

Received October 29, 2020, accepted November 19, 2020, date of publication November 24, 2020, date of current version December 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3040187

Saliency Improvement in Feature-Poor Surgical Environments Using Local Laplacian of Specified Histograms

ARTUR BANACH^{1,2}, MARIO STRYDOM², ANJALI JAIPRAKASH^{1,2}, GUSTAVO CARNEIRO³, CAMERON BROWN², ROSS CRAWFORD^{1,2}, AND AARON MCFADYEN^{1,2}, (Member, IEEE)

¹QUT Centre for Robotics, Queensland University of Technology, Brisbane, QLD 4000, Australia

²Science and Engineering Faculty, Queensland University of Technology, Brisbane, QLD 4000, Australia

³Faculty of Engineering, Computer and Mathematical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia

Corresponding author: Artur Banach (artur.banach@qut.edu.au)

This work was supported by the AISRF—Biotechnology Collaborative Research Project.

ABSTRACT Navigation in endoscopic environments requires an accurate and robust localisation system. A key challenge in such environments is the paucity of visual features that hinders accurate tracking. This article examines the performance of three image enhancement techniques for tracking under such feature-poor conditions including Contrast Limited Adaptive Histogram Specification (CLAHS), Fast Local Laplacian Filtering (LLAP) and a new combination of the two coined Local Laplacian of Specified Histograms (LLSH). Two cadaveric knee arthroscopic datasets and an underwater seabed inspection dataset are used for the analysis, where results are interpreted by defining visual saliency as the number of correctly matched key-point (SIFT and SURF) features. Experimental results show a significant improvement in contrast quality and feature matching performance when image enhancement techniques are used. Results also demonstrate the LLSHs ability to vastly improve SURF tracking performance indicating more than 87% of successfully matched frames. A comparative analysis provides some important insights useful in the design of vision-based navigation for autonomous agents in feature-poor environments.

INDEX TERMS Image enhancement, histogram specification, local Laplacian filtering, minimally invasive surgery.

I. INTRODUCTION

Minimally Invasive Surgery (MIS) has become a worldwide endeavour in surgical theatres over the last two decades. Reduced tissue damage, shorter procedure and recovery times, and improved pain management are the key benefits of MIS. Although MIS promises better health outcomes and efficiency improvements, it introduces a set of difficult challenges. Clinicians work under physically demanding ergonomics and, due to tremor and constrained precision, they may introduce unintentional damage [1]. Moreover, conventional MIS techniques do not provide explicit depth perception with respect to anatomies and introduce counter-intuitive hand-eye coordination between camera image and surgical tools, demanding extended training time. To help overcome these challenges the next generation

of MIS seeks to leverage robotic devices as an intermediary between the surgeon and the patient.

Medical robotic systems are designed to reduce tremor, and provide 3D vision and surgical tool active guidance. Accurate pose feedback of the distal end of the endoscope is necessary to provide the surgeon with either haptic [2] or enhanced visual information about relative distance with respect to surrounding tissues. To further automate such systems, it is required to accurately and robustly localize the surgical camera with respect to surrounding anatomies.

Visual localisation tasks are often tackled in several ways. Traditional methods defined image features as direct (intensity-based) [3] or indirect (key-point-feature based). Feature matching between frames and key-frames is then employed as a common front end processing task for Visual Odometry (VO) [4] and/or Simultaneous Localisation and Mapping (SLAM) frameworks [5]. These approaches then use the matched features to recover relative pose. Localisation is therefore challenging in surgical environments due

The associate editor coordinating the review of this manuscript and approving it for publication was Valentina E. Balas¹.

to the fundamental reliance on image feature correspondence. Surgical images are extremely feature-poor (low in texture/uninformative), contain numerous occlusions (smoke, blood, tools, floating debris, water bubbles), appear blurred (motion, debris on the lens), contain glare, and view highly deformable structures. As such, highly robust vision-based localisation in surgery presents an interesting and challenging problem for image enhancement.

In this article, we address the problem of feature paucity. We propose such an image enhancement method for low-texture gray-scale images and focus on the saliency improvement in the surgical environment. Our contributions focus on very feature-poor imagery and include:

- Evaluation of existing contrast enhancement methods: Contrast Limited Adaptive Histogram Specification (CLAHS) and Local Laplacian Filtering (LLAP).
- A combined image enhancement method coined Local Laplacian of Specified Histograms (LLSH) for contrast improvements.

To evaluate the image enhancement approaches, we visually analyse the processed images and choose key-point-feature based matching to validate their performance in the context of visual navigation. Multiple sets of arthroscopic images and the underwater seabed inspection Aqualoc [6], [7] are used in the evaluation.

II. BACKGROUND

We first introduce the context of visual saliency and image enhancement techniques and then consider their influence on visual navigation in surgery.

A. VISUAL SALIENCY

Visual saliency of an image indicates the amount of distinct visual features contained in a raw or enhanced image. Visual saliency strongly influences the likelihood of correct feature matching and hence successful pairwise camera measurements and localisation [8]. Therefore, visual saliency can be treated as a quantitative measure of the image registrability and an indicator of how successful subsequent localisation approaches (SLAM, VO etc.) may be. We express the saliency of arthroscopic images in terms of the number of matched Scale-invariant-feature-transform (SIFT) [9] and Speeded up Robust Features (SURF) [10] key-point features. SIFT and SURF are the state-of-the-art feature detection and description algorithms and they have been deployed in arthroscopic environment, with SIFT being the most successful [11]. RANSAC is the state-of-the-art algorithm for outlier rejection commonly used in visual navigation.

B. IMAGE ENHANCEMENT

Image Enhancement techniques can seek to improve the visual appearance of an image and/or to extract the hidden spatial or semantic information useful for machine input [12]. Enhancement techniques can be classified as linear or non-linear operations. Linear operations such as filtering in the spatial and frequency domain are simple and can be easily implemented. They are adequate for many applications but

tend to blur and distort the edges and details within the image [12]. Non-linear operations such as histogram modification, contrast stretching, noise clipping or pseudocoloring can improve image quality by effectively preserving or enhancing texture and edges whilst remaining robust to noise. Contrast enhancement is considered one of the biggest challenges in image processing.

Histogram Equalisation (HE) stretches the image histogram using the cumulative distribution function (CDF) of a given image as the mapping function and is one of the most common contrast enhancement techniques [13]. Due to its global nature, HE often fails to provide local enhancement and is prone to overenhancement when the CDF changes significantly. Adaptive histogram equalisation (AHE) and contrast limited adaptive histogram equalisation (CLAHE) extend HE and are discussed in detail in Section III. CLAHE was first applied as an enhancement technique for low contrast-medical imagery [14] and it is based on the principle of HE. Histogram Specification (HS) techniques map the intensity distributions to the desired shape and are commonly used for contrast enhancement of medical and feature-poor imagery [8], [14], [15]. HE-based techniques provide good contrast enhancement and are robust to non-uniform illumination but they suffer from amplified noise, especially in visually uniform regions.

Laplacian pyramids are a multi-scale image representation widely used for image analysis. However, they are built on spatially invariant Gaussian kernels, which made them believed to be unable to represent and maintain edges well. Anisotropic diffusion, wavelet bases and neighborhood filtering successfully tackled that challenge at the price of additional complexity and often higher computational cost. Paris *et al.* [16] proposed Local Laplacian filtering (LLAP) that utilizes Laplacian pyramids for edge-aware image processing, where small-scale details are differentiated from large-scale edges.

Yun *et al.* [17] combined global HE with Laplacian Pyramids. The role of the pyramid was to introduce local contrast enhancement and prevent overenhancement using just HE. A similar effect can be achieved with CLAHE. Lidong *et al.* [18] performed another fusion and combined CLAHE with discrete wavelet transform (DWT) to avoid contrast overstretching and noise enhancement. To the authors' knowledge there is no fusion of CLAHS and Laplacian Pyramids or Local Laplacian Filters proposed in the literature. In this work we combine the CLAHS' strong contrast enhancement and robustness to non-uniform illumination with LLAP's multi-scale Laplacian pyramid approach and detail enhancement while preserving the edges, to improve image saliency.

C. VISION-BASED SURGICAL NAVIGATION

Using image enhancement to obtain and use state-of-the-art SIFT/SURF features in navigation has mixed results. For already feature rich environments image enhancement brings limited benefit in terms of navigation [19], but for naturally feature-poor environments results suggest image

enhancement is of significant impact [8]. Therefore, it can play a meaningful role in vision-based MIS, impacting identification, matching and tracking of important features that could be used for navigation.

Surgical Vision

Various SLAM approaches using traditional computer vision techniques have been presented over the last two decades for the abdominal environment in MIS [20], [21]. Giannarou *et al.* [22] proposed a probabilistic framework to track affine-invariant anisotropic regions under contrastingly different visual appearances, and proved HS effective for changing lighting conditions. For the work directly related to this article, the literature provides laparoscopic image enhancement approaches but they filter hazing and non-uniform illumination for visually pleasant improvement [23], [24]. The aforementioned methods are currently not able to robustly solve the correspondence search problem and therefore they fail in localisation and mapping task in realistic surgical conditions.

The problem escalates in the arthroscopic environment as the cavity is smaller, visually more uniform and filled with water. The literature provides solutions for the uncertainty estimation in the internal knee joint measurement [25] and real-time joint motion analysis have been proposed, but they do not address the problem of the autonomous visual navigation. To this end, in the context of visual navigation, state-of-art feature detectors and descriptors for monocular knee-arthroscopic images were investigated [11]. Results showed that SIFT features could be best extracted and matched (compared to SURF and others) in knee-arthroscopy, but the study used only sequences containing six unrealistic images not representative of the complexity and length of the procedure. In later work, SIFT features proved to be insufficient for tracking due to the dynamic character (occlusions, blur, glare, deformation) of the environment [26]. To overcome that challenge, sensor-fusion using arthroscopic images, external camera and robot's odometry was employed to provide robust localisation for knee arthroscopy [26], [27]. Essentially, the non-visual sensory information enabled a dense feature mapping thanks to the underpinned localisation improvements.

The image enhancement in arthroscopic surgery has been addressed in [28] where the authors improved the surgical image of subpatellar vertebrae and achieved positive results in the treatment of infrapatellar plica. Histogram modification algorithm and high saturation color mapping have been also successfully deployed on arthroscopic images as image enhancement techniques [29]. The proposed method provides better brightness values while preserving color information. The authors of the two arthroscopic image enhancement works discussed above proved that the image enhancement in arthroscopy can provide clinical benefit. Interestingly, only those two works discussed above have made the effort to enhance arthroscopic images, yet it is reasonable to assume enhancement could improve localisation without the requirement for additional sensors.

As such, we approach the problem from a different perspective than the state-of-the art in arthroscopy [11] and consider the reason for our images to be so difficult to extract and subsequently track salient features. We take into consideration the underwater light degradation and then address the problem of feature-poor images by investigating contrast enhancement techniques. To target the low saliency of arthroscopic images we combine the strong contrast enhancement and illumination robustness properties of CLAHS with the detail enhancement and edge preserving attributes of LLAP. We propose a novel enhancement method based on a combination of the two techniques coined Local Laplacian of Specified Histogram (LLSH). Based on the observations from the enhancement methods we also implement the CLAHS and LLAP techniques on arthroscopic sequences for comparison. Furthermore, we evaluate the three enhancement techniques on an open source underwater Aqualoc dataset [6], [7] for completeness. The Aqualoc dataset has been deployed only in [30] where it was used to validate a positioning solution based on monocular visual odometry. To the best of our knowledge this work is the first attempt to enhance arthroscopic images with such techniques for the purpose of frames registration.

III. MATERIALS AND METHODS

In this section we describe the assumptions made, discuss the two chosen image enhancement methods (CLAHS and LLAP) and propose a novel image enhancement approach (LLSH).

A. ASSUMPTIONS

1) UNDERWATER DEGRADATION

The arthroscopic environment is filled with clear water, suggesting that scattering from water particles will degrade the intensity in the imagery taken in the joint cavity. The amount of degradation depends on the depth the light travels through [31], so for the small cavity making up the knee we expect minimal scattering and negligible effect on the amount of intensity information.

To verify this assumption, we assume that the maximum possible distance from the arthroscope to the knee anatomy is 5 cm and the light attenuation model is given by [31]

$$I(x) = J(x)e^{-\alpha \cdot d} \quad (1)$$

where $I(x)$ and $J(x)$ represent the degraded and actual pixel intensities respectively, α is the medium scattering coefficient ($\alpha = 0.005\text{m}^{-1}$ for pure water), and d is the depth ($d_{\text{max}} = 5\text{cm}$ for arthroscopy). The underwater light intensity degradation at the maximum possible distance is given by $I_{\text{min}}(x)$ and we calculate the ratio between the degraded and non-degraded intensity such that:

$$I_{\text{min}}(x) = J(x)e^{-0.005 \cdot 0.05} \quad (2)$$

$$\frac{I_{\text{min}}(x)}{J(x)} = 0.99975 \quad (3)$$

suggesting that the light degradation due to scattering can reduce pixel intensities in arthroscopic images by up to 0.025% which we consider negligible.

2) OCCLUSIONS AND DEFORMATIONS

We assume that the patients leg is static and there are no tissue deformations. This is a reasonable assumption as the test imagery contains negligible amount of movement. Even though one of our datasets contains occlusions caused by water bubbles, surgical tools and floating tissue, we only address the low saliency of the images and leave the challenge of occlusions for future work.

B. MATERIALS

1) DATASET

We evaluated our approach on two cadaveric datasets from knee arthroscopy and on images from an open source underwater seabed inspection dataset (Aqualoc) [6], [7], with each of them containing 1300-1500 images. The first arthroscopic dataset (dataset 1) was acquired with a Stryker arthroscope and represents a challenging and realistic scenario complete with occlusions (tissue, water bubbles or surgical tools) and feature-poor images (see Remark 1). The second arthroscopic datasets (dataset 2), is the sequence 'H' from [27]. It was acquired using a PointGrey Camera and represents a simple case of an arthroscopic sequence where the camera images are not subject to realistic challenges such as occlusions.

Remark 1: Importantly, in this article we do not address occlusions. They affect tracking, but we retain them in the datasets to provide for a realistic case. We note that occlusions could be managed pre or post image enhancement but do not explore it in this article.

2) IMAGE ENHANCEMENT ALGORITHMS

a: CONTRAST ADAPTIVE HISTOGRAM SPECIFICATION (CLAHS)

HE uses the cumulative distribution function (CDF) of a given image as the mapping function. AHE overcomes the limitations of HE by considering only the intensity distribution within the contextual/local region of each pixel. The idea behind this method is to subdivide the image into equal size regions before equalising the histogram of each region. AHE and HE achieve enhancement by spreading the grey levels of the input histogram over a wider range of the intensity scale [13]. The monotonically non-decreasing mapping function is calculated for each region and maps the local histogram to a desired (in case of HE - uniform) distribution such that [13]

$$s = T(r) = \int_0^r p_r(w)dw \quad (4)$$

where s , T , p_r , r , w are output intensity levels, mapping function, probability density function (PDF), input intensity levels and dummy variable of integration respectively. Note that the mapping function T is nothing else than the CDF.

Since the derivative of the CDF (cumulative histogram) is the PDF (histogram), the slope of the transformation function at any pixel intensity (contrast) is proportional to the height of the histogram at this pixel. Therefore clipping the height of the histogram is equivalent to limiting the slope of the mapping function (PDF). CLAHE builds on AHE in

this way by introducing a contrast enhancement limit (histogram clipping) that prevents noise overenhancement [14]. The clipping limit is specified as a multiple of the average histogram bin contents. Excessive pixels are afterwards uniformly distributed to the remaining bins. Eventually, neighboring regions are combined using bilinear interpolation to eliminate artificially induced boundaries.

HE does not always provide a successful outcome. In some applications it is much more useful to be able to specify the output histogram. Histogram Specification/Matching (HS) is a more general concept where the histogram of the output image region approximately matches the histogram specified prior to the operation [13]. That makes HE a specific case of HS, where the specified histogram is the uniform distribution. We introduced HE first as HS involves equalising the original and specified histogram. Suppose that we are looking for image intensity levels z of a specified density p_z , we can define a mapping function $H(z)$ such that

$$H(z) = \int_0^z p_z(w)dw = s \quad (5)$$

which transforms the desired (z) into the equalised (s) intensity levels. From (4) and (5), it follows that

$$z = H^{-1}(s) = H^{-1}[T(r)] \quad (6)$$

where we know $T(r)$ from (4), and we can find z as long as H is invertible. In the discrete domain invertability is guaranteed if $p_z(w)$ is a valid histogram (i.e. of unit area with no negative values or empty bins).

b: LOCAL LAPLACIAN FILTERING (LLAP)

LLAP is an edge-aware processing technique based on the standard laplacian pyramids that modifies the input image so that the edges (large discontinuities) remain in place with the retained intensity profiles in the neighboring pixels [16]. LLAP uses multiple mapping functions to distinguish the edges from the image details/texture. Consider the coefficients (x_0, y_0, l_0) , where the two former are the coordinates of the image pixel and the latter is the pyramid level. An intermediate image \tilde{I} can be created by applying a monotonic mapping function $r(i)$ to the original full-resolution image. This mapping function depends on the parameter σ_r and the local image value from the Gaussian pyramid $g_0 = G_{l_0}(x_0, y_0)$. The intensity variation threshold σ_r is used to help distinguish edges from details. The pyramid of the intermediate image $L[\tilde{I}]$ is computed and the corresponding coefficient is copied to the output $L[I']$. The mapping function can be represented as follows [16]:

$$r(i) = \begin{cases} r_d(i), & \text{if } |i - g_0| \leq \sigma_r \\ r_e(i), & \text{otherwise} \end{cases} \quad (7)$$

where two mapping functions $r_d(i)$ and $r_e(i)$ process details and edges of the image respectively. The function $r_d(i)$ alters the details/oscillations around the value g_0 such that

$$r_d(i) = g_0 + \text{sign}(i - g_0) \cdot \sigma_r \cdot f_d(\Delta) \cdot (|i - g_0|/\sigma_r) \quad (8)$$

where α represents the detail enhancement parameter, $sign$ is the signum function, and the smoothing function $f_d(\Delta) = \Delta^\alpha$ maps $[0, 1]$ to $[0, 1]$ and controls the edge modification. Similarly, the function $r_e(i)$ modifies the edge amplitude such that

$$r_e(i) = g_0 + sign(i - g_0)f_e(a) \cdot (|i - g_0| - \sigma_r) + \sigma_r \quad (9)$$

where the non-negative smoothing function $f_e(a)$ defined in $[0, \infty]$ and controls the modification of the edge amplitude (clips the edge). To focus on detail enhancement $f_e(a) = 1$.

c: LOCAL LAPLACIAN OF SPECIFIED HISTOGRAM (LLSH)

In describing CLAHS we relied on the notation describing intensity levels. Since LLAP operates on pixels, we use i_z , g_{0_z} , $r_{d_z}(i)$, $r_{e_z}(i)$ as the pixel, the center point and the mapping functions of an image with a previously specified histogram (processed by CLAHS).

In the first phase of our proposed method, we perform CLAHS on the original grey-scale image and obtain an intermediate image with pixel intensities i_z , similar to (5). In the second phase of our proposed method, LLAP is then applied to the histogram-specified image to further enhance the details. The mapping functions r_d and r_e from (8) and (9) can thus be reformulated to consider the CLAHS prior such that

$$r_{d_z}(i) = g_{0_z} + sign(i_z - g_{0_z}) \cdot \sigma_r \cdot f_d(\Delta)(|i_z - g_{0_z}|/\sigma_r) \quad (10)$$

$$r_{e_z}(i) = g_{0_z} + sign(i_z - g_{0_z})f_e \cdot (|i_z - g_{0_z}| - \sigma_r) + \sigma_r \quad (11)$$

C. METHODS

1) IMAGE ENHANCEMENT

First, we evaluate the CLAHS, LLAP and LLSH w.r.t visual assessment and histogram analysis. Before applying each of the enhancement methods, images were smoothed using a Gaussian filter with standard deviation $\sigma = 1$ to reduce the effect of noise. For the CLAHS, we chose the Rayleigh distribution as the desired histogram shape, and clipping limit $c = 0.1$. Changing the desired distribution to a *Uniform* (equalisation) or *Exponential* did not result in noticeable perceptual differences. For LLAP, the amplitude of edges parameter $\sigma_r = 1$ and the detail enhancement parameter $\alpha = 0.5$. The proposed LLSH approach uses the same parameters as for CLAHS and LLAP individually.

2) FEATURE DETECTION AND MATCHING

Second, we evaluate CLAHS, LLAP and LLSH w.r.t feature matching using SURF and SIFT key-point features. For SURF, the strongest feature threshold was set to 1000, and the number of octaves (Gaussian pyramid levels) to 4. For SIFT, the number of octaves was set to 4, the non-edge selection threshold to 3 and descriptor $d1$ was matched to a descriptor $d2$ only if the distance multiplied by a threshold $\eta = 2.7$ was not greater than the distance of $d1$ to all other descriptors. After detecting the successful SIFT and SURF feature

matches, the number of correct matches between frames was evaluated using RANSAC [32] for outlier rejection. Features within an euclidean distance threshold of $T = 0.01$ are matched otherwise determined to be outliers and rejected. We evaluate matching performance w.r.t the number of correctly matched (post-RANSAC) key-point features (CMF) and tracking performance (TP) w.r.t the percentage of successfully registered image frames. Successful registration occurs when the number of inliers allows the fundamental matrix to be estimated. Results are presented in Fig. 2 and Tab. 1.

We also verify the impact of the enhancement techniques on the precision and recall of the feature matching performance. We define precision as the ratio of post-RANSAC matched features to the sum of pre-RANSAC matched features. We define recall as the ratio of post-RANSAC features to the number of all detected features.

IV. RESULTS AND DISCUSSION

A. IMAGE ENHANCEMENT PERFORMANCE

In Fig. 1 we present the enhancement of three examples of arthroscopic images and their histograms. The image histograms are stretched and flattened after applying each of the techniques. In the first test image (row 1), the bottom part of the original image is partially occluded by debris. After applying CLAHS and LLSH we observe that the clear (upper) part of the image is enhanced regardless due to the local nature of the enhancement methodology. The second test image (row 2) is a non-occluded image with a high uniformity of illumination. The third testing image (row 3) does not contain occlusions either, but due to the wide distribution of distances between camera and tissue, the illumination is non-uniform.

CLAHS (column 2) and LLAP (column 3), when applied individually to all images, significantly increase the contrast of the image. CLAHS tackles the non-uniform scene illumination and enhances the image details by flattening the image histogram. LLAP clearly cannot handle the glare caused by the small distance between the light source and the tissue (rows 1 and 3) and does not cope well with the non-uniform illumination. This is a consequence of using a point-wise mapping function in LLAP, which constrains the output pixel intensity range to the intensity limits within the contextual region. That also explains the large number of outermost intensities in the LLAP histogram. LLSH (column 4) first enhances the image details and tackles the non-uniform scene illumination with CLAHS, then further enhances the contrast with LLAP to take advantage of the Laplacian Pyramid based processing. Contrast enhancement introduced by LLSH results in significant glare and blur reduction.

Remark 2: We can also see a post-enhancement artifact of noise effect coming from interpolation of highly uniform regions in CLAHE. A good example is the image in the third row and fourth column - in the lower left corner of the image we can see an artefact that visually resembles noise. A similar phenomenon can be noticed in the CLAHS representation of

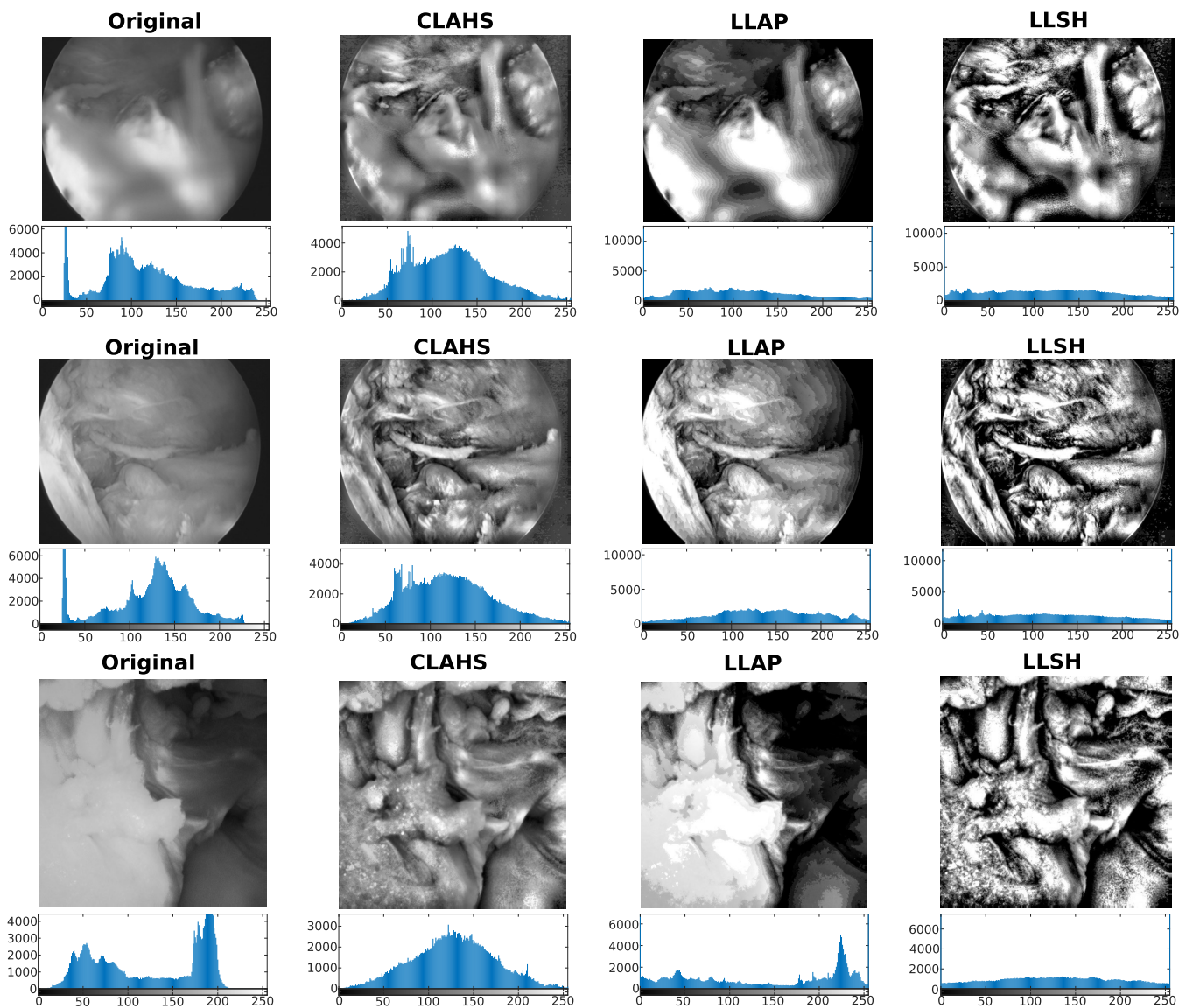


FIGURE 1. Example comparison of image enhancement approaches. The first two rows represent images picked from the first dataset (Row 1 - occluded, Row 2 - clear), and the third row shows a non-uniformly illuminated image from the second dataset. The four columns represent images and their histograms: original and processed by CLAHS, LLAP, LLSH respectively.

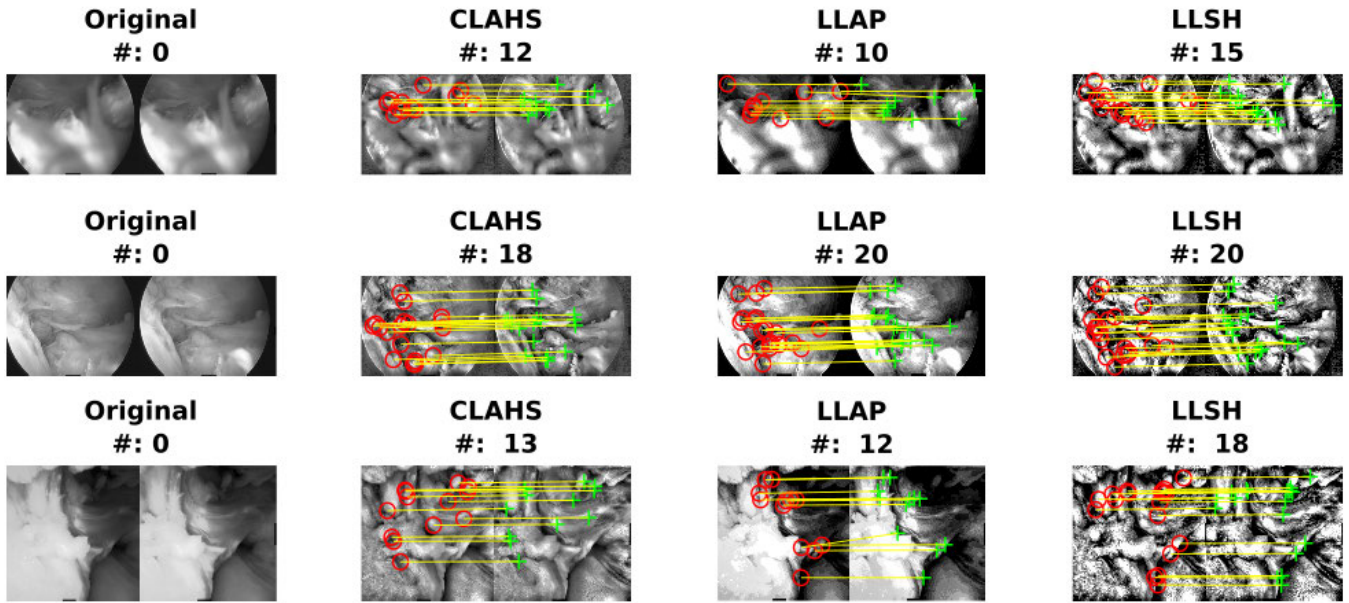
the same image (row 3, column 2). This artifact could be potentially tackled with DWT, similarly to [18].

B. FEATURE DETECTION AND MATCHING PERFORMANCE

Regarding feature detection, CMF evaluation and comparing to key-point detectors/descriptors performance analysis on arthroscopic, un-enhanced images [11] (see Remark 3), our results indicate that CLAHS, LLAP and proposed LLSH provide improvement in the number of detected SURF and SIFT features. The post-enhancement number of detected SURF features is comparable to the number of detected SIFT features, which suggests that the feature robustness in nearest neighbor matching will be the factor determining tracking performance. The analysed enhancement methods lead to an extraordinary SURF CMF performance but also degrade the SIFT CMF performance.

Regarding TP evaluation, our results indicate that CLAHS, LLAP and the proposed LLSH degrade TP of SIFT but significantly improve TP of SURF (87%,93% - LLSH). Importantly, it was shown that SIFT is unreliable in tracking long arthroscopic sequences [26] and also our results show improvement, TP using SIFT remains insufficient on such long datasets. All of the enhancement methods improve SURF from very poor to extraordinary TP, particularly when using the LLSH. In fact, the image enhancement coupled with SURF outperforms the best achieved SIFT outcomes (76%, 85%). The degradation in SIFT TP is outweighed by the improvement in SURF TP. The large gains in SURF TP is a result of the robustness of SURF features over the majority of image frames, even though the mean number of matched SIFT features is significantly higher (Tab. 1). We are also in agreement with [11] that SURF on its own

SURF



SIFT

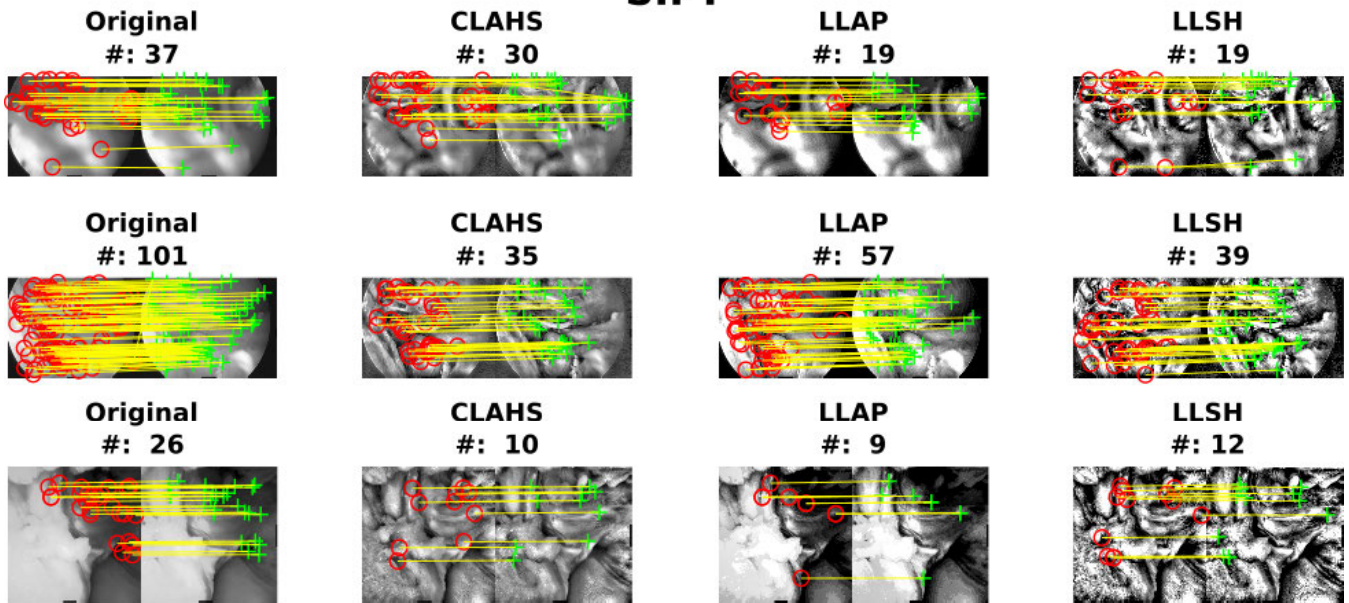


FIGURE 2. Matching Performance of SURF and SIFT key-point features on arthroscopic images. Each row represents a different pair of neighboring images picked from dataset 1 (rows 1,2,4 and 5) and from dataset 2 (rows 3 and 6), and each column represents its original or enhanced form (CLAHS, LLAP, LLSH respectively). We notice significant SURF matching improvement and degraded SIFT matching performance. It is also worth noting that for the occluded images (rows 1,4), after applying image enhancement, SURF features belong only to the non-occluded part of the image.

performs poorly on original arthroscopic sequences. SURF TP after applying LLSH on dataset 1 (93%) is higher than on dataset 2 (87%). This result might be initially surprising since dataset 1 contains occlusions. After image enhancement the occlusions could possibly be tracked together with the background, which would reduce the accuracy of the camera pose estimation.

Overall, our results highlight that for arthroscopic images, employing LLSH for image enhancement can not only improve SURF, but outperform SIFT TP. The implication is

that the localisation problem, hinged on the requirement for good features, may become less challenging and achievable without requiring sensor fusion (i.e vision only). Additionally, using our results in conjunction with those found in [26] provides greater insight regarding how to use SIFT or SURF features and input image enhancement (or lack thereof) for arthroscopic applications. We note, that the quality of the features has yet to be evaluated extensively.

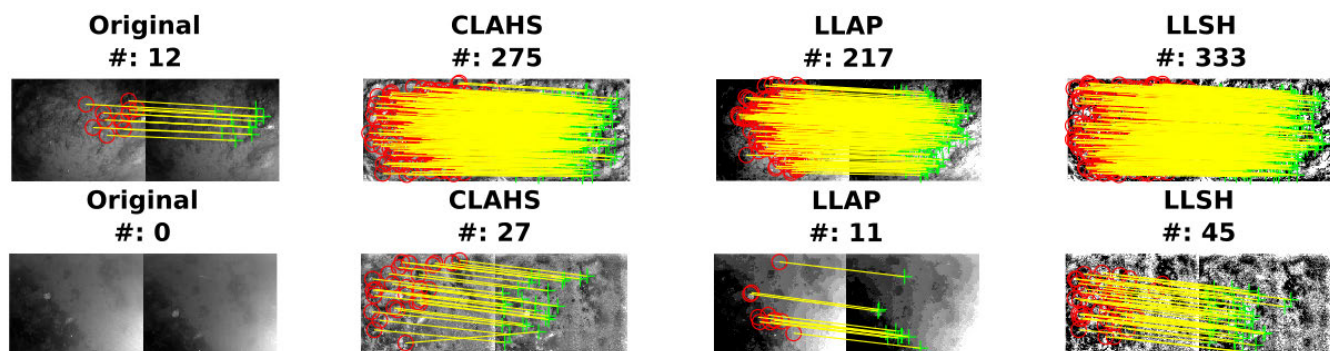
It is important to note that the image enhancement techniques significantly increased the number of detected

TABLE 1. Matching performance of SIFT and SURF for both arthroscopic sequences. %TP represents the percentage of successfully registered (tracked) frame-pairs. Prec. and recall stand for the precision and recall metrics respectively.

The other columns show the median and mean number of inliers and mean number of detected key-points for reference.

Seq.	Image	%TP SURF	median # SURF	mean # SURF	mean det. # SURF	prec. SURF	recall SURF	%TP SIFT	median # SIFT	mean # SIFT	mean det. # SIFT	prec. SIFT	recall SIFT
1	Original	12	0	1	9	0.46	0.01	76	36	77	897	0.95	0.13
	CLAHS	41	0	20	327	0.24	0.08	36	0	60	1461	0.97	0.08
	LLAP	33	0	14	246	0.24	0.1	39	0	49	1020	0.96	0.1
	LLSH	93	21	44	2202	0.35	0.03	40	0	58	1522	0.97	0.07
2	Original	2	0	0.1	12	0.02	0.11	85	115	106	833	0.99	0.09
	CLAHS	82	31	31	397	0.24	0.06	71	91	102	1245	0.97	0.04
	LLAP	85	30	31	328	0.25	0.06	81	90	87	896	0.99	0.05
	LLSH	87	51	56	1798	0.18	0.02	68	79	91	1255	0.99	0.04

SURF



SIFT

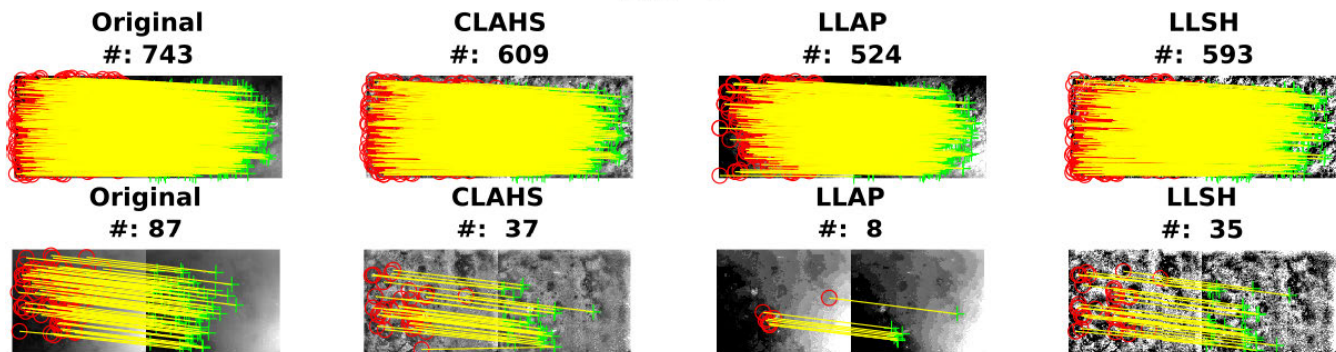


FIGURE 3. Matching Performance of SURF and SIFT key-point features on underwater harbour seabed inspection dataset.

(for SIFT and SURF) and matched (for SURF only) features, but the ratio of the features that survive RANSAC outlier rejection stayed low. Therefore, precision and recall do not show significant improvement but the image enhancement has in fact enabled tracking that may have otherwise not been possible (due to insufficient features). The LLSH enhancement allowed outstanding SURF tracking performance, outperforming SIFT.

Remark 3 Marmol et al. [11] within their small datasets (10 subsets of 6 images), claim performance of 92.4% TP with 110.4 mean # of correct matches for SIFT, and 51.3% TP with 59.9 mean # of correct matches for SURF [11]. In our experiments we use 2 datasets of more than 1300 images that resemble a realistic arthroscopic sequence. Using these datasets, the un-enhanced SIFT TP results are 72% (dataset 1)

and 86% (dataset 2) and SURF TP results are 12% (dataset 1) and 2% (dataset 2).

C. FURTHER RESULTS AND COMPARISON

To help verify the generality of our method beyond arthroscopy, we provide a brief comparison of our method’s performance on an open-source underwater Aqualoc dataset [6], [7] containing images of the seabed taken from a few meters away. The images resemble our arthroscopic sequences in terms of low saliency, but due to the large distance between the camera and seabed light degradation may no longer be negligible (see section III-A). We analyse the performance of CLAHS, LLAP and LLSH in terms of contrast enhancement and key-point feature matching performance on images from the Aqualoc dataset. Similarly to the

TABLE 2. Matching performance of SIFT and SURF for Aqualoc dataset. %TP represents the percentage of successfully registered (tracked) frame-pairs. The other columns show the median and mean number of inliers. *Prec.* and *recall* stand for the precision and recall metrics respectively. *Aq.* stands for Aqualoc dataset.

Seq.	Image	%TP SURF	median # SURF	mean # SURF	mean det. # SURF	prec. SURF	recall SURF	%TP SIFT	median # SIFT	mean # SIFT	mean det. # SIFT	prec. SIFT	recall SIFT
Aq.	Original	3	0	0.1	19	0.01	0.004	100	578	621.5	1324.9	0.99	0.47
	CLAHS	100	350	403.6	1411.4	0.41	0.29	100	853	881.63	1899.7	1.0	0.46
	LLAP	100	169	230.1	762.6	0.42	0.30	100	596	635.3	1461.5	0.99	0.44
	LLSH	100	441	492.95	2617.4	0.35	0.19	100	816	841.38	1950.6	1.0	0.43

performance on arthroscopic datasets, the discussed enhancement methods significantly improve CMF and TP SURF performance and, contrastingly to arthroscopic datasets, also improve CMF SIFT performance and keep SIFT TP on 100% (Tab. 2). For the Aqualoc dataset, similarly to arthroscopic data, precision and recall do not show significant improvement.

The post-enhancement performance improvements achieved on Aqualoc images indicates that CLAHS, LLAP and LLSH provide robust saliency improvement. Hence to achieve better performance on clinical data, other visual challenges need to be addressed.

D. FUTURE WORK AND FURTHER DISCUSSION

The main innovation of this work is the merging two enhancement techniques for low saliency images, like underwater or arthroscopic environments, to explicitly improve visual feature tracking for navigation. The main advantage of the proposed method is the improved SURF feature detection and matching. This increases tracking robustness such that SURF features can replace SIFT in surgical and underwater navigation. This broadens the feasible visual navigation strategies for surgery with potential to improve semantic recognition of the anatomical structures. However, the proposed approach can introduce artificial noise in homogenous regions (Remark 2) and enhance the artifacts/occlusions. This can lead to incorrect or artificial feature matching, the extent to which is largely dependent on the artifact.

This work investigated image enhancement using traditional processing methods. We aim to implement the proposed image enhancement technique in visual odometry tasks, to further evaluate its utility in navigation. Future work could combine additional techniques (deep learning, fuzzy logic etc.) to help address image artifacts (naturally occurring and introduced), and compensate for tissue deformation, allowing navigation in very challenging environments.

V. CONCLUSION

In this work we address the challenge of low contrast and visual saliency, which strongly hinders a successful pairwise camera measurement in arthroscopic images. We use two existing methods: CLAHS and LLAP, and combine them into a novel method LLSH to enhance the contrast of the images and hence increase their visual saliency. We exposed an interesting phenomenon whereby the proposed LLSH image enhancement enabled a large improvement in SURF feature tracking, and surprisingly degraded SIFT performance. We conclude that for arthroscopic images, the use of LLSH provides a significant improvement of tracking performance

using SURF. We also speculate that the enhanced images might also be useful as training data for visual navigation in low-saliency environment using deep learning solutions. Future work is planned to apply the enhanced images for accurate camera pose estimation with consideration of the other visual challenges.

ACKNOWLEDGMENT

Cadaveric experiments were approved by the Australian National Health and Medical Research Council (NHMRC) - Committee no. EC00171. The authors would like to thank Dr. Yu Takuda and Dr. Andres Marmol-Velez for fruitful discussions and help during the cadaveric experiments.

REFERENCES

- [1] A. et al Jaiprakash, "Orthopaedic surgeon attitudes towards current limitations and the potential for robotic and technological innovation in arthroscopic surgery," *J. Orthopaedic Surg.*, vol. 25, no. 1, pp. 1–6, 2017.
- [2] A. Banach, K. Leibrandt, M. Grammatikopoulou, and G.-Z. Yang, "Active constraints for tool-shaft collision avoidance in minimally invasive surgery," in *Proc. IEEE ICRA*, Nov. 2019, pp. 1556–1562.
- [3] J. Engel, J. Stuckler, and D. Cremers, "Large-scale direct SLAM with stereo cameras," in *Proc. IEEE/RSS Int. Conf. Intell. Robot. Syst.*, Dec. 2015, pp. 1935–1942.
- [4] D. Scaramuzza and F. Fraundorfer, "Tutorial: Visual odometry," *IEEE Robot. Autom. Mag.*, vol. 18, no. 4, pp. 80–92, Dec. 2011.
- [5] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 3, pp. 108–117, 2006.
- [6] M. Ferrera, J. Moras, P. Trouv, V. Creuze, and D. Denis, "The Aqualoc dataset: Towards real-time underwater localization from a visual-inertial-pressure acquisition system," 2018, *arXiv:1809.07076*. [Online]. Available: <https://arxiv.org/abs/1809.07076>
- [7] M. Ferrera, V. Creuze, J. Moras, and P. Trouv  Peloux, "AQUALOC: An underwater dataset for visual-inertial-pressure localization," *Int. J. Robot. Res.*, vol. 38, no. 14, pp. 1549–1559, Dec. 2019.
- [8] A. Kim and R. M. Eustice, "Real-time visual SLAM for autonomous underwater hull inspection using visual saliency," *IEEE Trans. Robot.*, vol. 29, no. 3, pp. 719–733, Jun. 2013.
- [9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [10] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. 9th Eur. Conf. Comput. Vis.*, vol. 3951, May 2006, pp. 404–417.
- [11] A. Marmol, T. Peynot, A. Eriksson, A. Jaiprakash, J. Roberts, and R. Crawford, "Evaluation of keypoint detectors and descriptors in arthroscopic images for feature-based matching applications," *IEEE Robot. Autom. Lett.*, vol. 2, no. 4, pp. 2135–2142, Oct. 2017.
- [12] A. C. Bovik, *The Essential Guide to Image Processing*. New York, NY, USA: Academic, 2009.
- [13] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*. Upper Saddle River, NJ, USA: Prentice-Hall, 2004.
- [14] S. Pizer, M. Austin, J. Safrid, D. Perry, and J. B. Zimmerman, "Adaptive histogram equalization for automatic contrast enhancement of medical images," *Appl. Opt. Instrum. Med.*, vol. 626, p. 242, Jun. 1986.
- [15] J. B. Zimmerman, S. M. Pizer, E. V. Staab, J. R. Perry, W. McCartney, and B. C. Brenton, "An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement," *IEEE Trans. Med. Imag.*, vol. 7, no. 4, pp. 304–312, Dec. 1988.

- [16] S. Paris, S. W. Hasinoff, and J. Kautz, "Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid," *Commun. ACM*, vol. 58, no. 3, pp. 81–91, Feb. 2015. [Online]. Available: <http://doi.acm.org/10.1145/2723694>
- [17] S.-H. Yun, J. Kim, and S. Kim, "Image enhancement using a fusion framework of histogram equalization and Laplacian pyramid," *IEEE Trans. Consum. Electron.*, vol. 56, no. 4, pp. 2763–2771, Nov. 2010.
- [18] H. Lidong, Z. Wei, W. Jun, and S. Zebin, "Combination of contrast limited adaptive histogram equalisation and discrete wavelet transform for image enhancement," *IET Image Process.*, vol. 9, no. 10, pp. 908–915, Oct. 2015.
- [19] S. G. Stanciu, D. E. Tranca, and D. Coltuc, "Contrast enhancement influences the detection of gradient based local invariant features and the matching of their descriptors," *J. Vis. Commun. Image Represent.*, vol. 32, pp. 246–256, Oct. 2015.
- [20] D. Stoyanov and G.-Z. Yang, "Soft tissue deformation tracking for robotic assisted minimally invasive surgery," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 254–257.
- [21] N. Mahmoud, Ó. G. Grasa, S. A. Nicolau, C. Doignon, L. Soler, J. Marescaux, and J. M. M. Montiel, "On-patient see-through augmented reality based on visual SLAM," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 1, pp. 1–11, Jan. 2017.
- [22] S. Giannarou, Z. Zhang, and G.-Z. Yang, "Deformable structure from motion by fusing visual and inertial measurement data," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 4816–4821.
- [23] C. Wang, V. Sharma, Y. Fan, F. A. Cheikh, A. Beghdadi, O. J. Elle, and R. Stiefelhagen, "Can image enhancement be beneficial to find smoke images in laparoscopic surgery?" in *Proc. Final Progr. Color Imag. Conf.*, Nov. 2018, pp. 163–168.
- [24] W. Xia, E. C. S. Chen, and T. Peters, "Endoscopic image enhancement with noise suppression," *Healthcare Technol. Lett.*, vol. 5, no. 5, pp. 154–157, Oct. 2018.
- [25] M. Strydom, A. Banach, L. Wu, A. Jaiprakash, R. Crawford, and J. Roberts, "Anatomical joint measurement with application to medical robotics," *IEEE Access*, vol. 8, pp. 118510–118524, 2020.
- [26] A. Marmol, P. Corke, and T. Peynot, "ArthroSLAM: Multi-sensor robust visual localization for minimally invasive orthopedic surgery," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1–5.
- [27] A. Marmol, A. Banach, and T. Peynot, "Dense-ArthroSLAM: Dense intra-articular 3-D reconstruction with robust localization prior for arthroscopy," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 918–925, Apr. 2019.
- [28] X. Tian, J. Wang, D. Du, S. Li, C. Han, G. Zhu, Y. Tan, S. Ma, H. Chen, and M. Lei, "Medical imaging and diagnosis of subpatellar vertebrae based on improved Laplacian image enhancement algorithm," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 105082, doi: [10.1016/j.cmpb.2019.105082](https://doi.org/10.1016/j.cmpb.2019.105082).
- [29] C. Kim, J. Han, G. Yoon, D. Kim, and S. Yang, "Novel framework for knee arthroscopic image enhancement," *J. Med. Imag. Health Informat.*, vol. 10, no. 6, pp. 1459–1465, Jun. 2020.
- [30] P. Dabove, V. Di Pietra, and M. Piras, "Monocular visual odometry with unmanned underwater vehicle using low cost sensors," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Apr. 2020, pp. 810–816.
- [31] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean. Eng.*, vol. 15, no. 2, pp. 101–111, Apr. 1990.
- [32] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, Jun. 1981.



MARIO STRYDOM received the bachelor's degree in electronics and the master's degree in business. He is currently pursuing the degree with the Australian Centre for Robotic Vision, Queensland University of Technology. He has authored or coauthored in the research area of image segmentation, monocular measurement, and uncertainty of the knee joint. He has a provisional approved patent for a robotic leg manipulator from the Australian patent office. He has 25 years of industry experience in the field of automation, electronics engineering, and information technology. His researches focus on the field of computer vision and robot kinematics applied for medical robotics.



ANJALI JAIPRAKASH received the Med.Sc. degree. She is currently a Life Sciences Scientist and an Advance QLD Research Fellow of medical robotics with the Australian Centre for Robotic Vision, Queensland University of Technology. She works at the intersection of medicine, engineering, and design, developing medical devices for diagnosis and surgery, including the patented light field retinal diagnostic systems and vision-based robotic leg manipulation system. She has experience in the field of orthopaedic research, optics, and design. She has extensive research experience in the hospital and clinical setting and the ethical conduct of research in compliance with the Australian Code for the Responsible Conduct of Research.



GUSTAVO CARNEIRO received the Ph.D. degree in computer science from the University of Toronto in 2004. He was with Siemens Corporate Research, The University of British Columbia, and the University of California at San Diego. He is currently a Professor with the School of Computer Science, The University of Adelaide, and the Director of medical machine learning with the Australian Institute of Machine Learning. His primary research interests are in the fields of computer vision, medical image analysis, and machine learning.



ARTUR BANACH received the B.Eng. degree in automatic control and robotics from the Poznan University of Technology, Poland, in 2016, the master's degree (research) in medical robotics and image-guided intervention from the Hamlyn Centre, Imperial College London, in 2017. He is currently pursuing the Ph.D. degree in surgical robotics with the Queensland University of Technology. He introduced active constraints for tool-shaft collision avoidance in minimally invasive surgery on the Da Vinci Surgical System at the Hamlyn Centre, Imperial College London. His interest covers innovating the field of surgery by looking for solutions to sublimely reduce patient suffering and improve quality of life.



CAMERON BROWN is currently the Director of the Medical Engineering Research Facility and the Head of the Photonics and Mechanics of Biomedical Materials Laboratory, Queensland University of Technology. His research interest is in the structure-property-function relationships in biomedical materials and systems and the development of frontier technologies for medicine.



ROSS CRAWFORD received the Ph.D. degree from Oxford University. He is currently a Professor of orthopaedic research with Queensland University of Technology (QUT) and undertakes private clinical practice at the Prince Charles and Holy Spirit Hospitals. He assists with cadaver surgery experiments with the QUT Medical Engineering Research Facility at the Prince Charles campus as an Expert Surgeon and brings significant knowledge of knee arthroscopy and the use of medical robotics to this research. He has mentored over 30 Ph.D. and M.Phil. students to the completion of their degrees. He has a wealth of experience in teaching and leading researchers at all levels. He has authored or coauthored more than 200 articles. He is currently a member of numerous medical committees.



AARON MCFADYEN (Member, IEEE) received the B.Eng. degree in aerospace avionics and the Ph.D. degree in robotics from the Queensland University of Technology (QUT). He is currently a Lecturer with the Science and Engineering Faculty, QUT, Australia. He has professional pilot qualifications (CASA) and engineering experience (Emirates, CAE). He was a recipient of multiple fellowship awards for his research in autonomous systems, including pioneering work on vision-based control (visual servoing) and complex system modeling (air traffic). Coupling these research strengths, he currently leads multiple industry-backed research projects on unmanned aircraft systems integration.

• • •