ASSESSMENT OF CANCER THERAPY EFFECTS USING TEXTON-BASED CHARACTERIZATION OF QUANTITATIVE ULTRASOUND PARAMETRIC IMAGES

Mehrdad J. Gangeh[1]*, Ali Sadeghi-Naini*[2, 3]*, Mohamed S. Kamel*[1]*, and Gregory J. Czarnota*[2, 3]

[1] Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada
[2] Depts. of Medical Biophysics and Radiation Oncology, University of Toronto, Toronto, ON, Canada
[3] Dept. of Radiation Oncology and Imaging Research, Sunnybrook Health Sciences Center, Toronto, ON, Canada

ABSTRACT

This paper proposes the application of texton-based approach for textural characterization of quantitative ultrasound parametric maps, in order to assess noninvasively the progressive effects of cancer treatment in preclinical animal models. Xenograft tumour-bearing animals were treated with chemotherapy. Ultrasound data were acquired from tumours prior to, and at different times after exposure, and quantitative ultrasound spectral parametric maps were generated. Texton-based features were extracted from 0-MHz Intercept parametric maps and applied to differentiate between pre- and post-treatment states. The classification error was then translated into a quantitative measure of the treatment effects. Obtained results demonstrated a very good agreement with histological observations, and suggested that the proposed approach can be used noninvasively to evaluate the progressive effects of cancer treatment.

Index Terms— Texture, quantitative ultrasound, texton, dictionary learning, cancer therapy, SVM

1. INTRODUCTION

Early assessment of the effects induced in tumours by cancer therapies is of great interests for oncologists, as it can facilitate treatment adjustments and/or alterations for patients, or even switching to a salvage therapy, on an individual basis [1]. Imaging techniques, in this regard, have considerably gained attentions as noninvasive methods that can provide quantitative evaluations of cancer treatment response [2]. In this context, ultrasound (US) imaging has been recently proposed for cancer treatment response monitoring. Ultrasound is the most frequently used clinical imaging modality in the world accounting for almost 25% of all imaging procedures [3]. It has the advantage of low cost, rapid imaging speed, portability, and high spatial resolution. However, due to many instrument parameters that can be chosen during an imaging session, it is difficult to compare standard B-mode images between different US machines and even for the same machine when different settings are used [4]. Quantitative ultrasound methods [5] have been proposed to overcome this limitation. Such methods use metrics that are predominantly independent of the instrument settings to analyze the data.

Responses developed in tumours as a result of cancer treatment are often heterogeneous [6]. This highlights potential advantages for textural analysis techniques which can characterize these features for a more accurate evaluation of therapy response. Textural analysis techniques have been previously applied for ultrasonic tissue characterization in order to discriminate between tissues with different intrinsic microstructures [7]. However, such techniques were frequently applied on US conventional B-mode images and thus highly dependent on the instrument and the scan settings. Quantitative ultrasound parametric maps, in this regard, are expected to provide more informative measures independent of settings applied in each scan.

As texture is a complicated phenomenon, there is no unique definition that is agreed upon by researchers [8]. Hence, there are many different approaches for texture analysis in the literature, each of which tries to model texture according to one or a few of its properties [9]. However, among these approaches, the techniques based on dictionary learning and sparse representation (DLSR) have shown great success in texture analysis [10, 11]. Texton-based approach is one of the DLSR-based methods, which is particularly tailored for texture analysis [10, 12].

In this paper, a texton-based approach has been investigated along with a support vector machine (SVM) to assess cancer treatment effects noninvasively, using conventional-frequency quantitative ultrasound spectral parametric maps. Ultrasound data were acquired from xenograft tumours before and at different times after chemotherapy exposure, and texton-based features were extracted from quantitative ultrasound spectral parametric maps. Obtained results were compared with histological analyses of tumour sections stained for cell death. Comparative investigations demonstrated that the proposed approach can effectively monitor development of treatment effects noninvasively, using quantitative ultrasound parametric images as a basis for texture analysis.

2. DICTIONARY LEARNING IN TEXTURE CLASSIFICATION

Dictionary learning and sparse representation are two closely related topics that have roots in the representation of data by bases. Although *predefined* bases such as those based on Fourier or wavelets transform can be used, recent investigations have shown that *learned* bases from data lead to state-of-the-art results in many applications such as face recognition [13], texture classification [11, 14], and biomedical tissue characterization [10, 12].

The first developed and customized dictionary learning technique for the application of texture analysis is the well-known texton-based approach [15]. This approach is mainly based on the Julesz' theory, which expresses that the textures can be represented by few primitive elements called textons [16]. Leung and Malik were the first who could propose a complete texture classification system based on textons [15]. Their main contribution was that they showed how the textons can be computed from textures by using k-means and the resulting cluster centers. The technique was further

refined by Varma and Zisserman [14] to a fully developed state-of-the-art system for texture classification that outperforms many other approaches in the field.

The classification systems based on texton-based approach is mainly consisting of three steps: 1) learning the dictionary, 2) sparse representation for learning the texture models, and 3) training a classifier. In following subsections, we explain each of these steps in more details.

2.1. Dictionary learning

In texton-based approach, textons correspond to dictionary elements. Although with the recent advances in the field of dictionary learning, there are several approaches to learn the dictionary from the data unsupervised, e.g., using K-SVD [17] or supervised, e.g., using discriminative K-SVD [18], texton-based approach can be considered as the first and tailored approach for learning the dictionary in the application of texture analysis with competent results compared to other approaches. To construct the dictionary in a texton-based approach, small patches are extracted from each texture image in the training set. These patches are then translated into proper patch representation such as filter banks or raw pixels. They are, finally, aggregated over all texture images of one class in the training set and submitted to a *k*-means algorithm. The resulting cluster centers are considered as the dictionary elements (textons) representing the texture images in the corresponding class. The subdictionaries learned from each class in the training set are eventually composed into one dictionary for representing all textures.

2.2. Sparse representation and model learning

The second step in a texton-based approach is sparse representation and model learning for each texture image. Sparse representation is a closely related topic to dictionary learning and it mainly implies that few elements in the dictionary are sufficient to represent the data. To represent an image in texton-based approach, patches of the same size as what was used during dictionary learning are systematically extracted from the top left to the bottom right of the image with one pixel sliding each time. Then the most similar element in the dictionary using a similarity measure such as Euclidean distance is considered as the dictionary element representing each patch. In other words, in texton-based approach, each patch in a texture image is represented by only one dictionary atom. The closest match in the dictionary is used to update a histogram of textons and by repeating this process over all patches extracted from an image, a histogram is eventually computed as the model for that image.

2.3. Classifier training

The final step in texton-based approach is to train a classifier. To this end, the models learned on training images in previous step are submitted to a classifier. It has been shown that SVM typically outperforms *k*-NN in texton-based approach [12]. Hence, SVM has been adopted in this research as the classifier.

3. EXPERIMENTAL SETUP

3.1. Data description and preparation

Human breast cancer cells (MDA-MB-231) were injected and permitted to grow to a size of 7-9 mm xenograft tumours in the hind leg of SCID mice. Animals were anaesthetized and treated with paclitaxel-doxorubicin through intravenous tail vein injection. Experimentation used five groups of animals. Each group was assessed at a different time after chemotherapy exposure, i.e., 0, 4, 12, 24, and 48 hours (named 0H, 4H, 12H, 24H, and 48H, respectively). Each mouse was imaged before and after the treatment. All tumours were excised following post-treatment imaging for histological examination.

Table 1. Data description.

Group	No. of Images per Subject		
	Subject	Pre-	Post-
"0H"	1	12	10
	2	11	10
"4H"	3	14	14
	4	14	15
	5	13	14
"12H"	6	12	12
	7	12	14
	8	15	13
	9	12	13
	10	12	12
"24H"	11	12	15
	12	13	14
	13	13	14
	14	14	16
"48H"	15	14	12
	16	11	12
	17	14	14

Ultrasound radiofrequency (RF) data were collected using a Sonix RP system (Ultrasonix, Vancouver, Canada) with a L14-5/38 transducer pulsed at 10 MHz with a central frequency of $\sim$7 MHz, focused at 1.5 cm depth, with data sampled at 40 MHz. The system was used to collect three dimensional data with scan plane separations of $\sim$0.5 mm.

Ultrasound data was analyzed across 11-15 scan planes with a size of 3.8 by 3.0 cm (Table 1). Standardized regions of interest (ROI) were used for analysis which were located at the tumour centre. Power spectra were calculated using a Fourier transform of the raw radiofrequency data for each scan line through the ROI using a sliding window method. Data were normalized with the averaged power spectrum obtained from an agar-embedded glass-bead phantom model [19]. Linear regression analysis was performed on the averaged power spectrum within a central-frequency based -6 dB window to generate a best-fit line [20]. 0-MHz intercept parametric maps were generated via the sliding window analysis within the ROI on a pixel by pixel basis, using a Hamming window of approximately 9-11 wavelengths.

Histological analysis was performed on tumour samples fixed in 5% formalin for 24 to 48 hours. Fixed tumour sections were cut in 3 representative planes with ISEL (in situ end nick labeling) staining for cell death. Microscopy was performed using a Leica DC100 microscope with a 20$\times$ objective and a Leica DC100 camera (Leica GmbH, Germany). Cell-death areas were quantified from immunohistochemistry-stained tumour sections using Image-J (NIH, Maryland, USA). At higher magnifications (40$\times$) apoptotic cells were also counted manually by identifying typical apoptotic bodies.

3.2. Dictionary and model learning

For the purpose of the computation of the dictionary, 500 patches of size 3×3 were extracted from each 0-MHz intercept parametric image in the training set. As suggested in [14], raw pixel representation was used and no filter bank was applied on the patches. Then all extracted patches from the training images of one class in each group were aggregated and submitted to the k-means clustering algorithm. This yields a subdictionary for this particular class, whose size depends on the number of cluster centers generated by the k-means. We have chosen $k = 10$ in k-means, i.e., there are 10 elements in each subdictionary. The subdictionaries were eventually merged to make the whole dictionary, with the total size of 20 in a two-class problem.

For the purpose of sparse representation of each patch and for learning the model for each image, patches of 3×3 were systematically extracted from the top left to the bottom right of each image with one pixel shift each time. The closest dictionary element (texton) to each extracted patch was computed using Euclidean distance to update a histogram of textons for each image. After normalization, this histogram was used as the model or feature set representing the image.

3.3. Classifier and evaluation

The main goal in the classification step is to measure how much the images in pre- and post-treatment in each group, e.g., "4H", are separable. This can indicate a quantitative measure of the treatment effects developed on average in each group. In other words, a higher percentage of classification error shows that the treatment effects have not been considerably apparent yet, while a lower percentage is an indication of more treatment effect development. In keeping with this, a two-class problem was defined, having all the pre-treatment images in one class, whereas the images in post-treatment were assigned to the other class.

Since there are several pre- or post-treatment images per subject (animal) in each group (Table 1), it should be ensured that these images are not mixed in the training and test sets. For this purpose, we considered each imaging session of each animal as a different case. The assignment of labels, subjects and cases are illustrated in Table 2 for group "0H" as an example. In this special case, to evaluate the effectiveness of the treatment, we have performed the evaluation of the classification system using leave-one-case-out (LOCO) approach. This means that in each group, e.g., "0H", all images of one case, i.e., pre- or post-treatment session of one subject, were used as the test set, and all other images in the group were considered as the training set. The process was repeated for all the cases in the group to obtain the classification error.

For the SVM classifier, an RBF kernel was adopted as suggested in [21]. To choose the optimal values for the kernel width (γ) and the trade-off parameter (C) for the SVM, we have performed grid search using 3-fold cross-validation at case level on the training set. In other words, the training set was divided three-fold at case level. One fold was used as the validation set and the other two folds as the training set. The dictionary was learned on these two folds each time. The experiments were repeated ten times and the average and standard deviation of the classification error was reported. The variation in classification error comes from selecting random patches from each image to learn the dictionary and also from using k-means to form the cluster centers as the dictionary elements. Recall that the cluster centers learned by the k-means are dependent on its initialization and can be different in each run.

Table 2. Subjects, labels, and cases for group "0H".

PRE/POST	PRE	PRE	POST	POST
Subject No.	1	2	1	2
Label	1	1	2	2
Case	1	2	3	4
No. of Images	12	11	10	10

Table 3. The classification error of the proposed approach and the extent of histological apparent cell death for different treatment groups. The experiments were repeated 10 times using a LOCO evaluation scheme, and the average and standard deviation of the error reported.

Group	**Error** (%)	**Average Histological Cell Death** (%)
"0H"	94.42 ± 2.50	8 ± 2
"4H"	62.14 ± 6.29	20 ± 12
"12H"	37.24 ± 7.62	22 ± 10
"24H"	12.70 ± 2.96	61 ± 24
"48H"	0.91 ± 0.62	42 ± 11

4. RESULTS AND DISCUSSION

Representative parametric images of the 0-MHz intercept overlaid on the corresponding US B-mode images obtained before and at different times after treatment are presented in Fig. 1. These images demonstrate the use of a quantitative ultrasound parameter to evaluate cancer treatment effects. Spatial heterogeneities are also apparent in the parametric images which can be characterized using textural analysis techniques.

The classification error obtained using the texton-based approach for textural analysis is provided in Table 3. As explained in previous section, the classification was performed by defining two-class problems in each group to discriminate between pre-and post-treatment cases. Since no treatment has been performed in control ("0H") group, the pre- and post-treatment images are very similar. Therefore, using a LOCO evaluation scheme, most of the images in a test set (case) are expected to be misclassified, since the classifier have been already trained on similar images labeled with the other class tag. The results given in Table 3 for this group demonstrated an encouraging performance of the proposed approach in this application. The classification error was reasonably decreased after longer times following the treatment, as therapy effects become progressively more apparent. This can be more appreciated considering the amounts of histological apparent cell death, as a principal effect of cancer therapy, where longer times after treatment typically resulted in larger areas of cell death within the tumor (Table 3).

5. CONCLUSION

In this paper, texton signatures along with the SVM classifier were proposed to derive objective measures from quantitative ultrasound parametric maps in order to assess noninvasively the progressive effects of cancer treatment in well-controlled preclinical animal tumour models. The proposed techniques incorporates the response heterogeneities frequently developed within tumours as a result of cancer therapy, by characterizing textural properties of ultrasonic spectral parametric maps. The obtained results demonstrated that the proposed approach can reasonably identify the gradual development of treatment effects within tumours. As such, this technique can be considered as an important step forward towards a precise

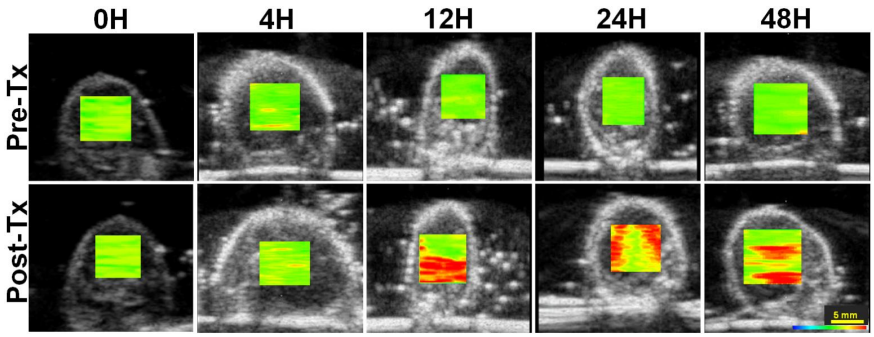

Fig. 1. Representative ultrasound B-Mode images with ROI parametric overlays of the 0-MHz intercept.

evaluation of cancer therapy regimens, early on during the treatment period. Such noninvasive assessment can facilitate addressing the necessities required for an efficacious implementation of personalized cancer therapy, which is expected to offer considerably better prognoses for the patients.

6. REFERENCES

[1] D. A. Mankoof *et al.*, "Monitoring the response of patients with locally advanced breast carcinoma to neoadjuvant chemotherapy using [technetium 99m]-sestamibi scintimammography," *Cancer*, vol. 85, no. 11, pp. 2410–23, June 1999.

[2] A. Sadeghi-Naini *et al.*, "Imaging innovations for cancer therapy response monitoring," *Imaging in Medicine*, vol. 4, no. 3, pp. 311–327, June 2012.

[3] F. Forsberg, "Ultrasonic biomedical technology; marketing versus clinical reality," *Ultrasonics*, vol. 42, pp. 17–27, 2004.

[4] D. P. Hruska, J. Sanchez, and M. L. Oelze, "Improved diagnostics through quantitative ultrasound imaging," in *Inetranational Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 1956–9.

[5] L. R. Taggart, R. E. Baddour, A. Giles, G. J. Czarnota, and M. C. Kolios, "Ultrasonic characterization of whole cells and isolated nuclei," *Ultrasound in medicine and biology*, vol. 33, no. 3, pp. 389–401, 2007.

[6] S. D. Rice, J. M. Heinzman, S. L. Brower, P. R. Ervin, N. Song, K. Shen, and D. Wang, "Analysis of chemotherapeutic response heterogeneity and drug clustering based on mechanism of action using an in vitro assay," *Anticancer Research*, vol. 30, no. 7, pp. 2805–11, 2010.

[7] Y. M. Kadah *et al.*, "Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images," *IEEE Trans. Medical Imaging*, vol. 15, no. 4, pp. 466–478, 1996.

[8] T. Ahonen and M. Pietikainen, "Image description using joint distribution of filter bank responses," *Pattern Recognition Letters*, vol. 30, no. 4, pp. 368–376, 2009.

[9] M. Mirmehdi, X. Xie, and J. Suri, *Handbook of Texture Analysis*, Imperial Collage Press, London, 2008.

[10] L. Sørensen, M. J. Gangeh, S. B. Shaker, and M. de Bruijne, "Texture classification in pulmonary CT," in *Lung Imaging and Computer Aided Diagnosis*, A. El-Baz and J. S. Sure, Eds., pp. 343–367. CRC Press, 2007.

[11] M. J. Gangeh, A. Ghodsi, and M. S. Kamel, "Dictionary learning in texture classification," in *Proceedings of the* 8^{th} *international conference on Image analysis and recognition - Volume Part I*, Berlin, Heidelberg, 2011, pp. 335–343, Springer-Verlag.

[12] M. J. Gangeh, L. Sørensen, S. B. Shaker, M. S. Kamel, M. de Bruijne, and M. Loog, "A texton-based approach for the classification of lung parenchyma in CT images," in *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Berlin, Heidelberg, 2010, pp. 595–602, Springer-Verlag.

[13] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[14] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2032–2047, Nov. 2009.

[15] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International Journal of Computer Vision*, vol. 43, pp. 29–44, June 2001.

[16] B. Julesz, "Textons, the elements of texture perception, and their interactions," *Nature*, vol. 290, no. 5802, pp. 91–97, 1981.

[17] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[18] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2691–2698.

[19] F. Dong, E. L. Madsen, M. C. MacDonald, and J. A. Zagzebski, "Nonlinearity parameter for tissue-mimicking materials," *Ultrasound in Medicine and Biology*, vol. 25, no. 5, pp. 831 – 838, 1999.

[20] F. L. Lizzi *et al.*, "Ultrasonic spectrum analysis for tissue assays and therapy evaluation," *International Journal of Imaging Systems and Technology*, vol. 8, no. 1, pp. 3–10, 1997.

[21] K. Duan, S. S. Keerthi, and A. N. Poo, "Evaluation of simple performance measures for tuning SVM hyperparameters," *Neurocomputing*, vol. 51, pp. 41–59, Apr. 2003.