EFFECT OF IMAGE VARIATION ON COMPUTER AIDED DETECTION SYSTEMS

SEYEDEH PARISA RABBANI

**KTH Technology
and Health**

Master of Science Thesis in Medical Engineering

Stockholm 2013

This master thesis project was performed in collaboration with
Diagnostic Image Analysis Group (DIAG), Radboud University
Medical Center, Nijmegen, the Netherlands
Supervisor at DIAG: Prof. B. van Ginneken

Effect of image variation on computer aided detection systems

Betydelsen av normalisering av bilder vid datorstödd bildanalys

SEYEDEH PARISA RABBANI

Master of Science Thesis in Medical Engineering
Advanced level (second cycle), 30 credits
Supervisor at KTH: Lars Gösta Hellström
Examiner: Martin Lindahl
School of Technology and Health
TRITA-STH. EX 2013:76

Royal Institute of Technology
KTH STH
SE-141 86 Flemingsberg, Sweden
<http://www.kth.se/sth>

Abstract

Computer Aided Detection (CAD) systems are expecting to gain significant importance in terms of reducing the work load of radiologists and enabling the large screening programs. A large share of CAD systems are based on learning from examples, to enables the decision making between the images with or without disease. Images are simplified to numerical descriptors (features vectors) and the system is trained with these features. The common practical problem with CAD systems is training the system with a data from a specific source and testing it on a data from a different source; the variations between sources usually affect the CAD system function. The possible solutions for this problem are (1) normalizing images to make them look more equal, (2) choosing less variation sensitive features and (3) modifying the classifier so that it classifies the data from different sources more accurately.

In this project the effect of image variations on the developed CAD system on chest radio graphs for Tuberculosis is studied at Diagnostic Image Analysis Group. Tuberculosis is one of the major healthcare problems in some parts of the world (1.3 million deaths in 2007) [1]. Although the system has a great performance on the train and test data from the same source, using different sub dataset for training and testing the system does not lead to the same result. To limit the effect of image variation of the CAD systems three different approaches are applied for normalizing the images: (1) Simple normalization, (2) local normalization and (3) multi band local normalization. All three approaches enhance the performance of the system in case of various sub datasets for training and testing purposes. According to the improvement achieved by applying normalization it is suggested as a solution for the stated problem above.

Although the outcome of this study has satisfactory result, there is always room for further investigations and studies; in specific testing different approaches for finding less variation sensitive features and modifying the classification procedure to a more variation tolerant process.

Acknowledgement

I wish to thank my supervisors, Prof. Bram van Ginneken at Diagnostic Image Analysis Group and Lars Gösta Hellström at KTH for continuous support during this work. I would like to thank CAD-for-TB project members, Laurens Hogeweg, Rick Philipsen and especially Pragnya Maduskar who has been my friend and mentor. This thesis would not be possible without the help of Anna Hellberg Gustafsson, Erasmus Coordinator at KTH who gave me the opportunity to become an exchange student at Radboud University.

I owe my deepest gratitude to my parents, family and friends who have always been there for me and supported me spiritually.

I would like to thank Medical Engineering master program coordinator, Mats Nilsson for his constant guidance and assistance during my studies at KTH.

In the end, thank you to every individual who helped me during my master thesis.

Table of Contents

1. Introduction	1
1.1 Computer Aided Diagnosis	1
1.2 Tuberculosis.....	1
1.3 Computer Aided Diagnosis for Tuberculosis	2
1.4 Problem definition	2
1.5 Purpose	2
1.6 Overview of the Project	3
2. Method and Materials	4
2.1 Materials	4
2.2. Simple Normalization.....	4
2.3. Local Normalization	5
2.4. Multi Band Local Normalization	5
3. Theory	10
3.1. Lung segmentation	10
3.1.1. Matching.....	10
3.1.2. Pixel classification (PC):	10
3.1.3. Features:.....	10
3.1.4. Classifier and feature transformation:	11
3.2. Lung Segmentation Configurations in this project:	12
3.3. Post Processing:.....	13
3.3.1. Gaussian Blurring:	13
3.3.2. Thresholding	13
3.3.3. Connected component labeling.....	13
3.3.4. Morphology	13
3.3.5. The hole filling.....	14
3.4. Evaluation:.....	14
3.5. Base line for the experiments:.....	15
4. Results.....	18
4.1 Simple Normalization.....	18
4.2 Local Normalization	20
4.3 Multi Band Local Normalization	25
5. Summary and overall discussion	33
6. Conclusion	36

7. Future works	37
8. Bibliography	38
Appendix I	40
Appendix II	43

1. Introduction

1.1 Computer Aided Diagnosis

The Computer Aided Diagnosis (CAD) can be defined as taking into account the computer analysis in a diagnosis procedure which is one of the broad concepts in artificial intelligence field [2]. The first attempt to apply computer diagnosis was made in 1960s, while the large scale investigation on CAD systems started in early 1980s with a fundamental change in the concept of computer output, from automated computer diagnosis to computer-aided diagnosis. CAD systems are gaining popularity nowadays both for reducing the workload of the radiologist and enabling the large screening program. They already became a part of clinical routines for breast cancer diagnosis in some hospitals and institutions in the United States. The current investigation for CAD on medical images is mostly concerned with three organs: the chest, breast and colon. [3]

Generally, CAD systems are trained by different examples, numerical descriptors, so that they learn how to make decisions. This process is accompanied by the feature vectors including the numerical descriptors and trained classifier systems. Once have the system trained it can be run on the test images, which are also defined by their corresponding feature vectors. Different pre and post-processing procedures could be applied on the CAD system to make it optimum. In the following paragraphs information about detection of Tuberculosis (TB), one of the world wide health problems and how it is subjected to CAD system can be found.

1.2 Tuberculosis

TB is one of the serious health problems in large area of the world, (1.3 million deaths in 2007) [1]. One of the problems of treating TB is that it does not have cheap and fast diagnosis process. The person with TB can transmit it by coughing to other people, so it is very important to be detected in a short while; none the less time is a key factor for the patient to start the treatment. The most common test to find an active TB is culture test on coughed-up sputum. This test has four main drawbacks: (1) it takes 4-6 weeks for the sputum to be ready to be decided whether it is infected with the bacteria or not (2) it needs a especial facilities which make the test to be expensive and not available everywhere (3) sputum should be studied under the microscope, the professional and skilled person is needed to handle the procedure, (4) there is an increasing rate for smear negative tests at the places with the high HIV incidences and (5) for the child is not easy to cough-up the sputum.

In addition to culture test, the chest radio graph is also common for the TB diagnosis. Although it was not recommended by World Health Organization in 60s due to the misinterpreting the chest radiographs by humans, it is now widely use for TB detection mostly because of the high rate of smear negatives. It has several advantages such as: (1) it is easy to access (2) it is low cost (3) it is ready in a second after the radiograph is

taken and (4) even in remote areas the acquired data can be sent to central locations to be read. The major problem with the chest radiographs as an option for detection of TB is that it requires an expert to interpret the chest radiograph; to deal with this problem making use of a CAD system is so advantageous.

1.3 Computer Aided Diagnosis for Tuberculosis

A CAD system for the detection of Tuberculosis (TB) is developed at the Diagnostic Image Analysis Group (DIAG). The CAD4TB system has the ability to detect from the posterior anterior chest radiographs the patients at high risk of having TB. CAD4TB has two major components: lung segmentation and texture analysis. In this project only the lung segmentation component of the CAD system is applied and the CAD4TB pipeline is implemented. This system has the ability to determine the potential patients at risk of having TB; however, it suffers from a serious problem which is common among CAD systems:

1.4 Problem definition

A major practical problem of CAD systems is: training the CAD system with data from one source gives the best output just for the images from the same source. In order to have the best results, one needs to always train the CAD system for each dataset accordingly; since the differences in the images from different sources distinctly affect the functionality of the CAD system. Not all images acquired under the exact same conditions which make them to vary case by case; there are several reasons for the variation between different images such as using different scanners, different scanner settings and also patient body size have influence on the image quality. Having these datasets coming from different sources, it is not feasible to control all these sources of variations. One possible solution is to train the system in separate sequences for different datasets but it greatly undermines the practicality of the CAD system and it affects its overall performance.

There are several approaches to be considered for handling this inevitable problem, and to make CAD system less sensitive to image variation. Each of the following procedures could provide less insensitive CAD system: First, the images could be normalized so that they become more similar. Second procedure could be done by implementing the invariant features. The third and final possible procedure is to modify the classification process in order to have the correct classification for images of different sources.

1.5 Purpose

The main purpose in this work is to treat the image variation problem by conducting a normalization method as a preprocessing step which enables us to train the CAD system with data from one dataset and run it successfully on another dataset from a different source. In other word, the ultimate goal is normalizing the images of different sources in a way that CAD could see and treat them as equals.

1.6 Overview of the Project

The three different normalization methods to tackle the defined problem are introduced in chapter 2 including: simple normalization, local normalization and multi band local normalization, a short description about the CAD for TB is mentioned in chapter 3, the result of normalization methods and the output of the CAD system on normalized images are shown and discussed in chapter 4. The overall conclusion on this work and the suggested future works can be found in chapter 5 and 6 respectively.

2. Method and Materials

As it has been mentioned in the previous chapter, image variation imposes practical problem for CAD system. A possible solution to limit or remove the dissimilarity between images from different sources is normalization. To achieve this goal number of different experiments were performed including implementation of normalization in different possible ways. Here three different methods for normalizing images are employed namely, simple normalization, the local normalization and multi band local normalization. The first two methods were suggested directly by Prof. van Ginneken, the third method was proposed by the author of this work which of course was not possible without the valuable guidance and the inspiration from Prof. van Ginneken and other members of the CAD4TB group. In the following sections of this chapter first the materials used during this project and then the three normalization methods are explained and reported in details.

2.1 Materials

Among different sites of collecting the chest radiograph for CAD4TB project, in this work frontal chest radiographs from six different sources are used. The reason that these specific images are chosen for this work is that they were annotated and could be applied as testing or training data to the CAD system. The name of the sites and the number of images from each site are:

- 1- Kanyama (100 images), note that whenever CAD system is implemented these images are implemented for training purpose.
- 2- Tilburg (100 images)
- 3- JSRT (20 images)
- 4- Gambia (20 images)
- 5- Radboud University Medical Center Nijmegen (20 images)
- 6- London (20 images)

The CAD for TB pipeline is the working platform in this project which is written in C++. The proposed methods are implemented as a preprocessing step prior to running the CAD system.

2.2. Simple Normalization

In simple normalization method for all images, test and train data, the mean value and the standard deviation are set to zero and one respectively. As a consequence the brightness and contrast will be adjusted homogeneously for images of different sources. The normalization is applied on the resized images to 1024 by 1024, and then the routine lung segmentation is applied including feature calculation, training step. Having the CAD system trained, lung segmentation could be tested on normalized test images. Instances of normalized images from different sources can be seen in section 4.1.

2.3. Local Normalization

In this experiment the images are locally normalized. The approach that is employed in this part is the same as the local normalization in the preprocessing steps for lung nodule detection in (Schilham et al, 2006). The chest radiograph includes several structures in different sizes and also in the lung field there are structures of different kinds which lead to variation in the intensity in the lung field. Applying the normalization locally leads to the fact that each point just gets influenced by its surrounding; which is determined by the scale parameter sigma (σ_{LN}). The local normalization filtering results in the global contrast equalization through the whole image it can be seen in Figure 8. In addition to global equalization of contrast this approach normalizes the edge strength as well. [4]

The local normalization on the image L is formulated as (1):

$$L_{LN} = \frac{(L - \tilde{L})}{(\tilde{L}^2 - (\tilde{L})^2)^{1/2}} \quad (1)$$

The tilde sign indicates the Gaussian blurring. The procedure can be explained as the image local deviation from its local average which is normalized on the local standard deviation. This process is controlled by a parameter σ_{LN} which is the scale for Gaussian blurring. The key property of local normalization is that it can be done locally with different scale values. Here the local normalization has been tested by implementing different scales from finer to coarser values of 30, 50, 128, 256 and 560 pixels on images of size 1024 by 1024 pixels. The examples of locally normalized images with different scales and their corresponding detected lung masks can be seen in section 4.2.

2.4. Multi Band Local Normalization

Multi band local normalization is defined as individually implemented local normalization on each energy band. The chosen algorithm for deriving the sub band images is a hierarchical unsharp masking as in Stahl et al. [5]. First each image is divided to five sub bands then each sub band is locally normalized with different scales. To remove the local normalization artifact which happens in homogenous areas a constraint is applied on each locally normalized sub band separately. A scaling gain is applied to all the sub bands and finally all of them are added together. The lung segmentation with the same configuration as in section 3.2 is applied on the output image of multi band local normalization. In this section every step of the multi band local normalization is described, for multi band locally normalized images and the lung segmentation results see section 4.3.

Figure 1 Deriving the sub band images process. Window level for four first sub band is (100/0) and for the last sub band and the original image is 2500/4000.

As Figure 1 shows, this is how the sub band images are derived: Original image is blurred by Gaussian blurring with scale 2 pixels; the blurred image is subtracted from the original image that gives the first sub band including the highest frequency components. The blurred image is blurred again with scale 4 and subtracted from the previous blurred image which gives the second sub band. This process continues with the scale getting doubled each time up to 16 pixels.

Two different images are derived from each sub band image: the variance and the locally normalized images. Each pixel in variance image equals to the variance value in the n by n neighbor of the corresponding pixel in the sub band image; the respected

pixel is the center of this n by n neighbor. This step will be discussed further in the succeeding paragraphs.

To limit blurring of the low frequency components of the image, the first sub band, which is the highest frequency sub band, is normalized by the largest scale and for each sub band the sigma is decreased by four. Therefore from the highest to lowest frequency sub bands images are locally normalized by the scale of 64, 60, 56, 52 and 48 pixels respectively.

A constraint is set in order to improve the locally normalized images. As it is shown in the Figure 1, there are some artifacts in homogenous areas caused by local normalization especially at smaller scales. In order to decrease these artifacts, the areas with low variance, homogenous areas, are slightly blurred. To be more specific if the pixel value in variance image is less than one ninth of the variance of the corresponding sub band image; then the corresponding pixel in the locally normalized image is replaced with the average value in its n by n neighborhood:

$$\text{Pixel value in variance image of sub band}_{(i)} < \frac{\text{variance of sub band image}_{(i)}}{9} \quad (2)$$

And here is how the variance for sub band is calculated:

$$\text{variance} = \frac{\sum_1^{x_{dimension}} \sum_1^{y_{dimension}} (\text{pixel}(x, y) - \text{mean})^2}{\text{Number of pixels}} \quad (3)$$

Figure 2 shows the five sub bands and their corresponding variances, locally normalized and improved locally normalized images. Note that the W and L in the figures' caption stand for "window" and "level" respectively.

W/L 100/0	W/L 100/50	W/L 6/0	W/L 6/0

W/L 100/0	W/L 15/5	W/L 6/0	W/L 6/0

W/L 100/0	W/L 10/3	W/L 6/0	W/L 6/0

W/L 100/0	W/L 10/3	W/L 6/0	W/L 6/0

W/L 2500/4000	W/L 500/150	W/L 6/0	W/L 6/0

Figure 2 Sub band images, variance images, locally normalized images and enhanced locally normalized images from left to right column wise.

At the end before adding up all the sub bands, a scaling gain is applied to each sub band to equalize the variance of all the images. For this purpose 100 Kanyama training images are considered as reference, the variance for each sub band for every image is calculated and the average variance value for these 100 images are calculated consequently. Each sub band is multiplied by the gain ratio, which is formulized as (4):

$$\text{Gain Ratio} = \frac{\text{reference average variance value for sub band "i"}}{\text{variance value of the sub band "i"}} \quad (4)$$

Finally these enhanced sub bands are added up. Figure 3 shows an example of a final image after all these steps.

Figure 3 Final image containing all sub bands (W/L 25/0)

This algorithm is tested for the variance images that are generated with 3 by 3, 5 by 5, 7 by 7 and 9 by 9 pixels neighborhood, see section 4.3.

3. Theory

3.1. Lung segmentation

First it is good to know what is considered as lung field in the lung segmentation process. One possible definition for lung field is the group of pixels in the chest radiograph that radiation has passed through the lung; although it is impossible outline this definition from the ordinary PA chest radiographs [2]. Therefore, the lung field is defined as the group of pixels in the chest radiograph that radiation is passed only through the lung but not mediastinum, the heart, the structure below the diaphragm and aorta [2]. It is estimated that %26 of lung field is overshadowed by one of these structures [6].

3.1.1. Matching

In segmentation methods matching or registration can be implemented in terms of pairing of an input image with a segmented reference image [2]. Matching is generally determined by four parameters [7]. The first parameter is the choice of feature space that ranges from choosing the reference image and the points in the respected image to use or possibly apply certain operation on. The second parameter is the allowed transformation. In the case of lung segmentation since the reference image and the input image are from different subjects so a global transformation should be used; such as an affine transformation, a local and elastic deformations. Third parameter is similarity metrics which indicates how well the transformed input image matches the reference image and finally the fourth parameter is an algorithm to determine the transformation that maximizes the similarity metrics.

3.1.2. Pixel classification (PC):

Having a feature vector for each pixel, segmentation can be treated as pixel classification problem; the output of the classification is the anatomical structure which encloses the pixel. Any kind of classifier can be utilized in this approach. Although different classifiers will not lead to the same result, the key factor in pixel classification is the choice of features. One of the advantages of pixel classification is that it does not inquire the user to set any parameter; the classifier will give the optimum result in some sense. Another interesting fact about applying this approach is implicitly employing knowledge in terms of using the training images. Unlike general segmentation methods such as watershed or region growing techniques that call for parameter(s) to be set by the user; implementation of this approach gives the optimum parameter by the classifier. However for pixel classification implementation the user choose the parameter set. [2]

3.1.3. Features:

As it has been mentioned before, lung segmentation depends on the chosen features. Due to the especial shape and location of the lung in the chest radio graphs, pixel location can be considered as one of the candidate features. Another specific

characteristic of lung field is that pixels inside the lung have lower values compared to that of surrounding tissues while inner lung pixels having higher values than pixels outside the body. This fact makes the intensity a good feature for lung segmentation [2]. Furthermore, overlaying ribs and vasculature makes the intensity inside the lung vary locally in small scales while the intensity outside the lung is more homogenous. It itself can be also formalized as a feature. [2]

Combination of Gaussian derivatives output at correct scale and direction is so close to the anatomical structures and boundaries which are not connected. The thorax central, the lung central lines, the edges of the mediastinum, the rib cage, the rib borders, and in the centerline of costal and intercostal spaces can be extracted this way. [2]

Pixels are transferred to their correct locations by use of the local structures that are obtained from Gaussian derivatives at different levels. This way the height and width of the lung and the location of the lung come to the correct place with correct size.

3.1.4. Classifier and feature transformation:

Among different classifiers the K nearest Neighbor classifier is employed for transformation of features. In KNN classification there is a feature space which derived from the training step. What happens in training step is that the training samples are labeled with the number of classes they belong to. The parameter k in this method is chosen by the user. For each query point q the k nearest neighbors are determined then the query point belongs to a class that encloses the most number of points among its k nearest neighbors. In the other word the class is defined by the majority of the votes by its neighbors. As the number of K increases, the noise effect decreases while the difference between the boundaries get less distinct [8]. One of the interesting characteristics of the KNN method is that under a specific statistical assumptions and infinite training data the error is $\left(1 + \frac{1}{K}\right) * R$ where R is the minimum achievable Bayes error. [8]

The tree-based KNN as in (Arya et al, 1998) has been implemented which is an optimal algorithm to approximate the nearest neighbors. This method puts the points of the set S in a certain data structure so for any given query point q, the nearest neighbor to q can be determined quickly. The nearest neighbor approximation is controlled by the factor ϵ . It is done in a way that the point p is $(1+\epsilon)$ -approximation of the nearest neighbor of q if its distance from q is a factor of $(1+\epsilon)$ of the distance of the true nearest neighbor of q. This can be extended for searching k nearest neighbor of the query point which means for $1 \leq j \leq k$, we have $p_1, p_2, \dots, p_k$ points that p_j is $(1+\epsilon)$ approximation of the jth nearest neighbor of the q. During searching it keeps track of k nearest neighbor approximations. Once the distance of the point p from q exceeds $(r_k / (1 - \epsilon))$, r_k is the distance of the latest kth nearest neighbor from q, the searching will terminate. The reason is that no other visited point can be closer than $r_k / (1+\epsilon)$ to the query point. The ϵ is set to 2 instead of 0 which gives the exact KNN; this installation does not decrease the accuracy [9]. The

features are also normalized which means the mean value is set to zero and the standard deviation to one.

3.2. Lung Segmentation Configurations in this project:

Here the state of the art *supervised lung segmentation* technique as in (B. van. Ginneken et al, 2006) is used [9]. The pixel classification method is the implemented approach for the lung segmentation. It is important that all the images have equal width of 1024 pixels, so the very first step would be to resize all the images from both training and testing dataset to 1024 pixels width size.

In the training sequence, first the image resolution is converted to the working resolution then for each image the features are calculated. By sampling the features, the feature vector is obtained for certain locations in the image. A suitable transformation for feature vector should be calculated, which follows by training the classifier by the feature vectors as inputs and the corresponding object classes as outputs. The classifier should be able to calculate the posterior probability for each input vector that belongs to each object class.

The first step in the testing and training sequence which is getting the image resolution at working resolution is the same. The second step is to calculate the feature vector for each pixel. The feature vectors are transformed then the posterior probability for feature vectors belonging to each object class is obtained by the trained classifier and the transformed feature vectors. The final step is thresholding the probability map at 0.5 which yields the binary segmentation.

In the multi resolution PC method, the training stage is performed for a group of working resolutions. At the testing stage the segmentation process starts with the coarsest resolution. The posterior probability p_i is calculated for each class i . In addition to p_i the $p_{\min}$, which is $\max(p_i, 1-p_i)$, is also derived. Then it continues with the next resolution and it just calculates the p_i for the pixels that their $p_{\min} < \text{Threshold } T$ at the previous resolution; otherwise, it just takes the interpolated p_i from the coarser resolution in. This procedure continues to the finest resolution. The advantage of multi resolution PC method is to decrease the computation time for classification.

The lower resolution images are created with Gaussian pyramids at 256 by 256 and 128 by 128 resolution levels and the T is set to 0.99 conservatively. [9]

The training images are sub sampled by a 64 by 64 grid which gives 4034 samples for each image and KNN classifier with $K=15$ is employed. The features that are considered for lung segmentation are x and y location, output of the Gaussian filter (of up to second order) for various scales of 0, 1, 2, 4, 8, 16 pixels which characterizes the local structures and finally the gray level value of the original image.

The features are calculated for each working resolution. Due to different pixel size for different resolution level, the filter apertures with different sizes are used to couple the

different resolution levels. In other words the larger aperture size is employed for coarser segmentation at coarser resolutions and the finer aperture size is employed to segment finer details consequently at higher resolutions. [9]

3.3. Post Processing:

The best possible lung mask can be obtained by applying other actions on the soft probability map (p_i) besides converting p_i to a binary mask by thresholding; the presence of some clouds of isolated pixels near the boundaries of the object does not ensure the connectedness of the detected object. Therefore the transition between the probability map and the lungs masks shown in succeeding sections includes some post processing actions. Post processing consists up to four different stages which are named and described elaborately in the next few paragraphs:

3.3.1. Gaussian Blurring:

Blurring is the very first procedure that is applied on the probability map. Blurring of the probability map results in more homogenous map and clouds the isolated pixels. For this purpose the Gaussian filter at the scale of 0.7 pixels is used which gives a smoother version of the raw probability map.

3.3.2. Thresholding

Thresholding is applied on blurred probability map in order to maximize the chance of capturing the true inner lung area within the lung mask the. In this work the threshold value is set to 0.5 for the first two applied methods and it is set to 0.3 for the third proposed method.

3.3.3. Connected component labeling

In general the connected component labeling definition dictates the scanning of the thresholded maps in hope of finding largest groups of connected pixels. Here the goal is to detect two largest groups of pixels that are connected. At this step in number of cases some information loss has been observed. During this project occasionally the upper part of lung had been missed; this is due to disconnection between the upper part of the lung and lower part of the lung below the clavicle. This type of information loss did not happen with case of varying threshold value in the section 1.4.2. Note that it is not practical to change every single configuration for each individual image.

3.3.4. Morphology

Closing is the morphological operation that is implemented in post processing. This operation simply enlarges the boundaries. One of the advantages of the closing operation is elimination of the holes in the foreground. Kernel size parameter controls the closing process and it is normally set to 10. Here different values for kernel size were tested in order to control any change or improvement in lung mask shape. Note the kernel size is 10 unless it is mentioned otherwise, See Appendix II.

All the post processing steps so far are illustrated in Figure 4.

Figure 4 The probability map before and after applying post processing sequences.

Some times after the closing process, the image ends up with a large number of holes in the lung, to remove this problem another post processing action is applicable which is stated in the following section 3.3.5.

3.3.5. The hole filling

A few numbers of visible holes in the lung mask could be observed in the succeeding sections. Beside Gaussian blurring, thresholding, connected component labeling and morphological operation some further processing action can be implemented to improve the detected lung mask. This post processing includes filling the holes inside the lungs which in most cases makes the lungs mask looks much better, see Appendix I and Appendix II. Figure 5 depicts the lung masks before and after the hole filling step.

Figure 5 Probability map before and after applying the hole- filling

3.4. Evaluation:

In order to evaluate the lung segmentation method an evaluation index is necessary. In our case that there are only two classes, the lungs and the outside the lungs area, there are four parameters that can be obtained and used in the evaluation process. (1) True positive (TP) area or correctly classified as object, (2) the False Positive (FP) area that is classified as object but it actually is in background, (3) the False Negative (FN) area that is classified as back ground although it is in the object and finally (4) the False Positive (FP) area that means it is classified as background and it really is in the background. Having these values different measures such as accuracy, sensitivity, specificity, kappa and overlap can be derived.

The evaluation of the lung segmentation is based on the overlap ratio which is defined as the Jaccard index or Jaccard similarity coefficient. In the other words it is the ratio of

intersection over the unity that is formulized as in (5) [2]. This is an acceptable measure but the fact that it is lower for smaller and more complex objects in comparison with large objects is inevitable [10].

$$\text{Overlap ratio} = \frac{\text{Number of true positives}}{\text{number of true positives} + \text{number false negatives} + \text{number of false positives}} \quad (5)$$

As you will see in the succeeding sections this ratio is used for evaluating each experiment and the results are reported correspondingly.

3.5. Base line for the experiments:

Deriving a base line for the project is the first and the most primary step to take. For this project the Computer Aided Detection system for lung segmentation that has been fully described in the preceding sections is employed. The data that are used for the base line experiment and later on for other experiments are gathered from six different sources: (1) 200 images from Kanyama, (2) 100 images from Tilburg, (3) 20 images from JSRT, (4) 20 images from London, (5) 20 images from Gambia and (6) 20 images from RUNMC.

For training of the CAD system, 100 images from the Kanyama dataset are employed. Having the system trained, the testing stage is performed on 100 test images from Kanyama, images from Tilburg, JSRT, London, Gambia and RUNMC dataset. In the Figure 3 the test results for several images from each dataset are illustrated.

Figure 6 shows the original image, the desired lung mask which are manually annotated and the lung segmentation by CAD system and the probability map. Table 1 shows the overlap ratio of the base line experiment.

Figure 6 First column is the original image from Five different Datasets, 2nd column is the desired lung mask, 3rd column is CAD lung segmentation and 4th column is the probability map.(W/L 3000/2800) Row-wise the images belong to Kanyama, JSRT, Tilburg, Gambia, RUNMC and London respectively.

Training dataset	Testing dataset	Pre-processing steps	Post Processing step	Overlap Ratio %
Kanyama Training dataset	Kanyama test	None	- Gaussian Blurring (Sigma=0.7pixels) - Thresholded at 0.5 - Connected component labeling - Morphology	91.19%
	Tilburg	None	Same as above	43.82%
		Contrast Inversion	Same as above	0.6%
	JSRT	None	Same as above	73.03%
	Gambia	None	Same as above	9.85%
	RUNMC	None	Same as above	0.02%
	London	None	Same as above	0.00%

Table 1 The evaluation result for the baseline experiment

As it can be seen in Figure 6, the CAD system fails to work properly for the datasets other than the Kanyama dataset. According to Table 1 the overlap ratio of 91.19% for kanyama testing data shows that the system works accurately when the data from the same source is used for training and testing steps. On the other hand for the data from the different source the overlap ratio is not as high as that of Kanyama. Among datasets from different sources than the training data, JSRT with the overlap ratio of 73.03% have an acceptable output while the overlap ratio for other datasets is almost zero. In Tilburg dataset the parts of the image that are classified as lung are the areas outside the lung; this may be because of the intensity range difference between the Tilburg and Kanyama datasets. This explains why when the Tilburg dataset contrast is not inverted, the inside lung have higher intensity while the ribs have lower intensities, the overlap ratio is 43.82% which of course is not of course acceptable.

The dissimilarity between the outcome of the lung segmentation of different sources does not only exist after the segmentation; if one take a look at the original images it is easy to see that the original images look different from the reference images too. We are looking at the images with the same window leveling setting so that they can be comparable to each other. The London image has the worst case of all; it is completely white in the appointed window and leveling set. It can be implied that the segmentation for JSRT turns out better than the other datasets because the original image looks more like the reference image as well which is not a surprise considering the fact that intensity is one of the implemented features in the lung segmentation see Figure 6.

4. Results

In this chapter the normalized images as a result of each of three proposed methods and their corresponding detected lung masks are shown and discussed upon.

4.1 Simple Normalization

It is expected that this simple normalization leads to greater overlap ratio in comparison with the results from the baseline experiment. The lung segmentation overlap ratio before and after applying normalization are presented in Chart 1. Figure 7 illustrates the normalized images, desired lung mask, detected lung mask by CAD and the probability map. Note that the W and L in the figures' caption stand for "window" and "level" respectively.

Chart 1 Comparing the overlap ratios before and after applying simple normalization

Figure 7 First column is the normalized images, 2nd column is the desired lung mask, and 3rd column is CAD lung segmentation. Each row belongs to the same data set from top to bottom Kanyama, Tilburg, JSRT, Gambia, RUNMC and LONDON respectively. (W/L 6/0).

As it can be seen from the images of Figure 7 there is a significant improvement in the detected lung mask in normalized images compared to the ones from the base line experiment. Compare to the result of Tilburg dataset in the baseline experiment, the

simple normalization method does not give the perfect result but it correctly classifies the areas within the lung as the lung area. By looking at the original images one can expect that there should be an improvement in the detected lung mask due to the obvious increased similarity between the original images. All the images are looking at with the same window and leveling set and all of them are visible, on the contrary to baseline experiment that the image from the London dataset could not be seen at same window level as the others. The overlap ratio supports the improvement that comes from applying the simple normalization, there is almost 50% leap in overlap ratio for Tilburg, RUNMC and London dataset comparing with the baseline experiment. Although there is slight decrease in overlap ratio for the Kanyama test images.

From the comparison between the lung masks from baseline experiment and after applying the simple normalization, it can be inferred that normalizing does enhance the overlap ratio and makes the CAD system see the similarity between different images. Since the CAD system depends extremely on features and intensity is one of the features for lung segmentation purpose so the normalization makes the intensity for different images be almost in the same range. Subsequently the variation between images is less than before and its influence can be seen directly by observing the improvement on the lungs masks.

4.2 Local Normalization

The following figures 8-14 and chart 2 illustrate results of local normalization and the overlap ratios for the lung segmentation with the same configuration as section 3.2, except that the hole filling step is skipped in the post processing procedure.

	Sigma:30 pixels	Sigma: 50 pixels	Sigma: 128 pixels	Sigma: 256 pixels	Sigma:512 pixels
Kanyama
Tilburg
JSRT
Gambia
RUNMC
London

Figure 8 Each column show locally normalized images with sigma 30, 50, 128, 256 and 512 pixels from left to right respectively and each row show the same image from one dataset with different normalization scales. Kanyama, Tilburg, JSRT, Gambia, RUNMC and London dataset from top to bottom. (W/L 6/0)

Figure 9 First row Probability map and the second row the detected lung mask for Kanyama locally normalized images with scale 30, 50, 128, 256 and 512 pixels, from left to right column wise.

Figure 10 First row Probability map and the second row the detected lung mask for Tilburg locally normalized images with scale 30, 50, 128, 256 and 512 pixels, from left to right column wise.

Figure 11 First row Probability map and the second row the detected lung mask for JSRT locally normalized images with scale 30, 50, 128, 256 and 512 pixels, from left to right column wise.

Figure 12 First row Probability map and the second row the detected lung mask for Gambia locally normalized images with scale 30, 50, 128, 256 and 512 pixels, from left to right column wise.

Figure 13 First row Probability map and the second row the detected lung mask for RUMCN locally normalized images with scale 30, 50, 128, 256 and 512 pixels, from left to right column wise.

Figure 14 First row Probability map and the second row the detected lung mask for London locally normalized images with scale 30, 50, 128, 256 and 512 pixels, from left to right column wise.

Chart 2 Comparing the overlap ratios of the lung segmentation on locally normalized images with baseline experiment

Chart 2 indicates that local normalization at any scale will result in better automated lung segmentation comparing the baseline experiment. Despite the significant enhancement in the overlap ratio at scale 30 pixels for most of the dataset, the JSRT and Kanyama test dataset have a slight decrease in overlap ratio from the baseline experiment. According to Chart 2, the best scale for applying local normalization is 50 pixels in which overlap ratio is more than 80% in most of cases. As it can be seen in Chart 2, the overlap ratio starts to decrease from sigma 128 pixels for all the data sets except for Kanyama dataset, which has the constantly increasing overlap ratio as the sigma increases. As for JSRT, there is no significant overlap ratio fluctuation as the scale changes.

The most similarity among the normalized images from different sources could be observed at sigma 50 and 128 pixels, see **Figure 8**. For coarser scales the similarity between different sources is diminished which directly affects the lung segmentation and reduces the overlap ratios. At a certain normalization scale this approach is expected to yield the same result as the simple normalization (average zero and standard deviation one); where the normalization scale is infinite. As it can be seen from the last columns of Figures 8-14 the probability maps start to have the pixelly-look patterns as in the probability maps of Figure 6; one may infer that as the scale grows the output starts behaving more similar to the simple normalization.

In conclusion, the local normalization could improve the CAD system performance on images from the different sources. Local normalization could be optimized by selecting the most suitable normalization scale value. As it has been noted earlier in this section, the scale of 50 pixels gives the best overlap ratio between the desired lung mask and the

automatically detected one and it results in the highest similarity between the images from the different sources in our case.

4.3 Multi Band Local Normalization

In this section the multi band local normalization result for different neighborhood sizes of 3 by 3, 5 by 5, 7 by 7 and 9 by 9 are shown. The normalized images are illustrated in Figure 15 and the output of the CAD system for these images is shown in Figures 16-21 and the corresponding overlap ratios can be seen in **Chart 3**. Note that during this experiment the threshold is set to 0.3 in the post processing procedure.

	Neighborhood scale:3 by 3	Neighborhood scale:5 by 5	Neighborhood scale:7 by 7	Neighborhood scale:9 by 9
Kanyama
Tilburg
JSRT
Gambia
RUNMC
London

Figure 15 Multi band locally normalized images with different neighborhood size o 3 by3, 5by5, 7by7 and 9by9 for variance image. (w/L 20/0)

Figure 16 First row Probability map and the second row the detected lung mask for multi band locally normalized Kanyama images with different variance images of 3by3, 5by5, 7by 7 and 9 by 9 neighborhood size, from left to right column wise.

Figure 17 First row Probability map and the second row the detected lung mask for Tilburg multi band locally normalized images with different variance images of 3by3, 5by5, 7by 7 and 9 by 9 neighborhood size, from left to right column wise

Figure 18 First row Probability map and the second row the detected lung mask for JSRT locally normalized images with scale 30, 50, 128, 256 and 512 pixels, from left to right column wise.

Figure 19 First row Probability map and the second row the detected lung mask for Gambia locally normalized images with scale 30, 50, 128, 256 and 512 pixels, from left to right column wise.

Figure 20 First row Probability map and the second row the detected lung mask for RUMCN locally normalized images with scale 30, 50, 128, 256 and 512 pixels, from left to right column wise.

Figure 21 First row Probability map and the second row the detected lung mask for London locally normalized images with scale 30, 50, 128, 256 and 512 pixels, from left to right column wise.

Chart 3 Comparison between multi band local normalization and baseline experiment

Comparing the multi band local normalization with the base line experiment, a significant leap in the overlap ratios of all the datasets could be observed. It indicates that implementing the multi band local normalization decreases the differences between the overlap ratios of the different datasets in other word it makes the images more similar to one another. The various neighborhood scales for variance image does not lead to notable changes in the output. There is a small increase in overlap ratio from neighborhood of size three to neighborhood of size five. The results are almost the same for the neighborhood five and seven, while the overlap ratio slightly decreases for neighborhood seven to nine.

Figure 15 supports the last paragraph statements as well. The images from all datasets look almost the same for all different scales variance images. The dissimilarity between images from one scale to another is so small that cannot be noticed. Consequently the lung segmentation result for all different scales is almost the same. See Figures 16-21.

To find out how different the results would be without the constraint for correcting the local normalization artifact, this experiment is repeated but this time the variance matrix and the correction step are skipped; the images are shown in Figure 22.

	Skipped Correction Step	Neighborhood scale: 7 by 7
Kanyama
Tilburg
JSRT
Gambia
RUNMC
London

Figure 22 Comparison between overlap ratio of multi band local normalization with and without correction (W/L 25/0)

From chart 4 it can be inferred that skipping the local normalization artifact correction does not lead to any notable difference in lung segmentation overlaps. Interestingly, the skipped correction and the neighborhood 3 by 3 have almost yielded the same overlap ratios. Skipping the correction step makes the algorithm simpler although the

appearance of the images with correction is better. In conclusion both procedures lead to excellent outputs both for the images to be segmented in the sense of similarity and the detected lung masks.

Chart 4 Comparison between multi band local normalization with or without correction.

5. Summary and overall discussion

Computer Aided Detection (CAD) systems are expected to be widely accepted by radiologists for decreasing the work load and also making large screening programs possible. A major part of CAD system is learning from examples, which is done by simplifying the images to numerical descriptors and train the system by them. Tuberculosis is a major healthcare problem in large parts of the world (1.3 million deaths in 2007). At Diagnostic image Analysis Group a system has been developed to detect TB on frontal chest radiographs. This system consists of several components including lung segmentation, texture analysis and etc. The lung segmentation component of this system is employed in this work to study the image variation effect on.

A key issue to be resolved in practical application of CAD systems is that usually image acquisition is not employed exactly at the same condition for all cases. Unequal initial condition causes variation in images from different sources and consequently affects the CAD system function. To limit this variation there are several options to be considered (1) the images can be normalized so that they look more equal, (2) to choose less variation sensitive features and (3) to modify the classification method to be able to correctly classify regardless of any variation in the image. In this project the effect of variation among images of different sources is studied on CAD for TB lung segmentation component. The lung segmentation feature vector consists of intensity value, the x and y position and output of the Gaussian filter (L_x , L_y , L_{xy} , L_{xx} , L_{yy}) for scales of 1, 2, 4, 8 and 16 pixels. The system is trained with 100 images from one source and tested on images from six different sources, including the test images from the same source as the training images.

As a solution to reduce the effect of image variation in the output of CAD system, three normalization approaches are implemented on the images including: (1) simple normalization, (2) local normalization and (3) multi band local normalization. Figure 23 shows instances of images before and after applying these three normalization methods. It also reveals the visible difference among the original images, as in some cases, for instance London original image in Figure 23, even with the same window leveling set it is not possible to observe images of different sources.

Implementation of each normalization method contributes to the equality among images from different sources up to a certain level. As it can be seen from Figure 23 local normalization and multi band local normalization result in higher level of similarity among images. The Chart 5 shows the overlap ratio for the detected lung mask by CAD system before and after applying normalization, this table supports the above statement; normal images achieved higher overlap ratios. According to the results local normalization and multi band local normalization lead to higher level of

similarity between images and consequently better and more accurate overlap ratio for lung segmentation.

Figure 23 Original images (W/L 3000/2800), images from simple normalization (W/L 6/0) , images from local normalization scale 50 pixels(W/L 6/0), and images from multiband local normalization(W/L 20,0) column wise from left to right

Chart 5 Comparison between the overlap ratio of the Simple normalization, local normalization, multi band normalization with and without correction and the baseline experiment.

There are other possible options to enhance the performance of the CAD system even more. First, thresholding the probability map can be done at a threshold level different than 0.5 considering the highest probability value for lung area. As easy and effective that this idea sounds there is a tradeoff between thresholding and desired lung area in the lung mask, since changing the threshold value may also include the unwanted areas outside the lung as lung field. Second, one can use mixture of data from different sources for training the system. By this way the system is trained with variety of instances and consequently will give better response to data from different sources. Training the system with data from various sources is a common approach and it can be considered as a way to improve the performance in case of different sub data.

6. Conclusion

Normalization is proposed as a possible solution for solving the effect of image variation on CAD systems. Three normalization methods are implied in this study, each of them improves the performance of the CAD system for lung segmentation up to some point. The first approach for normalization is normalizing the images to average zero and the standard deviation of one, which boosts the lung segmentation output although it does not give the best results. At the second approach images are normalized locally, this normalization leads to almost equally high performance of the CAD system on test images from both inside and outside the training images' database. The third implied normalization method is multi band local normalization which enhances the lung segmentation significantly compared to baseline experiment. Similar to the second approach, it also declines the difference between the lung segmentation on the test data from five different sources and the test data from the same source as training data.

The performance of the local normalization and the multi band local normalization is relatively the same and both lead to better lung segmentation than the simple normalization; however, the local normalization is less complicated and time consuming than the multi band local normalization. The fluctuation between the outputs of the CAD system for multi band locally normalized data from different sources is less than that of locally normalized data. Both methods are acceptable approaches for normalizing the images and help standardizing the CAD system performance; each has controlling parameters that can be set to optimize the result. In this project the best result of local normalization is achieved by setting the normalizing scale to 50 pixels. In case of the multi band local normalization there is a slight difference in the output by changing the neighborhood size for deriving the variance image. One may guess the neighborhood of size 9 by 9 could give the most effective result since there is a small increase in the overlap ratio for the test data of the same source as the training data. However the neighborhood of size 7 by 7 gives the most optimum result, since the overlap ratio for five test data from different sources is higher than that of 9 by 9 neighborhood size.

7. Future works

Based on the observations in this project normalizing the image reduces the variation effect up to a high level, although there are several possible solutions to be studied and investigated. As an alternative the CAD system can be trained with the mixture of instances from different sources. This way of training should give the system the intuition to classify the data from different sources more accurately, and it is expected to enhance the overall function of the system.

In addition to the statements above, the features considered for describing the images are good option to be carefully investigated. So far the output of Gaussian filter at different levels and scales plus the intensity are the employed features for describing the images and training the system. Gaussian derivatives are sensitive features; since we are aiming at make the system less sensitive to variations, it is would be effective if the alternative features were used instead of the Gaussian derivatives. Hence a possible substitute for Gaussian derivatives is considering the gray scale invariant gradient options.

In cases that the training data and the test data are not necessarily from the same space the knowledge transfer or transfer learning is a desirable solution. [11] If this process is carried out successfully it will enhance the performance of learning significantly. Transfer learning lets the features and domains in training and testing procedure to be different. The idea is inspired by human intelligence that by applying previously learned knowledge can solve the new problem. Transfer learning attracts a lot of attention in the field of data mining and machine learning and their different applications. The figure 24 shows the difference between the traditional training and the transfer learning.

Figure 24 (a) traditional machine learning (b) transfer learning [11]

Figure 24 illustrates how the transfer learning tries to use the previously learned knowledge to get to the new target; on the contrary the traditional machine learning learns each task from scratch. The above statements strongly suggest transfer learning as a potential solution for the problem defined in this project.

8. Bibliography

- [1] "Global tuberculosis control - epidemiology, strategy, financing," World Health Organization, 2009.
- [2] B. v. Ginneken, *Computer Aided Diagnosis in Chest Radiography*, Wageningen: Printed by Ponsen & Looijen, 2001.
- [3] S. Junji, L. Qiang, A. Daniel and D. and Kunio, "Computer-Aided Diagnosis and Artificial Intelligence in Clinical Imaging," *Seminar in Nuclear Medicine*, vol. 41, no. 6, pp. 449-462, 2011.
- [4] A. M. Schilham, B. v. Ginneken and M. Loog, "A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database," *Medical Image Analysis* 10, pp. 247-258, 2006.
- [5] M. Stahl, T. Aach and S. Dippel, "Digital radiography enhancement by nonlinear multiscale processing," *Medical Physics* 27, pp. 56-65, 2000.
- [6] H. Chotas and C. Ravin, "Chest radiography: estimated lung volume and projected area obscured by the heart, mediastinum, and diaphragm," *Radiology*, vol. 193, no2, pp. 403-404, 1994.
- [7] L. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325-376, 1992.
- [8] R. Duda, P. Hart and D. Stork, *Pattern classification*, second edition, New York: Wiley, 2001.
- [9] B. v. Ginneken, M. B. Stegmann and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, pp. 19-40, 2006.
- [10] M. Gerig, G. Jomier and M. Chakos, "Valmet: a new validation tool for assessing and improving 3D object segmentation.," in *MICCAI 2001, Lecture Notes in Computer Science*, vol. 2208., Berlin, 2001.
- [11] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, vol. 22, no. 10, pp. 1345-1359, 2010.
- [12] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code.," *IEEE Transactions on Communications OM-31 (4)*, pp. 532-540, 1983.
- [13] B. S. Everitt, S. Landau, M. Leese and D. Stahl, *Miscellaneous Clustering Methods*, 5th edition, UK: John Wiley & Sons, Ltd, Chichester, 2011.
- [14] B. v. Ginneken and B. M. t. H. Romney, "Automatic segmentation of lung fields in chest

- radiographs," *Medical Physics*, vol. 27, no. 10, pp. 2445-2455, 2000.
- [15] R. Fisher, S. Perkins, A. Walker and E. Wolfart, *Hypermedia Image Processing Reference*, 2009.
- [16] S. Dippe, M. Stahl, R. Wiemker and T. Blaffert, "Multiscale Contrast Enhancement for Radiographies:Laplacian Pyramid Versus Fast Wavelet Transform," *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 21, no. 4, pp. 343-353, 2002.
- [17] S. ARYA, D. M. MOUNT, N. S. NETANYAHU, R. SILVERMAN and A. Y. WU, "An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions," *Journal of the ACM*, vol. 45, no. 6, pp. 891-923, 1998.

Appendix I

Local normalization of different scales with applying hole- filling as a post processing procedure are shown in this chapter.

Figure 25 Comparing the overlap ratio of local normalization with scale 30 pixels with and without hole-filling.

Figure 26 Comparing the overlap ratio of local normalization with scale 50 pixels with and without hole-filling.

Figure 27 Comparing the overlap ratio of local normalization with scale 128 pixels with and without hole-filling.

Figure 28 the overlap ratio of local normalization with scale 128 pixels with different configuration for lung segmentation.

Figure 29 Comparing the overlap ratio of local normalization with scale 256 pixels with and without hole-filling.

Figure 30 the overlap ratio of local normalization with scale 512 pixels with and without hole-filling.

Appendix II

Multi band local normalization results scales with applying hole- filling as a post processing procedure and different lung segmentation configurations are shown in this chapter. The mentioned kernel size in the following figures refers to the kernel size for the Morphology step in the post processing.

Figure 31 the overlap ratio of multi band local normalization with neighborhood size of 3by3 with and without hole-filling.

Figure 32 the overlap ratio of multi band local normalization with neighborhood size of 5by5 with and without hole-filling.

Figure 33 the overlap ratio of multi band local normalization with neighborhood size of 7by7 with and without hole-filling.

Figure 34 the overlap ratio of multi band local normalization with neighborhood size of 9by9 with and without hole-filling.