




# Accurate Segmentation and Registration of Skin Lesion Images to Evaluate Lesion Change

Fulgencio Navarro , Marcos Escudero-Viñolo , and Jesús Bescós 

**Abstract**—Skin cancer is a major health problem. There are several techniques to help diagnose skin lesions from a captured image. Computer-aided diagnosis (CAD) systems operate on single images of skin lesions, extracting lesion features to further classify them and help the specialists. Accurate feature extraction, which later on depends on precise lesion segmentation, is key for the performance of these systems. In this paper, we present a skin lesion segmentation algorithm based on a novel adaptation of superpixels techniques and achieve the best reported results for the ISIC 2017 challenge dataset. Additionally, CAD systems have paid little attention to a critical criterion in skin lesion diagnosis: the lesion's evolution. This requires operating on two or more images of the same lesion, captured at different times but with a comparable scale, orientation, and point of view; in other words, an image registration process should first be performed. We also propose in this work, an image registration approach that outperforms top image registration techniques. Combined with the proposed lesion segmentation algorithm, this allows for the accurate extraction of features to assess the evolution of the lesion. We present a case study with the lesion-size feature, paving the way for the development of automatic systems to easily evaluate skin lesion evolution.

**Index Terms**—Lesion segmentation, image registration, lesion evolution feature, computer-aided diagnosis (CAD), superpixels, LF-SLIC, local features, SP-SIFT.

## I. INTRODUCTION

ACCORDING to the World Health Organization, one in every three cancers diagnosed is a skin cancer. The global incidence of skin cancer continues to rise.

Early detection of skin cancers boosts the effectiveness of the health care actions carried out. For instance, the 5-year survival rate for Melanoma can be increased over 90% if detected in its early stages of development [1].

Due to the widespread unavailability of equipment and qualified human resources required to screen every patient, there is a need for an automated system to assess skin lesions and classify them into melanoma, non-melanoma and benign. This

work presents contributions to the state-of-the-art (SOTA) in this direction.

Dermoscopy or Epiluminescence Microscopy (ELM) is a noninvasive imaging technique that helps diagnose skin lesions. ELM allows visualization of the subsurface structures of the skin revealing lesion details in colors and textures.

ELM improves the detection rate of skin lesions with respect to naked eye inspection, in which the highest accuracy is around 60% [2]. Nevertheless, diagnostic accuracy using ELM largely depends on the dermatologist's experience. Well-trained generalist computer-aided diagnosis (CAD) systems are designed to reduce this dependency. CAD systems may also be used to monitor benign skin lesions in order to prevent their evolution to malignant lessons. Generally, a CAD system is composed of three major stages: *image segmentation*, *feature extraction*, and *classification* [3], [4].

*Image segmentation* is used to locate the boundary between the lesion area and the surrounding skin. Obtaining an accurate segmentation of the lesion is important, especially to provide low error rates prior to later quantification of the shape, border and size features of the skin lesion [5]. In general, the segmentation process aims at the spatial discrimination of sets of inter-related pixels in a region of interest (ROI) to facilitate the detection of spatial transitions between these sets. Reported skin lesion segmentation methods are based on: edge extraction, image thresholding, region segmentation, artificial intelligence or active contours.

Edge based techniques [6], [7] are based on information about the image edges; more specifically, they search for abrupt changes in the intensity of neighboring image pixels.

The segmentation process may also depend on similarity criteria, such as similar grey levels, colors or textures. Thresholding [8], [9] and region-based [10], [11] segmentation are examples of methods that use similarity criteria to identify skin lesions in images.

Techniques based on artificial intelligence (AI) [12], [13] classify pixels as belonging to the lesion or to the background of the images. Neural networks, evolutionary computation and fuzzy logic are some examples of these techniques.

Algorithms based on active contours are also used for segmenting skin lesion images [14], [15]. In these algorithms, the initial curves evolve towards the boundaries of the lesion through appropriate automatic deformation.

*Feature extraction* plays a major role in automatic skin lesion diagnosis. In human-driven analysis of skin lesions, there are widely accepted templates to evaluate the evidence of particular

Manuscript received October 31, 2017; revised February 6, 2018 and March 26, 2018; accepted April 1, 2018. Date of publication April 10, 2018; date of current version March 6, 2019. This work was supported in part by the Spanish Government (HAVAide, TEC2014-53176-R) and in part by the TEC department (Universidad Autónoma de Madrid). (Corresponding author: Fulgencio Navarro.)

The authors were with the Video Processing and Understanding Lab, Universidad Autónoma de Madrid, Madrid 28049, Spain (e-mail: ful.navarro@estudiante.uam.es; marcos.escudero@uam.es; j.bescos@uam.es)

Digital Object Identifier 10.1109/JBHI.2018.2825251



Fig. 1. ELM skin lesion image (left), skin lesion segmentation (middle), and skin lesion registered over a subsequent ELM skin lesion image (right).

lesions. For instance, dermatologists have created the ABCDE rule for Melanoma lesions. Melanomas tend to be Asymmetric, have an irregular Border, present uneven Color distributions, their Diameter is greater than 6 mm and they Evolve in size, shape and color.

The (E)volve feature is a key element in the diagnosis of pigmented lesions. Its extraction is based on a prior image registration stage. Therefore, image registration is a critical task and an area that has been widely studied. Image registration can be done at a full-body level, i.e., full-body images are registered to detect the apparition of new moles or the growth of pre-existing ones [16]–[18]. Image registration can also be done at the level of individual single skin lesions [19], [20]. Skin lesions are registered with millimetric precision so even the smallest changes in the lesion can be observed. The main techniques tend to rely on points matching [21]–[23] or regions [20]. Some solutions include a prior skin lesion segmentation process [24].

Advances in the feature extraction stage in CAD systems have been focused on the automatic extraction of these cues. The spatial pixel area extracted from the segmentation process has been analyzed to derive asymmetry, shape, border and diameter features (ABCD rule) [25], [26].

Nevertheless, the automatic extraction of these features is problematic mainly due to inaccuracies at the segmentation stage and to the complexity in registering images of a skin lesion taken at different times.

*Classification* consists in recognizing and interpreting the information on the pigmented skin lesions based on the features extracted.

The main contributions of this paper are: a novel segmentation algorithm which obtains a highly precise isolation of the skin lesion; and a new strategy to match successive images of a skin lesion in order to measure its evolution. Segmentation is based on the superpixels technique, which provides a tight-to-boundaries result that enables the ABCD features to be reliably obtained from a single image. We then present a novel image registration process to measure the evolution (E) of the features describing a skin lesion, given two images capturing different stages of the lesion. An example of the above is shown in Fig. 1.

The rest of the paper is organized as follows. In Section II we present the background to the segmentation and registration techniques. The proposals for segmentation and registration are set out in Section III. Section IV lays out the experimental results. Section V includes a discussion about proposals and their results. The last section draws some conclusions.

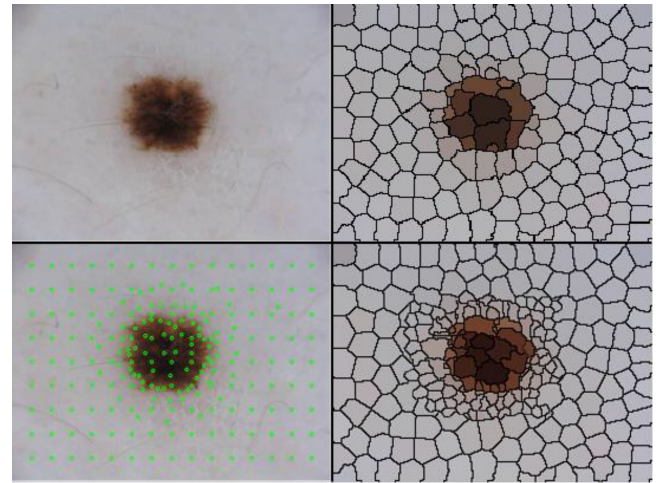


Fig. 2. SLIC and LF-SLIC comparison. Top row shows an input image (left) and the SLIC results (right). Bottom row shows LF-SLIC seeds in green superimposed on the image (left) and resulting LF-SLIC segmentation (right).

## II. BACKGROUND

### A. Superpixel Segmentation

Region segmentation techniques are pivotal for several computer vision applications including object segmentation [27], depth estimation [28], body model estimation [29], and object localization [30]. Traditional region segmentation approaches generate regions based on similarity measures. These measures may yield low levels of accurate regions if image borders are not well defined due to poor or smooth contrast [31]. Recently, superpixels have proven to be an efficient and effective solution for performing image segmentation under a conservative, low-error paradigm [32].

Superpixel algorithms group pixels into perceptually meaningful regions (see Fig. 2 top image) by capturing image redundancy and preserving boundaries. Superpixels provide a convenient primitive from which to compute image features, greatly reducing the complexity of subsequent image processing tasks [32]. They provide more reliable lower-error supports than SOTA alternatives [31]. In particular, a superpixel method segments a given image  $\psi$  into a set of superpixels  $\{\Omega_j, j = 1 \dots J\}$ , such that  $\psi = \cup_1^J \Omega_j$  and  $\Omega_j \cap \Omega_k = \emptyset, \forall j \neq k$ .

In spite of their widespread use in computer vision, their use in medical image applications is scant. However, the number of successful applications based on superpixels in this field is growing fast. They have been used as supporting regions for computational methods [33], as a tool for generic medical image segmentation [34], in detection tasks for segmenting ROIs [35], and prior to feature extraction tasks [36].

Among existing superpixel segmentation techniques, SLIC [32] has excelled due to its accuracy. SLIC depends on a single parameter,  $\tilde{J}$ , which represents the desired number of approximately equally-sized superpixels.  $\tilde{J}$  is initially used to set the superpixel centers by evenly sampling the image with a regular grid (see Fig. 2 top image). Centers are then moved to

seed locations corresponding to the lowest gradient position in a  $3 \times 3$  neighborhood. Then, following a scheme similar to the classical K-Means algorithm [37], pixels in the image are associated with the *nearest* center, attending to spatial distance and color similarity. After this first association, centers are repositioned in their respective clusters, the error between previous and current centers is obtained, and the association is performed again. This process is repeated until error convergence is reached.

### B. Local Features and Descriptors

Image registration has been tried with a wide variety of strategies. Some applications require high precision during registration. In those applications, local features have emerged as the most reliable technique in the SOTA.

A local feature is an image pattern which differs from its immediate neighborhood. It is usually associated with a change or singularity of an image property or set of properties, e.g., image intensity, color, or texture. Local features are usually image points, but they also can be edges or small image patches. A feature is typically described or characterized by indicators extracted from a region around it, which overall conform the *feature descriptor*. These descriptors are vectors which ideally characterize an image feature unequivocally.

Feature detectors can be categorized based on the type of image structures they extract: corners, blobs or regions. According to [38], feature descriptors can be categorized as local binary, spectral, basis spaces, polygons and volume. Local binary descriptors are the fastest approaches, and spectral descriptors are the most used in the SOTA. Generally, the prevalence of one scheme over the other depends on the target application.

The Scale Invariant Feature Transform (SIFT) [39] is a well-known and widely used technique which combines both a feature detector and a description methodology. The detector, based on scale-space theory [40], identifies robust-to-scale feature points. The descriptor, a 128-dimensional vector based on gradient distribution around the feature point, is invariant to image rotation.

When the SIFT feature points are located close to the boundaries of the object being characterized, the descriptor might be affected by non-object areas, which leads to poor characterization. In order to solve this, we propose the SP-SIFT [41] technique. It previously segments the image in superpixels isolating information areas. Then it computes the SIFT descriptor using just the pixels of the superpixel containing the feature point. The result is a full-foreground or full-background feature descriptor.

## III. MATERIALS AND METHODS

### A. Skin Lesion Segmentation

We present here an adaptation of the SLIC segmentation algorithm to ELM images of skin lesions, which is the first scientific contribution of this paper. We currently assume that the image captures a single lesion, and that the lesion is fully contained in

the image, a common acceptable situation; these requirements could, however, be removed in the future.

**SLIC guided by local features (LF-SLIC):** The original SLIC technique has been shown to be highly competitive for image segmentation. However, if there is a region of interest (ROI) in the image, segmenting the whole image is useless whereas accurately defining the contour of the ROI is convenient.

Initializing centers using a regular grid (see Fig. 2 top image) results in missing small details in the skin lesion boundaries while it extracts useless boundaries in the rest of the skin.

We propose to replace uniform center initialization with feature-driven initialization, so that superpixels are forced to be smaller around detected features. We first use the SIFT detector to identify feature points in the image. Then, we use these as anchor points to place initial centers using a Gaussian distribution (see Fig. 2 bottom image). Finally, we apply the SLIC algorithm to these centers. The result is a higher precision in the segmentation around the skin lesion at the expense of a lower precision in the segmentation in the rest of the image.

#### LF-SLIC region labeling via spatial continuity classification:

The image  $\psi$  is segmented into a set of LF-SLIC superpixels  $\{\Omega_j, j = 1 \dots J\}$ , where  $J$  is the number of superpixels. Two superpixels,  $\Omega_j$  and  $\Omega_{j'}$ , are neighbors if at least one of the pixels in  $\Omega_j$  is 8-connected with a pixel in  $\Omega_{j'}$ . Let  $\overline{bw}$  be the bandwidth of this partition, defined as the largest 5-dimensional distance vector (evaluating position and RGB color) between the centers of any two neighboring superpixels.

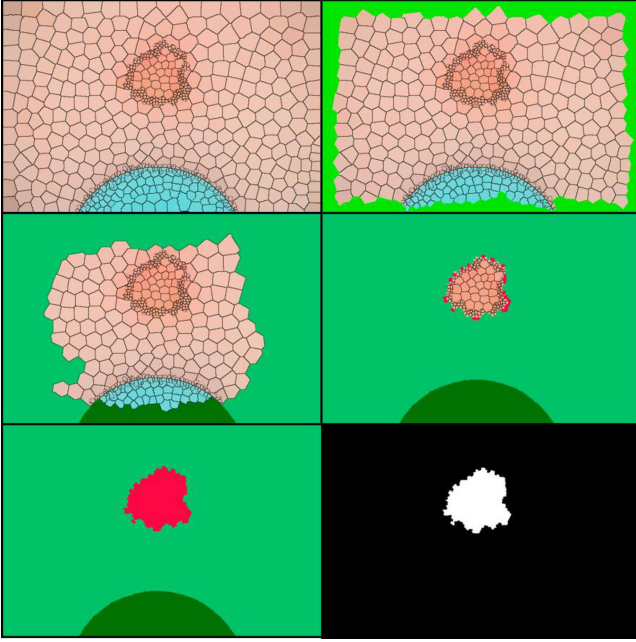
The final objective is to obtain two disjoint sets of superpixels: a subset of superpixels classified as non-lesion  $N = \{\Omega_p, p \in [1 \dots J]\}$ ,  $|N| = P$ ; and a subset of lesion superpixels  $L = \{\Omega_q, q \in [1 \dots J]\}$ ,  $|L| = Q$ , where  $P + Q = J$ . For this purpose, a greedy labelling scheme with connectivity restrictions is proposed.

First, assuming that the lesion is fully contained in the image, all superpixels that are 8-connected to the image boundary are assigned to the  $N$  set (see Fig. 3 top image), creating an initial estimation of non-lesion superpixels,  $N^0$ , and the complementary initial set of lesion superpixels,  $L^0$ . Then, superpixels in the  $N^0$  set are grouped into regions using a conservative mean-shift approach [42] with a bandwidth  $\overline{bw}_{MS} = \overline{bw}$ . This process merges superpixels in  $N^0$  into regions  $\{R_1 \dots R_m \dots R_M\}$ , each containing a subset  $\{\Omega_{p,m}\}$  of the  $N^0$  superpixels. Due to this conservative grouping, the set of colors  $\{\bar{c}(\Omega_{p,m})\}$  of the superpixels in every region can be assumed to define a close-to-Gaussian-distribution. Under this assumption, superpixels in  $L^0$  are reclassified by evaluating their likelihood to be part of any of their 8-connected regions in the set  $\{R_1 \dots R_m \dots R_M\}$ . For this purpose, for a superpixel in  $L^0$  with color  $\bar{c}(\Omega_q)$  that is connected to region  $R_m$ , a Grubbs' test is used to determine whether the superpixel is an outlier of the color distribution inside  $R_m$ :

$$G = \frac{|E[\bar{c}(\Omega_{p,m})] - \bar{c}(\Omega_q)|_2}{\sigma[\bar{c}(\Omega_{p,m})]} \quad (1)$$

with  $E[\bar{c}(\Omega_{p,m})]$  as the mean vector of the colors of the superpixels in  $R_m$  and  $\sigma[\bar{c}(\Omega_{p,m})]$ , its standard deviation. The hypothesis of  $\Omega_q$  where part of  $R_m$  is accepted at significant





**Fig. 3.** LF-SLIC labeling process. The top-left image shows the LF-SLIC superpixels segmentation. The top-right image shows the  $N^0$  set of superpixels in light green. The mid-left image shows an iteration  $t$ , where different green areas indicate different clusters formed in the  $N^t$  set. The mid-right image shows in red the superpixels classified into the  $L^t$  set for a later iteration. The bottom row shows the final classification: the left image describes the final clusters (green for the  $N$  set and red for the  $L$  one) while the right one depicts the final segmentation mask.

level  $\alpha$ , fixed in the experimental parameter setup. If:

$$G \leq \frac{M-1}{\sqrt{M}} \sqrt{\frac{M-1}{M-2 + ts^2_{\alpha/2M, M-2}}}$$

where  $M$  is the number of superpixels in the set, and  $ts$  the Student's t-distribution.

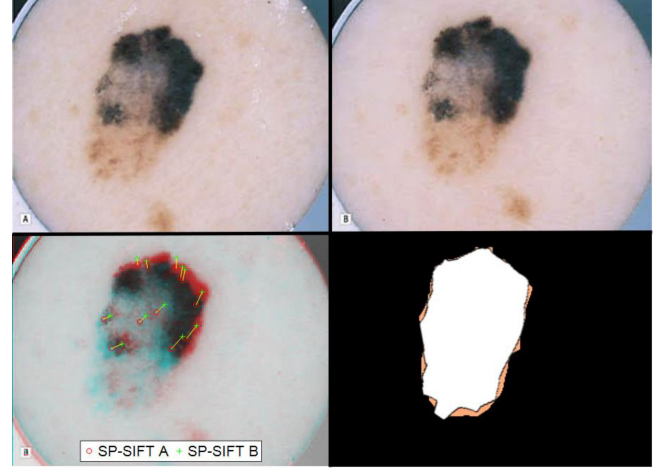
Reclassified superpixels are removed from  $L^0$  and assigned to  $N^0$ . This process is repeated for any superpixel in  $L^0$  which is a neighbor of at least one region  $\{R_1 \dots R_m \dots R_M\}$ , creating two new sets  $L^1$  and  $N^1$ .

The whole process is repeated until at a given iteration, say  $t$ , no further reassignments are performed. The sets at this iteration  $N = N^t$  and  $L = L^t$  define a tight-to-boundaries segmentation of the skin lesion (see Fig. 3 for iteration examples). However, artifacts in the image can also be segmented, so we propose an artifact removal method.

**Artifact removal:** The first step for the successful removal of artifacts is to define them precisely. According to SOTA reports, artifacts in ELM images mainly consist of hair and air bubbles.

These artifacts clearly differ from the skin lesions. Skin lesion boundaries show irregular shapes, and present smooth transitions with the surrounding skin; on the contrary, the artifacts identified show contours that contrast greatly with the surrounding skin and also very regular shapes: straight lines for the hairs and circles for the bubbles.

There are many established approaches to detect pre-defined and highly contrasted shapes. For this purpose, we propose to use the well-known Hough Transform (HT) [43], a voting



**Fig. 4.** Skin lesion registration and size evolution. The top row shows the first (A) and second (B) skin lesion images. The bottom left image shows the matched SP-SIFT feature points between both input images. The bottom right image shows the segmentation masks aligned or registered for easy use in size comparison.

scheme that obtains highly robust detection results in these situations. We apply the HT to detect pixels belonging to lines, circles or ellipses in the segmented image. Superpixels containing pixels voted as lines, circles or ellipses are re-classified into the  $N$  set.

## B. Skin Lesions Registration to Evaluate Lesion Change

The second contribution of this paper is the measurement the evolution of a skin lesion, given two images capturing different stages of the lesion, a crucial criterion for diagnosis.

The result of the proposed segmentation is a precise image of the isolated skin lesion, which allows for the extraction, for instance, of the ABCD features to further classify the lesion. If we have two images of the same lesion captured on different days, we could measure the change or evolution (E) in the ABCD features. However, in order for this process to be reliable and effective, both images should show the lesion with a comparable scale, orientation and point of view; that is, an image registration process should first be performed.

**Image registration with SP-SIFT features:** Registration requires identifying the same feature points in the two images to perform proper image alignment.

SOTA algorithms for skin lesion registration face the problem of aligning reference features that may have suffered remarkable changes (evolution).

We propose to use the SP-SIFT technique to detect and describe feature points in both images first, so that the evolution of the skin does not corrupt the characterization of the feature points. Detected features are used to establish matching points between these two images. These matches define a geometric transform (in this case, homography) between the pair of images. We use the transformation to align both images. An example of the image alignment process is depicted in Fig. 4.

**Evaluation of the lesion change:** In this paper we do not explore the extraction of features to characterize skin lesions.

Instead, we focus on obtaining a precise segmentation in order to extract the desired features more accurately, and on registering skin lesion images to allow comparison of features extracted at different times and then evaluate lesion change.

In order to illustrate and demonstrate the potential of our proposal, we present in the next section results on the evolution of the size of the lesion, one of the main characterization features. Variations could also be obtained for color, boundaries or asymmetry; however, this falls outside the scope of this work.

We use the image registration technique to align both skin lesions and their segmentations. We compare the segmented areas and calculate a pixel-level difference. The scale of the images is known, so we can map pixels to millimeters and provide the size-feature evolution in a comprehensive metric.

#### IV. EXPERIMENTAL RESULTS

We present here the results of a comparative analysis of the proposed segmentation method, using the recent benchmark proposed in the scope of the ISIC 2017 challenge.<sup>1</sup> Additionally, we evaluate the proposed image registration method against a modified version of the ISIC 2017 test set. Finally, we show how these methods allow a precise evaluation of the variation in the diameter of a skin lesion.

##### A. Evaluation of the Proposed Segmentation Method

**Data analyzed:** We have arranged the data according to the ISIC 2017 evaluation framework:

- Training data: 2000 dermoscopic images and their respective 2000 binary ground-truth masks.
- Validation data: 150 dermoscopic images and their respective 150 binary ground-truth masks.
- Test data: 600 dermoscopic images and their respective 600 binary ground-truth masks.

The training data are used to set the algorithm parameters; the validation data are used to assess the setup; the test data are used to evaluate the proposed algorithm and to perform the comparison with alternative state-of-the-art algorithms.

**Evaluation measures:** We first select the Jaccard Index ( $J$ ), which is one of the most widely-used metrics to evaluate segmentation methods, and the one used in the ISIC 2017 challenge.  $J$  is also known as *intersection-over-union*. It is defined as the ratio  $J(A, B) = |A \cap B| / |A \cup B|$ , where  $A$  and  $B$  are two binary masks; and it provides a normalized measure, the higher the better, of the overall performance of a segmentation method. We complement this indicator with the Dice coefficient ( $S$ ), also widely used to evaluate the similarity between two binary masks.  $S$  is usually considered to be a semi-metric version of  $J$ :  $S(A, B) = |A \cap B| / (|A| + |B|)$ . Additionally, as segmentation can be viewed as a pixel classification task, performance can also be measured by a classification quality indicator. We used Accuracy:  $ACC = TP + TN / (TP + TN + FP + FN)$ .

**System setup:** Default parameters are used for the SLIC, SIFT and SP-SIFT methods. The Mean-Shift bandwidth is set

TABLE I  
SEGMENTATION RESULTS ISIC 2017 CHALLENGE DATASET

Reference	Jaccard Index	Dice Coefficient	Accuracy
Top 1	0.765	0.849	0.934
Top 2	0.762	0.847	0.932
Top 3	0.76	0.844	0.934
Top 4	0.758	0.842	0.934
Top 5	0.754	0.839	0.934
<b>Proposed-1</b>	<b>0.769</b>	<b>0.854</b>	<b>0.955</b>
<b>Proposed-2</b>	<b>0.846</b>	<b>0.938</b>	<b>0.960</b>

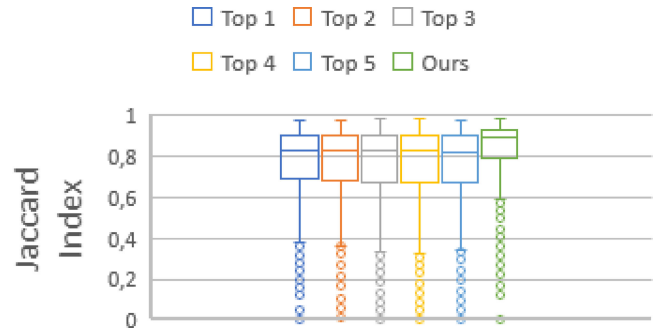


Fig. 5. Distribution of the Jaccard Index for all the images in the test set of the ISIC 2017 segmentation challenge. See text for discussion.

according to the LF-SLIC result. Hence, the training and validation data are just used to set the value of the significant level  $\alpha$  associated to the Grubbs' test. For this purpose, we have obtained  $ACC$  values for the range  $\alpha \in [0.7, 0.99]$ . We selected as a trade-off value the one that returns the highest value in both sets. In the experiments, this value was  $\alpha = 0.91$  achieving  $ACC = 0.998$  in both validation and training sets.

**Quantitative results:** The proposed method is compared (see Table I – *Proposed-1*) to the Top-5 algorithms in the ISIC 2017 Challenge; the Dice Coefficient and the Accuracy are also included. To better assess our method's performance, we also include our results (see Table I – *Proposed-2*) previously removing from the dataset those images that do not fulfill our assumptions (i.e., images where the skin lesion is not fully contained in the image). To further evaluate the operational range of the methods compared, Fig. 5 depicts box-plot diagrams of the Jaccard Index distribution: the vertical size of the box indicates result dispersion (standard deviation) and the horizontal lines represent average values; points outside the boxes are outliers.

##### B. Evaluation of the Proposed Lesion Registration Method

The aim of this experiment is to assess the effectiveness of the proposed registration method in the task of aligning skin lesion images. We compare the performance in this task of the SP-SIFT technique against two well-known feature detection-description algorithms: SIFT [39] and SURF [44].

**Data analyzed:** In order to carry out a systematic evaluation, we use the ISIC 2017 test dataset as the set of initial skin lesion images (i.e., those corresponding to the initial lesion capture),

<sup>1</sup><https://challenge.kitware.com/#challenge/583f126bcad3a51cc66c8d9a>

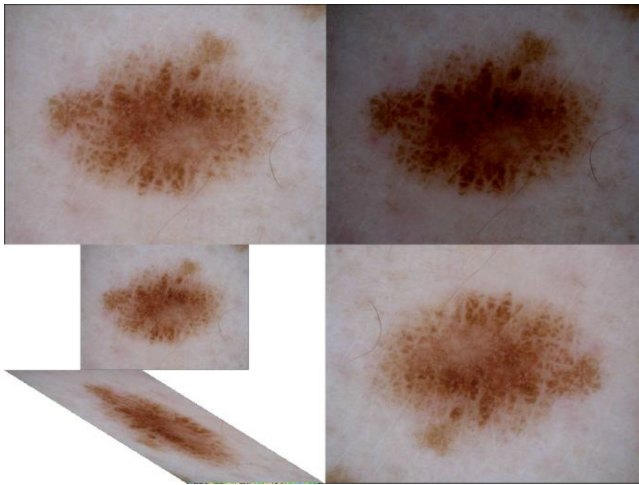


Fig. 6. Example of image distortion applied to the ISIC 2017 segmentation test set. First row, original image (left) and light change (right). Second row, scale change (top-left), view-point change (bottom-right), and orientation change (right).

and we then generate for each image in this test set, a new image simulating a capture in a different instant/conditions: we randomly generate one of the following modified images: a illumination change, a rotation or orientation change, a scale change, or a change in the point of view (see Fig. 6).

**Evaluation measures:** Each technique compared extracts local features from both the original image and each of the modified images, and matches them to establish correspondences between the initial and the modified image. The quality of the correspondence is then evaluated in terms of average precision and recall: if the correspondence is correct, a true positive is declared ( $TP$ ); if it is incorrect a false positive is declared ( $FP$ ); if no correspondence is established, a false negative is declared ( $FN$ ). Precision ( $P$ ) and recall ( $R$ ) of the matching process are then defined as  $P = TP / (TP + FP)$  and  $R = TP / (TP + FN)$ .

**Quantitative results:** Fig. 7 includes the results obtained for the three techniques on the modified version of ISIC 2017 test dataset in terms of average precision and recall. Results are given for each image modification.

### C. Case Study: Assessing the Evolution of the Lesion Diameter

In this experiment the objective is to present a potential application of the image registration process: measuring the evolution of the lesion's diameter.

**Data analyzed:** For this experiment, we use a subset of the [45] dataset. This subset contains 10 pairs of images from 10 different patients. Temporal distance between images of the same patient ranges between a few days (6) and a few months (4.5). Each pair of images has associated ground-truth information indicating the diameter variation between them.

**Evaluation measures:** We perform the evaluation based on two criteria. The average number of correctly matched points between the two temporally spaced samples and the error in mm ( $\varepsilon$ ) between the predicted and the annotated diameter change.

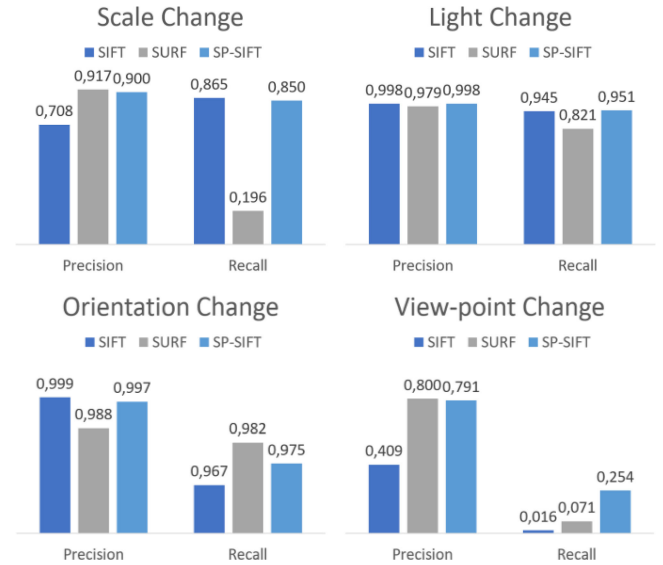


Fig. 7. Precision and Recall matching results for modifications of the images in the ISIC 2017 test set.

TABLE II  
SP-SIFT IMAGE REGISTRATION AND DIAMETER EVOLUTION RESULTS

Category	Ground-truth diameter evolution (mm)	Estimated diameter evolution (mm)	$\varepsilon$ (mm)	Average matched features
No Change	0	0.01	0.01	64.32
Short time	1.28	1.46	0.18	12.93
Mid/large time	4.51	5.15	0.60	3.23
Overall	<b>1.63</b>	<b>1.86</b>	<b>0.23</b>	<b>26.56</b>

Note that the image registration process, i.e., the homography estimation, requires at least three matched points.

**Quantitative results:** Table II shows the results of the evaluation on average. To evaluate the capabilities of the method better with respect to time variation, images are grouped into three categories:

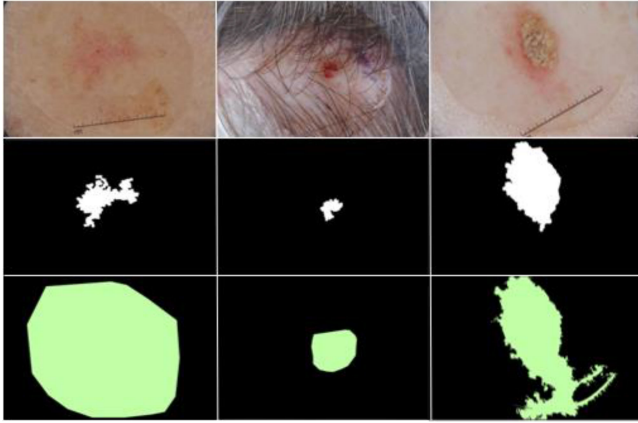
- No change: the skin lesion analyzed suffered no change between the first and the second picture.
- Short time: the time elapsing between the first and second images is less than 2 weeks. Changes are expected to be small.
- Medium to long time: the time elapsing between the first and the second image is more than 2 weeks. Changes are expected to be bigger than in the short time category.

## V. DISCUSSION

### A. Proposed segmentation method

In the segmentation stage, we extracted referenced results of state-of-the-art methods from the ISIC 2017 skin lesion segmentation challenge. Top-ranked algorithms present Jaccard Indexes ranging from 0.765 to 0.754, all very close (see Table I). The proposed segmentation method yields a Jaccard Index of 0.769, outperforming the other approaches. Besides, the proposed method also performs better in terms of the Dice





**Fig. 8.** Failure cases (three of the outliers in the Jaccard Index distribution presented in Fig. 5). First row, dermoscopic images. Second row, segmentation results obtained with the proposed method. Third row, ground-truth segmentation.

Coefficient and Classification Accuracy. Results are obtained using the whole test set, including images that do not meet the method’s prerequisite of having the skin lesion fully contained in the image. For a deeper understanding of the segmentation results, we include a box plot graphic in Fig. 5. The proposed method also outperforms the other methods by yielding a lower deviation, i.e., its operation is more stable for more images in the set. However, the distribution of the Jaccard Index achieved by the proposed method presents a higher number of outliers than the other methods. These outliers are basically the images which do not meet the prerequisite. If these images are removed, results improve up to 0.846 in Jaccard Index terms, 10.56% better than the top approach in the challenge (see Table I).

Results of the proposed approach (and of all the other approaches evaluated) are biased by the annotated ground-truth. Despite the high quality of the dataset, and the amount of data provided, the annotation of skin lesions is a subjective task. This can be observed in the failure cases presented in Fig. 8. The ground-truth annotations of the images in the two first columns are not tight to the lesion itself, but rather include a roughly affected spatial area around it which substantially differs from the proposed segmentation, which is tighter to the lesion. Differently, the third column depicts an example of an annotation mistake, in which the ruler is included in the ground-truth mask.

Despite the good results obtained, there is room for improvement. Superpixel segmentation provides a robust tool for skin lesion segmentation. However, the accuracy of the segmentation on the lesion boundaries is biased by the superpixels’ sizes. Despite the high accuracy achieved by the LF-SLIC, it can be improved by operating at pixel level.

### B. Proposed registration method and evolution assessment

Although it is a key stage for the extraction of feature evolution, to the best of our knowledge, there is no prior study dealing with skin lesion registration. We present a comparison between SOTA local features, as they have been shown to be successful

tools for image registration in other fields. According to Fig. 7, the SP-SIFT descriptor used for describing the superpixels obtained by the LF-SLIC segmentation, yields better results than the SIFT and SURF techniques. Light changes are well handled by both the proposed SP-SIFT scheme and the SIFT features. The scale changes affect the SIFT features slightly, but the proposed version of SP-SIFT is robust to these changes due to the tightness of the description supports. Finally, whereas geometric changes in terms of image orientation are handled well by all three methods, affine transformations or point-of-view changes are still challenging. Despite the proposed version of SP-SIFT yielding a recall 8.48% and 46.37% better than SIFT and SURF, its results can be still improved.

The image registration is presented as a tool to facilitate the extraction of feature evolution (E). Dermatologists agree on the relevance of the features’ evolution over time to detect potentially malignant lesions. Whereas there are some studies that describe strategies to extract this feature, the complexity of the process hinders the existence of robust automatic approaches and SOTA evaluations.

In this paper, the potential of image registration is exemplified by evaluating the variation in the diameter of 10 different skin lesions. The results obtained (see Table II) indicate that there is an average error of 0.23 mm between the estimated and the real evolution of lesion diameters. Considering that the critical diameter of a skin lesion is 6 mm, the error represents a deviation of 0.04% of this magnitude. However, results also suggest that accuracy degrades with the time elapsed between lesion samples, suggesting that a continuous observation of the lesion will be required for effective assessment of its evolution. The downgrading can be explained by the decrease in the average number of matched features. For large time lapses, the average number of correctly matched features is close to three, the minimum number required for image registration. In these situations, registration may be driven by incorrectly matched features.

## VI. CONCLUSION

The two main contributions of this paper are an algorithm for the accurate segmentation of skin lesions, and an algorithm for the accurate registration of two images of the same skin lesion. Moreover, these algorithms operate together to achieve a more challenging objective: a precise segmentation mask enables the extraction of precise features characterizing the skin lesion; precise registration further allows reliable measurement of the evolution of such features, which is also in a major contribution of this work.

We propose a segmentation algorithm that relies on a novel super-pixel segmentation method, which we refer to as LF-SLIC, combined with a robust artifact removal technique. Results demonstrate that this algorithm achieves top SOTA results with the dataset provided by the ISIC 2017 skin lesion segmentation challenge.

We also propose a technique for the registration of skin lesion images. The proposal uses a feature point detection and description technique, the SP-SIFT, which combines the SIFT detector

with a description guided by superpixels segmentation. The experimental results show that the proposal is able to perform the skin lesion registration under different capture conditions and lesion stages.

Finally, the combination of these techniques, an accurate segmentation and a reliable image registration, paves the road for the precise computation of features' evolution and automatic skin lesion classification.

## REFERENCES

- [1] G. Zongyuan *et al.*, "Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images," *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer-Verlag, 2017, pp. 250–258.
- [2] M. Zhen and J. M. R. S. Tavares, "A novel approach to segment skin lesions in dermoscopic images based on a deformable model," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 2, pp. 615–623, Mar. 2016.
- [3] B. Van Ginneken, B. M. Ter Haar Romeny, and M. A. Viergever, "Computer-aided diagnosis in chest radiography: A survey," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1228–1241, Dec. 2001.
- [4] H. Cheng *et al.*, "Computer-aided detection and classification of microcalcifications in mammograms: A survey," *Pattern Recognit.*, vol. 36, no. 12, pp. 2967–2991, 2003.
- [5] M. E. Celebi *et al.*, "Lesion border detection in dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 33, no. 2, pp. 148–153, 2009.
- [6] C. A. Z. Barcelos and V. B. Pires, "An automatic based nonlinear diffusion equations scheme for skin lesion segmentation," *Appl. Math. Comput.*, vol. 215, pp. 251–261, 2009.
- [7] D. H. Chung and G. Sapiro, "Segmenting skin lesions with partial-differential-equations-based image processing algorithms," *IEEE Trans. Med. Imag.*, vol. 19, pp. 763–767, Jul. 2000.
- [8] R. Kaur *et al.*, "Thresholding methods for lesion segmentation of basal cell carcinoma in dermoscopy images," *Skin Res. Technol.*, vol. 23, no. 3, pp. 416–428, 2017.
- [9] R. Garvani *et al.*, "Border detection in dermoscopy images using hybrid thresholding on optimized color channels," *Comput. Med. Imag. Graph.*, vol. 35, no. 2, pp. 105–115, 2011.
- [10] M. Silveira *et al.*, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *J. Sel. Topics Signal Process.*, vol. 3, pp. 35–45, 2009.
- [11] M. Rajab *et al.*, "Application of region-based segmentation and neural network edge detection to skin lesions," *Comput. Med. Imag. Graph.*, vol. 28, no. 1, pp. 61–68, 2004.
- [12] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 994–1004, Apr. 2017.
- [13] G. Schaefer *et al.*, "Colour and contrast enhancement for improved skin lesion segmentation," *Comput. Math. Methods Med.*, vol. 35, pp. 99–104, 2011.
- [14] H. Zhou *et al.*, "Skin lesion segmentation using an improved snake model," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, 2010, pp. 1974–1977.
- [15] H. Zhou *et al.*, "Mean-shift based gradient vector flow for image segmentation," *Comput. Vis. Image Understanding*, vol. 117, pp. 1004–1016, 2013.
- [16] H. Mirzaalian, T. K. Lee, and G. Hamarneh, "Skin lesion tracking using structured graphical models," *Med. Image Anal.*, vol. 27, pp. 84–92, 2016.
- [17] B. McGregor, "Automatic registration of images of pigmented skin lesions," *Pattern Recognit.*, vol. 31, no. 6, pp. 805–817, 1998.
- [18] K. Korotkov, J. Quintana, S. Puig, J. Malvehy, and R. Garcia, "A new total body scanning system for automatic change detection in multiple pigmented skin lesions," *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 317–338, Jan. 2015.
- [19] S. A. Pavlopoulos, "New hybrid stochastic-deterministic technique for fast registration of dermatological images," *Med. Biol. Eng. Comput.*, vol. 42, no. 6, pp. 777–786, 2004.
- [20] H. Huang and P. Bergstresser, "A new hybrid technique for dermatological image registration," in *Proc. IEEE Int. Conf. Bioinform. Bioeng.*, 2007, pp. 1163–1167.
- [21] J. Roning and M. Riech, "Registration of nevi in successive skin images for early detection of melanoma," *Pattern Recognit.*, vol. 1, pp. 352–357, 1998.
- [22] C. Anagnostopoulos, D. Vergados, and P. Mintzias, "Image registration of follow-up examinations in digital dermoscopy," in *Proc. IEEE Int. Conf. Bioinform. Bioeng.*, 2013, pp. 1–4.
- [23] D. A. Perednia and R. G. White, "Automatic registration of multiple skin lesions by use of point pattern matching," *Comput. Med. Imag. Graph.*, vol. 16, no. 3, pp. 205–216, 1992.
- [24] I. Maglogiannis, "Automated segmentation and registration of dermatological images," *J. Math. Model. Algorithms*, vol. 2, no. 3, pp. 277–294, 2003.
- [25] F. Nachbar *et al.*, "The ABCD rule of dermatoscopy," *Amer. Acad. Dermatol.*, vol. 30, no. 4, pp. 551–559, 1994.
- [26] H. Tsao *et al.*, "Early detection of melanoma: Reviewing the ABCDEs," *Amer. Acad. Dermatol.*, vol. 72, no. 4, pp. 717–723, 2015.
- [27] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.
- [28] B. Li, C. Shen, Y. Dai, A. van den Hengel, and M. He, "Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1119–1127.
- [29] A. C. Jalba, A. Sobiecki, and A. C. Telea, "An unified multiscale framework for planar, surface, and curve skeletonization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 129–141, Jan. 2016.
- [30] A. C. Jalba, A. Sobiecki, and A. C. Telea, "Salient object detection via structured matrix decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 818–832, Jan. 2017.
- [31] Z. Hao, Q. Wang, H. Ren, K. Xu, Y. K. Seong, and J. Kim, "Blur image segmentation using iterative super-pixels grouping method," in *Proc. IEEE Int. Conf. Mach. Learn. Cybern.*, 2013, vol. 3, pp. 1161–1167.
- [32] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [33] J. Yao *et al.*, *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*. Berlin, Germany: Springer-Verlag, 2015.
- [34] J. I. Shiyong *et al.*, "A new multistage medical segmentation method based on superpixel and fuzzy clustering," *Comput. Math. Methods Med.*, vol. 2014, Art. no. 747549.
- [35] E. Ahn *et al.*, "Saliency-based lesion segmentation via background detection in dermoscopic images," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 6, pp. 1685–1693, Nov. 2017.
- [36] Z. Hao, Q. Wang, H. Ren, K. Xu, Y. K. Seong, and J. Kim, "Multiscale superpixel classification for tumor segmentation in breast ultrasound images," in *Proc. IEEE Int. Conf. Image Process.*, 2012, pp. 2817–2820.
- [37] C. L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentation," *Int. J. Comput. Vis.*, vol. 75, pp. 49–65, 2007.
- [38] S. Krig, "Interest point detector and feature descriptor survey," *Computer Vision Metrics*. Cham, Switzerland: Springer-Verlag, 2016, pp. 187–246.
- [39] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [40] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79–116, 1998.
- [41] F. Navarro *et al.*, "SP-SIFT: Enhancing SIFT discrimination via superpixel-based foreground-background segregation," *Electron. Lett.*, vol. 50, no. 4, pp. 272–274, 2014.
- [42] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [43] D. C. W. Pao, H. F. Li, and R. Jayakumar, "Shapes recognition using the straight line Hough transform: Theory and generalization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 11, pp. 1076–1089, Nov. 1992.
- [44] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 404–417.
- [45] S. W. Menzies *et al.*, "Short-term digital surface microscopic monitoring of atypical or changing melanocytic lesions," *Archives Dermatol.*, vol. 137, no. 12, pp. 1583–1589, 2001.