

Symposium - Original Research

Immunohistochemical analysis of breast tissue microarray images using contextual classifiers

Stephen J. McKenna, Telmo Amaral¹, Shazia Akbar, Lee Jordan², Alastair Thompson³

School of Computing, University of Dundee, Dundee DD1 4HN, ¹Institute of Biomedical Engineering, Porto, Portugal, ²Pathology, Ninewells Hospital, Dundee, ³Dundee Cancer Centre, Ninewells Hospital, Dundee, UK

E-mail: *Stephen J. McKenna - stephen@computing.dundee.ac.uk

*Corresponding author

Received: 28 November 12

Accepted: 25 January 13

Published: 30 March 13

This article may be cited as:

McKenna SJ, Amaral T, Akbar S, Jordan L, Thompson A. Immunohistochemical analysis of breast tissue microarray images using contextual classifiers. *J Pathol Inform* 2013;4:S13.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2013/4/2/13/109871>

Copyright: © 2013 McKenna SJ. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Background: Tissue microarrays (TMAs) are an important tool in translational research for examining multiple cancers for molecular and protein markers. Automatic immunohistochemical (IHC) scoring of breast TMA images remains a challenging problem.

Methods: A two-stage approach that involves localization of regions of invasive and *in-situ* carcinoma followed by ordinal IHC scoring of nuclei in these regions is proposed. The localization stage classifies locations on a grid as tumor or non-tumor based on local image features. These classifications are then refined using an auto-context algorithm called spin-context. Spin-context uses a series of classifiers to integrate image feature information with spatial context information in the form of estimated class probabilities. This is achieved in a rotationally-invariant manner. The second stage estimates ordinal IHC scores in terms of the strength of staining and the proportion of nuclei stained. These estimates take the form of posterior probabilities, enabling images with uncertain scores to be referred for pathologist review. **Results:** The method was validated against manual pathologist scoring on two nuclear markers, progesterone receptor (PR) and estrogen receptor (ER). Errors for PR data were consistently lower than those achieved with ER data. Scoring was in terms of estimated proportion of cells that were positively stained (scored on an ordinal scale of 0-6) and perceived strength of staining (scored on an ordinal scale of 0-3). Average absolute differences between predicted scores and pathologist-assigned scores were 0.74 for proportion of cells and 0.35 for strength of staining (PR). **Conclusions:** The use of context information via spin-context improved the precision and recall of tumor localization. The combination of the spin-context localization method with the automated scoring method resulted in reduced IHC scoring errors.

Key words: Tissue microarrays, tumor localization, immunohistochemical scoring

Access this article online

Website:

www.jpathinformatics.org

DOI: 10.4103/2153-3539.109871

Quick Response Code:



INTRODUCTION

Tissue Microarrays (TMAs) have become an essential tool in translational research for examining multiple cancers for molecular and protein markers. However, skilled pathology review is required. Methods for automated

analysis of TMAs are under development with the aim of speeding up pathology-based research of clinical material and facilitating implementation of translational research into clinical practice. Available commercial annotation software often require a pathologist to partially annotate some tissue components in order for the software to

accurately analyze a whole mount slide. When such software is used to analyze TMA spots, typically 0.6 mm in diameter, regions are often mislabeled due to lack of context.

This paper reports on the development of methods for immunohistochemical (IHC) scoring of breast TMAs. These involve probabilistic labeling of invasive or *in-situ* carcinoma (to which we refer jointly as ‘tumor’ in this paper) and subsequent scoring using ordinal scales such as Quickscore or Allred. Tumor localization is performed using spin-context, a modification to the auto-context method that employs rotation-invariant, distribution-based context descriptors. An improvement to a previous scoring method^[1,2] is described that incorporates tumor localization. Empirical evaluations are reported on two different nuclear IHC stains, estrogen receptor (ER) and progesterone receptor (PR).

Related Work

Recent work on tumor segmentation in bright field microscopy images of histological sections includes that of Sertel *et al.*^[3] who segmented follicular lymphoma tissue slides based on mean shift and hierarchical grouping. Sieren *et al.*^[4] used clustering for segmentation of tissue types in sections of resected human lung cancer nodules. Wang *et al.*^[5] proposed a Markov random field tumor cell segmentation model applicable to TMA data.

Auto-context has been used for medical image segmentation. Morra *et al.*^[6] used AdaBoost with auto-context to segment hippocampus in 3D structural MRI. Tu *et al.*^[7] used auto-context to segment multiple structures in brain MRI. Tao *et al.*^[8] used Gaussian mixtures with simplified auto-context to segment ground glass nodules in 3D lung CT data. Montillo *et al.*^[9] segmented structures such as aorta, pelvis, and lungs in 3D CT data, proposing an extension of decision forest classifiers that incorporates semantic context in a manner similar to auto-context. However, context descriptors in the above were not distribution-based descriptors and appropriately for those applications, were not invariant under image rotation. To the best of our knowledge, auto-context has not been applied by other researchers to segmentation of 2D medical images such as those of histological sections. In this work we use a distribution-based, rotation-invariant context descriptor.

Tools are available to assist in scoring tissue sections subjected to nuclear IHC (e.g., Digital IHC Solution, Genie from Aperio Technologies, Ariol from Genetix/Applied Imaging, IHC score from Bacus Labs). Typically, they help determine the proportion of stained cells and staining strength, directly mapping those measures to different scoring systems (e.g., Allred^[10] and Quickscore^[11]), instead of performing a learned mapping based on training data. Turbin *et al.*^[12] trained Ariol software to analyze ER expression in breast carcinoma

TMAs; in their study automated and human scores were dichotomized between ER positive and ER negative, rather than directly compared. Sanders *et al.*^[13] developed a system to score TMA spots of various types and immunostained for each of several antibodies, where the mapping of a set of global features to discrete strength scores was learned from annotated data.

METHODS

Tumor Localization Method

We first address locating tumor in TMA spots. This is formulated as classifying each location on a grid as being tumor or non-tumor. The image patch around each location is characterized using local features extracted at full resolution, specifically differential invariants up to 2nd order^[14] and intensity spin image features.^[15] We experimented with multi-layer perceptron (MLP) and random forest classifiers based on these features.

We call our method for incorporating context in a rotationally-invariant fashion *spin-context*.^[16] Rotation invariance is potentially useful when analyzing histopathology images because the rotation of the tissue is arbitrary. Spin-context is a variant of auto-context^[7] inspired by the use of intensity spin features for texture representation.^[15] Auto-context is an iterative pixel labeling technique, in which some of the labels output by the classifier at a given iteration are used as contextual data that are concatenated with local image features to form the input vector for the classifier at the following iteration. Tu and Bai^[7] used a star-shaped ‘stencil’ for selection of locations at which to incorporate context. The resulting context feature vectors were not invariant under image rotation. Instead, spin-context computes context features for a grid location from label probability values within a circular support region centered at that location. This is done analogously to intensity spin features, computing a two-dimensional soft histogram reflecting the distribution of probabilities within the support region, with histogram rows representing probability intervals and columns representing radial distance intervals. The values in this histogram are concatenated with the image features for the subsequent classifier.

At the first iteration, context is not available from a previous iteration, so a uniform constant ‘context’ descriptor is adopted, and classification is based only on the local image features. At every subsequent iteration, context descriptors constructed from the probability map generated at the previous iteration are concatenated with local image features to form input to a classifier. This classifier produces an updated probability map. An iteration of spin-context is illustrated in Figure 1.

IHC Scoring Method

A previous method for IHC scoring^[1,2] estimated the

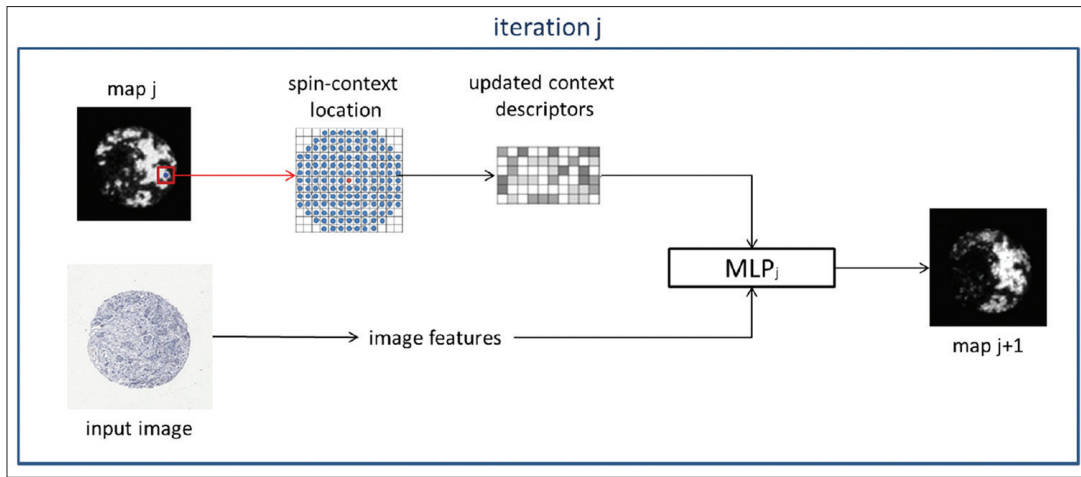


Figure 1: Spin-context constructs context descriptors for a point to be classified from a circular region on the previous iteration's classification map

proportion of epithelial nuclei that were stained and the strength of staining based on color and texture features. These values, referred to as formalized scores, were then mapped onto standard ordinal scoring scales using a classifier or ordinal regression model. Here the Quickscore ordinal scales are used. Quickscore assigns a proportion score in the range 0-6 and a strength score in the range 0-3.^[11] We refine the previous method^[1,2] by incorporating tumor localization probabilities obtained as above so that scoring focused on tumor regions. Equations (1) give the formalized scores incorporating tumor labeling where f_p reflects the proportion of nuclei that are stained and f_s reflects the staining strength; sums are over all pixels, and y_p^n , y_T^n , and y_N^n denote the posterior probabilities that the n^{th} pixel belongs to an immunonegative nucleus, an immunopositive nucleus, and a tumor region, respectively. t_p^n is 1 if y_p^n is greater than 0.5, 0 otherwise. In a further refinement of Amaral *et al.*,^[1,2] nuclear posterior probability ‘images’ were Gaussian smoothed prior to computation of IHC scores.

$$f_p = \frac{\sum_n y_T^n y_P^n}{\sum_n y_T^n (y_N^n + y_P^n)} \quad f_s = \frac{\sum_n y_T^n y_P^n t_P^n}{\sum_n y_T^n t_P^n} \quad (1)$$

Tumor labels were obtained by taking posteriors output by the method above. Nuclear labeling was carried out using the pixel classification technique described in^[17] that computes posteriors for each pixel that it is immunonegative epithelial nuclei, immunopositive epithelial nuclei, or non-nuclei. MLPs were trained to predict IHC scores from formalized scores. A decision to use MLPs was based on ten-fold cross-validation experiments on 200 ER spots (both normal and tumor) in which MLP classifiers gave better classification rates than Gaussian process ordinal regression models.

The scoring method outputs a posterior probability distribution over scores for each spot. Predictions for each spot are obtained by choosing the median score values.

The entropy of the posterior scoring distribution can be used as a measure of confidence, providing a mechanism by which to decide to refer spots with high uncertainty for pathologist review.

Data

TMA spots were subjected to nuclear staining for ER or PR. Spot images were 4000 × 4000 pixels. Data used in tumor labeling experiments consisted of 64 spots stained for ER, 32 of which contained tumor regions annotated by a highly experienced pathologist and 32 confirmed to contain only healthy tissue. Example pathologist annotations are shown in Figure 2. In addition, 20 circular regions stained for PR were annotated to identify immunonegative and immunopositive epithelial nuclei (approximately 700 nuclei). IHC scoring experiments focused on two sets of spots all known to contain tumor. The first set contained 188 spots stained for ER with Quickscores assigned by a pathologist. The second set consisted of 262 spots stained for PR with Quickscore assigned by a pathologist twice for each spot (during two sessions with 251 spots considered scorable in both sessions).

RESULTS

Tumor Labeling Experiments

Tumor labeling was evaluated using ten-fold cross-validation on the 64 ER spots. Each cross-validation experiment was repeated ten times to measure variability. MLP classifiers had five hidden units, a regularization constant of 0.1 and used scaled conjugate gradient optimization. Local and context features were computed at points on a 76 × 76 grid (a grid step of 50 pixels). Differential invariant features were computed at three scales using a Gaussian pyramid. Specifically, at each scale, images were convolved with a set of first- and second-order

two-dimensional Gaussian derivative kernels and the results combined to obtain four differential invariants at each location.^[18] The kernels had standard deviations of eight pixels, and thus, effectively 16 and 32 pixels at the second and third scales respectively. This compares with an average epithelial nuclear radius of approximately 16 pixels. These features exhibit rotation invariance as well as some invariance to absolute image intensity. Intensity spin local features were computed at two scales (again using a Gaussian pyramid) with a circular support region with a radius of 50 pixels. Spin-context used a circular support region with a radius of six grid points. We also tried auto-context (non-rotationally invariant context) using a stencil in which neighboring grid points within a radius of six grid spacings in each of the eight cardinal and inter-cardinal compass directions were used as context. Labeling obtained was compared to ground-truth segmentations provided by the pathologist to compute precision-recall curves. Figure 3 shows precision-recall curves. Figure 3a shows that spin-context improved precision and recall. Auto-context using a stencil also helped except at low recall. Figure 3b plots the

spin-context curves again, but this time with dotted curves at one standard deviation to indicate variability over runs of the experiment. For the important middle values of precision and recall, a second iteration of spin-context resulted in improvement. Figure 3c compares spin-context using MLP with spin-context using random forest classification. Except at low values of recall, MLP spin-context was superior.

Figure 2 shows three spots, two containing tumor and one not containing tumor, along with their expert annotations and the outputs of the spin-context method. In Figure 2a, posterior probabilities within tumor regions became higher at each iteration, so that after the final iteration they were above 0.5 for most tumor pixels. In Figure 2b, non-zero probabilities occur within regions of normal tissue at the first iteration; however, their values become lower after further iterations, so that a binarization of the labeling would result in an almost entirely empty (i.e., correct) output. Figure 2c shows a case of unsuccessful labeling. The region of inflammatory cells at the top-left is initially lightly detected and then correctly discarded;

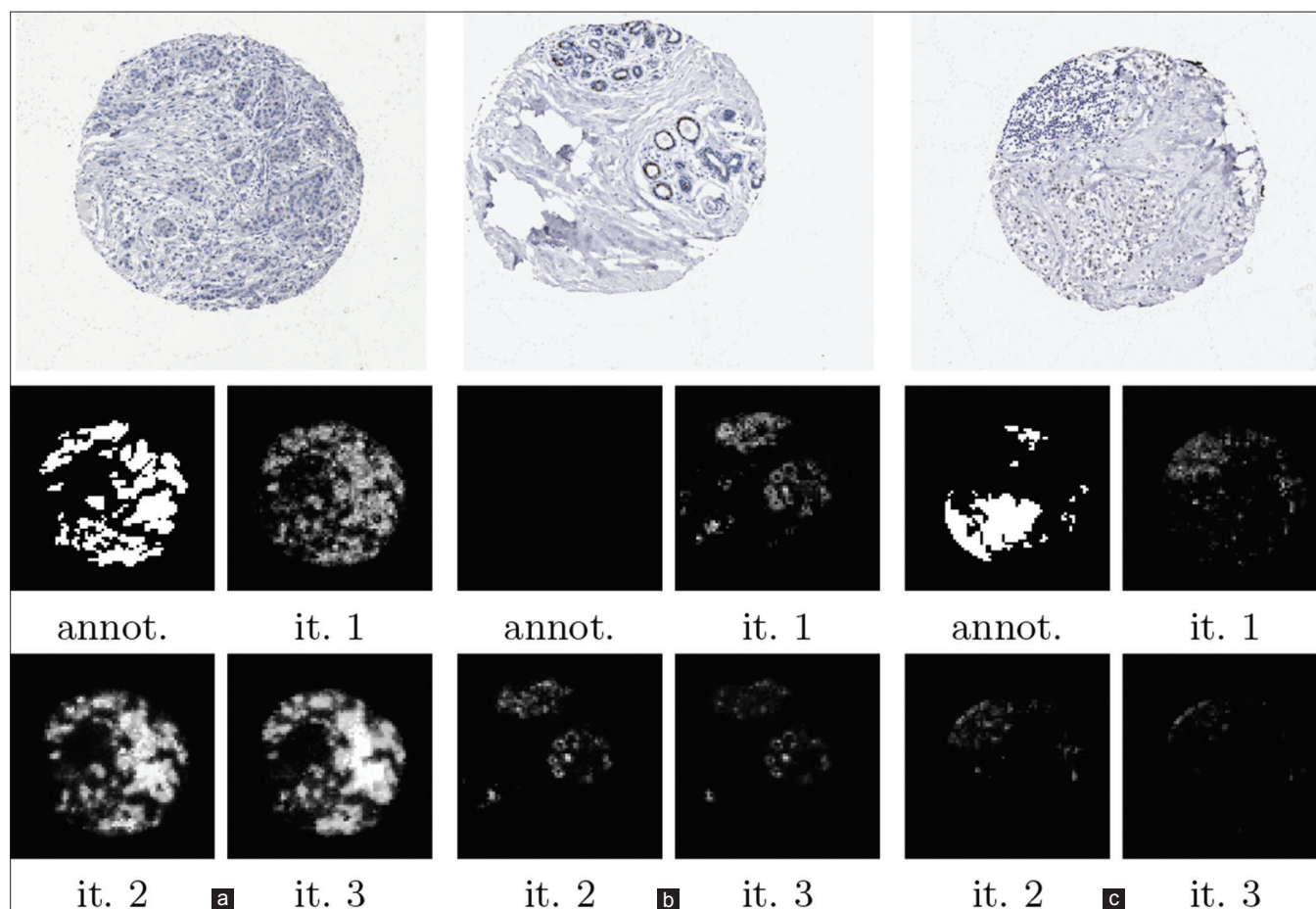


Figure 2: Tumor location probabilities obtained by spin-context. Shown for each TMA spot are the pathologist's annotation, the labeling obtained using local image features (iteration 1) and labeling obtained after incorporating label context (iteration 2 and iteration 3); (a) Shows invasive cancer labeled largely in agreement with the pathologist; (b) shows healthy tissue; (c) shows one of the worst results obtained

Table 1: Immunohistochemical scoring results. Error denotes the average absolute difference between predicted scores and the scores assigned by the pathologist. Errors are reported with and without the tumor localization step. (Proportion scores range from 0 to 6. Strength scores range from 0 to 3)

Dataset	Proportion				Strength		
	IOD	Error		IOD	Error		
		No localization	Localization		No localization	Localization	
PR (no smoothing)	0.28	0.93	0.83	0.12	0.48	0.40	
PR (smoothing)	0.28	0.88	0.74	0.12	0.39	0.35	
ER (smoothing)		1.27	1.19		0.54	0.47	

PR: Progesterone receptor, ER: Estrogen receptor, IOD: Intra-observer disagreement

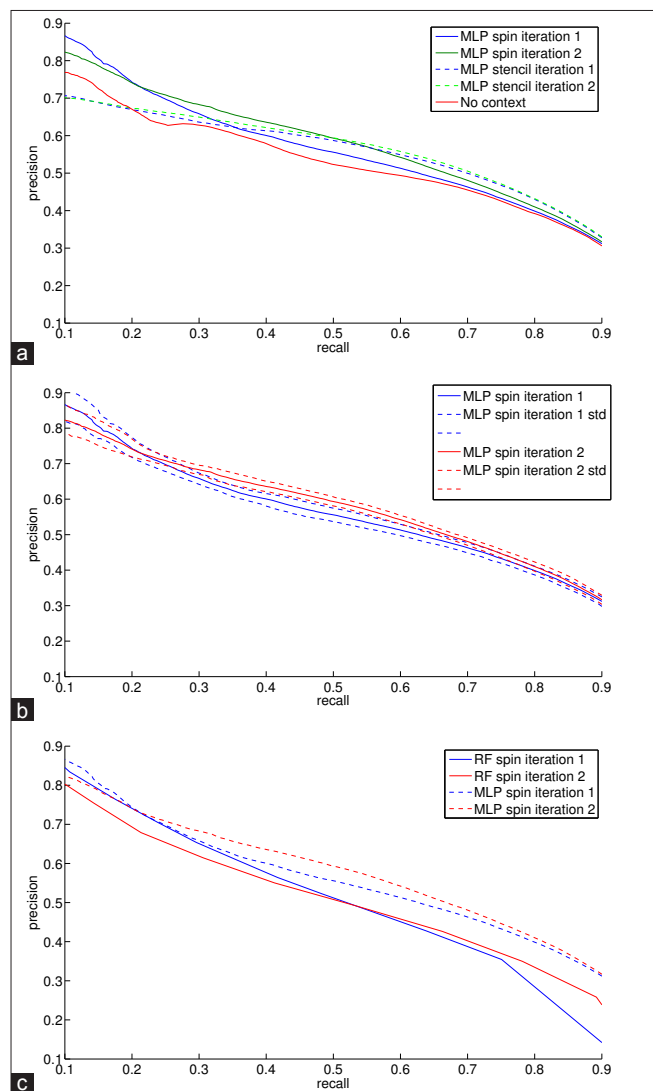


Figure 3: Precision-recall curves for tumor localization; (a) Effect of stencil and spin-context iterations on multi-layer perceptron (MLP) classification; (b) MLP spin-context with variation (dotted lines at one standard deviation); (c) Comparison of spin-context using random forest and MLP classifiers

but, most importantly, the tumor region at the lower left was missed. This may have been due to the fact that, in the region in question, tumor cells had large spatial separation giving a lower apparent density when

compared to most tumor regions in the other spots. This suggests the need for more training examples of this type.

IHC Scoring Experiments

A tumor model was trained using all 64 annotated ER spots and a nuclear model was trained using the 20 PR regions with nuclear annotations. These models were used in ten-fold scoring experiments using MLP to predict proportion and strength scores (as in Quickscores) based on the formalized scores in Equation 1. Mean absolute errors are reported, i.e., the average absolute differences between predicted scores and true scores. Table 1 details the proportion and strength score prediction errors with and without the use of tumor localization. Proportion error denotes the average difference between predicted scores and pathologist scores for percentage of positively stained cells. Similarly, strength error denotes the average difference between predicted scores and pathologist scores for the strength of IHC staining. Furthermore, shown are results obtained with the PR dataset when nuclear label smoothing was not used. Pathologist *intra*-observer disagreement (IOD) is shown for PR data. Figure 4 shows how the proportion of scored PR spots and the mean absolute error vary when only those spots predicted above a certain confidence threshold (based on posterior entropy) are considered. Spin-context provides a consistent trade-off between the proportion of TMA spots scored and error variation. This trade-off provides the means for selecting a suitable compromise between automation and accuracy when scoring TMA spots.

DISCUSSION

A method for tumor localization and IHC scoring was presented incorporating rotationally-invariant context features. The method was validated against manual pathologist scoring on two nuclear markers. Tumor localization and nuclear smoothing reduced scoring errors [Table 1]. Errors for PR data were consistently lower than those achieved with ER data. This may be partially due to the fact that the nuclear model was trained with PR data. We have also evaluated the

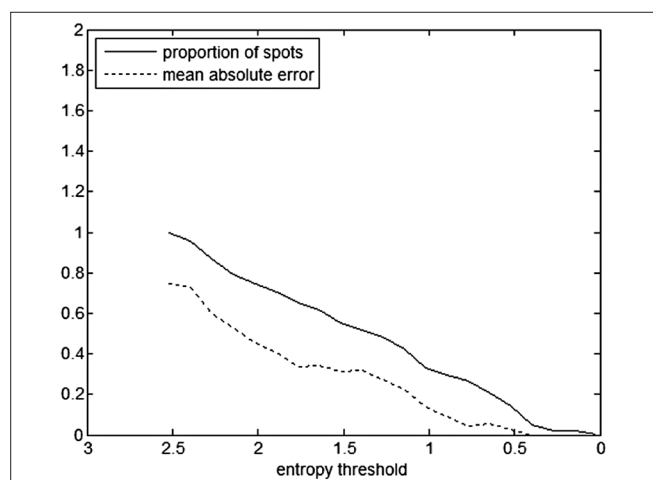


Figure 4: Effect of confidence threshold on the fraction of progesterone receptor spots scored and their mean absolute proportion score error

method with ER TMA data from a second, independent laboratory; in that experiment (not reported here) errors were at the same level even though no data from the second lab were used for training. Tumor maps such as those exemplified in Figure 1 could be useful in themselves to pathologists, helping to identify spots containing regions of tumor and to localize those regions. However, the tumor localization results should be regarded as preliminary. Some improvements in scoring obtained on the tumor data were modest. This is partly explained by the fact that most epithelial tissue present in the tumor spots (detected by nuclear labeling) constituted tumor tissue, limiting the benefits of the tumor labeling step. TMA images are diverse and localization errors on less frequently occurring structures suggest that even more annotated data would be helpful.

ACKNOWLEDGMENTS

This work was supported by the Chief Scientist Office, Scotland (grant no. CZB/4/761) and the UK Engineering and Physical Sciences Research Council (DTA grant). The authors are grateful to Dr. Katherine Robertson for her help in assembling and providing scores for some of the TMA images used.

REFERENCES

1. Amaral T, McKenna S, Robertson K, Thompson A. Scoring of breast tissue microarray spots through ordinal regression. In: Proc. Int. Conf. Computer

Vision Theory and Applications, Vol. 2. INSTICC Press; 2009. p. 243-8.

2. Amaral T, Sciarabba M, McKenna S, Robertson K, Thompson A. Scoring of breast tissue microarrays using ordinal regression: local patches versus nuclei segmentation. In: Proc. Medical Image Understanding and Analysis (MIUA), British Machine Vision Association (BMVA); 2009. p. 77-81.
3. Sertel O, Catalyurek U, Lozanski G, Shanaah A, Gurcan M. An image analysis approach for detecting malignant cells in digitized H&E-stained histology images of follicular lymphoma. In: Proc. International Conference on Pattern Recognition (ICPR) Istanbul, August. Los Alamitos, California: IEEE Computer Society; 2010. p. 273-6.
4. Sieren JC, Weydert J, Bell A, De Young B, Smith AR, Thiesse J, et al. An automated segmentation approach for highlighting the histological complexity of human lung cancer. *Ann Biomed Eng* 2010;38:3581-91.
5. Wang CW, Fennell D, Paul I, Savage K, Hamilton P. Robust automated tumour segmentation on histological and immunohistochemical tissue images. *PLoS One* 2011;6:e15818.
6. Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, et al. Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *Neuroimage* 2008;43:59-68.
7. Tu Z, Bai X. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2010;32:1744-57.
8. Tao Y, Lu L, Dewan M, Chen AY, Corso J, Xuan J, et al. Multi-level ground glass nodule detection and segmentation in CT lung images. *Med Image Comput Comput Assist Interv* 2009;12(Pt 2):715-23.
9. Montillo A, Shotton J, Winn J, Iglesias J, Metaxas D, Criminisi A. Entangled decision forests and their application for semantic segmentation of CT images. *Inf Process Med Imaging. Proceedings. Lecture Notes in Computer Science, Vol. 6801* New York: Springer; 2011. P. 184-96.
10. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* 1998;11:155-68.
11. Detre S, Saclani Jotti G, Dowsett M. A quickscore method for immunohistochemical semiquantitation: Validation for oestrogen receptor in breast carcinomas. *J Clin Pathol* 1995;48:876-8.
12. Turbin DA, Leung S, Cheang MC, Kennecke HA, Montgomery KD, McKinney S, et al. Automated quantitative analysis of estrogen receptor expression in breast carcinoma does not differ from expert pathologist scoring: A tissue microarray study of 3,484 cases. *Breast Cancer Res Treat* 2008;110:417-26.
13. Sanders T, Stokes T, Moffitt R, Chaudry Q, Parry R, Wang M. Development of an automatic quantification method for cancer tissue microarray study. In: Vol. 1, Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBS); Minneapolis. IEEE; 2009. p. 3665.
14. Schmid C, Mohr R. Local gray value invariants for image retrieval. *IEEE Trans Pattern Anal Mach Intell* 1997;19:530-5.
15. Lazebnik S, Schmid C, Ponce J. A sparse texture representation using local affine regions. *IEEE Trans Pattern Anal Mach Intell* 2005;27:1265-78.
16. Akbar S, Amaral T, McKenna SJ, Jordan L, Thompson A. Tumour segmentation in breast tissue microarray images using spin-context. In: Proc. Medical Image Understanding and Analysis (MIUA), Swansea: British Machine Vision Association; 2012. p. 25-30.
17. Amaral T, McKenna S, Robertson K, Thompson A. Classification of breast tissue microarray spots using colour and local invariants. *IEEE Int Symp Biomed Imaging (ISBI)* 2008:999-1002. IEEE.
18. Amaral T. Analysis of breast tissue microarray spots. PhD Thesis. Dundee, Scotland, UK: University of Dundee; 2010.