ORIGINAL ARTICLE

Image stitching based on human visual system and SIFT algorithm

Jindong Zhang[1,2] · Ying Xiu[1]

Abstract

The image stitching process often produces many undesirable effects. Solving problems such as discontinuity and dislocation of pictures has always been the focus of people's research. From the perspective of human vision, this dislocation situation can be easily perceived and found. In this paper, we propose a stitching strategy based on the human visual system (HVS) and scale-invariant feature transform (SIFT) algorithm. We preprocess the brightness difference and contrast of the stitched images, combining SIFT algorithm and HVS to divide the overlapping areas of the stitched images and establish an attribute relationship model. We use dynamic programming to find the optimal seamline according to the attribute relationship model, and the final result makes the optimal seamline almost invisible under the discriminative vision of human eyes. The experimental results show that our method has more advantages in the HVS.

Keywords Image stitching · Human visual system · Image fusion · Optimal seamline detection

1 Introduction

Image stitching technology is widely used. When using the ordinary camera to capture images, due to the limited shooting angle, it is impossible to take all elements in the scene, and the use of a panoramic camera has to face the objective phenomenon of expensive cost. To broaden the perspective of ordinary cameras, people have conducted in-depth research on image stitching technology and achieved some results and applied in various fields, including autonomous driving [1, 2], geoscience [3–5], electrical engineering [6], virtual reality, and other fields [7].

Since it is difficult to achieve consistent exposure of the original images for image stitching, there will be a discontinuous boundary, misalignment, ghosting, and many other factors that affect the stitching effect. The image fusion process is the key to solving these problems. Image fusion algorithms can split into two directions: the method based on smooth transition [8–10] and to find the optimal seamline. The former way eliminates artifacts by aligning the images as much as possible, and this class of methods generally divides the image into areas and calculates the corresponding homography matrix. Spatial variations are distorted over these areas to align the overlapping regions, resulting in a significantly reduced in artifacts. The latter method achieves the final result by optimizing the costs associated with the seams and sewing the areas on either side of the seamline together. Finding the optimal seamline method can avoid complex algorithms and optimize blurring, image misalignment, and other problems [11, 12], but the stitching performance decreases sharply when the number of feature matches is very little [13].

Many scholars have studied the optimal seamline method. Kerschner proposed the "twin snakes technique" method [14] and defined the energy as the sum of the mismatched pixels on the line. Li et al. fused information about the color, gradient magnitude, and texture complexity of the images into the data. They used a new multi-frame joint optimization strategy to find the seamline in multiple overlapping images at once [15]. An optimal seamline detection method based on a CNN-based semantic image segmentation and graph cut energy minimization framework was proposed by Li et al. [16]. Hejazifar et al. proposed FARSE, a fast and robust seam estimation method, to avoid visible seams and ghosting by defining grayscale weighted distances and gradient-domain difference regions [17]. Lin et al. introduced a new structure-preserving warping method to improve the effectiveness of stitching images with large parallax [18]. Li

✉ Jindong Zhang
zhangjindong_100@163.com

1   College of Computer Science and Technology, Jilin
    University, Changchun 130012, China

2   Key Laboratory of Symbol Computation and Knowledge
    Engineering of the Ministry of Education, Jilin University,
    Changchun 130012, China

et al. generated large-scale orthophotos by mosaicking multiple orthophotos, enabling the generation of a high-quality seamline network with fewer artifacts [19]. Zhang et al. found the optimal seamline for unmanned aerial vehicle (UAV) images based on the improved energy function by introducing optical flow into the energy function [20]. These methods have achieved good results in different fields. While optimizing and improving the quality of stitching images, the human visual system (HVS) is of great significance in evaluating the effect of image stitching. As early as 2017, HVS has proven to assess the quality of image stitching [21]. Li et al. proposed a new human perception-based stitching method [22], which considered the nonlinearity and inhomogeneity of human perception as energy minimization, finding seams that are more compatible with the HVS.

We propose an image stitching method based on the HVS and scale-invariant feature transform (SIFT) algorithm [23]. Find the optimal seamline to complete the stitching task. In the experiment, we evaluate the performance of image stitching. The experimental results show that our method conforms to the characteristics of the HVS, and the quality of the image stitching is greatly improved. The main contributions of our work are summarized as follows.

We propose an image stitching method based on optimal seamline, which is based on the SIFT algorithm and the HVS that quantifies the preprocessed image to find the optimal seamline, and finally uses a multi-scale fusion algorithm to make the seamline almost invisible.
We build an attribute relationship model based on the HVS to connect the properties of the HVS to make it more suitable for our method.

The remaining structure of this paper is organized as follows: Sect. 2 describes the related work, Sect. 3 introduces the methods used in this paper, Sect. 4 discusses and analyzes the experimental results, and Sect. 5 summarizes the whole paper.

2 Related work

2.1 Image stitching

The core goal of image stitching is to align the overlapping areas of images [10]. Image stitching consists of three parts: image preprocessing, image registration, and image fusion [24]. Many scholars have optimized the stitching process [25–29]. Recently, Jia et al. used the consistency of lines and points to preserve the linear structure in the stitching process to improve the image stitching quality [30]. Liao and Li proposed two types of single-perspective warping for natural-image stitching to reduce projection distortion [31].

The quality of image stitching is also closely related to the quality of the images [32]. External disturbances such as lens distortion, photorefractive effect, exposure, and differences in brightness are bound to be present when taking pictures [33]. Under these factors, there are many mismatches when feature point detection and feature matching are performed on images, and it further leads to low accuracy of the homography matrix used for stitching and shows a lot of image discontinuity problems [34]. These discontinuities are exactly the visual attention points of people, and how to avoid the seam problem to improve the image stitching quality is the key to the research.

2.2 Human visual system

The nervous system regulates human eye activity. Human perception of images is influenced by both physiological and psychological factors. The HVS, as an image processing system, perceives images non-uniformly and nonlinearly. For images, the main characteristics of the HVS are generally expressed in three aspects: brightness, frequency domain, and image type characteristics. The brightness characteristic is one of the most fundamental characteristics of the HVS, mainly about the sensitivity of the human eyes to changes in brightness. The human eyes are less sensitive to the noise attached to the region of high luminance. When the brightness of the area in the image is relatively high, the human eye is not sensitive to the change in gray value [35], so people can easily ignore the detail part in the background [36]. If the discontinuous edges of the stitched image are in high-attention areas of the human eyes, it will harm image quality. When the discontinuous edges are in the masked background, it will be difficult for the human eye to observe the undesirable areas, and image quality will improve.

3 Method

The process of image stitching is shown in Fig. 1. First, we compensate for image brightness, contrast, and saturation, aiming to remove significant brightness differences that may affect image registration and image fusion. Then, the processed image is extracted and registered features by SIFT algorithm and random sample consensus (RANSAC) algorithm to obtain the corresponding homography matrix [37]. The overlapping images are visually quantized, and the quantized characteristics include the following four types: brightness characteristic, brightness difference, masking characteristic, and visual attention. Based on the above, we establish the attribute relationship model and reference the edge detection result to find the candidate area of the optimal seamline. Adjust the parameters of the attribute relationship model to select the optimal seamline, at the
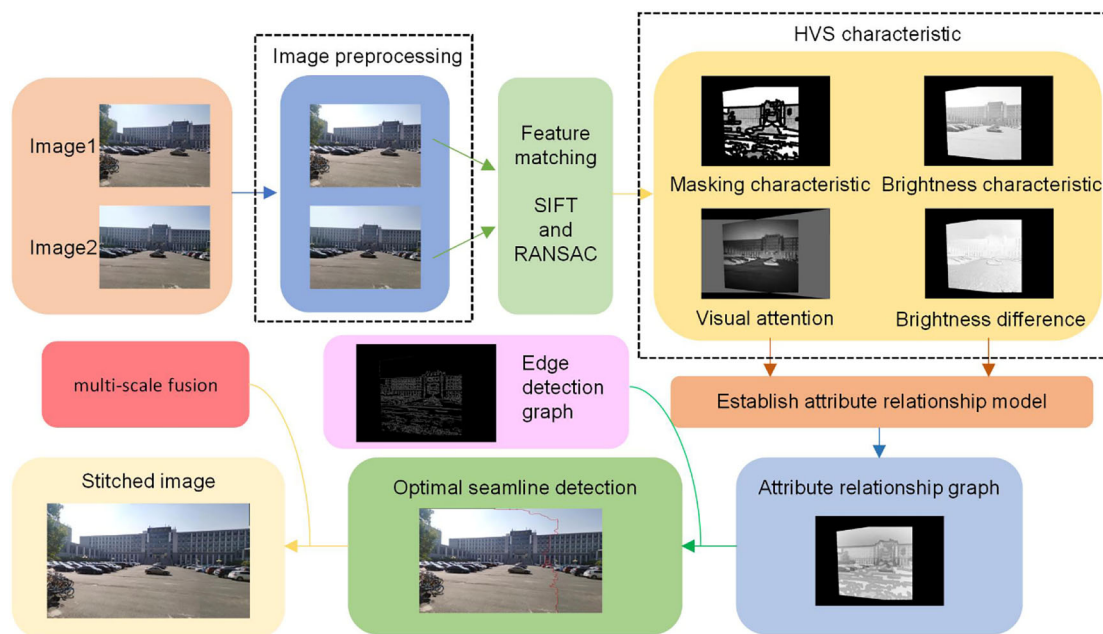
Fig. 1 Proposed image stitching method workflow

end fusion to get a seamless stitched image. These will be described in detail in later chapters.

3.1 Image preprocessing

In order to make the stitched image without a large color difference, we analyze the brightness, saturation, and contrast and compensate for the image with poor visual effect in the two images during data preprocessing. To make the brightness saturation contrast of the two images basically consistent under the visual effect of the human eye. And it does not affect the focus information of the image.

Set the original pixel grayscale as $f(i, j)$ and the transformed pixel grayscale as $g(i, j)$, and apply the linear transformation Eq. (1) to adjust the brightness contrast, where the coefficient con affects the image contrast and the coefficient lumin affects the brightness of the image.

$$g(i, j) = \text{con} * f(i, j) + lu \min \tag{1}$$

Adjust the parameters of another image according to one of the images, until the two images have an overlapping part of the brightness, the contrast is basically the same. The higher the saturation, the fuller the color. We fine-tune the saturation of the two images by adjusting them as follows.

First set a parameter p, which takes values from -100 to 100 and normalizes it to take values between -1 and 1. Calculate the maximum value LMax and the minimum value LMin for the RGB three-channel to obtain the values of Para1 and Para2, as shown in Eq. (2–3).

$$\text{Para1} = (L\text{Max} - L\text{Min})/255 \tag{2}$$

$$\text{Para2} = (L\text{Max} + L\text{Min})/255 \tag{3}$$

If LMax and LMin are the same, it indicates a gray point and processes the next pixel. The L and S in hue, saturation, lightness (HSL) color mode are calculated by Para2, as shown in Eq. (4–5).

$$L = (L\text{Max} - L\text{Min}/510) \tag{4}$$

$$S = \begin{cases} \dfrac{\text{Para1}}{\text{Para2}}, & L < 0.5 \\[2ex] \dfrac{\text{Para1}}{2 - \text{Para2}}, & L \geq 0.5 \end{cases} \tag{5}$$

At this point will be divided into two cases according to the p value, when $p \geq 0$, which means an increase in the color saturation, then the Para3 value is obtained from Eq. (6). Set the value of L*255 as K. The adjusted image RGB three-channel value M can be calculated from Eq. (7).

$$\text{Para3} = \begin{cases} S, & p + S \geq 0 \\ 1 - p, & p + S < 0 \end{cases} \tag{6}$$

$$M = M + (M - K) * \text{Para3} \tag{7}$$

If $p < 0$, reduce the color saturation, Para3 $= p$, adjusted the image RGB value M as shown in Eq. (8).

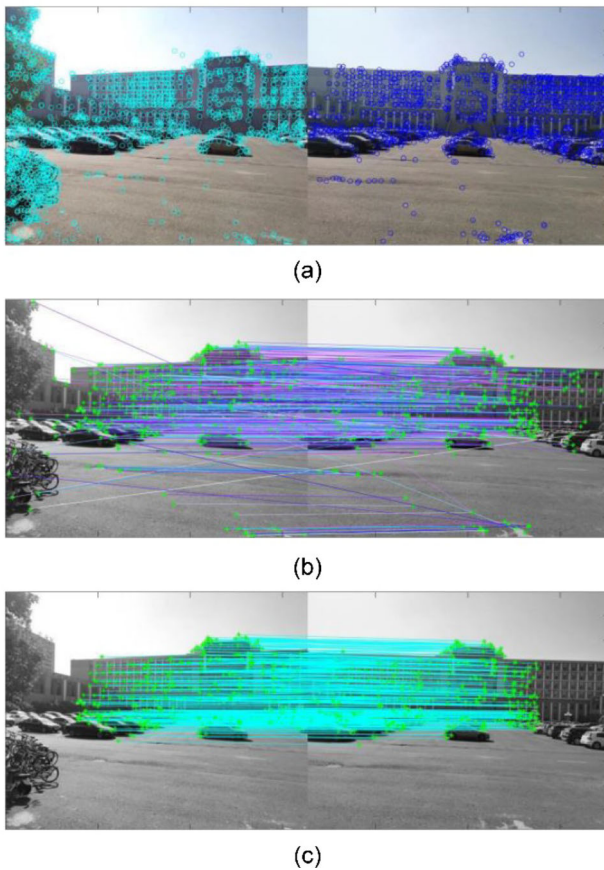$$M = K + (M - K) * (1 - \text{Para3}) \tag{8}$$

Fig. 2 Results of SIFT and RANSAC feature extraction: **a** key point detection, **b** SIFT key point matching, **c** RANSAC removing mismatched key points

The compensation results of brightness contrast and saturation can be seen in Fig. 1.

3.2 SIFT and RANSAC feature extraction and registration

To determine the homography matrix, we choose the SIFT and RANSAC algorithms to find and match feature points on images. The RANSAC algorithm is widely used in computer vision and mathematics fields, such as straight-line fitting, plane fitting, and computing transformation matrices between images or point clouds. Feature extraction and registration include the following steps. Build scale space and detect key points, determine the cardinal direction of key points and describe them, match key points and eliminate mismatching key points. Figure 2 shows the results of feature extraction and matching.

3.3 Human visual system characteristic

In 3.3 we will introduce the relevance of the four concepts of brightness characteristic, brightness difference, masking

characteristic, and visual attention in HVS to image stitching quality.

3.3.1 Brightness characteristic

The visual perception of brightness in humans is usually poor in the absolute brightness perception of objects and extremely sensitive to the relative difference in brightness, which is called high contrast sensitivity in academia. There are generally two definitions of the luminance contrast between an object and its surrounding background: Weber contrast [38] and Michelson contrast. The definition of Weber contrast C_{web} is shown in Eq. (9). L_b and L, respectively, the brightness value and background brightness value of the object.

$$C_{\text{web}} = \frac{L - L_b}{L} \tag{9}$$

Human eye brightness characteristics is a kind of study of the relationship between the brightness of the object and the brightness of human subjective perception. Existing research has proved that human perception of image brightness and object brightness is presented in the form of a logarithmic function. According to Weber's law it can be obtained: Human subjective perception of brightness is related to the brightness L of the image, as shown in Eq. (10), where T is constant, which is related to the average value of the brightness of the entire image, $t' = T \ln 10$, T_0 is a constant, L represents the objective brightness value. The discontinuous discrimination rate of the human eye in the dark is much higher than the resolution in a bright place, so positioning the image stitching in the bright area can significantly improve the stitching effect of the image. We define the brightness characteristic as BC.

$$\text{BC} = T \ln L + T_0 = t' \lg L + T_0 \tag{10}$$

3.3.2 Brightness difference

Differential Eq. (10) yields Eq. (11), and it can be seen that the difference in brightness conforms to the linear law with the change of actual brightness. Among them, $d(\text{BC})$ and dL represent subjective brightness and objective brightness, respectively. At the same time, we define the brightness difference as BD, as shown in Eq. (12).

$$d(\text{BC}) = T \frac{dL}{L} \tag{11}$$

$$\text{BD} = 255 - |(\text{average}(L_1) - \text{average}(L_2)| \tag{12}$$

where L_1 and L_2 represent the brightness of pixels with the same coordinates in the overlapping areas of the two

images to be stitched, respectively. The averange function is used to calculate the average brightness of pixels adjacent to the current pixel. Through the previous analysis, it can be obtained that compared with the high brightness difference area, the stitching in the low brightness difference area can obtain better stitching quality.

3.3.3 Masking characteristics

Masking characteristics is an important feature of the HVS, and masking characteristics play an important role in image processing. When there are multiple stimuli, the interaction between the stimuli will cause some of them to be unperceivable, especially when the characteristics of the stimuli are similar to the environmental characteristics [39], which is the masking characteristic. Visual masking characteristics are generally related to the spatial frequency, direction, and position of the stimuli.

The image regions can be divided into textured, smooth, and other regions. Mismatches in the smooth region have little effect on the results. The more cluttered the edge information in the textured region, the more it helps to improve the quality of the stitched image. The masking characteristics quantization formula is shown in Eq. (13). MC is the masking value for each pixel, where $x_{\max} - x_{\min}$ represents the difference between the maximum and minimum values of the grayscale value of the neighborhood window. α and β are constants, we set them to 0.8 and 1 according to the experience of Cao et al. Except for these two cases, the MC of the pixels in other regions is 0.

$$\text{MC} = \begin{cases} \alpha * (255 - x_{\max} + x_{\min}), & \text{pixels in smooth regions} \\ \beta * H, & \text{pixels in texture regions} \end{cases}$$

$$(13)$$

The textured region can mask the discontinuous edges during image stitching, and the complexity in each region is determined by Eq. (14), where H is the local entropy. l, w are the length and width of the window around the pixel being calculated. n_{ij} represents a grayscale pixel, and $l * w$ is the neighborhood window of the pixel.

$$H = -\sum_{i=0}^{l-1} \sum_{j=0}^{w-1} \frac{n_{ij}}{l * w} * \left(\log n_{ij} - \log(l * w) \right) \qquad (14)$$

3.3.4 Visual attention

When people observe images, they divide the images into different areas due to psychological factors, treat these areas separately, and sometimes focus on only a part of them. When the image is distorted in the areas the human eye focuses on,

distortion is more noticeable than in other areas. Every pixel of an image has a unique salience value, and pixels with higher significance have a greater impact on image quality. If the seamline is in an area with high visual attention, it will affect the quality of stitched images.

We choose the SDSP model proposed by Zhang et al. [40] to calculate the visual attention value VA of pixels, and the definition is shown in Eq. (15–18).

$$\text{VA}(x) = V_F(x) * V_D(x) * V_C(x) \qquad (15)$$

$V_F(x)$ is a saliency plot modeled by band-pass filtering, defining the image as $p(x)$, converting it to CIEL*a*b* opponent color space, and the resulting three channels are represented by $p_L(x)$, $p_a(x)$, and $p_b(x)$. $f(x)$ is the transfer function of the logarithmic filter Gabor.

$$V_F(x) = \sqrt[2]{(p_L * f)^2 + (p_a * f)^2 + (p_b * f)^2} \qquad (16)$$

$V_c(x)$ is color significant, θ_c is a parameter, $p_{an}(x)$ is a linear mapping of $p_a(x)$, and $p_{bn}(x)$ is a linear mapping of $p_b(x)$.

$$V_c(x) = 1 - e^{\left(\frac{p_{bn}^2(x) - p_{an}^2(x)}{\theta_C} \right)} \qquad (17)$$

$V_D(x)$ is location significant, and studies have shown that objects near the center of the image are more attractive to people, center is the center of the image $p(x)$. θ_D is a parameter.

$$V_D(x) = e^{\left(-\frac{\|x - \text{center}\|_2^2}{\theta_D} \right)} \qquad (18)$$

Selecting nonsignificant areas according to the VA model to find the optimal seamline can lead to better results.

3.4 Establish attribute relationship model

By image preprocessing we can get two images with similar brightness: image1 and transformed image2. Find the feature points through the SIFT algorithm and the RANSAC algorithm to extract and optimize the feature point pairs, filter out the mismatching results, and set the output homography matrix to H.

According to the analysis of Sect. 3.3, brightness characteristics, brightness difference, masking characteristics, and visual attention all affect the quality of image stitching. According to the magnitude of influence, by building an attribute relationship model to quantify the impact of these four attributes on image stitching, the attribute relationship model is shown in Eq. (19), where μ_1, μ_2, μ_3 are constants and $\mu_1 + \mu_2 + \mu_3 = 0.9$.

$$\text{ARM} = \mu_1 * \text{BC} + 0.1 * \text{BD} + \mu_2 * \text{MC} + \mu_3 * (255 - \text{VA}) \quad (19)$$

We mainly focus on the comprehensive feature of the SDSP model, the value setting of μ_3 will be higher than μ_1, μ_2, according to the image preprocessing results, the two images now have similar brightness, the brightness difference is not obvious, so the influence degree of the brightness difference is set to 0.1, the value of μ_1 is controlled between 0.1 and 0.2, and the splicing is positioned by adjusting the values of μ_2 and μ_3. The main steps are shown in Algorithm 1. We define P as the area where image1 and transformed image2 overlap.

Algorithm 1 Establish attribute relationship model

Input: P

Output: attribute relational graph $\boldsymbol{pic\_hvs}$.

1. *BC*: $\boldsymbol{t' \, lg \, L + T_0}$
2. *BD*: $\boldsymbol{255 - |(averange(L_1) - averange(L_2)|}$
3. *MC*: $\boldsymbol{\alpha * (255 - x_{max} + x_{min})}$,// pixels in smooth regions
   $\boldsymbol{\beta * H}$, // pixels in texture regions
   $\boldsymbol{0}$, // pixels in other regions
4. *VA*: $\boldsymbol{V_F(x) * V_D(x) * V_C(x)}$
5. *ARM*: $\boldsymbol{\mu_1 * BC + 0.1 * BD + \mu_2 * MC + \mu_3 * (255 - VA)}$
6. Process the P region according to *ARM*.

Return $\boldsymbol{pic\_hvs}$

3.5 Optimal seamline selection and image fusion

According to the previous analysis, seamline has the lowest impact on image quality when smooth and texture regions. We use the edge detection algorithm to perform edge detection on the image, according to the obtained edge map, to determine whether the edge pixels are splice line candidates. If the optimal seamline produces a small number of discontinuous strong edges, and the discontinuous weak edges appear in areas with low attention, the optimal seamline is almost invisible.

The main process of the optimal seamline selection algorithm is represented by Algorithm 2. When the two images are aligned, there will be an intersection point on the top edge of the image and the bottom edge of the image, which are the two endpoints of the optimal seamline, and the intersection points of the upper and lower edges are start and end, Find_edistance function is used to find the reciprocal of the Euclidean distance between the two parameters, point is the pixel currently being processed, next_point is the next candidate point, and the optimal seamline is a point set called optimal_seamline.

Algorithm 2 Choose the optimal seamline

Input: attribute relationship model ARM, **start**, **end**.

Output: $\boldsymbol{optimal\_seamline}$.

1. Candidate pixels are selected based on the *ARM*.
2. $\boldsymbol{point \leftarrow start}$, $\boldsymbol{b_i \leftarrow Find\_edistance}$ (candidate pixels $\boldsymbol{i}$ around $\boldsymbol{point}$, $\boldsymbol{end}$) ;
   $\boldsymbol{next\_point} \leftarrow \max(\text{ARM} + \boldsymbol{b_i})$ in candidate pixels
   Add $\boldsymbol{point}$ to the set $\boldsymbol{optimal\_seamline}$.
3. **While** ($\boldsymbol{next\_point}$!= $\boldsymbol{end}$) {
   Add $\boldsymbol{next\_point}$ to the set $\boldsymbol{optimal\_seamline}$.
   $\boldsymbol{point \leftarrow next\_point}$, $\boldsymbol{b_i \leftarrow Find\_edistance}$ (candidate pixels $\boldsymbol{i}$ around $\boldsymbol{point}$, $\boldsymbol{end}$)
   $\boldsymbol{next\_point} \leftarrow \max(ARM + \boldsymbol{b_i})$ in candidate pixels
   }
4. Add the pixel $\boldsymbol{end}$ to the set $\boldsymbol{optimal\_seamline}$
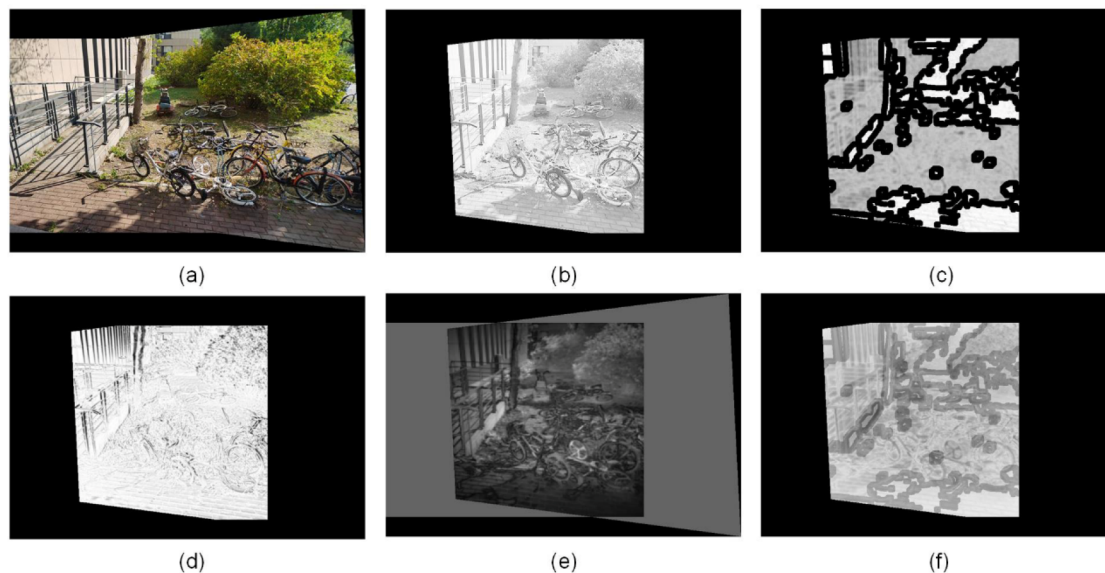
Return $\boldsymbol{optimal\_seamline}$

Fig. 3 HVS attribute graph **a** two images directly stitched **b** brightness characteristic, **c** masking characteristics, **d** brightness difference, **e** visual attention, **f** attribute relationship graph

After finding the optimal seamline for image fusion, divide the overlapping area of the two images into two parts based on the optimal seamline, these two parts are filled by two images separately. Fusing the images according to the Laplace pyramid fusion method. The main steps are as follows. Calculate the Gaussian pyramid and the Laplace pyramid of the input image. Merge the pyramids of Laplace, which are on the same level. Expand the upper pyramid of Laplace until it has the same resolution as the original image, then overlay the image one after the other. Finally, we obtain the output image.

4 Experimental results and analysis

The platform used for this method is Windows 10 on a PC with 3.33 GHz and 16 GB RAM. The program was written in MATLAB. The description of the algorithm proposed in this paper is an image stitching method based on HVS and SIFT algorithm. To verify the effectiveness of the proposed method, we evaluate the quality of image stitching through a series of experiments, the experimental results show that the proposed method can improve the quality of image stitching. We experimented with visually representative images, all of which were taken of natural scenes. We first compare the results under different parameters during the stitching process to find the influence of the parameters on the position of the optimal seamline. Second, we compared the experimental results with the experimental method of Li et al. and the experimental method of Cao et al., the results showed that our experiments were more effective. We also added a subjective evaluation of image quality. Finally, we performed a

limitation analysis of our experiment. To ensure the fairness of the comparison, all test methods use the same matching data and are performed on the same host.

4.1 Self-comparison experiments

Before conducting a comparative experiment, the images of the four HVS properties that appear during the experiment are first introduced.

In Fig. 3, we introduce the attribute relationship graphs of HVS, all of which are processed on the overlap of image1 and transformed image2. As the brightness difference graph shown in Fig. 3d, the white pixels indicate that there is a small brightness difference in the current image position, and the image is preprocessed according to the brightness difference graph, and a small brightness difference is expected in the area near the seamline to achieve the better stitching effect. In Fig. 3e the brightest part of the image is the visual attention point, and people will pay attention to these places first when looking at the image. Figure 3f shows the attribute relationship graph obtained by establishing the attribute relationship model based on the four graphs before this graph.

Next, according to the established attribute relational graph to find the optimal seamline, Fig. 4 shows the difference in seamline position when the three parameters are assigned differently.

In Fig. 4, the most eye-catching thing is the tall building in the middle and the sign on the building, and it can be seen that the position of the seamline is different when the attribute relationship model is different. When only MC or BC is concerned, the seamline will pass through the central
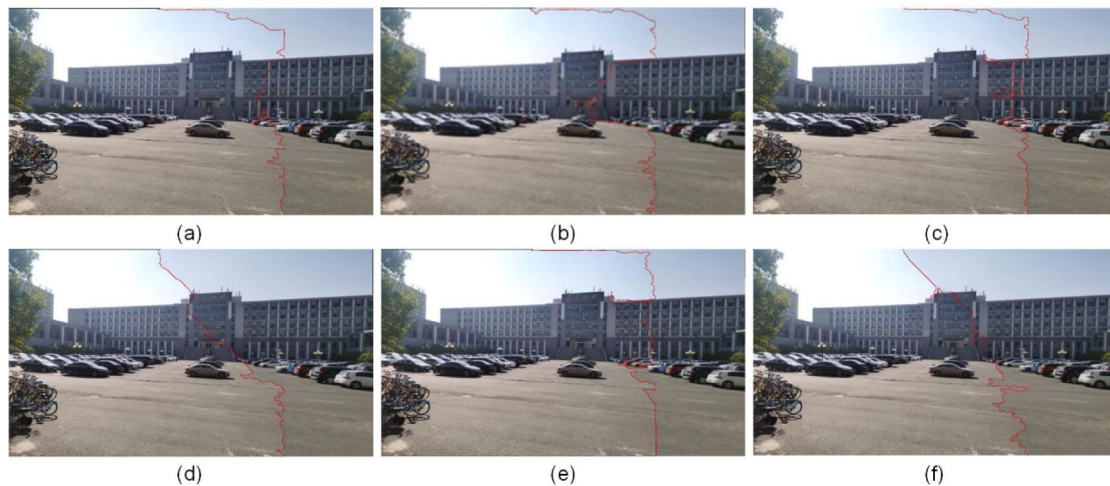
Fig. 4 The influence of parameters on the seamline position. **a** μ_3, μ_2 and μ_1 are 0.55 0.25 0.1, **b** μ_3, μ_2 and μ_1 are 0.5 0.2 0.2, **c** μ_3, μ_2 and μ_1 are 0.5 0.3 0.1, **d** $\mu_2 = 0.9$, **e** $\mu_3 = 0.9$, **f** $\mu_1 = 0.9$

building and the sign on the building, which will lead to poor stitching. When only focusing on the VA, the seamline is close to the central building. Observing the three graphs Figs. 4a, 4b, and 4c, it is found that Fig. 4b passes through the central building, Fig. 4a can avoid the central building, and Fig. 4c approaches the central building. The results of fusing according to different seamline show that the stitching effect of Fig. 4a is the best. The results of stitching according to Fig. 4a can be seen in Fig. 7.

The blind/referenceless image spatial quality evaluator (BRISQUE) algorithm is a reference-free spatial domain image quality assessment algorithm. The larger the score, the worse the quality of the image. We used the BRISQUE algorithm to evaluate the quality of the images under three different attribute relationship graphs, and the results showed that (a) had a result of 25.52, which was 1.15 lower than (b) result and 0.75 lower than (c) result.

In summary, the attribute relationship model influences the final stitching effect, finding the optimal seamline is the key to getting good experimental results.

4.2 Comparative experiment

We used multi-scale fusion to perform the final fusion processing of the image. To more clearly show the influence of image preprocessing and image fusion on the final result, we used the BRISQURE algorithm to compare the stitching results without preprocessing and the preprocessed results. Compare the direct fusion based on seamline with the multi-scale fusion results. It can be seen from Table 1 that preprocessing and multi-scale fusion have a certain improvement in experimental results.

In Figs. 5, 6, 7, 8, and 9, we list the direct fusion of SIFT algorithm, the method of Cao et al., the algorithm of Li et al.,

Table 1 Image quality module research

Dataset	Without preprocessing	Seam-based direct fusion	Ours
Stone	30.6890	30.1828	14.4051
Bike	19.2275	18.3640	11.8176
Building	26.2881	26.4091	25.5269
Stage	36.8865	35.4873	32.2650
Sunset	37.5081	37.8843	29.8874

and the stitching effect of our method under five datasets. From the highlighted content in the figure, it can be seen that SIFT splicing and direct fusion will produce an obvious dividing line. The method of Cao et al. will result in the problem of misalignment and ghosting, and the algorithm of Li et al. may cause a difference in brightness, and produce the problem of dislocation, affecting the visual effect. And our method can effectively avoid the problems of seams, ghosting, and misalignment. In summary, our stitching effect is better.

We give the results of scoring the five groups of images in Figs. 5, 6, 7, 8, and 9 using the BRISQUE algorithm, and the results are shown in Table 2.

In Fig. 10, to more visualize the results of our experiments, we provide the results of nine additional sets of experiments.

As can be seen in Table 2, our algorithm has the best-quality of stitching results. Figure 11 shows that our method scores line is under the other two methods with the best results. We use the BRISQUE algorithm to score the images in the dataset, excluding the failures, the average score of our method is lower than the method of Li et al. and the method of
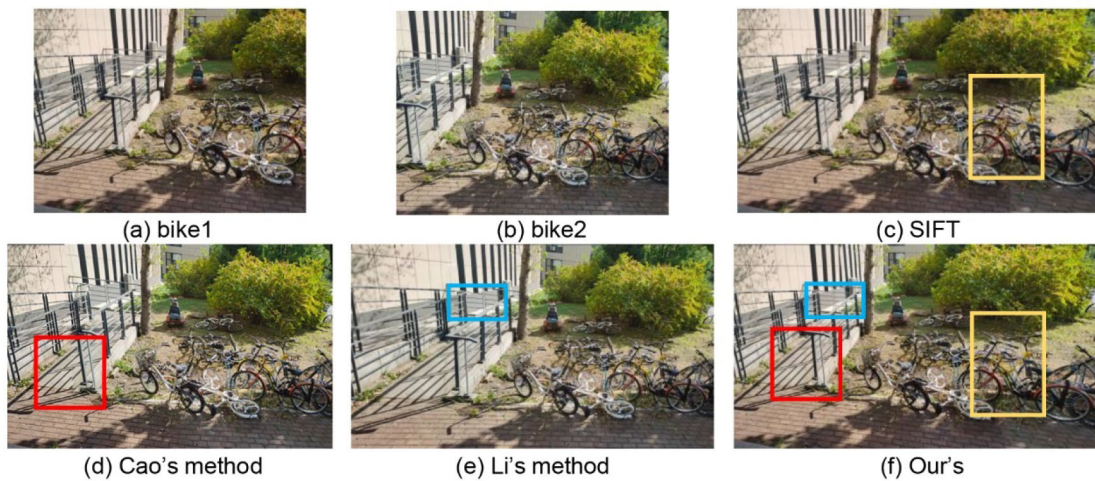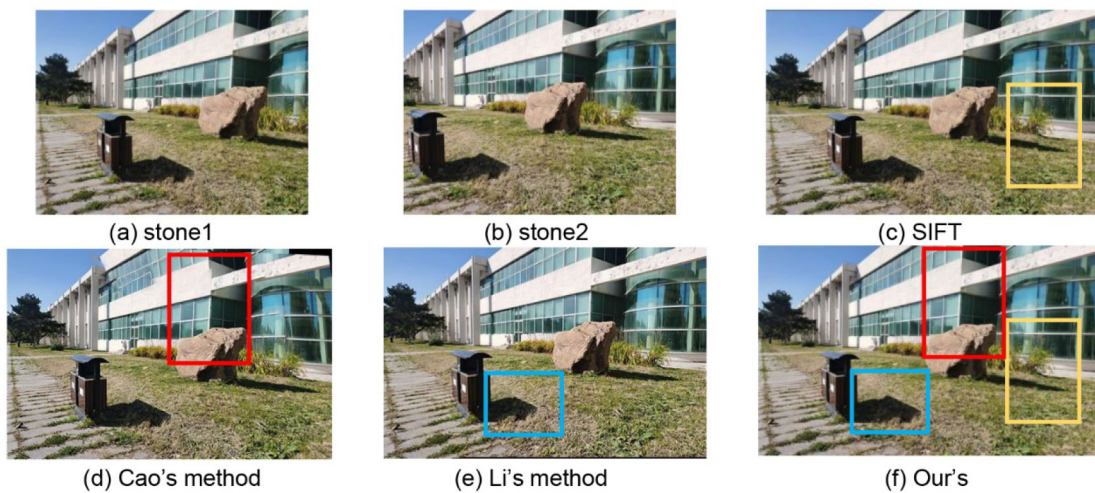
Fig. 5 Stitching effect of datasets bike
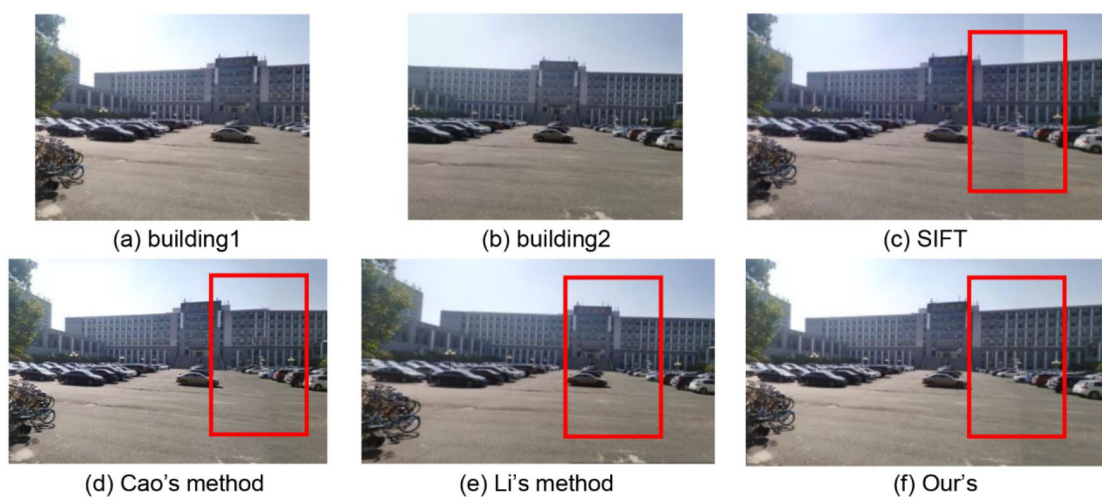

Fig. 6 Stitching effect of datasets stone

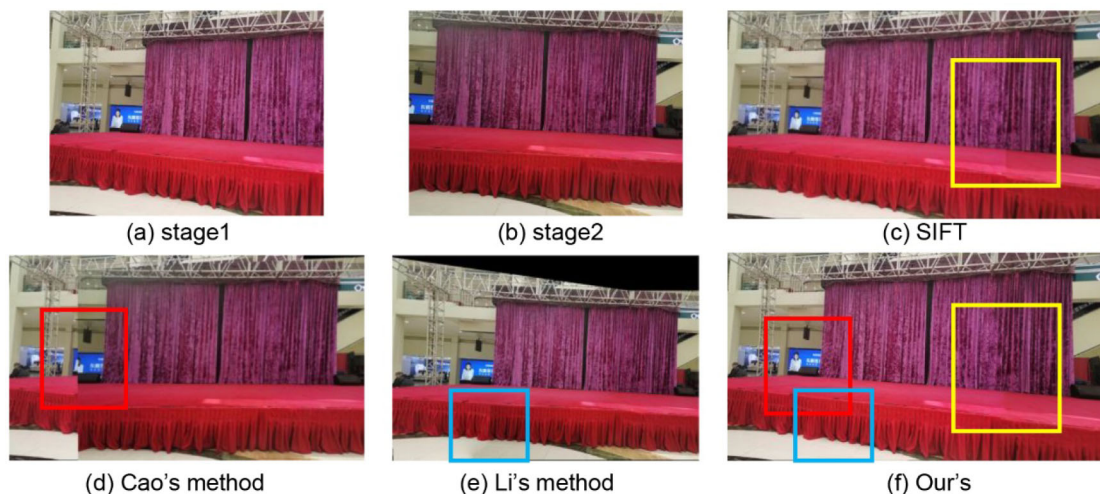**Fig. 7** Stitching effect of datasets building

Fig. 8 Stitching effect of datasets stage

Fig. 9 Stitching effect of datasets sunset

Table 2 BRISQUE algorithm scoring results

Dataset	Li's method	Cao's method	Ours
Stone	27.2629	20.1724	14.4051
Bike	26.3873	19.6064	11.8176
Building	33.0743	35.5858	25.5269
Stage	42.0465	39.7439	32.2650
Sunset	29.6381	36.8405	29.8874

Cao et al., which shows that the quality of our image stitching is high, and the algorithm is better than the other two algorithms.

In this paper, we focus on image stitching under HVS, and subjective evaluation is also an essential part of image quality evaluation. Accordingly, we introduce subjective evaluation. We conducted two user studies to compare the image stitching results of three different methods. We invited 20 participants to evaluate the unlabeled stitching results, including 10 researchers with computer vision backgrounds. We sequentially dropped two input images, our results, and the others' results on a large screen, each time with our results in a random order of placement with the others. The available results to the user are (1) A is better (2) B is better (3) both good (4) both bad. The evaluation results are shown in Fig. 12. It can be seen that our results are more favored by users.

4.3 Limitations

When using the suture-based image stitching method, the image must have a coincident area where the optimal seamline can be found. If the two images of the input have particularly large parallax or the significant structure of the image is very complex, when looking for the optimal seamline in the established attribute relationship graph, there will be no optimal seamline. We provide a failure case in Fig. 13, where the image has a large parallax and high-performance
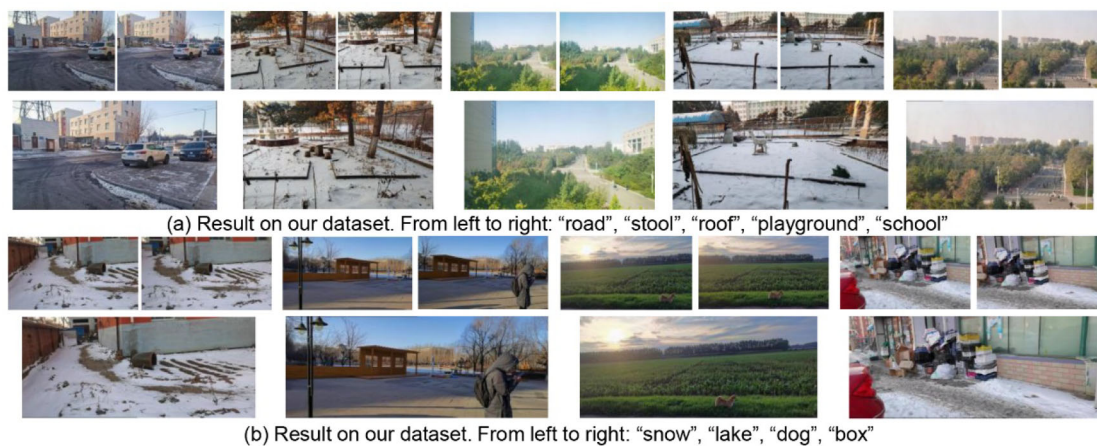
(a) Result on our dataset. From left to right: "road", "stool", "roof", "playground", "school"

(b) Result on our dataset. From left to right: "snow", "lake", "dog", "box"

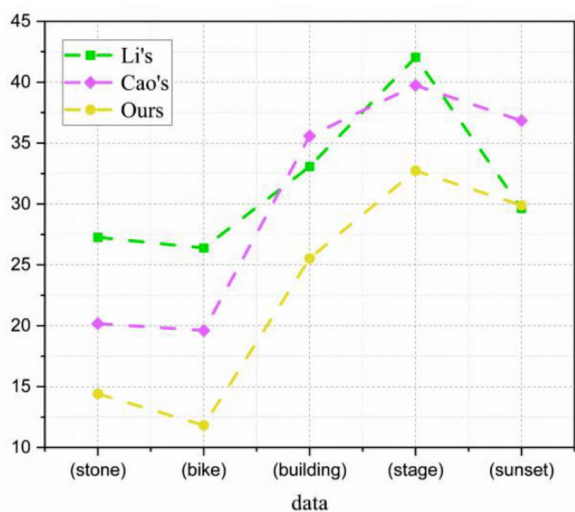**Fig. 10** More results of ours

Fig. 11 BRISQUE algorithm scoring line

5 Conclusion

The method of stitching images combined with SIFT algorithm and HVS proposed in this paper is a good solution to improve the image quality under the human vision. This method quantifies the visual characteristics of the human vision to locate the seamline of two images to be stitched, avoiding high perception area as much as possible. Before using the SIFT algorithm to obtain the homography matrix, preprocessing is used to minimize the brightness difference between the two images, and the saturation and contrast are fine-tuned to make them more suitable for subsequent processing. The next step is to build the attribute relationship model, determine the optimal seamline and perform multi-scale fusion to obtain the final results. Based on a series of comparative experimental results analysis, our method is evaluated and has a superior visual effect and good stitching effect under the human visual effect. In the following work, we will also develop more flexible adaptive brightness preprocessing methods to eliminate the disadvantage of manual

alignment is not possible during the image alignment phase, which directly affects the final result.

Fig. 12 user study on visual quality. The number are shown in percentage and averaged on 20 participants
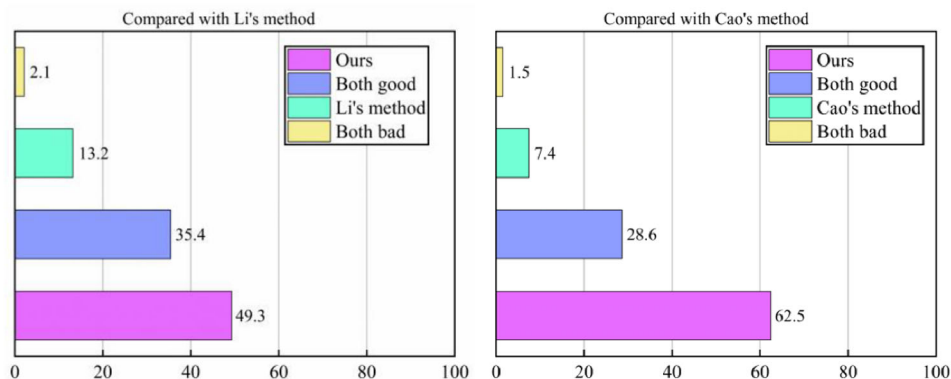
Fig. 13 A failure example. The red rectangle indicates the unsatisfying stitched areas

parameter adjustment for image preprocessing. By introducing machine learning methods, work efficiency can be further improved, which is also the research direction of future work.

Data availability All data, materials generated or used during the study appear in the submitted article.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Wang, L., Yu, W., Li, B.: Multi-scenes image stitching based on autonomous driving. IEEE 4th information technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 694–698 (2020)

2. Hu, F., Bai, L., Li, Y., Tian, Z.: Environmental reconstruction for autonomous vehicle based on image feature matching constraint and score. Pacific Rim International Conference on Artificial Intelligence, pp. 140–148 (2018)

3. Wang, B., Li, H., Hu, W.: Research on key techniques of multi-resolution coastline image fusion based on optimal seam-line. Earth Sci. Inform. **13**, 333–344 (2020)

4. Niu, C., Zhong, F., Xu, S., Yang, C., Qin, X.: Cylindrical panoramic mosaicing from a pipeline video through MRF based optimization. Vis. Comput. **29**, 253–263 (2013)

5. Chen, J., Fu, Z., Huang, J., Hu, X., Peng, T.: Boosting vision transformer for low-resolution borehole image stitching through algebraic multigrid. Vis. Comput. **38**, 3191–3203 (2022)

6. Zhu, C., Ding, W., Zhou, H., Yu, F.: Real-Time image mosaic based on optimal seam and multiband blend. IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), pp. 722–725 (2019)

7. Kim, H.G., Lim, H.-T., Ro, Y.M.: Deep virtual reality image quality assessment with human perception guider for omnidirectional image. IEEE Trans. Circuits Syst. Video Technol. **30**(4), 917–928 (2020)

8. Lee, K.-Y., Sim J.-Y. (2020): Warping residual based image stitching for large parallax. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8195–8203

9. Li, J., Wang, z., Lai, s., Zhai, y., Zhang, m.: Parallax-Tolerant Image Stitching Based on Robust Elastic Warping. IEEE Trans. Multimedia, vol. 20, no. 7, pp. 1672–1687 (2018)

10. Li, J., Deng, B., Tang, R., Wang, Z., Yan, Y.: Local-adaptive image alignment based on triangular facet approximation. IEEE Trans. Image Process. **29**, 2356–2369 (2020)

11. Liu, T., Zhang, J.: An improved path planning algorithm based on fuel consumption. J. Supercomput. **78**, 12973–13003 (2022)

12. Zhang, J., Gao, Y., Xu, Y. Huang, Y., Yu, Y., Shu, X.: A simple yet effective image stitching with computational suture zone. Vis. Comput. (2022)

13. Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.: Unsupervised deep image stitching: reconstructing stitched features to images. IEEE Trans. Image Process. **30**, 6184–6197 (2021)

14. Kerschner, M.: Seamline detection in colour orthoimage mosaicking by use of twin snakes. ISPRS J. Photogramm. Remote. Sens. **56**, 53–64 (2001)

15. Li, L., Yao, J., Lu, X., Tu, J., Shan, J.: Optimal seamline detection for multiple image mosaicking via graph cuts. ISPRS J. Photogramm. Remote. Sens. **113**, 1–16 (2016)

16. Li, L., Yao, J., Liu, Y., Yuan, W., Shi, S., Yuan, S.: Optimal seamline detection for orthoimage mosaicking by combining deep convolutional neural network and graph cuts. Remote Sens. **9**, 701 (2017)

17. Hejazifar, H., Khotanlou, H.: Fast and robust seam estimation to seamless image stitching. Signal Image Video Process **12**(5), 885–893 (2018)

18. Lin, K., Jiang, N., Cheong, LF., Do, M., Lu, J.: SEAGULL: seam-guided local alignment for parallax-tolerant image stitching. Comput. Vis. ECCV, pp. 370–385 (2016)

19. Li, L., Tu, J., Gong, Y., Yao, J., Li, J.: Seamline network generation based on foreground segmentation for orthoimage mosaicking. ISPRS J. Photogramm. Remote Sens. **148**, 41–53 (2019)

20. Zhang, W., Guo, B., Li, M., Liao, X., Li, W.: Improved seam-line searching algorithm for UAV image mosaic with optical flow. Sensors **18**(4), 1214 (2018)

21. Shi, Z., Chen, K., Pang, K., Zhang, J., Cao, Q.: A perceptual image quality index based on global and double-random window similarity. Digit. Signal Process. **60**, 277–286 (2017)

22. Li, N., Liao, T., Wang, C.: Perception-based seam cutting for image stitching, pp. 967–974. Signal, Image and Video Processing (2018)

23. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**, 91–110 (2004)

24. Zhang, J., Liu, T., Yin, X., Wang, X., Zhang, K., Xu, J., Wang, D.: An improved parking space recognition algorithm based on panoramic vision. Multimed. Tools Appl. **80**, 18181–18209 (2021)

25. Wang, Z., Yang, Z.: Review on image-stitching techniques. Multimed. Syst. **26**, 413–430 (2020)
26. Lee, H., Lee, S., Choi, O.: Improved method on image stitching based on optical flow algorithm. Int. J. Eng. Bus. Manag. **12**, 1847979020980928 (2020)
27. Sheng, M., Tang, S., Cui, Z., Wu, W., Wan, L.: A joint framework for underwater sequence images stitching based on deep neural network convolutional neural network. Int. J. Adv. Robot. Syst. **17**(2), 1729881420915062 (2020)
28. Pham, N.T., Park, S., Park, C.-S.: Fast and efficient method for large-scale aerial image stitching. IEEE Access **9**, 127852–127865 (2021)
29. Krishnakumar, K., Indira Gandhi, S.: Video stitching based on multi-view spatiotemporal feature points and grid-based matching. Vis. Comput. **36**, 1837–1846 (2020)
30. Jia, Q., Li, Z., Fan, X., Zhao, H., Teng, S., Ye, X., Latecki, L.: Leveraging line-point consistence to preserve structures for wide parallax image stitching. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12181–12190 (2021)
31. Liao, T., Li, N.: Single-perspective warps in natural image stitching. IEEE Trans. Image Process. **29**, 724–735 (2020)
32. Hossein-Nejad, Z., Nasri, M.: Clustered redundant keypoint elimination method for image mosaicing using a new Gaussian-weighted blending algorithm. Vis Comput **38**, 1991–2007 (2022)
33. Cao, Q., Shi, Z., Wang, P., Gao, Y.: A seamless image-stitching method based on human visual discrimination and attention. Appl. Sci. **10**, 1462 (2020)
34. Anzid, H., le Goic, G., Bekkari, A. Mansouri, A., Mammass, D.: A new SURF-based algorithm for robust registration of multimodal images data. Vis. Comput. (2022)
35. Sun, J., Wang, G., Goyal, V., Varshney, L.: A framework for Bayesian optimality of psychophysical laws. J. Math. Psychol. **56**, 495–501 (2012)
36. Goodhew, S., Dux, P., Lipp, O., Visser, T.: Understanding recovery from object substitution masking. Cognition **122**, 405–415 (2012)
37. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM. **24**(6), 381–395 (1987)
38. Dehaene, S.: The neural basis of the Weber-Fechner law: a logarithmic mental number line. Trends Cogn. Sci. **7**, 145–147 (2003)
39. Agaoglu, S., Agaoglu, M., Breitmeyer, B., Ogmen, H.: A statistical perspective to visual masking. Vision. Res. **115**, 23–39 (2015)
40. Zhang, L., Gu, Z., Li, H.: SDSP: A novel saliency detection method by combining simple priors. IEEE International Conference on Image Processing, pp. 171–175 (2013)

Jindong Zhang received the MSc and PhD degree in computer science and technology from Jilin University, in 2009. He is on faculty of Jilin University as an associate professor and doctoral supervisor of Computer Science and Technology, Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, and State Key Laboratory of Automobile Simulation and Control. He began a one-year visit to the Division of Computer Science, College of Computing, Hanyang University, Ansan, Republic of Korea, from 2017 to 2018. He has co-authored more than 60 research publications in peer reviewed journals and international conference proceedings

Ying Xiu received her Bachelor's degree in Software engineering from Jilin University, China, in 2021. She is currently a Master's student in the School of Computer Science and Technology at Jilin University, China. Her research interests include computer vision, artificial intelligence, image processing, image feature extraction, and image feature extraction