IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 19, 2018, accepted July 1, 2018, date of publication July 13, 2018, date of current version August 7, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2855264

An Evaluation on the Robustness of Five Popular Keypoint Descriptors to Image Modifications Specific to Laser Scanning Microscopy

DEVRİM ÜNAY[1] AND STEFAN G. STANCIU[2]
[1]Department of Biomedical Engineering, İzmir University of Economics, 35330 İzmir, Turkey
[2]Center for Microscopy-Microanalysis and Information Processing, University Politehnica of Bucharest, 060042 Bucharest, Romania

Corresponding authors: Devrim Ünay (devrim.unay@ieu.edu.tr) and Stefan G. Stanciu (stefan.stanciu@cmmip-upb.org)

This work was supported in part by a Short Term Scientific Mission of the EU COST Action CA15124 NEUBIAS, in part by the Scientific and Technological Research Council of Turkey under Grant 113E603, and in part by the Romanian Executive Agency for Higher Education, Research, Development and Innovation Funding under Grant PN-III-P2-2.1-PED-2016-1252 and Grant PN-III-P2-2.1-PED-2016-0450.

ABSTRACT Laser scanning microscopy (LSM) techniques are of paramount importance at this time for key domains such as biology, medicine, or materials science. Computer vision methods are instrumental for boosting the potential of LSM, providing reliable results for important tasks, such as image segmentation, registration, classification, or retrieval in a fraction of the time that a human expert would require (at similar or even higher accuracy levels). Image keypoint extraction and description represent essential building blocks of modern computer vision approaches, and the development of such techniques has gained massive interest over the past couple of decades. In this paper, we compare side-by-side five popular keypoint description techniques, scale invariant feature transform (SIFT), speeded-up robust features (SURF), binary robust invariant scalable keypoints (BRISK), fast retina keypoint (FREAK) and BLOCK, with respect to their capacity to represent in a reproducible manner image regions contained in LSM data sets acquired under different acquisition conditions. We evaluate this capacity in terms of descriptor matching performance, using data sets acquired in a principled manner and a thorough Precision-Recall analysis. We identify which of the five evaluated techniques is most robust to specific LSM image modifications associated to the laser beam power, photomultiplier gain, or pixel dwell, and show that certain pre-processing steps have the potential to enhance keypoint matching.

INDEX TERMS Keypoint descriptors, laser scanning microscopy, scale invariant feature transform (SIFT), speeded-up robust features (SURF), binary robust invariant scalable keypoints (BRISK), fast retina keypoint (FREAK), BLOCK.

I. INTRODUCTION

The Laser Scanning Microscopy (LSM) family incorporates a series of investigation techniques that play in current days essential roles in multiple fields of science. They rely on illuminating the specimen with a focused scanning laser beam (point scanning) and constructing an image pixel-by-pixel by exploiting various optical phenomena connected to either intrinsic optical and chemical properties of the sample or to those of exogenous agents, such as fluorescent dyes. Confocal Laser Scanning Microscopy (CLSM), probably the most popular LSM technique, took the realms of biology and medicine by storm soon upon its first physical implementation as a result of its superb capabilities to produce sharp optical sections of an investigated specimen. This is achieved by rejecting optical signals (reflectance or fluorescence) generated above and below the focal plane by means of a pinhole positioned in front of the detector [1], [2]. CLSM represents at this time maybe the most common choice for studying fluorescent cells [3], [4], and over the past years it has been demonstrated as well as a very valuable tool for investigating various tissues in-vivo [5]–[7]. Multiphoton microscopy (MPM) is likely to become in the near-future the default tool for probing tissue morphology, functionality and biochemical composition in both ex- and in-vivo assays; MPM techniques rely on optical configurations similar to CLSM - except for the fact that a femtosecond laser source is used for excitation instead of a continuous wave laser. Two-Photon Excitation Fluorescence Microscopy (TPEF) [8] or Second Harmonic Generation Microscopy (SHG) [9] can be used for probing optical

properties of tissues with deep implications for resolving important anatomical and physiological aspects that can be linked to disease predisposition or progression [10]–[13]. Coherent anti-Stokes Raman scattering (CARS) microscopy, another prominent LSM technique, can probe the vibrational signatures of molecules, which is important for understanding how the biochemical composition of tissues is modified across the progression of various diseases [14], and for identifying pathologies in very early stages. Besides the techniques that operate at micro-scale, the LSM family includes as well imaging techniques that can characterize biological specimens, and also materials, at optical resolutions that lie beyond the diffraction barrier, such as Stimulated Emission-Depletion Microscopy [15], Saturated Absorption Competition Microscopy [16] or Saturated Transient Absorption Microscopy [17], [18]. Without doubt, not far from now we will witness new LSM variants and strategies that will further contribute to expanding our knowledge on the structure and behavior of biological items and materials.

The power and potential of LSM techniques can be considerably amplified by intertwining their superb imaging capabilities with advanced computer vision methods, the field of science that combines artificial intelligence and computer science in the purpose of providing computers the ability to understand the content of digital images as a (trained) human being would, or even better [19], [20]. This has already been demonstrated in various experiments aimed at object segmentation and identification [21], object tracking [22], [23], image registration [24], image classification [25] and others. However, the combination of LSM and computer vision is still at an early stage, and progress in this area is achieved in a slower pace than desired by microscopy end-users. This relates to various reasons such as a not very well established community of bioimage computer vision scientists [26], and the current low speed by which latest hour computer vision techniques are transferred to the realm of microscopy. A reason behind this latter situation refers to the fact that computer vision scientists typically develop their algorithms taking into account specific properties and content-types found in natural images, and further on many of these algorithms are transferred to the realm of microscopy imaging in their original version. Because microscopy images differ from natural images in terms of image content specifics, this approach usually turns out to be less effective than expected. A solution in this regard might consist in replacing the direct transfer to microscopy applications of computer vision algorithms developed for natural images with the development of microscopy oriented algorithms which are designed from the very beginning by taking into account specific properties and content types. While very promising, this approach is quite difficult to implement as computer vision scientists usually have limited access to microscopy images. Another solution, which is more realistic (but also very laborious), consists in a thorough evaluation of computer vision algorithms developed for natural images with respect to microscopy specific aspects, before they are transferred and implemented in applications addressing this field. This essential stage can enable their optimization and hence consistent performance enhancements.

The work presented here is aligned to this latter philosophy, and deals with a side by side evaluation of five popular feature descriptors with respect to their robustness to LSM specific image modifications. The importance of this subject refers to the diverse roles played by keypoint descriptors in computer vision. Keypoint descriptors can be used to represent either keypoints at fixed locations [27] or keypoints identified using various detection strategies [28], and aim to represent an image patch in a reproducible manner. This reproducibility refers to a good similarity in terms of Euclidean distance (or other distance metrics) between keypoint descriptors extracted from corresponding locations in images collected under different acquisition conditions (e.g. scale, perspective, contrast, brightness, blur, etc). Because of their capacity to represent images in a very simple and concise manner, keypoint descriptors are now currently used in a wide variety of computer vision applications such as image retrieval [29], recognition [30], wide baseline matching [31], image registration [32], [33], image based localization [34] or image classification [35]. In these applications, the keypoint descriptors of one image are matched against the descriptors of others by means of various strategies, e.g. nearest-neighbor searches. The matched keypoint descriptors can subsequently be used to indicate the presence of a specific object, to vote for a particular image, to establish correspondences for epipolar geometry estimation, or to classify an image as belonging to a specific class. To date, keypoint descriptors have also been successfully used in various microscopy oriented computer vision applications [36]–[41], but earlier studies highlighted the fact that their robustness to typical modifications associated to microscopy specific acquisition parameters should be considered, as keypoint descriptors are sensitive in this regard [42], [43].

With this study we attempt to provide additional insights over the robustness of five consecrated keypoint description techniques: scale invariant feature transform (SIFT) [44], speeded-up robust features (SURF) [45], binary robust invariant scalable keypoints (BRISK) [46], fast retina keypoint (FREAK) [47] and BLOCK [48] to image modifications specific to LSM. For this, we evaluate and compare their matching performance in terms of Precision-Recall when extracted from fixed-grid locations in LSM images collected in a principled manner under various laser beam power, photomultiplier gain and pixel dwell settings.

II. BRIEF SURVEY OF THE FIVE KEYPOINT DESCRIPTORS USED IN OUR EXPERIMENT

According to the taxonomy presented in [49], feature descriptors can be categorized as local binary descriptors (e.g. LBP, BRISK, FREAK), spectra descriptors (e.g. SIFT, SURF), basis space descriptors (e.g. Fourier), and shape descriptors (e.g. MSER, shape context). In this study we selected four popular local binary and spectra descriptors (BRISK,

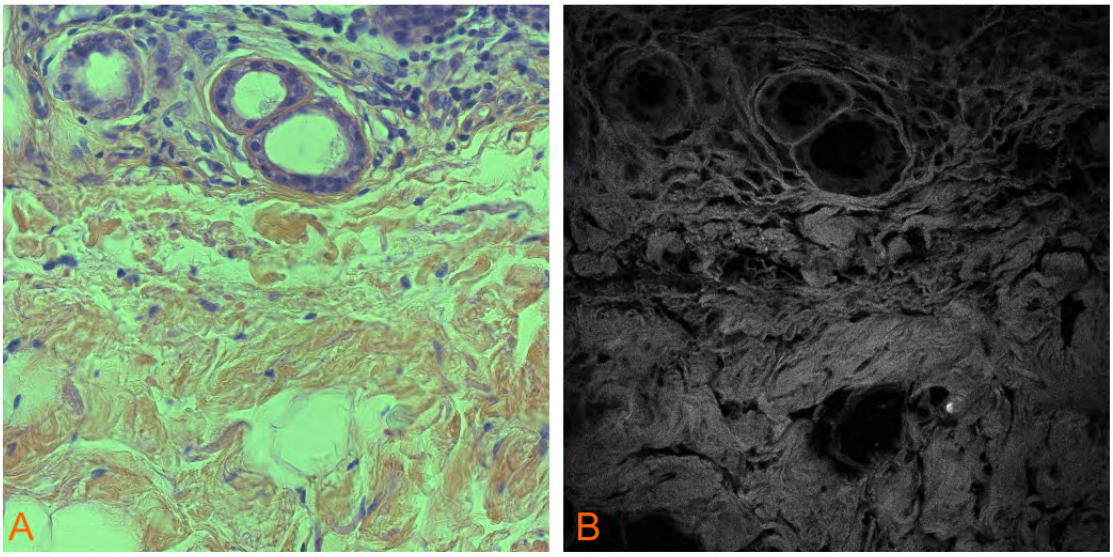

FIGURE 1. Images of a transverse section on human epithelial tissue collected with A) Brightfield microscopy, and B) Confocal Laser Scanning Microscopy.

FREAK, SIFT and SURF), and one elementary alternative (BLOCK) as reference.

In SIFT [44], the most popular keypoint description technique, the descriptor takes the form of a 128-element vector which is practically a histogram representation that combines local gradient orientations and magnitudes from a certain neighborhood around a keypoint. More precisely, the descriptor is in fact a 3D histogram of gradient locations and orientations, where location is quantized into a 4x4 location grid and the gradient angle is quantized into 8 orientations, one for each of the cardinal directions. SIFT was designed to be invariant to brightness, contrast, rotation, scale, affine transformation, and noise.

The SURF descriptor [45] is built in a similar fashion to SIFT, but it exceeds SIFT's speed by exploiting integral images and Haar wavelets concepts. The SURF descriptor is assembled by considering a grid around the keypoint and dividing each grid cell into sub grids. At each sub-grid cell the gradient is calculated and a histogram with angle corresponding bins is constructed. The gradient counts contributing to the histogram are increased by the magnitude of the gradient, all weighted by a Gaussian corresponding to the scale of the keypoint. The SURF descriptor, with a default size of 64 elements, is designed to ensure invariance to illumination, rotation, scale, and noise.

BRISK [46] is a binary descriptor invariant to brightness, contrast, rotation and scale. It computes three sets of pixel-pair binary comparisons over Gaussian smoothed data at multiple scales, where short pairs are aligned according to the gradient orientation estimated from long pairs.

FREAK [47] is a another binary descriptor invariant to brightness, contrast, rotation, scale, viewpoint, and blur, which consists of pixel-pair binary comparisons computed over Gaussian smoothed data at multiple scales.

The simplest descriptor of the five evaluated alternatives is BLOCK [48], which extracts small neighborhood blocks of the actual image centered at the keypoint locations.

III. THE BENCHMARKING FRAMEWORK

A. IMAGE ACQUISITION AND CONTENT

The image sets used as support in our experiment have been collected on epithelial tissue samples stained by Hematoxylin & Eosin (HE) with a Nikon C2+ Confocal Laser Scanning Microscope. Images of 512 x 512 pixels were acquired with a 10x objective (providing a Field of View of 1mm x 1mm), under a fixed pinhole aperture of 1.2 Airy Units. To generate the image sets, laser beam power (LB), photomultiplier gain (PMT) and pixel dwell (PD) have been modified in a principled manner using the proprietary C2+ acquisition software package NIS Elements, as discussed further in this article. For excitation, a 488nm laser beam was used. Under this illumination wavelength, the collected fluorescent signals were mainly generated by Eosin, but also by several endogenous fluorophores present in the human skin such as melanin, flavins, carotinoids or bilirubin; Eosin is not typically regarded as a fluorophore, but exhibits nonetheless high fluorescence emission. The absorption and emission maxima of Eosin are 527 and 550 nm, respectively, with both peaks laying in the green light range; however, highly satisfactory emission intensities have been observed in previous experiments under excitation with blue light (450-490 nm) [50], [51]. Fluorescence CLSM images collected on H&E stained samples using a 488nm laser beam recapitulate thus part of the features present in brightfield microscopy images, which are traditionally used in H&E histopathology workflows, Fig. 1. At the same time, fluorescence CLSM images collected on tissue with 488nm recapitulate as well part of the content of MPM images collected on tissues in vivo at typical excitation ranges.

The results of our experiment are thus relevant also for MPM imaging, where LB, PMT or PD are also routinely modified in typical imaging sessions.

The scene that was imaged for generating the test data sets was selected based on the grounds that it contains areas with various content densities. This allows a balanced evaluation of the five considered keypoint description techniques, with respect of their capacity to deal with cluttered image regions or areas with sparse content. The imaged tissue region contains a series of cues that are typically taken into consideration in the frame of histopathology exams aimed at epithelial tissue characterization. In the top-region of the brightfield image, three ''doughnut'' shaped structures can be observed, which correspond to eccrine sweat glands. The walls of these channels are made up of simple cuboidal epithelium without cellular atypia. Next to these, multiple inflamatory elements are present (mainly chronical). The upper part of the imaged scene includes also some elements presenting a form of atypia, represented by cells with large, hyperchromatic, nuclei and eosinophilic cytoplasm. The lower part of the imaged scene is mostly occupied by conjunctive tissue, which incorporates (acidophilic) collagen fibers and adipocytes. Several endothelial cells of capillaries (longitudinally sectioned) can be observed as well. In the CLSM image, the stroma (connective tissue consisting mainly in collagen fibers) is much better evidentiated compared to the case of the brightfield image. The collagen fibers in the middle of the image can be well differentiated which allows a better characterization of the stroma. The basal membrane of the sweat gland channels, which are barely noticeable in the brightfield image, are here clearly visible. Inflammatory elements (lymphocytes and hystiocytes) cannot be observed, and neither the nuclei of the cells, as these are labeled with Hematoxylyin which is not fluorescent.

1) DATA AVAILABILITY

The test datasets generated and analysed during the current study are fully available upon written request to the authors.

2) USE OF HUMAN TISSUE SAMPLES

The use of human tissue samples in this study was carried out in full compliance with the guidelines provided by the Ethics Commitees of University Politehnica of Bucharest and Izmir University of Economics. These samples have been supplied to the authors by a partner group at the Carol Davila University of Medicine and Pharmacy in Bucharest (UMF CD) after having been previously collected from patients for routine histopathologic examinations, and then de-identified. Access to the samples was provided in line with UMF CD's ocurring policies referring to informed patient consent.

B. GENERATION OF THE TEST DATA SETS

The purpose of our experiment was to compare side by side the robustness of SIFT, SURF, BRISK, FREAK and BLOCK, five popular keypoint descriptors, to image modifications specific to LSM. The image modifications considered in our experiment are those introduced by changes of the laser beam power, photomultiplier gain and pixel dwell time. These three acquisition parameters are routinely adjusted during LSM imaging sessions in order to increase the quality and overall aspect of the scene's 2D and 3D representations. Increasing the LB results in an increase of the intensity of the collected signal, but it can translate to photodamage and phototoxicity. Low LB levels can be compensated by high PMT gain settings, but using high PMT gain settings can lead to noise being introduced in the collected image. Longer PD times can be used to collect a sufficient amount of photons under low LB and PMT gain settings, but this can result of course in similar disadvantages as in the case of high LB, while also raising the image acquisition time which can be problematic in the case of scenes whose content rapidly evolves with time. The LB, PMT and PD need to be thus carefully tuned together so that good quality LSM images are collected, while avoiding sample alteration. Related aspects referring to the robustness of SIFT and SURF have been previously studied in [42] and [43]. It is important to mention that SIFT and SURF incorporate both a keypoint detection method and a keypoint description method, and these previous studies documented an overall effect of LSM image modifications over the SIFT and SURF algorithms (convoluted effect over both SIFT/SURF keypoint detection and description stages). In this current work, we address exclusively the problem of keypoint description, the locations of the keypoints (a.k.a interest points) being selected for all evaluated techniques according to a fixed grid, as explained in the next section. This approach allows an exact side-by-side assessment of the effects of specific LSM image modifications to the robustness of the five evaluated keypoint descriptors.

As discussed in the introductory part, keypoint descriptors (which typically take the form of one-dimensional arrays) aim to represent image patches in a reproducible manner. In the case of LSM, a keypoint description technique of high performance should represent the same image patch collected at different acquisition settings in a similar way. We compare thus SIFT, SURF, BRISK, FREAK and BLOCK with respect to their ability to achieve this for scenes that are acquired at different LB, PMT and PD settings. For this, we construct a series of test image sets in grid-like manner similar to the strategy introduced in [42]. For generating a test subset we kept all acquisition parameters fixed, except for two that were modified in a consecutive manner. For example, in the case of the test data sets generated by gradually modifying the LB and the PMT gain, under a specific LB setting we have collected images at all the considered PMT gain levels. In the next step the LB setting was modified (increased) and then CLSM images were collected again by sweeping the same PMT gain range. This led to the acquisition of a test image set with a grid like architecture, Fig. 2. For each position in the grid we acquired a z-stack of 11 optical sections. We performed the same routine for the PD vs. PMT pair. This strategy has led to the generation of four test data sets:

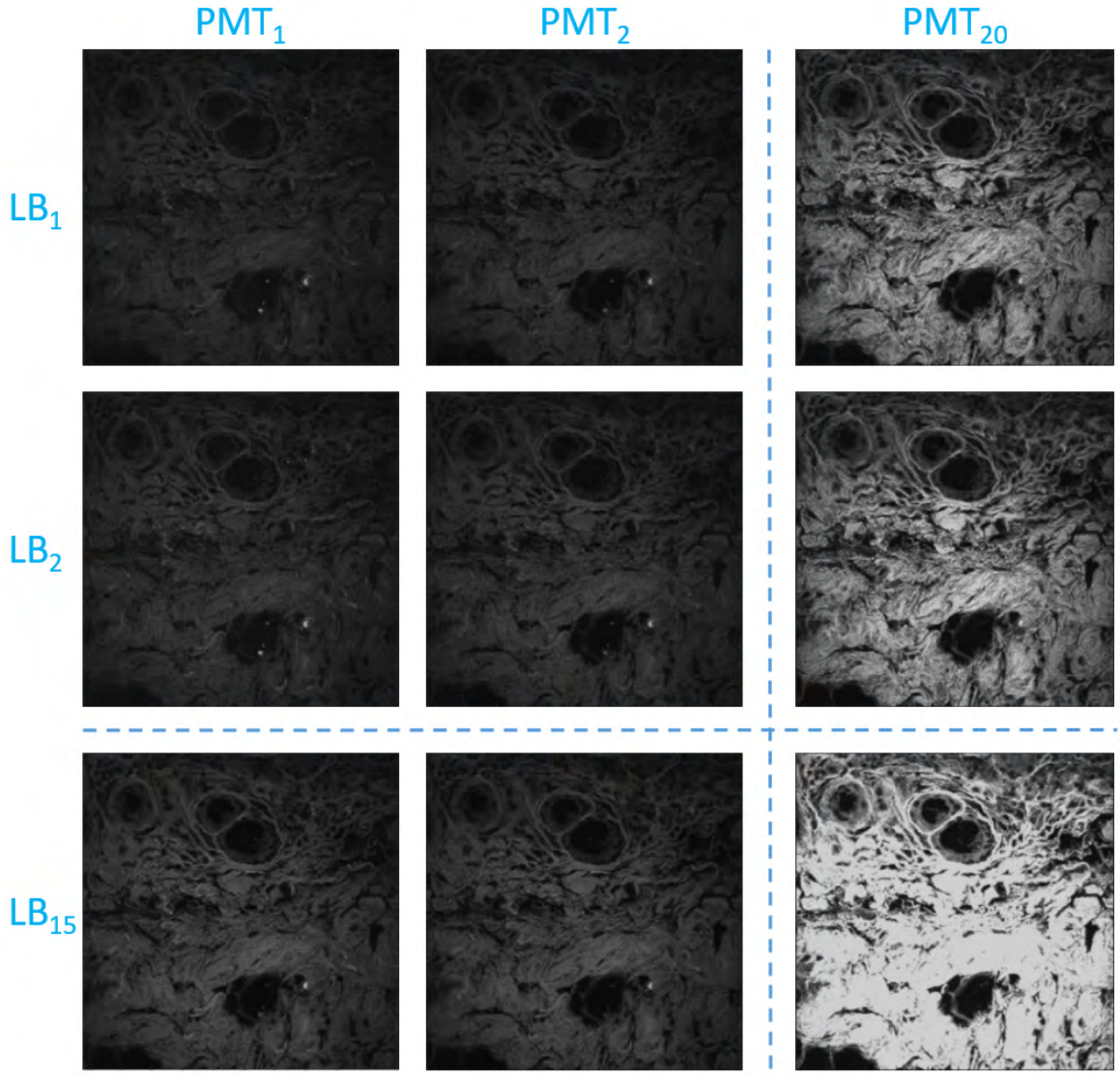

FIGURE 2. Architecture of the test data sets. In this example instances of TD#1a are displayed, a data set generated by gradually adjusting the LB and PMT.

1) TD#1a (LB vs. PMT, central image of the z-stack)
2) TD#1b (LB vs. PMT, maximum intensity projection of the z-stack)
3) TD#2a (PD vs. PMT, central image of the z-stack)
4) TD#2b (PD vs. PMT, maximum intensity projection of the z-stack)

For the generation of the TD#1 data sets we used 15 LB values and 20 PMT values, leading to a total of 300 images for each of the TD#1a and TD#1b sets. To generate the TD#2 data sets, we used 5 PD values (all the available values in the Nikon C2+ acquisition software) and 20 PMT values, which resulted in 100 images for each of the TD#2a and TD#2b sets.

C. EVALUATION OF THE KEYPOINT DESCRIPTORS

In this study the robustness of keypoint descriptors was evaluated in a principled manner by

1) Extraction of SIFT, SURF, BRISK, FREAK and BLOCK descriptors in all available images at fixed locations according to a grid,
2) Matching the descriptors of specific image pairs,
3) Quantification of the descriptor matching performance by Precision-Recall curves,
4) Generation of Area Under Precision Recall Curve (AUC PR) scores for the considered imaging parameter.

In the first step of this approach, dense image locations are generated by taking every i^{th} pixel of the image in 2D, where $i = 5$; SIFT, SURF, FREAK, BRISK, and BLOCK descriptors are extracted from each location. Descriptor matching is realized by computing Minkowski distances of order $p = 1$ between all the possible pairs of keypoint descriptors (same type) extracted from two images which are included in

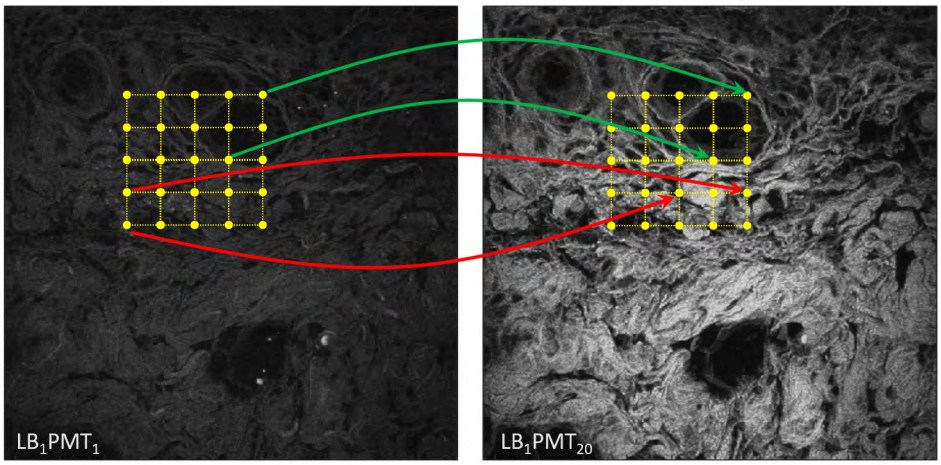

FIGURE 3. Illustration of correct (green) and incorrect (red) descriptor based keypoint matches between two images acquired at the same LB (LB_1) but with different PMT settings (PMT_1 and PMT_{20}).

the comparison. Two keypoint descriptors are considered to match based on a nearest neighbor ratio test over the distances computed between the available descriptor pairs. If for a considered pair of descriptors the nearest neighbor distance ratio is positioned below a threshold, these descriptors (and consequently, the keypoints they describe) are considered to match.

Say, two images acquired at the same LB (LB_1) but with different PMT settings (PMT_1 and PMT_5) are to be compared, namely the LB_1PMT_1 vs. LB_1PMT_5 images. Their keypoint descriptors are respectively represented by $D_{1,1}$ and $D_{1,5}$ matrices of size $m \times n \times k$, where m and n are the number of dense keypoint locations in horizontal and vertical directions and k is the descriptor length. We compute the Minkowski distance of order $p = 1$ between all possible pairs of descriptors of the two images and form the distance matrix as $S_{m,n}^{\Delta PMT=4} = (\sum_{j=1}^{k} |D_{1,1}(m,n,j) - D_{1,5}(m,n,j)|^p)^{1/p}$. Then for each descriptor location we calculate the nearest neighbor distance ratio as $\frac{\text{distance of nearest neighbor}}{\text{distance of 2nd nearest neighbor}}$. If the xy coordinates of a descriptor in the 1^{st} image and its nearest neighbor in the 2^{nd} image are the same, and the ratio is below a predefined threshold value (α), we assume a correct match (true positive). Conversely, a match established between two descriptors extracted from different xy coordinates, represents an incorrect match (false positive), Fig. 3.

We quantify the matching performance between the two images by means of *Precision* (ratio of the number of correct matches to the total number of matches that qualify from the nearest neighbor distance ratio test) and *Recall* (ratio of the number of correct matches to the total number of dense keypoints) pairs obtained by varying the threshold (α), and by computing the AUC PR via trapezoidal approximation. The threshold values are chosen as 20 evenly distributed points between upper and lower limits, which are calculated from the matrix of all image pairs compared across separate and joint variation of the considered imaging parameters.

TABLE 1. AUC PR scores for TD#1a.

	Change in laser beam power				Change in photomultiplier gain			
	Gauss smoothed		Not smoothed		Gauss smoothed		Not smoothed	
Method	mean	std	mean	std	mean	std	mean	std
SIFT	0.9934	0.0087	0.9322	0.1079	0.9631	0.0360	0.8369	0.1750
SURF	0.8104	0.0915	0.5917	0.1445	0.4936	0.1861	0.3536	0.1933
BRISK	0.8777	0.0898	0.8741	0.0852	0.4998	0.1880	0.5389	0.1729
FREAK	0.5968	0.1306	0.5020	0.1221	0.2019	0.1613	0.1957	0.1501
BLOCK	0.0562	0.1642	0.1789	0.2439	0.0296	0.0670	0.1459	0.2275

Fig. 4 displays the architecture of the evaluation system for the scenario of variation in PMT. Similar evaluation frameworks can be used to investigate the robustness of keypoint descriptors to additional LSM imaging parameters than those addressed in our experiment (e.g. sampling size, offset, polarization angle, etc) and also to acquisition parameters specific to other families of microscopy techniques (e.g. scanning probe microscopy, electron microscopy, etc.).

IV. RESULTS

A. ROBUSTNESS OF SIFT, SURF, BRISK, FREAK AND BLOCK DESCRIPTORS EXTRACTED FROM THE TD#1a AND TD#2a TEST DATA SETS

Average AUC PR scores for individual images of TD#1a and TD#2a sets are shown in Table 1 and Table 2 respectively, while the effect of imaging parameters' modifications on the AUC PR scores are graphically displayed in Fig. 5 and Fig. 6. Also in Table 1 and Table 2, we provide the results obtained for two additional test data sets that represent smoothed versions of TD#1a and TD#2a. To generate these, every image in the TD#1a and TD#2a has been processed by 2D Gaussian smoothing (sigma = 2.6). The reason behind our intention to investigate the effect of smoothing consists in a previous study that provided evidence that contrast enhancement detriments feature matching [52]. Hence, in our opinion smoothing, which represents the opposite of contrast enhancement (while also being of benefit for noise reduction), should boost up the matching performance.

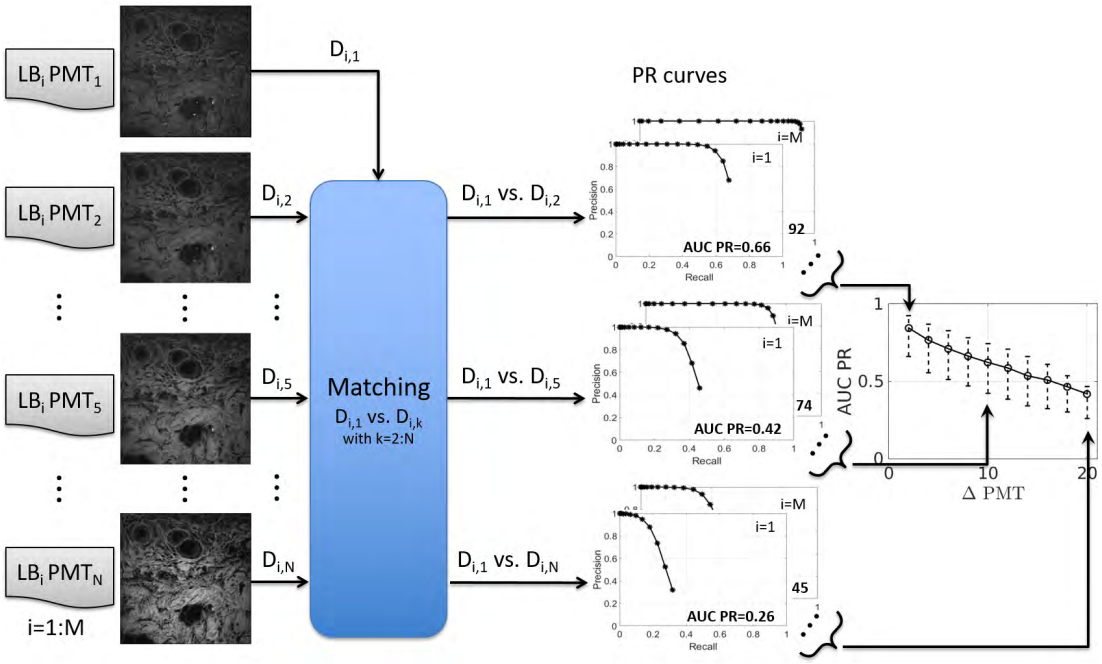

FIGURE 4. Architecture of the evaluation system for the scenario of variation in PMT. For each LB value the matching procedure is performed between keypoint descriptors extracted from dense locations in images collected at increasing PMT levels. The effect of PMT variation over descriptor matching is assessed using AUC PR scores. High AUC PR reflects good matching performance, and vice-versa. In the AUC PR vs ΔPMT graph, a line is plotted along the average AUC PR scores per PMT level. To generate the AUC PR vs. ΔLB results, we perform the same routine/procedure, by considering LB#PMT# pairs, from the TD#1a and TD#1b data sets.

TABLE 2. AUC PR scores for TD#2a.

	Change in pixel dwell				Change in photomultiplier gain			
	Gauss smoothed		Not smoothed		Gauss smoothed		Not smoothed	
Method	mean	std	mean	std	mean	std	mean	std
SIFT	0.9948	0.0046	0.7629	0.2112	0.9751	0.0366	0.9154	0.1199
SURF	0.7673	0.0378	0.4181	0.0912	0.8694	0.0811	0.6560	0.2261
BRISK	0.8433	0.0352	0.8281	0.0323	0.9113	0.0411	0.9154	0.0603
FREAK	0.4707	0.0570	0.3006	0.0783	0.7202	0.1522	0.6022	0.2221
BLOCK	0.0004	0.0010	0.0296	0.0552	0.2430	0.2773	0.5084	0.2980

From Table 1 and Table 2 we observe that SIFT, despite of being the ''oldest'' description technique of the evaluated ones, provides the highest average AUC PR score with the lowest deviation and therefore qualifies as the most robust descriptor against changes in LB, PMT and PD. BRISK and SURF (if the target images are previously smoothed) follow SIFT in robustness, while BLOCK performs the worst, which was somehow expected due to the simplicity of this descriptor. Generally, all descriptors perform better when 2D Gaussian smoothing is applied on the images prior to feature description. Future work will be important to identify the most optimal Gaussian convolution kernels for enhancing descriptor extraction and matching.

Graphical results displayed in Fig. 5 and Fig. 6 provide a more intuitive perspective of the performed measurements that underlie the values expressed in Table 1 and Table 2, providing at the same time additional insights; only the results obtained for the original, non-smoothed, data sets are provided in this format. We observe here that SIFT performs similarly even for large variations in LB, PMT gain and PD. Generally speaking, the other descriptors are increasingly more affected with larger variations in the imaging parameters.

B. ROBUSTNESS OF SIFT, SURF, BRISK, FREAK AND BLOCK DESCRIPTORS EXTRACTED FROM THE TD#1b AND TD#2b BENCHMARKING DATA SETS

In a second segment of our work, we repeated the experiments presented in the previous section, extracting this time SIFT, SURF, BRISK, FREAK and BLOCK descriptors from maximum intensity projected (MIP) versions computed from z-stacks collected on the same sample area. Table 3 and Table 4 report the corresponding average AUC PR scores for TD#1b and TD#2b, respectively, while Fig. 5 and Fig. 6 graphically display the effect of changes in the imaging parameter on the AUC PR scores.

Similar to the experiments on individual images, we notice that: 1) SIFT qualifies as the most robust descriptor against changes in pixel dwell, photomultiplier gain, and laser beam power; 2) all descriptors generally perform better when 2D Gaussian smoothing is applied on the images prior to feature description; and 3) in general, descriptors other than SIFT are increasingly more affected with larger variations in the imaging parameters. It is interesting to observe that in the case of the original, non-smoothed images, all five evaluated descriptors are more robust when extracted from MIP instances. In the case of the smoothed version the situation tends to be the same, except for the case of SIFT

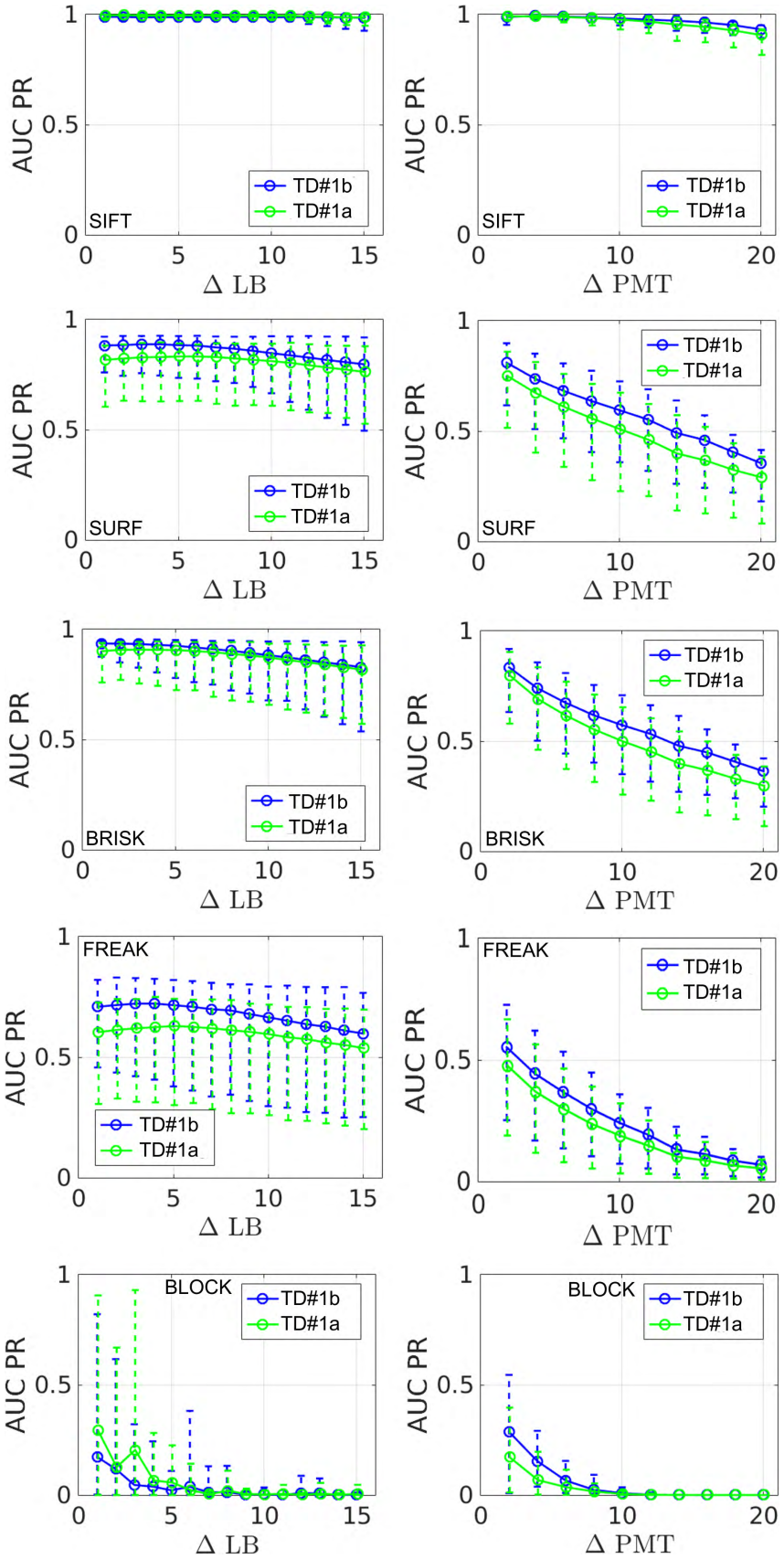

FIGURE 5. AUC PR scores of SIFT, SURF, BRISK, FREAK, BLOCK for TD#1a and TD#1b when laser beam power (left) and photomultiplier gain (right) are varied.

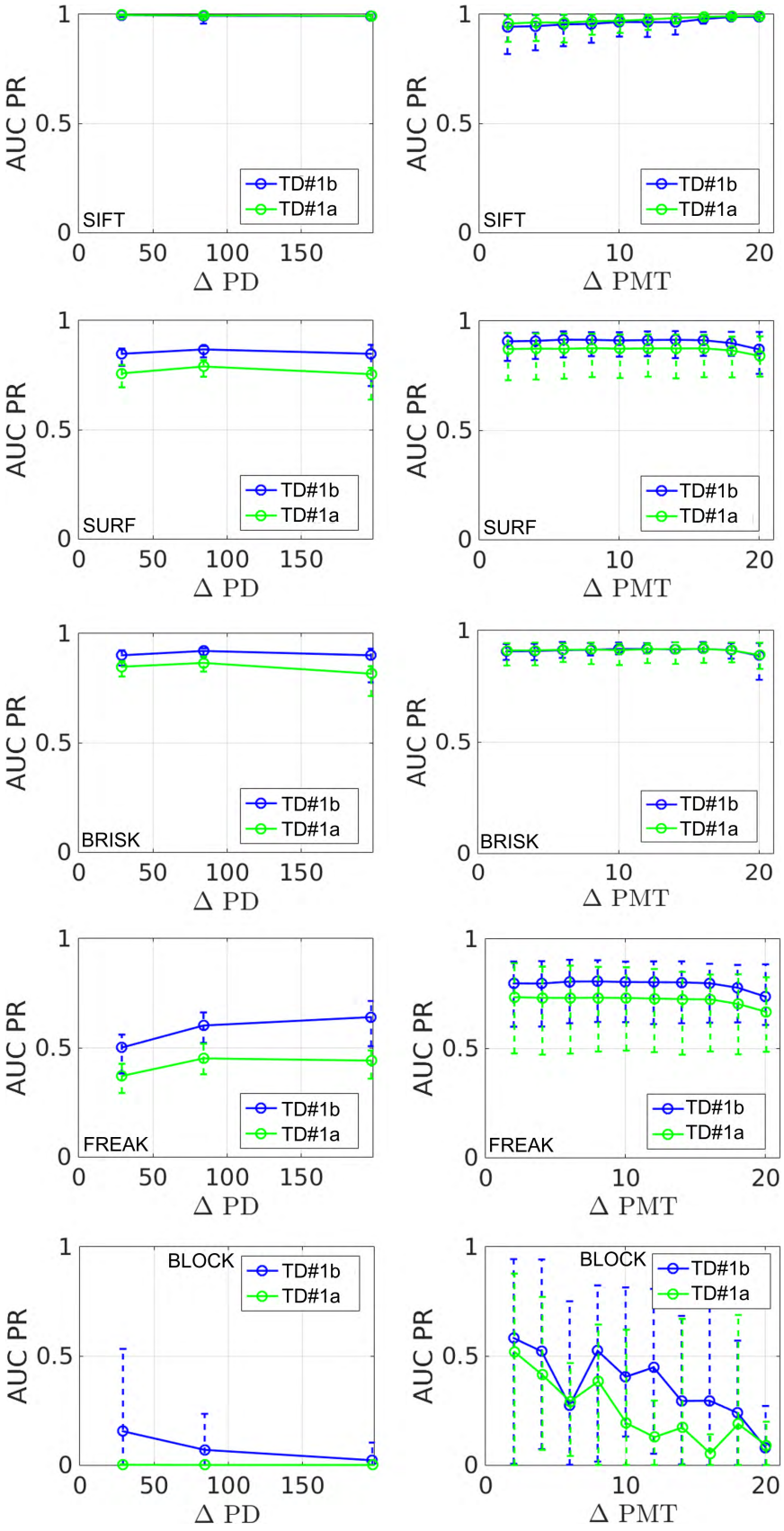

FIGURE 6. AUC PR scores of SIFT, SURF, BRISK, FREAK, BLOCK for TD#2a and TD#2b when pixel dwell (left) and photomultiplier gain (right) are varied.

TABLE 3. AUC PR scores for TD#1b.

	Change in laser beam power				Change in photomultiplier gain			
	Gauss smoothed		Not smoothed		Gauss smoothed		Not smoothed	
Method	mean	std	mean	std	mean	std	mean	std
SIFT	0.9878	0.0113	0.9591	0.0878	0.9737	0.0237	0.9305	0.0933
SURF	0.8560	0.0913	0.7215	0.1149	0.5724	0.1702	0.4876	0.1900
BRISK	0.8948	0.0923	0.9052	0.0847	0.5669	0.1714	0.6120	0.1549
FREAK	0.6770	0.1413	0.6298	0.1165	0.2492	0.1800	0.2694	0.1825
BLOCK	0.0338	0.1074	0.1826	0.2898	0.0543	0.1078	0.1985	0.2929

TABLE 4. AUC PR scores for TD#2b.

	Change in pixel dwell				Change in photomultiplier gain			
	Gauss smoothed		Not smoothed		Gauss smoothed		Not smoothed	
Method	mean	std	mean	std	mean	std	mean	std
SIFT	0.9938	0.0086	0.9361	0.0840	0.9644	0.0488	0.9666	0.0384
SURF	0.8544	0.0362	0.6502	0.0756	0.9062	0.0527	0.7724	0.1464
BRISK	0.9070	0.0305	0.8997	0.0281	0.9105	0.0309	0.9385	0.0349
FREAK	0.5808	0.0806	0.4731	0.0963	0.7922	0.1157	0.7202	0.1662
BLOCK	0.0818	0.1107	0.3487	0.2492	0.3657	0.3290	0.6726	0.3533

which is in general slightly less robust when extracted from MIP instances. Future work will be important to identify the impact of other more recent LSM image representation strategies, such as high-dynamic range [53] or smooth 2D manifold extraction [54], over the robustness of keypoint descriptors and their matching.

V. CONCLUSIONS

In this work we evaluated side-by-side five popular image keypoint description techniques: SIFT, SURF, BRISK, FREAK and BLOCK, with respect to their capacity to represent in a reproducible manner image regions belonging to confocal laser scanning microscopy data sets acquired under different acquisition conditions. We assessed this capacity in terms of matching performance, by means of an AUC PR based methodology used to evaluate the robustness of the five descriptors when dealing with modifications induced by laser beam power, photomultiplier gain and pixel dwell in CLSM images collected on a transverse section of human epithelial tissue. We feel important to emphasize that adapted versions of the proposed evaluation framework can also be used to investigate the robustness of keypoint descriptors to other microscopy imaging parameters (specific to LSM or to different microscopy families).

Three important conclusions that can be drawn from this study are: 1) SIFT outperforms the other evaluated keypoint description techniques; 2) image preprocessing with Gaussian smoothing prior to the descriptor extraction stage yielded an enhanced descriptor matching performance; 3) In the case of non-smoothed images, extraction of keypoint descriptors from MIP instances instead of single optical sections leads to better matching performance for all five evaluated techniques. The same trend was observed for SURF, BRISK, FREAK and BLOCK also when dealing with smoothed images.

Furthermore, we consider the reported results to be important for increasing the awareness of both microscopy and computer vision communities with respect to limitations and vulnerabilities of keypoint description techniques when dealing with LSM images. At the same time, our work shows that image acquisition and pre-processing protocols are instrumental for enhancing keypoint description and matching, which intrinsically translates to a boost in the performance of computer vision applications that rely on keypoint descriptors as building blocks.

ACKNOWLEDGMENT

(Devrim Ünay and Stefan G. Stanciu contributed equally to this work.) The authors thank Prof. Mariana Costache and Dr. Adrian Dumitru of the Carol Davila University of Medicine and Pharmacy in Bucharest (UMF CD), for providing the tissue samples and for interpreting the imaged scene.

REFERENCES

[1] C. J. R. Sheppard and C. J. Cogswell, "Three-dimensional image formation in confocal microscopy," *J. Microsc.*, vol. 159, no. 2, pp. 179–194, 1990.
[2] D. B. Murphy and M. W. Davidson, "Confocal laser scanning microscopy," in *Fundamentals of Light Microscopy and Electronic Imaging*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012, pp. 265–305.
[3] F. Li *et al.*, "Control of apoptosis and mitotic spindle checkpoint by survivin," *Nature*, vol. 396, no. 6711, pp. 580–584, 1998.
[4] N. S. Katheder *et al.*, "Microenvironmental autophagy promotes tumour growth," *Nature*, vol. 541, no. 7637, pp. 417–420, 2017.
[5] R. F. Guthoff, A. Zhivov, and O. Stachs, "*In vivo* confocal microscopy, an inner vision of the cornea—A major review," *Clin. Express Ophthalmol.*, vol. 37, no. 1, pp. 100–117, 2009.
[6] M. Rajadhyaksha, S. González, J. M. Zavislan, R. R. Anderson, and R. H. Webb, "*In vivo* confocal scanning laser microscopy of human Skin II: Advances in instrumentation and comparison with histology," *J. Investigative Dermatol.*, vol. 113, no. 3, pp. 293–303, 1999.
[7] R. Kiesslich, M. Goetz, A. Hoffman, and P. R. Galle, "New imaging techniques and opportunities in endoscopy," *Nature Rev. Gastroenterol. Hepatol.*, vol. 8, no. 10, pp. 547–553, 2011.
[8] P. T. So, C. Y. Dong, B. R. Masters, and K. M. Berland, "Two-photon excitation fluorescence microscopy," *Annu. Rev. Biomed. Eng.*, vol. 2, no. 1, pp. 399–429, 2000.
[9] R. Gauderon, P. Lukins, and C. J. R. Sheppard, "Three-dimensional second-harmonic generation imaging with femtosecond laser pulses," *Opt. Lett.*, vol. 23, no. 15, pp. 1209–1211, 1998.
[10] K. König, "Multiphoton microscopy in life sciences," *J. Microsc.*, vol. 200, no. 2, pp. 83–104, 2000.
[11] W. R. Zipfel, R. M. Williams, R. Christie, A. Y. Nikitin, B. T. Hyman, and W. W. Webb, "Live tissue intrinsic emission microscopy using multiphoton-excited native fluorescence and second harmonic generation," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 12, pp. 7075–7080, 2003.
[12] W. R. Zipfel, R. M. Williams, and W. W. Webb, "Nonlinear magic: Multiphoton microscopy in the biosciences," *Nature Biotechnol.*, vol. 21, no. 11, pp. 1369–1377, 2003.
[13] X. Chen, O. Nadiarynkh, S. Plotnikov, and P. J. Campagnola, "Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure," *Nature Protocols*, vol. 7, no. 4, pp. 654–669, 2012.
[14] C. L. Evans and X. S. Xie, "Coherent anti-Stokes Raman scattering microscopy: Chemical imaging for biology and medicine," *Annu. Rev. Anal. Chem.*, vol. 1, pp. 883–909, Jul. 2008.
[15] S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy," *Opt. Lett.*, vol. 19, no. 11, pp. 780–782, 1994.
[16] G. Zhao *et al.*, "Saturated absorption competition microscopy," *Optica*, vol. 4, no. 6, pp. 633–636, 2017.
[17] P. Wang *et al.*, "Far-field imaging of non-fluorescent species with subdiffraction resolution," *Nature Photon.*, vol. 7, no. 6, pp. 449–453, 2013.
[18] N. Liu *et al.*, "Far-field subdiffraction imaging of semiconductors using nonlinear transient absorption differential microscopy," *ACS Photon.*, vol. 3, no. 3, pp. 478–485, 2016.
[19] G. Danuser, "Computer vision in cell biology," *Cell*, vol. 147, no. 5, pp. 973–978, 2011.
[20] E. Meijering, A. E. Carpenter, H. Peng, F. A. Hamprecht, and J.-C. Olivo-Marin, "Imagining the future of bioimage analysis," *Nature Biotechnol.*, vol. 34, no. 12, pp. 1250–1255, 2016.

[21] M. U. Ghani *et al.*, "Dendritic spine classification using shape and appearance features based on two-photon microscopy," *J. Neurosci. Methods*, vol. 279, pp. 13–21, Mar. 2017.
[22] E. Meijering, O. Dzyubachyk, and I. Smal, "Methods for cell and particle tracking," *Methods Enzym.*, vol. 504, no. 9, pp. 183–200, 2012.
[23] V. Ulman, M. Maška, and C. Ortiz-de-Solorzano, "An objective comparison of cell-tracking algorithms," *Nature Methods*, vol. 14, pp. 1141–1152, Oct. 2017.
[24] P. Paul-Gilloteaux *et al.*, "eC-CLEM: flexible multidimensional registration software for correlative microscopies," *Nature Methods*, vol. 14, no. 2, pp. 102–103, 2017.
[25] S. Xu *et al.*, "Quantification of liver fibrosis via second harmonic imaging of the Glisson's capsule from liver surface," *J. Biophoton.*, vol. 9, no. 4, pp. 351–363, 2016.
[26] (2015). *Memorandum of Understanding for the Implementation of the COST Action A New Network of European Bioimage Analysts to Advance Life Science Imaging CA15124*. [Online]. Available: http://www.cost.eu/COST_Actions/ca/CA15124
[27] J.-G. Wang, J. Li, W.-Y. Yau, and E. Sung, "Boosting dense SIFT descriptors and shape contexts of face images for gender recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2010, pp. 96–102.
[28] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Fundations Trends Comput. Graph. Vis.*, vol. 3, no. 3, pp. 177–280, Jan. 2008.
[29] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: An experimental comparison," *Inf. Retr.*, vol. 11, no. 2, pp. 77–107, Apr. 2008.
[30] D.-N. Ta, W.-C. Chen, N. Gelfand, and K. Pulli, "SURFTrac: Efficient tracking and continuous object recognition using local feature descriptors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2937–2944.
[31] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.
[32] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, Aug. 2007.
[33] Z.-L. Song, S. Li, and T. F. George, "Remote sensing image registration approach based on a retrofitted SIFT algorithm and Lissajous-curve trajectories," *Opt. Express*, vol. 18, no. 2, pp. 513–522, 2010.
[34] A. R. Zamir and M. Shah, "Accurate image localization based on Google maps street view," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 255–268.
[35] T. Tamaki *et al.*, "Computer-aided colorectal tumor classification in NBI endoscopy using local features," *Med. Image Anal.*, vol. 17, no. 1, pp. 78–100, 2013.
[36] H. Li, J. Lu, G. Shi, and Y. Zhang, "Tracking features in retinal images of adaptive optics confocal scanning laser ophthalmoscope using KLT-SIFT algorithm," *Biomed. Opt. Express*, vol. 1, no. 1, pp. 31–40, 2010.
[37] D. Guo, A. L. Van de Ven, and X. Zhou, "Red blood cell tracking using optical flow methods," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 3, pp. 991–998, May 2014.
[38] R. M. Jiang, D. Crookes, N. Luo, and M. W. Davidson, "Live-cell tracking using SIFT features in DIC microscopic videos," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2219–2228, Sep. 2010.
[39] G. G. Lippolis, A. Edsjö, L. Helczynski, A. Bjartell, and N. C. Overgaard, "Automatic registration of multi-modal microscopy images for integrative analysis of prostate tissue sections," *BMC Cancer*, vol. 13, p. 408, Sep. 2013.
[40] S. G. Stanciu *et al.*, "Experimenting liver fibrosis diagnostic by two photon excitation microscopy and bag-of-features image classification," *Sci. Rep.*, vol. 4, Apr. 2014, Art. no. 4636.
[41] T. Saitou, H. Kiyomatsu, and T. Imamura, "Quantitative morphometry for osteochondral tissues using second harmonic generation microscopy and image texture information," *Sci. Rep.*, vol. 8, Feb. 2018, Art. no. 2826.
[42] S. G. Stanciu, R. Hristu, R. Boriga, and G. A. Stanciu, "On the suitability of sift technique to deal with image modifications specific to confocal scanning laser microscopy," *Microsc. Microanal.*, vol. 16, no. 5, pp. 515–530, 2010.
[43] S. G. Stanciu, R. Hristu, and G. A. Stanciu, "Influence of confocal scanning laser microscopy specific acquisition parameters on the detection and matching of speeded-up robust features," *Ultramicroscopy*, vol. 111, no. 5, pp. 364–374, Apr. 2011.
[44] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[45] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
[46] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2548–2555.
[47] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 510–517.
[48] MATLAB. (2015). *Computer Vision System Toolbox, Version 8.6.0 (R2015b)*. [Online]. Available: https://www.mathworks.com/products/computer-vision.html
[49] S. Krig, "Interest point detector and feature descriptor survey," in *Computer Vision Metrics*. Berkeley, CA, USA: Apress, 2014, pp. 217–282.
[50] M. R. Castellanos *et al.*, "Diagnostic imaging of cervical intraepithelial neoplasia based on hematoxylin and eosin fluorescence," *Diagnostic Pathol.*, vol. 10, p. 119, Jul. 2015.
[51] I.-C. Dragoi *et al.*, "Embedding complementary imaging data in laser scanning microscopy micrographs by reversible watermarking," *Biomed. Opt. Express*, vol. 7, no. 4, pp. 1127–1137, 2016.
[52] S. G. Stanciu, D. E. Tranca, 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.
[53] C. Vinegoni *et al.*, "Real-time high dynamic range laser scanning microscopy," *Nature Commun.*, vol. 7, Apr. 2016, Art. no. 11077.
[54] A. Shihavuddin *et al.*, "Smooth 2D manifold extraction from 3D image stack," *Nature Commun.*, vol. 8, May 2017, Art. no. 15554.

DEVRİM ÜNAY received the B.S. degree in electrical and electronics engineering and the M.S. degree in biomedical engineering from Boğaziçi University, Turkey, and the Ph.D. degree in applied sciences from the Faculté Polytechnique de Mons, Belgium. He was a Senior Scientist and a Marie Curie Fellow with the Video Processing and Analysis Group, Philips Research Eindhoven, The Netherlands, a Visiting Researcher with the VPA Laboratory, Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey, and a Faculty Member with Biomedical Engineering Department, Faculty of Engineering and Natural Sciences, Bahcesehir University, Istanbul. He is currently the Head of the Biomedical Engineering Department, İzmir University of Economics, Turkey. His research interests include medical image analysis, content-based information retrieval, machine learning, and machine vision and quality inspection.

STEFAN G. STANCIU received his Ph.D. degree in electronics engineering and telecommunications from the University Politehnica of Bucharest (UPB) in 2011. He served as a Postdoctoral Fellow at ETH Zürich and UPB. He is currently acting as a Scientific Researcher III (research track equivalent of Lecturer) at the Center for Microscopy-Microanalysis and Information Processing, UPB. His main research interests revolve around high- and super-resolution imaging by scanning laser and scanning probe microscopies. His current research agenda focuses on the optical characterization of biological species and advanced materials using combined photonic approaches, but also on the development of novel machine intelligence frameworks for automated disease diagnostics and microscope image analysis.