Studying the Effect of Digital Stain Separation of Histopathology Images on Image Search Performance

Alison K. Cheeseman[1], Hamid R. Tizhoosh[2], and Edward R. Vrscay[1(✉)]

[1] Department of Applied Mathematics, Faculty of Mathematics,
University of Waterloo, Waterloo, ON N2L 3G1, Canada
{alison.cheeseman,ervrscay}@uwaterloo.ca
[2] Kimia Lab, University of Waterloo, Waterloo, ON N2L 3G1, Canada
hamid.tizhoosh@uwaterloo.ca

Abstract. Due to recent advances in technology, digitized histopathology images are now widely available for both clinical and research purposes. Accordingly, research into computerized image analysis algorithms for digital histopathology images has been progressing rapidly. In this work, we focus on image retrieval for digital histopathology images. Image retrieval algorithms can be used to find similar images and can assist pathologists in making quick and accurate diagnoses. Histopathology images are typically stained with dyes to highlight features of the tissue, and as such, an image analysis algorithm for histopathology should be able to process colour images and determine relevant information from the stain colours present. In this study, we are interested in the effect that stain separation into their individual stain components has on image search performance. To this end, we implement a basic k-nearest neighbours (kNN) search algorithm on histopathology images from two publicly available data sets (IDC and BreakHis) which are: a) converted to greyscale, b) digitally stain-separated and c) the original RGB colour images. The results of this study show that using H&E separated images yields search accuracies within one or two percent of those obtained with original RGB images, and that superior performance is observed using the H&E images in most scenarios we tested.

Keywords: Digital histopathology · Encoded Local Projections (ELP) · Digital stain separation · Digital image retrieval and classification

1 Introduction

Histopathology, the examination of tissue under a microscope to study biological structures related to disease manifestation, has traditionally been carried out

manually by pathologists working in a lab. It is only in recent years that the technology has advanced to a point which allows for the rapid digitization and storage of whole slide images (WSIs). Consequently, digitized histopathology images are now widely available for both clinical and research purposes, and computerized image analysis for digital histopathology has quickly become an active area of research [8,11]. In this paper, we focus specifically on content-based image retrieval (CBIR) for histopathology images, which involves finding images which share the same visual characteristics as the query image. The identification and analysis of similar images can assist pathologists in quickly and accurately obtaining a diagnosis by providing a baseline for comparison.

While most radiology images (X-ray, CT, etc.) are greyscale, histopathology images are typically stained with dyes to highlight certain features of the tissue. In order to properly use the relevant colour information, a WSI analysis system should be able to process colour images and determine relevant biological information from the presence of different stain colours. In a previous work [4], we introduced a new frequency-based ELP (F-ELP) image descriptor for histopathology image search, which captures local frequency information and implemented this new descriptor on images which were separated into two colour channels based on chemical stain components using a digital stain separation algorithm. In [4], we found that both the ELP and F-ELP descriptors saw improved search accuracy when applied to the stain-separated images, as opposed to single-channel greyscale images. In this paper, we focus on studying the effectiveness of digital stain separation of histopathology images for image retrieval applications using a number of common handcrafted image descriptors. We compare the results of image retrieval for the stain-separated images to the results of the same experiment conducted on both greyscale and colour (three-channel RGB) images. Experiments are conducted using two publicly available breast cancer histopathology data sets, IDC and BreakHis.

2   Digital Stain Separation

The most common staining protocol for histology slides involves two dyes, namely hematoxylin and eosin (H&E). The hematoxylin stains cell nuclei blue, and eosin stains other structures varying shades of red and pink [16]. The colours which appear in a slide, and the size, shape and frequency at which they appear are all relevant factors a pathologist may assess when making a diagnosis. For this reason, we consider separating the input images into two stain components prior to the computation of an image descriptor.

In this paper, as in [4], we adopt the stain separation method proposed in [12], an extension of the wedge finding method from [10]. Unlike some previous methods for stain separation [16], this method does not require any calibration or knowledge of the exact stain colours, instead it works by using the available image data to estimate an H&E basis. Given that an image search algorithm should ultimately be applied to data from multiple sources, this is a desirable feature for the stain separation algorithm.

Figure 1 shows two sample images from the BreakHis data set which have been separated in their hematoxylin and eosin components. We can see that the algorithm is able to effectively separate the two components of both images, even though the stains appear as noticeably different colours in each image.
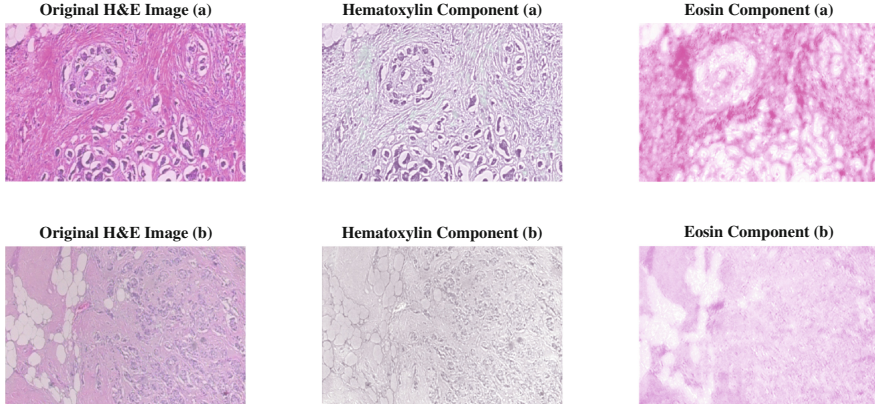

Fig. 1. Two sample images from the BreakHis dataset showing the resulting hematoxylin and eosin components after applying the stain separation algorithm from [12].

3   Proposed Study

The main purpose of this study is to investigate the effectiveness of digital stain separation in isolation from the other parameters of the image search process. To this end, we implement a basic image search algorithm, k-Nearest Neighbours (kNN), using input images that are: a) converted to greyscale (one colour channel), b) separated into their H&E components (two colour channels) by the method described above, and c) the original RGB colour images (three colour channels). For each colour channel of an input image having N total channels, an image descriptor, $\mathbf{h}_i$ is computed, and concatenated to form the final descriptor $\mathbf{h} = [\mathbf{h}_1 \; ... \; \mathbf{h}_N]$, which represents the entire image. We then implement the kNN algorithm for image search using a number of well-known distance functions to determine "nearest" neighbours and four different image descriptors of varying lengths and properties as inputs to the algorithm. The image descriptors and distance functions used are described in the following sections.

3.1   Image Descriptors

Feature extraction, or the computation of compact image descriptors, is an important part of many image analysis tasks, including image retrieval. As such, there has been plenty of research over the years into the design of image

descriptors for various imaging applications. Some of the most well known image descriptors include: local binary patterns (LBP) [1], the scale-invariant feature transform (SIFT) [9], speeded-up robust features (SURF) [3], and histograms of oriented gradients (HOG) [6]. Most of these methods, including SIFT, SURF, and HOG, perform well in more traditional applications such as object detection or tracking and face recognition, but perform poorly compared to LBP for the retrieval and classification of histopathology images [2,18]. LBP, which is based on computing binary patterns in local regions of the image, is generally thought to be a better image descriptor for textures, which may explain its superior performance on high resolution histopathology images, which resemble textures more than natural images.

In our study, we implement four image descriptors, including the LBP descriptor, along with the Gabor filter-based GIST descriptor [15] which computes the spatial envelope of a scene, the ELP descriptor (encoded local projections) from [18] which was designed with medical images in mind, and our proposed F-ELP descriptor from [4], designed specifically to be a compact descriptor for histopathology images.

3.2 Distance Functions for Image Search

In any image search algorithm, it is important to properly define what makes two images similar. Typically, that means one must choose an appropriate distance function between image descriptors. Six different distance functions were used in this study to determine the nearest neighbours for the kNN search, including the well-known L_1, L_2, cosine, correlation, and chi-squared metrics. We also consider the Hutchinson (also known as Monge-Kantorovich) distance [13], as it is thought to be a good measure of distance between histograms. In the finite one-dimensional case, the Hutchinson distance can be computed in linear-time using the method from [14].

4 Data Sets and Image Preprocessing

In this study, we used two publicly available data sets containing breast cancer histopathology images: IDC and BreakHis.

Invasive Ductal Carcinoma (IDC) Kaggle Data: The IDC dataset consists of digitized breast cancer slides from 162 patients diagnosed with IDC at the University of Pennsylvania Hospital and the Cancer Institute of New Jersey [5] (Fig. 2). Each slide was digitized at 40x magnification and downsampled to a resolution of $4\,\mu$m/pixel. The dataset provides each WSI split into patches of size $50\,\text{px} \times 50\,\text{px}$ in RGB colour space. The supplied data was randomly split into three different subsets of 84 patients for training, 29 for validation and 49 test cases for final evaluation. Ground truth annotation regarding the presence of IDC in each patch was obtained by manual delineation of cancer regions performed by expert pathologists.

Due to their small size, each individual image patch in the IDC data set may not contain both hematoxylin and eosin stains. Since the stain separation algorithm learns the stain colours from the data, both stains must be present in the image for accurate results. To ensure good performance on all image patches, we use the entire WSI to perform stain separation and then split the image back into the original patches to compute image descriptors. One further issue is that the stain separation algorithm used assumes that two (and only two) stain components (H&E in our case) exist in the image. However, some images are observed to have significant discolouration, such as large dark patches, and the introduction of other colours not caused by H&E staining. The prevalence of such artefacts negatively impacts the ability of the stain separation algorithm to provide good results for some patients, so we remove them by searching for images which have minimal variation in the RGB channels across the entire image. A total of 686 patches were flagged and removed from the total data set, all of which contain significant artefacts or discoloration.

Breast Cancer Histopathology Database (Breakhis): The Breast Cancer Histopathology Database (BreakHis) [17] was built as a collaboration between researchers at the Federal University of Parana (UFPR) and the P&D Laboratory - Pathological Anatomy and Cytopathology, in Parana, Brazil (Fig. 3). To date, it contains 9,109 images of breast tumour tissue from 82 patients using four different magnification factors: $40\times$, $100\times$, $200\times$, and $400\times$. The images are provided in PNG format (3-channel RGB, 8-bit depth/channel) and are 700×460 px. The data is divided into two classes, benign tumours and malignant tumours, with class labels provided by pathologists from the P&D Laboratory. Within each class, further labelling is provided to indicate tumour types. The data set consists of four histologically distinct benign tumours and four malignant tumour types. These additional intra-class labels are not used in the current study.

For this particular study, we use only the images taken at $40\times$ magnification. This subset of the data contains 1,995 images, of which 652 are benign and 1,370 are malignant. Using code provided by the authors of [17], the data is split into a training (70%) and testing (30%) set with the condition that patients in the training set are not used for the testing set. The results presented in this paper are the average of five trials, using the five data folds from [17].
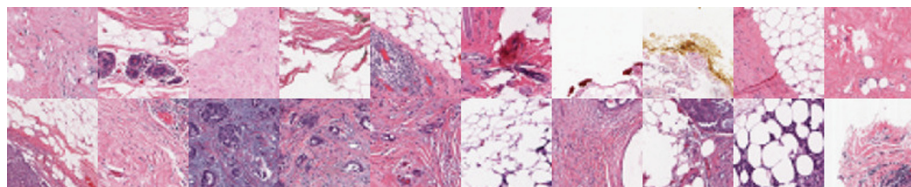

Fig. 2. Sample patches from the IDC data set. The top row shows negative examples (healthy tissue or non-invasive tumour tissue) and the bottom row shows positive examples (IDC tissue).
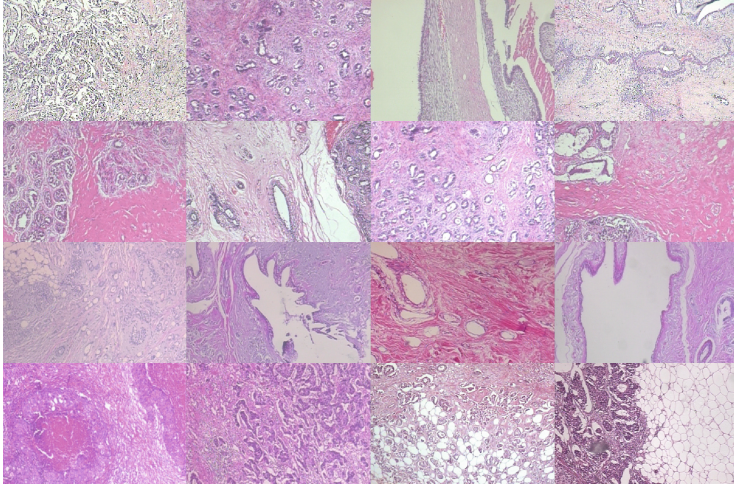
Fig. 3. Example of image patches from the BreakHis data set. The top two rows show examples of benign tumours and the bottom two rows show malignant tumours.

4.1 Accuracy Calculations

IDC: For consistency with previous works involving the IDC data, we use both the balanced accuracy (BAC) and F-measure (F1), which are defined as follows [5]:

$$\text{BAC} = \frac{\text{Sen} + \text{Spc}}{2},\ F1 = \frac{2 \cdot \text{Pr} \cdot \text{Rc}}{\text{Pr} + \text{Rc}}, \tag{1}$$

where Sen is sensitivity, Spc is specificity, Pr is precision and Rc, recall.

BreakHis: For the BreakHis data we compute patient scores and the global recognition rate, which were introduced in [17]. If we let N_P be the number of images of patient P and N_{rec} be the number of images of patient P that are correctly classified, then the patient score for patient P is defined as

$$\text{Patient Score} = \frac{N_{\text{rec}}}{N_P} \tag{2}$$

and the global recognition rate (GRR) as

$$\text{Global Recognition Rate} = \frac{\sum \text{Patient scores}}{\text{Total number of patients}}. \tag{3}$$

In addition to the global recognition rate we also compute the balanced accuracy as defined above in (1).

5   Experiment

In order to evaluate image retrieval performance we implement the kNN algorithm in MATLAB with each set of image descriptors as inputs. The kNN algorithm searches through the training data partition and classifies each image in the test set based on the class of its k nearest neighbours. Since there is no exact metric to quantitatively evaluate image retrieval performance, we measure the accuracy of classification using kNN. In this work, we test the kNN algorithm using three different values for k ($k = 1, 5$ and 15).

Table 1. A list of image descriptors used in this study and the corresponding number of features computed (i.e. the length of the feature vector).

Descriptor	Number of features		
	Greyscale	H&E stains	RGB image
ELP	1024	2048	3072
GIST	512	1024	1536
F-ELP	32	64	96
LBP	18	36	54

Table 1 lists the image descriptors used and their respective lengths on each set of input colour channels. We can see that as we increase the number of input colour channels from one to three, the length of the feature vectors increases. Given that the computation time for the kNN search algorithm has linear dependency on feature vector length [7], it is clear that for a fixed image descriptor, it is desirable to use fewer colour channels, so long as the overall search performance does not suffer significantly.

The length of each image descriptor is dependent on certain parameters of the algorithm. In this work the following parameters are used: the ELP and F-ELP descriptors are implemented with a window size of $w = 9$, the GIST descriptor, by default, divides the image into a 4×4 grid and uses a filter bank of 32 Gabor filters, and the LBP descriptor is computed with a radius of $R = 2$ and $P = 16$ neighbouring pixels. As a result, we have a wide variety of descriptor lengths, from the very short LBP descriptor to the long ELP histogram.

5.1   Comparing Input Image Colour Channels

In this section, we present results which compare the image search performance using greyscale, H&E stain-separated, and RGB images as inputs. For each descriptor, and each set of input colour channels, the best accuracy, taken over all distance functions, is presented. It should be noted here that there are some slight discrepancies between the results presented here for the IDC data set and those in our previous work [4]. This is due to a small error which was found in the code which slightly changes the numerical results, but does not change the overall conclusions of the previous study.

Table 2. The best KNN search ($k = 1$) accuracy for the IDC dataset taken over all distance functions. The top result in each column is highlighted in bold.

Colour channels	ELP		GIST		F-ELP		LBP	
	F1	BAC	F1	BAC	F1	BAC	F1	BAC
Greyscale	0.3987	0.5905	0.5086	0.6500	0.4183	0.6023	0.4625	0.6280
H&E stains	**0.4528**	**0.6235**	**0.5549**	**0.6923**	**0.5565**	**0.6932**	0.5860	0.7130
RGB image	0.4504	0.6219	0.5513	0.6890	0.5419	0.6836	**0.5926**	**0.7187**

Table 3. The best KNN search ($k = 5$) accuracy for the IDC dataset taken over all distance functions. The top result in each column is highlighted in bold.

Colour channels	ELP		GIST		F-ELP		LBP	
	F1	BAC	F1	BAC	F1	BAC	F1	BAC
Greyscale	0.4001	0.6080	0.5598	0.6968	0.4338	0.6208	0.5124	0.6645
H&E stains	**0.4880**	**0.6531**	**0.6052**	**0.7356**	**0.6303**	**0.7406**	0.6618	0.7614
RGB image	0.4832	0.6504	0.5918	0.7270	0.6028	0.7222	**0.6785**	**0.7750**

Table 4. The best KNN search ($k = 15$) accuracy for the IDC dataset taken over all distance functions. The top result in each column is highlighted in bold.

Colour channels	ELP		GIST		F-ELP		LBP	
	F1	BAC	F1	BAC	F1	BAC	F1	BAC
Greyscale	0.3839	0.6069	0.5910	0.7207	0.4218	0.6199	0.5396	0.6825
H&E stains	**0.4943**	**0.6589**	**0.6283**	**0.7570**	**0.6697**	**0.7665**	0.6887	0.7780
RGB image	0.4912	0.6569	0.6147	0.7448	0.6404	0.7462	**0.7125**	**0.7972**

IDC: Tables 2, 3 and 4 show the results for the IDC data set for kNN search with $k = 1, 5$ and 15, respectively. As expected, since coloured images contain relevant information which may be lost when converted to greyscale, we observe that using either the H&E stain separated image or the total RGB image is always an improvement over using the greyscale image. A more interesting comparison comes from looking at the bottom two rows of the tables, comparing the H&E images to the RGB images. We see that generally the F1 scores and balanced accuracies are similar (within one or two percent) for both H&E and RGB images. For all descriptors besides LBP, we actually observe an improved performance using the H&E image over RGB, despite the fact that the input image has less colour channels, and thus the feature vector is shorter.

We also, not surprisingly, observe that as k is increased, the search performance tends to improve, although the jump from $k = 1$ to 5 is quite a bit larger than the jump from 5 to 15. This may indicate that $k = 15$ is near an optimal value for k. We also see that for this particular data set, the LBP descriptor gives the highest accuracy, and the ELP descriptor performs the worst.

BreakHis: Similarly, for the BreakHis data set, Tables 5, 6 and 7 show the best global recognition rates and balanced accuracies for each image descriptor and set of input colour channels. Once again, we observe a general increase in search accuracy when using more than one input colour channel (H&E or RGB) as compared to the greyscale images. In many cases, in addition to the decreased computational cost of using fewer colour channels, for the BreakHis data set we see that there is another benefit to using stain separated images, which is a significant improvement in search performance.

Table 5. The best KNN search ($k = 1$) accuracy for the BreakHis dataset taken over all distance functions. The top result in each column is highlighted in bold.

Colour channels	ELP		GIST		F-ELP		LBP	
	GRR	BAC	GRR	BAC	GRR	BAC	GRR	BAC
Greyscale	0.6584	0.5812	0.6589	0.5602	0.6534	0.5577	0.6874	0.6212
H&E stains	**0.7532**	**0.7047**	**0.7083**	**0.6456**	**0.7433**	**0.6823**	**0.7397**	**0.6903**
RGB image	0.6604	0.5971	0.6578	0.5787	0.7358	0.6812	0.6689	0.6115

As before, we see that search performance tends to increase with increasing k, but does seem to level off around $k = 15$. Unlike our previous results on the IDC data set, we see here that the best search performance occurs using the ELP descriptor and F-ELP descriptors, while the GIST descriptor fares the worst on the BreakHis data. Due to the higher intra-class variation of the BreakHis data (multiple tumour types for benign and malignant classes), it makes sense that the balanced accuracies are generally lower on this data set.

Table 6. The best KNN search ($k = 5$) accuracy for the BreakHis dataset taken over all distance functions. The top result in each column is highlighted in bold.

Colour channels	ELP		GIST		F-ELP		LBP	
	GRR	BAC	GRR	BAC	GRR	BAC	GRR	BAC
Greyscale	0.6838	0.5844	0.6940	0.5808	0.6802	0.5637	0.6952	0.6049
H&E stains	**0.7602**	**0.7078**	**0.7340**	**0.6324**	**0.7557**	**0.6915**	**0.7323**	**0.6662**
RGB image	0.6744	0.5931	0.6884	0.6113	0.7480	0.6865	0.6977	0.6335

Table 7. The best KNN search ($k = 15$) accuracy for the BreakHis dataset taken over all distance functions. The top result in each column is highlighted in bold.

Colour channels	ELP		GIST		F-ELP		LBP	
	GRR	BAC	GRR	BAC	GRR	BAC	GRR	BAC
Greyscale	0.7085	0.5898	0.7128	0.5744	0.6957	0.5616	0.7051	0.5979
H&E stains	**0.7660**	**0.7060**	**0.7406**	**0.6286**	0.7737	**0.7033**	**0.7564**	**0.6885**
RGB image	0.6892	0.6008	0.7068	0.6090	**0.7749**	**0.7033**	0.7023	0.6219

5.2   Comparing Distance Functions

In this section, we consider the effect that the choice of distance function has on image search performance for each image descriptor. To do so, we introduce a ranking of each distance function, based on the balanced accuracy result. For a given image descriptor and choice of input colour channels (greyscale, H&E, or RGB) we rank each distance function based on the resulting balanced accuracy as a percentage of the maximum balanced accuracy for that particular search trial. Results presented here show the averaged distance ranking for each distance and each descriptor, averaged over the choice of input colour channels, and the choice of k for the kNN search algorithm. We present results only for the balanced accuracy, as the results for the F1 measure (IDC data) and global recognition rate (BreakHis data) follow similar trends.

IDC: Figure 4 shows the average ranking of all six distance functions tested for each image descriptor on the IDC data. We observe that, in general, the variation in search accuracy caused by the choice of distance function is relatively low, with the exception of the ELP descriptor, where a noticeable variation can be seen. For all but the ELP descriptor, it would be difficult to pinpoint which distance function is the "best" choice. In particular, for the F-ELP and LBP descriptors, the variation in performance across distance functions is almost non-existent.
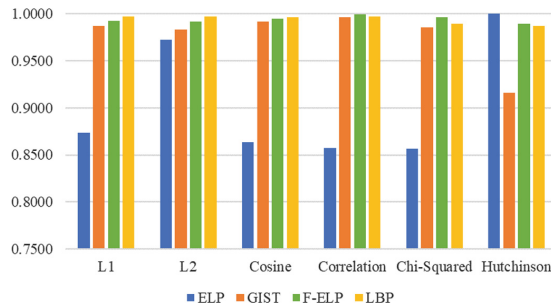

Fig. 4. A comparison of the average ranking of distance functions for each image descriptor on the IDC data set.

BreakHis: In Fig. 5 we show the average rankings of each distance function on the BreakHis data. Once again, the overall change in the accuracy as a result of the choice of distance function is surprisingly low for all descriptors. As with the IDC data, we see the most noticeable variation in performance with the ELP and GIST descriptors. The results for both the IDC and BreakHis data sets do not give any indication that one distance function is necessarily superior for image search even for a fixed image descriptor. Over many tests, we see only

one scenario (the ELP descriptor applied to the IDC data) where the choice of distance function significantly impacts the results. Further testing on more data is of course possible, however for good performance and generalisation to unknown data, it would seem that the best choice is simply to use the distance function which can be computed most efficiently.
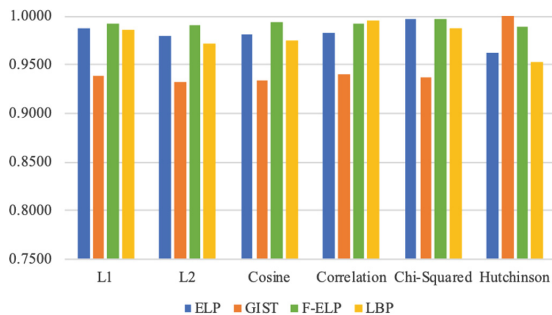

Fig. 5. A comparison of the average ranking of distance functions for each image descriptor on the BreakHis data set.

6   Conclusion

In this paper, we have investigated the effect of using digitally stain separated images, as compared to greyscale and RGB, for image retrieval applications. Our results are obtained through testing on two data sets containing breast cancer histopathology images. We find that separating images into their H&E stain components leads to a significant increase in search performance over simply using the greyscale images, as expected. More interestingly, we find that using H&E separated images yields search accuracies within one or two percent of those obtained with the original RGB images, despite the fact that the H&E images have only two colour channels. In fact, superior performance is observed using the H&E images in most tested scenarios. Given the improved computation speed afforded by using fewer image channels, it is reasonable to conclude that using H&E stain separated images is preferable to using the overall RGB images for image search. As well, ELP appears to benefit from investigations on choosing the distance metric, a factor that should be considered when using this descriptor.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
2. Alhindi, T.J., Kalra, S., Ng, K.H., Afrin, A., Tizhoosh, H.R.: Comparing LBP, HOG and deep features for classification of histopathology images (2018)

3. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (SURF). Comput. Vis. Image Und. **110**(3), 346–359 (2008)
4. Cheeseman, A.K., Tizhoosh, H., Vrscay, E.R.: A compact representation of histopathology images using digital stain separation and frequency-based encoded local projections. In: Karray, F., Campilho, A., Yu, A. (eds.) ICIAR 2019. LNCS, vol. 11663, pp. 147–158. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-27272-2_13
5. Cruz-Roa, A., et al.: Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: Progress in Biomedical Optics and Imaging - Proceedings of SPIE 9041 (2014)
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, pp. 886–893 (2005)
7. Friedman, J.H., Bentley, J.L., Finkel, R.A.: An algorithm for finding best matches in logarithmic expected time. ACM Trans. Math. Softw. **3**(3), 209–226 (1977)
8. Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B.: Histopathological image analysis: a review. IEEE Rev. Biomed. Eng. **2**(2), 147–171 (2009)
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
10. Macenko, M., et al.: A method for normalizing histology slides for quantitative analysis. In: Proceedings of International Symposium on Biomedical Imaging, Chicago, pp. 1107–1110 (2009)
11. Madabhushi, A.: Digital pathology image analysis: opportunities and challenges. Imaging Med. **1**(1), 7–10 (2009)
12. McCann, M.T., Majumdar, J., Peng, C., Castro, C.A., Kovacevic, J.: Algorithm and benchmark dataset for stain separation in histology images. In: Proceedings of 2014 IEEE International Conference on Image Processing (ICIP), Paris, pp. 3953–3957 (2014)
13. Mendivil, F.: Computing the Monge–Kantorovich distance. Comput. Appl. Math. **36**(3), 1389–1402 (2016). https://doi.org/10.1007/s40314-015-0303-7
14. Molter, U., Brandt, J., Cabrelli, C.: An algorithm for the computation of the Hutchinson distance. Inf. Process. Lett. **40**(2), 113–117 (1991)
15. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. **42**(3), 145–175 (2001). https://doi.org/10.1023/A:1011139631724
16. Ruifrok, A.C., Johnston, D.A.: Quantification of histochemical staining by color deconvolution. Anal. Quant. Cytol. Histol. **23**(4), 291–299 (2001)
17. Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L.: A dataset for breast cancer histopathological image classification. IEEE Trans. Biomed. Eng. **63**(7), 1455–1462 (2016)
18. Tizhoosh, H.R., Babaie, M.: Representing medical images with encoded local projections. IEEE Trans. Biomed. Eng. **65**(10), 2267–2277 (2018)