Optimizing LBP Structure For Visual Recognition Using Binary Quadratic Programming

Jianfeng Ren, *Student Member, IEEE*, Xudong Jiang, Junsong Yuan, *Senior Member, IEEE*, Gang Wang, *Member, IEEE*

***Abstract*—Local binary pattern (LBP) and its variants have shown promising results in visual recognition applications. However, most existing approaches rely on a pre-defined structure to extract LBP features. We argue that the optimal LBP structure should be task-dependent and propose a new method to learn discriminative LBP structures. We formulate it as a point selection problem: Given a set of point candidates, the goal is to select an optimal subset to compose the LBP structure. In view of the problems of current feature selection algorithms, we propose a novel Maximal Joint Mutual Information criterion. Then, the point selection is converted into a binary quadratic programming problem and solved efficiently via the branch and bound algorithm. The proposed LBP structures demonstrate superior performance to the state-of-the-art approaches on classifying both spatial patterns in scene recognition and spatial-temporal patterns in dynamic texture recognition.**

***Index Terms*—LBP structure optimization, maximal joint mutual information, binary quadratic programming, scene recognition, dynamic texture recognition**

I. Introduction

LOCAL binary pattern and its variants have wide applications, e.g. texture classification [1]–[3], dynamic texture (DT) recognition [4]–[6], scene recognition [7]–[9], facial analysis [10]–[16] and human detection [17]–[19]. LBP is popular because of its simplicity, ability to capture image micro-structures and robustness to illumination variations.

However, it remains challenging to derive the best LBP structure for a specific application. In the traditional pipeline, a handcrafted LBP structure was often utilized [8]–[10], [20]–[25]. The popular LBP structure consists of 8 nearest neighbors or P neighbors in a circle [8]–[10]. Other geometries such as line and disc were explored in Local Quantized Pattern (LQP) [25]. The handcrafted structure may not be optimal as it is often selected heuristically. More importantly, the LBP structure should be task-dependent because the intrinsic image characteristics of different applications or even image patches may be different. In [13], a heuristic hill-climbing technique was utilized to select the LBP structure. Lei et al. proposed to learn discriminant image filters and optimal neighborhood sampling strategy in a data-driven way [16]. Some other approaches aim to directly extract discriminative information from LBP-histogram bins, e.g. Adaboost bin selection [22], [23] and dictionary learning [24], [25].

In this paper, we propose a new method to learn the data-driven LBP (DDLBP) structure. Our motivation originates from volume-LBP [5]. The number of its histogram bins increases exponentially, e.g. $VLBP_{L,P,R}$ has 2^{3P+2} bins. Even for a small $P = 8$, it has $2^{26} = 67,108,864$ bins. It is intractable to reduce the dimensionality of such high-dimensional data. To solve this problem, we propose to find a small structure that generates a feature vector of a manageable size. We formulate it as a point selection problem, i.e. to select an optimal subset to compose the LBP structure that is most suitable for a specific application.

For feature selection, it is desirable to maximize the dependency of the target class on the data distribution, known as Max-Dependency scheme. It is difficult to directly estimate such a dependency. Thus, approximated algorithms such as Max Relevance and mRMR were often utilized [26]. We find that these may not closely approximate Max-Dependency criterion for LBP structure optimization. Thus, we propose to approximate it using joint mutual information between a feature pair and the classification variable. Then, a Maximal Joint Mutual Information (MJMI) scheme is proposed to optimize the LBP structure.

Given a feature selection criterion, greedy algorithms were often used [7], [26], which may only find a locally optimal solution. In this paper, we learn a globally optimal LBP structure by casting the point selection into a binary quadratic programming (BQP) problem [27] and solving it via the branch and bound algorithm [28].

Our contributions are three-fold: a) We propose a new formulation of LBP structure optimization by casting it as a point selection problem. b) We find that Max-Dependency criterion is better approximated using joint mutual information. Thus, a MJMI scheme is proposed for LBP structure optimization. c) Instead of greedy search, we cast the proposed MJMI as a BQP problem and derive a globally optimal structure. The proposed approach demonstrates superior performance on scene recognition and DT recognition.

II. The Proposed Data-Driven LBP

A. Overview

The block diagram is shown in Fig. 1. It consists of two steps: DDLBP structure optimization and DDLBP feature generation. We cast the DDLBP structure optimization as a point

This research is supported in part by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

J. Ren is with BeingThere Centre, Institute of Media Innovation, Nanyang Technological University, Singapore 637553. (e-mail: jfren@ntu.edu.sg).

X. Jiang, J. Yuan and G. Wang are with School of Electrical & Electronics Engineering, Nanyang Technological University, Singapore 639798. (e-mail: {exdjiang, jsyuan, wanggang}@ntu.edu.sg)

selection problem. Formally, the problem is defined as: given a set of potential candidates $\mathbf{x} = \{x_i, i = 1, 2, \ldots, n\}$ and target classification variable c, the goal is to find a subspace R^m of m candidates $\mathbf{x}_m \subseteq \mathbf{x}$ that optimally characterizes c.

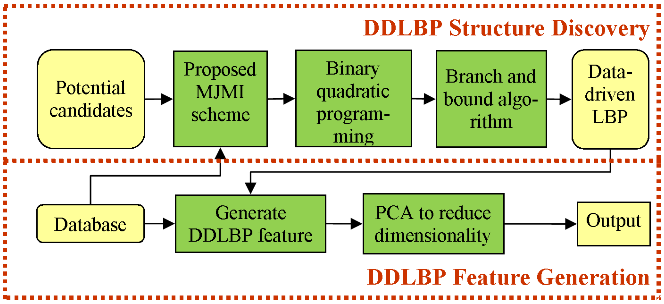

Fig. 1. Block diagram of the proposed method to extract LBP features.

In many scenarios, especially for spatial-temporal LBP (STLBP), the number of neighbors may be large, and hence it is not feasible to enumerate the histogram using such a large structure. Thus, we treat them as potential candidates and aim to find an optimal subset. The potential candidates can be natural extension of widely used handcrafted structures. Compared with deriving a good handcrafted structure, it is much easier to obtain a good set of potential candidates.

B. Feature Selection via Mutual Information

For feature selection, it is desirable to maximize the dependency of selected features on classification variable c (Max-Dependency) [26]. We use mutual information to characterize the dependency. The goal is to find $\mathbf{x}_m \subseteq \mathbf{x}$ so that:

$$\mathbf{x}_m^* = \underset{\mathbf{x}_m}{\operatorname{argmax}}\, I(\mathbf{x}_m; c), \tag{1}$$

$$I(\mathbf{x}_m; c) = \int\int p(\mathbf{x}_m, c) \log \frac{p(\mathbf{x}_m, c)}{p(\mathbf{x}_m)p(c)} d\mathbf{x}_m dc. \tag{2}$$

In general, it is difficult to reliably estimate $p(\mathbf{x}_m)$ and $p(\mathbf{x}_m, c)$ due to the limited number of samples available and the large number of joint states to be estimated. Alternatively, Max-Relevance is utilized, which approximates $I(\mathbf{x}_m; c)$ as:

$$\mathbf{x}_m^* = \underset{\mathbf{x}_m}{\operatorname{argmax}}\, \frac{1}{m}\sum_{i=1}^{m} I(x_i; c). \tag{3}$$

The features selected according to Max-Relevance may have rich redundancy, and hence Min-Redundancy criterion was added to select mutually exclusive features: $\mathbf{x}_m^* = \operatorname{argmin}_{\mathbf{x}_m} \frac{1}{m^2}\sum_{x_i,x_j \in \mathbf{x}_m} I(x_i; x_j)$. In [26], Min-Redundancy and Max-Relevance (mRMR) were combined:

$$\mathbf{x}_m^* = \underset{\mathbf{x}_m}{\operatorname{argmax}} \sum_{i=1}^{m} I(x_i; c) - \frac{1}{m}\sum_{x_i,x_j \in \mathbf{x}_m} I(x_i; x_j). \tag{4}$$

Recent research [29] shows that when high-order interaction information is negligible, $I(\mathbf{x}_m; c)$ can be approximated by:

$$\begin{aligned} I(\mathbf{x}_m; c) \approx &\sum_{x_i \in \mathbf{x}_m} I(x_i; c) - \sum_{x_i,x_j \in \mathbf{x}_m} I(x_i; x_j) \\ &+ \sum_{x_i,x_j \in \mathbf{x}_m} I(x_i; x_j | c), \end{aligned} \tag{5}$$

where $I(x_i; x_j|c)$ is conditional mutual information. [1] Only when $I(\mathbf{x}_m; c)$ in Eqn. (5) is dominated by the first term, Max-Relevance defined in Eqn. (3) is a good approximation of $I(\mathbf{x}_m; c)$. mRMR defined in Eqn. (4) differs from Eqn. (5) by a missing term $\sum_{x_i,x_j \in \mathbf{x}_m} I(x_i; x_j|c)$ and a weighting factor for the second term. In general, Max-Relevance and mRMR are not a close approximation of $I(\mathbf{x}_m; c)$.

C. The Proposed Maximal Joint Mutual Information Scheme

Our target is to derive a close approximation of $I(\mathbf{x}_m; c)$. Recall the chain rule for $I(\mathbf{x}_m; c)$,

$$I(\mathbf{x}_m; c) = \sum_{i=1}^{m} I(x_i; c | x_1, \ldots, x_{i-1}). \tag{6}$$

For $i \geq 3$, $I(x_i; c|x_1, \ldots, x_{i-1})$ is high-order conditional mutual information. $I(\mathbf{x}_m; c)$ is decomposed into m terms in Eqn. (6). In fact, including Eqn. (6) there are $m! = m(m-1)\cdots \times 2 \times 1$ different ways to do the decomposition. Averaging over these $m!$ decompositions, we have:

$$\begin{aligned} I(\mathbf{x}_m; c) = &\frac{1}{m}\sum_{i_1=1}^{m} I(x_{i_1}; c) + \frac{1}{m(m-1)}\sum_{i_1 \neq i_2} I(x_{i_1}; c | x_{i_2}) \\ &+ \cdots + \frac{1}{m!}\sum I(x_{i_1}; c | x_{i_2} x_{i_3} \ldots x_{i_m}), \end{aligned} \tag{7}$$

where $\{x_{i_k}\}, k = 1, 2, \ldots, m$ is an ordered set of $\mathbf{x}_m$ for the last term. We notice that all these terms are positive. When high-order conditional mutual information is negligible, $I(\mathbf{x}_m; c)$ can be approximated by:

$$\begin{aligned} I(\mathbf{x}_m; c) &\approx \frac{1}{m}\sum_{i=1}^{m} I(x_i; c) + \frac{1}{m(m-1)}\sum_{i \neq j} I(x_i; c | x_j) \\ &= \frac{1}{m(m-1)}\sum_{i \neq j} I(x_i, x_j; c), \end{aligned} \tag{8}$$

where $I(x_i, x_j; c) = I(x_i; c) + I(x_j; c|x_i)$ is joint mutual information between feature pair x_i, x_j and c. Then, we propose a Maximal Joint Mutual Information scheme for LBP structure optimization. Instead of maximizing intractable $I(\mathbf{x}_m; c)$, the goal is to find a subset $\mathbf{x}_m \subseteq \mathbf{x}$ that maximizes its approximation $\sum_{i \neq j} I(x_i, x_j; c)$, i.e.

$$\mathbf{x}_m^* = \underset{\mathbf{x}_m}{\operatorname{argmax}} \sum_{x_i,x_j \in \mathbf{x}_m, i \neq j} I(x_i, x_j; c). \tag{9}$$

D. Deriving a Globally Optimal DDLBP Structure

To derive a globally optimal solution to Eqn. (9), we convert the proposed MJMI scheme into a binary quadratic programming problem. Denote $\mathbf{a} = (a_1, a_2, \ldots, a_n)^T, a_i \in \{0, 1\}$ as the indication vector for $\mathbf{x}$, i.e. $a_i = 1$ means x_i is selected and $a_i = 0$ otherwise. Then, Eqn. (9) is equivalent to:

$$\mathbf{a}^* = \arg\max_{\mathbf{a}} \mathbf{a}^T \mathbf{M}\mathbf{a}, s.t. \sum_{i=1}^{n} a_i = m. \tag{10}$$

[1] For discrete random variables x, y, z, conditional mutual information $I(x; y|z) = \mathbb{E}_z\{I(x; y)|z\} = \sum_{x,y,z} p(x, y, z) \log \frac{p(z)p(x,y,z)}{p(x,z)p(y,z)}$, where $\mathbb{E}_z\{.\}$ is the expectation on z.

$\mathbf{M}$ is a matrix of size $n \times n$, whose diagonal elements are zero and off-diagonal elements $\mathbf{M}(i,j) = I(x_i, x_j; c)$. This optimization problem can be solved efficiently via the branch and bound algorithm [28].

We branch the feasible region $\mathcal{S}$ into k smaller subregions such that $\mathcal{S} = \bigcup_{i=1}^{k} \mathcal{S}_i$. These subregions naturally form a tree structure. More specifically, it is a branch-and-bound tree of n levels, and each level corresponds to one binary variable a_i. We bound the objective function of the subproblem in the node using quadratic relaxation created by relaxing the integer constraints to interval constraints, i.e. $a_i \in [0,1]$. If a partial solution from a subregion is less than the lower bound, it is discarded from the search. There are three possible causes of pruning a subtree: 1) Infeasibility, i.e. the subproblem has no feasible solution. 2) Optimality, i.e. an optimal solution to the sub-problem is found. 3) Dominance, i.e. the solution to the subproblem is no better than the current one. To reach a feasible solution fast, *Depth-first search* is employed. We utilize the Gurobi optimizer [30] to solve this BQP problem.

The joint probability mass function $p(x_i, x_j, c)$ can be estimated efficiently. Denote $h_{p,q}$ as the joint histogram for features x_i, x_j using q-th sample of p-th class. $p(x_i, x_j|c)$ is estimated as:

$$p(x_i, x_j|c = p) \longleftarrow \frac{1}{N_p} \sum_q h_{p,q}, \qquad (11)$$

where N_p is the number of samples for class p. Then, $p(x_i, x_j, c = p) = p(x_i, x_j|c = p)p(c = p)$, where $p(c = p) = N_p/N$ and N is the total number of samples. As we only need to estimate the joint pmf of three variables only, in which x_i, x_j are binary, the computational cost is low.

Image patches at different scales or locations may exhibit totally different characteristics. Instead of using a unified structure for all patches, we utilize the proposed MJMI scheme to learn the DDLBP structures on a patch-wise basis to better capture the characteristics of different patches. Then, PCA is applied on the LBP histogram of each patch to reduce the dimensionality. The features of all patches are concatenated to form the final feature vector, which is classified by a support vector machine with a RBF kernel [31].

III. Experimental Results

The proposed approach can be used in many applications. We show two examples: learning a set of patch-wise LBP structures for scene recognition and a STLBP structure for DT recognition. We use binarized pixel differences between 24 neighbors and the central pixel as potential candidates for spatial LBP as shown in Fig. 2(a), and those between 26 neighbors and the central pixel of frame t as potential candidates for STLBP as shown in Fig. 2(b).

A. Scene Recognition on the 21-Land-Use Dataset

The 21-land-use dataset contains 21 classes of aerial orthoimagery, and each class has 100 images of resolution 256×256 pixels [32]. Spatial pyramid [33] is utilized, i.e. each image is hierarchically divided into 31 patches, as shown in Fig. 3. We follow the same setup as in [8], [32], [34]. For each

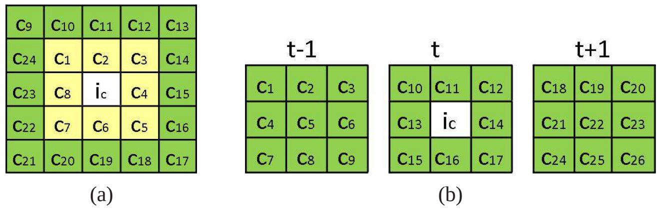

Fig. 2. Potential candidates for: (a) Spatial LBP, (b) Spatial-temporal LBP.

class, the images are randomly split into five equal-sized sets. Four of them are used for training and the held-out set is used for testing. We use CENTRIST [8] as the baseline algorithm, and construct the DDLBP structures using 8 neighbors, same as CENTRIST. Fig. 3 shows some learnt DDLBP structures. They are significantly different from each other, as the intrinsic image characteristic of each patch is different.

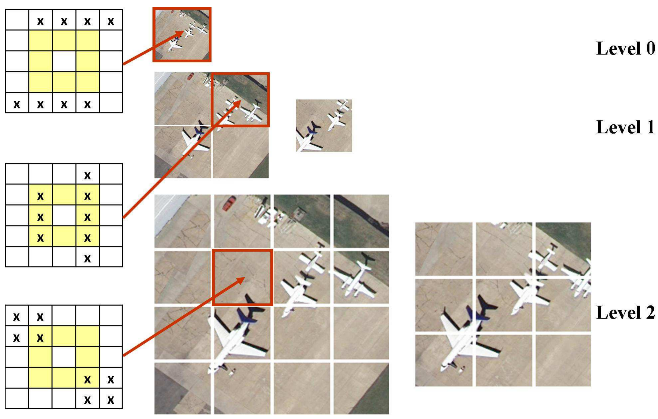

Fig. 3. DDLBP structures of the 21-land-use dataset.

We compare the proposed approach with the following: 1) Direct feature selection/extraction from the LBP-histogram bins: Adaboost bin selection [23], k-means bin clustering for LQP [25] and PCA dimensionality reduction for CENTRIST [8]. 2) Other LBP-structure-learning approaches: discriminant face descriptor (DFD) [16], and discriminative LBP structure learning using a heuristic hill-climbing technique [13]. 3) Other point selection algorithms under the proposed framework: the proposed DDLBP with Max-relevance and mRMR. 4) Other state-of-the-art solutions for scene recognition: SPCK, SPCK+, SPCK++ [32] and BRSP [34].

The results are summarized in Table I. BRSP achieves the best recognition rate of 77.8% in literature [34]. We improve it to 87.2%. The proposed approach also outperforms those directly extract features from the LBP-histogram bins using a handcrafted structure, in which the best one is CENTRIST [8] with a recognition rate of 85.9%. Our approach significantly outperforms DFD [16] and discriminative LBP [13]. This is partly because DFD [16] cannot well handle large image variations of scene images and discriminative LBP [13] cannot guarantee structure optimality. In summary, the proposed structure optimization consistently outperforms other approaches.

TABLE I
COMPARISONS WITH THE STATE OF THE ARTS ON THE 21-LAND-USE DATASET FOR SCENE RECOGNITION.

Method	Recognition Rate
SPCK [32]	73.1%
SPCK+ [32]	76.1%
SPCK++ [32]	77.3%
BRSP [34]	77.8%
AdaBoost bin selection [23]	82.7%
CENTRIST [8]	85.9%
LQP $Disc_5^{3*}$ [25]	83.0%
DFD [16]	62.8%
Discriminative LBP [13]	73.4%
Proposed DDLBP with Max-Relevance	86.3%
Proposed DDLBP with mRMR	87.0%
Proposed DDLBP with MJMI	**87.2%**

B. Scene Recognition on the 8-Event Dataset

The 8-event dataset [35] is composed of eight sport classes. Each class has 137 to 250 high-resolution images. To capture the image micro-structures at the same scale, we resize the image so that its minimum dimension (height or weight) is 600. The experiments are repeated 5 times. For each trial, we randomly select 70 images per class for training and 60 for testing, same as in [8], [34], [35]. Other setups are the same as for the 21-land-use dataset.

The experimental results are summarized in Table II. Compared with the published best recognition rate of 79.6% achieved by RSP + Boosting [34], the proposed approach significantly boosts it to 84.0%. Our approach also outperforms direct bin selection/extraction approaches, among which Adaboost bin selection [23] performs best but achieves a recognition rate of 80.2% only. The proposed approach also demonstrates a large performance gain over other structure-learning approaches, e.g. DFD [16] and discriminative LBP [13].

TABLE II
COMPARISONS WITH THE STATE OF THE ARTS ON THE 8-EVENT DATASET FOR SCENE RECOGNITION.

Method	Recognition Rate
Scene/Object Model + SIFT [35]	73.4%
RSP + Optimal Selection [34]	77.9%
RSP + Boosting [34]	79.6%
AdaBoost bin selection [23]	80.2%
CENTRIST [8]	78.3%
LQP $Disc_5^{3*}$ [25]	78.9%
DFD [16]	75.7%
Discriminative LBP [13]	66.5%
Proposed DDLBP with Max-Relevance	83.5%
Proposed DDLBP with mRMR	83.5%
Proposed DDLBP with MJMI	**84.0%**

C. DT Recognition on the DynTex++ Dataset

The recognition of dynamic texture involves the analysis of both spatial appearance of static texture patterns and temporal variations in appearance. The DynTex++ dataset [4] consists of 36 classes. Each class contains 100 sequences of size $50 \times 50 \times 50$. One STLBP structure is learnt and used to extract the histogram from each sequence. We use the same setup as in [4], [36]. For each trial, 50 sequences are randomly selected from each class for training, and the other 50 for testing. The experiments are repeated 5 times and the average performance is reported in Table III.

We use the binarized pixel differences between 26 neighbors and the central pixel of Frame t as potential candidates, as shown in Fig. 2(b). The DDLBP structures learnt using the proposed MJMI scheme are the same over 5 trials for $m = 4, 6, 8, 10, 12, 14$, which shows that our approach can find the underlying spatial-temporal structures for dynamic texture.

In literature, the best recognition rate reported on the DynTex++ dataset is 89.9% achieved by dynamic fractal analysis [36]. The proposed DDLBP built using 14 neighbors significantly boosts the performance by 5.9%. We also implement and test LBP-TOP [5] on this dataset, in which the large spatial-temporal LBP structure is broken into small handcrafted ones. Compared with LBP-TOP, the proposed approach improves the recognition rate by 2.6%. The hill-climbing technique [13] is utilized to select a spatial-temporal LBP structure of 14 neighbors. As the built structure is large (14 out of 26), many selected neighbors are the same as in the proposed approach. Even so, the proposed DDLBP with MJMI scheme still outperforms it by 1.4%.

TABLE III
COMPARISONS WITH THE STATE OF THE ARTS ON THE DYNTEX++ DATASET.

Method	Recognition Rate
DL-PEGASOS [4]	63.7%
Dynamic fractal analysis [36]	89.9%
LBP-TOP [5]	93.2%
Discriminative LBP [13]	94.4%
Proposed DDLBP with Max-Relevance	94.8%
Proposed DDLBP with mRMR	95.4%
Proposed DDLBP with MJMI	**95.8%**

IV. CONCLUSION

In this paper, we propose a new method of deriving the discriminative LBP structures by casting the structure optimization as a point selection problem. Existing algorithms such as Max-Relevance and mRMR may not well approximate Max-Dependency criterion. Thus, a MJMI scheme is proposed to better approximate Max-Dependency criterion. We then convert the proposed MJMI scheme into a binary quadratic programming problem and achieve a globally optimal solution via the branch and bound algorithm. The proposed approach is applied on scene recognition and DT recognition. For both tasks, it significantly outperforms the published best results. On the 21-land-use dataset, it boosts the recognition rate from 77.8% to 87.2%. On the 8-event dataset, it improves the recognition rate from 79.6% to 84.0%. On the DynTex++ dataset, it increases the recognition rate from 89.9% to 95.8%.

REFERENCES

[1] Z. Li, G. Liu, Y. Yang, and J. You, "Scale-and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2130–2140, 2012.

[2] K. Wang, C.-E. Bichot, C. Zhu, and B. Li, "Pixel to patch sampling structure and local neighboring intensity relationship patterns for texture classification," *IEEE Signal Processing Letters*, vol. 20, no. 9, pp. 853–856, 2013.

[3] T. Song, H. Li, F. Meng, Q. Wu, B. Luo, B. Zeng, and M. Gabbouj, "Noise-robust texture description using local contrast patterns via global measures," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 93–96, 2014.

[4] B. Ghanem and N. Ahuja, "Maximum margin distance learning for dynamic texture recognition," in *European Conference on Computer Vision*, 2010, pp. 223–236.

[5] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[6] G. Zhao, T. Ahonen, J. Matas, and M. Pietikainen, "Rotation-invariant image and video description with local binary pattern features," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1465–1477, 2012.

[7] J. Ren, X. Jiang, and J. Yuan, "Learning binarized pixel-difference pattern for scene recognition," in *IEEE International Conference on Image Processing (ICIP)*, 2013, pp. 2494–2498.

[8] J. Wu and J. Rehg, "CENTRIST: A visual descriptor for scene categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1489–1501, 2011.

[9] Y. Xiao, J. Wu, and J. Yuan, "mCENTRIST: A multi-channel feature generation mechanism for scene categorization," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 823–836, 2014.

[10] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[11] W. Zhang, S. Shan, X. Chen, and W. Gao, "Local gabor binary patterns based on kullback-leibler divergence for partially occluded face recognition," *IEEE Signal Processing Letters*, vol. 14, no. 11, pp. 875–878, 2007.

[12] X. Li, W. Hu, Z. Zhang, and H. Wang, "Heat kernel based local binary pattern for face representation," *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 308–311, 2010.

[13] D. Maturana, D. Mery, and A. Soto, "Learning discriminative local binary patterns for face recognition," in *IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, March 2011, pp. 470–475.

[14] X. Huang, G. Zhao, W. Zheng, and M. Pietikainen, "Spatiotemporal local monogenic binary patterns for facial expression recognition," *IEEE Signal Processing Letters*, vol. 19, no. 5, pp. 243–246, 2012.

[15] J. Ren, X. Jiang, and J. Yuan, "Noise-resistant local binary pattern with an embedded error-correction mechanism," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 4049–4060, 2013.

[16] Z. Lei, M. Pietikainen, and S. Li, "Learning discriminant face descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 289–302, Feb 2014.

[17] X. Wang, T. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *International Conference on Computer Vision*, 2009, pp. 32–39.

[18] J. Xu, Q. Wu, J. Zhang, and Z. Tang, "Fast and accurate human detection using a cascade of boosted MS-LBP features," *IEEE Signal Processing Letters*, vol. 19, no. 10, pp. 676–679, 2012.

[19] A. Satpathy, X. Jiang, and H. Eng, "LBP based edge-texture features for object recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 1953–1964, 2014.

[20] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, "Discriminative local binary patterns for human detection in personal album," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[21] C. Heng, S. Yokomitsu, Y. Matsumoto, and H. Tamura, "Shrink boost for selecting multi-LBP histogram features in object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3250–3257.

[22] C. Shan and T. Gritti, "Learning discriminative LBP-histogram bins for facial expression recognition," in *In Proc. British Machine Vision Conference*, 2008.

[23] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognition Letters*, vol. 33, no. 4, pp. 431–437, 2012.

[24] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2707–2714.

[25] S. ul Hussain and B. Triggs, "Visual recognition using local quantized patterns," in *European Conference on Computer Vision*, 2012, pp. 716–729.

[26] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[27] A. Chinchuluun, P. Pardalos, R. Enkhbat, and I. Tseveendorj, *Optimization and Optimal Control: Theory and Applications*, ser. Springer Optimization and Its Applications, 2010.

[28] R. Horst, P. M. Pardalos, and H. E. Romeijn, *Handbook of global optimization*, 2002.

[29] G. Brown, "An information theoretic perspective on multiple classifier systems," in *Multiple Classifier Systems*, 2009, pp. 344–353.

[30] I. Gurobi Optimization, "Gurobi optimizer reference manual," 2013. [Online]. Available: http://www.gurobi.com

[31] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[32] Y. Yang and S. Newsam, "Spatial pyramid co-occurrence for image classification," in *International Conference on Computer Vision*, 2011, pp. 1465–1472.

[33] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. 2169–2178.

[34] Y. Jiang, J. Yuan, and G. Yu, "Randomized spatial partition for scene recognition," in *European Conference on Computer Vision*, 2012, pp. 730–743.

[35] L. Li and L. Fei-Fei, "What, where and who? classifying events by scene and object recognition," in *International Conference on Computer Vision*, 2007, pp. 1–8.

[36] Y. Xu, Y. Quan, H. Ling, and H. Ji, "Dynamic texture classification using dynamic fractal analysis," in *International Conference on Computer Vision*, nov. 2011, pp. 1219 –1226.