An improved quantitative measurement for thyroid cancer detection based on elastography

Jianrui Ding\textsuperscript{a}, H.D. Cheng\textsuperscript{a,b,\ast}, Jianhua Huang\textsuperscript{a}, Yingtao Zhang\textsuperscript{a}, Jiafeng Liu\textsuperscript{a}

\textsuperscript{a} School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, PR China
\textsuperscript{b} Department of Computer Science, Utah State University, Logan, UT 84322, USA

1. Introduction

Thyroid nodules are very common. With fine-needle aspiration cytology, 9.2–14.8\% of the nodules diagnosed in clinical examination is malignant [1]. One of the key features of thyroid gland cancer for palpation is the degree of firmness: malignant lesions tend to be much harder than benign ones [2].

Elastography is a newly developed technique to evaluate the stiffness of the lesion. Currently, there are two criteria to evaluate the color elastogram for thyroid nodule. The first one is a 4-pattern criterion [3]. The second one is a semi-quantitative method [4]. These methods are quite subjective and may cause error.

In our previous study, an effective and robust metric to analyze and evaluate the elastograms was proposed [5]. The experimental results showed that the proposed method had higher accuracy (93.6\%) than that of the color score (83.2\%) and strain ratio methods (87.2\%), and it could classify the nodules more accurately and effectively [5].

In this paper, the previous method is improved by further considering how to measure the stiffness of a lesion better and more effectively. The degree and the range of the hard area are used to describe the stiffness of a lesion. The experimental results confirm such property and show that the performance of the proposed method is higher than that of the existing methods.

2. Materials and methods

2.1. Image acquisition

There are 125 patients (98 female, mean age 46.31 ± 9.79 years; range 11–67 years; and 27 male, mean age 54.9 ± 11.7 years, range 29–81 years) were selected for this study from January to November, 2009. The inclusion criterion was the presence of single solid lesion in one thyroid lobe. The mean size of the nodules was 1.74 cm (range 0.77–2.64 cm).

For testing the performance of the method, the number of malignant cases was tried to balance with the number of benign cases to avoid bias. The FNA and surgical results were the reference standards. The final pathological results of the nodules are 56 malignant and 69 benign.

Among malignant nodules, there were 44 papillary carcinomas, 7 microcarcinomas and 4 microcarcinomas coexisted with nodular...
goiters, and 1 medullar carcinoma. Among benign nodules, there were 56 nodular goiters and 13 adenomas.

Both the conventional ultrasonography (US) and real-time elastography were performed with the HITACHI Vision 900 system (Hitachi Medical System, Tokyo, Japan) equipped with a liner probe with central frequency of 6–13 MHz.

All the examinations were conducted and recorded by two experienced sonographers who were blind to the history and pathologic results. Both of them have more than 6 years’ experience in scanning and about 1 month special training in acquiring elastograms. On average, 4 dynamic sequences (range 2–6) and 9 static images (range 6–11) were obtained for each nodule. Each dynamic sequence is at least 5 s long. Totally, 1025 static images and 512 dynamic views were obtained from the 125 nodules.

2.2. Elasticity information extraction

For convenience, the color elastogram is overlaid on the B-mode image translucently, hence, physicians can refer the B-mode image at the same time. Due to the mixture of the color elastogram with the B-mode image, it is difficult to get the real elasticity information from color elastogram directly. According to the principle of color elastography, elasticity information is encoded with 256 pseudo color levels from red to blue. The variation of the color reflects different elasticity magnitude. To extract elasticity information, the original color elastogram is transformed from RGB color space to HSV color space.

Because the elasticity information was encoded with color; for decoding it, only the color information is needed. Therefore, the hue component is extracted to represent the elasticity value. The following formulas are utilized to compute the hue value.

\[
\begin{align*}
\text{if } R \geq G \geq B & \quad H = 60^\circ \times \frac{G - B}{R - B} \\
\text{if } G \geq R \geq B & \quad H = 60^\circ \times \left(2 - \frac{R - B}{G - B}\right) \\
\text{if } G \geq B \geq R & \quad H = 60^\circ \times \left(2 - \frac{R - G}{B - G}\right) \\
\text{if } B \geq G \geq R & \quad H = 60^\circ \times \left(4 - \frac{R - G}{B - G}\right) \\
\text{if } B \geq R \geq G & \quad H = 60^\circ \times \left(4 - \frac{R - G}{B - G}\right) \\
\text{if } R \geq G \geq B & \quad H = 60^\circ \times \left(6 - \frac{R - G}{B - G}\right)
\end{align*}
\] (1)

The hue value and its corresponding color are shown in Fig. 1. Since the elasticity magnitude from soft to hard is encoded from red to blue, it needs to eliminate the overlap between red and blue from 300° to 360°. In this paper, we set the hue value between 300° and 360° to the lowest hue value 0°. For example, Fig. 2(a)

is a color elastogram, in the upper part of the image, the pixels marked red represent the lower elasticities of soft tissues (such as strap muscles), and their hue value should be low. (b) is the hue component of (a) obtained by using Eq. (1). Due to overlap, some red pixels in (a) have high hue values in (b) which result errors (some bright spots), and (c) is the hue component after overlap problem is detected and corrected. Then it is consistent with the color encoding mechanism in which red means soft and blue means hard.

2.3. Feature extraction

In general, a benign nodule often is relatively homogenous and colored light green, or its center is colored green and the periphery is colored blue; whereas a malignant nodule often is displayed in blue or colored in a mixture of blue and light green [3].

To characterize the elastogram, histogram features such as mean and variance were extracted and used to differentiate malignant lesions from benign lesions [7–12].

In this paper, the representative static images were selected from the dynamic elastogram sequences, and lesion regions on the B-mode images were delineated manually by an experienced radiologist. The contours of the lesion regions were mapped to the corresponding elastograms automatically, and the features were extracted from the corresponding lesion regions.

The stiffness of the lesion is an important feature to differentiate malignant nodule from benign nodule. But how to measure the stiffness of the nodule from elastogram is problematic. In this paper, two aspects are considered, one is the degree of hardness and the other is the hard area. The hard pixel is defined if its hue value is larger than a threshold. According to Fig. 1, the threshold is set 128. The hard region area is defined as the largest connective hard area in the lesion. The “8-connected” neighbors are considered, as shown in Fig. 3. The hard area ratio is the hard region area divided by the lesion area; it is a quantitative metric for comparing the color pattern score.

\[
\text{hard pixel set } = \{(x, y) \mid h(x, y) > T_{hard}\}
\]

where \(h(x, y)\) is the hue value of pixel \((x, y)\) and \(T_{hard}\) is the hardness threshold.

\[
\text{hard area ratio } = \frac{A_{hard}}{A_{lesion}}
\]

where \(A_{hard}\) is the largest connective hard area in lesion and \(A_{lesion}\) is the lesion area.

---

By setting threshold \( T \text{hard} \), the pixels on elastogram are classified into hard pixels and soft pixels. When \( T \text{hard} \) is decreasing, the number of hard pixels will increase and the hard area ratio will increase, it will make the benign nodule seem harder and could cause false positives. When \( T \text{hard} \) is increasing, the number of hard pixels will decrease and the hard area ratio will decrease, it will make the malignant nodule seem softer, and could cause false negatives. The optimal accuracy can be reached when an appropriate \( T \text{hard} \) is selected.

2.4. Thyroid nodule classification

For comparing with previous study, the same classifier—support vector machine (SVM), is employed. SVM is a widely used pattern classification tool. It aims at maximizing the margin between the separating hyper plane and the data to minimize the upper bound of generalization error. Unlike traditional methods which minimize the empirical training error, SVM can be regarded as an approximate implementation of the Structure Risk Minimization principle. Properties of condensing information in the training data and providing a sparse representation by using a small number of data points (Support Vectors) make SVM quite attractive [13].

The training samples are mapped to higher dimension space with a kernel function, and an optimal decision plane can be created [14]. There are two advantages of SVM: the generalization ability is optimal by maximizing the margin distance, and it can solve nonlinear tasks by mapping samples to higher dimension space [15].

3. Results

3.1. Feature extraction

Firstly, the radiologist drew the lesion in the conventional B-mode image, and it was mapped to the corresponding elastogram automatically. Secondarily, the color elastogram was transformed from RGB space to HSV space and the overlap in the hue is eliminated; then the hard area was extracted. Finally, the optimal hard threshold \( T \text{hard} \) was selected.

The patients were selected for this study and the number of malignant and benign cases was almost balanced. In the experiments, leave-one-out method was utilized: one image is for testing, and the others are used for training, the process is iterated until all the images have been used for testing. The error rates or predictive accuracies obtained in different iterations are averaged to yield the overall error rate or predictive accuracy.

Define the number of correctly and incorrectly classified malignant nodules as true positive (TP) and false negative (FN), and the number of correctly and incorrectly classified benign nodules as true negative (TN) and false positive (FP), respectively, the classification accuracy is defined as: \( \frac{(TP + TN)}{(TP + TN + FP + FN)} \).

In our previous study, the statistical and textural features of the lesion region were extracted and the hard area ratio was computed. After feature selection, the top five features were determined and their combinations are input to the SVM to classify the thyroid nodules. Finally, the optimal two features (hard area ratio and energy property of the co-occurrence matrix) were chosen according to the classification accuracy. The results show that the top 1 feature is the hard area ratio having the best discrimination power among all features, and it is very sensitive to malignant cases; therefore, it can be utilized as a quantitative metric for diagnosing malignant lesions.

In the previous study, the hard area was defined by summing all hard pixels in the lesion. Considering the stiffness of a lesion should be represented by the homogenous hard region, in this paper, the hard area is redefined as the largest connected hard area in the lesion.

The hard area ratio feature is input to the SVM classifier and the performance is listed in Table 1.

The result shows that the hard area ratio is a very useful discriminative feature; and it can classify most of the benign and malignant thyroid nodules accurately. The distribution of the hard area ratio is illustrated in Fig. 4.

As shown in Fig. 4, when we consider a case whose hard area ratio \( > 0.5 \) as malignant, then all malignant nodules can be accurately classified; however, when we consider a case whose hard area ratio \( > 0.55 \) as the malignant case, there will be three false negatives. It confirms the discriminative power of the hard area ratio.

3.2. Test with different \( T \text{hard} \)

The hard area ratio will change as \( T \text{hard} \) varies. When the appropriate \( T \text{hard} \) is selected, the best performance will be reached. The performance variation related to \( T \text{hard} \) is listed in Table 2.

<table>
<thead>
<tr>
<th>( T \text{hard} )</th>
<th>TP</th>
<th>FN</th>
<th>TN</th>
<th>FP</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>114</td>
<td>50</td>
<td>6</td>
<td>62</td>
<td>7</td>
<td>89.6</td>
</tr>
<tr>
<td>116</td>
<td>51</td>
<td>5</td>
<td>62</td>
<td>7</td>
<td>90.4</td>
</tr>
<tr>
<td>118</td>
<td>51</td>
<td>5</td>
<td>62</td>
<td>7</td>
<td>90.4</td>
</tr>
<tr>
<td>120</td>
<td>52</td>
<td>4</td>
<td>62</td>
<td>7</td>
<td>91.2</td>
</tr>
<tr>
<td>122</td>
<td>52</td>
<td>4</td>
<td>62</td>
<td>7</td>
<td>91.2</td>
</tr>
<tr>
<td>124</td>
<td>52</td>
<td>4</td>
<td>62</td>
<td>7</td>
<td>91.2</td>
</tr>
<tr>
<td>126</td>
<td>53</td>
<td>3</td>
<td>62</td>
<td>7</td>
<td>92</td>
</tr>
<tr>
<td>130</td>
<td>53</td>
<td>3</td>
<td>62</td>
<td>7</td>
<td>92</td>
</tr>
<tr>
<td>132</td>
<td>53</td>
<td>3</td>
<td>63</td>
<td>6</td>
<td>92.8</td>
</tr>
<tr>
<td>134</td>
<td>53</td>
<td>3</td>
<td>64</td>
<td>5</td>
<td>93.6</td>
</tr>
<tr>
<td>136</td>
<td>53</td>
<td>3</td>
<td>64</td>
<td>5</td>
<td>93.6</td>
</tr>
<tr>
<td>138</td>
<td>53</td>
<td>3</td>
<td>64</td>
<td>5</td>
<td>93.6</td>
</tr>
<tr>
<td>140</td>
<td>53</td>
<td>3</td>
<td>65</td>
<td>4</td>
<td>94.4</td>
</tr>
<tr>
<td>142</td>
<td>53</td>
<td>3</td>
<td>65</td>
<td>4</td>
<td>94.4</td>
</tr>
<tr>
<td>144</td>
<td>53</td>
<td>3</td>
<td>66</td>
<td>3</td>
<td>95.2</td>
</tr>
<tr>
<td>146</td>
<td>53</td>
<td>3</td>
<td>66</td>
<td>3</td>
<td>95.2</td>
</tr>
<tr>
<td>148</td>
<td>53</td>
<td>3</td>
<td>66</td>
<td>3</td>
<td>95.2</td>
</tr>
<tr>
<td>150</td>
<td>53</td>
<td>3</td>
<td>66</td>
<td>3</td>
<td>95.2</td>
</tr>
<tr>
<td>152</td>
<td>53</td>
<td>3</td>
<td>66</td>
<td>3</td>
<td>95.2</td>
</tr>
<tr>
<td>154</td>
<td>52</td>
<td>4</td>
<td>65</td>
<td>4</td>
<td>93.6</td>
</tr>
<tr>
<td>156</td>
<td>49</td>
<td>7</td>
<td>65</td>
<td>4</td>
<td>91.2</td>
</tr>
<tr>
<td>158</td>
<td>46</td>
<td>10</td>
<td>64</td>
<td>5</td>
<td>88</td>
</tr>
<tr>
<td>160</td>
<td>45</td>
<td>11</td>
<td>64</td>
<td>5</td>
<td>87.2</td>
</tr>
</tbody>
</table>

The result confirms that the performance is affected by $T_{\text{hard}}$ value. According to Fig. 1, the color appears bluer when hue value is larger than 180°; i.e., the pixel appears harder than average elasticity. Utilizing hue value 180° as the hard threshold is a natural choice, and when the hue value is mapped to [0,255], using 128 as the hard threshold is quite reasonable. When the hard threshold increases, the hard area ratio will decrease; and the false positives will be reduced. In Table 2, when $T_{\text{hard}}$ is varying from 128 to 138, the number of FPs is decreasing from 7 to 5. When the hard threshold increases more, the performance is improved further. When $T_{\text{hard}}$ is increased from 144 to 152, the maximum accuracy is reached. The threshold range reveals that it will tolerate noise and has robust characteristics. As the hard threshold increases further, the hard area ratio will decrease more, and the difference between benign and malignant cases is reduced. It makes the accuracy dropped. When $T_{\text{hard}}$ increases from 154 to 160, the accuracy is dropped from 93.6% to 87.2%. The relation between hard threshold and classification accuracy is illustrated in Fig. 5.

Fig. 5 reveals that there exists a hard threshold range that can maximize the classification performance. In our experiments, such a range is from 144 to 152. When hard threshold is equal to 148, the distribution of hard area ratio is illustrated in Fig. 6.

Compared to Fig. 4, as $T_{\text{hard}}$ increases from 128 to 148, the distribution of each class is more condensed, and the distributions of two classes are more separate. Therefore, the classification accuracy is improved.

3.3. Comparing with the strain ratio calculated by HIVISION 900

The strain ratio was defined by comparing the mean elasticity in a surrounding normal tissue region with the mean elasticity in the lesion. The radiologist drew the lesion region and a surrounding normal tissue region on the elastogram, and the machine calculated the strain ratio. For eliminating operation error, the strain ratios of the three images of the same case were averaged as the strain ratio of the case. The value was used to classify the nodules into malignant and benign, and the performance was compared with that by using the hard area ratio as shown in Table 3.

The result shows that the strain ratio method has lower accuracy than that of the proposed method. This may be due to the error introduced by the selection of the lesion and normal tissue region subjectively.
3.4. Comparing with the color score method

For comparing with the color score method, three radiologists who were blind to the pathological results were invited to analyze the cases independently after the images were collected. All of them have more than 6 years’ experience of thyroid US examination. They were asked to evaluate the elastograms and score every lesion according to the color distribution [3]. Score 1 is for the nodules relatively homogenous and colored light green; score 2 is for the nodules whose centers colored green and their peripheries colored blue; score 3 for nodules colored the mixture of light green and red; and score 4 for nodules completely colored in blue.

The result shows that the mean scores for benign and malignant nodules are different (2.67 ± 0.86 vs. 3.86 ± 0.43, p < 0.001).
The classification result is listed in Table 4. The results confirm that the quantitative metric proposed in this paper has more discrimination power than that of the color score method.

4. Discussion

By further analyzing, the error cases can be classified into three groups. The first group contains cystic nodules which should be excluded, since they are mostly composed of fluid, and elastography cannot provide useful information [16]. An example is illustrated in Fig. 7, which produces a false positive. The second group contains the cases which are difficult to diagnose even for radiologists. An example is illustrated in Fig. 8, which is a false positive. The third group contains the cases which are hard to diagnose by using elastogram only, however, they can be correctly diagnosed by combining with the corresponding B-mode images. An example is illustrated in Fig. 9, which produces a false negative.

The results show that the elastography is an inspiring new modality in clinical practice, however, the B-mode features should also be integrated. The accuracy can be further improved by combining the features of elastograms and B-mode images. Since the quality of elastograms is highly dependent on operator’s skill, more effective and accurate criteria for elastogram evaluation should be further studied.

5. Conclusions

Real-time elastography is a newly developed medical imaging technique which measures the tissue biomechanics properties. But the evaluation criteria of elastography such as color score and strain ratio are quite subjective, and they are highly dependent on the radiologist’s experience.

The method proposed in this paper is a new quantitative metric for elastogram. The experimental results demonstrate that the proposed method is very predictive to the thyroid tumors. It is an objective evaluation of the elastograms and has higher accuracy than that of the color score and strain ratio metrics.

Conflict of interest

The authors have no conflicts of interest.

Acknowledgments

This work is supported, in part, by National Science Foundation of China; the Grant numbers are 60873142 and 30670546.

References
