Specular Region Detection and Covariant Feature Extraction

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. Endoscopy images pose a distinct set of challenges, such as specularity, uniformity, and deformation, which can obstruct surgeons' observations and decision-making processes. These hurdles complicate feature extraction and may ultimately lead to the failure of a surgical navigation system. To tackle these obstacles, we introduce a Modified Maximal Stable Extremal Region (MMSER) detector that specifically targets fine specular regions. Subsequently, we ingeniously fuse the capabilities of MMSER and saturation region properties to precisely identify specular regions within endoscopy images. Furthermore, our approach harnesses the shared properties of covariant features and endoscopic imaging to detect features in intricate regions, such as low-textured and deformed areas. Surpassing contemporary methods, our proposed technique demonstrates remarkable performance when evaluated on the available CVC-ClinicSpec datasets. Our method has shown improvements in accuracy, recall, f1-score, and Jaccard index by 0.21%, 25.42%, 7.77% snd 11.77%, respectively, when compared to recent techniques. Owing to its exceptional ability to accurately pinpoint specular regions and extract features from complex areas, our approach holds the potential to significantly advance surgical navigation.

Keywords: Endoscopy Imaging · Specular Region · Saturation Region · Feature Extraction · Feature Matching

1 Introduction

Endoscopic imaging systems have revolutionized medical procedures, enabling quicker recovery times and less invasive surgeries compared to traditional methods. Doctors use their experience to estimate spatial relationships and distances within the surgical environment [30]. However, the narrow field of view in endoscopy images often forces surgeons to perform multiple observations to

A. Antonacopoulos et al. (Eds.): ICPR 2024, LNCS 15312, pp. 170–186, 2025.
https://doi.org/10.1007/978-3-031-78198-8_12

gather information about the same area, which increases the risk and duration of the operation [21]. Additionally, the 2D images lack depth information, making it difficult for doctors to accurately determine the movements of surgical instruments [31]. Therefore, recognizing the 3D structure during the operation is pivotal for doctors and correct feature matching is essential to achieve this. However, endoscopy images present difficulties for feature extraction due to the presence of specular and uniform regions.

Specularity is a constant challenge in endoscopic images, as the angles of the lighting source and camera are nearly identical, causing valuable information like vessels and lesions to be concealed. Specular reflections lead to significant discontinuities, resulting in lost image texture and color information, which hinders the surgeon's observation and judgment [29]. Several existing studies focus on specularity detection in endoscopy imaging. Endoscopic image specularity detection methods can be broadly categorized into those based on different color spaces [12] and those employing classifiers [1].

Oh et al. [24] defined specular reflection areas as absolute bright areas and relative bright areas, determined through outlier detection. However, the detected relative bright areas may include not only specular highlights but also white tissues. Shen et al. [28] transformed endoscopic images into grayscale and detected specular regions using an empirical grayscale threshold, followed by mask region expansion through morphological techniques. However, this method is only suitable for endoscopic images with uniform brightness. Asif et al. [5] employed the Intrinsic Image Layer Separation (IILS) technique to identify specular regions, but this approach misidentifies edges and highly saturated areas as highlights in highly saturated, high-resolution images. Nie et al. [23] suggested a technique for detecting specular regions through brightness classification, enhancement, and thresholding. Although the concept of brightness enhancement is promising, the technique's reliance on different fixed thresholds based on image brightness is not ideal. This approach may fail to detect complex specular regions, such as larger white tissue regions containing specularity.

Extracting features from endoscopy images can be a daunting task, especially when the scene contains specularity, deformation, and low texture [4]. Existing feature extraction methods, such as Scale-Invariant Feature Transform (SIFT) [20], Speeded Up Robust Features (SURF) [8], Oriented Fast and Rotated BRIEF (ORB) [26] and Harris [34] are typically used in 3D reconstruction but are unable to compute enough good feature points from endoscopy images. As such, finding enough good features and correct matching in continuous endoscopy image frames is a critical aspect of recognizing the 3D structure during surgery.

There are only a few works that attempt to extract features in endoscopy imaging. Yan et al. [19] proposed using SIFT for feature extraction and improving the matching process through feature-point pair purification. Although this technique enhances matching performance, it overlooks the fact that having enough available features is crucial for improved matching. In a recent study, Barbed et al. [7] introduced a self-supervised SuperPoint [13] adaptation for the

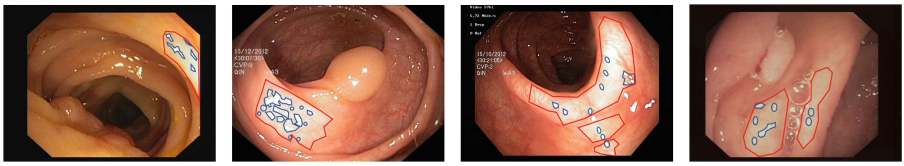
endoscopic domain. However, this learning-based technique has computational complexity, and the adapted model avoids features within specular regions.

Recent advances in specularity removal have yielded promising results.Pan et al. [25] introduced an accelerated adaptive non-convex robust principal component analysis (AANC-RPCA) method that enhances the efficiency and accuracy of highlight removal through adaptive threshold segmentation and quasi-convex function approximation. Zhang et al. [35] developed a partial attention network (PatNet) that employs highlight segmentation and image inpainting, significantly improving the visual quality of endoscopic images. Another innovative approach by Joseph et al. [17] presents a parameter-free matrix decomposition technique that decomposes the original image into a highlight-free pseudo-low-rank component and a highlight component, effectively removing specular reflections and boundary artifacts. These methods demonstrate significant progress in addressing the challenges of specular highlight removal in endoscopic imaging.

Existing techniques primarily focus on detecting specular regions. However, these methods often suffer from high computational complexity and insufficient detection in complex situations, such as when large white tissues overlap with densely specular regions, as shown in Figure 1. It displays white tissue regions enclosed by red boundaries and specular regions enclosed by blue boundaries from publicly available datasets CVC-ClinicSpec [27], Kvasir-Seg [16], Hyper-Kvasir [11], and CVC-ClinicDB [9]. Both these regions exhibit similar properties, including high intensity and low saturation, which often causes existing techniques to misidentify them as a single specular region, known as a false specular region. These false regions are generally larger than the actual specular regions because they include a combination of specular and white tissue regions. Saturation detectors can effectively identify these false specular regions, as they detect regions as specular if they possess high intensity and low saturation. However, the inadequate detection of false regions by existing methods makes their removal challenging, subsequently impacting feature detection. Moreover, covariant detectors can consistently detect affine-invariant frames in deformed and textureless regions, from which distinctive SIFT descriptors can be extracted for reliable matching.

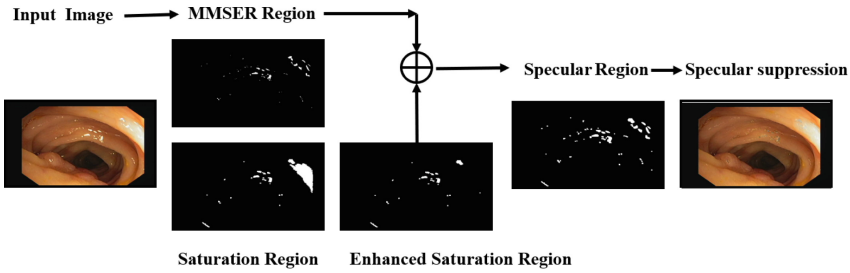
Our technical contributions are as follows:

- We introduce MMSER as a method for identifying fine specular regions.
- By integrating the specular regions identified using MMSER with those detected by the saturation detector, we effectively recognize complex specular regions and eliminate false regions.
- To tackle the complexity of feature extraction in deformed and low-texture regions of endoscopy images, we utilize the affine invariance properties of covariant detectors and the distinctiveness of SIFT descriptors.
- By employing adaptive distance thresholding and outlier rejection, we enhance the accuracy of matching.
- Our technique has shown improvements compared to recent techniques in accuracy 0.21%, recall 25.42%, f1-score 7.77%, and jaccard 11.77%.

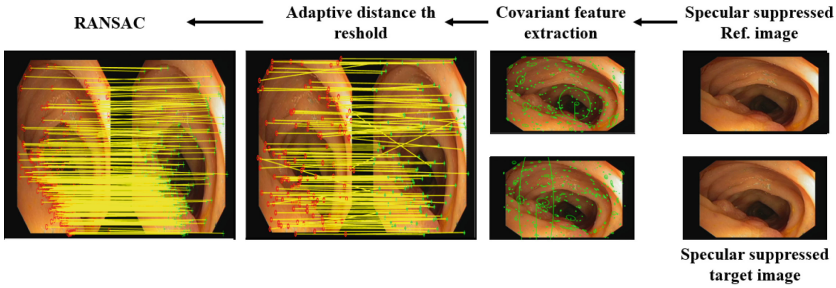


(a) CVC-ClinicSpec (b) Kvasir-Seg (c) HyperKvasir (d) CVC-ClinicDB

Fig. 1. Images with overlapping white tissue and specular regions from available public datasets.



(a) Specular Detection and Removal Module



(b) Covariant Feature Extraction and Matching Module

Fig. 2. Schematic of the proposed technique

2 Proposed Method

Capitalizing on the synergistic combination of MMSER and saturation regions, our proposed technique offers significant advantages over existing methods, as it can adeptly handle complex situations such as false, low-texture, and deformed regions. Furthermore, our affine adaptation of covariant features enables the detection of features within low-textured and deformed areas. The affine invariance properties of these features also enhance matching accuracy. With the capacity to extract features from intricate regions and improve matching accuracy, our method paves the way for precise 3D dense reconstructions in endoscopy

imaging, ultimately contributing to significant advancements in surgical navigation.

In essence, our proposed technique consists of the following steps: First, we separately detect MMSER and saturation regions. Next, we enhance the saturation region by suppressing false region areas, which typically correspond to white tissue regions containing specularity. We then merge the MMSER and the enhanced saturation regions to obtain our final detected specular region. Subsequently, we remove the detected specular region using existing techniques. Afterward, we extract covariant feature frames and SIFT descriptors with affine adaptation and match features between two images. Although there are some mismatched features, we employ adaptive distance thresholding to eliminate them. However, some mismatches may persist, so we use RANSAC to obtain accurate matches between features. The flowchart of our proposed technique is illustrated in Figure 2.

We can define the intensity function $\boldsymbol{I}$ using red (r), green (g) and blue (b) component of an image as follows:

$$\boldsymbol{I} = \frac{1}{3}(r + g + b). \tag{1}$$

2.1 Modified Maximally Stable Extremal Regions Detection

The MSER algorithm [22] identifies extremal regions (maximum or minimum intensity) as connected components within the level sets of an image. Among these extremal regions, locally maximally stable ones are chosen. The absence of smoothing enables the detection of both fine and large structures. While MSER's properties are valuable for extracting extremal regions and varying region sizes, our goal is to detect only maximum intensity and fine structures. Consequently, we modify the MSER detector's properties to suit our specific requirements for capturing higher intensity and fine regions in endoscopy images. The steps of our proposed MMSER detection method can be outlined as follows.

- We assume $\mathbf{I}(x, y)$ represent the intensity of an image at pixel location (x, y).
- We define an extremal region $\boldsymbol{M_R}$ as a connected component of an image $\boldsymbol{I}$, such that $\forall(x, y) \in \boldsymbol{M_R}$ and $\forall(x', y') \in \partial \boldsymbol{M_R}$ (the boundary of $\boldsymbol{M_R}$):

$$I(x, y) \geq I(x', y') + t, \tag{2}$$

  where, t is intensity threshold.
- We compute the area of the extremal region $\boldsymbol{M_R}$ for a range of intensity thresholds $t \in [0, 255]$. Let $A(t)$ denote the area of $\mathbf{R}$ at threshold t.
- Calculate the stability score $S(\boldsymbol{M_R})$ for each extremal region $\boldsymbol{M_R}$ as the absolute difference in areas over a range of intensity thresholds $\triangle t$:

$$S(\boldsymbol{M_R}) = A(t + \triangle t) - A(t), \tag{3}$$

  where $\triangle t$ is the sensitivity of stability.

– We consider an extremal region $\boldsymbol{M_R}$ is maximally stable $\boldsymbol{M}_{RS}$ if its stability score $S(\boldsymbol{M_R})$ is locally minimal compared to its neighboring regions in the intensity range, and its area $A(t)$ is smaller than a predefined maximum area A_{max}. Mathematically, this can be expressed as:

$$\boldsymbol{M}_{RS} = \begin{cases} \boldsymbol{M_R} & \text{if } S(\mathbf{R}) < S(\mathbf{R}') \text{ and } A(t) < A_{max} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where **R**' is a set of neighbouring regions.

2.2 Intensity-Saturation Regions Detection

Previously, in [32], the author established a correlation between intensity and saturation. In this study, specular regions in images were detected using the bi-directional histogram concept. The approach was based on the observation that specular areas exhibit higher brightness and lower saturation than surrounding regions. Although this technique was effective in detecting specular regions, it had limitations. For instance, it often incorrectly identified the white tissue region as specular because of its high intensity and low saturation, and it could not detect small regions. In this study, we leverage the intensity-saturation technique to detect the presence of white tissue in endoscopy images. Once we identify the white tissue region, we apply morphological operations to suppress it, resulting in an enhanced saturation region.The enhanced intensity-saturation regions $\boldsymbol{IS}_{en}$ are calculated as follows:

– We denote $\boldsymbol{S}$ as the saturation of an image, which can be expressed as:

$$\boldsymbol{S} = \begin{cases} \frac{1}{2}(2r - g - b) = \frac{3}{2}(r - m) & , if \quad (b + r) \geq 2g \\ \frac{1}{2}(r + g - 2b) = \frac{3}{2}(m - b) & , if \quad (b + r) < 2g \end{cases} \tag{5}$$

– To identify the specular region $\boldsymbol{IS}$ in the image, we consider each pixel p and check if it satisfies the following conditions:

$$\boldsymbol{IS} = \begin{cases} \boldsymbol{I}_p \geq \frac{1}{2}\boldsymbol{I}_{max} \\ \boldsymbol{S}_p \leq \frac{1}{3}\boldsymbol{S}_{max} \end{cases} \tag{6}$$

– We compute the connected component $\boldsymbol{c}$ in $\boldsymbol{IS}$.
– We compute the area $A_S(i)$ for each connected component.
– Next, we compute the enhanced intensity-saturation regions $\boldsymbol{IS}_{en}$ as follows:

$$\boldsymbol{IS}_{en} = \begin{cases} \boldsymbol{I}_m & , if \quad A_S(i) \leq t_h \quad \forall \boldsymbol{c} \\ 0 & otherwise \end{cases} \tag{7}$$

where t_h is an adaptive threshold that depends on the size of the image. For the CVC-ClinicSpec dataset, we evaluated with various threshold values and determined that the appropriate threshold value is 20.

2.3 Integration of Detected Regions

We combine the modified maximally stable extremal regions $\boldsymbol{M}_{RS}$ with the enhanced intensity-saturation regions $\boldsymbol{IS}_{en}$ to obtain the final specular regions $\boldsymbol{S}_R$.

$$\boldsymbol{S}_R = \boldsymbol{M}_{RS} + \boldsymbol{IS}_{en}. \tag{8}$$

Figure 2 top row appropriately illustrates the steps of our proposed technique for detecting specular regions, showcasing the effectiveness of our approach in complex situations. To highlight the need for our method, we carefully selected an image including false regions from the CVC-ClinicSpec dataset. Our process begins by converting the colored image to a grayscale image, followed by computing the MMSER ($\boldsymbol{M}_{RS}$) containing relatively fine specular regions. Subsequently, we compute the intensity-saturation region $\boldsymbol{IS}$, which encompasses relatively larger regions, including the crucial false regions (if present in the image). We refine the false regions based on their existence to obtain the enhanced intensity-saturation region $\boldsymbol{IS}_{en}$. Finally, we integrate the $\boldsymbol{M}_{RS}$ with the $\boldsymbol{IS}_{en}$ to obtain the final specular region $\boldsymbol{S}_R$ with remarkable precision and accuracy.

2.4 Specular Region Suppress

We use the technique [10] to remove specularity and achieve a clean image without specularity after detecting the specular region $\boldsymbol{S}_R$ in the previous step.

$$\boldsymbol{I}_{Refine} = \boldsymbol{I} \otimes \boldsymbol{S}_R, \tag{9}$$

where the operator $\otimes$ represents the specular removal operation implemented in [10].

2.5 Affine-Invariant Feature Detection

In the previous step, we obtained the specular-removed image $\boldsymbol{I}_{Refine}$, which we use to extract affine-invariant features. Using a covariant detector [33], we detect feature frames $\boldsymbol{F}$ that are defined by an affine matrix comprising a translation vector $\boldsymbol{t}_r$ and a linear map $\boldsymbol{L}$. These feature frames define elliptical regions in the image. Next, we extract description vectors $\boldsymbol{d}_v$ from these regions. An affine-invariant feature is associated with a matrix $\boldsymbol{F}$ and a vector $\boldsymbol{d}_v$.

$$\boldsymbol{F} = \left|\boldsymbol{L}\ \boldsymbol{t}_r\right| = \begin{vmatrix} l_{11}\ l_{12}\ t_{r1} \\ l_{21}\ l_{22}\ t_{r2} \end{vmatrix} \tag{10}$$

where the translation vector $\boldsymbol{t}_r$ represents the location of an image, while the linear map $\boldsymbol{L}$ represents the shape and orientation of the local features.

Following this, we extract a descriptor vector of dimension $128 - D$ from the detected region using the SIFT method. This descriptor will be used to calculate

Algorithm 1. Detection of specular highlights, extraction of affine-invariant features, and improvement of matching accuracy in endoscopy image pairs.

Require: Endoscopy image pair.
1: **for** each image **do**
2: Compute the modified MSE regions M_{RS}.
3: Compute the saturation regions IS.
4: Compute the enhanced saturation regions IS_{en}.
5: Integrate M_{RS} and IS_{en} to get S_R.
6: Compute I_{refine} using Arnold's method [30] to suppress specularity.
7: Extract the Affine-invariant features from I_{refine} by computing:
8: (i) Feature frame F
9: (ii) Feature descriptor d_v
10: **end for**
11: Compute the initial match between the image pair using the descriptors d_v and the Brute-Force technique [15].
12: Apply adaptive distance thresholding to improve the initial match.
13: Remove outliers using graph-cut RANSAC [6].

the matching confidence between two features by computing the distance ratio e_n.

$$e_n = \frac{d_{n,closest}}{d_{n,closest2}} = \frac{\| \boldsymbol{d}_{v,n} - \boldsymbol{d}_{v,m_{closest}} \|}{\| \boldsymbol{d}_{v,n} - \boldsymbol{d}_{v,m_{closest2}} \|} \tag{11}$$

In comparing features between two images, $\boldsymbol{d}_{v,n}$ and $\boldsymbol{d}_{v,m}$ denote the feature descriptors, while $d_{n,closest}$ is the descriptor distance between a particular feature and its nearest neighbor in the other image, and $d_{n,closest2}$ is the distance to the second-nearest neighbor. A smaller value of e_n indicates a greater similarity between the two features, increasing the likelihood that the correspondence is an inlier.

2.6 Feature Matching

Once the Affine-invariant features are computed, we can perform matching between corresponding images by comparing descriptors vector $\boldsymbol{d}_v$, using the Brute-force (BF) method [15].

Next, we will use the distance ratio e_n between the two best matching descriptors and only accept matches below an adaptive threshold th_{en}. After extensive evaluation of various endoscopy datasets, we determined the optimal threshold value to be 0.91.

Finally, we perform geometric verification on the previous matching results. The geometric verification technique Graph-Cut RANSAC [6] improves the matching from the previous step by removing outliers and mismatches that do not satisfy the geometric constraint.

3 Computational Procedures

Algorithm 1 outlines the complete steps of our proposed method for specular detection, affine-invariant feature extraction, and matching accuracy improvement.

4 Experimental Results and Analysis

4.1 Implementation and Data

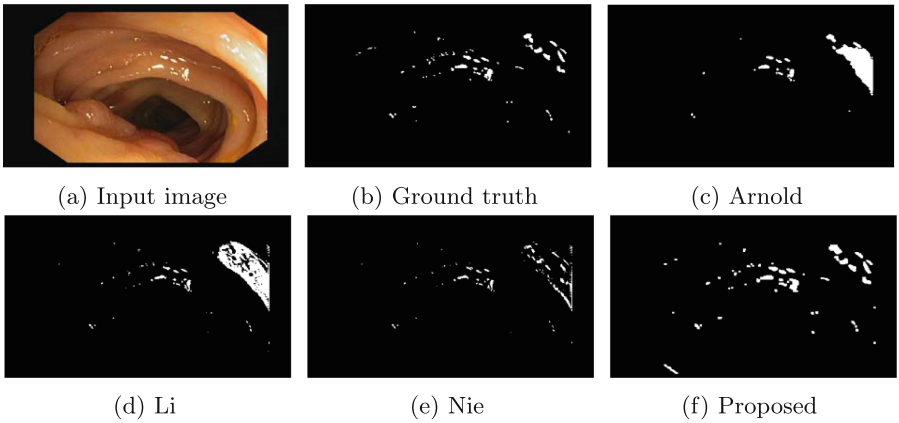
Our study involves evaluating the effectiveness of the proposed technique in detecting specular regions and extracting affine-invariant features in endoscopy imaging. To accomplish this, we compare our approach to state-of-the-art specular detection and feature extraction techniques. The experiments are conducted on a Windows 10 x64 system, utilizing OpenCV 4.0 for image processing. The hardware consists of an Intel i5-9400K with a 2.90 GHzX2 processor and 16 GB of RAM. Implementation of the algorithm is carried out using both Matlab 2022 and Python 3.8. Our evaluation is conducted on the publicly available CVC-ClinicSpec dataset, which contains colonoscopy images with annotated specular ground truth labels. To measure the effectiveness of our proposed approach, we use gold standard metrics such as Accuracy, Precision, Recall, F1-score, and Jaccard. We also evaluated the performance of affine-invariant feature extraction and matching in challenging conditions using the Hyper-Kvasir dataset. Specifically, we used the labeled videos in "lower-gi-tract/quality-of-mucosal-view/BBPS-2-3" from the dataset. We extracted and matched features across pairs of frames taken 1 second apart from each other within the sequences of the Hyper-Kvasir dataset (where 1 second $\bar{2}5$ frames). To evaluate the result, we used 10% of the frames from the videos. Our assessment of the detection algorithm's ability to segment the specular region involves the use of five metrics: Accuracy, Precision, Recall, F1-score, and Jaccard.

Our main contribution in this research is the detection of specular regions and the extraction of covariant features. Specifically, we focus on extracting covariant features, as specular regions often mislead feature extraction processes, complicating surgical navigation. We utilize an existing technique to suppress the detected specular regions, with the source code publicly available. In the following sections, we will evaluate our detection and feature extraction performance both qualitatively and quantitatively.

4.2 Visual Evaluation

In this part we will visually asses the specular region detection, covariant feature extraction and matching ability in challenging endoscopy imaging condition.

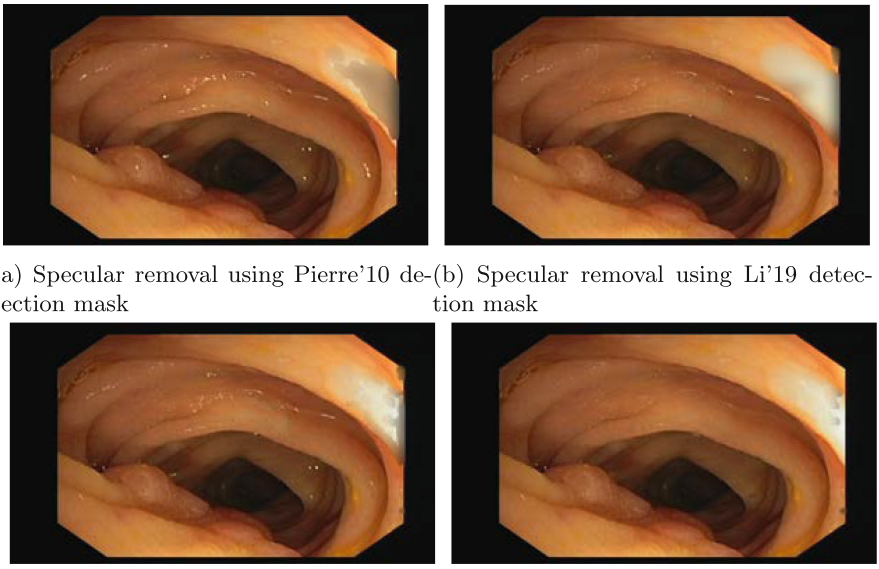
Specular Region Detection and Suppression Figure 3 presents a comparison between recent state-of-the-art techniques and our proposed method for detecting specular regions. The input image and ground truth from the CVC-ClinicSpec dataset are shown. Specifically, Arnold's method [3] fails to detect



(a) Input image (b) Ground truth (c) Arnold

(d) Li (e) Nie (f) Proposed

Fig. 3. Comparison between different detection methods

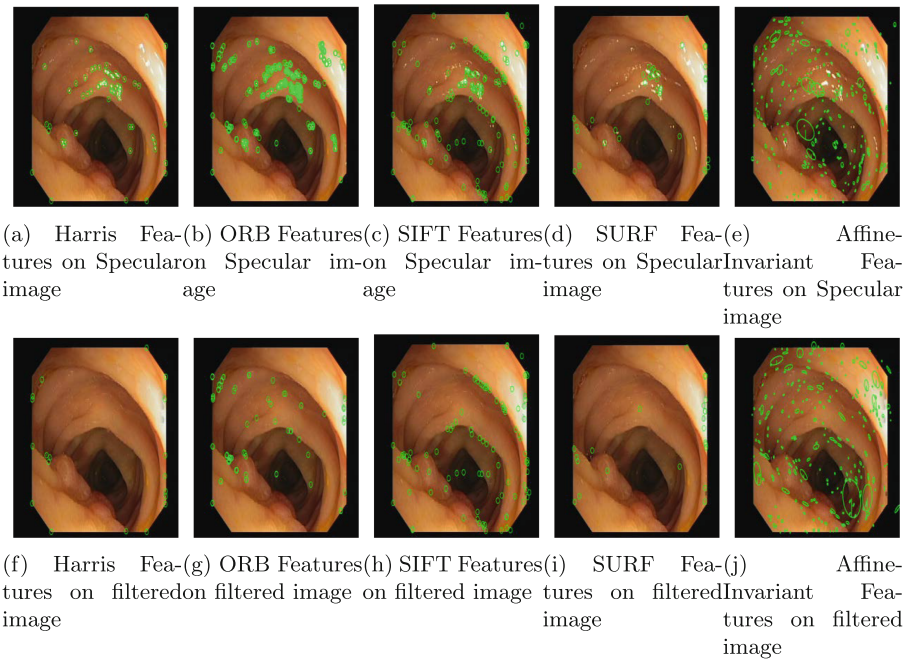


(a) Specular removal using Pierre'10 detection mask (b) Specular removal using Li'19 detection mask

(c) Specular removal using Nie'23 detection mask (d) Specular removal using Proposed detection mask

Fig. 4. Specularity removal using detected region from different techniques

fine specular regions and misidentifies the false region as a specular region. Li's method [18] is better at detecting fine specular regions, but also misclassifies the false region. Nie's method [23] can detect fine regions and specular regions on white tissue, but produces false edges that may affect the specular suppression process. In contrast, our proposed technique effectively detects fine specular regions and also accurately extracts specular regions from white tissue.

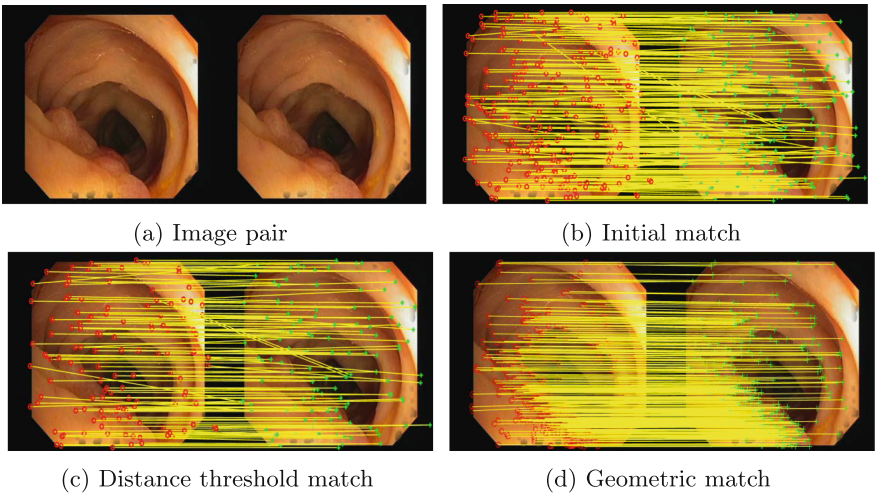


(a) Harris Features on Specular image
(b) ORB Features on Specular image
(c) SIFT Features on Specular image
(d) SURF Features on Specular image
(e) Affine-Invariant Features on Specular image
(f) Harris Features on filtered image
(g) ORB Features on filtered image
(h) SIFT Features on filtered image
(i) SURF Features on filtered image
(j) Affine-Invariant Features on filtered image

Fig. 5. Comparison between feature detectors

To suppress specular regions, we utilized Arnold's technique [3]. Figure 4 depicts the results of specular suppression using the detected specular region mask discussed in Figure 4. The results indicate that using Arnold's technique and regions detected by Li's method [18] do not effectively suppress the specular regions, and additionally introduce artifacts around the white tissue region. On the other hand, Nie's detected regions [23] perform better in removing the specular region compared to Arnold's and Li's methods, but still contain some specular regions and artifacts near the white tissue region. In contrast, our proposed detected region performs exceptionally well in suppressing the specular regions, with ignorable artifacts on the white tissue region.

Feature Detection and Matching The effectiveness of covariant detectors compared to commonly used detectors in endoscopy imaging is demonstrated in Figure 5. In Figure 5 top row, where the image contains specular regions, we observe that Harris, ORB, SIFT, and SURF detectors mostly detect features in specular highlight and corner-like areas, while avoiding low-texture areas. In contrast, covariant detector [33] detects affine-invariant frames throughout all the regions. Furthermore, in the filtered image in Figure 5 bottom row, we see a significant reduction in the number of features detected by Harris, ORB, SIFT, and SURF detectors, due to the absence of specular regions, while covariant detector detects a similar number of features throughout all the regions.



(a) Image pair (b) Initial match

(c) Distance threshold match (d) Geometric match

Fig. 6. Steps for improving matching accuracy

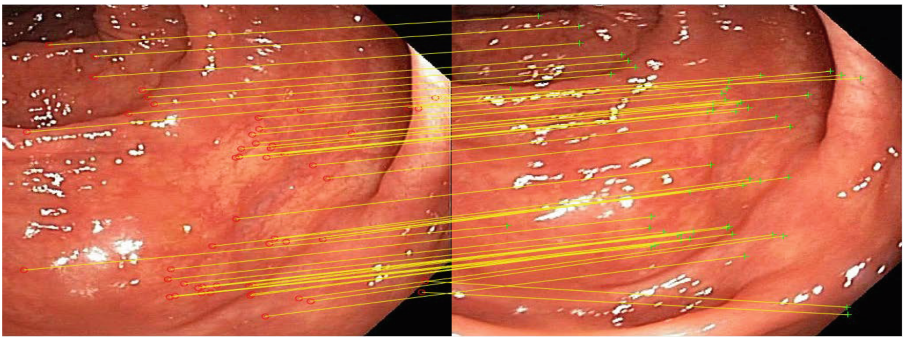


Fig. 7. Pair of 1-second-apart frames were used from the Hyper-Kvasir dataset to extract affine-invariant features and matches of inliers.

The matching process of affine-invariant features is illustrated in Figure 6. The filtered image sequence is represented in Figure 6(a). The initial matching results with numerous mismatches are depicted in Figure 6(b). By applying adaptive distance thresholding, the number of mismatches is significantly reduced, as shown in Figure 6(c). However, some mismatches still exist. Finally, Figure 6(d) demonstrates that there are no mismatches after applying the graph-cut RANSAC geometric verification technique.

Our evaluation uses the Hyper-Kvasir dataset with specularity, which allows us to compare our results with those in [7]. As shown in Figure 7, the detected affine-invariant features from the Hyper-Kvasir dataset are not confined to specific areas such as corners or specular regions but are also present in low-texture

areas. Additionally, the figure demonstrates the feature matching between two pairs of frames taken 1 second apart from the dataset.

4.3 Quantitative Evaluation

Table 1 presents a quantitative comparison of different techniques for detecting specular regions on the CVC-ClinicSpec dataset. The bold values indicate the best detection performance. Our proposed method outperforms the other methods in terms of Accuracy, Recall, F1-score, and Jaccard. Specifically, compared to the best-performing method Nie [23], our method achieves higher Accuracy, F1-score, and Jaccard by 0.21, 7.77, and 11.77 percent respectively. Additionally, our Recall value is 1.97 percent higher than that of Meslouhi [14].

These improvements are primarily due to our innovative combination of MMSER with saturation region properties, which enhances the detection and management of specular regions in endoscopy images. This dual approach allows for more precise identification of true positives, especially in low-textured and deformed areas, leading to significantly higher recall and better alignment with the ground truth, as reflected in the Jaccard Index. Although our precision (0.8516) is slightly lower than that of Nie et al., the substantial gains in recall and the balanced F1-Score underscore the robustness and effectiveness of our method in accurately detecting and managing specular regions, which is crucial in medical imaging for ensuring comprehensive and reliable feature extraction.

Table 1. Quantitative comparison between state-of-art techniques

Methods	Accuracy	Precision	Recall	F1-Score	Jaccard
Arnold et al. [3]	0.9837	0.8938	0.6739	0.7684	0.6589
Meslouhi et al. [14]	0.9920	0.3744	0.8961	0.5281	0.6616
Alsaleh et al. [2]	0.9699	0.6016	0.8020	0.6875	0.6016
Shen et al. [28]	0.9767	0.9064	0.6683	0.7694	0.6518
Asif et al. [5]	0.9584	0.6972	0.7151	0.6489	0.6333
Nie et al. [23]	0.9932	**0.9083**	0.7286	0.8085	0.7153
Proposed	**0.9953**	0.8516	**0.9138**	**0.8713**	**0.7995**

Table 2 provides a quantitative comparison between commonly used detectors and the covariant detector, demonstrating the efficacy of the latter in extracting features from complex, low-texture, and deformed regions, particularly in endoscopy imaging. As shown in the table, the covariant detector is capable of extracting a very large number of features, 4907 and 4847 in the specular and filtered image, respectively, outperforming ORB, SIFT, Harris, and SURF. Furthermore, the table shows that our method achieves a greater number of inliers, 1274, compared to ORB, SIFT, Harris, and SURF after applying distance thresholding and Graph-cut RANSAC on initial Brute-Force matching to the extracted affine-invariant features.

The performance of various state-of-the-art techniques is presented in Table 3 using the Hyper-Kvasir dataset. The proposed technique outperforms other techniques in terms of feature extraction and inlier matching and is not limited to specific areas such as corners or specular regions. The table shows that the proposed technique achieves superior quantitative results compared to other techniques.

Table 2. Quantitative comparison between feature detectors

Methods	Specular image		Filtered image	
	Features	Inliers	Features	Inliers
Harris [34]	59	2	19	1
ORB [26]	447	29	80	6
SIFT [20]	140	22	101	17
SURF[8]	40	14	23	12
Affine-Invariant	4907	1344	4847	1274

Table 3. Matching quality metrics for state-of-the-art techniques using the Hyper-Kvasir dataset [4].

Methods	Feat/Img	F Inl.
SIFT [20]	825.7	151.3
ORB [26]	361.3	137.2
SP Base [13]	211.8	51.3
E-SP [7]	591.3	200.4
Affine-Invarinat	**1125.7**	**281.3**

5 Conclusions and Future Work

Our study proposes a novel specular detection technique that leverages the affine invariance properties of covariant detectors to overcome challenges in complex endoscopy imaging scenarios. Our approach enhances the characteristics of the Maximal Stable Extremal Region (MSER) to detect fine specular regions, followed by the computation of saturation regions. By merging the Modified MSER (MMSER) and saturation regions, our method accurately pinpoints areas of specular reflection. Using an existing method for suppression of specular regions, we refine the image and compute affine-invariant features, resulting in the extraction of a significant number of high-quality features. Our technique excels at detecting intricate specular regions and extracting features from uniform and deformed areas in endoscopy images. This state-of-the-art approach has the potential to enhance surgical navigation precision significantly.

Our detection technique is evaluated on the CVC-ClinicSpec. Visual evaluation reveals that our approach can successfully extract specular regions in complex conditions where other state-of-the-art techniques fail. Furthermore, in terms of Accuracy, Recall, F-1 Score, and Jaccard, our technique outperforms other existing methods in quantitative evaluation. Our evaluation also shows that recent techniques fail to detect the false regions, leading to artifacts during specular region suppression, unlike our technique.

Our approach to specular detection can significantly impact surgical navigation by enhancing the clarity and reliability of endoscopic images through any efficient removal process. Specular reflections in these images can obscure critical anatomical details, leading to potential misinterpretations during surgical procedures. By accurately detecting and suppressing these specular regions, the

proposed technique ensures that important features are preserved and accurately extracted. This improved image quality facilitates better spatial understanding and decision-making for surgeons, thereby increasing the safety and efficacy of minimally invasive surgeries. Enhanced feature extraction and matching also contribute to more precise 3D reconstructions, further aiding in navigation and reducing the risk of complications.

The most robust aspect of the method is its combination of MMSER with saturation region properties to accurately detect and manage specular regions in endoscopy images. This technique significantly enhances the detection and extraction of features from low-textured and deformed areas, demonstrating notable improvements in metrics such as accuracy, recall, F1-score, and Jaccard index compared to existing methods. Conversely, the least effective aspect is the specular removal method used, which does not perform optimally. In the future, we plan to develop our own specular removal technique.

Acknowledgement. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00250742, 2022R1A4A3018824, RS-2024-00438248, RS-2022-00155885). This research was also supported by the MSIT(Ministry of Science and ICT), Korea under the ITRC(Information Technology Research Center) support program(IITP-2023-RS-2023-00259061) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

References

1. Ali, S., Zhou, F., Bailey, A., Braden, B., East, J.E., Lu, X., Rittscher, J.: A deep learning framework for quality assessment and restoration in video endoscopy. Med. Image Anal. **68**, 101900 (2021)
2. Alsaleh, S.M., Aviles-Rivero, A.I., Hahn, J.K.: Retouchimg: Fusioning from-local-to-global context detection and graph data structures for fully-automatic specular reflection removal for endoscopic images. Comput. Med. Imaging Graph. **73**, 39–48 (2019)
3. Arnold, M., Ghosh, A., Ameling, S., Lacey, G.: Automatic segmentation and inpainting of specular highlights for endoscopic imaging. EURASIP Journal on Image and Video Processing **2010**, 1–12 (2010)
4. Asif, M., Chen, L., Song, H., Yang, J., Frangi, A.F.: An automatic framework for endoscopic image restoration and enhancement. Appl. Intell. **51**, 1959–1971 (2021)
5. Asif, M., Song, H., Chen, L., Yang, J., Frangi, A.F.: Intrinsic layer based automatic specular reflection detection in endoscopic images. Comput. Biol. Med. **128**, 104106 (2021)
6. Barath, D., Matas, J.: Graph-cut ransac. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6733–6741 (2018)
7. Barbed, O.L., Chadebecq, F., Morlana, J., Montiel, J.M., Murillo, A.C.: Superpoint features in endoscopy. In: MICCAI Workshop on Imaging Systems for GI Endoscopy. pp. 45–55. Springer (2022)
8. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9. pp. 404–417. Springer (2006)

9. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized medical imaging and graphics **43**, 99–111 (2015)
10. Bertalmio, M., Bertozzi, A.L., Sapiro, G.: Navier-stokes, fluid dynamics, and image and video inpainting. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. vol. 1, pp. I–I. IEEE (2001)
11. Borgli, H., Thambawita, V., Smedsrud, P.H., Hicks, S., Jha, D., Eskeland, S.L., Randel, K.R., Pogorelov, K., Lux, M., Nguyen, D.T.D., et al.: Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. Scientific data **7**(1), 283 (2020)
12. Chu, Y., Li, H., Li, X., Ding, Y., Yang, X., Ai, D., Chen, X., Wang, Y., Yang, J.: Endoscopic image feature matching via motion consensus and global bilateral regression. Comput. Methods Programs Biomed. **190**, 105370 (2020)
13. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 224–236 (2018)
14. El Meslouhi, O., Kardouchi, M., Allali, H., Gadi, T., Benkaddour, Y.A.: Automatic detection and inpainting of specular reflections for colposcopic images. Central European Journal of Computer Science **1**, 341–354 (2011)
15. Jakubović, A., Velagić, J.: Image feature matching and object detection using brute-force matchers. In: 2018 International Symposium ELMAR. pp. 83–86. IEEE (2018)
16. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., De Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: MultiMedia modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26. pp. 451–462. Springer (2020)
17. Joseph, J., George, S.N., Raja, K.: Parameter-free matrix decomposition for specular reflections removal in endoscopic images. IEEE Journal of Translational Engineering in Health and Medicine **11**, 360–374 (2023)
18. Li, R., Pan, J., Si, Y., Yan, B., Hu, Y., Qin, H.: Specular reflections removal for endoscopic image sequences with adaptive-rpca decomposition. IEEE Trans. Med. Imaging **39**(2), 328–340 (2019)
19. 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)
20. Low, D.G.: Distinctive image features from scale-invariant keypoints. Journal of Computer Vision **60**(2), 91–110 (2004)
21. Marmol, A., Banach, A., Peynot, T.: Dense-arthroslam: Dense intra-articular 3-d reconstruction with robust localization prior for arthroscopy. IEEE Robotics and Automation Letters **4**(2), 918–925 (2019)
22. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image Vis. Comput. **22**(10), 761–767 (2004)
23. Nie, C., Xu, C., Li, Z., Chu, L., Hu, Y.: Specular reflections detection and removal for endoscopic images based on brightness classification. Sensors **23**(2), 974 (2023)
24. Oh, J., Hwang, S., Lee, J., Tavanapong, W., Wong, J., de Groen, P.C.: Informative frame classification for endoscopy video. Med. Image Anal. **11**(2), 110–127 (2007)
25. Pan, J., Li, R., Liu, H., Hu, Y., Zheng, W., Yan, B., Yang, Y., Xiao, Y.: Highlight removal for endoscopic images based on accelerated adaptive non-convex rpca decomposition. Comput. Methods Programs Biomed. **228**, 107240 (2023)

26. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: An efficient alternative to sift or surf. In: 2011 International conference on computer vision. pp. 2564–2571. Ieee (2011)
27. Sánchez, F.J., Bernal, J., Sánchez-Montes, C., de Miguel, C.R., Fernández-Esparrach, G.: Bright spot regions segmentation and classification for specular highlights detection in colonoscopy videos. Mach. Vis. Appl. **28**(8), 917–936 (2017)
28. Shen, D.F., Guo, J.J., Lin, G.S., Lin, J.Y.: Content-aware specular reflection suppression based on adaptive image inpainting and neural network for endoscopic images. Comput. Methods Programs Biomed. **192**, 105414 (2020)
29. Son, M., Lee, Y., Chang, H.S.: Toward specular removal from natural images based on statistical reflection models. IEEE Trans. Image Process. **29**, 4204–4218 (2020)
30. Song, J., Wang, J., Zhao, L., Huang, S., Dissanayake, G.: Mis-slam: Real-time large-scale dense deformable slam system in minimal invasive surgery based on heterogeneous computing. IEEE Robotics and Automation Letters **3**(4), 4068–4075 (2018)
31. Sui, C., Wu, J., Wang, Z., Ma, G., Liu, Y.H.: A real-time 3d laparoscopic imaging system: design, method, and validation. IEEE Trans. Biomed. Eng. **67**(9), 2683–2695 (2020)
32. Tchoulack, S., Langlois, J.P., Cheriet, F.: A video stream processor for real-time detection and correction of specular reflections in endoscopic images. In: 2008 joint 6th international IEEE northeast workshop on circuits and systems and TAISA conference. pp. 49–52. IEEE (2008)
33. Vedaldi, A., Fulkerson, B.: Vlfeat: An open and portable library of computer vision algorithms (2008) (2012)
34. Yuan, X.C., Pun, C.M.: Invariant digital image watermarking using adaptive harris corner detector. In: 2011 Eighth International Conference Computer Graphics, Imaging and Visualization. pp. 109–113. IEEE (2011)
35. Zhang, C., Liu, Y., Wang, K., Tian, J.: Specular highlight removal for endoscopic images using partial attention network. Physics in Medicine & Biology **68**(22), 225009 (2023)