

# 360° Panorama Stitching Method with Depth Information: Enhancing Image Quality and Stitching Accuracy

Tianjiao Wang, He Huang, Zhenglong Cai, Jialin Song, Junxing Yang

School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture, Beijing, China

**Keywords:** panoramic image stitching; depth information ; seamline detection ; graph Cut ; Markov random field

## Abstract

Panoramic images, prized for their capacity to capture a comprehensive 360-degree field of view, find extensive applications across various domains. However, the challenge arises when attempting to seamlessly stitch together images captured by multiple cameras onto a spherical surface, particularly in cases where the cameras lack concentric alignment. To mitigate this misalignment issue, our paper introduces a novel method that integrates depth information into panoramic image stitching, aiming to enhance registration accuracy and seamline detection. The proposed method encompasses two pivotal steps. Firstly, depth information is employed to rectify the image's placement on the panoramic sphere, facilitating a two-dimensional registration process. This initial step targets the elimination of misalignment problems between images while preserving image clarity. Subsequently, depth information is seamlessly integrated into the smoothing term of the Markov random field energy function to guide seamline detection. Leveraging depth information aids in circumventing foreground obstacles and directs the search through spatially smooth areas, thereby reducing the likelihood of misalignment issues. Experimental results substantiate the effectiveness of employing depth information for image correction on the spherical surface, especially in scenarios where cameras approximate concentric alignment. Furthermore, the integration of depth information into the stitching network construction markedly diminishes misalignment, leading to a notable improvement in panoramic image quality. This signifies a crucial advancement in the domain of panoramic image stitching from an academic perspective.

## 1. Introduction

360-degree panoramic images play a pivotal role in photogrammetry and remote sensing, serving as indispensable data and information resources for societal development. Providing a comprehensive 360° view, these images are essential for various applications. Central to the generation of panoramic images is the core technology of image stitching, involving critical steps such as image registration [Brown and Lowe, 2007, Zaragoza et al., 2013, Turner et al., 2012], seamline detection [Kwatra et al., 2003, Yan et al., 2006, Li et al., 2019], and image color blending [Park et al., 2016, Shen et al., 2017, HaCohen et al., 2013, Yang et al., 2021]. These steps are crucial for creating images with a wider field of view and higher resolution.

The equipment employed for capturing panoramic images typically comprises two or more cameras. However, inherent challenges arise due to variations in the projection centers of individual cameras and the use of a fixed-radius projection sphere during image projection. Achieving perfect geometric alignment in the images becomes unattainable, leading to inevitable misalignment issues.

In the field of photogrammetry and remote sensing, most methods are based on orthophoto correction, using auxiliary data such as Digital Elevation Model (DEM), Digital Surface Model (DSM), Ground Control Point (GCP), and Global Positioning System (GPS) to correct the image to a coordinate system [Deng and Yang, 2020, Laliberte et al., 2011, Turner et al., 2012, Xiang et al., 2019]. There are also many image stitching algorithms based on transformation models. These algorithms initially derive point correspondences between images through feature extraction and matching. Subsequently, they compute a transformation matrix for each image based on the matching information, followed by transforming the images to a common

coordinate system to complete the image stitching process. For instance, methods based on homography matrix, such as single homography [Brown and Lowe, 2007], APAP [Zaragoza et al., 2013], AANAP [Lin et al., 2015], GSP [Chen and Chuang, 2016] and Progressive Stitching [Xia et al., 2015], exhibit high precision in natural image registration. Some articles also use multiple registration models for hierarchical registration based on degrees of freedom. Models with smaller degrees of freedom reduce the deformation of the results, while models with higher degrees of freedom ensure registration accuracy [Caballero et al., 2007, Xia et al., 2015, Cheng and Zhang, 2016, Moussa and El-Sheimy, 2016, Tian et al., 2020].

However, these methods are primarily suitable for scenarios involving small data volumes and small disparities. In the presence of substantial disparities, such methods may still encounter considerable misalignment and shape distortion issues. When applied to 360-degree panoramic stitching, further image deformation is often required [Nie et al., 2022] to achieve the projection from a planar to a spherical surface. Nonetheless, this additional deformation may exacerbate the image distortion.

To address misalignment issues, seamline detection has emerged as an exceptionally effective approach. This method involves identifying a seamline in the overlapping region of two images and utilizing pixels from the respective images to fill the areas on either side of the seamline. Notably, when the seamline traverses regions with weak texture or high alignment precision, misalignment issues are significantly mitigated. Seamline detection is a pivotal component in generating high-quality panoramic images and has found widespread application in remote sensing [Li, 2019]. Seamline detection techniques can be broadly categorized into three main types: pixel-based seamline detection methods, auxiliary information-based methods, and object-based methods. There are also many algorithms for

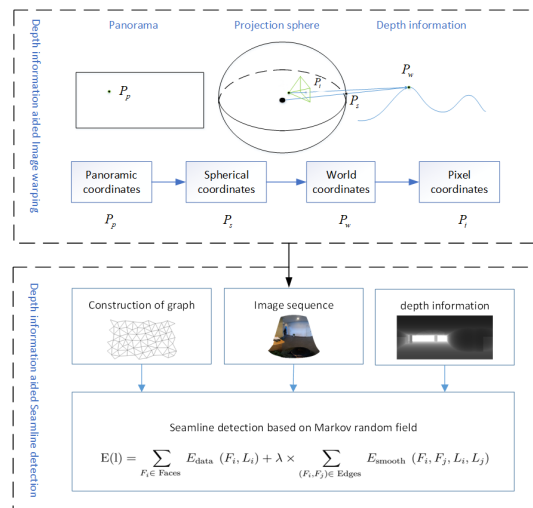
constructing concatenated networks [He et al., 2019, Li et al., 2019, Mills and McLeod, 2013, Pan et al., 2009, Pan et al., 2013, Zhu et al., 2018], which usually first construct a low precision concatenated network, and then use concatenated line search algorithms for optimization. Because this is not the focus of this article, it will not be elaborated here.

Pixel-based methods primarily determine the optimal seamline by analyzing pixel-level information. These methods typically involve computing difference maps between images and identifying the path with the minimum weighted cost. Common optimization techniques include ant colony algorithms, the twin-snake model [Kerschner, 2001], dynamic programming [Yu et al., 2012], and graph cut algorithms [Kwatra et al., 2003, Yan et al., 2006]. Standard evaluation metrics encompass grayscale differences, gradient disparities, texture discrepancies, and normalized cross-correlation coefficients. While pixel-based methods are straightforward and flexible, they often struggle to differentiate between foreground and background objects. Furthermore, foreground objects, such as buildings, may exhibit weak textures that are challenging for pixel-based methods to handle. Moreover, these methods demonstrate reduced efficiency when dealing with high-resolution orthoimages, limiting their applicability. To enhance efficiency, some techniques, such as the introduction of superpixels and triangulation [Yan et al., 2006], have been adopted. Methods based on auxiliary information utilize supplementary data sources such as digital surface models (DSM) [Zheng et al., 2017], digital line drawings [Wang et al., 2017], and point clouds [HaCohen et al., 2013] to guide the seamline detection process. These methods can efficiently and accurately determine the optimal seamline by leveraging features like roads and buildings. However, the reliance on additional data limits the applicability of these methods and increases their complexity. To enhance efficiency, techniques like block processing, parallel computation, image pyramids, and rapid stitching algorithms can be employed. Nevertheless, the dependence on additional data may constrain the suitability of these methods in certain scenarios. Methods based on objects [Li et al., 2017, Li et al., 2018] utilize techniques such as image segmentation, semantic analysis, and superpixel segmentation to predict and segment objects on the ground. This guidance helps the seamline avoid areas where objects are present. However, the effectiveness of object-based methods is heavily reliant on the accuracy of image segmentation, while deep learning methods may exhibit poor generalization, which constrains the applicability of such methods. In a paper by Pan [Pan et al., 2015], a new approach for orthoimage stitching seamline detection was proposed using Region Change Rate (RCR). This method enhances the algorithm's accuracy by incorporating RCR information into seamline feasibility region segmentation, built upon meanshift. Reference [Li et al., 2017] employed deep convolutional neural networks for image segmentation. Subsequently, energy costs for each pixel in the overlapping region were defined based on the classification probabilities for each specified category. Finally, Graphcut optimization was applied to obtain the final seamline. The advantage of this method is that, compared to meanshift, CNN offers higher segmentation accuracy, and when combined with Graphcut, it can produce high-quality seamlines. However, pixel-level seamline detection is inefficient and not suitable for large-scale high-resolution datasets. Additionally, because deep learning methods are utilized, a substantial amount of semantic data is required for network training. Addressing these issues, reference [Li et al., 2018] introduced a gradient domain seamline detec-

tion algorithm based on superpixels. This algorithm does not seek the optimal seamline in the overlap area among the entire pixel set via graph cut but instead begins by searching within superpixels created from input images and then refines them at the pixel level. This approach significantly improves the efficiency of global graph cut algorithm energy optimization, as the number of vertices in the graph cut model sharply decreases. Nevertheless, object-based methods depend on the precision of image segmentation, and both traditional segmentation methods and some deep learning-based methods exhibit limitations in accuracy and generalization, thereby restricting the scope of their application. Furthermore, there are stitching network construction algorithms, typically following a two-step strategy: initially creating an initial stitching network using Voronoi polygons and then optimizing each edge of the network using the seamline detection algorithm mentioned above. These algorithms heavily depend on the initial quality of the stitching network and the effectiveness of the seamline detection algorithm.

Seamline detection algorithms, while effective to a certain extent in mitigating misalignment issues, still struggle to handle severe misalignment problems resulting from significant disparities. To address this problem, a depth-assisted panoramic image stitching method has been proposed. This method combines depth information to improve registration accuracy and seamline detection. Firstly, by utilizing depth information, the images are rectified onto a spherical surface and unfolded into a two-dimensional representation. This process effectively reduces misalignment issues between images and maintains image clarity. Additionally, integrating depth information into the construction of the seamline network and optimizing it based on a Markov random field helps guide the seamline through regions with weak textures and gentle curvature. This further minimizes misalignment problems and significantly enhances the overall quality of panoramic images.

Figure 1. Overview of proposed method.



## 2. Method

Multi-lens panoramic cameras, with non-concentric shooting centers, often struggle with seamline challenges when projecting onto a spherical surface. Despite most cameras having approximately concentric centers, introducing depth information in the stitching process of 360-degree panoramic images significantly reduces seamlines. Our proposed algorithm employs depth information to improve image registration and seamline

detection, enhancing overall image quality. The algorithm, depicted in Figure 1, involves two key steps. Firstly, depth information corrects image positioning on the panoramic sphere, minimizing alignment errors within a narrow range. Subsequently, depth information seamlessly integrates into the Markov random field's energy optimization, guiding the seamline through areas with weak textures and smooth regions, avoiding abrupt disparities. This strategic integration mitigates alignment inaccuracies, ultimately elevating the panoramic image quality.

### 3. Image registration

Currently, with the aid of various sensors such as LIDAR, structured light, binocular vision, MVS algorithms, and more, it is convenient to obtain depth information of the scene. This depth information is then used to directly project images onto a panoramic sphere, facilitating the stitching process. Even in cases where the image sequence is approximately concentric, minor errors in the depth information do not lead to significant misalignment problems. This approach, which considers both depth information and visual data, offers substantial advantages over purely visual-based image stitching. In this study, we leverage this principle by introducing depth information to rectify images onto the projection sphere, achieving successful image registration. Our research utilized a 7-camera GoPro panoramic camera (GoPro, Inc., San Mateo, CA, USA) to capture panoramic images. Subsequently, we processed the images using PhotoScan (Agisoft LLC, St. Petersburg, Russia) to generate a 3D model of the scene and calculate depth values for the images. It's important to note that depth information can be easily obtained through various sensors such as time-of-flight (TOF) cameras, RGBD cameras, LIDAR, and stereo matching methods. As a result, the approach proposed in this paper can be extended to scenes captured using these devices. Following the calibration of multiple cameras, it becomes possible to determine the pose of each camera. The average position of the photographic centers of the  $n$  cameras is considered as the center  $O$  of the panoramic sphere. The formula for calculating the center  $O$  of the panoramic sphere is as follows:

$$O = \frac{\sum_{i=1}^n o_i}{n} \quad (1)$$

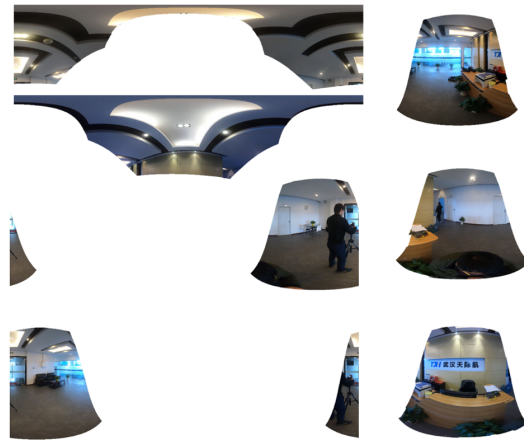
Once the center of the panoramic sphere is determined, it becomes possible to project the spatial three-dimensional points from the captured scene onto the panoramic sphere, resulting in a depth map of the panoramic sphere. However, devices like LIDAR and TOF cameras provide discrete point cloud data. To address this issue, our study employs interpolation methods to calculate the depth value of a pixel point  $p$ . The formula for calculating the depth value of a pixel point  $p$  is as follows:

$$C_g = \exp \left( \sqrt{\text{sobel}(m_x)^2} \right) \quad (2)$$

Among these,  $p_i$  refers to the  $i$ -th pixel of the captured image,  $Q$  is the  $p_i$  set of neighboring  $n$  pixel points,  $p_j$  represents the  $j$ -th pixel adjacent to  $p_i$ , and  $w_i$  denotes the weight coefficient of  $p_j$ .

Let the pixel point  $p$  correspond to a point  $P$  on the image, where the coordinates are  $(x, y)$ . The spherical coordinate system coordinates of  $p$  are represented as  $(r, \theta, \varphi)$ , where  $\theta$  denotes the zenith angle and  $\varphi$  represents the azimuth angle. Specifically,  $\theta = y/r$ ,  $\varphi = 2\pi - x/r$ .

Figure 2. Sequence images corrected by depth information.



The spherical coordinates of the projected point  $p$  on the panoramic sphere (in other words, in Cartesian rectangular world coordinates) are given by:

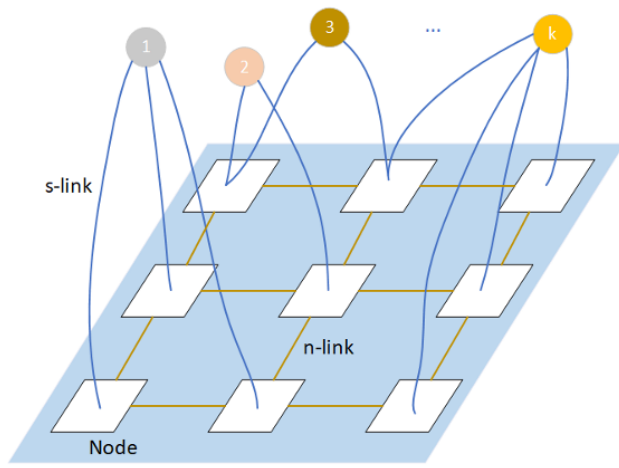
$$q = \left( r \begin{pmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{pmatrix} + O \right) \quad (3)$$

After obtaining the three-dimensional coordinates, it is possible to use the camera model and vectors to calculate the object coordinates corresponding to pixel point  $p$  in space. This information can be used to calculate the image coordinates corresponding to pixel point  $p$ , which in turn allows for the determination of the corresponding pixel value. This process completes the projection of the captured image onto the panoramic sphere. Finally, by unwrapping the panoramic sphere, a registered sequence of images can be obtained. Figure 2 illustrates the process of image projection onto the panoramic sphere.

### 4. Seamline detection

Once the images are projected onto the panoramic sphere to form a sequence of images, they are already registered on the spherical surface. However, using depth information for stitching alone does not fully exploit the advantages of depth information. Moreover, due to the presence of depth information errors and image pose errors, it is necessary to further employ seamline detection methods to eliminate stitching seams and achieve higher-quality images. Therefore, this paper proposes a method that utilizes depth information to assist seamline detection, aiming to enhance the quality of stitching. Before delving into the detailed discussion of seamline detection implementation, the theoretical foundations of seamline detection were introduced. This included Markov Random Field theory and Graph Cut Model theory. In computer vision research, there are many situations where objects in a model need to be labeled. For instance, in foreground-background segmentation, each pixel needs to be labeled as foreground or background; in stereo vision, each pixel needs to be labeled with a disparity value; in image segmentation, each pixel needs to be labeled as belonging to the object or the background, and so on. Graph Cut Models are highly effective tools for solving these types of problems. Seamline detection involves selecting which image provides the pixel values for each pixel, making it suitable to use Graph Cut Models to address this challenge.

Figure 3. Schematic diagram of graph cut.



In real-world scenarios, the problem is represented as an undirected graph  $G(V, E)$ . This graph differs slightly from a typical graph. In graph  $G$ , the vertices  $V$  represent labels and objects (e.g., pixels), and the edges  $E$  are divided into two categories: s-links and n-links. S-links connect label vertices to regular vertices, and their weights measure the quality of selecting a label. N-links connect regular vertices to each other, and their weights represent the cost of choosing different labels. Each edge in the graph has a non-negative weight. A graphical representation of a graph cut problem with two labels is shown in Figure 3. In the graph cut problem, a cut refers to a subset of edges in the graph, with a cost equal to the sum of the weights of the selected edges. When a cut divides the vertices into two disconnected subsets, each containing a label, it is referred to as a graph cut. If this cut has the minimum cost among all cuts, it is called the minimum cut. Ford-Fulkerson's theorem demonstrates the equivalence between the minimum cut and maximum flow. Therefore, the minimum cut can partition the vertices into two parts, thereby solving the label segmentation problem. When dealing with multiple labels ( $\geq 3$ ), alpha-expansion or swap algorithms can be used to find a solution. This problem can be expressed using the pairwise Markov Random Field energy formula. The Markov Random Field energy formula is as follows:

$$E(L) = E_{\text{data}}(L) + E_{\text{smooth}}(L) \quad (4)$$

In this formulation,  $E_{\text{data}}$  measures the cost incurred when a vertex selects a specific label, while  $E_{\text{smooth}}$  measures the cost when two adjacent vertices select different labels. These two components are commonly referred to as data and smoothness terms, respectively. The challenge lies in designing appropriate data and smoothness terms to achieve specific objectives. In the context of seamline detection, the ordinary vertices correspond to pixels, while the labels represent different images.

In this example, we illustrate the process of searching for a seamline between two sequential images. To begin, we generate a depth map from the images and then employ a graph cut algorithm to search for the seamline within the depth map. Each pixel in the sequential images is treated as a vertex within the graph, and we establish edges between adjacent pixels, assigning them appropriate weights. These weight values are determined based on color differences and gradients. The degree of color difference minimizes noticeable seamlines in the final

composite image, while the gradients ensure that the seamline follows the smoother areas within the depth map, avoiding object regions. In the overlapping area of the two sequential images, corresponding positions are regarded as nodes, and each node is assumed to have four potential neighbors: up, down, left, and right.

The process of searching for a seamline involves classifying the pixels in the overlapping region into two categories: one belonging to image A and the other to image B. This method is based on constructing a weighted graph structure from the depth map, where the edge weights represent the cost of the energy function. Subsequently, the maximum flow/minimum cut algorithm is employed to find the minimum cut within the graph, which represents the optimal seamline. Here, 'p' represents all the nodes in the graph, ' $N(p)$ ' denotes the adjacency relationships between nodes, and ' $M(p)$ ' represents the image labels for all nodes. The ultimate objective is to discover an image label mapping function that assigns each element 'p' to a label in such a way that it minimizes the energy value in equation (4).

In this study, for efficiency enhancement, the following data terms are set: If the 'i-th' triangle face is visible in the target image sequence ' $L$ ' (meaning the triangle face is covered by the sequence image ' $L$ '), the data term is set to 0; otherwise, it is set to infinity. The formula for setting the data terms is as follows:

$$E_{\text{data}}(F_i, L) = \begin{cases} 0 & \text{if } (F_i \in L) \\ \infty & \text{if } (F_i \notin L) \end{cases} \quad (5)$$

The smoothness term addresses adjacent pixels with different labels, and the gradient reflects the smoothness level of the image area. To encourage the seamline to pass through smooth regions, the smoothness term should be related to the gradient. Smoother regions have smaller gradients. Therefore, the gradient (depth gradient) for each pixel position in the overlapping region of the two sequence images is computed using the Sobel operator and used as the cost for the smoothness term. The cost formula for each pixel is as follows:

$$C_g = \exp \left( \sqrt{\text{sobel}(m_X)^2 + \text{sobel}(m_Y)^2} \right) \quad (6)$$

where  $\text{sobel}()$  represents the Sobel operator,  $\exp$  is a smoothing function,  $m_X$  is the depth gradient in the  $X$  direction, and  $m_Y$  is the depth gradient in the  $Y$  direction. For each pair of adjacent pixels  $(x', y')$ , its smooth energy cost is defined as the sum of the costs of the two pixels, and the calculation formula is as follows:

$$E_{\text{smooth}}(x', y') = C_g(x') + C_g(y') \quad (7)$$

The Markov random field energy term is ultimately minimized through  $\alpha$ -expansion to obtain the optimal solution. This involves selecting an image label for each pixel to obtain the final seamline.

## 5. Results and Discussion

In this section, the effectiveness of depth-assisted panoramic image seamline is analyzed from two perspectives. This study



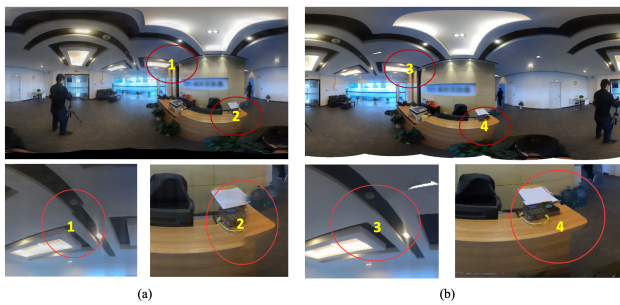


Figure 4. Comparison between the results of the algorithm proposed in this study and the results of the PhotoScan. (a) displays the output of the PhotoScan algorithm, while (b) exhibits the outcome of our proposed algorithm. Notably, regions annotated with numerical labels represent detailed areas

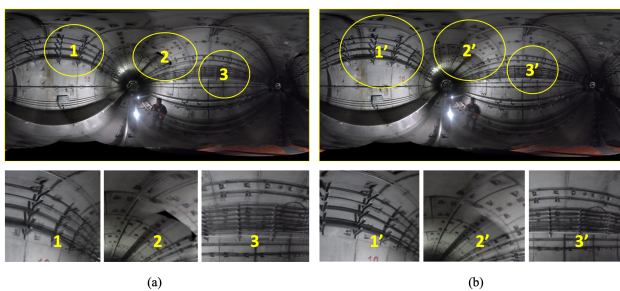


Figure 5. Comparison between fixed projection radius and dynamic radius in two scenarios. (a) depicts the experimental results with a fixed depth value, while (b) showcases the outcomes of our proposed algorithm. It is important to note that these comparisons pertain to the choice of projection radius and its impact on the results.

used a panoramic camera consisting of seven GoPro cameras to capture panoramic image data. The images were processed using PhotoScan software to obtain a 3D model of the scene, and then the depth values of the images were calculated based on the model. It is worth noting that the acquisition of depth information has become possible with the emergence of TOF cameras, RGBD cameras, LIDAR, and stereo matching methods. Therefore, the method proposed in this paper is applicable to scenes captured using these devices. Since indoor scenes have relatively small distances between the camera centers and the scenes, errors may lead to misalignment, so we selected two scenes with distinct line structures.

In traditional panoramic image stitching, it is typically assumed that the image depth is fixed. In some open scenes far from the camera, a fixed radius may not significantly affect image alignment. However, in narrow indoor settings, image misalignment becomes more pronounced. Figure 4 illustrates a comparison between fixed depth and dynamic depth in the experiments conducted in this study. At the end of the experiments, the images were merged into a single image using a linear blending method. When there is significant registration error between images, it can lead to blurriness in the overlapping regions. From the comparison results of (a) and (b), it can be seen that (a) has significant misalignment issues in some edge areas, because stitching with a fixed stitching radius cannot adapt to situations with large parallax, which is particularly common in indoor scenes. The problem of misalignment in the results of the algorithm proposed in this article has been greatly reduced, as the introduction of depth information can improve the registration accuracy between images. Even if there is some error in depth information, the registration accuracy is still high because the camera is roughly concentric, resulting in minimal

misalignment. This proves that the introduction of depth information can improve the quality of panoramic stitching.

We further analyzed the impact of the radius results in a narrow space, that is, conducting experiments in a tunnel scene. As shown in Figure 5, the experimental results using a fixed depth exhibit significant blurriness in some details, indicating a considerable error in the registration results. In contrast, the algorithm employed in this paper, which uses actual depth for registration, results in images without noticeable blurriness at the same locations. This demonstrates that the algorithm possesses a higher level of registration accuracy. Moreover, in the areas highlighted in Figure 5, where there are notable depth errors, the experimental results do not exhibit significant blurriness. This suggests that when the image sequence is roughly co-centric, even with some depth errors, the impact on the registration accuracy of panoramic images is relatively minor, and it does not lead to noticeable misalignment or blurriness issues, which can improve the quality of panoramic images.

Figure 6. Results of the seamline detection assisted by depth information. (a) shows the images after correction with depth information, while (b) presents the corresponding depth map. (c) displays the resulting seamline. This approach leverages depth information to refine the image alignment and seamline detection process, leading to improved results in panoramic image stitching

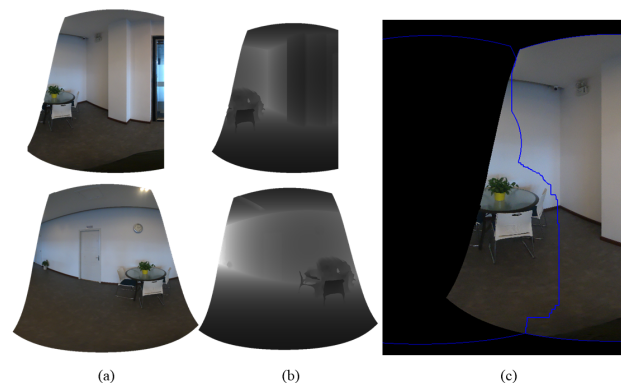
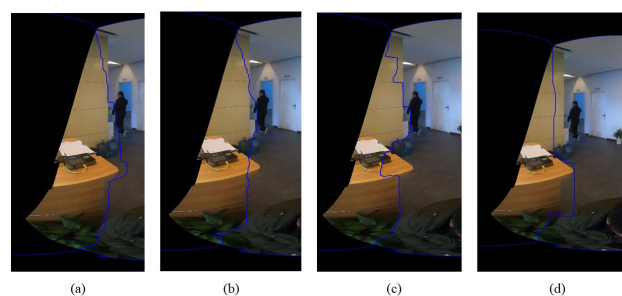


Figure 7. Comparison of results achieved by various methods. (a) represents the results obtained by Kwata et al. using the L1 color difference norm, while (b) corresponds to their results using the L2 color difference norm. (c) displays the outcomes obtained by integrating both color difference and gradient difference. (d) showcases the results achieved by our method, which utilizes depth information. This visual comparison highlights the performance differences among these methods



While the use of depth information can assist in the registration of panoramic images and significantly reduce misalignment issues, there may still be some misalignment problems

between images due to errors in depth information and non-collinear photographic centers in sequential images. Therefore, it remains necessary to employ panorama seamline detection algorithms to composite sequential images into panoramic images. This allows the seamline to navigate around foreground objects such as tables and pedestrians, minimizing misalignment. Depth information not only aids in the registration of sequential images but also guides the seamline detection during the image stitching process. Figure 6 shows the results of employing depth information to assist in the seamline detection. It is evident that the seamline perfectly avoids the table, preventing misalignment. This is because we utilize the gradient of image depth as a smoothing term in this work. In regions with foreground objects such as tables, the depth undergoes significant changes, leading to larger gradient values. In contrast, regions such as walls and floors exhibit relatively smooth depth changes, enabling the seamline to pass through without causing noticeable distortions.

Furthermore, we compared our algorithm in this paper with methods that solely rely on pixel information. Figure 7 illustrates how our approach incorporates depth information into the energy function's smoothing term, enabling the seamline to traverse smooth areas like walls and floors while bypassing foreground objects like tables for optimal stitching results. On the other hand, the other three methods consider only color information, and they fail to entirely circumvent foreground objects. The seamline passes through the table, resulting in misalignment. This is because real-world foreground objects may also possess textured regions with small gradients, which cannot be reliably identified as foreground areas based solely on color information. Hence, these methods have limitations and may be ineffective. Currently, depth information can be easily extracted from images and reflects the geometric structure of the real world. Significant depth changes occur in foreground regions. Therefore, the incorporation of depth information enables the seamline to smoothly navigate around these areas, preventing misalignment and enhancing the stitching quality.

## 6. Conclusions

This paper introduces a novel panorama image stitching method that leverages depth information to enhance stitching quality from two critical angles: image registration and seamline optimization. Initially, images are registered onto the panoramic spherical surface using depth information. Subsequently, an energy function based on a Markov Random Field (MRF) framework is constructed, employing depth gradient values as a smoothing term. This ensures that the seamline navigates regions with minimal depth variation, effectively avoiding foreground objects in the panorama. Experimental results demonstrate a significant improvement in alignment accuracy when depth information is incorporated, particularly in co-centered image scenarios, thereby mitigating image misalignment. Furthermore, the depth information is harnessed to guide the seamline detection, enabling it to adeptly circumvent foreground objects and smoothly traverse regions with low texture variations. This strategy further reduces misalignment in the final results, resulting in a remarkable enhancement in panorama quality, which can improve the visual experience of 360-degree panoramic images and be important for the applications based the panoramic images.

## Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant Numbers 42201483), the China Postdoctoral Science Foundation (Grant Numbers 2022M710332) and Funding for Postdoctoral Research Activities in Beijing (2023-zz-140).

## References

- Brown, M., Lowe, D. G., 2007. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74, 59–73.
- Caballero, F., Merino, L., Ferruz, J., Ollero, A., 2007. Homography based kalman filter for mosaic building. applications to uav position estimation. *Proceedings 2007 IEEE International Conference on Robotics and Automation*, IEEE, 2004–2009.
- Chen, Y.-S., Chuang, Y.-Y., 2016. Natural image stitching with the global similarity prior. *European conference on computer vision*, Springer, 186–201.
- Cheng, Z., Zhang, L., 2016. A Method of Aerial Image Mosaic Based on UAV Posture Information.
- Deng, F., Yang, J., 2020. Panoramic image generation using centerline-constrained mesh parameterization for arbitrarily shaped tunnel lining. *IEEE Access*, 8, 7969–7980.
- HaCohen, Y., Shechtman, E., Goldman, D. B., Lischinski, D., 2013. Optimizing color consistency in photo collections. *ACM Transactions on Graphics (TOG)*, 32(4), 1–10.
- He, J., Sun, M., Chen, Q., Zhang, Z., 2019. An improved approach for generating globally consistent seamline networks for aerial image mosaicking. *International Journal of Remote Sensing*, 40(3), 859–882.
- Kerschner, M., 2001. Seamline detection in colour orthoimage mosaicking by use of twin snakes. *ISPRS journal of photogrammetry and remote sensing*, 56(1), 53–64.
- Kwatra, V., Schödl, A., Essa, I., Turk, G., Bobick, A., 2003. Graphcut textures: Image and video synthesis using graph cuts. *Acm transactions on graphics (tog)*, 22(3), 277–286.
- Laliberte, A. S., Goforth, M. A., Steele, C. M., Rango, A., 2011. Multispectral remote sensing from unmanned aircraft: Image processing workflows and applications for rangeland environments. *Remote Sensing*, 3(11), 2529–2551.
- Li, L., 2019. Research on Orthophoto Image Stitching: Seamline Detection and Network Construction Methods. PhD thesis, Wuhan University.
- Li, L., Tu, J., Gong, Y., Yao, J., Li, J., 2019. Seamline network generation based on foreground segmentation for orthoimage mosaicking. *ISPRS Journal of Photogrammetry and Remote Sensing*, 148, 41–53.
- Li, L., Yao, J., Liu, Y., Yuan, W., Shi, S., Yuan, S., 2017. Optimal seamline detection for orthoimage mosaicking by combining deep convolutional neural network and graph cuts. *Remote Sensing*, 9(7), 701.

- Li, L., Yao, J., Shi, S., Yuan, S., Zhang, Y., Li, J., 2018. Superpixel-based optimal seamline detection in the gradient domain via graph cuts for orthoimage mosaicking. *International Journal of Remote Sensing*, 39(12), 3908–3925.
- Lin, C.-C., Pankanti, S. U., Natesan Ramamurthy, K., Aravkin, A. Y., 2015. Adaptive as-natural-as-possible image stitching. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1155–1163.
- Mills, S., McLeod, P., 2013. Global seamline networks for orthomosaic generation via local search. *ISPRS journal of photogrammetry and remote sensing*, 75, 101–111.
- Moussa, A., El-Sheimy, N., 2016. A fast approach for stitching of aerial images. *The international archives of the photogrammetry, remote sensing and spatial information sciences*, 41, 769–774.
- Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y., 2022. Deep rectangling for image stitching: a learning baseline. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5740–5748.
- Pan, J., Wang, M., Li, D., Li, J., 2009. Automatic generation of seamline network using area Voronoi diagrams with overlap. *IEEE Transactions on Geoscience and Remote Sensing*, 47(6), 1737–1744.
- Pan, J., Wang, M., Li, J., Yuan, S., Hu, F., 2015. Region change rate-driven seamline determination method. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 141–154.
- Pan, J., Wang, M., Ma, D., Zhou, Q., Li, J., 2013. Seamline network refinement based on area Voronoi diagrams with overlap. *IEEE Transactions on Geoscience and Remote Sensing*, 52(3), 1658–1666.
- Park, J., Tai, Y.-W., Sinha, S. N., So Kweon, I., 2016. Efficient and robust color consistency for community photo collections. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 430–438.
- Shen, T., Wang, J., Fang, T., Zhu, S., Quan, L., 2017. Color correction for image-based modeling in the large. *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20–24, 2016, Revised Selected Papers, Part IV 13*, Springer, 392–407.
- Tian, Y., Sun, A., Luo, N., Gao, Y., 2020. Aerial image mosaicking based on the 6-DoF imaging model. *International Journal of Remote Sensing*, 41(1), 74–89.
- Turner, D., Lucieer, A., Watson, C., 2012. An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote sensing*, 4(5), 1392–1410.
- Wang, D., Cao, W., Xin, X., Shao, Q., Brolly, M., Xiao, J., Wan, Y., Zhang, Y., 2017. Using vector building maps to aid in generating seams for low-attitude aerial orthoimage mosaicking: Advantages in avoiding the crossing of buildings. *ISPRS Journal of Photogrammetry and Remote Sensing*, 125, 207–224.
- Xia, M., Yao, M., Li, L., Lu, X., 2015. Globally consistent alignment for mosaicking aerial images. *2015 IEEE International Conference on Image Processing (ICIP)*, IEEE, 3039–3043.
- Xiang, T.-Z., Xia, G.-S., Zhang, L., 2019. Mini-unmanned aerial vehicle-based remote sensing: Techniques, applications, and prospects. *IEEE geoscience and remote sensing magazine*, 7(3), 29–63.
- Yan, J., Zhang, B., Liu, N., Yan, S., Cheng, Q., Fan, W., Yang, Q., Xi, W., Chen, Z., 2006. Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing. *IEEE transactions on Knowledge and Data Engineering*, 18(3), 320–333.
- Yang, J., Liu, L., Xu, J., Wang, Y., Deng, F., 2021. Efficient global color correction for large-scale multiple-view images in three-dimensional reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 209–220.
- Yu, L., Holden, E.-J., Dentith, M. C., Zhang, H., 2012. Towards the automatic selection of optimal seam line locations when merging optical remote-sensing images. *International Journal of Remote Sensing*, 33(4), 1000–1014.
- Zaragoza, J., Chin, T.-J., Brown, M. S., Suter, D., 2013. As-projective-as-possible image stitching with moving dlt. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2339–2346.
- Zheng, M., Xiong, X., Zhu, J., 2017. Automatic seam-line determination for orthoimage mosaics using edge-tracking based on a DSM. *Remote Sensing Letters*, 8(10), 977–986.
- Zhu, S., Zhang, R., Zhou, L., Shen, T., Fang, T., Tan, P., Quan, L., 2018. Very large-scale global sfm by distributed motion averaging. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4568–4577.