3D co-occurrence matrix based texture analysis applied to cervical cancer screening

Meng Liang

UPPSALA
UNIVERSITET

Abstract

3D co-occurrence matrix based texture analysis applied to cervical cancer screening

Meng Liang

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Cervical cancer is the second most common cancer in women worldwide, approximately 471,000 new cases are diagnosed each year. In 2005, there were about 500,000 cases of cervical cancer and 260,000 cases caused death in worldwide [1]. Cervical cancer starts as a precancerous condition, however the changes of precancerous are hardly detected by the naked eyes, special test such as Papanicolaou test are used to spot the conditions. These are time consuming to inspect visually. In the last 50 years there have been many projects to develop automated computer image analysis system for screening.

One of the most important changes in a cell when it becomes precancerous is a change in chromatin texture. The field of nuclear texture analysis gives information about the spatial arrangement of pixel gray levels in a digitized microscopic nuclei image. A well known method for quantifying textures in digital images is the gray level co-occurrence matrix(GLCM). This method tries to quantify specific pairwise gray level occurrence at specific relative positions. In this project, firstly we have developed and tested three image normalization methods : gradient based intensity normalization, histogram equalization and standardizing normal random variables; secondly we have developed 2D gray level co-occurrence matrix calculation and 3D gray level co-occurrence matrix calculation, thirdly compared Haralick features with adaptive feature vectors from class distance and class difference matrices (adaptive texture feature) based on the 2D gray level co-occurrence matrix; compared the Haralick features with adaptive feature based on the 3D gray level co-occurrence matrix. Our result shows that neither of the 3D results yields a significant improvement from 2D results.

Contents

1   Introduction

Cervical cancer [2] is the 2nd most common cancer in women worldwide, approximately 471,000 new cases are diagnosed in each year. 80% of cervical cancer deaths occur in the under resourced countries. Figure 1 shows worldwide incidence of cervical cancer per 100 000 females in all ages. In 2005, there were about 500,000 cases of cervical cancer and 260,000 deaths worldwide [1]. The rates of cervical cancer vary from 1 to 50 per 100,000 females [1]; rates are highest in Latin American and the Caribbean, Melanesia and South-East Africa. Most cervical cancer cases are diagnosed in women aged above 40 years. Countries with well organized programmes to detect and treat precancerous abnormalities and early stage of the cervical cancer can prevent about 80% of the cancers. However, effective screening programmes and follow up of women with abnormal screening tests are difficult to implement in low and middle-resource settings. Mortality rates are much higher in the developing world [3]. Cervical cancers starts on the surface of the
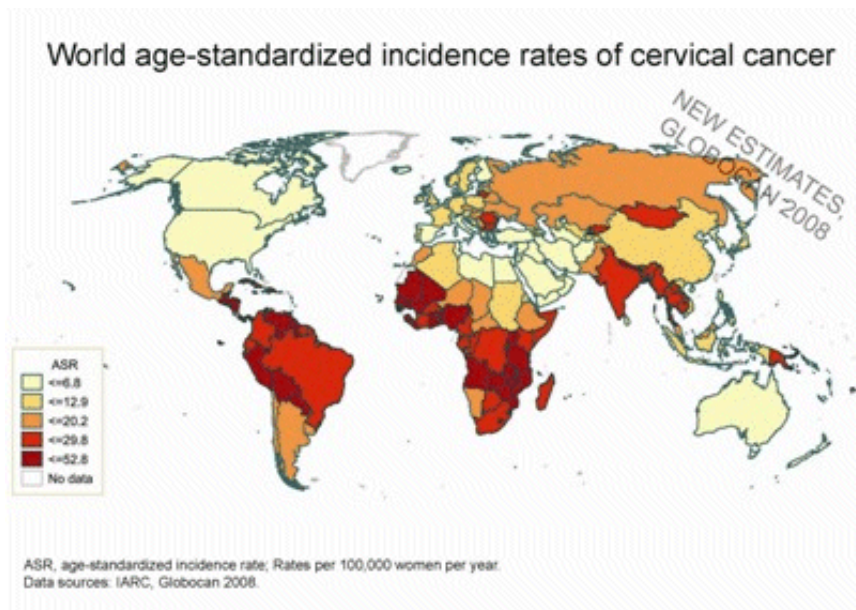
Figure 1: Worldwide incidence of cervical cancer per 100 000 females, age-standardized to the WHO standard population, 2005

uterine cervix in the cells. There are two types of cells on the cervix's surface which are squamous and columnar and most cervical cancers are from squamous cells. Usually, the cervical cancer are develops very slowly. It starts as a precancerous condition (dysplasia). This precancerous condition can be detected using a Papanicolaou (Pap) smear and is 100% treatable. It can take years for precancerous changes to turn into cervical cancer. Most women who are diagnosed with cervical cancer today have not had regular

Pap smears or they have not followed up on abnormal pap smear results. Almost all cervical cancers are caused by infection with oncogenic types of human papillomavirus (HPV). HPV is a common virus that is spread through sexual intercourse. There are many different types of HPV. Some strains lead to cervical cancer.

Precancerous changes of the uterine cervix and cervical cancer cannot be seen by the naked eye, special tests such as Pap smear test are used to spot conditions. After more then 50 years of research there are now some successful automated computer image analysis systems for screening. Since the mid 1950's, the several attempts to create an image analysis based automated system for cervical screening have been made [4]. It increase sensitivity and specificity of pap-smear screening, decrease the workload of cytotechnicians and cytopathologists, decrease the cost of the screening programs and decrease the incidence and mortality of cervical cancer. A key part of any automated screening system of cellular data is a robust cell description algorithm. The chromatin textures of the cell nucleus is the most important cell descriptor.

A well known method for quantifying textures in digital images is the gray-level co-occurrence matrix(GLCM). This method tries to quantify specific pairwise gray level occurrence at specific relative positions. This information is traditionally used to extract secondary statistical measures which describe the texture.

The purpose of this master thesis project is to implement, optimize and evaluate a number of extensions found in the literature to the standard gray level co-occurrence matrix (GLCM) method applied to the cellular data from the CerviScan project. This master thesis is being performed with in the CerviSCAN project which is aimed at developing a low cost cervical cancer screening, system for deployment, at least initially, in poorer regions around the globe. The CerviSCAN project started around 2007 and has since its start generated a large database of cell images acquired using a bright-field microscope. These images have also been expert annotated for algorithm development purpose. A specific goal is to extend the GLCM and its derivatives to operate in 3-dimension, this facilitate a pseudo-3D image acquisition that is currently being used in the project.

2 Background

2.1 The Papanicolaou test

The Papanicolaou test (Pap-test) [5] is a screening test used to detect potentially pre-cancerous and cancerous processes in the transformation zone of the female reproductive system. A standard Pap smear has a sample surface of at least 2×5 cm. For taking a Pap smear, a speculum is used to open the vaginal canal and allow the collection of cells from the outer opening

of the cervix of the uterus and the endocervix. Using the microscope to check abnormalities of the cells. This test aims to detect potentially precancerous changes (called cervical intraepithelial neoplasia (CIN) or cervical dysplasia), which are usually caused by sexually transmitted human papillomaviruses (HPV). The test is an effective, widely used method for early detection of pre-cancer, cervical cancer, detect infections and abnormalities in the endocervix and endometrium.

2.1.1 The Bethesda system

The Bethesda system [6] (TBS) is a system for reporting cervical or vaginal cytologic diagnoses, used for reporting Pap smear results. It was introduced in 1988, the name comes from the location Bethesda Maryland where the National Cancer Institute had their workshop.

Normal result is Negative for Intraepithelial Lesion or Malignancy (NILM). This means that the cells contained in the pap smear were normal when looked at under the microscope. Abnormal types of results:

- Atypical squamous cells

  - Atypical squamous cells of undetermined significance (ASC-US)
  - Atypical squamous cells – cannot exclude HSIL (ASC-H)

- Low grade squamous intraepithelial lesion (LGSIL or LSIL)

- High grade squamous intraepithelial lesion (HGSIL or HSIL)

- Squamous cell carcinoma (SCC)

- Atypical Glandular Cells not otherwise specified (AGC-NOS)

- Atypical Glandular Cells, suspicious for AIS or cancer (AGC-neoplastic)

- Adenocarcinoma in situ (AIS)

For our project, we only focus on Negative for Intraepithelial Lesion or Malignancy (NILM), Low grade squamous intraepithelial lesion (LSIL or LGSIL), High grade squamous intraepithelial lesion (HSIL or HGSIL) and Squamous cell carcinoma (SCC). LSIL indicates possible cervical dysplasia. The LSIL usually indicates mild dysplasia (CIN 1), more than likely caused by a human papillomavirus infection. HSIL indicates moderate or severe cervical intraepithelial neoplasia or carcinoma *in situ*. In some cases these lesions can lead to invasive cervical cancer, if not followed appropriately. SCC is a cancer of a kind of epithelial cell, it is a major forms of skin cancer. However, squamous cells also occur in cervix.

2.2 Automated Pap-smear analysis

Automation assisted screening [7] is aimed to increase sensitivity and specificity by finding atypical cells, both squamous and glandular cells. By excluding part of the slides from manual screening or by enriching the most atypical cells to images or to be studied by the microscope is designed to increase the performance of screening. By enhancing the effectiveness of the screening work, automation is allows more slides to be screened by the same staff. It would be an advantage, because in many countries there are severe shortage of cytotechnicians. As part of these automated devises are capable to process conventional and liquid based smears and it allows them to be used in different kinds of screening programmes.

The first idea about an automatic device for cancer cell screening appeared even before the digital computers were available in 1950s [8]. The machine was based on that a cancer cell could be distinguished from normal cell with nuclei size and optical density. The project of first fully automated microscope system was failed in the early sixties, because the system was not able to detect the artifacts in the sample.

Due to the bad reputation for cytology automation caused by expensive and early failure in US, the next couple of decades were moved to Europe and Japan. Another early project was the one one-parameter (nuclei size) automatic screener developed in Britain in the late sixties [9].

In Japan Watanabe and co-workers at Toshiba developed CYBEST [10]. Their first version used special purpose electronic circuit while later version was based on general purpose digital computers. They extracted nuclei area, nuclei density, cytoplasmic area and nuclei/cytoplasmic ratio from the nuclei image. The prototypes of CYBEST system were used in large field trials in the Japanese screening program and showed some promising results but it never became a product [11].

Since the first attempts at automation had limited success, much research on cell image analysis was carried out using interactive system around the world. Based on the progress made in this work there was a new wave of system designs in the early 1980s. System developed were BioPEPR [12], FAZYTAN [13], CERVISCAN [14], LEYTAS [15] and DIASCANNER [16]. In the middle 1980s, most of the systems reached an operational prototype stage. Some reported classification accuracies that were well within the range by the conventional visual screening. But none of them reached market, because of automated microscopes and computers with sufficient processing power were still too expensive causing lack of cost effectiveness.

In the late 1980s there was a great increase of interest in cytology automation in US and many new project are started [17]. Regarding to the new ways of specimens preparation technique and new computer display technology, the PAPNET system from Neuromedical Systems was the first to introduce interaction into automated-screening [18]. Another system Au-

toPap 300 from NeoPath was similar to PAPNET which used conventional Pap-smears and neural network classifier [19].

During the 1990s, there was a strong competition between the American companies developing screening technology as well as struggles to get solutions approved by Food and Drug Administration, FDA. No systems can be sold in US without FDA approval. The company Tripath, which merged NeoPath, Neuromedical and AutoCyte [20], calling Focal Point system, which has large extent based on AutoPap 300 [21]. the system is selling for routine use, so finally after 50 years research there was a computer supported system of Pap-smear screening on the market.

3 Materials and Methods

3.1 Microscope setup

Cells found on Pap-smears are not distributed as a flat 2D surface. Instead the thickness on the specimen can measure up to around 100 μm for areas with high amounts of debris and around 10-20 μm in more general fields. This fact creates problems when we are developing automated systems for diagnosing slides. In almost all image fields will exist nuclei that are not in the current focus layer. In order to handle that problem, we are registering focus stacks. We are using Olympus BX 51 optical microscope equipped with a E-662 Piezo server controller (Physik Instrumente GmbH & Co. KG, Karlsruhe, Germany) connected to a $40\times$, 0.95 NA objective for stack acquisition. Using this controller we can generate image stack which consists of 41 images with a 0.4 μm step length, thus covering 16 μm of material.

3.2 Dataset

In addition to a dataset with cervical cell nuclei images, we are using the Kylberg Texture dataset 1.0 [22] to test different image normalization approaches.

3.2.1 Kylberg Texture Dataset 1.0

The Kylberg Texture Dataset 1.0 contains 28 texture classes. Each class has 160 unique samples which has two versions: with rotated texture patches and without rotated texture patches. The format of the image are 8 bit gray scale PNG, texture patch size is $576 * 576$ pixel. Total size of dataset is 1.76 GB. Table 1 gives a short description of each class. For each class, a subset of 30 levels of additive Gaussian white noise [23] logarithmically spaced has been provided (see Figure 2 and Figure 3).

Class Name	Description
blanket1	Woven blanket.
blanket2	Woven blanket.
canvas1	Woven linen canvas.
ceiling1	Painted concrete ceiling.
ceiling2	Metal plate ceiling with small holes.
cushion1	Woven fabric on a cushion.
floor1	Plastic plate floor with stone imitation.
floor2	Linoleum floor with stone imitation.
grass1	grass from a lawn.
lentils1	Red lentils on a flat surface.
linesds1	Linseeds on a flat surface.
oatmeal1	Oatmeal on a flat surface.
pearlsugar1	Pearl sugar on a flat surface.
rice1	white rice on a flat surface.
rice2	Arborio rice on a flat surface.
rug1	Rya rug (long pile rug).
sand1	Sand surface.
scarf1	Woven scarf.
scarf2	Woven scart.
screen1	Woven fabric on a screen.
seat1	Woven fabric on chair.
seat2	Woven fabric on chair.
sesameseeds1	Sesame seeds on a flat surface.
stone1	Flat part of a granite base of a sculpture.
stone2	Flat part of stone base of a building.
stone3	Flat part of a stone wall.
stoneslab1	Stone slab from a building's entrance.
wall1	Painted lime plaster of an outdoor wall.

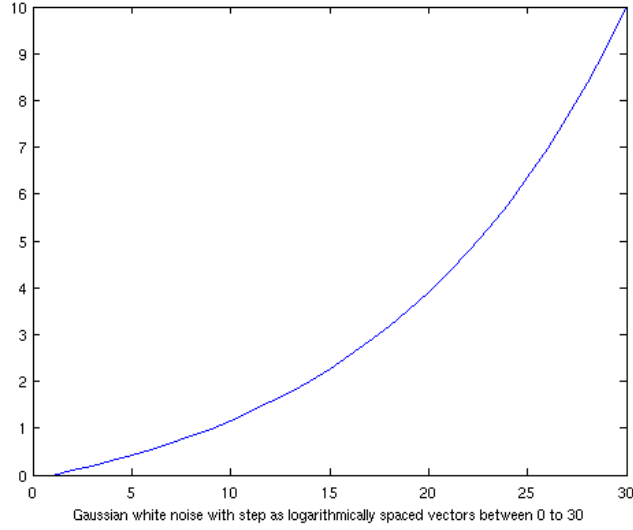**Table 1:** Names of texture classes and a short description of the material.
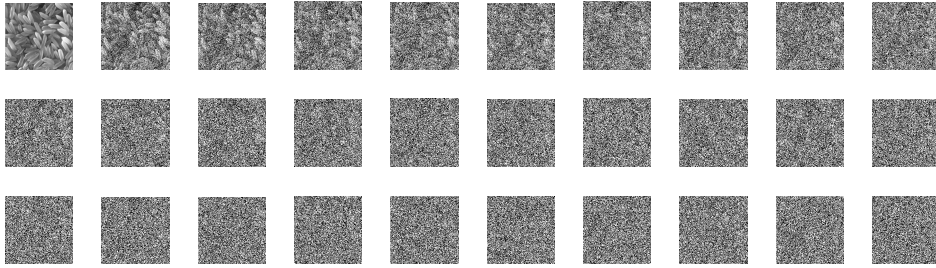
Figure 2: Gaussian white noise.

Figure 3: Images with gaussian white noise

3.2.2 Cervical cell nucleus dataset

Total number of NILM is 2000, LSIL is 215, HSIL is 568 and SCC is 744 (Table 2). Each cell nucleus has been segmented and imaged at 41 focus levels. For each cell nucleus volume, we have a cell diagnose and a sample diagnose, obtained from an expert cytologist. We also have a height-map and an extended depth of focus image (see chapter 4.3). For the focus image in Figure 4, we can clearly see the cell and its background, the gray area surrounding the cell. Having mask on the cell image we can easily separate cell from the background. As we can see in Figure 5 (c), some background pixels are around the boundary of the nuclei, which means the mask does not perfectly match the boundary of the nuclei. To solve this problem, we have applied mathematical morphology of erosion with a city-block metric to the mask (Figure 5 (b)). Then applied nuclei image with eroded mask (Figure 5 (d)). The height-map is used to identifying the focus layer; cell

diagnose and slide diagnose is identifying the cell class and slide class.

cervical cell nucleus dataset
NILM 2000
LSIL 215
HSIL 568
SCC 744

Table 2: cervical cell nucleus dataset.

(a) 1   (b) 15   (c) 20   (d) 30   (e) 41
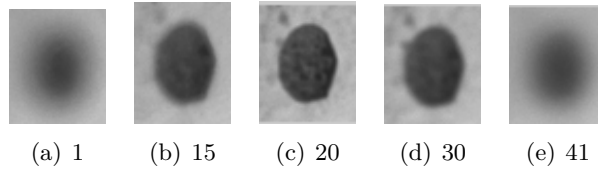
Figure 4: 5 images from an image stack. Image (c) depicts the cell nucleus in perfect focus in layer 20.
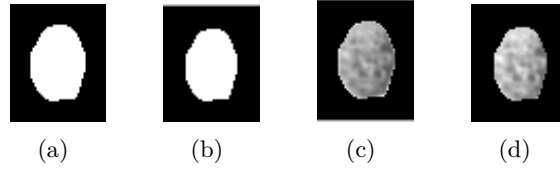

(a)   (b)   (c)   (d)

Figure 5: (a):Original mask; (b):Eroded mask; (c):Apply cell image with original mask; (d):Apply cell image with eroded mask.

3.3   Extended depth of focus

Microscopy image often suffers from the limited depth-of-focus: however, the specimen can be 'optically sectioned' by moving the object along the optical axis. Extended depth-of-focus [25] in a fusion algorithm that combines those images into a single sharp composite using wavelet transform. To extend the depth-of-focus, we should define an in-focus criterion. Typically, an focus image should have a maximum number of visible details. On the other hand, out of focus images are blurred by point-spread-function of the microscope. Therefore, we suppose that areas of an image which are focused contain more high frequency component than out-of-focus area. We are using complex Daubechies wavelet [26] [27] to locate high frequency. Taking the resulting image we can now create a height map by locating where in the focus stack we find the gray-value that is closest to the calculated one.

By taking the histogram of the height map, we can locate the optimal focus level. By combining the highest frequencies found at each pixel position in the focus stack er are able to create a single image where the object is perfectly focused. From the histogram in Figure 6 we can see that the focus position in the image stack is 16.
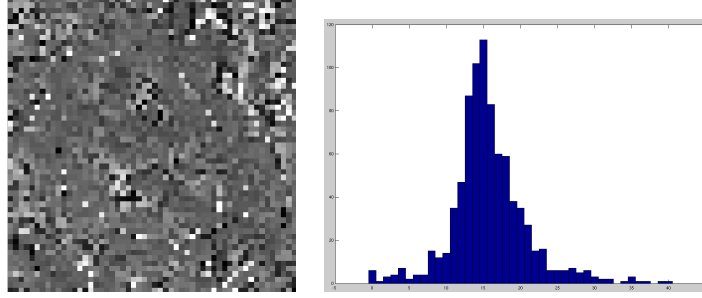

Figure 6: Height-map of the stack of cell images (left),histogram of the height-map (right).

3.4 Image normalization

In image processing, normalization is the process to change the range of the pixel intensity value, so that different images becomes more directly comparable. We have evaluated three normalization methods in our work.

3.4.1 Gradient based intensity normalization

Gradient intensity normalization [28] uses gray values as well as gradient information, and it is very robust to differences in object density. We have obtain a gradient image from the input image, then generate a bivariate histogram of image intensity and gradient. With each cell value b_{ig}, calculated the number of occurrences of image pixels with intensity i and gradient g. One-dimensional profile will be denoted by

$$\{P^X\}_i, i = 0, 1..., R \tag{1}$$

where X is the variant. Then we derived a profile from the bivariate histogram using average gradient (AG):

$$P_i^{AG} = \frac{\Sigma_{g=0}^{R-1} g b_{ig}}{\Sigma_{g=0}^{R-1} b_{ig}} \tag{2}$$

Cumulative sums of profile are calculated and rescaled to the range [0,1], it adjust saturation or other truncation effect. For our project, we do background correction (applying image with eroded mask) first, using ordinary histogram matching to generates a profile, which is summing bivariate histogram values over the gradient dimension.

3.4.2 Histogram equalization

Histogram equalization is a method used in the image processing of contrast adjustment. It increases the global contrast of many images, especially when usable data of the image is represented by close contrast value, the intensities can be better distributed on the histogram. A discrete gray scale image i, and n_x is the number of occurrences of gray level x. The probability of an occurrence of a pixel of level x in the image is

$$p_i(x) = p(i = x) = \frac{n_x}{n}, 0 \leq x < L \tag{3}$$

L is the total number of gray levels in the image, n is the total number of pixels in the image, and $p_i(x)$ is the histogram for pixel value x, normalized to [0,1]. *Cumulative distribution function* (CDF) corresponding to p_i is defined as

$$cdf_i(x) = \sum_{j=0}^{x} p_i(j) \tag{4}$$

A transformation of form $y = T(x)$ produces a new image k, such that its CDF will be linearised across value range,defined as

$$y = T(x) = cdf_x(x) \tag{5}$$

Applying the following simple transformation to map the values back to the original image

$$y' = y * (max(x) - min(x)) + min(x) \tag{6}$$

3.4.3 Standardizing normal random variables

Standardizing normal random variables is a property of normal distribution, because the normal distribution is a location-scale family, all random variables can be related to the standard normal. If X is normal with mean μ and variance σ^2 then

$$Z = \frac{X - \mu}{\sigma} \tag{7}$$

has mean zero and unit variance, this Z has standard normal distribution. Conversely, if Z is a standard normal random variable, it can construct another normal random variable with specific μ and variance σ^2:

$$X = \sigma Z + \mu \tag{8}$$

Because all the images in the cervical cancer cell nuclei are 8-bit gray level images, the gray level is between [0 255]. The mean value of the gray level is 127. So the mean value μ equals to 127. Because we have to set the gray level between [0 255], firstly we have to applied the image with the eroded mask to get an input image and then subtract with the mean value

from the input image.Afterwards we divide the input image by the standard deviation. Then checking the histogram of the image and finding out the standard deviation equals to 40 is the best choice to make the image gray scale value between [0 255].

3.5 Gray level co-occurrence matrix

Gray level co-occurrence matrix (GLCM) [29] is a second-order statistics methods, which is based on (local) information about gray levels in pair of pixels. The matrix defined over the image with distribution of co-occurring values of given offset. Let Q be an operator that defines the position of two pixels are relative (offset), and an image f, with L possible intensity levels. Let G be a matrix with element g_{ij} define number of times that pair of pixel with intensities z_i and z_j occur in f with specified position in Q $(1 \leq i, j \leq L)$. Mathematically we have a co-occurrence matrix $\mathbf{C}$ which defined over an $n \times m$ image I, offset $(\Delta x, \Delta y)$.

$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^{n} \sum_{q=1}^{m} \begin{cases} 1, & \text{if } (p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Table 3 contains the offset values that specify common values, given the inter pixel distance (IPD) D. As we can see in Figure 7, the centre pixel of a 3×3 matrix is pixel-of-interest.

Angle	Offset
0	[0 D]
45	[-D D]
90	[-D 0]
135	[-D -D]

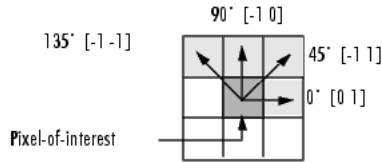**Table 3:** Offset value with specify distance D.

Figure 7: Illustrates the array of offset,offset = [0 1;-1 1;-1 0;-1 -1].

3.5.1 Gray level co-occurrence matrices for focal stack

A focal stack can be seen as a 3D volume data which contains 41 images with same size. For calculate GLCM in focal stack [30], we locate the focus

14

layer in stack first, calculate GLCM with specified offset, then we fetch one layer below the focus layer and one layer above the focus layer with equal Z-direction distance.

For these two pairs of layers: focus layer with one layer above the focus and focus layer with one layer below the focus, calculate GLCM with offset equals to Table 4.

Z-direction angle	Offset
90	[0 0 D]
270	[0 0-D]

Table 4: Offset value with specify distance D.

3.5.2   Haralick texture feature extraction

Haralick [31] described 14 statistic measurement that can be calculated from the co-occurrence matrix with the intent to describe the texture of the images. For our project we extracted four of those, contrast, correlation, energy and homogeneity from the cancer cells image.

The contrast measures the local variations in the GLCM, where i and j are pixels, K is the row (or column) dimension of square matrix, $p_{i,j}$ is probability of pixel pairs that satisfy the offset. When i and j are equal, so $(i - j) = 0$ and the cell is on the diagonal. The weight function $(i - j)^2 = 0$, represents pixels which are entirely similar to the neighbour. The weights continue to increase exponentially as $(i - j)$ increases. The function returns a measure of the intensity contrast between a pixel and its neighbour over the whole image. Range $= [0 \ (size(GLCM, 1) - 1)^2]$. If contrast equals to 0, it means the image is a constant. Contrast is also called *Inertia*.

$$Contrast = \sum_{i,j=0}^{K} P_{i,j}(i - j)^2 \tag{10}$$

The correlation texture measures the linear dependency of gray levels on neighbouring pixels. Returns a measure of how correlated the pixel is to its neighbour over the whole image. Where i and j are pixels, K is the row (or column) dimension of square matrix, $p_{i,j}$ is probability of pixel pairs that satisfy the offset. Range $= [-1 \ 1]$. The meaning to the actual calculated values: 0 is uncorrelated, 1 is perfectly correlated and NaN (Not a number) for a constant image.

$$Correlation = \sum_{i,j}^{K} \frac{(i - \mu_i)(j - \mu_j)p_{i,j}}{\sigma_i \sigma_j} \tag{11}$$

15

$$\mu_i = \sum_i^K iP(i) \tag{12}$$

$$\mu_j = \sum_j^K jP(j) \tag{13}$$

$$\sigma_i = \sum_i^K (i - \mu i)^2 P(i) \tag{14}$$

$$\sigma_j = \sum_j^K (j - \mu j)^2 P(j) \tag{15}$$

The energy function measures the homogeneity of an image, and provide the sum of squared elements in the GLCM.The homogeneous scene contains only a few gray levels, giving a GLCM with a few but relatively high values of $p(i,j)$. So the sum of squares will be high. Where $p_{i,j}$ is probability of pixel pairs that satisfy the offset. Range = [0 1]. Energy equals to 1 means the image is constant. It also called *Angular Second Moment.*

$$Energy = \sum_{i,j} p_{i,j}^2 \tag{16}$$

The homogeneity function returns a value that measures the closeness of the distribution of elements in the gray level co-occurrence matrix to the diagonal gray level co-occurrence matrix. Where i and j are pixels, K is the row (or column) dimension of square matrix, $p_{i,j}$ is probability of pixel pairs that satisfy the offset. Range = [0 1]. Homogeneity equals to 1 for a diagonal gray level co-occurrence matrix.

$$Homogeneity = \sum_{i,j=0} \frac{P_{i,j}}{1 + |(i - j)|} \tag{17}$$

3.5.3   Adaptive texture feature

By adaptive texture feature [32], it means the weighting functions adapt to the image material under study. Calculating adaptive texture feature vectors from class distance and class difference matrices extraction consists of the following two steps: (1) Computation of class distance and class difference matrix between two classes in the learning set. (2) Extraction of the adaptive texture features from each cell images.

For the given texture method, we first compute the probability matrix in each class ω_c of GLCM, $P_n(i,j|\omega_c)$, n = 1,2,...N(ω_c), then compute average matrix $\overline{P}(i,j|\omega_c)$ for each class ω_c, the class variance matrix $\sigma_p^2(i,j|\omega_c)$ and

the class different matrix $\Delta_p(i, j|\omega_1, \omega_2)$. Finally the Mahalanobis class distance matrix $J_p(i, j|\omega_1, \omega_2)$ between the two classes ω_1 and ω_2 is computed.

$$\overline{P}(i, j|\omega_c) = \frac{1}{N(\omega_c)} \sum_{n=1}^{N} P_n(i, j|\omega_c) \tag{18}$$

$$\sigma_p^2(i, j|\omega_c) = \frac{1}{N(\omega_c)} \sum_{n=1}^{N} (P_n(i, j|\omega_c) - \overline{P}(i, j|\omega_c)) \tag{19}$$

$$\Delta_p(i, j|\omega_1, \omega_2) = \overline{P}(i, j|\omega_1) - \overline{P}(i, j|\omega_2) \tag{20}$$

$$J_p(i, j|\omega_1, \omega_2) = 2\frac{(\overline{P}(i, j|\omega_1) - \overline{P}(i, j|\omega_2))^2}{\sigma_p^2(i, j|\omega_1) + \sigma_p^2(i, j|\omega_2)} \tag{21}$$

Figure 8 illustrates the computation of GLCM class distance and difference matrices form NILM and HSIL prognostic classes of Cervical Cancer Cell Nuclei dataset. The class distance matrix between a "good" and a "poor" prognosis class contains large areas of consistently high value (see Figure 8f). These area corresponding to separate areas of consistently positive or negative class difference values (see Figure 8e).
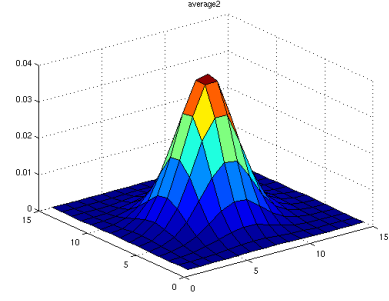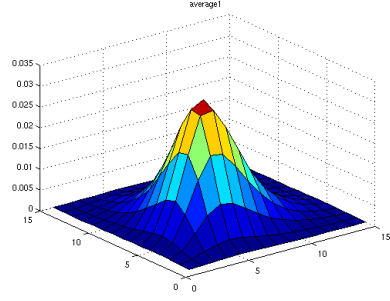
(a) average of GLCM of class ω_1,bad prognosis

(b) average of GLCM of class ω_2,good prognosis
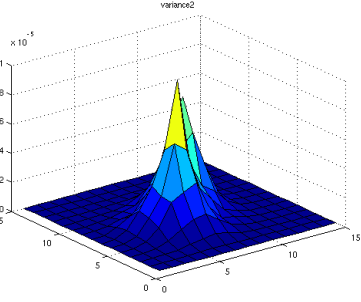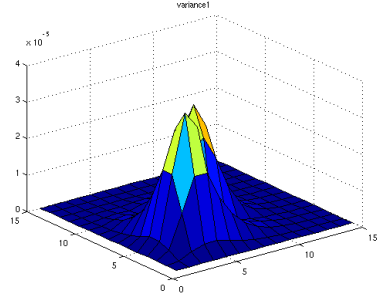
(c) variance of GLCM of class ω_1,bad prognosis

(d) variance of GLCM of class ω_2,good prognosis
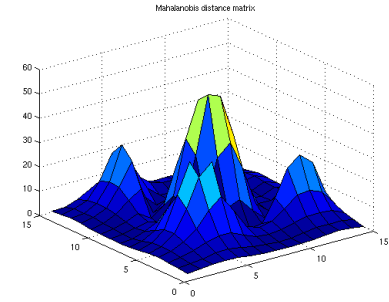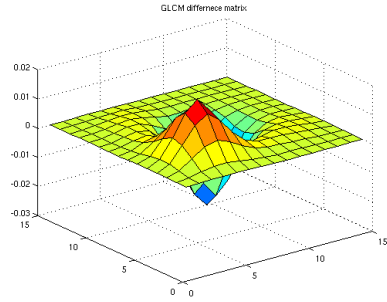
(e) GLCM class difference matrix

(f) GLCM Mahalanobis class distance matrix

Figure 8: On the basic of all training set images, we compute (a) (b) the average GLCM and (c) (d) the variance GLCM for each class. On the basic of these matrix, we then compute (e) difference matrix and (f) GLCM class distance. The class distance matrix contains information on which GLCM elements will provide high class discrimination.

18

For each texture image, we combine those GLCM elements that contribute the most to the class separability into two adaptive texture feature vectors from class distance and class difference matricess, using the squared class distance matrix values as summation weights. We use two disjoint positive and negative parts of the class difference matrix as the domains of the weighted summation (See Figure 9). While an image have a probability matrix $P_k(i,j)$ will give two adaptive texture feature vectors from class distance and class difference matricess values:

$$F_+ = \sum_{\Delta p(i,j|\omega_1,\omega_2) \geq 0} P_k(i,j|\omega_c)[J_p(i,j|\omega_1,\omega_2)]^2 \tag{22}$$

$$F_- = \sum_{\Delta p(i,j|\omega_1,\omega_2) < 0} P_k(i,j|\omega_c)[J_p(i,j|\omega_1,\omega_2)]^2 \tag{23}$$
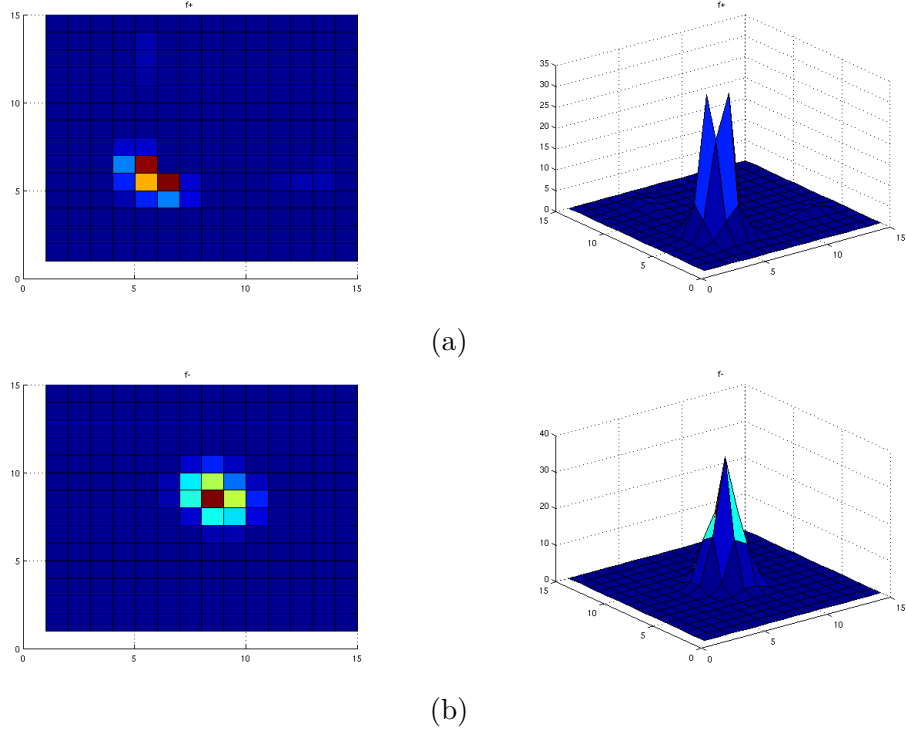

(a)

(b)

Figure 9: Weighting functions of two adaptive texture feature vectors from class distance and class difference matricess (a) F_+ and (b) F_-, computed from the training case of two different class of Kylberg texture.

19

3.6 Confusion matrices

A confusion matrix [33] is a specific table layout allowing to visualize the performance of the algorithm. It contains information about actual and predicted classifications. Each column of the matrix represents a predicted class, while each row represents the actual class. The Table 5 illustrate a confusion matrix which a is the number of the correct prediction with negative, b is the number of incorrect prediction with positive, c is the number of incorrect prediction with negative and d is the number of correct prediction with negative.

		Predicted class	
		Negative	Positive
Actual class	Negative	a	b
	Positive	c	d

Table 5: Confusion matrix.

3.7 Classification

Classification is used to assign corresponding levels with respect to homogeneous characteristics of groups, with the aim of discriminating multiple objects from each other. Image classification analyses the numerical properties of various image features into categories. Classification algorithm typically contains two phases: training and testing. In training phase, characteristic properties of image features are isolated, based on these, a unique description of each classification category is created. In the testing phase, using numerical properties of testing image to classify with classification category. For our project, we have used Mahalanobis distance to classify images. In statistics, Mahalanobis distance [34] is a distance measure based on correlation between variables by which different patterns can be identified and analysed. If we assume that all classes have the same covariance we get the discriminant function:

$$d_j(x) = -(x - m_j)^T C^{-1}(x - m_j) \tag{24}$$

where m_j is the mean of class w_j and C_j is the covariance matrix for class w_j. First step would be to find the average of the sample points. The closer the point to the center of the mass, the more likely it belongs to the set. However, we need to how the set is spread our, so we can decide the distance is significant or not. Estimate the standard deviation of the distances of the sample points from the center of mass would be the simplistic approach. We could conclude that it is highly probable that the test point belongs to the set if the distance between the test point and the center of mass is less than one standard deviation.

4 Evaluation of image normalization

The development and evaluation of different image normalization methods was perhaps the most time consuming part of this project. All three image normalization methods listed in Chapter 3.4 were exhaustively tested. Accordingly derived results are outlined below.

4.1 Gradient intensity normalization

Because no pre-written gradient based intensity normalization algorithm existed in MATLAB, the algorithm had to be implemented. For each nucleus image in the cervical cell nucleus dataset, we are using itself as reference image. Figure 10 is illustrate the result of taking 5 random images with background correction.
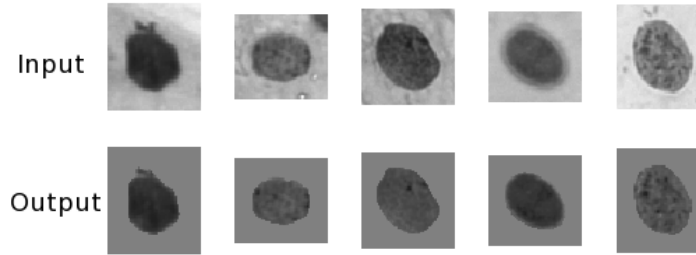

Figure 10: Result of gradient based intensity normalization with background correction, the first row of the image are original images, second row showing the result of gradient intensity normalization.

4.2 Histogram equalization

Because histogram equalization is such a common normalization method, MATLAB has implemented the algorithm for this method. The algorithm enhances the contrast of the images by transforming values in intensity image, so that the histogram of the output intensity image bins have the same number of pixels. We show the result in Figure 11.
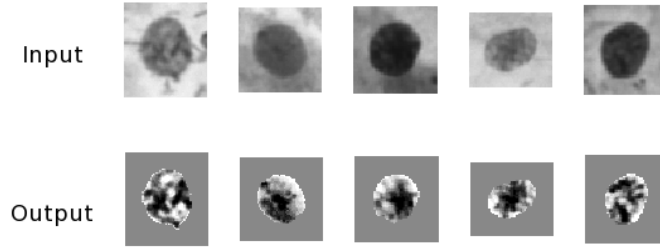
Figure 11: Result of histogram equalization,the first row contains 5 original images, the second row contains the result of histogram equalization.

4.3 Standardizing normal random variables

Because the image is normal random variable, so it only need to define the mean value μ (128) and variance σ^2 (40) for all the images. Applied eroded mask to the original image, showing the result in Figure 12.
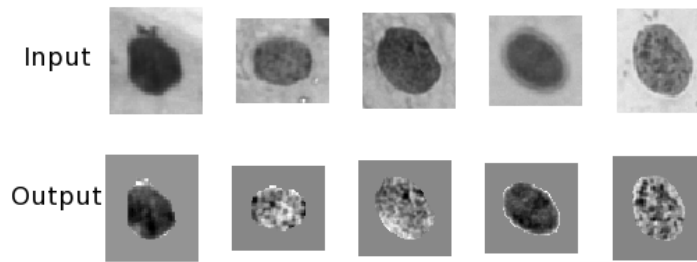

Figure 12: Standardizing normal random variables, the first row contains 5 original images, the second row contains the result of standardizing normal random variables.

4.4 Comparison between histogram equalization and standardizing normal random variables

It is hard to compare the result between histogram equalization and standardizing normal random variables visually. So we tested these two methods with the cervical cell nucleus dataset. Taking GLCM inter pixel distance (IPD) equals to 2. Using mahalanobis distance to classify different types of nuclei with Haralick features which we have mentioned in chapter 3.5.2. Table 6(a) shows the result of separating 4 classes with histogram equalization between each other, and mean accuracy is 55.17%. Table 6(b) shows the result of separating 4 classes with standardizing normal random variables between each other,and mean accuracy is 61.33%.

(a)

		Predicted class			
		NILM	LSIL	HSIL	SCC
	NILM	35	88	78	99
Actual class	LSIL	1	242	5	52
	HSIL	10	96	144	50
	SCC	1	53	52	194

(b)

		Predicted class			
		NILM	LSIL	HSIL	SCC
	NILM	97	93	49	61
Actual class	LSIL	0	201	51	48
	HSIL	56	52	147	45
	SCC	31	53	55	161

Table 6: Classification result between histogram equalization (a) and standardizing normal random variables (b).

5 Comparison between Haralick feature and adaptive texture feature for 2D data

5.1 Kylberg dataset

We are using Kylberg texture without rotated patches, it contains 30 subset. We are randomly taking 190 images from each texture classes, using 90 images for training set, the rest 100 images for testing set. We have chosen an IPD equals to 5, and the number of gray levels in each nuclei image was reduced by re-quantization to G = 16 to the accumulation of the matrix. We have compared the classification result between Haralick feature and the

adaptive texture feature.

For Haralick feature, as we have mentioned in 3.5.2, we are extracting Contrast, Correction, Energy and Homogeneity from the images. Firstly, we are looking for the best single feature from 30 sub-set. For non-noise level images, the average of classification accuracy of learning data was 89.93% (contrast), 91.42% (correlation), 90.20% (energy) and 89.47% (homogeneity). So correlation is the best single feature.
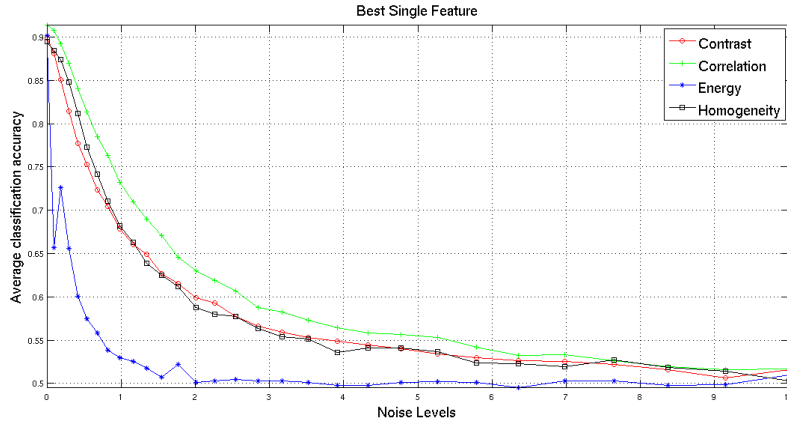
Figure 13: Best single feature extraction between 30 noise levels, solid line with circle performs contrast result, solid line with plus sign performs correction result, solid line with asterisk performs energy result and solid line with square performs homogeneity result.

Secondly, we are looking for the best pair-wise combination of features. For non-noise level images, the average of classification accuracy of learning data was 89.91% (correlation with contrast), 91.28% (correlation with energy), 91.25% (correlation with homogeneity).
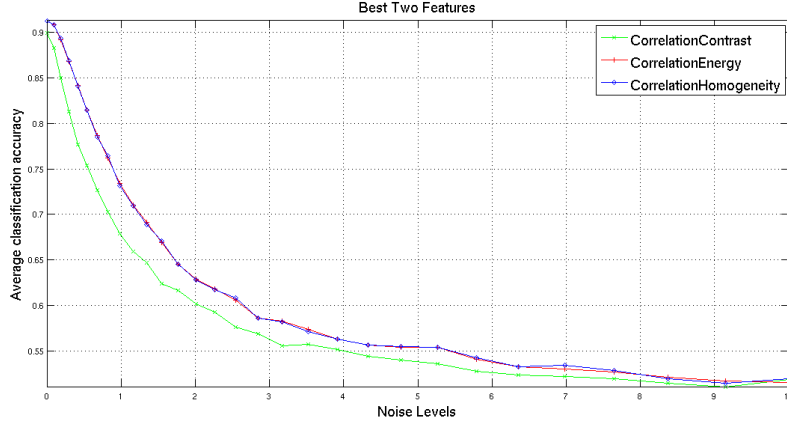
Figure 14: Best two feature extraction between 30 noise levels, solid line with plus sign performs correlation with energy result, solid line with asterisk performs correlation with contrast result and solid line with circle performs correlation with homogeneity.

For adaptive texture feature, as we mentioned in chapter 4.5.3, at non-noise level, the average of classification accuracy of learning data was 99.24%.
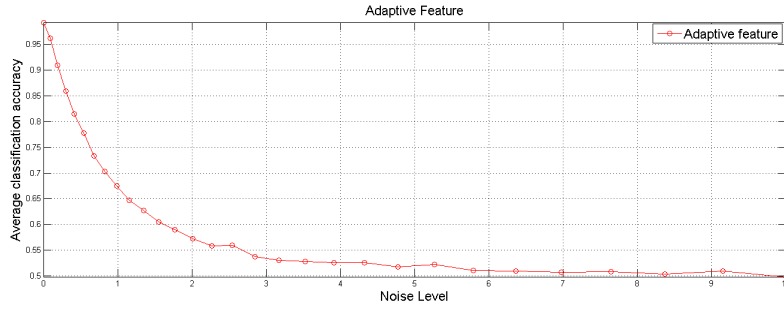

Figure 15: adaptive texture feature extraction between 30 noise levels

Figure 16 illustrate the classification result comparison between Haralick feature and adaptive texture feature. The solid line with circle shows the classification results when using four Haralick's features, solid line with plus sign shows the classification results when using adaptive texture feature. For non-noise level, Haralick feature average classification accuracy is 99.46%, adaptive texture feature is 99.24%, start from noise level equals to 0.2988, Haralick feature average classification accuracy is 88.65%, adaptive texture feature average classification accuracy is 85.88%, Haralick feature classification result is much better than adaptive texture feature. For a noise level of 10, Haralick feature average classification accuracy is 50.63%, adaptive texture feature average classification accuracy is 49.84%.
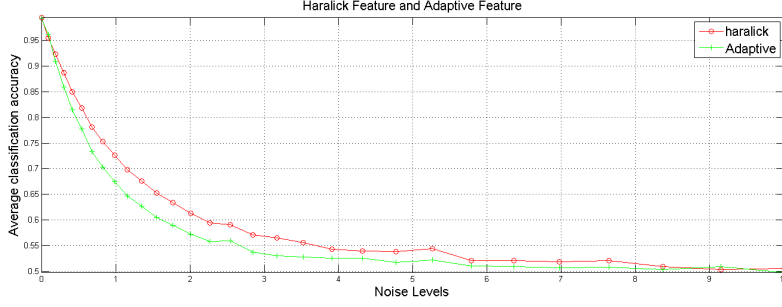
Figure 16: Haralick feature and adaptive texture feature extraction between 30 noise levels.

5.2   Cervical cell nucleus dataset

According to chapter 3.2.2, we know NILM have 2000 images, LSIL has 215 images, HSIL has 568 images and SCC has 744 images. We are randomly taking 190 images of NILM, LSIL, HSIL and SCC from cervical cell nucleus dataset, using 90 images for training set, the rest 100 images for testing set. For 2D approaches, firstly, we are optimized IPD of GLCM from 1 to 3, secondly, the number of gray levels in each nuclei image was reduced by re-quantization to $G = 16$ to the accumulation of the matrix, extract feature from GLCM, then we are using Mahalanobis distance to classify the image. Classification using Haralick features, we are running application 10 times to get 10 classification result for each IPD values. For IPD equals to 1, the mean classification result between four types of nucleus is 49.599%, mean classification result between NILM and others is 58.100%. For IPD equals to 2, the mean classification result between four types of nucleus is 49.941%, the mean classification result between NILM and others is 55.791%. For IPD equals to 3, the mean classification result between four types of nucleus is 49.475%, the mean classification result between NILM and others is 55.268%. Table 7 illustrate the best result of IPD equals to 2.

(a)

		Predicted class			
		NILM	LSIL	HSIL	SCC
Actual class	NILM	84	68	96	52
	LSIL	10	181	54	55
	HSIL	1	49	226	24
	SCC	40	48	74	138

(b)

		Predicted class	
		NILM	OTHER
Actual class	NILM	84	216
	Other	51	849

Table 7: Haralick features classification result of IPD equals to 2 between four types of nucleus (a) and NILM with other (b).

Classification using adaptive texture feature, we have 10 classification results for each IPD values. For IPD equals to 1, the mean classification result between four types of nucleus is 65.425%, the mean classification result between NILM and others is 81.431%. For IPD equals to 2, the mean classification result between four types of nucleus is 64.293%, the mean classification result between NILM and others is 80.122%. For IPD equals to 3, the mean classification result between four types of nucleus is 63.501%, the mean classification result between NILM and others is 81.028%. The IPD equals to 1 giving the best classification result between four types of nucleus. Table 8 illustrate the best result of IPD equals to 1.

		Predicted class			
		NILM	LSIL	HSIL	SCC
Actual class	NILM	208	30	29	33
	LSIL	35	195	39	31
	HSIL	24	36	186	54
	SCC	23	27	32	218

(b)

		Predicted class	
		NILM	OTHER
Actual class	NILM	208	92
	Other	82	818

Table 8: Adaptive texture feature classification result of IPD equals to 1 between four types of nucleus (a) and NILM with others (b).

6 Comparison between Haralick feature and adaptive texture feature for 3D data using cervical cell nucleus dataset

According to chapter 4.2.2, we are randomly taking 190 images of NILM, LSIL, HSIL and SCC from cervical cell dataset, using 90 images for training set, the rest 100 images for testing set. For 3D approaches, we first locate the focus layer in the volume. According to chapter 4.2.2, IPD equals to 2 for Haralick feature and IPD equals to 1 for adaptive feature, G equals to 16 gray levels. Then adding one layer above and one layer below the focus layer in Z direction with same distance from 1 to 10. After all, calculated average matrix of 3 layers, extract feature from GLCM, than using Mahalanobis distance to classify the image. We are running application 10 times to get 10 classification result for each Z direction distance. For Haralick feature, Z direction distance equals to 1 performs the best result for mean classification result between four types of nucleus is 50.492%, the mean classification result between NILM and others is 60.662% and Z direction distance equals to 7 performs best result for adaptive texture feature of mean classification result between four types of nucleus is 66.349%, the mean classification result between NILM and others is 82.404%. . Table 9 illustrate the best classification result between four types of nucleus for Z direction distance with Haralick feature and adaptive feature.

(a) Mean classification result is 50.492%

Haralick feature classification result
0.4925
0.5017
0.4975
0.5108
0.5092
0.5025
0.5158
0.4942
0.5042
0.5208

(b) Mean classification result is 66.349%

Adaptive feature classification result
0.6683
0.6750
0.6742
0.6525
0.6608
0.6558
0.6608
0.6617
0.6700
0.6558

Table 9: Classification result of 3D GLCM with IPD equals to 2 and Z direction distance equals to 1 of Haralick feature (a), Classification result of 3D GLCM with IPD equals to 1 and Z direction distance equals to 7 of adaptive texture feature (b).

6.1 Comparison of 2D and 3D classification result between Haralick feature and adaptive texture feature

According to the result we have, Table 10 illustrate Haralick feature and adaptive texture feature classification result between 2D and 3D.

We performed a t-test of null hypothesis for data in the first column and second column of Table 10 (a). There are independent random samples from normal distribution with equal means, alternative hypothesis is using the means are not equal. We got the p-value equals to 0.2127, meaning that the mean value of 3D result from Haralick feature is not significantly different from 2D result.

Then we performed a t-test of null hypothesis for data in the first column and

(a) Haralick feature		(b) Adaptive feature	
2D	3D	2D	3D
0.5242	0.4925	0.6925	0.6683
0.4917	0.5017	0.7017	0.6750
0.5058	0.4975	0.6933	0.6742
0.5050	0.5108	0.6808	0.6525
0.4933	0.5092	0.6833	0.6608
0.4700	0.5025	0.6967	0.6558
0.5208	0.5158	0.6717	0.6608
0.4858	0.4942	0.6767	0.6617
0.5008	0.5042	0.7092	0.6700
0.4967	0.5208	0.6850	0.6558
mean value	mean value	mean value	mean value
49.941%	50.492%	68.909%	66.349%

Table 10: 2D and 3D classification result with Haralick feature (a), 2D and 3D classification result with adaptive feature (b).

second column of Table 10 (b). We got the p-value equals to $1.9590*10^{-5}$, meaning that the mean value of 2D result from adaptive texture feature is significantly better from 3D result.

7 Discussion

7.1 Evaluation of image normalization

According to Chapter 4, three normalization methods are evaluated. Gradient based intensity normalization with ordinary histogram matching performs poorly (Figure 10). All five of the intensity histograms are different from each other, because ordinary intensity histogram is effected by both intensity levels and image content. From the result of histogram equalization (Figure 11) and standardizing normal random variables (Figure 12), we know the contrast of images are enhanced, according to the result in Chapter 5.1.4, standardizing normal random variables is the best normalization method for Cervical Cell Nuclei dataset.

7.2 Evaluation of 2D and 3D GLCM approaches

As we showed in Chapter 5 and 6, we have tested Haralick feature and adaptive texture feature using Kylberg Texture Database first. When noise level equals to 0, the classification result of these two feature extraction methods are performing equally well, Figure 16 (99.46% for Haralick feature, 99.24% for adaptive texture feature), while, adding Gaussian white noise with step as logarithmically spaced vector, Haralick feature performs better than the adaptive texture feature. When Gaussian white noise level equals to 10, two feature extraction methods perform bad (50.63% for Haralick feature, 49.84% for adaptive texture feature). This means that adaptive texture feature is more sensitive to the noise, and Haralick feature is more robust when extracting texture feature from GLCM.
2D GLCM with adaptive feature classification result have significant difference from 3D in cervical cell nucleus dataset. When we calculate the 3D GLCM, we adding the 7th layer above and below the focus layer, which are the blurring image, it is adding the redundancy for calculate class difference matrix.

7.3 Future improvement

For the future improvement, we could find the best two layers. Maybe the different distance in vertical direction layer can have significant difference from 2D GLCM. If we could have the volumetric data, we can calculated the GLCM with 26 directions, which should be more accurate and we could find the best four texture features from Haralick feature.

8   Acknowledgement

9   Reference

References

[1] Human papillomavirus vaccines. WHO position paper. Wkly Epidemiol Rec. 2009; 84 :118–31.

[2] Bosch, F. Xavier and Manos, M. Michele and Muñoz, Nubia and Sherman, Mark and Jansen, Angela M. and Peto, Julian and Schiffman, Mark H. and Moreno, Victor and Kurman, Robert and Shan, Keerti V. and International Biological Study on Cervical Cancer (IBSCC) Study Group, *"Prevalence of Human Papillomavirus in Cervical Cancer: a Worldwide Perspective."*. Journal of Journal of the National Cancer Institute, pp. 769 - 802, 1995.

[3] WHO, *"Weekly epidemiological record."*. No. 15, 2009, 84, 117 - 132.

[4] Ewert Bengtsson *"Recongnizing signs of malignancy - the quest for computer assisted cancer screening and diagnosis systems."*. 2010 IEEE International Conference, December 2010.

[5] Shield, P.W., Daunter, B. and Wright, R. G. , *"The Pap Smear Revisited."*. Australian and New Zealand Journal of Obstetrics and Gynaecology, 27: 269 - 283, 1987.

[6] Solomon D and Davey D and Kurman R and et al, *"The 2001 bethesda system: Terminology for reporting results of cervical cytology."*. JAMA: The Journal of the American Medical Association, vol. 287, no. 16, pp. 2114 - 2119, 2002.

[7] Koss, L. G., *"Cervical (pap) smear: New directions. Cancer."*. 71: 1406–1412. doi: 10.1002/cncr.2820710405, 1993.

[8] Mellor RC, Silver R, *"A microfluorometric scanner for the differential detection of cells: application to exfoliate cytology."*. Science 114: 356-360, 1951.

[9] Spriggs AI, Diamond RA, Meyer EW, *"A New Principle for Automated Screening of Cervical Smears."*. Lancet 1:359, 1968.

[10] Watanabe S and the CYBEST group, *"An Automated Apparatus for Cancer Prescreening: CYBEST."*. Comp Graph Image Proc 3:350-358, 1974.

[11] Tanaka N, Ueno T, Ikeda H, Ishikawa Am Yamauchi K, Okamoto Y, Hosoi S, *"CYBEST Model 4: Automated cytologic screening system for uterine cancer utilizing image analysis processing."*. Analytical and Quantitative Cytology and Histology 21(1): 79-84, 1977 .

[12] Zahniser DJ, *"Automation of Pap Smear Analysis: A review and States Report."*. In Haralick, RM (ed.): Pictorial Data Analysis. Springer, 1983.

[13] Erhardt R, reinhardt E, Schlipf W, Bloss WH, *"FAZYTAN: A System for Fast Automated Cell Segmentation, Cell Image Analysis and Feature Extraction Based on TV-image Pickup and Parallel processing."*. Analyt Quant Cytol 2:25-40, 1980.

[14] Tucker JH, Husaun OAN, *"Trials with the CERVISCAN Experimental prescreening Device on Polylysine-prepared Slides"*. Analyt Quant Cytol 3:117-120, 1981.

[15] ploem JS, van Driel-Kulker AMJ, Verwoerd NP, *"LEYTAS - A Cytology Screening System Using the New Modular Image Analysis Computer (MIAC) from Leitz."*. In Burger G, Ploem JS, Goerttler K (eds.): Clinical Cytometry and Hostometry: 24-35, Academic Press, London, 1987.

[16] Nordin B, *"The Development of an Automated Prescreener for the Early Detection of Cervical Cancer."*. Algorithms and Implementation. PhD Thesis, Uppsala univ.155 p, 1989.

[17] Koss LG, *"The Papanicolaou Test for Cervical Cancer Detection - A Triumph and a Tragedy."*. JAMA 261:737-743, 1989.

[18] DeCresce RP & Lifshitz MS, *"PAPNET Cytological Screening System."*. lab Medicine 22:276-280, 1991.

[19] Lee J, Nelson A, Wilbur DC, Patten SF, *"The Development of an Automated Papanicolaou Smear Screening System."*. Cancer, 81:332-336, 1998.

[20] Howell LP, Davis RL, Belk TI, Agdigos R, Lowe J, *"The AutoCyte preparation system for gynaecologic cytology."*. Acta Cytol., 42:171-177, 1998.

[21] Tripath website: http://www.tripathimaging.com/

[22] Gustaf Kylberg, *"The Kylberg Texture Dataset v. 1.0."*. Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, Uppsala, Sweden 2011.

[23] Marmarelis, V. Z. *"Nonlinear Dynamic Modeling of Physiological Systems, Appendix II: Gaussian White Noise."*. John Wiley and Sons, Inc., Hoboken, NJ, USA. 2004.

[24] Milan Sonka, Vaclav Hlavac, Roger Boyle *"Image Processing, Analysis, and Machine Vision."*. Thomson-Engineering ©2007, ISBN:049508252X

[25] Forster, B. and Van De Ville, D. and Berent, J. and Sage, D. and Unser, M. , *"Extended Depth-of-Focus for Multi-Channel Microscopy Images: A Complex Wavelet Approach."*. Proceedings of the Second *IEEE* International Symposium on Biomedical Imaging: From Nano to Macro *ISBI'04* Apr 2004, pp.660-663.

[26] W.Lawton, *"Extended Depth-of-Focus for Multi-Channel Microscopy Images: A Complex Wavelet Approach.". IEEE Transactions on Signal Processing*, IEEE, Vol.41, no. 12, pp. 3566-3568,1993

[27] J.-M. Lina and M. mayrabd, *"Complex Daubechies Wavelet"*. Applied and Computational Harmonic Analysis, Academic Press, Vol. 2, p. 219-229, 1995.

[28] SINTORN, I.-M., BISCHOF, L., JACKWAY, P., HAGGARTY, S. and BUCKLEY, M. , *"Gradient based intensity normalization."*, Journal of Microscopy 240: 249 - 258. doi: 10.1111/j.1365-2818.2010.03415.x, 2010.

[29] rafel C. Gonzalez, Richard E. Woods, *"Digital Image Processing."*, Pearson Education, Inc. Upper Saddle River, new Jersey 07458. Third Edition, pp.830-836.

[30] Arati S. Kurani, Dong-Hui Xu, Jacob Furst, Daniela Stan Raicu *"Co-occurrence matrices for volumetric data."*. Intelligent Multimedia Processing Laboratory, School of Computer Science, Telecommunications, and Information Systems, DePaul University, Chicago, Illinois, 60604, USA

[31] Robert M. Haralick, *"Statistical and structural approaches to texture."*, Proc. IEEE, vol. 67, no. 5, pp. 786 - 804, 1979.

[32] Nielsen, B. and Albregtsen, F. and Danielsen, H.E., *"Low dimensional adaptive texture feature vectors from class distance and class difference matrices."*, Medical Imaging, IEEE Transactions on 23: 73-84, 2004.

[33] Stephen V. Stehman, *"Selecting and interpreting measures of thematic classification accuracy."*, Journal of Remote Sensing of Environment, pp. 77 - 89, 1997.

[34] R. De Maesschalck and D. Jouan-Rimbaud and D.L. Massart, *"The Mahalanobis distance."*, Journal of Chemometrics and Intelligent Laboratory Systems, pp. 1 - 18, 2000.