Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 5804835, 15 pages
https://doi.org/10.1155/2020/5804835

Research Article

Dynamic Pavement Distress Image Stitching Based on Fine-Grained Feature Matching

Yuchuan Du, Zihang Weng, Chenglong Liu, and Difei Wu

The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Chenglong Liu; lcl_tongji@126.com

Received 7 November 2019; Revised 10 January 2020; Accepted 28 January 2020; Published 25 February 2020

Academic Editor: Rocío de Oña

Camera-based pavement distress detection plays an important role in pavement maintenance. Duplicate collections for the same distress and multiple overlaps of defects are both practical problems that greatly affect the detection results. In this paper, we propose a fine-grained feature-matching and image-stitching method for pavement distress detection to eliminate duplications and visually demonstrates local pavement distress. The original images are processed through a hierarchical structure, including rough data filtering, feature matching, and image stitching. The original data are firstly filtered based on the global position system (GPS) information, which can avoid full-dataset comparison and improve the calculating efficiency. A scale-invariant feature transform is introduced for feature matching based on the extracted key regions using spectral saliency mapping and bounding boxes. Two parameters: the mean Euclidean distance (MEuD) and the matching rate (MCR) are constructed to identify the duplication between two images. A support vector machine is then applied to determine the threshold of MEuD and MCR. This paper further discusses the correlation between the sampling frequency and the number of detection vehicles. The method provided can effectively solve the problem of duplications in pavement distress detection and enhances the feasibility of multivehicle pavement distress detection based on images.

1. Introduction

Pavement condition measurements are essential for maintenance decisions [1]. Pavement distress detection has traditionally been a highly laborious and time-consuming task [2]. Currently, the most commonly used detection vehicle is a specially modified car with precise but delicate instruments, and the process of detection is time-consuming, expensive, and inefficient [3]. With the increasing demand for real-time pavement maintenance, detection methods based on lightweight sensors and rough-set data mining are becoming popular. Automated pavement detections using cameras [4], lasers, and ultrasonic sensors [5] are widely used as replacements to manual work, which significantly improves the efficiency and lowers the cost [6]. Among them, the camera is the priority choice in pavement detection because of not only its low cost and intuitive data but also its lightweight and detachable features that satisfy the requirements of multiple-vehicle detection and rough-set data collection [7]. Therefore, pavement condition recognition based on video image has become a central issue [8].

With the development of deep learning and computer vision technology, image-processing algorithms lead to good performance in automatic identification of pavement distress [9]. Different kinds of pavement defects such as cracks, potholes, and nets can be identified with relatively high accuracy [10, 11]; thus, the image-based detection has been proved to be a reliable and efficient method [12]. Different approaches were employed for image analysis. The Sobel edge detector recognizes edges in an image by smoothing the image before computing the derivatives in the perpendicular direction to the derivative [13]. The Canny method is a multistep algorithm that can detect edges and concurrently suppress noise in an image [14]. The semantic texton forests (STF) algorithm is also used as a supervised classifier on a calibrated region of interest (ROI) in the detection of multiple pavement defects [15]. However, the results of convolution neural networks (CNNs) are significantly better than the aforementioned algorithms in image-based detection [16].

CNNs have become the most popular algorithm and have been constantly improved to better fit the distress detection [17]. CNNs have the advantage of performing feature extraction and predicting crack/noncrack conditions in an integrated and fully automated manner with good prediction performance and a classification accuracy rate (CAR) of 92.08% [18]. Gopalakrishnan et al. employed a deep CNN with transfer learning for pavement distress detection [19]. Jenkins et al. proposed a deep fully CNN to perform pixel-wise classification of surface cracks on roads and pavement images with 92.46% precision [20].

Besides, 3D laser-illuminated camera is also used to detect pavement deterioration. Li et al. applied a fully automated algorithm for segmenting and enhancing pavement crack based on 3D pavement images [21]. The depth information collected by 3D techniques helps to perform better in analyzing cracks, textures, rutting, etc.

However, there are still some practical problems remain unsolved during road detection using a 2D or 3D camera. High-acquisition frequencies are used to reduce the number of missing defects to the minimum, and at the same time, multiple overlaps of defects take place. Besides, it is always the case that the low vehicle speed or traffic congestion causes image duplications. Such duplication can greatly affect the statistical reliability of pavement health assessment and the calculation of relative indices like the pavement condition index (PCI) [22]. Moreover, length and area are used as units of summarization to better describe a crack and this problem is more of a concern.

For the comprehensive inspection cars, wheel encoders are adopted to avoid overlaps. However, this solution is not only expensive but also not suitable for our lightweight equipment that can install and work quickly on any car. Therefore, two existing problems are focused on in this paper as follows:

(1) A defect in different images might be misidentified as different ones due to a location and pixel-size discrepancy in different images, as shown in Figure 1(a).

(2) A longitudinal crack crossing different frames (Figure 1(b)) might be recognized as different cracks instead of one long crack.

To solve the problems mentioned above, we propose a pavement distress stitching method to preprocess detected data. On the one hand, stitching is a technology-neutral pattern to use in locating distress over multiple passes, especially over time. It eliminates duplications and orderly sorts the statistical summarizations such as number, length, and area. On the other hand, adjacent defects in consecutive images can be stitched to form a whole lane-level picture of pavement distress. Such panoramic pictures are conducive to manual verification while providing visualizations of the pavement condition.

One of the most crucial parts of image stitching is the feature-matching algorithm, which can be divided into three categories: global feature-based matching algorithms, local feature-based matching algorithms, and deep learning algorithms. Global feature-based matching algorithms such as the histogram of oriented gradient (HOG), local binary pattern (LBP), and Haar-like features performed well in human detection [23, 24]. Compared with global feature-based matching algorithms, local feature-based matching algorithms are more stable. Scale-invariant feature transform (SIFT) was first proposed by Lowe as a local feature description algorithm based on the analysis of existing invariance-based feature detection methods [25]. SIFT has good stability and invariance, but it imposes a large computational burden [26]. Speeded-up robust features (SURF) is the replacement to SIFT, which has lower computation cost for real-time systems at a tradeoff of poor relative performance [27]. The oriented FAST and rotated BRIEF (ORB) algorithm is rotation invariant and resistant to noise, and it performs almost as well as SIFT while being two orders of magnitude faster [28]. In the field of deep learning, deep matching (DM) is one of the most popular methods for establishing quasi-dense correspondences between images [29]. DM relies on a hierarchical, multilayer, correlational architecture designed for matching images that have high information dimensions and need sophisticated calculation. Moreover, if the feature matrix correlation parameter threshold control is too strict, the angular resolution will consequently decline. Therefore, SIFT is adopted in this paper because of its stability.

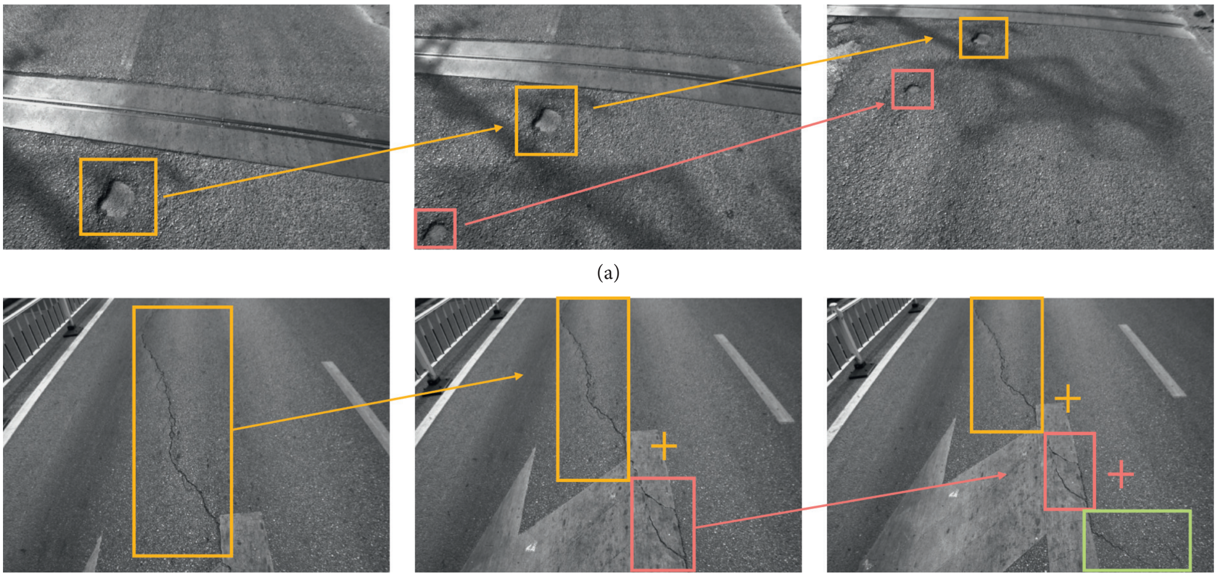
Image stitching is one of the main applications of SIFT. Lowe proposed an invariant feature-based approach to fully automatic panoramic image stitching [30], while Xiaoyan et al. created a large field of view for robot control and movement using dynamic image stitching when there was a moving object in the environment [31]. Qiu et al. proposed an image-stitching algorithm based on aggregated star groups to obtain a complete star map [32]. This paper applies the image-stitching method in pavement detection to solve engineering application problems.

Based on the above problems, we present a pavement distress image stitching method based on a feature-matching algorithm. Since the background of the pavement is monotonous and the algorithm can falsely match the features of the asphalt pavement, we propose the use of the spectral saliency mapping (SSM) method along with a pavement distress bounding box to extract information from dense regions. The scale-invariant features extracted from the key region serves as the stitching points between two images.

The remainder of this paper is organized as follows. In Section 2, we present the data processing methods. In Sections 3, 4, and 5, we describe the framework of the proposed approach where the feature matching, key region extraction, and image stitching are introduced, respectively. In Section 6, we discuss the correlation between the sampling frequency and the number of detection vehicles. In Section 7, we offer the conclusions of this study.

1.1. Data.

In our experiment, an integrated detection system was used to collect pavement images. An industrial camera



(a)

(b)

Figure 1: Typical problems of pavement distress images. (a) The same pothole with different sizes and locations in different pictures. (b) A longitudinal crack crossing three frames.

was fixed on the back of the vehicle, which faced obliquely downward. The vehicle also equipped with a GPS unit, which allows the images to match the corresponding locations on the road. Full videos were stored in a vehicle-mounted terminal while clipped images were uploaded at a frequency of 2 Hz.

Several typical pavement distress defects on the urban road in Shanghai are considered in this paper, including cracks, patched cracks, potholes, patched potholes, nets, patched nets, and manhole covers (Figure 2). A 13.2 km road section on Caoan Road in Shanghai was chosen for experiments and validation, as shown in Figure 3. The algorithm processed more than 6000 images and generated bounding boxes when the defects are recognized. At the same time, the results were artificially calibrated to guarantee accuracy.

1.2. Methodology. Figure 4 illustrates the flow chart of the proposed hierarchical framework for image processing, including rough data filtering, feature matching, and image stitching. The original images are firstly filtered according to the GPS information, which can exclude most of the irrelevant images. Through choosing the images that have the most overlap, a feature matching method is applied to extract the SIFT features in the key region using SSM and bounding boxes. After the feature-matching process, two or more images are stitched according to the features and the fitted perspective matrix.

2. Rough Data Filtering Using GPS to Reduce Computational Cost

The purpose of the preprocessing is to reduce the computation cost before further analysis. The basic idea is to select the images based on the GPS information because the location of the potential matched images must be close. GPS, though considered to be not accurate enough, excludes a large number of images that are geographically too far apart to be matched, thus serving as a rough data filtering to reduce the calculating amount.

The GPS module recorded the real-time locations during detection and then linked to images according to the timestamps [33]. The GPS information makes it easier to manage the statistical data at the level of road segment. Due to the instability of GPS, images within 10 meters (P_n) are selected as candidates for matching to make sure that no targeted picture is omitted. The chances that two defects within 10 meters are too similar to differentiate by a human or algorithm are negligible. If it happens, the number of the candidate images would be more than the detection times, and in this situation, the images need to be checked by a human. The Haversine equation [34] was adopted to calculate the distance between two points using their longitudes and latitudes, as formulated in the following equation:

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right), \tag{1}$$

where φ_1/φ_2 and λ_1/λ_2 are the latitude and longitude of point 1 and point 2, respectively and d is the distance between them. The same defects among P_n were searched and labeled by artificial identification to build the ground truth.

In most cases, the same defects can be found within 10 meters unless there exists a GPS deviation. Therefore, when P_n was an empty set, the GPSs of the retrieved images (P_x) were examined and the distances and time-lags from their adjacent and matched images (P_k) were calculated. Figure 5 describes the method of dealing with abnormal data.

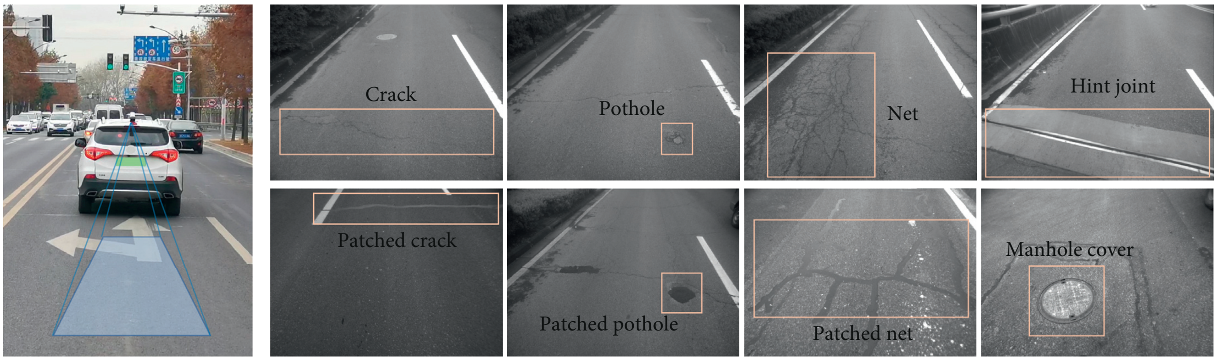


Figure 2: The targeted pavement distress.

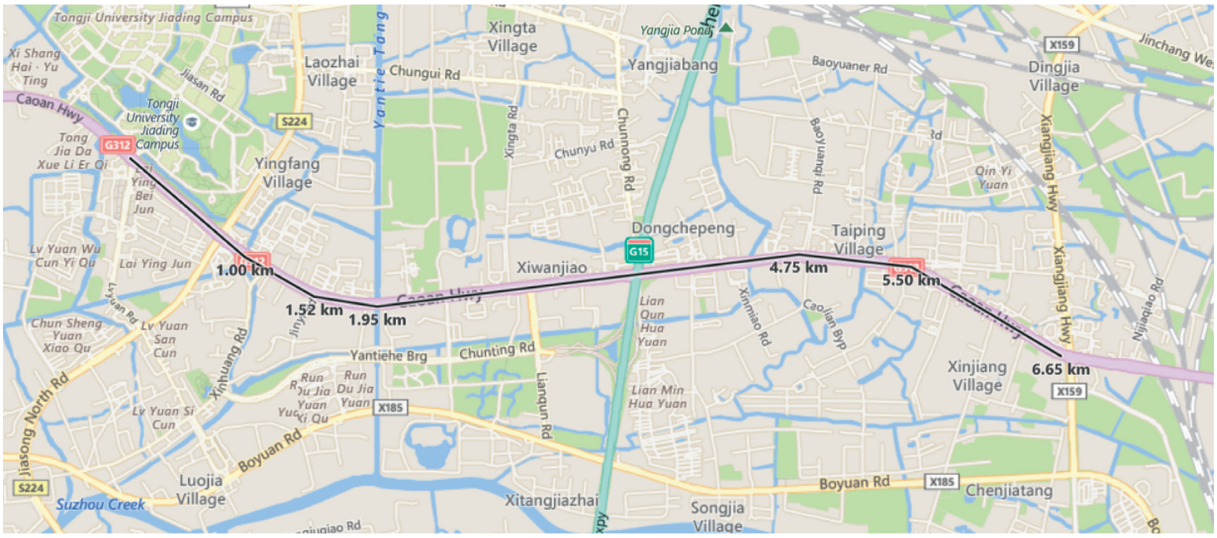


Figure 3: The experimental road section on Caoan Road in Shanghai.

The collection speed was used as a discriminative index. When the calculated value was more than 1.5 times the true value as formulated in equation (2), the location was considered as being in error and was redefined as the time-weighted average of P_k.

$$\frac{l\left(\mathrm{GPS}_X - \mathrm{GPS}_K\right)}{t_x - t_k} > 1.5v, \qquad (2)$$

where v is the true value of the velocity, l is the distance between two locations, GPS_X is the GPS location of P_x and t_x is the timestamp of P_x, and GPS_K is the GPS location of P_k and t_k is the timestamp of P_k.

3. SIFT Feature Matching

SIFT features are located at the scale-space maxima/minima of the differences between Gaussian functions, which keep the rotation, scale, or illumination invariant. They are robust in terms of vision changes, affine changes, and noise [35]. SIFT feature matching mainly includes the following three steps.

3.1. Feature Detection in Scale Space. This step involves searching for scale-invariant features from the multiscale images in scale space. The scale space is defined as the following convolution operation:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) = \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/\left(2\sigma^2\right)} * I(x, y), \qquad (3)$$

where σ is the scale-space factor, G is driven from a variable-scale Gaussian distribution, and I is the input image. The difference of Gaussian (DOG) function can be further established from the difference of the nearby scales with a constant multiplicative factor k as follows:

$$\begin{aligned} D(x, y, \sigma) &= \left(G(x, y, k\sigma) - G(x, y, \sigma)\right) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned} \qquad (4)$$

3.2. Feature Localization. The candidate feature points in the scale space extracted from the images are further refined to perform a detailed fit to the nearby data to determine the locations, scales, and ratios of principal curvatures. This information allows points to be rejected that have low contrast or are poorly localized along an edge. The DOG

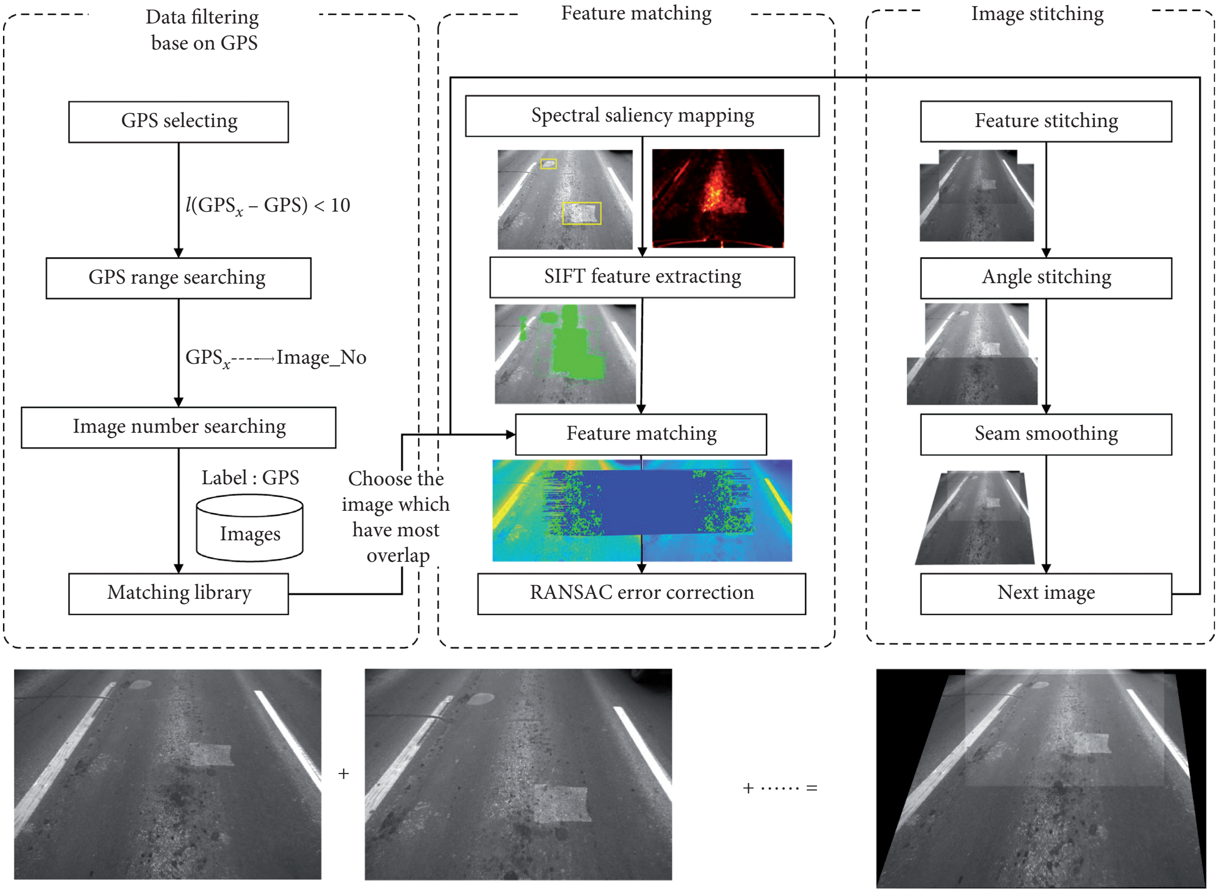


Figure 4: The pavement distress image-stitching pipeline.

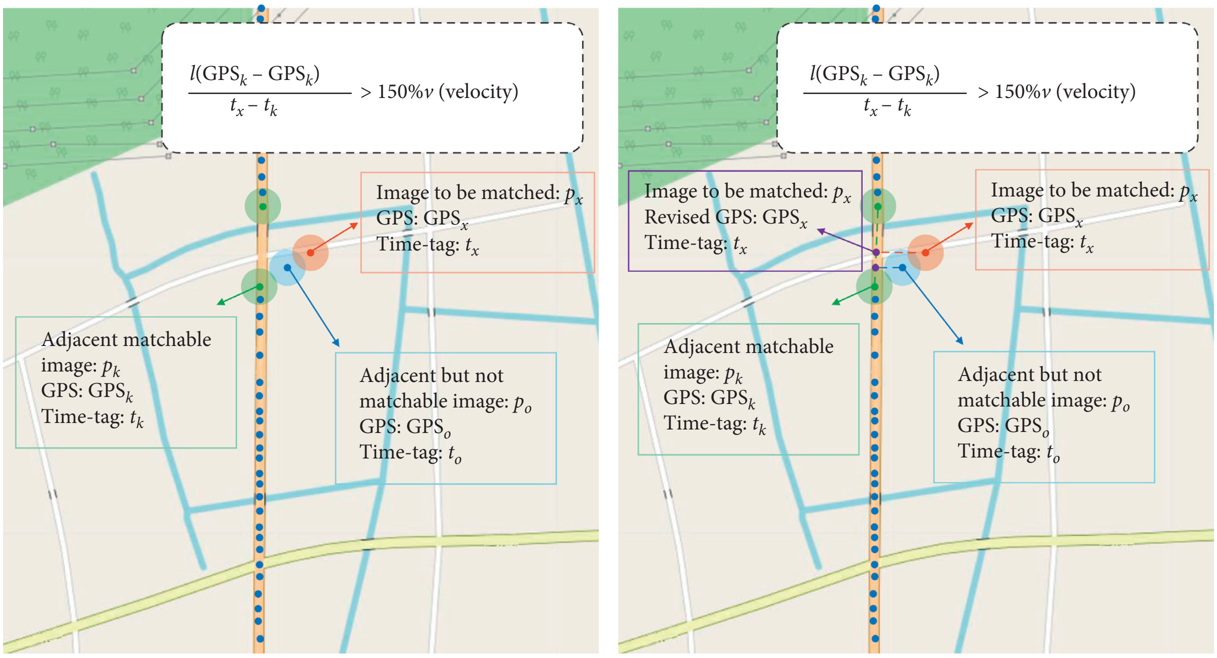

$\frac{l(\mathrm{GPS}_k - \mathrm{GPS}_k)}{t_x - t_k} > 150\% v$ (velocity)
Image to be matched: p_x
GPS: GPS_x
Time-tag: t_x
Adjacent matchable image: p_k
GPS: GPS_k
Time-tag: t_k
Adjacent but not matchable image: p_o
GPS: GPS_o
Time-tag: t_o
$\frac{l(\mathrm{GPS}_k - \mathrm{GPS}_k)}{t_x - t_k} > 150\% v$ (velocity)
Image to be matched: p_x
Revised GPS: GPS_x
Time-tag: t_x
Image to be matched: p_x
GPS: GPS_x
Time-tag: t_x
Adjacent matchable image: p_k
GPS: GPS_k
Time-tag: t_k
Adjacent but not matchable image: p_o
GPS: GPS_o
Time-tag: t_o

Figure 5: The method of dealing with abnormal data.

function at the candidate feature points $\widehat{X}$ is adopted here to discard unstable features with low contrast in the underwater images:

$$D(\widehat{X}) = D + 0.5\frac{\partial D^T}{\partial X}\widehat{X}, \quad (5)$$

where $\widehat{X}$ denotes the offset from the location of the extremum, and all extrema with a value of $D(\widehat{X})$ less than 0.03 are discarded. In this paper, the threshold of the principal curvature is set to 0.6 considering that the edge-detect results of the pavement distress are not obvious.

3.3. Orientation Assignment and Feature Description. The main direction and auxiliary direction of the key points are given according to the gradient direction histogram of the key feature points, where the resultant matrix of $2 * 2 * 8$ dimensions is mathematically described. SIFT features are calculated and the matching features are shown in Figure 6.

In Figure 6(a), the frame indicates the gradient direction of an extracted feature point. In Figure 6(b), a line indicates a link between two matched features. The more links exist, the greater the probability that the images share the same feature will be. However, the features of both pavement distress and normal pavement are extracted, as shown in Figure 7(a). Due to the similarity of the pavement structure and pavement markings therein, matching errors can easily arise. Therefore, a bounding box is needed to extract features in the designated area, which can greatly improve the matching accuracy and pertinence as shown in Figure 7(b).

Meanwhile, the random sample consensus (RANSAC) method was used once the feature matching is finished. RANSAC was firstly proposed by Fichler and Bolles as a robust estimation procedure that uses a minimal set of randomly sampled correspondences to estimate image transformation parameters and screens correct data [36]. In general, different perspectives can be transformed by a perspective matrix, and RANSAC was used to find parameters with the maximum likelihood in image matching. Theoretically, all the matched feature points should satisfy the matrix transformation. However, there will always be some errors, and RANSAC rejects abnormal values. The SIFT-matched results used in this paper were processed with RANSAC, which can effectively improve the reliability and robustness of feature points.

The mean Euclidean distance (MEuD) between two feature points and the matching rate (MCR) were used as indices to evaluate the matching degree. The Euclidean distance indicates the matching degree, and the matching rate illustrates the proportion of correctly matched points. The smaller the Euclidean distance is, the better two feature points match will be. When two defects are of the same type, there will be more matched features than those are not the same. However, the matched SIFT features do not fully indicate whether two objects are the same object. The shortest Euclidean distance can only illustrate the best match of the corresponding SIFT feature points on the other image. Hence, it is difficult to judge whether two matched features are the same defect with complete certainty using a numerical threshold or a threshold derived from the root mean square of the distance. In this paper, the MEuD and the MCR of the matched feature points were used as indicators for evaluating image similarity. The MEuD is defined as in the following formula:

$$\text{MEuD}(S,T) = \frac{1}{m}\sum_{i=1}^{m}\left(\min_{1\le j\le m}\sqrt{\sum_{k=1}^{128}\left(s_{ik}-t_{jk}\right)^2}\right), \quad (6)$$

where $\text{MEuD}(S,T)$ is the root mean square distance between two images (S and T), m is the number of matched SIFT feature points, j is the sequence number of a feature, and s_{ik}/t_{jk} is the SIFT matrix. Because the root mean square distance is affected by the size of the images, the MCR was also used as a similarity evaluation index. The MCR is defined as follows:

$$\text{MCR} = \frac{N_m}{N}, \quad (7)$$

where N represents the number of points and N_m is the number of matched points. The MCR indicates the proportion of all retrieved matched SIFT features. Cross-validation was used to calculate the matching accuracy of the SIFT features.

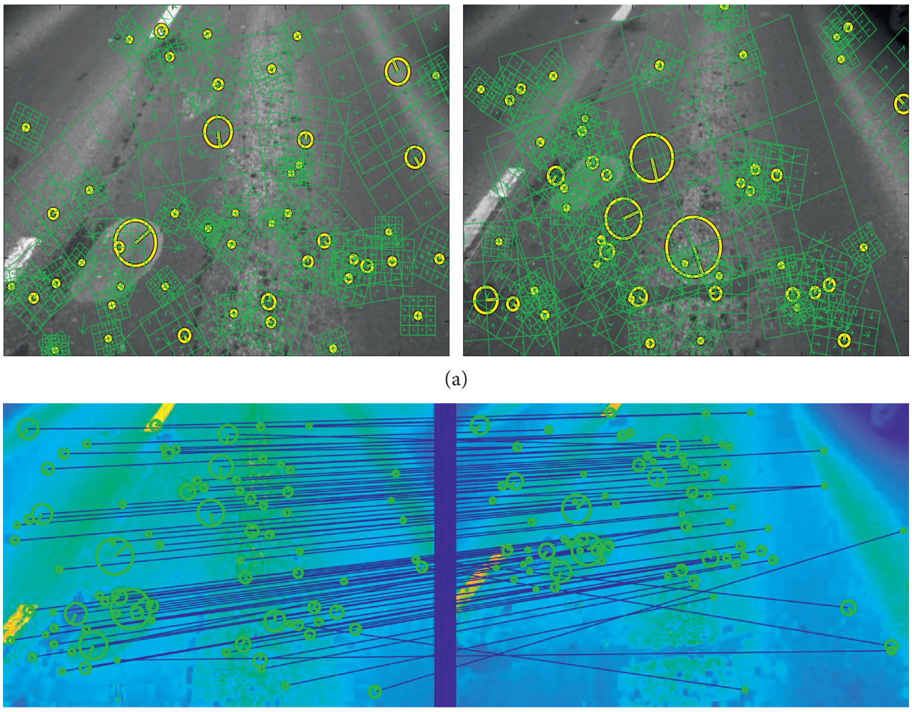
Table 1 shows the SIFT matching results of several selected images in a 10-m-long test section. Five of them are recognized as the same pothole by the algorithm.

Although SIFT is robust to the shooting angle, the MCR of the images with a large distance is only 37.39%, while the MCR of the images with similar angles is as high as 85.50%. The MEuD of the images is relatively stable, which reaches 10^4 orders of magnitude. As for the different types of defects, the MEuD does not exist and MCR equals zero because no matching features could be found.

A matching test of two hundred pairs of images was performed on the sample library to determine the SIFT-based image matching threshold. A support vector machine (SVM) was used to estimate the tangential plane to determine the model threshold. The matching accuracy, as determined with the five-fold cross-check method, of the SIFT features is 81.4%. Figure 8 shows the results of the binary classification based on SVM, in which the dots represent the good matching result, while the crosses represent the incorrect matching result. The SIFT model is more inclined to identify a mismatch as a correct match because SIFT has a certain degree of angular robustness. Unfortunately, this can easily cause errors due to the effect of shooting angles. As two different defects, which are highly similar, are less likely to be present in the same location, the matching accuracy was as high as 92% in the sample set test.

4. Key Region Extraction

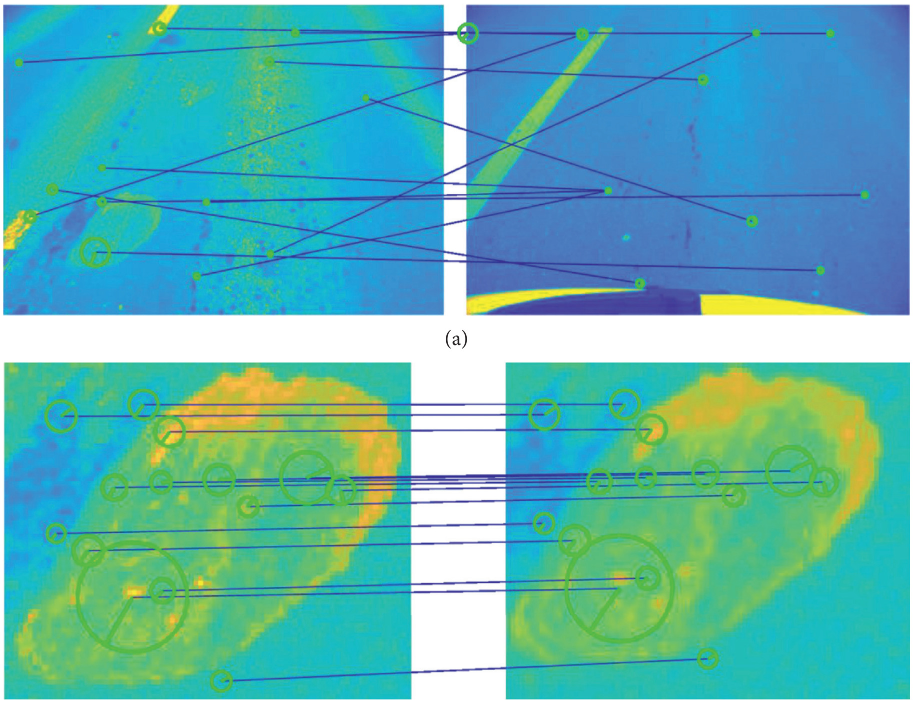
The monotonicity of a pavement results in matching errors of the SIFT features, as shown in Figure 9. To this end, we propose SSM along with bounding boxes generated by the



(a)

(b)

Figure 6: SIFT features matching. (a) The gradient direction of extracted features of two images. (b) Links between matched features of two images.

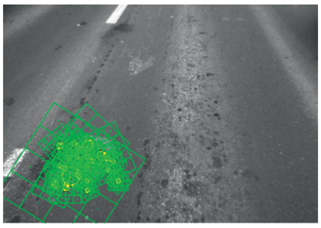
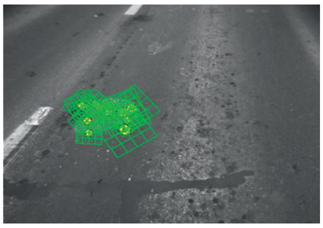
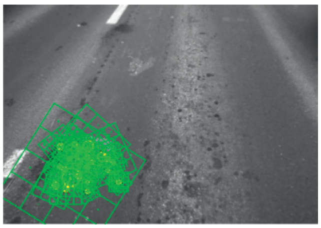
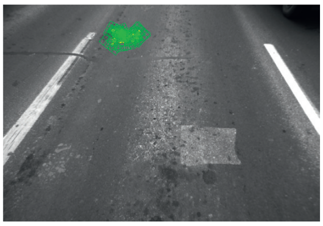
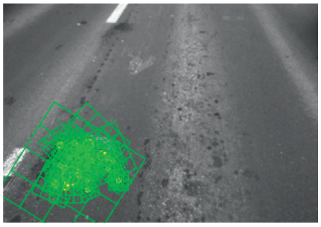
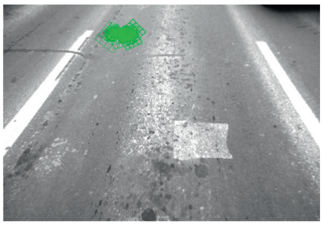
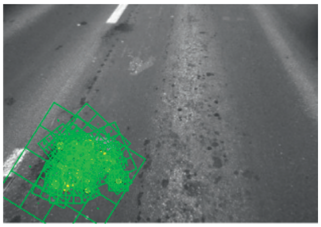
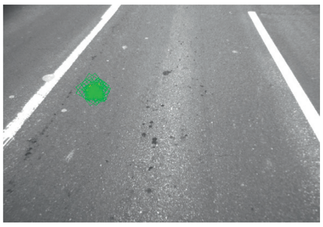
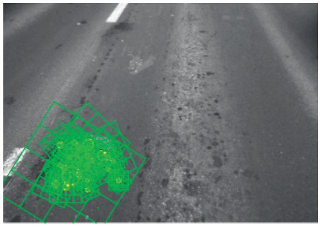
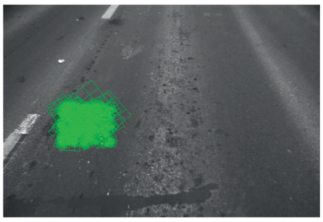


(a)

(b)

Figure 7: (a) The mismatched features. (b) SIFT feature matching with bounding box.

Table 1: SIFT matching results of the example images filtered by the specified GPS.

Image S	Image T	Distance (m)	MEuD	MCR (%)
		2.19	1.52×10^4	59.58
		2.19	5.16×10^4	37.49
		0.85	6.13×10^4	50.42
		6.44	NaN	0
		0.59	2.22×10^4	85.50

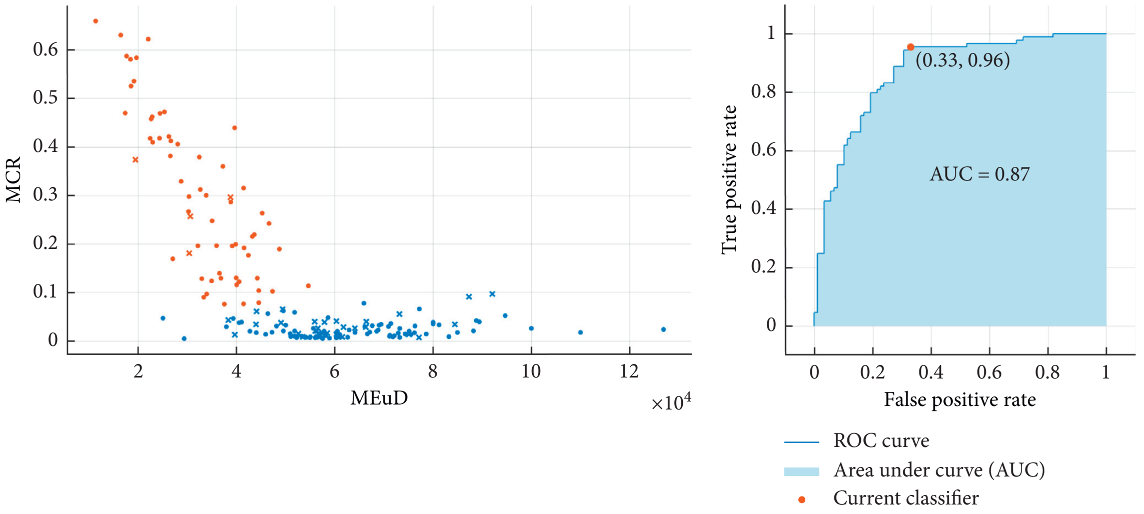


Figure 8: The results of the binary classification based on SVM.

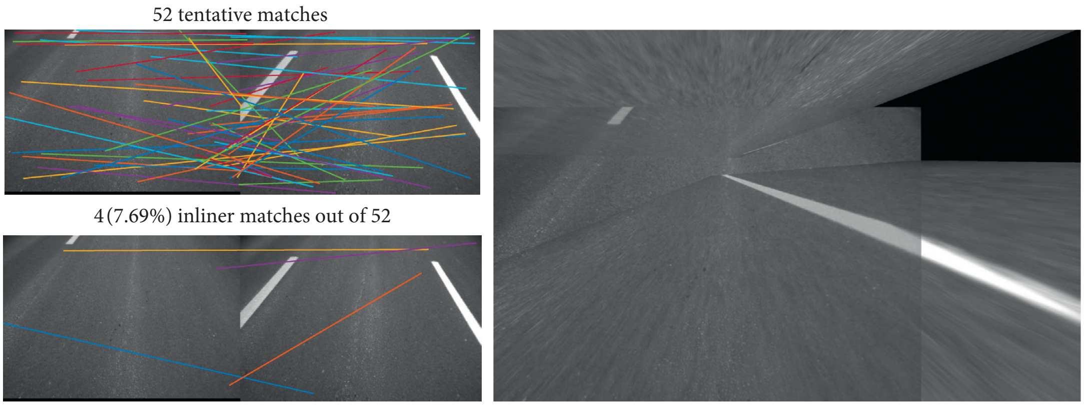


Figure 9: The bad matching and stitching results due to the monotonicity of a pavement.

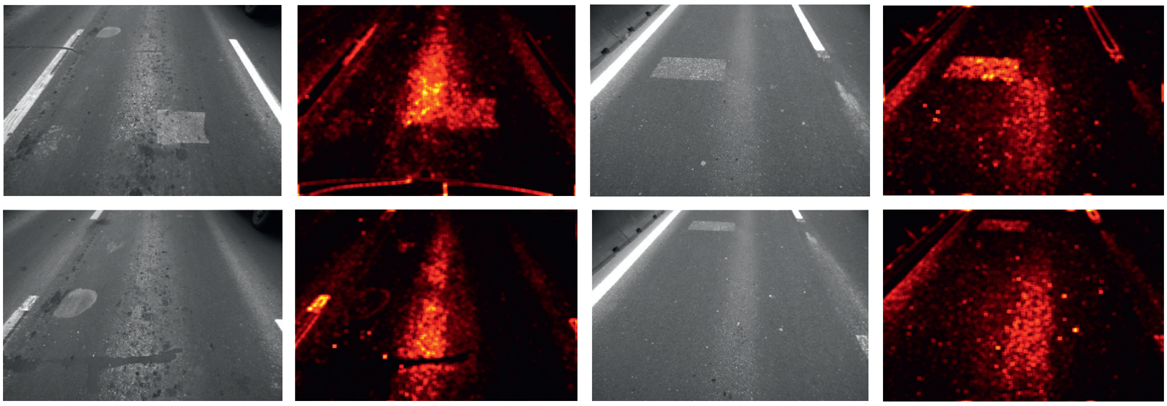


Figure 10: The key region extracted using SSM.

algorithm to extract the key regions to find prominent SIFT features.

SSM is a simulation of human visual attention characteristics, which can capture significant changes in an image. It is the dynamic visual attention that makes it easier for a human being to find important information in an image at first glance, instead of searching the elements one-by-one. From the perspective of information theory, the information processed by human beings is mainly divided into background information and changing information, the latter to which human vision is more sensitive. Although image incision technology and semantic segmentation can also segment background and subject information, they can only target specific objects and require a large amount of model training. Moreover, these methods will destroy the overall characteristics of an image, and it is difficult to reflect the overall characteristics of real human vision.

Xiaodi and Liqing found through a large amount of data analyses that the average log-spectrum of input images is positively correlated with the log frequency [37]. The spectral residual of an image in the spectral domain is extracted by subtracting the average log amplitude spectrum from the actual log amplitude spectrum of the image. In this paper, an FFT-based visual saliency model was used to extract the feature regions of the pavement, as shown in the following equation:

$$S(x) = g(x) * F^{-1}[\exp(L(f) - A(f) + P(f))]^2, \qquad (8)$$

where $S(x)$ represents the SSM of graph x, $g(x)$ is a Gaussian filter used to smooth the SSM graph, F^{-1} represents the inverse Fourier transform, $L(f)$ is the log vibration spectrum of the image, $A(f)$ represents the average log vibration spectrum, and $P(f)$ represents the phase spectrum of the image. Figure 10 shows the key region extracted using SSM.

According to Figure 10, the SSM method has a certain sensitivity to pavement distress, especially the patched distress, and the sensitivity is relatively stable, regardless of the location in an image. However, this method is not sensitive to potholes or cracks. Therefore, SSM was combined with a bounding box to form key regions. After selecting the key regions, SIFT feature extraction was performed on the region locations, and the SIFT factor was calculated in the selected region. Each image was rescaled to ensure that the directions were consistent. A K-dimension tree (KD Tree) was established, and the k-nearest-neighbors (KNN) algorithm was used to find the KNN for each feature, where K was set to 2. The validity needed to be verified when the K neighboring values were found. The valid verification threshold was 0.6, as is shown in the following inequality (9):

$$\frac{1 - \mathrm{NN}}{2 - \mathrm{NN}} < 0.6, \qquad (9)$$

where NN represents the nearest-neighbor.

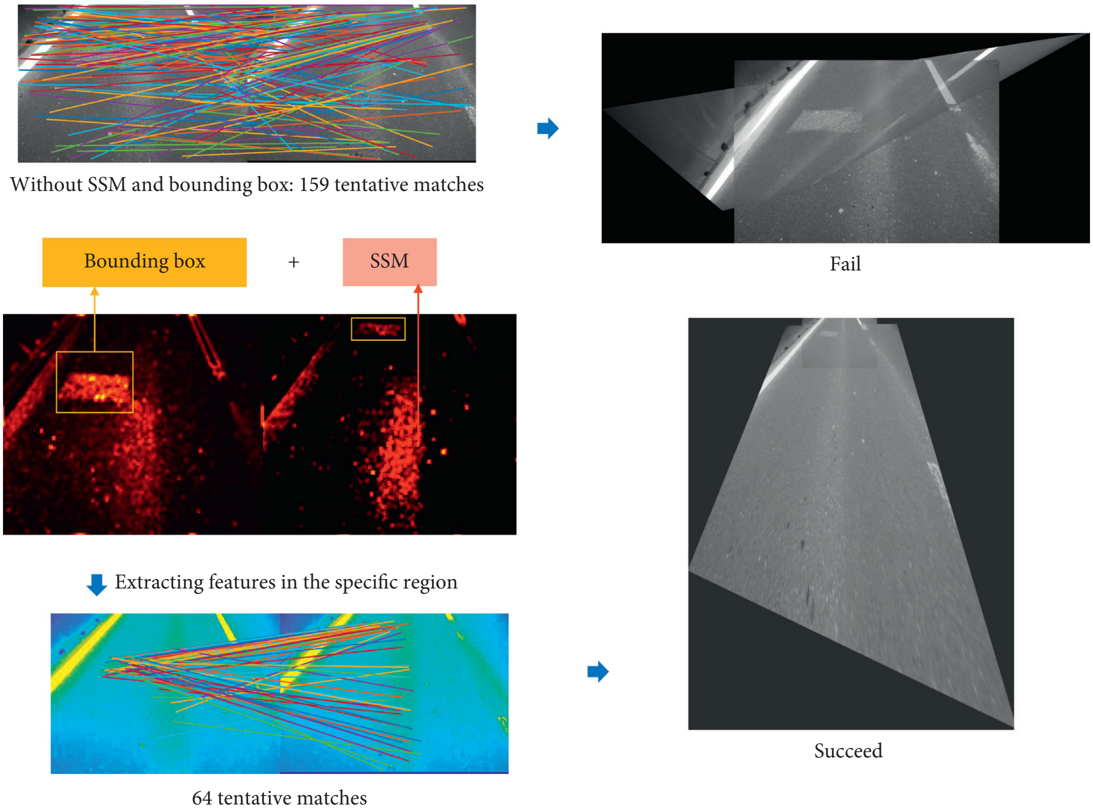

Without SSM and bounding box: 159 tentative matches
Fail
Bounding box
+
SSM
Extracting features in the specific region
64 tentative matches
Succeed

Figure 11: Feature extraction and image stitching using SSM and a bounding box.

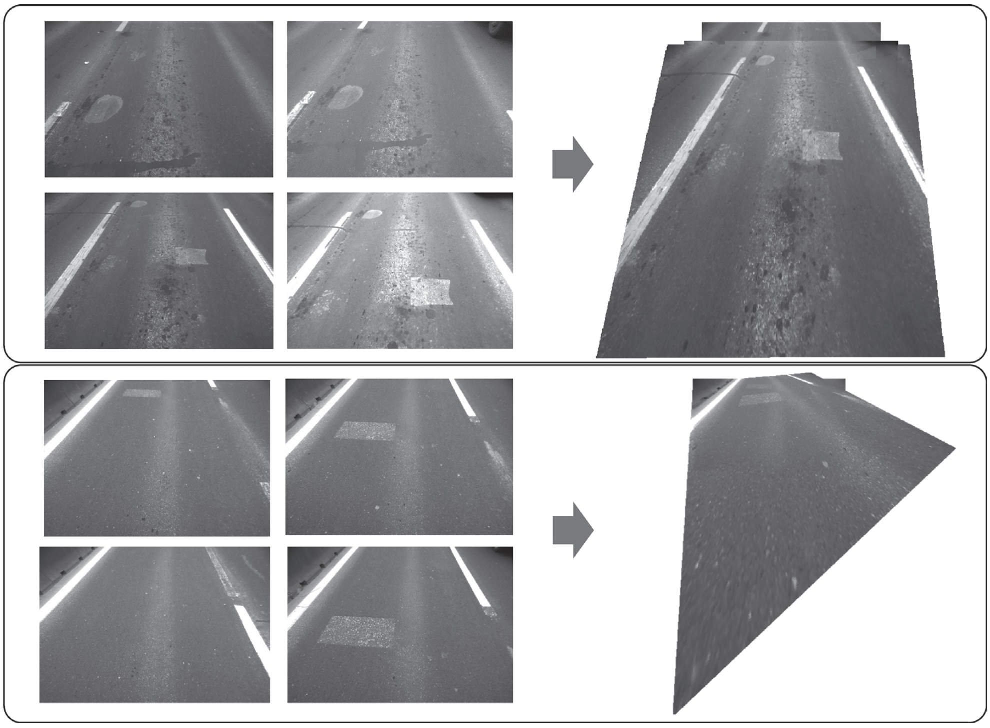


Figure 12: The stitching results.

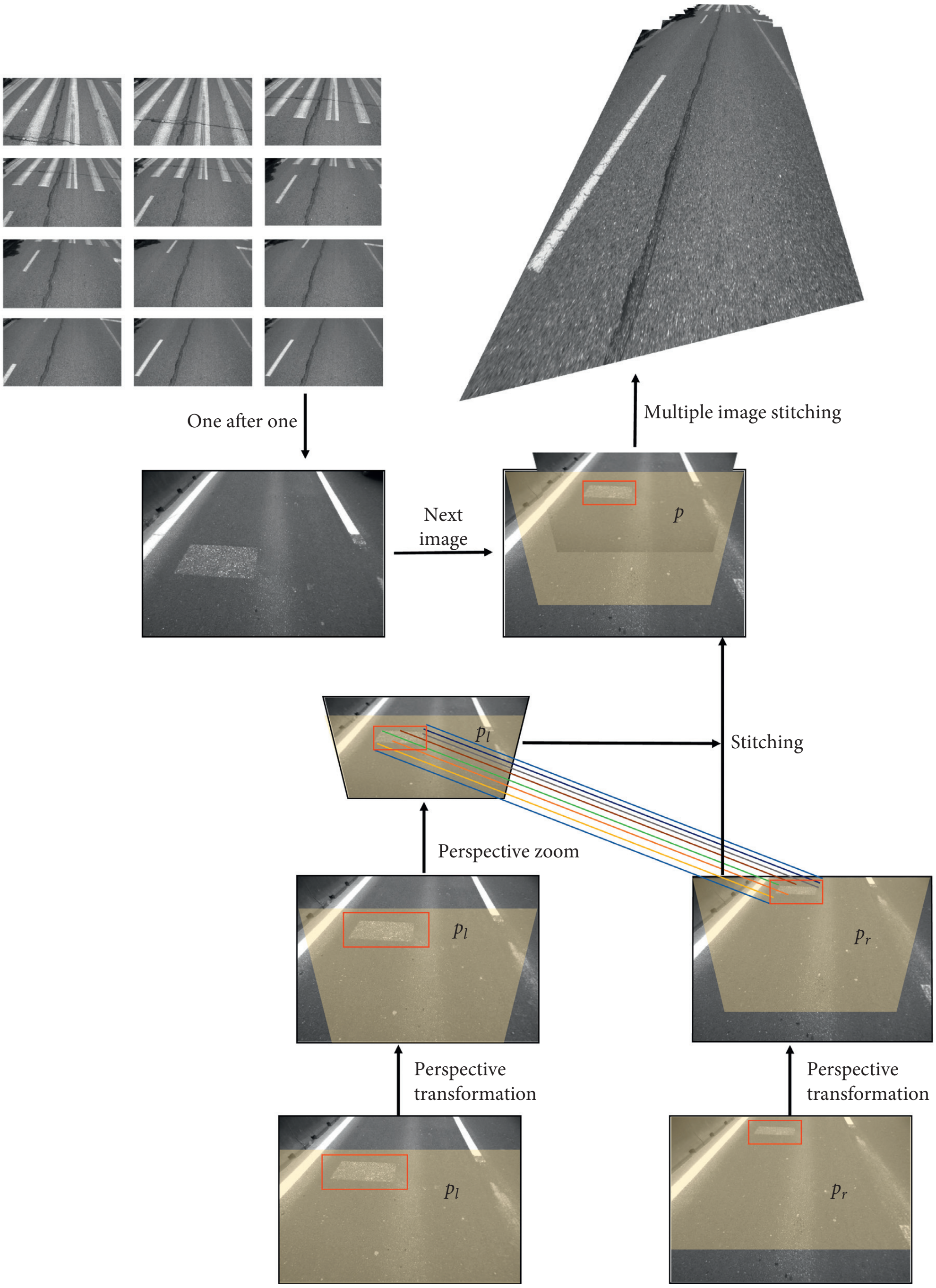


Figure 13: Flow chart of the image stitching algorithm.

Figure 11 shows the effect of the feature region on the results. When a feature region is not adopted, a large number of matching points exist in the normal pavement and more mismatches are caused due to the consistency of the pavement. However, when the SSM combined with the bounding box is applied, the matching accuracy improved.

In addition to the SSM method, the bounding boxes are generated to locate the region of interest by the object detection algorithm named "you only look once version 3 (YOLOv3) [38]." YOLO is one of the real-time deep CNN methods that aim at detecting objects and is widely applied in traffic management. YOLO reasons globally about the image when making predictions and learns generalizable representations of objects [39]. And it has been proved that YOLO performs well among other existing models, such as SSD or R-CNN in pavement defects recognition [40]. Moreover, YOLOv3 performs best especially in small object detection among the four versions of YOLO [41]. The precision of the algorithm was 0.7869 with 10,000 pavement images for training and 3,000 images for testing. Additionally, although YOLO consumes a lot of computational

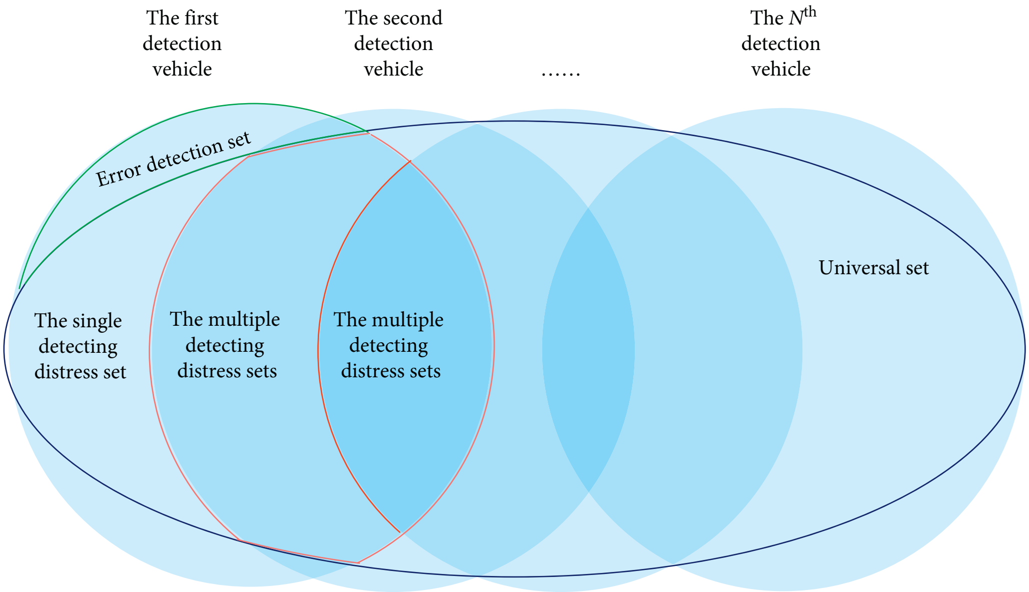

The first detection vehicle
The second detection vehicle
……
The N^{th} detection vehicle
Error detection set
Universal set
The single detecting distress set
The multiple detecting distress sets
The multiple detecting distress sets

Figure 14: The fusion result of multiple detection.

power when training the model, not much computational power is needed for prediction.

5. Image Stitching

After matching the SIFT features in the key region using the SSM and bounding box, two candidate images that had the most matched features were stitched according to the features and the fitted perspective matrix. After that, the next image was stitched on to the base of the previously stitched images. The stitching results are displayed in Figure 12.

The angle and size of the stitched portion can change within the perspective matrix, so the weighted average fusion approach was used, as shown in the following equation:

$$p = \frac{d_l}{d_l + d_r} p_l + \frac{d_r}{d_l + d_r} p_r, \quad (10)$$

where p represents the synthesized pixel coordinates and d_l and d_r represent the distances of p_l and p_r, respectively, from the left and right edges of the image.

According to the number of feature points matched in the image set, the images were preferentially stitched. The algorithm stopped when the ratio of inliers was less than 50%. A flow chart of the image stitching algorithm is shown in Figure 13. The current algorithm can process up to 12 images, and the result is shown in Figure 13. The perspective field of view exists in the original image, which makes it a challenge to stitch more images. The distortion becomes serious as the stitched images increase, and further study will be carried on to solve this problem.

6. Calculating the Minimum Number of Sampling Vehicles

It is difficult to obtain all the pavement distress characteristics with a single detection car. For one thing, it is always the case that some pavement defects are missed in the course of detection, in which the video sampling rate and vehicle speed are considered. For another, the algorithm could not completely identify the pavement defects, and the misdetections exist. Therefore, it is necessary to have multiple detection vehicles to superimpose and match the data to show the overall condition of the pavement. The minimum number of required vehicles is discussed here using probability theory as shown in Figure 14.

The precision p_t of the pavement detection algorithm used in this paper is 0.7869, which is the probability that we can correctly detect pavement distress. The parameter p_c represents the probability of collecting an image at a certain position on the pavement via detection with a single vehicle, which is the function $p_c(v, f)$ that is related to the traveling speed v and the camera sampling frequency f. When v is high and f is low, the detection vehicle could possibly miss some information at certain positions on the pavement, so the resultant value of p_c is low. Conversely, when v is low and f is high, the p_c value is high, but it can easily cause duplications. The number of detected pavement defects by the algorithm in one detection by a single vehicle is shown in the following expression (11):

$$M \times p_c \times p_t, \quad (11)$$

where M represents the actual number of pavement defects. Considering that v of different vehicles are basically the same in the same time period, and f are also the same, p is assumed to be a fixed value. In view of this, the pavement defects detected by each vehicle are consistent with the same distribution. Whether the pavement defects x can be detected conforms to the n-multiple Bernoulli trials $x \sim B(N, p_c \times p_t)$, as shown in the following equation:

$$P(x = k) = C_n^k p^k (1 - p)^{n-k}. \quad (12)$$

In order to meet the need that more than 95% of the defects are detected by multiple vehicles, the corresponding inequality is shown in the following inequality (13):

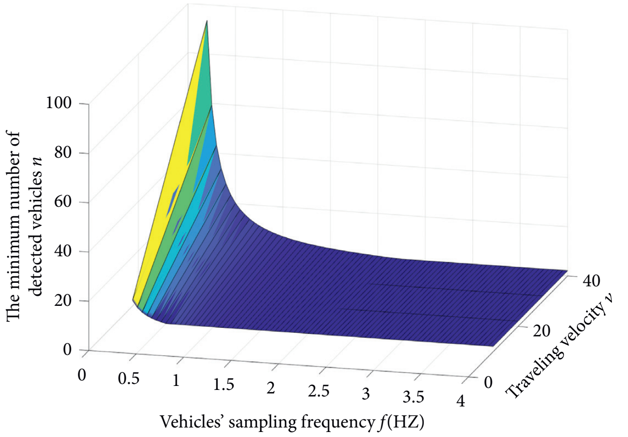


Figure 15: The relationship between speed, sampling frequency, and the minimum number of vehicles.

$$P(x \geq 1) = 1 - P(x = 0) = 1 - \left(1 - p_c \times p_t\right)^n \geq 0.95. \tag{13}$$

A pixel in a camera image is $i * j$, and the range that the camera can detect is $(\widehat{u}, \widehat{v})$. The matrix transformation relationship between the pixels and world coordinates is as shown in the following equation:

$$s \begin{bmatrix} \widehat{u} \\ \widehat{v} \\ 1 \end{bmatrix} == \mathbf{A} \left[r_1 \ \ r_2 \ \ t \right] \begin{bmatrix} i \\ j \\ 1 \end{bmatrix}. \tag{14}$$

Distance along the road is $\widehat{l} \in (\widehat{u}, \widehat{v})$, and the probability of collecting an image at a certain position on the pavement p_c can be expressed as follows:

$$p_c(v, f) = \min\left(\frac{\widehat{l} \cdot f}{v} - \varepsilon, 1 - \varepsilon\right), \tag{15}$$

where ε represents equivalent loss of the focused image. When the length of the road covered by the camera exceeds v multiplied by the collection interval, there will be duplicate areas between the pictures, so the detection probability is $1 - \varepsilon$. According to the conditions set in this paper, p_c is calculated to be 0.67, where $\varepsilon = 0.05$, $v = 50\,\text{km/h}$, $f = 2\,\text{Hz}$, and $\widehat{l} = 5\,\text{m}$. The minimum number of detected vehicles is five as calculated by the following formula (16):

$$n = \lceil \log_{0.473}(0.05) \rceil = 5. \tag{16}$$

According to the calculation result, at least five vehicles are needed to form the whole picture of the road surface. Based on the camera parameters used in this experiment, the relationship between speed, sampling frequency, and the minimum number of vehicles is shown in Figure 15. Figure 16 depicts the relationship between sampling frequency and the minimum number of vehicles at a speed of 50 km/h. The sampling frequency is determined by the traffic flow, the number of vehicles, the facilities, and the experimental environment. The purpose of this part is to indicate that the number of detecting vehicles is an essential parameter for further field implementation, and thus we conducted theoretical deductions to provide a recommended number of detecting vehicles, which can offer help for field applications.

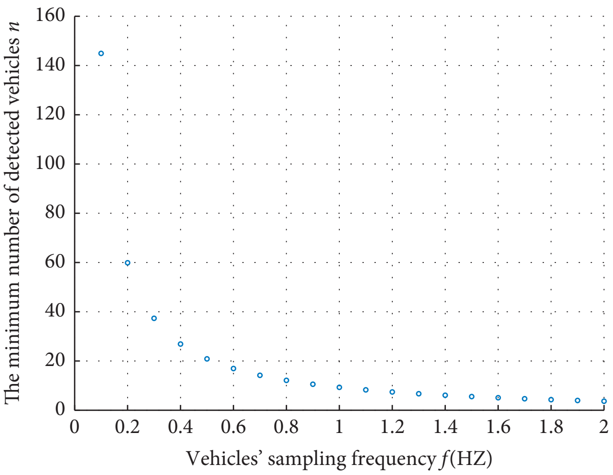


Figure 16: The relationship between sampling frequency and minimum number of vehicles at the detection speed of 50 km/h.

7. Conclusion

In this paper, we established a feature-matching and image-stitching method for pavement distress detection based on images obtained with multiple vehicles. A large number of pavement images and their corresponding time and positional information were obtained with detection vehicles under controlled acquisition conditions.

A hierarchical framework was built to process the images, including rough data filtering, feature matching, and image stitching. Duplications were effectively eliminated based on the three-layer structure that included GPS, bounding boxes, and SIFT features. GPS is used to avoid full-dataset comparison, which can reduce the calculating amount. SIFT was introduced to match features based on the extracted key regions using SSM and the bounding boxes. An SVM was used to analyze the influence of the output parameter thresholds of the MEuD and the MCR of the matching classification. The matching accuracy using the 5-fold cross-check method to calculate SIFT features is 81.4%, and the multilevel comprehensive matching accuracy can reach up to 92.0%. Images that have the most feature matches were stitched according to the matched features and the fitted perspective matrix. We then discussed the correlation between the sampling frequency and the number of detection vehicles and introduced a method to calculate.

Not only the whole lane-level pavement distress can be analyzed statistically by eliminating duplications and clustering according to the GPS tag and matched features, but local pavement distress can also be visually represented with the image-stitching algorithm. The algorithm provided in this paper effectively solves the problems of duplications of pavement distress and provides a reliable means for pavement distress detection in a collaborative, multivehicle environment.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to acknowledge the support provided by Ning Pan for processing the data. The authors are responsible for all views and opinions expressed in this paper. This study was supported by the Joint Science Foundation of the Ministry of Education of China & CMCC (2018202004). The first author was supported by the Program for Changjiang Scholars and Innovative Research Team in University.

References

[1] J. Ma, L. Cheng, and D. W. Li, "Road maintenance optimization model based on dynamic programming in urban traffic network," *Journal of Advanced Transportation*, vol. 2018, Article ID 4539324, 11 pages, 2018.

[2] W. Y. Yan and X.-X. Yuan, "A low-cost video-based pavement distress screening system for low-volume roads," *Journal of Intelligent Transportation Systems*, vol. 22, no. 5, pp. 376–389, 2018.

[3] S. Mathavan, K. Kamal, and M. Rahman, "A review of three-dimensional imaging technologies for pavement distress detection and measurements," *Ieee Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2353–2362, 2015.

[4] Y. Du, C. Liu, Y. Song, Y. Li, and Y. Shen, "Rapid estimation of road friction for anti-skid autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2019.

[5] T. B. J. Coenen and A. Golroo, "A review on automated pavement distress detection methods," *Cogent Engineering*, vol. 4, no. 1, p. 23, 2017.

[6] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth, "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," *Advanced Engineering Informatics*, vol. 29, no. 2, pp. 196–210, 2015.

[7] J. Masino, J. Thumm, G. Levasseur et al., "Characterization of road condition with data mining based on measured kinematic vehicle parameters," *Journal of Advanced Transportation*, vol. 2018, Article ID 8647607, 10 pages, 2018.

[8] J. D. Zhao, H. Q. Wu, and L. L. Chen, "Road surface state recognition based on SVM optimization and image segmentation processing," *Journal of Advanced Transportation*, vol. 2017, Article ID 6458495, 21 pages, 2017.

[9] K. Gopalakrishnan, "Deep learning in data-driven pavement image analysis and automated distress detection: a review," *Data*, vol. 3, no. 3, p. 19, 2018.

[10] C. Koch, G. M. Jog, and I. Brilakis, "Automated pothole distress assessment using asphalt pavement video data," *Journal of Computing in Civil Engineering*, vol. 27, no. 4, pp. 370–378, 2013.

[11] H. Oliveira and P. L. Correia, "Automatic road crack detection and characterization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 155–168, 2013.

[12] Q. Zhang and Z. Qin, "Application of machine vision technology IN road detection," *Stavební Obzor—Civil Engineering Journal*, vol. 27, no. 4, pp. 513–524, 2018.

[13] I. Sobel, "Neighborhood coding of binary images for fast contour following and general binary array processing," *Computer Graphics and Image Processing*, vol. 8, no. 1, pp. 127–135, 1978.

[14] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[15] S. C. Radopoulou and I. Brilakis, "Automated detection of multiple pavement defects," *Journal of Computing in Civil Engineering*, vol. 31, no. 2, p. 14, 2017.

[16] Y.-J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.

[17] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *Proceedings of the 2016 IEEE International Conference on Image Processing*, pp. 3708–3712, Phoenix, AZ, USA, September 2016.

[18] N. D. Hoang, Q. L. Nguyen, and V. D. Tran, "Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network," *Automation in Construction*, vol. 94, pp. 203–213, 2018.

[19] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection," *Construction and Building Materials*, vol. 157, pp. 322–330, 2017.

[20] M. D. Jenkins, T. A. Carr, M. I. Iglesias, T. Buggy, and G. Morison, "A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks," in *Proceedings of the 2018 26th European Signal Processing Conference*, pp. 2120–2124, Rome, Italy, September 2018.

[21] B. X. Li, K. C. P. Wang, A. Zhang, Y. Fei, and G. Sollazzo, "Automatic segmentation and enhancement of pavement cracks based on 3D pavement images," *Journal of Advanced Transportation*, vol. 2019, Article ID 1813763, 9 pages, 2019.

[22] G. Y. Baladi, E. C. Novak Jr., and W.-H. Kuo, "Pavement condition index—remaining service life," in *Pavement Management Implementation*, ASTM, West Conshohocken, PA, USA, 1992.

[23] M. Songyan and B. Lu, "A face detection algorithm based on AdaBoost and new Haar-like feature," in *Proceedings of the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, August 2016.

[24] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[25] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp. 20–27, Kerkyra, Greece, September 1999.

[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[27] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[28] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proceedings of the 2011 IEEE International Conference on Computer Vision*, pp. 2564–2571, Barcelona, Spain, November 2011.

[29] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "DeepMatching: hierarchical deformable dense matching," *International Journal of Computer Vision*, vol. 120, no. 3, pp. 300–323, 2016.

[30] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.

[31] G. Xiaoyan, P. Song, Y. Rao et al., "Dynamic image stitching for moving object," in *Proceedings of the 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, December 2016.

[32] S. Qiu, D. Zhou, and Y. Du, "The image stitching algorithm based on aggregated star groups," *Signal, Image and Video Processing*, vol. 13, no. 2, pp. 227–235, 2019.

[33] M. Munoz-Organero and R. Ruiz-Blazquez, "Detecting different road infrastructural elements based on the stochastic characterization of speed patterns," *Journal of Advanced Transportation*, vol. 2017, Article ID 3802807, 11 pages, 2017.

[34] A. Suryana, F. Reynaldi, F. Pratama, G. Ginanjar, I. Indriansyah, and D. Hasman, "Implementation of haversine formula on the limitation of E-voting radius based on android," in *Proceedings of the 2018 International Conference on Computing, Engineering, and Design (ICCED)*, Bangkok, Thailand, September 2018.

[35] L. Yanfang, Y. Wang, W. Huang, and Z. Zhang, "Automatic image stitching using SIFT," in *Proceedings of the 2008 International Conference on Audio, Language and Image Processing*, Shanghai, China, July 2008.

[36] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[37] H. Xiaodi and Z. Liqing, "Saliency detection: a spectral residual approach," in *Proceedings of the CVPR' 07. IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, June 2007.

[38] Y. C. Du, N. Pan, Z. Xu, F. Deng, Y. Shen, and H. Kang, "Pavement distress detection and classification based on YOLO network," *International Journal of Pavement Engineering*, pp. 1–14, 2020.

[39] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, Las Vegas, NV, USA, June 2016.

[40] V. Mandal, L. Uong, and Y. Adu-Gyamfi, "Automated road crack detection using deep convolutional neural networks," in *Proceedings of the 2018 IEEE International Conference on Big Data*, pp. 5212–5215, Seattle, WA, USA, December 2018.

[41] S. Luo, C. Xu, and H. Li, "An application of object detection based on YOLOv3 in traffic," in *Proceedings of the 2019 International Conference on Image, Video and Signal Processing, IVSP*, Shanghai, China.