CLASSIFICATION OF BREAST CANCER TUMOR BASED ON ULTRASOUND IMAGES.

DEVENDRAN PILLAI PERUMAL

A thesis submitted in fulfillment
of the requirements for the award of the degree of
Bachelor of Chemical Engineering (Biotechnology)

Faculty of Chemical & Natural Resources Engineering
Universiti Malaysia Pahang

April 2009

ABSTRACT

This research primarily focuses on the predictive technology of identifying the state of tumors in the breast tissues. In breast cancer diagnosis, patients are forced to undergo a series of biopsies just to identify and confirm on the state of tumor, as whether malignant or benign. In this research however, an algorithm will be developed using MATLAB Image Processing Toolbox to indentify the state of a tumor solely based on ultrasound images. Ultrasound images of breast tumors are imported into MATLAB and are passed through a set of filters to remove background noise. Next, the filtered images are run through a set of edge detection algorithms which identifies and defines the region of interest. The processed images are analyzed qualitatively and the following results are obtained; the analysis shows that malignant tumors have well defined boundaries while benign tumors have poorly defined boundaries. To test this theory, the algorithm is used to process another set of ultrasound images of unknown characteristics. The results were analyzed and classified into two groups; malignant and benign. The results are compared with the actual biopsy results from the IIUM Breast Cancer Research Institute, Kuantan and all the analyzed results matched the biopsy results. As a recommendation to improve this study, a quantitative analysis on the ultrasound images is carried out so that more accurate results can be obtained. If the development of this algorithm is proven to be a success, it would be used in every hospital throughout the country to diagnose patients with breast cancer.

ABSTRAK

Kajian ini memfukoskan teknologi peramalan bagi menentukan keadaan tumor dalam tisu payudara. Dalam rawatan kanser payudara, pesakit terpaksa menjalani beberapa biopsi bagi mengklasifikasikan tumor mereka. Sebaliknya, kajian ini menggunakan algorithma yang direka dalam MATLAB Image Processing Toolbox bagi mengklasifikasikan tumor berdasarkan imej ultrasound. Imej ultrasound diimport ke dalam MATLAB dan ditapis menggunakan beberapa algorithma penapisan untuk mengeluarkan bunyi akibat getaran. Imej tersebut kemudian dilalui oleh beberapa algorithma pengesan sisi bagi menentukan kawasan yang dikehendaki. Imej yand telah diproses dianalisa secara kualitatif dan keputusan berikut diperolehi; bagi tumor yang berkanser, kawasan sekitarnya adalah mudah diperhatikan manakala, bagi tumor yang tidak berkanser, kawasan sekitarnya tidak boleh diperhatikan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

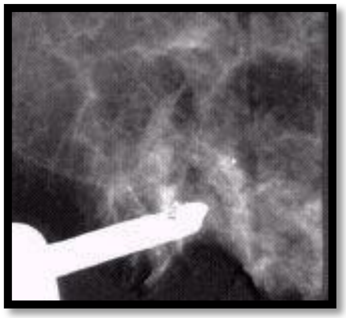
1.1 Background of Study

According to the U.S. National Cancer Institute, breast cancer is defined as cancer that forms in the tissues of the breast, usually the ducts (tubes that carry milk to the nipple) and lobules (glands that make milk). In the United States, 192 370 (female) and 1910 (male) new cases were estimated in 2009. From that, a total of 40 170 (female) and 440 (male) deaths were estimated (*Cancer trends progress reports -2007 update*, 2009). In Peninsular Malaysia, there were 3525 female breast cancer cases registered in National Cancer Registry (NCR) for the year 2006, accounting for 16.5% of all cancer cases registered (*Malaysian cancer statistics-data and figures*, 2006). According to the report, the overall age-standardised incidence rate (ASR) was 39.3 per 100 000 population. The age pattern for 2006 showed a peak ASR at the 50-59 age groups. The incidence of breast cancer was highest amongst Chinese; ASR of 46.4 per 100 000 population followed by Indian; ASR of 38.1 per 100 000 population and Malay: ASR of 30.4 per 100 000 population. These figures created a great deal of concern amongst Malaysians and the Malaysian Government.

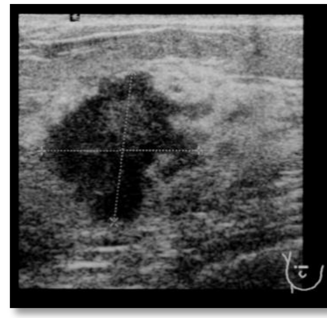
Breast cancer occurs as a result of mutations, or abnormal changes, in the genes responsible for regulating the growth of cells in the breast tissues and keeping them healthy. These abnormal changes give cells the ability to keep dividing without control or order, producing more cells just like it and forming a tumor. The tumor can be classified into two categories, benign (non-cancerous) or malignant (cancerous). Benign tumors are not injurious to health; their cells have close resemblance to normal in appearance, they grow relatively slowly, and do not invade nearby tissues or spread to other parts of the body. Malignant tumors on the other hand are injurious to health. Without proper diagnosis and treatment, malignant cells eventually can spread beyond the original tumor to other parts of the body. Breast cancer however is curable at early stages. The current procedure involving breast cancer diagnosis initiates with screening test, followed by breast biopsy and treatment. Screening tests are done to identify the presence of bodily masses within the breast tissues. It determines the presence of the tumor and also locates the tumor's position in the breast. Photographs of the breast's internal structure are taken and the images are studied by oncologists to identify any abnormalities within the breast tissues.

These images provide oncologists with an exceptional amount of information such as the size of the tumor, shape of the tumor, and the area of spread together with the location of the tumor. Formerly, the most commonly used screening method was mammography, which uses X-rays to capture images of the breast. This method is hazardous due to the amount of radiation which is incident to patient during each screening. Exposure to these radiations contributes to leukemia and other long term diseases (Chang *et al.*, 2005). Hence, patients and radiologists prefer ultrasound scans (US) as safer screening alternative. Ultrasound involves the exposure of body parts to high frequency sound waves to produce pictures of the body's internal structure. Ultrasound does not use ionizing radiation as in mammogram, making it extremely safe. Plus, it also produces images with relatively high resolution (Chen *et al.*, 2008).

Figure 1.1 shows the images of a tumor taken from two different screening methods. Image (a), is a mammogram where the image is less clear and the tumor is poorly defined. Image (b), is the ultrasound image of the same tumor; here the tumor and the surrounding tissues are well defined.



(a)

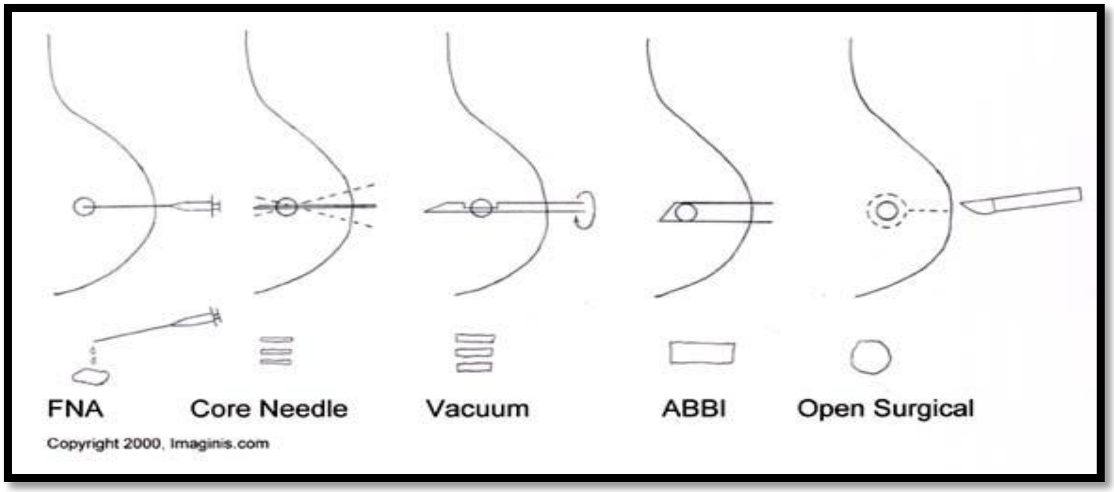


(b)

Figure 1.1 Comparison between mammogram and ultrasound image of breast tumor a) Mammogram of breast tumor b) Ultrasound image of breast tumor

The next step in breast cancer diagnosis is the breast biopsy procedure, which involves removing a sample of breast tissue to determine whether the tumor is malignant (cancerous) or benign (non-cancerous). While ultrasound and other imaging methods can help detect breast abnormalities, biopsy followed by pathological analysis is the only definite way to determine if cancer is present. There are several methods of biopsies available depending on the location, size, appearance and characteristics of the abnormalities. Such available methods are fine needle aspiration (FNA), core needle biopsy, vacuum – assisted biopsy (MIBB), large core biopsy (ABBI) and open surgical biopsy.

Figure 1.2 shows the different forms of biopsies performed to identify the types of tumor. Image (a), fine needle biopsy is a diagnostic procedure used to investigate superficial lumps or masses that is located just below the skin layer. In this technique, a fine, hollow needle is inserted into the mass to extract cells which are then stained and examined under a microscope. This procedure is considered to be very safe and is also regarded as a minor surgical procedure. Image (b), core needle biopsy follows a similar procedure to fine needle aspiration except, fine needle aspiration is used to investigate masses located just below the skin layer whereas, core needle biopsy is used to investigate masses located in tissues deep below the skin layer. Image (c), vacuum – assisted biopsy is new procedure where an incision is made on the breast skin and a vacuum needle is inserted into the breast which then vacuums out the mass from the breast into a sampling chamber. This procedure is done for masses larger than that could be handled by fine needle and core needle biopsies. Image (d), large core biopsy is similar to core needle biopsy except that this procedure is for the extraction of masses larger than that of vacuum – assisted biopsy. Image (e), Open surgical biopsy is an old and common form of procedure where a 1.5 to 2.0 inches long incision is made on the skin and a wire with a hook on its end is inserted into the breast. Once the mass is located, the wire is hooked onto the mass and the wire is pulled out, removing along with it the entire tissue around the lesion.



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

Figure 1.2 Types of biopsy methods a) Fine needle aspiration, FNA, b) Core needle biopsy, c) Vacuum – assisted biopsy, MIIB, d) Large core biopsy, ABBI e) Open surgical biopsy

Apart from breast biopsy, image analysis is currently being tested as an alternative method of diagnosis. Image analysis uses images from screening tests to classify the tumor as to whether malignant or benign (Shen *et al.*, 2007). The results from these analyses however are taken only as secondary references rather than a definitive result to determine the state of the tumor. This research however, hopes to produce results from image analyses which are more definitive and credible.

In this research, MATLAB Image Processing Toolbox is used as the image analysis tool. MATLAB (MATtrix LABoratory) is a numerical computing environment developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, and Fortran (Shi *et al.*, 2009). Although MATLAB was initially designed for numerical computing, an additional package, Simulink, provides graphical multi – domain simulation and Model – Based Design for dynamic and embedded systems. Image Processing Toolbox software provides a comprehensive set of reference-standard algorithms and graphical tools that allows image processing, analysis, visualisation, and algorithm development. It also allows the restoration of noisy or degraded images, enhancement images for improved intelligibility, extraction features, analysis of shapes and textures, and registration of two images. Most toolbox functions are written in the open MATLAB language, giving the ability to inspect the algorithms, modify the source code, and create custom functions.

By using MATLAB Image Processing Toolbox, oncologists can analyze the images from screening tests and immediately classify of the tumor. It is predicted, by using this technique; oncologists should be able to take the analysis results as primary reference during diagnosis and consequently reducing the number of biopsies.

1.2 Problem Statement

At the present, oncologists solely depend on biopsy results to determine whether a tumor is cancerous or benign before they could proceed with the treatment. This has proven to be a major inconvenience to patients and oncologists as breast biopsies pose certain risks and disadvantages. First of all, breast biopsies cause bruising or scarring to the patients breasts, which consequently may cause infections. Besides that, there is high risk probability that this procedure may cause hematoma; a phenomenon referred to pooling of blood trapped inside the biopsy area.

Apart from health issues, financial issues also cause concern amongst patients due to the high cost of the procedure. The cost of a single biopsy ranges from RM2750 to RM6600. The bad news is the state of a tumor cannot usually be determined by a single biopsy; it usually requires a series of biopsies to confirm the presence of cancerous tumor. Commonly, patients are forced to undergo at least three biopsies before oncologists could determine for certain the malignance of the tumor. This directly increases cost of the procedure as well as amplifying the chances of patients to be affected by the risks and side effects of undergoing biopsies.

In some cases, it is possible for biopsies to reveal false positive or false negative results. False positive is a phenomenon when the biopsy results indicate a cancer when it is not actually present, whereas false negative is the phenomenon when the biopsy results miss a cancer when it is actually present. This in turn reduces the accuracy and the reliability of biopsies.

Due to these disadvantages and inconveniences, physicians are finding the new means for reducing the number of biopsies, at the same time acquiring accurate results. Though it is impossible to completely eliminate biopsy procedure, it is possible to use image analysis tools to reduce the number of biopsies as well as provide reliable results as it would in a normal biopsy.

1.3 Ojective

This research aims to classify breast tumors into two categories; malignant and benign, by analysing ultrasound images of breast scan using MATLAB Image Processing Toolbox.

1.4 Scope of Research

This research primarily focuses on three major aspects to classify breast tumors:

1.4.1 Qualitative analysis on ultrasound images of breast tumors

1.4.2 Utilization of Digital Imaging and Communication in Medicine (DICOM) format images of ultrasound scans for analysis

1.4.3 Computing of algorithm using MATLAB Image Processing Toolbox for the purpose of analysis

1.5 Rationale & Significance of Research

The significance of this research paper is to create a completely new method of image analysis that would produce accurate and credible results that would be on par with results produced by biopsies. With that, oncologists and physicians could be able to confidently rely on the analysis results; the results would be a support to the physician's decision in proceeding with the treatment for breast cancer. Consequently, patients would not be required to undergo as many biopsies as before this just to confirm the presence of cancer. This would reduce the patients' risks of the side effects of biopsies and also reduce the overall costs for diagnosis.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Many attempts have been made to come up with an efficient alternative method to classify breast tumors which could outride the requirement of performing biopsies. Unfortunately none of such alternatives have been found to date. However, great deals of efforts have been made to come up with efficient methods to reduce the number of biopsies in which the alternative methods are able to yield results that are identical to biopsy results.

Currently, the most popular method being developed is the image analysis technique. In this method, the tumors are classified through calculations using only the images obtained from screening tests. The upside to this technique is that it requires no physical surgery, thus does not pose any harm or risk to patients, in contrary to biopsies. Unfortunately, this method is far from perfect, affecting the accuracy of the results. This downside is the main reason which contributes to the reluctance of oncologists to rely on the results provided by image analysis method. Despite all odds, researches around the world are striving to make this technology to be more accurate and reliable.

2.2 Numeric Modeling of Breast Tumor

Previous studies have illustrated that; changes in the electrical properties of abnormal breast are more pertinent compared to the breast of normal tissues. In the present study done by Ng *et al.,* 2007, a simple 2D models of breast (close to realistic), with and without artificially inserted malignant cancer were simulated, based upon electrical activity within the breast.

An inhomogeneous female breast model, with close resemblance to the actual one, was developed by considering a breast as a hemisphere with layers of various thicknesses in supine condition. In order to determine the potential distribution developed due to a dipole source, isotropic homogeneous conductivity was assigned to each of these compartments and the volume conductor problem was solved using finite element method.

Significant changes in the potential distribution were recoded in the malignant and normal breast regions. The surface potential decreases about 0.5%, for the small malignant region of surface area $13mm^2$ and the surface potential decreases about 16.4% for large malignant surface area of $615mm^2$. Hence, the results show that, the sizes of tumours result in the reduction of surface potential and follow a fourth order polynomial equation. Thus, biofield analysis yields promising results in the detection of the breast cancer of various sizes.

2.3 Classification of breast ultrasound images using fractal feature

Based on a research done by Chen *et al.*, 2005, fractal analyses have been applied successfully for the image compression, texture analysis, and texture image segmentation. The fractal dimension could be used to quantify the texture's information.

According to Chen, the differences of gray value of neighbouring pixels are used to estimate the fractal dimension of an ultrasound image of breast lesion by using the fractal Brownian motion. In addition to that, a computer-aided diagnosis (CAD) system based on the fractal analysis is applied to classify the breast lesions into two classes: benign and malignant. To improve the classification performances, the ultrasound images are preprocessed by using morphology operations and histogram equalization. Finally, the k-means classification method is used to classify benign tumors from malignant ones. The ultrasound breast image databases include only histologically confirmed cases: 110 malignant and 140 benign tumors, which were recorded.

2.4 Removal of Speckle Noise from Ultrasound Medical Image based on Special Filters: Comparative Study

Removing background noise from the original image is an important aspect in image processing. Basically, there is no common enhancement approach for noise reduction. Previously, several approaches have been introduced and applied. Each approach however has its own assumption, advantages and disadvantages. The speckle noise being studied is commonly found in the ultrasound medical images.

Thangavel *et al.*, 2009 have proposed different filtering techniques based on statistical methods for the removal of background noise from ultrasound images. Various filtering models are used for the purpose of this paper, including max filter, min filter, harmonic mean filter, contra harmonic min filter and geometric mean filter. A number of success full experiments validate the proposed filtering model. The quality of the enhanced images is measured by the statistical quantity measures: Signal-to-Noise Ratio (SNR), Peak Signal-to-Noise Ratio (PSNR), and Root Mean Square Error (RMSE).

2.5 Development of Matlabfiltering techniques in digital speckle pattern interferometry

In interferometric fringe pattern analysis, specular and speckle fringe patterns are the two main divisions. While specular fringes are characterized by quality fringes, speckle (that obtains due to the diffuse scattering of the coherent radiation from an optically rough surface) fringe patterns are characterized by noisy fringes.

This paper, written by Murukeshan *et al.*, 2003, focuses on this aspect and the Matlab based filtering techniques to improve and enhance the quality of speckle fringe patterns by developing an appropriate algorithm. Furthermore, the newly developed algorithm ‘‘Macurv’’ will be presented which can give the second order derivative (curvature) fringe information. An algorithm with several functions is written using Matlab. The algorithm includes, digital shearing, Fourier transform, subtraction and Lowpass filtering. The objective of the algorithm is to provide a more effective way for the post-processing of speckle interferometric fringes.

2.6 Finite Element Analysis and Acoustic Modeling

The relationship between the geometry of the vocal tract and the speech sound produced has been studied using acoustic models of the human vocal tract by Yeung, 2005. This project involves 3D finite element analysis (FEA) and acoustic tube modeling of the human vocal tract phonating the /r/ sound which requires the extraction of 3D tract geometry information from magnetic resonance images (MRI) of the tract using medical image processing techniques.

Results obtained from the FEA simulation can identify a more accurate area function extraction method and the correct acoustic tube configuration, permitting important improvements to acoustic tube modeling for the liquid sound /r/. A better understanding of the vocal tract articulation for liquid sound /r/ would be beneficial to clinical applications as well as advancement in speech technology such as speech therapy, speech and speaker verification systems, and speech synthesis technology.

2.7 TNM Classification of T4 Breast Cancer

Guth *et al.,* 2007, discovered that the presence of skin involvement in breast cancer results in the classification of the tumor into the highest tumor category and accordingly into the highest non-metastatic disease stage (current TNM classification: T4/stage III). This traditional view is no longer justifiable, as tumors that show non-inflammatory skin involvement (T4b) comprises of a considerably heterogeneous group with a high percentage of small-sized tumors.

Classifying all lesions demonstrating this feature together results in the combination of tumors with widely differing prognostic and therapeutic implications into a single group. This contradicts the basic principle of the TNM concept in that only tumors exhibiting similar extension and prognosis should be grouped into one category/stage.

Furthermore, the currently valid definitions of non-inflammatory skin involvement are misconceived for the substantial group of small tumors which often have ambiguous morphologic findings: the clinical classification depends on the subjective perception of the individual observer, and the pathologic staging considers histologic criteria that are not justifiable from a functional–morphological point of view. Consequently, it is highly suggested that there is a necessity to revise the current T4 category.

Guth recommend that breast carcinomas currently classified as T4a–c should be eliminated from the T4 category and classified simply according to their tumor size (T1–3). The prognostically very unfavorable inflammatory carcinoma (T4d) should be maintained as the only clinicopathologic entity in the T4 category. This proposal, which will also lead to a revision of the stage III group, adheres more closely to the goals and principles of the TNM classification than do the current classification guidelines. Through the revision of the T4 category, the definitions and guidelines of inflammatory breast carcinoma should be adapted to the internationally accepted nomenclature.

2.8 Computerised Segmentation of Breast Lesions

Chen *et al.,* 2006 stated in their paper that accurate quantification of the shape and extent of breast tumors has a vital role in nearly all applications of breast magnetic resonance imaging (MRI). Specifically, tumor segmentation is a key component in the computerised assessment of the likelihood of malignancy. However, manual delineation of lesions in four-dimensional magnetic resonance (MR) images is labor intensive and subject to interobserver and intraobserver variations. They developed a computerized lesion segmentation method that has the advantage of being automatic, efficient, and objective.

Their simulation was initiated by presenting a fuzzy c-means (FCM) clustering-based method for the segmentation of breast lesions in three dimensions from contrast-enhanced MR images. The proposed lesion segmentation algorithm consists of six consecutive stages: region of interest (ROI) selection by a human operator, lesion enhancement within the selected ROI, application of FCM on the enhanced ROI, binarization of the lesion membership map, connected-component labeling and object selection, and hole-filling on the selected object. We applied the algorithm to a clinical MR database consisting of 121 primary mass lesions. Manual segmentation of the lesions by an expert MR radiologist served as a reference in the evaluation of the computerized segmentation method. They also compared the proposed algorithm with a previously developed volume-growing (VG) method.

From the results obtained, it is observed that for the 121 mass lesions in their database, 97% of lesions were segmented correctly by means of the proposed FCM-based method at an overlap threshold of 0.4, whereas 84% of lesions were correctly segmented by means of the VG method. This summarises that their proposed algorithm for breast-lesion segmentation in dynamic contrast-enhanced MRI was shown to be effective and efficient.

2.9 Segmenting Tumors in Ultrasound Images

Parveen, 2008, stated that at the present breast cancer is the leading cause of death amongst women. Currently the effective method for early detection and screening of breast cancers is Ultrasonography. Unfortunately the high rate of false positives in mammography causes patients to undergo a large number of unnecessary biopsies.

Ultrasonography is an important counterpart to mammography in breast cancer detection. The accuracy rate of breast ultrasound can reach a high level in the diagnosis of simple benign cysts and reduce the number of false positives. This work helps to segment the breast tumors in ultrasound images. This will help in the computer-aided evaluation of the tumors and the distinction of benign and malignant nodules. The results extracted for the different features are coherent with the assessment of the specialists and represent a great help for the examination of the images and the decision making process.

2.10 Automatic Medical Ultrasound Strain Image Segmentation for Breast Tumors

Currently, the most popular underlying imaging modality used is ultrasound. While a variety of stress field schemes have been applied to create contrast, the images used for this project were created by exposing the tissue to quasi-static compression (McCormick *et al.*, 2005). This is implemented by compressing the tissue with ultrasound transducer. 'Quasi-static' denotes that the compression was applied very slowly.

The stress causes a change from the pre-compression image to the postcompression image. The local displacement of the object is determined by finding the lag of a region that corresponds to the maximum normalized cross correlation of that region. The gradient of the displacement is then taken as an estimation of local deformation; this makes a strain image.

An elastic modulus image displays a spatial map of tissue solid mechanical properties. A strain image is a result of the energy used to excite the object's response (the stress field) and the object's properties (the elastic modulus field). Assuming the stress field is uniform during quasi-static compression, the strain image is a rough estimate of the elastic modulus image.

The strain image reveals the same mechanical properties as would be elicited during manual palpation. For example, a lump felt in breast self-examination would show up as an area of contrast in a strain image. The mechanical properties of tissues are highly correlated with disease states, such as breast cancer.

2.11 Adaptive Finite Element Method for Image Processing

Bazan *et al.,* 2005, has developed an adaptive finite element strategy which is employed to solve the Perona-Malik model as modified by Catté, Lions, Morel and Coll for image processing by (often highly) nonlinear diffusion. FEMLAB® and MATLAB® are used to implement the experiments and they prove to be very suitable tools to run this type of problem. Refinement and coarsening of the grids are used as needed and the approach leads to unstructured grids where the efficiency of the remeshing strategy is demonstrated by obtaining very similar results as in the regular grid case, though with fewer unknowns.