34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

CROSS-DEVICE AUTOMATED PROSTATE CANCER LOCALIZATION WITH MULTIPARAMETRIC MRI

Yusuf Artan[1], *Aytekin Oto*[2], *Imam Samil Yetik*[1]

[1] Medical Imaging Research Center, Illinois Institute of Technology, Chicago, IL, USA
[2] Department of Radiology, University of Chicago, Chicago, IL, USA

ABSTRACT

Automated cancer localization with supervised techniques plays an important role in guiding biopsy, surgery and treatment. It is crucial to have an accurate training dataset for supervised techniques. Since different devices with e.g. different protocols and/or field strengths cause different intensity profiles, each device/protocol must have an accompanying training dataset which is very costly to obtain. In this paper, we propose a novel method that has the ability to design classifiers obtained from one imaging protocol and/or MRI device to be used on a dataset from another protocol and/or imaging device. As an example problem we consider prostate cancer localization with multiparametric MRI. We show that simple normalization techniques such as z-score are not sufficient to allow for cross-device automated cancer localization. On the other hand, the methods we have originally developed based on relative intensity allows us to successfully use a classifier obtained from one device to be applied on a test patient imaged with another device.

Index Terms— Intensity Normalization, Magnetic Resonance Imaging (MRI), Prostate Cancer, Linear Discriminant Analysis (LDA)

1. INTRODUCTION

Automated cancer localization with supervised techniques plays an important role in guiding biopsy, surgery and treatment. In the past, many detection and localization methods for a broad range of cancer types have been developed using supervised classification techniques. It is crucial for supervised methods to have accurate training data, otherwise constructed classifier will result in high error rates. Different devices with e.g. different protocols and/or field strengths cause different intensity profiles; therefore, each device/protocol must have an accompanying training dataset which is very costly to obtain. Therefore it would be a significant advantage if we could apply a classifier obtained from device A on test data obtained from device B. Cross-device/protocol training and testing has never been successfully implemented in the past. Unfortunately, simple normalization techniques such as z-score are not sufficient to allow for cross-device automated cancer localization. In this study, we propose a novel relative intensity based method that has the ability to design classifiers obtained from one imaging protocol and/or MRI device to be used on a test patient imaged with another protocol and/or imaging device.

As an example problem we consider prostate cancer localization with multiparametric MRI. Prostate cancer is one of the most prevalent and leading causes of cancer related death for men in the United States [1]. Several fully automated localization methods have been developed in the past [2]-[5]. However, none of these earlier proposed techniques has considered cross device-protocol training. We develop a novel method based on relative intensities to solve the problem of cross-device training. To evaluate the proposed method, we use two datasets collected from a 1.5-T Excite, GE and a 1.5-T Philips Healthcare scanners at the University of Chicago [1].

Figure 1 presents the motivation for our cross-device automated classification method. Figure 1.(A)-(B) show the classifier trained using tumor (green) and normal (red) pixels of several patients for Philips and GE devices, respectively. Figure 1.(C)-(D) shows the z-score normalized Philips data with GE classifier (classifier from part (B)) and GE data with Philips classifier (classifier from part (A)). From parts (C) and (D), we notice that if we directly apply classifier developed for Philips machine to GE, or vice versa, it results in high error rates even though z-score normalization is used for individual devices. Reducing this error rate and thereby allowing cross-device training is the overall objective of this paper. Figure 1.(E) demonstrates the Philips data with GE classifier after the proposed relative intensity transformation is applied. Similarly, Figure 1.(F) shows the GE data with Philips classifier after relative intensity transformation is applied. Parts (E)-(F) show that it is possible to design an accurate classifier from one MRI device to be used on another MRI device using the proposed method.

Relative intensity as described in Section 2.2 can be computed efficiently for training data, but can not be calculated directly for testing data without the knowledge of the labels of tumor and normal. Therefore, we propose an iterative algorithm to estimate the relative intensity based on the current estimate of the class memberships. In this study, we use linear discriminant analysis (LDA) for localization. Note that the proposed segmentation method can also be implemented with other popular classifier such as support vector machines, perceptrons, or neural networks [6].

The rest of the paper is organized as follows. In Section 2, we briefly describe z-score normalization, discriminant analysis, and the proposed method. Section 3 presents the quantitative, qualitative results for the z-score normalization and the proposed methods. Finally, conclusions are presented in Section 4.

2. METHODOLOGY

In this section, we first describe z-score normalization and its shortcomings in cross-device segmentation applications. Next, we explain the proposed relative intensity method, and also briefly describe linear discriminant analysis used for classification.

2.1. Z-score normalization

Z-score normalization has been ubiquitously used in medical imaging applications [7]. This technique brings intensities of different types of MR images within the same dynamic range, improving the stability of the classifiers. It is applied such that intensities have zero

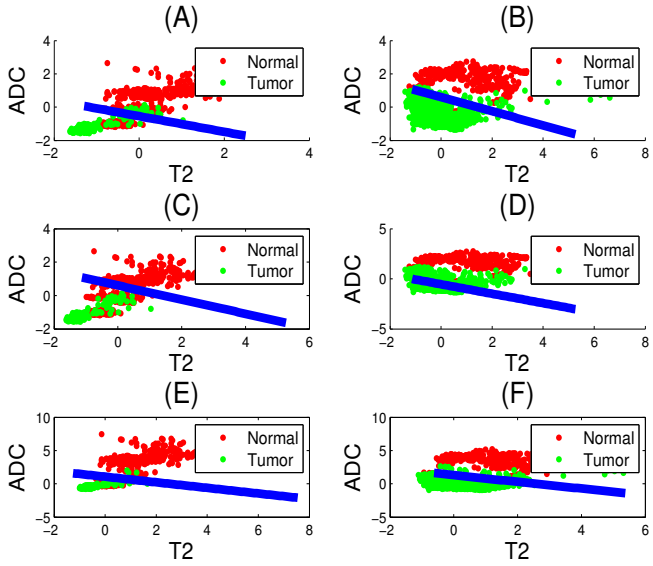

Fig. 1. Figure presents the outline of our cross device automated classification study. (A)-(B) shows the classifier learnt using tumor and normal pixels of several patients for Philips and GE devices, respectively. Fig. (C)-(D) shows the z-score normalized Philips data with GE classifier and the z-score normalized GE data with Philips classifier. Fig. (E) demonstrate the Philips data with GE classifier after relative contrast transformation applied to it. Similarly, Fig. (F) shows the GE data with Philips classifier after relative contrast transformation applied to it.

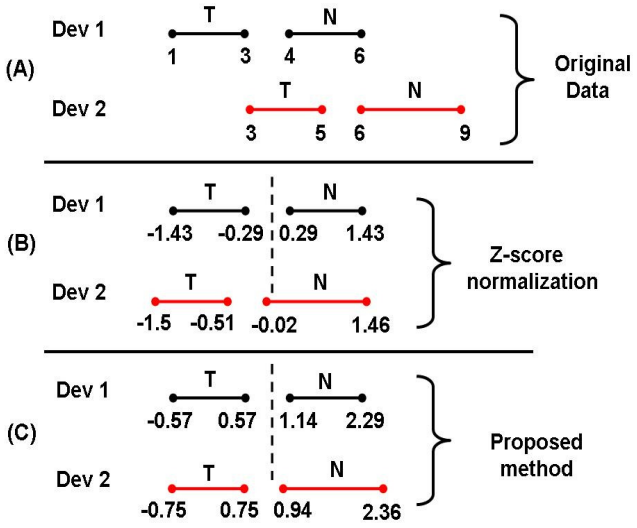

Fig. 2. Intensity distributions in (A) for data coming from two different devices, (B) z-score normalization applied to data, and (C) proposed method.

mean and unit standard deviation for all the training and testing subjects for a particular image type:

$$Y_{ij} = \frac{X_{ij} - \mu_i}{\sigma_i} \quad , \tag{1}$$

where X_{ij} is the image type i for patient j, Y_{ij} is the normalized image type i for patient j, and μ_i, σ_i refers to the mean and standard deviation of multiparametric image type i. As noted in earlier studies, z-score normalization works well for patients with similar intensity distributions. However, this assumption usually fails due to the high variation in cross-device/protocol intensities. The z-score normalization induce a bias by shifting the data by a global mean. However, this method is not sufficient to obtain satisfactory results as apparent from e.g., Figure 1.(D).

For further explanation, Figure 2.1 shows a toy example to illustrate the difference between the z-score normalization and the proposed method. Figure 2.1.(A)-(B) illustrates the original intensities and z-scored normalized distributions for two patients from two different devices. A classifier constructed using device 1 (see the dashed line in Figure 2.1.B) will not properly classify test data from device 2. Note that proposed method shown in Figure 2.1.(C) is able to classify test data coming from device 2 without errors.

Potential problems associated with z-score normalization has not received much attention in the past other than a few recent studies [3], [7]. However, none of these available methods consider the problem of cross-device training for localization problems. To avoid the potential problems caused by z-score normalization and to be able to successfully apply cross-device training, we introduce a novel approach using iterative normalization based on the idea of relative contrast between patients as explained next.

2.2. Proposed Method

In this section, we propose a novel normalization technique to mimic manual human segmentation. Human readers typically compare the contrast between the classes without knowing the actual intensity values. In this study, we utilize relative intensity idea to perform the segmentation task. We define the relative intensity for the i^{th} pixels as follows:

$$x_{im}^{rc} = (x_{im} - \mu_{1m})/\sigma \quad \forall \text{ i} \tag{2}$$

where x_{im} is the pixel intensity of the m^{th} feature at pixel i, μ_{1m} is the mean intensity of the m^{th} feature of tumor class, and σ is the standard deviation of the m^{th} feature. Notice that the relative intensity for one class depends on the other class requiring the ground truth and this is not a simple affine transformation of data.

For the training data, labels are known a priori and we can compute relative contrast directly using (2). However, for the test data, we can not simply apply Eq. (2) since we do not know the label information. Therefore we propose an iterative method to estimate the class memberships. Initially, we start with an initial estimate of class labels obtain from available techniques such as LDA with z-score normalization. Next, we update the μ_k and σ values

$$\mu_k = \frac{1}{N_k} \sum_{x \in S_k} x_i, \sigma^2 = \frac{1}{N} \sum_{x \in S} (x_i - \mu)^2 \quad , \tag{3}$$

where k denotes the class number, S_k the set of elements in class k (S denotes all elements) and N_k the number of pixels in class k set by LDA classifiers. Then, we calculate the relative contrast for this initial labeling. Next, we classify our initial estimates into two groups by the LDA classifier. The method described here offers an elegant way of transforming images such that cross device training is now possible.

The main steps of the proposed algorithm are as follows:

(1) Given the training data with their labels from imaging device/protocol A, create new intensity values according to Eq. (2).
(2) Using the training data obtained from Step 1, construct a classifier using linear discriminant analysis (LDA) algorithm.
(3) Apply the given test data from imaging device/protocol B on the classifier model obtained in Step 2.
(4) Update μ_k, and σ according to Eq. (3).
(5) Update the relative contrast using Eq. (2).
(6) Repeat steps (3)-(5) until convergence.
The above iterative process stops when there is no change in relative intensity values. In our experiments, this process takes less than 20 iterations, hence, there is no heavy computation involved.

2.3. Linear Discriminant Analysis (LDA)

In this paper, we use LDA for classification; however, note that the developed relative intensity method is applicable to any type of classifier. Given a finite sample data set (x_i, y_i) for $i = 1, 2, \cdots, N$ where $x_i \in R^d$ is a d dimensional input (feature) vector and $y_i \in \{1, 2\}$ is a class label, the objective is to estimate the parameter vector w of the discriminant function $f(x) = \langle w, x\rangle + b$ in order to classify future test samples. Linear Discriminant Analysis (LDA) searches for those vectors in the underlying space that best discriminate among classes. In LDA framework, (for all the samples of all classes) we define within-class scatter matrix given by $S_W = \sum_{i\in c}\sum_{j\in k}(x_j - \mu_i)(x_j - \mu_i)^T$, and between-class scatter matrix $S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$, where μ_i represents the mean of class i. Linear Discriminant determines w vector by maximizing the ratio $arg\max_{w'} \frac{<w', S_B w'>}{<w', S_W w'>}$. The solution is given by $w = S_W^{-1}(\mu_1 - \mu_2)$. It is often useful to see this w in geometrical terms: the observation belongs to y if corresponding is located on a certain side of a hyperplane perpendicular to w. The location of the plane is defined by the threshold b, which is equal to $b = w \cdot \frac{(\mu_1+\mu_2)}{2}$.

3. EXPERIMENTS

3.1. Description of Data

In this section, we show that we are able to perform successful classification of the test data from one device by using a classifier obtained from another device. We also show that this was not possible with available methods. In order to assess the performance of the proposed method, we use a dataset that is collected by two different MRI machines. These two devices are : (i) 1.5-T scanner (Excite HD) GE Healthcare, (ii) 1.5-T scanner (Achieva) Philips Healthcare. An endorectal coil (Medrad) combined with a phased-array surface coil was used for all examinations.

Protocol for GE Healthcare Scanner is as follows: Array spatial sensitivity-encoding technique (parallel imaging) factor of 2 was used in all sequences. T2-Weighted Imaging Parameters; TR range/TE range $= 3200 - 3500/90 - 100$, Matrix size $= 192 \times 256$, Echo-train length $= 19$, Number of signals acquired$= 4$, Section thickness$= 3$ mm, Intersection gap $= 0$ mm, FOV $= 14 - 16$ cm. Diffusion-Weighted Imaging Parameters; TR range/TE range $= 7000 - 8000/80 - 90$, Matrix size $= 128 \times 128 - 224$, b values $= 0, 1000, \text{and } 1500$ s/mm^2, Number of signals acquired $= 4$, Slice thickness $= 4$ mm, Gap = 0 mm, FOV = 1418 cm.

Protocol for Philips Healthcare Scanner is as follows: Effective sensitivity-encoding (parallel imaging) factor of 2 was used in all sequences. T2-Weighted Imaging Parameters; Resolution $= 0.8 \times 0.8 \times 3$ mm, TR range/TE $= 4300 - 5000/120$, Matrix size $= 204 \times 256$, Echo-train length $= 24$, Number of signals acquired $= 4$, Section thickness = 3 mm, Intersection gap $= 0$ mm, FOV $= 14 - 18$ cm. Diffusion-Weighted Imaging Parameters; TR range/TE range $= 3800 - 4200/80 - 90$, Matrix size $= 128 \times 128$, b values $= 0, 1000$, and 1500 s/mm^2, Number of signals acquired $= 4$, Slice thickness $= 4$ mm, Gap $= 0$ mm, FOV $= 14 - 18$ cm.
Ground truth: Ground truth is required for both training the classifiers, and for evaluation. After radical prostatectomy, the surgical specimen of the entire prostate was fixed in 10% buffered formalin for 24 hours. After dehydration, the specimen was cut serially into 3-mm-thick blocks from apex to base in transverse planes. Each block was then either halved or quartered (depending on its size), and microtome slices were cut and stained with hematoxylin-eosin. A genitourinary pathologist reviewed hematoxylineosin stained slices with cancer, and, by using a four-quadrant (right anterior, right posterior, left anterior, and left posterior) approach, recorded on a schematic prostate diagram the size, location, gleason score, and presence or absence of carcinoma. A radiologist determined the locations of the carcinoma and normal regions on T2-weighted images on the basis of these diagrams and consultation with the pathologist. Each hematoxylineosin stained slice was then visually matched to a corresponding T2-weighted MR image on the basis of the location of the ejaculatory ducts, the dimension of the prostate and the approximate distance from the base or apex.

We conducted experiments for the following MR images; T2-weighted, apparent diffusion coefficient (ADC) maps. We normalize T2-weighted image using a uniform region from the bladder, this allows us to bring T2-weighted images into a comparable range. Apparent diffusion coefficient (ADC) maps are derived from DWI parametric maps. Prostate cancer tissue intensity is lower than healthy regions in both T2 and ADC images. It is an indicator of the movement of water within the tissue and provides a measure of the flow and distance water molecule has moved. We have used a dataset obtained from 8 patients from both devices with ground truth.

3.2. Experimental Results

In addition to visual results, number of misclassified pixels (error rate) and dice coefficients are used to evaluate the performance quantitatively. Dice measure (DSC) is a common metrics used by many researchers in prostate cancer segmentation [2], [3]. It is defined as,

$$Dice(A, B) = 2 \cdot \frac{|A \cap B|}{|A| + |B|} \quad ,$$

where A is the segmentation result, B is the ground truth for the tumor and the operation $|.|$ means the number of segmented pixels.

In our experiment, we perform a cross-device classification in which we design a classifier using data from device A and test it on the data from device B. Our evaluation consists of two parts. First, we look at the classification error rate using normal regions and tumor regions for 8 patients (4 patients data collected with GE machine, and 4 from Philips machine). This error rate values are reported in Table 1 for the proposed method and z-score normalization method. Figure 1.(C)-(E) corresponds to the error rates listed in this table. Note that number of misclassified pixels using proposed method, 128, is much lower than the case for z-score normalization 271 when we design a classifier using Philips device and test it on the GE device. Similarly, number of misclassified pixels reduces from 718 to 141 when we train a classifier using GE device and test it on the Philips device comparing the proposed method with z-score normalization.

Next, we apply the same technique on the whole peripheral zone

Table 1. Comparison of (misclassified pixels)/(number of test pixels) using data from 8 patients (4 patients of GE machine, 4 patients of Philips machine) for z-score normalization and the proposed method. [Ph(train), GE(test)] denotes that classifier is constructed using Philips device's training data, tested on the GE's data.

	Error rate
Proposed: GE (test) and Ph (train)	141/1959
z-score: GE (test) and Ph (train)	718/1959
Proposed: Ph (test) and GE (train)	128/776
z-score: Ph (test) and GE (train)	271/776

to be able to visually compare segmented tumors. Note that we compare the proposed method to an LDA classifier constructed using z-score normalized data. Figure 3 demonstrates the visual results obtained using the proposed method and the z-score normalization. For patient 1 and patient 3 (both obtained from Philips scanner), there is a large amount of false positives with z-score normalization. However, results in the fourth column show that proposed method has better sensitivity, specificity rates compared to z-score normalization. Table 2 demonstrates the mean and standard deviation of DSC, sensitivity and specificity values for 8 patients for these two methods. Proposed technique results in a dice measure of 0.4369, whereas z-score normalization generates a dice measure of 0.4029. In terms of sensitivity and specificity rates, proposed method again results in improved performance.

Table 2. Mean $\pm$ std of DSC, sensitivity and specificity rates for 8 patients with z-score normalization and the proposed method.

Method	z-score	Proposed
Dice	0.4029 ± 0.1855	0.4369 ± 0.1914
Sens	0.8151 ± 0.2100	0.8651 ± 0.2066
Spec	0.5261 ± 0.2095	0.5342 ± 0.1545

Our preliminary experiments indicate that a simple normalization technique such as z-score is not sufficient to allow for cross-device automated cancer localization. On the other hand, the method we have proposed using relative intensity allows us to successfully use a classifier obtained from Philips device applied on a test patient imaged with GE device, and vice versa.

4. CONCLUSION

In this paper, we presented a novel image segmentation method to allow for cross-device automated cancer localization. This iterative algorithm based on relative intensity enables us to successfully utilize a classifier trained with a classifier obtained a GE MRI device to be applied on a dataset obtained from a Philips MRI device, vice versa. This method is highly advantageous when it is costly to obtain training data for an MRI device. We showed that simple normalization techniques such as z-score are not sufficient to allow for cross-device automated cancer localization. Proposed technique generates a dice measure of 0.4369, whereas z-score normalization generates a dice value of 0.4029.

In the future, we will test our method on a larger dataset. However, our results based on the toy data and 8 patients show that the proposed method is a promising technique for cross-device automated classification.

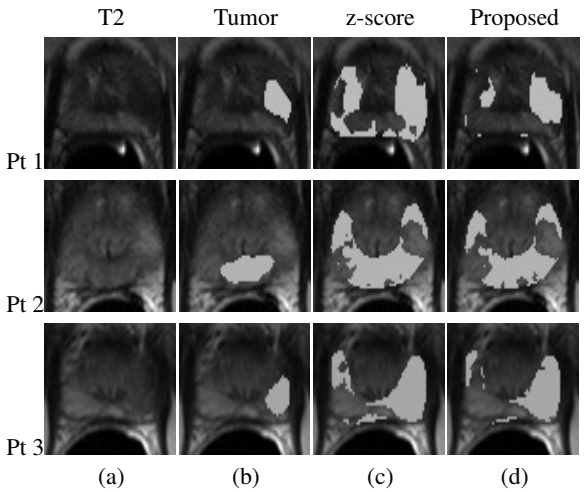

Fig. 3. Comparison of LDA segmentation results of the three patients obtained from the normalized data and z-score normalized data. The first column is the normalized T2 image, second column is the ground truth outlined by a radiologist, the third column is result obtained with the z-score normalized image, and the fourth column is obtained by the proposed normalized data. We observe that significant improvement is obtained when our method is used with cross-device training.

5. REFERENCES

[1] A. Oto, A. Kayhan, Y. Jiang, M. Treiakova, *et al.*, "Prostate cancer: differentiation of central gland cancer from benign prostatic hyperplasia by using diffusion-weighted and dynamic contrast-enhanced MR imaging" *Radiology*, vol. 257 no. 3, pp. 715-23, 2010.

[2] Y. Artan, I. S. Yetik, *et al.*, "Prostate Cancer Localization with multispectral MRI using cost-sensitive Support Vector Machines and Conditional Random Fields," *IEEE Trans. on Image Processing*, vol. 19, pp. 2444-2455, Oct. 2010.

[3] X. Liu, M. Haider, D. Langer, I. S. Yetik, "Using relative contrast and iterative normalization for improved prostate cancer localization with multispectral MRI", ISBI, pp. 1369-1372, 2010.

[4] I. Chan, W. Wells, R.V. Mulkern, S. Haker, *et al.*, "Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier," *Medical Physics*, vol. 30, no. 9, pp. 2390-2398, Sep. 2003.

[5] A. Madabhushi, M. D. Feldman, *et al.*, "Automatic Detection of prostatic adenocarcine from high-resolution ex vivo MRI," *IEEE Trans. of Med. Imaging.*, vol. 24, no. 12, pp. 1611-1625, 2005.

[6] K. Fukunaga, Introduction to Statistical Pattern Recognition, 2^{nd} Ed., Academic Press, 1990.

[7] J. L. R. Andersson, "How to estimate global activity independent of changes in local activity," *NeuroImage*, vol. 6, no. 4, pp. 237-244, 1997