Learning LBP Structure by Maximizing the Conditional Mutual Information

Jianfeng Ren*

BeingThere Centre, Institute for Media Innovation, Nanyang Technological University, 50 Nanyang Drive, Singapore 637553.

Xudong Jiang, Junsong Yuan

Electrical & Electronic Engineering, Nanyang Technological University, Nanyang Link, Singapore 639798.

Abstract

Local binary patterns of more bits extracted in a large structure have shown promising results in visual recognition applications. This results in very high-dimensional data so that it is not feasible to directly extract features from the LBP histogram, especially for a large-scale database. Instead of extracting features from the LBP histogram, we propose a new approach to learn discriminative LBP structures for a specific application. Our objective is to select an optimal subset of binarized-pixel-difference features to compose the LBP structure. As these features are strongly correlated, conventional feature-selection methods may not yield a desirable performance. Thus, we propose an incremental Maximal-Conditional-Mutual-Information scheme for LBP structure learning. The proposed approach has demonstrated a superior performance over the state-of-the-arts results on classifying both spatial patterns such as texture classification, scene recognition and face recognition, and spatial-temporal patterns such as dynamic texture recognition.

Keywords: LBP Structure Learning, Scene Recognition, Face Recognition, Dynamic Texture Recognition, Maximal Conditional Mutual Information

*Corresponding author. Tel.: +65 6790 5018

Email addresses: jfren@ntu.edu.sg (Jianfeng Ren), exdjiang@ntu.edu.sg (Xudong Jiang), jsyuan@ntu.edu.sg (Junsong Yuan)

1. Introduction

Local binary pattern (LBP) encodes the signs of the pixel differences between a pixel and its neighbors to a binary code [1]. LBP and its variants have been widely used in many applications, e.g. image texture classification [1–3], dynamic texture (DT) recognition [4–7], scene recognition [8, 9], facial analysis [10–19], and others [20–26]. Its popularity arises from its simplicity, the ability to capture image micro-structures and robustness to illumination variations.

One potential problem surfaces with the wide application of LBP features, *i.e.* the feature dimensionality increases exponentially with the number of LBP bits. The histogram of original LBP has 256 bins only [1]. This LBP only utilizes the pixel differences between a pixel and its 8 nearest neighbors, which cannot capture the image structures of a larger scale. In [27], LBP features were extracted using P neighbors uniformly sampled on a circle at the radius of R to the center pixel, denoted as $LBP_{P,R}$. By varying R, micro-structures at different scales are captured. $LBP_{P,R}$ has 2^P bins. Due to the storage and computational complexity constraints, the number of LBP bits is in general limited to 24, *i.e.* $2^{24} = 16,777,216$ bins. Instead of using circular neighbors, some other geometries were explored in Local Quantized Pattern (LQP) [3], *e.g.* horizontal line, vertical line, horizontal-vertical cross, diagonal cross and disc shape. Recent researches [3, 11, 27] show that LBP features using more bits extracted in a larger neighborhood have higher potential to capture complex patterns than using the basic ones, at the cost of high feature dimensionality. It imposes a big challenge to handle such high-dimensional LBP features, especially when we need to extract these features from a large-scale database.

In the literature, many algorithms were proposed to reduce the feature dimensionality by extracting a good set of features from the LBP histogram. In [27], uniform LBP was defined for circular/rectangular structure towards the objective of capturing fundamental image structures such as bright/dark spot and edges of various positive/negative curvatures. The feature dimensionality is significantly reduced from 2^P to $P(P-1)+3$. However, uniform patterns are not clearly defined towards this objective for other geometries such as cross or disc structure in [3]. A global dominant pattern set was learned from the LBP histogram through a 3-layered framework in [28]. Shan et al. utilized Adaboost algorithm to learn discriminative LBP-histogram bins for facial expression recognition [29]. In LQP [3], k-means was utilized to cluster the LBP features into a small number of visual words. However, as the spatial information is not fully utilized, spatially similar patterns may not be clustered into the same group, and hence LQP may not yield a desirable performance. For CENTRIST feature [8], PCA was applied on the LBP histogram to derive a compact feature representation. Yuan et al. improved CENTRIST by mining discriminative co-occurrence patterns [30]. In [31], Cao et al. utilized a random projection tree to encode LBP features, and used PCA to reduce the dimensionality. PCA can be also applied on the concatenated LBP histogram of all patches [32]. However, these approaches may not be applicable for a large LBP structure as it is not feasible to enumerate 2^P bins for a large P.

Instead of constructing a high-dimensional LBP histogram, some algorithms tackle the problem by breaking the large structure into small ones, or simply replacing it by a small one. Volume-LBP (VLBP) [33] has a typical large structure, in which a joint histogram of patterns in three successive frames is built. The dimensionality is as high as 2^{3P+2}. Thus, Zhao et al. proposed LBP-TOP [5] to extract LBP features from three orthogonal planes. The feature dimensionality is reduced to 3×2^P. However, the co-occurrence information of patterns extracted from these three planes is sacrificed. In LQP [3], LBP features were extracted using handcrafted structures such as line, cross and disc shapes. In Center-Symmetric LTP (CS-LTP) [34], the pixel differences between diagonal neighbors were utilized. LBP-TOP [5], CS-LTP [34] and LQP [3] are initial attempts to reduce the dimensionality by directly reducing the size of LBP structure. A heuristic hill-climbing technique was used to select the LBP structures in [35]. However, these heuristically selected structures may not be optimal. In [36], Lei et al. proposed to iteratively learn discriminant image filters and neighborhood sampling strategy. Their approach works well for face recognition, but cannot effectively handle large image variations in other applications.

In this paper, we propose to reduce the dimensionality of LBP features by optimizing the LBP structure directly. We formulate it as a point-selection problem. Given a neighborhood, the goal is to select an optimal subset of neighbors to compose the LBP structure. For each point, its binarized pixel difference with respective to the center pixel is treated as a feature. For feature selection, it is often desirable to maximize the dependency of target classification variable on data distribution (known as Max-Dependency scheme). However, it is difficult to directly calculate such a dependency as it requires to estimate high-order multivariate probability density functions [37]. Thus, approximated algorithms were often utilized, e.g. Max-Relevance, Min-Redundancy-Max-Relevance [9, 37], Max-Min-Conditional-Mutual-Information [38] and Maximal-Joint-Mutual-Information [7]. One important characteristic of the LBP-structure-learning problem is that these binarized-pixel-difference features are strongly correlated. However, previous approaches assume certain degree of feature independence, e.g. the high-order interaction is assumed negligible in [7, 9]. In view of this, we propose to first approximate the dependency as closely as possible by a set of low-order conditional mutual information, and then achieve Max-Dependency criterion through maximizing its approximation. In such a way, we derive a more accurate approximation of Max-Dependency criterion. Then, we propose a Maximal-Conditional-Mutual-Information (MCMI) scheme for LBP structure learning. It is difficult to seek a globally optimal solution for MCMI scheme. For simplicity, we utilize sequential forward selection (SFS) [39].

After deriving the LBP structures, the LBP histograms are generated using these structures. PCA is applied on the histogram of each patch to further reduce the dimensionality. The final concatenated feature vector is classified by a support vector machine with a RBF kernel or a nearest-neighbor classifier with Chi-squared distance. We use LIBSVM package [40] in the implementation.

BRIEF feature [41] and ORB feature [42] also utilize binarized pixel differences. They randomly sample pixel differences in a patch, binarize them and finally form a feature vector of a long bit stream. These two features are essentially different from the proposed approach and other LBP-based features since they do not utilize the histogram of the bit stream as the final feature vector.

2. Overview of Proposed Approach

In general, it is beneficial to extract LBP features using more binarized pixel differences in a larger neighborhood. As the feature dimensionality grows exponentially, those approaches that directly extract/select histogram-bin features quickly become infeasible. An optimal LBP structure needs to be determined first to produce a histogram of a reasonable size. Some initial attempts utilize handcrafted structures [3, 5, 34]. To develop a good handcrafted structure, a trial-and-error process is often involved. Even so, in many situations the handcrafted structure cannot yield a desirable result due to the followings: a) Optimality cannot be guaranteed as the handcrafted structure is often selected heuristically. b) The handcrafted structure is not scalable. Given a good handcrafted structure, it is unclear how to compress it to a compact one that achieves a comparable performance at a higher speed, or how to extend it to a larger one to achieve a better performance. c) The handcrafted structure is not universal. For different applications or even different patches of an image, the intrinsic image characteristics may be different. It is impossible to develop one structure that works well for all. Tremendous effort is needed to design a good handcrafted structure for every application and every patch.

The block diagram of the proposed approach is shown in Fig. 1. We derive the LBP structure first, and then use it to generate LBP-histogram bins. Formally, denote $z_i = C_i - I_c$ as the pixel difference between the neighboring

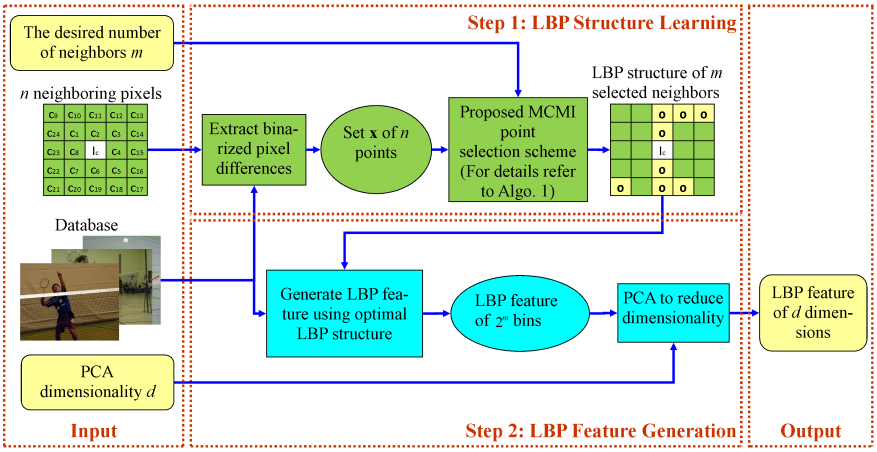

Step 1: LBP Structure Learning
The desired number of neighbors m
n neighboring pixels
C9 C10 C11 C12 C13
C24 C1 C2 C3 C14
C23 C8 Ic C4 C15
C22 C7 C6 C5 C16
C21 C20 C19 C18 C17
Extract binarized pixel differences
Set **x** of n points
Proposed MCMI point selection scheme (For details refer to Algo. 1)
LBP structure of m selected neighbors
o o o
o
Ic
o
o o o
Database
Generate LBP feature using optimal LBP structure
LBP feature of 2^m bins
PCA to reduce dimensionality
LBP feature of d dimensions
PCA dimensionality d
Input
Step 2: LBP Feature Generation
Output

Figure 1: Block diagram of the proposed approach. The neighbors of learned LBP structure are marked as "o".

pixel C_i and the center pixel I_c. The binarized pixel difference is defined as:

$$x_i = \begin{cases} 1 & \text{if } z_i \geq 0, \\ 0 & \text{if } z_i < 0. \end{cases} \tag{1}$$

Now point C_i is represented by its binarized pixel difference x_i. We cast the LBP structure learning as a point subset selection problem: given binarized pixel differences $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ and target classification variable c, the goal is to find a subset of m binarized pixel differences $\mathbf{x}_m \subseteq \mathbf{x}$ that "optimally" characterize c. We solve the problem via the proposed Maximum-Conditional-Mutual-Information scheme, which will be elaborated in detail in the next section. As a result, the feature dimensionality is reduced from 2^n to 2^m. Then, we apply PCA to further reduce it to d.

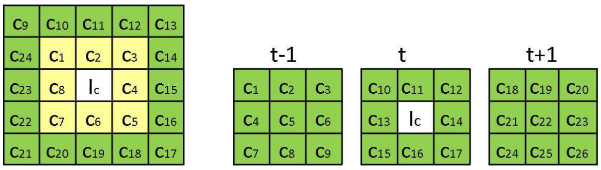
Potentially, the proposed approach could handle a large set of binarized pixel differences, whereas approaches that directly extract features from the histogram bins [3, 8, 31, 32, 43–45] cannot as it is not feasible to enumerate 2^n bins for a large n. This is even more crucial when handling a large-scale database.

Even in the scenario that it is possible to enumerate all histogram bins, the proposed approach may still yield a better performance. The joint probability $p(\mathbf{x})$ of set $\mathbf{x}$ can be estimated by the LBP histogram of 2^n bins generated from these n points. If point x_j is not selected, the marginal probability of the remaining point set is $\sum_{x_j} p(\mathbf{x})$. Correspondingly, each resulting bin is the summation of two original histogram bins. Eventually, when we derive the LBP structure of m points out of these n candidates, we reduce the number of histogram bins from 2^n to 2^m, and each resulting bin is the summation of 2^{n-m} original bins. Thus, the resulting histogram is less noisy, more statistically significant and reliable than the original histogram. In contrast, for those direct-bin-selection approaches [3, 8, 31, 32] the feature set is directly extracted from the histogram of 2^n bins, which may be statistically insignificant and error-prone. Thus, we expect a better generalization performance for the proposed approach.

Now we briefly discuss how to determine the potential candidates. In LQP [3], disc structure in a neighborhood of 5×5 pixels has demonstrated a superior performance over other geometries. Thus, we use the same neighborhood, and the binarized pixel differences between 24 neighbors and the center pixel as potential candidates, as shown in Fig. 2(a). The LBP structure of CENTRIST feature [8] consists of 8 neighbors in the neighborhood of 3×3 pixels, as highlighted in yellow in Fig. 2(a). For spatial-temporal LBP (STLBP), we consider a spatial-temporal neighborhood of 3×3 pixels in 3 successive frames, resulting in 26 binarized pixel differences, as shown in Fig. 2(b). It is not feasible to directly use all 26 neighbors to construct the histogram as $2^{26} = 67,108,864$. Thus, we treat them as potential candidates and aim to find the optimal subset.

3. An Incremental Maximal-Conditional-Mutual-Information Scheme for LBP Structure Learning

Due to the spatial dependence among image pixels within a small neighborhood, the binarized pixel differences are strongly correlated. Previous feature-



(a) (b)

Figure 2: Potential candidates. (a) The binarized pixel differences between 24 neighbors and center pixel in a neighborhood of 5 × 5 pixels as potential candidates. LQP $Disc_5^{3*}$ [3] utilizes the same set of neighbors. CENTRIST feature [8] utilizes 8 neighbors highlighted in yellow. (b) The binarized pixel differences between 26 neighbors and center pixel of Frame t as potential candidates for STLBP.

selection approaches assume certain degree of feature independence, and hence may not be suitable in our scenario. Thus, we propose an incremental Maximal-Conditional-Mutual-Information scheme. We begin with problem analysis of previous algorithms.

3.1. Review of Feature-Selection Algorithms Based on Mutual Information

In Max-Dependency scheme [37], the dependency of target classification variable c on data distribution is maximized. Mutual information is often employed to characterize the dependency. Given two random variables x and y, their mutual information is defined in terms of probability density functions $p(x)$, $p(y)$ and $p(x, y)$:

$$I(x; y) = \int\int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dxdy. \tag{2}$$

Given a set of features $\mathbf{x}$, the goal is to find a subset of m features $\mathbf{x}_m \subseteq \mathbf{x}$, which jointly have the largest dependency on target classification variable c:

$$\mathbf{x}_m^* = \underset{\mathbf{x}_m \subseteq \mathbf{x}}{\operatorname{argmax}} I(\mathbf{x}_m; c). \tag{3}$$

$$\begin{aligned} I(\mathbf{x}_m; c) &= \int\int p(\mathbf{x}_m, c) \log \frac{p(\mathbf{x}_m, c)}{p(\mathbf{x}_m)p(c)} d\mathbf{x}_m dc \\ &= \int \dots \int p(x_1, \dots, x_m, c) \log \frac{p(x_1, \dots, x_m, c)}{p(x_1, \dots, x_m)p(c)} \\ & dx_1 \dots dx_m dc. \end{aligned}$$

It is difficult to reliably estimate $p(x_1, \ldots, x_m)$ and $p(x_1, \ldots, x_m, c)$ due to limited training samples available and the large number of joint states to be estimated. Thus, approximated algorithms were often utilized [37, 38]. In Max-Relevance scheme, $I(\mathbf{x}_m; c)$ is approximated by the mean value of mutual information between individual feature x_i and c:

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

Max-Relevance has a strong assumption that all features are independent to each other [37], which in general does not hold. The features selected according to Max-Relevance may have rich redundancy. To solve this problem, the criterion of Min-Redundancy was added to select mutually exclusive features.

$$\mathbf{x}_m^* = \underset{\mathbf{x}_m}{\operatorname{argmin}} \frac{1}{m^2} \sum_{x_i, x_j \in \mathbf{x}_m} I(x_i; x_j). \tag{5}$$

In [37], Min-Redundancy and Max-Relevance (mRMR) were combined:

$$\mathbf{x}_m^* = \underset{\mathbf{x}_m}{\operatorname{argmax}} \sum_{x_i \in \mathbf{x}_m} I(x_i; c) - \frac{1}{m} \sum_{x_i, x_j \in \mathbf{x}_m} I(x_i; x_j). \tag{6}$$

mRMR has two weak assumptions that high-order interaction information is negligible and the desired features are conditional independent given classification variable c [46]. In many cases, these two assumptions do not hold. Particularly in our case, the binarized-pixel-difference features are strongly correlated, and hence the high-order interaction information among these features is non-negligible.

Both Max-Relevance and mRMR only approximate Max-Dependency criterion intuitively. Recent research [46] shows that when high-order interaction information is negligible, $I(\mathbf{x}_m; c)$ can be approximated by:

$$I(\mathbf{x}_m; c) \approx \sum_{x_i \in \mathbf{x}_m} I(x_i; c) - \sum_{x_i, x_j \in \mathbf{x}_m} I(x_i; x_j) + \sum_{x_i, x_j \in \mathbf{x}_m} I(x_i; x_j | c), \tag{7}$$

where $I(x_i; x_j | c)$ is conditional mutual information. For discrete random variables x, y, z, the conditional mutual information is defined as:

$$\begin{aligned} I(x; y|z) &= \mathbb{E}_z\{I(x; y)|z\} \\ &= \sum_{x,y,z} p(x, y, z) \log \frac{p(z)p(x, y, z)}{p(x, z)p(y, z)}, \end{aligned} \tag{8}$$

where $\mathbb{E}_z\{.\}$ is the expectation on z.

It can be observed that only when Eqn. (7) is dominated by its first term, $\sum_{x_i \in \mathbf{x}_m} I(x_i; c)$ defined in Eqn. (4) for Max-Relevance is a good approximation of $I(\mathbf{x}_m; c)$. It is equivalent to the feature-independence assumption of Max-Relevance. mRMR defined in Eqn. (6) differs from Eqn. (7) 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. It has the assumption that the last term in Eqn. (7) and high-order interaction information are negligible. In our scenario, these two assumptions do not hold. A better approximation, Maximal-Joint-Mutual-Information (MJMI) scheme, is given in [7]. However, MJMI scheme also assumes that high-order interaction is negligible.

A Max-Min-Conditional-Mutual-Information (MmCMI) scheme is proposed in [38], which iteratively selects the feature x_i^* so that:

$$x_i^* = \underset{i}{\operatorname{argmax}} \left\{ \min_j I(x_i; c|x_j) \right\}, \tag{9}$$

where x_j is one of selected features. As we will show shortly, MmCMI implicitly assumes negative interaction information, which may not be true in our case.

3.2. A Close Approximation to $I(\mathbf{x}_m; c)$

In this paper, we propose to approximate $I(\mathbf{x}_m; c)$ closely by a set of low-order conditional mutual information. Our formulation is inspired by [47]. Chow and Liu approximated high-order discrete probability distributions $P(\mathbf{x})$ with dependence trees $\tilde{P}(\mathbf{x})$ as follows:

$$\tilde{P}(\mathbf{x}) = \prod_{i=1}^{n} P(x_{a_i}|x_{a_{j(i)}}), 0 \leq j(i) < i, \tag{10}$$

where $(a_1, a_2, \ldots, a_n)$ is an unknown permutation of integers $1, 2, \ldots, n$ and $P(x_{a_1}|x_{a_0})$ is by definition equal to $P(x_{a_1})$.

Similarly, we would like to approximate $I(\mathbf{x}_m; c)$ by a set of conditional mutual information of the form $I(x_i; c|x_j)$:

$$\tilde{I}(\mathbf{x}_m; c) = \sum_{i=1}^{m} I(x_{b_i}; c|x_{b_{j(i)}}), 0 \leq j(i) < i, \tag{11}$$

where $\{b_1, b_2, \ldots, b_m\}$ is an unknown permutation of integers $1, 2, \ldots, m$ and $I(x_{b_1}; c|x_{b_0})$ is by definition equal to $I(x_{b_1}; c)$. In fact, such a formulation is feasible. Recall the chain rule for $I(\mathbf{x}_m; c)$:

$$I(\mathbf{x}_m; c) = \sum_{i=1}^{m} I(x_i; c|x_1, \ldots, x_{i-1}). \tag{12}$$

For $i \geq 3$, $I(x_i; c|x_1, \ldots, x_{i-1})$ is high-order conditional mutual information. Compare Eqn. (11) with Eqn. (12), we can see that if we could approximate high-order conditional mutual information $I(x_i; c|x_1, \ldots, x_{i-1})$ closely by low-order conditional mutual information of the form $I(x_i; c|x_j)$, we could approximate $I(\mathbf{x}_m; c)$ closely by $\tilde{I}(\mathbf{x}_m; c)$ as defined in Eqn. (11).

Now the question is: how to derive a close approximation of $I(\mathbf{x}_m; c)$ using $\tilde{I}(\mathbf{x}_m; c)$ defined in Eqn. (11)? More specifically, which term $I(x_i; c|x_j)$ should be chosen? At a first glance, we may assume that $I(x_i; c|x_j) \geq I(x_i; c|x_1, \ldots, x_{i-1}), j = 1, 2, \ldots, i-1$. Thus, $I(\mathbf{x}_m; c)$ is upper-bounded by $\tilde{I}(\mathbf{x}_m; c)$, and we should minimize $I(x_i; c|x_j)$ in order to achieve a tighter upper bound. However, this assumption may not hold always. Let us re-examine it:

$$I(x_i; c|x_1, \ldots, x_{i-1}) - I(x_i; c|x_j) = I(x_i; \mathbf{x}_k; c|x_j), \tag{13}$$

where $I(x_i; \mathbf{x}_k; c|x_j)$ is conditional interaction information and $\mathbf{x}_k = \{x_k\}, k = 1, 2, \ldots, i-1, k \neq j$. The interaction information is the gain (or loss) in information among a set of variables due to additional knowledge of the other variables, *i.e.* $I(x; y; z) = I(x; y|z) - I(x; y)$. The conditional interaction information can be obtained as: $I(x; y; z|w) = \mathbb{E}_w\{I(x; y; z)|w\}$.

In general, negative interaction information seems much more natural than positive interaction information as it explains typical common-cause structures. However, in our case we are dealing with positive interaction information. Let us start with a simplified case. The interaction information among x_i, x_j, c can be derived as: $I(x_i; x_j; c) = I(x_i, x_j; c) - I(x_i; c) - I(x_j; c)$, where $I(x_i, x_j; c)$ is joint mutual information. One feature alone, *i.e.* each individual binarized-pixel-difference feature, is rather weak and does not have much discriminative power. Thus, both $I(x_i; c)$ and $I(x_j; c)$ are small. If we consider two features together, the classification capability increases significantly, *i.e.* $I(x_i, x_j; c) > I(x_i; c) + I(x_j; c)$. This is a typical example of positive interaction information. We can view it from another aspect. By definition $I(x_i; x_j; c) = I(x_j; c|x_i) - I(x_i; c)$. In our case, $I(x_i; c)$ is very small and approaches 0. Then, $I(x_i; x_j; c) \approx I(x_j; c|x_i) \geq 0$. We calculate $I(x_i; x_j; c)$ for the 21-land-use dataset [48] and dyntex++ dataset [49, 50]. We find that 92.9% and 96.4% are positive interaction information, respectively. Similarly, $I(x_i; c; \mathbf{x}_k|x_j)$ may also very likely be positive.

Therefore, we conclude that in general for binarized pixel differences, $I(x_i; c|x_j) \leq I(x_i; c|x_1, \ldots, x_{i-1})$. Thus, rather than an upper-bound, $\tilde{I}(\mathbf{x}_m; c)$ is actually a lower bound for $I(\mathbf{x}_m; c)$. We will further verify this in the experimental section. In order to obtain a tighter lower bound, $\tilde{I}(\mathbf{x}_m; c)$ is derived as:

$$\tilde{I}(\mathbf{x}_m; c) = \sum_{i=1}^{m} \max_{j(i)} I(x_{b_i}; c|x_{b_{j(i)}}), 0 \leq j(i) < i. \tag{14}$$

3.3. Proposed Incremental MCMI Scheme

After we approximate $I(\mathbf{x}_m; c)$ closely by $\tilde{I}(\mathbf{x}_m; c)$ defined in Eqn. (14), we achieve Max-Dependency criterion by maximizing $\tilde{I}(\mathbf{x}_m; c)$:

$$\mathbf{x}^*_m = \underset{\mathbf{b}_m}{\text{argmax}} \left\{ \sum_{i=1}^{m} \max_{j(i)} \{ I(x_{b_i}; c|x_{b_{j(i)}}) \} \right\}, \tag{15}$$

where $\mathbf{b}_m = \{b_1, b_2, \ldots, b_m\}$ is an subset of m integers out of $1, 2, \ldots, n$.

Equivalently, our objective function is:

$$\mathbf{b}^*_m = \underset{\mathbf{b}_m}{\text{argmax}} \left\{ \sum_{i=1}^{m} I(x_{b_i}; c|x_{b_{j(i)}}) \right\}, 0 \leq j(i) < i. \tag{16}$$

We call this as Maximal-Conditional-Mutual-Information (MCMI) scheme. As we only need to estimate the joint probability mass function of three variables only, in which $x_{b_i}, x_{b_{j(i)}}$ are binary, the computational cost is low.

It is worth noting that our method is different from MmCMI in [38]. Eqn. (9) of MmCMI can be equivalently rewritten as:

$$\mathbf{x}_m^* = \underset{\mathbf{b}_m}{\operatorname{argmax}} \left\{ \sum_{i=1}^{m} \min_{j(i)} \{ I(x_{b_i}; c | x_{b_{j(i)}}) \} \right\}. \tag{17}$$

Compared it with Eqn. (15), we can see that the proposed MCMI aims to Max-Max the conditional mutual information, whereas MmCMI aims to Max-Min the conditional mutual information. MmCMI implicitly assumes that it is an upper bound of $I(\mathbf{x}_m; c)$ in Eqn. (12), i.e. $I(x_{b_i}; c | x_{b_1}, \ldots, x_{b_{i-1}}) \leq I(x_{b_i}; c | x_{b_{j(i)}})$. Equivalently, it assumes negative conditional interaction information, i.e. $I(x_{b_i}; \mathbf{x}_i; c | x_{b_{j(i)}}) = I(x_{b_i}; c | x_{b_1}, \ldots, x_{b_{i-1}}) - I(x_{b_i}; c | x_{b_{j(i)}}) \leq 0$. Then, MmCMI aims to minimize $I(x_{b_i}; c | x_{b_{j(i)}})$ to achieve a tighter upper bound of $I(\mathbf{x}_m; c)$. In Section 3.2, we have shown that this assumption does not hold true in our case. Later in Section 4.1, we will show using a real dataset that MmCMI is a lower bound rather than an upper bound of $I(\mathbf{x}_m; c)$, and the proposed MCMI achieves a tighter lower bound than MmCMI. By better approximating $I(\mathbf{x}_m; c)$, the proposed MCMI consistently outperforms MmCMI on all 6 datasets in the experimental section.

It is challenging to seek a globally optimal solution to MCMI scheme as it is a NP-hard problem. Even a small pool of 100 features will result in 1.86×10^{11} subsets for selecting just 8 features. Sequential forward selection (SFS) [39] is widely used in feature selection, e.g. MmCMI [38] and mRMR [37]. For simplicity and for a fair comparison with MmCMI, we also utilize SFS in this work.

The proposed algorithm starts with a set of possible point candidates. We initialize the selected point subset as a null set. Denote $\mathbf{s}^i = \{x_j^i\}$ as the current point subset, and $\mathbf{v}^i = \mathbf{x} - \mathbf{s}^i$ as the available feature pool. Then, we aim to find a point in such a way that:

$$x_{j^*}^{i+1} = \underset{x_j^{i+1} \in \mathbf{v}^i, x_s^i \in \mathbf{s}^i}{\operatorname{argmax}} \quad I(x_j^{i+1}; c | x_s^i). \tag{18}$$

Here, we assume the conditional mutual information $I(x_j^1; c | x_s^0) = I(x_j^1; c)$. The proposed incremental MCMI scheme is summarized in Algorithm 1.

Algorithm 1 Proposed incremental MCMI scheme to learn LBP structure

Input: a set of potential candidates $\mathbf{x}$, the desired number of neighbors m

Output: an optimal subset $\mathbf{x}_m$

Initialization: Set the selected feature subset as $\mathbf{s}^0 = \emptyset$ and the feature pool $\mathbf{v}^0 = \mathbf{x}$.

```
for $i \leftarrow 1 : m$ do
    for $j \leftarrow 1 : n - i + 1$ do
        Calculate the conditional mutual information $I(x_j^i; c|x_s^{i-1})$ by Eqn. (8),
        where $x_j^i \in \mathbf{v}^{i-1}, x_s^{i-1} \in \mathbf{s}^{i-1}$.
    end for
    Choose $x_j$ that maximizes $I(x_j^i; c|x_s^{i-1})$.
    Update $\mathbf{s}^i \leftarrow \mathbf{s}^{i-1} \cup x_j$
    Update $\mathbf{v}^i \leftarrow \mathbf{v}^{i-1} - x_j$
end for
$\mathbf{x}_m = \mathbf{s}^m$
```

Image patches at different scales or locations may exhibit totally different characteristics. Thus, the "spatial pyramid" [8, 51] was proposed to preserve the patch-wise location information and capture image characteristics at different scales. Similarly, we propose to derive the LBP structure on a patch-wise basis instead of using a unified one for all patches. Each patch is processed separately. The feature vectors of all patches are concatenated to form the final feature vector.

4. Experimental Results

The proposed approach can be used in many applications. In this paper, we show three examples: learning a single LBP structure for texture classification, a set of patch-wise LBP structures for scene/face recognition and a spatial-temporal LBP structure for dynamic texture recognition. The proposed approach is compared with approaches that directly extract features from the

LBP histogram [3, 8], those utilizing handcrafted LBP structures [5], as well as the state of the art reported in solving these problems.

4.1. Comparison in Approximating $I(\mathbf{x}_m; c)$

We first show that we derive a closer approximation to high-order mutual information $I(\mathbf{x}_m; c)$. Max-Min Conditional Mutual Information scheme (MmCMI) [38] is similar to the proposed one. Thus, we compare it with the proposed scheme in approximating $I(\mathbf{x}_m; c)$.

We use the 21-land-use dataset [48] for illustration. It contains 21 classes of aerial orthoimagery, and each class has 100 images of resolution 256 × 256 pixels. We randomly choose 80 images from each class as the training set, and randomly select 8 points from 24 point candidates as the LBP structure. We calculate the ground-truth value of high-order mutual information $I(\mathbf{x}_m; c)$ using this 8-bit LBP structure. Then, we approximate $I(\mathbf{x}_m; c)$ using $\max\{\sum_{i=1}^{m} I(x_{b_i}; c|x_{b_{j(i)}})\}, 0 \leq j(i) < i$ for the proposed MCMI scheme, and using $\sum_{i=1}^{m} \max_i\{\min_{j(i)}\{I(x_{b_i}; c|x_{b_{j(i)}})\}\}, 0 \leq j(i) < i$ for MmCMI [38]. We repeat it 100 times, and plot the ground-truth and approximated values of $I(\mathbf{x}_m; c)$ in Fig. 3.

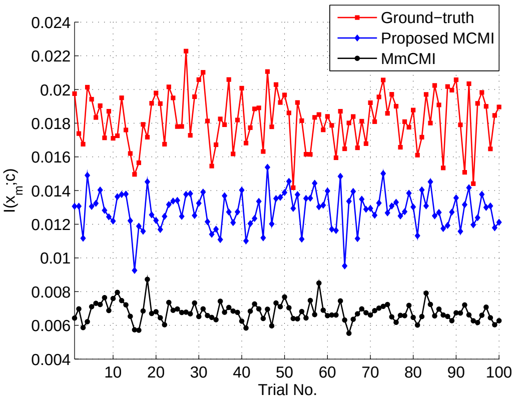


Figure 3: The proposed MCMI achieves a closer approximation to high-order mutual information $I(\mathbf{x}_m; c)$ than MmCMI [38]. The proposed MCMI is a tighter lower bound of $I(\mathbf{x}_m; c)$.

We can observe from Fig. 3 that even we greedily maximize $\tilde{I}(\mathbf{x}_m; c)$ through the proposed incremental MCMI, $\tilde{I}(\mathbf{x}_m; c) < I(\mathbf{x}_m; c)$. Thus, $\tilde{I}(\mathbf{x}_m; c)$ is a lower bound rather than an upper bound of $I(\mathbf{x}_m; c)$. The proposed MCMI achieves a tighter bound than MmCMI [38]. It justifies our Max-Max scheme as defined in Eqn. (15), rather than Max-Min scheme [38]. To compensate possible scaling factor for $\tilde{I}(\mathbf{x}_m; c)$, we calculate the correlation coefficients between the approximated value $\tilde{I}(\mathbf{x}_m; c)$ and the ground-truth value $I(\mathbf{x}_m; c)$ in 100 trials, and show them in Table 1. The proposed MCMI scheme exhibits a larger correlation coefficient than MmCMI. It suggests that the proposed MCMI scheme varies more closely with high-order mutual information $I(\mathbf{x}_m; c)$ than MmCMI. In the rest of the experiments, we will utilize the proposed approach to derive the LBP structures using both MCMI and MmCMI schemes, and show their performances for different applications.

Table 1: Correlation coefficient between $\tilde{I}(\mathbf{x}_m; c)$ and $I(\mathbf{x}_m; c)$ for MmCMI and the proposed MCMI scheme.

Method	Correlation Coefficient
MmCMI [38]	0.20
Proposed MCMI	0.47

4.2. Texture Classification on the KTH-TIPS-2a Dataset

We now compare the proposed approach with others on the KTH-TIPS-2a dataset [52] for texture classification using basic experimental settings. This dataset contains 11 classes, and each class has four samples (groups). In each sample group, there are 72-108 images taken at different scales from different orientations. We follow the same experimental settings as in [3]. Three samples of each class are used for training and the fourth for testing. We report average classification rate over four random partitions. Same as in [3], we use simple 3-nearest-neighbor classifiers with Chi-squared distance. Better results using more sophisticated classifiers such as SVM are available in [52].

We compare the proposed approach with LQP [3], which is an alternative to reduce the LBP feature dimensionality by directly extracting features from a large pool of histogram bins. Many handcrafted structures were proposed in LQP [3], among which $Disc_5^{3*}$ performs the best. $Disc_5^{3*}$ means features extracted in a disc-shape region of 5×5 pixels and $3*$ means split ternary coding [53]. For a fair comparison to LTP and $Disc_5^{3*}$ LQP, a variant of the proposed approach is used, i.e. after deriving the LBP structure, we use it to generate split ternary codes, similarly as in [3, 53]. We use the default setting as in [53] for LTP, e.g. the threshold $t = 5$. We also utilize the proposed approach to extract LTP features using MmCMI scheme [38]. The results are summarized in Table 2. The results for LBP, LTP, $Disc_5^{3*}$ LQP, Weber Law Descriptors (WLD) [54] and color WLD [54] are taken from [3].

Table 2: Comparisons with other approaches on the KTH-TIPS-2a dataset for texture classification.

Method	Recognition Rate
WLD [54]	59.4%
Color WLD [54]	56.5%
LBP [27]	58.7%
LTP [53]	60.7%
$Disc_5^{3*}$ LQP [3]	64.2%
Proposed approach with MmCMI	70.2%
Proposed approach with MCMI	**71.1%**

We can see that the proposed approach outperforms LTP by more than 10%, which shows that the proposed approach is able to derive a structure that has more potential to capture the intrinsic image structures than the handcrafted one. Compared with $Disc_5^{3*}$ LQP that directly extracts features from the LBP-histogram bins, the proposed approach with MCMI scheme improves the classification rate from 64.2% to 71.1%.

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

We conduct experiments on the 21-land-use dataset [48] for scene recognition using the same experimental setup as in [48, 55]. The spatial pyramid [8, 51] is utilized. Each image is divided into 1 patch for level 0, 5 patches for level 1 and 25 patches for level 2, respectively. For each class, we randomly split it into five equal-size sets. Four of the sets are used for training and the held-out set is used for testing. We use CENTRIST feature [8] as the baseline algorithm. As CENTRIST utilizes 8-bit LBP, we derive LBP structures of 8 points for a fair comparison. The learned LBP structures for the first patch of each pyramid level of trial 1 are shown in Fig. 4. These LBP structures are significantly different from each other. For each patch, the intrinsic image characteristic is different, and hence a different LBP structure is needed.

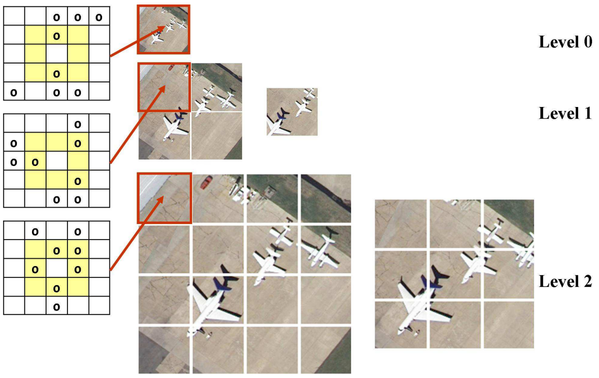


Figure 4: The learned LBP structures for the first trial of 21-land-use dataset.

Then, LBP histograms are generated using these 8-bit patch-wise LBP structures. The concatenated feature vector of 31 patches is of size 256 × 31 = 7936. To derive a compact feature representation, PCA is applied on each patch to reduce the dimensionality from 256 to 50. Image statistics are shown helpful for scene recognition [8], and hence included as features. For CENTRIST, different weights are assigned to features of different pyramid levels. In our experiments, we use $w_0 = 2.4, w_1 = 1.2, w_2 = 1$ for level 0, 1, 2, respectively.

The proposed approach is also compared with $Disc_5^{3*}$ LQP [3]. We implement $Disc_5^{3*}$ LQP according to [3], in which unsupervised k-means algorithm is used to cluster LBP-histogram bins into visual words and linear SVM is used for classification. For a fair comparison, we also include the results for $Disc_5^{3*}$ LQP using RBF SVM. The results of other state-of-the-art approaches are also included in Table 3, such as spatial pyramid co-occurrence kernel (SPCK) [48], extended SPCK (SPCK+) [48], second extended SPCK (SPCK++) [48] and randomized-spatial-partition-based classifier via boosting (BRSP) [55].

Table 3: Comparisons with the state of the art on the 21-land-use dataset for scene recognition.

Method	Recognition Rate
SPCK [48]	73.1%
SPCK+ [48]	76.1%
SPCK++ [48]	77.3%
BRSP [55]	77.8%
CENTRIST [8]	85.9%
$Disc_5^{3*}$ LQP Linear SVM [3]	83.0%
$Disc_5^{3*}$ LQP RBF SVM [3]	85.6%
Proposed approach with MmCMI	87.4%
Proposed approach with MCMI	**88.2%**

The proposed approach extracts features in the same 5 × 5 neighborhood, whereas it outperforms $Disc_5^{3*}$ LQP by 2.6%. It demonstrates the advantages of the proposed approach over directly extracting features from LBP histogram. CENTRIST feature achieves a recognition rate of 85.9%. The proposed approach outperforms it by 2.3%, which shows that the learned LBP structures can better capture the image characteristics than handcrafted structures. Clearly, the proposed approach to extract LBP features demonstrates a superior performance over others. The confusion matrix of the proposed approach with MCMI scheme is given in Fig. 5.

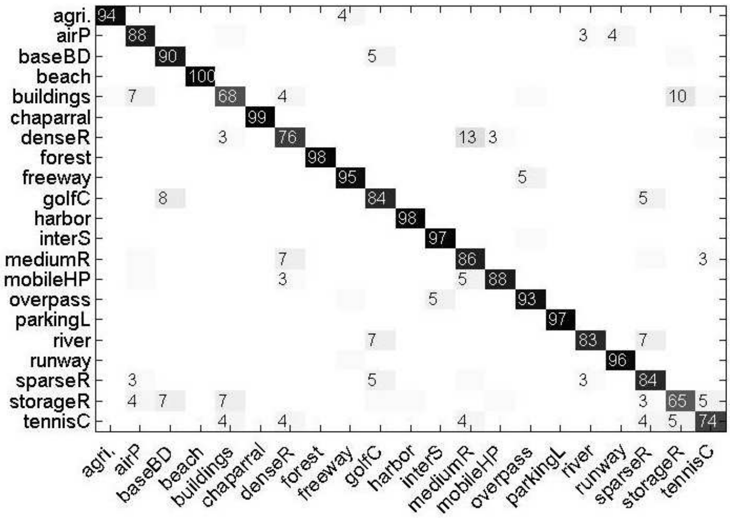


Figure 5: Confusion matrix of the proposed approach with MCMI scheme for the 21-land-use dataset. Only entries with count at least 3 are shown.

4.4. Scene Recognition on the 8-Event Dataset

The 8-event dataset [56] is composed of eight sport classes. Each class has 137 to 250 high-resolution images (from 800 × 600 to thousands of pixels per dimension). To capture the image micro-structures at the same scale, we resize the images so that the minimum image dimension (height or width) is 600 pixels. We use the same setup as in [8, 55, 56]. For each class, we randomly select 70 images for training and 60 for testing. The experiments are repeated 5 times. PCA is applied to reduce the dimensionality of LBP histogram from 256 to 40, same as in [8]. The other parameters are the same as for the 21-land-use dataset.

The learned LBP structures are shown in Fig. 6. These structures are significantly different from those in Fig. 4. As the discriminative information for different applications may not be the same, different LBP structures are needed.

The comparisons with the state of the art are summarized in Table 4. The baseline algorithm CENTRIST built upon handcrafted LBP structure achieves

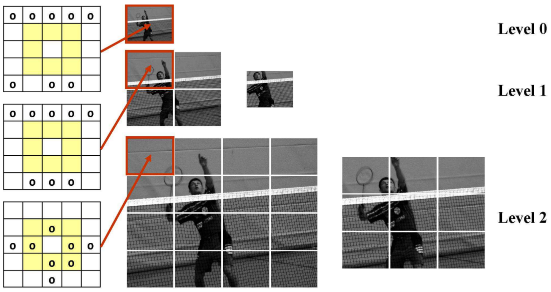


Figure 6: The learned LBP structures for the first trial of 8-event dataset.

a recognition rate of 78.3%. The proposed approach using MCMI scheme significantly improves it to 84.3%. It also outperforms $Disc_5^{3*}$ LQP, which achieves a recognition rate of 78.9% using linear SVM and 77.8% using RBF SVM. The confusion matrix of the proposed approach with MCMI scheme is shown in Fig. 7.

Table 4: Comparison with the state of the art on the 8-event dataset for scene recognition.

Method	Recognition Rate
Scene/Object Model + SIFT [56]	73.4%
RSP + Optimal Selection [55]	77.9%
RSP + Boosting [55]	79.6%
CENTRIST [8]	78.3%
$Disc_5^{3*}$ LQP linear SVM [3]	78.9%
$Disc_5^{3*}$ LQP RBF SVM [3]	77.8%
Proposed approach with MmCMI	83.0%
Proposed approach with MCMI	**84.3%**

4.5. Face Recognition on the HKPU-NIR Dataset

The Hong Kong Polytechnic University near-infra-red (HKPU-NIR) face database [57] consists of around 40,000 near-infra-red face images of 335 subjects. We combine the training, gallery and probe subsets of Experiment 1, 2,

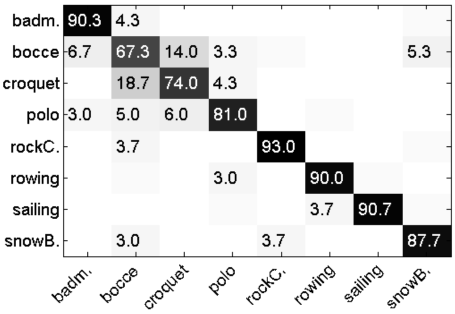


Figure 7: Confusion matrix of the proposed approach with MCMI scheme for the 8-event dataset. Only entries with count at least 3 are shown.

3 used in [57], which results in a large subset of 7,778 images of 298 subjects. The images are normalized to 64 × 64 pixels, the same as in [57]. To reduce illumination variations, the images are photometrically normalized as in [53]. We randomly choose 80% of images from each subject as the training set and the rest as the testing set. The experiments are repeated 5 times and the average performance is reported.

We compare the proposed approach with CENTRIST [8] and LQP [3]. For a fair comparison with CENTRIST, the same spatial pyramid [8, 51] is utilized. For the proposed approach, we use 24 neighbors in Fig. 2(a) as potential candidates, and nearest-neighbor classifier with Chi-squared distance (NNC-Chi2D) for classification. For CENTRIST and LQP, we report the results using linear SVM, RBF SVM and NNC-Chi2D. The results are summarized in Table 5.

As shown in Table 5, NNC-CHi2D performs the best for CENTRIST and linear SVM performs the best for LQP. The proposed approach with MCMI improves the recognition rate by 2.0% compared with CENTRIST + NNC-Chi2D, and by 2.2% compared with LQP + linear SVM.

4.6. Dynamic Texture Recognition on the DynTex++ Dataset

Dynamic texture is sequences of images of moving scenes that exhibit certain stationarity properties in time [49, 50]. The dynamics of texture elements are statistically similar and temporally stationary. The recognition of DT in-

Table 5: Comparison with the state of the art on the HKPU-NIR face dataset.

Method	Recognition Rate
CENTRIST [8], Linear SVM	95.4%
CENTRIST [8], RBF SVM	93.9%
CENTRIST [8], NNC-Chi2D	96.4%
$Disc_5^{3*}$ LQP [3], Linear SVM	96.2%
$Disc_5^{3*}$ LQP [3], RBF SVM	96.0%
$Disc_5^{3*}$ LQP [3], NNC-Chi2D	94.3%
Proposed approach with MmCMI	98.1%
Proposed approach with MCMI	**98.4%**

volves the analysis of both the spatial appearance of static texture patterns and temporal variations in appearance.

The DynTex++ dataset proposed in [4] aims to provide a rich benchmark for DT recognition. It consists of 36 classes of DTs. Each class contains 100 sequences of $50 \times 50 \times 50$ pixels. It mainly contains tightly cropped sequences. Thus, we treat each DT sequence as a whole and do not divide it into patches. We use the same setup as in [4, 50]. For each class, 50 sequences are randomly chosen as the training set, and the other 50 sequences are used in testing. The experiments are repeated 5 times and the average performance is reported. We use 26 binarized pixel differences as potential candidates, as shown in Fig. 2(b). The learned spatial-temporal LBP structures for the DynTex++ dataset are shown in Fig. 8. The number indicates the order of the neighbor being selected. We select up to 16 neighbors. The STLBP structures for 5 trials are fairly consistent. Specifically, the same model is built for Trial 1, 2, 3 and 5, and the first 13 neighbors are the same for 5 trials.

There is one free parameter, *i.e.* the desired number of neighbors m. The proposed approach is scalable. We could vary m to obtain different LBP structures according to the requirements on computational complexity and accuracy.

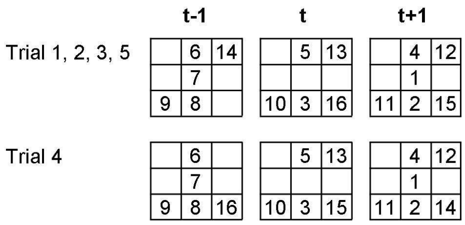


Figure 8: The learned spatial-temporal LBP structures of the DynTex++ dataset for 5 trials.

The respective LBP structures for different m can be easily derived from Fig. 8. In contrast, the handcrafted structure is not scalable. It is difficult to define a consistent rule to build the handcrafted structure for different m. The average recognition rates over 5 trials vs. m are shown in Fig. 9. The highest recognition rate is 95.9% when $m = 16$. The recognition rate does not increase significantly after $m = 10$.

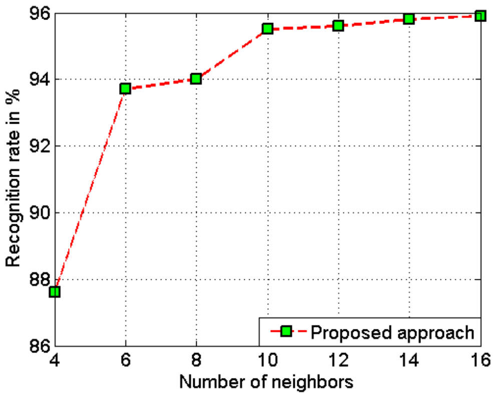


Figure 9: The average recognition rate over 5 trials for the proposed approach built using different number of neighbors.

It is not feasible to enumerate $2^{26} = 67,108,864$ bins of the LBP histogram built using all 26 neighbors. This is exactly the case that direct feature selection/extraction from histogram bins is not applicable. Alternatively, we compare the proposed approach with LBP-TOP [5], in which the large LBP structure is broken into small ones. In LBP-TOP, LBP features are extracted from three orthogonal planes, *i.e.* XY-plane (spatial LBP), XT-plane and YT-plane [5]. The feature vectors of three planes are concatenated to form the final feature vector. We summarize the performance comparisons with the state of the art in Table 6.

Table 6: Comparison with the state of the art on the DynTex++ DT dataset.

Method	Recognition Rate
DL-PEGASOS [4]	63.7%
Dynamic fractal analysis (DFS) [50]	89.9%
LBP-TOP [5]	93.2%
Proposed approach with MmCMI	95.7%
Proposed approach with MCMI	**95.9%**

We implement and test LBP-TOP on the Dyntex++ dataset, which achieves a recognition rate of 93.2%. The proposed approach with MCMI scheme improves the recognition rate to 95.9%, which demonstrates a superior performance to other approaches.

4.7. Dynamic Texture Recognition on the UCLA Database

The UCLA dynamic texture database [58, 59] has been widely used as a benchmark dataset for DT categorization [4, 50, 60, 61]. It consists of 50 classes of DTs, each with 4 sequences captured from different viewpoints. There are several breakdowns when evaluating the dataset [50]. Among them, we choose the following three representative settings:

50-Class: This is the original setting for the UCLA dataset. 4-fold cross-validation is used. The average recognition rate over 4 trials is reported. It is rather a simple setting and a recognition rate of 100% is achieved in [50].

9-Class: Those 50 classes can be clustered to 9 classes by combining the sequences of different viewpoints. We use the same experimental settings as in [4, 50]. We randomly choose half of the dataset as the training set and the other half as the testing set. The experiments are repeated 10 times and the average recognition rate is reported. This setting is challenging and used to evaluate DT recognition under viewpoint changes.

Shift-invariant Recognition(SIR)-class: In [61], each original sequence is spatially partitioned into non-overlapping left and right halves and 400 sequences are obtained. The "shift-invariant recognition" [61] was implemented to compare the sequences only between halves to test the shift-invariant property. This setting is very challenging. Thus, in general rank-5 recognition rate was reported [50, 61]. [1]

We try different ways to divide DTs into patches. The best performance is achieved using the following settings: for 50-class and SIR, we divide DTs along time axis into two patches of equal size; for 9-class, we spatially divide the sequences into 3×3 patches. We use the same potential candidates as shown in Fig. 2(b). As the number of training samples is limited, to avoid over-fitting we train one 12-neighbor STLBP structure for all patches. Nearest-neighbor classifier with Chi-squared distance is utilized. The proposed approach is compared with published methods in literature [4, 50, 61], as well as LBP-TOP [5] built upon handcrafted structures. The results are summarized in Table 7.

Compared with previous results [4, 50, 61] in literature, the proposed approach demonstrates a superior performance. On 9-class setting, it reduces the error rate from 2.5% to 1.6%. On the most challenging SIR setting, it improves the recognition rate from 73.8% to 94.5%. Compared with LBP-TOP [5], the proposed approach achieves a much better performance on 50-class and 9-class setting, and a slightly better performance on SIR setting. The performance of LBP-TOP on SIR setting is also significantly better than previous results. This is partially because LBP-TOP feature is less affected by shift operation, especially the LBP histograms extracted from XT-plane and YT-plane.

[1] If among top-5 matches there is at least one gallery sample with the same label as the probe sample's, it is counted as a successful recognition. The recognition rate calculated in this way is rank-5 recognition rate.

Table 7: Comparison with the state of the art on the UCLA DT dataset for 50-class, 9-class and SIR settings.

Method	50-Class	9-Class	SIR
Distributed spacetime orientation [61]	81.0%	-	60.0%
DL-PEGASOS[4]	99.0%	95.6%	-
DFS [50]	100.0%	97.5%	73.8%
LBP-TOP [5]	87.5%	85.8%	93.8%
Proposed approach with MmCMI	99.5%	98.4%	92.3%
Proposed approach with MCMI	**100.0%**	**98.4%**	**94.5%**

4.8. Analysis of Computational Complexity

We analyze the computational complexity in both off-line training stage and online testing stage. In the training stage, compared with handcrafted LBP/LTP, we need an additional step to estimate the conditional mutual information $I(x_i; c|x_j)$ in order to determine the LBP structure. To derive $I(x_i; c|x_j)$, we need to estimate the joint probability mass function (PMF) $p(x_i, x_j, c)$. When incremental search is utilized, we need to estimate roughly $n(m-1)$ such PMFs, in which x_i, x_j are binary. Each PMF corresponds to a histogram of $4N_c$ bins, where N_c is the number of classes. The computational complexity is $O(mnN_c)$. Two points are worth to mention. Firstly, we only need to store the estimated histogram of a class, but not histograms of all samples in this class. Thus, much less memory is required, which is particularly important when dealing with a large-scale database. Secondly, the computational complexity grows linearly with the number of candidates n and the desired number of neighbors m. In general, we expect $m \leq 24$ for binary encoding and $m \leq 16$ for ternary encoding. Thus, ideally we could utilize a large number of potential candidates, and hence deal with high-dimensional LBP features. For those applications we tested, we find that $n = 24$ for spatial LBP and $n = 26$ for spatial-temporal LBP yield a satisfactory performance. In the testing stage, the LBP structures derived in the training stage can be used in the same way as handcrafted ones.

The training and testing time of different algorithms is summarized in Table 8. The 21-land-use dataset has 1680 training images and 420 testing images of size 256 × 256 pixels, whereas the HKPU-NIR dataset has 6097 training images and 1681 testing images of size 64 × 64 pixels. We report the testing time using RBF SVM on the 21-land-use dataset and using NNC-Chi2D on the HKPU-NIR dataset. In our experiment, we use Matlab 2013a on Intel(R) Core(TM) i7-3930K CPU @ 3.20 GHz with 16Gb memory.

Table 8: Comparisons of time consumption for CENTRIST, $Disc_5^{3*}$ LQP and the proposed approach on the 21-land-use dataset and the HKPU-NIR dataset.

Method	21-land-use		HKPU-NIR	
	Training (s)	Testing (s)	Training (s)	Testing (s)
CENTRIST [8]	65.7	18.6	109.6	109.2
$Disc_5^{3*}$ LQP [3]	4779.5	53.3	1747.3	465.7
Proposed approach	1417.7	38.9	1535.5	604.1

CENTRIST [8] utilizes handcrafted structures, which can be viewed as the baseline for time comparison. The proposed approach needs additional time to derive the LBP structures. Even so, the extra time required is still reasonable. We also include the time cost of LQP for reference. We use FKMEANS package [62] for k-means clustering of LQP. It can be seen that both sample size and image size affect the training time. Although the HKPU-NIR dataset has about 4 times of training samples of the 21-land-use dataset, the training time does not significantly increase for the proposed approach. The testing time is much less than the training time, particularly when SVM is used for classification.

5. Conclusion

In general, LBP features using more bits improve the performance, at the cost of high feature dimensionality. The dimensionality easily goes beyond the capability of approaches that directly extract features from the histogram bins as it increases exponentially with the number of LBP bits. Alternatively, the large structure was broken into small handcrafted ones. However, handcrafted LBP may not be optimal. Instead, we learn discriminative LBP structures from image and video data. The LBP structure learning is casted as a point-selection problem, where the optimal subset of neighboring pixels rather than histogram bins are selected. Due to the high dependency among image pixels in a local neighborhood, existing feature-selection algorithms may not be suitable in our scenario. We thus propose a novel feature-selection method by approximating the high-order mutual information with a set of low-order conditional mutual information, and achieving Max-Dependency via maximizing the approximated one. As shown in the experiments, our proposed incremental MCMI scheme can well solve the LBP-structure-learning problem. Moreover, the proposed approach is readily incorporated to spatial pyramid framework that can better capture the intrinsic characteristics of image patches at different locations and scales. The proposed approach outperforms the state-of-the-art solutions to scene and dynamic texture recognition significantly on several benchmark datasets.

Acknowledgment

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.

References

[1] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognition 29 (1) (1996) 51–59.

[2] X. Qian, X.-S. Hua, P. Chen, L. Ke, PLBP: An effective local binary patterns texture descriptor with pyramid representation, Pattern Recognition 44 (1011) (2011) 2502 – 2515.

[3] S. ul Hussain, B. Triggs, Visual recognition using local quantized patterns, in: European Conference on Computer Vision, Springer, 716–729, 2012.

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

[5] G. Zhao, M. Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (6) (2007) 915–928.

[6] J. Ren, X. Jiang, J. Yuan, Dynamic texture recognition using enhanced LBP features, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2400–2404, 2013.

[7] J. Ren, X. Jiang, J. Yuan, G. Wang, Optimizing LBP Structure for Visual Recognition Using Binary Quadratic Programming, IEEE Signal Processing letters 21 (11) (2014) 1346 – 1350.

[8] J. Wu, J. Rehg, CENTRIST: A visual descriptor for scene categorization, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (8) (2011) 1489–1501.

[9] J. Ren, X. Jiang, J. Yuan, Learning binarized pixel-difference pattern for scene recognition, in: IEEE International Conference on Image Processing (ICIP), 2013.

[10] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: Application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) 2037–2041.

[11] S. U. Hussain, T. Napoléon, F. Jurie, et al., Face recognition using local quantized patterns, in: British Machive Vision Conference, 2012.

[12] A. Hadid, M. Pietikinen, Combining appearance and motion for face and gender recognition from videos, Pattern Recognition 42 (11) (2009) 2818 – 2827.

[13] Z. Liu, C. Liu, Fusion of color, local spatial and global frequency information for face recognition, Pattern Recognition 43 (8) (2010) 2882 – 2890.

[14] B. Jun, T. Kim, D. Kim, A compact local binary pattern using maximization of mutual information for face analysis, Pattern Recognition 44 (3) (2011) 532 – 543.

[15] B. Jun, D. Kim, Robust face detection using local gradient patterns and evidence accumulation, Pattern Recognition 45 (9) (2012) 3304 – 3316.

[16] J. Ren, X. Jiang, J. Yuan, Relaxed local ternary pattern for face recognition, in: IEEE International Conference on Image Processing (ICIP), 2013.

[17] J. Ren, X. Jiang, J. Yuan, Noise-Resistant Local Binary Pattern with an Embedded Error-Correction Mechanism, IEEE Transactions on Image Processing 22 (10) (2013) 4049–4060.

[18] M. Bereta, P. Karczmarek, W. Pedrycz, M. Reformat, Local descriptors in application to the aging problem in face recognition, Pattern Recognition 46 (10) (2013) 2634 – 2646.

[19] R. Mehta, J. Yuan, K. Egiazarian, Face recognition using scale-adaptive directional and textural features, Pattern Recognition 47 (5) (2014) 1846 – 1858.

[20] L. Nanni, A. Lumini, Local binary patterns for a hybrid fingerprint matcher, Pattern Recognition 41 (11) (2008) 3461–3466.

[21] M. Heikkilä, M. Pietikäinen, C. Schmid, Description of interest regions with local binary patterns, Pattern Recognition 42 (3) (2009) 425–436.

[22] L. Nanni, S. Brahnam, A. Lumini, A simple method for improving local binary patterns by considering non-uniform patterns, Pattern Recognition 45 (10) (2012) 3844 – 3852.

[23] C. Zhu, C.-E. Bichot, L. Chen, Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition 46 (7) (2013) 1949 – 1963.

[24] D. T. Nguyen, P. O. Ogunbona, W. Li, A novel shape-based non-redundant local binary pattern descriptor for object detection, Pattern Recognition 46 (5) (2013) 1485 – 1500.

[25] A. Satpathy, X. Jiang, H.-L. Eng, Human detection by quadratic classification on subspace of extended histogram of gradients, IEEE Transactions on Image Processing 23 (1) (2014) 287–297.

[26] A. Satpathy, X. Jiang, H.-L. Eng, LBP based edge-texture features for object recognition, IEEE Transactions on Image Processing 23 (5) (2014) 1953–1964.

[27] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002) 971–987.

[28] Y. Guo, G. Zhao, M. Pietikinen, Discriminative features for texture description, Pattern Recognition 45 (10) (2012) 3834 – 3843.

[29] C. Shan, T. Gritti, Learning discriminative LBP-histogram bins for facial expression recognition, in: British Machine Vision Conference, 10, 2008.

[30] J. Yuan, M. Yang, Y. Wu, Mining discriminative co-occurrence patterns for visual recognition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2777–2784, 2011.

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

[32] S. Yan, H. Wang, X. Tang, T. Huang, Exploring Feature Descritors for Face Recognition, in: International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 1, 629–632, 2007.

[33] G. Zhao, M. Pietikäinen, Dynamic texture recognition using volume local binary patterns, in: Dynamical Vision, Springer, 165–177, 2007.

[34] R. Gupta, H. Patil, A. Mittal, Robust order-based methods for feature description, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 334–341, 2010.

[35] D. Maturana, D. Mery, A. Soto, Learning discriminative local binary patterns for face recognition, in: IEEE International Conference on Automatic Face Gesture Recognition and Workshops, 470–475, doi:\bibinfo{doi}{10.1109/FG.2011.5771444}, 2011.

[36] Z. Lei, M. Pietikainen, S. Li, Learning Discriminant Face Descriptor, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (2) (2014) 289–302, ISSN 0162-8828, doi:\bibinfo{doi}{10.1109/TPAMI.2013.112}.

[37] H. Peng, F. Long, 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 27 (8) (2005) 1226–1238.

[38] F. Fleuret, Fast binary feature selection with conditional mutual information, Journal of Machine Learning Research 5 (2004) 1531–1555.

[39] A. Jain, R. Duin, J. Mao, Statistical pattern recognition: A review, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (1) (2000) 4–37.

[40] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, TIST 2 (2011) 27:1–27:27.

[41] M. Calonder, V. Lepetit, C. Strecha, P. Fua, Brief: Binary robust independent elementary features, in: Computer Vision–ECCV 2010, Springer, 778–792, 2010.

[42] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: an efficient alternative to SIFT or SURF, in: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, 2564–2571, 2011.

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

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

[45] C. Shan, Learning local binary patterns for gender classification on real-world face images, Pattern Recognition Letters 33 (4) (2012) 431–437.

[46] G. Brown, An information theoretic perspective on multiple classifier systems, in: Multiple Classifier Systems, Springer, 344–353, 2009.

[47] C. Chow, C. Liu, Approximating discrete probability distributions with dependence trees, IEEE Transactions on Information Theory 14 (3) (1968) 462–467.

[48] Y. Yang, S. Newsam, Spatial pyramid co-occurrence for image classification, in: International Conference on Computer Vision, 1465–1472, 2011.

[49] G. Doretto, A. Chiuso, Y. Wu, S. Soatto, Dynamic textures, International Journal of Computer Vision 51 (2) (2003) 91–109.

[50] Y. Xu, Y. Quan, H. Ling, H. Ji, Dynamic texture classification using dynamic fractal analysis, in: International Conference on Computer Vision, 1219 –1226, 2011.

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

[52] B. Caputo, E. Hayman, P. Mallikarjuna, Class-specific material categorisation, in: International Conference on Computer Vision, 1597–1604, 2005.

[53] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, IEEE Transactions on Image Processing 19 (6) (2010) 1635–1650.

[54] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, W. Gao, WLD: A robust local image descriptor, Pattern Analysis and Machine Intelligence, IEEE Transactions on 32 (9) (2010) 1705–1720.

[55] Y. Jiang, J. Yuan, G. Yu, Randomized Spatial Partition for Scene Recognition, in: European Conference on Computer Vision, 730–743, 2012.

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

[57] B. Zhang, L. Zhang, D. Zhang, L. Shen, Directional binary code with application to PolyU near-infrared face database, Pattern Recognition Letters 31 (14) (2010) 2337–2344.

[58] P. Saisan, G. Doretto, Y. Wu, S. Soatto, Dynamic texture recognition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, 58–63, 2001.

[59] A. Chan, N. Vasconcelos, Probabilistic kernels for the classification of auto-regressive visual processes, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 846–851, 2005.

[60] A. Ravichandran, R. Chaudhry, R. Vidal, View-invariant dynamic texture recognition using a bag of dynamical systems, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1651–1657, 2009.

[61] K. G. Derpanis, R. P. Wildes, Dynamic texture recognition based on distributions of spacetime oriented structure, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 191–198, 2010.

[62] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, in: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 1027–1035, 2007.