Semi-automated techniques for the retrieval of dermatological condition in color skin images

Ranxi Huang

Semi-automated Techniques for
the Retrieval of Dermatological Condition
in Color Skin Images

by

Ranxi Huang

B.E. Huazhong University of Sci. & Tech, Wuhan, China (2004)

M.E. Huazhong University of Sci. & Tech, Wuhan, China (2006)

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in the Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology

May 2009

Signature of the Author _____

Accepted by _____

Coordinator, M.S. Degree Program   Date

Chester F. Carlson Center for Imaging Science
College of Science
Rochester Institute of Technology
Rochester, New York

Certificate of Approval

_____

M.S. Degree Thesis

_____

The M.S. Degree Thesis of Ranxi Huang has been examined and approved
by the research committee as satisfactory for the thesis
requirements for the Master Degree in Imaging Science

_____

Dr. John Kerekes, Thesis Advisor

_____

Dr. Eli Saber

_____

Dr. Maria Helguera

_____

Date

ii

Thesis Release Permission Form

Chester F. Carlson Center for Imaging Science
College of Science
Rochester Institute of Technology
Rochester, New York

Title of Thesis: Semi-automated Techniques for the Retrieval of
Dermatological Condition in Color Skin Images

I, Ranxi Huang, hereby grant permission to the Wallace Memorial Library of
Rochester Institute of Technology to reproduce my thesis in whole or in part.
Any reproduction will not be for commercial use or for profit.

_____

Signature of the Author

_____

Date

iii

Semi-automated Techniques for
the Retrieval of Dermatological Condition
in Color Skin Images

Ranxi Huang

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Imaging Science in the Center for Imaging Science, Rochester Institute of Technology

Abstract

Dermatologists base the diagnosis of skin disease on the visual assessment of the skin. This fact shows that correct diagnosis is highly dependent on the observer's experience and on his or her visual perception. Moreover, the human vision system lacks accuracy, reproducibility, and quantification in the way it gathers information from an image. So, there is a great need for computer-aided diagnosis.

We propose a content-based image retrieval (CBIR) system to aid in the diagnosis of skin disease. First, after examining the skin images, pre-processing will be performed. Second, we examine the visual features for skin disease classified in the database and select color, texture and shape for characterization of a certain skin disease. Third, feature extraction techniques for each visual feature are investigated respectively. Fourth, similarity measures based on the extracted features will be discussed. Last, after discussing single feature performance, a distance metric combination scheme will be explored.

The experimental data set is divided into two parts: developmental data set used as an image library and an unlabeled independent test data set. Two sets of experiments are performed: the input image of the skin image retrieval algorithm is either from developmental data set or independent test data set.

The results are top five candidates of the input query image, that is, five labeled images from image library. Results are laid out separately for developmental data set and

independent test data set. Two evaluation systems, both the standard precision vs. recall method, and the self-developed scoring method are carried out. The evaluation results obtained by both methods are given for each class of disease.

Among all visual features, we found the color feature played a dominating role in distinguishing different types of skin disease. Among all classes of images, the class with best feature consistency gained the best retrieval accuracy based on the evaluation result. For future research we recommend further work in image collection protocol, color balancing, combining the feature metrics, improving texture characterization and incorporating semantic assistance in the retrieved process.

Acknowledgements

This dissertation would never be possible without the generous help and support from many people. I would like to express my deepest respect to my supervisor, Dr. John Kerekes for his invaluable guidance and inspiring advice. He taught me the skills I needed to finish my M.S. degree with great patience and loving kindness. Without his considerate help, I could not come to this point successfully as an international student. Thanks also go to my committee members, Dr. Eli Saber and Dr. Maria Helguera, for their helpful suggestions in my research.

Also, I would like to express my gratitude to the people from Logical Images, who provided and selected the image data for me to work on this project.

At last, I would like to thank my family and friends for their encouragement and support throughout the year. Especially to my dear mom, who came to the US to accompany me in the last few months and made my life much easier.

Table of Contents

List of Figures

x

List of Tables

1. Introduction

Skin disease is one of the top 15 groups of medical conditions for which prevalence and health care spending increased the most between 1987 and 2000, with approximately 1 of 3 people in the United States with a skin disease at any given time. Even so, a national data profile on skin disease has not been conducted since the late 1970s. A study by Bikers, et al, 2004, closed the gap by estimating the prevalence, economic burden, and impact on quality of life for 22 leading categories of skin disease. The estimated annual cost of skin disease in 2004 was $39.3 billion, including $29.1 dollars in direct medical costs (costs of health services and products) and $10.2 billion in lost productivity costs (defined as costs related to consumption of medical care, costs associated with impaired ability to work, and lost future earning potential because of premature death). Based on a methodology of willingness to pay for symptom relief, the additional economic burden of skin disease on quality of life amounted to an estimated $56.2 billion. Including the economic burden on quality of life, the total economic burden of skin disease to the US public in 2004 was approximately $96 billion.

Skin diseases are well known to be a large family. The identification of a certain skin disease is a complex and demanding task for dermatologist. Thus, it is important to develop efficient schemes for clinical diagnosis and to support the dermatologists with computer-aided diagnosis systems. Also, a computer aided system can reduce the work load of the dermatologists, especially when the image database is immense (Schmid-Saugeon et al, 2003; Takemae et al, 2000; Casitello et al, 2004).

However, most contemporary work on computer aided analysis skin disease focuses on the detection of malignant melanoma (Fiorini et al, 2004; Ganster et al, 2001;

1

Maglogiannis et al, 2005). Thus, the features they used are very limited. The goal of our work is to build a retrieval algorithm for the more general diagnosis of various types of skin diseases. Our database is provided by a local company (Logical Images, 2009), and contains images from eight categories. It can be very complex to define the features that can best distinguish between classes and yet be consistent within the same class.

There are mainly two kinds of methods for the application of a computer assistant. One is text query. A universally accepted and comprehensive dermatological terminology is created, and then example images are located and viewed using dermatological diagnostic concepts using a partial or complete word search. But the use of only descriptive annotation is too coarse and it is easy to make different types of disease fall into same category. (Dermatology Lexicon Project @ University of Rochester, 2005)

The other method is to use visual features derived from color images of the diseased skin. The ability to perform reliable and consistent clinical research in dermatology hinges not only on the ability to accurately describe and codify diagnostic information, but also complex visual data. Visual patterns and images are at the core of dermatology education, research and practice. As we move into the next millennium, advances in digital imaging techniques and processing will deliver new and powerful methods to measure outcomes in clinical research. These new imaging and computing technologies will also facilitate remote diagnosis through telemedicine. Digital technology is also creating new opportunities to enhance medical education and care through digital image databases, computerized medical records and knowledge sources.

Visual features are broadly used in melanoma research, skin classification and segmentation. But there is a lack of tools using content-based skin image retrieval. So we

developed a content-based image query system for the assignment of skin disease to eight categories. Our approach uses five steps: First, after examining the whole database, visual features are identified as indicators to discriminate of each image class. Second, features are extracted by image processing according to the definition in the expression of mathematical formulae. Third, distance metrics are calculated for a similarity measure of each feature. Four, a distance metric combination is applied taking advantage of the complementary nature of the distance metrics. Lastly, using the distance metrics, the top five similar images from the image library are retrieved.

The following summarizes the rest of the document. Chapter 2 lists the objectives of this research. Chapter 3 investigates the background of this research, including general content based image retrieval, medical image retrieval and skin disease identification. Chapter 4 introduces the dataset of this research and illustrates the eight kinds of skin diseases involved. Chapter 5 elaborates on the methodology of skin image retrieval: feature identification, feature extraction, similarity measure, distance metric combination and indexing. Chapter 6 provides the results and their evaluation. Chapter 7 presents the summary and conclusion with a discussion of future work to be done.

2. Objectives

The overall objective of this project is to assist dermatologists with a computer-aided recognition technique to achieve skin disease diagnosis with higher accuracy. The desired outcome is to develop a dermatological image query system to provide doctors with the best candidates of the pending image among the existing comprehensive image database.

The intermediate objectives of this work are the following:

- Elucidate the background of this thesis and find the relative literature in the field. Lay out the necessary background for a reader to understand this thesis. In order to proceed in this research, certain reflections on previous work should be investigated, such as general image query systems and medical image retrieval methods, especially as applied to the field of skin image query techniques. Also, related fields, such as skin disease identification, were researched.

- Determine the visual features to be used for dermatological image query. Image query systems can be based on visual features such as color, texture, and shape. Color plays the most significant role in dermatological images. In homogenous skin patterns, texture demonstrates the character of the skin. Some dermatological conditions have a certain consistent geometric shape that can be easily distinguished.

- Determine the feature extraction method for each visual feature. For color features, determine which color space is suitable for the skin image database. Texture information can be obtained by investigating some operators and descriptors that can qualitatively demonstrate the patterns of the skin. To

5

extract a shape description, the method of object recognition and contour extraction should be investigated.

- Determine distance metrics for feature similarity measure.

  For each visual feature, different distance metrics should be selected according to the property of the feature space.

- Determine distance metric combination method.

  If distance metrics for each feature are not of the same type and cannot be integrated at the level of similarity measure, the combination of different distance metrics should be considered.

- Develop a skin image query algorithm and provide results.

  After pre-processing (if necessary), perform the feature extraction for both the test image and library images. Then, compute distance metrics for each feature between the test image and each library image. After distance metric combination, select top five library images most similar to the test image as the retrieval results.

- Evaluate the querying accuracy for each class.

  Perform the standard *precision* vs. *recall* evaluation of the retrieval system for each image category. Also, build a statistical scoring algorithm based on the five result images to count the accuracy of this query system for each image category.

  This project will develop a skin image retrieval system that will produce five best candidates most similar to the query image from the image library, thus largely narrowing down the search and comparison work of the dermatologists. The results of two evaluation methods will be given for each image category.

6

3. Background and previous work

Dermatology can be described as the branch of medicine primarily concerned with skin and its diseases. It is one of the most difficult areas in medicine. It demands detailed knowledge and experience. Fortunately, dermatologists are now aided by a number of modern inventions which makes their work easier and produces better results for patients (Antkowiak, 2006).

Computer assisted diagnosis can help doctors in remote diagnosis, or interaction with large image database. However, in most cases, doctors make the diagnosis by their knowledge and rich experience using perceptual judgment, and lab tests or radiographs.

Half of all diseases exhibit visual or skin characteristics. These perceptual skills are critical in a health care system with a 15 percent rate of misdiagnosis (New York Times, 2009). Thus, computer-aided image processing could play an important role in assisting perceptual judgment of a given case. There are lots of ways that computers can help in diagnosis, such as pre-processing of the medical image, justification of a certain disease, alignment of different sources of image, and classification of several different diseases. In our case, the computer will produce a group of candidates generated by a retrieval system from a large image database containing various kinds of skin disease.

In this chapter, we will first review the basic method of general image retrieval, and then narrow down to medical image retrieval. At last, we will elaborate in the field of computer-aided skin disease diagnosis.

3.1 General image query methods

This section will introduce algorithms for general content-based image retrieval. Content-based image retrieval (CBIR), also known as query by image content (QBIC) is the application of computer vision to the image retrieval problem; that is, the problem of searching for images in large databases (Niblack et al, 1993). The general structure of CBIR system is illustrated in Fig. 3-1.
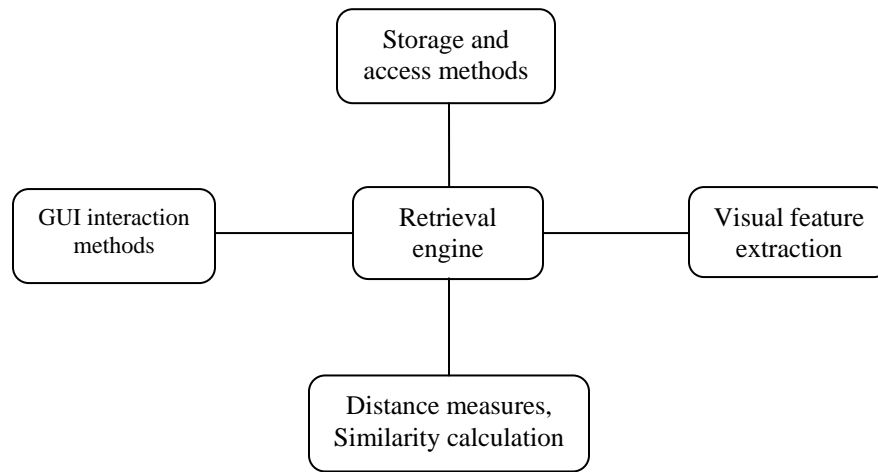
Figure 3-1   Principal components of all content-based image retrieval systems(Muller et al, 2004)

The following will introduce each component of CBIR techniques respectively including the software systems as well as the querying techniques.

3.1.1. Content used for image description

This sub-section lists some common features of image content (Vasconcelos and Vasconcelos, 2004; Hoiem et al, 2004; Veltkam and Tanase, 2002).

Color: Since color contains significant information of an image that can be used for
discrimination from other images and due to its independence of image scaling
and orientation, color query methods are among the most widely used techniques.

8

Commonly used color features are: eigen images, global/subimage histograms, correlation histograms, color moments, color coherence vectors, region histograms, and dominant colors.

Texture: Defined as a function of the spatial variation in pixel intensities, texture is the most important visual cue in identifying a certain texture pattern. The most commonly used texture features are: wavelets, Gabor filters, Fourier descriptors, co-occurrence matrices, atomic texture features, random field decompositions, local binary patterns, and edge statistics.

Shape: This is defined in our work as the geometric information of a particular region that is being investigated. But shape often depends on segmentation, which is difficult to perform automatically. So, shape detection usually involves human interaction. The common shape features are: edge direction histograms, template matching, Fourier descriptors, curvature scale space, bounding box/ellipses, and elementary description techniques.

3.1.2 Similarity measures

The similarity measure between a query image and data set must be adaptive to the feature metrics extracted above. For example, histogram intersection often serves as the similarity measure for color histogram extraction, which is further discussed in Appendix A. Hausdorff distance is a similarity metric for shape contours.

Also, mutual information (Viola, 1997; Pluim et al, 2003; Hao et al, 2000; Chua and Tischer, 2003; Alvarez et al, 2005) is a measure of the degree of dependence between corresponding pixels in the images being compared. It makes uses of both spatial information and feature characteristics. It is invariant with position, rotation and scaling,

9

avoiding the difficulties in image segmentation. It has outstanding performance in registration between multi-modal images. Although mutual information is widely applied as a criterion in image registration, it can also be used as a similarity measure between image pairs, while simple spatial correlation methods fail to demonstrate the maximum likelihood. Thus, mutual information can serve as similarity measure of those content features described above.

3.1.3   Indexing techniques

To efficiently search from hundreds of thousands of images to a few that the user can quickly browse, images and their associated features need to be properly indexed. There are several multidimensional indexing techniques for capturing the low-level features like feature based or distance based techniques, each of which can be further classified as a data-partitioned or space-partitioned based algorithm. Feature based indexing techniques project features of an image as a vector in a feature space and then indexes the space. The basic feature based index structures are R-tree (data partitioning based indexing structure), KDB-tree (space partitioning based indexing structure), and Hybrid tree (combine these two structures together) (Chatterjee and Chen, 2006; Liu et al, 1998).

Distance based indexing structures are built based on the distances or similarities between two data objects.

3.1.4   Querying techniques

Query by example: use an example image to search the database, and the resulting images should all share common elements with the provided example. The example image may be supplied by the user or chosen from a random set. The example can also be

10

a sketch drawn by the user. This query technique avoids difficulties when employing words to describe the image

Query by semantic text: user makes a request by a phrase to describe the image needed; usually those phrases are generally regarded as the high-level features, compared to the low-level features such as texture, color, and shape. However, the accuracy of identifying high-level features often relies on the success of extracting low-level features.

3.1.5 Software systems

There are dozens of CBIR systems; for space limitation, we just list some major systems in Tables 3-1, 3-2, and 3-3.

Table 3-1   Commercial CBIR Systems

System name	Description
IBM:QBIC	Query large on-line image databases using the images' content as basis of queries (Niblack et al, 1993)
Virage	Rich media management software and intelligent video analytics (Virage, 2009)
Corbis	Text-base image retrieval system for photographers (Corbis, 2009)

Table 3-2   Free and Open Source of CBIR

System name	Description
imgSeek	Open source photo collection manager and viewer with content-based search and many other features (ImgSeek, 2009)
GIFT	The GNU Image Finding Tool : an open source query by example system (GNU, 2005)

Table 3-3   Academic CBIR System

System name	Description
VisualSEEk	Uses color and spatial layout, with sketch tool (Smith and Chang, 1996)
WebSEEk	Utilizes text and visual information; relevance feedback (Webseek, 2009)
MetaSEEk	Content-based meta-search engine for finding images on the Web (Benitez, 1998)

11

3.2   Medical image retrieval

The following sub-sections will explore how image retrieval methods can be applied in the medical field.

Content-based Image Retrieval (CBIR) from medical image databases does not aim to replace the physician by predicting the disease of a particular case but to assist the doctor in diagnosis. The visual characteristics of a disease carry diagnostic information. In many cases, visual similarity between different medical images indicates the correspondence to the same disease category. By consulting the output of a CBIR system, the physician can gain more confidence in his/her decision or even consider other possibilities (Dreisetl et al, 2001; Park et al, 2007). Fig.3-2 illustrates how a medical image retrieval system works.
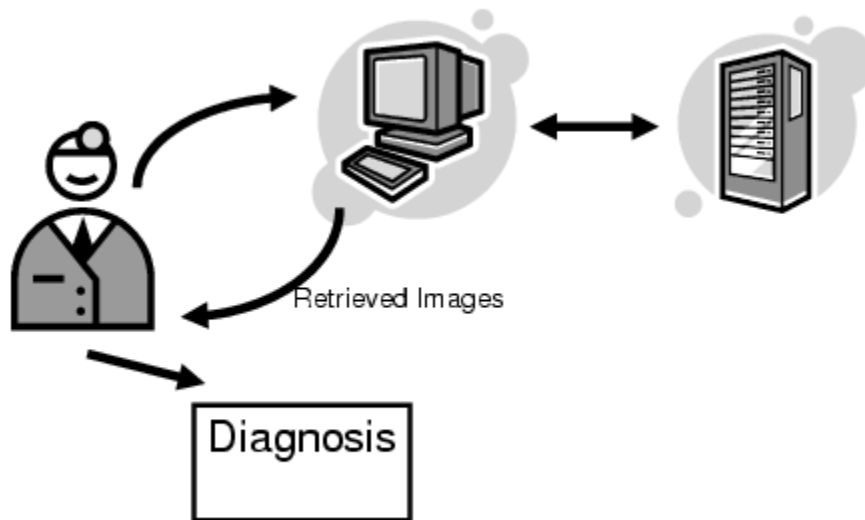
Figure 3-2   Medical image retrieval system (Park et al, 2007)

12

3.2.1 Image types used for content-based image retrieval

There are a variety of medical departments where the use of content-based access methods has been implemented or proposed. Most applications are centered on images produced in radiology departments, but some other departments also employ CBIR method. Table 3-4 list the various image types and systems that are using these images.

Table 3-4   Image Types and the Applied Systems

Images used	Name of the systems
HRCTs of the lung	ASSERT
Functional PET	FICBDS
Spine X-rays	CBIR2, MIRS
Pathologic images	IDEM, I-Browse, PathFinder, PathMaster
CTs of the head	MIMS
Mammography	APKS
Images from biology	BioImage, BIRN
Dermatology	MELDOQ[1]
Breast cancer biopsies	BASS
Varied images	I^2C, IRMA, KMed, COBRA, MedGIFT, ImageEgnie

3.2.2 Features selected for medical images

We consider two kinds of features used in medical image research: visual features and textual features.

3.2.2.1 Visual features: (Muller et al, 2004)

a) *Color*: If available, color has been a most effective feature. While most of the images are in RGB color space, many indexing and querying techniques do not employ this space due to the poor correspondence to human color perception. Other spaces such as HSV, CIE and LUV are much better with respect to human

---

[1] The MELDOQ system has been developed by an interdisciplinary team to improve the early recognition of malignant melanoma (Delventhal, 1998).

13

perception and more frequently used. However, most medical images are only grey scale so that the application of color information is limited.

b) *Texture*: There are various ways to measure texture. Some of most common measures are wavelets and Gabor filters, since they perform better and correspond well to the properties of human visual cortex for edge detection. These texture measures try to capture the characteristics of the image with respect to changes in certain directions and the scale of the changes. Other texture descriptors derive from co-occurrence matrices.

c) *Local and global features*: For precise description, features can be used on a global level or on a local level. To define a local feature, one simple method is to use blocks of fixed size and location. Usually, local feature selected contains more information about the image objects or underlying structures.

d) *Shape*: Processes dealing with the geometric character of the image need to extract shape of the object by segmentation and contour extraction technology. Segmentation and contour extraction are interrelated, but fully automated algorithm for both is an unsolved research problem. There are several common techniques: Sobel, Laplacian, LOG, and DOG descriptor to find the edges of object, deformable template (active contour) (McInerney and Terzopoulos, 1996) and statistical threshold segmentation. In the medical image area, professionals usually play a role in segmentation together with computer, because accuracy is more important than speed in medical situation. In most cases, doctors identify an area of interest, and then computer can find a more accurate contour.

3.2.2.2 Textual features

For effective clinical decisions, one needs to use semantic or cognitive information from an image rather than those low-level features. Semantic features are defined according to certain image object properties, but still it is difficult to extract semantic information from primitive visual features. Most of the annotation combines visual features with text, which is most useful in medical applications because of good annotated atlases of medical images containing objective knowledge (Xu et al, 2004; Hammiche et al, 2004; Han et al, 2004).

Textual information can be collected from patient records or studies. Some define a context-free grammar, a standardized vocabulary for image description. Also, some structured information containing a whole set of patient information can also provide some context information of the image. (Muller et al, 2004)

3.2.3 Similarity measures used

Similarity measures compute the corresponding feature distance between two images to be compared. One category of this is vector distance, such as Euclidean distance or Mahalanobis distance (Wikipedia, 2009). Another category is a probabilistic framework to measure the probability that an image is relevant, such as support vector machines. Other statistical approaches use Bayesian networks (Niedermayer, 1998) and hidden Markov Models (Hidden Markov Model, 2009). For image retrieval, this is based on two principles: one, a feature frequently in an image describes this image well; two, a feature frequent in the collection is a weak indicator to distinguish images from each other (Iakovidis et al, 2006).

15

3.2.4 Indexing methods

Since medical image retrieval is an interactive system, an efficient and quick access to the data is very important. To reduce the feature space, principal component analysis (PCA) and independent component analysis (ICA) are mentioned in Muller, 2004. Other indexing techniques such as KD-tree and R-tree are also mentioned in Muller, 2004.

3.3 Skin disease recognition

The following sub-sections will explore computer processing methods in the field of skin disease. Since malignant melanoma is a form of skin cancer that is on the rise in most parts of the world, computer assisted identification of malignant melanoma has become a topic of great interest in contemporary literature of this field (Holmstrom, 2005). So we focus on this problem in our review. Also, other literature specified on skin classification and segmentation will also be investigated.

3.3.1 Melanoma research

In order to recognize melanoma more automatically, much research has been done in computer aided image analysis of melanoma. In this sub-section, we will first introduce the classical melanoma recognition strategy. Then, other methods to distinguish melanoma are mentioned.

Most commonly images are collected by Epiluminescence microscopy (ELM). ELM has proven to be an important tool in the early recognition of malignant melanoma. This imaging technique reveals most of the pigmented structures by allowing the light to penetrate deeper into the skin, thus rendering the surface translucent and making subsurface structures visible.

16

First, we shall introduce the classical ABCD diagnosis system, ABCD represents the asymmetry, border structure, variegated color, and dermatoscopical structures and defines the basis for a diagnosis by a dermatologist. The computer aided melanoma diagonosis contains four steps: (Binder et al, 2000; Schmid-Saugeon et al, 2003; Castiello et al, 2004; Ercal et al, 1993; Fiorini et al, 2004; Holmstrom, 2005).

a) *Image Collection*: Epiluminescence microscopy (ELM) has proven to be an important tool in the early recognition of malignant melanoma. This imaging technique reveals most of the pigmented structures by allowing the light to penetrate deeper into the skin, thus rendering the surface translucent and making subsurface structures visible.

b) *Segmentation*: For skin lesion segmentation, mainly region-based segmentation methods are applied, and within this category a thresholding operation is most often used (Ganster et al, 2001; Fischer et al, 1996). There are also other kinds of segmentation method, like color segmentation developed by Schmidt and Fischer, 1997. Combining the different methods resulted in further improvement of correctly identified tumor boundaries.

c) *Feature Calculation*: For asymmetry, different shape features were used. Some groups use the sharpness of the transition from the lesion interior to the skin, as descriptors of the structure and irregularity of the border. The descriptors of color are mainly statistical parameters calculated from different color channels (Ganster et al, 2001; Fischer et al, 1996). Kreutz and Gehlen, 2001 employed a hybrid method that combines a statistical clustering of the color space and a hierarchical region growing method.

17

d) ***Feature Selection and Classification***: a variety of statistical and machine learning approaches to classification tasks are currently available: k-nearest neighbors, logistic regression, artificial neural networks (Binder et al, 2006), decision trees,support vector machines, and mixture-of-experts architectures (Kreutz et al, 2001).

There are other criteria to differentiate benign melanoma from malignant melanoma.

Green et al, 1994, proposed several features: means and standard deviations of values from red, green and blue planes within the lesion boundary; the means and standard deviation of the gradient at each boundary pixel in the RGB planes; area of lesion; and perimeter; fragment index which has a value of unity if a lesion is a perfect circle and smaller values if it is irregular; change in area when the color thresholds were reduced by 10%, as a broad indication of gradient of color change around the lesion boundary; and shape change.

Round et al, 1998 and 2001, found that skin line patterning tends to be disrupted by malignant but not non-malignant skin lesions, which could be used as an aid to lesion differentiation. Skin pattern has been extracted from optical images by high-pass filtering and profiles of local line strength variation with the angle estimated using a new consistent high-value profiling technique. A measure based on the relationship between the classification results and an intensity-based segmentation was calculated, and this represented the disruption of the skin line patterning.

Claridge, 1992 employed 'bulkiness' as a measure of the shape. Irregularity of the border is expressed by two fractal dimension measures, one for the 'structural' aspect of the shape and the other for the 'textural' aspect.

18

Handels et al, 1999 proposed several types of features for characterizing the structure of skin surface profiles: texture features based on co-occurrence matrices, Fourier features and fractal features. Then, a genetic algorithm was applied to determine suitable feature subsets for the recognition process.

3.3.2 Skin disease classification

Antkowiak, 2006 used Artificial Neural Networks (3 layers) and Support Vector Machines separately for classification of 7 types of skin diseases taken under visible light. The feature space was obtained by Fast Fourier Transformation from the spatial domain to frequency domain.

A texture classification approach (Abdel Wahab el al, 2005) employed a back propagation neural network to classify texture features: gray level co-occurrence matrices, wavelet analysis, Law's Texture energy measurements, gray level gradient distribution and statistical moments (mean, variance, kurtosis, skewness). These features were gathered into one feature vector and underwent feature reduction process using PCA.

A color classifier approach (Cheng and Umbaugh, 2005) used 17 color features: binary features and histogram features in R, G, B bands. These features were classified by two classification models: Discriminant Analysis and Multi-layer Perceptron.

Note that none of those methods used the combination of color feature, textural information and shape characteristics which forms the basis of our proposed method.

3.3.3 Skin disease segmentation

Color segmentation proposed by Umbaugh et al, 1989 was performed strictly on color information. The color space was transformed from rectangular RGB space to a two

19

dimensional spherical coordinates, and then a center split method was performed in this space. This algorithm was shown to be a useful aid in the identification of tumour border, ulcer, and other features of interest. Umbaugh et al, 1993 added another 3D Principal Component Transformed space and performed median split in the space for segmentation. Hance et al, 1996 explored several types of color segmentation methods: adaptive thresholding, fuzzy c-means, Spherical Coordinates Transform/center split, Principal Component Transform/median cut, split and merge and multi-resolution segmentation.

Phung et al, 2005 examine three important issues of color pixel calssification approach to skin segmentation: color representation, color quantization and calssification algorithm, and found out skin segmentation based on color pixel classification is largely unaffected by the choice of the color space.

Cell nucleus segmentation proposed by Tanaka et al, 2001, extracted nuclei regions which were surrounded by edges of certain strength, and under this restriction, segmentation of arbitrary shaped nuclear regions. Segmentation of arbitrary shaped nuclear regions and weakly stained nuclear region is made. A dynamic thresholding method with combining Laplacian histogram with Ostu's method is used for segmentation.

3.3.4　Skin disease retrieval by text

Recently, the Dermatology Lexicon Project (University of Rocheter, 2005) has developed a system that can query a skin image by word search. Just enter the term into the entry window and a single diagnosis or list of diagnoses will be returned. The user can select the diagnosis and view its placement in a pathophysiology hierarchy.

20

3.3.5 Discussion on literature review

To sum up the literature review, we can see there is a lack of references devoted to skin image retrieval by computer extracted visual features for various types of skin disease. Skin image retrieval differentiates from disease justification between only two possibilities like melanoma research, or image classification – classification of a group of images into different categories, or retrieval by text. However, we can still borrow ideas from those techniques, such as the use of color, texture feature extraction, and segmentation to extract shape feature.

3.4   Performance evaluation of content-based image retrieval

Muller et al, 2001, gave a review on evaluation methods in CBIR systems. They listed several kinds of performance evaluation methods: rank of the best match, average rank of relevant images, precision and recall, target testing, error rate, retrieval efficiency and correct and incorrect detection.

Precision and recall, proposed by Squire et al, 1999 is the most common evaluation measure used in image retrieval.

$$precision = \frac{\#relevant\ images\ retrieved}{total\ \#\ images\ retrieved} \qquad \textbf{Equation 3-1}$$

$$recall = \frac{\#relevant\ images\ retrieved}{Total\ \#relevant\ images\ in\ the\ collection} \qquad \textbf{Equation 3-2}$$

They are usually presented as a *precision vs. recall* graph (PR graph) as shown in Fig. 3-3.

21
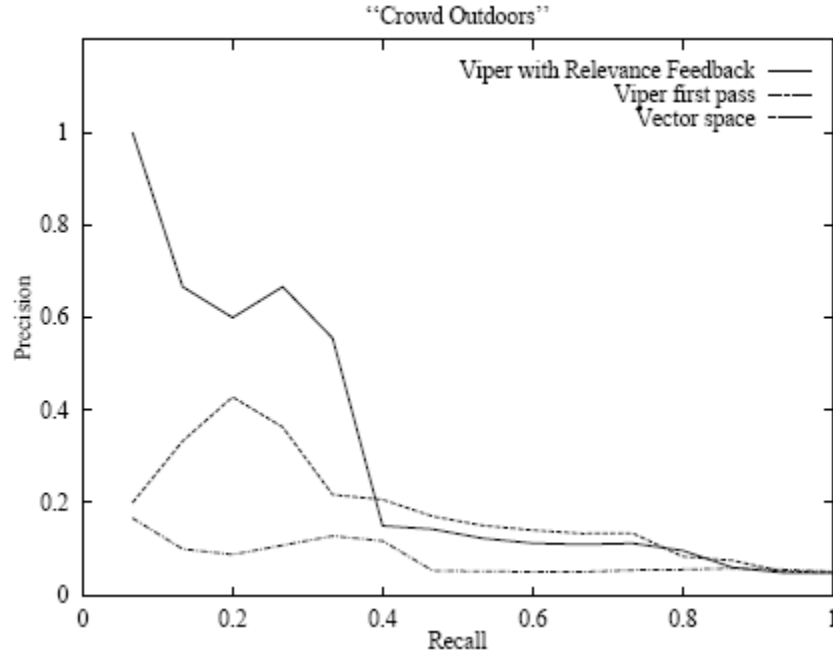
Figure 3-3   An example of *precision* vs. *recall* graph (Squire et al, 1999)

An ideal case is: precision=1 for all values of recall when all the relevant images

are retrieved before any irrelevant ones.
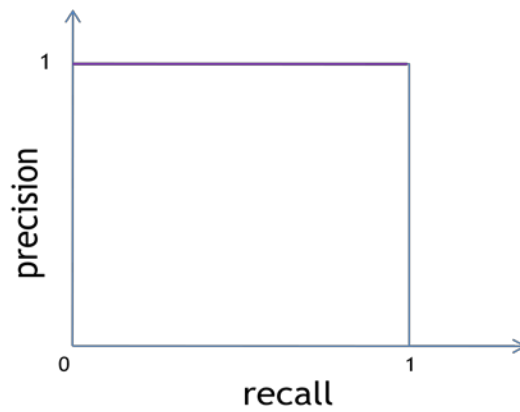
Figure 3-4   An idea case of *precision* vs. *recall* graph

4. Skin color images data set

In this chapter, the data set used for the project is introduced.

Our data set contains eight classes of skin disease, provided by a local company (Logical Images, 2009). It has been divided into two sets: developmental data with a total of 139 labeled images, and an independent test data with a total of 76 unlabeled images.

The eight classes of skin disease are listed in Table 4-1 (The Free Dictionary, 2009).

All images for the data set are included in the Appendix B. The sample size of images for each disease in the developmental data is listed in Table 4-2 .Examples of each kind of disease are shown in Table 4-3.

All the images are in JPEG format. No pixel dimension information was given, except for a few images with a ruler on it, indicating the actual size per pixel. The illumination condition was also unknown for each image. Also, the images were collected with various backgrounds.

23

Table 4-1   Definition of the Diseases

Skin disease	Definition
Fluid-filled	A blister filled with fluid (an amorphous substance whose molecules move freely past one another).
Gangrene-necrotic tissue	A mortified or gangrenous part or mass, (pathology) gangrene that develops in the presence of arterial obstruction and is characterized by dryness of the dead tissue and a dark brown color.
Pigmentation	Coloration of tissues by pigment.
Purpura–violaceous	Purpura: a small hemorrhage in the skin, mucous membrane, or serosal surface; a group of disorders characterized by the presence of purpuric[2] lesions, ecchymoses, and a tendency to bruise easily.
	Violaceous: having a violet color, usually describing a discoloration of the skin.
Raised with color change	A convex area with a different color from normal skin.
Redness – general	Red patches on the skin.
Ulcerated	A lesion of the skin or a mucous membrane such as the one lining the stomach or duodenum that is accompanied by formation of pus and necrosis of surrounding tissue, usually resulting from inflammation or ischemia; a corrupting condition or influence.
Warty-crusty-scabby	A hard rough lump growing on the skin, caused by infection with certain viruses and occurring typically on the hands or feet; A similar growth or protuberance, as on a plant.

Table 4-2   Size of the Developmental Data Set

Skin disease	sample size
fluid filled	20
gangrene-necrotic	12
pigmentation	17
purpura - violaceous	18
raised with color change	21
redness - general	19
ulcerated-eroded	17
warty-crusty-scabby	15
Sum	139

.

---

[2] Purpuric: a condition characterized by hemorrhages in the skin and mucous membranes that result in the appearance of purplish spots or patches, also called *peliosis*.

24

Table 4-3   Example Images for Each Skin Disease

Skin disease	Examples of the disease			
Fluid filled				
Gangrene-necrotic				
Pigmentation				
Purpura-violaceous				
Raised-with-color-change				
Redness-general				
Ulcerated-eroded				
Warty-crusty-scabby				

www.manaraa.com

5. Methodology

A general overview of the approach will be provided. Then, we will elaborate on each part of it in the subsections. First, the pre-processing steps are given. Second, a visual feature identification study is discussed by analyzing the character of each skin disease. Third, we discuss the feature extraction and similarity measure technique used for each selected feature. Four, due to different distance metric used for each feature, a distance metric combination will be employed. However, before determining the scheme of distance metric combination, each visual feature will be analyzed separately to see their performances. Finally, a self-developed scoring system for performance evaluation will also be introduced.

5.1   General approach

The general approach is based on content based image retrieval methods. Fig.5-1 illustrates an overview of the retrieval process.
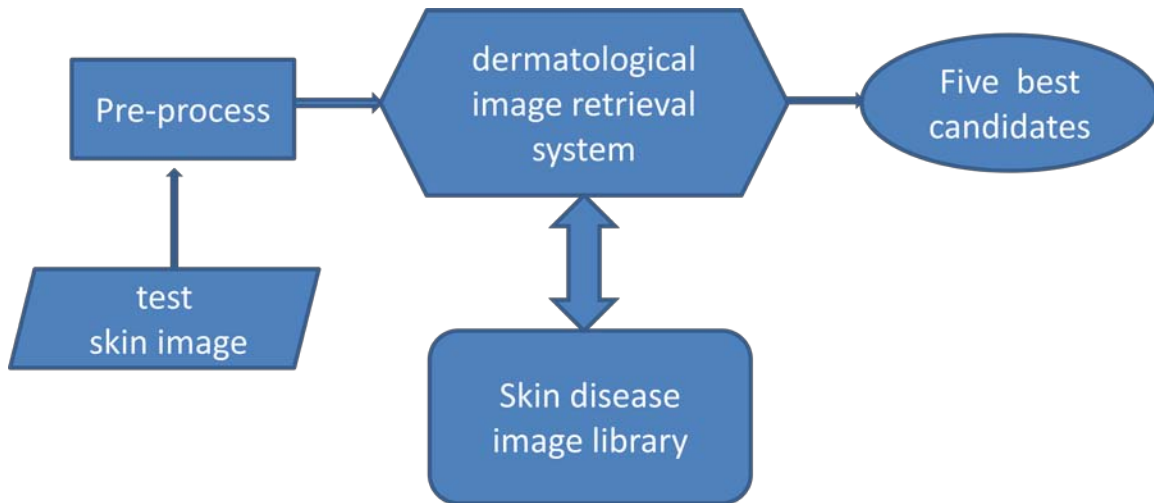
Figure 5-1   Overview of the skin image retrieval process

The input test image is a single image either from the developmental data set or the independent test data. The skin disease library consists of the labeled developmental data. The dermatological image retrieval system is the main algorithm whose scheme will

27

be discussed next. The outputs are five best candidates from the skin disease image library that are most similar to the test image.

The dermatological image retrieval system consists of the following parts. First, pre-process both the test image and image library. Second, after feature identification, color, texture and shape feature are extracted. Third, for the color feature, histogram intersection serves as the similarity measure; for the texture and shape feature, Euclidean distance is used as the distant metric. Four, distance metric combination is used to colligate the two distant metrics. Fig.5-2 shows the content of dermatological image retrieval system.
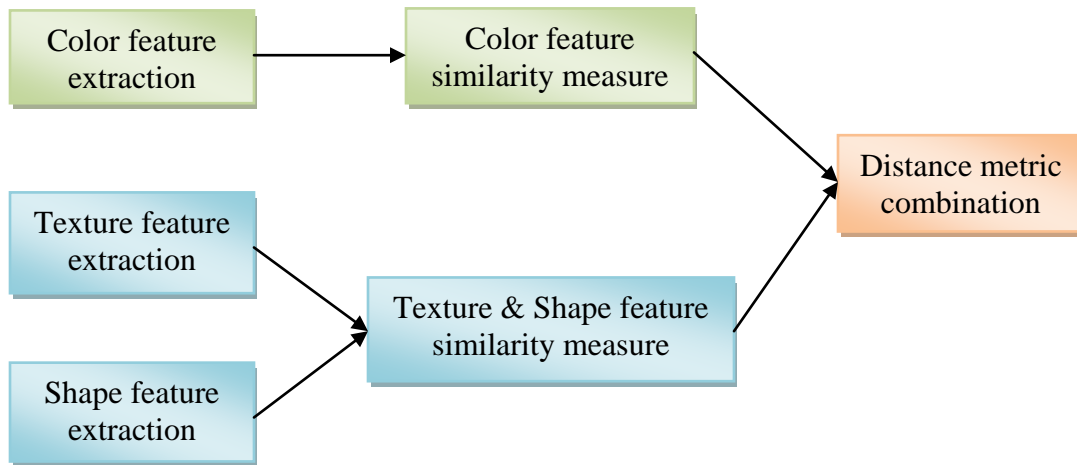
Figure 5-2   Content of the dermatological image retrieval system

5.2   Pre-process

All the images are in JPEG format. In order to perform the retrieval algorithm on image, having the same level of actual size, one should either know the pixel dimension beforehand or make sure all the images have the same pixel dimension. So, a rescaling must be performed to both the developmental and independent testing data.

In our case, we rescale both the test and library images to the approximately same scale level using human visual estimation. The pixel dimension is set to be around 100 pixels/ 1 cm.

From the sample images shown earlier, one can see they have various backgrounds, which could influence the performance of the algorithm. What is more, different types of skin color may also have an impact on the result of the method. So, as to minimize all the factors except the disease itself, a region of interest (ROI) should be extracted to make sure it contains the maximum information of disease and minimum information of the background.

So, the second step in pre-processing is to manually select the region of interest (ROI) to contain only the disease and a little surrounding skin. Thus, a non-skin region like a table cloth is excluded, which could disturb the performance of the algorithm. One image may produce several ROIs, for the sake of a full representation of the disease. For record keeping, ROIs excerpted have the same image number with an additional sequential number appended. The examples are shown in Fig.5-3.

The rescaled image	ROIs	
Fluid 1903	Fluid 1903-1	
Purpura-violaceous 10501	Purpura-violaceous 10501-1	Purpura-violaceous 10501-2
	Purpura-violaceous 10501-3	

Figure 5-3   Two examples of ROI selection

The size of developmental data set after pre-processing is given in Table 5-1; the

size of independent test data set after pre-process is given in Table 5-2.

Table 5-1   Size of Developmental Data Set after Pre-process

Skin disease	Original sample size	Sample size after pre-process
fluid filled	20	50
gangrene-necrotic	12	24
pigmentation	18	42
purpura - violaceous	19	52
raised with color change	22	63
redness - general	20	40
ulcerated-eroded	18	29
warty-crusty-scabby	15	42
Sum	**144**	**342**

Table 5-2   Size of Independent Test Data Set after Pre-process

Original sample size	Sample size after pre-process
76	167

5.3   Feature identification

In this sub-section, important visual feature factors for image data retrieval are investigated, based on which image processing techniques can be implemented.

Color: Since most skin diseases have their specific color, color plays a key role in this whole process of identification, especially for those diseases that have a characterized color in a large homogenous region.

Texture: Normal skin usually has a uniformed texture pattern, distinguished from skin disease whose texture may vary a lot. Some skin diseases display a high

31

homogeneity, while some others show a vivid pattern whose intensity varies quickly in the neighborhood.

Shape: in the data set, elliptical shape or circular shape will be identified.

Table 5-3 gives a full description of visual features for every type of skin disease in the data set based upon a visual examination of all labelled images.

Table 5-3   Visual Feature Description for Each Skin Disease

disease	Color	Texture	Shape
fluid filled	Red or white	Homogenous	Elliptical
gangrene-necrotic	Black	Homogenous	No shape
pigmentation	Contrast with normal skin	Homogenous	No shape
purpura - violaceous	Dark red or reddish purple	Heterogeneous	No shape
raised with color change	Contrast with normal skin	Heterogeneous	Some are elliptical
redness-general	Red or pink	Homogenous	No shape
ulcerated-eroded	Mixed color	Heterogeneous	Some are elliptical
warty-crusty-scabby	Contrast with normal skin	Heterogeneous	Some are elliptical

5.4   Feature extraction

In this sub-section, we will discuss how to extract those features identified above.

Color: the question for using the color feature is to choose the suitable color space for our database. The most common color spaces are RGB, HSV (stands for hue, **s**aturation, **v**alue), Lab (dimension **L** for luminance and **a** and **b** for the color-opponent dimensions, based on nonlinearly-compressed CIE XYZ color space coordinates).Color models of RGB, HSV and Lab are shown in Fig.5-4, 5-5 and 5-6 separately.
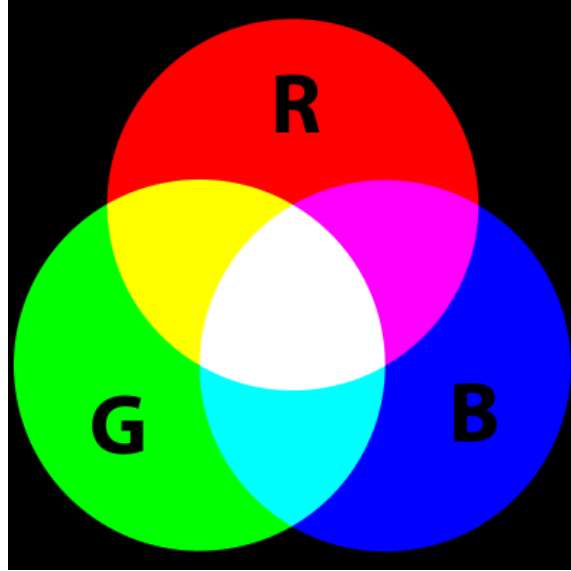
Figure 5-4   Additive primary colors (Wikipedia, 2009)
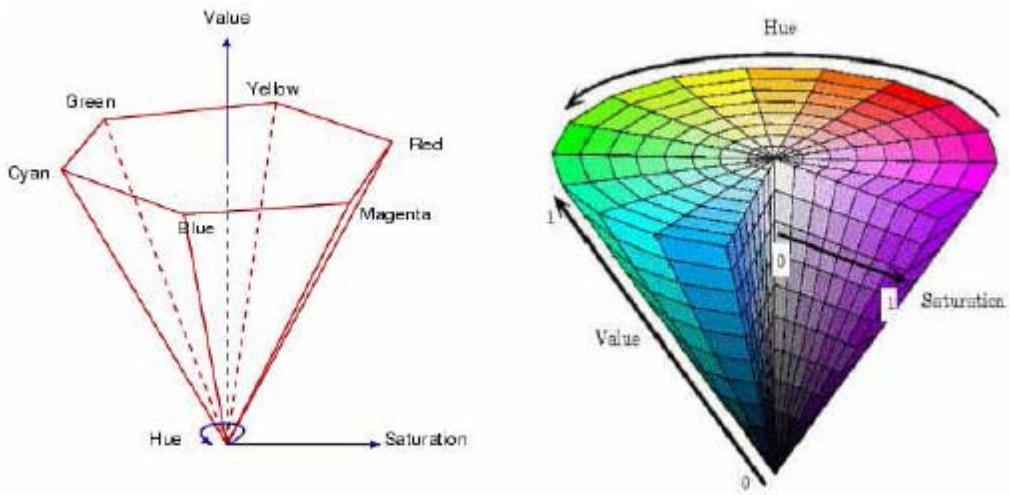
Figure 5-5   HSV coordinate system and color model (Khan, 2005)

33
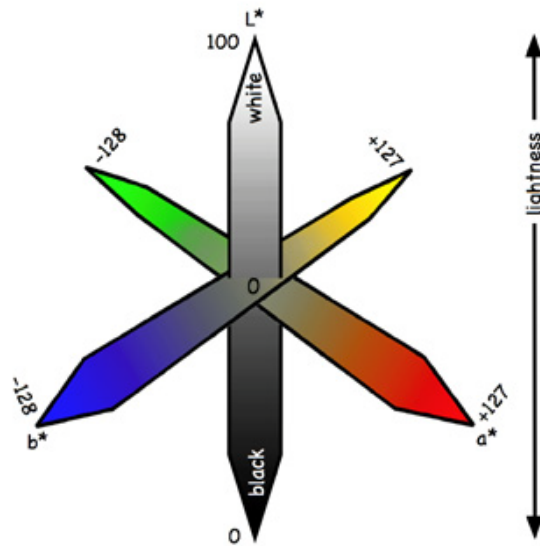
Figure 5-6   LAB color model (Khan, 2005)

In the development of the algorithm, three different color spaces were tested: HSV, Lab, RGB. Images are read in format of RGB. Therefore, we needed to perform RGB to HSV and RGB to LAB transformation (Appendix A). Table 5-4 gives the scoring results (which will be discussed later) by testing the three different color spaces in the full image retrieved process. The higher the score the better the performance is. The intermediate result shows HSV space has the best outcome.

Table 5-4   Scoring Results of Three Different Color Space

Disease	HSV space	Lab space	RGB space
fluid filled	0.64	0.54	0.60
gangrene-necrotic	0.73	0.55	0.55
pigmentation	0.38	0.29	0.33
purpura-violaceous	0.62	0.58	0.45
raised with color change	0.52	0.45	0.52
redness-general	0.67	0.49	0.50
ulcerated-eroded	0.33	0.31	0.33
warty-crusty-scabby	0.67	0.52	0.57
mean	0.57	0.47	0.48

34

In HSV color space, the "hue" parameter is the color blend, usually characterized by eight principal colors: black, white, red, yellow, green, cyan, blue and magenta. Thus, in order to make finer quantization, the Hue channel is further quantized unevenly into 20 blocks according to these eight principal colors. The quantization scheme is listed in Table 5-5. "Hue" value in Table 5-5 is normalized between 0 and 1.

The "Saturation" parameter selects how grey or pure the color will be. The "Value" parameter defines the brightness of the color. Since "saturation" and "value" do not have preferred value like "hue", they are equally quantized into 20 blocks. Thus, each channel will have 20 blocks.

Table 5-5   Quantizing Result of Hue Channel

Principal colors	Quantizing result
black	value < .2
white	saturation < .05 & value > .85
red	hue > .9167 & hue <= .96
	hue > .96 | hue <= .033
	hue > .033 & hue <= .083
yellow	hue > .083 & hue <= .125
	hue > .125 & hue <= .175
	hue > .175 & hue <= .25
green	hue > .25 & hue <= .3
	hue > .3 & hue <= .35
	hue > .35 & hue <= .4167
cyan	hue > .4167 & hue <= .47
	hue > .47 & hue <= .52
	hue > .52 & hue <= .5833
blue	hue > .5833 & hue <= .64
	hue > .64 & hue <= .7
	hue > .7 & hue <= .75
magenta	hue > .75 & hue <= .80
	hue > .80 & hue <= .85
	hue > .85 & hue <= .9167

35

Figure 5-7   An example image for color feature extraction

Table 5-6   An Example of Color Feature Vector

color space components	Extracted feature vector
Hue	[0,0,0,0.16,0.84,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
Saturation	[0,0,0,0,0,0.02,0.04,0.08,0.12,0.21,0.22,0.18,0.09,0.03,0,0,0,0]
Value	[0,0,0,0,0,0.01,0.03,0.08,0.13,0.15,0.13,0.11,0.10,0.11,0.10,0.03,0.01,0,0,0]

As an example, for the image shown in Fig.5-7, the extracted color

features are listed in Table 5-6.

Texture: Haralick's texture feature (explained in Appendix A) is employed to describe

the texture pattern. First, the co-occurrence matrix is computed. The number of

gray-levels is scaled into 64 levels, so as to reduce the computation. Considering

all directions equally, we generate four co-occurrence matrixes at equally spaced

directions with a displacement of two pixels, as shown in Table   5-7 and Fig.5-8.

Table 5-7   Displacement Vector of Co-occurrence Matrix in Four Direction

Angle	Displacement vector d
0°	[0, 2]
45°	[-2, 2]
90°	[-2, 0]
135°	[-2,- 2]

36
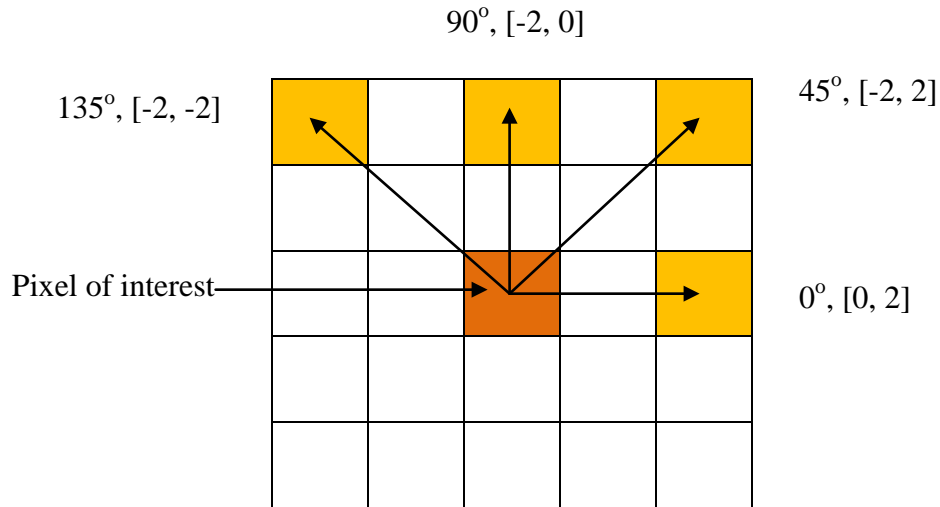
Figure 5-8   Illustration of the four displacement vectors

Since we want to focus on the homogeneity property of the texture pattern,

contrast and *homogeneity* features of the gray level co-occurrence matrix

(GLCM) are selected to represent the texture property. (The definitions of

contrast and *homogeneity* are shown in Appendix A). Thus, the texture feature,

we have 2×4 = 8 feature parameters. Since the range of *homogeneity* is [0, 1], it

does not need normalization. The theoretical range of *contrast*

is $[0\ (size(GLCM) - 1)^2]$. $size(GLCM)$ is 64 in our case, so the range is [0,

3969]. However, the practical results of *contrast* are far below this maximum

value. As a result, the normalization of *contrast* is to divide the current *contrast*

value by the maximum value among the current *contrast* value and those of the

library images.

As an example, the extracted texture features of Fig.5-7 are listed in

Table 5-8.

Table 5-8   An Example of Texture Feature Vector

Texture property	Texture feature vector			
	0°	45°	90°	135°
Contrast	6.75	7.08	2.40	7.07
Normalized *contrast*	0.07	0.07	0.02	0.07
Homogeneity	0.51	0.50	0.62	0.49

Shape: since the images to be analyzed are all extracted ROIs, they just contain the diseased area and normal skin as background. So, a two class K-means clustering (explained in Appendix A) resolution is manipulated to separate the object from background. For conformity with color feature extraction, HSV space is used to generate vectors for classification. K-means clustering is performed on the *Value* channel.

After K-means clustering, the image is divided into two classes: the object of interest and the background. The *Canny* descriptor (Green, 2002) is employed to extract edges of the object.

For those images containing elliptical features (such as fluidfilled or ulcerated), the edges are more close to an ellipse than those images that have no evident elliptical features (such as pigmentation, gangrene and redness). Using this observation, we perform elliptical regression technique, which finds the optimization ellipse based on least square errors. Fig.5-9 illustrates an example of the whole process.

a) The image after pre-process	b) The result of 2-means clustering
c) The edge map extracted by Canny descriptor	d) The regressed ellipse
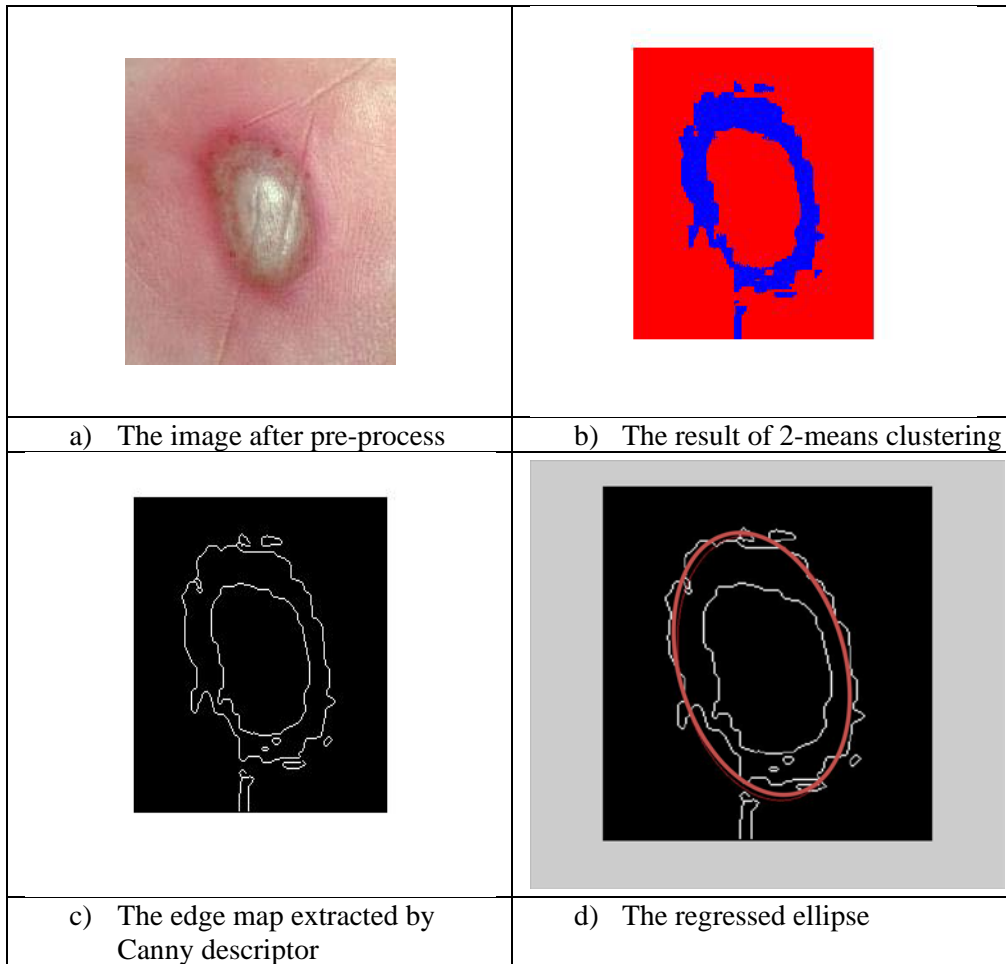
Figure 5-9   An example of elliptical regression

To characterize this elliptical shape feature, one wants to know how close is the contour of the object to the regressed ellipse. The most commonly used method to measure the distance between two contours is the Hausdorff Distance (see in Appendix A). So, Hausdorff Distance was selected initially as the final shape feature. However, to reduce the misinterpretation of the contour caused by some outliers, Improved Hausdorff Distance is also investigated in the test. In the final algorithm, the Improved Hausdorff Distance was chosen to represent the shape feature parameter.

Considering the impact of image size on the Hausdorff Distance, even if two contours have the same similarity to an ellipse, the Hausdorff Distance between the contour and the regressed ellipse differs due to different image size - the larger the size, the greater the Huasdorff Distance. To eliminate the influence caused by image's size, Huasdorff Distance is divided by one half of the number of pixels along the image's diagonal as normalization.

As an example, the extracted texture features of Fig.5-7 are listed in Table 5-9.

Table 5-9   An Example Shape Feature Parameter

Shape property	Shape feature parameter
Hausdorff Distance	15.33
Normalized Hausdorff Distance	0.13
Improved Hausdorff Distance	41.23
Normalized Improved Hausdorff Distance	0.36

5.5   Similarity measure

The extracted color features are three vectors – the quantized histogram blocks for each channel of HSV space. Histogram intersection is carried out to measure the similarity of histogram distribution in each channel between the test image and the reference image from the image library. The intersection result is normalized so that the outcome of each channel is a value between 0 and 1.

Now, we define the histogram intersection result of Hue, Saturation and Value as D_h, D_s, D_v. And the distant metric for color feature is defined as:

$$D_{color} = \frac{1}{3} \cdot D_h + \frac{1}{3} \cdot D_s + \frac{1}{3} \cdot D_v \qquad \text{Equation 5-1}$$

We give the same weight to each channel, so that each component of the HSV space will have an equal impact on the result. The final distant metric is also normalized

40

into [0, 1]. The more similar the color of the test image is to that of the reference image, the larger the color distant metric.

The outcomes of texture and shape features are 9 scalar values normalized from 0 to 1. The feature f vector for texture and shape consists of 4 *contrast* features (c_1, c_2, c_3, c_4), 4 *homogeneity* features (h_1, h_2, h_3, h_4), and 1 Improved Hausdorff Distance (I). Thus, regarding these 9 scale values as a vector in a 9 dimensional space:

$$f = [c_1, c_2, c_3, c_4, h_1, h_2, h_3, h_4, I]' \qquad \textbf{Equation 5-2}$$

The similarity of texture and shape feature between the test image and library image is the Euclidean distance of their vectors. The greater the similar they are, the smaller the distance is.

5.6   Self-developed scoring system

In section 3.4, the standard evaluation method for content-based image retrieval was introduced. However, the *Precision vs. Recall* graph is drawn while adjusting the number of images retrieved from 1 to the size of the whole image library. So, the *Precision vs. Recall* graph can be regarded as an evaluation for the image retrieval system, but not an evaluation based on a given number of retrieved images. In our case, the number of images retrieved is fixed at five, so we develop a scoring system particular to this preset range without changing the number of images retrieved.

Each result in the five best candidates should be weighted differently. Because the higher in the order a given image is in the five best candidates, the larger weight the image has in the doctor's judgment. So, we weight the five best matches in the following order: 5/5, 4/5, 3/5, 2/5, 1/5.

41

Besides this weight of order, we set another weight according to the content of the image. For each image of the five best candidates, if the image is in the same class of the input image, it is considered as a correct result, and its weight is 1. Else, it is considered as a wrong result, and its weight is 0.

To combine these two kinds of weights above, we multiply the weights correspondingly for each image in the five best candidates, and average them as the score for this particular input test image. So, the maximum of the score is :

$1*(5/5+4/5+3/5+2/5+1/5) = 3$, and the minimum of the total weight is

$0*(5/5+4/5+3/5+2/5+1/5) = 0$. So, a normalized score is obtained by dividing the score by 3.

If the input is a classified data set, then we can evaluate the retrieval system performance by computing the score for each class of disease. Suppose a single input image c_i is from class $C: \{c_1, c_2, c_3 \ldots, c_I\}$. After image retrieval, the individual input image has five best candidates $d_i: \{d_{i1}, d_{i2}, d_{i3}, d_{i4}, d_{i5}\}$. The score of each candidate is:

$$score_{in} = W_{in} * W_g \qquad \textbf{Equation 5-3}$$

W_{in} is the weight of order, $W_{i1} = \frac{5}{5}$, $W_{i2} = \frac{4}{5}$, $W_{i3} = \frac{3}{5}$, $W_{i4} = \frac{2}{5}$, $W_{i5} = \frac{1}{5}$.

W_g is the weight of content. If the candidate is from the same class of the input image, $W_g = 1$; else, $W_g = 0$.

So, the normalized score of this particular input image is:

$$score_i = \sum_{n=0}^{5} score_{in} / range \qquad \textbf{Equation 5-4}$$

$$range = max - \min = 1 * \left(\frac{5}{5} + \frac{4}{5} + \frac{3}{5} + \frac{2}{5} + \frac{1}{5}\right) - 0 * \left(\frac{5}{5} + \frac{4}{5} + \frac{3}{5} + \frac{2}{5} + \frac{1}{5}\right) = 3$$

The averaging score of this class is:

42

$$score_c = \frac{1}{I}\sum_{i=1}^{I} score_i \qquad \textbf{Equation 5-5}$$

I is the sample size of the class.

The whole scoring process for all images in each class C is shown in Fig.5-10.
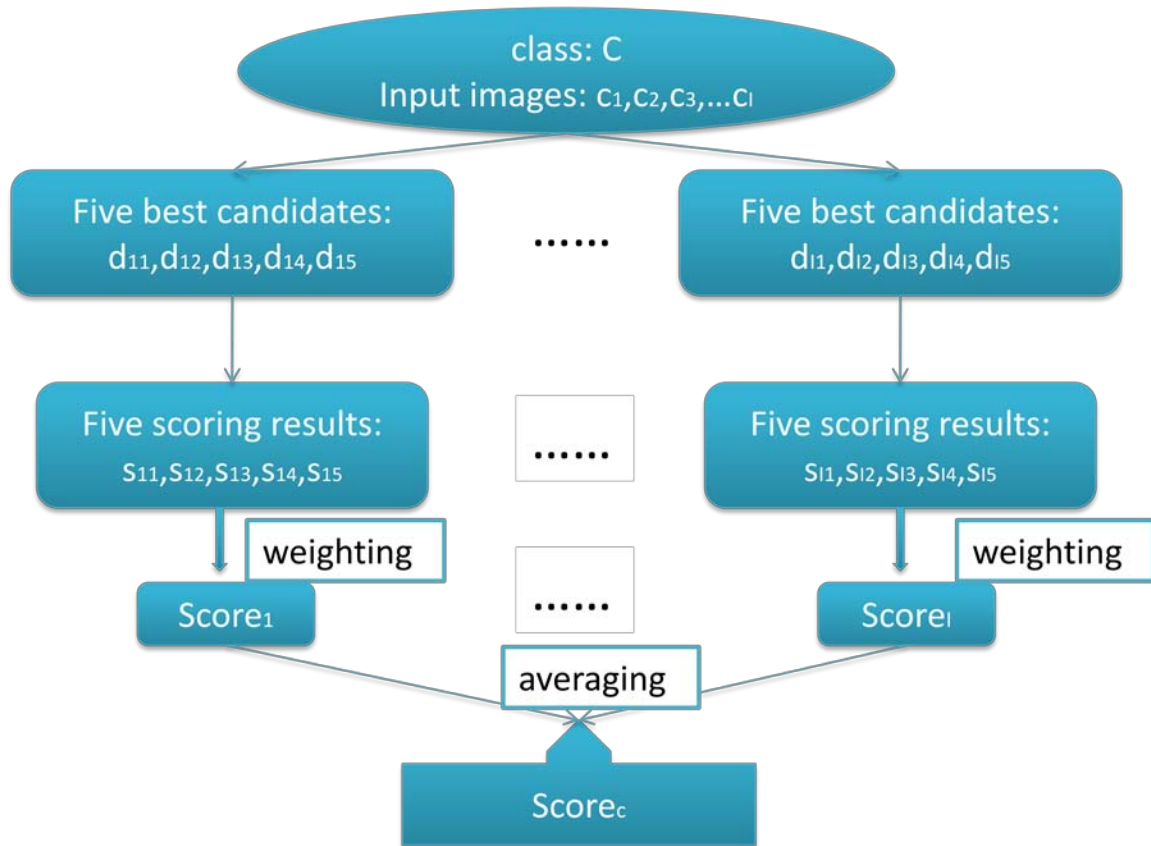
Figure 5-10   The flowchart of scoring process

5.7   Individual feature performance

The algorithms for image retrieval only based on the color, texture and shape feature individually are illustrated in Fig. 5-11, Fig.5-12 and Fig.5-13.
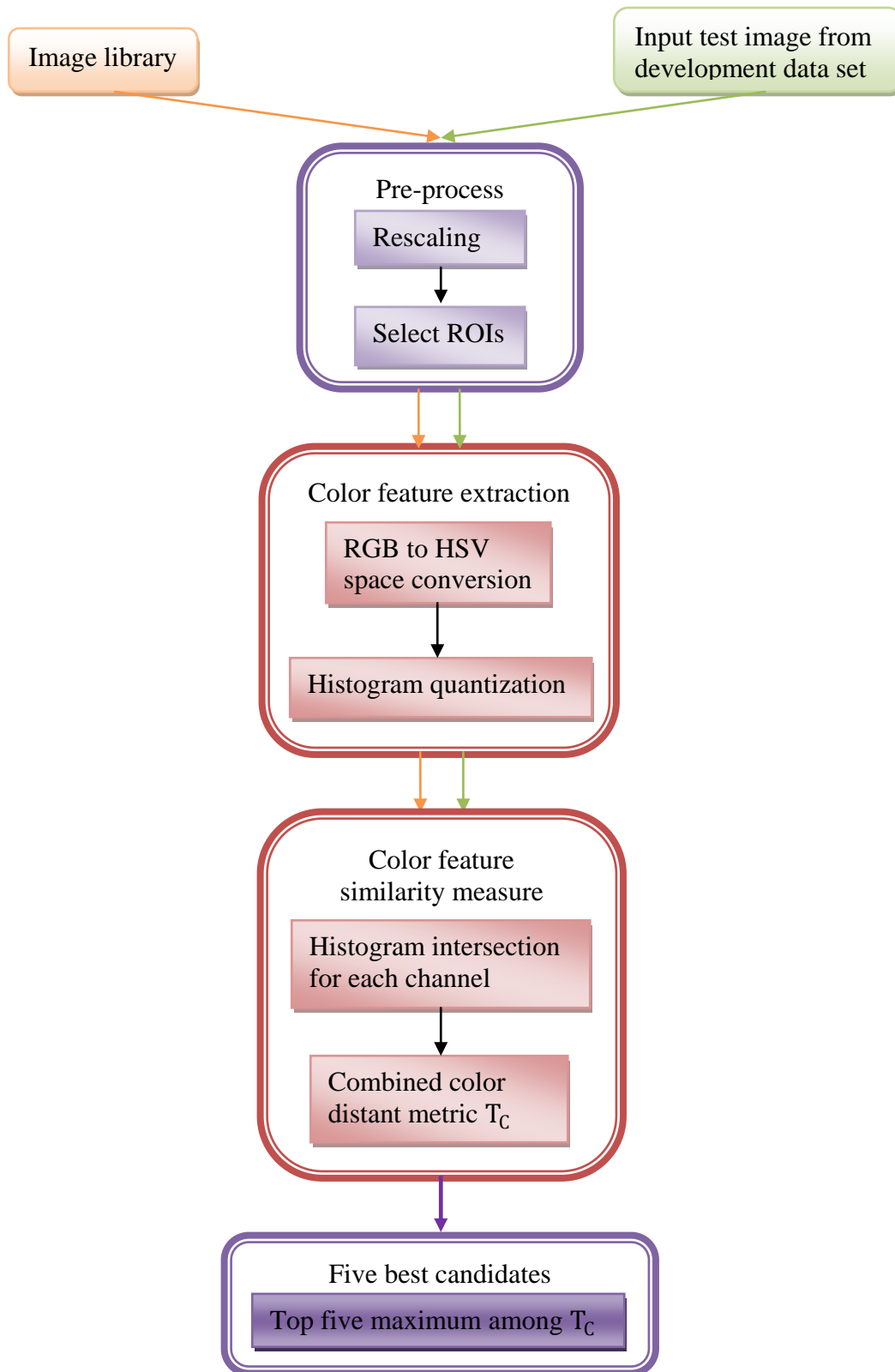
Figure 5-11   Flowchart of algorithm with color feature only
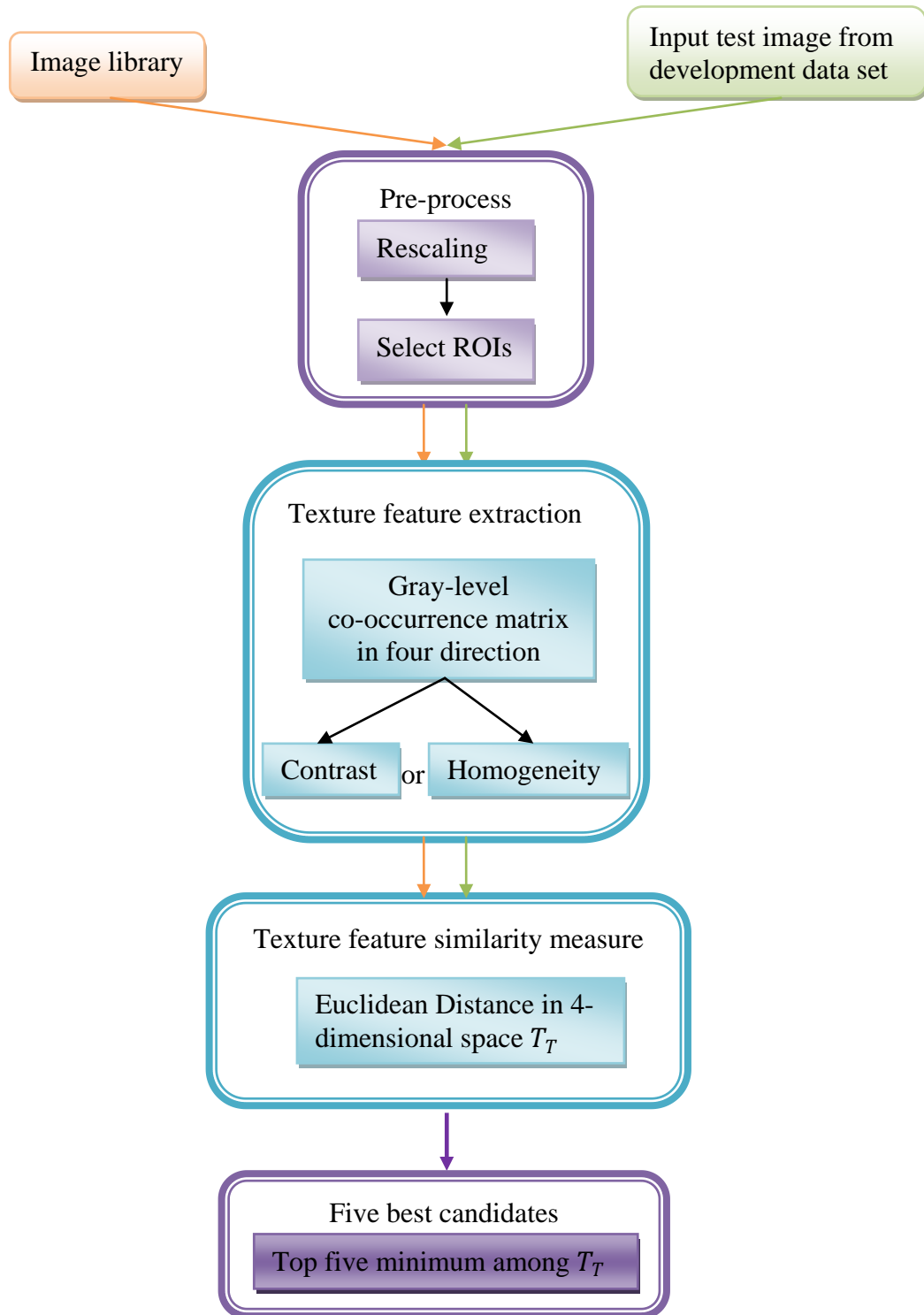
Figure 5-12   Flowchart of algorithm with texture feature only
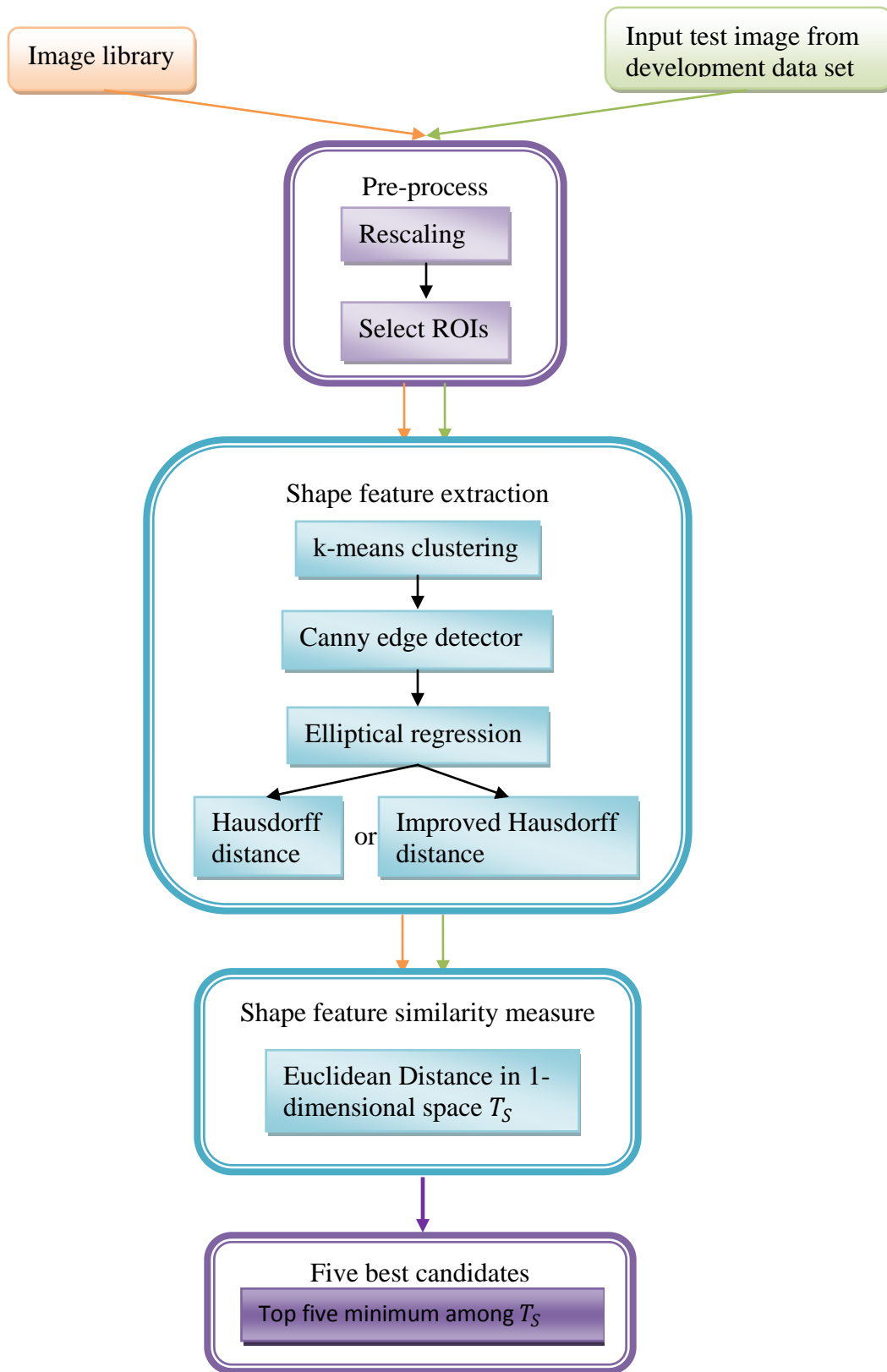
Figure 5-13   Flowchart of algorithm with shape feature only

The scores of each disease for the three systems are listed in Table 5-10. And Fig.

5-14 gives the chart of the result.

Table 5-10   The Scoring Results of Each Individual System

Class	color	texture		shape	
		contrast	homogeneity	Hausdorff distance	Improved Hausdorff distance
fluid filled	0.33	0.22	0.22	0.07	0.14
gangrene-necrotic	0.41	0.08	0.10	0.04	0
pigmentation	0.29	0.16	0.17	0.14	0.22
purpura - violaceous	0.36	0.17	0.21	0.14	0.18
raised with color change	0.45	0.29	0.28	0.15	0.14
redness-general	0.61	0.37	0.34	0.14	0.18
ulcerated-eroded	0.51	0.10	0.14	0.12	0.15
warty-crusty-scabby	0.45	0.29	0.21	0.13	0.11
sum	3.41	1.68	1.67	0.93	1.12
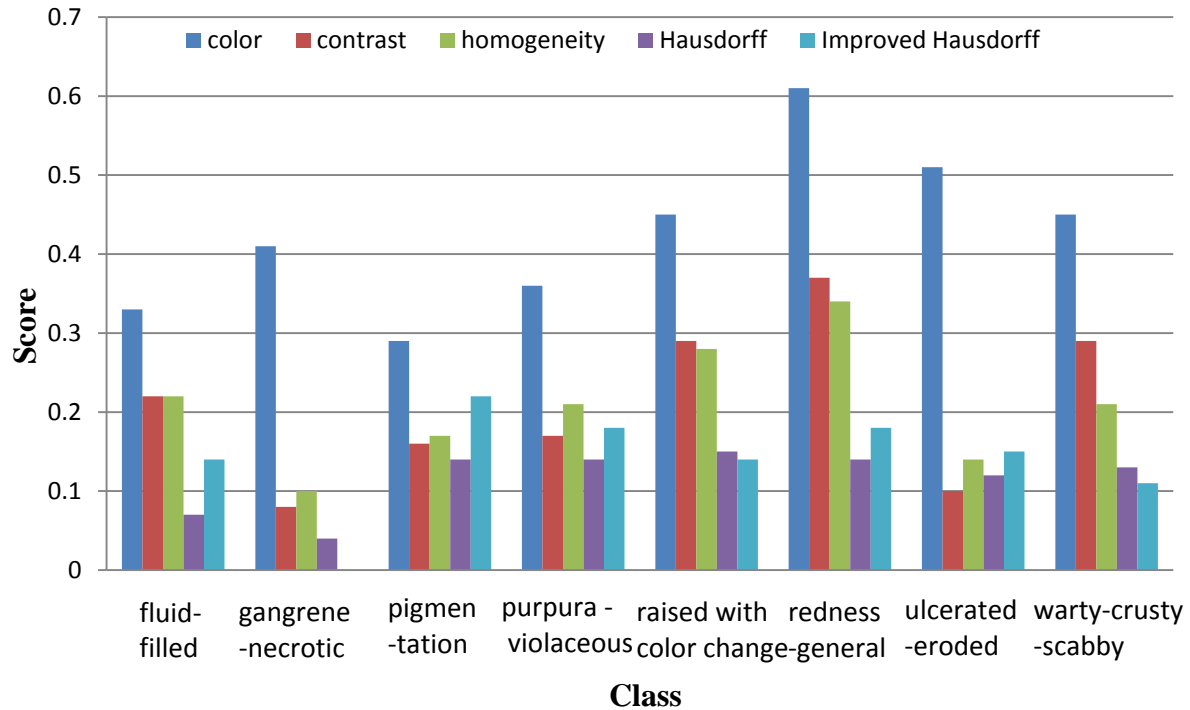
Figure 5-14   The chart of score for individual systems

From the table above, the performance of Improved Hausdorff distance is a little better than Hausdorff distance. So we choose Improved Hausdorff distance for the shape feature.

From the chart, we can see the system using only the color feature performed best and high above the other two features. Next is the system with only the texture feature. The worst is the system with only the shape feature.

5.8   Distance metric combination

The task of image retrieval algorithm is to find the five best candidates for the test image. For the color feature separately, they are the five library images that can produce the top five maximums in the similarity measure step. For the texture and shape features

www.manaraa.com

separately, there are the five library images that can produce the top five minimums in the similarity measure step.

So, since texture and shape features are best when minimum, they can be integrated into a vector of Euclidean space in the similarity measure as discussed in section 5.4. The system with combined texture and shape features is shown in Fig. 5-15.
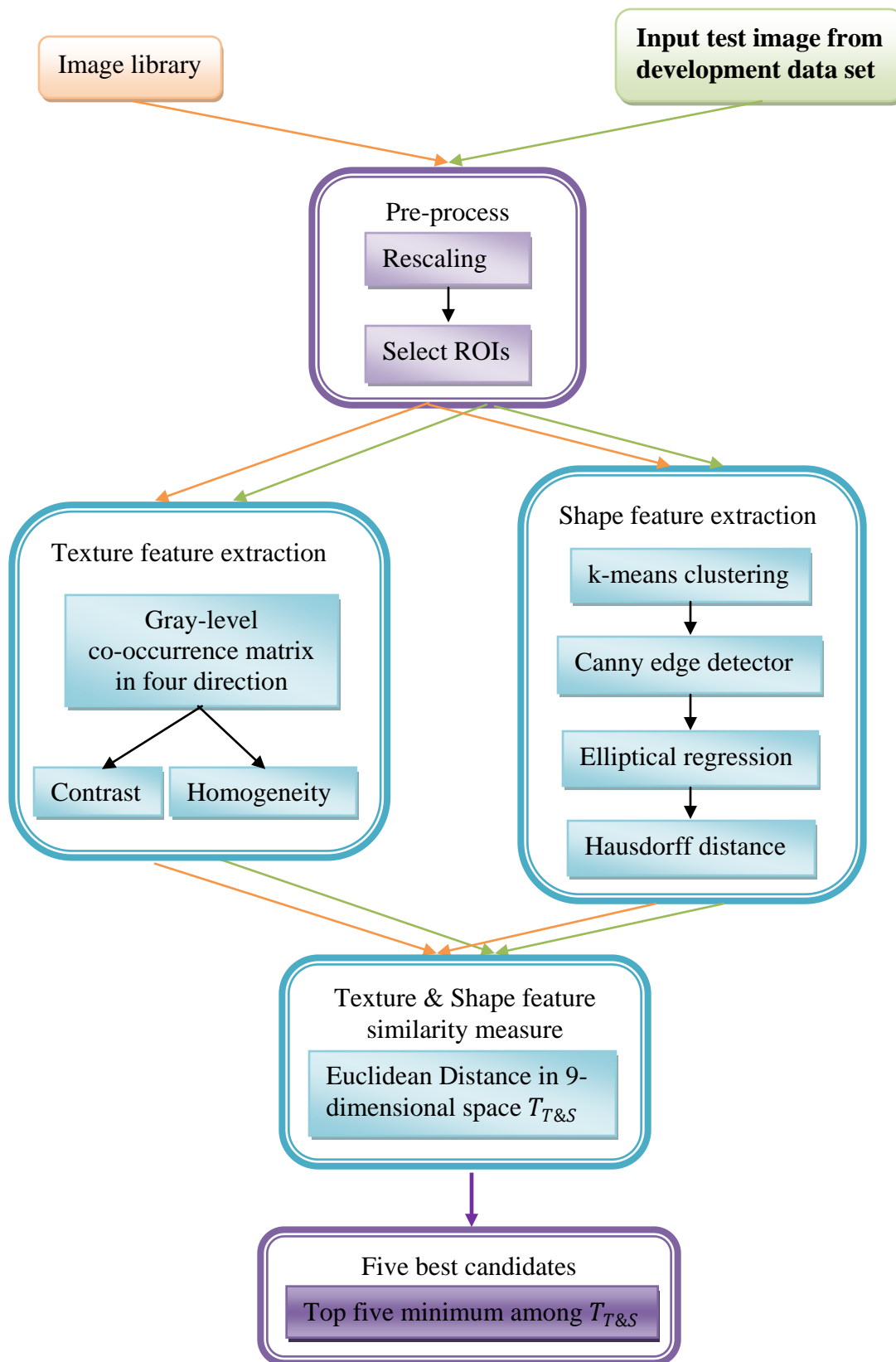
Figure 5-15   Flowchart of algorithm with texture and shape feature

Since the similarity measure methods for the color feature varies from that of texture and shape feature, a combination of the two distance metrics was developed. We implemented an ad hoc approach to combine these two distance metrics based on the result of the developmental data set.

Suppose we have two distance metrics A and B. The similarity metrics of developmental data set generated by A are $a_1, a_2, ..., a_{na}$, the training similarity metric generated by B are $b_1, b_2, ..., b_{na}$. The range of them are R_A, R_B. The score of distance metric A and B are S_A, S_B. The similarity metrics of an independent test data generated by A and B are t_A, t_B. These parameters are illustrated in Fig. 5-16.
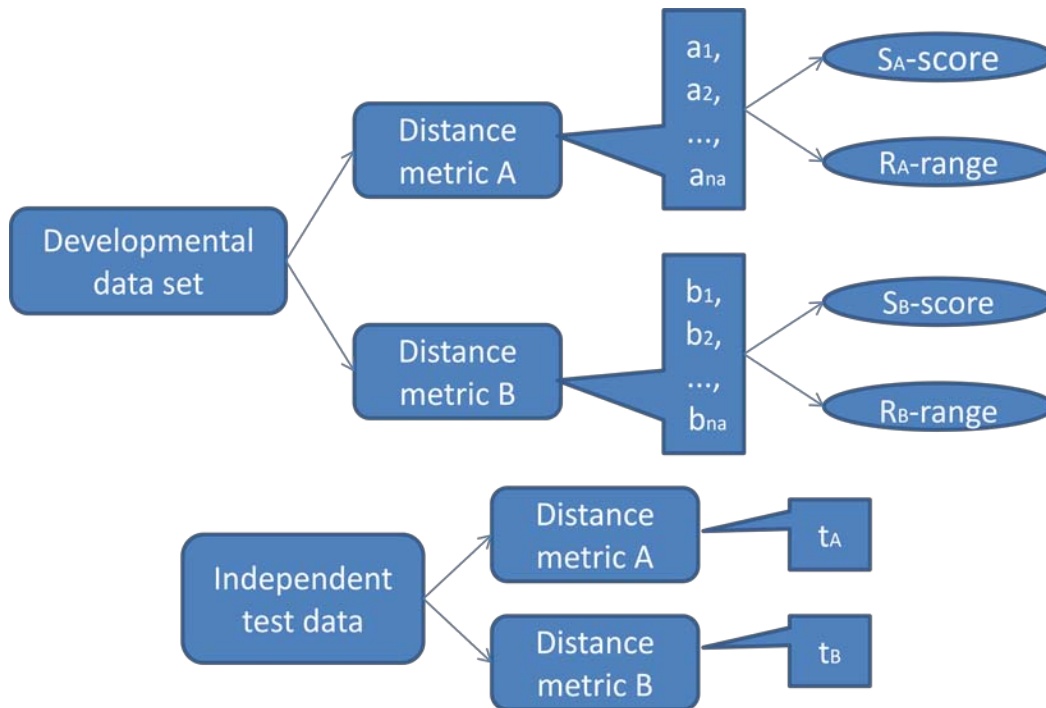
Figure 5-16   Illustration of distance metric combination scheme

Then, the combined parameter for choosing the five best candidates is:

$$c_{A,B} = S_A \cdot \frac{t_A}{R_A} + S_B \cdot \frac{t_B}{R_B}$$

Equation 5-6

51

It is based on the idea that the higher the score of the distance metric's performance in the training data set, the more weight it carries in the combined distance metric. And each component in the combined distance metric is normalized by dividing by their data range.

In our case, equation 5-6 would be:

$$C_{C,T\&S} = S_C \cdot \frac{T_C}{R_C} - S_{T\&S} \cdot \frac{T_{T\&S}}{R_{T\&S}}$$ **Equation 5-7**

The coefficients S_C and $S_{T\&S}$ are the weights for color and texture & shape component. The higher score of the component the more weight the component in combination. Thus, coefficient S_C is defined as the sum of the scores for all the classes using the image retrieval system based on color feature only. $S_{T\&S}$ is defined as the sum of the score for all the classes using the image retrieval system based on texture & shape features only.

$$S_C = \sum score_C$$ **Equation 5-8**

$$S_{T\&S} = \sum score_{T\&S}$$ **Equation 5-9**

$score_C$ and $score_{T\&S}$ are obtained by flowchart in Fig.5-17.
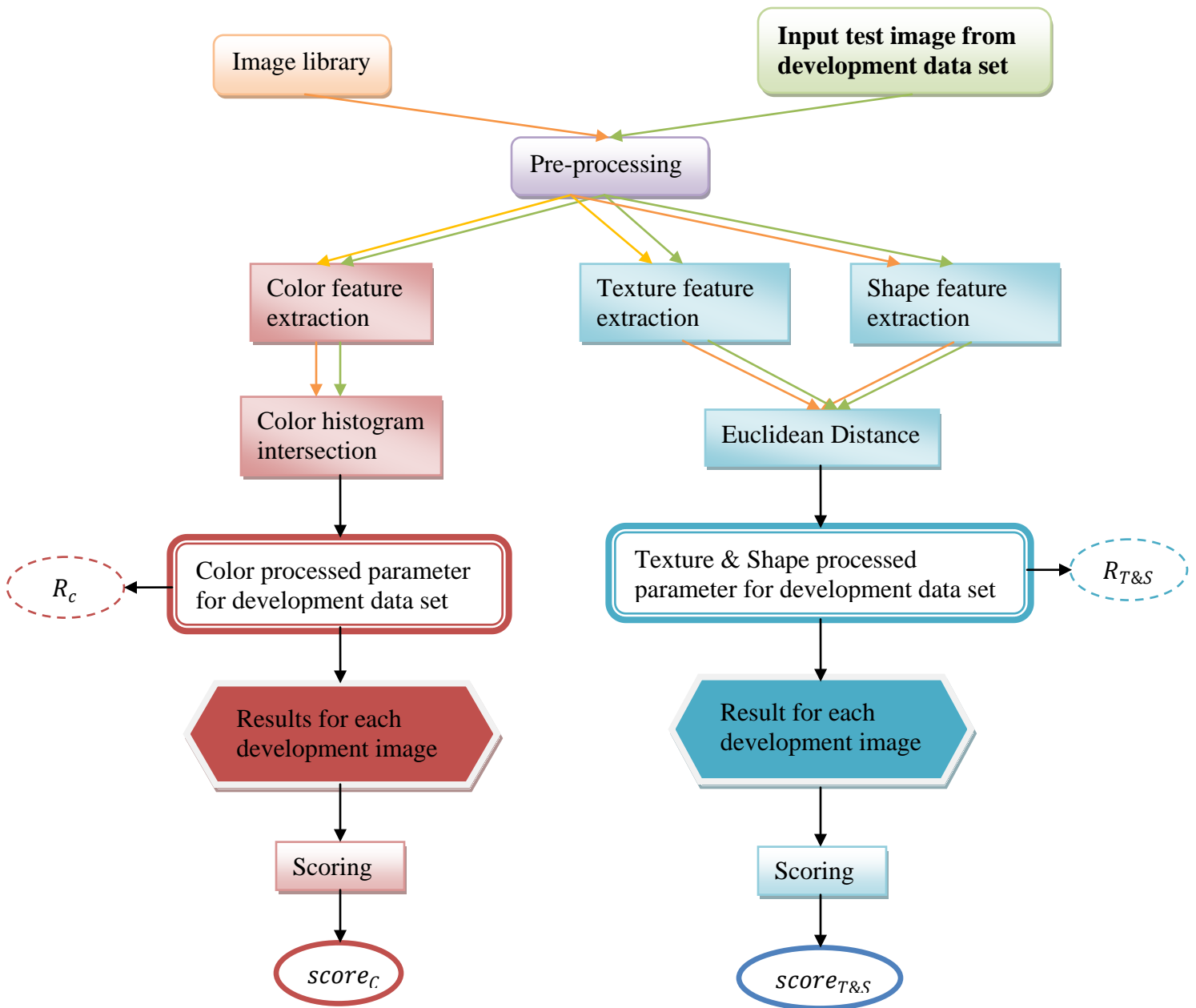
Figure 5-17   Flowchart of procedure for score$_C$ and score$_{T\&S}$

T_C is the similarity metric generated by color feature similarity measure, $T_{T\&S}$ is the similarity metric generated by texture and shape features similarity measure. They are obtained through the flowchart in Fig.5-18.
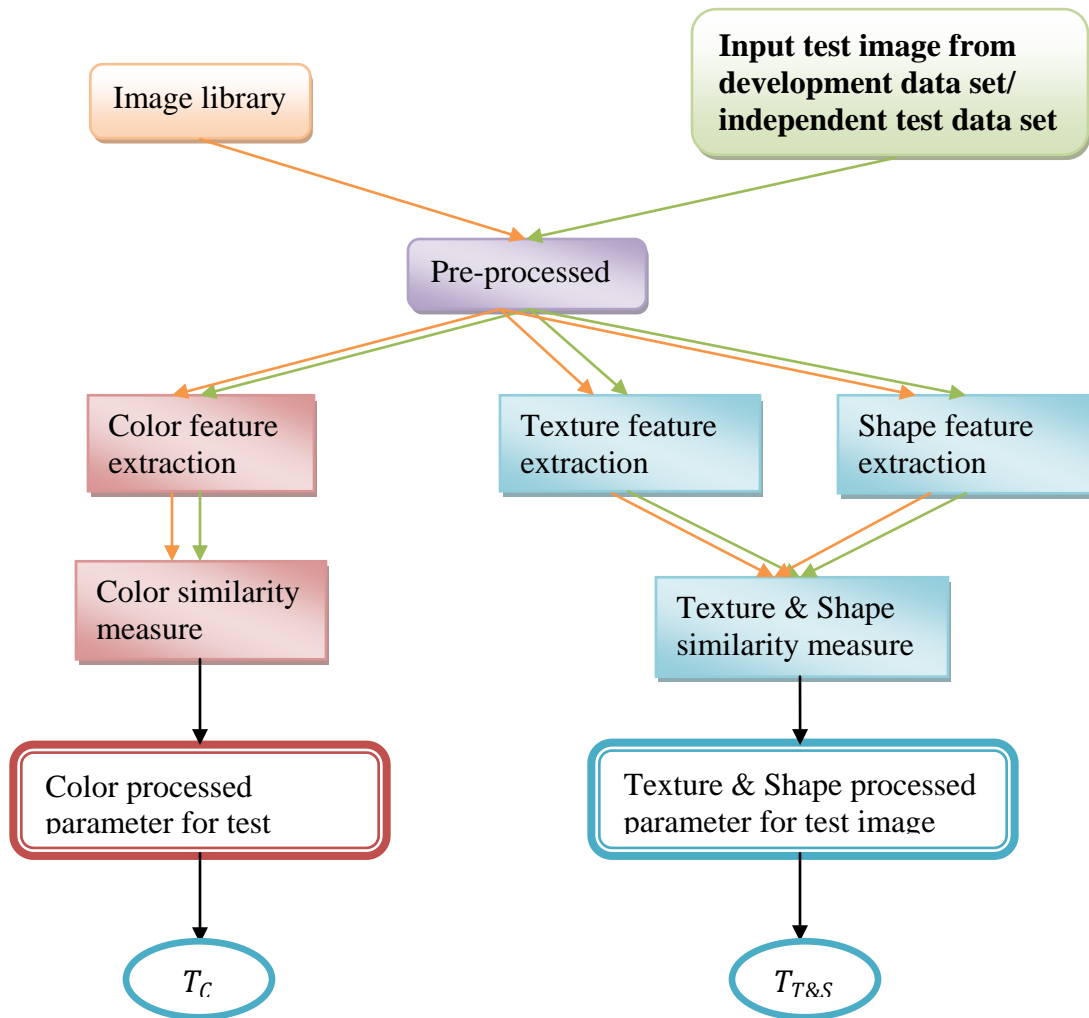
Figure 5-18   Flowchart of procedure for T_C and $T_{T\&S}$

Color processed parameter T_C is the normalized histogram intersection result, so the range of T_C: $R_C = [0\ 1]$. Texture & Shape processed parameter $T_{T\&S}$ is the Euclidean Distance between two vectors in a nine dimensional space, each dimension of the vector has a range between 0 and 1. So, the range of T_C: $R_{T\&S} = \left[0\ \sqrt{9}\right] = [0\ 3]$.

The larger T_C is, and the smaller $T_{T\&S}$ is, the closer are the input test image and the library image. So, we add a minus sign in front of the Texture & Shape component, so that the larger $C_{C,T\&S}$ is, the more agreement between the input test image and the library image. After searching through the whole image library, the five images that produce the

54

www.manaraa.com

top five maximum $C_{C,T\&S}$ scores with the input test image are chosen as the five best candidates.

The scoring results $score_C$ and $score_{T\&S}$ for developmental data set are listed in Table 5-11.

Table 5-11   Scoring Results of Systems with Color and Texture & Shape Feature

Class	$score_C$	$score_{T\&S}$
fluid filled	0.33	0.22
gangrene-necrotic	0.41	0.10
pigmentation	0.29	0.32
purpura - violaceous	0.36	0.25
raised with color change	0.45	0.35
redness-general	0.61	0.40
ulcerated-eroded	0.51	0.17
warty-crusty-scabby	0.45	0.34
sum	**3.41**	**2.15**

Therefore, $S_C = 3.41$, $S_{T\&S} = 2.15$. So Equation 5-7 will be:

$$C_{C,T\&S} = 3.41 \cdot \frac{T_C}{1} - 2.15 \cdot \frac{T_{T\&S}}{3} \qquad \textbf{Equation 5-10}$$

6. Results

This section is divided into two parts: results for the developmental data set and results for the independent test data set. In each part, we will give an overview of the algorithm, and then provide the five best candidates. At last, the algorithm performance evaluation will be discussed.

6.1   Results for developmental data set

In this sub-section, an overview of the algorithm is given in Fig.6-1. Then, a sample result is shown in Fig.6-2. At last, the results using the standard precision vs. recall method and the self-developed scoring method will be presented.

57
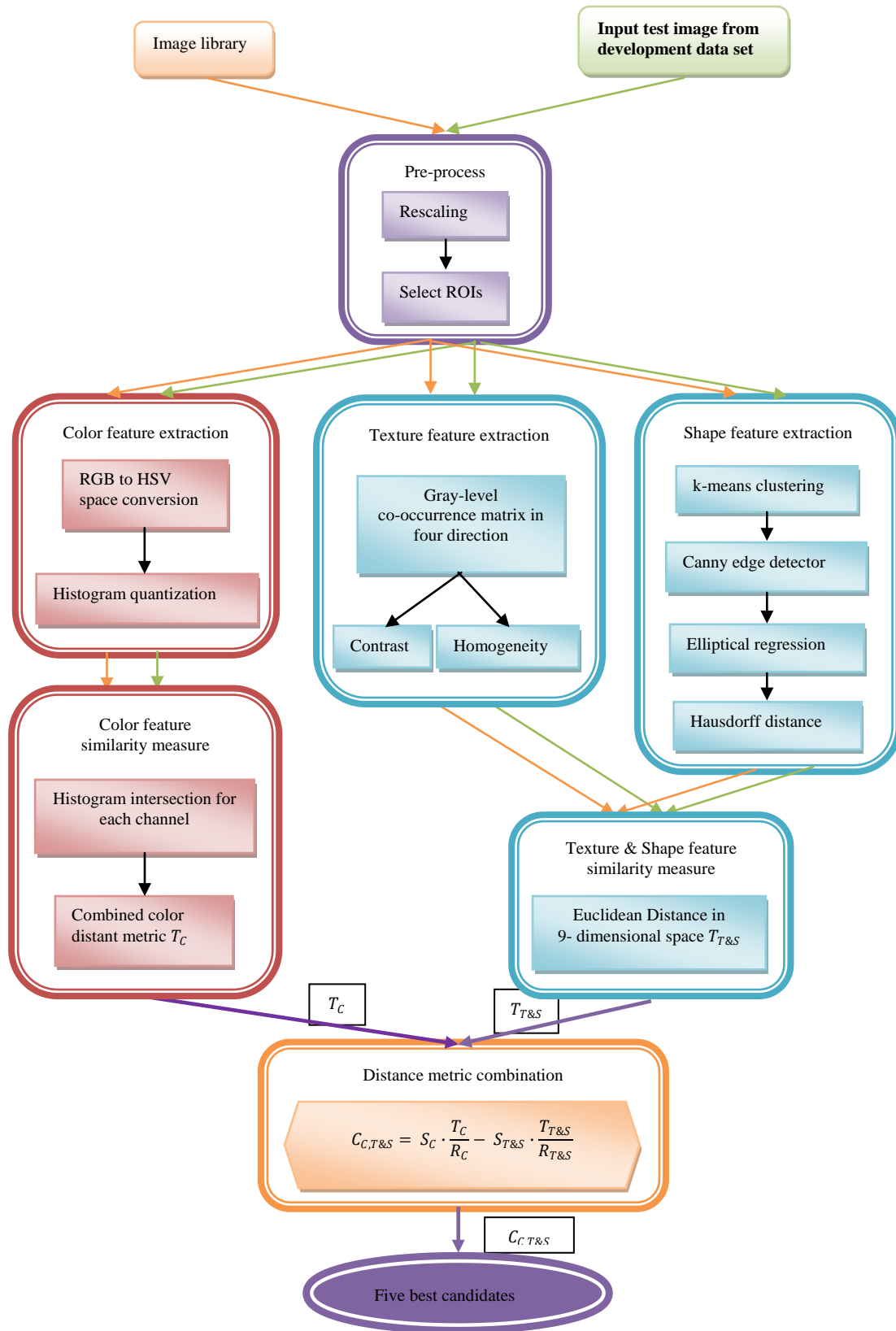
Figure 6-1   The image retrieval algorithm for developmental data set

58

Figure 6-2   Results for "gangrene-necrotic\0074-1.jpg" from developmental data set

The top left image in Fig. 6-2 is the input test image after pre-processing. The top right image is the best candidate from the library image, the left image in the second row is the second best candidate, the right image in the second row is the third best candidate, the left bottom image is the fourth best candidate, and the right bottom image is the fifth best candidate. Doctors can make judgment based on these candidates.

We notice that the best match "gangrene-necrotic\0074-2" is a ROI excerpted from the same original image as input "Gangrene-necrotic\0074-1", since they have the same root number.

www.manaraa.com

"Raised with color change\05059-1" ranks in the 2nd place, for its similarity in texture and color. "Purpura-violaceous\24051-1" ranks in the 3rd place as it also contains black and red colors. However, "gangrene-necrotic\07948-1" has neither red color as "Purpura-violaceous\24051-1", nor similar texture pattern as "raised with color change\05059-1", so it ranks in the 4th place.

The score for this example is shown in Table 6-1.

Table 6-1   **Score for the Result in Fig.6-2**

Retrieved images	W_n	W_g	$Score_n$
Gangrene-necrotic\0074-2	1	1	1
Raised with color change\05059-1	4/5	0	0
Purpura-violaceous\24051-1	3/5	0	0
Gangrene-necrotic\07948-1	2/5	1	2/5
Purpura-violaceous\27644-1	1/5	0	0
Score			0.47

6.1.1   The standard performance evaluation

In section 3.4, we introduced the standard performance evaluation method. We revise Equation 3-1 and 3-2 into our case:

$$precision = \frac{\#retrieved\ images\ from\ the\ same\ class\ of\ the\ query\ image}{\#retrieved\ images} \qquad \textbf{Equation 6-1}$$

$$recall = \frac{\#retrieved\ images\ from\ the\ same\ class\ of\ the\ query\ image}{the\ size\ of\ the\ class\ to\ which\ the\ query\ image\ belongs} \qquad \textbf{Equation 6-2}$$

To draw the *precision* vs. *recall* graph, we set the range of "# retrieved images" to vary from 1 to the size of the whole developmental data set. For each class of disease, we take average *precision* and *recall* among all images of the class as the representative for the class. The *precision* vs. *recall* graph for each disease is shown in Fig.6-3 to Fig. 6-10.

60
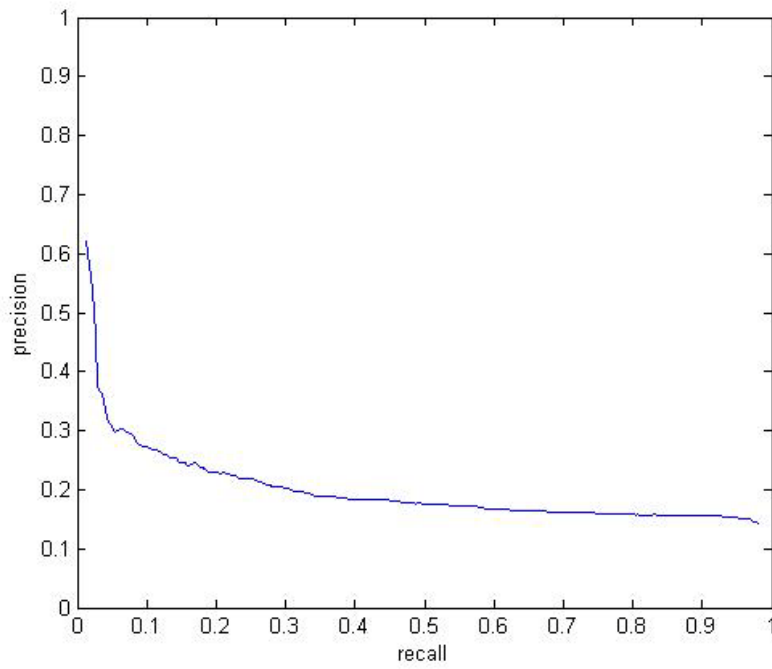
Figure 6-3 The *precision* vs. *recall* graph for fluidfilled
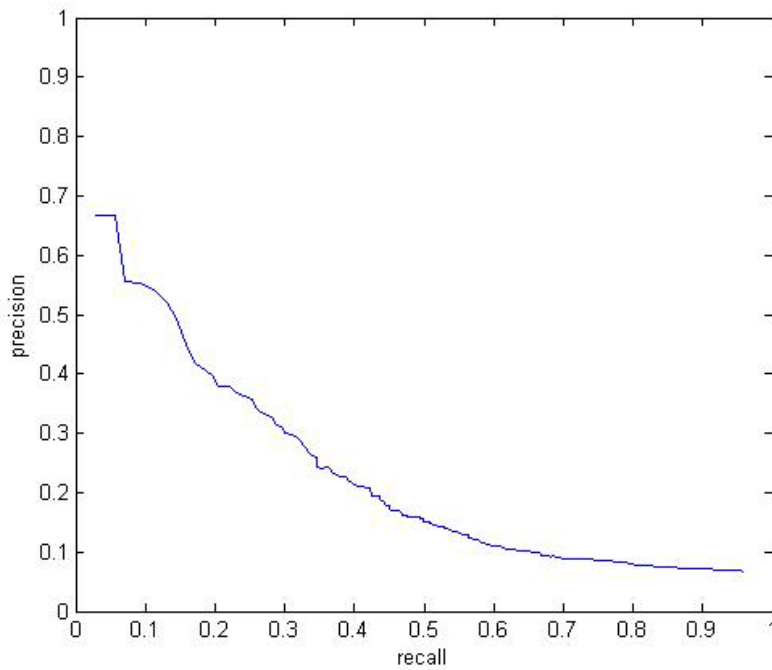
Figure 6-4 The *precision* vs. *recall* graph for gangrene-necrotic

61
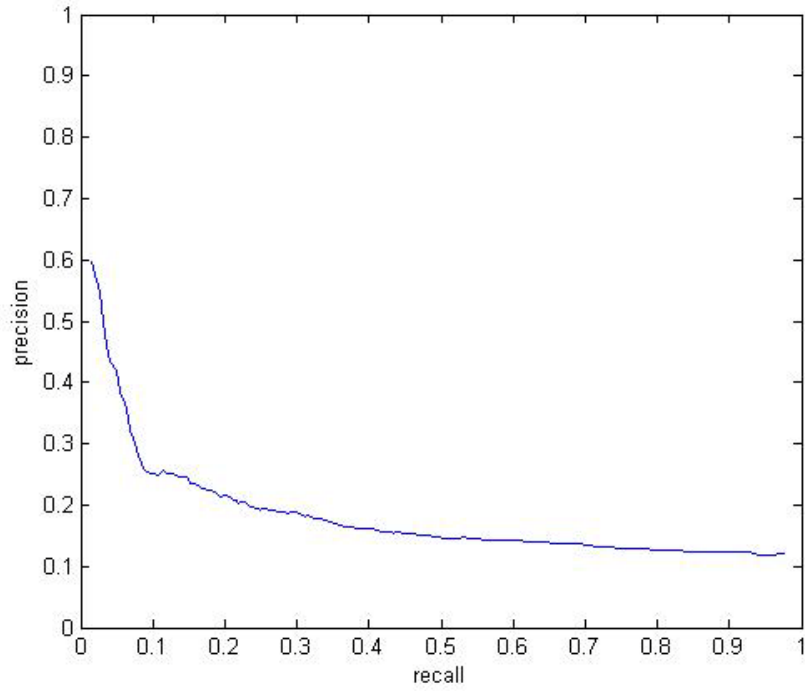
Figure 6-5 The *precision* vs. *recall* graph for pigmentation
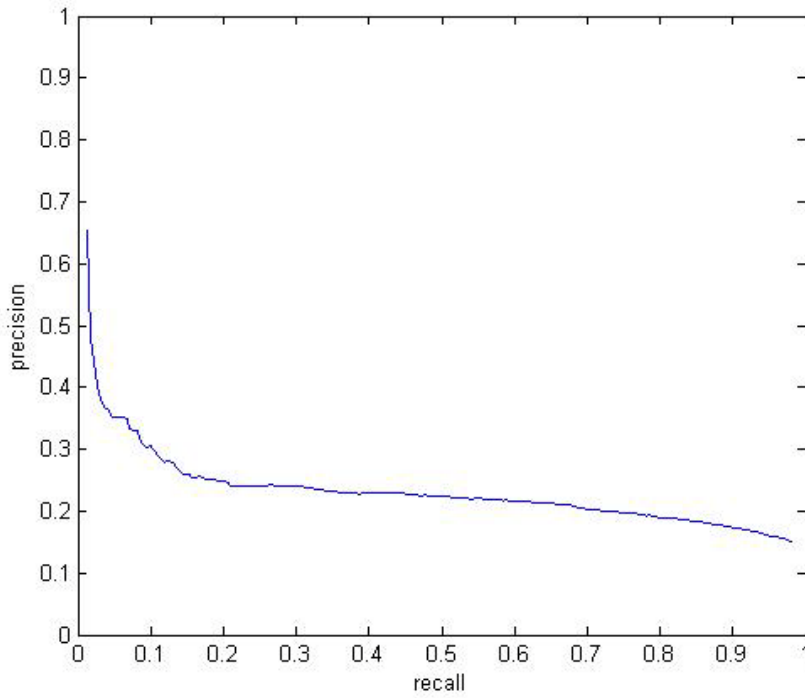
Figure 6-6 The *precision* vs. *recall* graph for purpura-violaceous

62
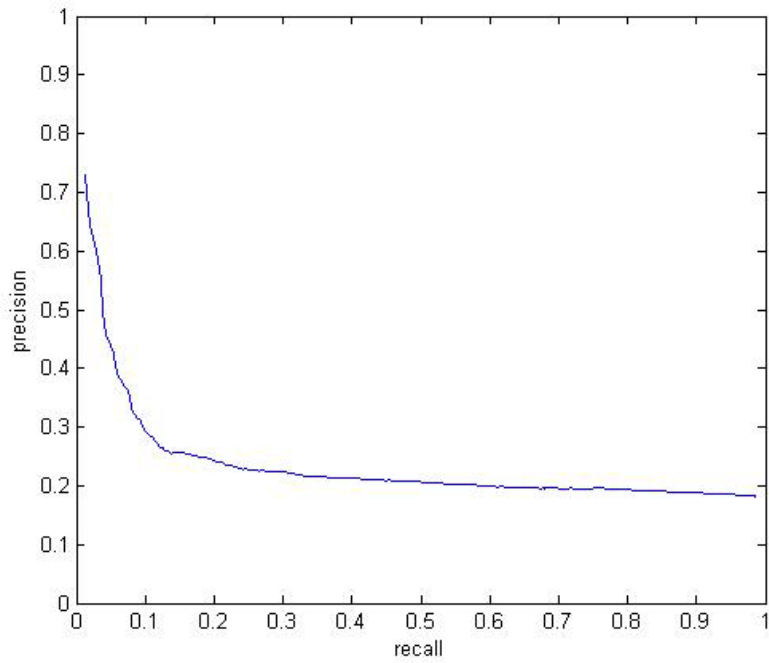
Figure 6-7   The *precision* vs. *recall* graph for raised_with_color_change
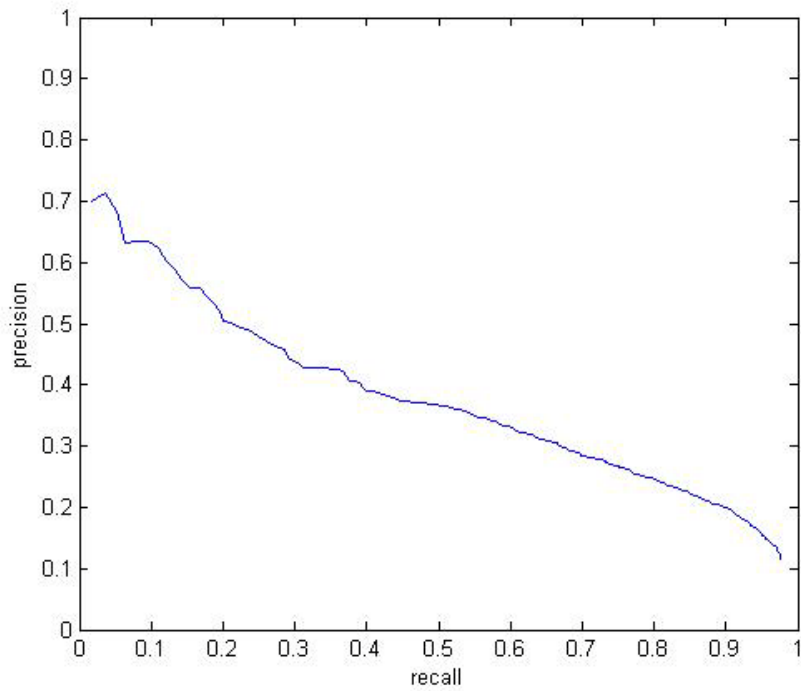
Figure 6-8   The precision vs. recall graph for redness-general

63
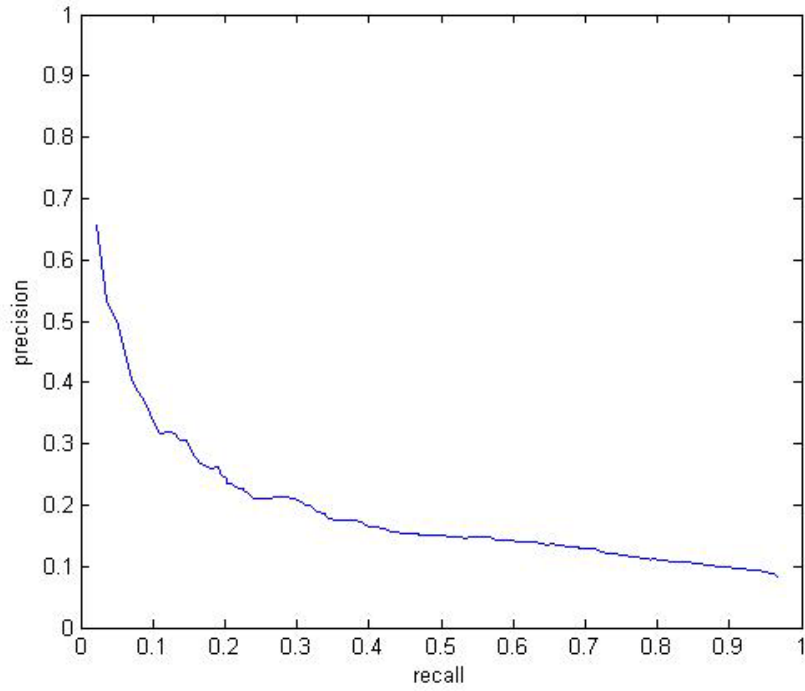
Figure 6-9   The *precision* vs. *recall* graph for ulcerated-eroded
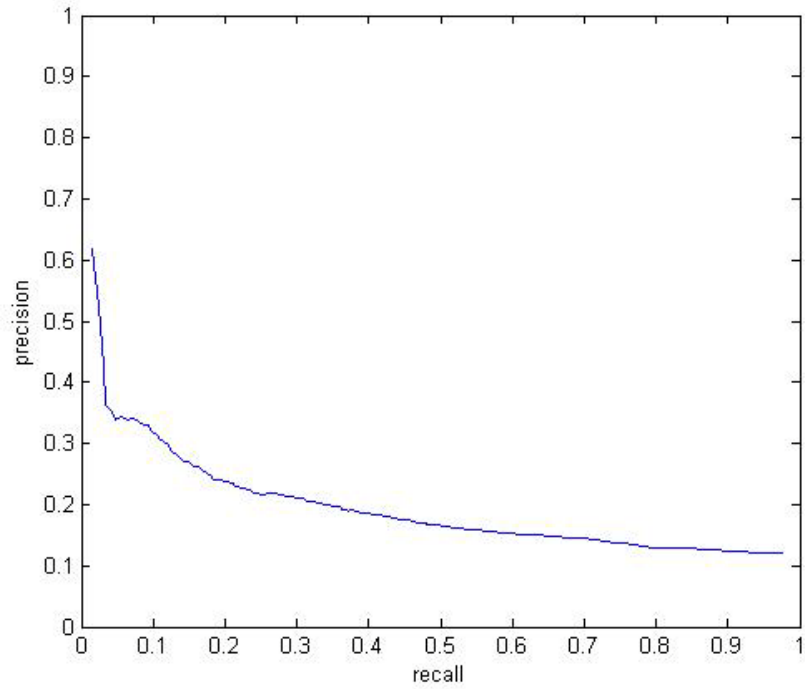
Figure 6-10   The *precision* vs. *recall* graph for warty-crusty-scabby

From these graphs above, we can see all of them are in accordance with the theoretical case, in that, the *precision* drops while *recall* increases. Also, except for gangrene-necrotic and redness-general, the graphs show a deceleration of drop rate at around *recall* = 0.1. Among all classes, the result for "redness-general" is the best, since *precision* is higher than for other classes at same *recall* value.

6.1.2 Self-developed scoring result

For the scoring method introduced in section 5.5., the final combination scoring is listed in Table 6-1, shown in comparison with scoring using only the color feature and only the texture & shape features. Fig. 6-11 gives the chart of the scoring result.

Table 6-2   Scoring Result of System with Only Color Feature, Texture & Shape Feature and Distance metric Combination

Class	color	Texture & shape	combination
fluid filled	0.33	0.22	0.43
gangrene-necrotic	0.41	0.10	0.57
pigmentation	0.29	0.32	0.46
purpura - violaceous	0.36	0.25	0.42
raised with color change	0.45	0.35	0.57
redness-general	0.61	0.40	0.66
ulcerated-eroded	0.51	0.17	0.47
warty-crusty-scabby	0.45	0.34	0.41
average	**0.43**	**0.27**	**0.50**

65
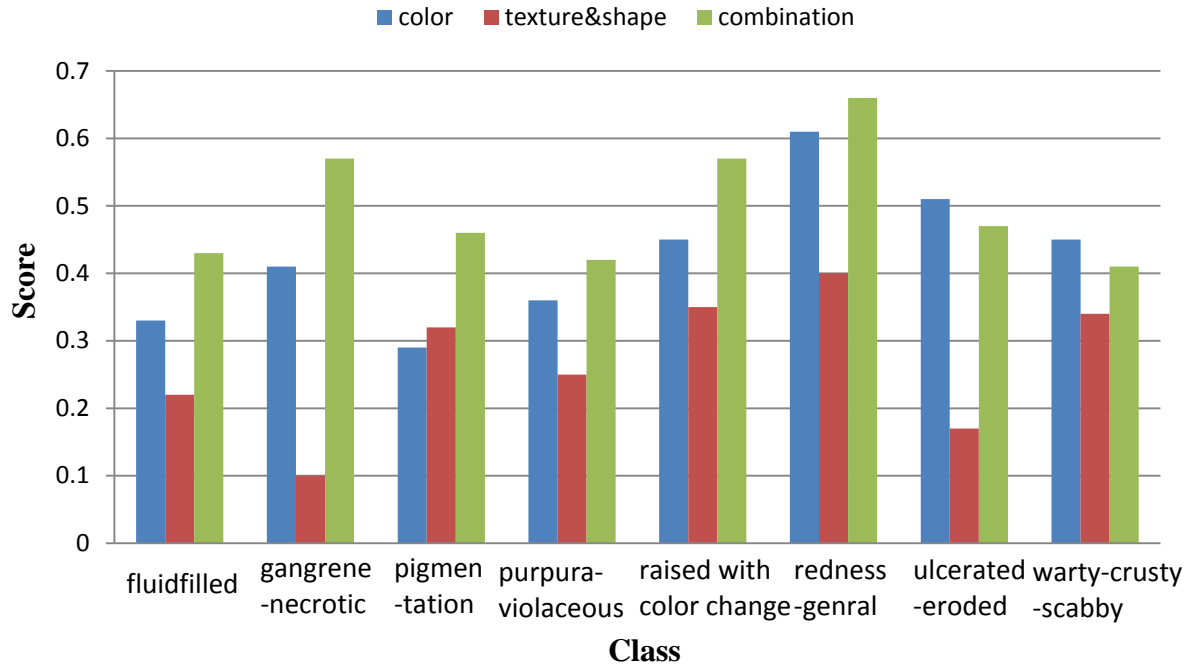
Figure 6-11   Chart of score of developmental data set for individual systems and combined system

From the table and chart above, overall, the combination scheme performs the best, but for some classes, it is a little worse than the system using only the color feature.

Among all the classes of disease, "redness-general" has the highest score, which is in agreement with the standard evaluation result.

The color feature alone dominates over the texture and shape feature for most diseases except for pigmentation. This unintuitive result can be explained by recognizing that pigmentation refer to a relative change in skin color, while our algorithm is based on a absolute color value.

In our experiment, the most computationally demanding part was shape feature extraction, especially the elliptical regression.

The retrieved candidates often contained the ROI images excerpted from the same original one, like the one in Fig. 6-2. This is because those excerpted images from the same source usually have the highest similarity.

6.2 Results for independent test data set

In this sub-section, an overview flowchart of the algorithm for independent test data is shown in Fig.6-12. An example result is shown in Fig.6-13. The self-developed scoring results are given in Table 6-2.
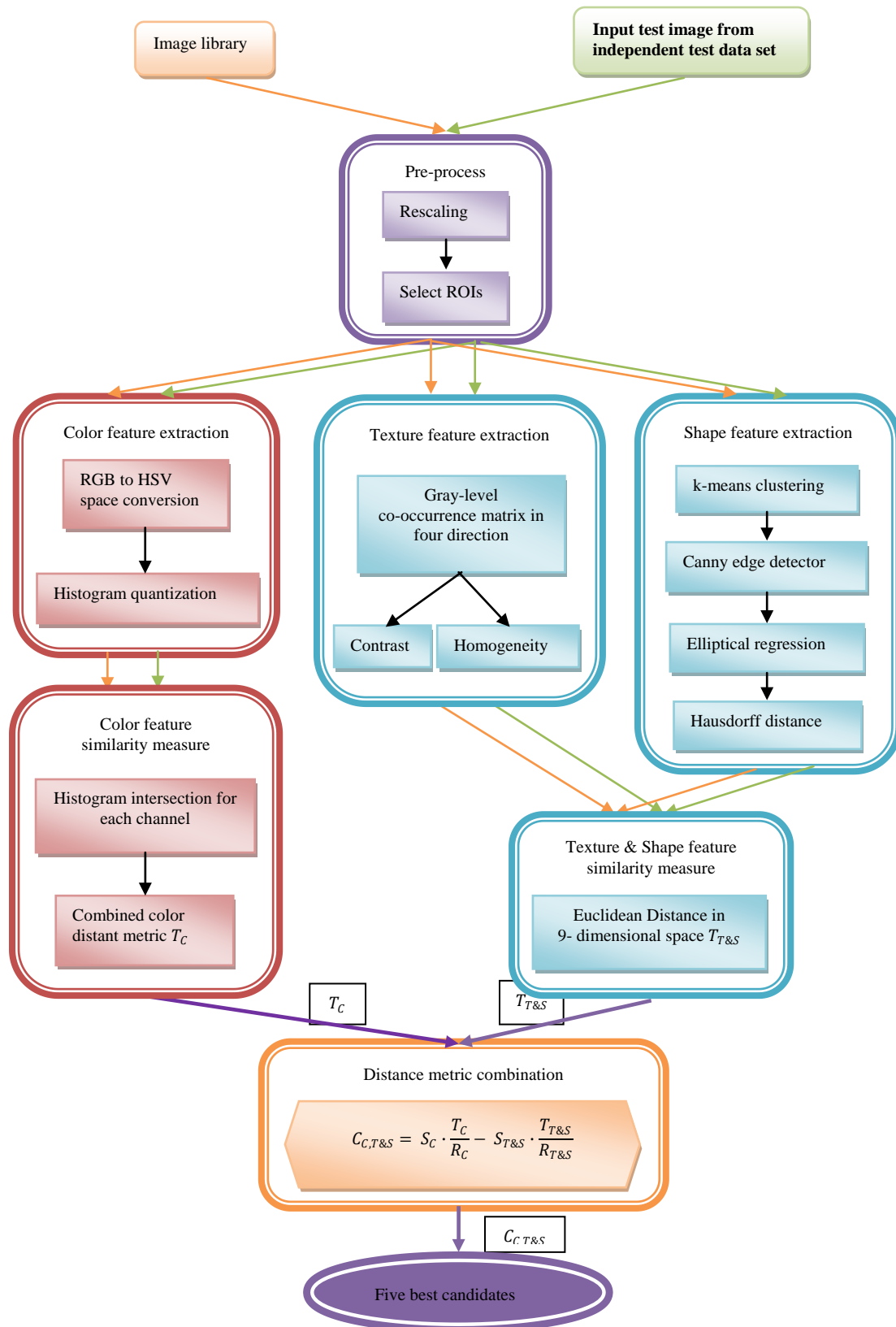
Figure 6-12   The algorithm for independent test data
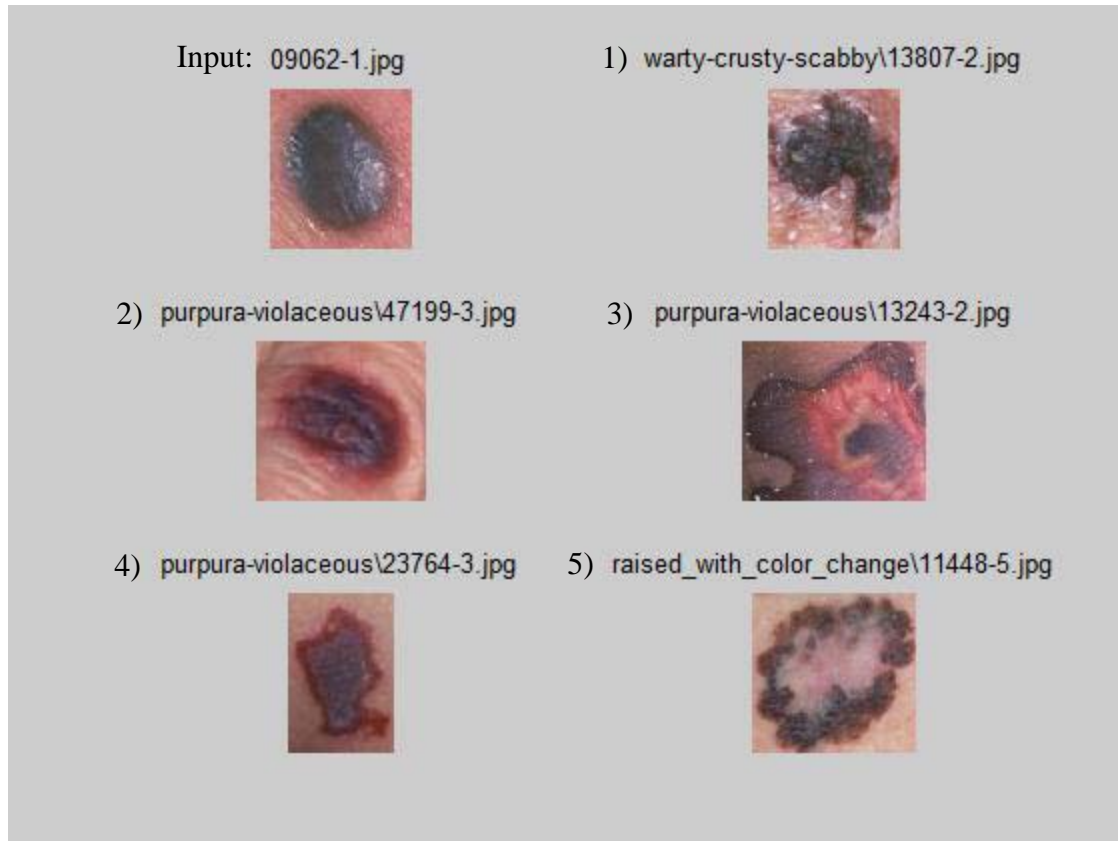
Figure 6-13   The result of "09062-1.jpg" from independent data set

The top left image in Fig. 6-13 is the input test image after pre-processing. The top right image is the best candidate of the input test image, the left image in the second row is the second best candidate, the right image in the second row is the third best candidate, the left bottom image is the fourth best candidate, and the right bottom image is the fifth best candidate. Doctors can make judgment based on these candidates.

Based on these results of the independent test data, my advisor, Dr. Kerekes calculated the score for each class. The self-developed scoring results for each class are listed in Table 6-3 and Fig. 6-14 gives the chart of the scoring results for the independent test data set in comparison with those of developmental data set.

Table 6-3   Scoring Results of Independent Test Data Set

Class	scoring
fluid filled	0.22
gangrene-necrotic	0.55
pigmentation	0.10
purpura - violaceous	0.37
raised with color change	0.25
redness-general	0.75
ulcerated-eroded	0.11
warty-crusty-scabby	0.28
average	0.30

Figure 6-14   Chart of score for developmental and independent test data set

The retrieval result of "redness-general" is again the best. However, the overall performance for independent test data set is poorer than that of developmental data set. One reason may be that for the developmental data set, the knowledge of the disease type will drop a hint in the ROI selection, and thus the selected ROI can represent the disease more accurately.

Besides, some images have multiple diseases. If it is unlabeled, we might have chosen a different diseased region other than the region showing the presumed one. This could also cause the mismatch and reduce the accuracy.

Also, we know for previous analysis, the best matched candidates in developmental data set are often from the same original image with the query image. However, there is no chance for this to happen in the independent test data set, because the candidates can only be selected from the labelled developmental data set. This is also an important reason for the drop of accuracy in independent test data set.

One more reason of the score drop is that the coefficients in distance metric combination are obtained using the developmental data set. Thus, the coefficients fit the developmental data set better than independent test data set.

The "pigmentation" score of independent data is much lower than that of developmental data. After examination, we find most accurate candidates of "pigmentation" in developmental data set are the family members of the input data. That is, "pigmentation" gains its score by its family members, while this is not possible in the case of independent test data.

The fall of "ulcerated-eroded" score is due to a difference in the observed features between the image in the developmental data and independent test data. Some of the "ulcerated-eroded" images in the independent data set are closer to images labelled "redness" in the developmental data set.

The "redness-general" score of the independent test data set is a little higher than that of developmental data set, since the "redness-general" samples in the independent test data set closely resemble those of the developmental data set in all visual features.

71

7. Summary and conclusions

7.1 Summary

Due to the difficulty of skin disease identification for dermatologists, a computer aided image system utilizing an image library for image retrieval is needed. The aim of our project is to develop a skin image retrieval system that can assist doctors to make better judgment. Computer aided image querying saves a lot of effort of the doctors and provides results through quantitative methods.

Because of lack of contemporary work on skin image retrieval for different classes of disease, this paper gives a literature review on content-based image retrieval, medical image retrieval and skin disease query. We adapted those techniques for general image retrieval into our case of study. Also, computer aided skin disease recognition has also been explored, so we can refer to those techniques employed in skin feature extraction, classification and segmentation.

Our data set contained two parts: a developmental data set for use as an image library with 139 images, and an independent test data set for use as unlabeled test data with 76 images. They both included eight kinds of skin diseases: fluidfilled, gangrene-necrotic, pigmentation, purpura-violaceous, raised with color change, redness-general, ulcerated-eroded and warty-crusty-scabby.

Our method consisted of several parts: pre-process, feature identification, feature extraction, similarity measure, and distance metric combination. In pre-process, rescaling and ROI selection are performed. Color, texture and shape feature are chosen in feature identification. HSV space, contrast and homogeneity of co-occurrence matrix and improved Hausdorff distance are extracted respectively. Histogram intersection is

manipulated for the color similarity measure and Euclidean distance is selected for texture and shape similarity measure. Then, several experiments on systems with individual feature were carried out, and we studied their performances according to the self-developed scoring result. At last, we proposed a distance metric combination scheme to integrate the two similarity measure methods.

The results are the top five candidates for the input query image; that is, five labelled images from the image library, so that the doctors can not only know the name of disease but also refer to pictures of it. We choose the number five arbitrarily, simply because it is a reasonable number of images for doctors to consider. Results are laid out separately for the developmental data set and the independent test data set. Two evaluation systems, both the standard *precision* vs. *recall* method, and the self-developed scoring method are carried out. The evaluation results obtained by both methods are given for each class of disease.

7.2 Conclusion

Among all visual features, we found the color feature plays a dominating role in distinguishing different types of skin disease.

The retrieved candidate images often contain the ROI images excerpted from the same original one when testing with the developmental data set.

The retrieval results of "redness-general" had the best accuracy, since it is has the best consistency in color, texture and shape. The second best was "gangrene-necrotic", which also had a good uniformity in color, texture, and shape. In Table 7-1, we perform a subjective evaluation of the feature consistency among each class with reference to

74

Appendix B. Non-consistency in features impedes the work of image retrieval and reduced back on the algorithm's accuracy.

Table 7-1   Subjective Evaluation of Feature Consistency

disease	Color	Texture	Shape
fluid filled	Bad	Good	Good
gangrene-necrotic	Very good	Moderate	Moderate
pigmentation	Very bad	Moderate	Moderate
purpura - violaceous	Bad	Very bad	Bad
raised with color change	Bad	Good	Good
redness-general	Very good	Very good	Good
ulcerated-eroded	Bad	Very bad	Moderate
warty-crusty-scabby	Bad	Bad	Moderate

The score of the developmental data set is higher than that of independent test data set due to the pre-knowledge based selection of ROIs, the sibling effect of retrieved candidates and the developmental data set determined coefficients in distance metric combination.

7.3   Future work

Since the lack of contemporary research in this specific area of skin disease retrieval, this is preliminary research and could benefit from many improvements. These include the following three aspects: image collection, feature extraction and distance metric.

7.3.1 Image collection

There might be some improvements in the image collection section. If the illumination condition for each image is given, color balancing may be performed in the pre-processing step, in order to reduce the impact of mismatched color balance between

75

the test and library images. As to "pigmentation", since it really depends on the skin color to define the symptom, the original skin color should be taken into account while performing color balancing.

The images are recommended to be collected at the same physical scale level to eliminate the impact of spatial scale.

To make those visual features more consistent in a disease group, the developmental data set may be divided into different sub-categories. Also, the sample size should be increased so as to get more precise statistical analysis results.

Additionally, collecting images in more spectral channels (ie, hyper spectral imaging) could lead to additional features to improve performance.

7.3.2 Feature extraction

The texture feature was only based on homogeneity of selected region. The dispersed nature of the disease may be explored as another texture feature. One example of a dispersed pattern is in shown in Fig.7-1 in comparison with single pattern.
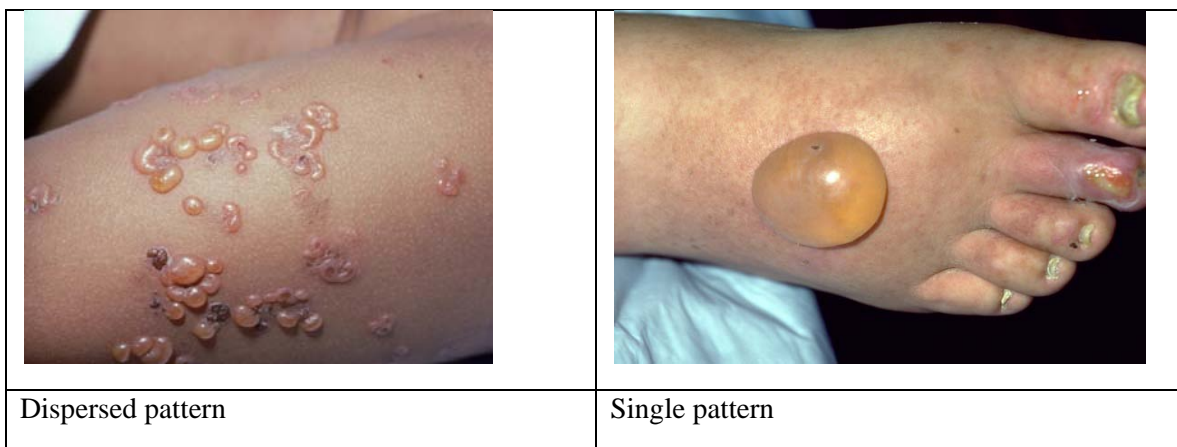
| Dispersed pattern | Single pattern |

Figure 7-1   Dispersed pattern and single pattern

In our texture feature extraction method, only *contrast* and *homogeneity* properties were chosen to characterize the homogeneity property of texture. In the future, some other properties of GLCM, such as *entropy*, could be investigated in order to depict the texture in various aspects.

During the shape feature extraction, we used K-means clustering to perform segmentation so as to extract the shape. Some other segmentation algorithms may generate a more accurate boundary. Also, we just tried to extract the ellipse for shape feature, while some other shape could be studied in future research.

At last, if the feature identification and extraction can be associated with some medical knowledge of those skin diseases as a semantic feature, it could significantly improve the precision of the disease description.

7.3.3 Distance metric

In our methodology, we used Euclidean distance for the similarity measure of the texture and shape feature. In further study, some other distance metric, such as the Mahalanobis distance, could be explored.

The distance metric combination scheme may be further investigated. Maybe some auto-adjusted strategy can take place of the empirical parameters in equation 5-10.

77

Appendix A: Mathematical tools

A.1.   RGB to HSV transformation (Wikipedia, 2009)

Given a color defined by (R, G, B), where R, G, and B are between 0.0 and 1.0, with 0.0 being the least amount and 1.0 being the greatest amount of that color, an equivalent (H, S, V) color can be determined by a series of formulas.

Let MAX be the maximum of the (R, G, B) values, and MIN be the minimum of those values. The formula can then be written as:

$$H = \begin{cases} \frac{G-B}{MAX-MIN} \times 60, & if\ R = MAX \\ \left(2 + \frac{G-B}{MAX-MIN}\right) \times 60, & if\ G = MAX \\ \left(4 + \frac{G-B}{MAX-MIN}\right) \times 60, & if\ B = MAX \end{cases}$$ **Equation A-1**

$$S = \frac{MAX-MIN}{MAX}$$ **Equation A-2**

$$V = MAX$$ **Equation A-3**

The resulting values are in (H,S,V) form, where H ranges from 0.0 to 360.0, represents the <u>angle</u> in <u>degrees</u> around the color circle where the hue is located. S and V vary from 0.0 to 1.0, with 0.0 being the least amount and 1.0 being the greatest amount of saturation or value, respectively.

A.2.   RGB to LAB transformation (Wikipedia, 2009)

$$L^* = 116 f(Y/Y_n)$$ **Equation A-4**

$$a^* = 500[f(X/X_n) - f(Y/Y_n)]$$ **Equation A-5**

$$b^* = 500[f(Y/Y_n) - f(Z/Z_n)]$$ **Equation A-6**

where

$$f(t) = \begin{cases} t^{1/3} & t > (6/29)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} & otherwise \end{cases}$$ **Equation A-7**

Here X_n, Y_n and Z_n are the CIE XYZ values of the reference white point (the subscript n suggests "normalized").

A.3.   Histogram intersection

Histogram intersection simply uses the minimum values of corresponding bins between two histograms. Because of its easy implementation, it is widely used in image retrieval area. (Swain and Ballard, 1991; Stanley et al, 2003; Funt and Finlayson, 1995)

Given a pair of histograms, I and M, each containing *n* bins, the intersection of the histograms is defined to be

$$\sum_{j=1}^{n} min(I_j, M_j)$$ **Equation A-8**

The result of the intersection of a model histogram with an image histogram is the number of pixels from the model that have corresponding pixels of the same color in the image. To obtain a fractional match value between 0 and 1 the intersection is normalized by the number of pixels in the model histogram. The match value is then

$$H(I, M) = \frac{\sum_{j=1}^{n} min(I_j, M_j)}{\sum_{j=1}^{n} M_j}$$ **Equation A-9**
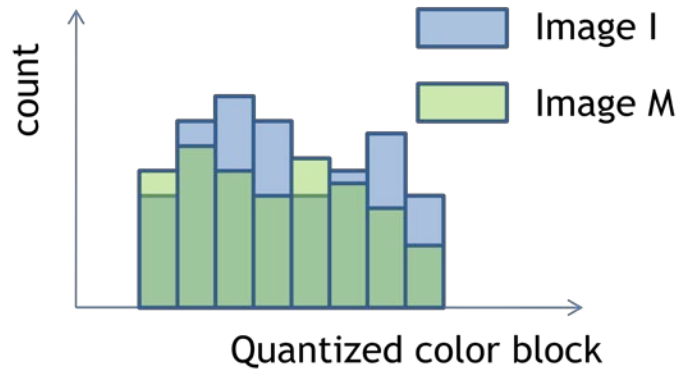
Figure A-1   Illustration of histogram matching scheme

As an example, consider two 4×4 images with 4 gray-level values:

1	3	3	2
0	0	1	1
3	3	2	1
3	2	0	0

Image *I*

0	0	1	2
3	2	2	2
3	3	1	0
0	0	1	1

Image *M*

The histogram of this is:
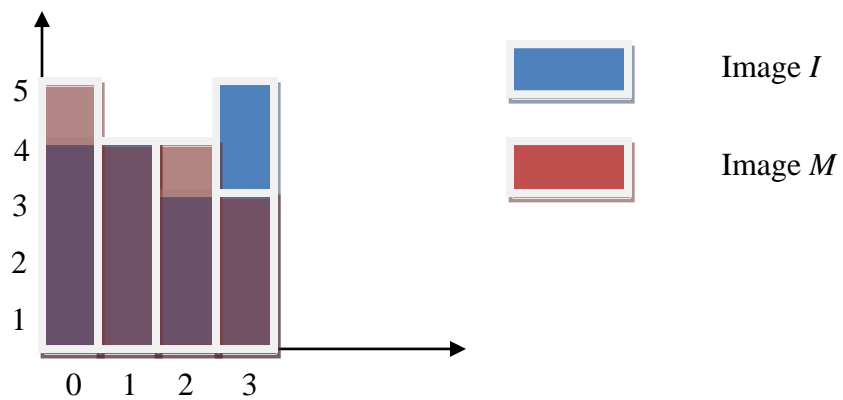
Figure A-2   Histogram of Image *I* and Image *M*

The normalized histogram intersection between the two images is:

$$H(I, M) = \frac{4 + 4 + 3 + 3}{5 + 4 + 4 + 3} = 0.875$$

A.4. Gray level co-occurrence matrix

Spatial gray level co-occurrence estimates image properties related to second-order statistics. Haralick, 1973 suggested the use of gray level co-occurrence matrices (GLCM) which have become one of the most well-known and widely used texture features. The gray level co-occurrence matrix for a displacement vector is defined as follows. The entry (i, j) is the number of occurrences of the pair of gray levels i and j which are a distance d apart. Formally, it is given as

$$P_d(i, j) = \left|\{((r, s), (t, v)): I(r, s) = i, I(t, v) = j\}\right| \qquad \textbf{Equation A-10}$$

Where $(r, s), (t, v) \epsilon N \times N, (t, v) = (r + dx, s + dy)$ and $|.|$ is the cardinality of a set. (Chen, 1998)

As an example, consider the following 4×4 image containing 3 different gray values:

1	1	1	0
1	1	0	0
0	0	2	2
0	0	2	2

The 3×3 gray level co-occurrence matrix for a displacement vector of d = (1, 0) is given as follows:

$$P_d = \begin{bmatrix} 3 & 0 & 2 \\ 2 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Here the entry (0, 0) of P_d is 3 because there are four pixel pairs of (0, 0) that are off-set by (1, 0) amount. Examples of P_d for other displacement vectors are given below:

d = (0, 1)	$P_d = \begin{bmatrix} 3\ 2\ 0 \\ 0\ 3\ 0 \\ 2\ 0\ 2 \end{bmatrix}$
d = (1, 1)	$P_d = \begin{bmatrix} 3\ 1\ 1 \\ 1\ 2\ 0 \\ 1\ 0\ 1 \end{bmatrix}$

The co-occurrence matrix reveals certain properties about the spatial distribution of the gray levels in the texture image. Haralick, K. 1973 has proposed a number of useful texture features that can be computed from the co-occurrence matrix. Table A-1 lists some of these features.

Here μ_x and μ_y are the means and σ_x and σ_y are the standard deviation of $P_d(x)$ and $P_d(y)$, respectively, where $P_d(x) = \sum_j P_d(x, j)$ and $P_d(y) = \sum_i P_d(i, y)$.

Table A-1   Texture Features Derived from Co-occurrence Matrix

Texture Feature	Formula
Energy	$\sum_i \sum_j P_{\mathbf{d}}^2(i, j)$
Entropy	$-\sum_i \sum_j P_{\mathbf{d}}(i, j) \log P_{\mathbf{d}}(i, j)$
Contrast	$\sum_i \sum_j (i - j)^2 P_{\mathbf{d}}(i, j)$
Homogeneity	$\sum_i \sum_j \dfrac{P_{\mathbf{d}}(i, j)}{1 + |i - j|}$

Contrast is a measure of the local variations present in an image. If there is a large amount of variation in an image, the $P_d(i, j)$ will be concentrated away from the main diagonal and contrast will be high. It has a range between $[0\ (size(GLCM) - 1)^2]$. $size(GLCM)$ is the number of gray levels.

83

Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. A homogeneous image will result in a co-occurrence matrix with entries having both high and low $P_d(i,j)$ values. If the range of gray levels is small, the $P_d(i,j)$ will tend to be clustered around the main diagonal. A heterogeneous image will result in an even spread of $P_d(i,j)$. It has a range of [0 1].

Entropy is a measure of information content. It measures the randomness of the intensity distribution. Such a matrix corresponds to an image in which there are no preferred gray level pairs for the distance vector *d*. Entropy is highest when all entries in $P_d(i,j)$ are of similar magnitude, and small when the entries in $P_d(i,j)$ are unequal.

Energy is the sum of squared values in the GLCM. It has a range of [0 1].

A.5.   K-means clustering

K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple way to classify a data set by a certain number of clusters (assume K clusters). The main idea is to define K centroids, one for each cluster. The better choice is to place the centroids as much as possible far away from each other. Then, take each point belonging to a given data set and associate it to the nearest centroid. When performing this to all the points, the first step is completed. At this point we need to re-calculate K new centroids of the clusters resulting from the previous step. After we generate these K new centroids, a new binding has to be done between the same data set points and the nearest new centroid. As a result of this loop, we notice that the K centroids change their location step by step until no more changes are done. Finally, this algorithm comes to minimize an *objective function*, in this case a squared error function. The objective function is

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad \text{Equation A-11}$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, and is an indicator of the distance of the n data points from their respective cluster centers.

A.6. Hausdorff distance

Hausdorff distance is the "maximum distance of a set to the nearest point in the other set" (Rote, G. 1991). Given two finite point sets $A = \{a_1, \dots, a_p\}$ and $B = \{b_1, \dots, b_p\}$, the Hausdorff distance is defined as

$$H(A, B) = max\big(h(A, B), h(B, A)\big) \qquad \text{Equation A-12}$$

Where

$$h(A, B) = max_{a \in A} \, min_{b \in B} \|a - b\| \qquad \text{Equation A-13}$$

$\|.\|$ is some underlying norm on the points of A and B, usually the Euclidean norm.

The function $h(A, B)$ is the directed Hausdorff distance from A to B. It chooses the point $a \epsilon A$ that is farthest from any point of B and measures the distance from a to its nearest neighbor in B.

The Hausdorff distance $H(A, B)$ is the maximum of $h(A, B)$ and $h(B, A)$. Thus, it measures the degree of mismatch between two sets by measuring the distance of the point of A that is farthest from any point of B and vice versa. Intuitively, if the Hausdorff distance is d, then every point of A must be within a distance d of some point of B and

85

vice versa. Thus, the notion of resemblance encoded by this distance is that each member of A be near some member of B and vice versa. (Huttenlocher et al, 1993).

An Improved Hausdorff distance is proposed by Dubuisson and Jain, 1994:

$$h(A, B) = \frac{1}{N_a} \sum_{a \in A} \min_{b \in B} \|a - b\| \qquad \text{Equation A-14}$$

$$H(A, B) = max\big(h(A, B), h(B, A)\big) \qquad \text{Equation A-15}$$

The Improved Hausdorff distance is best for matching two objects based on their edge points among the class of distance measures based on the Hausdorff distance. It has the following desirable properties: (1) its value increases monotonically as the amount of difference between the two sets of edge points increases, (2) it is robust to outlier points that might result from segmentation errors.

Appendix B: Data set

B.1.   Developmental data set

Figure B-1   Fluidfilled

Figure B-2   Gangrene-necrotic

Figure B-3   Pigmentation

88

Figure B-4   Purpura-violaceous

89

Figure B-5   Raised with color changes

Figure B-6   Redness – general

91

Figure B-7   Ulcerated-eroded

Figure B-8   Warty-crusty-scabby

B.2. Independent test data set

Figure B-9   Independent data set

Bibliography:

Abdel Wahab; N. M. Abdel Wahed, and A.S.A. Mohamed. 2005. Texture features neural classifier of some skin diseases. *Micro-NanoMechatronics and Human Science, 2005 IEEE International Symposium on*. 1: 380-382.

Alvarez, N. etc. 2005. Contour-Based Image Registration Using Mutual Information. *Lecture notes in computer science*. 3522: 227-234.

Antkowiak, M. 2006. Artificial Neural Networks vs. Support Vector Machines for Skin Diseases Recognition. *Master's Thesis in Computing Science, Umea University, Sweden*.

Benitez, A. B.; M. Beigi and S.-F. Chang. 1998. Using relevance feedback in content-based image metasearch. *IEEE Internet Computing*. 2(4):59-69.

Bickers, D. R.; H. W. Lim; D. Margolis, etc. 2004. The Burden of skin disease: a joint project of the American Academy of Dermatology Association and the Society for Investigative Dermatology. *J Am Acad Dermatol*. 55(3):490-500.

Binder, M. etc. 2000. Computer-aided epiluminescence microscopy of pigmented skin lesions: the value of clinical data for the classification process. *Melanoma Research*. 10(6):556-561.

Binder, M. etc. 2006. Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study. *British Journal of Dermatology*. 130(4): 460-465

Castiello, C.; G. Castellano and A. M. Fanelli. 2004. Neuro-Fuzzy Analysis of Dermatological Images. *2004 IEEE International Joint Conference on Neural Networks*. 4(25-29):3247-3252.

Chatterjee, K. and S. Chen. 2006. Affinity Hybrid Tree: An Indexing Technique for Content-Based Image Retrieval in Multimedia Databases. *Proceedings of the Eighth IEEE International Symposium on Multimedia*: 47-54.

Chen, C.H.; L. F. Pau and P.S.P. Wang. 1998. *The Handbook of Pattern Recognition and Computer Vision (2nd Edition)*. World Scientific Publishing Co.

Cheng, Y. and S. E. Umbaugh. 2005. Color-based Diagnosis: Clinical Images. www.ee.siue.edu/~sumbaug/SkinTumorClassificationUsingRelativeColorFeatures07132005.ppt (accessed Jan 28, 2009).

Chua, J.J. and P.E. Tischer. 2003. A Similarity Measure Based on Causal Neighbours and Mutual Information. *Design and application of hybrid intelligent systems*: 842-851.

Claridge, E.; P. Hall etc. 1992. Shape analysis for classification of malignant melanoma. *Journal of Biomedical Engineering*. 14(3):229-34.

Corbis. 2009. http://pro.corbis.com/ (accessed Jan 5, 2009).

Delventhal, H. and R. Pompl. 1998. MELDOQ-Computer assisted early recognition of malignant melanoma. http://www.imse.med.tumuenchen.de/mi/derma/English/index.html (accessed May 9, 2009)

Description of dermatology. 2009. http://www.wrongdiagnosis.com/medical/dermatology.htm (accessed Jan 2, 2009)

Dreisetl, S.; L. Ohno-Machado, etc. 2001. A comparison of Machine Learning Methods. *Journal of Biomedical Informatics*. 34: 28-36.

98

Dubuisson, M.P. and A.K., Jain. 1994. A modified Hausdorff distance for object matching. *Pattern Recognition, Proceedings of the 12th IAPR International Conference on Computer Vision & Image Processing*. 1(9-13): 566 – 568.

Ercal, F.; M. Moganti, etc. 1993. Detection of Skin Tumour Boundaries in Color Images. *IEEE Transactions on Medical Imaging*. 12(3): 624-627.

Fiorini, R. A.; G. Dacquino and G. Laguteta. 2004. A New Melanoma Diagnosis Active Support System. *Proceedings of the 26th Annual International Conference of the IEEE EMBS*. 2(1-5): 3206 - 3209.

Fischer, S.; P. Schmid and J. Guillod. 1996. Analysis of skin lesions with pigmented networks. *International Conference on Image Processing*. 1(16-19): 323-326.

Funt, B. V. and G. D. Finlayson. 1995. Color Constant Color Indexing. *IEEE Trans on Pattern Analysis and Machine Intelligence*. 17 (5): 522-529.

Ganster, H.; A. Pinz, etc. 2001. Automated Melanoma Recognition. *IEEE Transaction on Medical Imaging*. 20(3): 233-239.

GNU. 2005. http://www.gnu.org/software/gift/ (accessed Jan 5, 2009).

Green, A., etc. 1994. Computer image analysis in the diagnosis of melanoma. *American Academy of Dermatology*. 31: 958-64.

Green, B. 2002. Canny Edge Detection Tutorial.

http://www.pages.drexel.edu/~weg22/can_tut.html (accessed Mar 27, 2009).

Hammiche, S.; S. Benbernou, etc. 2004. Semantic Retrieval of Multimedia Data. *Proceedings of the 2nd ACM international workshop on Multimedia databases*: 36-44.

99

Han, J. K.; N. Ngan; M. Li and H. Zhang. 2004. Learning semantic concepts from user feedback log for image retrieval. *IEEE International Conference on Multimedia and Expo*. 2(30): 995-998.

Hance, G.A.; S.E. Umbaugh; R.H. Moss and W.V. Stoecker. 1996. Unsupervised color image segmentation: with application to skin tumor borders. *Engineering in Medicine and Biology Magazine, IEEE*. 15(1): 104-111.

Handels, H, etc. 1999. Feature selection for optimized skin tumor recognition using genetic algorithms. *Artificial Intelligence in Medicine*. 16(3): 283–297.

Hao, W.; Z. Li and W. Guowei. 2000. Image Search Using Multiresolution Matching with a Mutual Information Model. *IEEE 5th International Conference on Signal Processing Proceedings*. 2: 951-954.

Hidden Markov Model. 2009.
http://www.csse.monash.edu.au/~lloyd/tildeMML/Structured/HMM/ (accessed Jan 9, 2009)

Hoiem, D.; R. Sukthankar, etc. 2004. Object-Based Image Retrieval Using the Statistical Structure of Images. *Computer Vision and Pattern Recognition*. 2: 490-497.

Holmstrom, T. 2005. A survey and Evaluation of Features for the Diagnosis of Malignant Melanoma. *Master Thesis in Computing Science*.

Huttenlocher, D.P.; G.A. Klanderman and W.J. Rucklidge. 1993. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 15(9): 850 – 863.

Iakovidis, D.K., N. Pelekis, etc. 2006. A Pattern Similarity Scheme for Medical Image Retrieval. *IEEE Transactions on Information Technology in Biomedicine*. 99(0): 1-1.

ImgSeek. 2009. http://www.imgseek.net/index.html (accessed Jan 5, 2009).

Khan, N. 2005. Multimedia Database Search and Techniques. ENG4112 Research Project. *Faculty of Engineering and Surveying, University of Southern Queensland*.

Kreutz, M. and S. Gehlen. 2001. Automated Diagnosis of Skin Cancer Using Digital Image Processing and Mixture-of-Experts, Biomedizinische Technik-Biomedical Engineering-Supplement: 357-361.

Logical Images. 2009. http://www.logicalimages.com/ (accessed April 6, 2009).

Liu, Y.; R. Dellaert and W.E. Rothfus. 1998. Classification Driven Semantic Based Medical Image Indexing and Retrieval. *Research thesis, The Robotic Institute, Carnegie Mellon University.*

MacQueen, J. B. 1967. Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability.* 1:281-297.

Maglogiannis, I.; S. Pavlopoulos, etc. An Integrated Computer Supported Acquisition, Handling, and Characterization System for Pigmented Skin Lesions in Dermatological Images. *IEEE Transactions on Information Technology in Boimedicine*. 9(1): 86-98.

McInerney, T. and D. Terzopoulos. 1996. Deformable Models in Medical Image Analysis: A Survey. *Medical Image Analysis*. 1(2): 91-108.

101

Müller, H.; W. Müller and D. McG. Squire. 2001. Performance Evaluation in Content-Based Image Retrieval: Overview and Proposals. *Pattern Recognition Letter.* 22(5): 593-601.

Muller, H.; N. Miachoux, etc. 2004. A review of content-based image retrieval systems in medical applications——clinical benefits and future directions. *International Journal of Medical Informatics.* 73: 1-23.

Niblack, W.; R. Barber, etc. 1993. The QBIC Project: Querying Images by Content Using Color, Texture, and Shape. *SPIE.* 1908: 173-187.

Niedermayer, D. 1998. An Introduction to Bayesian Networks and their Contemporary Applications. http://www.niedermayer.ca/papers/bayesian/ (accessed Jan 9, 2009).

Papier, A. 2009. The Perceptual Skills of Your Dermatologist. http://query.nytimes.com/gst/fullpage.html?res=9F0CE5DD163EF935A3575BC0A96E9 C8B63 (accessed Jan 2, 2009).

Park, J.; G. Kim and H. Choi. 2007. Image retrieval using by Skin Color and Shape Feature. *Lecture Notes in Computer Science*. 4705: 1045-1053.

Phung, S.L.; A. Sr. Bouzerdoum; D. Sr. Chai. 2005. Skin Segmentation Using Color Pixel Classification: Analysis and Comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 27(1): 148 – 154.

Pluim, J.; J. B. Antoine Maintz and M. A. Viergever. 2003. Mutual-Information-Based Registration of Medical Images: A Survey. *IEEE Transactions on Medical Imaging*. 22(8): 986-1004.

Rote, G. 1991. Computing the minimum Hausdorff distance between two point sets on a line under translation. *Information Processing Letters*. 38: 123-127.

Round A. J.; A.W.G. Duller; P. J. Fish. 1998. A comparison of skin patterning feature analysis methods for lesion classification. *SPIE proceedings series: image processing*. 3338 (2): 202-210.

Round A. J.; A.W.G. Duller; P. J. Fish. 2001. Lesion classification using skin patterning. *Skin Research and Technology*. 6(4): 183-192.

Schmid-Saugeon, P.; J. Guillod and J. Thiran. 2003. Towards a computer-aided diagnosis system for pigmented skin lesions. *Computerized Medical Imaging and Graphics*. 27: 65-78.

Schmidt, P. and S. Fischer. 1997. Color Segmentation for the Analysis of Pigmented Skin Lesions. *Proceedings of the Sixth International Conference on Image Processing and its Applications*. 2: 688-692.

Schmidt, P. 1999. Segmentation of Digitized Dermatoscopic Images by Two-Dimensional Color Clustering. *IEEE Transaction on Medical Imaging*.18(2): 164-171.

Smith, J.R. and S. Chang. 1996. VisualSEEk: a fully automated content-based image query, In Proceedings of ACM Multimedia: 87-98.

Squire, D.M.; W. Muller, etc. 1999. Content-based query of image databases: inspirations from text retrieval. *Pattern Recognition Letters*. 21(13-14): 1193-1198.

Stanley, R. J.; R. H. Moss, etc. 2003. A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images. *Computerized Medical Imaging and Graphics*. 27: 387-396.

Swain, M. J. and D. H. Ballard. 1991. Color Indexing. *International Journal of Computer Vision*. 71: 32.

Takemae, Y.; H. Saito and S. Ozawa. 2000. The Evaluating System of Human skin Surface Condition by Image processing. *System, Man, and Cybernetics, 2000 IEEE International Conference on*. 1: 218-223.

Tanaka, T.; T. Joke and T. Oka. 2001. Cell nucleus segmentation of skin tumour using image processing. *Proceedings of the 23rd Annual International Conference of the IEEE in Medicine and Biology Society*. 3: 2716-2719.

The Free Dictionary. 2009. http://medical-dictionary.thefreedictionary.com/ (accessed Jan 28, 2009).

Umbaugh, S.E.; R.H. Moss; W.V. Stoecker and G.A. Hance. 1989. Automatic color segmentation Automatic color segmentation of images with application auto-detection of variegated coloring in skin tumors. *Engineering in Medicine and Biology Magazine, IEEE*; 8(4): 43-50.

Umbaugh, S.E.; R.H. Moss; W.V. Stoecker and G.A. Hance. 1993. Automatic color segmentation algorithms-with application to skin tumor feature identification. *Engineering in Medicine and Biology Magazine, IEEE.* 12(3): 75-82.

University of Rochester. 2005. Dermatology Lexicon Project.

http://www.dermatologylexicon.org/ (accessed Jan 20, 2009).

Vasconcelos, N. and M. Vasconcelos. 2004. Scalable Discriminant Feature Selection for Image Retrieval and Recognition. *IEEE Conf. on Computer Vision and Pattern Recognition*. 2:770-775.

104

Viola, P. 1997. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*. 24(2): 137-154.

Virage Corporation. 2009. http://www.virage.com/home/index.en.html (accessed Jan 5, 2009).

Webseek. 2009. http://ei.cs.vt.edu/~mm/cache/WebSeek.htm (accessed Jan 5, 2009).

Wikipedia. 2009. HSL and HSV. http://en.wikipedia.org/wiki/HSL_color_space (accessed May 19, 2009).

Wikipedia. 2009. Lab color space. http://en.wikipedia.org/wiki/Lab_color_space#RGB_and_CMYK_conversions (accessed May 19, 2009).

Wikipedia. 2009. Primary color. http://en.wikipedia.org/wiki/Additive_primaries#Additive_primaries (accessed May 6, 2009).

Wikipedia. 2009. Mahalanobis distance. http://en.wikipedia.org/wiki/Mahalanobis_distance (accessed Jan 9, 2009).

Xu, Y.; E. Saber and A. M. Tekalp. 2004. Dynamic learning from multiple examples for semantic object segmentation and search. *Computer Vision and Image Understanding*. 95(3): 334-353.