Skin Lesion Classification using Bag-of-3D-Features

Pedro M. M. Pereira*†, Lucas A. Thomaz†‡, Luis M. N. Tavora‡, Pedro A. A. Assuncao†‡,
Rui Fonseca-Pinto†‡, Rui Pedro Paiva*, and Sergio M. M. Faria†‡
*University of Coimbra, CISUC, DEI, Coimbra, Portugal
†Instituto de Telecomunicações, Leiria, Portugal
‡Polytechnic of Leiria, Leiria, Portugal

***Abstract*—Computer-aided diagnostic has become a thriving research area in recent years, namely on the identification of skin lesions such as melanoma. This work presents a novel approach to this field by exploiting the 3D characteristics of the skin lesion surface, advancing beyond common features such as, shape, colour, and texture, extracted from dermoscopic RGB images. To this end, a relevant set of features was investigated to obtain 3D skin lesion characteristics from images with depth information. These features were used to train a Bag-of-Features model to distinguish between malignant and benign lesions, also discriminating melanoma from all other lesion types. Despite the large class imbalance, often present in medical image datasets, the feature set achieved a top accuracy of 73.08%, comprising 75.00% sensitivity and 66.67% specificity when classifying between malignant and benign lesions, and 88.46% accuracy (100.00% sensitivity and 86.96% specificity) when discriminating melanoma from all other lesion images, using only depth information. The achieved experimental results indicate the existence of relevant discriminative characteristics in the 3D surface of skin lesions which allow the improvement of existing classification methods based on 2D image characteristics only.**

***Index Terms*—Medical Image Analysis, 3D Features, Classification, Melanoma, Skin Lesion**

I. INTRODUCTION

The importance of computational methods for skin lesion classification arises from the fact that one of the most lethal types of skin cancers, the melanoma, is developed from pigmented melanocytes, which are becoming increasingly common in the global population [1], and yet they are hard to distinguish from other types of benign lesions. Computer vision techniques to automatically identify melanoma have been under study for decades and automatic techniques for detection and classification are becoming increasingly useful to assist dermatologists [2]. In the pursuit of a solution, current systems tend to use the same type of information that is used by dermatology experts, dermoscopic images (2D). However, other image modalities or data dimensions exist which are fairly unexplored. One of these modalities is 3D imaging (e.g., stereo), which has already proven to enhance skin lesion classification performances due to the added depth information [3], [4]. In this work, only the 3D information is used for feature extraction and classification. Such information is computed from dense light fields captured by a light field camera, which provides multiple views of the visual scene with small disparities.

This work was supported by the Fundação para a Ciência e a Tecnologia (FCT), Portugal, under PhD Grant SFRH/BD/128669/2017, Programa Operacional Regional do Centro, project PlenoISLA POCI-01-0145-FEDER-028325 and by FCT/MCTES through national funds and when applicable co-funded by EU funds under the project UIDB/EEA/50008/2020.

Various feature extraction methods resorting to 3D information are already described in the literature, as presented in Section II. From these, a total of 11 were selected to be used in this work to either characterise the underlying 3D image globally or characterise it in key local regions. The adequate selection of these regions is also a research topic, therefore two different methods were studied in this work.

The main contribution of this paper is to demonstrate that 3D information from skin lesions contains relevant discriminative features capable of providing high classification precision of melanoma versus nevus. Such third dimension, that is beyond conventional colour, texture, and shape, proves to be beneficial for the classification process. This work does not aim to use RGB information nor improve existing algorithms, it is exclusively to evidence that skin surface topology has potential discriminative information.

The remainder of the paper is organised as follows: Section II presents the current state of the art, including previous works that inspired this work. Section III provides brief descriptions of feature extraction and keypoint detection methods relevant to this research. Section IV presents the proposed approach, describing feature extraction and relevant classification details. Section V presents and discusses the attained results and Section VI highlights the conclusions and future work.

II. BACKGROUND

Most available skin lesion datasets used in related literature include a single modality – 2D dermoscopic images. Although significant performances have already been achieved using these single modality datasets [5], the low granularity of the information might still pose limitations to the classification problem. Alternative modalities have already shown to be efficient in identifying the type of skin lesion using, for example, 3D stereo imaging techniques [3], [4]. However, the research on 3D features of skin surface in either melanoma identification or any other related studies involving skin lesions is very scarce. Nevertheless, recent studies indicates that depth information can be useful and methods to synthetise 3D information have been investigated in 2D datasets, such as [6]. Real 3D information of skin lesions was also made available through a dataset named Skin Lesion Light-fields (SKINL2) [7].

Since Bag-of-Features (BoF) models [8], [9] were proposed for skin lesion classification in 2008 [10], several research

2021 Telecoms Conference (ConfTELE) | 978-1-6654-1588-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/CONFTELE50222.2021.9435509

978-1-6654-1588-0/21/$31.00 ©2021 IEEE

works have been published resorting to it [11]. In image classification, a bag of features is a vector of occurrence counts of a dictionary of local image features, which can also be understood as a histogram over extracted image features. This type of structure can be employed for classification resorting to a Support Vector Machine (SVM) classifier. The image features that compose the BoF models are designed for object detection, classification, or retrieval generally fall under two base types: signature or histogram. Signature-based feature extractors aim to register specific object characteristics or attributes capable of providing discrimination against other objects or scenes – effectively, a signature of a set is a lossless representation of its histogram. Examples of signature based features extractors are: Normal Aligned Radial Features (NARF) [12], Radius-based Surface Descriptor (RSD) [13], [14], Global RSD (GRSD) [15], and Principal Curvatures (PC) [16]. In contrast, histogram-based extractors aim to produce a summarised representation of the underlying data – typically, the presence of a set of features and their occurrence count. Examples of histogram based features extractors are: Rotation Invariant Feature Transform (RIFT) [17], Point Feature Histogram (PFH) [18], Fast PFH (FPFH) [19], [20], Signature of Histograms of OrienTations (SHOT) [21], [22], Ensemble of Shape Functions (ESF) [23], 3D Shape Context (SC3D) [24], and Unique Shape Context (USC) [25]. More details on these features ([12]–[25]) are described in Section III.

In some cases, having a large number of features can be a problem – e.g., when several features are extracted but their relevance for the intended solution is unknown, or when there are insufficient data samples. A useful method to reduce the number of features, by selecting the most meaningful ones, is the Neighborhood Component Analysis (NCA) [26]. NCA is a non-parametric algorithm that enables feature selection with the goal of maximising prediction accuracy of regression and classification algorithms.

III. Keypoint detectors and descriptors

As mentioned above, BoF classification operates over extracted image features. Some of these features may require additional methods (hereinafter "keypoint detector") to predetermine keypoint locations to operate on. Examples of keypoint detectors and relevant feature extractors for 3D characterisation are detailed in the following:

- In [12], the authors define the NARF, which comprises two distinct algorithms: a keypoint detector and a feature extractor that operates on found keypoints. The keypoint detector has two major characteristics. Firstly, keypoints are extracted in areas where the direct underlying surface is stable and the neighbourhood contains major surface changes. The resulting keypoints are located in the local environment of significant geometric structures and not directly on them. Secondly, NARF takes object borders into account, which arise from view dependent non-continuous transitions from the foreground to the background. Thus, the silhouette of an object has a strong influence on the resulting keypoints. The NARF keypoint detector pipeline is as follows: *(i)* transform point cloud to range image; *(ii)* find object borders; *(iii)* compute normals to border points; *(iv)* compute principal curvature for non-border points; *(v)* compute interest value for all points; *(vi)* isolate keypoints. Having found areas of interest, the NARF feature extractor can now take place. The feature descriptor is computed by defining a normally aligned range value patch around the feature point, computed by constructing a local coordinate system, where the observer looks at the point along the normal. At this point, a star-shaped pattern is projected into the patch (where each beam corresponds to a value in the final descriptor) capturing how much the pixels under the beam change. Then, a unique orientation is extracted from the projection and the values are shifted accordingly, to make this rotation invariant.
- In [27], the author defines the Intrinsic Shape Signatures (ISS) keypoint detector, which employs a saliency measure based on the eigenvalue decomposition of a scatter matrix of the points belonging to a support value. These points are only retained if the ratio between two successive eigenvalues is below a predefined threshold. Their saliency is determined by the magnitude of the smallest eigenvalue, in order to only include points with large variations along each principal direction. The rationale behind this pruning stage is that points exhibiting a similar spread along the principal directions (where a repeatable canonical reference frame cannot be established) should be avoided because a subsequent description stage would hardly turn out effective. Afterwards, a point will be considered a keypoint if it has the maximum saliency value on a given neighbourhood. Contrary to the NARF detector, the ISS is much more selective and inherently produces less keypoints, reducing the computation time.
- In [13], [14], authors define the Radius-based Surface Descriptor (RSD) as a descriptor that depicts the geometric property of a point by estimating the radial relation with its neighbour points. First the radius is modelled as a relation between distance of two points and the angle between their normals. Then, the maximum radius and minimum radius are recorded as the final features for each point.
- In [15], the RSD extractor is extended to the Global RSD (GRSD), which computes a global histogram for the whole point cloud. First, the input point cloud is voxelised and the RSD descriptor is generated for every voxel neighbourhood. Then, voxel surfaces are categorised into six possible surfaces based on a set of defined rules using the two RSD features. After categorising all voxels, the GRSD histogram relies on the number of transitions between all of these local categories, which results in 21 dimensions/features. GRSD allows the use of depth images with or without colour information.
- In [16], authors define the Principal Curvatures (PC), which returns the eigenvector of the largest eigenvalue along with both the largest and the smallest eigenvalues after performing a Principal Components Analysis on the point normals of a surface patch (in the tangent plane of the

given point normal).

- In [17], the Rotation Invariant Feature Transform (RIFT) is defined such that, given a point, it extracts a sparse set of affine covariant elliptical regions of the surrounding texture using the Harris affine or Laplacian blob detectors, which detect complementary types of structures, and normalise each elliptical region into a unit circle to reduce the affine ambiguity to a rotational one. Then, the method divides the circular normalised patch into four concentric rings with equal width and compute a gradient orientation histogram with eight orientations/bins within each ring. This results in a descriptor of 32 features that is later adjusted for rotational invariance by the radial outward direction at each point.
- In [18], a method named Point Feature Histogram (PFH) that encodes the geometric properties of the k-nearest-neighbours of a point is defined, by using the average curvature of the multidimensional histogram around such point. This is done by calculating, for each pair of points, the difference of three angular variables (obtained from a Darboux frame where the third angular variable is normal to the point's plane) and their euclidean distance. Finally a histogram is created with the 4 variables along each computed pair.
- In [19], [20], a variant of PFH, the Fast Point Feature Histogram (FPFH), is proposed as a computational simplification of PFH. In comparison, first, for each point, FPFH uses a method similar to PFH to calculate the three angular variables and obtain a simplified PFH. Then, a weighted neighbouring pairing is used to calculate the final value of the histogram, where the weights depend on the centre point and a neighbour point at a given distance metric space.
- In [21], [22], a method defined as the Signature of Histograms of OrienTations (SHOT) is proposed, based on disambiguated eigenvalue decomposition of the covariance matrix of points within the neighbourhood region, where an isotropic spherical grid defines the signature structure. These locations produce local histograms by counting the number of points within a region of the spherical grid. The juxtaposing of all local histograms with quadrilinear interpolation generates the final collection of features.
- In [23], authors define the Ensemble of Shape Functions (ESF), which comprises ten 64-sized histograms: three angle related histograms, three area related histograms, three distance related histograms, and one histogram of distance-ratio. The first nine histograms are created by, respectively, classifying an angle formed by randomly sampled three points, the area created by such three points, and a shape function. While the last is built on the paring-lines generated during the shape function execution.
- In [24], authors define the 3D Shape Context (SC3D) as a descriptor that captures the local shape of a point cloud at a centre point using the distribution of points in a spherical support. Within this support, a set of bins is formed by equally dividing the azimuth and elevation, and logarithmically spacing the radial dimension. Then, the final descriptor is computed as the weighted sum of the number of points falling into bins.
- In [25], authors define the Unique Shape Context (USC) as an improvement over the SC3D descriptor by adding a unique and unambiguous local reference frame, with the purpose of avoiding computation of multiple features at each keypoint. Given a query point and its spherical support region, a weighted covariance matrix is defined so that three unit vectors of a local reference frame can be computed from the Eigen Vector Decomposition of this matrix. The eigenvectors corresponding to the maximum and minimum eigenvalues are reoriented in order to match the majority of the vectors they depicted, while the sign of the third eigenvector is determined by the cross product. Once the local reference frame is built, the subsequent steps are analogous to those in SC3D.

IV. Proposed Classification Experiment

The main goal of the work described here is to perform the classification of malignant skin lesions based on 3D surface information. To this end, the utilised methodology comprises a BoF approach, as in [8]–[10], with a dataset holdout of 30% on the SKINL2 dataset where, as a pre-processing stage, pixel values in the RGB channels of all images were replaced with zeros (since some of the selected feature extractors also consider colour information). This means the colour information is not used, only the depth. The following subsections provide added details to the pipeline. Section IV-A details about the selected keypoint detectors and feature extractors. Section IV-B provides some information about the BoF model.

A. Features

A total of 5215 features were extracted from each image using 11 features extractors. These extractors were selected based on the relevance of their characteristics for the input signal (3D information). RIFT (32 features) was selected because it provides invariant to illumination, viewpoint, scale, and rotation. Like RIFT, NARF (42 features) and PFH/FPFH (125/33 features) also possess some of these characteristics, PFH/FPFH, in particular, provides robustness against outliers and noise. Other features extractors as SHOT (361 features), SC3D (1989 features), and USC (1969 features) also provide robustness against noise. Additionally, both SHOT and USC are reported to provide uniqueness amongst detection, as well as unambiguous representations. Finally, ESF (640 features), PC (5 features), RSD (2 features), and GRSD (21 features) were selected for being descriptive, simple, and intuitive shape descriptions. ESF has proven to be efficient and expressive, while GRSD adds expressiveness to the simple RSD by partitioning the image point cloud into several voxel-surfaces of understandable shapes.

Apart from ESF and GRSD, the other feature extractors operate on specified image keypoints, which must be predetermined. In order to provide such set of locations, two keypoint detectors were selected: NARF and ISS. The NARF detector

seems specially suited for skin lesion imagery since it selects locations of high surface changes and takes object borders into account, as in skin to lesion borders, which have already been noticed to have relevant information [28]). Then, ISS is also selected because, like NARF, it produces keypoints which tend to be at saliency regions, like the lesion border or texture-full regions inside the lesion, but in a more selective manner (outputting less keypoints).

B. BoF model

The dataset images were divided into a training and a testing set, the former with 70% and the latter with 30%. From the training set keypoints are extracted, producing 5215 features each, and a SVM model is trained on the histograms produced after applying k-means clustering to those features. A SVM model is selected for this work because the mentioned dataset provides fewer images than typically necessary for Deep Learning approaches. When building the BoF model, for classification of Malignant versus Benign lesions or for Melanoma versus All other lesions, the used SVM classifier is of polynomial kernel of second order and has box constraint of 1. Since all experiments are defined for binary classification, the SVM solver is the Iterative Single Data Algorithm, which minimises by a series of one-point minimisations and does not respect the linear constraint nor explicitly includes the bias term [29].

Because some of the features might not contribute for the adequate label separation, or might effectively injure the model's capability, during the classification training process the feature selection is also performed with NCA (on the targeted 70%) before training the BoF model. Fitting of the NCA model is done with all training samples and using a stochastic gradient descent solver. The NCA algorithm is susceptible to overfitting but possesses a parameter to prevent it through regularisation. This value is fine tuned via grid-search in the range of $[0; 0.003]$, where 20 equidistant grid-points are selected. At the end, the NCA model provides a relevance-weight for each of the 5215 features. In this experiment, only those with a relevance superior to $0.02 * \max(1, \max(f_w))$ are selected, where f_w is a vector with all provided NCA feature weights.

V. EXPERIMENTAL ASSESSMENT

The proposed pipeline was applied to the publicly available SKINL2 dataset [7]. The skin lesion light-fields were captured at a hospital facility (Centro Hospitalar de Leiria, Portugal), with a Raytrix R42 camera, from patients previously screened by a medical doctor during dermatology clinical appointments. Each image has 3858×2682 pixels per RGB channel, as well as the relative depth of each pixel. The skin lesions were classified and organised based on clinician diagnosis according to ICD10 (International Classification of Diseases) and on histopathological analysis. Procedures related to the image acquisition, storage, and publication were evaluated and approved by a health ethics committee. The procedure and purpose of the study was explained to all volunteers, who also signed an informed consent form. Particularly in this work, the second version of this dataset was used [30], due to its increase in lens magnification of $\approx 30\%$ (which means more detail) in comparison to its first version. This dataset currently comprises 19 malignant lesion images (9 melanomas, 9 basal cell carcinomas, and 1 squamous cell carcinoma) and 66 benign lesion images (32 nevi, 13 angiomas, and 21 seborrheic keratoses), which undergo the pre-processing, feature extraction and classification processes, described in Section IV.

TABLE I: Overall Top Results.

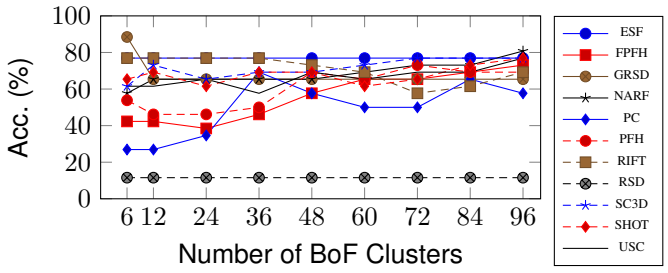
Experiment	Detector	#Clusters	Acc.	SEN	SPE	BAC
Ma vs Be	NARF	84	73.08	75.00	66.67	70.83
		96	80.77	100.00	15.00	57.50
Me vs All	NARF	48	84.62	66.67	86.96	76.81
		96	84.62	66.67	86.96	76.81
	ISS	96	84.62	66.67	86.96	76.81

The results obtained from these assessments are recorded in terms of percentage of classification accuracy (*Acc.*), specificity (*SPE*), and sensitivity (*SEN*), similarly to [4]–[6], [10], [11], [28]. In addition, because this is an unbalanced problem, the named balanced-accuracy (*BAC*) (1) is also used (as in [11]). It corresponds to the average value between sensitivity and specificity.

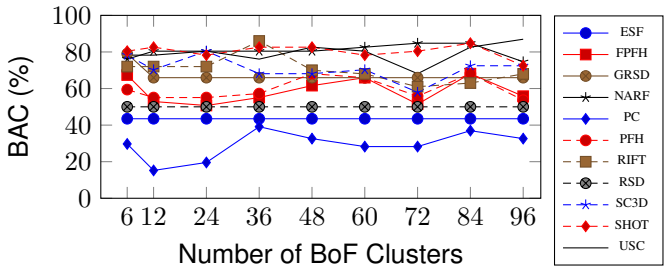
$$BAC = \frac{SEN + SPE}{2} \tag{1}$$

The main experimental results are shown in Table I and the individual behaviour of the different extracted features is plotted in Fig. 1. Table II adds to Table I with results resorting to the features selected after running the NCA algorithm. Only best *BAC* results are shown in the tables. In these tables, results are shown without background shading while values above 75% are highlighted in grayscale towards 100%. Given the available dataset samples previous described in this section, in these tables the column "Experiment" indicates the classification objectives, being either "Ma vs Be" for malignant versus benign lesions or "Me vs All" for melanoma versus all other skin lesion types. Additionally, as mentioned in Section IV-A, a keypoint detector is necessary for most of the feature extractors, therefore column "Detector" is present to indicate which of the two selected keypoint extractors was used. As the BoF model pipeline uses a k-means clustering algorithm, column "#Clusters" expresses the number of specified clusters. In this work the number of clusters was defined as: either 6 or multiples of 12 up to 96, in a total of nine variations, as represented by x-axis of Fig. 1.

From Table I, the highest accuracy (*Acc.*) result obtained for detecting malignant skin lesions is 80.77% (when the BoF pipeline uses 96 histogram bins for classification). Although this is not the best overall result, some clinicians find it appealing as it presents 100% *SEN*, meaning that no life-threatening condition goes unchecked. It is important to notice that, in this case, the *SPE* metric indicates that benign lesions are correctly classified only 15.00% of the time, meaning that 83.33% of the benign lesions are incorrectly labelled as malig-



(a) Accuracy



(b) Balanced-Accuracy

Fig. 1: Performance of each feature extractor for Me vs All classification problem: (a) Acc. metric and (b) BAC metric.

nant instead of benign. A more balanced solution is achieved with 73.08% accuracy when merging some of the data points, by using less clusters (84). This solution comprises 75.00% *SEN* and 66.67% *SPE*, meaning that 25% of the malignant lesions pass as benign but only 33.33% of the benign lesions get get classified as malignant, in comparison to the previous 83.33%. Focusing on the melanoma lesion type, a higher accuracy result of 84.62% is achieved when performing direct comparison between melanoma and the remaining skin lesion images. In this case, the best results achieved using the NARF detector are also attainable using the ISS detector. However, balanced *SEN* and *SPE* results show that: the algorithm is only capable of correctly classify melanomas 66.67% of the time (*SEN*), while the benign lesions are correctly classified 86.96% of the time (*SPE*).

The achieved balance of the metrics shows that some information exists in the 3D surface that enables a level of discrimination between skin lesion types. By observing the individual behaviour of the different feature extractors in Fig. 1, it is possible to infer their contributions towards the current melanoma classification results shown in Table I. Such behaviour is not uniform across all features extractors, but from the accuracy metric in Fig. 1a it can be seen that there is a trend to provide superior accuracy results as the number of clusters increases, although not all feature extractors follow this rule. An exception to this trend occurs, for instance, for RSD features, which provide a constant 11.54% accuracy, resulting from the classification of every sample as melanoma. This also means that the *BAC* metric for the RSD, in Fig. 1b, is 50% $((100+0)/2)$. Another example of a non-discriminative set of features is the PC, which always presents a *BAC* performance lower than RSD, e.g. 32.61% for 96 Clusters, despite presenting a higher accuracy. A possible reason for this type of contradictory results is the use of a unbalanced dataset. RSD always labels samples as melanoma, the smaller class (lower accuracy), while PC mostly labels samples as nevus, the larger class (higher accuracy). Apart from the mentioned outliers, what stands out the most in Fig. 1b is that several individual feature extractor results (26, at different cluster settings) achieve *BAC* performances that are superior to the recorded 76.81% in Table I, which are only achievable with balanced *SEN* and *SPE* settings, thus also generating high accuracy values. In particular, results above 80% are attained when using either NARF, RIFT, SC3D, SHOT, or USC. Specifically, USC and RIFT are able to reach a top performance of 86.96% *BAC*.

TABLE II: Overall Top Results after NCA.

Experiment	Detector	#Clusters	Acc.	SEN	SPE	BAC
Ma vs Be	NARF	6	61.54	100.00	50.00	75.00
		84	69.23	83.33	65.00	74.17
Me vs All	NARF	24	84.62	66.67	86.96	76.81
		48	88.46	100.00	86.96	93.48

With the previously mentioned insights, it becomes clear that the BoF model is not able to withstand the presence of non-discriminative features and performs poorly in comparison to use only a feature extractor's individual-set of features. Nevertheless, the combination of various subsets of different feature extractors could still yield even higher performance results. For this reason, the NCA feature selector is introduced in the experimental setup. Results obtained after applying the NCA selection are depicted in Table II. Only the experiments with the highest performing *BAC* are shown in the table (changing the selected "#Clusters" column from Table I to Table II). When applying NCA to the Malignant versus Benign problem, 21 features are selected and the performance of the *BAC* metric increases from the previous 70.83% to 75.00%. Also, when applying NCA to the Melanoma versus All other lesions problem, fewer (14) features are selected, but the *BAC* metric achieves the best performance of 93.48% (from the previous 76.81% in Table I). Independently of the problem, the best training results were achieved with a lambda of 0.0028. In comparison to the results presented in Table I, the NCA selected feature subset provides significant improvements (in Table II). With 48 clusters, the BoF model achieves 88.46% accuracy by only using depth-based features, in the Melanoma versus All other lesions problem. In addition, the generated model presents the capability of correctly identifying all melanoma samples (100.00% *SEN*), while only incorrectly labelling 13.04% of benign lesions as malignant ones (86.96% *SPE*). As expected, these results are far superior than using an individual set of features with only one feature extractor, e.g. 93.48% *BAC* for NCA, in Table II, against 86.96% *BAC* for USC at 96 Cluster, in Fig. 1b.

VI. CONCLUSIONS AND FUTURE WORK

The pursuit of a solution to automatically identify melanoma or malignant lesions has been under research for many years. Currently, automatic techniques for detection and classification are becoming increasingly useful. However, classification (or discrimination) of melanoma versus nevus still remains difficult to achieve, due to its similarity at an early stage of the lesion development. A reliable solution might depend on the use of new acquisition modalities instead of the widely available, and utilised, 2D dermoscopic images, which may introduce new fairly unexploited dimensions.

The main contribution of this work is exploitation of depth information from light fields images for classification of skin lesions. This newly introduced type of 3D data was specifically acquired for this purpose and has shown the ability to provide rich information for image classification. There are already several methods for classification using 3D surface information in the literature, even if originally developed for other purposes. These works address the definition of features extraction methods and classification models, as is the Bag-of-Features model for classification of the image feature-collections.

As previously exposed, classification between benign and malignant lesions achieved 75.00% *BAC*, comprising 61.54% accuracy, 100.00% *SEN*, and 50.00% *SPE*. In a more explicit setting, discrimination of melanomas against all other available skin lesions was achieved with 88.46% accuracy, 100.00% *SEN*, and 86.96% *SPE*, with a *BAC* of 93.48%. These results evidence the usefulness of unexploited 3D lesion surface information in the classification process of skin lesions.

The insight drawn from the experiments in this paper may foster further research that takes advantage of all the information provided by the light-field cameras, namely embracing depth information with texture information (2D) to improve lesion discrimination algorithms.

REFERENCES

[1] H. Kaufman, *The melanoma book: a complete guide to prevention and treatment*, Gotham Books, New York, NY, USA, 1st edition, May 2005.

[2] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artif. Intell. in Medicine*, vol. 56, no. 2, pp. 69–90, Oct. 2012.

[3] S. McDonagh, R. Fisher, and J. Rees, "Using 3D information for classification of non-melanoma skin lesions," in *Med. Image Understanding and Anal.*, Dundee, United Kingdom, July 2008, pp. 164–168.

[4] L. Smith, M. Smith, A. Farooq, J. Sun, Y. Ding, and R. Warr, "Machine vision 3D skin texture analysis for detection of melanoma," *Sensor Rev.*, pp. 111–119, Mar. 2011.

[5] S. Pathan, K. Prabhu, and P. Siddalingaswamy, "Techniques and algorithms for computer aided diagnosis of pigmented skin lesions - A review," *Biomed. Signal Process. and Control*, vol. 39, pp. 237–262, Jan. 2018.

[6] T. Satheesha, D. Satyanarayana, M. Prasad, and K. Dhruve, "Melanoma is skin deep: a 3D reconstruction technique for computerized dermoscopic skin lesion classification," *IEEE J. Transl. Eng. Health Med.*, vol. 5, pp. 1–17, Jan. 2017.

[7] S. Faria, J. Filipe, P. Pereira, L. Tavora, P. Assuncao, M. Santos, R. Fonseca-Pinto, F. Santiago, V. Dominguez, and M. Henrique, "Light field image dataset of skin lesions," in *Annu. Int. Conf. of the IEEE Eng. in Medicine and Biol. Soc.*, Berlin, Germany, July 2019, pp. 3905–3908.

[8] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *IEEE Int. Conf. on Comput. Vision*, Nice, France, Oct. 2003, pp. 1470–1477.

[9] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learn. in comput. vision*, Prague, Czech Republic, May 2004, pp. 1–2.

[10] N. Situ, X. Yuan, J. Chen, and G. Zouridakis, "Malignant melanoma detection by bag-of-features classification," in *Annu. Int. Conf. of the IEEE Eng. in Medicine and Biol. Soc.*, Vancouver, BC, Canada, Aug. 2008, pp. 3110–3113.

[11] K. Hu, X. Niu, S. Liu, Y. Zhang, C. Cao, F. Xiao, W. Yang, and X. Gao, "Classification of melanoma based on feature similarity measurement for codebook learning in the bag-of-features model," *Biomed. Signal Process. and Control*, vol. 51, pp. 200–209, May 2019.

[12] B. Steder, R. Rusu, K. Konolige, and W. Burgard, "Point feature extraction on 3D range scans taking into account object boundaries," in *IEEE Int. Conf. on Robot. and Automat.*, Shanghai, China, May 2011, pp. 2601–2608.

[13] Z. Marton, D. Pangercic, N. Blodow, J. Kleinehellefort, and M. Beetz, "General 3D modelling of novel objects from a single view," in *IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Taipei, Taiwan, Dec. 2010, pp. 3700–3705.

[14] Z. Marton, D. Pangercic, N. Blodow, and M. Beetz, "Combined 2D-3D categorization and classification for multimodal perception systems," *The Int. J. of Robot. Res.*, vol. 30, no. 11, pp. 1378–1402, Aug. 2011.

[15] A. Kanezaki, Z. Marton, D. Pangercic, T. Harada, Y. Kuniyoshi, and M. Beetz, "Voxelized shape and color histograms for RGB-D," in *IROS Workshop on Active Semantic Perception*, San Francisco, CA, USA, Sept. 2011, pp. 1–6.

[16] R. Rusu and S. Cousins, "3D is here: Point cloud library (pcl)," in *IEEE Int Conf on Robot. and Automat.*, Shanghai, China, May 2011, pp. 1–4.

[17] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 27, no. 8, pp. 1265–1278, June 2005.

[18] R. Rusu, N. Blodow, Z. Marton, and M. Beetz, "Aligning point cloud views using persistent feature histograms," in *IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Nice, France, Sept. 2008, pp. 3384–3391.

[19] R. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *IEEE Int. Conf. Robot and Automat.*, Kobe, Japan, May 2009, pp. 3212–3217.

[20] R. Rusu, A. Holzbach, N. Blodow, and M. Beetz, "Fast geometric point labeling using conditional random fields," in *IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, St. Louis, MO, USA, Oct. 2009, pp. 7–12.

[21] F. Tombari, S. Salti, and L. Stefano, "Unique signatures of histograms for local surface description," in *Comput. Vision - ECCV*, Heraklion, Crete, Greece, Sept. 2010, pp. 356–369.

[22] F. Tombari, S. Salti, and L. Stefano, "A combined texture-shape descriptor for enhanced 3D feature matching," in *IEEE Int. Conf. on Image Process.*, Brussels, Belgium, Dec. 2011, pp. 809–812.

[23] W. Wohlkinger and M. Vincze, "Ensemble of shape functions for 3D object classification," in *IEEE Int. Conf. on Robot. and Biomimetics*, Karon Beach, Phuket, Thailand, Dec. 2011, pp. 2987–2992.

[24] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *Comput. Vision - ECCV*, Prague, Czech Republic, May 2004, pp. 224–237.

[25] F. Tombari, S. Salti, and L. Stefano, "Unique shape context for 3D data description," in *ACM Workshop on 3D Object Retrieval*, New York, NY, USA, Oct. 2010, pp. 57–62.

[26] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *J. of Comput.*, vol. 7, no. 1, pp. 161–168, Jan. 2012.

[27] Y. Zhong, "Intrinsic shape signatures: A shape descriptor for 3D object recognition," in *IEEE Int. Conf. on Comput. Vision Workshops*, Kyoto, Japan, Oct. 2009, pp. 689–696.

[28] P. Pereira, R. Fonseca-Pinto, R. Paiva, P. Assuncao, L. Tavora, L. Thomaz, and S. Faria, "Skin lesion classification enhancement using border-line features - The melanoma vs nevus problem," *Biomed. Signal Process. and Control*, vol. 57, pp. 101765, Mar. 2020.

[29] V. Kecman, T. Huang, and M. Vogt, "Iterative single data algorithm for training kernel machines from huge data sets: Theory and performance," in *Support Vector Machines: Theory and Appl.*, pp. 255–274. Springer, Apr. 2005.

[30] S. Faria, M. Santos, P. Assuncao, L. Tavora, L. Thomaz, P. Pereira, R. Fonseca-Pinto, F. Santiago, V. Dominguez, and M. Henrique, "Dermatological imaging using a focused plenoptic camera: the SKINL2 light field dataset," in *Conf. on Telecommun.*, Lisbon, Portugal, June 2019.