Imperial College London

Department of Computing

Image processing and classification algorithm to detect cancerous cells morphology when using in-vivo probe-based confocal laser endomicroscopy for the lower gastrointestinal tract

by

Pijika Watcharapichat [pw610]

Submitted in partial fulfilment of the requirements for the MSc Degree in Computing Science of Imperial College London

September 2012

Abstract

Confocal laser endomicroscopy has recently emerged as a novel technology that brings major improvement to endoluminal imaging especially endoscopy for gastrointestinal tract. On the other hand, it could be a challenge for endoscopists as they need another specialized discipline, which is histology, and should be able to integrate this new domain knowledge into their clinical practice to make better medical judgement. This challenge is the motivation for us to initialize a computer-aided decision making system that has the ability to quantitatively analyze confocal images and return informative output which can assist endoscopists to improve their medical diagnosis. Primary objective of this project is to create a computer-aided decision support system that has the ability to assist endoscopists in establishing the final diagnosis on the information of confocal laser endomicroscopic images of the colon. We are particularly interested in confocal images of the colon because colon-related diseases are relatively common regardless ethnic.

We develop a system based on "content-based image retrieval" or CBIR method which has the ability to perform image search and finally present endoscopists a number of similar images from which they are able to decide themselves on the final diagnosis. The search according to CBIR analyzes image similarity based on image signatures derived from texture analysis using Gabor filters bank and texture distance between images is computed under Earth Mover's distance with scale- and rotation-invariant features. We employ a "patch technique" which is to select only certain parts of example images and search according to them instead of the conventional method which usually processes the entire example image. Moreover, we incorporate "relevance feedback" which has potential to adapt the search behavior towards the user's need. The mechanism of relevance feedback fundamentally operates on embedded manifold obtained from "Isomap".

We conducted experiments on a database of 1000 confocal images which are all classified by clinical experts based on Miami classification. The final results showed that the patch technique outperformed the conventional method. Our system achieved the retrieval rate of 87.78% when the ability to distinguish between neoplastic and benign categories is the criteria of accuracy. Interestingly, it achieved the retrieval rate of 58.89% when the ability to return images with correct classification is the criteria to determine the accuracy of the system. In addition, there is also an ongoing preliminary experiment to study the feasibility to integrate an Eye Tracking system into our computer-aided decision support system and the result will be reported in the near future.

Acknowledgement

I would like to gratefully acknowledge my supervisor Professor Guang-Zhong Yang for the inspiration, smart guidance, and continuous support throughout my project period.

I wish to thank Hamlyn Centre research team especially Vasiliki Simaiaki, Dr. Tou-Pin Chang, and Dr. Michael Hughes, for sharing the source of confocal images, introducing new knowledges, and valuable technical discussions. I would like to thank Kenko Fujii and Johannes Totz for providing me with a great opportunity to study more about the Eye tracking system.

Finally, I would like to thank my family and my best friends for continuous encouragement and positive feedbacks. Without them, I would not have great courage to pursue my dream in Computer Science.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Colon cancer is one of the leading health problems worldwide and it is one of the most common cause of cancer-related death. Screening endoscopy for the colon or "colonoscopy" is an imaging-based screening test that is widely accepted as it has been proven to reduce the incidence of colon cancer. During the endoscopy, tissue biopsy[1] are routinely performed if there is any sign of cancer. Histology[2] examination after tissue biopsy is considered as the gold standard for the final judgement of tissue pathology[3] and definite diagnosis.

Nowadays, there are several novel technologies emerging within the field of endoluminal imaging and confocal laser endomicroscopy is counted as major improvement. In terms of functionality, endomicroscopes provide highly-magnified images that facilitate microscopic analysis of mucosal[4] surface structure and it has become more widely-used because it can be integrated into standard video endoscopes.

Confocal laser endomicroscopy can be used to diagnose several conditions especially gastrointestinal diseases like, such as colon cancer. It allows endoscopists to see in vivo tissue architechure and characteristics of disease during ongoing endoscopy. This benefit will lead to more targeted biopsies and random specimen can be avoided. As it potentially provides higher accuracy on tissue biopsy, the number of false positive or noninformative biopsy tissue will become less. Thus workload on histology examination, which is a time-consuming process, will decrease.

On the other hand, endoscopists are required to interpret imaging in order to make a diagnosis during the ongoing endoscopy so that medical treatment can be planned and promptly performed. This could be a real challenge as they have to combine knowledge of both macroscopic and microscopic appearance of the in vivo tissue concurrently. The next critical issue could be the interpretation of images obtained from the endomicroscopy because such interpretation is essentially examiner-dependent and immediate in vivo diagnosis could be a difficulty. Computer-aided decision support system can be a solution to address this problem.

Computer-aided decision support system or CADSS has become one of the major research topics in medical imaging field during the past two decades. In contrast to the concept of automated computer diagnosis, the performance of CADSS does not have to be comparable or superior to that of clinicians but it needs to be "complementary" to them. Basically, digital medical images are analyzed quantitatively by computers based on their specific algorithms and clinicians can use the output from CADSS as a "second opinion" and make the final medical decision. The concept of CADSS has been widely applied into real clinical practice because the correct computer output has the ability to assist clinicians in improving their medical decisions and at the same time, they can still maintain their own clinical opinions. So far, CADSS can play 2 important roles in medical

[1]The removal and examination of a sample of tissue from a living body for diagnostic purposes

[2]The anatomical study of the microscopic structure of animal and plant tissues.

[3]The anatomic or functional manifestations of a disease

[4]Related to the membranes lining the passages of the body, such as the respiratory and digestive tracts, that open to the outside. Cells in the mucous membranes secrete mucus, which lubricates the membranes and protects against infection.

imaging field; one for the detection of disease lesions and another for helping diagnosis based on differentiation between neoplastic and non-neoplastic lesions. [1]

However, due to the fact that confocal laser endomicroscopy is the most recent technique in the endoscopic field, the number of CADSS related works and publications on confocal images in gastrointestinal tract is still limited.[2] Hence we would like to investigate the potential of CADSS in the confocal laser endomicroscopic field and initiate a new CADSS framework which has potential to assist endoscopists in establishing the final diagnosis on the information of confocal laser endomicroscopic images of the colon. To the best of our knowledge, there exist only a few CADSS related works published and they particularly investigate the use of content-based image/video retrieval (CBIR/CBVR) method to classify endomicroscopic images of the colon. In these works, the authors [3] proposed a system which aims to distinguish between benign and neoplastic colon polyps by using the Bag of Visual Words method in combination with multiple SIFT descriptors to obtain a scale invariant texture descriptor. They also extended their work by using a co-occurrence matrix based on the adjacency of visual words to exploit the relationships between local feature regions and their approach yields highly accurate results.

In our project, we would like to develop a framework which has the ability to perform image search and finally present endoscopists a number of similar images from which they are able to decide themselves on the final diagnosis. Our framework is based on CBIR method but, different to previous publications, texture analysis using Gabor filters bank will be our main tool to analyze image content because confocal images of the colon can be seen as diversely textured images. Moreover, we integrate relevance feedback feature into our CBIR framework so endoscopists can interact with our system and this allows great opportunities to refine the retrieved results and makes the image search even more accurate and powerful.

1.2 Objectives

The objectives of this project are listed as below.

- To create a computer-aided decision support framework using CBIR method to facilitate image search for confocal images of the colon
- To experiment whether Gabor filters bank is a valid method to analyze confocal images of the colon
- To explore whether our technique using 3 small patches as a query can outperform conventional method using the entire example image as a query
- To explore whether the integration of relevance feedback to the main framework can improve the retrieval results with efficiency
- To explore whether Isomap is a valid method to make our relevance feedback successful
- To build a GUI for the user to operate our system

1.3 Report structure

The rest of this report is divided into 7 chapters and organized as follows. Chapter 2 introduces clinical background of colon cancer in several aspects including the cause of the disease, the disease nature, as well as how the disease is investigated in clinical practice. Technical knowledge on confocal laser endomicroscopy is also provided in the subsequent section giving basic ideas how this advanced technology works and how it becomes an important procedure in routine clinical practice. Moreover, classification of confocal images of the colon are described and illustrated in this chapter.

Chapter 3 provides technical background of content-based image retrieval (CBIR) when using Gabor wavelets as the principal texture-based method for feature extraction along with full

description of Earth Mover's distance method that is used to calculate image similarity in this project.

In chapter 4, relevance feedback technique is described together with the details of a dimensionality reduction method, Isomap, which we employ as the core implementation for the relevance feedback. Chapter 2,3 and 4 all include references of the literature that have influenced the development and are related to this project.

Chapter 5 provides the details of software design and several implementation techniques we have used to develop our application. We also present experimental results of image retrieval by demonstrating examples of final outcomes in the subsequent section of this chapter.

In chapter 6, we display results from our mini visualization tool along with discussion on the relationship between different images in terms of visual property.

Our comments and discussion on each implementation step can be found in chaper 7 and thereafter, chapter 8 concludes the project.

Chapter 2

Clinical background

2.1 Colon cancer

Colon cancer is one of the leading health burdens worldwide. According to the report of the World Health Organization (WHO) in 2003, there were approximately 940,000 patients being diagnosed with colon cancer and nearly half of them died from the disease. In 2008, it was reported that colon cancer was ranked as the third most common cause of cancer-related death. [5]

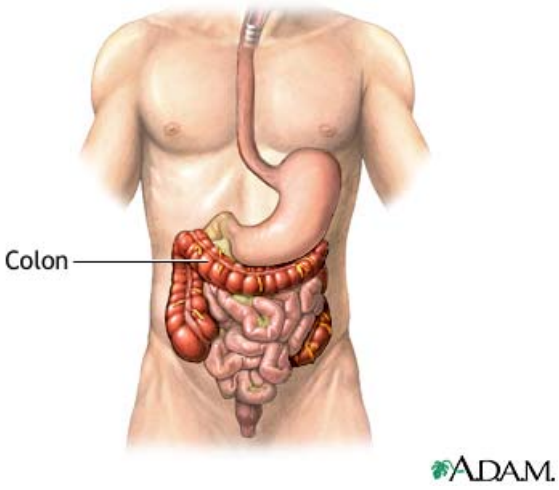


Figure 2.1: Picture of the Colon [6]

Colon cancer originates from epithelial[1] cells lining of the colon. Most colon cancers are silent tumors. They progress slowly and do not produce symptoms until they become a large size. Colon cancer usually begins as a polyp[2]. Polyps are often benign but some have potential to develop into invasive cancer. There are 2 most common types of polyps found in the colon including

- **Hyperplastic and inflammatory polyps**

  These kinds of polyps usually do not carry a risk of cancer development. However, hyperplastic polyps with a large size, especially on the right-side colon, are of concern and should be completely removed.

- **Adenomas or adenomatous polyps**

  These polyps have potential to turn malignant. In other words, they are considered as precancerous stage.

[1] Related to the thin tissue forming the outer layer of a body's surface and lining the alimentary canal or other hollow structures

[2] A non-specific term to describe a growth on the inner wall of the colon

Not all of colon polyps become cancer but all colon cancers virtually start from these growths. Apart from polyps, other risk factors that are associated with the cancer include nutritional, environmental, genetic, and familial factors, as well as certain preexisting inflammatory diseases of the bowel such as Crohn's disease or ulcerative colitis. [7]

However, colon cancer is a preventable and curable if detected early. Early detection through screening programs is the most important factor in the recent decline of colon cancer. One of the most commonly used screening procedures is colonoscopy. Basically this procedure provides information on the mucosa of the entire colon and the sensitivity in detecting tumors is very high. Screening colonoscopy can also be used to obtain biopsy specimens of the examined colon and permits the excision of polyps. There is direct clinical evidence supporting that screening procedures can significantly reduce death related to colon cancer.[11] Additionally, colonoscopy has been proven to reduce the incidence of colon cancer in patients with pre-cancerous disease e.g. adenomatous polyps.[12]

When performing colonoscopy, if polyps are discovered, they tend to be excised for further histology examination and the definite diagnosis can be made. Histological analysis of biopsy tissue is the gold standard for the final diagnosis of colon disease. However, such examination and evaluation require time and medical specialists e.g. pathologists [3]. Furthermore, people tend to have more cancer awareness and are encouraged to receive a regular screening test so any disease can be detected earlier. The number of tissue specimens has increased but the number of pathologists is limited. In this case, more selective biopsy can be the best solution for this limitation. Clearly, more information is required to perform such selective procedure.

Recently, a confocal laser endomicroscopy has been developed and integrated with conventional colonoscopes. This endomicroscopy provides highly-magnified images of living tissue during ongoing colonoscopy. Apparently, more information can be collected as real-time in vivo tissue architechure can now be observed. Random specimen can be reduced and workload on histology examination, which is a time-consuming process, will subsequently decrease.

2.2 Confocal laser endomicroscopy (CLE)

2.2.1 Overview

Confocal laser microscope is considered as one of the most significant advances in optical microscope field as it offers significant advantages over conventional microscopes. Not only can it provide better lateral and axial resolution, but also it has the ability to visualize selectively in depth by significantly rejecting out-of-focus light through the use of a confocal pinhole.

The term "confocal" was coined by Marvin Minsky (1955) because the objective lens and condenser have the same focal point. A illumination source is focused at a point in the sample and illuminates only a single point at a time. The pinhole aperture passes light from the focus but it substantially rejects light from depths in the sample above and below the focal plane as shown in figure 2.2. Due to this property, imaging of thick biological samples can be easily taken without the need for tissue excision, section, or fixation as is required for the conventional one. [14] Hence it has subsequently become a popular tool for high resolution imaging in biomedical field including in vivo imaging of human tissue. Currently, there is a wide array of confocal imaging applications and one of them is "confocal laser endomicroscopy".

Confocal laser endomicroscopy (CLE) represents a rapid emerging field of endoluminal imaging that can bridge the interface between endoscopy and histology because it offers extremely high magnification in vivo images. It can visualize living tissue in a real time so it has the potential to change the clinical paradigm of diagnosis including the role of biopsy in gastrointestinal[4] endoscopy. Although histological analysis of biopsy tissue is highly valuable and currently considered as the gold standard of diagnosis, biopsy is expensive and it may prolong treatment process because there is usually a long gap between the time when endoscopy is performed and the time when diagnosis

[3] a doctor who specializes in medical diagnosis

[4] Related to the stomach and intestines

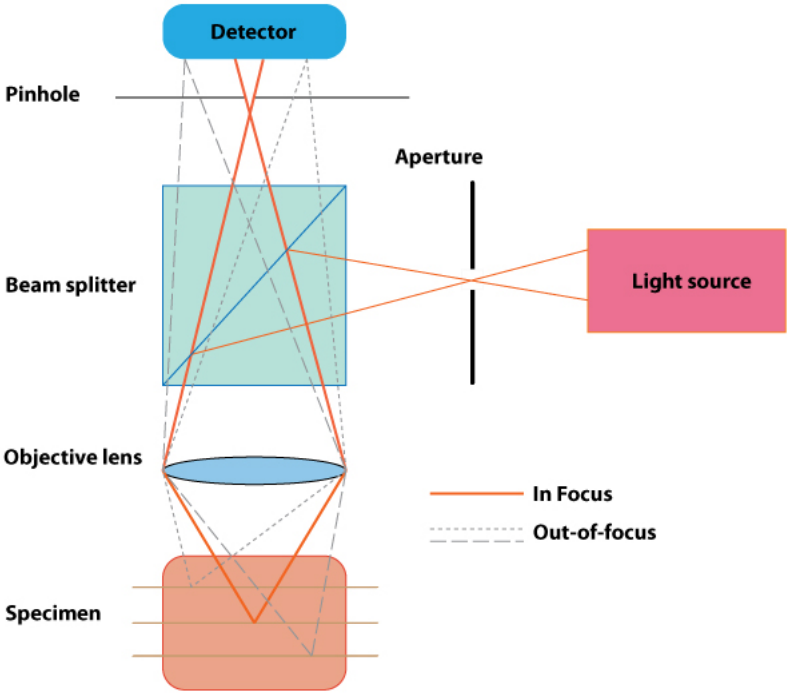


Figure 2.2: This figure demonstrates a simple diagram for confocal microscope. The solid orange line coming from the focal spot (in the specimen) is the only ray that can pass through the confocal pinhole and reach the detector. The out-of-focus rays, which are represented by grey dash lines, will be rejected by the pinhole. As a result, we can capture a clear image particularly from the focus region.

is histologically confirmed.

2.2.2 General techniques for CLE

CLE captures microscopic images of in vivo tissue. The microscope employs focused laser light of one specific wavelength and passes it through a confocal aperture. Confocal images are then reconstructed in two dimensions.

Since CLE is also used based on a through-the-scope probe, it can be called probe-based CLE or pCLE. This probe-based type is developed and produced by Mauna Kea Technologies. With pCLE, the laser scanning unit is mounted outside the conventional endoscope, and an optical fiber bundle delivers light and collects data to form images. There is also a built-in mechanism using "Proprietary algorithms" to correct for image distortion resulting from the long fiber bundle. The pCLE system offers high resolution of approximately 1 micron and a fixed imaging depth of 50 microns. It also allows fast video-rate scanning (12 frames per second) and adjustability with any endoscopic system. Figure 2.3

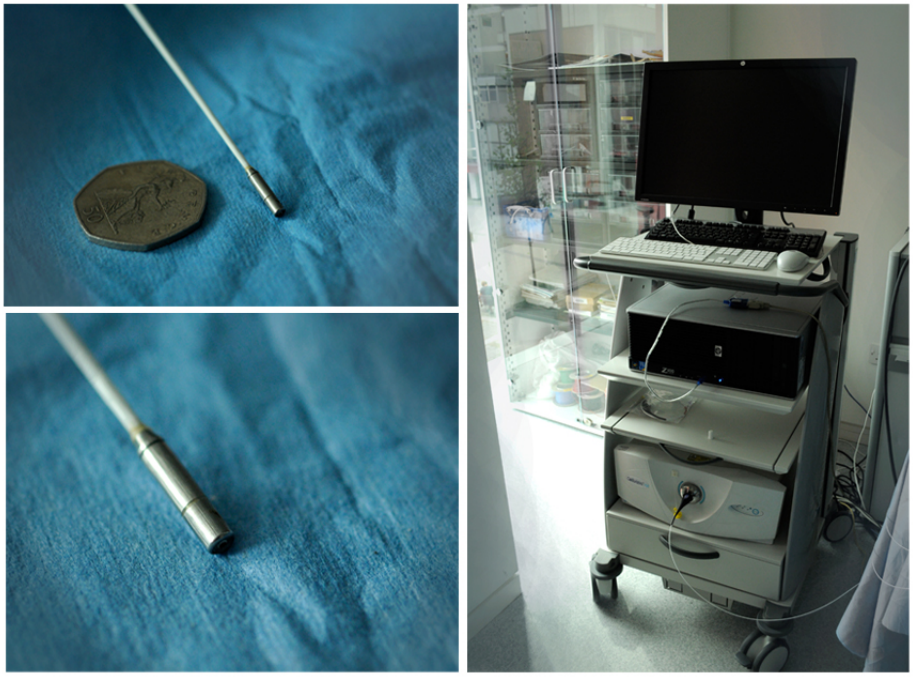


Figure 2.3: The confocal laser endomicroscopy system.

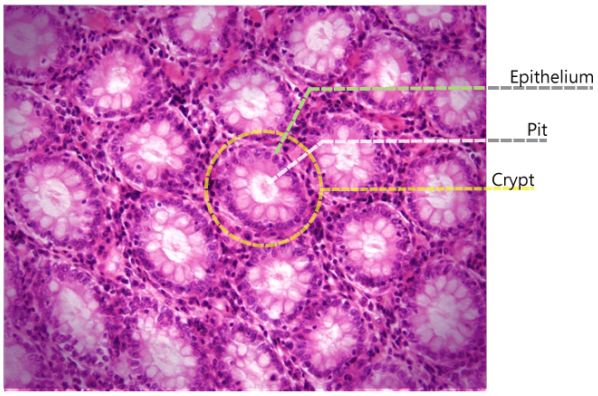
2.2.3 Image acquisition

All CLE imaging systems are optimized by certain contrast agents and the most commonly used agent is "fluorescein sodium". The optimal timing to obtain good images should be within the first 10 minutes after contrast agent is injected into the blood circulation and it is recommended that the probe should be placed perpendicular to the surface of colon wall and maintaining probe stability is essential for good image acquisition.

2.2.4 CLE classification of the colon

In 2009, there was a group of gastrointestinal experts proposing the `Miami Classification System` for normal and abnormal states in several gastrointestinal diseases using pCLE. The classification system distinguished neoplastic from benign disease of the colon based on a color, regularity and thickness of epithelial layer of the colon. Typical confocal images of 4 different colon conditions are shown in figure 2.5, 2.6, 2.7, and 2.8.

Figure 2.4: This image illustrates colon structure at tissue level. [15] `Epithelium` is defined as the thin tissue which forms the lining the alimentary canal e.g. intestines. `Crypt` is a gland structure which is normally formed by the epithelial lining of the alimentary canal especially small intestine and colon. The function of the structure is to produce and secrete enzyme. `Pit` is the opening of a thin hollow located in the middle of crypt structure. Normally, dense network of blood vessels are found surrounding the crypt structure. However, it is invisible in the illustration.



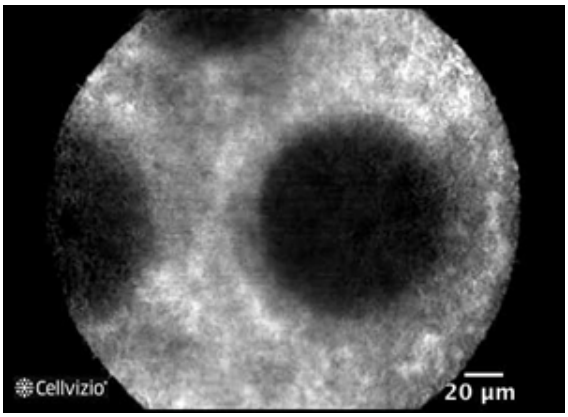


Figure 2.5: Confocal pattern of normal colon tissue shows round crypt structures consisting of dark cells and regular narrow blood vessels surrounding crypt structures.

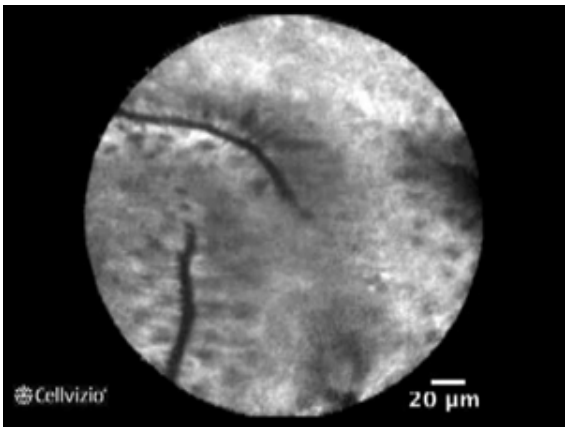


Figure 2.6: Confocal pattern of hyperplastic polyp shows crypts with slit or stellate openings (pits) and bright non-thickened uniform epithelium.

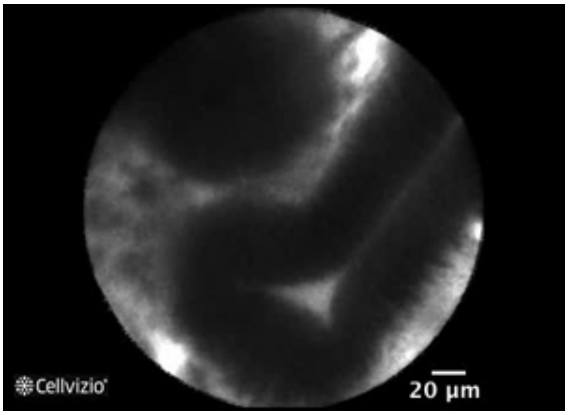


Figure 2.7: Confocal pattern of adenoma polyp shows irregular or villiform structures and dark irregularly thickened epithelium.

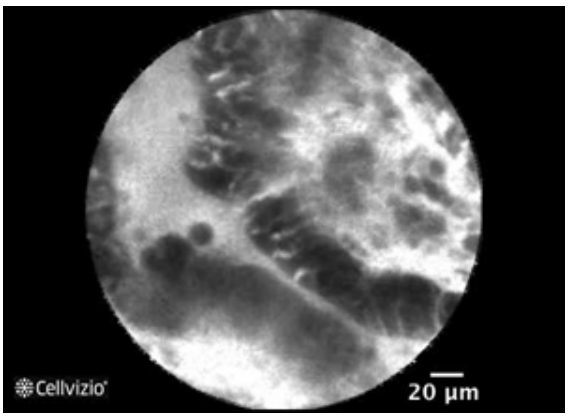


Figure 2.8: Confocal pattern of adenocarcinoma shows disorganized villiform or lack of structure, dark irregularly thickened epithelium and dilated blood vessels.

In summary, colon cancer is considered as one of the leading health burdens worldwide. Nowadays, the disease surveillance is an important strategy to reduce such burden due to early disease detection. Disease screening with colonoscopy is widely used in clinical practice because not only it can provide informative imaging results but also it allows endoscopists to give medical treatment to patients during ongoing endoscopy. Confocal laser endomicroscopy has recently emerged and successfully integrated with conventional endoscopes. This new imaging technique enables endoscopists to visualize histological characteristics of the colon in addition to macroscopic images routinely obtained from conventional endoscopy. Recently, a group of gastrointestinal experts have developed a clinical practice guideline that suggests how to distinguish different colon diseases based on confocal laser endomicroscopic images. In practice, pattern recognition is the key skill required in order to interprete and make a diagnostic judgement. Due to this basis, the pattern recognition is apparently the key problem that has to be solved and in this project, we experiment whether our approach with "texture analysis" can be a valid method to recognize different patterns in confocal images of the colon.

Chapter 3

Content-based image retrieval (CBIR)

3.1 Overview

Content-based image retrieval (CBIR), also known as query by image content (QBIC), is an application of computer vision techniques to deal with the problem of searching images in large databases.

The word "content-based" means that image retrieval algorithm will depend on the actual contents of the image, not metadata (e.g. keywords or annotation). The term "content" in this context refers to any information that can be derived from the image properties or low-level visual features, e.g. color, texture ,and shape.

The simple work flow of CBIR is shown in figure 3.1.

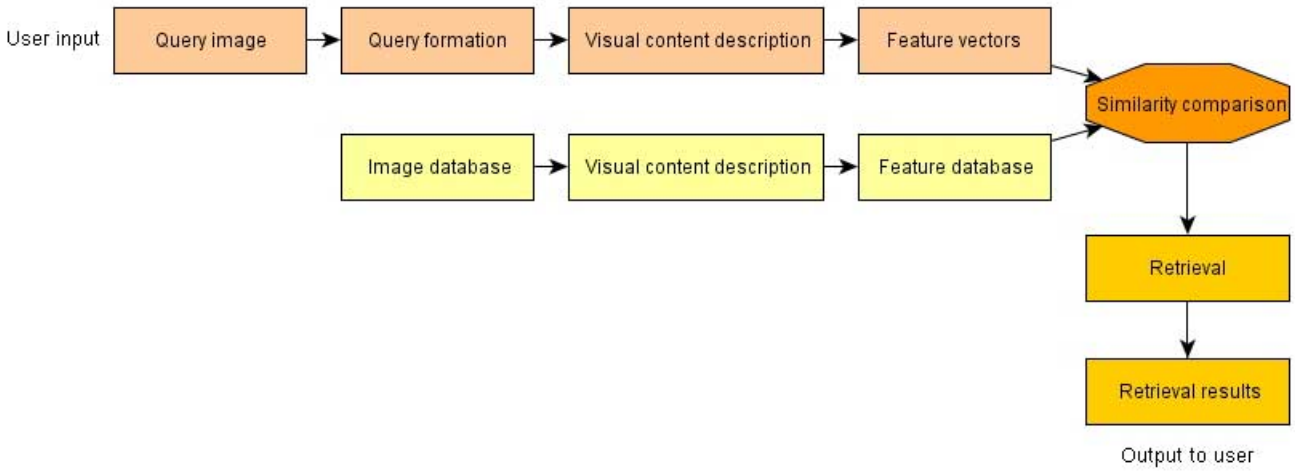


Figure 3.1: Work flow of general content-based image retrieval system

Basically, the CBIR will retrieve images that are similar to user-defined specifications according to its image analysis algorithms. The user can explain their specifications to the system by simply providing example image or sketches.

Image content can be analysed in several ways depending on low-level visual features of interest. These features will eventually form an `image descriptor` which can represent the original image. This concept is very important because the search algorithm will entirely rely on the image descriptor. An image descriptor is normally characterized by

- An extraction algorithm to encode image features into feature vectors
- A similarity measure to compare two images

The image analysis algorithms are usually distinct for different image domains i.e. narrow and broad domain. This project mainly focuses on confocal laser endomicroscopic images and tends to have a limited and predictable variability. Hence it is classified as narrow domain.

As seen from the previous example images, the confocal endoscopic image provides grayscale images, not color-based, and we can notice that confocal endoscopic images have an appearance similar to texture images. As a result, the main feature on which this project focuses is "texture".

3.2 Texture feature

Texture features are intended to capture repetitive patterns within images. The role of texture features in domain-specific retrieval systems, such as medical imaging, is particularly crucial because they tend to relate closely to the underlying semantics in such cases. Texture features have long been studied in image processing and computer vision fields. [17] So far, the commonly-used techniques include second-order statistics, fractals, wavelet transforms and multi-orientation filter banks [18].

Among several texture-based approaches, `Gabor` wavelet [19] has been shown to give superior results when compared to other texture approaches [20]. Interestingly, Gabor function also has a strong evidence that simple cells in the primary visual cortex of mammalian brains can be modeled by Gabor function. [21] This means that image processing using Gabor function may offer results that are similar to the perception in the human visual sytem. Thus, we employed Gabor filters to represent textures in our work.

3.2.1 Gabor Filters

Gabor filter (or Gabor wavelet) has been widely adopted to analyze texture features for image retrieval ([22], [23]) and its efficiency has already been proven. Basically, Gabor filters are a group of two-dimensional wavelets and each wavelet captures energy at a specific frequency and a specific direction. Expanding a signal using this basis gives a localized frequency description, hence it can capture local features or energy of the signal. After applying these wavelets, texture features can then be extracted according to the resulting energy distributions. The scale (or frequency) and orientation tunable property of Gabor filter makes it especially useful for texture analysis.

For a given image I(x, y), its discrete Gabor wavelet transform can be derived by a convolution

$$G_{mn}(x,y) = \sum_s \sum_t I(x-s, y-t)\psi *_{mn} (s,t)$$

where $\psi *_{mn}$ is the complex conjugate of ψ_{mn} which is class of self-similar functions which can be generated from one mother wavelet by dilation and rotation. The mother wavelet is shown as following.

$$\psi(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp[-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2})] \exp(j2\pi Wx)$$

The self-similar Gabor wavelets can be obtained by this generating function

$$\psi_{mn}(x,y) = a^{-m}\psi_(\tilde{x}, \tilde{y})$$

`m` specifies scale or frequency and `n` specifies orientation

$$\tilde{x} = a^{-m}(x\cos\theta + y\sin\theta)$$

$$\tilde{y} = a^{-m}(-x\sin\theta + y\cos\theta)$$

where by θ can be calculated by $\theta = \frac{n\pi}{N}$, $a = (\frac{U_h}{U_l})^{\frac{1}{M-1}}$, and $W_{mn} = a^m U_l$

The rest variables shown in the abovementioned equation are also defined as follow.

$$\sigma_{x,m,n} = \frac{(a+1)\sqrt{2ln2}}{2\pi a^m(a-1)U_l}$$

$$\sigma_{y,m,n} = \frac{1}{2\pi \tan(\frac{\pi}{2N})\sqrt{\frac{U_h^2}{2ln2} - (\frac{1}{2\pi\sigma_{x,m,n}})^2}}$$

In our implementation, we use the constants $U_l = 0.05$ and $U_h = 0.15$ according to our observation. The abovementioned method is implemented in this project and applying these Gabor filters to an image results for every image pixel in a texture vector

$$\boldsymbol{t} = [t_1, t_2, \ldots, t_d]^T$$

where d is the number of scales times the number of orientations that are used in the filter bank. In this project, we use four scales (M = 4) and six orientations (N = 6) so that $d = 24$. The figure 3.2 roughly depicts our Gabor filters bank. After applying 24 different Gabor filters, 24 resulting filtered images can be produced as shown in figure 3.3.

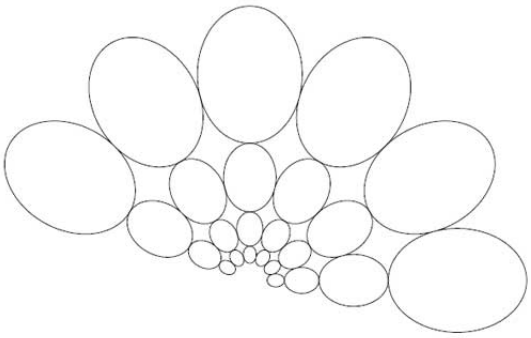


Figure 3.2: This Gabor bank uses four scales and six orientations.

3.3 Texture signature

A single texture descriptor, which is directly derived from an image, can lead to good distance measures only among homogeneous-texture images. As previously mentioned, the confocal endomicroscopic images of the colon is likely to be non-homogeneous and they can be seen as `multi-textured image`. Hence, the approach with only single texture descriptor tends to "overlook" some information of the confocal images.

An image can actually produce nearly as many texture vectors as it has pixels. In order to represent the full distribution of an image in a compact way, we should find dominant clusters in the M x N dimensional texture space, where M and N are the number of scales (M = 4) and orientations (N = 6) ,respectively. Hence this project uses 24 dimensional space for this step. The texture content of an entire image can then be represented by a distribution of texture features, i.e. a cloud of points in a space of 24 dimensional space.

To find dominant clusters, we use a K-means clustering algorithm because it is simple and fast. The number of clusters "K" is predefined as 8 in this project because the confocal images of the colon tend to have multiple textures within one image and a large number of cluster will be able to represent the distribution in texture space more reliably compared to lower cluster numbers. This algorithm will then return a varied number of clusters (up to 8) depending of the complexity of original images i.e. an image can contain a various range of different textures inside it. As the number of clusters increases, more subtle details will be captured. [24] Finally, the resulting set of

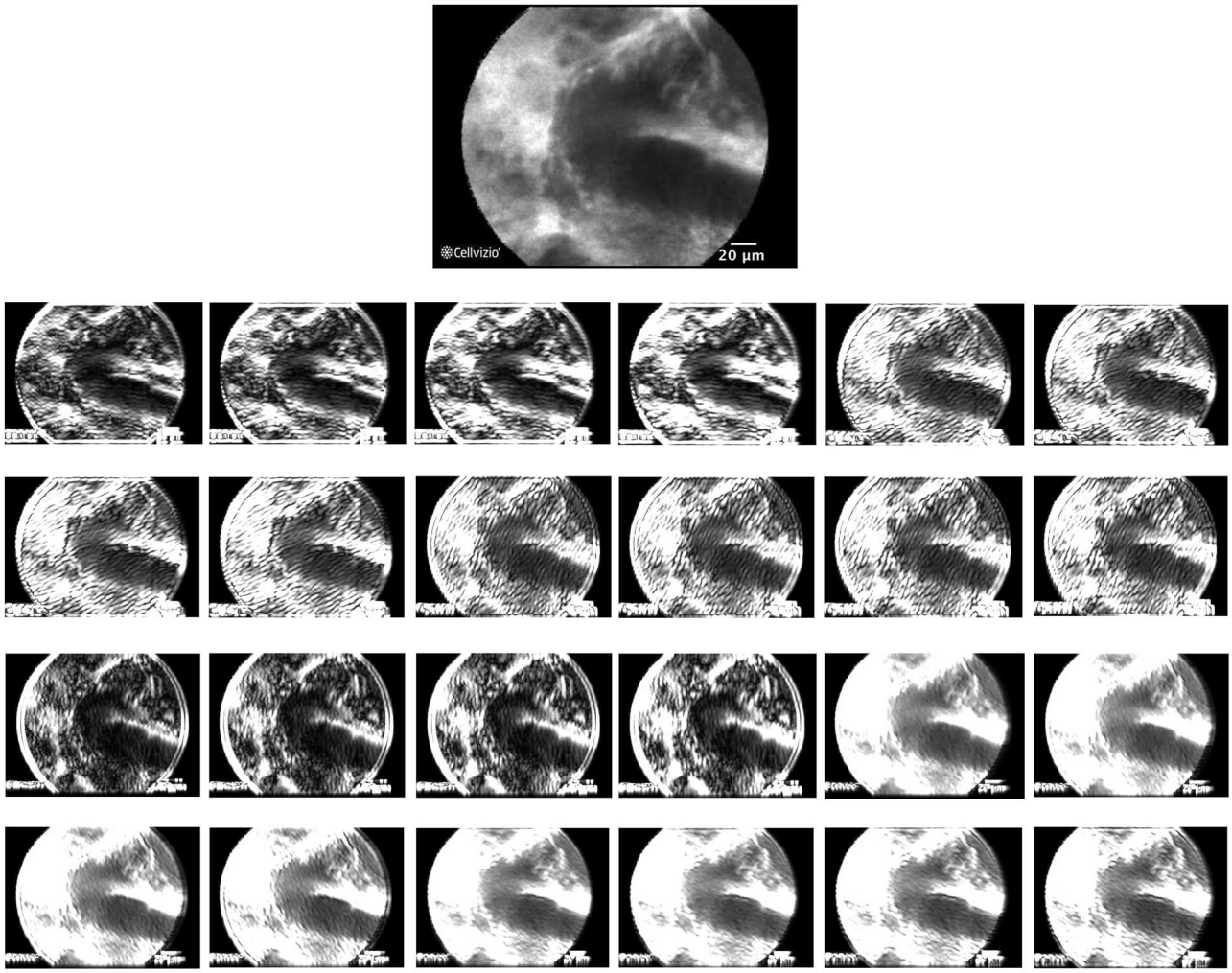


Figure 3.3: This is a set of resulting filtered images after applying our Gabor filters bank that contains 4 different scales and 6 different orientations.

cluster centers along with the fractional cluster weights then become the `texture signature` (or image descriptor) of the image. Figure 3.4 shows an example of texture feature representing one image.

Below shows a simple format of texture signature with m clusters.

$$P = \{(\boldsymbol{p}_1, \omega_{\boldsymbol{p}1}), \ldots, (\boldsymbol{p}_m \omega_{\boldsymbol{p}m})\}$$

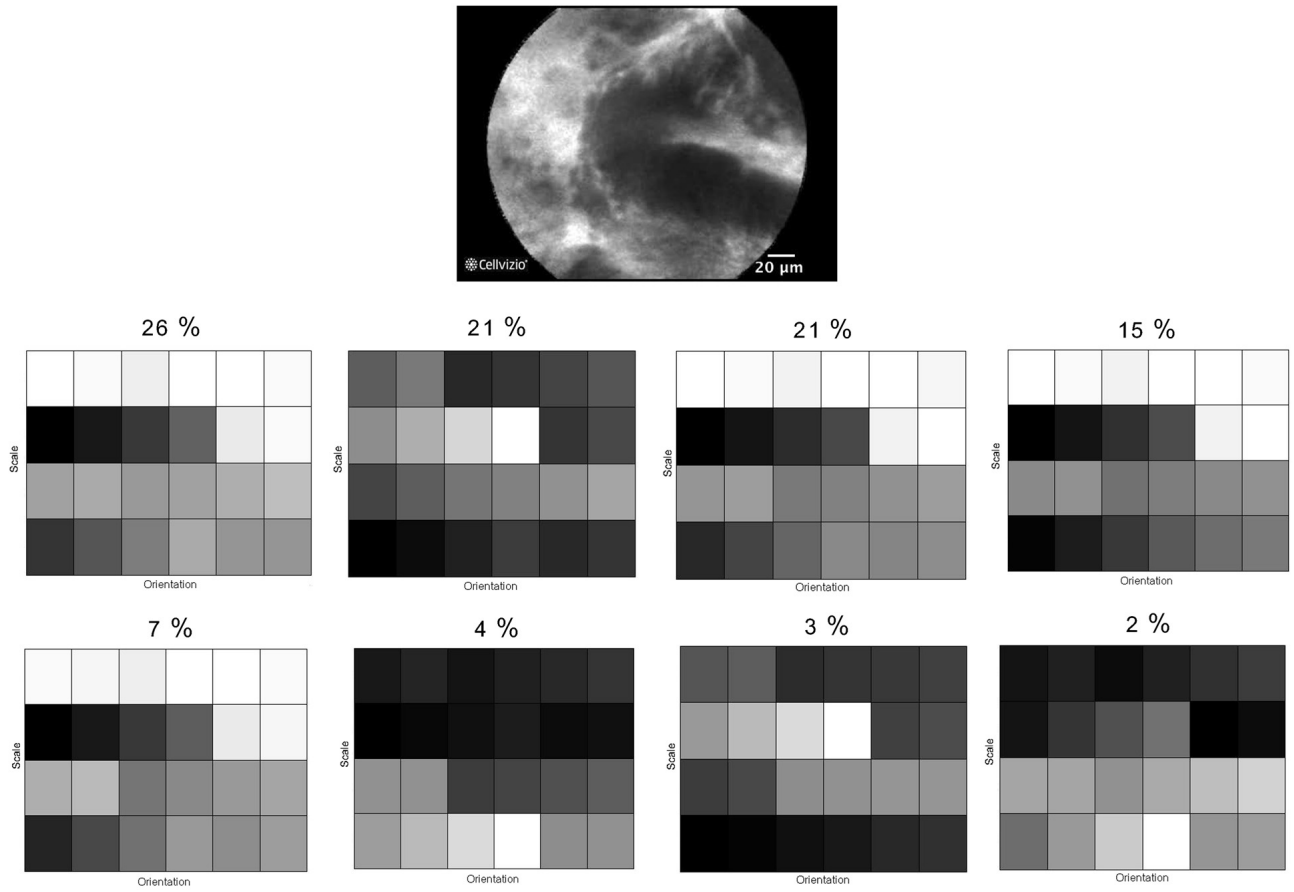


Figure 3.4: The texture signature of this confocal image consists of 8 clusters. The numbers in percentage labelled on top of each energy map represent the weight of each cluster. Brighter squares represent stronger responses.

3.4 Texture distance with Earth Mover's Distance (EMD)

3.4.1 Approach to multi-texture image problem

As mentioned earlier, the confocal images of the colon are likely to be non-homogeneous texture and contain multiple textures. This project will compare a multi-texture query image with other multi-texture images in the database and, based on our literature review, "texture segmentation" is one of the commonly-used approach to resolve this kind of problem. Basically, images are first split into several regions of uniform textures, and a similarity measure is subsequently applied between pairs of such regions. However, two major concerns can occur based on this segmentation approach [24]:

- Texture segmentation is a complicated task and the notion of "uniform texture" is not always well defined.

- Image retrieval based on comparing texture segments is usually sensitive to both over-segmentation and under-segmentation.

Due to these drawbacks, this project, instead, employs the `Earth Mover's Distance` (EMD) to handle individual image descriptors. Based on this similarity measure, both complete and partial queries on multi-texture images can be effectively achieved.

3.4.2 Earth Mover's Distance (EMD)

The Earth Mover's Distance (EMD) is an effective method that aims to evaluate dissimilarity between two multi-distributions in certain feature spaces where a distance measure between features is given. Normally the distance in the feature space is called `ground distance`. The EMD uplifts this distance from individual features to full distributions. [24]

When we talk about two distributions represented by signatures, one can be seen as a "mass of earth" spread in a space while the other can be seen as a group of holes in the same space. Then, the EMD measures the least amount of work needed to fill those holes with earth. A unit of work can be derived from the product of a unit of earth and a unit of ground distance, which is a distance in the feature space. The EMD is originally based on the transportation problem proposed by Hitchcock in 1941 [25] and can be solved efficiently with linear optimization algorithms using its special structure.

Let $P = \{(\boldsymbol{p}_1, \omega_{\boldsymbol{p}1}), \ldots, (\boldsymbol{p}_m, \omega_{\boldsymbol{p}m})\}$ be the first signature containing m clusters, where p_i is the representative of each cluster and $\omega_{\boldsymbol{p}i}$ is the weight of each cluster, and let $Q = \{(\boldsymbol{q}_1, \omega_{\boldsymbol{q}1}), \ldots, (\boldsymbol{p}_n, \omega_{\boldsymbol{q}n})\}$ be the second signature with n clusters. $DIST = [dist(\boldsymbol{p}_i, \boldsymbol{q}_j)]$ is the ground distance matrix where dist(p_i, q_j) is the distance between clusters $\boldsymbol{p}_i$ and $\boldsymbol{q}_j$. The EMD between P and Q is then

$$EMD(P,Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} dist(p_i, q_j)}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}}$$

where F = $[f_{ij}]$, with $f_{ij} \geq 0$ the flow between p_i and q_j , is the optimal flow from P to Q that minimizes the numerator subject to the constraints below

$$\sum_{j=1}^{n} f_{ij} \leq \omega_{\boldsymbol{p}i}, \sum_{i=1}^{m} f_{ij} \leq \omega_{\boldsymbol{q}j}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = min(\sum_{i=1}^{m} \omega_{\boldsymbol{p}i} \sum_{j=1}^{n} \omega_{\boldsymbol{q}j})$$

EMD provides several useful properties such as

- Each image can have "adaptive representation" of high-dimensional distributions of features independently
- It offers robustness to small variations of feature values
- There is no need for segmentation. Due to the representation with a finite number of clusters, it does not suffer from over and under-clustering
- Partial match is possible

In terms of ground distance for texture signature, the following equation is used

$$dist(p,q) = 1 - \exp^{-\frac{\|p-g\|_1}{D}}$$

where p and q are two texture vectors and D is a constant that differentiate between "far" and "close" distances in feature space. This project uses

$$D = d(0, \frac{1}{2}\sigma)$$

where 0 is a zero vector, $\sigma = [\sigma_1 \ldots \sigma_d]^T$ is a vector of standard deviations of the feature components in each dimension from the overall distribution of images in the entire database, and d is the dimensionality of feature space.[24]

3.4.3 Rotation and scale invariance

Rotation and scale invariance can be achieved by an `exhaustive search` for the "minimal" distance over all possible shifts in both orientation and scale.

$$EMD(E_1, E_2) = min_{\substack{n_s=0,\ldots,N-1 \\ m_s=0,\ldots,M-1}} EMD(E_1, E_2, n_s, m_s)$$

where M is the number of scales and N is the number of orientation.

In summary, our main objective of this project is to create an application which has the ability assist endoscopists to intepret confocal images of the colon by providing more information regarding image patterns. The CBIR is a one of the most promising techniques that analyzes patterns of images and eventually return similar images to the user. It can be a difficulty for clinicians when they have to make a judgement based on information of a single confocal image. Due to this rationale, the retrieved images from our CBIR framework can contribute as the additional information for clinicians when medical judgement has to be made from confocal images of the colon. In this project, our approach towards the pattern recognition will be texture analysis. Having reviewed the literature, we learned that Gabor filters bank is a powerful method to distinguish different patterns especially texture-based images. We hereby explore whether Gabor approach can work successfully on confocal images of the colon. Fundamentally, Gabor filters can easily be tuned and adapted to match the details of image contents. We can also experiment on different designs of Gabor filters bank to find out what design matches best to our project. In terms of similarity calculation, we employ a well-known technique "Earth Mover's distance" to calculate texture distance along with rotation- and scale- invariant features so that the search will be less sensitive to the changes of scales and orientation.

Chapter 4

Relevance feedback with manifold embedding

4.1 Relevance feedback

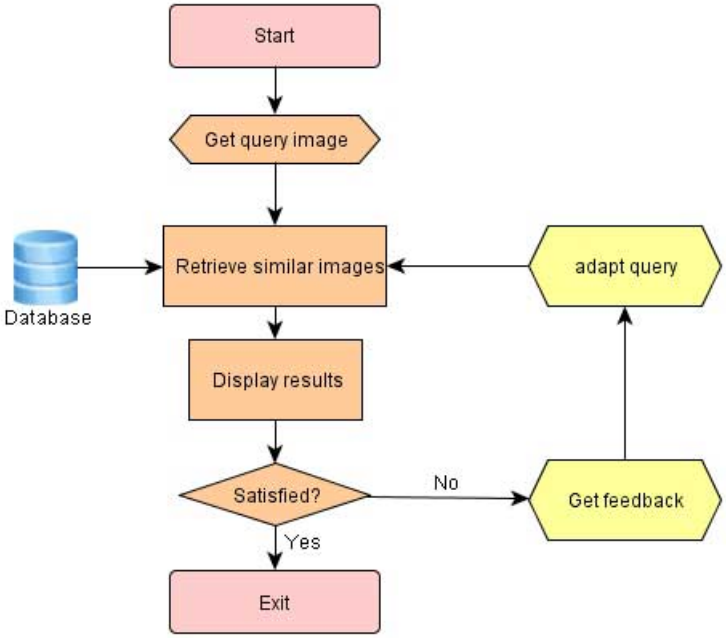


Figure 4.1: Relevance feedback loop

Since early 1990's, when CBIR was firstly proposed, several techniques in this area have been continuously developed. At its early stage, research primarily focused on exploring various visual feature representations, hoping to find the right representation for each feature and two concerns were relatively ignored [28]

- The gap between high-level concepts and low-level features or `semantic gap`:

  Humans tend to use their high-level concepts in everyday life whereas what current computer vision techniques can extract from image are mostly low-level features. In a general setting, the low-level features do not have a direct link to the high-level concepts.

- The subjectivity of human perception of visual content:

  Different people, or the same person under different timing or circumstances, may perceive the same visual content differently.

Due to these two reasons, the performance of CBIR system in its early stage was not satisfactory. To narrow down such `semantic gap`, it was proposed that "human intervention" should be integrated into the CBIR and the system should allow the user to easily provide their evaluation of the retrieval results to the system in order to improve the retrieval performance. That is relevance feedback technique.

`Relevance feedback` is a long-standing powerful concept which has been used in traditional text-based information retrieval system before. It is the process of automatically adjusting the existing query using the information fed-back by the user about the relevance of previously retrieved objects such that the adjusted query can be a better approximation to the user's information need. [26] [27] In the relevance-feedback-based approach, the retrieval process becomes interactive between the computer and user. Under the assumption that high-level concepts can be captured by low-level features, the relevance feedback technique tries to bridge the gap between high-level concepts and low-level features with information from the users feedback. The figure 4.1 shows the work flow of CBIR with relevance feedback loop.

As relevance feedback strongly relies on the user's interaction, response time becomes an important concern next to the accuracy of the retrieval. The relevance feedback is a process that inevitably involves with the user's patience because they need to go through the system. Having taken this concern into account, we are then interested in employing the embedded manifold into the relevance feedback process in order to improve the retrieval efficiency.

4.2 Manifold

4.2.1 Overview

The need of dimensionality reduction is unavoidable step when working with large volumes of high-dimensional data. Usually, the real-world data, such as digital images, has high dimensionality and in order to deal with this kind of data more efficiently, its dimensionality requires to be reduced. Dimensionality reduction is a technique that aims to find meaningful structures hidden inside the original high-dimensional data and then construct a new (reduced) data representation by embedding them into a low-dimensional manifold. Ideally, the reduced data representation should have a dimensionality that corresponds to the intrinsic dimensionality of the original data. The intrinsic dimensionality of data is defined as the minimum number of parameters needed to account for the observed properties of the data.[30] Dimensionality reduction is important in several domains, since it alleviates the curse of dimensionality and other undesired properties of high-dimensional spaces. [31]

Dimensionality reduction was traditionally performed under linear techniques such as Principal Components Analysis (PCA). However, such linear technique cannot adequately handle complex nonlinear data. Therefore a large number of nonlinear techniques for dimensionality reduction have been developed and proposed. These new techniques have the ability to deal with complex nonlinear data. This is very important because the real-world data is likely to be nonlinear or highly nonlinear.

The nonlinear technique of particular interest to this project is `Isomap`, which have been investigated and successfully applied in the field of medical image analysis [32] [33] [34] [35].

4.2.2 Isomap

Isomap stands for "isometric feature mapping". [29] It is an powerful technique to solve dimensionality reduction problems by using easily measured local metric information to learn the underlying global geometry of a data set. In other words, it is a technique that tries to preserve "global properties" of the original data in the low-dimensional representation.

Normally, the classical techniques for dimensionality reduction , such as PCA and MDS, are simple to implement, efficiently computable, and guaranteed to discover the true structure of data lying on a linear or near linear subspaces of the high-dimensional input space [36]. Unlike those classical techniques , Isomap has capability to discover the "nonlinear" degrees of freedom that

underlie complex natural observations of the original data. As most of real-world data sets contain nonlinear structures, they are invisible to those classical techniques [37] [38].

Isomap essentially combines major algorithmic features of PCA and MDS (computational efficiency, global optimality, and asymptotic convergence guarantees) with the flexibility to learn a broad class of nonlinear manifolds. Figure 4.2 illustrates an example of the nonlinear data lying on a two-dimensional "Swiss roll". It can be observed that points far apart on the underlying manifold, as measured by their geodesic distances, may visually mislead that they are close in the high-dimensional input space, as measured by their straight-line "Euclidean" distance. Only the geodesic distances effectively reflect the true low-dimensional geometry of the manifold, not the Euclidean.

Isomap builds on classical MDS but, at the same time, seeks to preserve the intrinsic geometry of the original data, as captured in the geodesic manifold distances between all pairs of points. The crux of the Isomap is to estimate the geodesic distance between faraway points, given only input-space distances. For neighboring points, input-space distance provides a desirable approximation to geodesic distance. For faraway points, geodesic distance can be approximated by adding up the "short hops" between neighboring points. These approximations can be computed efficiently by finding shortest paths in a graph with edges connecting neighboring points.

The complete isometric feature mapping (Isomap) algorithm has three important steps:

- Construct neighborhood graph

  This step determines which points are neighbors on the manifold M based on the distances $d_X(i,j)$ between pairs of data points. Here, a pair of two data points are represented by i and j in the input space X.

  Tenenbaum [29] proposed two simple methods used to connect each point to its neighborhood. The first one is to connect each point within a fixed radius $\in$ and the other is to connect to all of its K nearest neighbors. Afterthat, these neighborhood relations are then represented as a weighted graph G over all the data points, with edges of weight $d_X(i,j)$ between neighboring points.

- Compute shortest paths

  In this second step, the estimated geodesic distances $d_M(i,j)$ between all pairs of data points on the manifold M are derived by computing the shortest path distances $d_G(i,j)$ in the graph G from the first step. The algorithm to find the shortest paths is explained as below.

  Initialize $d_G(i,j) = d_X(i,j)$ if i, j are linked by an edge and $d_G(i,j) = \infty$ otherwise. Then for each value of $k = 1, 2, ..., N$ in turn, replace all entries $d_G(i,j)$ by $min\{d_G(i,j), d_G(i,k) + d_G(k,j)\}$. The matrix of final values $D_G = \{d_G(i,j)\}$ will contain the shortest path distances between all pairs of points in G.

- Construct d-dimensional embedding

  The final step essentially applies the classical MDS method to the matrix of graph distances $D_G = \{d_G(i,j)\}$, and then construct an embedding in a d-dimensional Euclidean space Y that best preserves the manifolds intrinsic geometry. The coordinate vectors y_i for data points in space Y are chosen to minimize the below cost function

  $$E = \|\tau(D_G) - \tau(D_Y)\|_{L^2}$$

  where D_Y represents the matrix of Euclidean distances $\{d_Y(i,j) = \|y_i - y_j\|\}$ and $\|A\|_{L^2}$ the L^2matrix norm $\sqrt{\{\Sigma_{i,j} A_{ij}^2\}}$.

  The τ operator is defined by by $\tau(D) = -\frac{HSH}{2}$, where S is the matrix of squared distances $\{S_{ij} = D_{ij}^2\}$ and H is the centering matrix $\{H_{ij} = \delta_{ij} - \frac{1}{N}\}$.

  The τ operator has the function to convert distances to inner products which characterize the intrinsic geometry of the data in a form that supports efficient optimization. The global

minimum of $E = \|\tau(D_G) - \tau(D_Y)\|_{L^2}$ is achieved by setting the coordinates y_i to the top d eigenvectors of the matrix $\tau(D_G)$.

In terms of eigenvectors, let λ_p be the p-th eigenvalue (decreasing order) of the matrix $\tau(D_G)$, and let v_p^i be the i-th component of the p-th eigenvector. Then set the p-th component of the d-dimensional coordinate vector y_i equal to $\sqrt{\lambda_p v_p^i}$.

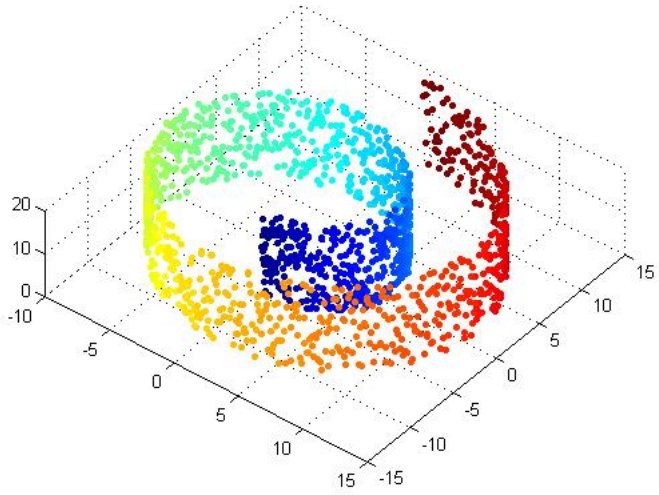


Figure 4.2: An example of nonlinear data "Swiss roll".

In all cases, when the dimensionality of embedded space is increased, the residual variance will decrease. The intrinsic dimensionality of the data can be estimated by findind the "elbow" sign at which the curve ceases to decrease significantly with added dimensions. For example, in figure 4.3, the true or approximate intrinsic dimensionality according to this graph should be at 2D.

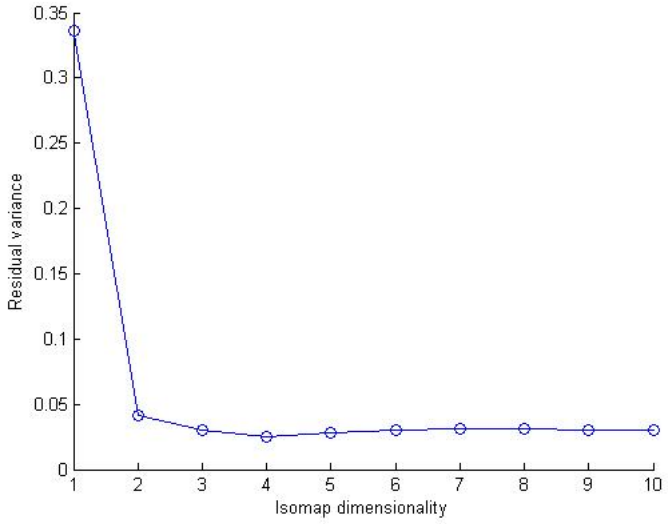


Figure 4.3: The relationship between residual variance and different dimensionalities.

Additionally, according to [29], Isomap is guaranteed asymptotically to recover the true dimensionality and geometric structure of a strictly larger class of nonlinear manifolds. It is also proven that Isomap can even find globally meaningful coordinates and non-linear structure from a complex dataset which does not have a clear manifold geometry.

To conclude this chapter, we expect that relevance feedback will play an important role to our CBIR framework as we believe it can reduce semantic gap. The information from user's feedback will be the principal input for subsequent search and it is expected that the retrieved results in the next search will become more and more satisfactory and meet user's need. In this project, we decided to employ a dimensionality reduction into our relevance feedback mechanism because it has potential to enhance the performance of image search and finally bring an efficiency to our relevance feedback process. The technique of interest here is Isomap which proves to be highly efficient to preserve global data properties and generally applicable to a broad range of dimensionalities and data sources including images. In this work, Isomap is applied to an database of 1000 confocal images and our relevance feedback will operate on the embedded space that is derived from Isomap.

Chapter 5

Software implementation and final outcome

5.1 Initial design decisions

Programming language The language of choice for this project is `MATLAB` which can be integrated with `C++` using MEX-files.

Platform We choose to primarily target Windows based machines during our application development. However other platforms can also be supported because we can use MATLAB Compiler^{TM} to build a standalone executable and then run with MATLAB Compiler Runtime(MCR).

Graphical user interface (GUI) design We use `MATLAB GUIDE` (Graphical user interface Development Environment) to design and program GUIs for this project because this development tool can easily create plots and effectively handle images at the same time.

5.2 Project development tools

Below is a brief summary of the development tools used in this project.

MATLAB This is our main development environment.

Doxygen This is used to automatically generate well-organized documentation from the source code. The pdf format and HTML documentation can easily be generated.

yED Graph Editor We use this application to generate high-quality diagrams for this report.

Adobe Illustrators CS5 This is another application we use to generate high-quality diagrams for this report.

Adobe Photoshop CS5 This image editor software enables us to produce beautiful-looking figures for this report.

DVDVideoSoft Free Studio We use this free software to split confocal endomicroscopic video files into several still JPEG images.

5.3 User interface

Figure 5.1 shows the main user interface of our program. From the menu the user can ...

Database related process

- Load database

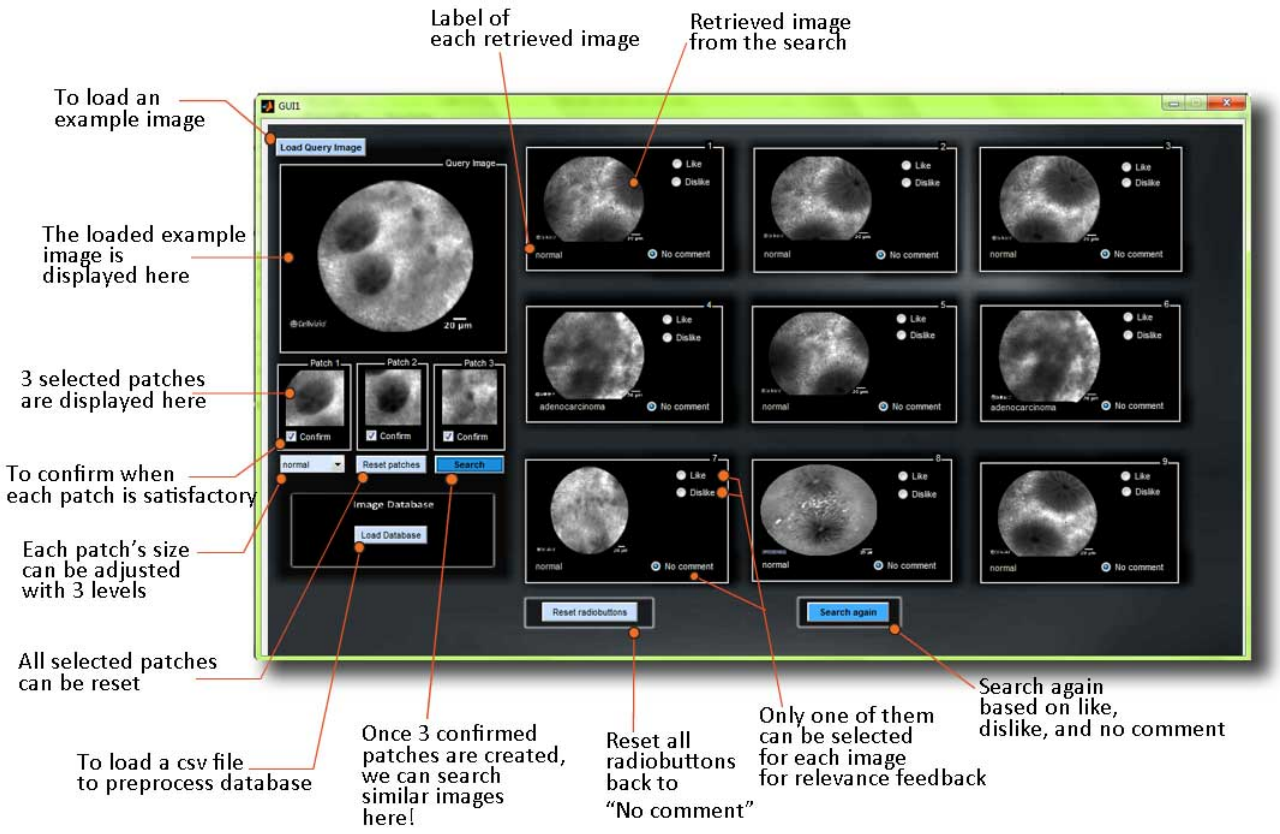


Figure 5.1: The main user interface

Query image related process

- Load a query image
- Pick three arbitrary patches (small square area) within the query image
- Adjust the size of patches (3 different sizes)
- Check boxes to confirm the patch chosen until 3 patches are completely chosen
- Clear all the patches and begin to pick 3 patches again
- Search for similar images once 3 patches are completely chosen

Relevance feedback related process

- Visualize the nine most similar results
- Click `Like`, `Dislike` radiobutton, or simply leave as `No comment`
- Reset all `Like` and `Dislike` radiobuttons to `No comment`
- Search again according to like-dislike information

5.4 Principal solutions for each process

Here, we particularly discuss about details of certain implementations which are considered as core methods for each process.

5.4.1 Database related process

Image signature generation

This section will give details about how confocal image database is preprocessed. In this step, we use `csv` file to feed data into the program. The sequence of data in the `csv` file consists of

- Image index
- Image filepath
- classification of the image i.e. normal, hyperplastic, adenoma,or adenocarcinoma

Our aim in this step is to preprocess all of the confocal images in database and store in a single `mat` file. In other words, this `mat` file will contain all of the processed information and can eventually represent the entire database. The structure inside this `mat` file is an array of MATLAB `struct` datatype.

Currently, this project works on 1000 different confocal images of the colon classified by clinical experts based on Miami classification. Our database consists of

- 300 images of normal colon
- 300 images of hyperplastic colon
- 300 images of adenomatous colon
- 100 images of adenocarcinoma

The format of data in the abovementioned `struct` datatype consists of 4 different information which are

- Image filepath (`string`)
- Classification (`string`)
- Image signature (`cell`)
- A row vector (`double`, 1-by-24 matrix)

In order to produce image signatures, we apply Gabor filters bank so that useful features can be extracted from original images. In this project, Gabor filters bank consists of 4 different scales (frequencies) and 6 different rotations. As a result, 24 different filtered images are then created. In terms of parameters and constants for Gabor filters, we have experimented on different set of values and, eventually, $U_h = 0.15$ and $U_l = 0.05$ are selected.

We first tried our Gabor filters bank on simple uniform textures such as wooden wall (figure 5.2). The good point of this example is that we can clearly notice the shifted energy pattern in the energy map after we apply 90 degree rotation to the original image. In energy map, brighter squares represent stronger responses to Gabor filter.

After trying the Gabor filters bank with the uniform texture, we then experimented further on confocal images. The example of adenocarcinoma classification is shown in figure 5.3.

Applying these 24 Gabor filters to an image results for every image location in a texture vector i.e. every pixel of an image can represent as one point in 24-dimension space.

In order to represent the image signature in a compact way, we then employ K-means clustering algorithm to find dominant clusters from the cloud of points in the previously mentioned 24-dimension space. In this project, we predefine K as 8 because the confocal images of the colon are likely to contain multiple textures and a large number of cluster (K) is required to reliably represent the distribution in texture space.

After clustering, two important information will be collected, the center of each cluster (in 24-dimension space) and each cluster's weight. The weight here means the number of points which

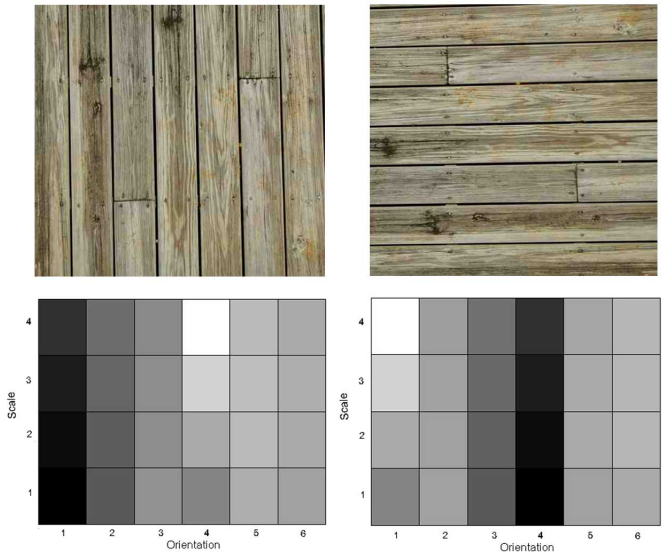


Figure 5.2: The left picture is a "wooden wall" image with its energy map and the right picture is a rotated version along with its energy map. The shifted energy pattern can also be observed.

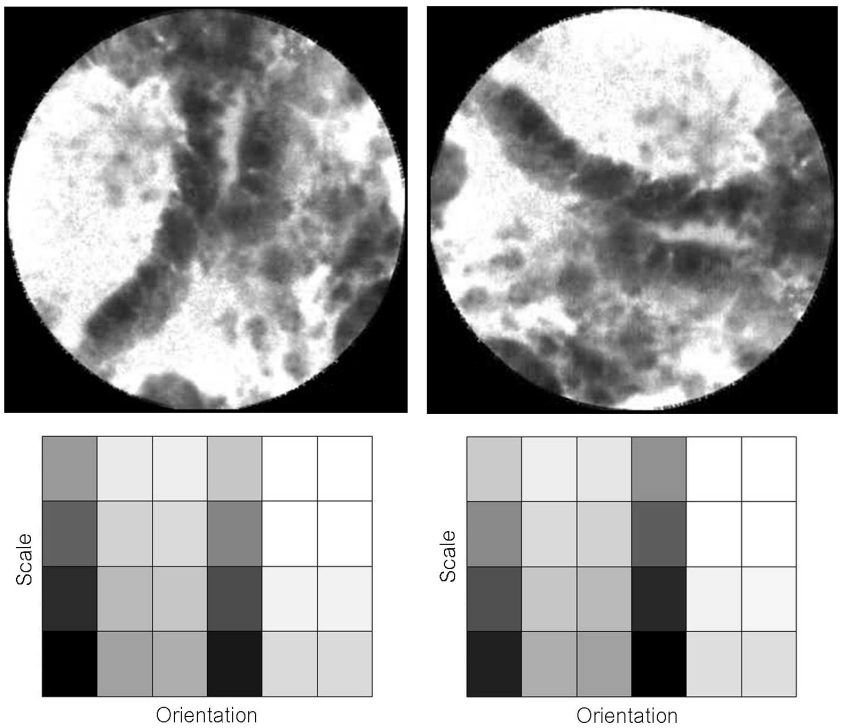


Figure 5.3: The left picture is a confocal image of adenocarcinoma of the colon with its energy map and the right picture is a rotated version along with its energy map. The shifted energy pattern can also be observed.

belong to the each cluster. The resulting set of cluster centers together with the fractional cluster weights eventually becomes the signature (or texture signature) of an image.

This is an example of image signature with m clusters.

$$P = \{(\boldsymbol{p}_1, \omega_{\boldsymbol{p}1}), \ldots, (\boldsymbol{p}_m, \omega_{\boldsymbol{p}m})\}$$

where each $\boldsymbol{p}$ is a row vector and each ω is an integer.

For the row vector in the abovementioned `struct` datatype, it is 1-by-24 matrix. Each element in the row vector represents the mean value of all pixels in each resulting image after applying Gabor filters bank. As 24 different filtered images are created, there are 24 elements in the row vector. This vector will be used later on when texture distance is calculated under the EMD method.

To conclude, a `mat` file, which is created as previously described, will then contain the information of 1000 different confocal images under the MATLAB `struct` datatype.

After the signature generation step, there are still some steps required to be precalculated in order to prepare the database before the similarity search step.

Ground distance of EMD for similarity calculation

For our project, we use the EMD to calculate similarity and the ground distance is shown in the following equation.

$$dist(p, q) = 1 - \exp^{-\frac{\|p-g\|_1}{D}}$$

p and q are texture vectors and D is a constant that differentiate between "far" and "close" distances in feature space. This project uses the constant

$$D = d(0, \frac{1}{2}\sigma)$$

where 0 is a zero vector and $\sigma = [\sigma_1 \ldots \sigma_d]^T$ is a vector of standard deviations of the feature components in each dimension from the overall distribution of images in the entire database, and d is the dimensionality of feature space which is "24" in our project.

The vector of standard deviations can be derived from row vector (1-by-24 matrix) in the abovementioned `struct` datatype. Inside each row vector, each component represents mean value of pixel energy in each dimension, i.e. the j^{th} element of a row vector of an image is the mean energy value of axis j in 24-dimension space.

In dimension j, the standard deviation σ_j can be calculated from element j^{th} of row vectors of all images in the database. Once the vector of standard deviations is successfully obtained, the constant D can be calculated from L_1 norm with a zero vector. Finally, ground distance can be identified and naturally adapted to the variation within our database. In another word, the constant D can dynamically change according to the image pool in the database.

5.4.2 Query image related process

Based on our literature review, majority of image retrieval systems tend to use the entire example image directly and calculate image similarity. However, most of confocal images of the colon are non-homogeneous and tend to have more than one (uniform) textures within one image, processing the whole example image as a single query image may not return the result that the user expects to see. Moreover, some confocal images even contain multiple conditions of the colon. Due to these reasons, we explore another approach by focusing on particular part of example images in which the user is interested and then treat each of these part as a query image. In this project, we use 3 "image patches" to represent these small part. Hence there are 3 small query images to be handled each time when we search. We do not process the example image as a whole.

Image patches

We use MATLAB function `getpt` to pinpoint a part of image. This step entirely depends on the user's preference and interest. A patch can be easily constructed by gathering pixels surrounding the picked point within a particular range. Our program offers 4 different patch sizes (ranges) which are squares of 80x80, 150x150, 200x200, or 250x250 pixels as depicted in figure 5.4.

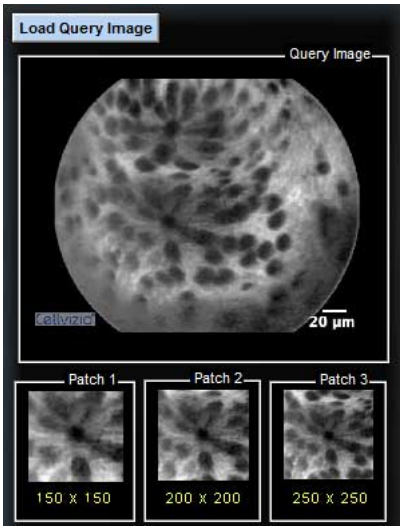


Figure 5.4: A query image of the normal colon with different patch sizes.

Once 3 patches are successfully constructed, we can apply the Gabor filters bank to these patches to create an image signature for each patch. For each patch signature, distances between itself and each image in the database will be computed later on during similarity search process.

Similarity measure

Distance between each patch to images in the database is computed based on Earth Mover's distance method. During the EMD calculation, scale and rotation invariant are also included by implementing with the "exhaustive search" technique. Basically, this method shifts feature elements in all direction. For example, if the original vector is *abcde*, exhaustive search will lead to 4 more vectors which are *bcdea*, *cdeab*, *deabc*, and *eabcd*. We use all of these possible vectors to measure the distance between two images and select the minimal distance from the 5 resulting distance to represent as the final distance.

After the calculation, we are particularly interested in the 9 most similar images returned from each patch. Up to 27 different resulting images can be returned from 3 image patches because it is also possible that different patches can return some identical resulting images. Among these resulting images, we give different scores to them based on the rule that the closer receives higher score. Within the set of 9 resulting images derived from a patch, the highest score is granted to the first in rank and the lower scores are given to the rest in descending order. Similarly but in the patch level, higher score is granted to the first patch and the lower scores are given to the rest in descending order.

Once the scores are assigned to all resulting images, the next step is to search through the list of 27 images and find out whether different patches return identical resulting images. If so, the scores of such cases will be summed up and this reasonably gives these resulting images higher scores.

At the final step, a new list of resulting images is reconstructed and re-ranked according to the new scores. High score guarantees that resulting images either ...

- are very close to the first patch, which the user pays more attention to, `or`
- contains information that is close to more than one patches

The nine best images with the highest scores can eventually be retrieved according to information derived from the selected patches.

5.4.3 Relevance feedback related process

After the resulting images are returned to the user, it is still possible that the results are far from what users expect due to the `semantic gap` between human perception and low-level visual feature like texture signature. With the aim to narrow the gap, we integrate the relevance feedback mechanism into our retrieval system and also explore embedded manifold as method of choice to improve the system performance.

Embedded manifold with Isomap

First of all, we need to precalculate a distance matrix which represents the "distance" relationship between all pairs of images in the database. The distance matrix will be a very important input for manifold calculation. As abovementioned, the distance is calculated based on EMD method with the ground distance that can be adapted according to the variety of images in the database. Again, scale and rotation invariant are also included in this step. As our current database contains 1000 confocal images, the distance matrix consequently has a size of 1000-by-1000. The embedded manifold technique of interest to this project is `Isomap`.

At the initial phase, we experimented the efficiency of Isomap by running two pilot tests with small databases. The two databases include

- the one containing 80 different confocal images with mixed classifications
- the other containing 40 small (image) patches of one confocal image classified as hyperplastic colon

In our first test, there are 4 classifications of the colon and each contributes 20 images in this database. The figure 5.5 clearly shows that Isomap can efficiently reduce dimensionality of our image data into 2- and 3-dimension space. In this figure, the connectivity between points is calculated base on k-nearest neighbor algorithm and here we predefine k as 6.

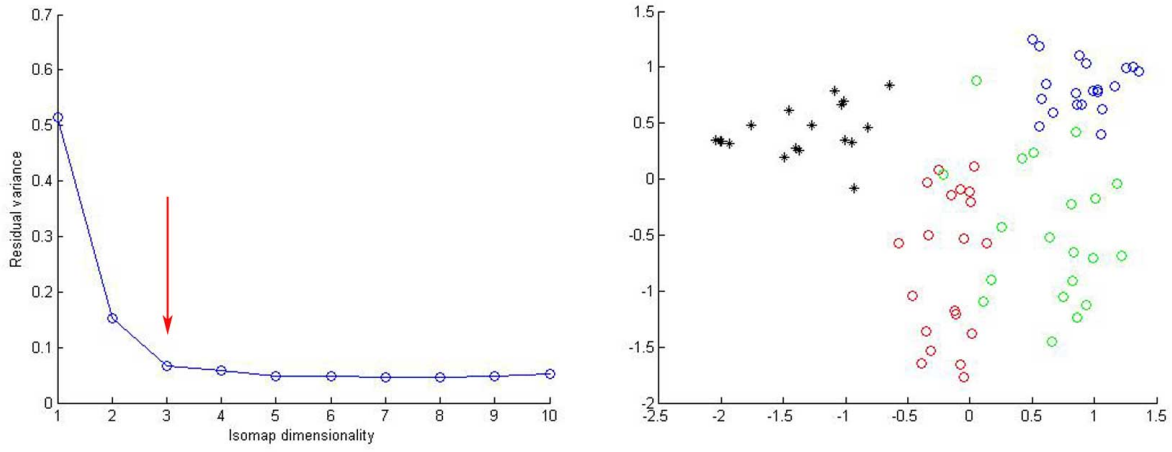


Figure 5.5: Isomap can efficiently reduce dimensionality of complex image data (texture signatures) and represent them in 3-dimension space with significantly low residual variance as indicated by the red arrow sign on the left-hand side. For the k-nearest neighbor algorithm inside, K is predefined as 6 in this case. The right picture shows how the image data are organized in 2-dimension space after applying Isomap. (Color representation: Black-Adenoma, Red-Adenocarcinoma, Green-Normal colon, and Blue-Hyperplastic colon)

We also explore further the relationship between images in this initial database by projecting this image set onto 2-dimension Euclidean space and visualizing with a sample of the original input images. The visualization is shown in figure 5.6. Here we construct the mapping with different parameter from the figure 5.5 by defining k as 4 so that the best residual variance can be obtained for the 2-dimensional projection.

In our second test, the database contains 40 small image patches. These small patches are taken from one large image which is classified as hyperplastic colon. The procedure of taking these patches is illustrated in figure 5.7.

Here, we would like to experiment whether different patches, that are taken from one single direction in the same image, will show any pattern or particular result in embedded manifold after Isomap is applied. We plot 1-dimension coordinates of 40 points against the number of patches. The 1-dimension coordinates can be derived by applying Isomap to our image dataset and reducing the dimensionality to 1-dimension. The top plot of figure 5.8 demonstrates a trend indicating that if two images are quite different, they will subsequently be represented by faraway points in embedded space. The same trend can be observed in 2-dimension embedded space as shown by the buttom plot of figure 5.8. Similar images can be represented by neighboring points in Euclidean space.

In the next phase, we then explored further on our large database which contains 1000 confocal images of the colon. According to Tenenbaum's technical comments [39], the success of Isomap essentially depends on the neighborhood size (`k` in k-nearest neighbours for neighbourhood graph construction). The neighborhood size (`k`) should be neither too large nor too small. If `k` is too large, it will potentially lead to "short circuit" edges in the neighborhood graph, which makes the low-dimensional embeddings not preserving the true topology. If it is too small, the graph will become too sparse to calculate geodesic paths accurately. Having experimented on different values of `k`, we decided to define `k` as 5 and created embedded manifold at 7-dimensionality because optimal residual variance can be observed at this point. The residual variance and a 2-dimensional projection is visualized in figure 5.9.

Once 1000 confocal images are mapped to data points in 7-dimension space, we can utilize these points to calculate similarity between images, based on their coordinates in the Euclidean space. Similarity between images can simply be measured by calculating Euclidean distance between points with $L1norm$, i.e. Two similar images are represented by two points that are close to each other in the 7-dimension space. We integrate this method into our relevance feedback mechanism because the calculation of Euclidean distance is simple and, as a result, it provides the user fast response when they want to search again.

During the relevance feedback step, the user can feedback whether they are satisfied with the results from the first search. If the user is not fully satisfied, the system will be asked to re-calculate similarity but this time onwards the calculation method will be different because it is mainly based on the user's feedback, not the selected patches. The user can simply interact with the system by clicking `like` or `dislike` radio buttons. If they have no comment, they can just leave radio buttons ticked as `no comment`, which is automatically provided by the system from the beginning. (Figure 5.10) Mutual exclusion is also implemented to ensure that only one choice is made for each resulting image.

5.5 Final outcome of program

Comparison between searching with and without image patches

In order to describe the retrieval performance of our system when patch technique is implemented, we demonstrate two examples in figure 5.11 and 5.12 as well as show retrieval results based on an experiment on 20 different confocal images of the colon.

Figure 5.11 shows 2 different image set retrieved from different search approaches. The retrieved images at the top are based on the entire original image which is classified as normal colon whereas the buttom one takes 3 square patches into search calculation instead. The first approach returns 5 normal-colon images and 3 of them contain similar pattern to the example image i.e. consisting of

dark round objects in the middle of smooth grey background. After we apply the second approach with small patches picked from area of interest, 7 normal-colon images are returned and 6 of them contain the pattern of interest.

Figure 5.12 is another example that shows improvement of retrieval system when patches are integrated into the searching. For this time, adenocarcinoma is used as the example image. The first approach returns 2 adenocarcinoma images while the patch approach returns 3 adenocarcinoma images plus 3 adenoma images that contain similar pattern we are looking for, i.e. high brightness at the background with dark thickening rod-like object as the foreground. The high brightness is considered as an important character to describe adenocarcinoma because such brightness is pathologically originated from the leakage of contrast agents from blood vessels. The vessel leakiness is the key manisfestation of most cancers due to the fact that such leakiness can enhance the environment for cancers to grow. [40] Moreover, adenoma and adenocarcinoma have close relationship with each other. Aadenomatous polyp is considered as the pre-cancerous stage and has potential to develop into adenocarcinoma in the future, especially when the polyp size is large.

In addition to the previous examples, we also conducted an experiment on a set of 20 test images, in which each classification contributes 5 different confocal images. We measure the retrieval performance by using `retrieval rate` i.e. the percentage of retrieved images with correct classification in each search. In here, we would like to assess system's retrieval performance using 2 different criteria to determine "accuracy".

- The ability to distinguish between neoplastic and benign colons i.e. If a query image is adenocarcinoma, an accurate result can be either adenocarcinoma or adenoma.

- The ability to retrieve images with correct classification i.e. If a query image is adenocarcinoma, an accurate result has to be adenocarcinoma only.

In table 5.1, we show retrieval results when the ability to distinguish between neoplastic and benign colons is the criteria of accuracy. There are 2 different categories here which is benign and neoplastic.

Retrieval rate based on the ability to distinguish between neoplastic and benign		
Classification	Entire image as query	3 patches as query
Overall	66.67%	87.78%
Normal	91.11%	91.11%
Hyperplastic	82.20%	84.44%
Adenoma	66.67%	91.11%
Adenocarcinoma	26.67%	84.44%

Table 5.1: This table shows the retrieval results when the ability to distinguish between neoplastic and benign colons is the criteria o te accuracy. We compare results between the conventional method using entire example image as a query and our approach using small patches to represent as a query. Overally, our approach outperforms the conventional method (87.78% Vs 66.67%). Similar trend can also be observed in each confocal classification, apart from the normal classification where exactly the same retrieval rate can be achieved by both methods. Normal colon demonstrates the best retrieval rates in both methods and can achieve above 90% retrieval rate when our patch approach is implemented. In addition, adenoma can also reach the same rate to the normal under our approach. Interestingly, adenocarcinoma type shows to improve remarkably when 3 patch method is applied.

Next in table 5.2, we demonstrate retrieval results when the ability to retrieve images with correct classification is the criteria to determine the accuracy of the system.

Here, we conclude that retrieval rate can be improved after the patch technique is implemented into the CBIR framework. The retrieval performances both in overall and in each classification are better. Moreover, adenocarcinoma seems to gain significant benefit from the patch technique.

Retrieval rate based on the ability to retrieve images with correct classification		
Classification	Entire image as query	3 patches as query
Overall	43.89%	58.89%
Normal	48.89%	64.44%
Hyperplastic	53.33%	60.00%
Adenoma	66.67%	88.89%
Adenocarcinoma	6.67%	22.22%

Table 5.2: This table shows the retrieval results when the ability to retrieve images with correct classification is the criteria to determine the accuracy of the system. Again, we compare results between the conventional method using entire example image as a query and our approach using 3 small patches to represent as a query. Overally, our approach shows to outperform the conventional method (58.89% Vs 43.89%). The result also shows similar trend in each confocal classification. Adenoma shows the best retrieval rates in both methods. It also achieves nearly 90% retrieval rate when our patch approach is implemented. Meanwhile, adenocarcinoma shows the least retrieval rates in both approaches but this cancerous type gains the most significant improvement.

Results comparison between initial search and subsequent search with relevance feedback

To further describe the retrieval performance when the relevance feedback step is involved, we demonstrate two examples in figure 5.13 and 5.14.

Figure 5.13 is an example showing results from subsequent search according to user's feedback. This example continues from the previous adenocarcinoma image and 3 images are considered as satisfactory results (labelled with green tick at the top of figure 5.13). After the system re-calculates similarity according to user's information, 5 satisfactory results with correct classification are successfully returned to user. Although the rest of retrieved results are not classified as adenocarcinoma, they all are adenoma.

We can also demonstrate how relevance feedback improves our system performance with benign classification or "hyperplastic colon". Figure 5.14 shows that more satisfactory images can be retrieved after searching again with relevance feedback. In the initial search, the pattern we are interested is black long streak surrounded by moderately thickening margin. At the top of figure 5.14, we can observe that most of returned images obviously contain the thickening pattern but only 2 of them also contain "streak". Hence we vote `like` only to those two images (labelled with green tick). For the subsequent search, 8 hyperplastic colon images are returned and we consider 7 of them as satisfactory due to the combination of characters of interest.

In addition to the previous examples, we also conducted another experiment on 5 testing adenocarcinoma images. We opted to test with this classification because its retrieval rate is the lowest in our previous experiment. We then would like to see whether relevance feedback feature can eventually improve the system performance. Again, we measure the retrieval performance by using `retrieval rate` in the same way as in previous experiment on 20 testing images. In table 5.3, we demonstrate the comparison of retrieval results before and after relevance feedback is called. Similar to the previous experiment, we used 2 different criteria to determine the system's accuracy.

- The ability to distinguish between neoplastic and benign colons i.e. If a query image is adenocarcinoma, an accurate result can be either adenocarcinoma or adenoma.

- The ability to retrieve images with correct classification i.e. If a query image is adenocarcinoma, an accurate result has to be adenocarcinoma only.

Although low retrieval rates are observed among testing adenocarcinoma images in the initial image search, this experiment can demonstrate that the relevance feedback feature remarkably helps to improve the retrieval rates for this particular classification of the colon.

Criteria of accuracy	Before relevance feedback	After relevance feedback
Retrieve adenocarcinoma	13.33%	48.89%
Retrieve either adenocarcinoma or adenoma	62.20%	93.33%

Table 5.3: This table shows the comparison of retrieval results between "before" and "after" the relevance feedback feature is called. Apparently, higher retrieval rates can be achieved when using relevance feedback regardless accuracy criteria of retrieval rates.

In addition to the better retrieved results our system can achieve, the relevance feedback also shows to respond to the user requests quickly and this truly demonstrates a good quality of relevance feedback mechanism.

5.6 Further extension to Eye Tracking system integration

As the ultimate goal for this computer-aided decision support system is to assist endoscopists to make a diagnosis based on information of confocal laser endomicroscopic images in clinical routine practice. The system ideally should be able to integrate into real endoscopic sessions. However, the way endoscopists interact with the system will become another important issue as "mouse" manipulation can no longer be a convenient choice for the user-computer interaction process. We would like to explore further whether manipulation based on a mouse can be replaced by other methods. Here, the method of interest is `Eye tracking system` which has the ability to estimate the direction of eye gaze of the user.

Recently, we have initiated a preliminary experiment and the objective of our experiment is to find out how feasible the eye tracker can effectively be utilized to capture image patches from query images. For the very first step, we study about the correlation between "the image locations where the user looks at" and "the image locations where he/she selects patches". Our methodology of experiment is designed to collect 2 set of data, which are

- Cursor movement history which can be represented by arrays of coordinates along the time when the user interacts with the CBIR system
- Time history (in milli-second) according to each cursor event e.g. cursor movement and mouse clicking

These 2 kinds of data have to be collected from both the CBIR and the eye tracking system. Once we obtain the data from both side, we can analyze whether these 2 set of data from 2 different systems can match and synchronize. The result of this data analysis will provide a useful information whether the eye tracking system can be considered as a promising tool to help capturing image patches from query images in our CBIR system. The result of data analysis will be reported in the near future.

5.7 Chapter summary

To conclude this chapter, we summarize our implementation methods along with final outcome here in short.

Regarding database-related processes, 1000 confocal images in the database are transformed into compact image signatures by applying our Gabor filters bank along with finding dominant clusters in texture distribution. One thousand signatures from this step can then represent 1000 confocal images in the database. These signatures are essential when image search is requested according to an example image. Signatures are further processed and dimensionality reduction is applied in this step. By applying Isomap, complex image signatures are projected onto low dimensionality and each point in the low dimension space represents one image in the database. Coordinates of these points will be utilized in the relevance feedback step.

In terms of search-related process, patch technique is implemented in the query image step. Each patch is transformed into an image signature by Gabor filters bank just like the way we process each image in database. Similarity (texture distance) will be calculated for each patch using Earth Mover's distance with rotation- and scale- invariant features. More scores wil be given to images with lower distance. After summing the scores up, 9 images from database with the highest scores will be returned as the 9 most similar images according to the contents of the 3 patches.

For the relevance feedback step, the user is asked to select satisfactory and/or unsatisfactory images that are retrieved from the first search. Distances in low-dimension Euclidean space are then calculated according coordinates of those satisfactory images. Again, high scores wil be given to points close to satisfactory ones. After summing the scores up, 9 images with the highest scores will be returned as the 9 most similar images according to the user's feedback. The relevance feedback can be operated as many times as the user wants until high satisfaction is reached.

Regarding to the final outcome and system's retrieval performance, our experiment shows that implementation of patch approach to the initial phase of image search is beneficial and yields more retrieval rate when compared to the conventional method using the whole example image as a query. Furthermore, the confocal images of adenocarcinoma, which contain strong non-homogeneity within images, appear to benefit most. Retrieval performance can also be further increased with the help of relevance feedback mechanism. This step shapes the retrieval results towards the user's requirement with high efficiency.

For the eye tracking integration, our preliminary experiment is ongoing.

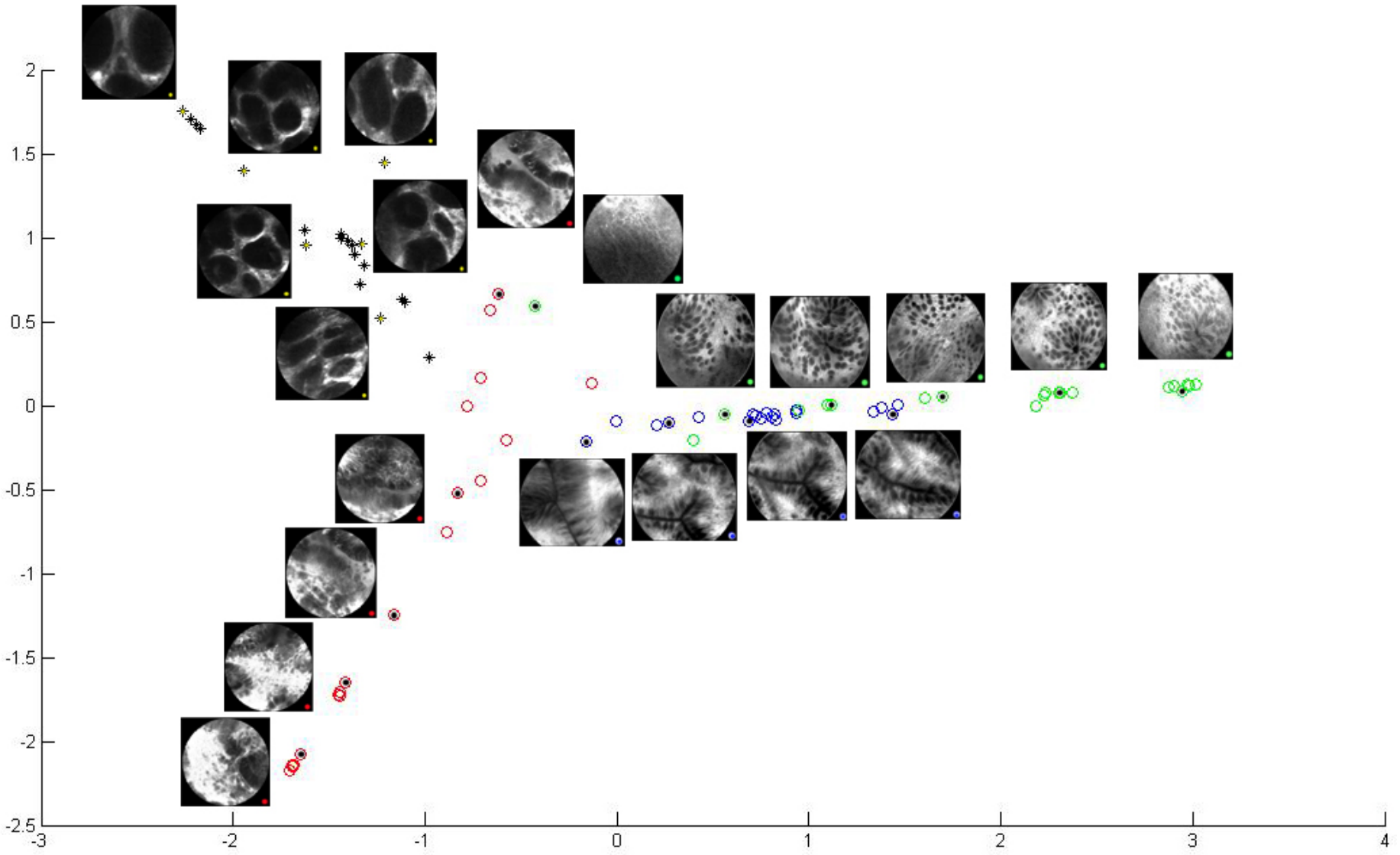


Figure 5.6: We demonstrate here that 80 different confocal images from our initial database are successfully mapped into the low-dimensional space. There are 4 different classifications which are adenoma(Black), adenocarcinoma(Red), normal(Green), and hyperplastic colon(Blue). We also show a sample of the original input images with the 2-dimensional projection.

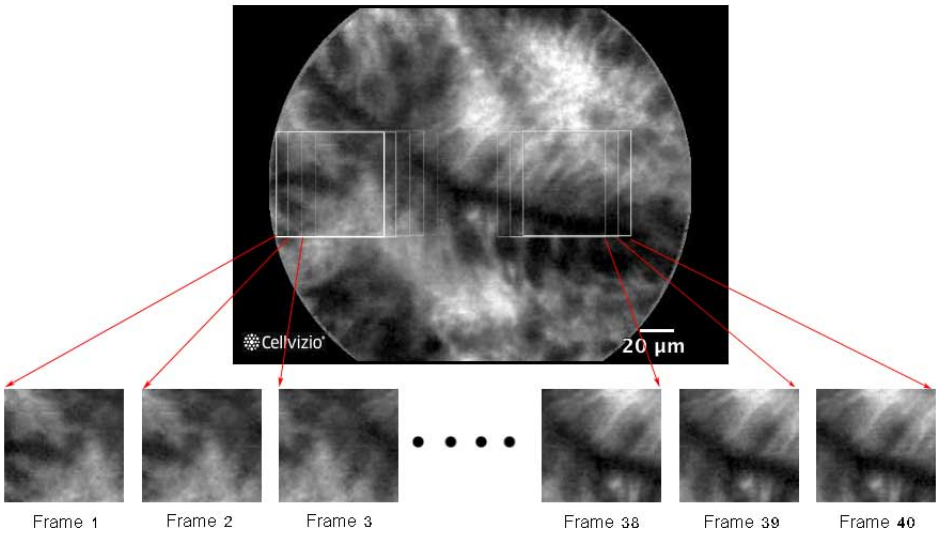


Figure 5.7: The figure illustrates how 40 image patches can be taken from the original image. One single direction (parallel to x-axis of the image) is applied in order to obtain image patches.

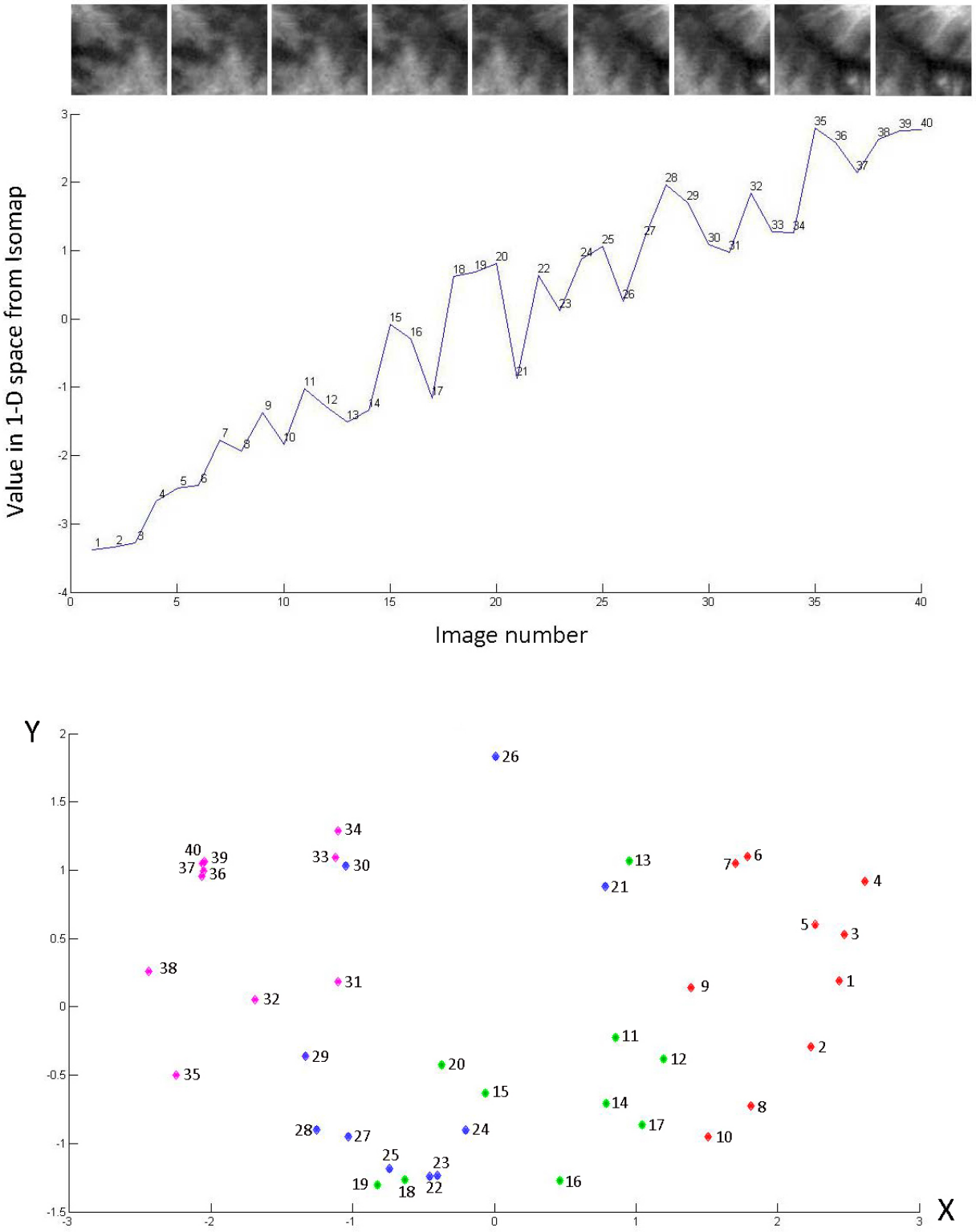


Figure 5.8: Top: 1-dimension coordinates of 40 points are plotted against the number of patches. Buttom: 2-dimension projection of the same image set after dimensionality reduction is applied. Both plots demonstrates the same trend indicating that the similarity of images can be represented by the distance in Euclidean space.

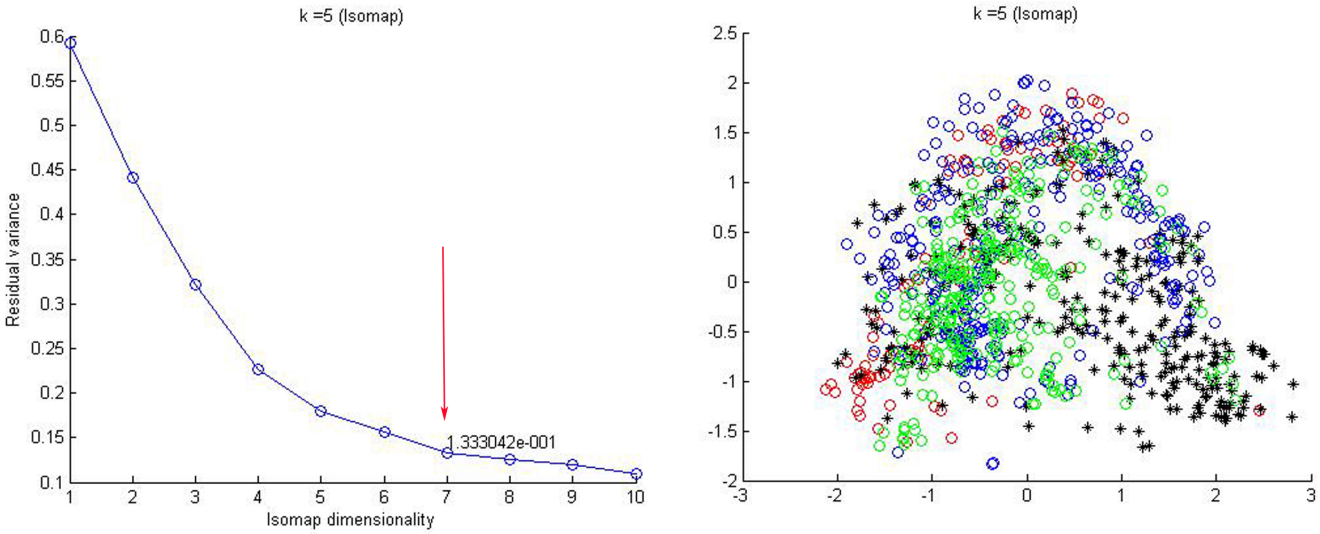


Figure 5.9: The left graph shows that the curve stops to decrease significantly with added dimensions when 7-dimensionality is reached. The residual variance is approximately 0.1333042 and we define `k` as 5. On the right, two-dimensional projection of 1000 data points is visualized with 4 different classifications which are adenoma(Black), adenocarcinoma(Red), normal(Green), and hyperplastic colon(Blue).

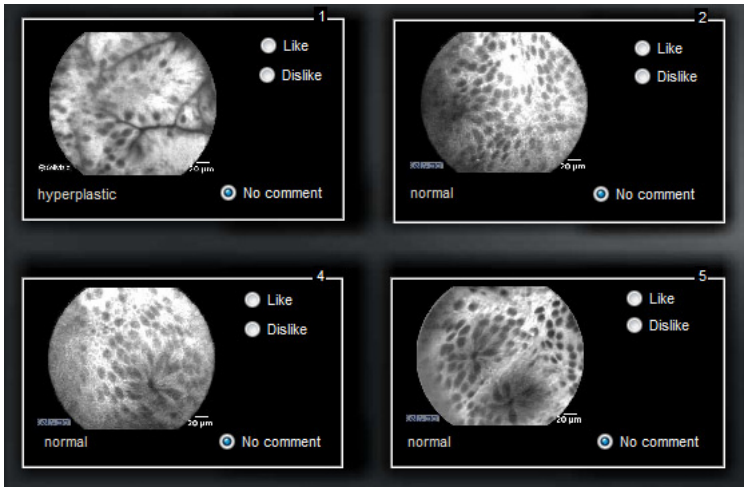


Figure 5.10: This figure illustrates the radio buttons, the feature that allows the user to express their satisfaction for each resulting images. The `no comment` buttons are automatically provided by the system from the start to simplify process for the user.

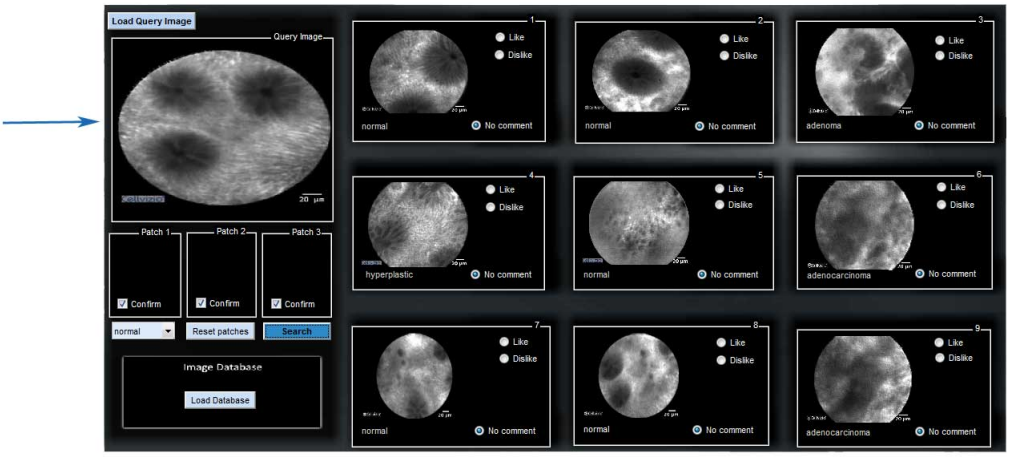


Image search based on the original example image

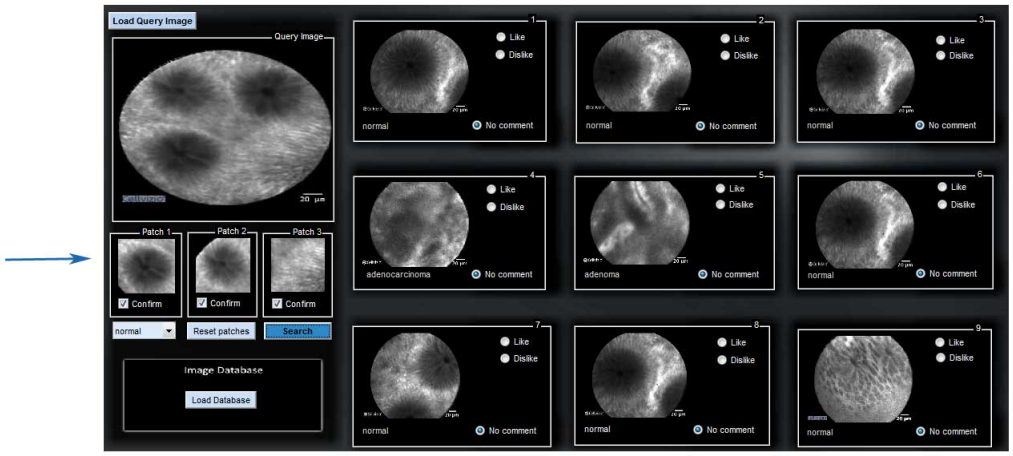


Image search based on 3 selected small patches

Figure 5.11: Normal tissue of the colon is used as the example image here. The result at the top is based on similarity calculation from the entire original image. It returns 5 normal-colon images and 3 of them contain similar pattern to the example image (Pattern: dark round objects floating in the middle of smooth grey background). The result at the buttom is, instead, based on 3 square patches partially selected from the original image. This approach returns 7 normal-colon images and 6 of them contain the pattern of interest.

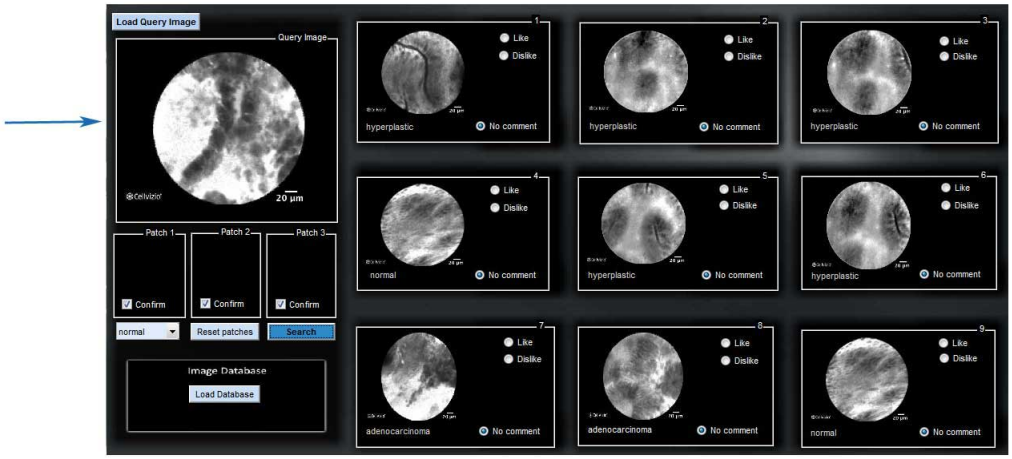


Image search based on the original example image

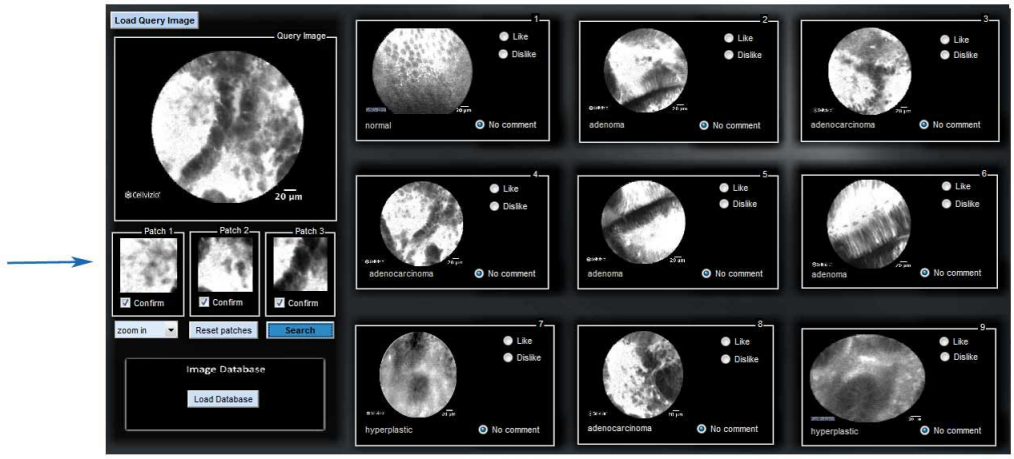
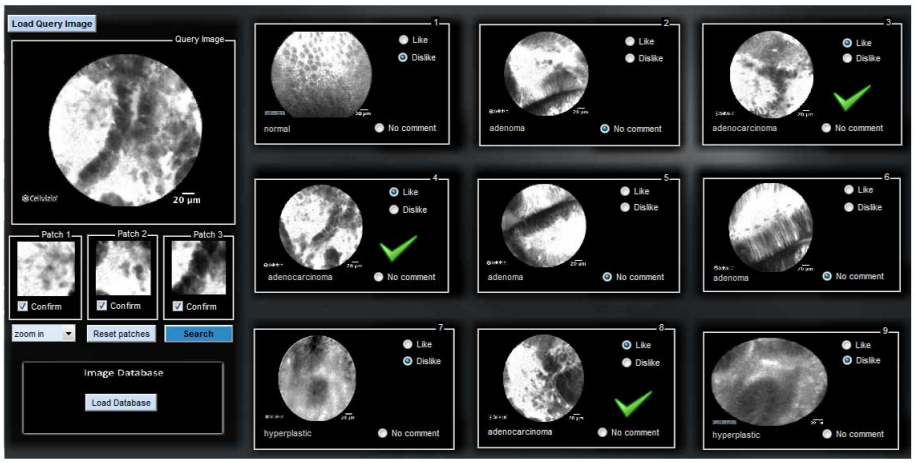
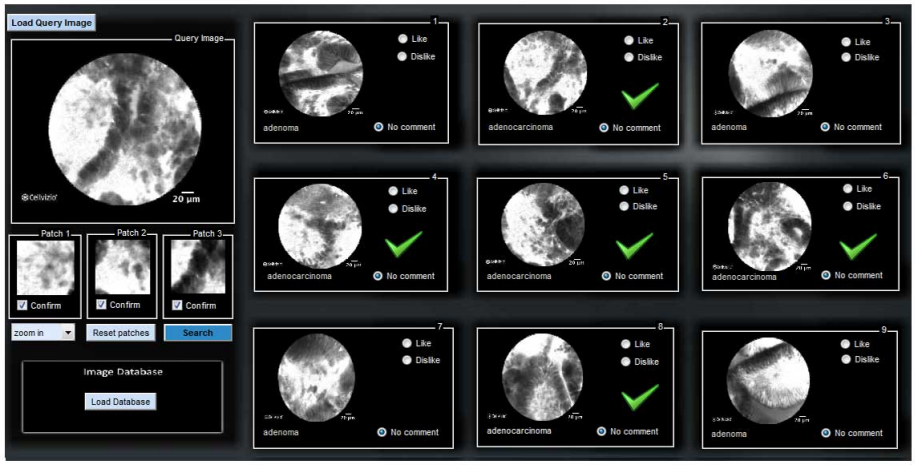


Image search based on 3 selected small patches

Figure 5.12: Adenocarcinoma is used as the example image here. The result at the top is based on similarity calculation from the entire original image and it returns 2 adenocarcinoma images. Instead, the result at the buttom is based on the patch approach and it returns 3 adenocarcinoma images along with 3 adenoma images with the pattern of interest. (Pattern: dark thickening rod-like objects floating on very bright background.)

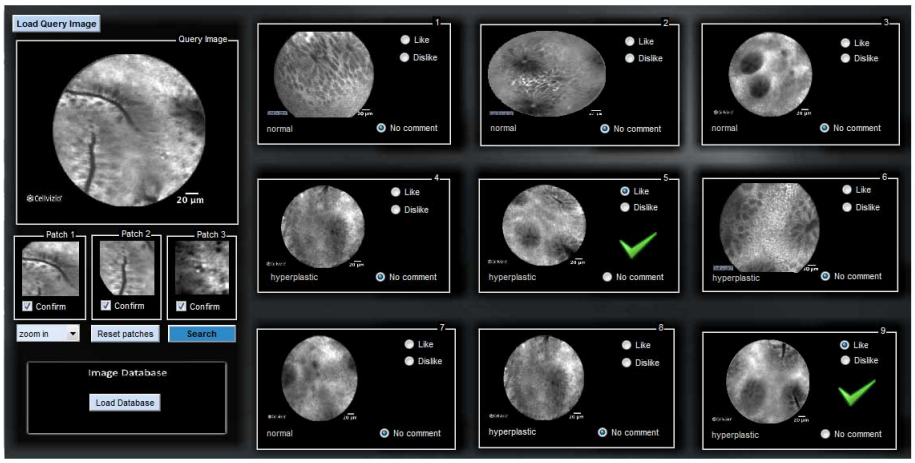


First search based on 3 selected small patches



Search again with relevance feedback using dimensionality reduction

Figure 5.13: Further search is performed on the previous adenocarcinoma example image. The results at the top are based on our previous search according to 3 small patches and 3 images are considered as satisfactory results. They will be an important input for the next search. The buttom demonstrates the results from the subsequent search based on user's feedback and the similarity is calculated on the 7-dimension embedded manifold as previously described.



First search based on 3 selected small patches

Search again with relevance feedback using dimensionality reduction

Figure 5.14: Hyperplastic colon is used as the example image here. The result at the top is derived from the initial search and it returns 5 hyperplastic colon images. However, only 2 of them are considered as satisfactory results because we look for a pattern with black long streak in particular. The result at the buttom is derived from subsequent search with relevance feedback. It returns 8 hyperplastic colon images and we consider 7 of them as satisfactory.

Chapter 6

Visualization

In addition to the visualization in our pilot test (80 confocal images) as shown in figure 5.5, we also created a mini visualization tool which allows us to understand more about the relationship between images in our final database with 1000 images. In this mini program, we can plot data in two-dimension and handle the plot with standard features including zooming, panning, and picking points. The GUI of the program is shown in figure 6.1.

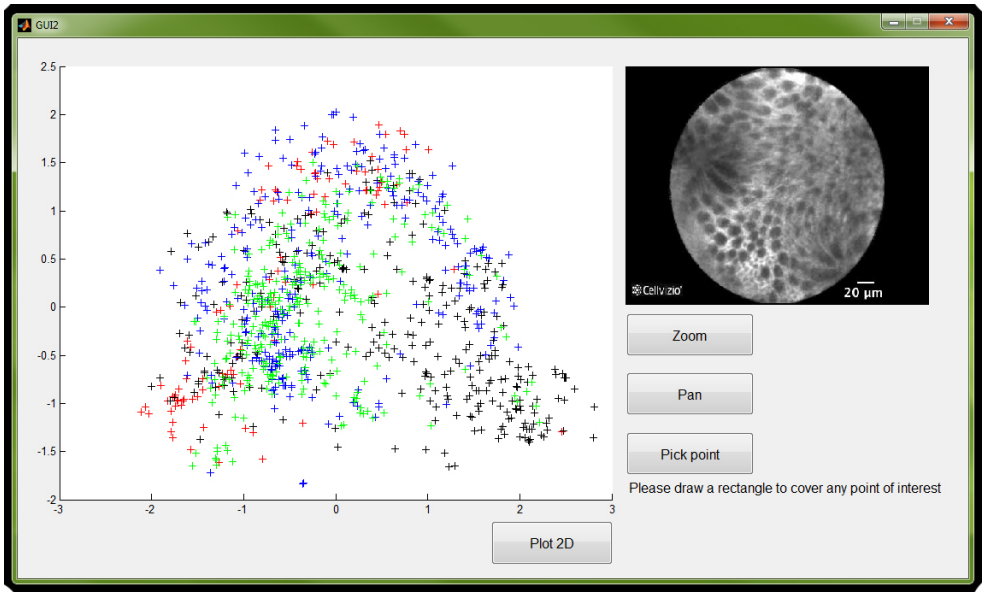


Figure 6.1: The GUI of this mini visualization program provides basic image-handling features including zooming, panning, and picking points.

Although the experiment in the previous chapter showed that embedded manifold in 7-dimensionality best preserves the geometry of original image data, it is difficult to understand how points are organized in 7-dimensional Euclidean space. Hence we decided to visualize in 2-dimension. However we are aware that this may not completely represent most of the variability of our original dataset because we can only observe data under 2 degrees of freedom.

We firstly divided data into 4 set according to Miami classification of the colon i.e. normal, hyperplastic, adenoma, and adenocarcinoma and then projected each set of data separately onto 2-dimension space. Under this division, we expect to observe the relationship among data in 2-dimension space more clearly compared to the entire set of 1000 data.

6.1 Normal colon

Three hundred images which are clinically classified as normal colon are presented as 2-D points in figure 6.2. We can clearly see that the majority of data are located in the left buttom part of

the distribution. Along the X-axis, left images have fine granular texture and gradually become coarse with well-circumscribed round structures. From the clinical perspective, the fine granular pattern normally can be observed from the superficial layer of healthy colon tissue whereas the well-circumscribed round structure indicates deeper layer of normal colon. Along the Y-axis, contrast agent inside blood vessels is noticeable in the right buttom part of the plot and the vessel shape becomes harder to identify and gradually blend with the gray background when going upward.

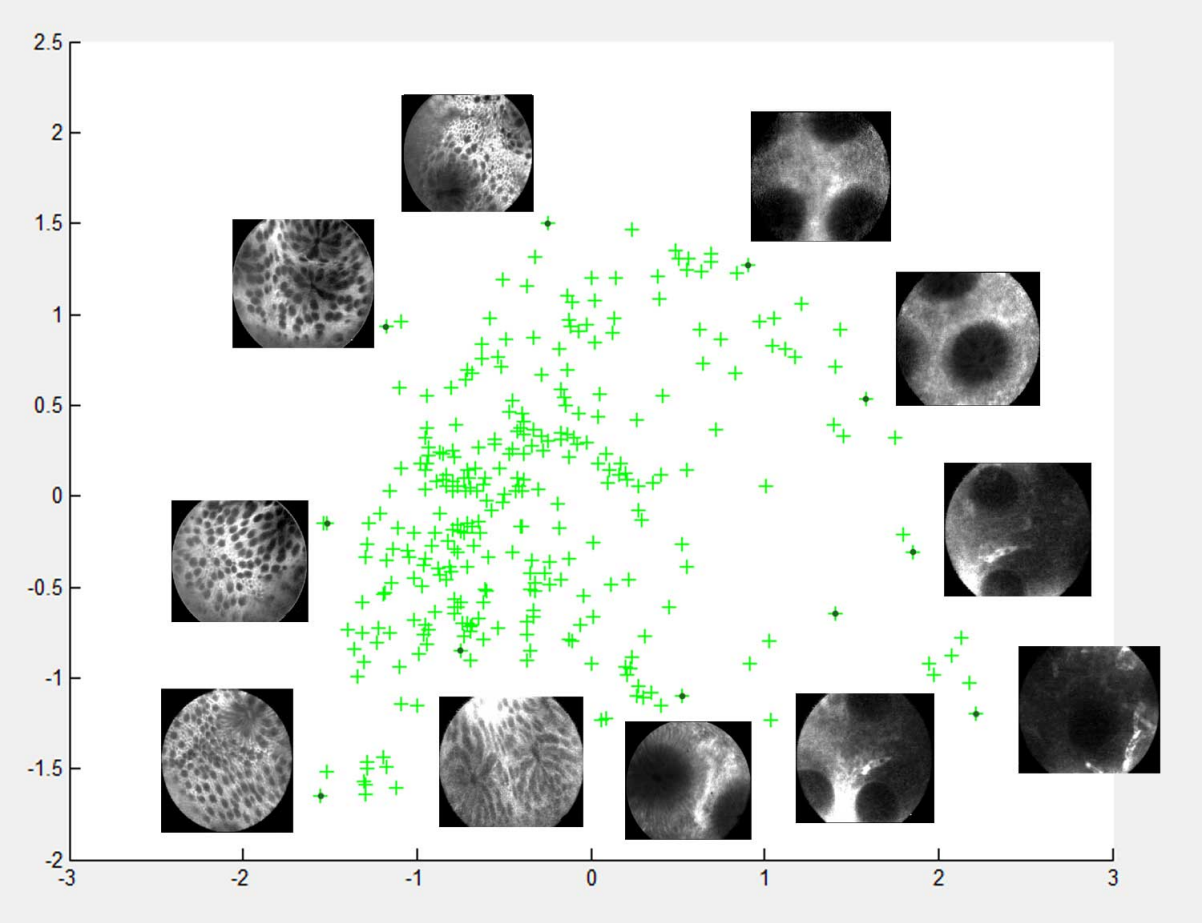


Figure 6.2: This plot demonstrates how 300 data points representing normal colon images organize in 2-dimension space after Isomap is applied. Along the X-axis,fine granular texture on the left side gradually become coarse with well-circumscribed round structures on the right side. Along the Y-axis, the shape of blood vessels filled with contrast agent can be easily identified at the buttom part and become blurred when going toward the top part of the plot.

6.2 Hyperplastic colon

Three hundred images of hyperplastic colon are represented by data points in figure 6.3. Along the X-axis, the distinctive point of left images is that the fine texture covers the majority of image area and there are no well-formed structures. However, dark thickening walls of colon structure are very obvious in the right-sided images and this enables us to easily identify the structure of the colon crypt. Along the Y-axis, the image contrast between the foreground and background are higher in the buttom area.

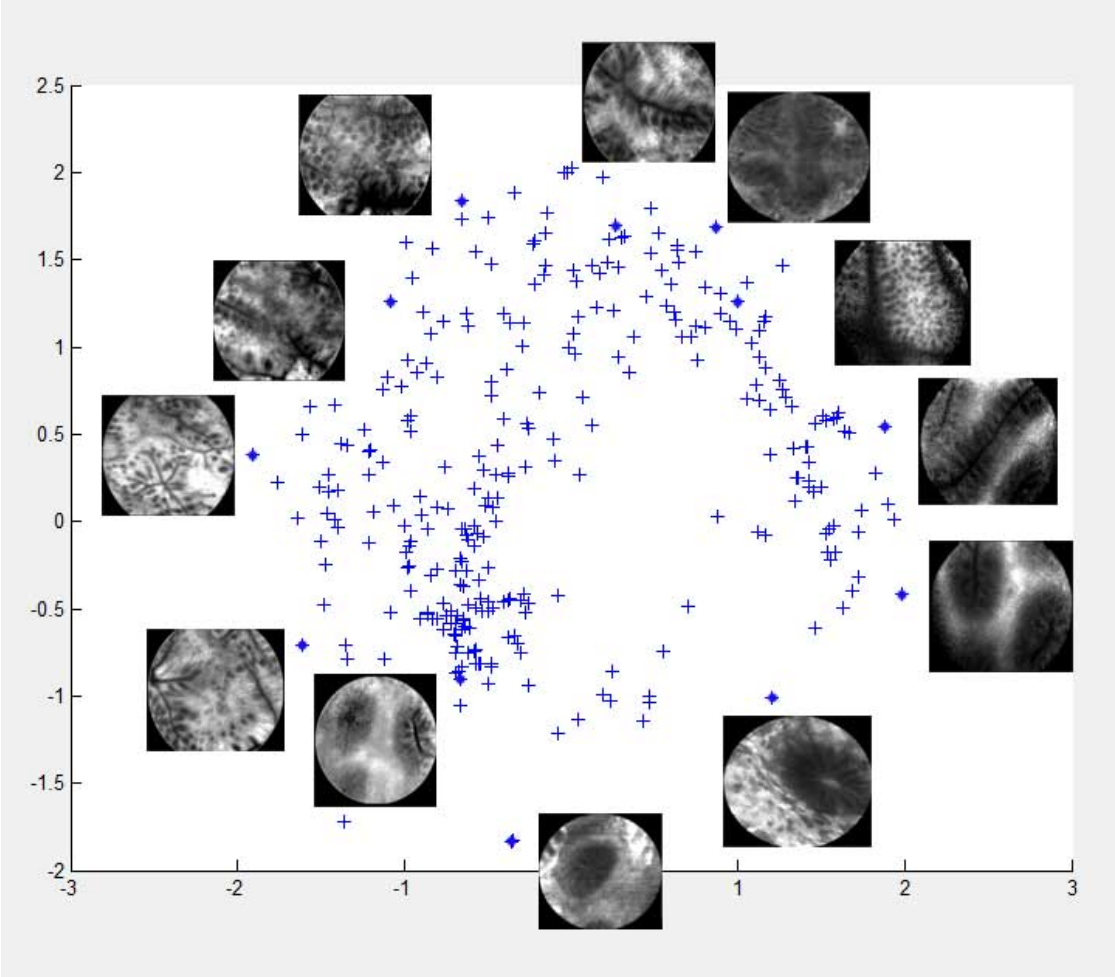


Figure 6.3: This plot demonstrates how 300 data points representing hyperplastic colon images organize in 2-dimension space after Isomap is applied. Along the X-axis,fine granular texture covers the majority of image area whereas dark thickening walls of colon structure are dominant in the right-sided images. Along the Y-axis, the image contrast between the foreground and background are higher in the buttom area compared to the top part.

6.3 Adenoma

Three hundred different images of adenoma classification are represented by data points in figure 6.4. We can notice that the density of data points is quite high in the right buttom part of the plot. On the right hand side, dark thickening structures are quite packed while images on the left side look looser. Due to the high density of the dark structures in right-sided images, the entire images become dark-looking. Again, along the Y-axis, the image contrast between the foreground and background are higher in the buttom area. The background of buttom images tend to be very bright compared to the top, which is gray. From the clinical perspective, very dark thickening structure is the key to characterize adenomas. Hence all of the images shown in the figure 6.4 contains this unique character.

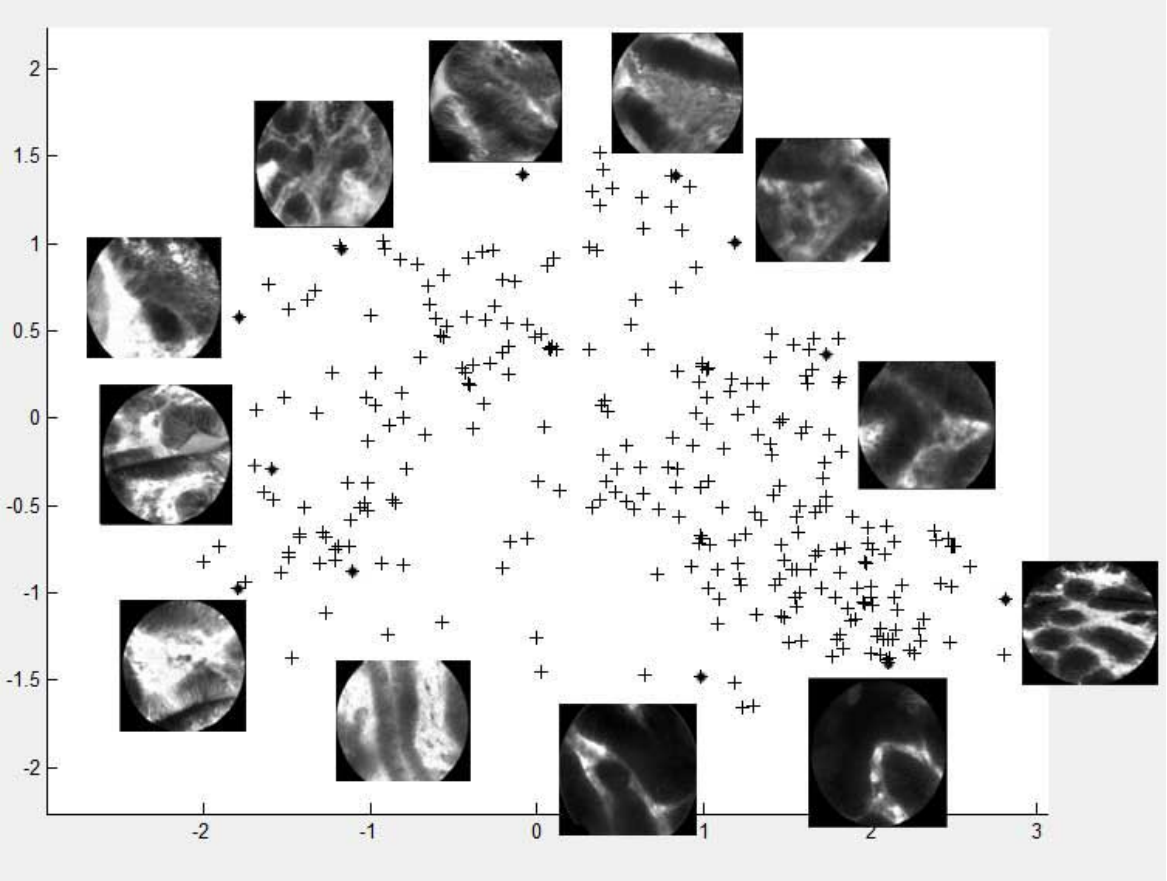


Figure 6.4: This plot represents 300 data points of adenoma images organize in 2-dimension space. On the right hand side, dark thickening structures look more packed than the left side ones. Along the Y-axis, the image contrast between the foreground and background tend to be higher in the buttom area.

6.4 Adenocarcinoma

There are 100 images of colon cancer included in this project. Hence only 100 data points are plotted and shown in figure 6.5. It can be seen that the distribution of the data can be divided into 2 poles, the left buttom and the right top. The left buttom images consist of disorganized black granules mixed with rod-like structures on very bright background whereas the right top contains chunky structures on gray background. From the clinical perspective, very bright background comes from the leakage of contrast agent from blood vessels and then disperse all over the colon tissue. Also, the disorganized structure and the lack of pattern are the key characteristics of cancer.

In summary, after we applied Isomap to our database with 1000 confocal images and reduced

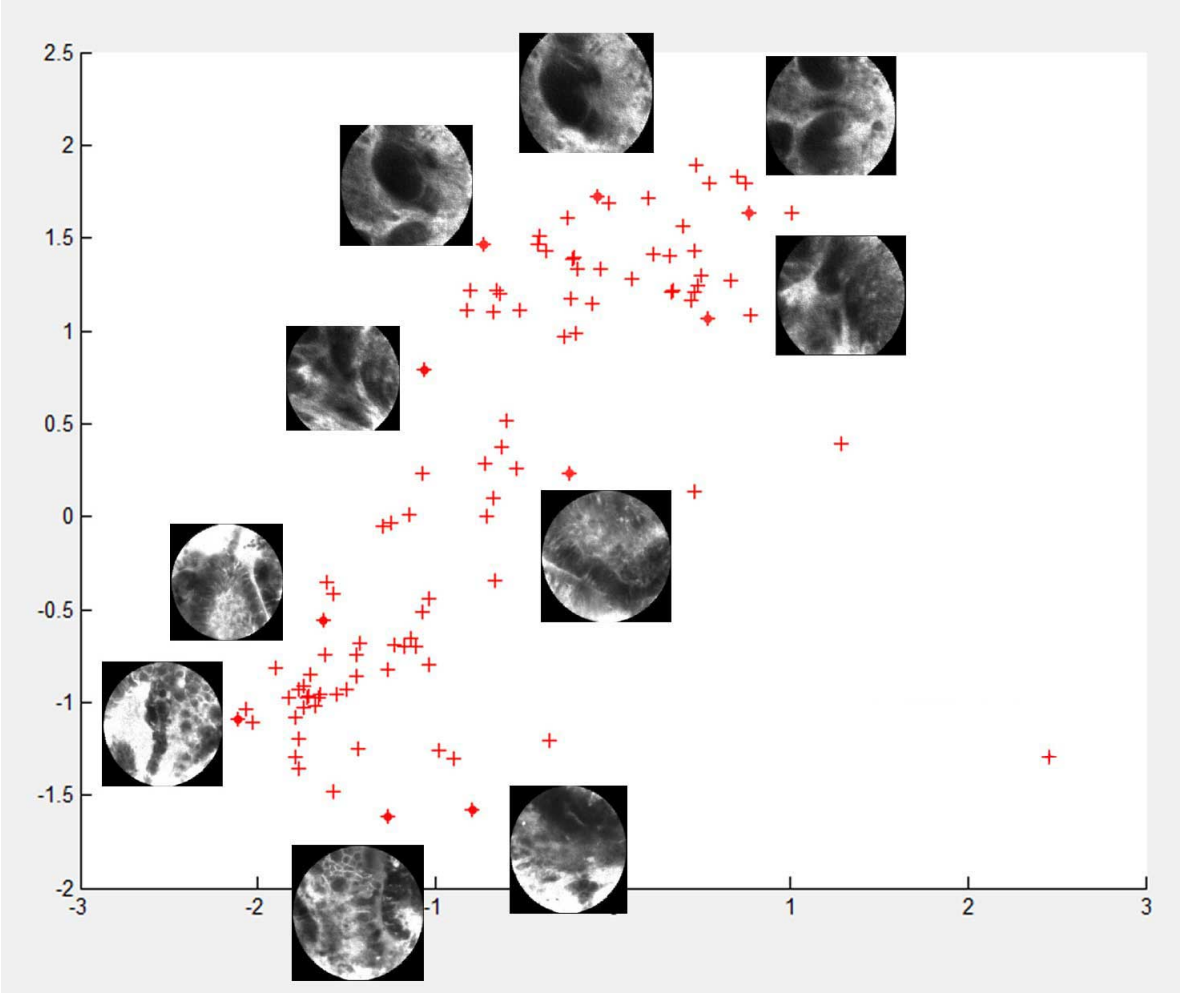


Figure 6.5: This plot represents 100 data points of adenocarcinoma images organize in 2-dimension space. There are 2 poles within the point distribution,the left buttom and the right top. The left buttom images consist of uneven black granules mixed with long thin structures on very bright background (Brightness comes from contrast agent.) whereas the right top contains chunky structures without any particular shape on top of gray background.

dimensionality to 2-dimension, we found that we could observe a gradual transformation of image pattern in terms of fineness and coarseness along the x-axis and also the different level of image contrast along the y-axis. However, this interpretation has its own limitation because 7-dimensionality is shown in this project that it can best preserve the intrinsic geometry of the original data, not 2-dimensionality.

Chapter 7

Discussion

7.1 Texture analysis

Although there are several kinds of low-level visual features , we consider that texture feature suits this project best.

Color feature apparently is not an appropriate choice for this project because confocal images of the colon are greyscale. Greyscale can be another option if we can control the environment when confocal images are captured. The brightness and color range are not only determined by the structure of colon tissue itself, but also depend on timing relative to the level of contrast agent in the blood stream i.e. An image of a colon structure appears overally brighter if it is captured at the early phase of contrast agent (High blood level of contrast) but the same structure can look overally darker if it is captured at the very late phase of contrast agent (Low blood level of contrast).

We consider that shape feature could be another suitable choice but microscopic structures of the colon can have a wide range of contours, even within the same classification. This could be a difficulty to identify what shape belongs to what classification in terms of shape.

In terms of texture signature, we decided to employ Gabor filters bank as it proves to be very useful texture analysis and is widely used in the literature. In this project, we experimented what values we should define for the constant U_l and U_h for our confocal images. Although $U_l = 0.05$ and $U_h = 0.4$ are ones of the most commonly used values in the literature to determine the range of frequency of Gabor wavelets in the filters bank, we found that signatures produced under these values appear indistinguishable among different images and we can hardly detect shift pattern in the energy map when images are rotated. Eventually we defined U_l and U_h as 0.05 and 0.15 respectively for this project. We also considered whether we should expand the frequency range by including more frequencies into our filters bank. However, the computation time turned to be very long and it seems impractical when Gabor filters bank is applied to user's example images.

In the aspect of texture distribution, we predefined the number of clusters for texture signatures as 8 to support the multi-textured property of the confocal images. This cluster number could be increased in order to capture more details. However, the computation time will also inevitably increase.

7.2 Texture distance

7.2.1 Earth Mover's distance

We consider Earth Mover's distance (EMD) is a suitable technique to calculate texture distance in this project due to several reasons including

- It has the ability to handle multi-textured queries
- The ground distance inside EMD adapts itself to the range of variation within the database
- EMD can easily handle image signatures with the format of distributions of descriptors.

7.2.2 Rotation- and scale-invariant features

Rotation invariant is very important and highly desirable to image retrieval because, in practice, we can neither control the environment to guarantee a zero rotation angle between images nor make an assumption that all the images are captured under the same orientations. In reality, no matter how textured images are rotated, they are still perceived as the same texture. Hence we integrated rotation invariant feature into our work. At the same time, scale invariant feature is included during the distance calculation in order to make our application less sensitive to the changes of scales of the texture pattern. According to our experiment, we found that the retrieval results improved when rotation- and scale-invariant features were included.

7.3 Technique with 3 small patches as the query image

As previously described, confocal images of the colon are likely to be non-homogeneous and consist of multiple textures within one image. Hence searching based on the entire example image could bring the problem of "wide semantic gap" because what the user focuses may be just certain parts of the example image. We experimented whether our patch approach can outperform the conventional method using entire image as a query by observing the retrieved results and calculating final retrieval rates. We found that our patch implementation improved the retrieval performance and outperformed the conventional method both overall and for each individual classification.

To increase the accuracy for the adenocarcinoma, we did many experiments by adjusting patch size and also built a new additional patch feature from which the user can customize the size. During our experiment on customized patch, we found that retrieval rate in certain test adenocarcinoma images can be further improved when very small patch is applied. This truly our idea that confocal images of the colon is non-homogeneous and consist of multiple textures within one image. We consider that adenocarcinoma is an extreme case which contains the highest heterogeneity degree in terms of textures. The result demonstrate that the more specific patch we can determine for adenocarcinoma, the more accurate the system can retrieve. Hence, we have added another option in the patch-size dropdown feature with a very small patch. We call this option as `zoom in++`.

7.4 Relevance feedback

In the relevance feedback process, we opted to implement dimensionality reduction using Isomap because we expect short response time with highly efficient retrieval behavior during this user-related process. We employed Isomap to create embedded manifold due to several reasons as below

- It clearly provides us with an information of residual variances to decide which dimensionality best preserves the intrinsic geometry of our original image data.
- In the pilot test with small database, we can easily observe 4 separate clusters, which represent 4 different classifications, being located in 2-D embedded space as shown in figure 5.5.
- It has been investigated and successfully applied in the field of medical image analysis before

During the dimensionality reduction step, we came across one major challenge of selecting the suitable Isomap parameter so that desirable embedded manifold can be achieved. According to Tenenbaum [29], the neighborhood size (The parameter `k` for k-nearest neighbours for neighbourhood graph construction) should be neither too large nor too small. We then tested different values from 4 to 9 and finally decided to use $k = 5$ and the optimal residual variance can be observed at 7-dimensionality. However, during our test, we did not notice any substantial differences between different k values. Moreover, we slightly take an advantage of our GUI to test whether similar images are mapped close by in 7-dimensional Euclidean space. So far, we found that the majority of similar images are plotted in vicinity to one another.

In addition to our work in 7-dimensional space, we also explored whether we can work on 2-dimension with additional approach. In the 7-dimension, our approach is to calculate similarity

based on Euclidean distance using L1-norm. However, in our experiment with 2-dimension, we asked the user to always select 3 satisfactory retrieved images in the relevance feedback process. Afterthat, we built an "imaginary triangle" (Figure 7.1) which is essentially created from 3 points in 2-dimensional Euclidean space and these 3 points are originally from the 3 retrieved images that the user selected in the relevance feedback process.

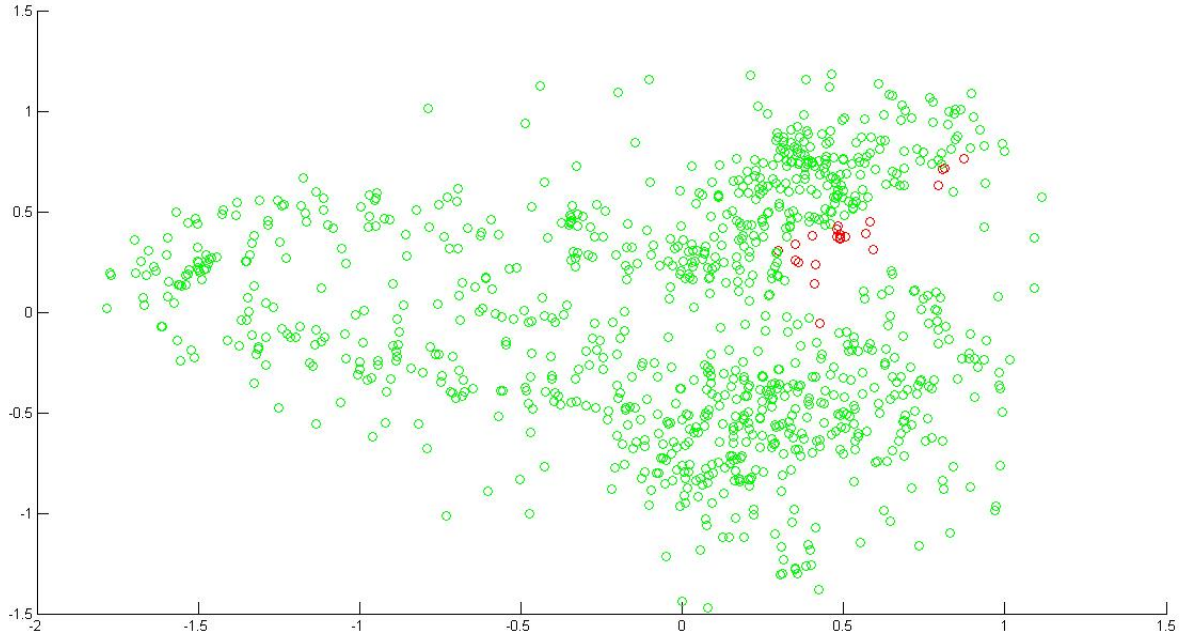


Figure 7.1: This figure explains the technique using our imaginary triangle in relevance feedback step and, in here, each small circle represents an image in our database. Basically, a triangle is an area in an embedded space and it can be formed and becomes "noticeable" according to 3 separate points in the 2-D space. The triangle represents one specific area containing several points inside, which are represented by red small circles here and it will come to play a role when the subsequent search is requested. The search will only work on the points located inside the area of the triangle. This makes the subsequent search very quick and also returns very close results to the previous images marked as satisfactory by the user.

Our assumption is that the user would expect to see the next retrieved results that contain similar properties to what they previously selected. Hence, the next set of retrieved results should be searched simply not far away from those 3 points. With this rationale, an imaginary triangle is constructed. Similarity calculation in the next retrieval can be done on a small set of points inside the triangle, not all the 1000 points. We also integrated an algorithm to help enlarging the triangle area in case that less than 9 points appear in the triangle. As far as we concern, this method would work very well if we can project our 1000 complex image data onto 2-dimensionality. This method can be reconsidered in the future with other dimensionality reduction techniques that has ability to reduce more dimensionalities i.e. t-Distributed Stochastic Neighbor Embedding or t-SNE.

In terms of retrieval outcome, our relevance feedback feature proves to enhance the system's retrieval performance as well as demonstrate very high efficiency during its subsequent search.

7.5 Limitation of the project

As fas as we concern, there are 2 major limitations in this project.

7.5.1 The variety of confocal images within database

The 1000 still images in our database have been extracted from

- 2 videos of adenocarcinoma (contribute to 100 still images)
- 23 videos of adenoma (300 still images)
- 14 videos of hyperplastic colon (300 still images)
- 11 videos of normal colon (300 still images)

From the resource we have, we tried our best to extract still images by avoiding repeated frames and images that look nearly identical. Each video contributes different numbers of still images depending on the its recording time. We cannot extract more than 100 still images out of 2 adenocarcinoma videos; otherwise, images will look repetitive. This problem also affected the amount of images that we can use as a test set in our experiments. We have no problem with 11 videos of normal colon because normal videos all have long recording time.

As far as we concern, wider range of confocal images of the colon is really important because our database will contain diversified confocal images and images that appears atypical can be excluded.

7.5.2 Computation time

Another challenge that still has room for improvement is how we can reduce computation time when database is processed and even more importantly, when the search operates. The computation is huge during image search because similar images are obtained from 1000 confocal images along with scale and rotation features. In our Gabor filters bank design, there are 4 different scales and 6 different rotations. Twenty-four shiftings are required when a pair of images are compared. A query, which is based on 3 small patches, then requires 3*1000*24 EMD's calculations in order to return the 9 most similar results. We actually tried to solve this problem by using MATLAB `parfor` to split works for different MATLAB workers expecting that computation time can substantially decrease with this paralell mechanism. However, we can only implement this MATLAB function to certain parts of our program and we still do not observe any significant improvement so far.

Multithreading programming apparently has potential to solve this problem. As MATLAB does not seem to support multithreading, we can consider using `MEX` function so `C++` codes can be combined with our current MATLAB code.

Chapter 8

Conclusion and future work

This chapter summarizes about our key achievements of this project as well as a brief description of possible future work.

Our major achievement of the project is the development of a computer-aided decision making system that has the ability to assist endoscopists to improve their differential diagnosis on the information of confocal laser endomicroscopic images of the colon. Our system is a content-based image retrieval framework which utilizes Gabor filters bank as the principal technique to analyze visual properties of confocal images of the colon. Gabor filters prove to be effective and able to capture distinct texture patterns of different classifications of confocal images of the colon. The distance calculation in our retrieval algorithm essentially operates on the Earth Mover's distance with rotation- and scale- invariant features which also shows to be effective. Experiments on a database of 1000 confocal images with mixed classification are carried out. As expected, confocal images with similar texture patterns can be successfully retrieved and better results can be observed when small patch technique is applied instead of taking the entire example image as a query. With the integration of relevance feedback technique, this framework also shows the ability to interact with the user and capably adapt itself towards user's interest. This feature shows to operate with high efficiency when interacting with the user. All of the abovementioned features and functionalities can be easily handled via our friendly-to-use GUI.

Another accomplishment of our project is the visualization tool that enables us to understand more on how texture characteristics of different confocal images transform from one to another. This tool essentially allows us to visualize high-dimensional image data by means of a simple two-dimensional map.

In our future work, there are many aspects we would like to explore and improve our computer-aided decision making system. One promising next step is to continue and extend our preliminary experiment on integrating an eye-tracking system into our project. This new combination will make the our system become "touchless" interactive CBIR and it will be very practical in real clinical practice because "hand-free" image search can be easily performed in surgical theatre without concern on sterility issue. Secondly, we would like to improve the search speed during the system's first search. This improvement has great potential to bring our system close to the real-world practice. Thirdly, we would like to add another functionality that facilitates the growth of database and subsequently augments its flexibility. "Out-of-Sample extensions for Isomap" will be considered as the principal solution for this challenge. Last but not least, our computer-aided decision making system should be evaluated quantitatively in the clinical setting how much endoscopists can use the system results to improve their overall performance of diagnosis based on information of confocal laser endomicroscopic images of the colon.

We have a great hope that our framework can be a good start and further development on this project will enable this computer-aided decision making system to be employed successfully as a useful tool for diagnostic examinations in daily clinical work in the near future.

Bibliography

[1] Kunio D. "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential"; Comput Med Imaging Graph 2007; 31(4-5): 198211

[2] Liedlgruber M., Uhl A., "Computer-aided Decision Support Systems for Endoscopy in the Gastrointestinal Tract: A Review", IEEE Reviews in Biomedical Engineering, 2011

[3] Andre B., Vercauteren T., "Endomicroscopic image retrieval and classification using invariant visual features", Proc. ISBI09, 2009, pp. 346349.

[4] Buchner AM, Gomez V, Kanwar R et al, "The Learning Curve for In Vivo Probe Based Confocal Laser Endomicroscopy (pCLE) for Prediction of Colorectal Neoplasia", Gastrointestinal Endoscopy, 2009 : Pages AB364-AB365

[5] Tomislav D, http://emedicine.medscape.com/article/277496-overview

[6] http://www.umm.edu/presentations/100158.htm

[7] WebMD Image Collection: Human Anatomy, http://www.webmd.com/digestive-disorders/picture-of-the-colon

[8] Grady WM, "Genetic testing for high-risk colon cancer patients", Gastroenterology 2003;124:1574-1594

[9] Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. "Genetic alterations during colorectal-tumor development", N Engl J Med 1988;319(9):525-32

[10] Chung DC, "The genetic basis of colorectal cancer: insights into critical pathways of tumorigenesis", Gastroenterology 2000;119:854-865

[11] Selby JV, Friedman GD, Quesenberry CP, Weiss NS, "A case control study of screening sigmoidoscopy and mortality from colorectal cancer", N Engl J Med 1992;326:653-657

[12] Winawer SJ, Zauber AG, Ho MN, OBrien MJ, Gottlieb LS, Sternberg SS, Waye JD, Schapiro M, Bond JH et al, "Prevention of colorectal cancer by colonoscopic polypectomy" N Engl J Med 1993;329:1977-1981

[13] Kiesslich R, Juergen B, Michael V et al, "Confocal Laser Endoscopy for Diagnosing Intraepithelial Neoplasias and Colorectal Cancer In Vivo", Gastroenterology 2004, 127:706-713

[14] JABBOUR JM , Saldua MA, et al, "Confocal Endomicroscopy: Instrumentation and Medical Applications", Annals of Biomedical Engineering 2012, Vol. 40, No. 2, 378397

[15] http://php.med.unsw.edu.au/embryology/

[16] Wallace M ,"Miami classification for probe-based confocal laser endomicroscopy", Endoscopy 2011

[17] Haralick R,"Statistical and structural approaches to texture", Proc. IEEE 1979, 67, 5, 786804

[18] Malik J, Perona P, "Preattentive texture discrimination with early vision mechanisms", J. Optical Soc. Amer. 1990, A 7, 5, 923932.

[19] Gabor D, "Theory of communication", The Journal of the Institute of Electrical Engineers, Part III, 93(21):429-457, January 1946

[20] Mao J, Jain A, "Texture classification and segmentation using multiresolution simultaneous autoregressive models", Pattern Recognition, 25:173-188, 1992

[21] J. D. Daugman, "Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression", IEEE ASSP, 36:1169-1179, 1988

[22] Manjunath B. S. and Ma W. Y. , "Texture features for browsing and retrieval of large image data IEEE Transactions on Pattern Analysis and Machine Intelligence 1996, (Special Issue on Digital Libraries), Vol. 18 (8), 837-842

[23] Alexander D. , "Rotation Invariant Texture Description using General Moment Invariants and Gabor Filters. In Proc. Of the 11th Scandinavian Conf. on Image Analysis Vol I. June 1999, 391-398

[24] Yossi R, Tomasi C, "Texture-Based Image Retrieval Without Segmentation" IEEE International Conference on Computer Vision (ICCV) 1999

[25] Hitchcock FL, "The distribution of a product from several sources to numerous localities" J. Math. Phys., 20:224230, 1941

[26] Shaw WM, "Term-relevance computations and perfect retrieval performance" Information Processing and Management

[27] Salton G, McGill MJ, "Intorduction to modern information retrieval", MCGraw-Hill book Company, 1983

[28] Rui Y, Huang TS et al, "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval", IEEE transaction on circiuts and systems for video technology 1998

[29] Tenenbaum JB, Silva V, Langford JC, "A global geometric framework for non-linear dimensionality reduction", Science 2000, Vol. 290, 2319-2323

[30] Fukunaga K, "Introduction to Statistical Pattern Recognition", Academic Press Professional, Inc., San Diego, CA, USA, 1990

[31] L.O. Jimenez LO, Landgrebe DA, "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data", IEEE Transactions on Systems, Man and Cybernetics, 28(1):3954, 1997

[32] Lim IS, Ciechomski PH, Sarni S, Thalmann D, "Planar arrangement of high-dimensional biomedical data sets by Isomap coordinates. In Proceedings of the 16th IEEE Symposium on Computer-Based Medical Systems, pages 50-55, 2003

[33] Samuel G, Tolga T et al, "On the Manifold Structure of the Space of Brain Images", Medical Image Computing 2009

[34] Zhang Q, Souvenir R, Pless R, "On manifold structure of cardiac MRI data: Application to segmentation", CVPR 2006, IEEE, 1092-1098

[35] Rohde G, Wang W, Peng T, Murphy R, "Deformation-based nonlinear dimension reduction: Applications to nuclear morphometry" (May 2008) 500-503

[36] Mardia KV, Kent JT, Bibby JM, "Multivariate Analysis", Academic Press, 1979

[37] Murase H, Nayar SK, Int. J. Comp. Vision 1995, 14;5

[38] McClurkin JW, Optican LM, Richmond BJ, Gawne TJ, Science 1991, 253; 675

[39] Tenenbaum JB, Silva V, Langford JC, "The Isomap algorithm and topological stability", Technical Comments Response in Science 2002, Vol.295

[40] D M McDonald and P Baluk, "Significance of Blood Vessel Leakiness in Cancer", Cancer Res 2002;62:5381-5385

Appendix A

Experiment results

A.1 An experiment on 20 testing confocal images

We conducted the first experiment on 20 different images as a test data set. There are 4 different classifications and each contributes 5 images.

Table A.1: Normal colon classification

Techniques + criteria	NM 1	NM 2	NM 3	NM 4	NM 5
Conventional + same class	5	0	7	6	4
Conventional + benign/neoplasm	9	9	8	9	9
Patch + same class	8	4	5	7	5
Patch + benign/neoplasm	9	7	8	9	9

Table A.2: Hyperplastic colon classification

Techniques + criteria	HP 1	HP 2	HP 3	HP 4	HP 5
Conventional + same class	4	8	0	8	4
Conventional + benign/neoplasm	7	9	6	9	6
Patch + same class	4	6	6	6	7
Patch + benign/neoplasm	7	9	9	6	7

Table A.3: Adenoma classification

Techniques + criteria	AD 1	AD 2	AD 3	AD 4	AD 5
Conventional + same class	5	0	7	9	9
Conventional + benign/neoplasm	5	0	7	9	9
Patch + same class	9	6	7	9	9
Patch + benign/neoplasm	9	6	8	9	9

A.2 An experiment on 5 testing adenocarcinoma images

We conducted the second experiment on 5 adenocarcinoma images as a test data set. It is the same set as one in the first experiment. However, this experiment was conducted in the different time to the first experiment.

Table A.4: Adenocarcinoma classification

Techniques + criteria	ADC 1	ADC 2	ADC 3	ADC 4	ADC 5
Conventional + same class	1	0	1	0	1
Conventional + benign/neoplasm	2	8	1	0	1
Patch + same class	3	0	3	1	3
Patch + benign/neoplasm	7	9	7	7	8

Table A.5: Adenocarcinoma classification: Before and after relevance feedback

	ADC 1		ADC 2		ADC 3		ADC 4		ADC 5	
Techniques + criteria	-RF	+RF	-RF	+RF	-RF	+RF	-RF	+RF	-RF	+RF
Patch + same class	2	8	0	0	2	6	1	4	1	4
Patch + benign/neoplasm	7	9	4	6	7	9	5	9	5	9