IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 29, 2018, accepted August 7, 2018, date of publication August 10, 2018, date of current version September 5, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2864754

TCvBsISM: Texture Classification via B-Splines-Based Image Statistical Modeling

JINPING LIU[1,2], JIEZHOU HE[1], WUXIA ZHANG[1], PENGFEI XU[1], AND ZHAOHUI TANG[3]

[1]College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China
[2]Key Laboratory of High Performance Computing and Stochastic Information Processing (Ministry of Education of China), College of Mathematics and Statistics, Hunan Normal University, Changsha 410081, China
[3]School of Information Science and Engineering, Central South University, Changsha 410083, China

Corresponding author: Pengfei Xu (pengfeixu@sina.com)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61501183 and Grant 61771492, in part by the Hunan Natural Science Foundation under Grant 2018JJ3349, in part by the Joint Fund of NSFC and Guangdong Provincial Government under Grant U1701261, in part by the Young Teacher Foundation of Hunan Normal University under Grant 11405, and in part by the MOE Key Laboratory of Image Processing and Intelligence Control under Grant IPIC2017-03.

ABSTRACT This paper presents an image statistical modeling-based texture classification (TC) approach via the Bayesian-driven B-splines probability density estimation of the image textural surface appearance (ITSA), termed TCvBsISM. It approximates the probability density functions (pdfs) of the marginal distribution and joint distribution, involving the global organization and the locally structural layout of local homogeneous patches in the texture surface, respectively, of both the image raw pixel space and the filter response space, by the linear combination of B-spline basis functions (BsBFs) for ITSA feature characterization. The corresponding linear weighting coefficients (LWCs) are determined by an entropy-based optimization criterion with a prior smooth constraint over the LWCs. By leveraging the B-spline-based pdf modeling, distinctive ITSA structural features of texture images are characterized by the LWCs of the pre-defined BsBFs, which are then embedded in an integrated statistical feature dictionary learning, texture pattern representation, and discrimination model to perform TC. Extensive confirmative and comparative experiments on three different texture databases and one natural environmental scene database demonstrate that the proposed TCvBsISM is very promising, especially when the images of different texture patterns appear to be quite similar with the limited training samples. The effects of various parameters on TCvBsISM, such as the choice of the filter bank, the size of the image statistical feature dictionary, as well as the number of BsBFs, are also discussed.

INDEX TERMS Image statistical modelling, B-spline probability density estimation, minimum entropy, Bayesian probability density estimation, texture classification.

I. INTRODUCTION

Texture, ubiquitously existing in all-natural images, is an important visual perception clues to a variety of image analyses and computer vision tasks [1], such as image segmentation, objection detection, image retrieval, and many others more cannot be enumerated exhaustively. Hence, texture analysis, especially texture classification (TC) constitutes one of the fundamental issues in computer vision. Nevertheless, it is even difficult to get a precise definition for computer texture modeling and texture analysis, though human beings can discriminate different texture patterns effortlessly. Hence, TC is still an indispensable but challenging task in computer vision and some related intelligent information processing fields [2].

A common TC system involves two modules, texture feature extraction (or feature learning) and pattern identification [1], [3]. It is believed that distinctive texture feature extraction is a bottleneck in automatic TC [1], hence that has attracted researchers' great interest. During the past decades, local descriptor-based texture analysis methods have become a mainstream, which can be categorized into two classes [4]. One is the sparse descriptor approach based on some detectable interest points, and the other is the dense descriptor approach extracting pixel-wise features of local patches.

Representative sparse descriptor is the scale invariant feature transform(SIFT) proposed by Lowe [5] with its variants, e.g., the speed up version, SURF [6] and the

dimension-reduction version, PCA-SIFT [7]. Whereas the most popular dense descriptors include Gabor wavelet [8] and local binary pattern(LBP) [9] with its variants, e.g., the pairwise rotation invariant co-occurrence local binary pattern (PRICoLBP) [10], dual-tree complex wavelet transform (DTCWT)-fused LBP feature termed LBDTCWT as well as the energies in the log-polar(LP) transform domain termed LPDTCWTE [11], diamond sampling structure-based local adaptive binary pattern (DLABP) [12], and so on. Though ample studies have demonstrated that the LBP-based descriptors can achieve promising TC performance on some specific data sets, they still exhibit limitations in fully capturing the discriminative texture pattern information as analyzed in [12], since they mainly concern the change of the stimulus between the center pixel and its neighbor pixels and ignore the pixel change trends and the original intensities of the local homogeneous patches (LHPs). Hence, some researchers attempted to extract more discriminant local pattern information beyond the local paired pixels. For instance, Chen *et al.* [4] proposed a robust descriptor termed Weber Local Descriptor (WLD), encoding both differential excitations and orientations at certain locations based on the fact that human perception of a texture pattern depends not only on the change of a stimulus but also on the original intensity of the stimulus. Analogically, Pan *et al.* [13] proposed a feature-based LBP (FbLBP), taking the magnitude of the difference vector computed based on the traditional LBP as a complementary. Whereas Yu *et al.* [14] proposed a multitrend binary code descriptor (MTBCD) fused the gray level co-occurrence matrix (GLCM) texture statistics computation method, attempting to reveal the trend of pixels change in local region and the spatial relation of local patterns.

In addition, some researchers combined the ideas of sparse and dense descriptors in a crossway. For instance, Li et al. used the SIFT in a dense sampling way [15] and Heikkila extracted the LBP base on the interest regions [16] of the sparse way.

All these local descriptor-based texture analysis methods attempt to choose a limited subset of the texture features from local patches [1] of raw pixel space or filtering response space. Despite ample evidence has shown the good performance of these methods on some specific texture databases, "*no single approach did perform best or very close to the best for all images, thus no single approach may be selected as the clear winner*" as it was concluded in a comparative study [17].

One big limitation of these local descriptor methods is the neglecting of the global characteristics of the holistic image surface appearance. As stated in [1], the Bag-of-words(BoW) model, borrowed from natural language processing and document classification areas, is an effective local-global texture characteristics representation approach, gaining great interest in the academic community and leading a new prospect in TC. It encodes both the local texture pattern by using LHPs-based pre-learned *textons* [18], [19] and the global texture appearance by an orderless histogram of the occurrence frequencies of *textons*. However, it is worth noting that the patch size is a non-negligible influence factor. As analyzed in [1], smaller sized patches can achieve more local characteristics, but cannot effectively capture larger scale spatial structural characteristics, which may be a dominant texture feature but sensitive to capture pose, illumination change and noise contamination. Whereas larger sized patches will lead to a quadratic increase in the dimension of the patch space and bring about great pressure in the *textons* learning.

Intuitively, the ITSA essentially dues to the albedo variation on a flat or three-dimensional surface, which exhibits a special organizational behavior of the LHPs. The local structural characteristics and the global organization behavior of the LHPs should attribute to the spatial variation of the image pixel intensity values and finally be boiled down to the variant statistical distribution characteristics. In other words, the random organization of the LHPs generates the global visual appearance, while the regular or irregular exhibition or layout of each LHP results in the local characteristics. Hence, how to characterize the distinctive global organizations and the local exhibition of the ITSA is greatly important to TC [20].

Many statistical modeling-related methods were introduced to describe these global organizations and local exhibition characteristics for texture analysis. The early widespread methods are the statistics-based methods, such as the first-order and second-order statistics, and co-occurrence matrix-based joint statistics, as well as their variants [21], [22]. The statistics-based texture analysis methods do not concern the latent probability distribution model (LPDM) of the ITSA, which mainly depend on the measured values of the predefined statistics. That may achieve analogous statistical values from different texture images from different texture patterns with different LPDMs and consequently lead to misclassifications. Hence, investigating the LPDM of the ITSA for accurate TC is important and urgent [23].

Some empirical statistical models (ESMs) were introduced to approximate the LPDM of the ITSA especially in a specific image transform domain. These ESMs are generally sharp-peaked distribution models with long tail on left, right or both sides of the distribution, e.g., Gaussian scale mixture model (GSMM) [24], symmetric Weibull distribution model [20], [25]–[27], which mainly capture the global characterizations of texture images based on the evidence of the sparse behavior of texture images in the multi-scale transform spaces [28]. To characterize the local exhibition characteristics of the ITSA, the MSD models can be extended to multivariate statistical distribution (MvSD) models, e.g., the joint distribution, representing the statistical dependence of the pixels with a specific distance and orientation. These MSD and MvSD models can be also used as a prior to improve the capabilities of the image processing technologies. For instance, Hammond and Simoncelli [29] proposed a GSMM-based image denoising method with Bayesian least squares estimation and demonstrated to achieve the

best performance among the state-of-the-art image denoising methods.

However, ESMs-based texture analysis methods are difficult to approximate the statistical model of texture images with unexpected distribution profiles. For instance, the commonly-used ESMs are generally unimodal, which cannot approximate LPDMs of ITSAs with multimodal distribution profiles. Another negative influence is that the ESMs-based texture analysis methods generally depend on the researchers' knowledge experience. Hence, it is bound to achieve poor TC performance if the unreasonable empirical model is introduced. Recently, the basis distribution or polynomial based PDF estimation (PDFE) methods attract increasing interest, which can approximate any complicated distribution model without any prior knowledge of the true distribution by introducing Bayesian smooth constraints [30], [31].

This paper presents a B-spline image statistical modeling-based TC method, termed TCvBsISM. It derived the one-dimensional and two-dimensional BsBFs-based PDFE for ITSA feature characterization by introducing a prior smooth constraint on the distribution model, which can effectively avoid the fluctuation of the PDFE results inspired by the literature [31] and consequently achieve the distinct LPDMs of any texture pattern. Based on the pre-fixed BSBFs, it converted the global organization and the local layout of LHPs in a texture image into the corresponding linear weighting coefficients (LWCs) of the pre-fixed BSBFs, which is beneficial to the automatic TC. The performance of TCvBsISM is verified by extensive experiments on three different texture image databases and one natural environmental scene data set, which achieved superior performance compared with the state-of-the-art TC methods.

The remainder of this paper is organized as follows: the Bayesian-driven image statistical modeling method based on B-splines PDFE is presented in Section II. Detailed TCvBsISM is presented in Section III. Experimental results on three different textural image databases and one natural environmental image data set with the effects of various parameters in TCvBsISM approach are discussed in Section IV and conclusions are given in Section V.

II. BAYESIAN-DRIVEN IMAGE STATISTICAL DISTRIBUTION MODELING

ESMs, e.g., Gaussian distribution, Weibull distribution or exponential distribution, cannot effectively capture the complicated and irregular distribution profiles of ISTAs for the diversity and unpredictability of the texture patterns. Common nonparametric PDFE methods, e.g., histogram, kernel density estimation, either approximate the PDF by a non-smooth bar chart with low estimation precision or by a continuous function without convenient and applicable expression form for the following pattern classification.

Recently, some researchers attempted to use the mixture of polynomials, e.g., normal polynomials, Hermit polynomials, to approximate the distribution model. It demonstrated that the high order polynomials were not stable and the lower order polynomials may not achieve the PDF model with sufficient precision. Alternatively, the using of basis distributions or positive kernels with some constraints is demonstrated to achieve the PDF approximation result with enough preciseness [31], [32].

Hence, we approximate the PDF of ITSA by the linear weighting combination of some pre-defined BsBFs, which converts the PDFE into a weight optimization issue. By introducing a prior smooth constraint on the linear weighting coefficients(LWCs), we derived a stable and reliable Bayesian-driven BSBFs-based PDFE method, including the one-dimensional PDF modeling and the two-dimensional modeling, which can be adaptive to various random samples with no prior distribution information of the true statistical models.

A. BsBFs-BASED ONE-DIMENSIONAL PDF ESTIMATION

1) ONE-DIMENSIONAL B-SPLINE PDF MODELLING

Suppose a random variable $\boldsymbol{x}$ within the range of $[t_{\min}, t_{\max}]$, the PDF of $\boldsymbol{x}$ can be approximated by the linear weighting combination of BSBFs, given by

$$f_X(x|\mathbf{a}) = \mathbf{a}^T \mathcal{B}_k = \sum_{i=1}^{\mathcal{M}} a_j \beta_{j,k}(x)\, t_{\min} \le x \le t_{\max} \tag{1}$$

where $\mathbf{a} = (a_1, a_2, \cdots, a_{\mathcal{M}})^T$ is the LWCV, $\beta_{j,k}(x)$ is the j-th BsBF of k-order, $\mathcal{M}$ is the number of BSBFs. The BsBF can be defined in a recursive form as

$$\beta_{j,k}(x) = \frac{x - t_i}{t_{i+k-1} - t_i}\beta_{j,k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}}\beta_{j+1,k-1}(x) \tag{2}$$

where $\beta_{j,0}(x) = \begin{cases} 1, & if\ t_i \le x \le t_{i+1} \\ 0, & \text{otherwise} \end{cases}$ with a specific definition $0/0 = 0$, $\mathcal{T} = (t_0, t_1, \cdots, t_n)^T$ is a knot vector, consisting of $n+1$ non-descending real-valued knots, satisfying $t_{\min} = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = t_{\max}$. The second-order BsBF is the most commonly-used basis function, which can be defined as

$$\beta_{i,2} = \begin{cases} \dfrac{(x - t_i)^2}{(t_{i+2} - t_i)(t_{i+1} - t_i)} & \text{if } t_i \le x \le t_{i+1} \\ \dfrac{x - t_i}{t_{i+2} - t_i} \cdot \dfrac{t_{i+2} - x}{t_{i+2} - t_{i+1}} + \dfrac{t_{i+3} - x}{t_{i+3} - t_{i+1}} \cdot \dfrac{x - t_{i+1}}{t_{i+2} - t_{i+1}} & \text{if } t_{i+1} \le x \le t_{i+2} \\ \dfrac{(t_{i+3} - x)^2}{(t_{i+3} - t_{i+1})(t_{i+3} - t_{i+2})} & \text{if } t_{i+2} \le x \le t_{i+3} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

As can be seen from (3), four extended nodes, $t_{-2}, t_{-1},, t_{n+1}, t_{n+2}$ are needed to make a meaningful computation on the end points. For convenience, it usually takes $t_{-2} = t_{-1} = t_0$ and $t_{n+1} = t_{n+2} = t_n$.

Since $f_X(x|\mathbf{a})$ is a PDF, two constraints must be satisfied. One is that all the coefficients in $\mathbf{a}$ must be nonnegative to make sure the estimated PDF being always positive, and the integral of $f_X(x|\mathbf{a})$ over $(-\infty, +\infty)$ must be one, namely,

$$\int_{-\infty}^{+\infty} f_X(x|\mathbf{a})dx = \int_{-\infty}^{+\infty} \sum_{i=1}^{\mathcal{M}} a_i \beta_{i,k}(x)dx = \sum_{i=1}^{\mathcal{M}} a_i \frac{t_{i+3} - t_i}{3} = 1 \tag{4}$$

2) MLE-BASED B-SPLINE PDF MODEL PARAMETER ESTIMATION

Suppose a random sample $\mathbf{X}_s = \left(x_1, x_2, \cdots, x_{n_s}\right)^T$ is sampling from an unknown probability model $f(x)$, the key step of the BsBFs-based PDFE is to determine the LWCV $\mathbf{a}$ based on the observed sample $\mathbf{X}_s$. As reported in [33], the cross entropy can be used to evaluate the PDFE results, given by

$$J(g, f) = -\int f(x) \log g(x)dx \tag{5}$$

where $f(x)$ is the true distribution model and $g(x)$ is the candidate distribution model. To facilitate application, the measurable entropy (*ME*) can be used to make an asymptotically unbiased estimation of $J(g, f)$ [33], given by

$$ME = -\int f_X(x|\mathbf{a}) \log f_X(x|\mathbf{a})\, dx + \frac{3n_f}{2n_s} \tag{6}$$

where n_f is the number of the free parameter in BsBFs-based PDF model in (1). The model complexity $\mathcal{M}$ and LWCV $\mathbf{a}$ can be obtained by minimizing *ME*.

In this paper, since the number of the BsBFs is fixed, the estimation of the PDF model parameter degrades into a maximum likelihood estimation (MLE) method. Namely, we can solve the following optimization problem to obtain $\mathbf{a}$,

$$\begin{aligned} \hat{\mathbf{a}} &= \underset{\mathbf{a}}{\operatorname{argmax}} \left\{ \log \left[\prod_{i=1}^{n_s} f_X(x_i|\mathbf{a}) \right] \right\} = \underset{a}{\operatorname{argmax}} \sum_{i=1}^{n_s} \log f_X(x_i|\mathbf{a}) \\ s.t. &\sum_{i=1}^{\mathcal{M}} a_i \frac{t_{i+3} - t_i}{3} = 1 \\ &a_i \geq 0 \quad (i = 1, 2, \cdots, \mathcal{M}). \end{aligned} \tag{7}$$

Many optimization methods can be used to solve the abovementioned optimization problem. As derived in [33], a fast-convergent iteration scheme was given by

$$\hat{a}_j^{(n+1)} = \frac{1}{\boldsymbol{n}_s c_j} \sum_{i=1}^{n_s} \frac{\hat{a}_j^{(n)} \beta_j(x_i)}{f_X\left(x_i|\hat{\mathbf{a}}^{(n)}\right)} \tag{8}$$

where $c_j = \left(t_{j+3} - t_j\right)/3$, $\hat{a}_j^{(n)}$ denotes the n-step iteration results of the j-th LWC. Given an initial estimation of the LWCV $\hat{\mathbf{a}}^{(0)}$, it reported that it can achieve a stable estimation of LWCV $\hat{\mathbf{a}}$ within a small number of iteration steps.

3) PRIOR SMOOTH CONSTRAINT(PSC)-BASED PARAMETER REVISING

The MLE-based parameter estimation result generally exists a certain error, resulting from the overemphasizing the complexity or overfitting of the B-spline model.

Suppose the estimation deviation of each coefficient a_i is w_i, the MLE-based PDFE result $\hat{\mathbf{a}}$ can be expressed as,

$$\hat{\mathbf{a}} = \mathbf{a} + \mathbf{w} \tag{9}$$

where $\mathbf{a}$ is the true LWCV and $\mathbf{w}$ is the error vector. Based on the central limit theorem, $\mathbf{w}$ should follow a normal distribution with zero-mean and a common covariance $\sigma^2\mathbf{I}_{\mathcal{M}\times\mathcal{M}}$, ($\mathbf{I}_{\mathcal{M}\times\mathcal{M}}$ is a $\mathcal{M}\times\mathcal{M}$ identity matrix [31]). Hence, conditioned on the true LWCV $\mathbf{a}$, the likelihood of $\hat{\mathbf{a}}$ can be computed as

$$\begin{aligned} p\left(\hat{\mathbf{a}}|\mathbf{a}\right) &= \frac{1}{(2\pi)^{\mathcal{M}/2}} \frac{1}{\left|\sigma^2\mathbf{I}_{\mathcal{M}\times\mathcal{M}}\right|} \\ &\quad \times \exp\left\{-\frac{1}{2}\left(\hat{\mathbf{a}} - \mathbf{a}\right)^T \left(\sigma^2\mathbf{I}_{\mathcal{M}\times\mathcal{M}}\right)^{-1} \left(\hat{\mathbf{a}} - \mathbf{a}\right)\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\mathcal{M}/2} \exp\left\{-\frac{1}{2\sigma^2}\left\|\hat{\mathbf{a}} - \mathbf{a}\right\|_2^2\right\} \end{aligned} \tag{10}$$

where $\|\cdot\|_2^2$ means the square value of 2-norm. As stated in [31], since $\hat{\mathbf{a}}$ is estimated directly from the samples, $p\left(\hat{\mathbf{a}}|\mathbf{a}\right)$ contains all the necessary information of the sampling data related to the real distribution.

One of the underlying and useful hypotheses to govern the statistical modeling is the model *smoothness*, which can be defined by introducing the approximated first derivatives on both the left and right side of a checked point on the true model parameter $\mathbf{a}$ in mathematics, namely, $\frac{\partial \mathbf{a}}{\partial i}\big|_+ = \frac{\partial \mathbf{a}}{\partial i}\big|_-$, where the subscript "+" and "−" represent the left and the right derivatives, respectively. In the discrete form, it can be expressed as $a_{i+1} - a_i = a_i - a_{i-1} + e_i$, where e_i denotes the smoothness residual error of the i-th coefficient in $\mathbf{a}$, alternatively, $e_i = a_{i+1} + a_{i-1} - 2a_i$ and $i = \{2, 3, \ldots, \mathcal{M}-1\}$. Based on the smoothness assumption, elements in the smoothness residual error $\mathbf{e}$ should obey an independent and identical Gaussian distribution with zero mean and the same variance ρ^2, namely,

$$\begin{aligned} p(\mathbf{e}) &= p(e_2, e_3, \cdots, e_{\mathcal{M}-1}) \\ &= \left(\frac{1}{2\pi}\right)^{(\mathcal{M}-2)/2} \frac{\exp\left\{-\frac{1}{2}\mathbf{e}^T\left(\rho^2\mathbf{I}_{(\mathcal{M}-2)\times(\mathcal{M}-2)}\right)^{-1}\mathbf{e}\right\}}{\left|\rho^2\mathbf{I}_{(\mathcal{M}-2)\times(\mathcal{M}-2)}\right|} \\ &= \left(\frac{1}{2\pi\rho^2}\right)^{(\mathcal{M}-2)/2} \exp\left\{-\frac{1}{2\rho^2}\mathbf{e}^T\mathbf{e}\right\} \end{aligned} \tag{11}$$

By introducing a transform matrix, we can rewrite $\mathbf{e} = \mathbf{D}\mathbf{a}$, where $\mathbf{D}$ is

$$\mathbf{D} = \begin{bmatrix} 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & \vdots & \\ & & 1 & -2 & 1 & & \\ & & & & \cdots & & \\ & & \cdots & & & 1 & -2 & 1 \end{bmatrix}_{(\mathcal{M}-2)\times\mathcal{M}} \tag{12}$$

Then substitute $\mathbf{e} = \mathbf{Da}$ back into (11), we can obtain a prior of the true LWCs $\mathbf{a}$ in (13), which is a prior constraint to restrict the fluctuation of the B-spline based statistical model.

$$p(\mathbf{a}|\sigma,\rho) = \left(\frac{1}{2\pi\rho^2}\right)^{(\mathcal{M}-2)/2} \exp\left\{-\frac{1}{2\rho^2}(\mathbf{Da})^T(\mathbf{Da})\right\} = \left(\frac{1}{2\pi\rho^2}\right)^{(\mathcal{M}-2)/2} \exp\left\{-\frac{1}{2\rho^2}\|\mathbf{Da}\|_2^2\right\} \quad (13)$$

In Zong's study [31], it introduced a variance ratio parameter $\omega^2 = \sigma^2/\rho^2$ representing the smooth factor, then the prior distribution $p\left(\mathbf{a}|\omega^2\right)$ can be represented as

$$p\left(\mathbf{a}|\omega^2\right) = \omega^{(\mathcal{M}-2)}\left(\frac{1}{2\pi\sigma^2}\right)^{(\mathcal{M}-2)} \exp\left\{-\frac{\omega^2}{2\sigma^2}\|\mathbf{Da}\|_2^2\right\} \quad (14)$$

To obtain the true LWCV $\mathbf{a}$, we can use the Bayes' theorem to obtain the posterior distribution $p\left(\mathbf{a}|\hat{\mathbf{a}}\right)$ as

$$p\left(\mathbf{a}|\hat{\mathbf{a}}\right) \propto p\left(\mathbf{a}|\omega^2\right)p\left(\hat{\mathbf{a}}|\mathbf{a}\right) = \omega^{\mathcal{M}-2}\left(\frac{1}{2\pi\sigma^2}\right)^{2\mathcal{M}-2} \times \exp\left\{-\frac{1}{2\sigma^2}\left[\left\|\hat{\mathbf{a}}-\mathbf{a}\right\|_2^2 + \omega^2\|\mathbf{Da}\|_2^2\right]\right\} \quad (15)$$

Then we can achieve the point estimation of $\mathbf{a}$ by maximizing the posterior $p\left(\mathbf{a}|\hat{\mathbf{a}}\right)$. Namely, the PSC-based model parameter estimation $\mathbf{a}_{Bayes}$ should be revised as

$$\mathbf{a}_{\text{Bayes}} = \arg\max\left\{-\left\|\hat{\mathbf{a}}-\mathbf{a}\right\|_2^2 - \omega^2\|\mathbf{Da}\|_2^2\right\} = \left(\mathbf{I}_{\mathcal{M}\times\mathcal{M}} + \omega^2\mathbf{D}^{\mathrm{T}}\mathbf{D}\right)^{-1}\hat{\mathbf{a}} \quad (16)$$

Figure 1 displays an artificial image with its B-spline-based statistical modeling results. As can be seen from Fig.1, the fluctuation of the direct MLE-based PDFE result can be alleviated effectively by introducing a prior smooth constraint on the LWCs.

Apparently, the smooth factor ω^2 will greatly affect the PDFE result. We can obtain the best PDF model by a more Bayesian treatment, namely, obtaining the best smooth factor ω^2 based on the samples. As reported in [31], the best smooth factor can be estimated by the Type II maximum likelihood Principle with a new measured factor, named Bayesian Measured Entropy (MEB), given by

$$\text{MEB}\left(\omega^2\right) = -(\mathcal{M}-2)\log\omega^2 + (n_s-2)\log q^2 + \log\left|F^TF\right| + \text{constant} \quad (17)$$

The optimal smooth parameter ω^2 can be obtain by minimizing MEB $\left(\omega^2\right)$, where $q^2 = -\left\|\hat{\mathbf{a}}-\mathbf{a}\right\|_2^2 - \omega^2\|\mathbf{Da}\|_2^2$, and $F = \begin{bmatrix}\mathbf{I}_{\mathcal{M}\times\mathcal{M}} \\ \omega\mathbf{D}_{(\mathcal{M}-2)\times\mathcal{M}}\end{bmatrix}_{(2\mathcal{M}-2)\times\mathcal{M}}$.

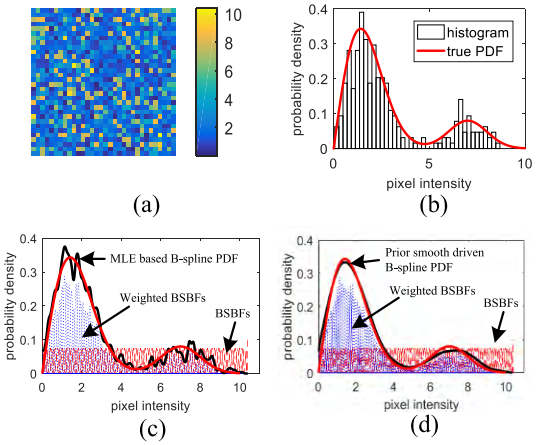


FIGURE 1. BSBFs-based one-dimensional PDFE. (a) An artificial image with pixel intensities randomly sampling from a known mixed probability model, which is $p_{\mathbf{X}}(x) = g(x)/K$, where $g(x) = 0.8 \times \left[\frac{x}{2}\exp\left\{-\left(\frac{x}{2}\right)^2\right\}\right] + 0.2 \times \left[\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x-7)^2}{2}\right\}\right]$ and $K = \int_0^{10} g(x)dx \approx 1$,is a normalized parameter. (b) The true PDF model with its histogram distribution; (c) The MLE- based B-spline statistical modeling results and (d) Bayesian-driven B-spline statistical modeling result with a prior smooth constraint on LWCs, where the smooth factor $\omega^2 = 90$.

B. BSBFS-BASED TWO-DIMENSIONAL PDF MODELING

1) TWO-DIMENSIONAL B-SPLINE MODEL DESCRIPTION

The two-dimensional PDF can be approximated as [33]

$$f_{XY}(x,y|\mathbf{a}) = \sum_{i=1}^{\mathcal{M}}\sum_{j=1}^{\mathcal{N}} a_{i,j}\beta_{i,k}(x)\beta_{j,k}(y) \quad (18)$$

where $\mathbf{a} = \left(a_{1,1}, a_{1,2}, a_{1,3}, \cdots, a_{\mathcal{M},\mathcal{N}}\right)^T$, $\beta_{i,k}(x)$ is the k-order BsBF. Analogizing to the one-dimensional PDF model, two constraints must be considered to make a reasonable PDF, the nonnegative LWCs and the integral of n $f_{XY}(x,y|\mathbf{a})$ must be one, namely,

$$\int_{-\infty}^{+\infty} f_{XY}(x|\mathbf{a})dx = \int_{-\infty}^{+\infty}\sum_{i=1}^{\mathcal{M}}\sum_{j=1}^{\mathcal{M}} a_{i,j}\beta_{i,2}(x)\beta_{j,2}(y)dxdy = \sum_{i=1}^{\mathcal{M}}\sum_{j=1}^{\mathcal{N}} a_{i,j}\frac{x_i - x_{i-3}}{3}\times\frac{y_j - y_{j-3}}{3} = 1 \quad (19)$$

Given the random sampling pair $\mathrm{X}_s = \left((x_1,y_1)^T, (x_2,y_2)^T, \cdots, \left(x_{n_s},y_{n_s}\right)^T\right)^T$, LWCV $\mathbf{a}$ can be estimated by solving the following optimization problem,

$$\hat{\mathbf{a}} = \underset{\mathbf{a}}{\text{argmax}}\left\{\log\left[\prod_{i=1}^{n_s} f_{\mathrm{XY}}\left(x_i,y_j|\mathbf{a}\right)\right]\right\} = \underset{\mathbf{a}}{\text{argmax}}\sum_{i=1}^{n_s}\log f_{\mathrm{XY}}\left(x_i,y_j|\mathbf{a}\right)$$

$$s.t.\ \sum_{i=1}^{\mathcal{M}}\sum_{j=1}^{\mathcal{N}} a_{i,j}\frac{x_i - x_{i-3}}{3}\times\frac{y_j - y_{j-3}}{3} = 1$$
$$a_{i,j} \geq 0 \quad i=1,2,\cdots,\mathcal{M};\ j=1,2,\cdots,\mathcal{N} \tag{20}$$

Zong and Lam [30], [33] has derived an iteration formula to solve the abovementioned optimization problem, namely,

$$\hat{a}_{ij}^{(n+1)} = \frac{1}{\boldsymbol{n}_s c_{i,j}}\sum_{l=1}^{n_s}\frac{a_{i,j}^{(n)}\beta_{i,2}(x_l)\,\beta_{j,2}(y_l)}{f_{\mathrm{XY}}\left(x_l, y_l|\hat{\mathbf{a}}^{(n)}\right)} \tag{21}$$

where $c_{i,j} = \frac{x_i - x_{i-4}}{3}\times\frac{y_j - y_{j-4}}{3}$.

2) PSC-BASED PARAMETER REVISING

Suppose the estimation deviation vector is $\mathbf{w}$, the MLE-based PDF model parameter estimation result $\hat{\mathbf{a}}$ can be expressed as

$$\hat{\mathbf{a}} = \mathbf{a} + \mathbf{w} \tag{22}$$

Like the one-dimensional PDFE, the estimation error w_{ij} should follows an independent identically Gaussian distribution model with zero mean and a common σ^2 variance. Hence, the likelihood function of $\hat{\mathbf{a}}$ is given by

$$p\left(\hat{\mathbf{a}}|\mathbf{a}\right) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\mathcal{MN}}\exp\left\{-\frac{1}{2\sigma^2}\left\|\hat{\mathbf{a}} - \mathbf{a}\right\|_2^2\right\} \tag{23}$$

According to the Bayes's theorem, the posterior probability $p\left(\mathbf{a}|\hat{\mathbf{a}}\right) \propto p(\mathbf{a})\,p\left(\hat{\mathbf{a}}|\mathbf{a}\right)$, where the prior distribution term $p(\mathbf{a})$ is still unknown, which is a prior used to govern the smoothness of the statistical model.

According to the Zong's definition [30], the smoothness of the statistical model can be approximated by the second-order difference of $\mathbf{a}$, defined as

$$e_{i,j} = a_{i,j-1} + a_{i-1,j} - 4a_{i,j} + a_{i,j+1} + a_{i+1,j} \tag{24}$$

while on the four boundaries, the parameter differences $e_{1,j}, e_{i,1}, e_{\mathcal{M},j}, e_{i,\mathcal{N}}$ are defined with a little change, which are defined as

$$\begin{aligned}
e_{i,1} &= a_{i-1,1} + a_{i+1,1} - 2a_{i,1}; \quad i = 2,\cdots,\mathcal{M}-1\\
e_{1,j} &= a_{1,j-1} + a_{1,j+1} - 2a_{1,j}; \quad j = 2,\cdots,\mathcal{N}-1\\
e_{i,N} &= a_{i-1,1} + a_{i+1,1} - 2a_{i,1}; \quad i = 2,\cdots,\mathcal{M}-1\\
e_{\mathcal{M},j} &= a_{\mathcal{M},j-1} + a_{\mathcal{M},j+1} - 2a_{\mathcal{M},j}; \quad j = 2,\cdots,\mathcal{N}-1
\end{aligned} \tag{25}$$

where $\mathbf{e} = \left(e_{1,1}, e_{1,2},\cdots,e_{\mathcal{MN}}\right)$. As reported in [30], based on the smooth hypothesis, $e_{i,j}$ should follow an independent and identical Gaussian distribution with zero mean and a common ρ^2 variance, namely,

$$\begin{aligned}
p(\mathbf{e}) &= p\left(e_{1,2}, e_{1,3},\cdots,e_{1,\mathcal{N}-1}, e_{2,1},\cdots,e_{\mathcal{M}-1,\mathcal{N}}\right)\\
&= \prod_{i=1}^{\mathcal{M}-1}\prod_{j=2}^{\mathcal{N}}\frac{1}{\sqrt{2\pi}\rho}\exp\left\{-\frac{e_{i,j}^2}{2\rho^2}\right\}\\
&= \left(\frac{1}{\sqrt{2\pi}\rho}\right)^{\mathcal{MN}-4}\exp\left\{-\frac{1}{2\rho^2}\sum_{i=1}^{\mathcal{M}-1}\sum_{j=2}^{\mathcal{N}}e_{i,j}^2\right\}\\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^{\mathcal{MN}-4}\frac{1}{\left|\rho^2\mathbf{I}_{\mathcal{MN}-4}\right|^{1/2}}\\
&\quad\times\exp\left\{-\frac{1}{2}\mathbf{e}^T\left(\rho^2\mathbf{I}_{\mathcal{MN}-4}\right)^{-1}\mathbf{e}\right\}
\end{aligned} \tag{26}$$

By introducing a new transform matrix $\mathbf{D}_2$, we can obtain a new expression $\mathbf{e} = \mathbf{D}_2\mathbf{a}$ and substitute it back into (26) we can obtain a prior model of $\mathbf{a}$, given as

$$p(\mathbf{a}|\rho) = \left(\frac{1}{\sqrt{2\pi}\,\rho}\right)^{\mathcal{MN}-4}\exp\left\{-\frac{1}{2\rho^2}\left\|\mathbf{D}_2\mathbf{a}\right\|_2^2\right\} \tag{27}$$

where

$$\mathbf{D}_2 = \begin{bmatrix}
\mathbf{D} & & & & \\
\mathbf{I}_{(\mathcal{M}-2)\times\mathcal{M}} & \mathbf{D}_1 & \mathbf{I}_{(\mathcal{M}-2)\times\mathcal{M}} & & \vdots \\
 & \mathbf{I}_{(\mathcal{M}-2)\times\mathcal{M}} & \mathbf{D}_1 & \mathbf{I}_{(\mathcal{M}-2)\times\mathcal{M}} & \\
 & & \cdots & & \cdots \\
 & & \cdots & & \mathbf{D}
\end{bmatrix}_{(\mathcal{MN}(4)\times\mathcal{MN}}$$ (28)

$\mathbf{D}$ is defined in (12), and

$$\mathbf{D}_1 = \begin{bmatrix}
-2 & & & & \\
1 & -4 & 1 & & \vdots \\
 & 1 & -4 & 1 & \\
 & & \cdots & & \cdots \\
 & & \cdots & & 2
\end{bmatrix}_{(\mathcal{M}-2)\times\mathcal{M}}$$ (29)

If we import another new parameter $\omega^2 = \sigma^2/\rho^2$, representing the smooth factor, the prior distribution of the true model parameter $\mathbf{a}$ can be expressed as

$$p(\mathbf{a}|\omega) = \omega^{\mathcal{MN}-4}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\mathcal{MN}-4}\exp\left\{-\frac{\omega^2}{2\sigma^2}\left\|\mathbf{Da}\right\|_2^2\right\} \tag{30}$$

Hence, based on the Bayes' theorem, the prior smooth-driven BsBFs-based statistical model parameter can be estimated by maximizing the following posterior distribution,

$$\begin{aligned}
p\left(\mathbf{a}|\hat{\mathbf{a}},\omega\right) &\propto p(\mathbf{a}|\omega)\,p\left(\hat{\mathbf{a}}|\mathbf{a}\right)\\
&\propto \omega^{\mathcal{MN}-4}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\mathcal{MN}-4}\exp\left\{-\frac{\omega^2}{2\sigma^2}\left\|\mathbf{Da}\right\|_2^2\right\}\\
&\quad\times\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\mathcal{MN}}\exp\left\{-\frac{1}{2\sigma^2}\left\|\hat{\mathbf{a}} - \mathbf{a}\right\|_2^2\right\}\\
&= \omega^{\mathcal{MN}-4}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{2\mathcal{MN}-4}\\
&\quad\times\exp\left\{-\frac{1}{2\sigma^2}\left[\omega^2\left\|\mathbf{Da}\right\|_2^2 + \left\|\hat{\mathbf{a}} - \mathbf{a}\right\|_2^2\right]\right\}
\end{aligned} \tag{31}$$

Given a specific ω^2, we can achieve the point estimation of the true model parameter $\mathbf{a}$ by maximizing $p\left(\mathbf{a}|\hat{\mathbf{a}},\omega\right)$

$$\mathbf{a}_{\mathrm{Bayes}} = \left(\mathbf{I}_{\mathcal{MN}\times\mathcal{MN}} + \omega^2\mathbf{D}^{\mathrm{T}}\mathbf{D}\right)^{-1}\hat{\mathbf{a}} \tag{32}$$

Furthermore, we can perform a more Bayesian treatment to obtain the best smooth factor ω^2 by the Type II maximum likelihood Principle as reported in [30].

Figure 2 displays the scatter plot of 3600 point-pairs, sampling from a mixture distribution model. The PDFE results indicate that more stable and accurate statistical model will be achieved by introducing a prior smooth constraint over the model parameter.

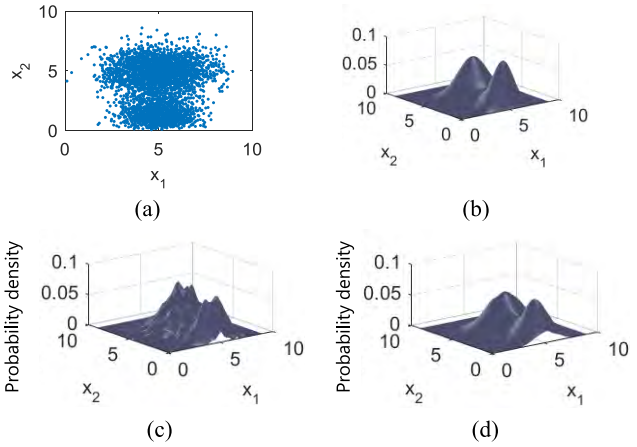


FIGURE 2. BsBFs-based two-dimensional PDF modeling. (a) Scatter plot of 3600 pair-points, sampling from a mixture distribution, whose PDF is $p_{XY}(x,y) = g(x,y)/K$, where $g(x,y) = 0.3 \times \mathcal{N}\left(x|u_0, (\beta y)^{-1}\right)\Gamma(y|a,b) + 0.7 \times \mathcal{N}(x|u,\Sigma)$, where $\mathcal{N}(\cdot)$ and $\Gamma(\cdot)$ denote the normal distribution and Gamma distribution, respectively, and $x = (x,y)^T$, $\mu_0 = 5$, $\beta = a = b = 2$, $u = (5,5)^T$, $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ and $K = \int_0^9 g(x,y)dxdy \approx 1$ is a normalized parameter. (b) The true PDF of the random sampling pair-points; (c) The MLE-based B-Spline PDFE results and (d) Bayesian driven B-Spline PDFE result with the smooth factor $\omega^2 = 12$.

III. PROPOSED TCvBsISM

The proposed TCvBsISM consists of three stages, offline B-spline LWCs-based image statistical feature dictionary learning (ISFDL) phase, frequency histogram-based texture pattern representation training (FH-TPRT) phase and online TC (OTC) phase, displayed in Fig.3. The BsBFs-based image statistical modeling is performed on the original pixel space as well as on the multi-channel transform space, filtered with the orientation and spatial-frequency selective linear filter banks.

A. ORIENTATION AND SPATIAL-FREQUENCY SELECTIVE LINEAR FILTERING

Three widely-used filter banks, the steerable isotropic Gaussian derivative filter (SIGDF) [34] bank, orientation anisotropic Gaussian derivative filter (OAGDF) [35] bank and Gabor wavelet filter bank (GWFB) [21] are considered for MS&MO ITSA feature representation.

1) SIGDF

SIGDFs are implemented by introducing a steerable operation on the isotropic Gaussian derivative filters (IGDFs) with low computation complexity. Suppose a k-order GDF template is

$$G_{k,\sigma}(x,y) = \sum_{m=1}^{k}\sum_{i=0}^{m} K_{m,i}\frac{\partial^i}{\partial x^i}\frac{\partial^{m-i}}{\partial x^{m-i}}G_\sigma(x,y) \tag{33}$$

where $G_\sigma(x,y)$ is the Gaussian function. As stated by Freeman and Adelson [36], the Gaussian derivative filter (GDF) template $G_{k,\sigma}(x,y)$ is steerable, thus the filtering response of image I at a special orientation θ can be expressed by the following linear combination formula [20],

$$I(\mathbf{X}) * G_{k,\sigma}(\mathbf{R}_\theta\mathbf{X}) = \sum_{m=1}^{k}\sum_{i=0}^{m}\alpha_{m,i}I_{m,i}(\mathbf{X}) \tag{34}$$

where $\mathbf{X} = (x,y)^T$, $\mathbf{R}_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ is the rotation transform matrix, $I_{m,i}(\mathbf{X})$ is the filtering response of image I with the mixed GDF $G_{m,i,\sigma}$ with the derivative orders of i and m-i with respect to the variable x and y respectively.

If we obtain the linear weighting coefficient $\alpha_{m,i}$, we can achieve the filtering response at any orientation at a low computational cost. As reported in the literature [20], [37], $\alpha_{m,i}$ is a trigonometric polynomial function of θ given by,

$$\alpha_{m,i} = \sum_{j=0}^{m} K_{m,j}\sum_{t=0}^{j}\sum_{l=0}^{m-j}(-1)^l C_j^t C_{m-j}^l (\cos\theta)^{t+m-j-l}(\sin\theta)^{j-t+l} \tag{35}$$

2) OAGDF

The commonly-used two-variable elongated Gaussian kernel function (eGKF) is given by [38]

$$EG_{\sigma,\rho,\theta}(\mathbf{x}) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{R}_{-\theta}\begin{bmatrix}\rho^2 & 0 \\ 0 & \rho^{-2}\end{bmatrix}\mathbf{R}_\theta\mathbf{x}\right\} \tag{36}$$

where ρ is the anisotropic factor, satisfying $\rho \geq 1$.

The anisotropic GDF (AGDF) is the first-order partial derivative of an eGKF with respect to the first variable x, and OAGDF is the rotating expression of the AGDF, which reflects the structure variation of a texture image along the direction θ,

$$\begin{aligned}\nabla EG_{\sigma,\rho,\theta}(\mathbf{x}) &= \frac{\partial EG_{\sigma,\rho,\theta}(\mathbf{x})}{\partial\theta} \\ &= -\frac{\rho^2[\cos\theta, \sin\theta]\,\mathbf{x}}{\sigma^2}EG_{\sigma,\rho,\theta}(\mathbf{x})\end{aligned} \tag{37}$$

It is worth noting that the OAGDF defined in (37) is not steerable, hence the computation of OAGDF in a specific direction θ cannot be achieved by the linear weighting summation of some fixed eGKF bases in analogous to the generation of SIGDF. However, it reported that eGKF can be decomposed into two Gaussian linear filters of nonorthogonal directions [35]. Thus, the recursive algorithm or common convolution operation can be applied to achieve the filter response of OAGDF with low computation complexity, as reported in [35].

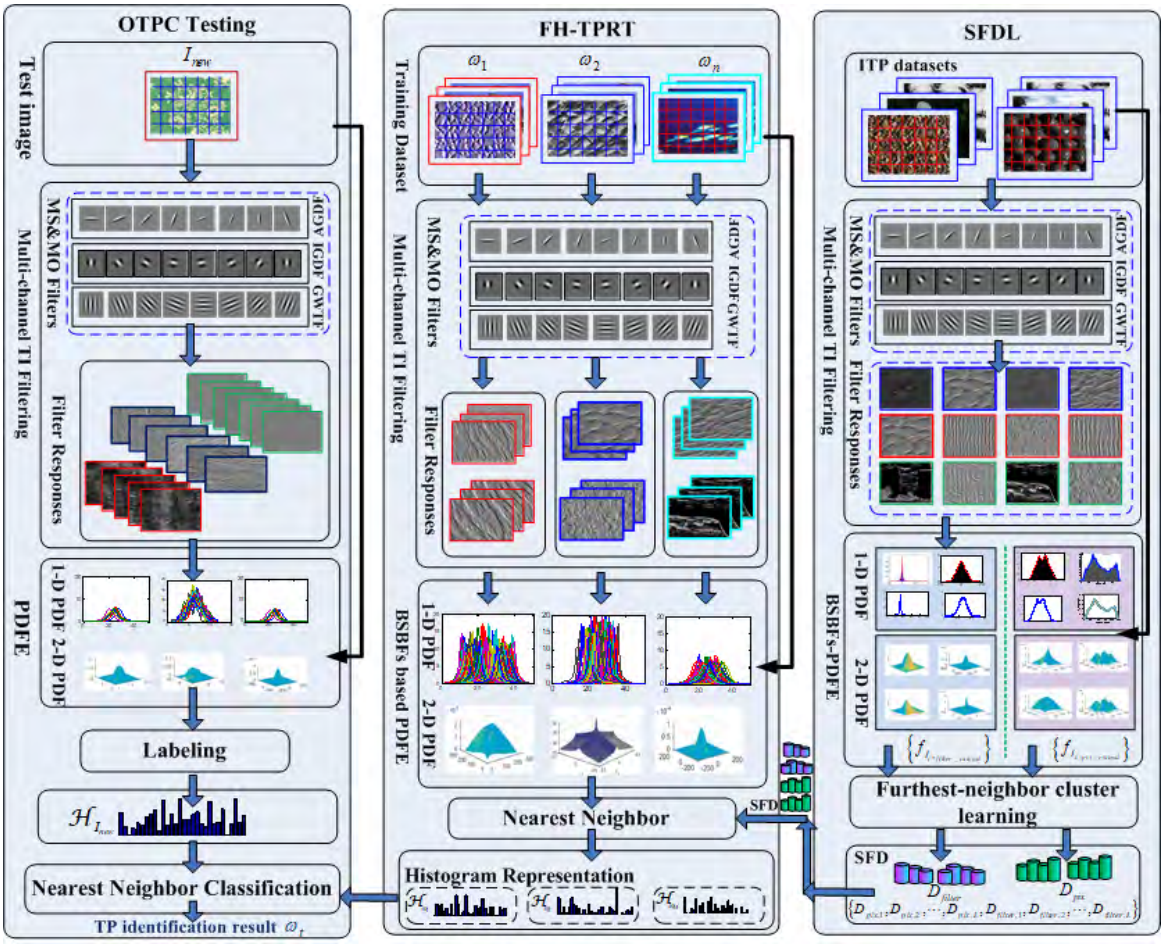


FIGURE 3. Overview of the proposed TCvBsISM.

3) GWFB

A commonly-used GWFB is computed as follows [39],

$$G_{u,v}(x,y) = \frac{K\omega_0}{\sqrt{\pi}d\kappa} e^{-\frac{\left(\frac{\omega_0}{f^u}\right)^2}{2d^2\kappa^2}\left(d^2(x)^2_{\theta_v}+(y)^2_{\theta_v}\right)} \cdot \left[e^{i\frac{\omega_0}{f^u}(x\cos\theta_v+y\sin\theta_v)} - e^{-\frac{\kappa^2}{2}}\right] \quad (38)$$

where $\theta_v = \pi v/8$, ω_0 is the maximum frequency, and f^u is the spacing ratio among Gabor kernels in the frequency, u and v represent the scale factor and orientation factor, respectively. The standard deviation κ reflects the half magnitude spatial frequency bandwidth [40], [41], with $\kappa = 2.4653$ for a frequency bandwidth of 1.5 octave.[1] The scale parameter K is usually set as 1 so that its L^2 norm $< G_{u,v}, G_{u,v} >= 1$. Fig.4 displays the ensemble of the real parts and the imaginary parts of Gabor wavelets with spatial frequency width of 1.5 octave.

[1]Based on the Gabor filter definition in (38), the relation between κ and the half-amplitude bandwidth of the Gabor filter can be expressed as $\kappa = \sqrt{2\log 2}\left(\frac{2^\phi+1}{2^\phi-1}\right)$ [39], [41], where ϕ is the bandwidth in octaves. Hence, for $\phi = 1$ octave, κ should be 2.0393; for $\phi = 1.5$ octave, $\kappa \approx 2.4653$.

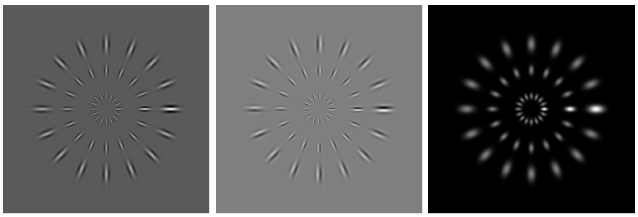


FIGURE 4. Ensemble of real parts (a), imaginary parts (b) and the magnitude (c) of the Gabor Wavelets with spatial frequency width of 1.5 octave. The magnitudes of the Gabor filters are normalized by set the scale parameter $K = 1$. The aspect ratio d of the elliptical Gaussian is 2:1 and the other parameters are set as follows, $\omega_0 = \pi$, $f = \pi/2$, the filters on the three rings represents three different scale, namely, $u = 3, 2, 1$ from outside to inside.

B. STATISTICAL FEATURE EXTRACTION

The statistical features are obtained by the LWCV of the BsBFs-based PDF model. The ultimate statistical feature of an image I is achieve as

$$f_{I_{pix}} = \left(\boldsymbol{\alpha}^T_{1,I_{pix}}, \boldsymbol{\alpha}^T_{2,I_{pix}}\right)^T \quad (39)$$

where $\boldsymbol{\alpha}_{1,I_{pix}}$ and $\boldsymbol{\alpha}_{2,I_{pix}}$ represent the LWCVs of the one-dimensional and two-dimensional BSBFs-based PDF model. Simultaneously, if D filters are considered for the MS&MO characteristics representation, we can obtain D filtering responses of image I, thus the filter banks-based statistical feature can be expressed as

$$f_{I_{\text{filter}}} = \left(\boldsymbol{\alpha}^T_{1,I_{filter1}}, \boldsymbol{\alpha}^T_{1,I_{filter1}}, \cdots, \boldsymbol{\alpha}^T_{1,I_{filterD}}, \boldsymbol{\alpha}^T_{2,I_{filter1}}, \boldsymbol{\alpha}^T_{2,I_{filter2}}, \cdots, \boldsymbol{\alpha}^T_{2,I_{filterD}}\right)^T \quad (40)$$

where $\boldsymbol{\alpha}^T_{1,I_{filteri}}$ and $\boldsymbol{\alpha}^T_{2,I_{filteri}}$ represent the one-dimensional and two-dimensional BsBFs-based LWCVs of the filtering responses based on the i-th filter. Hence, the statistical feature of an image I can be expressed as

$$f_I = \left(f^T_{I_{pix}}, f^T_{I_{\text{filter}}}\right)^T \quad (41)$$

In addition, if we partition the original image I into K non-overlapped sub-images, the extended statistical feature, f_{I_extend}, can be expressed as

$$f_{I\_extend} = \left(f^T_{I_{pix\_extend}}, f^T_{I_{filter\_extend}}\right)^T \quad (42)$$

where $f_{I_{pix\_extend}}$ and $f_{I_{filter\_extend}}$ stand for the extended BsBFs-based statistical features of the raw pixel and the filter bank responses respectively, given by

$$f_{I_{pix\_extend}} = \left(f^{\mathrm{T}}_{I_{pix}}, f^{\mathrm{T}}_{I_{pix,sub_1}}, f^{\mathrm{T}}_{I_{pix,sub_2}}, \cdots, f^{\mathrm{T}}_{I_{pix,sub_K}}\right)^{\mathrm{T}} \quad (43)$$

$$f_{I_{\text{filter\_extend}}} = \left(f^{\mathrm{T}}_{I_{\text{filter}}}, f^{\mathrm{T}}_{I_{\text{filter},sub_1}}, f^{\mathrm{T}}_{I_{\text{filter},sub_2}}, \cdots, f^{\mathrm{T}}_{I_{\text{filter},sub_K}}\right)^{\mathrm{T}} \quad (44)$$

where $f_{I_{pix,sub_i}}$ records the BsBFs-based statistical feature vector of the i^{th} sub-image of the original image, $f_{I_{\text{filter},sub_j}}$ represents the BsBFs-based statistical feature vectors of the j^{th} sub-blocks of the filter bank responses of image I.

C. DETAILS OF TCvBsISM

As can be seen from Fig.3, the proposed TCvBsISM mainly includes three phases, ISFDL, FH-TPRT and OITC phase. Suppose there are M-class of texture patterns in the training database, $\{\omega_i; 1 \le i \le M\}$, the i-the texture pattern consists of S_{ω_i} samples, where the j-th sample is dented as $I_{\omega_i,j}$. The main steps of the proposed TCvBsISM can be expressed as follows.

Stage I: ISFDL. Learn the statistical feature dictionary (SFD), illustrated in the right of the Fig.3.

I-i). For each image I_i in the training image database, use the BsBFs-based PDFE method to obtain the one-dimensional and two-dimensional statistical feature of the raw pixel intensities, $f_{I_{i,pix}} = \left(\boldsymbol{\alpha}^T_{1,I_{i,pix}}, \boldsymbol{\alpha}^T_{2,I_{i,pix}}\right)^T$.

I-ii). Prepare a multichannel filter bank $filter_{bank} = \{filter_{SIGDF}, filter_{OAGDF}, filter_{GWFB}\}$ to obtain the MS&MO filter responses of image I_i, where $filter_{SIGDF}$, $filter_{OAGDF}$ and $filter_{GWFB}$ represents the filter banks of SIGDF, OAGDF and GWFB, respectively.

I-iii). For each filter response $I_{i,filter_j}$, obtain the one-dimensional and two-dimensional statistical features $f_{I_{i,filter_j}}, f_{I_{i,filter_j}} = \left(\boldsymbol{\alpha}^T_{1,I_{i,filter_j}}, \boldsymbol{\alpha}^T_{2,I_{i,filter_j}}\right)^T$.

I-iv). Partition the current image I_i into K non-overlapped sub-images and obtain the extended statistical feature $f_{I_i\_extend}$ of image I_i as expressed in (42). Specifically, if $K = 0$, it means the only the statistical features of the holistic image are considered. In this work, images are partitioned into 32×32 non-overlapping sub-blocks for texture analysis.

I-v). Repeat the same statistical feature extraction steps from I-i to I-iv to obtain the statistical feature vectors $F_{Training}$ of all images in training database, $F_{Training} = \left\{f_{I_{i\_extend}} | i = 1, \cdots, \sum_{i=1}^{M} S_{\omega_i}\right\}$.

I-vi). Obtain the SFD $\mathcal{D}_{\text{Training}}$ based on the statistical feature vectors of the image raw pixel space and the filter bank response space, respectively, by the furthest-neighbor cluster learning approach (the complete linkage agglomerative clustering).

$$\begin{aligned}\mathcal{D}_{\text{Training}} &= \left\{D_{pix}, D_{\text{filter}}\right\} \\ &= \left\{D_{pix,1}, D_{pix,2}, \cdots, D_{pix,L}, D_{filter,1}, D_{filter,2}, \cdots, D_{filter,L}\right\}\end{aligned} \quad (45)$$

where D_{pix} and D_{filter} represent the *textons* learned from the raw pixel space and the filtering responses space, respectively; L denotes the number of entries of D_{pix} or D_{filter}. In the ISFDL, the dissimilarity between two sets is measured by the KL-divergence and the mean value of each cluster is achieved as an item (*texton*) of SFD.

Stage II: FH-TPRT. Obtain the histogram-based texture pattern feature representation based on the frequency occurrence of the learned *textons,* illustrated in the middle of Fig.3.

I-i). For each texture pattern ω_i, obtain the extended statistical feature $f_{I_{\omega_{i,j}\_extend}}$ of each image $I_{\omega_i,j}$ in the data set of the texture pattern ω_i by using the same processing steps from I-i to I-iv and then achieve the statistical feature vectors of all of images in the training database, $F_{\omega_i} = \left\{f_{I_{\omega_i,j\_\text{extend}}} | j = 1, \cdots, S_{\omega_i}\right\}$.

I-ii). Obtain the histogram representation of the texture pattern ω_i, denoted as $\mathcal{H}_{\omega_i}$. We firstly label each of the statistical feature $f_{I_{\omega_i,j\_\text{extend}}}$ with a closest pre-learned *texton* based on the nearest-neighbor method and a normalized $2L$-tuple occurrence frequencies histogram is achieved.

I-iii). Repeat the steps II-i and II-ii to obtain the histogram-based representation of all texture patterns, $\mathcal{H} = \left\{\mathcal{H}_{\omega_1}, \mathcal{H}_{\omega_2}, \cdots, \mathcal{H}_{\omega_M}\right\}$.

Stage III: OTC. Identify the texture pattern of a given texture image, illustrated in the left part of Fig.3.

Given a new testing image I_{new}, obtain its histogram $H_{I_{new}}$ representation with the same steps from II-i to II-ii. Then I_{new} is labelled by the nearest neighbor classification method by the histogram distance measurement, $\omega_t = \arg\min_t \left\{\text{distance}\left(\mathcal{H}_{I_{new}}, \mathcal{H}_{\omega_i}\right) | i = 1, 2, \cdots, M\right\}$, where the χ^2 statistics is used for distance measurement, namely, $\text{distance}\left(H_{I_{new}}, H_{\omega_i}\right) = \chi^2\left(H_{I_{new}}, H_{\omega_i}\right) = \frac{1}{2}\sum_{k=1}^{\mathcal{N}} \frac{\left[H_{I_{new}}(k) - H_{\omega_i}(k)\right]^2}{H_{I_{new}}(k) + H_{\omega_i}(k)}$.

D. RELATIONS TO CLASSIC TEXTONS-BASED TEXTURE CLASSIFICATION METHODS

The proposed TCvBsBFs is a member of the family of *textons*-based TC methods, originally proposed in [42], where the authors opened up the possibility of generating a universal operational definition of *textons* with the supervised clustering learning approach. Then, flourishing variants of the *textons*-based TC approaches can be found in the literature. Amongst these methods, VZ algorithm [18] and their subsequent approaches, PATCH and PATCH-MRF [19], are representative methods. We hereby analyze the difference of the TCvBsISM with VZ algorithm and PATCH approaches.

The VZ algorithm also includes a *textons* learning stage, a histogram-based texture pattern modeling stage and a TC stage. In the *textons* learning stage, training images in each texture pattern are firstly convolved with filter banks, e.g., the rotationally invariant filter bank, then the filter responses of all images in the corresponding texture pattern are concatenated and aggregated for *textons* learning based on the K-means clustering method. Regarding the number of *textons* in the feature dictionary, it reported the experimental results of 10 *textons* per class on the texture databases. Since the entire *textons* space is relative large (about 10*Nt, where Nt means the number of texture class), as reported in the literature [42], these *textons* can be pruned down, e.g., retaining 100 *textons*, to achieve slimmer *texton*s space by merging centers or getting rid of those centers with too few data assigned to them. The texture pattern is then modeled by a discrete histogram of the learned *textons*, and the number of the histogram model is directly proportional to the number of training images in each texture pattern. Finally, a test texture image is also represented by a histogram model, and its texture pattern category is labeled by the distance measurement of the histogram models. Apparently, the discriminant feature regarding the viewpoint, illumination and the small-scale change of the imaged texture surface results from the ability of the filter banks. However, in the later literature [19], it directly learned *textons* on the raw pixel space and use the patch-based representation in VZ algorithm to generate a new method PATCH. The neighboring structure, patch-based thought is also applied in the literature [1], the difference is the latter author attempted to seek *non-adaptive, information-preserving, universal-dimensionality* reduction of texture patches for TC.

Though the dominant role of the filer banks is questioned in the Varma and Zisserman's literature [19], we believed that the filter bank response information is a necessary complementary of the image raw pixel space-related ISTA characteristics, and there is ample evidence demonstrating the strong ability of the orientation and spatial-frequency selective filter banks.

Hence, the first difference is that the proposed method takes full account of both the image raw pixel space information and the filter bank responses information. We introduced three kinds of orientation and spatial-frequency selective filter banks, including SIGDF, obtaining the multiscale and omnidirectional isotropic structural features, OAGDF, used to obtain the anisotropic structures, and Gabor wavelets, which can minimize the uncertainty of signal both in time and frequency domain.

Another difference is that the *textons* in TCvBsISM are neither directly learned from the original pixel space as PATCH nor directly learned from the filter bank responses as VZ algorithm. TCvBsISM learned *textons* based on the probability density models of the texture images. As summarized in the literature [19], the success of the canonical texture research areas, such as synthesis, classification, segmentation, compression, was largely due to a fuller statistical representation, which was the entire distribution profile representation as opposed to some low or middle-high-order moments of the statistical distribution. The proposed TCvBsISM employs the B-splines-based PDFE approach to approximate the marginal and joint distribution models of ITSA, which can make a fuller consideration of the statistical distribution of the raw pixel space and the filter response space, including the marginal distribution, which can be attributed to the global origination of the micro-structures, and the joint statistical model, which can extract the local spatial structured characteristics of ITSA. In addition, B-splines-based PDFE can approximate any complicate distribution model without any prior knowledge of the true distribution model. By introducing a prior smooth constraint on the PDF model, the essential structure feature of texture images can be expressed by the LWCs of the pre-defined BsBFs with enough high preciseness. Hence, the proposed TCvBsISM is not texture pattern-specific and it can make effective texture analysis for any kind of texture pattern without any prior information of the true image distribution model.

The third difference is that the *textons* in TCvBsISM are not learned class by class, which are learned directly on the stack of all training images of all texture patterns. The big advantage of this processing fashion is it can achieve discriminant *textons* for the successive texture pattern modeling and representation. The class-wise learning mode achieves the same number of *textons* for every texture pattern, which is not conducive to histogram-based texture pattern representation. For instance, the isotropic texture pattern intuitively needs less *textons* than the anisotropic texture pattern to characterize its ISTA. Furthermore, the theoretical number of *textons* for best TC is also affected by the dissimilarity of different TPs,

for instance, if all the textures look considerably different, a small number of *textons* can achieve good discriminant performance, on the contrary, if two texture patterns look very similar, small number of *textons* may not have strong distinguishing ability, because the *texton*s learned from similar textures will inevitably exhibit only gentle difference, which lead to a low classification accuracy for the similar representation of texture patterns. The class-wise *textons* learning cannot effectively exhibit the intra difference of each texture pattern, since it treats all kind of texture pattern indiscriminatingly. The proposed TCvBsISM achieves textons by making an integrated consideration of the texture images of all texture patterns, the learned textons will be adaptive to intra-variation and inter-similarity of all kinds of texture patterns, hence, it can achieve more representative models for texture pattern representation and classification.

IV. EXPERIMENTAL EVALUATION

A. DATA SETS AND EXPERIMENT SET-UP

Three commonly-used texture databases, the normalized Brodatz texture database Brodatz [43], Columbia-Utrecht Reflectance and texture database CURet [44], UIUC [45], a 8-class outdoor natural scene categories data set (8-CONS) [46] were used to validate the effectiveness of the proposed TCvBsBFs method.

The **small-scale Brodatz subset $\mathcal{B}_{\mathrm{Sc}}^{24}$ (24 classes)** consists of 24 homogeneous texture patterns from the normalized Brodatz texture (NBT) database [43]. Images in $\mathcal{B}_{\mathrm{Sc}}^{24}$ were treated as reported in literature [1].

The **large-scale Brodatz data set $\mathcal{B}_{Ls}^{108}$ (108 classes)** contains almost all the texture patterns in the NBT database, except the texture patterns D14, D55, D78 and D105, since the visual appearance of the texture pairs D14 and D16, D53 and D55, D78 and D79, D105 and D106 are nearly the same.

$\mathcal{B}_{Ls}^{108}$ is a challenging testing data set, because (1) There are a large number of texture patterns but only a small number of training samples. (2) Some different texture patterns visually belong to the same visual scene captured with different illumination, imaging distance or viewing angles, e.g., D6 and D14, D23 and D27, D25 and D26, D79 and D78. (3) Some texture patterns are prone to incur visual ambiguities, e.g., D40 and D42, D103 and D104, etc. Some confusing texture patterns can be seen in Fig.5.

The **CURet subset C_{Sub}^{20} (20 classes)** includes 20 classes of different texture patterns, where texture images are captured from the real-world surface with different capturing scales, surface height variations and orientations. The 20-class texture patterns are displayed in Fig.6.

The **UIUC data set $\mathcal{T}_{D}^{25}$ (25 classes)** includes 25 texture patterns. The database includes surfaces whose texture is due mainly to albedo variations (e.g., wood and marble), 3D shape, as well as a mixture of both(e.g., brick and carpet) [45]. In each texture pattern, images have significant variations on scale differences and viewpoint as well as the uncontrolled illumination variation. Some texture images can be seen from Fig.7.

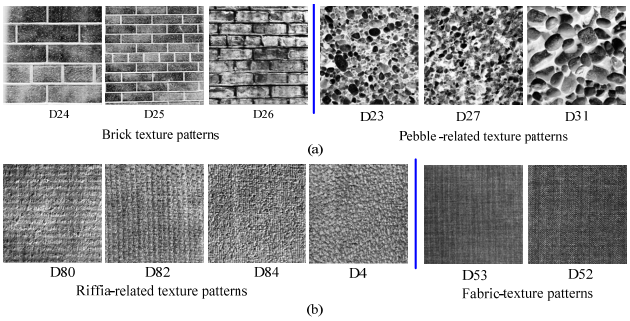


FIGURE 5. Some confusing texture patterns in $\mathcal{B}_{Ls}^{108}$. (a) Texture images seem to be captured from the same scene; (b) Image captured from different texture patterns with visual ambiguities.

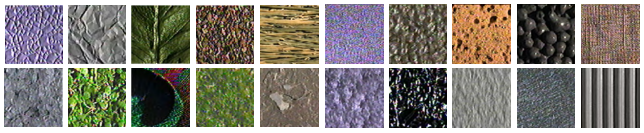


FIGURE 6. 20-class CURet subset C_{Sub}^{20}.

The **natural environmental scene database $\mathcal{N}_{out}^{8}$ (8 classes)** consists of 8 natural outdoor categories scene patterns [46], coast, mountain, forest, open country, street, inside city, tall buildings and highways. There are 2688 color images, each of the image is 256×256 pixels. Typical image samples can be seen in Fig.8.

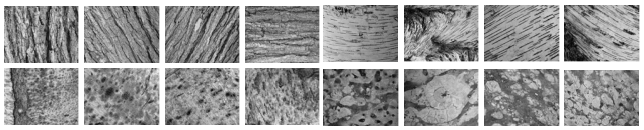


FIGURE 7. UIUC data set $\mathcal{T}_{D}^{25}$.

FIGURE 8. Images in the data set $\mathcal{N}_{out}^{8}$.

Table 1 summarizes the related information of these data sets used in the experiments. Only the gray value of the texture images is considered in the TC experiments. In each independent experiment, we randomly select a fixed number of images as the training samples, the remaining samples as the test samples.

B. EXPERIMENTAL VALIDATION

Three kinds of key parameters affect the classification performance of the proposed TCvBsISM approach, the *textons* number (L), the number of BSBFs for the one-dimensional (M) and two-dimensional (M, N) PDFE, and the number of

TABLE 1. Summary of the data sets used in the experiments.

Data set	viewpoint change	scale difference	illumination variation	classes	sample size	samples per class	Training	Test
$\mathcal{B}_{Sc}^{24}$			√	24	128×128	25	13	12
$\mathcal{B}_{Ls}^{108}$		√	√	112	128×128	25	13	12
$\mathcal{C}_{Sub}^{20}$	√	√	√	61	128×128	90	45	45
$\mathcal{T}_{D}^{25}$	√	√	√	25	640×480	40	20	20
$\mathcal{N}_{out}^{8}$	√	√	√	8	256×256	details displayed in Tab.II		

TABLE 2. Details of data set $\mathcal{N}_{out}^{8}$.

data set	sample size	Samples(Training/Test)							
		Coast	Forest	Highway	Inside city	Mountain	Open country	Street	Tall building
$\mathcal{N}_{out}^{8}$	256×256	180/180	164/164	130/130	154/154	187/187	205/205	146/146	178/178

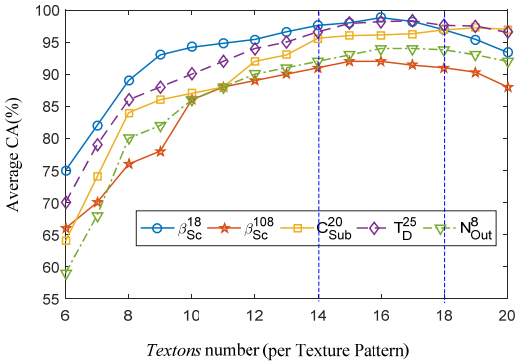


FIGURE 9. Average CA on the five data sets with different *textons* number *L*. The number of BsBFs for the one-dimensional and two-dimensional PDFE are 80 and (80, 80), and the parameters of the first-order SIGDF are set as scales $\sigma \in \{2, 2\sqrt{2}, 4, 4\sqrt{2}, 8\}$ with 9 uniform distributed-orientations in the $[0, \pi]$, namely, a total of 45 filters in the SIGDF. The scale and orientation parameters of the OAGDF are set the same as the SIGDF and the anisotropic factor $\rho = 2$. There 1.5 octave[2] Gabor wavelet filters are adopted to generate the GWFB with 5 scales and 8 orientations and $\kappa = 2.4653$.

filters in MS&MO filter banks, mainly determined by the derivative orders (d) and the number of the orientations (O) and the number of the scales (S) or the choice of the scales of the SIGDF, OAGDF and GWFB.

1) EFFECT OF THE PARAMETER OF TEXTON NUMBER IN SFD

The classification results with different *texton* number L is shown in Fig.10, where L is denoted as a ratio of *textons* number per texture patterns. Namely, the total number of *textons* is $L \times N_t$, where N_t is the number of textures. As can be seen from Fig.9, the classification accuracy (CA) increases with the increase of L and keeps stable with relative high CA on a certain range. Specifically, when L ranges from 14 to 18, the average CAs on the five data sets are relative higher than both the smaller and higher value of L, and it ultimately decreases for a sufficiently large number.

Theoretically, number L represents the richness of the SFD. Intuitively, the larger value of the *textons* number, more complicated texture patterns can be expressed and relative higher CA will be achieved. However, the number of training samples is limited which restricts the diversity expression of the texture pattern. When the *textons* number L exceeds the diversity expression ability of the training samples, the CA of the proposed TCvBsISM will decrease.

Extensive experiments demonstrated that the proposed TCvBsISM reached its best classification performance as the L ranges in 14 to 18. Hence, we fix the *textons* number L as 16 per texture pattern in the following experiments except with a specific statement.

2) EFFECT OF THE NUMBER OF BSBFs

The number of BsBFs directly affects the fitting goodness of the PDF and finally effects the classification performance. Fig.10 displays the PDFE results of a testing texture image (D27 in Brodatz database). It can be seen clearly that the smaller M generate the smoother PDF model, whilst the larger M results in the complicated PDF model.

The goodness of the estimated probability density model was tested by the one-sample Kolmogorov-Smirnov(KS) test method. In this work, extensive numeric KS statistics have demonstrated when the number of BsBFs is larger than 40, the BsBFs-based PDF model on any test database can pass the KS test. Hence, the number of the BsBFs for precise enough PDFE should be larger than 40.

Without doubt, the larger number of BsBFs, the complicated PDF model will be achieved. However, the larger number of BsBFs may not mean the better classification performance, because the larger BsBFs number would bring about the unstable and complicated PDFE results, which would consequently have an adverse impact on the following texture pattern representation and texture classification.

We conducted 30 independent trails with the same parameter setting rule on the five data sets, the average CAs with the standard deviations of the proposed TCvBsISM method with different BsBFs numbers are displayed in Fig.11. The parameters of the MS&MO filter banks are set up the same as experiments in Fig.9 and the *textons* number L stays for 16 per texture pattern for every data set.

As can be seen from Fig.12, given the fixed parameters except of the BsBFs number M, the CAs don't always increase with the increasing of the BsBFs number. When M is over a certain value (about 65), CAs will decrease. The experimental results are generally consistent to our speculation that unstable or complicated PDFE results would generate an adverse impact on TC. However, CAs on data sets, $\mathcal{T}_{D}^{25}$ and $\mathcal{N}_{out}^{8}$, will improve again when the BsBFs number M

[2]The vivo biological experiments have demonstrated that the half magnitude spatial frequency widths of most simple cells in the visual cortex have a half magnitude spatial frequency width between 1 and 1.5 octaves, with a median of 1.4 octaves [41].

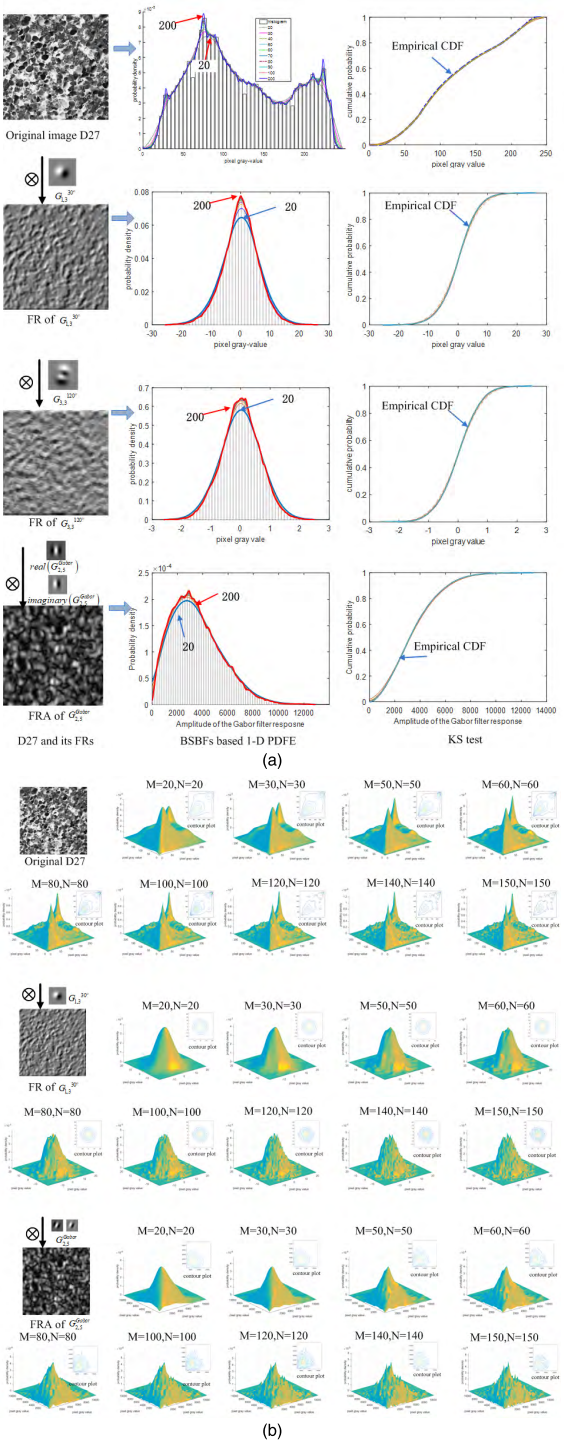


FIGURE 10. BSBFs based PDFE of texture image D27 and its filter responses (FRs) with SIGDF of different derivative orders and the filter response amplitude (FRA) of a Gabor filter, where $G_{1,3}^{30^\circ}$ represents the first-order GDF of the orientation of 30°, $G_{3,3}^{120^\circ}$ stands for the third-order GDF of orientation of 120° in the counter-clockwise direction, $G_{2,5}^{Gabor}$ is the Gabor wavelet filter of the subband $(u, v) = (2, 5)$ in (38). (a) BSBFs based 1-D PDFE. (b) BSBFs based 2-D PDFE.

increases to a relative large value (in the experiment, it is bigger than 120). The reason for this result mainly results from the difference of the numbers of the training samples.

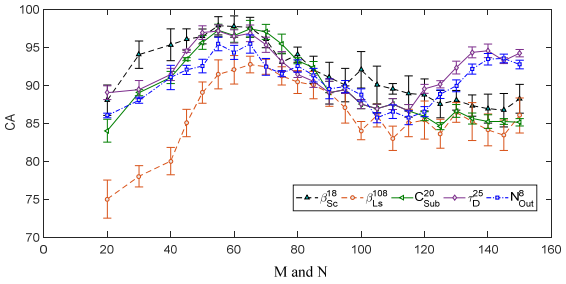


FIGURE 11. CAs (%) on the five data sets with different number of BsBFs.

The first three test data sets, $\mathcal{B}_{\mathrm{Sc}}^{18}$, $\mathcal{B}_{\mathrm{Ls}}^{108}$ and $\mathcal{C}_{Sub}^{20}$, have fewer training samples than the last two data sets, $\mathcal{T}_{D}^{25}$ and $\mathcal{N}_{out}^{8}$. Hence, when the complicated or unstable PDFE results achieved due to a relative large BsBFs number, the adverse effect will be alleviated by the increasing training samples and it can be expected that when the training samples is sufficient, a relative larger BsBFs number (however, it should not be too large, which will bring about unstable PDFE results and leads to too many LWCs adverse to the following texture pattern representation and classification) would achieve better classifier performance at a certain extend. However, when the training samples is limited, just like the experiments in this work, the proper BsBFs number should be ranged from 50 to 70 as displayed in Fig.12.

3) EFFECT OF THE SETTING OF FILTER BANKS

The generation of the MS&MO filter banks is theoretically another great impact factor to the classification performance. To facilitate description, we used G_1, G_2 and G_3 represent the first-order, second-order and third-order SIDGFs, respectively, where $G_i = \left\{G_{i,\sigma}^{\theta} | 1 \le i \le 3, \sigma \in \boldsymbol{\sigma}, \theta \in \boldsymbol{\theta}\right\}$. And we denote $\nabla EG = \left\{\nabla EG_{\sigma,\rho,\theta} | \sigma \in \boldsymbol{\sigma}, \theta \in \boldsymbol{\theta}, \rho \in \boldsymbol{\rho}\right\}$ as the OAGDF and $G^{gabor} = \{G_{u,v}^{gabor} | u \in \{0, 1, \cdots, U-1\}, v \in \{0, 1, \cdots, V-1\}\}$ as the GWFB.

Table III displays the CAs of TCvBsISM with different combinations of the filter banks, where the scale parameter σ of the SIGDF and the OAGDF are set as $\boldsymbol{\sigma} = \left\{\sqrt{2}, 2, 2\sqrt{2}, 4, 4\sqrt{2}, 8\right\}$, the orientation parameter $\boldsymbol{\theta} = \left\{0, \frac{\pi}{90}, \frac{2\pi}{90}, \cdots, \pi\right\}$, the anisotropic factor $\boldsymbol{\rho} = \{2, 3, 4, 5, 7, 9\}$. 40 Gabor wavelet filters under 5 scales and 8 orientations with $\kappa = 2.4653$. The number of BsBFs for the one-dimensional and two-dimensional PDFE are 60 and (60,60). In Tab.3, *None* means only the raw pixel are considered for texture feature extraction, Com_1 means the combination of the all the first-order, second-order and third-order SIGDF.

As can be seen from Tab.III, the proposed TCvBsISM can achieve quite high and stable CAs on the five data sets. Especially, when it considers the ITSA features of both the image raw pixel space and the filter response space of the full combination of SIGDFs, OAGDFs and GWFB, the average CA on the small scale Brodatz data set $\mathcal{B}_{Sc}^{18}$ is nearly 99%

TABLE 3. CAs (%) with different combination of filter banks(FBs).

FBs	$\mathcal{B}_{Sc}^{8}$	$\mathcal{B}_{Ls}^{108}$	C_{Sub}^{20}	$\mathcal{T}_{D}^{25}$	$\mathcal{N}_{out}^{8}$
None	92.67±4.54	90.12±4.34	95.12±2.56	94.32±3.52	89.23±4.56
G_1	95.34±1.34	93.34±2.34	95.34±4.34	95.34±4.34	92.34±4.34
G_2	94.02±3.20	92.01±2.45	94.78±3.89	95.01±3.56	91.45±2.89
G_3	95.21±2.56	93.56±2.45	95.28±3.40	95.30±3.45	92.34±4.12
∇EG	95.89±2.45	94.02±1.23	95.89±1.02	94.89±0.28	93.45±2.45
G^{gabor}	95.02±2.13	93.23±3.45	94.56±4.45	96.32±2.34	93.20±3.21
G_1+G_2	96.12±2.02	94.84±2.12	95.12±3.45	96.46±3.45	93.80±2.32
G_1+G_3	96.67±2.30	94.96+3.45	95.68±2.14	96.04±3.02	93.98±1.23
G_2+G_3	95.89±2.45	95.02±2.34	94.32±2.13	96.54±2.80	93.89±2.45
Com_1	97.12±1.23	95.81±1.68	95.68±1.24	96.89±2.30	93.96±2.34
Com_1+∇EG	97.45±2.34	95.84±1.24	96.82±2.12	96.80±1.48	94.08±2.68
Com_1+G^{gabor}	98.67±1.23	95.87±3.34	97.67±2.33	97.07±2.64	92.07±2.34
∇EG + G^{gabor}	97.02±2.34	96.02±2.23	97.02±1.38	97.82±1.34	94.21±1.32
Com_1+∇EG+G^{gabor}	**98.80±0.89**	**96.20±1.39**	**98.72±0.79**	**98.04±1.02**	**94.31±0.79**

with low fluctuation (reflected by the standard derivation of the 30 independent trails, as low as 0.89 in percentile). The best CA on the challenging data set $\mathcal{B}_{Lc}^{108}$ is over 96%, which is a promising classification result comparing with the state-of-the-art methods in terms of the confusing visual texture patterns existing this testing data set. The best average CA on C_{Sub}^{20} and $\mathcal{T}_{D}^{25}$ are both over 98%. The accurate identification of the texture pattern in data sets C_{Sub}^{20} and $\mathcal{T}_{D}^{25}$ are also challenging tasks since the great inter-similarities existing on the different texture patterns, though the texture pattern numbers in these two data sets are not as large as that in $\mathcal{B}_{Lc}^{108}$. The proposed method can also achieve good classification performance on the natural environmental scene data set, $\mathcal{N}_{out}^{8}$, which is over 94% though the images in a specific scene exhibit dramatically variant ITSA due to the great variability in albedo, illumination, camera and viewing position, etc.

In all, the proposed TCvBsISM can achieve relatively good CAs because of the full consideration of the global organization and the local structural layout of the ITSA. The global and local structural features of texture images are transformed to the LWCV of the pre-fixed BSBFs, which can effectively obtain the essential visual structural feature of the texture image and is beneficial to the statistical feature dictionary learning and pattern representation.

C. COMPARATIVE EVALUATION

We compared TCvBsISM with six classic TC methods, GLCM [22], Gabor wavelet transform (GWT) [22], Multi-resolution Markov random field (MRMRF) model [47], VZ algorithm [18], PATCH [48], PATCH-MRF [48] and six more recent methods including Random-Projection (RP) [1], FbLBP [13], LBPDTCWT [11], LETRIST [49], SWM [50] and PRICoLBP [10]. Details of these methods are expressed as follows.

GLCM [22], [51]. It characterizes texture structures by calculating how often pairs of pixel intensity in a specified spatial relationship occur by generating GLCMs. Based on each GLCM, 14 statistical measures can be computed. The extended GLCM-based texture analysis approach, termed multi-scale GLCM, as reported in [22], is taken for comparison.

GWT [22]. It applies the Gabor wavelet transform with 5 scales and 8 directions on the texture image and we obtain 40 subbands, then a total of 80 parameters including the mean and variance of the energy spectrum of each Gabor subband are computed to constitute the Gabor texture feature vector.

MRMRF [47]. A 3-layer Haar wavelet decomposition is applied to construct the multiresolution image representation firstly, and then the MRF model in each decomposition subband is established. Each subband image is modeled with 4 non-zero MRF model parameters. For the 9 detailed subbands, a total of 36 parameters can be obtained. At the same time, 10 wavelet energy signatures (including the lowpass subimage energy signature) are collected. Hence, we can obtain $36 + 10$ parameters to generate the MRMRF feature vector. If we divided the image into nonoverlapped subimages of 32×32 pixels and each subimage is processed the same as an independent image.

VZ algorithm [18]. Texture pattern is represented by the joint distribution of image filtering responses (MR8 filter bank is used in the experiments, since it achieves the best classification accuracies as reported in the original literture [18])) The distribution is represented by the frequency histogram of filter response cluster centers (*textons*) of the filter resposnes.

PATCH [19]. Raw pixel intensities of local patches from the selected training images in a texture class are aggregated and clustered. The set of cluster centers from all the classes comprises the *texton* dictionary. Both the training and the testing is performed in the patch domain. The classification is still achieved by nearest neighbor classifier with the χ^2 statistic.

PATCH-MRF [19]. A texture image is represented by a two-dimensional histogram: one dimension for the quantized bins of the patch center pixel, the other dimension for the learned *textons* from the patch with the center pixel excluded. As reported in [19], it chooses 200 bins for the central pixel with $10 \times N_{TP}$ textons, which can achieve relative better performance, where N_{TP} means the number of texture patterns.

RP [1]. It also a patch-based texture classification approach and implemented directly in the compressed domain based on the random projection for dimension reduction of the loach patches. The only difference bewtten the PATCH method is the random measurements are used as feature for compressed dictionary learning.

FbLBP [13] is a LBP-extended texture feature analysis approach. The texture features are extracted as follows. Firstly, the difference vector of the local patches based on the traditional LBP approach is decomposed into sign part and magnitude part, the sign part is described by conventional LBP, while the magnitude part is described by two features of the mean and the variance of the magnitude vector.

LBPDTCWT [11]. The local feature of a texture pattern is represented by the classic LBP feature by applying the

dual-tree complex wavelet transform (DTCWT) and the global texture features are extracted the energies of the DTCWT coefficients on the detail subbands of the log-polar (LP) transform image. Finally, the local feature histograms and the global energy-based feature of all the subbands can be concatenated to generate the fused texture feature.

LETRIST [49]. It is also a histogram-based texture feature representation that encodes the joint information within an image across feature and scale spaces.

SWM [50]. It encodes the class-specific local organizations of image directions (LOIDs) based on steerable circular harmonic wavelets(CHWs).

PRICoLBP [10]. It is an extended LBP feature. According to the processing method reported in literature [10], it extracts six co-occurrence patterns from three scales and two orientations and thus achieve a $2 \times 3 \times 590$-dimensional feature vector.

TCvBsISM. Both the statistical features of the original pixel intensity space and the filter bank responses space are considered. In the comparative experiments, the full combination of the three filter banks are applied, and the parameters are set as the experiments in Tab.III.

TABLE 4. CAs (%) of the comparative texture classification methods on the first Four data sets.

Methods	Data sets				Published In
	$\mathcal{B}_{Sc}^{24}$	$\mathcal{T}_{D}^{25}$	C_{Sub}^{20}	$\mathcal{B}_{Ls}^{108}$	
GLCM	87.19^{+}	81.19	N/A	N/A	NC2013[22]
GWT	89.40^{+}	69.90	N/A	N/A	NC2013[22]
MRMRF	90.35$^{\triangle}$	N/A	N/A	N/A	$^{\triangle}$PRL[47]
VZ algorithm	N/A	N/A	96.16(0.37)	N/A	PAMI2009[19]
PATCH	N/A	92.94*	97.47*	91.8(0.49)$^{\triangle}$	*TIP2017 [50] $^{\triangle}$PAMI2012[1]
PATCH-MRF	N/A	N/A	97.97	N/A	PAMI2012[1]
RP	96.90(0.35)	N/A	98.52	94.2(0.49)	PAMI2012[1]
FbLBP	N/A	94.14	97.02	N/A	ESWA2017[13]
LBPDTCWT	**99.75**	N/A	86.39	N/A	ACCESS2018[11]
LETRIST	N/A	**97.63**	**98.54**	N/A	TCSVT2017[49]
SWM	N/A	89.12	96.86	N/A	TIP2017[50]
PRICoLBP	N/A	80.38$^{\triangle}$	98.40*	**96.90**	*PAMI2014[10] $^{\triangle}$PR2017[52]
TCvBsISM(14)	98.45(1.32)	97.98(0.92)	96.73(2.02)	95.98(0.52)	
TCvBsISM(15)	98.57(1.44)	97.96(1.34)	97.02(0.97)	95.90(2.02)	
TCvBsISM(16)	**98.80**(0.89)	**98.04**(1.02)	**98.72**(0.79)	**96.20**(1.39)	
TCvBsISM(17)	97.67(0.88)	97.89(2.02)	98.34(0.79)	95.89(2.02)	
TCvBsISM(18)	97.85(1.12)	97.84(2.72)	97.34(2.02)	95.02(2.02)	

The CAs on $\mathcal{B}_{Sc}^{24}$, $\mathcal{T}_{D}^{25}$, C_{Sub}^{20} and $\mathcal{B}_{Ls}^{108}$ are shown in Tab.IV, where the classification results of the twelve comparative methods are quoted the best results(with different parameter setting) directly from the published papers as denoted in the last column. The classification results of GLCM and GWT method on the data set $\mathcal{B}_{Sc}^{24}$, marked with plus, means there is a little difference of the data set between this work and the published paper [22]. In [22], the number of the texture patterns used for experiments is 32, but the size of the image is smaller as 64×64. The numbers followed the CAs in the brackets are the standard deviations of the repeated replacement experiments. The numbers in the brackets followed TCvBsISM stands for the number of the *textons* of each texture pattern learned in the propose method.

As can be seen from Tab.IV, the proposed TCvBsISM can achieve CAs of 98.80%, 96.20%, 98.72%, 98.04% on average base on the 30 independent experiments with the *textons* number of 16 per texture pattern on the first four data sets, $\mathcal{B}_{Sc}^{24}$, $\mathcal{B}_{Ls}^{108}$, C_{Sub}^{20} and $\mathcal{T}_{D}^{25}$. Compared with the classic and the more recent TC methods, TCvBsISM outperformances all of them on these two data sets, $\mathcal{T}_{D}^{25}$ and C_{Sub}^{20}. Seemingly, LBPDTCWT can achieve the highest CA on the data set $\mathcal{B}_{Sc}^{24}$. However, $\mathcal{B}_{Sc}^{24}$ is a small-scale data set and the visual appearances of different texture patterns look different, hence almost all the state-of-the-art methods can achieve satisfactory classification results. Abundant literature reports have demonstrated that almost all the state-of-the-art method can achieve high CAs (generally higher than 96%) and significantly higher than the early classic TC methods, GLCM, GWT and MRMRF. For instance, the proposed method can also achieve as high as 98.80% on average with a low standard deviation based on this data set. However, the comparative

TABLE 5. CAs (%) of the comparative texture classification methods on the challenging data sets $\mathcal{N}_{out}^{8}$ and $\mathcal{B}_{Ls}^{108}$.

L	method	$\mathcal{B}_{Ls}^{108}$	$\mathcal{N}_{out}^{8}$
14	GLCM	86.42(±2.58)	79.22(±2.45)
	GWT	85.83(±2.88)	77.02(±2.68)
	MRMRF	90.48(±2.02)	83.48(±4.02)
	VZ algorithm	93.23(±3.52)	89.89(±3.52)
	PATCH	93.83(±2.12)	90.12(±2.65)
	PATCH-MRF	93.92(±1.52)	90.23(±2.71)
	RP	94.02(±1.34)	90.23(±2.13)
	FbLBP	90.12(±2.01)	87.34(±2.32)
	LBPDTCWT	89.92(±0.45)	89.23(±1.23)
	LETRIST	94.23(±2.34)	90.34(±1.26)
	SWM	93.45(±0.89)	92.34(±1.45)
	PRICoLBP	93.56(±0.58)	**93.90**(±2.34)
	TCvBsISM	**95.98**(±0.52)	**93.90**(±1.02)
15	GLCM	89.43(±2.45)	84.90(±3.56)
	GWT	88.98(±1.98)	75.78(±2.34)
	MRMRF	87.78(±2.76)	83.34(±3.41)
	VZ algorithm	90.12(±3.01)	89.45(±2.32)
	PATCH	93.91(±2.62)	90.12(±3.72)
	PATCH-MRF	93.89(±3.12)	91.23(±2.42)
	RP	95.45(±1.31)	92.34(±1.23)
	FbLBP	89.45(±1.45)	88.34(±1.45)
	LBPDTCWT	88.78(±2.03)	87.34(±2.35)
	LETRIST	92.34(±1.23)	91.45(±1.45)
	SWM	92.45(±1.56)	90.45(±2.45)
	PRICoLBP	93.45(±1.34)	93.65(±1.04)
	TCvBsISM	**95.90**(±2.02)	**94.87**(±2.02)
16	GLCM	90.12(±2.34)	84.56(±2.45)
	GWT	89.45(±1.56)	75.45(±2.76)
	MRMRF	87.34(±2.37)	82.72(±3.02)
	VZ algorithm	93.98(±2.76)	88.98(±2.09)
	PATCH	94.02(±2.52)	89.90(±2.62)
	PATCH-MRF	94.18(±1.82)	91.02(±1.82)
	RP	94.67(±2.34)	91.45(±1.34)
	FbLBP	88.45(±3.04)	86.04(±2.04)
	LBPDTCWT	87.34(±1.23)	89.34(±1.45)
	LETRIST	92.30(±1.67)	90.02(±2.59)
	SWM	91.45(±1.59)	90.45(±2.49)
	PRICoLBP	92.14(±1.45)	93.32(±2.34)
	TCvBsISM	**96.20**(±1.39)	**94.31**(±0.79)
17	GLCM	89.98(±2.33)	86.35(±2.03)
	GWT	90.01(±2.39)	75.56(±1.38)
	MRMRF	90.03(±1.69)	81.32(±0.98)
	VZ algorithm	90.58(±2.60)	87.08(±2.03)
	PATCH	92.34(±2.90)	89.45(±3.13)
	PATCH-MRF	93.02(±2.87)	90.45(±2.62)
	RP	94.56(±2.34)	92.45(±2.19)
	FbLBP	90.23(±2.34)	87.89(±2.30)
	LBPDTCWT	89.45(±1.63)	89.45(±1.89)
	LETRIST	91.23(±1.23)	91.12(±0.91)
	SWM	92.12(±2.90)	91.24(±1.39)
	PRICoLBP	94.34(±2.58)	92.45(±2.73)
	TCvBsISM	**95.89**(±2.02)	**94.45**(±2.02)
18	GLCM	90.12(±2.02)	86.90(±1.98)
	GWT	89.49(±1.90)	76.79(±1.24)
	MRMRF	89.98(±2.05)	82.82(±1.29)
	VZ algorithm	90.01(±2.39)	88.82(±2.30)
	PATCH	91.45(±2.90)	89.49(±2.62)
	PATCH-MRF	92.02(±1.32)	90.45(±2.78)
	RP	**95.60**(±2.56)	91.49(±1.56)
	FbLBP	88.45(±2.45)	87.34(±2.56)
	LBPDTCWT	87.56(±1.34)	86.34(±1.23)
	LETRIST	92.56(±2.90)	89.23(±1.69)
	SWM	91.45(±3.40)	90.12(±1.29)
	PRICoLBP	93.23(±0.89)	90.34(±2.45)
	TCvBsISM	95.02(±2.02)	**95.02**(±2.02)

TABLE 6. Comparison of the Highest classification performance on data sets $\mathcal{N}_{out}^{8}$.

GLCM	GWT	MRMRF	VZ	PATCH	PATCH-MRF	RP	FbLBP	LBPDTCWT	LETRIST	SWM	PRICoLBP	TCvBsISM
85.35	76.79	83.48	89.89	90.12	91.23	92.45	88.34	89.45	91.45	92.34	93.90	**95.02**

method LBPDTCWT achieves much lower CA on data set C_{Sub}^{20} based on the original literature reports of this method. Hence it indicates that performance of this TC method is vulnerable, which is a data set-specific method. Analogically, PRICoLBP can achieve higher CA on data set $\mathcal{B}_{Ls}^{108}$ but it did not perform well on the data set $\mathcal{T}_{D}^{25}$. Whereas the proposed method TCvBsISM can achieve relative high CAs on all the first four data sets.

In terms of the aforementioned description of these testing data sets, $\mathcal{B}_{Ls}^{108}$ and $\mathcal{N}_{out}^{8}$ are two challenging data sets. $\mathcal{B}_{Ls}^{108}$ includes a large number of texture patterns, but the sample number is relative small. Even worse, some different texture pattern exhibit very similar visual appearance. $\mathcal{N}_{out}^{8}$ is comprised of the natural outdoor scene patterns. Due to the large intra-variation resulting from the uncontrollable illumination change and the viewing points of the camera, $\mathcal{N}_{out}^{8}$ is also a challenging testing data set. Hence, we mainly concern the classification performance of these comparative methods on these two data sets, $\mathcal{B}_{Ls}^{108}$ and $\mathcal{N}_{out}^{8}$, of the following experiments.

In the following comparative experiments, we use the identical TC framework of the proposed method TCvBsISM for all the comparative TC methods (six classic and six more recent TC methods). In other words, we carry out three stages, including dictionary learning, histogram-based texture pattern representation and texture pattern classification, for all of these methods, except that, instead of using the statistical distribution model parameter feature of the proposed method, these comparative TC methods use their own texture feature and all the other parameters of different TC methods stay the same in the experiments. The *textons* number in the statistical feature dictionary is set ranging from 14 to 18 per texture pattern. Each texture pattern is modeled by the occurrence frequencies (normalized histogram) of the learned *textons*. The CAs on the data sets, $\mathcal{B}_{Ls}^{108}$ and $\mathcal{N}_{out}^{8}$, are reported in Tab.V, where L means the number of textons per texture pattern.

As can be seen from Tab.V, TCvBsISM can achieve the best CA on the challenging data set $\mathcal{B}_{Ls}^{108}$ as high as 96.02% with the *textons* number of 16 per texture pattern on average, which is apparently superior over the other comparative methods with the same parameter setting. Though PR method can achieve a little higher (higher about 0.58% on average) CA on $\mathcal{B}_{Ls}^{108}$ with the *textons* number of 18 per texture pattern, which underperforms on data set $\mathcal{N}_{out}^{8}$ (lower 3.53% than the proposed method on that data set). The superior classification results of the proposed method TCvBsISM mainly due to the elaborate texture pattern representation, including the B-spline-based marginal and joint probability distribution estimation, the joint information exhibition both on raw pixel space and the filter response space and the joint *textons* distribution-based texture pattern representation, which can effectively approximate any ITSA for texture pattern classification without any prior knowledge of the texture pattern. Thus the proposed method TCvBsISM has strong discriminant ability of the visually confusing texture patterns and achieve higher classification performance on the testing data set.

Besides the pure texture data bases, TCvBsISM can also achieve better classification results (TCvBsISM gives the highest classification, reaching as high as 95.02%) on the natural scene database $\mathcal{N}_{out}^{8}$ according to the classification results displayed in Tab.V. To facilitate performance comparison, Table VI summarizes the preceding table, presenting the overall best classification performance achieved by each method on data set $\mathcal{N}_{out}^{8}$ for any parameter setting.

The natural scene classification is a much more complex issue because: (1) There are thousands of scenes exist in nature. (2) The surface appearance of the captured images is unpredictable with the unknown and unpredictable illumination, viewpoint, and the imaging distance and so on. Hence, even for a specific natural scene, e.g., street, the captured images will exhibit great intra-variation. TCvBsISM can well fit any marginal and joint distribution of texture images for the global and local visual characteristics representation, hence it can achieve much higher CAs on this natural scene data set..

V. CONCLUSIONS

This paper presents TCvBsISM, a B-spline PDFE-based TC scheme, which converts the representation of the complicated ITSA into the estimation of the LWCs of the pre-defined BSBFs. The proposed TCvBsISM approach can achieve promising classification performance on the identification of the visually similar texture images with limited training samples mainly results from the following aspects:

(1) TCvBsISM fully considers the basic features of the global organization and the local spatial structural exhibition of LHPs in texture image, which are believed to be the essential factor to determine the ITSAs of different texture patterns.

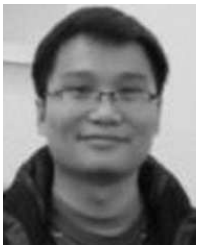
(2) It demonstrated that the global organization and the local spatial structural characteristics of LHPs are boiled down to the 1-D and 2-D probability distribution models, respectively, which are approximated by the BSBFs-based PDFE, rather than the ESMs or some low or middle-high-order moments of the statistical distribution. Corresponding LWCs are determined by an entropy-based optimization criterion with a prior smooth constraint over the LWCs. Hence, TCvBsISM is not texture pattern-specific and it can make effective texture analysis for any kind of texture pattern without any prior knowledge of the image distribution.

(3) Texture pattern is modeled by the joint probability distribution of B-spline statistical feature-based *textons*, which has taken advantage of the merits of the-state-of-the-art TC classification methods, e.g., VZ algorithm and PATCH method. Both the raw pixel space information and filter response space information are considered comprehensively and the patch-based texture pattern modeling is adopted. Hence it is easy to implement and can achieve higher and more stable classification accuracy.

REFERENCES

[1] L. Liu and P. W. Fieguth, "Texture classification from random features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 574–586, Mar. 2012.
[2] X. Sun, Z. Liu, Y. Hu, L. Zhang, and R. Zimmermann, "Perceptual multi-channel visual feature fusion for scene categorization," *Inf. Sci.*, vol. 429, pp. 37–48, Mar. 2018.
[3] L. Liu, L. Shao, and X. Li, "Evolutionary compact embedding for large-scale image classification," *Inf. Sci.*, vol. 316, pp. 567–581, Sep. 2015.
[4] J. Chen *et al.*, "WLD: A robust local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1705–1720, Sep. 2010.
[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[6] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
[7] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun./Jul. 2004, pp. 506–513.
[8] L. Yu, Z. He, and Q. Cao, "Gabor texture representation method for face recognition using the Gamma and generalized Gaussian models," *Image Vis. Comput.*, vol. 28, pp. 177–187, Jan. 2010.
[9] S. Liao, M. W. K. Law, and A. C. S. Chung, "Dominant local binary patterns for texture classification," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 1107–1118, May 2009.
[10] X. Qi, R. Xiao, C.-C. Li, Y. Qiao, J. Guo, and X. Tang, "Pairwise rotation invariant co-occurrence local binary pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2199–2213, Nov. 2014.
[11] P. Yang, F. Zhang, and G. Yang, "Fusing DTCWT and LBP based features for rotation, illumination and scale invariant texture classification," *IEEE Access*, vol. 6, pp. 13336–13349, 2018.
[12] Z. Pan, X. Wu, Z. Li, and Z. Zhou, "Local adaptive binary patterns using diamond sampling structure for texture classification," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 828–832, Jun. 2017.
[13] Z. Pan, Z. Li, H. Fan, and X. Wu, "Feature based local binary pattern for rotation invariant texture classification," *Expert Syst. Appl.*, vol. 88, pp. 238–248, Dec. 2017.
[14] L. Yu, L. Feng, H. Wang, L. Li, Y. Liu, and S. Liu, "Multi-trend binary code descriptor: A novel local texture feature descriptor for image retrieval," *Signal Image Video Process.*, vol. 12, pp. 247–254, Feb. 2018.
[15] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 524–531.
[16] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognit.*, vol. 42, no. 3, pp. 425–436, 2009.
[17] T. Randen and J. H. Husøy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 291–310, Apr. 1999.
[18] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *Int. J. Comput. Vis.*, vol. 62, pp. 61–81, Apr. 2005.
[19] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2032–2047, Nov. 2009.
[20] J. Liu, Z. Tang, P. Xu, W. Liu, J. Zhang, and J. Zhu, "Quality-related monitoring and grading of granulated products by Weibull-distribution modeling of visual images with semi-supervised learning," *Sensors*, vol. 16, no. 7, p. 998, 2016.
[21] J. Liu, W. Gui, Z. Tang, H. Hu, and J. Zhu, "Machine vision based production condition classification and recognition for mineral flotation process monitoring," *Int. J. Comput. Intell. Syst.*, vol. 6, no. 5, pp. 969–986, 2013.
[22] F. R. de Siqueira, W. R. Schwartz, and H. Pedrini, "Multi-scale gray level co-occurrence matrices for texture description," *Neurocomputing*, vol. 120, pp. 336–345, Nov. 2013.
[23] J. Fan, Y. Gao, H. Luo, and G. Xu, "Statistical modeling and conceptualization of natural images," *Pattern Recognit.*, vol. 38, pp. 865–885, Jun. 2005.
[24] A. Lagrange, M. Fauvel, and M. Grizonnet, "Large-scale feature selection with Gaussian mixture models for the classification of high dimensional remote sensing images," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 230–242, Jun. 2017.
[25] J. Liu, Z. Tang, J. Zhang, Q. Chen, P. Xu, and W. Liu, "Visual perception-based statistical modeling of complex grain image for product quality monitoring and supervision on assembly production line," *PLoS ONE*, vol. 11, p. e0146484, Mar. 2016.
[26] J. Liu, Z. Tang, W. Gui, W. Liu, P. Xu, and J. Zhu, "Application of statistical modeling of image spatial structures to automated visual inspection of product quality," *J. Process Control*, vol. 44, pp. 23–40, Aug. 2016.
[27] J. Liu, Z. Tang, Q. Chen, P. Xu, W. Liu, and J. Zhu, "Toward automated quality classification via statistical modeling of grain images for rice processing monitoring," *Int. J. Comput. Intell. Syst.*, vol. 9, no. 1, pp. 120–132, 2016.
[28] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imag. Vis.*, vol. 18, no. 1, pp. 17–33, 2003.
[29] D. K. Hammond and E. P. Simoncelli, "Image modeling and denoising with orientation-adapted Gaussian scale mixtures," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2089–2101, Nov. 2008.
[30] Z. Zong and K. Y. Lam, "Bayesian estimation of 2-dimensional complicated distributions," *Struct. Saf.*, vol. 23, no. 2, pp. 105–121, 2001.
[31] Z. Zong and K. Y. Lam, "Bayesian estimation of complicated distributions," *Struct. Saf.*, vol. 22, pp. 81–95, Feb. 2000.
[32] J. Han, C. Yang, X. Zhou, and W. Gui, "Entropy-based estimation of bubble size distributions in froth flotation using B-spline functions," *IFAC-PapersOnLine*, vol. 49, no. 20, pp. 96–101, 2016.
[33] Z. Zong and K. Y. Lam, "Estimation of complicated distributions using B-spline functions," *Struct. Saf.*, vol. 20, pp. 341–355, Dec. 1998.
[34] J.-M. Geusebroek, A. W. M. Smeulders, and J. V. de Weijer, "Fast anisotropic Gauss filtering," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 938–942, Aug. 2003.
[35] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.
[36] M. Jacob and M. Unser, "Design of steerable filters for feature detection using canny-like criteria," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1007–1019, Aug. 2004.
[37] C. Lopez-Molina, G. V.-D. de Ulzurrun, J. M. Baetens, J. Van den Bulcke, and B. De Baets, "Unsupervised ridge detection using second order anisotropic Gaussian kernels," *Signal Process.*, vol. 116, pp. 55–67, Nov. 2015.
[38] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, Oct. 1996.
[39] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A*, vol. 2, no. 7, pp. 1160–1169, 1985.
[40] R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vis. Res.*, vol. 22, no. 5, pp. 545–559, 1982.
[41] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 29–44, Feb. 2001.
[42] S. Abdelmounaime and H. Dong-Chen, "New Brodatz-based image databases for grayscale color and multiband texture analysis," *ISRN Mach. Vis.*, vol. 2013, Jan. 2013, Art. no. 876386.
[43] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Trans. Graph.*, vol. 18, no. 1, pp. 1–34, Jan. 1999.
[44] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.

[45] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.
[46] L. Wang and J. Liu, "Texture classification using multiresolution Markov random field models," *Pattern Recognit. Lett.*, vol. 20, pp. 171–182, Feb. 1999.
[47] T. Song, H. Li, F. Meng, Q. Wu, and J. Cai, "LETRIST: Locally encoded transform feature histogram for rotation-invariant texture classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1565–1579, Jul. 2018.
[48] A. Depeursinge, Z. Puspoki, J. P. Ward, and M. Unser, "Steerable wavelet machines (SWM): Learning moving frames for texture classification," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1626–1636, Apr. 2017.
[49] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[50] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recognit.*, vol. 62, pp. 135–160, Feb. 2017.



JINPING LIU received the B.Eng. degree in computer science and technology, the M.Eng. degree in computer application, and the Ph.D. degree in control science and engineering from the School of Information Science and Engineering, Central South University, in 2006, 2009, and 2013, respectively. He is currently an Associate Professor with the College of Information Science and Engineering, Hunan Normal University. He is also a Researcher with the Key Laboratory of High Performance Computing and Stochastic Information Processing (Ministry of Education of China), Hunan Normal University. His research interests include digital signal processing and pattern recognition.

JIEZHOU HE received the B.Eng. degree in computer science and technology from Hunan Normal University, where he is currently pursuing the master's degree with the College of Information Science and Engineering. His research interest covers computer vision and pattern recognition.

WUXIA ZHANG received the B.Eng. degree in computer science and technology from Henan Agricultural University in 2015. She is currently pursuing the master's degree in computer application with the College of Information Science and Engineering, Hunan Normal University. Her research interest covers computer vision and pattern recognition with their industrial applications.

PENGFEI XU received the B.Eng. degree in computer application from Hunan Normal University in 2000, the M.Eng. degree in computer science and technology from Nankai University in 2003, and the Ph.D. degree in computer application technology from Central South University in 2012. He is currently an Associate Professor with the College of Information Science and Engineering, Hunan Normal University. His research interest includes computer vision and pattern recognition.

ZHAOHUI TANG received the Ph.D. degree from Central South University (CSU), China, in 2008. He was a Visiting Scholar with the University of Duisburg-Essen, Germany, from 2005 to 2006. He is currently a Professor with the School of Information Science and Engineering, CSU. His research interests include signal processing and industrial process fault diagnosis.