threshold.

The pixel coordinates of the feature matching pairs, which are supposed to be inliers, should satisfy the following equation or be within the allowable error threshold. The pairs without meeting the equation will be regarded as unqualified feature matching pairs, and then back to implement the MSAC algorithm again.

$$\boldsymbol{u}^T \boldsymbol{H} \boldsymbol{v} = \boldsymbol{0} \tag{3}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are the pixel coordinates of one feature matching pair in the adjacent images, and can be represented in the form of $[x \quad y \quad 1]^T$ and T is the transpose operator. The homography matrix, $\boldsymbol{H}$, is a 3-by-3 matrix and describes the projective transformation of feature point correspondences in the same plane. The matrix can be derived from four intrinsic parameters of the physical camera setup, namely the focal length, pitch, height, and principle point of the pinhole camera model.

Then the parameters of the homography transformation matrix are optimized by iterating the above procedures for specified times. There is a cost function to judge whether the parameters are optimal, that's also the difference from the RANSAC algorithm. In the RANSAC algorithm, the loss function is shown as the following formula

$$Loss_{\text{RANSAC}}(e^2) = \begin{cases} 0, & e^2 < T^2 \\ constant, & e^2 \geq T^2 \end{cases} \tag{4}$$

where T is the error threshold which is used to divide the

(a) One SIFT descriptor of crack area

(b) One SIFT descriptor of spalling area

Fig. 4 Illustration of SIFT keypoint descriptors within two distinct damage scenarios

inliers and outliers, and e^2 is the square of maximum likelihood estimation error (Torr and Zisserman 2000).

One of the shortcomings of the RANSAC algorithm is the sensitivity of the results regarding the selection of error threshold T. Therefore, the MSAC algorithm is developed to optimize the loss function in order to reduce the sensitivity to the threshold. The loss function of the MSAC algorithm can be calculated as the following formula

$$Loss_{\mathrm{MSAC}}(e^2) = \begin{cases} e^2, & e^2 < T^2 \\ T^2, & e^2 \geq T^2 \end{cases} \tag{5}$$

Although the MSAC algorithm can filter out most of the potential outliers and calculate the homography transformation matrix acceptably, the simple dependence on the MSAC method may not fully satisfy the image stitching demands of the building exteriors, especially the building exteriors with mostly similar or even identical architectural texture or appearance. If the inliers filtered by the MSAC algorithm are not evenly distributed in the overlap regions, the local image misalignment may occur and the accumulative error will cause the failure of image stitching eventually.

To avoid this problem and assure the inliers are distributed evenly for the particular study of building exteriors, a multi-level constraint criterion is proposed, which is the major distinguishing characteristic between the currently developed method and the existing image stitching algorithms, such as the image stitching available in OpenCV (2011). Regarding the geometric scale and cover similarity of building exteriors, both the accurate registration and evenly distribution of feature points are the fundamental motivation and concerns for the proposed method. Accordingly, the multi-level constraint criterion includes the following four formulas

$$width_{inliers} \geq \alpha_1 \times Width_{image} \tag{6a}$$

$$height_{inliers} \geq \alpha_2 \times Height_{image} \tag{6b}$$

$$\mathbf{max}\,(distance_{inliers}) \geq \alpha_3 \times Diagonal_{overlap} \tag{6c}$$

$$convhull_{inliers} \geq \alpha_4 \times Area_{overlap} \tag{6d}$$

where $width_{inliers}$ and $height_{inliers}$ are the width and height of the boundary enclosed by inliers, respectively, while $Width_{image}$ and $Height_{image}$ are the width and height of the target image, respectively; $distance_{inliers}$ is the pixel-level distance between inliers with the maximum operator $\mathbf{max}\,()$, while $Diagonal_{overlap}$ is the maximum distance of the overlap region in the diagonal direction; $convhull_{inliers}$ is the area of convex hull enclosed by the inliers, while $Area_{overlap}$ is the area of the overlap region.

In the present study, four controlling coefficients, assigned as α_1, α_2, α_3, and α_4, are proposed. As relevant to the overlap ratio ϕ, the parameter value ranges between zero and unity, where the unity value of 1.0 indicates the equivalence to the variable ϕ. The four coefficients are designed to guarantee the desired even distribution of the inliers, as previously illustrated in Fig. 2. α_1 and α_2 are relevant to the horizontal and vertical overlap ratio respectively, α_3 represents the constraint to the diagonal distribution, and α_4 is relevant to the ratio of overlap area to the whole image. Especially, the increase of the value of α_3 indicates the stricter constraint of the diagonal distribution.

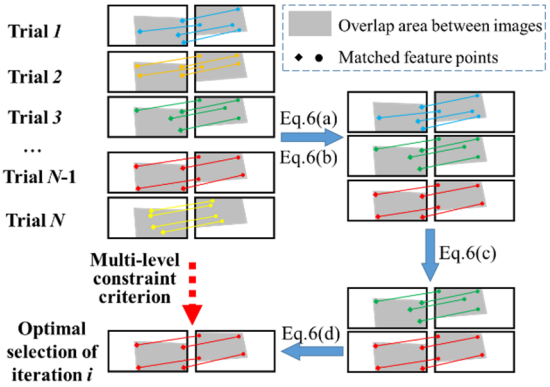


Fig. 5 Multi-level constraint criterion for optimizing inlier distribution

Screening processing of optimizing inlier distribution is illustrated in Fig. 5. Owing to the randomness of the initial selection of feature points in the MASC algorithm, in a single iteration of outliers elimination processing, the MSAC algorithm will trial N times to improve the image transformation robustness, and every individual trial will generate a set of inliers and a homography transformation matrix. Usually, it needs about 500 times iteration to obtain stable results. And to avoid occasionality, we used to set 1000 as the iteration times.

Multi-level constraint criterion plays an important role to pick out the most evenly distributed set of inliers, which has the maximum number of inliers as well. Eqs. (6a) and (6b) can pick out the trials whose distribution of inliers occupied most of the width and height of the ROI enclosed by inliers. Eqs. (6c) and (6d) can eliminate the set of inliers that distributes intensively in a triangular area. Finally, the trials with the evenly distributed inliers are retained, then the specific trail with picking the maximum quantity of inliers becomes the optimal selection of ith iteration. After the assigned iterations are performed, the set of inliers with the evenest distribution and the maximum quantity of inliers is proposed to be the final and converged feature matching pairs. Meanwhile, the resultant homography transformation matrix will be applied in the follow-up step.

3.3 Image blending

After obtaining the optimal feature matching pairs consisting of inliers only and the homography transformation matrix, the exposure difference can be compensated in two steps. (1) In the HSV (Hue, Saturation, Value) color space, firstly, computing the scale factor of V value between the feature matching pairs in corresponding images. Then multiplying the V channel of the corresponding images by the scale factor of V value to compensate for the saturation difference. (2) In the RGB (Red, Green, Blue) color space, the scale factors of R

channel, G channel and B channel are then computed, respectively. The scale factors are multiplied to the target images in the corresponding channels to correct the color difference. Then the paired images are sequentially transformed by applying the homography transformation matrix. In the particular case of the present study, if the building in the panorama result is not shown in the front view, the conversion of the panorama into an orthophoto is then implemented for better evaluation.

3.4 Image quality assessment

To evaluate the stitched panorama quality, a classic index out of the full reference (FR) image evaluation indices, as named of SSIM (Zhou *et al*. 2004), is then adopted. As compared to the evaluation indices of no reference (NR) image and reduced reference (RR) image, FR indices can relatively objectively assess the quality of image stitching results. The calculation of SSIM is based on the image luminance, contrast and structure, and is in accordance with human feelings to some extent. The value of SSIM ranges from -1.0 to 1.0, and the increase of the SSIM value indicates the increased similarity of the two target images as a good quality assessment. Theoretically, the SSIM value of 1.0 corresponds to the scenario of two images being identical. SSIM can be calculated as following formula

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1) + (\sigma_x^2 + \sigma_y^2 + c_2)} \quad (7)$$

where x, y represent reference image and panorama, respectively; μ_x, μ_y are the mean intensity of x and y, respectively; σ_x, σ_y are the standard deviation of x and y, respectively; σ_{xy} represent the covariance of x and y; c_1, c_2 are the constant to avoid instability.

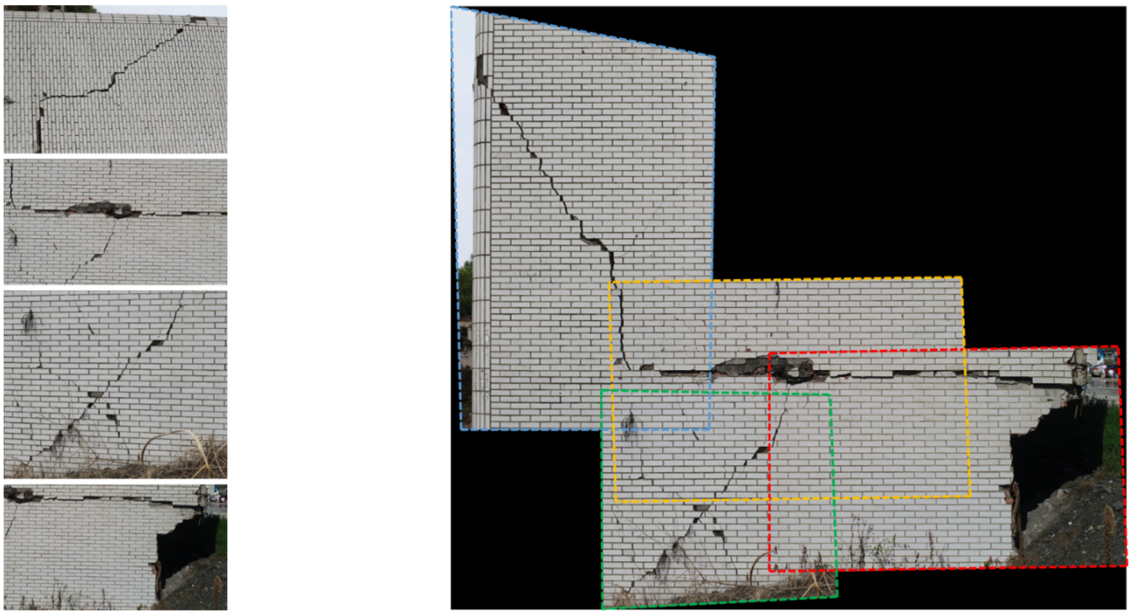
4. Experimental investigation

The present feature-based image stitching method was systematically evaluated by two illustrative examples. The first example is the exteriors of an earthquake-damaged masonry wall in Beichuan, Sichuan province, while the other example is the stone curtain walls of structural exteriors of a high-rise building located in the Jiading campus of Tongji University.

4.1 Earthquake-damaged masonry exterior

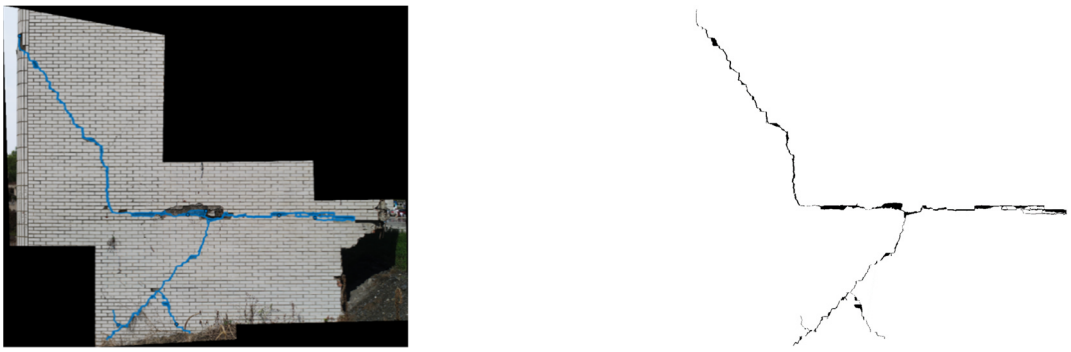
In recent years, with the development of computer vision, the quantitative assessment of seismic damage has become a popular field, including wall cracks, steel exposure and concrete spalling. Structural cracks, as the most direct and distinct indication of seismic damages, can always be visually observed on the building exteriors after the earthquakes. As for the early-stage cracks with small geometry scale, the mapping of the target crack is relatively convenient to with the whole scenes while restoring high-resolution. Such vision data of cracks may support the quantitative assessment with high precision. On the contrary, for the severe cracks with relatively large-scale length and distribution, the high-resolution details of the seismic cracks are of necessity and informative, with a cluster of close-up images taken manually.

Illustratively, Fig. 6 presents a case of the stitching results of a large-scale wall crack distributing over the whole facade of the masonry structure. As shown in Fig. 6(a), the four individual images shot different parts of the long-length and multi-crossing crack. Obviously, neither the crack length nor the development orientation can be easily concluded from just one single view. The pixel resolutions of the input images are 5184×3456, 2736×1539, 3500×2957, 2736×1539 from top to bottom, respectively.



(a) Input images (b) Output panorama result

Fig. 6 Panorama result of earthquake-damaged masonry exterior



(a) Panorama-based crack edge detection (b) Pixel-level crack area extraction

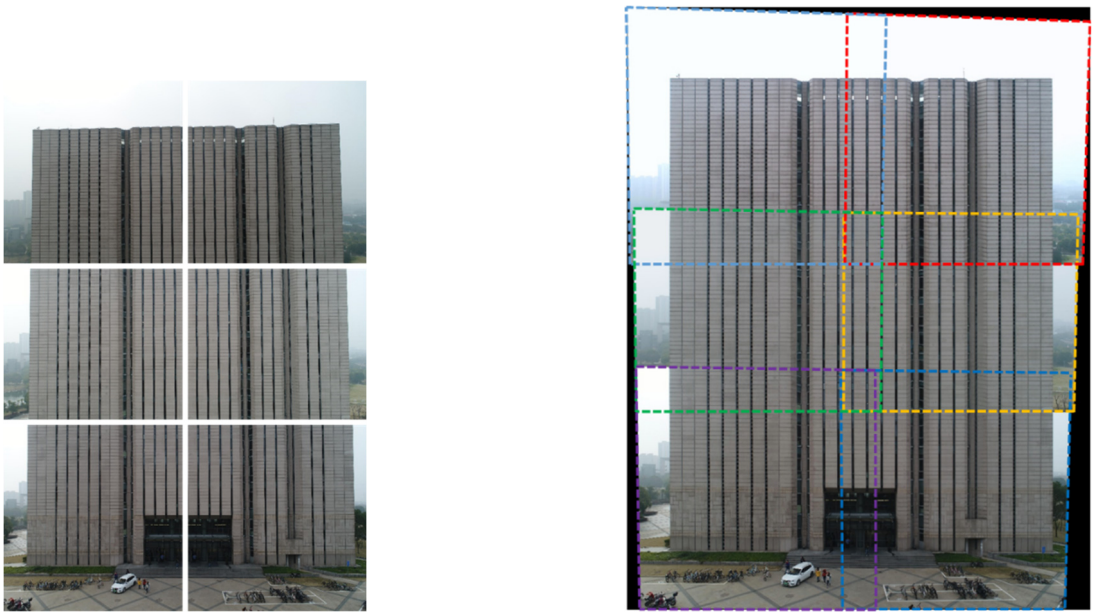
Fig. 7 Illustration of the computer-aided quantitative assessment of the wall crack that is fully visible in the panorama image

The proposed feature-based stitching method was then implemented to the four crack images. The overlap ratio between the neighboring images from top to bottom is approximately estimated to be 35%, 45% and 55% between the blue and orange boxes, the orange and green boxes, and the orange and red boxes in Fig. 6(b). The vector of the controlling coefficients $[\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ is assumed to be [0.40, 0.35, 0.90, 0.80], [0.40, 0.55, 0.90, 0.80], and [0.60, 0.50, 0.90, 0.80] in sequence. With different resolutions of the same in the input images, various input orders will lead to getting diverse output results which differ from the resolutions, in which processing the down-sampling or up-sampling are executed automatically.

Finally, the maximum resolution of the resultant panorama is illustrated in Fig. 6(b), with an overall pixel resolution of 22266×18708. During the optimization step, the converged number of inlier pairs are 395, 2168, and 398 for the overlap regions intersected between the blue and orange boxes, the orange and green boxes, and the orange and red boxes in Fig. 6(b), respectively. Furthermore, with the help of the image stitching technique, all the visual quantitative assessment can be calculated, such as the edge of the crack can be semantically segmented and the total area of the extracted crack is calculated, as illustrated in Fig. 7.

4.2 Stone curtain wall of high-rise building exteriors

The curtain wall unit, as an important symbol of modern architecture, has been widely used as an external enclosure and decorative structure due to its lightness and structure beauty. In practice, the current safety-state evaluation and damage detection are relatively time-consuming and requiring intense human-labor. Targeting on this particular maintenance problem, a high-resolution and global-view image would be significantly necessary as the fundamental vision data for inspection. The feature-based image stitching method with the UAV-based data acquisition may provide a low-cost and fast technique for preliminary visual damage detection.



(a) Input images (b) Output panorama result

Fig. 8 Panorama result of stone curtain wall of high-rise building exteriors

A field test of visual inspection over the high-rise building was performed by using a consumer-grade UAV, DJI Phantom 4 Pro. Fig. 8(a) was an illustration of panorama result with 25% as horizontal and vertical overlap ratio, the six input images, with blue, red, green, orange, purple and dark blue dashed boxes in sequence, were then stitched together, and the panorama of stone curtain walls of building exteriors was generated with the resolution of 5905×7682, as shown in Fig. 8(b).

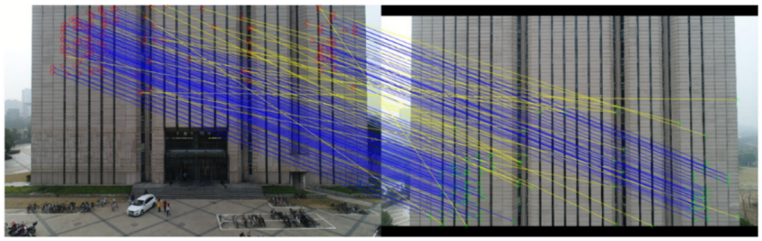
Compared to the exposure difference of input images as shown in Fig. 8(a), the performance of exposure compensation is apparent in Fig. 8(b). Then, the roles MSAC algorithm and multi-level constraint criterion played in the whole method are illustrated in Fig. 9, and the paired images of IMG-1, IMG-2 and IMG-3 are stitched from the two images in the top, middle and bottom line of the input images in Fig. 8(a), respectively. The two groups of detected features linked by blue and yellow lines, respectively, represent the inliers and the outliers, respectively. The final paired features between IMG-1 IMG-2 are 104 out of the total matched number of 532, including the outliers, while that between IMG-2 and IMG-3 is 92 out of total 798. It is observed that the final paired features number is not relevant to the total matched number. With the help of MSAC algorithm and multi-level constraint criterion, outliers are eliminated, and the inliers are relatively distributed evenly in the shared overlap region.

To illustrate the stitching performance, the panorama obtained by the proposed method is further compared to the results generated by two commercial software. The selected stitching software are AutoStitch (Brown and Lowe 2007) and Photoshop with the Photomerge function (Adobe 2006). Two local regions of illustration are compared in Fig. 10, where the regions enclosed by the red and blue dashed box are Region I and II, respectively. In Region I as shown in Fig. 10(d), the misalignment of the stone curtain wall bricks is shown in the result of AutoStitch, while the edges of the curtain wall bricks are acceptably aligning as shown in the subplots by Photoshop of Fig. 10(e) and our proposed method of Fig. 10(f). Similarly in Region II, the stitched results by AutoStitch in Fig. 10(g) and Photoshop in Fig. 10(h) exhibit the apparent misalignment as marked by orange dashed ellipse, while our proposed method presents an accurate alignment result as shown in Fig. 10(j).

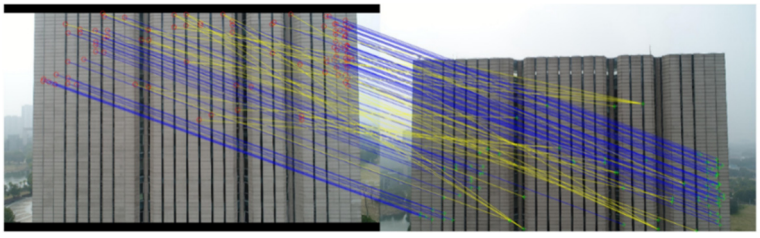
The reasons that lead to the quality difference in panorama results are not only the multi-level constraint criterion but also the potential limitation of built-in algorithms in AutoStitch and Photoshop. The RANSAC algorithm, which is proved to show relatively poorer performance than the MASC algorithm, is adopted by AutoStitch. Furthermore, the homography transformation matrix is emphasized with the minimization of the cost function, while the quantity and distribution of inliers in the overlap area may not be the focus. The user guide of Photomerge in Photoshop explained that the images input should have a sufficient overlap ratio, which is recommended to be no less than 40% but no larger than 70%. The position and exposure to take the series of input images are also suggested to be the same.

Besides comparing the local details, the accuracy of the global outline is also compared for performance illustration, as shown in Fig. 11. The red lines, blue lines, and green lines indicate the outline edges of the high-rise building exteriors within the panoramas generated by AutoStitch, Photoshop, and our proposed method, respectively. The yellow line represented the vertical and horizontal lines. As compared in Fig. 11, it is observed that the building panoramas output by AutoStitch and Photoshop were noticeably crooked and distorted. Meanwhile, the proposed image stitching method presents the picture of building exterior close to the practical reality.

To explore the potential influence of image overlapping on the stitching performance of the proposed method, two scenarios of input images with six overlap ratio setups within each scenario are assumed and assigned for

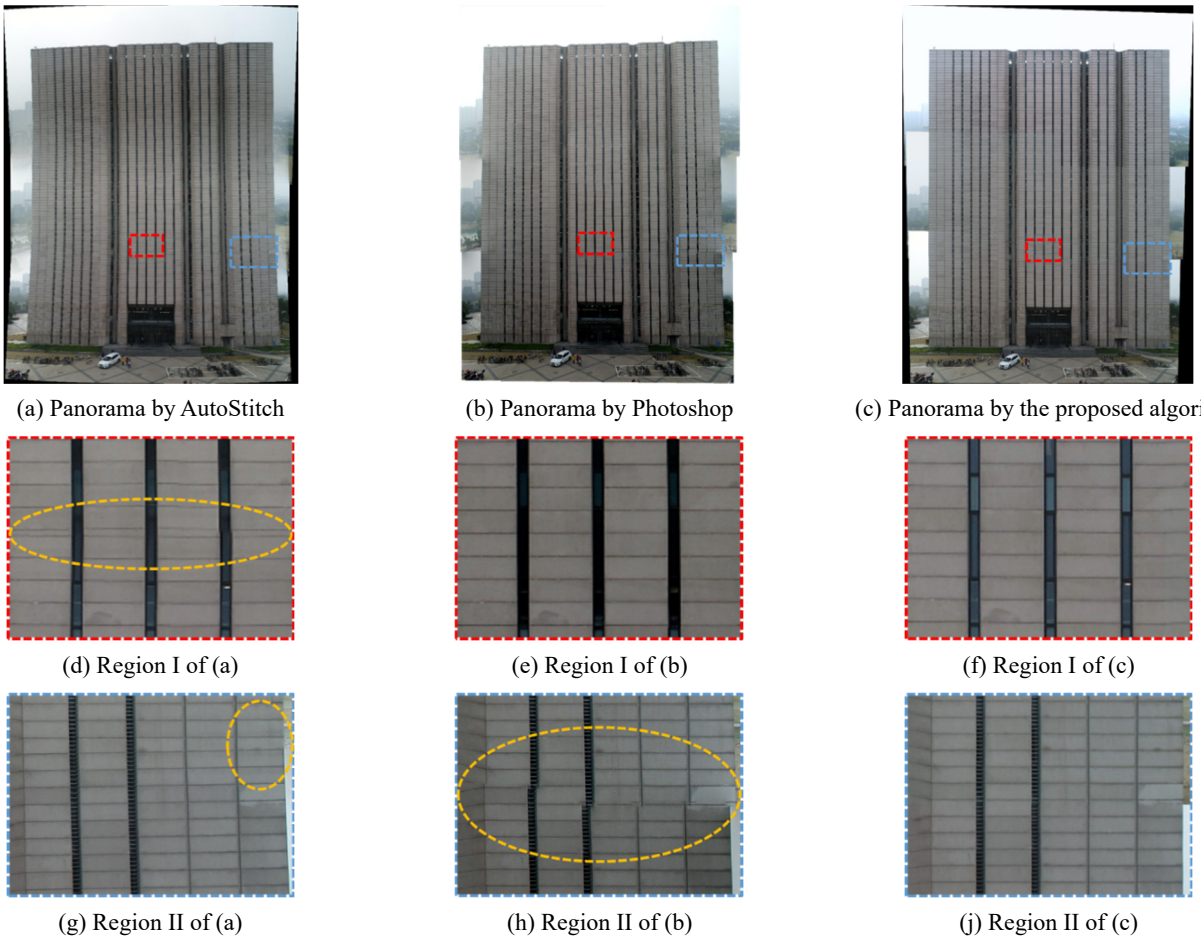


(a) Matching results between IMG-1 and IMG-2



(b) Matching results between IMG-2 and IMG-3

Fig. 9 Illustration of inliers and outliers distinguished by the adopted MSAC algorithm and the proposed multi-level constraint criterion

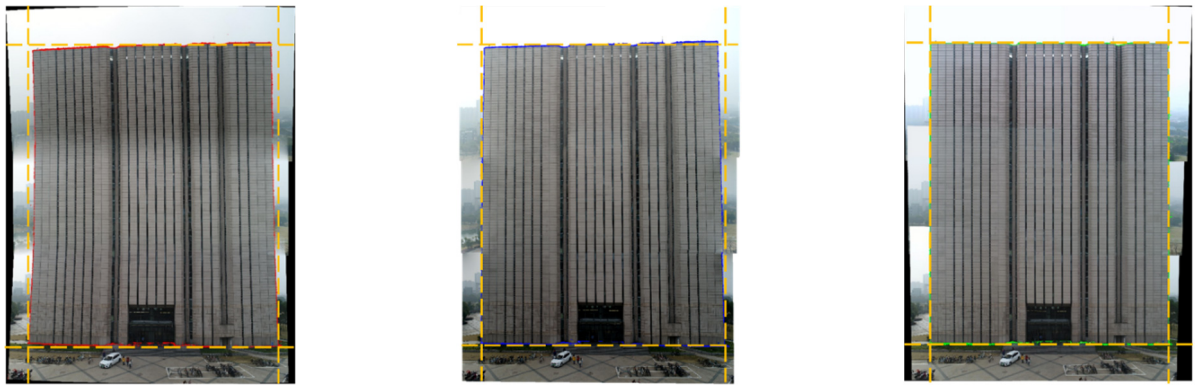


(a) Panorama by AutoStitch (b) Panorama by Photoshop (c) Panorama by the proposed algorithm

(d) Region I of (a) (e) Region I of (b) (f) Region I of (c)

(g) Region II of (a) (h) Region II of (b) (j) Region II of (c)

Fig. 10 Comparison of the stitching quality of three panorama images in illustrative local regions



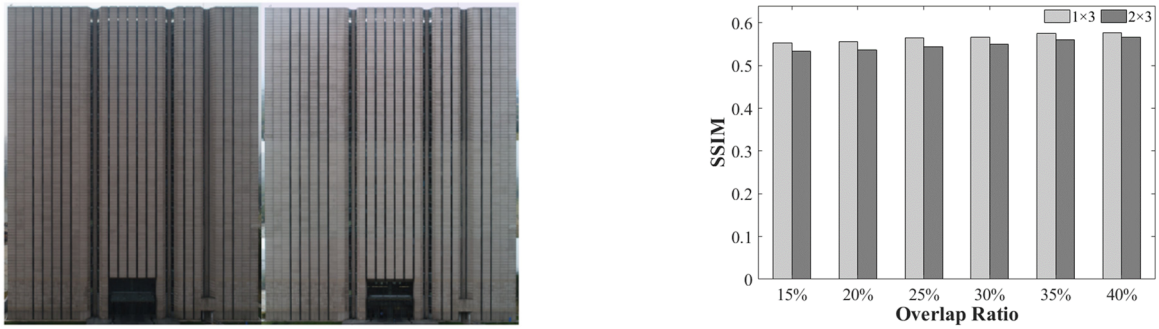
(a) Panorama by AutoStitch (b) Panorama by Photoshop (c) Panorama by the proposed algorithm

Fig. 11 Comparison of the fundamental rectangular shape of panorama images by inserting global outlines

parametric study. A set of input images, including three photos shoot from three aircraft altitudes by the UAV, are prior cropped. Several sequences of input images with horizontal and vertical overlap ratio gradients are obtained, and the fundamental information of the image sets are listed in the following Table 1. Obviously, the first scenario with the 1×3 setup was not split at each aircraft altitude, and the vector of the controlling coefficients $[\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ is assigned to be [0.10, 0.70, 0.90, 0.80], [0.15, 0.70, 0.90, 0.80], [0.20, 0.70, 0.90, 0.80], [0.25, 0.70, 0.90, 0.80], [0.32, 0.70, 0.90, 0.80], and [0.36, 0.70, 0.90, 0.80] in sequence. The second scenario with the 2×3 setup represents that original images were cropped into two images at each aircraft altitude with different horizontal overlap ratio, while different vertical overlap ratio gradient were also cropped. The overlap ratio was determined regarding the practical implementation with the range between 15% and 40%with 5% as the step size, and the vector of the controlling coefficients $[\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ is assigned to be [0.12, 0.12, 0.90, 0.80], [0.17, 0.17, 0.90,

Table 1 Basic information of image sets

Input image sets	Overlap ratio	Pixel resolution		
		Upper view	Middle view	Lower view
Original set of input images	45% (approximately)		5472×3678	
1×3	15%	5472×3029	5472×2501	5472×3087
	20%	5472×3119	5472×2681	5472×3177
	25%	5472×3211	5472×2865	5472×3269
	30%	5472×3302	5472×3048	5472×3361
	35%	5472×3393	5472×3230	5472×3452
	40%	5472×3485	5472×3413	5472×3543
2×3	15%	2958×3029	2958×2501	2958×3087
	20%	3040×3119	3040×2681	3040×3177
	25%	3127×3211	3127×2865	3127×3269
	30%	3219×3302	3219×3048	3219×3361
	35%	3317×3393	3317×3230	3317×3452
	40%	3420×3485	3420×3413	3420×3543



(a) Reference image and panorama image (b) Image Quality Assessment: SSIM

Fig. 12 Image quality assessment by employing different sets of input images

0.80], [0.21, 0.21, 0.90, 0.80], [0.26, 0.26, 0.90, 0.80], [0.32, 0.32, 0.90, 0.80], and [0.36, 0.36, 0.90, 0.80] in sequence, basing on the overlap ratio.

After stitching the input images with different overlap ratio setups, the image quality assessment was implemented by using the adopted SSIM as an image evaluation index, as illustrated in Fig. 12(a). It is observed that the SSIM slightly decreases from 0.5776 for the most optimal ratio of 40% to 0.5531 for the most challenging ratio of 15%, regarding the scenario with the 1×3 setup. In addition, the proposed method seems to be relatively reliable in stitching performance with the illustration of similar SSIM values between the two assigned scenarios and even between the different overlap ratio setups.

5. Conclusions

A feature-based image stitching method is logically designed and proposed for visual inspection of building exteriors, with the inspection-oriented requirements on global view and local resolution highlighted. The concept of multi-level constraint criterion is introduced for optimized inlier distribution and is adopted in the proposed method framework for better panorama construction.

Two illustrative examples were presented to experimentally investigate the proposed feature-based iterative image stitching method. It is shown that a severely developed structural crack on a masonry wall can be globally extracted and locally segmented with satisfactory pixel-level accuracy, by using four ordinary images. In the second case, the overview of the curtain wall of a high-rise building is successfully generated by using high-quality UAV-based close-up photos. Two local regions and the global rectangular edges are illustrated for comparison with the proposed method and two commercial stitching software. The observed misalignment of curtain bricks may indicate the improved stitching performance of the proposed approach over the general stitching software in this particular texture-similar case. Furthermore, the proposed method seems to be relatively reliable in stitching performance with a detailed parametric study regarding different setups of input images by using an evaluation index of SSIM.

Furthermore, in terms of image registration, the structure from motion (SfM) (Schönberger and Frahm

2016) can transform a sequence of images into a three-dimensional point cloud model, as compared to the image stitching generating the two-dimensional panorama. The procedures of SfM may be similar to image stitching, more computationally expensive, but can handle arbitrary surfaces with an arbitrary camera motion. However, the 3D point clouds generated by SfM may be sparse regarding zooming in the specific area, which may not provide sufficient local information. Differently, as illustrated in the current study, the 2D panorama from the presented image stitching method may accelerate the process and present satisfactory accuracy and details if the target structural region is manually or automatically determined. In addition, we can integrate the SfM with the current method for two-level image registration for a combined advantage of global point cloud segmentation and local image segmentation.

Currently, the proposed feature-based method and the stitching performance may be relatively influenced by the manual inputs and operation during the processing of image stitching. To continue the present feature-based image stitching study, a logical next step appears to be the automatic determination of the overlap region and more implementations over building, bridge, and tunnel inspection, to be embedded into the general and operational maintenance procedure.

Acknowledgments

This study is sponsored by the National Natural Science Foundation of China (Grant No: 51878483), the Key Laboratory of Shock and Vibration of Engineering Materials and Structures, Sichuan Province (No. 19kfgk03), the Shanghai Qi Zhi Institute (Grant No. SYXF0120020109), and the Peak Discipline Construction Project of Shanghai (No. 2021-CE-03).

References

Adobe Photoshop (2006), Create and edit a panorama in Photoshop, Adobe Photoshop, San Jose, USA. https://helpx.adobe.com/photoshop/using/create-panoramic-images-photomerge.html

Agdas, D., Rice, J.A., Martinez, J.R. and Lasa, I.R. (2016), "Comparison of visual inspection and structural-health monitoring as bridge condition assessment methods", *J. Perform. Constr Facil.*, **30**(3), 04015049. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000802

Akbar, M.A., Qidwai, U. and Jahanshahi, M.R. (2018), "An evaluation of image-based structural health monitoring using integrated unmanned aerial vehicle platform", *Struct. Control Health Monitor.*, **26**(1), e2276. https://doi.org/10.1002/stc.2276

Bang, S., Kim, H. and Kim, H. (2017), "UAV-based automatic generation of high-resolution panorama at a construction site with a focus on preprocessing for image stitching", *Automat. Constr.*, **84**, 70-80. https://doi.org/10.1016/j.autcon.2017.08.031

Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L. (2008), "Speeded-up robust features (SURF)", *Comput. Vis. Image Underst.*, **110**(3), 346-359. https://doi.org/10.1016/j.cviu.2007.09.014

Brown, M. and Lowe, D.G. (2007), "Automatic panoramic image stitching using invariant features", *Int. J. Comput. Vis.*, **74**(1), 59-73. https://doi.org/10.1007/s11263-006-0002-3

Choi, J., Yeum, C.M., Dyke, S.J. and Jahanshahi, M.R. (2018), "Computer-aided approach for rapid post-event visual evaluation of a building façade", *Sensors*, **18**(9), 3017. https://doi.org/10.3390/s18093017

Fischler, M.A. and Bolles, R.C. (1981), "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography", *Commun. ACM*, **24**(6), 381-395. https://doi.org/10.1145/358669.358692

Harris, C.G. and Stephens, M. (1988), "A combined corner and edge detector", *Alvey vision conference*, Manchester, UK, September.

Herrmann, C., Wang, C., Bowen, R.S., Keyder, E., Krainin, M., Liu, C. and Zabih, R. (2018), "Robust image stitching with multiple registrations", *Proceedings of European Conference on Computer Vision*, Munich, Germany, September.

Hoskere, V., Park, J.W., Yoon, H. and Spencer, B.F. (2019), "Vision-based modal survey of civil infrastructure using unmanned aerial vehicles", *J. Struct. Eng.*, **145**(7), 04019062. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002321

Hsu, T.Y., Pham, Q.V., Chao, W.C. and Yang, Y.S. (2020), "Post-earthquake building safety evaluation using consumer-grade surveillance cameras", *Smart Struct. Syst., Int. J.*, **25**(5), 531-541. https://doi.org/10.12989/sss.2020.25.5.531

Iacovino, C., Ditommaso, R., Ponzo, F.C. and Limongelli, M.P. (2018), "The interpolation evolution method for damage localization in structures under seismic excitation", *Earthq. Eng. Struct. Dyn.*, **47**(10), 2117-2136. https://doi.org/10.1002/eqe.3062

Jahanshahi, M.R., Masri, S.F. and Sukhatme, G.S. (2011), "Multi-image stitching and scene reconstruction for evaluating defect evolution in structures", *Struct. Health Monitor.*, **10**(6), 643-657. https://doi.org/10.1177/1475921710395809

Jiang, X., Ma, J., Jiang, J. and Guo, X. (2020), "Robust feature matching using spatial clustering with heavy outliers", *IEEE Transactions on Image Processing*, **29**, 736-746.

Jung, H.J., Lee, J.H., Yoon, S. and Kim, I.H. (2019), "Bridge inspection and condition assessment using unmanned aerial vehicles (UAVs): major challenges and solutions from a practical perspective", *Smart Struct. Syst., Int. J.*, **24**(5), 669-681. https://doi.org/10.12989/sss.2019.24.5.669

Lai, Z., Alzugaray, I., Chli, M. and Chatzi, E. (2020), "Full-field structural monitoring using event cameras and physics-informed sparse identification", *Mech. Syst. Signal Process.*, **145**, 106905. https://doi.org/10.1016/j.ymssp.2020.106905

Lee, K.Y. and Sim, J.Y. (2020), "Warping residual based image stitching for large parallax", *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, Washington, USA, June.

Lowe, D.G. (2004), "Distinctive image features from scale-invariant keypoints", *Int. J. Comput. Vis.*, **60**(2), 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Luo, J. and Oubong, G. (2009), "A comparison of SIFT, PCA-SIFT and SURF", *Int. J. Image Process.*, **3**(4), 143-152.

Microsoft: Image Composite Editor (2008), Image Composite Editor - Microsoft Research, Microsoft, Redmond, USA. https://www.microsoft.com/en-us/research/product/computational-photography-applications/image-composite-editor/

Moore, M., Phares, B.M., Graybeal, B., Rolander, D. and Washer, G. (2001), "Reliability of Visual Inspection for Highway Bridges", Final Rep. No. FHWA-RD-01-020, Federal Highway Administration (FHWA).

Nie, L., Lin, C., Liao, K., Liu, S. and Zhao, Y. (2021), "Unsupervised deep image stitching: reconstructing stitched features to images", *IEEE Transact. Image Process.*, **30**, 6184-6197. https://doi.org/10.1109/TIP.2021.3092828

OpenCV (2011), OpenCV: Images stitching, OpenCV team, USA. https://docs.opencv.org/master/d1/d46/group__stitching.html

Rosten, E., Porter, R. and Drummond, T. (2010), "Faster and better: a machine learning approach to corner detection", *IEEE Transact*. *Pattern Anal*. *Mach*. *Intell*., **32**(1), 105-119. https://doi.org/10.1109/TPAMI.2008.275

Rousseeuw, P.J. (1984), "Least median of squares regression", *J*. *Am*. *Statist*. *Assoc*., **79**(388), 871-880.

Rublee, E., Rabaud, V., Konolige, K. and Bradski, G. (2011), "ORB: An efficient alternative to SIFT or SURF", *Proceedings of 2011 International Conference on Computer Vision*, Barcelona, Spain, November. https://doi.org/10.1109/ICCV.2011.6126544

Schönberger, J.L. and Frahm, J. (2016), "Structure-from-motion revisited", *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, Nevada, USA, June.

Spencer, B.F., Hoskere, V. and Narazaki, Y. (2019), "Advances in computer vision-based civil infrastructure inspection and monitoring", *Engineering*, **5**(2), 199-222. https://doi.org/10.1016/j.eng.2018.11.030

Strauss, A., Bien, J., Neuner, H., Harmening, C., Seywald, C., Österreicher, M., Voit, K., Pistone, E., Spyridis, P. and Bergmeister, K. (2020), "Sensing and monitoring in tunnels testing and monitoring methods for the assessment of tunnels", *Struct*. *Concrete*, **21**(4), 1356-1376. https://doi.org/10.1002/suco.201900444

Sun, Y., Zhao, L., Huang, S., Yan, L. and Dissanayake, G. (2014), "L2-SIFT: SIFT feature extraction and matching for large images in large-scale aerial photogrammetry", *ISPRS J*. *Photogram*. *Remote Sens*., **91**, 1-16. https://doi.org/10.1016/j.isprsjprs.2014.02.001

Sun, L., Shang, Z., Xia, Y., Bhowmick, S. and Nagarajaiah, S. (2020), "Review of bridge structural health monitoring aided by big data and artificial intelligence: from condition assessment to damage detection", *J*. *Struct*. *Eng*., **146**(5), 04020073. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002535

Szeliski, R. (2007), *Image Alignment And Stitching*: *A Tutorial*, Now Publishers Inc., Boston, MA, USA.

Torr, P.H.S. and Zisserman, A. (2000), "MLESAC: a new robust estimator with application to estimating image geometry", *Comput*. *Vis*. *Image Underst*., **78**(1), 138-156. https://doi.org/10.1006/cviu.1999.0832

VLFeat - Home (2008), VLFeat: An open and portable library of computer vision algorithms, Vedaldi A. and Fulkerson B. http://www.vlfeat.org

Wu, R.T. and Jahanshahi, M.R. (2020), "Data fusion approaches for structural health monitoring and system identification: past, present, and future", *Struct*. *Health Monitor*., **19**(2), 552-586. https://doi.org/10.1177/1475921718798769

Xiang, T., Xia, G. and Zhang, L. (2019), "Mini-unmanned aerial vehicle-based remote sensing: techniques, applications, and prospects", *IEEE Geosci*. *Remote Sens*. *Magaz*., **7**(3), 29-63. https://doi.org/10.1109/MGRS.2019.2918840

Ye, X.W., Dong, C.Z. and Liu, T. (2016), "Image-based structural dynamic displacement measurement using different multi-object tracking algorithms", *Smart Struct*. *Syst*., *Int*. *J*., **17**(6), 935-956. https://doi.org/10.12989/sss.2016.17.6.935

Yeum, C.M., Choi, J. and Dyke, S.J. (2017), "Autonomous image localization for visual inspection of civil infrastructure", *Smart Mater*. *Struct*., **26**(3), 035051. https://doi.org/10.1088/1361-665X/aa510e

Yuan, M., Lei, T., Liu, X. and Li, S. (2019), "Fast image stitching of unmanned aerial vehicle remote sensing image based on SURF algorithm", *Proceedings of the 11th International Conference on Digital Image Processing* (*ICDIP 2019*), Guangzhou, China, May.

Zhou, W., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004), "Image quality assessment: from error visibility to structural similarity", *IEEE Transact*. *Image Process*., **13**(4), 600-612. https://doi.org/10.1109/TIP.2003.819861

BS