

## DISCRIMINATIVELY WEIGHTED MULTI-SCALE LOCAL BINARY PATTERNS: APPLICATIONS IN PROSTATE CANCER DIAGNOSIS ON T2W MRI

Haibo Wang, Satish Viswanath and Anant Madabhushi\*

Department of Biomedical Engineering  
Case Western Reserve University  
{haibo.wang<sup>3</sup>, satish.viswanath, anant.madabhushi}@case.edu

### ABSTRACT

In this paper, we present discriminatively weighted Local Binary Patterns (DWLBP), a new similarity metric to match Multi-scale LBP (MsLBP) in Hamming space. While MsLBP is widely used in image processing on account of its extremely fast bitwise operations on modern CPU, identifying a good metric that measures the dissimilarity of MsLBP remains an open problem. The Hamming score is typically computed at each individual scale and the scores across scales are summed up. This approach however often results in underestimating salient patterns. In this paper we seek to learn a vector obtained by optimally weighing the contribution of each individual scale when performing MsLBP based matching. Inspired by supervised learning, our methodology aims to learn the multi-scale, weight vector by minimizing the Hamming scores between positive class samples and jointly maximizing the scores between positive and negative class samples. This objective function leads to a convex formulation with equality and inequality constraints, which can then be solved via the *interior-point* optimization method. In this paper we evaluate the efficacy of the DWLBP scheme in detecting prostate cancer from T2w MRI and demonstrate that the approach statistically significantly outperforms MsLBP.

**Index Terms**— Prostate Cancer, MRI, Image Processing, Local Binary Patterns, multi-scale

### 1. INTRODUCTION

Pixelwise template matching is often utilized to perform an exhaustive search of an entire image to find pixels similar to a query pixel. A major challenge to overcome when matching a pair of pixels lies in the conflict between matching speed and accuracy. To address the challenge, pixel representation should be distinctive and compact, and computing its dissimilarity metric should be achievable quickly.

Local Binary Pattern (LBP) [1] has been shown to be powerful to match a local pixel feature to another. The LBP de-

scriptor of a pixel is a string of binary bits, each of which is obtained by comparing the gray value of the pixel with a number of others sampled on a ring centered on the pixel. The Hamming distance of LBP refers to the number of bits that are different. This only requires carrying out bitwise XOR operations, which can be computed extremely quickly on modern CPUs. Finding a distinctive ring radius is critical to extract salient LBP descriptions. However, the existing approach that detects local Laplacian extrema [2, 3] is computationally costly, hence, largely negating the benefit of LBP. Sampling multiple radii can statistically guarantee measuring textual content at the salient scale. By assuming independent sampling, measuring this multi-scale LBP is defined as the sum of the Hamming scores across individual scales (MsLBP) [1]. Nevertheless, salient patterns are under-emphasized while insignificant patterns are over-emphasized during the process. It is necessary to define a weight vector so as to account for the statistical significance of information at the salient scale by measuring the dissimilarity between a pair of multi-scale LBPs. We anticipate that this strategy will significantly improve the matching quality and hence detection accuracy.

Recent work shows that *learned binary projections* are a powerful way to index large image collections based on content [4]. *Unsupervised hashing* [5, 6] leads to binary codes that can be as poor as random binarization. With supervised learning imposed, *supervised hashing* [4] explicitly learns a mapping that maximizes the distances among different classes. However, due to the non-differentiable *sign* function, one has to relax the objective function, resulting in a suboptimal solution.

In this paper, we present a new method, discriminatively weighted Local Binary Patterns (DWLBP), to tackle the problem of combining the multi-scale Hamming scores for matching MsLBP. Inspired by *supervised hashing*, we seek to learn a weight vector by minimizing the squares of Hamming distances between positive class samples and jointly maximizing the Hamming distances between positive and negative class samples. Since each element of the vector must be normalized and their sum equals one, we get a final objective function that is convex and constrained by linear equality and in-

\*This work was made possible by grants from the National Institute of Health (R01CA136535, R01CA140772, R43EB015199, R21CA167811), National Science Foundation (IIP-1248316), and the QED award from the University City Science Center and Rutgers University.

equality constraints. Literature provides many solutions to the problem and we use the interior-point method. The main advantages of DWLBP are two-fold:

1. Compared to MsLBP [1] matching, we impose supervised learning so as to optimally weigh the significance of each scale in combining Hamming scores. This yields a principled approach to obtaining an optimal weight vector, one that is more general than the *ad hoc* approach employed in [7].
2. The concern of non-differentiability associated with *supervised hashing* is addressed by shifting binarization to an earlier step of LBP extraction, which results in a smooth, differentiable and globally convex problem.

In this paper we evaluate the efficacy of the DWLBP scheme in detecting prostate cancer from T2w MRI. The cancer groundtruth was first annotated on ex vivo histological sections obtained from patients undergoing radical prostatectomy who receive a pre-operative MRI. Deformable co-registration was then applied to non-linearly align the pre-operative MRI with the post-operative histologic sections to map spatial extent of cancer onto corresponding MRI sections [8]. DWLBP is then run on the T2w MRI images to generate a probabilistic heatmap on which higher probability indicates the presence of suspected cancer.

## 2. METHODOLOGY

### 2.1. Framework

As Figure 1 shows, the proposed methodology consists of two stages. The first stage is a learning procedure. Given images with labelled positive and negative samples, we extract multi-scale LBPs for each pixel and compute their Hamming scores at each scale. Then we learn a weight that best combines the multi-scale scores in the sense that positives and negatives are separated as far as possible.

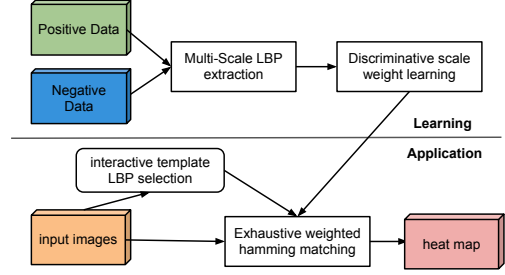
Once the learning is done, salient feature detection can proceed in such a way: a small number of distinct image pixels are selected by the user and the LBPs code pertaining to the pixels are extracted. A exhaustive search of the entire image area is then performed to match the LBPs of each pixel with the template LBPs at each scale. Finally, weighted sum of the Hamming scores of all scales yields a statistical probability heatmap, where coordinates having higher probabilities indicate candidate salient features.

### 2.2. Local Binary Pattern

LBP operates on the intensity values of image pixels and models a single pixel via its local neighbours. For an arbitrary pixel with intensity value  $f_c$ , its Local Binary Pattern (LBP) number consists of the signs of the gray-level value discrepancies between  $f_c$  and the gray values of  $p$  equally spaced pixels on a circle of radius:

$$\mathbf{x} = \sum_{i=0}^{p-1} \text{sign}(f_i - f_c)2^i,$$

where  $f_i$ ,  $i \in \{0, 1, \dots, p-1\}$ , is the gray value of  $i^{\text{th}}$  sampled pixel. The coordinates of  $f_i$  are given by  $(-r \sin(2\pi \frac{i}{p}),$



**Fig. 1:** Flowchart of DWLBP. With representative positive and negative class samples, we extract MsLBP codes to discriminatively train an optimal vector that weighs the Hamming score at each scale. Then with an interactively selected template, exhaustive search over the entire input image is performed to generate a heatmap based on the similarity between the DWLBP scores between the query and template pixels.

$r \cos(2\pi \frac{i}{p}))$ , all of which together form a circularly symmetric neighbor set.  $r$  is the radius of circle and  $\text{sign}()$  is:

$$\text{sign}(y) = \begin{cases} 1 & \text{if } y \geq 0; \\ 0 & \text{if } y < 0. \end{cases}$$

The binomial factor  $2^i$  transforms LBP to a number whose bits are measured in Hamming distance. By its definition LBP is invariant to local gray-scale shift. In order to be rotation-invariant as well, a circular bitwise right shift on the  $p$  bits is performed  $p$  times and the minimum resulting number is retained as the final LBP output.

### 2.3. Multi-Scale LBP

By combining multiple operators by varying  $(p, r)$ , a LBP can capture the property of multi-resolution texture, leading to multi-scale LBP. The metric for measuring similarity between a pair of multiscale LBPs is via a simple kernel,

$$\mathcal{H}(\mathbf{x}, \mathbf{x}') = \sum_{n=1}^N d_H(\mathbf{x}_n, \mathbf{x}'_n), \quad (1)$$

where  $N$  is the number of operations of varying  $(p, r)$  and  $d_H(\mathbf{x}, \mathbf{x}')$  is the Hamming distance:  $d_H = \sum_{i=0}^{p_n-1} (x_i \neq x'_i)$ , where  $x_i$  and  $x'_i$  are the  $i^{\text{th}}$  bit of  $\mathbf{x}$  and  $\mathbf{x}'$ , respectively.

### 2.4. Discriminatively Weighted LBP

The above dissimilarity measurement assigns a uniform weight to each scale. Here we present a more general scale selection scheme that learns a weight  $w \in [0, 1]$  for each scale. Supposing LBP involves  $N$  operations by altering  $(p, r)$ , scale selection seeks a vector  $\mathbf{w} \in R^{N \times 1}$  such that the dissimilarity metric turns to a weighted sum

$$\mathcal{H}(\mathbf{x}, \mathbf{x}') = \sum_{n=1}^N w_n d_H(\mathbf{x}_n, \mathbf{x}'_n) = \mathbf{w}^T \mathbf{b}_H, \quad (2)$$

where  $\mathbf{b}_H$  is a column vector of  $d_H(\mathbf{x}_n, \mathbf{x}'_n)$ .

Our goal is to learn an optimal  $\mathbf{w}$  from a set  $\mathcal{P}$  of *positive* representative samples (extracted from cancer pixels in our case) and a set  $\mathcal{N}$  of *negative* LBP descriptors (non-cancer pixels in our case), by simultaneously minimizing the weighted hamming distances between all samples of  $\mathcal{P}$  and maximizing that distances of samples between  $\mathcal{P}$  and  $\mathcal{N}$ . Hence we get the following objective function to minimize:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \sum_{\mathbf{x} \in \mathcal{P}, \mathbf{x}' \in \mathcal{P}} \mathcal{H}^2(\mathbf{x}, \mathbf{x}') - \alpha \sum_{\mathbf{x} \in \mathcal{P}, \mathbf{x}' \in \mathcal{N}} \mathcal{H}^2(\mathbf{x}, \mathbf{x}'), \\ \text{s.t.} \quad & 1 \geq w_n \geq 0, n = 1, \dots, N, \text{ and } \sum_{n=1}^N w_n = 1, \end{aligned} \quad (3)$$

where  $\alpha$  is to balance the in- and inter-class distances,  $1 \geq w_n \geq 0$  constrains that  $w_n$  is a weight and satisfies  $\sum_{n=1}^N w_n = 1$ . Combining Eq. 2 and Eq. 3 yields

$$\begin{aligned} & \sum_{\mathbf{x} \in \mathcal{P}, \mathbf{x}' \in \mathcal{P}} \mathbf{w}^T \mathbf{b}_H \mathbf{b}_H^T \mathbf{w} - \alpha \sum_{\mathbf{x} \in \mathcal{P}, \mathbf{x}' \in \mathcal{N}} \mathbf{w}^T \mathbf{b}_H \mathbf{b}_H^T \mathbf{w} \\ & = \mathbf{w}^T \Sigma_{\mathcal{P}} \mathbf{w} - \alpha \mathbf{w}^T \Sigma_{\mathcal{P}\mathcal{N}} \mathbf{w}, \end{aligned} \quad (4)$$

where  $\Sigma_{\mathcal{P}} = \sum_{\mathbf{x} \in \mathcal{P}, \mathbf{x}' \in \mathcal{P}} \mathbf{b}_H \mathbf{b}_H^T$  and  $\Sigma_{\mathcal{P}\mathcal{N}} = \sum_{\mathbf{x} \in \mathcal{P}, \mathbf{x}' \in \mathcal{N}} \mathbf{b}_H \mathbf{b}_H^T$  are the intra-class and inter-class Hamming distance kernel, respectively. Eq. 3 is now arranged in a matrix form:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^T (\Sigma_{\mathcal{P}} - \alpha \Sigma_{\mathcal{P}\mathcal{N}}) \mathbf{w} \\ \text{s.t.} \quad & \mathbf{b}^T \mathbf{w} = 1, \text{ and } \mathbf{0} \leq \mathbf{w} \leq \mathbf{1}. \end{aligned} \quad (5)$$

where  $\mathbf{b} = [1, \dots, 1]^T$ ,  $\mathbf{b} \in \mathcal{R}^{N \times 1}$ .

## 2.5. Optimization

Eq. 5 is a quadratic function with both linear equality and inequality constraints. By choosing  $\alpha$  appropriately, we can keep the symmetric matrix  $\Sigma_{\mathcal{P}} - \alpha \Sigma_{\mathcal{P}\mathcal{N}}$  positive definite. Since the two constraints are both convex set, we have a convex problem in hand. In practice, we solved the convex optimization problem using the *interior-point* optimization method implemented in the MATLAB *quadprog* function.

Different to learning-based hashing technique [4], our approach works on direct binary strings and learns a weight vector rather than a projection matrix. The benefits of our technique include (1) reduced overhead in computing  $\Sigma_{\mathcal{P}}$  and  $\Sigma_{\mathcal{P}\mathcal{N}}$  via bitwise comparisons, and (2) our solution is optimal in the sense that the non-differentiable binary function is not a concern for our problem.

With appropriate weights, we can apply the weighted Hamming metric in Equation 2 to index the entire image surface with a query pixel picked by the user. Details of the entire procedure are presented in Algorithm 1.

## 3. EXPERIMENTAL RESULTS

### 3.1. Dataset

The experimental dataset is a T2w MRI prostate dataset consisting of a total of 80 slices at a size of  $512 \times 512$  pix-

**Data:**  $\mathcal{P}, \mathcal{N}$ , Input image  $\Omega(f)$ , query LBP vector  $\mathbf{x}_0$

**Result:**  $\mathbf{w}$ , heatmap  $\Theta(f)$

**training phase:**

1. compute  $\Sigma_{\mathcal{P}}$  with  $\mathcal{P}$  set;
2. compute  $\Sigma_{\mathcal{P}\mathcal{N}}$  with  $\mathcal{N}$  and  $\mathcal{P}$  sets;
3. compute  $\mathbf{w}$  according to Eq. 5;

**testing phase:**

**while** scan  $\Omega(f)$  at each  $(i, j)$  position **do**

1. extract multi-scale LBP  $\mathbf{x}_{ij}$  at  $(i, j)$ ;
2. find  $d_H(\mathbf{x}_{ij}^n, \mathbf{x}_0^n)$  at each of  $n$  scales;
3. compute the weighted sum  $\mathbf{w}^T \mathbf{b}_H$ ;
4. assign the weighted sum to  $\Theta(f)$  at  $(i, j)$ ;

**end**

**Algorithm 1:** Summary of DWLBP

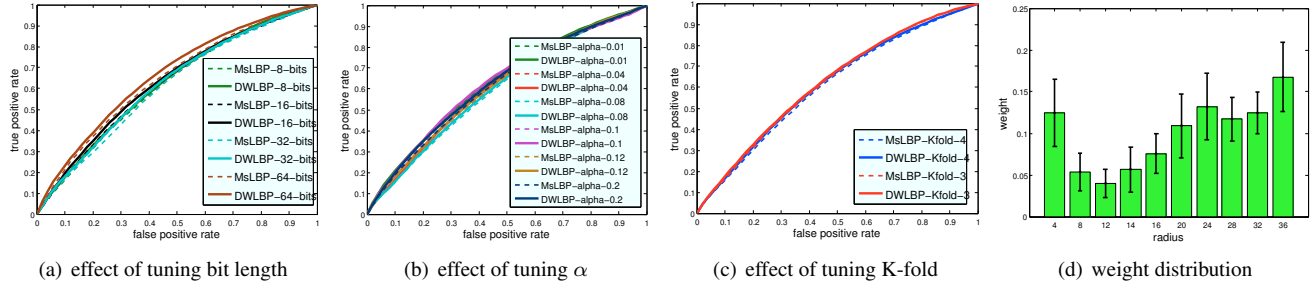
els. They were scanned from 22 patients who had previously been diagnosed with prostate cancer via core needle biopsies. Since all of these patients underwent radical prostatectomy, ex vivo histological sections were available. Deformable co-registration was applied to non-linearly align the pre-operative MRI with the post-operative histologic sections to map spatial extent of cancer onto corresponding MRI sections [8]. The goal of this experiment was to distinguish between cancer and benign regions on a per pixel basis.

	8	16	32	64
MsLBP	0.617	0.621	0.608	0.646
DWLBP	<b>0.623</b>	<b>0.633</b>	<b>0.622</b>	<b>0.664</b>

**Table 1:** The area value under ROC curve (AUC) of DWLBP and MsLBP for different bit lengths. These values are obtained by using  $\alpha = 0.04$  and 3-fold cross-validation over 80 MRI images and 22 patient studies. The average AUC of DWLBP is larger than that of MsLBP by an average of 1.25%.

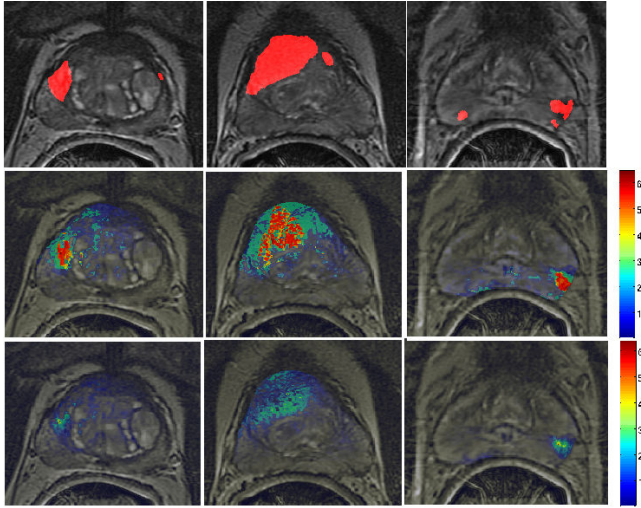
### 3.2. Experimental Design

A randomized K-fold cross validation was performed by dividing the 80 image slices into K parts on a per patient basis, where  $(K-1)/K$  of the dataset was used for training and  $1/K$  for testing. This is repeated until all the samples have been classified within each dataset. The resulting ROC curves are then averaged. For balanced classification, on each slice we randomly sampled 40 cancer pixels as positive data and 40 non-cancer pixels as negative data. The template pixel was randomly picked on the labelled cancer region of each slice for both the purposes of weight learning and evaluation. The radii of MsLBP are fixed as  $r \in \{4, 8, 12, 14, 16, 20, 24, 28, 32, 36\}$ . As for the number of sample  $p$  (length of bits), we tested  $p \in \{8, 16, 32, 64\}$ . The K-fold is set as 3 or 4, and we tested several values for  $\alpha$ : 0.01, 0.04, 0.08, 0.1, 0.12 and 0.2.



(a) effect of tuning bit length (b) effect of tuning  $\alpha$  (c) effect of tuning K-fold (d) weight distribution

**Fig. 2:** ROC curves and learned weight distribution. (a) Evaluating the effect of varying bit number among 8, 16, 32, 64 by fixing  $\alpha = 0.04$  and  $K = 3$ . 64-bits shows better performance than others. (b) Evaluating the effect of varying  $\alpha$  among 0.01, 0.04, 0.08, 0.1, 0.12 and 0.2 by using  $K = 4$  and 8 bits. We see that performance is robust to  $\alpha$ . (c) Evaluating the effect of varying  $K$  between 3 and 4, when using  $\alpha = 0.04$  and 8 bits. (d) Showing the distribution (mean and standard variation) of the learned weight vector. We see that LBP patterns at  $r = 8, 12, 14, 16$  are less discriminating than others.



**Fig. 3:** Top row: cancer ground-truth (red) obtained via deformable co-registration of cancer maps from ex vivo histology onto pre-operative MRI for 3 different patient studies. Second and third rows: heatmaps showing likelihood of cancer occurrence via DWLBP and MsLBP, respectively. The color bar shows the suspected probability of being cancer.

### 3.3. Results

We compared DWLBP mainly with the multi-scale LBP matching approach (termed as MsLBP) that treats each scale equally. The resulting ROCs are depicted in Fig. 2 (a)-(c) and the corresponding AUC values are shown in Table 1. We see that the accuracy of DWLBP is always higher than that of MsLBP. This is also verified in Fig. 3. The heatmaps of DWLBP demonstrate statistically better cancer prediction than MsLBP.

Fig. 2 (a)-(c) also illustrates the affect of tuning parameters, including the length of bit at each scale, the weight threshold  $\alpha$  and the number of K-fold for cross-validations. When using 64 bits, DWLBP has a visibly better ROCs than MsLBP. Tuning  $\alpha$  and K-fold did not appear to have visible effect on the results.

Fig. 2 (d) demonstrates the distribution (mean and vari-

ance) of the learned weights obtained with over one thousand iterations. It shows that the LBP patterns at  $r \in \{8, 12, 14, 16\}$  are less discriminating, which reveals that small and large scales tend to be more discriminating compared to the medium scales.

## 4. CONCLUDING REMARKS

This paper presented DWLBP that attempts to tackle the problem of how to combine the Hamming scores of LBPs at different scales. By learning a discriminative vector that weighs the matching score at each individual scale, it leads to a metric that better represents cancer region. Applied in prostate cancer diagnosis, the new DWLBP approach improves the accuracy of cancer prediction as compared to MsLBP. Experiments reveal that LBP patterns at small and large scales are more important than those at medium scales.

## 5. REFERENCES

- [1] Timo Ojala and et al, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE TPAMI*, vol. 24, pp. 971–987, 2002.
- [2] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, pp. 91–110, 2004.
- [3] Herbert Bay and et al, "Speeded-up robust features (surf)," *CVIU*, vol. 110, pp. 346–359, 2008.
- [4] Christoph Strecha and et al, "Ldhash: Improved matching with smaller descriptors," *IEEE TPAMI*, vol. 34, pp. 66–78, 2012.
- [5] Loïc Paulevé and et al, "Locality sensitive hashing: A comparison of hash function types and querying mechanisms," *PRL*, vol. 31, pp. 1348–1358, 2010.
- [6] Yair Weiss and et al, "Spectral hashing," in *NIPS*, 2008, pp. 1753–1760.
- [7] Jason Hipp and et al, "Optimization of complex cancer morphology detection using the sivq pattern recognition algorithm," *ACP*, vol. 35, pp. 41–50, 2012.
- [8] John Chappelow and et al., "Elastic registration of multimodal prostate mri and histology via multiattribute combined mutual information," *Medical Physics*, vol. 38, pp. 2005–2018, 2011.