# DERMOSCOPIC INTEREST POINT DETECTOR AND DESCRIPTOR

*Howard Zhou*[†*], *Mei Chen*[§], *James M. Rehg*[†]

[†]Georgia Institute of Technology, [§]Intel Research Pittsburgh

## ABSTRACT

Dermoscopy is an imaging technique dermatologists use to better visualize pigmented skin lesions (*PSLs*) and determine their malignancy. *Dermoscopic features* revealed by this technique have been shown to correlate with histopathology features, and are used as diagnosis indicators by many dermatologists. Hence, automated detection and classification of these features is the first step toward computer-aided diagnosis of melanoma in dermoscopy. In this paper, we present a novel scale- and rotation-invariant feature detector and descriptor specifically designed as a general visual vocabulary of dermoscopic features. We compare our feature detector and descriptor to the popular interest point detectors in the vision community, namely, SIFT, and a more recent fast variant, SURF. We demonstrate that our feature detector is more discriminative and reliable for dermoscopic features.

***Index Terms***— Computer-Assisted Image Interpretation, Dermoscopy, Feature detection, Pigmented Skin Lesion

## 1. INTRODUCTION

Skin cancer is the most common form of malignancy in humans. Among all its variations, melanoma is the leading cause of mortality. The incidence of melanoma is increasing at a rate greater than any other form of cancer in the United States [1]. However, when melanoma is caught in its early stage, it can often be cured with a simple excision. Dermoscopy is a non-invasive imaging technique to aid in melanoma detection. During image acquisition, the clinician places an incident light magnification system (a.k.a. dermatoscope) at the area of interest, frequently with a liquid contact medium applied at the skin-scope interface. This allows the incident light to penetrate the top layer of the skin tissue and permits a detailed examination of the pigmented structures beyond what would be visible to the naked eye. Studies have shown that using dermoscopy can improve the diagnostic accuracy of dermatologists by as much as 30% over clinical examination [1]. This improvement in diagnostic accuracy, however, is seen primarily when dermoscopy is used by a trained expert. Consequently, there has been increasing interest in computer-aided analysis of dermoscopy images.

Automated detection and classification of indicative dermoscopic features is a key step toward computer aided diagnosis. Previous work focused on specific types of dermoscopic features. Betta *et al.* [2] developed techniques to detect and analyze atypical pigmented network and vascular patterns. Grana *et al.* [3] detects curvilinear features to characterize network patterns. Tanaka *et al.* [4] applied texture analysis techniques to classify three common global patterns typical to melanocytic lesions. Iyatomi *et al.* [5] developed a procedure to classify parallel furrow and ridge patterns that are characteristic to melanocytic acral lesions. However, these methods do not generalize to the large number of features typically present in a dermoscopy dataset. A PSL typically has several dermoscopic features, and there are over one hundred of these features commonly seen in clinical studies. Upon close examination, most of them share low level image characteristics such as ridges, blobs, and streaks, yet each has a distinct spatial configuration and concentration of them. In this paper, we propose a feature detector designed to extract these low level building blocks (a.k.a. dermoscopic interest points) and a descriptor for constructing a general visual vocabulary for dermoscopic features. To the best of our knowledge, this is the first time the concept of interest points for dermoscopic features is introduced.

These dermoscopic interest points (*DIP*) can be used in various computer-aided PSL diagnosis applications. For instance, we can use DIP as a general vocabulary for dermoscopic feature extraction, matching, and classification. To detect DIP, candidate points are selected at distinctive locations such as corners, blobs, and streaks in the dermoscopy image. For an interest point detector, repeatability is its most important merit, i.e. whether it reliably finds the same interest points under different viewing conditions, or within different instances of the same type of lesion. Next, the neighborhood of every interest point is represented by a feature vector, an *interest point descriptor*. This representation not only has to be distinctive, but also robust to noise as well as geometric and photometric deformation. A number of detectors and descriptors have already been introduced in the computer vision literature (e.g. [6, 7, 8]). Recent work in this area has also proposed learning image descriptors from training samples [9, 10]. We have built our dermoscopic interest point detector and descriptor on insights gained from this previous work. The first issue for a local feature detector is the level of

[*]The first author performed the work at Intel Research Pittsburgh

invariance. The same feature can appear in different images and undergo both geometric and photometric deformations. Due to the imaging technique used, geometric deformations are mostly restricted to a plane; therefore, we focus on scale and rotation invariant detectors and descriptors.[1] We use a linear model with a scale factor and offset to model photometric deformations. Many existing interest point detectors select distinctive locations from corners, blobs, and junctions because of their robustness to various deformations. In dermoscopy images, due to the common presence of curvilinear features such as fibrillar pattern and radial streaming, we include locations where strong curvilinear features are present. Moreover, unlike previous descriptors which discard color information, we introduce a color component in our descriptor since color information is important for discriminating many dermoscopic features in PSLs. We compare our feature detector and descriptor to some of the most popular interest point detectors in computer vision. We demonstrate that DIP are discriminative and reliable for dermoscopic features.

## 2. DETECTOR

We select interest points at distinctive locations in the dermoscopy image, such as corners, blobs, junctions, and streaks. We want our interest point detector to be repeatable, i.e. it should reliably find the same interest points under different viewing conditions. In general, vision interest point detectors such as SIFT [7] and SURF [8] focus on corners and blobs because these image characteristics are robust to various image transformations. This applies to dermoscopy images as well; therefore, we start with detecting corners and blob structures. In addition, many dermoscopic features, such as the pigmented networks shown in Figure 2, have strong curvilinear components. In order to capture those features, we augment our detector with a component that specifically locates curvilinear structures.

### 2.1. Corners and blobs

We adopt the fast-Hessian detector proposed in Bay, *et al.* [8] to locate corner and blob structures. Given a dermoscopy image $J$, we first convert its pixels from *RGB* to *L*a*b** values since the *CIELAB* space is more perceptually uniform, and the $L$ channel roughly captures the luminous component of the image. We use the determinant of the Hessian matrix for selecting both the location and scale of each interest point. Given a point $\vec{x} = (x, y)$ in the intensity channel $L$, the Hessian matrix $\mathcal{H}(\vec{x}, \sigma)$ in $\vec{x}$ at scale $\sigma$ is defined as follows

$$\mathcal{H}(\vec{x}, \sigma) = \begin{bmatrix} L_{xx}(\vec{x}, \sigma) & L_{xy}(\vec{x}, \sigma) \\ L_{xy}(\vec{x}, \sigma) & L_{yy}(\vec{x}, \sigma) \end{bmatrix}, \qquad (1)$$

where $L_{xx}(\vec{x}, \sigma)$ is the image $I$ convoluted with the Gaussian second order derivative $\frac{\delta^2}{\delta x^2} g(\sigma)$, and likewise for $L_{xy}$ and $L_{yy}$. The detector approximates second order Gaussian derivatives with box filters (Fig. 1), denoted as $D_{xx}$, $D_{yy}$, and $D_{xy}$, which can be evaluated efficiently using integral images. We use $9 \times 9$ box filters at the lowest scale $\sigma_0 = 1.2$. The filter response is computed as $det(\mathcal{H}_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2$, where $0.9$ is used to balance the relative weights from the approximations. The responses are normalized with respect to the mask size to guarantee a constant Frobenius norm for any filter size. Scale space computation is also made easy by the use of box filters and integral images. We build each image layer, referred as $s = \sigma$, by filtering the original image with gradually increasing masks (e.g. $9 \times 9$, $15 \times 15$, $21 \times 21$) instead of iteratively downsampling the image. As a result, the same integral image can be reused and the speed is exactly the same for any filter size. The Gaussian derivatives scale along with the filters. For example, the $27 \times 27$ filter corresponds to $\sigma = 3 \times \sigma_0 = 3.6 = s$. After applying non-maximum suppression in a $3 \times 3 \times 3$ neighbourhood of the image layers, we interpolate the Hessian determinant in both scale and image space. We then threshold the interpolated maximum responses to exact candidate sites for corner and blob structures.

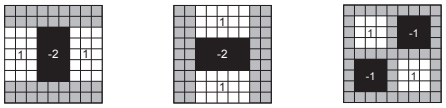


**Fig. 1**. Gaussian second order partial derivatives in $x$-, $y$-, and $xy$-direction approximated using box filters. Grey is zero.

### 2.2. Curvilinear structures

To locate strong curvilinear components, we apply Steger's line point detection algorithm [11] to the intensity channel $L$. Line points are points in an intensity image where the first directional derivative in the direction of the line vanishes, and the second directional derivative has a large absolute value. A point $\vec{x} = (x, y)$ is a line point if it satisfies $(tn_x, tn_y) \in [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$, where $(n_x, n_y)$ is the normalized eigenvector[2] that corresponds to the maximum absolute eigenvalue of the local Hessian matrix $\mathcal{H}(\vec{x}, \sigma)$, and $t$ is evaluated as follows:

$$t = -\frac{L_x(\sigma)n_x + L_y(\sigma)n_y}{L_{xx}(\sigma)n_x^2 + 2L_{xy}(\sigma)n_x n_y + L_{yy}(\sigma)n_y^2}, \qquad (2)$$

where $L_x$, $L_y$, $L_{xx}$, $L_{xy}$, $L_{yy}$ are partial derivatives of the image estimated by convolving the image with discrete two-dimensional Gaussian partial derivative kernels. The standard deviation $\sigma$ of these kernels is directly tied to the expected line width. Therefore, we apply the line detection algorithm at multiple scales (with different $\sigma$'s) to extract line segments within a certain width range. The saliency of a line point

[1] Deformations such as skew and anisotropic scaling are covered to some degree by the overall robustness of the descriptor.

[2] $(n_x, n_y)$ points in the direction perpendicular to the line direction at point $(x, y)$.

$(x, y)$, i.e. the absolute value of the second directional derivative along $(n_x, n_y)$, is proportional to its intensity. We find local maximum filter responses in both scale and image space to locate strong curvilinear structures.

## 3. DESCRIPTOR

The distinctive power of state-of-the-art interest point descriptors such as SIFT or SURF relies on a combination of approximately spatially localized information and the distribution of gradient-related features. Relative strengths and orientations are often used instead of absolute ones to reduce the effect of photometric changes. The proposed dermoscopic feature descriptor is based on similar properties, with the addition of linear and color components. We first identify a reproducible orientation based on local statistics calculated from a circular region around each interest point. We then construct a square region aligned to this orientation and extract a feature vector from it.

### 3.1. Orientation

Once the location of an interest point is determined, we identify a reproducible orientation in order to achieve rotation invariance. We first compute the Haar-wavelet responses in both $x$ and $y$ direction in a circular neighbourhood of radius $6s$ around the interest point, with $s$ being the scale of the interest point. The sampling step and the scale at which we compute wavelet responses are chosen as $s$, so the wavelets are large at high scales. Using the same integral images introduced in the last section, only six operations are needed to compute the wavelet response in both directions at any scale. The resulting responses, after being weighted by a Gaussian, are represented as vectors in a $2D$ space with the horizontal and vertical response strength as coordinate axes. The dominant orientation is estimated by summing up all the respective horizontal and vertical responses in a sliding window covering an angle of $\frac{\Pi}{3}$. The two summed responses then yield a new vector, and the orientation of the longest such vector is chosen as the orientation of the interest point.

### 3.2. Descriptor components

Once a reproducible orientation is identified, we construct a square region oriented along this direction, and centered around the interest point. The size of the square is $20s$ and it defines the context of our descriptor. After dividing the context region uniformly into $4 \times 4$ sub-regions, we compute four simple features at $5 \times 5$ regularly spaced sample points for each sub-region. We use $d_x$, $d_y$ to denote the Haar wavelet responses in horizontal and vertical directions (filter size $2s$). We also extract the absolute values of the responses, denoted by $|d_x|$ and $|d_y|$, to register the polarity of the intensity changes. After summing up these individual measurements within each sub-region, we obtain a four-dimensional descriptor vector $v$ with underlying intensity structure $v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$. Therefore, for each interest point, the resulting descriptor vector is of length 64 for all $4 \times 4$ sub-regions. The wavelet responses are already invariant to illumination changes, and we can further achieve contrast invariance by normalizing the descriptor into a unit vector.

Popular interest point descriptors, such as SIFT and SURF, only compute intensity statistics and discard color information. However, color is an important diagnostic cue in dermoscopy since most PSLs arise from pigmented skin cells (*melanocytes*). The appearance of melanocytes, which varies with depth, is vital to discriminating dermoscopic features. For instance, black globules have an identical gradient profile as the brown ones; however, they are very different dermoscopic features, and it is important to distinguish them. Accordingly, we include color statistics to augment our descriptor. For each region surrounding an interest point, we compute a coarse color histogram (6 bins in dimension) in the $\alpha$ and $\beta$ channels of the L*a*b* representation of the image. The resulting vector is of length 36, and it captures the color distribution of the region. After normalization, this color component is concatenated to the intensity component to form a DIP descriptor in $\mathbb{R}^{100}$.

## 4. VALIDATION AND RESULTS

We evaluate our representations on a dataset of 150 dermoscopy images containing common PSLs. At least one dermoscopic feature is present within each lesion boundary. The features are outlined and annotated by our collaborating dermatologists. For the detectors and descriptors used in comparison, SURF is based on the original implementation of the authors, and SIFT is from a relatively efficient implementation [12] based on the original publication. We first compare how sensitive these detectors are to dermoscopic features. We then check their repeatability on dermoscopy images undergoing common transformations.

For each image, all the detector responses within the lesion boundary are retrieved. Those points that land inside dermatologists' manual feature outlines are considered relevant. The starting threshold for each detector is set to a level low enough to generate a large number of responses. These responses at the lowest threshold are used as the relevant feature set for each detector. As we gradually increase the threshold, fewer responses are produced. We plot the precision-recall graph in Figure 3(a). Our experiment shows that the DIP detector consistently outperformed both the SURF and SIFT detector on dermoscopy images. Figure 2 shows an illustrative comparison between SURF and DIP detection results. The lesion in the images exhibits a common dermoscopy feature called pigmented networks. There are only a few corners and blobs strong enough to trigger SURF detector responses, and the majority of the interesting feature region is overlooked.[3]

[3]Similar results are observed when the SIFT detector is used. The image is not shown due to space limitation.

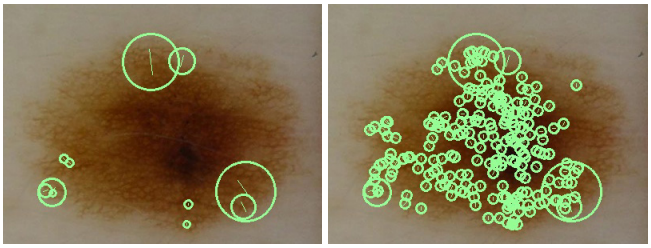

**Fig. 2**. Interest points detected by the SURF (left) and DIP (right) detectors.

Under similar settings, the DIP detector captures more interest points on the same lesion. This is true in general, and according to the precision-recall graph, a higher percentage of these DIP responses are relevant to dermoscopic features.

An effective interest point detector and descriptor should reliably find the same interest points under different viewing conditions. For example, reliable feature correspondences are needed in applications such as lesion registration and change detection for longitudinal studies. To demonstrate the repeatability of the DIP detector and descriptor, we set up the following experiment. We perform a set of scale, rotation, and lighting change operations on each image. The detector responses at each location before and after the change are matched; any inconsistency indicates a miss detection. When both responses are present (within a three pixel radius), we calculate the Euclidean distance between their descriptor vectors. If this distance is larger than a preset threshold, we consider the pair as a miss match. Figure 3(b), 3(c), and 3(d) shows the repeatability comparison results for decreasing light, scale, and rotation change, respectively. Although under similar settings, the DIP detector often extracts more interest points than SIFT and SURF, and the additional color components in its descriptor can increase matching difficulties, its repeatability, as evidenced in the graph, is on par with SURF, and slightly superior to SIFT. All the experiments were performed on a standard Windows PC (Pentium D, 3.20 GHz).

## 5. CONCLUSION

We presented a novel scale- and rotation-invariant feature detector and descriptor specifically designed for dermoscopic features. We have demonstrated that our detection-description scheme is more effective in retrieving dermoscopic interest points while achieving a comparable level of invariance to lighting, scale, and rotation changes when measured against the popular SIFT and SURF detectors and descriptors.

### 6. REFERENCES


[1] D.S. Rigel, R.J. Friedman, and A.W. Kopf, "The incidence of malignant melanoma in the united states: Issues as we approach the 21st century," *J Am Acad Dermatol.*, vol. 34, pp. 839–847, 1996.


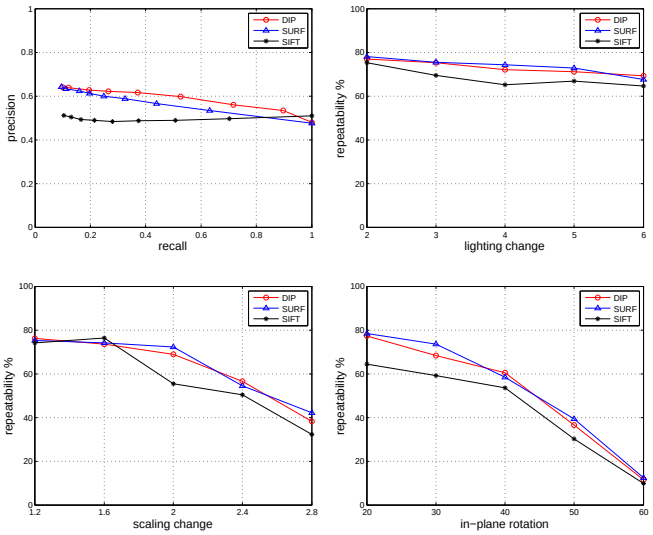


**Fig. 3**. (a), (b), (c), and (d) are from left to right and top to bottom. The precision-recall graph (a) demonstrates that the DIP detector is more sensitive to dermoscopic features. (b)-(d) show that DIP achieves a comparable level of invariance to lighting change, scale change, and in-plane rotation when measured against SIFT and SURF.


[2] G. Betta, G.D. Leo, G. Fabbrocini, A. Paolillo, and P. Sommella, "Dermoscopic image-analysis system: estimation of atypical pigment network and atypical vascular pattern," in *MEMEA '06: Proc. of the IEEE Intl. Workshop on Medical Measurement and Applications*, 2006, pp. 63–67.

[3] C. Grana, R. Cucchiara, G. Pellacani, and S. Seidenari, "Line detection and texture characterization of network patterns," in *Proc. of Intl. Conf. on Pattern Recognition (ICPR)*, 2006, pp. 275–278.

[4] T. Tanaka, S. Torii, I. Kabuta, K. Shimizu, and M. Tanaka, "Pattern classification of nevus with texture analysis," *IEEJ Trans. on Electrical and Electronic Eng.*, vol. 3, no. 1, pp. 143–150, 2008.

[5] H. Iyatomi, H. Oka, M.E. Celebi, K. Ogawa, G. Argenziano, H P. Soyer, H. Koga, T. Saida, K. Ohara, and M. Tanaka, "Computer-based classification of dermoscopy images of melanocytic lesions on acral volar skin," *J Invest Dermatol.*, vol. 128, no. 8, pp. 2049–2054, Aug 2007.

[6] T. Lindeberg, "Feature detection with automatic scale selection," *Intl. J. of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.

[7] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. J. of Computer Vision*, vol. 60, pp. 91–110, 2004.

[8] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," in *Eur. Conf. on Computer Vision (ECCV)*, 2006, pp. 404–417.

[9] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 9, pp. 1465–1479, 2006.

[10] S. A. Winder and M. Brown, "Learning local image descriptors," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.

[11] Carsten Steger, "An unbiased detector of curvilinear structures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 2, pp. 113–125, 1998.

[12] Andrea Vedaldi, "A lightweight c++ implementation of david lowe's scale invariant feature transforms," Available from: <http://www.cs.ucla.edu/ vedaldi/>.