Advanced Endoscopy Imaging with Automatic Feedback

D. M. Bappy[1], Donghwa Kang[1], Jinkyu Lee[2], Youngmoon Lee[3], Minsuk Koo[1], and Hyeongboo Baek[4(✉)]

[1] Department of Computer Science and Engineering, Incheon National University, Incheon, Republic of Korea
[2] Department of Computer Science and Engineering, Sungkyunkwan University, Suwon, Republic of Korea
[3] Department of Robotics, Hanyang University, Ansan, Republic of Korea
[4] Department of Artificial Intelligence, University of Seoul, Seoul, Republic of Korea
hbbaek359@gmail.com

Abstract. As we move towards a future where minimally invasive methods become the norm for surgeries and diagnostic procedures, it is increasingly vital to improve our strategies for viewing the organs and complex structures within our bodies. Image stitching presents an enticing solution, expanding our field of view by seamlessly weaving together a sequence of images. While existing stitching techniques do lean on the capabilities of endoscopy imaging, they, unfortunately, overlook the critical need for automated feedback when grappling with the complexities and challenges innate to endoscopy imaging. these methods struggle to stand firm against deformations and regions with low texture. In this paper, we introduce a robust endoscopic image-stitching algorithm designed to thrive in adversity. Its unique resilience to deformations and low-texture regions is reinforced by the inclusion of a radial basis function weighting that is paired harmoniously with location-dependent homography based on the corresponding locations of the strong features extracted by affine shape-adapted Hessian-Laplace detector. Crucially, this algorithm is steered by a sophisticated automatic feedback mechanism. This feedback system makes astute evaluations based on an image quality metric and the structural comparison between the sequences of endoscopy images. We have thoroughly validated the efficacy of our new approach using two public datasets, namely EndoSLAM and EndoAbS, under demanding conditions. The results eloquently illustrate the superior benefits of our technique. Our proposed method surpasses commonly employed techniques, delivering superior performance in quantitative metrics, including precision at 30.07%, recall at 114.89%, F1-score at 84.62%, and TRE at 46.07%.

Keywords: Endoscopy Imaging · Endoscopy Stitching · Homography · Feature Extraction · Feature matching.

1   Introduction

Endoscopy is a common and essential tool in medical diagnostics and research, playing a pivotal role in identifying and treating diseases, particularly tumors [29]. Yet, there are limitations such as the inability to obtain the best field of view and magnify an image simultaneously [15]. Increasing magnification enhances detail but reduces the visual field, complicating comprehensive organ examination and disease evaluation [26]. With the advancement of computer technology, medical images can be shared online, facilitating remote consultations [27]. However, a single endoscopic image might not suffice for accurate judgment. Thus, the challenge lies in stitching together successive images for complete visualization, which is crucial for remote diagnosis and data measurement.

Image stitching involves addressing several interconnected problems. Firstly, there is the issue of data association, which involves finding common scene elements across various views, a topic discussed in-depth by Huang [9]. Secondly, we have the task of estimating a geometric transformation. This transformation should align with the data association and unify disparate views into a single, seamlessly stitched image [8]. These two tasks are usually undertaken concurrently [4]. Lastly, it is necessary to blend the individual images' intensities to ensure a consistent, smooth final image [28].

Of these, the sub-problem of data association is the most complex, particularly in surgical scenarios, and hence attracts significant research focus. A classic strategy for addressing this issue is to identify and extract image point features that correspond to unique landmarks within the scene and then match them across different views. This approach, known as feature-based stitching [18], has been thoroughly researched in recent years. Various well-known hand-crafted feature methodologies, such as Harris [20], SIFT [14], SURF [24], ORB [7], and FAST [12] have been applied to it. In more recent developments, data-driven features derived from deep neural networks are being employed for image stitching [2].

There are also other strategies for stitching that do not hinge on feature extraction. Direct and dense pixel-based registration techniques can be developed as an iterative optimization problem, with the goal of maximising similarity as calculated by mutual information [19] or other photometric similarity/difference measures [13]. With the increased use of deep learning across various fields, there are now proposed end-to-end stitching algorithms that rely on deep learning to deduce registration parameters [3].

Technology's evolution has substantially enhanced endoscopic instruments. Luo et al. [16] proposed a method that uses an Auto-Regressive (AR) model to predict human motion intentions for teleoperated tasks. This algorithm anticipates human movements, updating and adjusting robotic actions during Human-Robot Interactions (HRI) as necessary. Moreover, Su et al. [25] developed an innovative technique. They employed a swivel motion reconstruction method to mimic human-like behavior through kinematic mapping in robot redundancy.

Their groundbreaking framework merges an incremental learning approach with a deep convolutional neural network, facilitating efficient and speedy learning.

However, these kinds of tasks require automated system feedback regarding image quality. Owing to the endoscope's movement, limited field of view, and the complexity of human body structures, there is a high likelihood of missing crucial features in the image sequence. Therefore, it is essential to have real-time updates from the imaging process if a vital structure is not captured in the images due to the endoscope camera's unpredictable movement. For this reason, we provide automated system feedback during endoscopy regarding image quality, which is otherwise impossible to process manually.

Based on the analysis of existing literature, it becomes evident that traditional image stitching techniques have been primarily used for endoscopic imagery, despite their inherent limitations. Such approaches do not take into account the specific attributes of endoscopic images, resulting in significant difficulties in solving the stitching problem. In scenarios such as large motion, a blurred environment, low texture, and deformations, the existing techniques are prone to stitching failures, often without providing any insightful feedback.

In Section 3, we substantiate our proposed methodology by employing publicly accessible datasets. The comparison of stitched images in Figure 4 and 6 with recent techniques attests to the advantages gained from extracting a substantial number of robust features in texture-less and deformed regions. These features are then leveraged to compute a weighted local homography for stitching image sequences. The improvements compared to best performers in quantitative metrics, as depicted in Table 1–precision at 30.07%, recall at 114.89%, F1-score at 84.62%, and TRE at 46.07% undeniably affirm that our technique significantly outperforms commonly employed methods.

This paper offers several unique contributions:

– The introduction of a fully automated, feedback-oriented, robust stitching algorithm designed specifically for endoscopic image stitching.
– Our algorithm not only provides feedback to surgeons but also executes blur removal on endoscopic images based on image quality assessment.
– The implementation of an affine-shape adapted Hessian-Laplace detector to extract robust features from the images.
– Our approach uses these robust features to apply a weighted local homography designed with endoscopic imaging properties in mind.

2   Approach

In this section, we introduce our proposed approach for robust endoscopic image stitching with automated feedback. Figure 1 illustrates the various components of our technique.
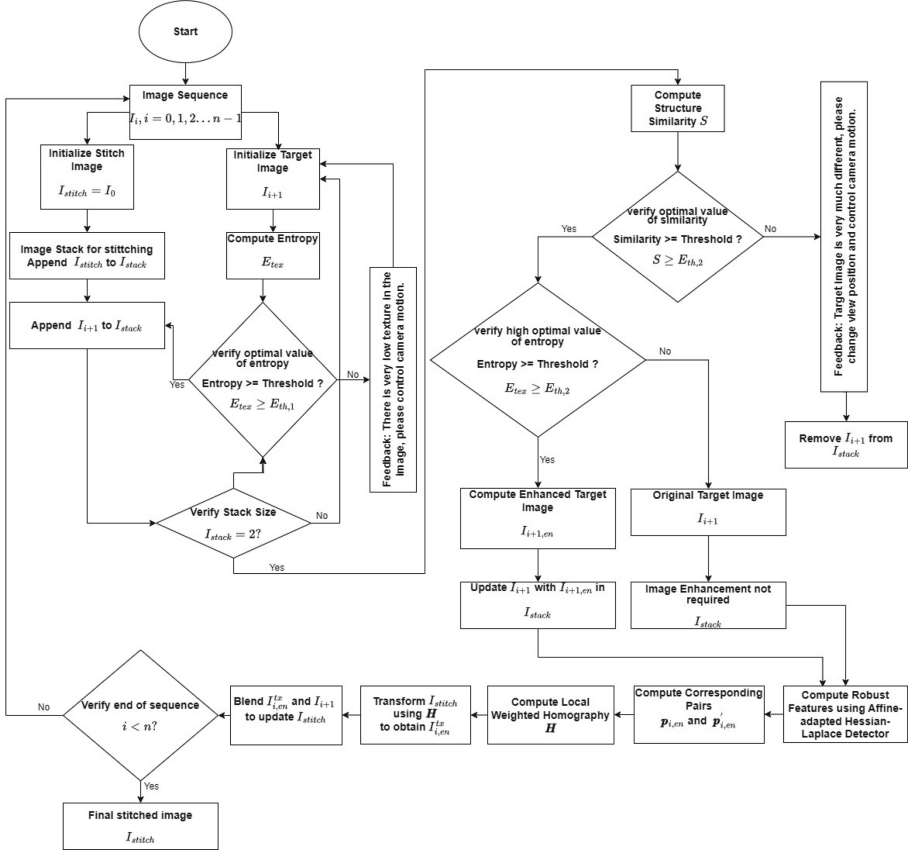
Fig. 1. Schematic of the proposed technique

2.1 Statistical Analysis-driven Automated Feedback

We make the assumption that the endoscopic images, denoted as I_i with $i = 1, 2, 3, ..., n$, are already aligned. Our processing of these images follows a sequential approach. To assess the level of texture availability, we employ a measure based on uncertainty or randomness within an image. This measure, denoted as $E_{i,tex}$, is calculated as follows:

$$E_{i,tex} = \sum_{k=1}^{255} (-p_k \log_2 p_k). \tag{1}$$

In Equation (1), p_k represents the probability associated with the gray level k of the image.

In order to assess the structural similarity between two consecutive images, I_i and I_{i+1}, we consider their luminance, contrast, and structure. This comparison is based on the following equations:

$$S(I_i, I_{i+1}) = [l(I_i, I_{i+1})]^\alpha \cdot [c(I_i, I_{i+1})]^\beta \cdot [s(I_i, I_{i+1})]^\gamma. \tag{2}$$

Here, the terms are defined as follows:

$$l(I_i, I_{i+1}) = \frac{2\mu_{I_i}\mu_{I_{i+1}} + \epsilon_1}{\mu_{I_i}^2 + \mu_{I_{i+1}}^2 + \epsilon_1},$$

$$c(I_i, I_{i+1}) = \frac{2\sigma_{I_i}\sigma_{I_{i+1}} + \epsilon_2}{\sigma_{I_i}^2 + \sigma_{I_{i+1}}^2 + \epsilon_2},$$

$$s(I_i, I_{i+1}) = \frac{\sigma_{I_i I_{i+1}} + \epsilon_3}{\sigma_{I_i}\sigma_{I_{i+1}} + \epsilon_3}$$

where μ_{I_i} and $\mu_{I_{i+1}}$, σ_{I_i} and $\sigma_{I_{i+1}}$, and $\sigma_{I_i I_{i+1}}$ represent the local means, standard deviations, and cross-covariance for images I_i and I_{i+1}.

2.2 Blur Removal for Enhanced Image Quality

To ensure that the texture information of a good quality endoscopic image is not lost during the blur removal process, we adopt a selective approach based on the score obtained from Equation (1). Randomly removing blur may result in the degradation of image quality and loss of texture details. The blur removal technique discussed in [5] is employed for this purpose. Consequently, the computation of the enhanced images is carried out as follows:

$$I_{i+1,en} = \frac{I_{i+1} - [1 - t(x)]A}{t(x)}. \tag{3}$$

Here, $t(x)$ represents the transmission map, and A denotes the single color in the image where the transmission map is 0.

2.3 Robust Feature estimation using Affine-Shape Adapted Hessian-Laplace Detector

To initiate our procedure, we employ a detector based on the Hessian matrix to identify salient points within the scale space. The Hessian matrix is essentially a matrix of second-order partial derivatives, obtained from the Taylor series expansion. It is articulated as:

$$H_{i,en} = \begin{bmatrix} I_{i,en,xx}(x; \sigma_d) & I_{i,en,xy}(x; \sigma_d) \\ I_{i,en,yx}(x; \sigma_d) & I_{i,en,yy}(x; \sigma_d) \end{bmatrix} \tag{4}$$

Here, $I_{i,en,xx}$, $I_{i,en,xy}$, and $I_{i,en,yy}$ denote second-order derivatives that are calculated using Gaussian kernels of scale σ_d.

Next, we employ a scale-normalized Laplacian, which helps us choose the appropriate scale for a detected point. It is expressed as:

$$Laplacian(x; \sigma_d) = \sigma_d^2 \left| I_{i,en,xx}(x; \sigma_d) + I_{i,en,yy}(x; \sigma_d) \right| \tag{5}$$

We can estimate the affine shape of the neighborhood surrounding the detected point by leveraging the eigenvalues of the second-moment matrix. This matrix is represented as:

$$M_{i,en} = \sigma_d^2 g(\sigma_I) * \begin{bmatrix} I_{i,en,x}^2(x;\sigma_d) & I_{i,en,x}I_{i,en,y}(x;\sigma_d) \\ I_{i,en,x}I_{i,en,y}(x;\sigma_d) & I_{i,en,y}^2(x;\sigma_d) \end{bmatrix} \tag{6}$$

In this scenario, the derivatives are averaged over the detected point's neighborhood by applying a smoothing process using a Gaussian window of scale σ_I.

2.4 Location-Based Weighted Homography Estimation for Feature Correspondence

Consider reference and target images denoted as $I_{i,en}$ and $I_{i+1,en}$. We compute matching pairs $\boldsymbol{p}_{i,en} = [x,y]^T$ and $\boldsymbol{p}_{i,en}' = [x',y']^T$ from detected features x_r in previous section using these images. The pairs are derived from robustly identified feature points discussed in the previous section, using the Brute-Force (BF) matching algorithm [10]. The homographies between these pairs can be defined as

$$\boldsymbol{p}_{i,en}' = \boldsymbol{h}(\boldsymbol{p}_{i,en}) \tag{7}$$

From Equation (7), we can perform a simple matrix expansion resulting in the following expressions:

$$x' = \frac{h_1 x + h_2 y + h_3}{h_7 x + h_8 y + h_9} \tag{8}$$

$$y' = \frac{h_4 x + h_5 y + h_6}{h_7 x + h_8 y + h_9} \tag{9}$$

In homogeneous coordinates, Equation 7 is represented as

$$\tilde{\boldsymbol{p}}_{i,en}' = \boldsymbol{H}(\tilde{\boldsymbol{p}}_{i,en}) \tag{10}$$

where $\tilde{\boldsymbol{p}}_{i,en}' = [x',y',1]^T$, $\tilde{\boldsymbol{p}}_{i,en} = [x,y,1]^T$ and $\boldsymbol{H}$ is a 3×3 matrix.

The columns of $\boldsymbol{H}$ are denoted by $\boldsymbol{h}_1 = [h_1,h_4,h_7]^T$, $\boldsymbol{h}_2 = [h_2,h_5,h_8]^T$, and $\boldsymbol{h}_3 = [h_3,h_6,h_9]^T$. By taking a cross-product on both sides of Equation 10, we get:

$$\boldsymbol{0}_{3\times1} = \tilde{\boldsymbol{p}}_{i,en}' \times \boldsymbol{H}\tilde{\boldsymbol{p}}_{i,en} \tag{11}$$

This equation (11) can be reformulated as follows:

$$\boldsymbol{0}_{3\times1} = \begin{bmatrix} \boldsymbol{0}_{3\times1} & -\tilde{\boldsymbol{p}}_{i,en}^T & y'-\tilde{\boldsymbol{p}}_{i,en}^T \\ -\tilde{\boldsymbol{p}}_{i,en}^T & \boldsymbol{0}_{3\times1} & -x'\tilde{\boldsymbol{p}}_{i,en}^T \\ -y'\tilde{\boldsymbol{p}}_{i,en}^T & x'\tilde{\boldsymbol{p}}_{i,en}^T & \boldsymbol{0}_{3\times1} \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \\ \boldsymbol{h}_3 \end{bmatrix} \tag{12}$$

The 9×1 vector in Equation 12 is referred to as $\boldsymbol{h}$. Considering that only two rows of the 3×9 matrix in Equation 12 are linearly independent, we can determine $\boldsymbol{h}$ utilizing a collection of N corresponding points, denoted as $\tilde{\boldsymbol{p}}i, en, kk = 1^N$ and $\tilde{\boldsymbol{p}}i, en, k^{'}k = 1^N$.

$$\boldsymbol{h} = \operatorname*{argmin}_{\boldsymbol{h}} \sum_{k=1}^{N} \left\| \begin{bmatrix} \boldsymbol{a}_{k,1} \\ \boldsymbol{a}_{k,2} \end{bmatrix} \boldsymbol{h} \right\|^2 = \operatorname*{argmin}_{\boldsymbol{h}} \left\| \boldsymbol{A}\boldsymbol{h} \right\|^2 \tag{13}$$

The components $\boldsymbol{a}_{k,1}$ and $\boldsymbol{a}_{k,2}$ correspond to the two rows of the matrix specified in Equation 12. We will also apply a constraint of $\|\boldsymbol{h}\|^2 = 1$ since a homographic transformation possesses only 8 degrees of freedom.

The paper [30], introduced the Moving DLT (Direct Linear Transform) framework to approximate local homography. This method incorporates locality-enforcing weights into the objective of Equation 13. The estimation of local homography at the position $\boldsymbol{p}_{i,en,j}$ is conducted as follows:

$$\boldsymbol{h}_j = \operatorname*{argmin}_{\boldsymbol{h}_j} \sum_{k=1}^{N} w_{j,k} \left\| \begin{bmatrix} \boldsymbol{a}_{k,1} \\ \boldsymbol{a}_{k,2} \end{bmatrix} \boldsymbol{h} \right\|^2 \tag{14}$$

The calculation of scalar weights, denoted as $w_{j,k}k = 1^N$, is determined by changes relative to $\boldsymbol{p}_{i,en,j}$.

$$w_{j,k} = \left(\left\| \boldsymbol{p}_{i,en,k} - \boldsymbol{p}_{i,en,j} \right\|^2 * \log \left\| \boldsymbol{p}_{i,en,k} - \boldsymbol{p}_{i,en,j} \right\| \right) / \sigma^2 \tag{15}$$

We can reformulate Equation 14 as follows:

$$\boldsymbol{h}_j = \operatorname*{argmin}_{\boldsymbol{h}_j} \left\| \boldsymbol{W}_j \boldsymbol{A}\boldsymbol{h} \right\|^2 \tag{16}$$

The weight matrix $\boldsymbol{W}_j \in \mathbb{R}^{2N \times 2N}$ is structured as follows:

$$\boldsymbol{W}_j = diag\left(\begin{bmatrix} w_{1,j}w_{1,j}...w_{N,j}w_{N,j} \end{bmatrix} \right) \tag{17}$$

In the research [30], they propose the concept of allotting increased weights to data that is closer to $\boldsymbol{p}_{i,en,k}$. This enhances the ability of the projective warp, $\boldsymbol{H}$, with respect to the local structure around $\boldsymbol{p}_{i,en,k}$. They also incorporate an offset parameter, γ, to circumvent numerical issues.

A key aspect of this concept is that the computation of local homography is only applicable to the areas of the target image that overlap with the reference image. For each pixel in regions that do not overlap, the transformation is computed as a weighted linear combination of the local homographies in the overlapping areas. Thus, an appropriate set γ is crucial to prevent extrapolation.

This overarching idea is highly effective in the context of real-world images, which often contain abundant textures and rigid structures. However, complications arise during endoscopy imaging, due to the presence of low texture and numerous deformations. To address these challenges, we consider an endoscopy

imaging property adapted weight function in Equation 15 that tackles deformations and outliers caused by low texture. This is achieved by combining squared distance and log distance. This combined approach allows for the flexible fitting of deformation patterns, while the log distance alone aids in reducing outliers.

When $\boldsymbol{p}_{i,en,k}$ shifts continuously in the source image domain $I_{i,en}$, the corresponding homography estimation $\boldsymbol{H}$ adjusts smoothly. This process results in a dynamic warp that can flexibly adapt to data and transform the source image into $I_{i,en}^{tx}$.

Blending Warped Images

In this section, we merge the transformed source image $I_{i,en}^{tx}$ with the target image $I_{i+1,en}$ to generate the final stitched image I_{stitch}. The computation for the stitched image [23] is as follows:

$$I_{stitch} = \alpha I_{i,en}^{tx} + (1 - \alpha)I_{i+1,en} \tag{18}$$
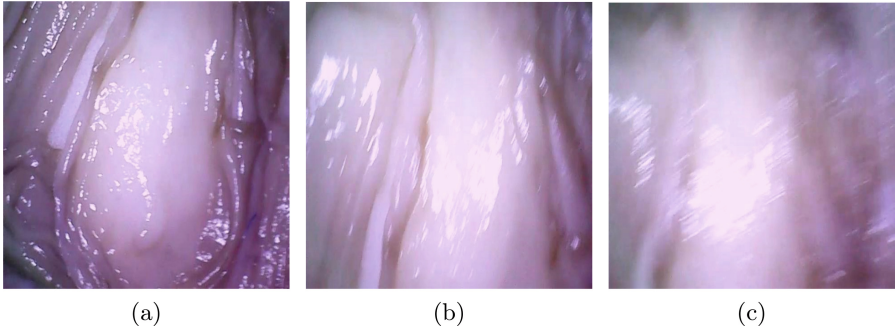

| (a) | (b) | (c) |

Fig. 2. Images with different entropy: (a) higher entropy (b) moderate entropy, and (c) lower entropy

3   Experiments

We have utilized three diverse, publicly accessible datasets: EndoSLAM [21], EndoAbs [22], and Hamlyn [16]. These datasets were selected purposefully to encompass a variety of conditions, including differing lighting circumstances, small fields of view, low-texture regions, and deformed areas, thereby providing a comprehensive evaluation of our method. The texture of an image is crucial for image-guided systems as it enhances feature extraction reliability and boosts computational precision.
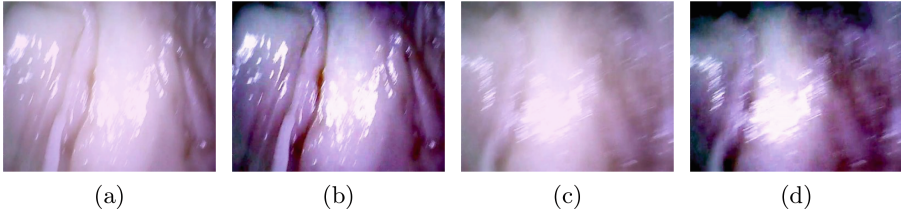
<center>(a)   (b)   (c)   (d)</center>

Fig. 3. Image Enhancement: (a) image with moderate blur, (b) enhanced image from moderate blur, (c) image with extreme blur and (d) enhanced image from extreme blur

3.1   Qualitative Evaluation

Figure 2 depicts three images from the EndoSLAM dataset, each presenting varying levels of texture $E_{i,tex}$. In Figure 3(a), the image displays a moderate blur with an entropy value of 7.21. After deblurring, an enhanced image is produced, as seen in Figure 3(b), which possesses a richer texture and an improved entropy value of 7.87. This refined image is more conducive to robust feature extraction. Conversely, the image in Figure 3(c) exhibits an intense blur with an entropy of 7.13. Following deblurring, the resultant image on Figure 3(d) shows only a slight improvement in texture quality, with an entropy of approximately 7.17. Despite the enhancement, this image remains below the optimal threshold and is not ideal for stitching.

Figure 4 illustrates that our proposed method generates higher quality and more accurate stitching in comparison to other cutting-edge techniques. To further scrutinize the performance, a region in Figure 5 where robust feature extraction and subsequent image frame warping proved challenging was cropped. The APAP technique, as illustrated in Figure 5(a) and outlined in [30], presents a comparable outcome; specifically, it fails to accurately merge the tissue area, resulting in a warped region. The AutoStitch method [6] generates a malformed area and even excludes some parts entirely from the original image sequence. As we shift our focus to Figure 5(c), the latest wide parallax technique [11] misplaces the tissue region entirely due to the lack of robust features and incorrect warping in that particular area. In stark contrast, our proposed method delivers perfect alignment of the tissue region. This is achieved by successfully extracting robust features in these areas thanks to the affine adaptation of the Hessian-Laplace detector. Subsequently, these features are used to calculate a weighted homography, which enables precise warping. This is made possible by the appropriate distribution of weight based on whether a randomly located feature is sufficiently close to the target feature location.

Figure 6 and 7 showcase the exemplary results achieved when stitching using the stereo pairs from the Hamlyn and EndoAbs datasets, respectively. The Hamlyn dataset, while offering images of high resolution, introduces intricate challenges due to the presence of elements such as blood, deformations, and surgical instruments. Impressively, our proposed method managed to stitch the Hamlyn stereo pairs with remarkable accuracy, highlighting its resilience and effec-
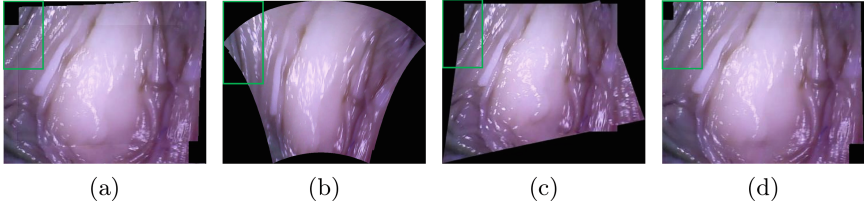
Fig. 4. Image stitching using state of art techniques: (a) APAP (b) AutoStitch (c) Wide Parallax, and (d) Proposed
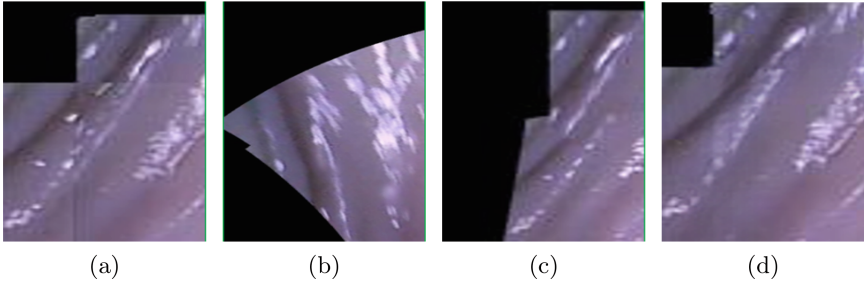

Fig. 5. Magnified stitched region using state-of-the-art feature based techniques: (a) APAP (b) AutoStitch (c) wide parallax, and (d) Proposed

tiveness even when faced with occlusions caused by surgical tools. Turning our attention to the EndoAbs dataset, it's worth noting that our technique consistently delivered despite the challenging low-light conditions. Such challenging scenarios are not uncommon in surgical environments. The successes displayed herein emphasize the potential of our method to furnish an expanded field of view, characterized by both precision and quality, from a stereo image sequence. This expanded perspective can be invaluable in enhancing surgical operations and diagnostic procedures.

Figure 8 depicts an instance of a failure during the image stitching process. As evident from the figure, the extreme blur in Figure 8(a) (the left image) impedes the algorithm's capability to establish correspondences between image sequences. The particular image lacks sufficient texture information, making it challenging to extract robust features. Consequently, the stitching process fails noticeably as shown in Figure 8(b). In such scenarios, our algorithm offers feedback to the experts, suggesting adjustments in the endoscope's movement to capture higher-quality images.

Figure 9 illustrates the scenario of the maximum angular limit in our proposed technique. Beyond this limit, the stitching process is affected as the angle increases. To assess the maximum allowable angle for flawless stitching, we set the scale to 0.5 and varied the angle values to 5, 10, 15, 45, and 50 degrees. Table 1 includes the maximum limit angle of 45 degrees, beyond which the stitching process begins to degrade. In Figure 9(a)(b)(c), the green box region demonstrates

nearly perfect stitching. However, in Figure 9(d)(e), we observe a degradation in the stitching process within the green box region as the angle increases from 45 to 50 degrees. These quantitative results are reflected in Table 1, where the Target Registration Error (TRE) is 0.05 for angles 5, 10, and 15 and increases to 1.3 for the 45-degree angle.
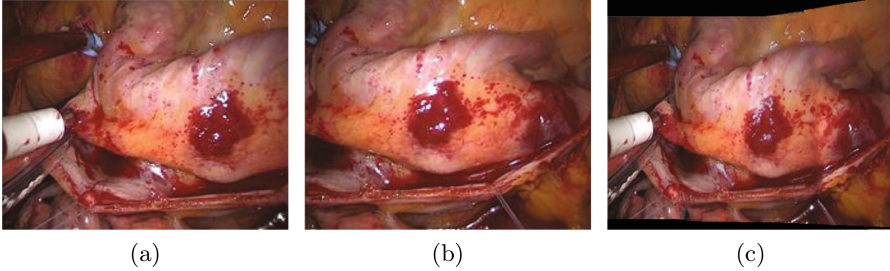

| (a) | (b) | (c) |

Fig. 6. Image Stitching using stereo pairs: (a) Hamlyn left view (b) Hamlyn right view (c) Hamlyin stitched

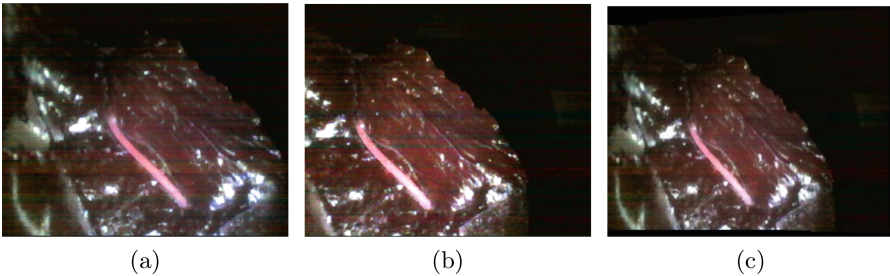| (a) | (b) | (c) |

Fig. 7. Image Stitching using stereo pairs: (a) EndoAbs left view (b) EndoAbs right view, and (c) EndoAbs stitched

3.2   Quantitative Evaluation

To evaluate the efficacy of the affine-adapted Hessian-Laplace detector, we compared it with two recently adapted feature detectors for endoscopy imaging: Pyramid ORB [31] and Improved SIFT [16]. We employed the same image sequence these detectors use for feature extraction. When compared to [31] with 800 key points, our proposed method extracts 17,261 key points, which is approximately 21.5 times greater. Furthermore, when compared to [16] with 113 corresponding points, our technique maintains 1,505 correspondences, which is 13.3 times greater. The large number of key points that our affine-adapted detector extracts features from most regions, and the high number of correspondences allows for dense matching.
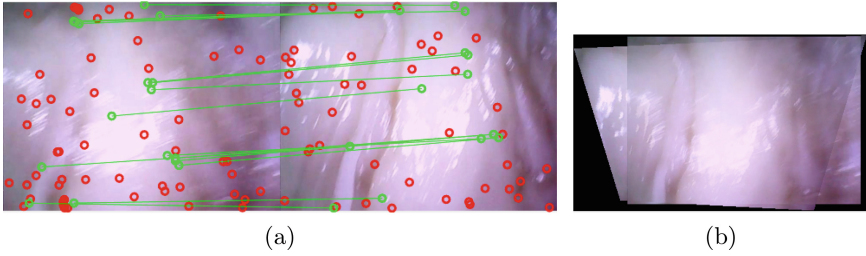
Fig. 8. Failure case of image stitching: (a) matching pairs, and (b) stitched image
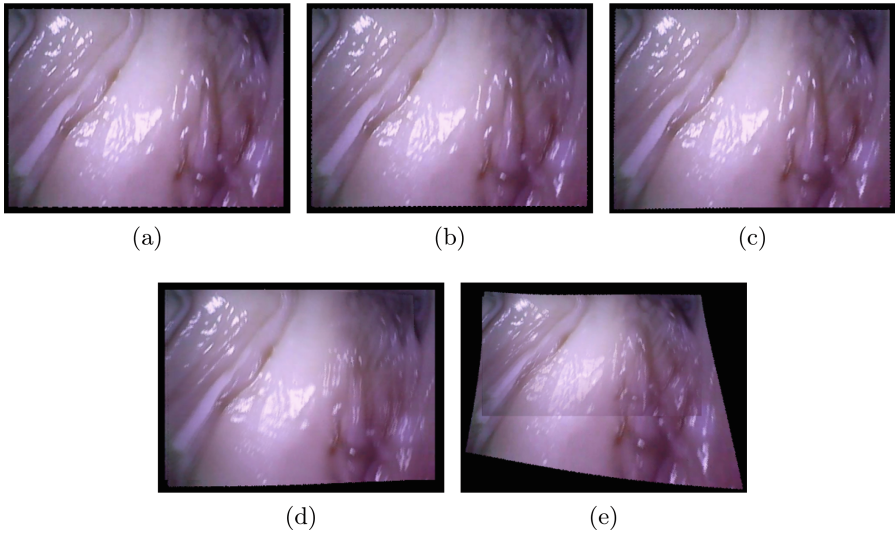

Fig. 9. Stitching images using the proposed method at a scale of 0.5 with various rotations: (a) 5-degree rotation (b) 10-degree rotation (c) 15-degree rotation (d) 45-degree rotation, and (e) 50-degree rotation

Feedback Evaluation In our research, we utilized the EndoSLAM dataset to determine optimal feedback threshold values. After a thorough examination of multiple trials, we identified optimal thresholds for both entropy and structural similarity, which stood at 7.25 and 0.76, respectively. These identified thresholds play a pivotal role in facilitating feedback for our proposed algorithm. Specifically, an endoscopy image with entropy below the threshold suggests that the image might lack sufficient texture, while a lower structural similarity value compared to the set threshold suggests a lack of adequate structure similarity in the consecutive image sequence for successful stitching. Consider the images shown in Figure 2(a)(b)(c) with entropies of 7.45, 7.21, and 7.13 respectively. The progressive increase in entropy value signifies an image enriched in texture. Based on these observations, our method leverages these entropy values to provide the

surgeon with feedback regarding the optimal positioning and movement of the endoscope. Additionally, the structural similarity value obtained from equation 2 assists in determining when there is an inadequate overlapping region for effective comparison and feature extraction. For instance, Figure 3(b) depicts how improved image quality can accentuate the texture in an image. Yet, Figure 3(d) illustrates that even with enhanced image quality, extracting meaningful information for subsequent processes remains a challenge.

We need the feedback system to operate in real time during the endoscopy procedure to decide whether to accept or reject an image. However, real-time stitching is not necessary, as it is time-consuming and not crucial for the surgeon's immediate evaluation. The priority is to obtain high-quality images during the procedure, ensuring they can be stitched later without information loss. This will provide the surgeon with a high-quality stitched image for better diagnosis. Our feedback system can evaluate and correct endoscopy images at a rate of 10 frames per second.

Table 1. Evaluating the Efficacy of Various Techniques

Orientation		000000 Methods																				
Scale	Rotation	SIFT				SURF				ORB				AKAZE				Proposed			
		P	R	F1	TRE	P	R	F1	TRE	P	R	F1	TRE	P	R	F1	TRE	P	R	F1	TRE
0.9	5	0.65	0.17	0.27	0.45	0.59	0.29	0.39	8.88	0.59	0.15	0.23	3.87	0.75	0.32	0.3	3.70	0.95	0.77	0.85	0.06
0.9	10	0.64	0.15	0.25	0.85	0.60	0.15	0.23	10.7	0.69	0.12	0.20	6.71	0.82	0.34	0.48	8.02	0.93	0.66	0.77	0.06
0.9	15	0.65	0.12	0.20	1.15	0.50	0.11	0.19	17.8	0.72	0.11	0.20	7.03	0.81	0.35	0.49	10.6	0.88	0.56	0.68	0.05
0.8	5	0.48	0.10	0.13	0.41	0.41	0.19	0.26	9.46	0.73	0.22	0.34	3.62	0.67	0.25	0.36	3.53	0.95	0.89	0.92	0.06
0.8	10	0.47	0.10	0.13	0.79	0.48	0.11	0.18	10.7	0.68	0.20	0.31	7.01	0.69	0.25	0.37	7.31	0.93	0.75	0.84	0.06
0.8	15	0.51	0.10	0.13	1.11	0.46	0.10	0.16	16.6	0.58	0.16	0.25	7.20	0.73	0.25	0.38	10.8	0.89	0.63	0.74	0.05
0.7	5	0.40	0.10	0.10	0.36	0.60	0.23	0.33	8.21	0.67	0.14	0.23	3.13	0.71	0.21	0.33	3.66	0.93	0.89	0.91	0.05
0.7	10	0.39	0.10	0.11	0.70	0.55	0.14	0.23	10.2	0.61	0.13	0.22	6.18	0.76	0.23	0.35	7.02	0.92	0.87	0.90	0.06
0.7	15	0.41	0.10	0.11	1.10	0.49	0.11	0.18	15.8	0.55	0.14	0.22	7.22	0.80	0.22	0.34	10.4	0.90	0.76	0.83	0.05
0.6	5	0.39	0.10	0.11	0.31	0.33	0.11	0.17	7.49	0.73	0.10	0.16	2.88	0.55	0.11	0.19	3.32	0.93	0.88	0.90	0.04
0.6	10	0.44	0.10	0.12	0.61	0.37	0.11	0.16	14.5	0.56	0.10	0.11	5.33	0.66	0.14	0.23	6.22	0.93	0.86	0.89	0.05
0.6	15	0.45	0.10	0.13	0.89	0.44	0.10	0.17	14.9	0.69	0.10	0.16	7.84	0.66	0.14	0.23	6.24	0.90	0.82	0.86	0.05
0.5	5	0.48	0.10	0.15	0.27	0.60	0.12	0.21	6.13	0.71	0.10	0.10	2.33	0.46	0.10	0.10	2.52	0.84	0.78	0.81	0.04
0.5	10	0.45	0.10	0.13	0.51	0.51	0.10	0.17	11.8	0.73	0.10	0.12	4.43	0.43	0.10	0.10	4.57	0.93	0.86	0.89	0.05
0.5	15	0.47	0.10	0.14	0.74	0.53	0.10	0.16	17.1	0.70	0.10	0.10	6.73	0.41	0.10	0.10	7.64	0.86	0.77	0.81	0.05
0.5	45	0.36	0.08	0.11	2.10	0.29	0.08	0.06	18.3	0.42	0.05	0.08	7.95	0.38	0.04	0.07	8.71	0.31	0.26	0.28	1.30
Average		0.47	0.10	0.14	0.77	0.48	0.13	0.20	12.4	0.64	0.12	0.19	5.58	0.65	0.19	0.28	6.51	**0.88**	**0.74**	**0.79**	**0.12**

Assessment Through Simulated Transformations Endoscopy imaging encounters various obstacles, with the random motion of the endoscope being one of the primary challenges. This motion may introduce minor rotations, and the tissues encountered can also undergo deformation. Bearing this scenario in mind, we select an arbitrary frame from the EndoSLAM dataset and subject it to transformations such as rotation followed by scaling. This procedure yields 15

unique transformation combinations denoted as T_{GT}. The transformation matrix T_{est} can then be estimated based on the matched feature points identified in the feature correspondence section.

We used a specific threshold value of 0.8 for all methods, including our proposed method, as suggested by [17] in Table 1 to evaluate quantitative performance. After numerous observations and experiments with different thresholds on endoscopy images, we determined that a threshold value of 0.8 was the most effective.

For quantitative evaluation with simulated data, we adopt an approach akin to that presented in [1]. To assess the results, we utilized metrics such as precision, recall, and F1-score to quantify feature detection and matching accuracy. Additionally, the TRE was used to gauge the precision of the registration.

Table 1 displays the quality evaluation metrics for various methods. It is evident from the table that our proposed method significantly surpasses other best performers in commonly used techniques, delivering superior performance in quantitative metrics. Specifically, we achieved a precision of 30.07%, recall of 114.89%, F1-score of 84.62%, and TRE of 46.07%. The superior performance of our method is primarily attributed to the innovative affine-shape adapted detector, which excels in detecting a substantial number of robust features, especially in low-texture regions. In contrast, other methods under comparison tend to identify features primarily around specular reflection-affected areas, leading them to estimate less accurate matches. Remarkably, our method managed to extract a commendable 78 percent of accurate matches from the vast number of features identified. Such a substantial figure is pivotal for our stitching algorithm, as it relies on local homography.
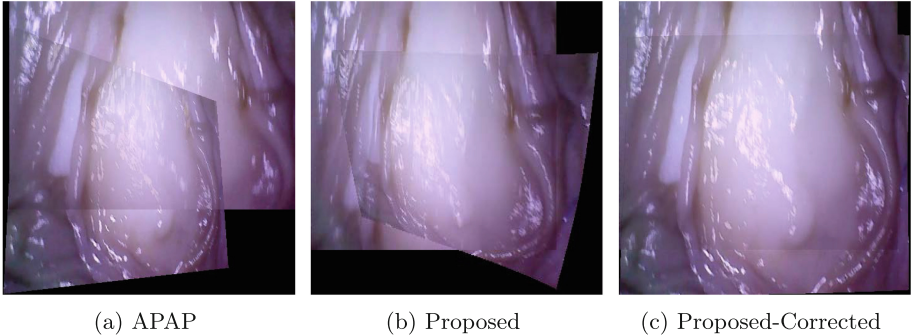

(a) APAP   (b) Proposed   (c) Proposed-Corrected

Fig. 10. Ablation Study

4   Ablation Study

We conducted an ablation study to verify the importance of reducing image blur, which can significantly affect stitching and feature extraction. In both

Table 2. Ablation Study

methods	P	R	F1	TER
SIFT	0.41	0.53	0.55	1.04
SURF	0.21	0.016	0.03	NA
ORB	0.19	0.01	0.02	NA
Proposed	0.65	0.71	0.69	0.08
Proposed-Corrected	0.79	0.76	0.77	0.02

Figure 10 and Table 2, "Proposed-Corrected" refers to the case of using reduced-blur images. Figures 10a and 10b demonstrate that the APAP and our proposed methods perform poorly without blur reduction. Conversely, after reducing blur, the "Proposed-Corrected" method in Figure 10c stitches images almost perfectly. Wide Parallax and Autostitch results are not included because these methods failed to stitch the images due to insufficient feature detection in blurred images.

Additionally, we performed a quantitative evaluation of feature extraction techniques using both blurred and corrected images, as shown in Table 2. The metrics indicate that SIFT, SURF, and ORB methods perform poorly with blurred images. Even our proposed technique performs averagely with blurred images. However, after reducing blur, our "Proposed-Corrected" method shows significantly improved performance.

5   Conclusions and Future Work

In our study, we introduce an innovative automatic feedback system critical to contemporary endoscopic tools used in medical surgeries and diagnostics. This automatic feedback notifies the surgeon and assists the robot in adjusting the endoscope's motion during the imaging process. Moreover, it provides crucial information to experts, enabling them to acquire meaningful images necessary for diagnosis and surgery.

Our technique successfully eliminates blur based on feedback, ensuring our algorithm does not degrade the quality of high-grade images by inappropriately applying the smoke removal technique. The affine shape-adapted Hessian-Laplace detector incorporated in our approach also extracts robust features from complex endoscopic images, particularly in deformed and low-texture areas. Finally, we implemented an endoscopy property-adapted weighting to estimate local homography, enhancing the stitching in deformed and low-texture regions.

References

1. Ali, S., Rohr, K., Axer, M., Amunts, K., Eils, R., Wörz, S.: Registration of ultra-high resolution 3d pli data of human brain sections to their corresponding high-resolution counterpart. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). pp. 415–419. IEEE (2017)

2. Bano, S., Vasconcelos, F., Shepherd, L.M., Vander Poorten, E., Vercauteren, T., Ourselin, S., David, A.L., Deprest, J., Stoyanov, D.: Deep Placental Vessel Segmentation for Fetoscopic Mosaicking. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (eds.) MICCAI 2020. LNCS, vol. 12263, pp. 763–773. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_73

3. Bano, S., Vasconcelos, F., Tella Amo, M., Dwyer, G., Gruijthuijsen, C., Deprest, J., Ourselin, S., Vander Poorten, E., Vercauteren, T., Stoyanov, D.: Deep Sequential Mosaicking of Fetoscopic Videos. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 311–319. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_35

4. Bartoli, A.: Groupwise geometric and photometric direct image registration. IEEE Trans. Pattern Anal. Mach. Intell. **30**(12), 2098–2108 (2008)

5. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1674–1682 (2016)

6. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. Int. J. Comput. Vision **74**, 59–73 (2007)

7. Chaudhari, K., Garg, D., Kotecha, K.: An enhanced approach in image mosaicing using orb method with alpha blending technique. International Journal of Advanced Research in Computer Science **8**(5) (2017)

8. Chum, O., Matas, J.: Homography estimation from correspondences of local elliptical features. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). pp. 3236–3239. IEEE (2012)

9. Huang, T.S., Netravali, A.N.: Motion and structure from feature correspondences: A review. Proc. IEEE **82**(2), 252–268 (1994)

10. Jakubović, A., Velagić, J.: Image feature matching and object detection using brute-force matchers. In: 2018 International Symposium ELMAR. pp. 83–86. IEEE (2018)

11. Jia, Q., Li, Z., Fan, X., Zhao, H., Teng, S., Ye, X., Latecki, L.J.: Leveraging line-point consistence to preserve structures for wide parallax image stitching. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12186–12195 (2021)

12. Konen, W., Tombrock, S., Scholz, M.: Robust registration procedures for endoscopic imaging. Med. Image Anal. **11**(6), 526–539 (2007)

13. Levin, A., Zomet, A., Peleg, S., Weiss, Y.: Seamless Image Stitching in the Gradient Domain. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3024, pp. 377–389. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24673-2_31

14. Li, Y., Wang, Y., Huang, W., Zhang, Z.: Automatic image stitching using sift. In: 2008 International Conference on Audio, Language and Image Processing. pp. 568–571. IEEE (2008)

15. Liu, S., Wang, L., Liu, H., Su, H., Li, X., Zheng, W.: Deriving bathymetry from optical images with a localized neural network algorithm. IEEE Trans. Geosci. Remote Sens. **56**(9), 5334–5342 (2018)

16. Liu, Y., Tian, J., Hu, R., Yang, B., Liu, S., Yin, L., Zheng, W.: Improved feature point pair purification algorithm based on sift during endoscope image stitching. Front. Neurorobot. **16**, 840594 (2022)

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

18. Milgram, D.L.: Computer methods for creating photomosaics. IEEE Trans. Comput. **100**(11), 1113–1119 (1975)

19. Miranda-Luna, R., Daul, C., Blondel, W.C., Hernandez-Mier, Y., Wolf, D., Guillemin, F.: Mosaicing of bladder endoscopic image sequences: Distortion calibration and registration algorithm. IEEE Trans. Biomed. Eng. **55**(2), 541–553 (2008)

20. Okumura, K.i., Raut, S., Gu, Q., Aoyama, T., Takaki, T., Ishii, I.: Real-time feature-based video mosaicing at 500 fps. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 2665–2670. IEEE (2013)

21. Ozyoruk, K.B., Gokceler, G.I., Bobrow, T.L., Coskun, G., Incetan, K., Almalioglu, Y., Mahmood, F., Curto, E., Perdigoto, L., Oliveira, M., et al.: Endoslam dataset and an unsupervised monocular visual odometry and depth estimation approach for endoscopic videos. Med. Image Anal. **71**, 102058 (2021)

22. Penza, V., Ciullo, A.S., Moccia, S., Mattos, L.S., De Momi, E.: Endoabs dataset: Endoscopic abdominal stereo image dataset for benchmarking 3d stereo reconstruction algorithms. The International Journal of Medical Robotics and Computer Assisted Surgery **14**(5), e1926 (2018)

23. Porter, T., Duff, T.: Compositing digital images. In: Proceedings of the 11th annual conference on Computer graphics and interactive techniques. pp. 253–259 (1984)

24. Rong, W., Chen, H., Liu, J., Xu, Y., Haeusler, R.: Mosaicing of microscope images based on surf. In: 2009 24th International Conference Image and Vision Computing New Zealand. pp. 271–275. IEEE (2009)

25. Su, H., Qi, W., Hu, Y., Karimi, H.R., Ferrigno, G., De Momi, E.: An incremental learning framework for human-like redundancy optimization of anthropomorphic manipulators. IEEE Trans. Industr. Inf. **18**(3), 1864–1872 (2020)

26. Tang, Y., Liu, S., Deng, Y., Zhang, Y., Yin, L., Zheng, W.: Construction of force haptic reappearance system based on geomagic touch haptic device. Comput. Methods Programs Biomed. **190**, 105344 (2020)

27. Tang, Y., Liu, S., Deng, Y., Zhang, Y., Yin, L., Zheng, W.: An improved method for soft tissue modeling. Biomed. Signal Process. Control **65**, 102367 (2021)

28. Tian, F., Shi, P.: Image mosaic using orb descriptor and improved blending algorithm. In: 2014 7th International Congress on Image and Signal Processing. pp. 693–698. IEEE (2014)

29. Yang, B., Liu, C., Zheng, W., Liu, S., Huang, K.: Reconstructing a 3d heart surface with stereo-endoscope by learning eigen-shapes. Biomed. Opt. Express **9**(12), 6222–6236 (2018)

30. Zaragoza, J., Chin, T.J., Brown, M.S., Suter, D.: As-projective-as-possible image stitching with moving dlt. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2339–2346 (2013)

31. Zhang, Z., Wang, L., Zheng, W., Yin, L., Hu, R., Yang, B.: Endoscope image mosaic based on pyramid orb. Biomed. Signal Process. Control **71**, 103261 (2022)