

REGION-GROWING BASED SEGMENTATION AND BAG OF FEATURES
CLASSIFICATION FOR BREAST ULTRASOUND IMAGES

LEE LAY KHOON

Thesis submitted in fulfilment of the requirements
for the award of the degree of
Master of Computer Science

Faculty of Computer System and Software Engineering
UNIVERSITI MALAYSIA PAHANG

JANUARY 2017

ABSTRACT

A precise segmentation of medical image is an important stage in contouring throughout radiotherapy preparation. Medical images are mostly used in the hospital to assist doctor for patient's diagnosis and conduct treatment for patient. Ultrasound is one of the prominent tools used to detect breast tumor in the early stage. As the number of cases for breast cancer raises from year to year, segmentation play a vital role in the analysis of tumor. Tumor analysis usually has to be completed by very experience doctor or a lab test, where segmentation can help the surgeon to identify the location and the shape of tumor. Region growing method has been widely used to detect the presence of tumor in MRI (Magnetic Resonance) images and mammography, however there is not much research done on ultrasound segmentation by using region growing. Therefore, there appears to be a gap between the knowledge of region growing segmentation and ultrasound tumors segmentation. The purpose of this study is to investigate the modality and methodologies of segmentation and classification. This study aims to develop a scheme (algorithm) to segment and classify the type of tumor in ultrasound. The proposed scheme is consisting of three important stages, which is preprocessing, segmentation and classification. For the preprocessing stage, median filtering has been used to reduce the noise in ultrasound. In the next stage, which is the segmentation stage, region growing algorithm is used to automatically detect tumors in ultrasound images. After that, next stage, which is the classification stage, bag of feature (BoF). After segmentation done, the classification will take place when ultrasound is input. The algorithm has been utilized in the experiment to classify the type of tumor. Results show that, the region growing algorithm actually can works on the segmentation of ultrasound. To measure the result of algorithm developed, dice coefficient (DC) is the metric that is chosen to measure the accuracy of algorithm; Dice similarity coefficient (DSC) was used as a statistical validation metric to evaluate the performance of both the reproducibility of manual segmentations and the spatial overlap accuracy of automated probabilistic fractional segmentation of ultrasound images. Eventually a mean and standard deviation value of 0.949 ± 0.00147 is obtained as a result. Overall, a total of 116 ultrasound images have been used in the experiment where 43 are benign and 73 are malignant. Additional, result of accuracy 87.07% has been obtained from the classification experiment. Lastly, MIAS database (with total 322 images) has been included in the comparison section. By includes of MIAS database in the experiment allow a fair comparison with previous work. In conclusion, region growing segmentation and Bag of features classification able to perform well in ultrasound image.

ABSTRAK

Segmentasi di perubatan merupakan satu peringkat yang penting dalam pengkonturan sepanjang penyediaan radioterapi. Imej perubatan kebanyakannya digunakan di hospital untuk membantu doktor mengesan lokasi kanser. Selain itu, ia juga digunakan sebagai rujukan bagi pesakit menjalankan rawatan. Ultrasound adalah salah satu alat yang penting digunakan untuk mengesan tumor di payudara pada peringkat awal. Kes-kes kanser payudara meningkat dari tahun ke tahun, segmentasi memainkan peranan penting dalam analisis tumor. Analisis Tumor biasanya dianalisis oleh doktor yang sangat berpengalaman atau menjalankan ujian di makmal. Segmentasi boleh membantu pakar pembedah lebih memahami lokasi dan bentuk tumor. Kaedah “Region growing” makin digunakan untuk mengesan kehadiran tumor dalam MRI (Magnetic Resonance) imej dan mamografi. Bagaimanapun tidak banyak kajian yang dilakukan pada segmentasi ultrasound dengan menggunakan “Region-growing”. Tujuan tesi ini adalah untuk mengkaji modaliti dan kaedah segmentasi dan klasifikasi. Skim algorithm mempunyai tiga peringkat yang penting, pertama pra-pemprosesan, kedua adalah segmentasi dan ketiga adalah klasifikasi. Di peringkat pra pemprosesan, “Median Filtering” telah dilaksanakan untuk mengurangkan bunyi dalam ultrasound. Pada peringkat seterusnya iaitu peringkat segmentasi, algoritma berkembang rantau digunakan untuk mengesan tumor secara automatik di dalam imej ultrasound. Selepas itu, pada peringkat seterusnya iaitu peringkat klasifikasi, BOF algoritma telah digunakan dalam percubaan untuk mengklasifikasikan jenis tumor. Keputusan menunjukkan bahawa, algoritma rantau yang semakin meningkat sebenarnya boleh berfungsi pada pembahagian ultrasound. Untuk memeriksa keberkesanan algoritma maju, “Dice coefficient (DC) value” adalah metrik yang dipilih untuk mengukur ketepatan algoritma; akhirnya nilai sisihan min dan taraf $0,949 \pm 0,00147$ telah diperolehi dalam eksperisi. Secara keseluruhan, sebanyak 116 imej ultrasound telah digunakan dalam kajian ini di mana 43 adalah benigna dan 73 adalah malignan. Untuk prestasi klasifikasi, ketepatan 87,07% telah diperolehi daripada 116 imej ultrasound yang digunakan dalam eksperimen ini. Dari segi perbandingan di bahagian klasifikasi, MIAS pangkalan data (dengan jumlah 322 imej) telah ditambah dalam membuat kerja-kerja perbandingan dengan penyelidik lain. Kesimpulannya, penambahan peringkat pra pemprosesan yang (penapisan median) dalam algoritma meningkatkan prestasi segmen yang secara langsung akan memberi kesan kepada hasil nilai DC.