Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(9): 17356–17383.
DOI: 10.3934/mbe.2023772
Received: 04 July 2023
Revised: 05 August 2023
Accepted: 27 August 2023
Published: 08 September 2023

Research article

A global optimization generation method of stitching dental panorama with anti-perspective transformation

Ning He, Hongmei Jin*, Hong'an Li and Zhanli Li

College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, China

* **Correspondence:** Email: jinhm@xust.edu.cn; Tel: +86-135-7227-8766.

Abstract: To address the limitation of narrow field-of-view in local oral cavity images that fail to capture large-area targets at once, this paper designs a method for generating natural dental panoramas based on oral endoscopic imaging that consists of two main stages: the anti-perspective transformation feature extraction and the coarse-to-fine global optimization matching. In the first stage, we increase the number of matched pairs and improve the robustness of the algorithm to viewpoint transformation by normalizing the anti-affine transformation region extracted from the Gaussian scale space and using log-polar coordinates to compute the gradient histogram of the octagonal region to obtain the set of perspective transformation resistant feature points. In the second stage, we design a coarse-to-fine global optimization matching strategy. Initially, we incorporate motion smoothing constraints and improve the Fast Library for Approximate Nearest Neighbors (FLANN) algorithm by utilizing neighborhood information for coarse matching. Then, we eliminate mismatches via homography-guided Random Sample Consensus (RANSAC) and further refine the matching using the Levenberg-Marquardt (L-M) algorithm to reduce cumulative errors and achieve global optimization. Finally, multi-band blending is used to eliminate the ghosting due to unalignment and make the image transition more natural. Experiments show that the visual effect of dental panoramas generated by the proposed method is significantly better than that of other methods, addressing the problems of sparse splicing discontinuities caused by sparse keypoints, ghosting due to parallax, and distortion caused by the accumulation of errors in multi-image splicing in oral endoscopic image stitching.

Keywords: panorama generation; image stitching; affine invariance; perspective transformation; global optimization

1. Introduction

Nowadays, oral health has become an important indicator for measuring people's overall health and quality of life. The oral endoscope has brought a new mode for the examination and treatment of the oral cavity. However, oral endoscopic imaging usually can only capture one or two teeth on a certain tooth surface, and dentists need to repeatedly review partial images to understand and evaluate the overall oral health condition during diagnosis. Image stitching is a commonly used method to solve the problem of the narrow field of view of partial images under microscopic photography, which helps dentists make detailed judgments on the scope and specific conditions of oral lesions, thus improving the accuracy and efficiency of oral treatment diagnosis.

Due to the limited and similar texture features of dental images, existing stitching algorithms have sparse matching key points, leading to incorrect estimation of homography matrices and inaccurate representation of mapping relationships, resulting in problems such as misalignment, unevenness, and stitching interruption at the seams. Hand-held shooting cannot fix the optical center position, and even slight shaking can cause large depth changes and viewing angle differences between images, resulting in the incorrect matching of corresponding pixel points in the overlapping area between the reference image and the image to be stitched. In multi-image stitching, error accumulation becomes increasingly significant, leading to severe and uneven stretching and deformation of images far from the reference image. Therefore, panoramic stitching of intraoral endoscopic images faces many challenges. Traditional feature-based image stitching methods heavily rely on the quality of feature extraction, and the scarcity and uneven distribution of dental image features also pose difficulties for these methods. Yan et al. [1] proposed a multi-constrained super pixel feature-based laryngeal ultrasound image stitching algorithm for local image stitching on both sides of the larynx. Ghanoum et al. [2, 3] proposed an improved method for low-texture region matching for frame stitching of intraoral images that extracts normal information from the tooth surface to generate feature-rich normal maps, uses normal maps to detect, extract, and match corresponding features, and estimates an approximate projection transformation model. However, this method does not perform well in high-texture regions, and the data used are simulated images rather than real intraoral environments, making it difficult to evaluate its effectiveness. One of the commonly used image stitching methods is the combination of edge contour recognition and segmentation of subregions with high-detail information. This is achieved by performing edge detection on the image of the region of interest. However, traditional integer-order derivatives may lead to loss of detail and interference from noise in the edge detection process. In contrast, fractional-order derivatives provide a more accurate representation of edge information in the image. By constructing appropriate partial differential equation models of fractional order, such as Caputo fractional differential equations [4, 5], it is possible to effectively extract subtle edge and texture features in the image. This enhances the robustness of edge detection and ultimately reduces biases and artifacts in the stitching process.

Compared with traditional methods, the deep learning stitching method has a stronger feature extraction ability for images with fewer texture features and has advantages in solving transformation parameter issues. Due to the high cost of manually labeled data acquisition, supervised [6–9] methods mainly use synthetic datasets for training, and weakly supervised [10] or unsupervised [11] approaches adaptively learn in a data-driven mode; however, these methods cannot adapt to the parallax and illumination variations of real oral images. Zhang et al. [12] and Liu et al. [13] solved the problem of insufficient feature correspondence to a certain extent by learning content-aware masks to select reliable regions for homography estimation. However, by extracting features from background regions and regions with rich textures for image registration, important lesion details in oral cavity images are easily overlooked. Nie et al. [11] proposed the first unsupervised deep learning image stitching framework, which consists of two stages: unsupervised coarse image alignment and unsupervised image reconstruction. The first stage estimates a global homography matrix to coarsely align input images, and the second stage reconstructs the coarsely aligned results to obtain the stitched image, which can effectively eliminate artifacts. However, the coarse image alignment stage of this method has high requirements for overlap ratio and disparity. Huang et al. [14] based on the above unsupervised deep image stitching framework, proposed an unsupervised endoscope image stitching algorithm based on overlap region extraction and deep feature loss that extracts the overlap region of input images through polygon intersection sketches. The estimation of isomorphism from coarse to fine is accomplished using a similar approach to the one described in the literature [15], which operates on a three-layer feature pyramid structure. The final step involves reconstructing the stitched image by mapping features back to pixels. The network relies on the tooth dataset and has high requirements for its quality, which reduces the network's generalization ability. In addition, the reconstruction method used is still the same as that in the literature [11] , and the accuracy of homography estimation in oral cavity image stitching tasks still needs to be improved.

Although image stitching is a very classical and complete system in the field of computer vision, it is still in its infancy in the field of oral endoscopy, and there is very little research and literature on image stitching in dentistry. Advanced stitching techniques for non-medical applications have not yet been tested in RGB oral image environments, and there is no targeted image stitching solution for oral endoscopic images. At the same time, these algorithms lack publicly available datasets that can be used to evaluate oral endoscopic images. Traditional feature point-based image stitching methods rely on the quality of feature extraction, and deep learning-based methods require processing large-scale data; therefore, small samples of RGB oral image data cannot achieve the desired stitching results.

To overcome the above limitations, we construct the first small-scale real dataset for oral endoscopic panorama image stitching (to the best of our knowledge) and design the first method for generating natural-looking dental panoramas for oral endoscopic image stitching, which consists of two stages: the anti-perspective transformation feature extraction and the coarse-to-fine global optimization matching. The panorama results obtained by the proposed method are experimentally verified to maintain the integrity of the oral images and look more natural both locally and globally. In general, the contributions of this work are summarized as follows:

1) We design a two-stage method for generating natural dental panoramas based on oral endoscopic imaging based on the characteristics of oral images, which effectively mitigates the problems of sparse splicing discontinuities caused by sparsely spaced keypoints, heavy shadows caused by parallax, and distortions caused by the accumulation of multiple-image splicing errors in the splicing of oral endoscopic images.

2) In the first stage,we increase the number of matched pairs and improve the robustness of the algorithm to viewpoint transformation by normalizing the anti-affine transformation region extracted from the Gaussian scale-space and using log-polar coordinates to compute the gradient histogram of the octagonal region to obtain the set of perspective transformation resistant feature points.

3) In the second stage, we design a global optimization matching strategy from coarse to fine by

adding motion smoothing constraints, improving the fast nearest-neighbor algorithm by using neighborhood information to achieve coarse matching, eliminating mismatches via homography-guided Random sample consensus, and combining with the Levenberg-Marquardt algorithm to reduce the cumulative error to achieve global optimization. The strategy is to ensure that the resulting panorama looks more natural locally and globally.

The rest of the paper is organized as follows: Section 2 presents related work, including the core steps of image stitching image alignment and the current mainstream spatial transformation distortion methods to solve the parallax problem. Section 3 describes the proposed dental panoramic stitching strategy in detail, and Section 4 describes the experimental results and analysis. Finally, the paper is summarized in Section 5.

2. Related work

2.1. Image registration

Feature-based image alignment is a crucial process that provides important support for subsequent image stitching work. This process consists of two steps: feature extraction and feature matching. Initially, the focus of feature extraction was on corner point detection, including algorithms such as Harris corner point, Moravec corner point, and accelerated segment test FAST corner point. However, due to the limited information contained in corner points, Lowe et al. [16] proposed the Scale Invariant Feature Transform (SIFT), which utilizes Gaussian fuzzy to construct the scale space, determines the location and scale of feature points through Gaussian differential function and model fitting, and applies local gradients as feature point directions to construct 128-dimensional feature descriptors, as shown in Figure 1. Nonetheless, in regions with weak or repeated textures, the sparse nature of SIFT feature descriptors may result in insufficient feature information. The Speeded Up Robust Features (SURF) [17] algorithm is an improvement of SIFT, which uses Haar wavelets to approximate the gradient operation in SIFT and improve the efficiency of the algorithm. BRIEF, ORB [18] and other binary descriptors are commonly used to balance the robustness of features and computational efficiency but do not have scale invariance or rotation invariance.

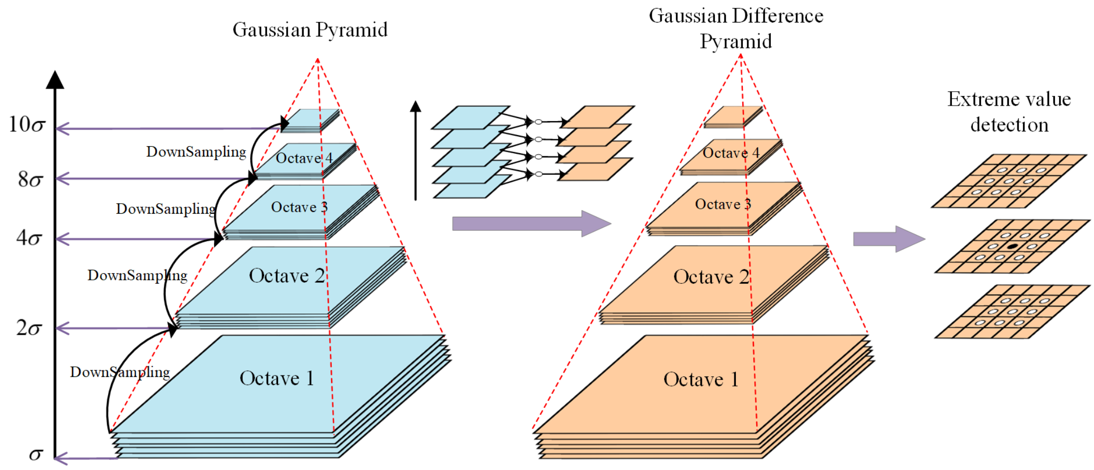


Figure 1. Gaussian scale space.

2.2. Spatially-varying warping

In the feature matching stage, establishing the exact correspondence between two sets of image points is typically achieved through coarse matching and fine matching. Coarse matching involves rough matching based on the similarity between feature descriptors, such as violent matching. The outcome of rough matching generally contains many false matches. Therefore, the Random Sample Consensus (RANSAC) algorithm [19] is commonly used as fine matching in image stitching to further reject false matches and obtain the best single correspondence matrix between two images. However, RANSAC filters feature points by global homography, which can lead to missed detections in multiple single-response scenarios. Additionally, the threshold parameter of projection error needs to be adjusted for different input data to optimize the number and quality of inner point pairs on various data, resulting in the low robustness of the image stitching algorithm. Bian et al. [20] proposed a GMS algorithm based on grid motion statistics. This method removes erroneous matches by analyzing the number of features that exhibit matching relationships in neighboring areas around coarse matches. This is accomplished using grid partitioning and motion statistics properties, which transform feature point matching from a quantitative to a qualitative process. The Thin Plate Spline (TPS) [21, 22] technique has been widely used in non-rigid deformation methods for medical image alignment, with Li et al. [23] proposing a robust elastic warping method using TPS to address the parallax problem. Nie et al. [24] extended the applicability of TPS to deep learning frameworks by proposing a flexible warping approach that models global single-strain to local thin plate spline motion. Despite its effectiveness in medical image alignment, the flexibility of TPS renders it unsuitable for aligning dental structures, which are harder and more rigid than human organs and tissues.

Image alignment is the core of image stitching and the key to solving the parallax problem, mainly by extracting salient features to solve the transformation model parameters. The most widely used image alignment model is the homography transformation, which accurately considers the transformations between two 2D planes. such as the AutoStitch algorithm [25], which utilizes a global homography aligned image. However, this model is limited in its ability to align images at different depths and perspectives and is unsuitable for non-coplanar scenes. Traditional feature-based methods use a series of homography matrices to distort images, collectively referred to as spatially varying distortion methods.

Gao et al. [26] proposed the Dual-Homography Waping (DHW) algorithm, which aligns the background and foreground using two homography matrices but cannot handle multiple planes. Lin et al. [27] presented the Smoothly Varying Affine (SVA) algorithm, which utilizes conventional affine transformation parameters as global transformation parameters. This approach provides flexibility in handling parallax while maintaining the desirable extrapolation and occlusion processing properties of parametric transformation. However, it does not have the ability to impose global projectivity. Zaragoza et al. [28] proposed the As-Projective-As-Possible image stitching (APAP) algorithm, which introduces grid deformation and establishes a Moving DLT mathematical model. This solves the overdetermined equations to generate grid homography matrices for perspective transformation, mapping the grid onto the panorama canvas to obtain a distorted image. APAP performs well in general image stitching situations but extrapolates the projection transformation beyond the non-overlapping regions, causing severe perspective distortion in the far-off areas from the boundary. Chang et al. [29] proposed the Shape-Preserving Half-Projective (SPHP) algorithm, which smoothly extrapolates the projection transform of overlapping regions to non-overlapping regions from a shape correction perspective, resulting in a globally aligned image while retaining its original perspective. However, in situations where the overlapping regions consist of multiple intermediate planes, deriving a single global similarity transform from the global homography may not be adequate and may lead to unnatural visual effects in the stitching process for large parallax scenarios. Subsequent researchers have expanded upon the APAP and SPHP methods in various ways. One such method is the Adaptive As-Natural-As-Possible (AANAP) algorithm, which was proposed by Lin et al. [30]. This algorithm combines the best similarity transformation with a local single strain to reduce perspective distortion and correct shape. Additionally, some algorithms overlap regions to provide more natural parallax and alleviate distortion in non-overlapping regions [31]. Other algorithms add linear constraints to preserve the content contour structure [32, 33]. However, despite these improvements, all of these algorithms still use the traditional RANSAC algorithm to reject mismatching points. As a result, these improved algorithms are less robust and do not easily generalize to multiple images to obtain natural panoramas. Du et al. [34] proposed a stitching method to protect the geometric structure by first extracting various types of large-scale edges using a deep learning-based edge detection method and then sampling the extracted edges and constructing multiple sets of triangles to represent the geometric structure, which produces a panoramic image with a natural visual effect and less distortion. In multi-mono-strain parallax scenes, the approach of combining feature extraction and mismatch elimination, such as SIFT and RANSAC, may result in inadequate or erroneous matching, leading to noticeable ghosting. Moreover, these methods are incapable of addressing distortions in parallax scenes, such as intraoral images characterized by sharp variations in depth of field or abrupt changes. Hence, it is crucial to rectify parallax distortions and incorporate post-processing steps in oral image stitching to ensure a high degree of consistency in the output.

3. Proposed method

3.1. Overall architecture

The method proposed in this paper for generating naturalistic dental panoramas based on oral endoscopic imaging comprises two main stages: anti-perspective transformation feature extraction and coarse-to-fine global optimization matching. Figure 2. illustrates the overall architecture. Initially, a set of images designated for stitching is input. Subsequently, the improved SIFT feature extraction algorithm is utilized to obtain the anti-perspective transformation feature point set. Next, adaptive FLANN coarse matching based on neighborhood information (Vicinity-FLANN) and the RANSAC algorithm guided by homography are used to obtain a set of matched point pairs, effectively rejecting mismatching. Global optimization is then achieved using the L-M algorithm, ensuring a globally optimal panoramic result and computing the image transformation model with the refined matched point pairs. Finally, a multi-band hybrid image fusion algorithm is applied to generate the final panoramic stitched image. The primary objective of this method is to address issues like stitching breakage caused by sparse keypoints, parallax-induced ghosting, and distortion resulting from the accumulation of errors in multi-image stitching.

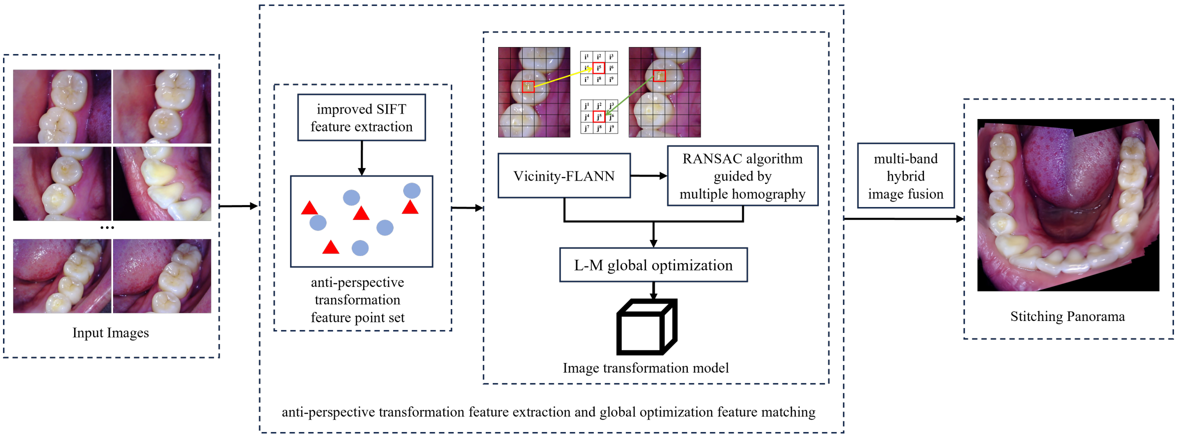

Input Images
improved SIFT feature extraction
anti-perspective transformation feature point set
Vicinity-FLANN
RANSAC algorithm guided by multiple homography
L-M global optimization
Image transformation model
anti-perspective transformation feature extraction and global optimization feature matching
multi-band hybrid image fusion
Stitching Panorama

Figure 2. The architecture of the proposed method is in this paper. The core part of the method, represented by the middle dashed box, consists of two stages: the anti-perspective transformation feature extraction and the coarse-to-fine global optimization feature matching.

3.2. Feature extraction algorithm based on anti-perspective transformation

The SIFT algorithm has been widely used for image feature extraction, which can find stable key points at various scales and orientations and calculate their descriptors. However, the performance of the SIFT algorithm can be compromised when dealing with handheld devices such as oral endoscopes, which leads to significant depth changes and perspective differences in the captured images due to slight shaking. Additionally, the SIFT algorithm does not possess full affine invariance, limiting its ability to extract image features with large spatial variations in the shooting angle. To address these issues, a novel feature extraction approach is proposed in this study that constructs an affine scale space and normalizes the elliptic neighborhood to obtain viewpoint transformation-resistant feature points. The proposed method demonstrates a significant improvement in matching point pairs, which enhances the SIFT algorithm's performance in resisting perspective changes.

3.2.1. Normalized elliptic region extraction

The shape of the local neighborhood can be articulated through the gradient distribution in terms of the second-order moment matrix of the grayscale gradients within the proximity of the pixel point. Consequently, the second-order moment matrix can be used to estimate the elliptical neighborhood structure of the feature points within the image, thereby facilitating the acquisition of affine invariance for the descriptor. For any given grayscale image $I(x)$, its affine Gaussian scale space is expressed.

$$L(x,\Sigma) = g(x,\Sigma) * I(x) \tag{3.1}$$

where $x \in R^2$; Σ is a symmetric semi-positive definite covariance matrix and corresponds to the scale of the eigenpoints. The non-uniform Gaussian kernel function $g(x,\Sigma)$ is defined as (3.2).

$$g(x,\Sigma) = \frac{1}{2\pi\sqrt{\det\Sigma}}\exp(-\frac{x^T\Sigma^{-1}x}{2}) \tag{3.2}$$

The second-order moment matrix in scale space at any point in the image is defined as (3.3).

$$\mu\left(x,\Sigma_I,\Sigma_D\right)=\left(\begin{array}{cc}\mu_{11} & \mu_{12}\\ \mu_{21} & \mu_{22}\end{array}\right)=\det\left(\Sigma_D\right)g\left(x,\Sigma_I\right)\times\left(\left(\nabla L\right)\left(x,\Sigma_D\right)\left(\nabla L\right)\left(x,\Sigma_D\right)^T\right) \tag{3.3}$$

where T is the transpose operator; Σ_I and Σ_D are the integral Gaussian kernel and differential Gaussian kernel of Gaussian kernel; ∇L is the gradient operator, defined in equation (3.4).

$$\nabla L\left(x,\Sigma_D\right)=\left(\begin{array}{c}L_x\left(x,\Sigma_D\right)\\ L_y\left(x,\Sigma_D\right)\end{array}\right) \tag{3.4}$$

To find the second order matrix by iterative method.

$$M=\mu\left(x,\Sigma_I,\Sigma_D\right) \tag{3.5}$$

$$\Sigma_I=\sigma_I M^{-1} \tag{3.6}$$

$$\Sigma_D=\sigma_D M^{-1} \tag{3.7}$$

To ensure accurate mapping of each sample point in the elliptical vicinity of feature points to their respective blocks, this study employed the second-order matrix of feature points to determine the parameters of the elliptical region. The transformation of all data in the elliptical image region to the circular region is achieved by applying the square root of said second-order matrix.

$$x'=M^{\frac{1}{2}}x \tag{3.8}$$

3.2.2. Affine Scale Space Construction

For a feature point on an already normalized circular neighborhood $I^{'}(x)$, the scale of the affine scale space image is close to the local scale σ of this feature point and can therefore be generated by convolution with the standard Gaussian kernel function.

$$L\left(x,\sigma\right)=G\left(x,\sigma\right)*I'\left(x\right) \tag{3.9}$$

$$G\left(x,\sigma\right)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^Tx}{2\sigma^2}} \tag{3.10}$$

This Gaussian kernel is symmetric, and its parameters are determined by a scale factor σ.The scale image can be obtained by Gaussian smoothing of the image, and this process can be repeated on the image that has been smoothed at the previous level to obtain a set of continuously varying images at different resolutions. The detection points are compared with the 26 adjacent pixel points above and below them. When a detection point is identified as a local extreme point, it is temporarily considered a feature point.

In the actual computation, the affine Gaussian difference scale space of the image is computed by subtracting two Gaussian images from the same set of adjacent scales, which can be expressed as follows.

$$DoG\left(x,\sigma\right)=\left(G\left(x,k\sigma\right)-G\left(x,\sigma\right)\times I'\left(x\right)\right)=L\left(x,k\sigma\right)-L\left(x,\sigma\right) \tag{3.11}$$

The points with smaller intensity values in the Gaussian difference scale space are removed, i.e.:$|DoG(x,\sigma)|<K$, where K is the threshold value, indicating the interval between two adjacent scale spaces.

3.2.3. Logarithmic polar coordinate mapping

Each pixel point on the normalized image can be expressed in right-angle coordinates (x, y) or in polar coordinates (r, θ). Let the coordinate origin be $(0, 0)$, then the two coordinates satisfy the following relationship.

$$z = x + yi = r(cos\theta + isin\theta) = re^{i\theta} \tag{3.12}$$

$$r = \sqrt{x^2 + y^2} \tag{3.13}$$

$$\theta = tan^{-1}\left(\frac{y}{x}\right) \tag{3.14}$$

Suppose the logarithmic polar coordinates of the points $z(x, y)$ are (ρ, φ) , and let $\omega = \ln z$, then the relationship between the right angle coordinates and the logarithmic polar coordinates can be deduced as (3.15-3.16).

$$\rho = lnr + i\theta = \frac{1}{2}ln\left(x^2 + y^2\right) \tag{3.15}$$

$$\varphi = \theta \tag{3.16}$$

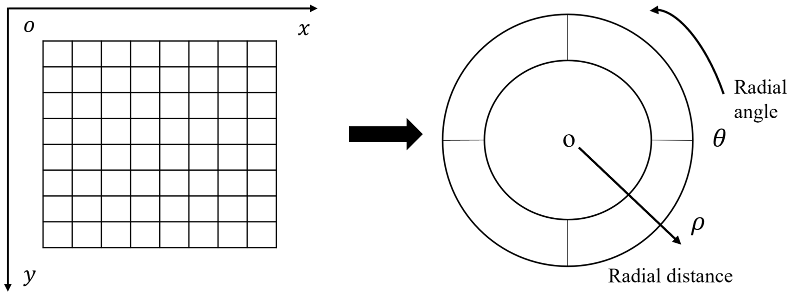
From the nature of logarithmic polar coordinates, it is known that the point $z(x, y)$ is assumed to be scaled r_0 times and rotated θ_0 degrees, then the transformed point is $z^{'}(x^{'} + y^{'})$, ρ, φ satisfying the following relation.

$$z'(x' + y') = r_0re^{i(\theta+\theta_0)} \tag{3.17}$$

$$\rho = lnr + lnr_0 \tag{3.18}$$

$$\varphi = \theta + \theta_0 \tag{3.19}$$

The equations (3.17-3.19) demonstrate that the manipulation of the scaling or rotation of the image in Cartesian coordinates corresponds to the translation of the image in log-polar coordinates in the vertical or horizontal direction. This relationship is depicted in Figure 3 and effectively resolves the issue of maintaining scale and rotation invariance in the image.



(a) adjacent region around feature (b) Sub region under polar space

Figure 3. Principle of logarithmic polar coordinate transformation.

3.2.4. Anti-perspective transformation descriptor generation

The SIFT descriptor calculates gradient histograms over square regions. However, when the image undergoes rotation, directly selecting the same square region to construct the descriptor will result in significant errors, as shown in Figure 4. To address this issue, we consider that the pixels contained in circular regions are entirely consistent across the different main directions. Therefore, we utilize the logarithmic polar coordinate method to compute the gradient histogram over an octagonal region. The main steps of this approach are described as follows.

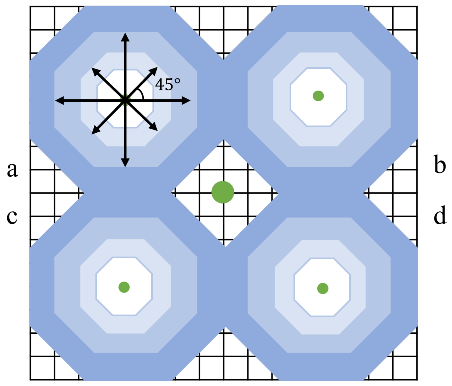


Figure 4. Schematic diagram of key point descriptors, a set of 16 × 16 panes encircling the feature points are selected. Each pane is representative of a single-pixel point. The neighboring 16×16 pixel points encompassing the key point that corresponds to the Gaussian image are separated into 4 sub-regions.

Step 1. To achieve rotation invariance of feature points, a direction reference is assigned to each local feature of the image. The direction information of the feature point is described using the gradient of the feature point's neighborhood. The formula for calculating the magnitude and direction of the pixel gradient is as follows.

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \tag{3.20}$$

$$\theta(x,y) = arc\tan\left[\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right] \tag{3.21}$$

where $m(x,y)$ is the gradient modulus of the pixel point and $\theta(x,y)$ is the direction of the pixel point.

Step 2. To ascertain the requisite image area for the descriptor, a set of 16×16 panes encircling the feature points are selected. Each pane is representative of a single pixel. The neighboring 16×16 pixel points encompassing the key point that corresponds to the Gaussian image are separated into 4 sub-regions. The key point descriptors are presented in Figure 4 for further clarification.

Step 3. Take the subregion a with 8×8 pixels in the upper left corner as an example. First, find the center pixel point of the sub-region, consider the pixel point as the origin, and construct four octagonal rings emanating. Second, use 0,45,90,135,180,225,270,315,360 as the direction of the descriptor feature vectors, calculate the feature vectors of 8 directions in the 4 octagonal rings respectively, and superimpose the $4 \times 8 = 32$ dimensional feature vectors to the center pixel point of the sub-region. Finally, the feature vectors of the 4 sub-regions are superimposed on the key points according to the distance weights to generate the key point descriptors.

$$H_a(m,\theta) = \sum_{i=1}^{32} h_i\left(m_i, \theta_i\right) \tag{3.22}$$

Where $h_i(m_i, \theta_i)$ is any vector of the octagonal ring of the subregion; $H_a(m,\theta)$ is the vector descriptor of the subregion a.

$$H_0(m,\theta) = H_a + H_b + H_c + H_d = \sum_{j=1}^{128} h_j(m_j, \theta_j) \tag{3.23}$$

Where $H_0(m,\theta)$ is the 128-dimensional vector of key point descriptors.

The key point descriptors within a sub-region are produced through the utilization of four octagonal rings, each with varied distance weights in eight directions. Compared with the original SIFT algorithm, which uses the generated 16 seed points to calculate the key point descriptors. Therefore, this method enhances the overall accuracy of feature point matching.

3.3. Coarse-to-Fine Global Optimized Feature Matching Algorithm

After describing the feature points, it is necessary to match the feature vectors to determine the specific parameters of the image transformation model. However, the traditional SIFT approach for one-way feature matching through nearest neighbors leads to a high false matching rate and adversely impacts the image alignment due to one-to-many matching. This paper presents a feature-matching algorithm that is founded on global optimization. Firstly, the feature points are roughly matched by the adaptive FLANN matching algorithm with neighborhood information, The homography matrix is then estimated robustly through the RANSAC algorithm, and the matching is guided by the homography matrix under a given parallax threshold until a stable number is reached. This matching process is repeated for the remaining feature points until the RANSAC algorithm estimate exceeds the limit. Finally, the L-M algorithm is employed for achieving global optimization, where the cumulative error can be reduced and the splicing distortion can be corrected by continuously adjusting the model parameters of each image according to the matched pairs of points between neighboring images to minimize the mean-square error of the distance between matched pairs in order to minimize the global mean-square error and achieve global optimization.

3.3.1. Vicinity-FLANN coarse matching

The conventional FLANN algorithm typically employs a fixed threshold to determine whether matching point pairs are suitable, requiring multiple traversals of the entire index to locate matching point pairs that satisfy the threshold condition, which is rather time-consuming. As a result, this paper proposes an adaptive threshold that is designed by analyzing the data of the initial feature point set. Moreover, the Vicinity-FLANN algorithm is improved by utilizing the constraint information of neighboring points.

Suppose the set of feature points of the two images are P_1 and P_2. For each feature point in P_1 , find the two closest Euclidean distances to P_1 ,denoted as d_i' and d_i'' , respectively, and the number of feature matching pairs is denoted as N . The average of the differences in Euclidean distance is calculated as follows.

$$avgd = \frac{(\sum_{1}^{N}(d_i'' - d_i'))}{N} \tag{3.24}$$

If the difference between the nearest Euclidean distance and the second nearest Euclidean distance of the detected points is less than the average of the distance differences, they are retained. After all the feature points are detected, a set of matched point pairs is obtained.

$$d_i' > d_i'' - avgd \tag{3.25}$$

Adding the motion smoothness constraint to feature matching reduces the feature matching area. As shown in Figure 5. In the motion space, correct matches are smooth, and adjacent features with consistent motion have adjacent areas for correspondingly matched features. Therefore, after searching for matching features in the reference image for the given feature point in the image to be registered, all the feature points in the neighborhood of that feature point only need to be searched for the matching point in the image to be registered. This constraint information can improve the accuracy of matching and can better handle complex situations such as noise, overlap, and shelter in dental images.

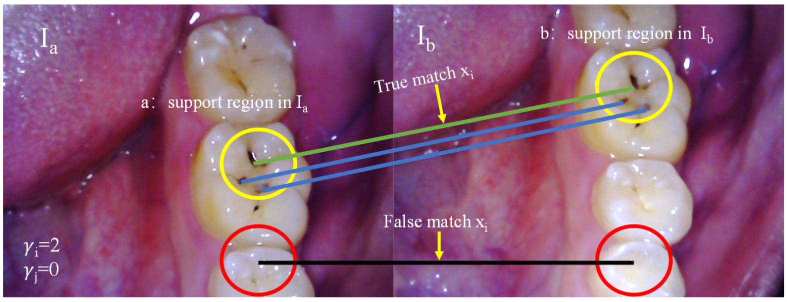


Figure 5. Distribution of correct matching and incorrect matching features. The motion smoothness constraint principle states that true correspondences often have more similar neighbors than false correspondences, so we count the number of similar neighbors to separate them.

To facilitate neighborhood selection, a uniform grid of $n \times n$ is applied to both the reference image and the image to be matched, whereby matching points are sought through the grid. The matching process is graphically represented in Figure 6, and feature points that fall within the same neighborhood as feature point i necessitate searching for matching points only within the 3×3 vicinity of feature point j. As a result, the matching search range is reduced, thereby ensuring the real-time performance of the coarse matching algorithm.

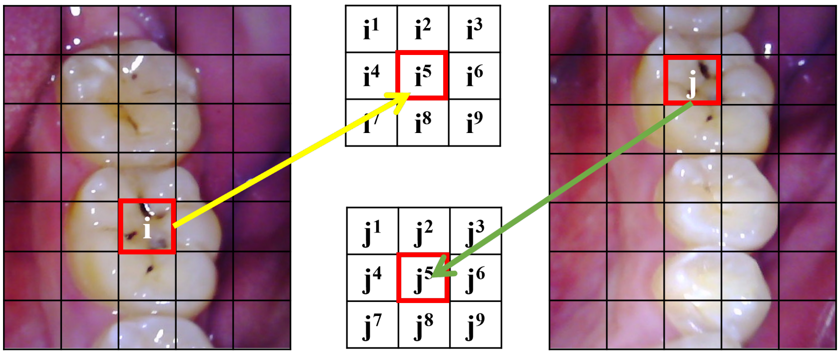


Figure 6. Schematic diagram of feature matching. In calculating the feature points in the same neighborhood of the feature point i, we consider the 3×3 neighborhood as the feature point j.

3.3.2. Homography-guided RANSAC fine matching

Regardless of the feature-matching methods employed, mismatching is difficult to avoid. In the present study, we propose the use of the single-strain guided RANSAC algorithm to accurately estimate the single-strain H and its corresponding inner points while effectively eliminating the outer points. Additionally, the homography model is selected to accommodate the transformation of image pairs captured with varying imaging models, such as scaling, translation, and rotation. The point $p(x, y)$ is multiplied with the projection matrix H to obtain the point $P'(x', y')$, and the homography matrix can be expressed accordingly.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3.26}$$

where the projection matrix H is a 3×3 matrix.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \tag{3.27}$$

The RANSAC algorithm calculates a homography matrix H' from the set of N pairs of matching points after coarse matching with adaptive FLANN based on neighborhood information. Then it randomly selects four pairs of data points and calculates the distance between them and the transfer points in accordance with H' for the remaining $N - 4$ data points in the dataset. The obtained inner points are subsequently removed, and the remaining feature points are matched using H' for bootstrap matching to obtain additional matching pairs. Notably, the search area of the transfer points is defined using the solved $H^{'}$ given the maximum parallax, to reduce the search range of matching. This enables guided matching, which significantly increases the number of matching points and improves overall accuracy.

$$P(x, y)' = H'p \pm \gamma \tag{3.28}$$

where $P(x, y)$ is the coordinate of the point in the image, $P(x, y)' = [H'p - \gamma, H'p + \gamma]$ is the search area of the transfer point.

Owing to the constraints of oral space size, parallax, and depth, changes are easily produced by translation and rotation operations during the shooting process. By suitably adjusting the parallax threshold, a considerable number of matching pairs can be accurately obtained. Subsequently, $H^{'}$ prime guidance is applied to Vicinity-FLANN coarse matching. This results in the acquisition of many correct matching pairs owing to the narrowed search range and the high uniqueness of the improved SIFT features. The previously outlined steps are iterated until the number of interior points obtained falls below the pre-set threshold, at which point the current feature point matching is considered complete.

3.3.3. Global tuning of the L-M optimization algorithm

The process of sequence image stitching may result in substandard quality because of the accumulation of matching errors. To ensure the attainment of a dependable and highly accurate transformation matrix, this research paper proposes the use of the homography-guided RANSAC algorithm, thereby achieving a fast and robust estimation of the initial transformation matrix H. In order to reduce the splicing distortion caused by the cumulative error, this paper proposes a global optimization method based on the L-M algorithm. The proposed method leverages the initial projection parameters of each image, employing iterative computation through the L-M algorithm. Additionally, it dynamically adjusts the damping factor μ throughout the iteration process to consistently update both the iteration direction and step size. Consequently, this method identifies the optimal set of projection parameters by minimizing the sum of error distances across all feature points in all images following the projection transformation, so that global adjustments can be made to reduce the accumulation of errors and improve the quality of panoramic images.

The average geometric distance offset E of the matched feature point pairs optimized by the RANSAC algorithm is chosen as the criterion for the optimized transformation matrix, and this input error function $E(H_k)$ is calculated with the following equation.

$$E(H_k) = \sum_i \frac{d(q_i', Hq_i)}{n} = \frac{1}{2}\sum_{i=1}^{n} e_i^2(H) \tag{3.29}$$

Where i is the pixel point serial number,q_i is the coordinate value of the pixel point with serial number i ,q_i' prime is the coordinate value of the pixel point corresponding to q_i , $e_i(H)$ is the single residual of the error function, n is the number of all matched feature point pairs, $d(q_i', Hq_i)$ and denotes the Euclidean geometric distance between q_i' and q_i , and the smaller the average geometric The smaller the distance offset, the more correct the matching relationship is proved. H is the homography matrix obtained after matching, which can be transformed into the internal and external parameters of the camera.

In order to avoid the drawback of matrix singularities in Eq. and thus the breakage of the algorithm, the L-M algorithm improves the Gauss-Newton method, which takes the form of.

$$\Delta H = -(J(H)^T J(H) + \mu I)^{-1} J(H)^T e(H) \tag{3.30}$$

where μ is the scale factor, whose presence allows the L-M algorithm to approximate the global optimal solution more accurately. I is the unit matrix introduced to prevent the integrability of the iteration step matrix, and $J(H)$ is the Jacobian matrix of $e(H)$.

$$J(H) = \begin{bmatrix} \frac{\partial e_1(H)}{\partial m_0} & \cdots & \frac{\partial e_1(H)}{\partial m_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial e_N(H)}{\partial m_0} & \cdots & \frac{\partial e_N(H)}{\partial m_N} \end{bmatrix} \tag{3.31}$$

Determine if the iteration step is smaller than the pre-set iteration step threshold; if it is smaller, stop outputting H_k and vice versa continue. Use the iteration step to update the input parameters of the error function. Assuming that H_k is the vector consisting of the weight and threshold at the kth iteration, the update can be performed by calculating equation (3.32).

$$H_{k+1} = H_k + \Delta H \tag{3.32}$$

Specifically, the L-M algorithm computes the partial derivatives of the error function with respect to the parameters, i.e., the Jacobian matrix, at each iteration. Then, by modifying and weighting the Jacobian matrix and combining it with the error vector, an augmented matrix is formed. The iterative formula of the L-M algorithm is used to calculate the parameter increments, which are then added to the current parameter estimate to obtain a new parameter estimate. By continuously updating the parameters, the objective function is minimized, thereby minimizing the error between the predicted values of the model and the actual observed values.

3.4. Multiband hybrid image fusion

After using the L-M algorithm to optimize the camera parameters of multiple images, the spatial position and rough stitching results of these images can be obtained. However, the stitching results yielded from this approach cannot be directly implemented as an outcome. This is due to the lack of assurance that all pixel points in the overlapping areas of multiple images are perfectly aligned, leading to the presence of stitching traces and breaks at the image boundaries, ultimately compromising the quality of dental panoramas and the visual experience. Direct blending often results in undesirable stitching lines, discontinuous color transitions, and blurry ghosting, rendering it unsuitable for practical applications. Linear blending, while effective in ideal conditions, is often problematic in parallax images, with a tendency to cause loss of oral image details and an unnatural appearance of the fused image. In view of these limitations, this paper proposes the utilization of multi-band blending to fuse the overlap. By segmenting the images into high and low frequencies in the frequency domain, which correspond to the details and contours in the images. Multi-band blending constructs a Laplace pyramid for the input image to obtain a better fusion effect and then fuses the images in the same layer according to the Alpha blending/feathering rules. The resultant image is obtained by reconstructing the fused pyramid, with the Laplacian pyramid of the final image formed as equation (3.33).

$$Y_k(i, j) = X_{1,k}(i, j)M_k(i, j) + X_{2,k}(i, j)(1 - M_k(i, j)) \tag{3.33}$$

$X_{1,k}$ and $X_{2,k}$ denote the kth level of the Laplacian pyramid decomposition of two images, which are adjusted for coordinates.Y_k represents the kth level of the Laplacian pyramid decomposition of the resulting fused image, while Y_k stands for the kth level of the Gaussian pyramid decomposition of the image mask.

The technique of pyramid blending smoothly blends the image's low frequencies while ensuring a more distinct transition for the high frequencies. This process effectively minimizes ghosting effects in the result and enhances the smoothness of the stitched image, resulting in a more natural transition in the stitched region.

4. Experiments

The testing environment for this experiment is Windows 10 (a 64-bit operating system), AMD Ryzen 7-5800H 3050Ti @ 1.90 GHz, 16 GB of memory, and OpenCV 3.4.5 integrated with PyCharm 2021.

In this experiment, the image size is 800 x 600 pixels, and the grid is set to 10 x 10. After conducting several tests, the thresholds for the ratio of the nearest neighbor distance and the next nearest neighbor distance of the feature points have been set to 0.7. The value of the residual terms per match used in the bundle adjustment is set to 2. Additionally, the damping factor in the L-M optimization algorithm is set to 5, and the maximum number of iterations in the optimization process is limited to 100. Furthermore, the initial error threshold for matching is set to 150. During the bundle adjustment process, matches with an initial error (residual) greater than this threshold are considered potential mismatches and may be removed from the optimization to enhance the accuracy of the final solution.

4.1. Dataset

In this study, we have constructed a small-scale oral endoscopic image dataset called S-EDD (Self-Established Endoscopic Dental Image Dataset). The dataset was created using a wireless oral endoscope to capture localized image samples on the surface of the mandibular teeth. The setup for capturing these images is illustrated in Figure 7. The endoscope lens used in the dataset has a diameter of 7.0 mm and a viewing angle of 60 degrees. To achieve optimal image quality, the focal length for capturing the images was set to 1.5 cm.

The S-EDD dataset comprises a total of 27 sets of image data, consisting of 2116 images, and 70 sets of video data. These images and videos cover various dental perspectives, including molars, premolars, incisors, and multiple other angles. The data modality for this dataset is RGB. Each group of video clips can be divided into picture frames at different rates. The image size is set to 800*600 pixels, and all images are saved in jpg format. Each image in the dataset contains 2-7 teeth, and we ensured an overlap rate exceeding 70% between adjacent images to contribute to more effective and reliable oral dental image stitching.

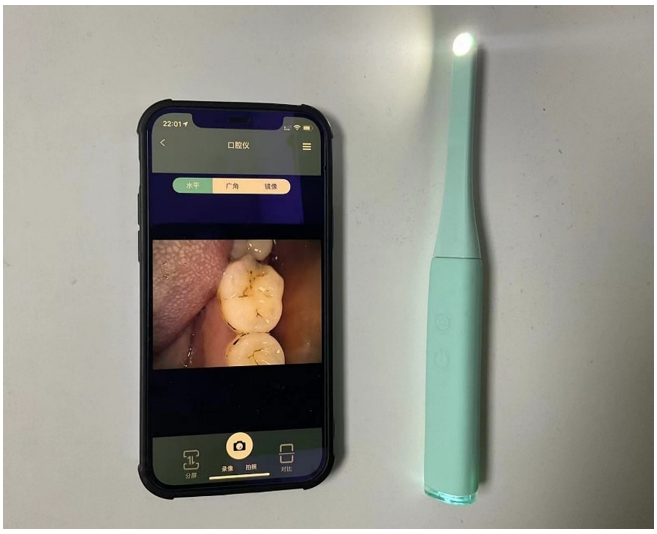
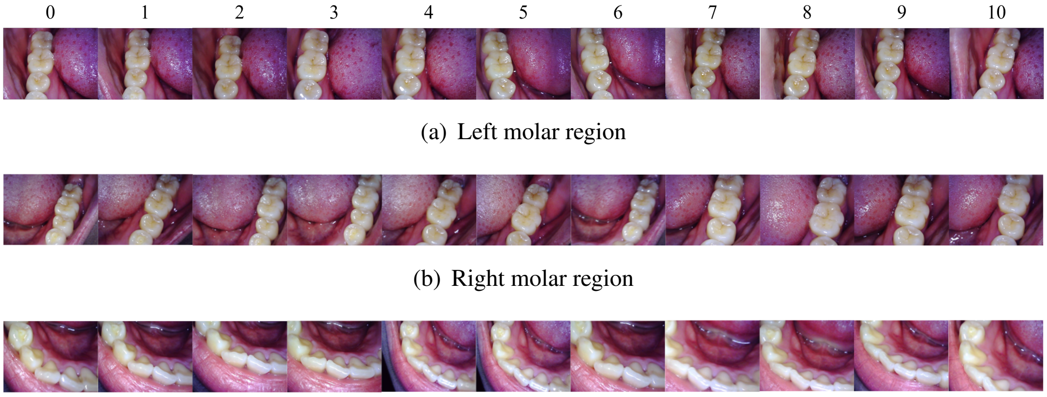


Figure 7. Oral endoscopy model. The datasets tested in this experiment are all derived from the acquisition of localized images captured by the device.

4.2. *Analysis of image matching results*

To verify the practicality of the proposed algorithm, many real images were taken using an oral endoscope, and two images with different characteristics were selected for multiple sets of experiments. Part of the test dataset is shown in Figure 8, containing 10 sets each for the left molar region (first premolar, second premolar, first molar, second molar), the right molar region, and the anterior incisor region (lateral incisor, middle incisor, canine), including changes in translation, rotation, perspective, and overlap rate. Figure 9 shows the matching accuracies of SIFT [16], ORB [18], GMS [20] and our proposed method for 10 sets of mandibular teeth images with different angles. Our proposed method has higher matching accuracy than the other three algorithms, and the average correct rate is over 87%.

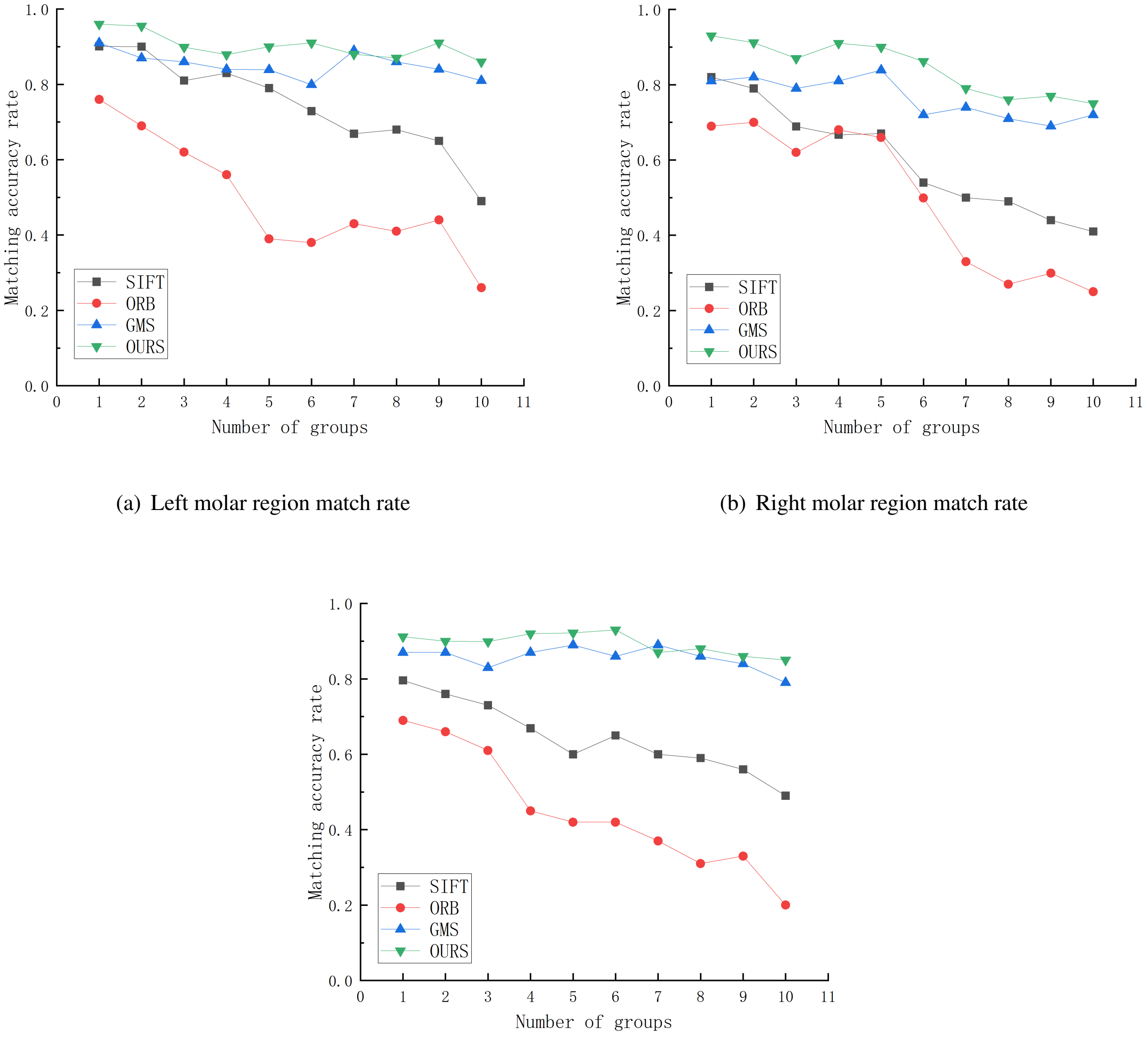


(a) Left molar region

(b) Right molar region

(c) Anterior incisor region

Figure 8. Example of test images from the S-EDD dataset. (a) Left molar region; (b) Right molar region; and (c) Anterior incisor region. The three different region images using image number 0 as the base image demonstrate the variations in translation, rotation, viewing angle, and overlap rate contained between the 10 sets of images from number 1 to number 10.



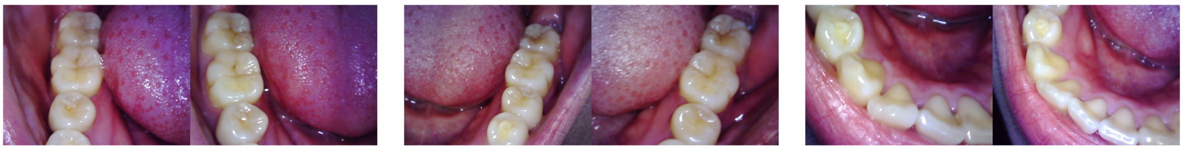
(a) Left molar region match rate

(b) Right molar region match rate

(c) Anterior incisor region match rate

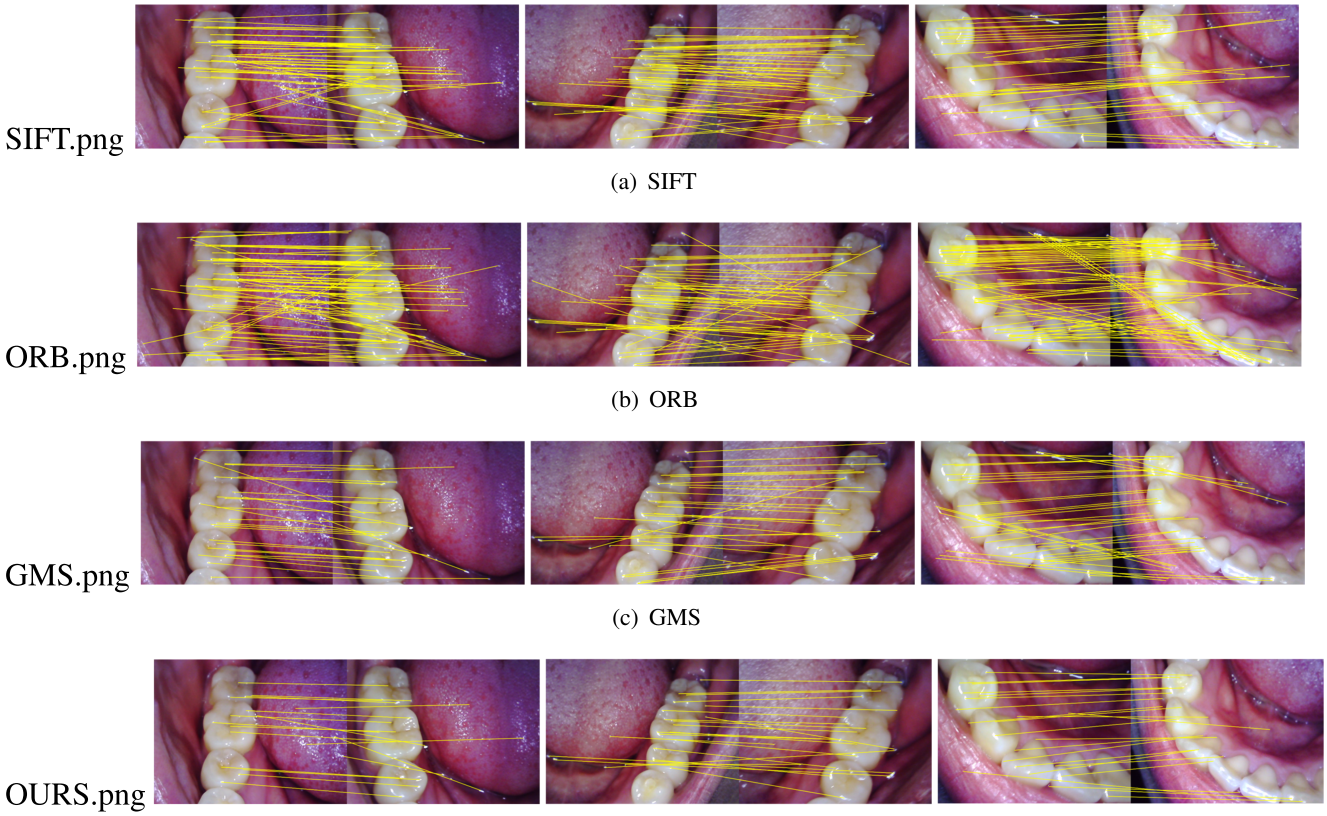
Figure 9. Correct matching rate of SIFT, ORB, GMS, and the method of this paper in (a) Left molar region; (b) Right molar region; and (c) Anterior incisor region of dental images in the oral cavity. The horizontal axes 1 to 10 correspond to the number of sets matched with different matching algorithms in the 0th and the next 10 sets of each test image in Figure 8, and the vertical axes are the matching accuracy obtained.

In Figure 10, Set A is the left molar image to be matched, Set B is the right molar image to be matched, and Set C is the anterior incisor image to be matched. All with an image size of 800600 pixels. Figure 11 shows the matching results of the three groups of different regions of tooth images using SIFT [16], ORB [18], GMS [20], and the algorithm of this paper.



(a) Left molar to be matched (b) Right molar to be matched (c) Anterior incisor to be matched image

Figure 10. The input images of three sets of different regions were used to verify the image matching effect.



(a) SIFT

(b) ORB

(c) GMS

(d) OURS

Figure 11. The matching results of SIFT, ORB, GMS, and our proposed method on three sets of different dental image regions are shown. Taking the first column as an example, it illustrates the feature matching results of the four methods on the left molar region in Figure 10(a). Similarly, taking the first row as an example, it represents the feature matching results of the SIFT algorithm on the three sets of different dental images in Figure 10.

The results presented in Figure 11 demonstrate that the SIFT algorithm yields evenly distributed matching point pairs in the three sets of dental images from different regions. However, the correct point pairs are relatively sparse, and false matches are more conspicuous. On the other hand, the ORB algorithm, when applied without inner point filtering, generates a higher number of feature points, but this comes at the cost of a substantial number of false matches. In contrast, both the GMS algorithm and the proposed method yield evenly distributed matching points and exhibit satisfactory performance.

However, our proposed method outperforms the GMS algorithm in terms of generating fewer false matches and producing stable matching points, thereby yielding an overall better matching result.

To further quantify the superiority of the proposed method objectively, the test results obtained by the proposed method and three other algorithms were analyzed using the correct matching rate (CMR). According to Table 1, the proposed method has demonstrated a 2-7% increase in the correct matching rate compared to the GMS algorithm across all three image sets. The proposed method leveraged the RANSAC algorithm guided by homography to eliminate incorrectly matched points, which ultimately enhanced the accuracy and stability of the matching process. Compared to the ORB algorithm, the proposed method has shown an impressive 30% increase in the correct matching rate across all three image sets. This is attributed to the proposed method's ability to construct a grid using constraint information from neighboring points, which allows for a better distinction between correct and incorrect matches. Compared to the SIFT algorithm, the proposed method has demonstrated a more than 10% increase in the correct matching rate across all three image sets, as the RANSAC algorithm guided by homography helps to eliminate incorrectly matched points after the FLANN coarse matching based on neighborhood information. The GMS algorithm has the shortest running time, whereas the proposed method has a slightly higher running time than the GMS algorithm yet maintains the simplicity and speed characteristics of the GMS algorithm while having a significantly shorter running time than the SIFT and ORB algorithms. In the oral endoscopic environment, the proposed algorithm has a higher false-match rejection rate, which provides assurance for the subsequent application of image matching.

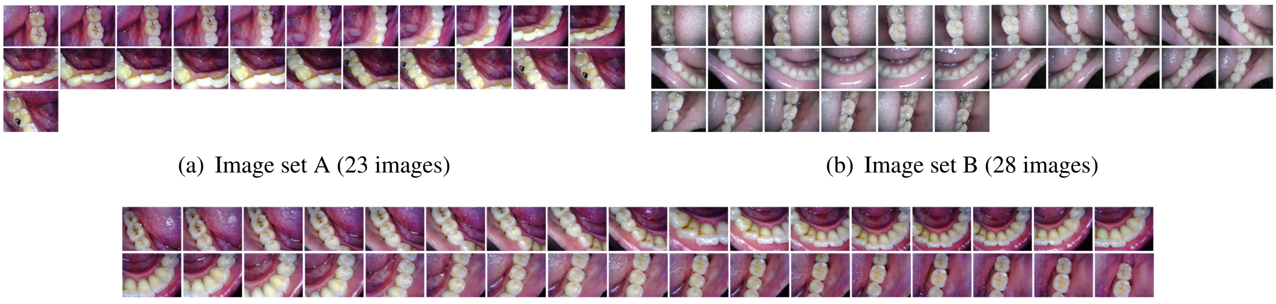
Table 1. Comparison of CMR of different methods.

Experimental	Algorithm	Matching	Correct	CMR /%	False match	Matching
images		Points	Match Points		rejection time(s)	Total time(s)
Set A	SIFT	108	86	79.63	0.0047	0.5214
	ORB	198	79	39.90	0.0052	0.1327
	GMS	73	62	84.93	0.0049	0.0403
	OURS	65	59	**90.77**	0.0046	0.0620
Set B	SIFT	113	75	66.37	0.0073	0.9213
	ORB	176	122	69.32	0.0083	0.2840
	GMS	84	69	82.14	0.0077	0.0467
	OURS	70	63	**90.00**	0.0073	0.682
Set C	SIFT	107	71	66.36	0.0053	0.4879
	ORB	203	88	43.35	0.0049	0.1928
	GMS	92	82	89.13	0.0050	0.0279
	OURS	79	72	**91.14**	0.0049	0.0458

Note:Bold font is the best value for each column.

4.3. Analysis of panorama stitching results

The original image sequence of 3 sets of input images acquired by the oral endoscope, consisting of 23 image sets A, 28 image sets B, and 34 image sets C, is presented in Figure 12. The image dimensions for all sets are 800 600 pixels. The shaking of the handheld camera during the shooting process results in a change in the image view angle, and the overlapping position and overlapping area size between each image are not fixed. Current image stitching algorithms are limited to two-image stitching, and the accumulation of deformation errors during multi-image stitching can cause a breakdown in the stitching process, leading to an incomplete oral panoramic image, so the classical SIFT [16], Autostitch [25], SPHP [29], Method of Ref.[11], and GES-GSP [34] algorithms and the proposed method are selected here for three sets of experiments.



(a) Image set A (23 images)

(b) Image set B (28 images)

(c) Image set C (34 images)

Figure 12. Three sets of input image sequences.

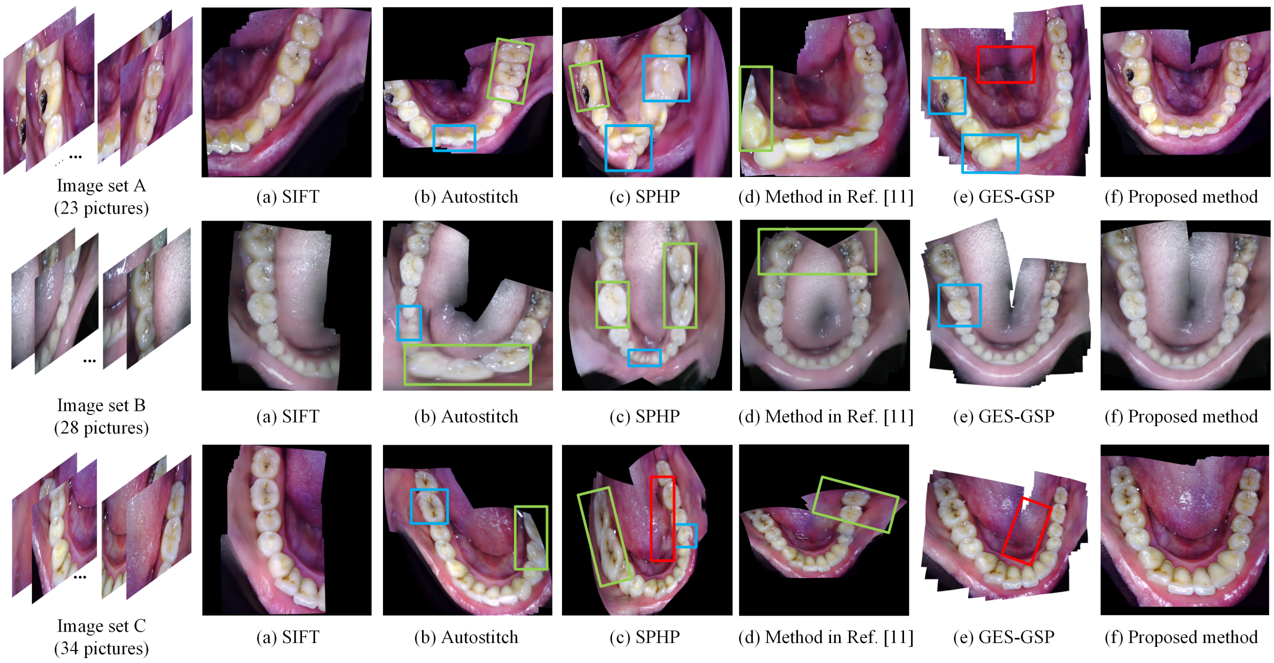


Image set A (23 pictures)

(a) SIFT

(b) Autostitch

(c) SPHP

(d) Method in Ref. [11]

(e) GES-GSP

(f) Proposed method

Image set B (28 pictures)

(a) SIFT

(b) Autostitch

(c) SPHP

(d) Method in Ref. [11]

(e) GES-GSP

(f) Proposed method

Image set C (34 pictures)

(a) SIFT

(b) Autostitch

(c) SPHP

(d) Method in Ref. [11]

(e) GES-GSP

(f) Proposed method

Figure 13. The stitching effect of a panorama of three sets of sequence images. Blue boxes were assigned to mark misalignments and overlays; green boxes were employed to identify obvious distortions; and red boxes were used to mark incorrectly filled pixels. (a) The SIFT [16] resulted in a splicing fracture and was unable to generate a panorama. (b) AutoStitch [25], (c) SPHP [29], (d) Method of Ref.[11] and (e) GES-GSP's [34] results have different degrees of dislocation, deformation, artifacts, and so on. (f) Our results obtained a complete and natural oral dental panorama.

When this study is compared with other methods, a multi-band hybrid algorithm is used as a post-processing technique in the image fusion stage for a fair comparison. As depicted in Figure 13, all three sets of results were obtained through the SIFT algorithm's splicing break at the Canine and incisors. The oral environment comprises relatively single and similar feature points, and few key points are matched in the image alignment session, which results in the computed projection transformation matrix not being able to characterize its mapping relationship well. Notably, when the number of matched point pairs is less than 4, the necessary conditions for the simulation of the projection transformation matrix model cannot be met, and thus there is an interruption and the panoramic effect cannot be synthesized. Similarly, all three sets of results obtained through the AutoStitch algorithm underwent varying degrees of fracture. The first and third sets of experiments identified misalignment and overlap in the incisal region and the molar region, respectively, due to insufficient alignment accuracy, whereas in the second set of experiments, significant distortion was observed. The SPHP algorithm globally aligns the images while maintaining the original perspective by combining the projection transform with a similar transform; however, it does not perform well in the oral environment. The first set of experiments generated a complete panorama; nevertheless, there were apparent misalignments and distortions, and the diseased teeth were stretched longitudinally. In the second set of experiments, misalignments of the tongue and teeth were evident, resulting in the teeth and tongue overlapping incorrectly, and the teeth in the middle of the molar region were stretched longitudinally. In the third set of experiments, the teeth with lesions on the left side were noticeably distorted, and the teeth and tongue on the right side were partially filled with the wrong pixels. The Method of Ref. [11] uses local feature point symmetry constraints for image stitching, and in the first and third sets of experiments, there were different degrees of breakage as well as deformation of the molar region. In the second set of results, although a complete panorama was obtained, the molar region on both sides was contracted and deformed, which deviated from the normal Dental Arch morphology. The GES-GSP algorithm, a geometrically protected splicing method, is based on the edge of the deep-learning detection method to extract tooth edges; however, it also suffers from slight misalignment and is prone to erroneous pixel filling in soft tissues such as the tongue. However, the proposed method generates panoramic images that do not distort the number of images, exhibit better stability, and better preserve dental information and lesion details. Thus, the proposed method is more suitable for dental auxiliary diagnosis.

To provide a more objective assessment of the efficacy of the proposed method, Root Mean Square Error (RMSE) was utilized as a metric to quantify the alignment accuracy of SIFT, AutoStitch, SPHP, Method of Ref.[11], GES-GSP and the proposed method feature point pairs. RMSE is a widely accepted measure of image alignment within the realm of image stitching. Table 2 provides an account of the RMSE values associated with the three test image sets depicted in Figure 13.

From the data in Table 2, it is evident that the performance of the proposed method surpasses that of the customary algorithm. The limited and similar texture attributes of dental images have resulted in mapping errors during stitching, which manifest as discontinuous transitions in the image. While the Autostitch technique exhibits a marginally lower RMSE than our method in the first set of experimental results, it produces suboptimal visual outcomes with stitching discontinuity. The SPHP algorithm rectifies global deformation to ensure the completeness of panoramic stitching, but its quantitative evaluation outcomes are unsatisfactory. Unsupervised splicing networks in the literature [11] rely on dental datasets and have high quality requirements, and multi-graph splicing is prone to distortion, which reduces the generalization ability of the network. GES-GSP has excellent performance in preserving geometric structures but is slightly less effective than our method when applied to the oral environment. In conclusion, the proposed method achieves high stitching accuracy and superior visual outcomes and outperforms SIFT, Autostitch, SPHP, Method of Ref. [11] and GES-GSP in oral dental images.

Table 2. Comparison of RMSE of different methods.

Experimental images	SIFT	AutoStitch	SPHP	Method in Ref. [11]	GES-GSP	Proposed method
Image set A (23 pictures)	23.1516	**4.2636**	11.6853	9.3365	5.9861	4.3511
Image set B (28 pictures)	20.4127	12.4709	10.3231	5.6004	4.0988	**3.4289**
Image set C (34 pictures)	28.8942	15.7361	13.0995	7.4435	4.6992	**4.6925**

Note:Bold font is the best value for each column.

Furthermore, this study selects a set of video data to implement the frame-taking operation and obtains 248 frame images. Six different methods were used to stitch them together to compare the success rate of stitching. To assess the image stitching outcomes in a quantitative manner, the results were classified into three grades, namely A, B, and C. Successful stitching was defined as achieving a grade of B or higher. Grade A means an $RMSE \leq 5$, Grade B means $5<RMSE \leq 15$, whereas Grade C means $RMSE<15$ or cannot be spliced. The stitching success rate is presented in Table 3.

Table 3. Comparison of stitching success rate of different methods.

Algorithm	Grade A / Picture	Grade B / Picture	Grade C / Picture	Stitching success rate %
SIFT	47	9	192	22.58
AutoStitch	105	58	85	65.73
SPHP	147	51	50	79.84
Method in Ref. [11]	198	29	21	91.53
GES-GSP	231	9	8	96.77
Proposed method	234	8	6	**97.58**

Note:Bold font is the best value for each column.

From the data in Table 3,we can see that the success rate of the existing stitching methods is very low for dental images, the quality of the stitched images is poor, and some of the images cannot be stitched because the number of matching pairs of feature points is less than 4 pairs. In contrast, the success rate of the global optimal generation method of the dental panorama against perspective transformation proposed in this paper is much higher than the other five methods, at 97.58%.

4.4. Ablation studies

The anti-perspective transformation feature extraction and the coarse-to-fine global optimization matching are the core parts of our proposed method. Therefore, we compare the homology estimation performance of the two modules with and without (w/o) on the self-constructed oral endoscopic image dataset S-EDD, and the evaluation metric is 4pt-Homography RMSE. As shown in Table 4 ”(w/o) anti-perspective transformation”, the application of anti-perspective transformation feature extraction can effectively increase the number of matched pairs of points, which improves the homology solving accuracy. Obtaining the set of anti-perspective transformation feature points can help the model select feature points that are more reliable in homology estimation. In addition, the effectiveness of the global optimization strategy designed in the second stage is tested by removing the coarse-to-fine feature matching and L-M global optimization adjustment. As shown in Table 4 ”(w/o) coarse-to-fine global optimization”, the addition of coarse-to-fine feature matching and L-M algorithm global optimization helps to improve the matching accuracy, reduce the cumulative error, and ensure the optimal solution of the panorama results globally.

We rank the stitching results obtained from our experiments, where Top 0-30% denotes the results ranked in the top 0% to 30%, i.e., good quality splicing results, 30%-60% denotes medium quality splicing results, and 60%-100% denotes poor quality splicing results.

Table 4. Ablation studies on the anti-perspective transformation feature extraction and the coarse-to-fine global optimization matching.

Level	(w/o) anti-perspective transformation	(w/o) coarse-to-fine global optimization	Proposed method
Top 0-30%	4.8544	4.0083	**3.2810**
30%-60%	5.9025	4.9724	**4.0557**
60%-100%	7.3964	7.0818	**6.3732**
Average	6.2006	5.3761	**4.7339**

Note:Bold font is the best value for each column.

5. Conclusions

This article proposes a method for generating panoramic images of teeth through image stitching. The problem of low registration accuracy resulting from the lack of features and large disparities in oral endoscopic images is addressed. The effectiveness of the proposed method is verified through metrics such as CMR, RMSE, and stitching success rate. Anti-perspective transformation feature descriptors are used to extract features, and a coarse-to-fine global optimization matching approach is employed to improve feature matching accuracy. The average matching accuracy exceeds 87%, which is significantly higher than that of traditional SIFT and ORB algorithms. The proposed method also has fewer mismatching points and more stable matching points than the GMS algorithm. The RANSAC algorithm, guided by homography, is used to remove mismatching points, and the L-M algorithm is used to optimize feature point pairs, which improves the accuracy of the image transformation model. The global optimization strategy used in this article leads to better visual effects than SIFT, Autostitch, SPHP, Method of Ref. [11] and GES-GSP algorithms. The experimental results demonstrate that the proposed method effectively addresses issues such as stitching breaks, unnatural stretching, and deformation, and meets the requirements of practical applications in diagnostic assistance.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was partly supported by the Natural Science Basis Research Plan in Shaanxi Province of China under Grant 2023-JC-YB-517, and the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under Grant VR-LAB2023B08.

Conflict of interest

All of the authors declare that there is no conflict of interest regarding the publication of this article and would like to thank the anonymous referees for their valuable comments and suggestions.

References

1 Y. Yan, L. Xu, Y. Y. Liu, X. Su, J. L. Gao, M. X. Wan, Larynx ultrasound image stitching based on multiconstraint super-pixel feature, *IEEE Trans. Instrum. Meas.*, **72** (2023), 1–11. https://doi.org/10.1109/TIM.2023.3243675

2 M. Ghanoum, A. M. Ali, S. Elshazly, I. Alkabbany, A. A. Farag, Frame stitching in human oral cavity environment using intraoral camera, in *IEEE International Conference on Image Processing (ICIP)*, (2019), 1327–1331. https://doi.org/10.1109/ICIP.2019.8803071

3 M. Ghanoum, A. M. Ali, S. Elshazly, I. Alkabbany, A. A. Farag, Panoramic view of human jaw under ambiguity intraoral camera movement, in *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, (2020), 1–4. https://doi.org/10.1109/ISBI45749.2020.9098562

4 T. Jin, F. Z. Li, H. J. Peng, B. Li, D. P. Jiang, Uncertain barrier swaption pricing problem based on the fractional differential equation in Caputo sense, *Soft Comput.*, **27** (2023), 11587–11602. https://doi.org/10.1007/s00500-023-08153-5

5 T. Jin, H. X. Xia, Lookback option pricing models based on the uncertain fractional-order differential equation with Caputo type, *J. Ambient Intell. Human. Comput. (JAIHC)*, **14** (2023), 6435–6448. https://doi.org/10.1007/s12652-021-03516-y

6 C. W. Shen, X. Y. Ji, C. L. Miao, Real-Time image stitching with convolutional neural networks, *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, (2019), 192–197. https://doi.org/10.1109/RCAR47638.2019.9044010

7 L. Nie, C. Y. Lin, K. Liao, M. Q. Liu, Y. Zhao, A view-free image stitching network based on global homography, *J. Vis. Comun. Imag. R*, **73** (2019), 102950. https://doi.org/10.1016/j.jvcir.2020.102950

8 L. Nie, C. Y. Lin, K. Liao, Y. Zhao, Learning edge-preserved image stitching from multi-scale deep homography, *Neurocomputing*, **491** (2022), 533–543. https://doi.org/10.1016/j.neucom.2021.12.032

9 D. Y. Song, G. M. Um, H. K. Lee, D. Cho, End-to-end image stitching network via multi-homography estimation, *IEEE Signal Process Lett.*, **28** (2021), 763–767. https://doi.org/10.1109/LSP.2021.3070525

10 D. Y. Song, G. Lee, H. K. Lee, G. M. Um, D. Cho, Weakly-Supervised stitching network for real-world panoramic image generation, *Lect. Notes Comput. Sci.*, **13676** (2022), 54–71. https://doi.org/10.1007/978-3-031-19787-1_4

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

12 J. R. Zhang, C. Wang, S. C. Liu, L. P. Jia, N. J. Ye, J. Wang, et al., Content-Aware unsupervised deep homography estimation, *Lect. Notes Comput. Sci.*, **12346** (2020), 653–669. https://doi.org/10.1007/978-3-030-58452-8_38

13 S. C. Liu, N. J. Ye, C. Wang, J. R. Zhang, L. P. Jia, K. M. Luo, et al., Content-Aware unsupervised deep homography estimation and its extensions, *IEEE Trans. Pattern Anal. Mach. Intell.*, **45** (2022), 2849–2863. https://doi.org/10.1109/TPAMI.2022.3174130

14 R. Huang, Q. Chang, Y. Zhang, Unsupervised oral endoscope image stitching algorithm, *J. Shanghai Jiaotong University (Science)*, (2022), 1–10. https://doi.org/10.1007/s12204-022-2513-7

15 S. Wang, F. Y. Yuan, B. Chen, H. F. Jiang, W. Q. Chen, Y. Wang, Deep homography estimation based on attention mechanism, in *2021 7th International Conference on Systems and Informatics (ICSAI)*, (2021), 1–6. https://doi.org/10.1109/ICSAI53574.2021.9664027

16 L. David, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.*, **60** (2004), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

17 H. Bay, T. Tuytelaars, L. Van Gool, SURF: speeded up robust features, *Lect. Notes Comput. Sci.*, **3951** (2006), 404–417. https://doi.org/10.1007/11744023_32

18 E. Rublee, V. Rabaud, K. Konolige, ORB: An efficient alternative to SIFT or SURF, in *2011 International Conference on Computer Vision (ICCV)*, (2011), 2564–2571. https://doi.org/10.1109/ICCV.2011.6126544

19 G. F. Zhang, Y. He, W. F. Chen, J. Y. Jia, H. J. Bao Multi-viewpoint panorama construction with wide-baseline images, *IEEE Trans. Image Process*, **25** (2016), 3099–3111. https://doi.org/10.1109/TIP.2016.2535225

20 J. W. Bian, W. Y. Lin, Y. Liu, L. Zhang, S. K. Yeung, M. M. Cheng, et al., GMS: Grid-Based motion statistics for fast, ultra-robust feature correspondence, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 4181–4190. https://doi.org/10.1109/CVPR.2017.302

21 R. Sprengel, K. Rohr, H.S. Stiehl, Thin-plate spline approximation for image registration, in *Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (1996), 1190–1191. https://doi.org/10.1109/IEMBS.1996.652767

22 K. Rohr, H. S. Stiehl, R. Sprengel, W. Beil, T. M. Buzug, J. Weese, et al., Point-based elastic registration of medical image data using approximating thin-plate splines, *Lect. Notes Comput. Sci.*, **1131** (1996), 297–306. https://doi.org/10.1007/BFb0046967

23 J. Li, Z.M. Wang, S.M. Lai, Y. P. Zhai, M. J. Zhang, Parallax-tolerant image stitching based on robust elastic warping, *IEEE Trans. Mult.*, **20** (2018), 1672–1687. https://doi.org/10.1109/TMM.2017.2777461

24 L. Nie, C. Y. Lin, K. Liao, S. C. Liu, Y. Zhao, Learning Thin-Plate spline motion and seamless composition for parallax-tolerant unsupervised deep image stitching, *Comput. Vision Pattern Recogn. (CVPR)*, (2023), https://doi.org/10.48550/arXiv.2302.08207

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

26 J. H. Gao, S. J. Kim, M. S. Brown, Constructing image panoramas using dual-homography warping, *Comput. Vision Pattern Recogn. (CVPR)*, (2011), 40–56. https://doi.org/10.1109/CVPR.2011.5995433

27 W. Y. Lin, S. Y. Liu, Y. Matsushita, T. T. Ng, L. F. cheong, Smoothly varying affine stitching, *Comput. Vision Pattern Recogn. (CVPR)*, (2011), 345–352. https://doi.org/10.1109/CVPR.2011.5995314

28 J. Zaragoza, T. J. Chin, M. S. Brown, D. Suter, As-projective-as-possible image stitching with moving DLT, *Comput. Vision Pattern Recogn. (CVPR)*, (2013), 2339–2346. https://doi.org/10.1109/CVPR.2013.303

29 C. H. Chang, Y. Sato, Y. Y. Chuang, Shape-preserving half-projective warps for image stitching, *Comput. Vision Pattern Recogn. (CVPR)*, (2014), 3254–3261. https://doi.org/10.1109/CVPR.2014.422

30 C. C. Lin, S. Pankanti, K. Ramamurthy, A. Aravkin, Adaptive as-natural-as-possible image stitching, *Comput. Vision Pattern Recogn. (CVPR)*, (2015), 1155–1163. https://doi.org/10.1109/CVPR.2015.7298719

31 Y. S. Chen, Y. Y. Chuang, Natural Image Stitching with the Global Similarity Prior, *Lect. Notes Comput. Sci.*, **9909** (2016), 186–201. https://doi.org/10.1007/978-3-319-46454-1_12

32 T. Z. Xiang, G. S. Xia, L. P. Zhang, N. N. Huang, Locally warping-based image stitching by imposing line constraints, in *2016 23rd International Conference on Pattern Recognition (ICPR)*, (2016), 4178–4183. https://doi.org/10.1109/ICPR.2016.7900289

33 C. He, J. Zhou, Mesh-based image stitching algorithm with linear structure protection, *J. Image Graph.*, (2018), 973–983. https://doi.org/10.11834/jig.170653

34 P. Du, J. F. Ning, J. G. Cui, S. L. Huang, X. C. Wang, J. X. Wang, Geometric Structure Preserving Warp for Natural Image Stitching, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2022), 3688–3696 https://doi.org/10.1109/CVPR52688.2022.00367