

Article

# Real-Time Video Stitching for Mine Surveillance Using a Hybrid Image Registration Method

Zongwen Bai <sup>1,2,\*</sup>, Ying Li <sup>1,\*</sup>, Xiaohuan Chen <sup>2</sup>, Tingting Yi <sup>2</sup>, Wei Wei <sup>3</sup>, Marcin Wozniak <sup>4</sup>   
and Robertas Damasevicius <sup>4,5</sup> 

<sup>1</sup> School of Computer Science, National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, Shaanxi Provincial Key Laboratory of Speech & Image Information Processing, Northwestern Polytechnical University, Xi'an 710129, China

<sup>2</sup> School of Physics and Electronic Information, Shaanxi Key Laboratory of Intelligent Processing for Big Energy Data, Yan'an University, Yan'an 716000, China; chenxiaohuan@yau.edu.cn (X.C.); yitingting@yau.edu.cn (T.Y.)

<sup>3</sup> School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; weiwei@xaut.edu.cn

<sup>4</sup> Faculty of Applied Mathematics, Silesian University of Technology, 44-100 Gliwice, Poland; marcin.wozniak@polsl.pl (M.W.); robertas.damasevicius@polsl.pl (R.D.)

<sup>5</sup> Department of Applied Informatics, Vytautas Magnus University, 44404 Kaunas, Lithuania

\* Correspondence: ydbzw@yau.edu.cn (Z.B.); lybyp@nwpu.edu.cn (Y.L.)

Received: 6 July 2020; Accepted: 13 August 2020; Published: 19 August 2020



**Abstract:** Video stitching technology provides an effective solution for a wide viewing angle monitoring mode for industrial applications. At present, the observation angle of a single camera is limited, and the monitoring network composed of multiple cameras will have many overlapping images captured. Monitoring surveillance cameras can cause the problems of viewing fatigue and low video utilization rate of involved personnel. In addition, current video stitching technology has poor adaptability and real-time performance. We propose an effective hybrid image feature detection method for fast video stitching of mine surveillance video using the effective information of the surveillance video captured from multiple cameras in the actual conditions in the industrial coal mine. The method integrates the Moravec corner point detection and the scale-invariant feature transform (SIFT) feature extractor. After feature extraction, the nearest neighbor method and the random sampling consistency (RANSAC) algorithm are used to register the video frames. The proposed method reduces the image stitching time and solves the problem of feature re-extraction due to the change of observation angle, thus optimizing the entire video stitching process. The experimental results on the real-world underground mine videos show that the optimized stitching method can stitch videos at a speed of 21 fps, effectively meeting the real-time requirement, while the stitching effect has a good stability and applicability in real-world conditions.

**Keywords:** video stitching; image registration; hybrid image feature detection; mine video surveillance

## 1. Introduction

According to the 2018 statistical data, China is the biggest global producer of coal [1], while coal remains the main energy source in China. Most of the coal is produced from the underground coal mines. Despite recent technological innovations and advances in labor safety, coal mining remains a high-risk industry, while accidents in mines still happen frequently [2]. Therefore, preventing coal mine accidents is critically important for ensuring work safety and sustainable economic development. Prevention of lethal accidents in underground mines require the adoption of intelligent video surveillance methods to facilitate underground work safety and health [3]. Providing clear and uninterrupted video surveillance

images is the principle basis for ensuring safe operation of mining industry and timely emergency alerting [4]. Moreover, such underground video surveillance systems are a critical part of the concept of digital underground mines supported by cyber-physical-systems (CPS) [5], Internet of Things (IoT) [6] and smart sensors [7], which enable proactive real-time safety monitoring and risk assessment [8].

Image and video stitching have become increasingly popular research fields [9,10], and their application range is becoming more and more extensive [11]. For example, image stitching has been widely used in security monitoring [12], scenic viewing [13], object tracking [14], and autonomous car driving [15]. Image stitching forms an important part of the intelligent video surveillance process allowing to provide seamless monitoring from multiple video cameras. The basic process of image stitching involves image acquisition, image preprocessing, image registration, and image fusion. Before video stitching, the video must be stitched frame by frame, so the implementation of video stitching depends on the implementation of image stitching. Image stitching is mainly divided into image registration and image fusion. Image registration methods can be based on grayscale and template methods, image transform methods, and image feature-based methods [16]. The latter methods have the best comprehensive performance and have strong general-purpose and a wide range of application characteristics [17].

Feature-based image registration is a hot topic in current image processing technology research [18]. The main process is to extract the features that meet certain conditions from the two images to be matched, and then match the features of the two images to generate the geometric transformation relationship between the two images for image stitching. Features such as points, lines, edges, contours, and geometric moments are commonly used.

Numerous research studies have been done on image stitching and video stitching technology, recently. For example, Alomran et al. extracted the feature points of the image by using the speeded up robust features (SURF) method, and tested the minimum overlap area for image stitching to work normally and the stitching effect under different focal lengths of the camera [19]. Ho et al. suggested an interpolation grid image stitching method that uses rigid moving least squares algorithm to solve the fisheye lens 360° image stitching and solve the problem of video jitter through a graphics processing unit (GPU) acceleration [20]. Yeh et al. proposed a real-time video stitching method and accelerated the system speed through parallel programming using Compute Unified Device Architecture (CUDA). This method performs well in scenes with better lighting conditions and has real-time performance [21]. Babu et al. used the Harris algorithm to extract feature points of the image, and applied feature matching to the geometric differences of feature vectors through biorthogonal wavelet transform to achieve deep sea image registration and stitching [22]. Ruan et al. proposed an effective image stitching algorithm for SURF feature matching and wavelet transform for image fusion, which has a better effect on image stitching [23]. Aung et al. described the process of creating a panoramic image from multiple frames with a narrow horizontal angle of view, and proposed methods to reduce the processing time of stitched video images [24]. Lu et al. [25] used a modified fast and rotated brief (ORB) algorithm, combined with a local feature distance constraint based rough selection and geometric constraints-based image matching algorithm to find local features based on geometric invariances. Nie et al. used a method to recognize the background of input videos in order to deal with large parallax problems, and the false matching elimination scheme and optimization are encapsulated into a loop to prevent the optimization from being affected by poor functional matching [26]. Fang et al. used weighted interpolation algorithms to calculate color changes for all pixels in the base image based on the color difference of the pixels on the best stitching line, and then added the calculated color changes to the base image to eliminate color inconsistencies in order to achieve seamless image stitching [27]. Lee et al. used adaptive warping followed by inter-view and inter-frame correspondence matching by energy minimization for multi-view videos stitching [28]. Yang et al. minimized an energy function on the indirect graph Markov random field to stitch real-world undersea images [29]. Park et al. used the homography based method to stitched videos taken from static cameras [30].

In the underground mine video surveillance systems, the recorded video images are usually noisy and blurry and have low contrast and poor illumination. Therefore, the image denoising and detail enhancement methods must be applied, such as the method based on L0 norm minimization and low rank analysis [31]. Due to the limited viewing angle of a single camera, the data provided often cannot meet the application requirements. In this complex context, the SIFT feature extraction combined with multiband lifting wavelet transform [32] were used. To compensate for image distortions, such as the ones introduced by the fisheye lenses, the latitude and longitude geographic coordinate approach, spherical projection, and panoramic stitching can be used [33]. Usually, in a video surveillance system, multiple cameras are used to shoot from different perspectives and then stitch the pictures into the panoramic image. Panoramic image stitching uses images taken by different sensors at different times and different viewpoints, to find their overlapping areas, and obtain a panoramic image. To generate panoramas, Kim et al. used dynamic programming based on energy minimization, followed by content-aware adaptive blending to reduce color discontinuity and a partial seam-update scheme to account for the moving objects [34]. Liu et al. adopted the graph-cut algorithm in the overlapped regions of the panoramic videos to calculate the optimal seams [35]. Li et al. constructed an analytical warping function for computing image deformations and directly reprojecting the warped images, whereas a Bayesian model was used delete invalid local matches [36]. Chen et al. used the locality preserving matching, the robust elastic warping function to eliminate the parallax error, and the global projectivity preserving method for drone image stitching [37]. Kang et al. adopted a convolutional neural network (CNN) to assess shifts between images, followed by a sub-pixel refinement based on a camera motion model [38]. Kang et al. proposed an edge-based weighted minimum error seam method to defeat limitations brought by parallax and video stitching, and an image matching method based on triangular ratios to reduce computational complexity [39].

Krishnakumar and Gandhi [40] suggested a video stitching method for keeping spatial-temporal consistency of videos captured from multiple cameras. First, they detected a feature point found in the first frame and then tracked it in the next frames using Kalman Filter combined with an interacting model, which allowed to reduce jitter and improve computational complexity. Krishnakumar and Indira Gandhi [41] used region-based statistics for detection and registration of multi-view features, which allowed to preserve spatial and temporal consistency. Lin et al. [42] suggested a method for stitching synchronized video streams acquired from multiple fixed cameras. First, they aligned the video background using line-preserving alignment, and used the constructed mesh for stitching, thus obtaining smooth transition videos without any artifacts and ghosting. Du et al. [43] use L-ORB algorithm for extraction of video features, LSH based method for matching feature points and CUDA based parallel video stitching approach for panoramic real-time video stitching. Kakli et al. [44] proposed using geometric look-up tables (G-LUT) to stitch a reference frame from multiple videos, which is used to map video frames to panorama. The neural network is used to recognize moving objects and identify G-LUT control points. The position of these points is adjusted using patch-match based optical flow. Liu et al. [35] used the improved graph-cut technique to obtain optimal seams in the overlapped regions. They used foreground detection combined with Gaussian filtering to achieve smooth seams of videos while avoiding blurring, distortion, and ghosting. Park et al. [30] used homography estimation from multiple video frames. The identified representative feature points are matched using random sample consensus (RANSAC) combined with probability-based sampling.

However, these methods work poorly in low lighting and high noise conditions, such as the ones present in the underground coal mine video surveillance images. Low-textured and low-quality images are difficult to align due to the unreliable and insufficient point matches. Moreover, they often fail on multi-camera images and videos with large parallax and wide baselines. Deep learning, as the most popular feature in engineering, is used in most image processing task, but it is very hard used in a noisy environment [45]. Most deep learning based methods use good quality images for training and ignore the fact that image quality can vary greatly in the testing data. As a result, deep learning

methods often fail on noisy images due to its tendency to overfit data [46]. However, our task is a typical representative of the noisy image processing task.

To address these problems, we propose an underground video surveillance image stitching framework that includes image defogging, image feature extraction, and image registration. Our novelty and contribution is a novel hybrid image feature detection and extraction method based on Moravec optimization [47] and the SIFT [48] operator for underground surveillance video stitching.

As we all known, the 360-degree panoramic camera is very popular in video detection and as an alternative solution, it has a broader perspective. However, in our research, we consider the more common situation of occlusion, because the underground roadway is always zigzag-shaped.

The organization of the remaining sections of this paper is as follows: The proposed image stitching method for underground coal mine video surveillance images is presented in Section 2. The results are revealed and discussed in Section 3. Finally, the conclusions are delivered in Section 4.

## 2. Materials and Methods

### 2.1. Image Preprocessing for Defogging

When using a detection algorithm to detect the feature points of an image, in order to improve the detectability of the image, it is generally necessary to pre-process the acquired image. Due to the complex conditions in the coal mine, the collected images usually are blurred, which is not conducive to image processing in the subsequent steps. In our research, the camera is fixed since, our application scenario is the underground mine, but our video stitching is dynamic; in other words, more new videos will be added over time.

The fog (haze) is considered as the main cause of image degradation due to water vapor and coal dust present in the underground air [49]. Therefore, the image dehazing algorithm is used to pre-process the images in order to improve image quality. A physical model of the foggy image degradation based on the atmospheric scattering model and its deformation form is created, and the image degradation process is reversed to restore a fog-free image. Currently, a dark channel prior algorithm [50] based on the prior laws of dark primary colors, is recognized as the best method for defogging (dehazing) [51,52].

Atmospheric scattering [53] is a traditional model which is widely in practice to describe the image restoration process as follows:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (1)$$

where  $I(x)$  is a foggy image,  $J(x)$  is a target image to be finally recovered,  $A$  is an atmospheric light value, and  $t(x)$  is a transmittance,  $t(x) = e^{-\beta d(x)}$ , where  $\beta$  is the scattering coefficient, and  $d(x)$  is the image depth.

According to the physical scattering model, only by accurately estimating the values of atmospheric light  $A$  and transmittance ratio  $t(x)$  can a restored image  $J(x)$  with better dehazing effect be obtained. The dark channel prior law states that for an image, the dark channel can be determined by the following expression:

$$J^{dark} = \min_{y \in \Omega(x)} \left( \min_{c \in \{R,G,B\}} J^c(y) \right) \quad (2)$$

where  $J^c$  is the brightness value of each channel of the color image,  $c$  is the color channel in the image, and, according to the dark channel theory, the lower value  $J^{dark}$  approaches zero.

We consider that the atmospheric light value of an image is constant and the transmittance is also a constant within  $\Omega(x)$ . Therefore, the minimum value channel is first taken on both sides of the physical scattering formula, and then the local minimum value is filtered. The rough estimation expression of transmittance after finishing is as follows:

$$t(x) = 1 - \min_{y \in \Omega(x)} \left( \min_{c \in \{R,G,B\}} \frac{I^c(y)}{A^c} \right) \quad (3)$$

In order to make the restored image more consistent with the subjective visual perception of the human eye, a small amount of fog must be retained in the restored image, so the degree of defogging is controlled by constraining the parameter  $w$ , that is,

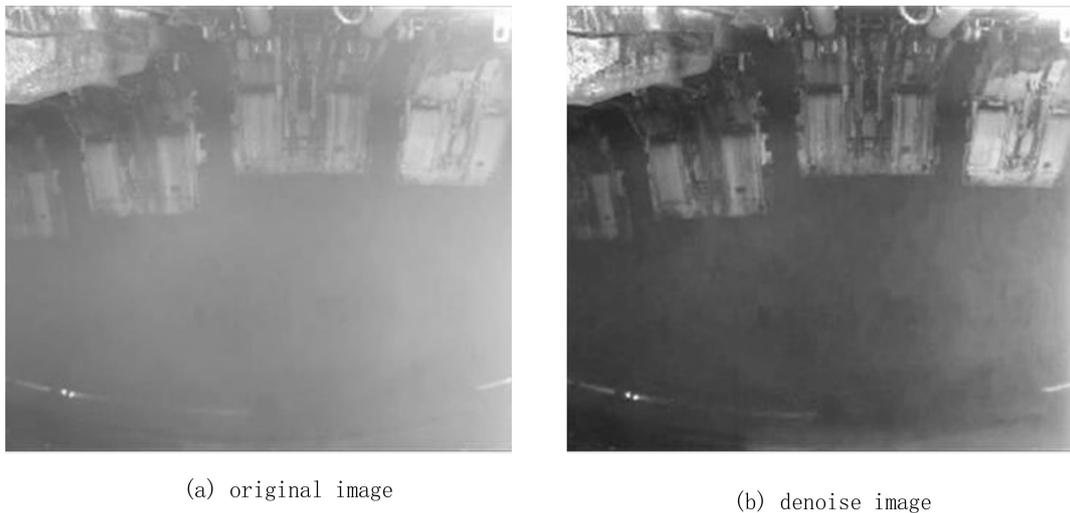
$$t(x) = 1 - w \min_{y \in \Omega(x)} \left( \min_{c \in \{R,G,B\}} \frac{I^c(y)}{A^c} \right) \quad (4)$$

where  $w = 0.95$  is generally taken, and the restored image is obtained as:

$$J(x) = \frac{I(x) - A}{\max(t(x), t_0)} + A \quad (5)$$

Equations (2) to (5) provide the details to calculate process of each term of Formula (1), and the formula has a limiting constant, generally  $t_0 = 0.1$ , because when it is equal to 0, the first term in the physical model of atmospheric scattering  $J(x)t(x)$  is also equal to 0, which results in the final restored image to be noisy. Therefore, the transmittance must be limited.

Through the above methods, we can process the image of the coal mine to reduce the impact of dust in the image, as an example is shown in Figure 1.



**Figure 1.** An example of image defogging: original image (a) and denoised (defogged) image (b).

## 2.2. Hybrid Detection Method Based on Moravec Optimization

In the complex underground scene, to help the mine staff to monitor the actual situation, the Moravec operator [47] can be used for feature extraction from images. However, there is no corresponding feature point descriptor in the Moravec operator to describe the feature points, which provides the experimental basis for the next step of feature point matching. Therefore, we propose the use of the hybrid feature extraction method based on the Moravec operator to extract the feature points, combined with the feature description using the SIFT operator.

The implementation steps are as follows:

- (1) Calculate image corner response function for a set of image pixel interest points:

The pixel  $p$  block window created in the image with the pixel as the center uses the following formula to calculate the sum of the squared differences of the gray differences of the adjacent pixels in the four directions  $v_H, v_B, v_V, v_D$ :

$$\begin{cases} v_H = \sum_{i=-K}^{K-1} (g_{x+i,y} - g_{x+i+1,y})^2 \\ v_B = \sum_{i=-K}^{K-1} (g_{x+i,y+i} - g_{x+i+1,y+i+1})^2 \\ v_V = \sum_{i=-K}^{K-1} (g_{x,y+i} - g_{x,y+i+1})^2 \\ v_D = \sum_{i=-K}^{K-1} (g_{x+i,y-i} - g_{x+i+1,y-i-1})^2 \end{cases} \quad (6)$$

where  $g(x, y)$  is the pixel value of the image at the position  $(x, y)$ , and  $K$  is  $[N \div 2]$ . Take the minimum of the four values of  $v_H, v_B, v_V, v_D$  as the corner response function, that is  $I_V = \min(v_H, v_V, v_D, v_B)$ , and set the value  $I_V$  to the point of interest.

(2) Determine the threshold value of pixel points and filter the candidate points that meet the conditions.

The selection criteria of the threshold range should meet two basic conditions: the required points can be extracted from the candidate points; redundant corner points can be removed to avoid many feature points, thus resulting in excessive calculations. After the threshold is set, compare the interest value with the threshold, and select the points, which are larger than the threshold, as candidate points.

(3) Determine corner points by local non maximum suppression method.

Local non-maximum suppression selects feature points. In a certain size window, the candidate feature points selected in the second step are removed from the point, whose response value is not a maximum value, and the remaining feature points are the feature points of the area are sought.

(4) Use the feature descriptors of the SIFT operator to describe the corner points selected in step (3). The implementation steps are the following:

(4a) Find the scale space structure.

Perform scale transformation on the source image, and then obtain the scale space representation sequences of the image at multiple scales. The scale sequence-based main contour extraction is performed on these representation sequences in order to find multi-scale features. A Gaussian function is used to convolve the image. The scale space of a 2D image is:

$$L(x, y, \delta) = G(x, y, \delta) * I(x, y) \quad (7)$$

where  $G(x, y, \delta)$  is a two-dimensional Gaussian function with a variable scale:

$$G(x, y, \delta) = \frac{1}{2\pi\delta^2} e^{-\frac{(x^2+y^2)}{2\delta^2}} \quad (8)$$

where  $\delta$  is the scale space factor, and the size of  $\delta$  determines the smoothness of the image. The bigger the value is, the more it corresponds to the outline feature of the 2D image, and the lesser the value is, the less detailed features of the 2D image are.

The Gaussian difference scale space is generated by using different scales of the Gaussian difference kernel and image convolution, and the DOG calculation is as follows:

$$D(x, y, \delta) = (G(x, y, k\delta) - G(x, y, \delta)) * I(x, y) = L(x, y, k\delta) - L(x, y, \delta) \quad (9)$$

The SIFT algorithm introduces a Gaussian pyramid model to reflect the continuity of scale, that is, a Gaussian filter is added based on downsampling of the image pyramid. Gaussian filtering is performed on the first layer of the image pyramid with parameters of  $0, \delta, k\delta, 2k\delta$  to obtain the first

set of images. Downsampling the first layer of the first group of images to obtain the first layer of images in the second group, and then Gaussian smoothing on the second group of images is performed. After repeated several times, the image pyramid of the  $O$  group layer  $L$  is obtained.

(4b) Obtain stable feature points.

The Gaussian difference ( $DOG$ ) pyramid is used to obtain stable feature points. The first layer of the first batch of images in the Gaussian difference pyramid is obtained by subtracting the first layer of the first batch from the second layer of the first batch of Gaussian pyramids, and so on. Each group of Gaussian pyramids can obtain a layer height is  $L - 2$  of  $DOG$  pyramid. In order to determine the extreme points, in the Gaussian differential pyramid model, each pixel in the middle layer is compared with its eight adjacent pixels on the same layer and each adjacent pixel in the upper and lower adjacent layers. To ensure that local extremums are detected in both scale space and image space. If the  $DOG$  value of a pixel  $P$  marked as a point is higher than the neighboring pixels, the position of the point  $P$  and the corresponding scale are recorded. Most of the detected feature points still have noise effects and corresponding edges. So, we need to further extract the feature points to get feature points that are more stable. First, use the Taylor expansion  $D(x, y, \delta)$  to expand as shown:

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \tag{10}$$

where

$$X = \begin{bmatrix} x \\ y \\ \delta \end{bmatrix} \frac{\partial D}{\partial X} = \begin{bmatrix} \frac{\partial D}{\partial x} \\ \frac{\partial D}{\partial y} \\ \frac{\partial D}{\partial \delta} \end{bmatrix} \frac{\partial^2 D}{\partial X^2} = \begin{bmatrix} \frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial xy} & \frac{\partial^2 D}{\partial x\delta} \\ \frac{\partial^2 D}{\partial yx} & \frac{\partial^2 D}{\partial y^2} & \frac{\partial^2 D}{\partial y\delta} \\ \frac{\partial^2 D}{\partial \delta x} & \frac{\partial^2 D}{\partial \delta y} & \frac{\partial^2 D}{\partial \delta^2} \end{bmatrix} \tag{11}$$

where  $D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X}$  is the derivative of  $D(X)$ .

Find the point  $\hat{X}$  whose derivative is 0,  $\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X}$ . If  $\left| D(\hat{X}) \right| \geq 0.03$ , keep this point as a feature point.

Define a Hessian matrix of size  $2 \times 2$ ,  $H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$ . The determined principal curvature requires the ratio of the eigenvalues of the matrix  $H$ . Assume that the maximum eigenvalue of the matrix  $H$  is  $\lambda_1$ , and the minimum eigenvalue is  $\lambda_2$ ,  $\gamma = \frac{\lambda_1}{\lambda_2}$  where,  $Tr(H) = D_{xx} + D_{yy} = \lambda_1 + \lambda_2$ ,  $Det(H) = D_{xx}D_{yy} + (D_{xy})^2 = \lambda_1\lambda_2$ . In order to eliminate edge response points, the reference value of  $\lambda_1$  is set to 10 in this paper, and when  $\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$ , this point is considered an extreme point.

(4c) Feature point description.

The feature point extraction part of the SIFT algorithm has obtained stable feature points. It is necessary to establish a descriptor for each feature point so that it has the characteristics of invariance of illumination and rotation.

Set  $p(x, y)$  as a feature point, then the gradient value  $g(x, y)$  and gradient direction  $\theta(x, y)$  of the point  $p$  are:

$$g(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2}, \tag{12}$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)} \tag{13}$$

where  $L$  is the scale space, where the point  $p$  is located.

The feature point  $p$  is the center, a  $2 \times 2$  window is selected, and the gradient value and direction of each pixel in the window are calculated on the scale, where the feature point is located, and expressed in the histogram. The histogram has 36 directions, each of which contains  $10^\circ$  out of a gradient

direction of  $0^\circ$  to  $360^\circ$ . The maximum value of the gradient direction in the histogram is taken as the principle direction of the feature point, while the direction greater than 80% of the max value in the main direction is retained as the auxiliary direction to improve the accuracy of the feature description. To ensure that the feature points detected by the SIFT algorithm are rotationally invariant, after determining the principle direction of the feature point  $p$ , the main direction is used as the direction of the coordinate axis.

Finally, the Gaussian weighting is performed on the pixels except the row and column, where the feature points are located within a  $16 \times 16$  window with the eigen-point as the center. The  $16 \times 16$  size window is divided into  $16 \times 4 \times 4$  small windows, and the gradients in eight directions of each small window are accumulated. A vector of size  $4 \times 4 \times 8 = 128$  is generated as the feature point descriptor.

### 2.3. Feature Point Matching Using KD Tree

After extracting the feature points of two images, we need to match the points extracted using the SIFT feature descriptor and determine the matching feature point pairs. By calculating the relationship between the matched feature points, the transformation matrix between two images is obtained. Due to the scale and shift transformations between two images to be stitched, at least four pairs of matching points need to be obtained in order to get the transformation matrix between two images. To achieve this, we use the nearest neighbor matching algorithm based on the  $k$ -dimensional (KD) tree. First, features are extracted and feature description vectors are generated. Next, the KD tree index is established and search matching is completed by the nearest neighbor algorithm.

The KD tree is a  $k$ -dimensional binary index tree. The feature space of the KD tree-based feature approximate nearest neighbor matching is the  $n$  dimension real number vector space  $R_n$ , which uses the Euclidean distance to find the neighbor of the instance. The KD tree is used to divide data points into specific parts in the  $n$ -dimension space  $R_n$ . The purpose of the KD tree is to retrieve the nearest Euclidean distance from the query point. After all, Euclidean distances  $d(p, q)$  are stored in the KD tree structure, the nearest point can be searched effectively.

The search of the nearest neighbor based on the KD tree index is performed as follows:

(1) Feature points are extracted by the SIFT operator, the feature point descriptor vector is generated, and the KD tree index is established according to the set of the  $k$ -dimension data points.

(2) Using the priority queue, search the KD tree for a given target point, find the nearest neighbor data points  $m$ , scan from the root node to the leaf node, and assign missed nodes to the queue; then, take out the point with the smallest distance from the current dimension from the queue, and rescan to the leaf node until the queue is empty, or the number of scans reaches the set upper limit.

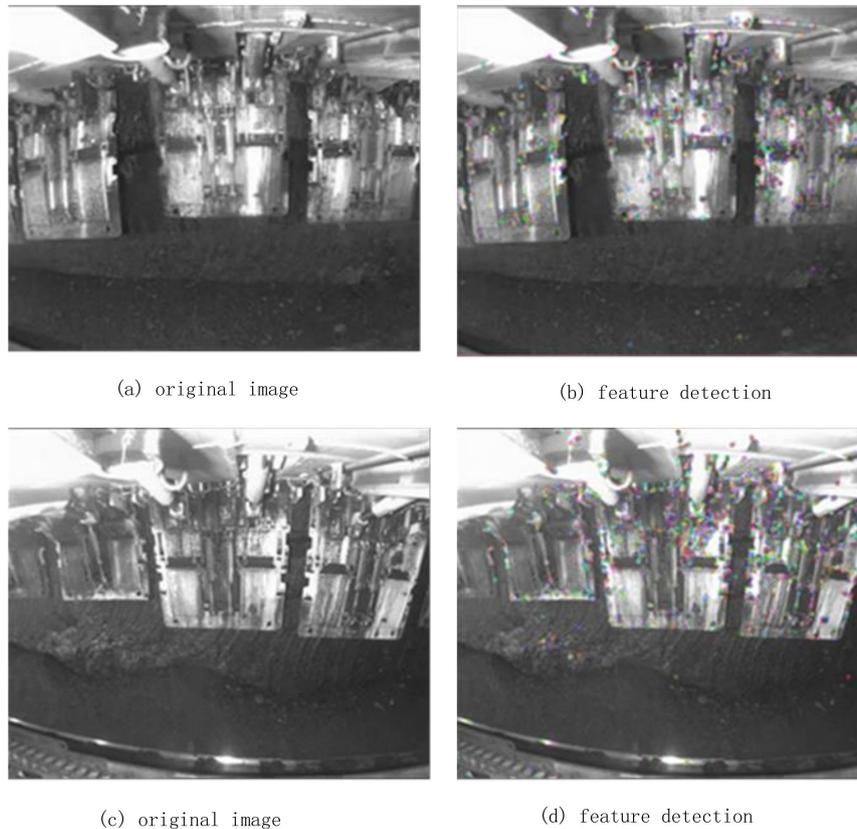
(3) If  $M$  nearest neighbors are found, compare the distance between the closest and the second closest neighbor to the target point. If it is less than the given threshold, select the nearest neighbor as the search result; if it is not found, return to null.

Figure 2a,c represent the images in the actual scene of the coal mine, respectively. The detected and labeled features are shown in Figure 2b,d, respectively. The feature points are mostly concentrated in the areas, where the image features are obvious, indicating that the detection algorithm works well on complex coal mine underground scenes.

### 2.4. Image Registration

After getting the feature matching pairs between the images to be stitched, we need to use the transformation model between the images to determine the pixel moving relationship between the images. In order to get a high-quality mosaic image, we need to refine the feature matching pairs. Here we use the random sample consensus (RANSAC) algorithm [54] to purify the feature point pairs. The RANSAC algorithm uses the iterative method to calculate the best mathematical model and parameters from a group of discrete data groups. The sample data group, whose data can be described by the same parameter model, also includes some data that deviates from the normal distribution and cannot be described by the parameter model. These data points deviating from

the normal distribution may be caused by the application of the incorrect measurement and calculation method. In the underground coal mine images, due to poor lighting, equipment, human operation, and other reasons, when the image is matched with feature points after feature point extraction, there are mismatched points, so it is necessary to purify the data set using the following algorithm.



**Figure 2.** Results of feature point detection in coal mine underground images: (a,c) original images, and (b,d) feature detection results.

For a group of observation data used for line estimation, first, randomly select two points to determine a straight line; then, set an allowable error threshold for the straight line; for other points except the selected two points; when the distance from the straight line is less than the threshold, set it as the interior point; otherwise set, it as the outer point. Repeat the above two steps until the number of interior points is enough, and the model obtained is the model to be estimated.

When solving the projection transformation matrix with the RANSAC algorithm, the following criteria are used to determine whether the obtained matching point is the interior point of the model: assuming that  $(p, p')$  is a matching point pair, the distance between the two points is calculated by transforming the projection  $p$  to  $p'$  the coordinate system in which it is located, similarly, the distance  $d_1$  between the two points is calculated by transforming the projection  $p'$  to  $p$  in the coordinate system in which it is located, and the sum of the distance between the two points  $d_1$  and  $d_2$  is called a pair by calculating the formula It is called transformation error  $d_i$ , and it is used as the basis of the interior point as follows:

$$d_i = \sum (d(\vec{p}_i, H^{-1}\vec{p}'_i)^2 + d(\vec{p}'_i, H\vec{p}_i)^2) \quad (14)$$

Set a threshold value  $T$ , if  $d_i$  is greater than the threshold value  $T$ , the matching point pair is determined as the outer point of the model, and the matching point pair is removed; otherwise, it is determined as the inner point of the model, and the matching point pair is retained.

The concrete implementation steps of the RANSAC algorithm are:

(1) In order to calculate the current parameter model  $H_{cur}$ , four pairs of matching feature points were randomly selected from the selected matching point pairs;

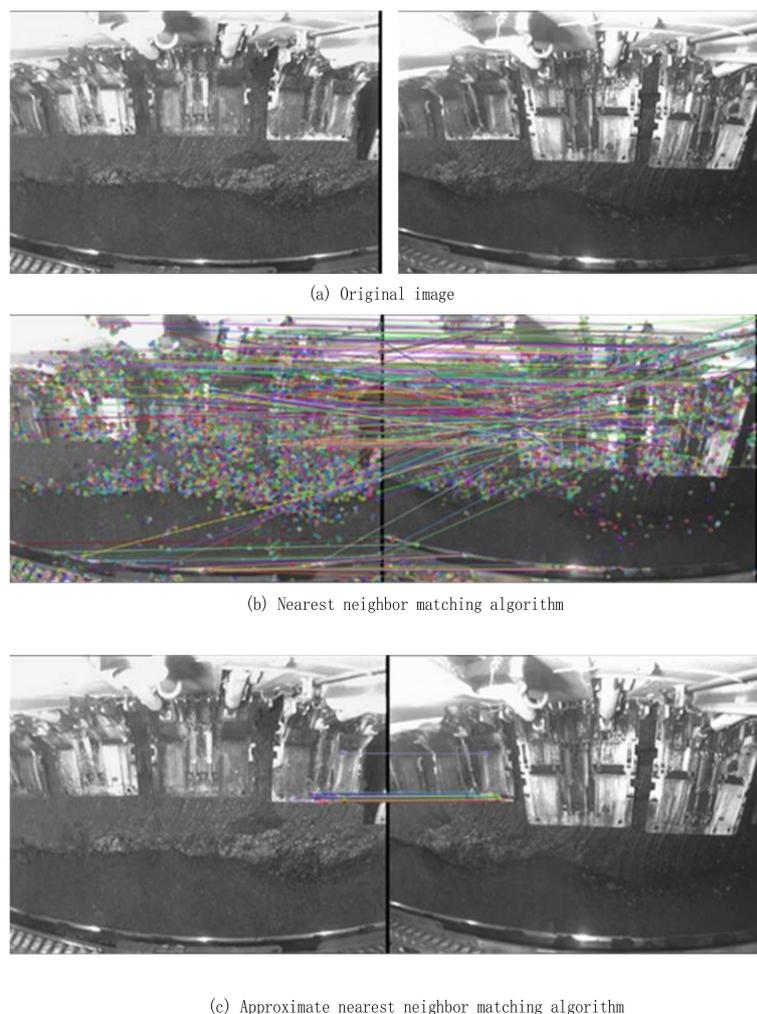
(2) For the other matching point pairs that are not selected, the symmetrical transformation error is calculated by  $H_{cur}$ . The inner point is set when the symmetrical transformation error is less than the threshold value, and the number of inner points  $m$  is counted;

(3) If  $M \geq M_{inliner}$ , and  $H_{cur}$  is a better parameter model, and  $H = H_{cur}$ ;

(4) Repeat steps (1) through (3) until completed.

If there are too few internal points, the model is discarded; otherwise if the new model performs better than the previous one, then the new model is saved.

For an example of image registration results, see Figure 3, which uses the same picture as in Figure 2. The feature points, extracted from the original images (Figure 3a), are described by the SIFT descriptors, and rough matching can be obtained by using nearest neighbor matching algorithm based on the KD tree. However, because there are many mismatches in the registration image, it is easy to cause errors in the calculation of transformation matrix, this paper uses random sampling consistent algorithm to match the registered feature points. One step is to delete the mismatching and get Figure 3c, to get the correct feature point registration sequence and calculate the transformation matrix. In Figure 3c, the correct feature point registration pair is more than four pairs, so the transformation matrix can be more accurate.



**Figure 3.** Results of coal mine image feature point registration: (a) original image, (b) nearest neighbor matching algorithms, and (c) approximate nearest neighbor matching algorithm.

In Figure 3, we can see that the RANSAC algorithm together with the nearest neighbor matching algorithm based on the KD tree, can effectively purify the feature point pairs of registration, providing an effective experimental basis for the next step of calculating the transformation matrix. After the feature point detection and registration algorithm, the effective feature point pair is used to calculate the parameter model, and the obtained projection transformation matrix is the optimal transformation model.

### 3. Experiment

#### 3.1. Dataset

We used the videos from four datasets: Tytyri mine, Finland; Sonderhausen mine, Germany [55]; Târgu Ocna Salt Mine, Romania; and Xiaoyadaokou tunnel, China [56]. The examples of images from the dataset are presented in Figures 4 and 5. We perform the experiments with MATLAB (MathWorks, Natick, MA, USA) implementation of our method on a PC with Intel Core i5-8265U 1.6GHz CPU and 8GB memory (Intel Corporation, Santa Clara, California, CA, USA). We use the SIFT implementation provided by the vlfeat library.

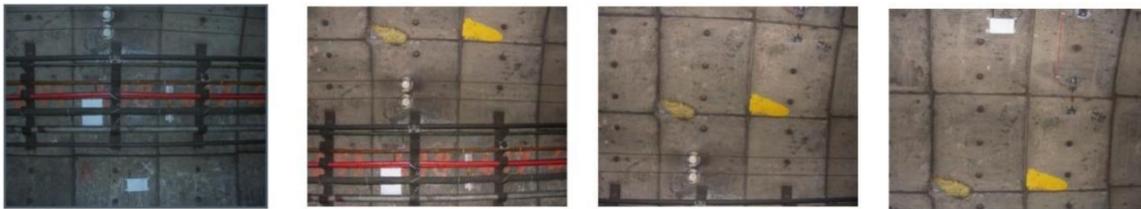


Figure 4. Example of video frames from Xiaoyadaokou tunnel, China [56].



Figure 5. Example of still images from Sonderhausen mine, Germany [55].

#### 3.2. Evaluation

Following Guo et al. [57], we adopted the stability score and the stitching score, while the third score, the trajectory stability score was adopted from [26], to evaluate the generated panoramic videos. We also use the MVSQA metric [58] for panoramic video quality assessment based on the distortion maps of the stitched images. We describe the used metrics in more detail as follows:

(1) Stability metric assesses the smoothness of the stitched video. The tracks are identified on the stitched video and analyzed using Fast Fourier Transform (FFT) in the domain of frequency. Then the energy of the lowest frequencies (from 2nd to 6th) are calculated and divided by full energy of the spectrum. The good value of the stability score should be close to 1.

(2) Stitching metric assesses the stitching quality [59]. First, for each video frame, the distance between matching features within a small area that is close to the stitching boundary is computed. The stitching score of one frame is calculated as the average of distances between all feature pairs [36]. Average distance reflects the difference of the two features, the smaller the score, the closer the features.

The gives final stitching score is the largest value among all frames as the good value of the stitching score should be close to 0, indicating good alignment of the frames.

(3) Trajectory stability score. We acquire all feature trajectories of the video, and segment it as suggested in [26], and for all segments we calculate the trajectory stability scores. The final stability score is computed as the average trajectory stability score of all the segments.

(4) The MVSQA score [58] uses a weighting mask around image stitching edge. In order to create the desired mask, a distance transform from the stitching seam based on the structural similarity (SSIM) index is calculated for each of the stitched images. This value is weighted by three maps, each representing local feature to which the human eye is most sensitive.

### 3.3. Results

We can evaluate the quality of the defogging using three evaluation metrics: standard deviation, average gradient, and information entropy. We can see from Table 1 that the image quality of the coal mine underground image is improved after the defogging process as the mean values of the evaluation metrics have been increased. Table 1 shows that the gradient have improved from 1.6242 to 2.3639, and information entropy is increased from 6.8060 to 7.1968. All the performance improving is because that the proper method is taken for the special tasks. In subjective evaluation, it can also be seen that the processed image is sharper than the original image. This facilitates monitoring of the actual situation in an underground coal mine tunnel.

**Table 1.** Evaluation of defogging algorithm: average standard deviation, gradient mean and information entropy values on four analyzed video datasets.

	Standard Deviation	Gradient Mean	Information Entropy
Original images	32.1948	1.7294	6.7359
Defogged images	32.8739	2.4861	7.2056

Table 2 summarizes the stability, stitching, and trajectory stability scores and MVSQA metric value of the analyzed underground mine and tunnel video datasets.

**Table 2.** Quantitative evaluation of stitching quality on four video datasets.

Video Dataset Number	#1 Tytyri	#2 Sonderhausen	#3 Târgu Ocna	#4 Xiaoyadaokou
Resolution	540p	720p	540p	540p
Number of frames	280	450	300	120
Processing time (s)	13.03	21.02	14.26	5.58
Stability score	0.638	0.529	0.482	0.572
Stitching score	1.02	1.11	0.89	0.94
Trajectory stability score	1.25	1.13	1.32	1.18
MVSQA	0.24	0.29	0.31	0.28

The average processing time of our method for a single video frame is 46.5 ms (i.e., about 21.5 fps), which means that our method is fitting for real-time mine surveillance systems. Finally, we present the examples of the generated stitched panoramas in Figure 6.

Table 3 presents a comparison between the algorithms [60,61] and our method. Note that in order to compare, we run the algorithms [60,61] on our dataset, but we must agree that these algorithms have their own scope of application.

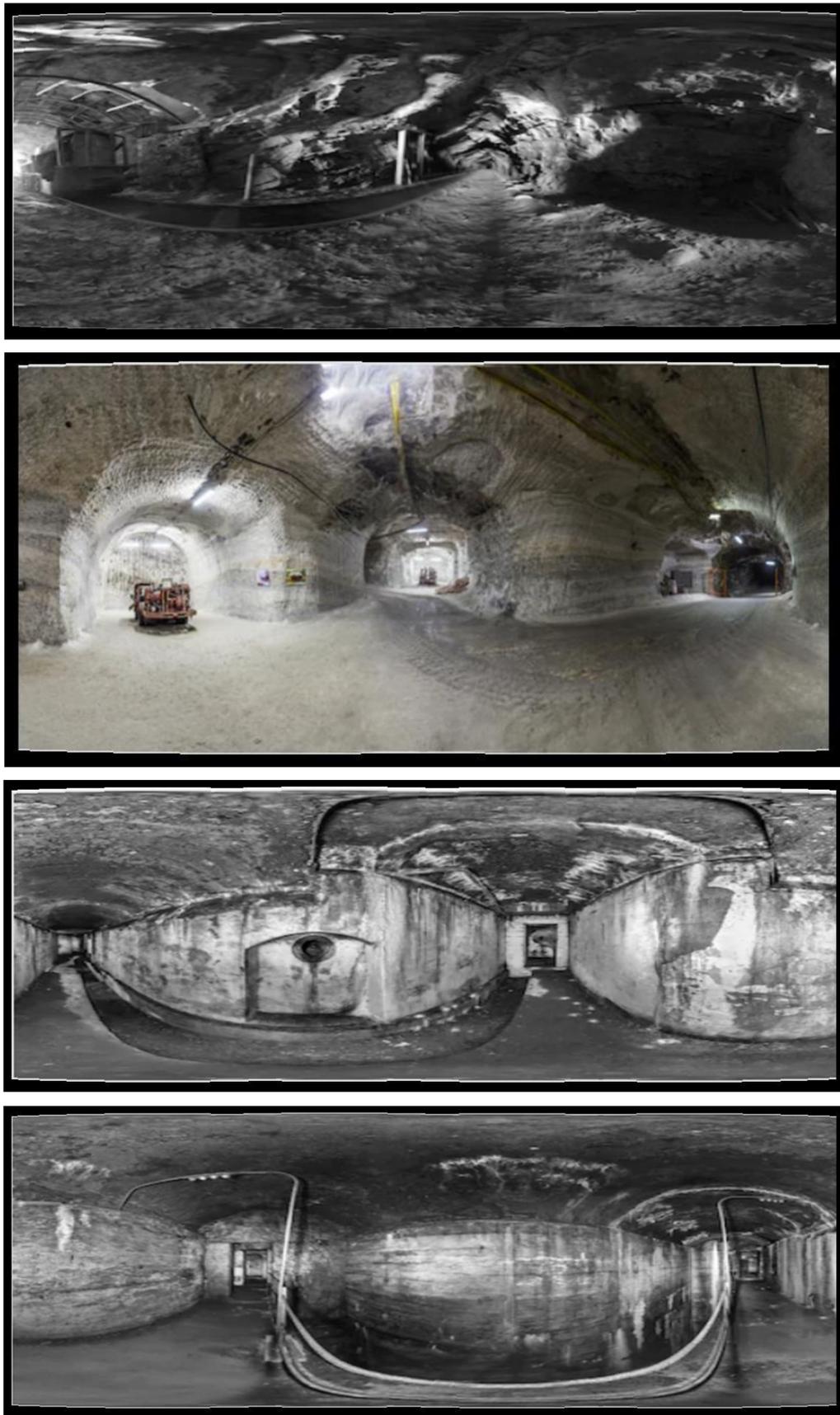


Figure 6. Example of generated panoramas of underground tunnel images.

**Table 3.** Comparison of results with related work.

Performance	Processing Time (s)	Stability Score	Stitching Score	Trajectory Score	MVSQA
Parallax-Tolerant [60]	6.62	0.545	0.91	1.20	0.25
Region based	6.58	0.562	0.88	1.16	0.27
Parallax-Tolerant [61]	5.58	0.572	0.94	1.18	0.28
Our method	5.58	0.572	0.94	1.18	0.28

### 3.4. Discussion

Unfortunately, there has been not much research on underground mine or tunnel video stitching. A vast majority of the related work use static images of underground tunnels and not perform real time processing of images. For example, Zhu et al. [62] used image stitching to make layout panorama of lining using images captured in the Changsha Metro tunnel and the Aizizhai tunnel using a hand-held camera. Huang et al. [63] performed visual inspection of metro shield tunnel for crack and leakage defects. Image stitching is implemented via splicing the overlapping images from six cameras. Kim et al. [64] stitched tunnel lining images to produce longitudinal panoramic images of tunnel wall surface. Konishi et al. [65] conducted image stitching of Shanghai metro line tunnels for water leakage detection. The images, however, are static and contain only parts of tunnel surfaces without any other objects. Similarly, Zhao et al. [66] analyzed the moisture marks of metro shield tunnel lining. Image stitching was used to obtain larger images from smaller captured images, whereas the quality of stitching is not evaluated. Finally, Deng et al. [56] used images from the Beijing metro tunnel and Xiaoyadaokou tunnel to stitch the images of tunnel surfaces.

The method proposed in this paper uses video images from the closed-circuit television (CCTV) cameras obtained in real-world underground conditions. In many cases, such images have poor quality due to low lighting, shadows and hazing, and can have blurry moving objects. Moreover, the images can have overlapping areas and camera angle misalignments and parallax.

The proposed method addresses the main requirement of processing such videos in real time (we achieved 21 fps image processing speed), thus allowing the adoption of the method for real-time remote surveillance and intelligent security monitoring applications for the safe operation of underground mines. As such, our method contributes to the development of digital mine infrastructure [67,68].

## 4. Conclusions

We have proposed a hybrid video image stitching algorithm for coal mine video surveillance. The corner detection operator is used to find the feature points, and the nearest neighbor algorithm based on the k-dimensional (KD) tree is used to match the feature points. In order to remove the mismatches, the random sampling consistent (RANSAC) algorithm is employed to filter and purify, and the transformation model parameters can be effectively obtained. Thus, the gradual in and out image fusion method is used to fuse the image without obviously visible image seams, which is more convenient for viewing by human personnel. The proposed method reduces the image stitching time and solves the problem of feature re-extraction due to the change of observation angle thus optimizing the video stitching process. The count of feature points is reduced, which reduces the calculation amount and speeds up the image stitching, while the influence of noise on the image can be largely ignored during the detection of feature points. The experimental results on the real-world mine video sequences show that the presented video stitching method can stitch 21 frames per second, effectively meeting the requirements for real-time, and the stitching effect has a good stability, consistency, and applicability.

**Author Contributions:** Conceptualization, Z.B. and Y.L.; methodology, Z.B. and M.W.; software, X.C., and T.Y.; validation, Z.B., W.W. and M.W.; formal analysis, Z.B., W.W., and M.W.; investigation, Y.L., Z.B., X.C., T.Y., W.W., M.W. and R.D.; resources, Z.B. and W.W.; data curation, X.C., and T.Y.; writing—original draft preparation, Z.B., and W.W.; writing—review and editing, Z.B., M.W. and R.D.; visualization, Y.L., Z.B. and R.D.; supervision, Z.B.

and W.W.; project administration, Z.B. and Y.L.; funding acquisition, Z.B., Y.L. and W.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (Grant No. 61941112, 61761042, 61763046), Natural Science Foundation of China of Shaanxi (No. 2020JM-556), Key Research and Development Program of Yan'an (Grant No. 2017KG-01, 2017WZZ-04-01), Key Research and Development Program of Shaanxi Province (No. 2018ZDXM-GY-036) and Shaanxi Key Laboratory of Intelligent Processing for Big Energy Data (No. IPBED7, IPBED10).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. IEA. *Coal Information 2019*; IEA: Paris, France, 2019; Available online: <https://www.iea.org/reports/coal-information-2019> (accessed on 6 August 2020).
2. Wang, X.; Meng, F. Statistical analysis of large accidents in China's coal mines in 2016. *Nat. Hazards* **2018**, *92*, 311–325. [[CrossRef](#)]
3. Jo, B.W.; Khan, R.M.A. An Event Reporting and Early-Warning Safety System Based on the Internet of Things for Underground Coal Mines: A Case Study. *Appl. Sci.* **2017**, *7*, 925.
4. Zhang, F.; Xu, Z.; Chen, W.; Zhang, Z.; Zhong, H.; Luan, J.; Li, C. An Image Compression Method for Video Surveillance System in Underground Mines Based on Residual Networks and Discrete Wavelet Transform. *Electronics* **2019**, *8*, 1559. [[CrossRef](#)]
5. Wei, W.; Xia, X.; Wozniak, M.; Fan, X.; Damaševičius, R.; Li, Y. Multi-sink distributed power control algorithm for cyber-physical-systems in coal mine tunnels. *Comput. Netw.* **2019**, *161*, 210–219. [[CrossRef](#)]
6. Singh, A.; Kumar, D.; Hötzel, J. IoT based information and communication system for enhancing underground mines safety and productivity: Genesis, taxonomy and open issues. *Ad Hoc Netw.* **2018**, *78*, 115–129. [[CrossRef](#)]
7. Kumar, D. Application of modern tools and techniques for mine safety disaster management. *J. Inst. Eng. Ser. D* **2016**, *97*, 77–85. [[CrossRef](#)]
8. Dong, G.; Wei, W.; Xia, X.; Woźniak, M.; Damaševičius, R. Safety risk assessment of a pb-zn mine based on fuzzy-grey correlation analysis. *Electronics* **2020**, *9*, 130. [[CrossRef](#)]
9. Zhou, B.; Duan, X.; Ye, D.; Wei, W.; Woźniak, M.; Damaševičius, R. Heterogeneous image matching via a novel feature describing model. *Appl. Sci.* **2019**, *9*, 4792. [[CrossRef](#)]
10. Zhou, B.; Duan, X.; Wei, W.; Ye, D.; Wozniak, M.; Damasevicius, R. An adaptive local descriptor embedding Zernike moments for image matching. *IEEE Access* **2019**, *7*, 183971–183984. [[CrossRef](#)]
11. Wei, L.; Zhong, Z.; Lang, C.; Yi, Z. A survey on image and video stitching. *Virtual Real. Intell. Hardw.* **2019**, *1*, 55–83. [[CrossRef](#)]
12. He, B.; Yu, S. Parallax-Robust Surveillance Video Stitching. *Sensors* **2016**, *16*, 7. [[CrossRef](#)] [[PubMed](#)]
13. Zhu, M.; Wang, W.; Liu, B.; Huang, J. Efficient Video Panoramic Image Stitching Based on an Improved Selection of Harris Corners and a Multiple-Constraint Corner Matching. *PLoS ONE* **2013**, *8*, e81182. [[CrossRef](#)] [[PubMed](#)]
14. Zhou, B.; Duan, X.; Ye, D.; Wei, W.; Woźniak, M.; Połap, D.; Damaševičius, R. Multi-level features extraction for discontinuous target tracking in remote sensing image monitoring. *Sensors* **2019**, *19*, 4855. [[CrossRef](#)] [[PubMed](#)]
15. Chen, X.; Ma, H.; Wan, J.; Li, B.; Xia, T. Multi-view 3D Object Detection Network for Autonomous Driving. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [[CrossRef](#)]
16. Bonny, M.Z.; Uddin, M.S. Feature-based image stitching algorithms. In Proceedings of the 2016 International Workshop on Computational Intelligence (IWCI), Dhaka, Bangladesh, 12–13 December 2016. [[CrossRef](#)]
17. Adel, E.; Elmoggy, M.; Elbakry, H. Image Stitching based on Feature Extraction Techniques: A Survey. *Int. J. Comput. Appl.* **2014**, *99*, 1–8. [[CrossRef](#)]
18. Zhu, H.; Zou, K.; Li, Y.; Cen, M.; Mihaylova, L. Robust Non-Rigid Feature Matching for Image Registration Using Geometry Preserving. *Sensors* **2019**, *19*, 2729. [[CrossRef](#)]

19. Alomran, M.; Chai, D. Feature-based panoramic image stitching. In Proceedings of the 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), Phuket, Thailand, 13–15 November 2016. [[CrossRef](#)]
20. Ho, T.; Schizas, I.D.; Rao, K.R.; Budagavi, M. 360-degree video stitching for dual-fisheye lens cameras based on rigid moving least squares. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017. [[CrossRef](#)]
21. Yeh, S.-H.; Lai, S.-H. Real-time video stitching. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Honolulu, HI, USA, 21–26 July 2017. [[CrossRef](#)]
22. Babu, V.M.M.; Santha, T. Efficient brightness adaptive deep-sea image stitching using biorthogonal multi-wavelet transform and harris algorithm. In Proceedings of the 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, India, 23–24 June 2017. [[CrossRef](#)]
23. Ruan, J.; Xie, L.; Ruan, Y.; Liu, L.; Chen, Q.; Zhang, Q. Image Stitching Algorithm Based on SURF and Wavelet Transform. In Proceedings of the 7th International Conference on Digital Home (ICDH), Guilin, China, 30 November–1 December 2018. [[CrossRef](#)]
24. Aung, N.L.; Victor, D.K.; Ye, K.Z.; Htet, Z.W. The study of the process of stitching video images in real time. In Proceedings of the IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIconRus), Moscow, St. Petersburg, Russia, 29 January–1 February 2018. [[CrossRef](#)]
25. Lu, Y.; Gao, K.; Zhang, T.; Xu, T. A novel image registration approach via combining local features and geometric invariants. *PLoS ONE* **2018**, *13*, e0190383. [[CrossRef](#)]
26. Nie, Y.; Su, T.; Zhang, Z.; Sun, H.; Li, G. Dynamic Video Stitching via Shakiness Removing. *IEEE Trans. Image Process.* **2018**, *27*, 164–178. [[CrossRef](#)]
27. Fang, F.; Wang, T.; Fang, Y.; Zhang, G. Fast Color Blending for Seamless Image Stitching. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1115–1119. [[CrossRef](#)]
28. Lee, K.; Sim, J. Stitching for multi-view videos with large parallax based on adaptive pixel warping. *IEEE Access* **2018**, *6*, 26904–26917. [[CrossRef](#)]
29. Yang, X.; Liu, Z.; Qiao, H.; Su, J.; Ji, D.; Zang, A.; Huang, H. Graph-based registration and blending for undersea image stitching. *Robotica* **2020**, *38*, 396–409. [[CrossRef](#)]
30. Park, K.; Shim, Y.; Lee, M.; Ahn, H. Multi-frame based homography estimation for video stitching in static camera environments. *Sensors* **2020**, *20*, 92. [[CrossRef](#)] [[PubMed](#)]
31. Zheng, J.; Wang, D.; Geng, Z. Coal Mine Video Data Detail Enhancement Algorithm Based on L0 Norm and Low Rank Analysis. *Eur. J. Electr. Eng.* **2019**, *21*, 55–60. [[CrossRef](#)]
32. Li, D.; Qian, J.; Liu, Z.; Yang, P. Stitching technology of coal mine video with complex environment. *J. China Coal Soc.* **2011**, *36*, 878–884.
33. Xu, M. Comparison and research of fisheye image correction algorithms in coal mine survey. *Iop Conf. Ser. Earth Environ. Sci.* **2019**, *300*. [[CrossRef](#)]
34. Kim, B.; Choi, K.; Park, W.; Kim, S.; Ko, S. Content-preserving video stitching method for multi-camera systems. *IEEE Trans. Consum. Electron.* **2017**, *63*, 109–116. [[CrossRef](#)]
35. Liu, Q.; Su, X.; Zhang, L.; Huang, H. Panoramic video stitching of dual cameras based on spatio-temporal seam optimization. *Multimed. Tools Appl.* **2018**. [[CrossRef](#)]
36. Li, J.; Wang, Z.; Lai, S.; Zhai, Y.; Zhang, M. Parallax-tolerant image stitching based on robust elastic warping. *IEEE Trans. Multimed.* **2018**, *20*, 1672–1687. [[CrossRef](#)]
37. Chen, J.; Wan, Q.; Luo, L.; Wang, Y.; Luo, D. Drone image stitching based on compactly supported radial basis function. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4634–4643. [[CrossRef](#)]
38. Kang, L.; Wei, Y.; Jiang, J.; Xie, Y. Robust cylindrical panorama stitching for low-texture scenes based on image alignment using deep learning and iterative optimization. *Sensors* **2019**, *19*, 5310. [[CrossRef](#)]
39. Kang, J.; Kim, J.; Lee, I.; Kim, K. Minimum Error Seam-Based Efficient Panorama Video Stitching Method Robust to Parallax. *IEEE Access* **2019**, *7*, 167127–167140. [[CrossRef](#)]
40. Krishnakumar, K.; Gandhi, S.I. Video stitching using interacting multiple model based feature tracking. *Multimed. Tools Appl.* **2019**, *78*, 1375–1397. [[CrossRef](#)]
41. Krishnakumar, K.; Indira Gandhi, S. Video stitching based on multi-view spatiotemporal feature points and grid-based matching. *Visual Comput.* **2019**. [[CrossRef](#)]
42. Lin, L.; Ding, Y.; Wang, L.; Zhang, M.; Li, D. Line-preserving video stitching for asymmetric cameras. *Multimed. Tools Appl.* **2019**, *78*, 14591–14611. [[CrossRef](#)]

43. Du, C.; Yuan, J.; Dong, J.; Li, L.; Chen, M.; Li, T. GPU based parallel optimization for real time panoramic video stitching. *Pattern Recognit. Lett.* **2020**, *133*, 62–69. [[CrossRef](#)]
44. Kakli, M.U.; Cho, Y.; Seo, J. Parallax-tolerant video stitching with moving foregrounds. In *Asian Conference on Pattern Recognition, ACPR 2019; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; Volume 12047*, pp. 625–639. [[CrossRef](#)]
45. Nazaré, T.S.; da Costa, G.B.P.; Contato, W.A.; Ponti, M. Deep Convolutional Neural Networks and Noisy Images. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications; Springer: Cham, Switzerland, 2018; pp. 416–424.* [[CrossRef](#)]
46. Algan, G.; Ulusoy, I. Image Classification with Deep Learning in the Presence of Noisy Labels: A Survey. *arXiv* **2019**, arXiv:1912.05170.
47. Moravec, H.P. Towards Automatic Visual Obstacle Avoidance. In Proceedings of the 5th International Joint Conference on Artificial Intelligence, Cambridge, MA, USA, 22 August 1977; p. 584.
48. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the International Conference on Computer Vision, 2, Kerkyra, Corfu, Greece, 20–25 September 1999; pp. 1150–1157. [[CrossRef](#)]
49. Dongmei, W.; Siqi, Z. Research on image enhancement algorithm of coal mine dust. In Proceedings of the 2018 International Conference on Sensor Networks and Signal Processing, SNSP 2018, Xi'an, China, 28–31 October 2018; pp. 261–265.
50. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353. [[CrossRef](#)]
51. Liu, S.; Rahman, M.A.; Wong, C.Y.; Lin, S.C.F.; Jiang, G.; Kwok, N. Dark channel prior based image de-hazing: A review. In Proceedings of the 5th International Conference on Information Science and Technology (ICIST), Istanbul, Turkey, 21–23 March 2015. [[CrossRef](#)]
52. Lee, S.; Yun, S.; Nam, J.-H.; Won, C.S.; Jung, S.-W. A review on dark channel prior based image dehazing algorithms. *J. Image Video Proc.* **2016**, *4*. [[CrossRef](#)]
53. Li, Y.; You, S.; Brown, M.S.; Tan, R.T. Haze visibility enhancement: A Survey and quantitative benchmarking. *Comput. Vis. Image Underst.* **2017**, *165*, 1–16. [[CrossRef](#)]
54. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
55. Chaiyasarn, K.; Kim, T.-K.; Viola, F.; Cipolla, R.; Soga, K. Distortion-Free Image Mosaicing for Tunnel Inspection Based on Robust Cylindrical Surface Estimation through Structure from Motion. *J. Comput. Civ. Eng.* **2016**, *30*, 04015045. [[CrossRef](#)]
56. Deng, F.; Yang, J. Panoramic Image Generation Using Centerline- Constrained Mesh Parameterization for Arbitrarily Shaped Tunnel Lining. *IEEE Access* **2020**, *8*, 7969–7980. [[CrossRef](#)]
57. Guo, H.; Liu, S.; He, T.; Zhu, S.; Zeng, B.; Gabbouj, M. Joint video stitching and stabilization from moving cameras. *IEEE Trans. Image Process* **2016**, *25*, 5491–5503. [[CrossRef](#)] [[PubMed](#)]
58. Nabil, S.; Balzarini, R.; Devernay, F.; Crowley, J. Designing Objective Quality Metrics for Panoramic Videos based on Human Perception. In Proceedings of the Irish Machine Vision and Image Processing Conference, IMVIP 2018, Ulster, UK, 29–31 August 2018; pp. 189–192.
59. Yoon, J.; Lee, D. Real-Time Video Stitching Using Camera Path Estimation and Homography Refinement. *Symmetry* **2018**, *10*, 4. [[CrossRef](#)]
60. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
61. Zhao, C.; Zhang, H.; Chen, J.; Fu, W. Region-based parallax-tolerant image stitching. In Proceedings of the Tenth International Conference on Graphics and Image Processing (ICGIP 2018), Chengdu, China, 12–14 December 2018. [[CrossRef](#)]
62. Zhu, Z.-H.; Fu, J.-Y.; Yang, J.-S.; Zhang, X.-M. Panoramic Image Stitching for Arbitrarily Shaped Tunnel Lining Inspection. *Comput. Aided Civ. Infrastruct. Eng.* **2016**, *31*, 936–953. [[CrossRef](#)]
63. Huang, H.; Li, Q.; Zhang, D. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. *Tunn. Undergr. Space Technol.* **2018**, *77*, 166–176. [[CrossRef](#)]
64. Kim, C.N.; Kawamura, K.; Shiozaki, M.; Tarighat, A. An image-matching method based on the curvature of cost curve for producing tunnel lining panorama. *J. JSCE* **2018**, *6*, 78–90. [[CrossRef](#)]

65. Konishi, S.; Imaizumi, N.; Enokidani, Y.; Nagaya, J.; Machijima, Y.; Akutagawa, S.; Murakoshi, K. Effective water leakage detection by using an innovative optic fiber sensing for aged concrete lining of urban metro lines in Tokyo. In *Proceedings of the Tunnels and Underground Cities: Engineering and Innovation meet Archaeology, Architecture and Art, Naples, Italy, 3–9 May 2019*; CRC Press: Boca Raton, FL, USA, 2019; pp. 2383–2392. [[CrossRef](#)]
66. Zhao, S.; Zhang, D.M.; Huang, H.W. Deep learning–based image instance segmentation for moisture marks of shield tunnel lining. *Tunn. Undergr. Space Technol.* **2020**, *95*, 103156. [[CrossRef](#)]
67. Liu, Y.; Song, J. Using the internet of things technology constructing digital mine. In *Procedia Environmental Sciences, 10(PART B)*; Elsevier BV: Amsterdam, The Netherlands, 2011; pp. 1104–1108. [[CrossRef](#)]
68. Wang, J.; Wang, Y.; Fu, J. Crucial technology research and demonstration of digital mines. *J. China Coal Soc.* **2016**, *41*, 1323–1331. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).