Detection of Prostate Abnormality within the Peripheral Zone Using Grey Level Distribution

Andrik Rampun¹, Paul Malcolm² and Reyer Zwiggelaar¹

¹Department of Computer Science, Aberystwyth University, Aberystwyth, UK
²Department of Radiology, Norfolk & Norwich University Hospital, Norwich, UK

Email:{yar,rrz}@aber.ac.uk , {paul.malcolm}@nnuh.nhs.uk

Abstract—Development of CAD systems for detection of prostate cancer has been a recent topic of research and remains a challenging task. In this paper, we propose a novel method of prostate cancer detection within the peripheral zone. The key idea is to assume that every grey level could be associated with malignant or normal tissues by using a weighted probability. Based on the weighting, we use specific metrics to determine abnormality. We show experimental results to illustrate the performance of this method in comparison to some previous studies. Initial results show that our method achieved 81% correct classification result and 9% and 10% false positive and false negative results, respectively (sensitivity specificity: 0.85/0.72).

Index Terms—Prostate Cancer Detection, Computer Aided Detection, Histogram Analysis, Grey Level Distribution, Prostate Abnormality

I. INTRODUCTION

With an estimate of 1.7 million cases globally by 2030 [1], prostate cancer is one of the most common cancers affecting men, and is a leading cause of mortality in males [2]. In 2013, there were approximately 280,000 and 40,000 cases in the United State and United Kingdom, respectively [1]. Clinical diagnostic tools such as prostate-specific antigen (PSA) level, digital rectal examination (DRE), transrectual ultrasound (TRUS) and biopsy tests are used globally despite their inconsistent results (0.51-0.89 and 0.67-0.87 sensitivity and specificity, respectively) [3]. Some of these methods are invasive and patients can suffer stress from false-positive test results, subsequent evaluation, therapy, and "over-treatment” [2].

Prostate magnetic resonance imaging (MRI) has the potential to improve the accuracy of clinical diagnostic tests [4]. Therefore the main goal of our research is to develop computer aided diagnosis (CAD) tools towards the detection of prostate abnormalities. Since 80-85% of the cancers arise within the peripheral zone [5], the proposed method only considered abnormalities that occur within this region. There have been several studies which used only the peripheral zone of the prostate [6-8] and allows us to compare our quantitative results to these previous studies. Fig. 1 shows an example of prostate

Figure 1. The ground truth of prostate gland, central zone and tumor and represented in red, yellow and green, respectively.

MRI image with its ground truth delineated by an expert radiologist and Fig. 2 shows a schematic overview of a prostate containing a tumor.

Figure 2. CZ = central zone, PZ = peripheral zone, TZ=transitional zone T = tumor.

The proposed method compares grey level histograms from each slice with models (e.g. normal and malignant histograms models) which were constructed based on training data the distribution of grey level. Subsequently, we use specific metrics to determine abnormality. There are several methods in the literature which have used histogram analysis techniques [9-12]. A method proposed in [9] characterised each suspicious region of interest by performing histogram analysis on multiparametric MR images. On the other hand, [10] used colour channel histograms to capture the pattern of malignancy tissues in Gleason graded images. Moreover, [11] reports that features based histograms can achieve high sensitivity and specificity which is similar to some other features such as Gray Level Co-occurrence matrix (GLCM) and Grey Level Run-Length matrix (GLRLM). Finally, [12] used T2-weighted signal intensity histogram skewness as one of their features in differentiating between malignant and normal tissues. In terms of using grey level properties in detecting abnormalities, several studies have been done in medical image analysis application. The authors in [13] proposed a method for recognition of lung abnormalities based on four different templates describing typical geometry and grey level distribution of lung modules.
[13]. On the other hand, Yu [14] proposed a method which is relying on spatial grey level dependence for detecting and locating brain abnormality. Finally, the authors in [15] proposed a method using the distribution of grey level variations for abnormality detection in mammography.

In this paper we propose a new method for detecting prostate cancer within the peripheral zone using the distribution of grey levels. In comparison to the existing methods (detecting prostate cancer in medical images) in the literature, our method is different in the sense that we do not rely on texture features such as blobs and statistical features. The methods in [16], [17] and [18] used texture features to classify malignant and normal tissue. Secondly, our method uses a single modality (T2-Weighted MRI) unlike the method in [19], which used multimodality, i.e. diffusion MRI and MR Spectroscopy. Similarly, the method proposed in [9] used multiparametric MR, i.e. T1- and T2-weighted imaging. Finally, the proposed method is purely based on the grey levels information whereas the method in [9] used additional clinical diagnostic information such as biopsy tests in deciding whether cancer is truly present or not.

II. MODELLING THE PERIPHERAL ZONE

Since the proposed method only considers malignant tissues/regions within the PZ, it is important to define the PZ region. We use a simple method proposed in [20, 21] which uses a quadratic equation \( y = ax^2 + bx + c \) to define the boundary of PZ based on three vertex coordinates \( v_1, v_2 \) and \( v_3 \) (which can be found based on the prostate’s boundary). Fig. 3 shows an example how the PZ region is defined. The coordinates of \( v_1, v_2 \) and \( v_3 \) can be determined based on the \( Cp \) (central coordinate), minimum and maximum \( x \) and \( y \) coordinates \((x_{\text{min}}, y_{\text{min}}, x_{\text{max}} \text{ and } y_{\text{max}})\). A detailed description of the method can be found in [20, 21].

![Figure 3: Prostate gland (red) and the defined PZ (indigo region) and its boundary (green) which goes through \( v_1, v_2 \) and \( v_3 \).](image)

III. METHODOLOGY

Fig. 4 shows the overview of the proposed algorithm. In summary, the methodology relies on the distribution of grey level within the histograms which were constructed based on the grey level occurrence within 44 malignant and normal regions. The histogram models represent the weighting values for every grey level. Based on the constructed histograms (88 slices) for each new slice, the histogram will be compared with the histogram models constructed in the early phase (yellow region in Fig. 4).

![Figure 4: The overview of the proposed methodology](image)

A. Construction of Histograms

For every slice, the proposed method constructs the first histogram \( (H_m) \) by taking every pixel intensity within the malignant region (note that all malignant regions were delineated by an expert radiologist) and each pixel is classified according to its intensity into an appropriate grey level. The second histogram \( (H_b) \) is constructed using the same process with \( H_m \) but taking normal regions instead. Normal regions are taken from the whole PZ region (with condition there is no tumors found within the PZ) as shown in Fig. 3 indigo region. This means \( H_m \) and \( H_b \) contain the distribution of malignant and normal grey level, respectively. Since every image size is 16 bits, there is a total of 65,536 grey. There were 44 malignant and normal regions (in total 88 regions) taken from 44 slices (in total 88 slices all from 20 patients) to construct \( H_m \) and \( H_b \).

Mathematically, this process can be presented in (1), (2), (3) and (4).

\[
\begin{align*}
    m_n & = \{i_1, i_2, i_3 \ldots i_n\} \quad \text{(1)} \\
    b_n & = \{i_1, i_2, i_3 \ldots i_n\} \\
    H_m & = \{m_1 + m_2 + \ldots + m_n\} \quad \text{(3)} \\
    H_b & = \{b_1 + b_2 + b_3 + \ldots + b_n\} \\
\end{align*}
\]

where \( n, i, m \) and \( b \) represent the \( n^{th} \) region (or grey level), every grey level, malignant and normal region, respectively. This means, there are 44 histograms in total from 44 malignant slices (another 44 histograms from normal slices). Finally, we combine all histograms using
(3) and (4) to produce the resulting models. Fig. 5 illustrates an example of this process.

![Image](image1.png)

**Figure 5**: An example of histogram result ($H_m$). The x and y axis represent the grey level location and the frequency of occurrence for every grey level.

Fig. 5 shows an example of the construction of $H_m$ based the $n$ slices (in our case 44 slices). Note that the pixel intensities are only taken within the malignant region which is within the red boundary in Fig. 5. The process is similar when constructing $H_b$ but taking the PZ region as defined in Fig. 3. To this point we have constructed two histogram models which are $H_m$ and $H_b$ where both histograms store the frequency of grey level occurrence from malignant and normal regions, respectively. However, since many grey levels occur in both regions, we need to identify the distinctive grey levels which occur only in malignant or normal regions, and grey levels that occur both in malignant and normal regions. These allow us to differentiate malignant and normal grey tissues. These can be found by finding the intersection of $H_m$ and $H_b$ as shown in (5). In our case, intersection means a particular grey level can be found in both malignant and normal regions (i.e. grey level at the position 3990th occurs in $H_m$ and $H_b$, see Fig. 9).

$$H_o = H_m \cap H_b$$  \(5\)

where $H_o$ is the histogram contains all the grey levels occur both in malignant and normal regions. Each grey level’s frequency in $H_o$ is the mean of corresponding frequencies in $H_m$ and $H_b$ (e.g. $H_o(i) = (H_m(i) + H_b(i))/2$) with condition that the element $i$ is in $H_m$ and $H_b$. Fig. 6 shows an illustration of this process in a Van diagram. From the van diagram, we can see that $H_m$ and $H_b$, have their unique grey levels. The values in $H_o$ represents grey levels 3, 50, 234 and 941 occur in both $H_m$ and $H_b$. This means, prostate tissues which are fall in these grey level values are classified as malignant or normal.

![Image](image2.png)

**Figure 6**: $H_o$ contains all the overlap grey levels from $H_m$ and $H_b$. Each element in the Van diagram represents the grey level.

By the end of this phase, we have three histogram models which present the only malignant ($H_m$), only normal ($H_b$) and malignant and normal ($H_o$) grey levels. Fig. 7 and 8 show examples of malignant and normal grey level distributions from grey level 38,000 to 39,000, respectively. We chose only 1000 grey levels in the following examples so the differences are visible.

![Image](image3.png)

**Figure 7**: Malignant grey level distribution for 1000 grey levels.

![Image](image4.png)

**Figure 8**: Normal grey level distribution for 1000 grey levels

On the other hand, Fig. 9 shows an example of grey level distribution of $H_o$ for the same 1000 grey levels.
Based on the examples in Fig. 7-9 we can clearly see that normal regions have higher occurrence within this range and very small occurrence for malignant grey levels within this range. However, it is possible that all grey levels in malignant regions only occur in $H_o$ (e.g. if all grey levels occur in $H_o$ we need further steps to determine abnormality). To reduce this problem we need to extract two more histograms from $H_o$. Both histograms ($H_t$ and $H_u$) represent the distribution of overlapping malignant and benign grey levels, respectively. In contrast to $H_o$, $H_t$ and $H_u$ were constructed based on the exact number of occurrence of grey levels whereas $H_o$ is based on the average occurrence number of grey levels of $H_m$ and $H_b$. Therefore, every grey level now has a weighted value and using these histogram models ($H_m$, $H_b$, $H_o$, $H_t$ and $H_u$) we could estimate the probability of abnormality using specific metrics.

B. Histogram Normalisation

Since the sum of $H_o$ outweighs considerably high the sum of $H_m$ and $H_b$ (caused false negative results), all models need to be normalised to ensure the weighting values for all grey levels are distributed evenly (used for the first metric only). By normalising all histogram models the sum for every histogram is equal to 1. The normalisation can be done using the following equation

$$h_k(i) = \frac{H_k(i)}{\sum H_k(i)}$$

where $k \in \{m, n, o, t, u\}$ and $h$ (small $h$) indicates a normalised histogram. Therefore we get an equal summation of histogram but still different distributions and weighting value for every grey level’s location.

C. Abnormality Detection

In the proposed method, abnormality detection is performed by calculating specific information from every histogram which is extracted from PZ. We use the following metrics to measure abnormality.

1) The sum of histogram multiplication ($H_{j \times k}$) for each $H_j$ with each of $h_k$ (e.g. $h_m$ and $h_b$). This can be calculated using the following equation

$$H_{j \times k} = \sum_{i=1}^{\text{size}} H_j(i) \times h_k(i)$$

This metric indicates the product of probability when every element in $H_j$ is multiplied with every element (frequency) in $h_k$. This means, the higher the value of $H_{j \times k}$ the more chance the slice has a tumor. For instance, $H_{j \times m} > H_{j \times b}$ means the malignant product of probability is higher when multiplied with $h_m$ compared to $h_b$. A large number of $H_{j \times m}$ indicates there are many grey levels in $H_j$ have the same grey level in $h_m$ (similar case in $H_{j \times b}$ and $H_{j \times o}$).

2) The histogram intersection ($H_{j \times k}$) [22] between $h_j$ and $h_k$ which can be calculated using the following

$$H_{j \times k} = \sum_{i=1}^{\text{size}} \min\{H_j(i), H_k(i)\}$$

Based on (8), we do not use the normalised histograms ($h_k$) but $H_k$ (denormalised histogram) instead, because normalised histograms could affect the value of histogram intersection due to the small values in $h_k$ which lead to incorrect results. We used this metrics as it has been successfully used for similarity measure in many different applications including medical image analysis. This metric measures the closeness of match between two histograms (in our case $H_j$ and $H_k$). Higher value of $H_{j \times k}$ indicates higher probability of the histograms is similar. In the proposed method, this metric is used to measure the closeness of a match between every $H_j$ with the histogram models. The closer the match with model the higher the value of $H_{j \times k}$. Finally, higher value means most grey levels in both histograms are distributed equally.
Finally, now we have seven variables ($H_{jkm}$, $H_{jnt}$, $H_{jkt}$, $H_{jku}$, $H_{jhx}$ and $H_{jko}$) which will be used to determine as whether abnormality is present or not based on the decision rules in Fig. 10. It should be noted that there are other metrics in the literature [23] available but these will be investigated in the future. Based on Fig. 10, it is clearly shows that if $H_{jkm}$, $H_{jnt}$ or $H_{jhx}$ value is higher than $H_{jkt}$, $H_{jku}$ or $H_{jko}$ indicates higher probability of the prostate being abnormal. Similarly when the prostate’s slice is more likely to be benign, the values of $H_{jkm}$, $H_{jnt}$ or $H_{jhx}$ are greater than $H_{jkm}$, $H_{jnt}$ or $H_{jhx}$ in the next section we will present our experimental results.

IV. EXPERIMENTAL DATA

In this study, we used 243 slices of T2-Weighted MRI images taken 35 different patients aged 47 to 79. The data contains 88 slices of training data from 20 patients and 50% of the training data are malignant and the other half is normal. For evaluation purpose we used 155 slices of MRI images from 35 patients and 105 slices of them are malignant and 50 slices are normal. Our data were collected from Norfolk and Norwich Hospital University and for every slice all ground truths (prostate gland, central zone and tumor) were delineated by an expert radiologist.

V. EXPERIMENTAL RESULTS

This section presents the experimental results based on 155 slices T2-Weighted MRI images with 105 slices are malignant and 50 slices are normal from 35 different patients aged 49 to 74. The prostate, cancer and central zones were delineated by an expert radiologist on each of the images. Each slice was analysed and classified as to whether the prostate contains abnormality based on the methodology described in section three. Next, we compared the result with the ground truth as to whether the prostate contains cancer regions or not. An abnormality is considered to have been detected if the classification result is correct in comparison with the ground truth. We use several quantitative measures to evaluate the results such as sensitivity (Sen), specificity (Spe) and Accuracy (Acc). Each of these metrics can be calculated using the following equations

\[
Sen = \frac{TP}{TP+FN} \tag{9}
\]

\[
Spe = \frac{TN}{TN+FP} \tag{10}
\]

\[
Acc = \frac{TN+TP}{TN+TP+FP+FN} \tag{11}
\]

where $TP$ and $FP$ denote the numbers of true positive and false positive, respectively. Similarly, $TN$ and $FN$ show the numbers of true negatives and false negatives. Accuracy means the number of correct classified slices out of the total number of slices. Sensitivity measures the proportion of actual positives which are correctly identified (in this case the percentage of malignant slice which are correctly identified) whereas specificity measures the proportion of actual negatives which are correctly identified (in this study the percentage of normal slice which are correctly identified).

The proposed method achieved 81% correct accuracy, which means 126 slices were classified correctly with 0.85 and 0.72 sensitivity and specificity, respectively. On the other hand, the proposed method produced 9% and 10% false positive and false negative results, respectively. In comparison with existing methods in the literature the proposed method achieved similar results. Nevertheless, it is extremely difficult to make a qualitative comparison due to the differences in datasets (different modalities such as T2-weighted (T2-W) MRI, diffusion-weighted (DWI) MRI, dynamic contrast enhanced (DCE) MRI, Magnetic resonance spectroscopy (MRS), etc.) and frameworks used by the other methods. However, to compare the proposed method, we cite several methods which have similar goals (detecting prostate cancer). There are many other methods in the literature but it is difficult to gather all of those methods (also space limitation) and we selected these methods (see Table 1) because they have at least one of the qualitative results (e.g. sensitivity) and it is clearly stated the number of cases used in the evaluation.
achieved the highest accuracy but based on DCE and T2-W+DCE modalities, respectively. Therefore it is extremely difficult to make a direct comparison either quantitatively or qualitatively. However, for indirect comparison purpose our method achieved comparable results with the state of art. One obvious drawback of this method is since it entirely relying on the grey level distribution, the metrics could give inaccurate results if the data is heavily affected by noise (due to change of pixel intensity affected by noise).

VI. CONCLUSIONS

We have introduced a novel method for automated prostate cancer detection using grey level distribution. The proposed method achieved similar results to some of the methods in the literature. The proposed method shows that prostate abnormalities could be detected using grey level distribution by giving a weighted value (by normalising the histogram models) for each of the grey levels. Moreover, in this paper we have shown the importance of grey-level values in detecting prostate abnormality by assigning every grey level into different classes (e.g. malignant or benign) and assigning a weighting value for every single grey level’s location. In short, with 9% and 10% false positive and false negative results, respectively we have achieved comparable results (81% accuracy out of 35 patients). Although it is difficult to make a quantitative comparison with the methods in Table 1 due the differences in datasets and frameworks, the main objective to show the potential of grey level distribution in detecting prostate cancer because it has been showed its potential in different human’s body such as breast [15], lung [13] and brain [14]. Finally, the next stage of this research is to test it on a larger dataset with several combination methods [20, 21] and applying a robust noise reduction method to improve its sensitivity and specificity.

ACKNOWLEDGMENT

Andrik Rampun would like to thank for awards given by Aberystwyth University under the Departmental Overseas Scholarship (DOS) and Doctoral Career Development Scholarships (DCDCS).

REFERENCES


<table>
<thead>
<tr>
<th>Authors</th>
<th>#</th>
<th>Acc</th>
<th>Sen</th>
<th>Spe</th>
<th>Mod</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sung et al.[24]</td>
<td>42</td>
<td>89</td>
<td>89</td>
<td>89</td>
<td>DCE</td>
</tr>
<tr>
<td>Vos et al.[9]</td>
<td>29</td>
<td>89</td>
<td>-</td>
<td>-</td>
<td>T2-W</td>
</tr>
<tr>
<td>Ampeliotis at al.[3]</td>
<td>10</td>
<td>87</td>
<td>-</td>
<td>-</td>
<td>T2-W+DCE</td>
</tr>
<tr>
<td>Rampun et al.[21]</td>
<td>19</td>
<td>85</td>
<td>82</td>
<td>87</td>
<td>T2-W</td>
</tr>
<tr>
<td>Tiwari et al.[25]</td>
<td>19</td>
<td>84</td>
<td>-</td>
<td>-</td>
<td>T2-W+MRS</td>
</tr>
<tr>
<td>Artan and Yetik.[8]</td>
<td>15</td>
<td>82</td>
<td>76</td>
<td>86</td>
<td>DCE</td>
</tr>
<tr>
<td>Our method</td>
<td>35</td>
<td>81</td>
<td>85</td>
<td>72</td>
<td>T2-W</td>
</tr>
<tr>
<td>Castaneda et al.[26]</td>
<td>15</td>
<td>80</td>
<td>67</td>
<td>86</td>
<td>CrW</td>
</tr>
<tr>
<td>Reinsberg et al.[27]</td>
<td>42</td>
<td>-</td>
<td>81-93</td>
<td>64-73</td>
<td>DWI+MRS</td>
</tr>
<tr>
<td>Litjens et al.[28]</td>
<td>188</td>
<td>-</td>
<td>84</td>
<td>-</td>
<td>DWI+DCE</td>
</tr>
<tr>
<td>Futterer et al.[29]</td>
<td>6</td>
<td>-</td>
<td>83</td>
<td>83</td>
<td>T2-W</td>
</tr>
<tr>
<td>Giroun et al.[30]</td>
<td>46</td>
<td>-</td>
<td>78-81</td>
<td>32-56</td>
<td>T2-W</td>
</tr>
<tr>
<td>Llobet et al.[18]</td>
<td>303</td>
<td>-</td>
<td>57</td>
<td>61</td>
<td>Ultrasound</td>
</tr>
</tbody>
</table>

Table 1 presents the experimental results of thirteen different methods including the proposed method and their accuracy, sensitivity, specificity and modalities. All methods were ordered based on accuracy, sensitivity and specificity accordingly. Note that some of the authors did not include one/two of these qualitative results (indicated as ‘-‘). The method proposed in [9] and [24] achieved the highest correct classification rate (89%) followed by the method in [3] with 87% accuracy. Our method has similar accuracy result with the method in [26] with just 1% higher. In addition, the proposed method reported similar sensitivity with the methods proposed in [21], [28] and [29] and the method of [24] reminds 89%. Although the method proposed in [27] achieved the highest sensitivity but the authors reported inconsistency of 81%-93%. The method in [18] achieved the lowest accuracy which is 57% whereas the method in [8] and [26] produced 76% and 67% sensitivity, respectively. On the other hand, the method in [21], [26], [8] and [29] achieved high specificity 87%, 86%, 86% and 83%, respectively. The proposed method achieved only 72% but it is still higher than the methods proposed in [18] and [30].

These comparisons are subjective because accuracy, sensitivity and specificity are highly influenced by the number of datasets, different modalities and methods’ framework. For instance, although the method proposed in [26] and [29] achieved 86% and 83% specificity, respectively; the evaluation is based on smaller numbers of dataset (6 and 15 patients, respectively). On another study [31] shows higher sensitivity and specificity of 93% and 96%, respectively but based on 46 ultrasound images. Similarly, the method proposed in [24] and [9] has achieved the highest accuracy but based on DCE and T2-W+DCE modalities, respectively. Therefore it is extremely difficult to make a direct comparison either quantitatively or qualitatively. However, for indirect comparison purpose our method achieved comparable results with the state of art. One obvious drawback of this method is since it entirely relying on the grey level distribution, the metrics could give inaccurate results if the data is heavily affected by noise (due to change of pixel intensity affected by noise).


[22] https://sites.google.com/site/dashiwuzhuka/toolkits/overlapping.histograms.


Andrik Rampun received his B.Sc. and M.Sc. in Software Engineering from MARA University of Technology, Malaysia and Hertfordshire University, UK in 2008 and 2011, respectively. He is currently a second year PhD student at the Department of Computer Science, Aberystwyth University, UK. His current research interests are Prostate Imaging particularly in abnormality detection and localisation as well as Texture Based Segmentation. On the other hand, he is also interested in e-learning on social networks.

Paul Malcolm received a BSc in Biochemistry and basic medical sciences in 1983 from Kings College, London and qualified in Medicine from the University of London in 1986. He is a fellow of the Royal College of Radiologists and currently a consultant radiologist and research lead in radiology at the Norfolk & Norwich University Hospital. His clinical interests include imaging of prostate cancer and his research interests are in new MRI techniques for measurement of body compartments and function.

Reyer Zwiggelaar received the Ir. Degree in Applied Physics from the State University Groningen, Groningen, The Netherlands, in 1989, and the Ph.D. Degree in Electronic and Electrical Engineering from the University College London, London, UK, in 1993. He is currently a Professor at the Department of Computer Science, Aberystwyth University, UK. He is the author or co-author of more than 180 conference and journal papers. His current research interests include Medical Image Understanding, especially Focusing on Mammographic and Prostate Data, Pattern Recognition, Statistical Methods, Texture-Based Segmentation, and Feature-Detection Techniques.