33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

An Integrated Texton and Bag of Words Classifier for Identifying Anaplastic Medulloblastomas

Joseph Galaro[a], Alexander R. Judkins[b], David Ellison[c], Jennifer Baccon[d], Anant Madabhushi[a]

[a]Rutgers, Department of Biomedical Engineering. Piscataway, NJ, USA.
[b]Children Hospital of L.A., Department of Pathology Laboratory Medicine. Los Angeles, CA, USA
[c]St. Jude Children's Research Hospital from Memphis, TN, USA
[d]Penn State College of Medicine, Department of Pathology. Hershey, PA, USA

Abstract— **In this paper we present a combined Bag of Words and texton based classifier for differentiating anaplastic and non-anaplastic medulloblastoma on digitized histopathology. The hypothesis behind this work is that histological image signatures may reflect different levels of aggressiveness of the disease and that texture based approaches can help discriminate between more aggressive and less aggressive phenotypes of medulloblastoma. The bag of words approach attempts to model the occurrence of differently expressed image features. In this work we choose to model the image features via textons which can quantitatively capture and model texture appearance in the images. The texton-based features, obtained via two methods, the Haar Wavelet responses and MR8 filter bank, provide spatial orientation and rotation invariant attributes. Applying these features to the bag of words framework yields textural representations that can be used in conjunction with a classifier (κ-nearest neighbor) or a content based image retrieval system. Over multiple runs of randomized cross validation, a κ-NN classifier in conjunction with Haar wavelets and the texton, bag of words approach yielded a mean classification accuracy of .80, an area under the precision recall curve of .87 and an area under the ROC curve of .83 in distinguishing between anaplastic and non-anaplastic medulloblastomas on a cohort of 36 patient studies.**

I. INTRODUCTION

Medulloblastoma neoplasms are embryonic tumors that grow in the posterior fossa and have greatest occurrence along the mid-line of the cerebellum of young children. Medulloblastoma accounts for 25% of all pediatric brain tumors. Motivation behind classifying medulloblastoma arises because intense and early treatment is required to overcome aggressive cancers, whereas delayed treatment is preferable for less aggressive cases. Grouping the cancer into phenotypes is useful in determining the potential outcome of the disease. Manual classification schemes are wrought with difficulty because the subtypes of medulloblastoma are hard to differentiate and new types are still being described [1]. Different variants of medulloblastoma include: classical, large cell, anaplastic, and desmoplastic. In anaplastic medulloblastoma, the degree of anaplasia may correlate to disease aggressiveness and is of clinical value when evaluating a patient prognosis [2]. Anaplastic medulloblastoma is indicated by the presence of large, irregular cells that lack organization and in some cases begin to wrap around each other.

Texture based feature extraction strategies have been used in the context of computer aided diagnosis and prognosis of disease from digital pathology [3]. Features generated from Gabor and Haralick filters have been shown to be successful in capturing the differences between phenotypes of breast cancer [4]. Texture based features have been shown to be highly representative of visual cues that pathologists employ when attempting to diagnose cancer regions across a range of magnifications [5].

Previous work has been done to leverage the combination of textons and the bag of words model. The texton based approach is a generative method for building representative pixel-level textural features. This approach is fundamentally different from other texture based classifier approaches (e.g. classifiers based off statistical features computed within local neighborhoods) which are bound by the spatial ordering of pixel intensities. Bag of words (BoW) is a classification scheme used originally for text classifiers which was introduced to the imaging domain [6] by treating textons as words. A BoW strategy disassociated from textons [7] met with limited success classifying histopathology images. Khurd et al. combined the BoW and texton based approach for classification of prostate histology images [8].

The methodology presented in this paper, while similar to [8] is different in that (1) it employs the nearest neighbor (κ-NN) classifier, and that (2) it employs the Haar wavelet filter bank as opposed to the traditional MR8 filter bank [6]. The BoW, texton classifier employed in this work is modeled as a two class dichotomizer, where the objective is to differentiate between anaplastic and non-anaplastic medulloblastomas on digital pathology. The texton, BoW approach diminishes the importance of the spatial context of the features, thereby lending the texton based classifier greater flexibility. The MR8 filter bank, which emphasizes local, rotationally invariant features has been shown to perform well

This work was made possible by grants by the Walter H. Coulter Foundation, National Cancer Institute (Grant Nos. R01CA136535, R01CA140772, and R03CA143991), Department of Defense Prostate Cancer Research Program, The Cancer Institute of New Jersey and the Society for Imaging Informatics in Medicine.

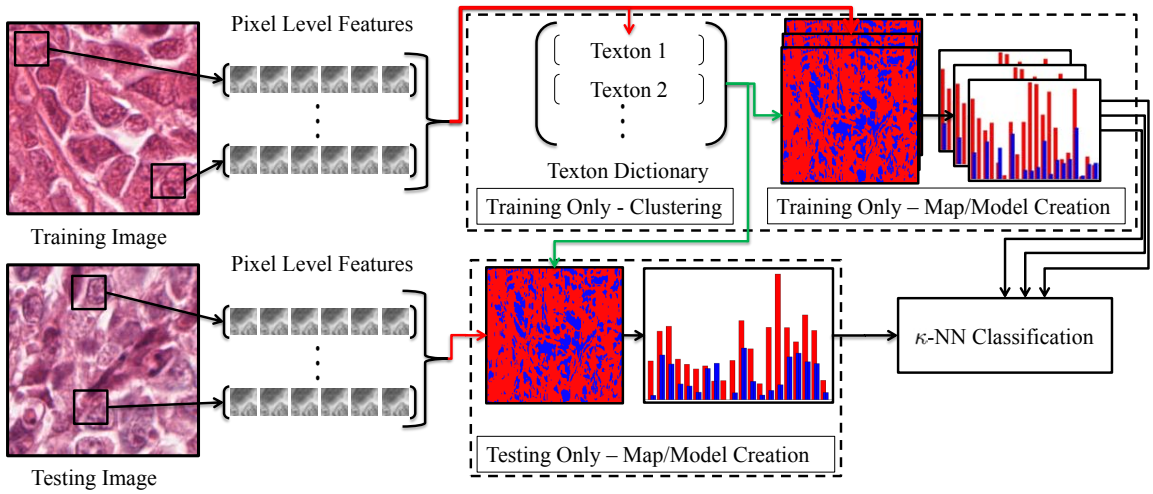

Fig. 1. Schematic for the construction of a texton based classifier using the Haar wavelet filter bank. The initial features are extracted on a pixel-wise scale. The features are then clustered via K-means to identify the cluster centroids. The cluster centroids are reflective of the most significant features for a given class. The feature space representing an image is then reduced to the index of the nearest neighbor texton. The final step is the bag of words tally of the textons, forming the histogram model for an image. Classification of a new test image involves first constructing the corresponding texton signature and then using a κ-NN classifier to identify the κ closest database (annotated) images in terms of the L2 norm. The test image is then assigned the class label of the majority of the closest retrieved instances by the κ-NN classifier.

on synthetic data [6]. However, given that some researchers have reported less than satisfactory results on histopathology image analysis [9] it is not clear whether the MR8 filter bank would be appropriate for disease classification where the textural differences are often very subtle. Haar Wavelets on the other hand, in addition to yielding textural features that the MR8 filter bank is able to capture, also produce textural responses at different resolutions which are able to better capture the gradients across the images. This enables the Haar transform to better account for torosity and misshapen cells, defining features of anaplastic medulloblastoma. The novel contributions of this work are:

1) An integrated texton, bag of words scheme which in conjunction with the Haar filter bank is used for texture classification, and
2) application of this integrated classifier for an important prognostic problem in medulloblastomas – namely, distinguishing anaplastic from non-anaplastic medulloblastomas.

II. Methodology

A. Haar Wavelet Filter Bank

The Haar Wavelet Filter allows for image representation in terms of an orthonormal basis [10]. Wavelet analysis involves solving for the dot product of a function and a translational wavelet in order to represent the data as a superimposition of basis wavelets. In the case of Haar Wavelets, the translational wavelet is a block function, The process allows for the formation of an image basis by subtracting the intensity mean across multiple image pixels. The Haar wavelet filter bank comprises of wavelets from low to high scales which when convolved with an image can be used to extract multi-scale image features. For instance Figures 2(b) and 2(c) represent the finer detail textural images obtained by convolving the original medulloblastoma Figure 2(a) with higher scale filters.

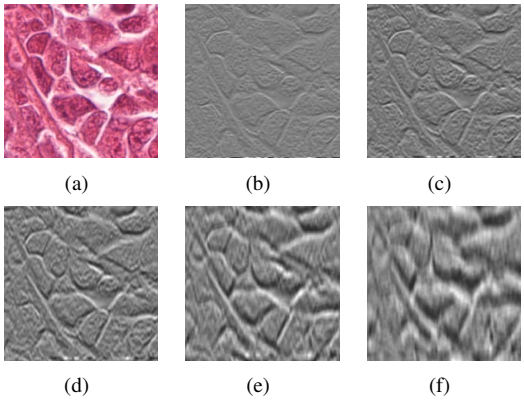

Fig. 2. Illustration of the resulting Haar wavelet images. (a) Original medulloblastoma image. (b)-(f) corresponding Haar wavelet based texture representations of the original images obtained by convolving the filter bank with (a). These images are able to both capture fine features at the higher scale (b),(c) and macro-transitional features at the lower scale (d)-(f).

B. Feature Generation

We denote an image scene $\mathcal{C} = (C, f)$ where C is an ordered set of pixels, $f(c)$ is the grayscale intensity at a pixel $c \in C$, and ℓ is the corresponding label of the image. The class label of the image is defined as $\ell(\mathcal{C}) = \omega, \omega \in \{+1, -1\}$. The feature space of a pixel consists of the corresponding value in each of 7 Haar wavelet responses, where at each pixel, the texture based features are denoted by $F(c)$. In order to account for differences in illumination of the sample during image capture and outlying filter responses, we normalize $F(c)$ via the following equation [6]:

$$F(c) \leftarrow \frac{F(c)}{|F(c)|} \log\left(1 + \frac{|F(c)|}{0.03}\right) \quad (1)$$

where $|F(c)|$ represents the cardinality of set $F(c)$. After calculating $F(c)$ at each pixel, the vectors are aggregated into groups by class in the form $S_\omega = \{F(c)|c \in C, \ell(\mathcal{C}) = \omega\}$.

C. Creation of Texton Dictionary

The texton dictionary is an aggregation of the centroids received from K-means clustering on S_ω for all $\omega \in \{+1, -1\}$. The steps towards building the texton dictionary begin by pre-specifying the number of textons per class, K. $T_{k,\omega}$ is a centroid for a cluster, $k \in \{1...K\}$, obtained via K-means clustering of S_ω. The texton dictionary is then defined as the set $\{T_{k,\omega}|\omega \in \{+1, -1\}, k \in \{1, ..., K\}\}$. It is important to note that there will only be K textons per class, because an under- or over-represented class in the texton dictionary will degrade the classifier.

D. Construction of the Texton Map & Model

Using the texton dictionary created in Section II-C, any pixel c can be assigned a texton label

$$h(c) = \{k, \omega\} = \underset{k,\omega}{\operatorname{argmin}} \|F(c) - T_{k,\omega}\|_2 \qquad (2)$$

based on its nearest texton (via Euclidean distance) in the feature space.

In this manner, a texton map can be generated for an image $\mathcal{C}$ by finding the corresponding texton label $h(c)$ for each pixel $c \in C$. Subsequently, a model (i.e. histogram of texton frequency) comprising K bins for each class can be constructed to represent each image $\mathcal{C}$ via its relationship with the texton dictionary.

E. Content Based Image Retrieval via κ-NN

For a new image to be classified, the Haar wavelet filter bank is first convolved with the image and its corresponding texton dictionary constructed. A κ-NN classifier is then used to identify the κ closest database (annotated) images in terms of the L2 norm. The query image is then assigned the class label of the majority of the closest retrieved instances by the κ-NN classifier. The strategy is evaluated via classifier accuracy, area under the receiver operating characteristic (ROC) curve and also content based image retrieval performance measures such as area under the precision-recall (PR) curve.

III. Experimental Results and Discussion

	κ	3	5	7	9	10
Haar	Accuracy	0.78	0.79	0.79	0.80	0.80
	AUC	0.81	0.82	0.83	0.83	0.83
	AUPRC	0.85	0.86	0.87	0.86	0.86
MR8	Accuracy	0.60	0.59	0.61	0.61	0.62
	AUC	0.61	0.59	0.61	0.63	0.62
	AUPRC	0.70	0.69	0.68	0.69	0.68

TABLE I

The results of the κ optimization using Haar wavelet and MR8 filter features. Values displayed are accuracy, area under the receiver operating characteristic curve (AUC) and area under the precision-recall curve (AUPRC).

A. Dataset

36 labeled histopathological patient studies were received from the St. Jude Children's Research Hospital of which, 5 were labeled anaplastic and 31 were non-anaplastic. For each study the cancerous regions were manually annotated by an expert neuro-pathologist. Each image is 10000×10000 pixels in size, and was split into 750 individual image scenes $\mathcal{C}$, each of which are 200×200 pixels in size.

B. MR8 Filter Bank for Comparison

The Maximum Response 8 (MR8) filter bank was employed as a comparative strategy and evaluated against the Haar wavelet filters to generate the initial texture features. The MR8 filter bank was generated using a Gaussian derivative function, and by varying the parameters in order to create a robust set of filters [6]. A total of 4 distinct filters were applied: 2 isotropic filters, an edge and a line. The filters are duplicated at three different scales and 6 different rotations. The rotated sets for each scale of filter are compared and the largest value becomes the representative response for that filter at that scale.

C. Quantitative Results

Multiple trials cross validation were conducted to determine the system parameters. The testing set for each iteration consisted of the image scenes of one anaplastic and one non-anaplastic patient and the training set was images scenes of the remaining patients. Clustering the textons with K-means $K = 20$ clustering was found to be optimal for binary classification. From earlier analysis, it was found that for optimal performance, the value of K should vary inversely with the number of classes in the study [6]. Comparing the results by varying the weight of members in κ-NN classification was implemented to no effect, so an unweighted κ-NN was used. Table I shows the corresponding classifier accuracy for the Haar wavelet based filter in distinguishing between anaplastic and non-anaplastic medulloblastomas. Corresponding values for the MR8 filter bank are also shown. The Haar wavelet filter bank outperforms the MR8 filter bank, achieving a classification accuracy of 0.80 as opposed to the MR8 maximum accuracy of 0.62. Note also that the classifier statistics are remarkably consistent over different values of κ for the κ-NN classifier using the Haar Wavelets, shown in Figure I, yielding a mean accuracy of .79 and standard deviation of .0068. Additionally we also evaluated the texton based approach (Haar and MR8) in terms of (a) area under the precision recall (PR) curve, and (b) the area under the receiver operating characteristic curve (ROC). The values obtained were .87 and .83 for the AUPRC and AUC, respectively, of the Haar wavelet implementation. The MR8 scheme, however, yielded a AUPRC of .68 and an AUC of .62, as shown in Table I. An unpaired t-test was performed under the null hypothesis that there were no statistically significant differences between the AUC values for the Haar wavelet and MR8 based classifiers. The null hypothesis was rejected with a p ¡= 0.0001.

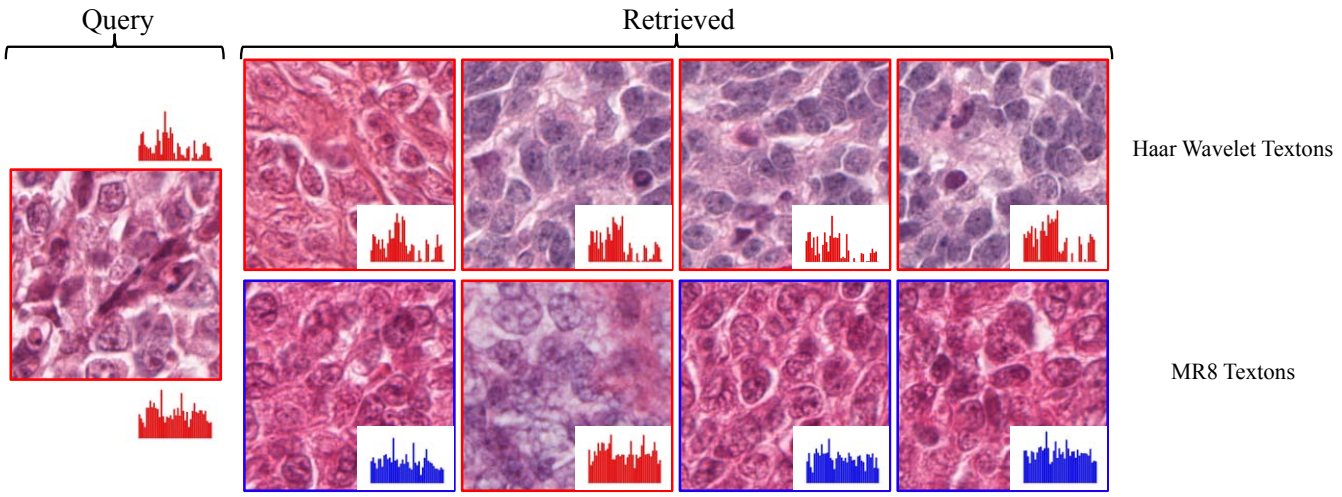

Fig. 3. The image patches bordered in red are anaplastic and the image bordered in blue are non-anaplastic. The color of the respective histogram also reflects the class of the image. The Haar wavelet filters resulted in retrieval of images that were identical to the class label of the query instance, while the corresponding MR8 filter bank (in this case) was only able to retrieve 1 of 4 instances that matched the class label of the query instance.

D. Qualitative Results

The content based image retrieval in Figure 3 shows for a given training set and test image, the 4 closest neighbors retrieved when using Haar wavelets and MR8 filters. Along with a quantitative analysis, the difference between the MR8 and Haar Wavelets can be appreciated by viewing the difference between the nearest neighbors retrieved by the κ-NN classifier with respect to a query image (Figure 3). The distance between the texton signature of the query image and different database images is computed via the L2 norm. The database images with the minimum distance with respect to the training image are retrieved. Note that while in the case of the Haar wavelet response, the retrieved images correspond to the same class of the query images, fewer images corresponding to the class of the query image are retrieved in the context of the MR8 wavelet filter bank.

IV. Concluding Remarks

In this paper we presented a texture based classifier approach for identifying anaplastic medulloblastomas. These tumors typically have poor prognosis and there is a need for finding less invasive, cheap prognostic strategies for identifying them. By using a rotationally invariant filter bank in conjunction with the bag of words architecture, the problems associated with attempting macro-geometric feature based classification are avoided. By using textons to generate features, local spatial properties are the driving force behind the classification scheme. Many differences between anaplastic and non-anaplastic medulloblastoma are limited to cellular characteristics, explaining the success of leveraging local texture features in this classifier. The texton-based bag of words approach in conjunction with the Haar wavelet filter bank allowed us to achieve a classification accuracy of 0.80 in distinguishing between anaplastic and non-anaplastic medulloblastomas. The results were found to be superior to a MR8 filter bank based texton classifier. In future work we intend to validate our approach on a much larger data cohort.

References

[1] D. Ellison, "Classifying the medulloblastoma: insights from morphology and molecular genetics," *Neuropathology and Applied Neurobiology*, vol. 28, no. 4, pp. 257–282, 2002.

[2] C. G. Eberhart, J. L. Kepner, P. T. Goldthwaite, L. E. Kun, P. K. Duffner, H. S. Friedman, D. R. Strother, and P. C. Burger, "Histopathologic grading of medulloblastomas," *Cancer*, vol. 94, no. 2, pp. 552–560, 2002.

[3] A. Madabhushi, S. Agner, A. Basavanhally, S. Doyle, and G. Lee, "Computer-aided prognosis: Predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data." *Computerized medical imaging and graphics*, 2011 Feb 16 2011.

[4] S. Doyle, S. Agner, A. Madabhushi, M. D. Feldman, and J. Tomaszewski, "Automated grading of breast cancer histopathology using spectral clusteringwith textural and architectural image features," in *ISBI*, 2008, pp. 496–499.

[5] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "A boosted bayesian multi-resolution classifier for prostate cancer detection from digitized needle biopsies," *Biomedical Engineering, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2010.

[6] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, no. 1–2, pp. 61–81, Apr. 2005.

[7] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, "Histopathology image classification using bag of features and kernel functions," in *Proceedings of the 12th Conference on Artificial Intelligence in Medicine: Artificial Intelligence in Medicine*, ser. AIME '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 126–135.

[8] P. Khurd, C. Bahlmann, P. Maday, A. Kamen, S. Gibbs-Strauss, E. Genega, and J. Frangioni, "Computer-aided gleason grading of prostate cancer histopathological images using texton forests," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, april 2010, pp. 636 –639.

[9] A. Basavanhally, S. Ganesan, S. Agner, J. Monaco, M. Feldman, J. Tomaszewski, G. Bhanot, and A. Madabhushi, "Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology." *IEEE transactions on bio-medical engineering*, vol. 57, pp. 642–53, 2010 Mar 2010.

[10] Z. Struzik and A. Siebes, "The haar wavelet transform in the time series similarity paradigm," in *Principles of Data Mining and Knowledge Discovery*, ser. Lecture Notes in Computer Science, J. Zytkow and J. Rauch, Eds., vol. 1704. Springer Berlin / Heidelberg, 1999, pp. 12–22, 10.1007/978-3-540-48247-52.