Similar Image Retrieval for Dermoscopy Images using Interest Point Detection

by

Ardalan Benam

B.Sc., University of Tehran, 2014

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in the School of Computing Science Faculty of Applied Sciences

© Ardalan Benam 2017
SIMON FRASER UNIVERSITY
Spring 2017
Approval

Name: Ardalan Benam
Degree: Master of Science (Computing Science)
Title: Similar Image Retrieval for Dermoscopy Images using Interest Point Detection

Examinining Committee:

Chair: Brian Funt
Professor

Stella Atkins
Senior Supervisor
Professor

Mark Drew
Supervisor
Professor

Ze-Nian Li
Internal Examiner
Professor

Date Defended: March 28, 2017
Abstract

Providing physicians with a set of pathology-confirmed similar images to a new difficult case can efficiently assist towards a more confident diagnosis; this concept is called Content-Based Image Retrieval. We used SURF interest point detection to find and match similar dermoscopy images from a labelled dermoscopic image database. SURF automatically finds points of interest with the shape of blobs, dots. Haar - wavelet responses and local color histograms are locally extracted from each detected key point. The similarity of two images is decided by matching their key points and finding the Euclidean distance between them. We evaluated our system’s performance based on its ability for retrieving images with the same texture features and similar diagnosis. For query images containing a pigment network the precision with retrieval of 9 images, P(9), is 75%; for dots and globules, the precision P(9) is 80%. The precision P(9) for Melanoma diagnosis is 72%, which is acceptable for such systems.

Keywords: Dermoscopy; Content-based Image Retrieval; CBIR; Image retrieval; Interest Point Detection
Dedication

To my parents.
   You are the reason of whom I become today.
To my life long friend, my sister, Niloufar.
   Thank you for always being there right where I needed you.
Acknowledgements

I would like to thank all the people who contributed in some way to the work described in this thesis. First and foremost, I wish to express my deepest gratitude to my supervisor, Dr. Stella Atkins. She contributed to a rewarding graduate school experience by giving me intellectual freedom in my work, supporting my attendance in various conferences, engaging me in new ideas, and demanding a high quality of work in all my endeavours. Additionally, I want to thank Dr. Mark Drew for all of his support and feedback on my work. Moreover, I want to thank Dr. Maryam Sadeghi for all her supports in my graduate studies. For sure I wouldn’t be where I am if it were not because of her. I would also want to thank my parents and my sister for their support. I owe them everything. Finally, special thanks are reserved for my girlfriend, Gabriela Rodriguez, for her support, kindness and love.
# Table of Contents

Approval ii

Abstract iii

Dedication iv

Acknowledgements v

Table of Contents vi

List of Tables viii

List of Figures ix

1 Introduction 1

2 Background 3

  2.1 Biology of the skin ......................................................... 3
  
  2.1.1 Epidermis ................................................................. 3
  
  2.1.2 Dermis ................................................................. 3
  
  2.1.3 hypodermis ............................................................ 3

  2.2 Dermoscopy ................................................................. 4

  2.3 Content-based Image Retrieval ........................................ 6

  2.4 Interest Point Detection ................................................. 6

  2.4.1 Scale Invariant Feature Transform ................................... 8

  2.4.2 Speeded-Up Robust Features ......................................... 9

3 Method 13

  3.1 Dermoscopy Point Location Detection .................................. 13

  3.2 Descriptor Calculation .................................................. 13

  3.3 Distance Calculation and Matching ..................................... 15

4 Evaluation Methods 18

  4.1 Clinical Feature Retrieval ............................................... 18
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>CBIR evaluation results for PN detection using IAD Atlas data set (1011 images)</td>
<td>22</td>
</tr>
<tr>
<td>5.2</td>
<td>CBIR evaluation results for dots and globules detection using IAD Atlas data set (1011 images)</td>
<td>23</td>
</tr>
<tr>
<td>5.3</td>
<td>Accuracy results for diagnosis retrieval using IAD Atlas data set (1011 images)</td>
<td>23</td>
</tr>
<tr>
<td>5.4</td>
<td>Accuracy results for diagnosis retrieval using ISIC2017 and combined (ISIC2017 and IAD Atlas data set)</td>
<td>23</td>
</tr>
<tr>
<td>5.5</td>
<td>Comparison of performance of detecting images with pigments networks and dots and globules using different interest point detectors in CBIR system on the IAD Atlas data set of 1011 images</td>
<td>24</td>
</tr>
</tbody>
</table>
List of Figures

Figure 2.1  structure of the skin.[10] ............................................. 4
Figure 2.2  A lesion containing pigment network.[1] ......................... 5
Figure 2.3  A lesion containing dots highlighted by arrows. [20] ........... 5
Figure 2.4  A lesion containing streaks highlighted by arrows.[1] ........... 6
Figure 2.5  The left set of images are the different octaves of scale space. The right set of images are the difference in Gaussian which is an approximation of LoG operator.[18] ................................. 8
Figure 2.6  In order to calculate the key point descriptor at first it is needed to compute the gradient magnitude and orientation around the key point (left figure). After being weighted by the Gaussian window (the blue circle in the left figure), then the samples are accumulated in orientation histograms that summarize the content of 4 x 4 subregions.[18] ....................................................... 9
Figure 2.7  The first two pictures from the left demonstrate the second order partial derivative in y- and xy- direction. The second two pictures on the right are the approximated filters using box filters.[3] ....... 10
Figure 2.8  Area computation using integral images.[3] ......................... 10
Figure 2.9  Non-maximum suppression: each pixel is being compared with the surrounding pixels in scale-space.[12] ................................. 11
Figure 2.10 Orientation detection: as the circle segment slides around the origin the sum of all responses covered in the segment is calculated which is the vector arrow. The largest such vector determines the orientation of the key point.[12] ................................................... 12
Figure 2.11 Descriptor: the highlighted green square is one of the subregions of the 20s square at the key point. The blue circles inside the highlighted square indicates Haar-wavelets of size 2s.[12] ............. 12

Figure 3.1  General work flow of the designed CBIR system: First, Query image’s dermoscopy key points and their descriptors get calculated. Then, the set of key points of the query image is compared to each image in the data base and a distance for each pair is generated. Then the top most similar images is shown. ......................... 14
Figure 3.2 Examples of (a) dots and (b) pigment network detection using the proposed algorithm .......................... 15

Figure 3.3 Error in matching the keypoints when using the traditional approach. Most of the keypoints in the right image are matched to one or two keypoints in the left image. .......................... 16

Figure 3.4 Matching with new algorithm. Holes are matched to each other .......................... 17

Figure 4.1 An example of Melanoma lesion containing pigment network. The black arrows shows the visually obvious pigment network.[20] .......................... 19

Figure 4.2 An example of Cutaneous melanoma containing irregular brown dots and globules. The black arrow points to the brown dots.[2] .......................... 20

Figure 5.1 Pigment Network Retrieval: The image in the first row is the query image. Other images are ranked based on the order that they are retrieved. In this example, all of the retrieved images contain pigment network. .......................... 25

Figure 5.2 Dots and globules retrieval: The image in the first row is the query image. Images in the other rows are ranked based on the order that they are retrieved. .......................... 26

Figure 5.3 Diagnosis Retrieval: The image in the first row is the query image. Images in the other rows are ranked based on the order that they are retrieved. .......................... 27

Figure 6.1 Two lesions with PNs with different sizes. The red annotations in the right picture are the areas that keypoints are incorrectly chosen due to artifact and small number of pigment holes. .......................... 29

Figure 6.2 Two lesions with PNs of similar size. All of the holes are matched to each other. .......................... 30

Figure 6.3 Examples of cases that the CBIR system fails to detect the correct dermoscopic features .......................... 31

Figure 6.4 First 100 strongest interest points calculated by SURF. Green circles surround the interest points. .......................... 32

Figure 6.5 The effect of shaved hair in the CBIR system: The image in the first row is the query image. Other rows consist of retrieved images that are ranked based on their retrieval order. .......................... 33

Figure 6.6 First 100 strongest interest points detected by SURF on a dermatofibroma with a uniform structure. Most of the keypoints are detected outside of the lesion. The red circles represent interest points with pale centre and darker surrounding. The blue circles represent interest points with dark centre and pale surrounding. .......................... 34
Figure 6.7 Screen shots of the proposed educational CBIR app.
Chapter 1

Introduction

Skin cancer is one of the most common type of cancers and its worldwide mortality rate is increasing [16]. Malignant melanoma can be deadly; however, if it is detected in an early stage, it can be often cured. There are many initiatives for early detection including lesion screening, educational campaigns and computer aided diagnosis that are aimed to detect malignant melanoma earlier.

There are many types of skin lesions and diagnosing the exact nature of a particular skin lesion can be a difficult task for dermatologists. Dermoscopy is a non-invasive technique for imaging skin lesions using a magnifier and polarized light which aids diagnosis of skin lesion images, although expert knowledge is needed for diagnosis of malignant melanoma from dermoscopic images [31]. Therefore, a second diagnosis using Computer-Aided Diagnosis (CAD) is beneficial to the dermatologists and it can reduce the work load of the dermatologists [28].

Most of the contemporary CAD systems are focused on providing the physician a definitive second diagnosis of whether the skin lesion is cancerous or not. However, a study [9] shows that the second diagnosis is not always welcomed by physicians when it does not match their own initial diagnosis. In fact, the study showed that only 24% of physicians changed their decision when they encounter a second different diagnosis proposed by the CAD system.

Another approach using CAD systems is to assist the physicians to a precise diagnosis instead of giving a crisp second decision. Doi [7] believed that providing physicians with similar looking lesions with a confirmed pathology diagnosis would efficiently assist the physician to a more confident diagnostic decision. This concept of providing a source image and retrieving similar-looking images from a database is known as Content-Based Image Retrieval (CBIR) in computer vision. Retrieving similar looking cases can provide a diagnostic support environment rather than a single second diagnosis.

Chung et al. developed one of the very early basic concepts of image retrieval in dermoscopy. They designed a system allowing the user to query the database using feature attribute values - but the user had to input the feature values manually and then a web-
based data browser retrieved all of the images with those feature values [6]. Another image retrieval system was proposed by Rahman et al. based on the contents on the query. In this algorithm the lesion was segmented and dermoscopic features were extracted. Then the similarity between different images was found by comparing their feature vectors using the Euclidean and Bhattacharyya metrics [22]. Dorileo et al. introduced CBIR to dermatological lesions by using color histogram and texture features for similar-image retrieval [8]. A modern approach is using deep learning to classify images according to their features [21], but many thousands of labelled images would be required to train the system. For instance, an important new paper presents a machine learning system for diagnosing the melanagny of images. To train this system 129,450 clinical images including 2,032 different diagnosis with 757 different disease classes were used [11].

Furthermore, a closer inspection of the published literature in dermoscopy CBIR shows that they all suffer from not matching meaningful dermoscopic structures. In all of the proposed systems the traditional approach is taken, i.e. a feature vector is extracted from the query image and the most similar images according to a distance metric are retrieved. This approach is different from what humans do in order to determine similarity. Instead, a more meaningful approach can be taken, which is to match the high level characteristics such as blobs, dots, streaks etc. Locating and matching the similar looking dermoscopic interest points (DIP) will lead to a more meaningful image retrieval. The concept of DIP was first introduced by Zhou et al. in [32]. They proposed using blob detectors and a curvilinear structure detector for finding the meaningful points with important dermoscopic information.

This thesis extends Zhou’s work [32] and investigates the use of interest point detection in dermoscopy image retrieval. In chapter 2, we review skin biology, dermoscopy and interest points. Our method and modifications over standard interest point detection algorithms are described in chapter 3. In Chapter 4 we discuss methods for evaluating a dermoscopy CBIR system. Chapter 5 shows the results of our proposed system and we discuss the results in Chapter 6. Chapter 7 concludes this thesis with a brief summary and future work.
Chapter 2

Background

2.1 Biology of the skin

Skin is the largest organ of the body with the role of protecting the internal system from outside hazards. Moreover, maintaining body temperature, insulation and production of vitamin D are the other roles of this organ. Skin consists of 3 layers shown in figure 2.1: epidermis, dermis and hypodermis from the outer most layer to the inner most layer respectively.

2.1.1 Epidermis

Epidermis is the outer layer of the skin that consists of flattened dead cells, mainly dead keratinocytes on top of a thick layer of keratinocytes. Also, the epidermis contains melanocytes which are responsible for the pigmentation of the skin. This pigmentation protects skin from sun damage. Additionally, Langerhans cells exist in epidermis which are a part of the body’s immune system.

2.1.2 Dermis

The middle layer of the skin is called the dermis. It consists mainly of collagen and elastin that are in charge of strength and the elasticity of the skin respectively.

2.1.3 Hypodermis

The deepest layer of skin is called hypodermis. It mainly consists of fat which is used as an energy source. This layer provides heat insulation for the body.
2.2 Dermoscopy

Dermoscopy is a non-invasive imaging technique for physicians that allows a better in vivo inspection of epidermis to the papillary dermis, the upper most layer of the dermis. Dermoscopy imaging consists of imaging the skin with dermoscope consisting of a magnifier and illuminator, with a polarizer or a skin surface gel to remove the reflection of light and helps the penetration of light into the epidermis. Studies have shown that the diagnosis accuracy can be improved by dermoscopy in comparison with clinical diagnosis without a dermoscope; however, the success is directly dependent on the experience of the observer. Consequently, several dermoscopy diagnostic algorithms have been proposed in order to increase the accuracy of inexperienced doctors. The ABCD rule, 7-point checklist and 3-point checklist are examples of these methods [2]. Research on detection and implementation of algorithms that mimic these clinical algorithms has been motivated by the recent developments in computer vision, as reviewed by Celebi et al. in [5]. Moreover, a review by Rosado et al. [24] on computer aided diagnosis (CAD) systems of dermoscopic images shows that the results achieved by the computer are statistically the same as human diagnosis. Most of the literature of CAD systems of dermoscopic images are focused in detecting dermoscopic structures. In this research our goal is to match similar looking dermoscopic structures together. Some of the most common dermoscopic structures are described below.

Pigment Network

Pigment networks (PNs) are one of the most common and important structures in dermoscopy images. Pigment networks can be defined as a set of pale blobs with dark borders attached to each other in a rectilinear pattern. In other words, A typical pigment network
is defined as “a light-to-darkbrown network with small, uniformly spaced network holes and thin network lines distributed more or less regularly throughout the lesion and usually thinning out at the periphery” [30]. A sample of pigment network is shown in figure 2.2. Sadeghi et. al. proposed an algorithm for detection of such structures using image processing, which is accurate enough to be used as a CAD support system.[26]  

Figure 2.2: A lesion containing pigment network.[1]  

Dots and Globules  
Dots are dark circular structures in various sizes within a skin lesion and globules are just big dots. In other words, Dots are defined as “black, brown or blue round structures irregularly distributed within the lesion”[2]. These structures can be detected by well-known interest point detection algorithms such as SIFT and SURF [29]. An example of dots and globules can be found in figure 2.3. F  

Figure 2.3: A lesion containing dots highlighted by arrows. [20]  

5
Streaks

Streaks are linear brown/black structures that are usually at the periphery of the lesion. They are found both in malignant and benign lesions. Streaks can be irregular or regular. An example of streaks is shown in figure 2.4. Sadeghi et. al. proposed an algorithm for detection of streaks using image processing, which is accurate enough to be used as a CAD support system. [27]

![Figure 2.4: A lesion containing streaks highlighted by arrows.][1]

2.3 Content-based Image Retrieval

A Content-based Image Retrieval (CBIR) system uses image content to search and retrieve similar images to the query image. CBIR retrieves similar images by automatically deriving features from images and matching relevant images together based on those features. Although the details of CBIR systems are different in details, the general concept is the same. For most of the CBIR systems extracting features and comparing them according to a specific metric are the main stages.

2.4 Interest Point Detection

Features are a set of values extracted from a given image that can represent the image content. One of the mostly used type of features in literature of image matching are the point based features or interest point descriptors. There are many available algorithms for interest point detection. Most of these algorithms focus on detecting corners, points that lie on edges and blob-like structures.[17] In addition, more advanced interest point
detectors provide scale invariant and rotation invariant keypoints.\cite{18}\cite{3}. We will talk more about the scale and orientation of the interest points in sections 2.3.1. and 2.3.2. There are several categories of interest point detectors. Basically, they may fall into categories of edge detectors, corner detectors and blob detectors. In this chapter we will give a brief overview of these categories.

**Edge Detectors**

Edge detectors usually detect points that are on linear structures. There are many different algorithms for edge detection and Canny, Frangi and Steger are among the most popular ones. Most of the linear structure detection algorithms work by detecting a big gradient magnitude in of any set of points in the image.\cite{13}\cite{4}

A threshold on the response of these filters leads to interest points that lie on the linear structures in the image. \cite{29}

**Corner and Blob Detectors**

Harris is one of the earliest and most common interest point detectors\cite{14}. It analyses the image intensity change in a local window. The interest points are detected by comparison of the computed eigenvalues of matrix $M$ defined as equation 2.1.

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$ (2.1)

In the equation above, $w(x, y)$ is a window function. Also, $I_x$ and $I_y$ are the components of the gradient of the image.

Harris corner detector is not scale-invariant. Furthermore, Lindeberg \cite{17} introduced the concept of scale selection which allowed each keypoint to have their own characteristic of scale.

Known modern blob detectors such as SIFT and SURF \cite{18}\cite{3} work by smoothing the image by taking the Gaussian to remove noise and high frequency data. Then, a Laplacian of Guassian of the image in different scales is used to build the scale-space. Using the scale space for each image, corners and blobs in different scales are detected. Scale invariance is one of the most important features of the recent corner and blob detectors. For an image, if the Laplacian of the Gaussian (LoG) is taken at different blurring values we will find certain corners and blobs. But if other blobs and corners with other sizes are wanted, the LoG filter should be applied to the down sampled version of the image. If this is done for several down samplings of the image then the collection of such responses is called scale space.
2.4.1 Scale Invariant Feature Transform

Lowe introduced a new algorithm for detecting key points and their features that are invariant to image scaling and rotation.[18] The algorithm is called Scale Invariant Feature Transform (SIFT). The properties of these features are suitable for matching different images. Moreover, the detected features are distinctive enough that a single keypoint can be matched to a correct keypoint with high probability against a large database of keypoints. Scale space is built by repeatedly convolving the Gaussians with the initial image (see fig. 2.5). Then adjacent Gaussian images are subtracted to produce the difference-of-Gaussian as an approximation of Laplacian of Gaussian. This makes one octave of scale space. Furthermore, for making the next octave of scale space the Gaussian image is down-sampled, by a factor of 2 and the process is repeated.

![Scale Invariant Feature Transform Diagram](image)

Figure 2.5: The left set of images are the different octaves of scale space. The right set of images are the difference in Gaussian which is an approximation of LoG operator.[18]

In order to find the potential keypoints, the scale-space is then searched for local maxima or local minima. A key point is defined as a pixel which is the minimum or the maximum among all its 26 neighbours. (See figure 2.9).

Then, based on the local image gradients, one or more orientations is computed and assigned to each key point. In the final stage a region of 16x16 in the neighbourhood of the
key point is chosen and divided into 16 subregions (4 x 4). Furthermore a 8 bit orientation histogram is built for each of these subregions which will lead into a 128 bit feature for each key point. Figure 2.6 better clarifies this process.

![Image of key point descriptor](image)

Figure 2.6: In order to calculate the key point descriptor at first it is needed to compute the gradient magnitude and orientation around the key point (left figure). After being weighted by the Gaussian window (the blue circle in the left figure), then the samples are accumulated in orientation histograms that summarize the content of 4 x 4 subregions.[18]

### 2.4.2 Speeded-Up Robust Features

Speeded-Up Robust Features (SURF) interest point detector is another detector that claims to outperform SIFT with a lower computation cost[3]. SURF uses the determinant of Hessian matrix to calculate the location of the interest point and scale. The Hessian matrix $\mathcal{H}(x, \sigma)$ is defined as below

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

In the equation above $L_{xx}(x, \sigma)$ refers to the convolution of the image I with the second order Gaussian derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$, and similarly for $L_{xy}$ and $L_{yy}$. SURF detector approximates the Laplacian of Gaussian with box filters shown in figure 2.7, which are efficiently calculated using the concept of integral image. These approximations are referred to as $D_{xx}$, $D_{xy}$, and $D_{yy}$. The filter response is then calculated by the approximate determinant of the Hessian matrix as $\text{det}(\mathcal{H}_{\text{approx}} = D_{xx}D_{yy} - (0.9D_{xy}^2))$. Here, 0.9 is to balance the approximation of the filter. Moreover, the computation of the scale space has been done using the integral images and box filters. Given an image $I$ the integral image $I_{\Sigma}(x, y)$ is defined as follows:
Using the integral image, the task of calculating the area of an upright rectangular region is reduced to four operations. If we consider a rectangle bounded by vertices A, B, C and D as in Figure 2.8, the sum of pixel intensities is calculated by:

$$\sum = A + D - (C + B)$$  \hspace{1cm} (2.4)

Where A, B, C and D are the integral image produced by their vertex. This means that, computation of summation of intensity values inside the rectangle created by these four vertices would be as follow:

$$\sum = \sum_{i=0}^{i<x_A} \sum_{j=0}^{j<y_A} I(i,j) + \sum_{i=0}^{i<x_B} \sum_{j=0}^{j<y_B} I(i,j) - \left( \sum_{i=0}^{i<x_D} \sum_{j=0}^{j<y_D} I(i,j) + \sum_{i=0}^{i<x_C} \sum_{j=0}^{j<y_C} I(i,j) \right)$$  \hspace{1cm} (2.5)

Figure 2.7: The first two pictures from the left demonstrate the second order partial derivative in y- and xy- direction. The second two pictures on the right are the approximated filters using box filters.[3]

Figure 2.8: Area computation using integral images.[3]
Now in order to build the scale-space we need to convolve Gaussian kernels of increasing size with the original image. Since we are using integral images the processing time is invariant to the size of the kernel. Therefore, we can build each image layer by filtering the original image with increasing masks (e.g. 9x9, 15x15, 21x21). This allows calculating multiple layers of scale space simultaneously. Then, the non-maximal suppression is applied in order to find the candidate points. Therefore, all of the pixels in scale-space are compared with their 26 neighbours that is 9 in each scale above or below and 8 in the same scale.

![Non-maximum suppression](image)

Figure 2.9: Non-maximum suppression: each pixel is being compared with the surrounding pixels in scale-space.[12]

Once the location of the interest point is determined, the orientation of each key point is calculated in order to have rotation invariant results. SURF calculates the orientation by computing the Haar-wavelet responses with the help of integral images introduced in the previous section in a circular neighborhood with radius 6s where s is the scale at which the wavelet responses are calculated, centered at the key point, so that responses are large in higher scales. The calculated responses are represented as a vector in 2D after being weighted by Gaussians centered at the interest point. The dominant orientation is calculated by sliding a segment of a circle of size $\pi/3$ around the origin. Then a new vector will be formed each time with the circle segment sliding that is the sum of all horizontal and vertical responses covered in that segment. Then the largest of such vectors will decide the dominant orientation. The process is illustrated in figure 2.10

After assigning the orientation for each key point, then a square with the size of 20s will be formed along the direction of the calculated dominant orientation. This square is divided uniformly into 4 x 4 sub-regions. Then 4 features are calculated in each region. In each sub-region the Haar-wavelet is calculated. Moreover, in each sub-region, Haar-wavelets of
Figure 2.10: Orientation detection: as the circle segment slides around the origin the sum of all responses covered in the segment is calculated which is the vector arrow. The largest such vector determines the orientation of the key point.[12]

Size 2s are calculated that provides 25 wavelet responses for each sub region. We annotate the wavelet responses in horizontal and vertical direction by $d_x$ and $d_y$ respectively. Also, we assume that $|d_x|$ and $|d_y|$ are the absolute value of the Haar-wavelets calculated in each horizontal and vertical direction, respectively. (See figure 2.11) Therefore, the feature vector for each sub-region would be $v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$. This provides a vector of size $4 \times 4 \times 4 = 64$ for each key point.

Figure 2.11: Descriptor: the highlighted green square is one of the subregions of the 20s square at the key point. The blue circles inside the highlighted square indicates Haar-wavelets of size 2s.[12]
Chapter 3

Method

Pigment networks and dots and globules are the common dermoscopic structures of malignant melanoma, and these structures can be detected by interest point detection algorithms. Moreover, the descriptor of each interest point can be used for matching image contents. In this chapter we discuss our method in detail. Our CBIR system consists of three phases that are “dermoscopy point location detection”, “descriptor calculation” and “image retrieval”. When a query image is given to the CBIR system the interest point locations are detected and their descriptors are calculated. Then, the closest matching images are retrieved from the database based on the distance of their keypoint descriptors. The workflow of this process is shown in figure 3.1.

3.1 Dermoscopy Point Location Detection

SURF is a known blob detector that works with grey scale images. Dermoscopic structures are proven to be more detectable on green channel [25]. Therefore, the green channel of the image has been chosen as the grey scale image. Then, histogram equalization is applied to the image for more contrast. This processed greyscale image will be the input to the SURF key point detector.

SURF is a known blob detector. Therefore it can easily find pigment network holes (pale blobs with a dark background) and dots (dark blobs with a pale background). Figure 3.2 shows an example of pigment network and dots and globules detection.

3.2 Descriptor Calculation

Once the location of a Dermoscopy Interest Point (DIP) is determined, a descriptor for each key point is to be calculated. SURF calculates these descriptors using Haar-wavelets as discussed int the previous chapter. But, this is not enough for dermoscopy use. Color
Figure 3.1: General work flow of the designed CBIR system: First, Query image’s dermoscopy key points and their descriptors get calculated. Then, the set of key points of the query image is compared to each image in the data base and a distance for each pair is generated. Then the top most similar images is shown.

plays an important role in dermoscopy image analysis. For instance, the profile of a dark globule or dot might be the same as the profile of the holes of pigment network and color is one of the important components in differentiating them. Also, color helps to retrieve more similar looking pigments and helps disambiguate visual similarity. Therefore, we have included color to the SURF descriptor. The size of the hole or blob detected by our detector is proportional to the scale of the key point. Therefore, as the extraction of color information from all of the structure’s region is desired, we calculate the color histogram of a circle with radius 7s centered at the keypoint. The value radius is optimized using IAD atlas database and 7s gave the best result both in pigment network and dots and globules retrieval. We calculate the three dimensional color histogram in the interest region in channels R, G, and B (4 bins in each dimension). This results in a vector of length 64 which records the color distribution of the region. As the last step we normalize the color component and concatenate it to the intensity component. This results in a descriptor of length 128. Also, a coefficient for each component is considered in order to weigh the importance of each component. These coefficients are calculated using the retrieval results of 10 test images.
The strength of a keypoint is defined as the determinant of the approximated Hessian as shown in the equation below.

\[ \text{det}(H_{\text{approx}} = D_{xx}D_{yy} - (0.9D_{xy})^2). \] (3.1)

In this algorithm, the strength of each detected keypoint is calculated and sorted by the strength of its descriptor, and the first 100 strongest keypoints are kept for matching. The number of keypoints to keep was determined by optimizing the precision results of pigment network and dots and globules retrieval. In other words, average precision for dots and globules retrieval is calculated for different values for number of keypoints and the number of keypoints that gave the highest precision was chosen.

### 3.3 Distance Calculation and Matching

Traditionally, in most CBIR systems, in order to find a match for a keypoint, the descriptor of that keypoint is compared to other descriptors by using a distance metric (e.g. Euclidean distance). Then the keypoint with the least distance is declared a match. Also, Lowe[18] introduced K-nearest neighbors (KNN) to matching keypoints. In his proposed method, the first two closest matches to each key point are retained. Then the ratio of the second match to the second match is calculated. If that ratio is higher than a certain threshold the matches are considered bad matches and eliminated from matches category. The reason for this is that a match is ambiguous if the first two closest matches to the keypoint are very close to each other.
However, we propose that this approach does not work for matching dermoscopy images containing pigment networks and dots and globules. This is because most of the dermoscopy structures are similar to each other and therefore it is hard to have a one to one match. This results in having multiple matches for a keypoint. For instance, as most of the pigment networks’ holes are the same it is challenging to find a single match for a detected pigment network by deploying the traditional method. For if we do so, we encounter cases where many keypoints in one image are matched to a single keypoint in the other image. For example, consider two images with pigment networks as shown in figure 3.3. If the query image is on the left, it is seen that many of the keypoints in the right image are matched to just a couple of keypoints in the query image. To solve this problem, we calculate the distance of a keypoint in the query image to all of the keypoints in all of the images in the data base using the Euclidean distance metric as described in equation 3.2.

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$  \hspace{1cm} (3.2)

Where \( \mathbf{p} = (p_1, p_2, ..., p_n) \) and \( \mathbf{q} = (q_1, q_2, ..., q_n) \) are two points in Euclidean n-space and \( d(\mathbf{p}, \mathbf{q}) \) is the Euclidean distance between those two points. Then all these values are stored in a 100x100 matrix for further evaluation as shown below.

$$M = \begin{bmatrix}
    d_{1,1} & d_{1,2} & \ldots & d_{1,100} \\
    d_{2,1} & d_{2,2} & \ldots & d_{2,100} \\
    \vdots & \vdots & \ddots & \vdots \\
    d_{100,1} & d_{100,2} & \ldots & d_{100,100}
\end{bmatrix}$$  \hspace{1cm} (3.3)
Once we have done this for all of the keypoints we search for the minimum distance among all of the distances in the matrix, which forms the most promising match among all of the possible matches. We then remove those points from our matrix so that they can’t be matched to any other point. We continue until all of the keypoints are matched, as shown in figure 3.4.

![Figure 3.4: Matching with new algorithm. Holes are matched to each other](image)

After matching all of the key points of one image to the other, we calculate the similarity of two images by taking the average distance of all the matches. Also, a threshold is set on the distance between two images. Therefore, if two images are very different to each other they will not be considered as a match in the matching list.
Chapter 4

Evaluation Methods

Evaluating a CBIR system for dermoscopy is a difficult task as there is no obvious quantitative measurement for similarity measurement of two dermoscopy images, unlike natural images where, for example, if a query image contains a cat then the retrieved images can be correctly identified, and the system’s performance accurately measured. Ideally, expert dermoscopists would evaluate the system, but it was not possible for our research project. Therefore, we had to devise novel evaluation criteria for our CBIR system. Our CBIR system was evaluated by using 1011 labelled dermoscopic images from the IAD Atlas[1] and ISIC2017[15] dermoscopy dataset. The labels of 1011 Atlas images include the lesion’s diagnosis, and a curated list of dermoscopic features present in the image. The labels of the 2000 ISIC2017 database include only the diagnosis. We evaluated the retrieval performance by two methods: in terms of dermoscopic features retrieval, and retrieval by diagnosis.

4.1 Clinical Feature Retrieval

If two dermoscopy images have the same dermoscopic features with the same color and size then they are more likely to be alike. Therefore, one evaluation method can measure the precision of retrieval of images with the same dermoscopic structures.

4.1.1 Retrieving Pigment Networks

Presence of a pigment networks (PN) is an important diagnostic parameter for melanoma [30].

It would be important to retrieve images with pigment networks when the query image contains a pigment network, as the PN feature is visually obvious, as shown in Figure 4.1. Therefore we have focused on matching and retrieving similar looking lesions containing pigment networks. Of the 1011 dermoscopic images in the Atlas database, there are 611 images labelled as having a PN present, i.e. 60.4 %. The evaluation starts by querying all 611 lesion images in the database with pigment network to the CBIR system. The first 10 most
similar retrieved images were found; in every case the first most likely image was the target image, so we observed the 9 remaining most similar images. The images retrieved with a pigment network were counted. The precision for each feature category was calculated as the average percentage of the number of correctly retrieved images over the total number of retrieved images i.e. 9. Also, the recall is calculated as the average number of correctly retrieved images in a category to the number of images of that category in the database. Moreover, the same process was done for images without pigment network.

![Figure 4.1: An example of Melanoma lesion containing pigment network. The black arrows shows the visually obvious pigment network.][20]

### 4.1.2 Retrieving Dots and Globules

Dots and globules are one of the medical features of Lentigo maligna, Lentiginous melanoma and Cutaneous melanoma shown in figure 4.2. Therefore dots and globules also have importance in clinical diagnosis. In this section we focused on retrieving dots and globules when the query image contains them. The database of 1011 labelled images contains 782 lesions with dots and globules, i.e. 77.3% of the database.

The ability of correctly identifying and matching dots and globules by this CBIR system was evaluated as for the previous section. All 782 images with dots and globules were randomly chosen as test query images. We calculated the precision and recall in the same way as for evaluating pigment network retrieval, as the percentage of the nine most similar images retrieved with the same features for each query image.
4.2 Diagnosis Retrieval

Another approach to assessing a dermoscopy image retrieval system is to calculate the accuracy of the system based on the diagnosis labels of the retrieved images. As similar dermoscopy features get located and matched, it is expected that the diagnosis of the retrieved images would be mostly similar to the query image. We have evaluated our system on 4 different diagnosis categories: Melanoma, Clark Nevus, Seboherric Keratosis and Spitz Nevus, as PNs or dots are one of their important clinical features. The Atlas database of 1011 labelled images contains 251 images for melanoma (23%), 399 images of Clark Nevus (33%), 45 images for Seboherric Keratosis (2.2%), and 79 images for Spitz Nevus.

In order to assess the diagnosis retrieval accuracy of the CBIR system, all of the images in the database in each category were used as query image and the 9 most similar images were retrieved. The precision of the system for each category is calculated as before by the percentage of images correctly retrieved out of the first 9 retrieved images. Also, the recall is calculated as the average number of correctly retrieved images in a category to the number of images of that category in the database.
4.2.1 New Evaluation Criteria

We devised a new metric for calculating the retrieval difficulty for each category for a more fair evaluation results. Difficulty of a category is calculated by the ratio of the number of all images in the database number to the number of images in that category. This means that if a diagnosis doesn’t have many image samples in the database, that diagnosis has a higher difficulty metric (DM), and it is considered a more challenging category to retrieve.
Chapter 5

Results

As discussed in the previous chapter, we have evaluated our CBIR system by its ability to retrieve similar structures and similar diagnosis. Here, we present the results.

5.1 Dermoscopic Structure Retrieval

5.1.1 Pigment Network Retrieval

An example of retrieving similar images from a query image with a pigment network can be found in figure 5.1. The accuracy of the system has been evaluated both when the query images contain, and also do not contain, a PN. Table 5.1 shows the accuracy result of querying the database with all of the images in the Atlas database (399 absent and 611 present).

5.1.2 Dots and Globules Retrieval

An example of retrieving similar images from a query image with dots and globules can be found in figure 5.2. The accuracy of the system was evaluated both when the query images contain, and also do not contain, dots and globules. Table 5.2 shows the accuracy result of querying the database with all the images in the data base (228 without dots and 782 with dots).

Table 5.1: CBIR evaluation results for PN detection using IAD Atlas data set (1011 images).

<table>
<thead>
<tr>
<th>Category</th>
<th>Precision P(9)</th>
<th>Recall at P(9)</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query image contains PN</td>
<td>75.4%</td>
<td>1.23%</td>
<td>1.7</td>
</tr>
<tr>
<td>Query image does not contain PN</td>
<td>81.2%</td>
<td>2.03%</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Table 5.2: CBIR evaluation results for dots and globules detection using IAD Atlas data set (1011 images).

<table>
<thead>
<tr>
<th>Category</th>
<th>Precision P(9)</th>
<th>Recall at P(9)</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query image contains dots and globules</td>
<td>80.21%</td>
<td>1.02%</td>
<td>1.3</td>
</tr>
<tr>
<td>Query image does not contain dots and globules</td>
<td>59.12%</td>
<td>2.59%</td>
<td>4.4</td>
</tr>
</tbody>
</table>

Table 5.3: Accuracy results for diagnosis retrieval using IAD Atlas data set (1011 images).

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Precision P(9)</th>
<th>Recall at P(9)</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanoma</td>
<td>72.3%</td>
<td>2.64%</td>
<td>3.7</td>
</tr>
<tr>
<td>Clark Nevus</td>
<td>85.1%</td>
<td>2.52%</td>
<td>3.0</td>
</tr>
<tr>
<td>Seboherric Keratosis</td>
<td>25.1%</td>
<td>11.40%</td>
<td>45.4</td>
</tr>
<tr>
<td>Spitz Nevus</td>
<td>39.3%</td>
<td>6.14%</td>
<td>15.6</td>
</tr>
</tbody>
</table>

5.2 Diagnosis Retrieval Results

The precision and difficulty metric of each diagnosis category using the Atlas images is shown in table 5.3. An example of diagnosis retrieval when the target image is Clark Nevus is shown in figure 5.3.

Also, for further evaluation of the performance of the system we have tested this algorithm on the ISIC2017 challenge database which consists of 2000 images [15]. We evaluated the precision of the system by querying all 375 melanomas present in this database. The results are shown in the first row of table 5.4.

In another experiment, the two databases of ISIC2017 and IAD Atlas have been combined to form a data set of 3011 images. Then, all of the images labeled as melanoma were fed into the CBIR system. The average number of correctly retrieved images using every melanoma in the data base as query image is shown in table 5.4.

5.3 Method Comparison

Further investigations show that our proposed method for content based image retrieval improves other standard interest point detection methods (i.e. SIFT and SURF). One of the main reasons is that SIFT and SURF only work on grey scale images. Therefore, color, which is one of the most important information in dermoscopy images is missing in these

Table 5.4: Accuracy results for diagnosis retrieval using ISIC2017 and combined (ISIC2017 and IAD Atlas data set).

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Precision P(9)</th>
<th>Recall at P(9)</th>
<th>DM</th>
<th>Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanoma</td>
<td>64.65%</td>
<td>1.72%</td>
<td>5.33</td>
<td>ISIC2017 (2000 images)</td>
</tr>
<tr>
<td>Melanoma</td>
<td>69.21%</td>
<td>1.13%</td>
<td>4.92</td>
<td>IAD and ISIC2017 (3011 images)</td>
</tr>
</tbody>
</table>
Table 5.5: Comparison of performance of detecting images with pigments networks and dots and globules using different interest point detectors in CBIR system on the IAD Atlas data set of 1011 images.

<table>
<thead>
<tr>
<th>Method</th>
<th>Performance measure $P(9)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed method</td>
<td>77.8%</td>
</tr>
<tr>
<td>SURF</td>
<td>72.2%</td>
</tr>
<tr>
<td>SIFT</td>
<td>70.6%</td>
</tr>
</tbody>
</table>

interest point detectors. Table 5.5 compares the overall performance of our system, SIFT and SURF.
The performance of the system for each method is defined as the average precision of pigment network detection and dots and globules detection on IAD Atlas data set.
Figure 5.1: Pigment Network Retrieval: The image in the first row is the query image. Other images are ranked based on the order that they are retrieved. In this example, all of the retrieved images contain pigment network.
Figure 5.2: Dots and globules retrieval: The image in the first row is the query image. Images in the other rows are ranked based on the order that they are retrieved.
Figure 5.3: Diagnosis Retrieval: The image in the first row is the query image. Images in the other rows are ranked based on the order that they are retrieved.
Chapter 6

Discussion

Each year over 5.4 million new cases of non-melanoma skin cancer are treated in United States. This number is higher than the combined incidences of cancers of breast, prostate, lung and colon [23]. These facts show the importance of skin cancer and the necessity of a system to help physicians make a more accurate treatment decision and diagnosis. The system that we have proposed in this thesis helps physicians by showing them similar looking lesions with known pathology. This concept has been implemented by the use of interest point detection.

6.1 Dermoscopy Structure Size Matching

As mentioned in section 3.2, the 100 strongest key points are kept for matching. We need to keep enough key points to find structures with approximately the same size, as well to match similar looking dermoscopy structures. For instance, we may have two images with PNs—image one only contains a PN with 50 pigment holes (a so-called smaller pigment network) and image two contains a lesion with 100 pigment holes (a so-called bigger pigment network). So if we try to match these two images together, only the maximum number of 50 key points from the pigment in the first image could be possibly matched to PN holes in the second image and the rest would be matched with higher distance of non-pigment pixels in the first image. Since the distance of two images is defined as the average of the distance of their matches, these two images are more probable to have a higher distance compared with when both images have the same number of holes; in this example, 100 pigment holes.

Figure 6.1 shows an example of this concept. As it can be seen the right image contains a smaller pigment network compared with the one on the left. To find the similarity of these two images we extract the strongest 100 keypoints and we try to match the keypoints together. Since the smaller pigment network has some artifacts on the side and it doesn’t have many pigment networks, some of these keypoints have fallen outside of the pigment.
network (marked in red rectangles). Therefore matching these interest points to the ones to the left image will make a big distance as the neither the structure not the color is the same. As a result these two images will have a higher distance compared with two lesions with PNs of the same size. (See fig. 6.2)

Of course in the given example for simplicity we have just assumed that PN is the only dermoscopic structure present in the image. In fact, if there are other structures present, each size difference, i.e. the number of pigment holes and the number of dots and globules adds a penalty to the distance comparing to when all structures have the same size.

6.2 Limitations

Based on our experimental results there are certain type of images that are hard for our system to find and match their dermoscopic structures. In most of the cases artifacts are the main problem that makes an image difficult to understand for computer vision algorithms. Artifacts can be hair, air bubbles etc. (See fig. 6.3a and fig. 6.3b). Moreover, as our system works by detecting and matching features in an image, featureless images do not generate accurate results. An example of a featureless image can be seen in figure 6.3c. Moreover, one of the other limitations of this system is having a fixed number approach for the number of used keypoints in each image. This fixed number is tuned based on the current data base using the ability of the system for retrieving pigment networks and dots and globules. Therefore, this system has to be tuned each time used with a new dataset for the best performance.
The designed system works by detecting the strongest features of a query image. Therefore, if the stronger features appear to exist on the surrounding skin of the lesion instead of the lesion itself then the retrieved results would be matches of the skin. In this section we will discuss all of these issues in detail.

6.2.1 Hair artifact

Shaved hair, as shown in 6.3a, are generally small black shapes surrounded by a skin. This pattern is considered as a part of image with high value of information as there is sudden change in the gradient of image in that point. Therefore, these areas will be strong interest points found by any kind of interest point detection algorithm. As an example, the result of first 100 interest points detected by SURF is shown in figure 6.4 and it is seen that most of them are around hair pixel areas. Therefore, presence of hair will affect the retrieval process.

Dermoscopy Hair removal algorithms are a very effective preprocessing step in this case and can often improve the accuracy of the system. Maglogiannis et al. proposed a hair removal algorithm that detects hair and restores image using interpolation.[19]. Examples of lesions with hair retrieved in our CBIR system can be seen in figure 6.6.

6.2.2 Featureless lesions

The dermoscopy interest point detects keypoints based on the detection of change in intensity of the image. But some lesions like dermatofibroma, due to their uniform color will not have any interest point detection inside the lesion. Therefore, all of the interest points will
Figure 6.3: Examples of cases that the CBIR system fails to detect the correct dermoscopic features

(a) Example of hair artifact  
(b) Example of bubble artifact  
(c) Example of feature-less lesion with hair artifact and pigmented skin
be chosen from the surrounding skin. This will result in retrieved images with the similar skin rather than similar lesions. These images are considered hard cases for our algorithm as we detect points with noticeable intensity change. Figure 6.3c shows an example of these type of lesions. Moreover, figure 6.6 shows the first 100 strongest interest points by SURF. As it can be seen most of the interest points are located outside of the lesion.

6.3 An Educational and Diagnostic Tool

As discussed before, content-based image retrieval would be a very helpful tool for physicians to have a more accurate and confident diagnosis. Moreover, this tool would be very beneficial to students learning dermatology. Our CBIR system can not only retrieve similar images but also can find the position of the matched dermoscopy structures. This feature will be very effective when dermoscopy structures are being studied. Screen shots of our proposed CBIR app are shown in figure 6.7. Moreover, as shown in figure 6.7b different statistical reports of the retrieval results can be extracted and shown. These statistical reports will help the user to understand the query image better.
Figure 6.5: The effect of shaved hair in the CBIR system: The image in the first row is the query image. Other rows consist of retrieved images that are ranked based on their retrieval order.
Figure 6.6: First 100 strongest interest points detected by SURF on a dermatofibroma with a uniform structure. Most of the keypoints are detected outside of the lesion. The red circles represent interest points with pale centre and darker surrounding. The blue circles represent interest points with dark centre and pale surrounding.
Figure 6.7: Screen shots of the proposed educational CBIR app.

(a) Query submission

(b) CBIR results and statistical reports
Chapter 7

Conclusion and Future work

We have implemented a CBIR approach toward dermoscopic CAD systems focusing on detecting and matching dermoscopic structures. The main goal of this system is to locate and retrieve images with similar dermoscopic structures. As the designed CBIR system matches the meaningful dermoscopic structures, it can also retrieve similar diagnosis to the query image. Since most of the key points are always inside of the lesion, the color of the skin does not play a significant role in the CBIR result, unlike in traditional CBIR methods. To the best of our knowledge, this system is the first usage of interest point detection in dermoscopy image retrieval. Moreover, one of the other contributions of our system is the novel matching system. This matching method is used for finding a more meaningful similarity between two dermoscopy images. This system can be used to help doctors to have more accurate diagnosis by showing them visually similar images to the query image with known pathology. Also, this system can be useful to general practitioners or medical students who are studying dermoscopy.

Future work includes expanding this approach to other clinical dermoscopic structures such as streaks. This can be possible by adding a curvilinear point detector to our interest point detector. It would also be important to have many more images in the database for more accurate image retrieval. Also the system should be evaluated by expert dermatologists, to assess the appropriateness of the retrieved images. Finally it would be key to perform a user study to gauge whether the CBIR results would affect and improve the initial diagnosis.
Bibliography


