STATISTICAL COLOR TEXTURE DESCRIPTORS FOR HISTOLOGICAL IMAGES ANALYSIS

Nicolas Hervé[a], Aude Servais[b], Eric Thervet[b], Jean-Christophe Olivo-Marin[a], Vannary Meas-Yedid[a]

[a] Institut Pasteur, Quantitative Image Analysis Unit, URA CNRS 2582, 75015 Paris, France
[b] Departments of Nephrology and Renal Transplantation, Necker Hospital, 75015 Paris, France

ABSTRACT

In this paper we compare different approaches to combine color and statistical texture descriptors. Previous studies on this topic were conducted on natural images only. We focus on the particular case of histological datasets where color plays an important role due to the staining process of the biological samples. We also introduce two new variants of the well-known Local Binary Patterns (LBP) operator. We test these approaches on three diversified histological datasets. We show that combining color and texture features extracted separately is preferable on datasets having a large variability in the staining, while simultaneous extraction of color and texture information is recommended only for more standardized stainings.

Index Terms— Color texture, LBP, Co-occurrence matrix, stained tissue, classification, histology

1. INTRODUCTION

In the field of computer aided diagnosis based on microscopy image assessment, an objective quantitative approach is required in order to increase reproducibility and predictive accuracy. In a biopsy image, biologically different parts of the tissue are characterized by a spatial organization of its cellular and connective tissues components that give raise to the tissue texture. Color is also of great importance for analyzing histology images. Indeed, the samples are usually prepared with some chemical solution that will enhance the contrast and stain in very distinctive colors specific parts of the cells or tissue that have been sampled. Common staining processes include H&E (hematoxylin and eosin) or Masson's trichrome. Although there exists many works applied to histological images in texture [1, 2, 3] or color analysis [4], only a few of them concerns both texture and color [5, 6]. Whatever the final application, the quality of low-level features extractors is essential for all image analysis processes.

In [7], the authors discuss pros and cons of using color and texture description jointly or separately for natural images. They conclude that using color version of texture descriptors should only be used in controlled environment regarding illumination conditions. Otherwise, extracting feature vectors separately (e.g. color histograms and texture histograms) and concatenating them works better. In [8], Palm has grouped the methods combining color and texture into three approaches: parallel, sequential and integrative, and claims that the integrative approach provides the best performances unlike the conclusions in [7].

The objective of this paper is to compare different combinations of color and texture features for the particular case of histological images in order to assess the best approach in this domain. We focus on statistical texture features as they provide very good results and also introduce two new variants for the LBP operator. We test these approaches on three different datasets. For comparison purposes, performances obtained with Gabor features are also reported.

2. APPROACHES

2.1. Local Binary Patterns

The Local Binary Patterns operator (LBP) [9] is now a standard method for analyzing textures. We consider a gray scale image $\mathbf{G}$ of $w \times h$ pixels. The basic idea is to describe locally the texture at a given pixel (x, y) by thresholding its neighbourhood by its intensity value $\mathbf{G}(x, y)$ and produce an index. Usually, the neighbourhood is defined by a number P of equally spaced pixels on a circle of radius R. Each neighbour pixel is defined by $N_p(x, y) = \left(x + Rcos(\frac{2\pi p}{P}), y - Rsin(\frac{2\pi p}{P})\right), p \in [0, P-1]$. Intensities of neighbour pixels $\mathbf{G}(N_p(x, y))$ are obtained by bilinear interpolation. Some other neighbourhood have also been introduced [3]. The LBP index is then defined as

$$LBP_{P,R}(\mathbf{G}, x, y) = \sum_{p=0}^{P-1} s\left(\mathbf{G}(x, y) - \mathbf{G}(N_p(x, y))\right) 2^p$$

with the standard function s for thresholding (equal to one if the argument is positive, zero otherwise). By accumulating LBP indexes in an histogram we obtain, for any given region of interest, a vector of size 2^P that is used as a feature. Each bin of this histogram is defined by

$$H(\mathbf{G}, i) = \sum_{x,y} \delta(i, LBP_{P,R}(\mathbf{G}, x, y)), i \in [0, 2^P - 1]$$

where δ is the Kronecker delta.

Several extensions have been proposed [9, 10] in order to improve the initial approach : rotation invariance, uniformity, contrast taken into account, multi-resolution, opponent colors, fuzzification, ternary patterns. Rotation invariance is achieved by reducing the number of possible LBP indexes, regrouping together similar local pattern that only differ by a rotation. In the same way, uniformity only considers simple local structures. More complex structures are accumulated in an extra bin of the histogram. Considered simulteanously, these two improvements lead to the $LBR^{riu}_{P,R}$ variant that can produce $P + 1$ distinct indexes.

A straightforward approach to extend texture descriptors to color images is to extract them separately on the three color channels, each one of them being considered as a grayscale image, and then combine the features obtained by concatenation. This is the single channel integrative approach described in [8] where cross-channels

is also considered (multiple channel integrative approach). For example, with $\mathbf{C}$ a RGB color image, we concatenante the three channelwise histograms :

$$H^{RGB}(\mathbf{C}, i) = H(\mathbf{C}^R, i) \oplus H(\mathbf{C}^G, i) \oplus H(\mathbf{C}^B, i)$$

In [10], soft histograms of LBP are introduced. Instead of hard-thresholding the intensity difference of the center pixel and its local neighborhood with s, a fuzzy function f is applied in order to smooth the LBP operator response.

$$f_\alpha(z) = \begin{cases} 0, & z < -\alpha \\ \frac{1+\frac{z}{\alpha}}{2}, & |z| \leq \alpha \\ 1, & z > \alpha \end{cases}$$

The amount of fuzzification is controled by adjusting the parameter α. The contribution of a given pixel (x, y) to bin i of the histogram is then

$$SLBP(\mathbf{G}, x, y, \alpha, i) = \prod_{p=0}^{P-1} [(2b_p(i) - 1) f_\alpha(\mathbf{G}(x, y) - \mathbf{G}(N_p(x, y))) + 1 - b_p(i)]$$

with $b_p(i) \in \{0, 1\}$ being the p-th bit of binary representation of i. The fuzzy histogram is then defined as

$$H_\alpha(\mathbf{G}, i) = \sum_{x,y} SLBP(\mathbf{G}, x, y, \alpha, i), i \in [0, 2^P - 1]$$

2.2. Proposed LBP variants

We propose to extend the rotation invariant and uniform approach to the soft LBP version. Indeed, the equations of ROR and U detailled in [9] cannot be applied directly on the $SLBP$ as they are now real values and not integers anymore. Nevertheless, they can efficiently be used for the histogram computation by using a look-up table LUT and marginalizing the accumulation of $SLBP$ in bins according to the rotation invariant and uniform version of the corresponding LBP index.

$$H_{\alpha,riu}(\mathbf{G}, i) = \sum_{x,y,j} SLBP(\mathbf{G}, x, y, \alpha, j), j \in LUT(i)$$

We have already shown in [11] that color quantization is an effective approach when segmenting some specific structures in histological images due to the staining process. A natural extension to such image analysis is thus to consider texture features on quantized colors. The objective we have in mind is to highlight some tissue structures that will be first isolated by some quantized colors before being described by texture features. Any quantization algorithm tries to find a set $\mathbf{Q}_n = \{q_b\}$ of n representative colors that minimizes the global error between the original image and its quantized version where each pixel color is replaced by the closest one in $\mathbf{Q}_n$ according to the distance function d. The value of n is either a parameter of the algorithm or is determined automatically during the quantization process. A membership function $\mathbf{M}$ may indicate the degree of similarity of the pixel color and its associated q_b. When a fuzzy clustering process is used (e.g. FCM), this membership function is part of the algorithm. Otherwise, as with KMeans or Split and Merge approaches, it may be computed afterwards. For example, we can define for each representative color $b \in [1, n]$ an image $\mathbf{M}^b(\mathbf{C})$ of memberships

$$\mathbf{M}^b(\mathbf{C}, x, y) = \frac{1}{d(\mathbf{C}(x, y), q_b) \sum_{k=1}^{n} \frac{1}{d(\mathbf{C}(x,y), q_k)}}$$

If a pixel color is the same as a representative color, $d(\mathbf{C}(x, y), q_b) = 0$, then the membership is 1 for the color b and 0 for all the other ones.

We propose to merge color and texture information extraction by applying fuzzy LBP on the membership images. This is a sequential approach as described in [8]. The Softly Quantized Color Local Binary Pattern, that provides our global feature vector, is therefore defined as the concatenation :

$$SQCLBP_\alpha(\mathbf{C}, i) = \bigoplus_{b=1}^{n} H_{\alpha,riu}(\mathbf{M}^b(\mathbf{C}), i)$$

2.3. Co-occurence matrices

Co-occurence matrices are also often used to extract texture informations [12]. Following our previous notations, we can define, for a quantized color image $\mathbf{C}_{Q_n}$, the co-occurence matrix COM. This is also a sequential approach. For a given spatial relationship, the co-occurence of two quantized colors (a and b) is mesured by a cell of this matrix. Here, we use the neighbourhood as defined by the LBP operator. We thus have :

$$COM_{P,R}(\mathbf{C}_{Q_n}, a, b) = \frac{1}{w \times h} \sum_{x=1}^{w} \sum_{y=1}^{h} \sum_{p=0}^{P-1} \delta\left(a, \mathbf{C}_{Q_n}(x, y)\right) \times \delta\left(b, \mathbf{C}_{Q_n}(N_p(x, y))\right)$$

This matrix can be used directly as a descriptor or to extract features, known as Haralick features. We choose to use the generalized contrast feature to marginalize the co-occurence matrix. Note that the distance we use in the contrast measure is computed in the color space. The result is a n-dimensional feature vector :

$$CCOM_{P,R}(\mathbf{C}_{Q_n}, a) = \sum_{b=1}^{n} COM_{P,R}(\mathbf{C}_{Q_n}, a, b) \times d(q_a, q_b)$$

3. EXPERIMENTAL RESULTS

3.1. Datasets

We have tested the different color and texture descriptors individually and in combination on three histological datasets. Two of them come from a publicly available benchmark suite [13]. The last dataset has been assembled internally and is also provided publicly to foster comparison of different approaches[1].

We used the Liver Gender dataset for 6-month old mices under ad libitum diet (LG6MAL). The goal is to distinguish the samples extracted from male and female mices. These samples are stained using H&E. A single person has prepared all the staining and imaging, leading to a small variability. We reduced the original image dimensions by a factor 4 (leading to 347×260 pixels). As the acquisition was done using 12 bits per color channel, we also streched the histograms so as to fully cover the 16 bits encoding. This set has 265 images.

The Lymphoma dataset (LYMPH) contains three types of malignant lymphoma (chronic lymphocytic leukemia, follicular lymphoma and mantle cell lymphoma). The goal is to distinguish these three classes (CLL, FL and MCL). The biopsies have been prepared by different pathologists at different sites. Thus, we face a huge variability in the staining and imaging quality. The samples are stained using H&E. As with the previous datased, we reduced the image dimensions by a factor 4. This set has 374 images.

[1] GlomDB : http://www.bioimageanalysis.org/glomdb.html

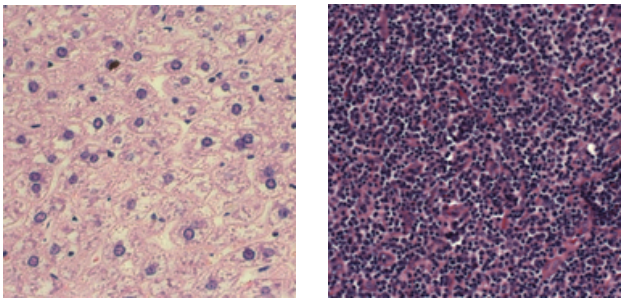

Fig. 1. (L) Male liver - (R) FL type lymphoma

The GLOMDB dataset has been specifically designed by us to test color and texture descriptors. We are interested in quantifying interstitial fibrosis on renal biopsies [11], and need to assess the performance of joint color and texture information. In order to finely segment the images, we need to properly detect the glomeruli structures of the kidney. We used 15 different biopsies, stained using Masson's trichrome, that were collected and processed by different operators at different times in the same hospital. All the imaging was done using the same microscope (Zeiss Mirax Scan, 20X objective, NA 0.8) under identical illumination context. We manually segmented all the glomeruli and then automatically extracted 16×16 pixels non-overlapping square patches. This set has 1976 images of texture, half of which are glomeruli.

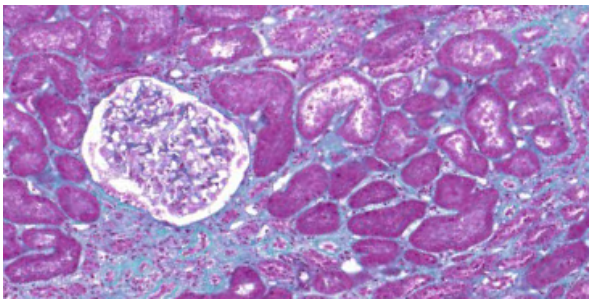

Fig. 2. Part of a renal biopsie with a glomerulus on the left

3.2. Experimental setup

On all three datasets, we used soft margin support vector machines (SVM) as classifiers with a triangular kernel. It has the great advantage of being non-parametric while able to adapt to any scale in the feature space. We choose the simple approach that assimilates the score of the SVM decision function in a confidence level, following the intuitive idea that being far from the decision boundary is equivalent to being less ambiguous for a concept. This setup has proven to perform well with statistical visual descriptors [14].

We use the Mean Average Precision measure (MAP) to evaluate the performances as it encompasses both precision and recall in a single figure. Datasets are randomly split in a training and a testing set. We choose to have only a few training examples so as to simulate the difficult conditions we usually face. For LG6MAL and LYMPH, 5% of the images are used in the training set. We keep only 1% for the GLOMDB dataset. Once the SVM has been trained, we rank the full testing set according to the score provided by the classifier and compute the average precision. We choose a one-against-all configuration. For LG6MAL, we test the male class. For GLOMDB, we test the glomerulus class. For LYMPH, we test the three classes and average the results. All experiments are launched 5000 times, the reported results are averaged. As some datasets are unbalanced, we also report the MAP obtained for random ranking of the sets.

	Desc.	GLOMDB	LG6MAL	LYMPH
	Random	**0.503**	**0.572**	**0.341**
Color	$I1H2H3_3$	0.850	0.736	0.546
	$I1H2H3_5$	0.886	**0.820**	0.578
	$I1H2H3_{10}$	0.911	0.784	0.594
	SM_8	0.899	0.757	0.532
	SM_{16}	0.904	0.791	0.577
	SM_{20}	**0.917**	0.777	0.580
	SM_{32}	0.913	0.776	0.590
	SM_{64}	0.909	0.775	**0.604**
Texture	$Gabor$	0.863	0.922	0.517
	LBP	0.705	0.945	0.529
	LBP_{riu}	0.723	0.922	0.506
	$SLBP$	0.919	0.949	0.535
	$SLBP_{riu}$	**0.922**	**0.954**	**0.540**
Sep.	$I1H2H3 \oplus SLBP_{riu}$	0.946	**0.875**	0.617
	$SM \oplus SLBP_{riu}$	**0.952**	0.842	**0.631**
Join.	$Gabor, I1H2H3$	0.894	0.962	0.583
	$SLBP_{riu}, I1H2H3$	**0.974**	**0.973**	0.576
	$SQCLBP$	0.937	0.954	**0.632**
	$CCOM$	0.933	0.825	**0.633**

3.3. Results

We first test separately color and texture descriptors to assess the respective importance of these two feature categories and to establish baseline results. We extract color features with simple histograms. We use two types of color space quantization. First, we use a standard fixed grid on the $I1H2H3$ color space that splits each axis in n intervals, producing n^3 bins. The hybrid colorspace $I1H2H3$ is defined by $I1 = \frac{R+G+B}{3}$, $H2 = R - G$ and $H3 = \frac{R+G}{2} - B$. It was shown to be accurate for Masson trichrome histology slides segmentation [11]. The results are only reported for this color space, but we obtained very similar performances with HSV, RGB and $L^*a^*b^*$ spaces. Then we use a clustering approach in order to quantize finely the colors of the full datasets. As in [11], we choose a Split and Merge algorithm (SM) and report results for different number of clusters. We have unified the notations to define the neighbourhood of LBP variants and co-occurence matrices. In the following, we choose to work with 8 neighbours per pixel, thus we have $P = 8$ for all descriptors and datasets. The other parameters of the texture descriptors have been set by cross-validation. The radius R has been determined with LBP and the same value has been used with the other descriptors. This parameter depends on the intrinsic scale of the tissue structures and the acquisition context (microscope objective, image resizing). We have for GLOMDB : $R = 1, \alpha = 25$, for LG6MAL : $R = 4, \alpha = 25$ and for LYMPH : $R = 1.5, \alpha = 20$. We can notice that the fuzzification parameter α is almost the same for the three datasets. We use the Gabor wavelet features described in [15] as a reference : normalized mean and standard deviation of the magnitude with 8 orientations and a single scale (as LBP approaches).

We have tested the two approaches described in [7] to combine

color and texture informations. We first extract each type of information separately and concatenate the features vectors (this approach is called parallel in [8]). For each dataset, we use the best version of color and texture descriptor taken alone. Then, we extract color and texture jointly. We use the best texture descriptor, $SLBP_{riu}$, on the three color channels of the $I1H2H3$ color space (single channel integrative method). The feature vectors are then concatenated. The other approach is first to quantize the datasets with the Split an Merge algorithm and extract texture information with our LBP variant ($SQCLBP$) and with the contrast version of co-occurence matrices ($CCOM$). For each dataset, we use the number of cluster n that gives the best performances for the SM color histogram taken alone.

4. DISCUSSION AND CONCLUSION

The three histological datasets we used are quite diversified (different organ tissues, purposes, staining, imaging, variability). We first analyze the results of the descriptors alone to know if one feature category is dominant over the other one. When looking at the first half of the results table, we can easily see that for GLOMDB, both color and texture taken alone achieve the same performances. On LG6MAL, texture is clearly the dominant feature, unlike LYMPH where color alone provides better results. Moreover, using the Split and Merge quantized histogram works better than fixed grid splitting of the color space for GLOMDB and LYMPH. For the texture on the grayscale images, the proposed $SLBP_{riu}$ approach provides the best performances on the three datasets. When combined with color histograms, we find the same conclusions as before (SM works better for GLOMDB and LYMPH, and grid split for LG6MAL). More surprisingly, if the performances are slightly improved for GLOMDB and LYMPH over the descriptors taken alone, this is not the case for LG6MAL. This is probably explained by the gap between the poor performances of the color descriptors and the good results of the texture on this dataset. The combination by concatenation provides an average and the classifiers are not able to take any advantage or, at least, keep the base performances. When color and texture are considered jointly, the best overall performances are reached on all three datasets. $SLBP_{riu}$ on the color channels performs best for GLOMDB and LG6MAL. For LYMPH, the two approaches based on quantized color spaces obtain the same performances as with the previous combination approach. On all datasets, both on grayscale and color images, the $Gabor$ features achieve slightly worse performances than $SLBP$.

If we average the performances on the three datasets, the descriptors $SLBP_{riu}$ on $I1H2H3$ and $SQCLBP$ obtain exactly the same results (MAP 0.841). As a first conclusion, this would seem to indicate that no method outperform the other when no *a priori* information is known about the dataset. However, in the case of histological images, we draw the same conclusions as Mäenpää and Pietikäinen in [7]. Indeed, based on the datasets descriptions and a manual inspection, we can say that there is no variability of the staining in LG6MAL. In GLOMDB, the variability exists but is limited and in LYMPH it is obviously present. Using the color version of $SLBP$ works better than the histograms concatenation for the two datasets that have limited variability. However considering color and texture separately provides better performances for LYMPH, the set that has the highest variability in images.

5. ACKNOWLEDGEMENT

We thank L.H. Noël from the Necker hospital in Paris for the renal biopsy samples. This project is supported by ATN.

6. REFERENCES

[1] K. Masood and N. Rajpoot, "Texture based classification of hyperspectral colon biopsy samples using clbp," in *IEEE Int. Symp. Biomedical Imaging*, 2009, pp. 1011–1014.

[2] L. Sørensen, S. B. Shaker, and M. de Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Medical Imaging*, vol. 29, no. 2, pp. 559–569, 2010.

[3] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis.," *Artificial Intelligence in Medecine*, vol. 49, no. 2, pp. 117–125, 2010.

[4] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N.E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *IEEE Int. Symp. Biomedical Imaging*, 2009, pp. 1107–1110.

[5] A. Tabesh, M. Teverovskiy, H. Pang, V. Kumar, D. Verbel, A. Kotsianti, and O. Saidi, "Multifeature prostate cancer diagnosis and gleason grading of histological images.," *IEEE Trans. Medical Imaging*, vol. 26, no. 10, pp. 1366–1378, 2007.

[6] O. Sertel, J. Kong, G. Lozanski, A. Shana'ah, U. Catalyurek, J. Saltz, and M. Gurcan, "Texture classification using nonlinear color quantization: Application to histopathological image analysis," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2008, pp. 597–600.

[7] T. Mäenpää and M. Pietikäinen, "Classification with color and texture: jointly or separately?," *Pattern Recognition*, vol. 37, no. 8, pp. 1629 – 1640, 2004.

[8] C. Palm, "Color texture classification by integrative co-occurrence matrices," *Pattern Recognition*, vol. 37, no. 5, pp. 965 – 976, 2004.

[9] T. Mäenpää and M. Pietikäinen, *Handbook of Pattern Recognition and Computer Vision*, chapter Texture analysis with local binary patterns., pp. 197–216, World Scientific, 2005.

[10] T. Ahonen and M. Pietikäinen, "Soft histograms for local binary patterns," in *Finnish Signal Processing Symp.*, 2007.

[11] N. Hervé, A. Servais, E. Thervet, J.-C. Olivo-Marin, and V. Meas-Yedid, "Improving histology images segmentation through spatial constraints and supervision," in *IEEE Int. Conf. Image Processing*, 2010.

[12] O. Sertel, J. Kong, U. Catalyurek, G. Lozanski, Saltz J., and M. Gurcan, "Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading," *Signal Processing Systems*, vol. 55, issue 1-3, pp. 169–183, 2009.

[13] L. Shamir, N. Orlov, D. Eckley, T. Macura, and I. Goldberg, "Iicbu 2008: a proposed benchmark suite for biological image analysis," *Medical and Biological Engineering and Computing*, vol. 46, pp. 943–947, 2008.

[14] N. Hervé and N. Boujemaa, "Image annotation : which approach for realistic databases ?," in *ACM Int. Conf. Image and Video Retrieval*, 2007.

[15] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, 1996.