This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2018.2800008, IEEE Transactions on Pattern Analysis and Machine Intelligence

Compressive Binary Patterns: Designing a Robust Binary Face Descriptor with Random-Field Eigenfilters

Weihong Deng, Jiani Hu, Jun Guo

***Abstract*— A binary descriptor typically consists of three stages: image filtering, binarization, and spatial histogram. This paper first demonstrates that the binary code of the maximum-variance filtering responses leads to the lowest bit error rate under Gaussian noise. Then, an optimal eigenfilter bank is derived from a universal assumption on the local stationary random field. Finally, compressive binary patterns (CBP) is designed by replacing the local derivative filters of local binary patterns (LBP) with these novel random-field eigenfilters, which leads to a compact and robust binary descriptor that characterizes the most stable local structures that are resistant to image noise and degradation. A scattering-like operator is subsequently applied to enhance the distinctiveness of the descriptor. Surprisingly, the results obtained from experiments on the FERET, LFW, and PaSC databases show that the scattering CBP (SCBP) descriptor, which is handcrafted by only 6 optimal eigenfilters under restrictive assumptions, outperforms the state-of-the-art learning-based face descriptors in terms of both matching accuracy and robustness. In particular, on probe images degraded with noise, blur, JPEG compression, and reduced resolution, SCBP outperforms other descriptors by a greater than 10% accuracy margin.**

***Index Terms*— Face Recognition, Local Binary Patterns, Binary Code Learning, Face Descriptor.**

I. Introduction

Local descriptors are at the core of many computer vision tasks. For example, local descriptors of regions of interest are widely used to find correspondences between image regions (patches), which is a key factor in a wide range of applications, ranging from stereo matching [1] and multi-view reconstruction [2] to object detection and alignment [3][4]. Furthermore, encodings of local descriptors are predominantly used for feature representation in image and video retrieval [5][6], as well as in object and scene recognition [7][8]. Due to the importance of these issues, various descriptors have been proposed with the aim of improving accuracy and efficiency. For example, regions of interest are typically represented by handcrafted SIFT [3], SURF [9] and BRIEF [10] descriptors and their variants. By end-to-end optimizing for available data, deep learning techniques, such as autoencoder and convolutional network, have recently become dominant for both local descriptors [11][12] and holistic representations [13][14].

For face recognition, local binary patterns (LBP) is one of the most popular local descriptors [15][16], and it has motivated a large family of successful handcrafted and learning-based face descriptors. Some variants of LBP improve the representational power by decomposing an image into sub-band images before LBP description [17][18], whereas others change the topology of the neighborhood to obtain greater diversity in sampling pattern shapes and sizes [19][20][21][22]. To enhance the discriminatory ability, ensemble descriptors are designed by concatenating the histograms at landmark points and regular spatial cells [23][24] or the local features are extracted in multi-scale manners [20][25]. With their adaptation to specific datasets, learning-based descriptors have generally become preferred in recent years. For example, local quantized patterns (LQP) [26] apply a clustering-based codebook to encode the long binary codes from extensive sampling pattern shapes and sizes. Discriminant face descriptor (DFD) [27] and compact binary face descriptor (CBFD) [28] also learn the local filters with objective functions on discrimination, reconstruction, and code distribution.

Despite their success, many of the previous descriptors are vulnerable to image noise or degradation. Some pioneering works on robust LBP descriptors [29][30][31][32] have skilfully designed robust encoders for binary patterns, but these works lacked sufficient theoretical analyses. In this work, we revisit Ahonen and Pietikainen's interpretation of the LBP histogram as an approximation of the joint distribution of local derivative filtering responses [33]. The framework helps us analyze the bit error rate of the LBP-like descriptor, based on which we further demonstrate that the filters with maximum-variance responses lead to the most robust binary code under additive Gaussian white noise.

Motivated by this optimality justification, we design a new random-field eigenfilter (RF eigenfilter) bank by selecting the orthonormal filters that produce the maximum-variance responses under the assumption that the local patches are stationary random fields. The novel compressive binary patterns (CBP) is proposed by simply replacing the local derivative filters of LBP with a set of *6 RF eigenfilters*, which characterize the most common local edge, wedge, and bar structures that are stably preserved during image contamination and degradation. Furthermore, a scattering operator [34] is applied to extend the scope of the 6 eigenfilters and generate a scattering CBP (SCBP) histogram to characterize more complex and "fine-grained" structures.

Although our method is simple and handcrafted, it is very effective for enhancing the robustness and informativeness of face descriptor. On the standard FERET and LFW databases, the proposed SCBP achieves better face matching accuracy than state-of-the-art handcrafted and learned face descriptors using a relatively low feature dimension. More importantly, to systematically evaluate the robustness of the face descriptor, we extend the standard FERET evaluation by superposing four types of common degradations, including *Gaussian noise, Gaussian blur, JPEG compression, and reduced resolution*, on the probe images. In this evaluation, the proposed RF-eigenfilter-based descriptors exhibit strong robustness to all types of degradation, leading to a 10%–30% accuracy gain compared to the up-to-date learning-based descriptors, such as DFD [27], CBFD [28], and handcrafted descriptors such as MD-DCP [24]. Furthermore, on the challenging PaSC database with real-world degraded images, a high-dimensional SCBP descriptor achieves superior accuracy compared to the deep autoencoder method and accuracy comparable to the VGG deep face descriptor.

II. Error Analysis of Local Binary Patterns

The LBP operator labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary code. This operator can be interpreted as a three-stage local feature description framework [33]: image filtering, binary encoding, and spatial histogram. Under this framework, Ahonen and Pietikainen [33] showed that the LBP operator is equivalent to *sign-based binary encoding of the convolution output of a set of local derivative filters*. Unfortunately, due to the high correlation between neighboring pixels of natural images, the responses of a local derivative filter are mostly close to zero. These low-amplitude responses make the sign-based binary code of LBP highly unstable under noise turbulence, resulting in a noise-sensitive descriptor. Under the LBP-based description framework, we explore how to design filters that derive the optimal robust binary code.

Consider the problem of matching two image patches with a robust binary code. Let $\mathbf{X} \in \mathbb{R}^d$ denote a vectorized image patch of the template (gallery) image, and let $\mathbf{Y} \in \mathbb{R}^d$ denote the corresponding patch (within the same spatial cell for histogram counting) of the

The authors are with the Pattern Recognition and Intelligent System Laboratory, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China. E-mail: whdeng@bupt.edu.cn.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2018.2800008, IEEE Transactions on Pattern Analysis and Machine Intelligence

test image. We assume that the difference between the template image and test image can be modeled by the following additive noise model: $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$, where $\mathbf{Z}$ is the variation term due to image sensor or encoding issues, such as Gaussian noise, blur, compression, and low resolution. In face recognition, these noise or degradations are superposed on many intra-class variations in poses, expressions, illumination, makeup, ages, etc [35][36][37].

In the image filtering stage, there are K filters denoted as a stacked filter matrix $F = [f_1, \ldots, f_K] \in \mathbb{R}^{d\times K}$, where f_i is the i-th vectorized image filter. For the i-th filtering response, we have $f_i^T\mathbf{Y} = f_i^T\mathbf{X}+f_i^T\mathbf{Z}$. In the binary encoding stage, the LBP descriptor simply uses the component-wise sign function

$$\mathbf{B} = \text{sgn}(F^T\mathbf{X}) \in \{1, 0\} \tag{1}$$

$$\mathbf{B'} = \text{sgn}(F^T\mathbf{Y}) \in \{1, 0\} \tag{2}$$

with $B_i = \text{sgn}(f_i^T\mathbf{X})$ and $B_i' = \text{sgn}(f_i^T\mathbf{Y})$, $1 \le i \le K$, as the i-th bit of the binary patterns. The sign function divides each dimension of the filter bank output into two bins, and the K-dimensional output space is uniformly divided into 2^K subspaces. Whether the corresponding patch of the test image can match the template patch is determined by the amplitude of the noise component. Let $\tilde{X}_i = f_i^T\mathbf{X}$ and $\tilde{Z}_i = f_i^T\mathbf{Z}$ denote the i-th filtered random variables. The error probability of two corresponding bits is equal to the probability that the sign of $\tilde{X}_i$ is altered by the additive noise $\tilde{Z}_i$, i.e.,

$$p_i = P_{B_iB_i'}\{B_i \neq B_i'\} \tag{3}$$

$$= P\{\tilde{X}_i > 0, \tilde{X}_i + \tilde{Z}_i < 0\} + P\{\tilde{X}_i < 0, \tilde{X}_i + \tilde{Z}_i > 0\} \tag{4}$$

To conduct an optimal analysis, we assume that the vectorized image patch follows a Gaussian distribution $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma_X)$, and in the testing stage, the patches are contaminated by additive Gaussian white noise $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \lambda^2 I)$, where I is an identity matrix.Then, their filtering responses also have a Gaussian distribution, i.e., $\tilde{X}_i \sim \mathcal{N}(0, \sigma_i^2)$ and $\tilde{Z}_i \sim \mathcal{N}(0, \lambda^2)$, which gives rise to the signal-to-noise ratio of the i-th filter response as follows:

$$SNR_i = \frac{\sigma_i^2}{\lambda^2},\ i = 1 \cdots K \tag{5}$$

where σ_i^2 is the variance of the i-th filter response. Based on these assumptions, we can compute the i-th bit error rate, denoted as p_i, as follows:

$$p_i = P_{B_iB_i'}\{B_i \neq B_i'\} \tag{6}$$

$$= 2\int_0^\infty \left[\int_{-\infty}^{-\tilde{x}_i} \phi(\tilde{z}_i, \lambda^2))d\tilde{z}_i\right] \phi(\tilde{x}_i, \sigma_i^2))d\tilde{x}_i \tag{7}$$

$$= 2\int_0^\infty Q\left(\frac{\tilde{x}_i}{\lambda}\right) \phi(\tilde{x}_i, \sigma_i^2)d\tilde{x}_i \tag{8}$$

$$= 2\int_0^\infty Q\left(t\sqrt{SNR_i}\right) \phi(t, 1)dt \tag{9}$$

where $\phi(\tilde{x}_i, \sigma_i^2)$ is the pdf of the distribution $\mathcal{N}(0, \sigma^2)$ and the last step is due to the change of variable: $\tilde{x}_i = \sigma_i t$. Since $Q(\cdot)$ is a decreasing function, p_i is a decreasing function of SNR_i. According to Eq. 5, *the filter with the maximum-variance response leads to an optimal robust binary code with the lowest bit error rate*.

If the filters are orthogonal, then the Gaussian-distributed filtering responses $\tilde{X}_i$ are uncorrelated and also independent. The probability that the K-bit binary codes are fully matching can be approximated as follows:

$$P\{\mathbf{B} = \mathbf{B'}\} \approx \Pi_{i=1}^K(1 - p_i) \tag{10}$$

The binary code matching rate is apparently a decreasing function of p_i and thus an increasing function of SNR_i. For face description, the image is divided into spatial cells, and histograms of each cell are computed independently [15], which are then concatenated to form a global description. A high pattern matching rate between the template and degraded patches would typically lead to robust matching of the histogram sequences between images, leading to a robust image matching algorithm.

III. From LBP to CBP (Compressive Binary Patterns)

This section introduces our design principle and implementation of CBP, which is a generalized form of LBP that aims to address its limitations on noisy and low-quality images. CBP is dedicated to maintaining the simplicity, low-dimensionality and learning-free advantages of LBP, which differentiates CBP from the sophisticated learning-based descriptors.

A. Optimal Design of Filter Bank for Binary Code

Our design principle of the filter bank for binary patterns considers both the robustness for noise resistance and the compactness for information preservation. First, the variances of the filter responses should be as large as possible to minimize the error rate of the binary codes under noise disturbance. The SNR of each response $SNR_i = var(f_i^T\mathbf{X})/var(f_i^T\mathbf{Z})$ is maximized. Under the Gaussian patch and noise assumptions, the SNR can simply be represented as follows:

$$SNR_i = \frac{f_i^T\Sigma_X f_i}{\lambda^2} \tag{11}$$

Second, to facilitate the following spatial histogram, the number of filters used to describe the image patch must be as small as possible to make the histogram compact. To achieve this goal, we aim to design a filter bank $f_1, \ldots, f_K$ such that the filter responses are statistically uncorrelated. This can be naturally fulfilled by restraining the filters to be mutually orthogonal. Therefore, robustness and compactness can be simultaneously optimized by the K eigenvectors corresponding to the first K largest eigenvalues of the following eigenproblem:

$$\Sigma_X f = \gamma f \tag{12}$$

where γ is the eigenvalue (indicating the SNR) associated with eigenvector f.

This kind of eigenvector is known as "eigenfilter" in the literature. The concept of eigenfilter was initially proposed by Ade [38] in 1983, and has since been widely used in texture analysis [39][40] and object tracking [41]and recognition. A comparative study [42] indicated that the eigenfilter is optimized with respect to image representation but not discrimination. However, our analysis reveals that they are robust to noise and degradation, particularly when used for binary encoding. Many works have proposed extending the basic eigenfilter. Binarized statistical image feature (BSIF) [43] applies independent component analysis after whitened PCA to learn the independent binary codes. PCANet applies eigenfilters to learn the feature maps in the deep architecture [44]. CBFD imposes additional constraints on the code distribution for enhanced compactness. Gabor-PCA filters are learned by PCA on the local patches of Gabor-filtered images [45]. In contrast to these variants of eigenfilters that learn from image patches, we aim to *design the filters through general knowledge of the pixel correlations*.

To reduce the chance of overfitting, our design of the optimal filters begins with the basic assumption that the local image patch is a realization of a *random field*, where the correlation coefficient between adjacent pixel values is ρ and the variance of each pixel is σ^2. Without loss of generality, we can also assume that $\sigma^2 = 1$. The pixel covariance σ_{ij} depends on the distance between pixel locations P_i and P_j. Given the ρ representing the correlation between

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2018.2800008, IEEE Transactions on Pattern Analysis and Machine Intelligence

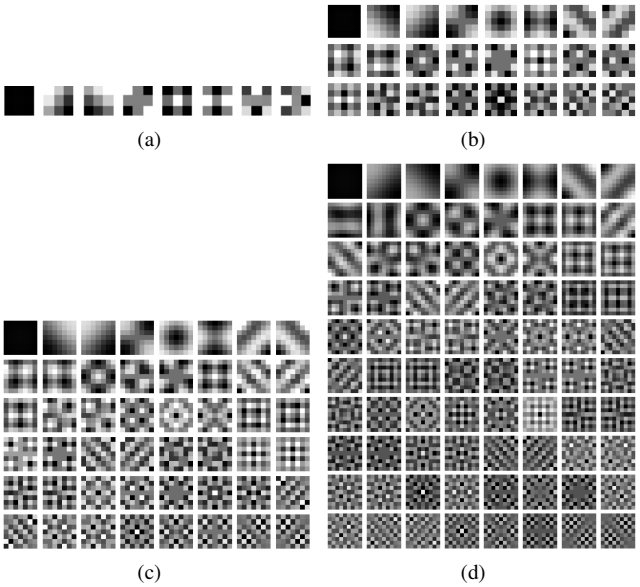


Fig. 1. The eigenvectors (displayed in image form) computed from 3×3, 5×5, 7×7, and 9×9 random fields with a neighboring correlation coefficient of 0.95. The eigenvectors are arranged according to decreasing eigenvalues. Regardless of the size of the random field, the first few eigenfilters display identical primitive structures that are useful for robust and compact feature description.

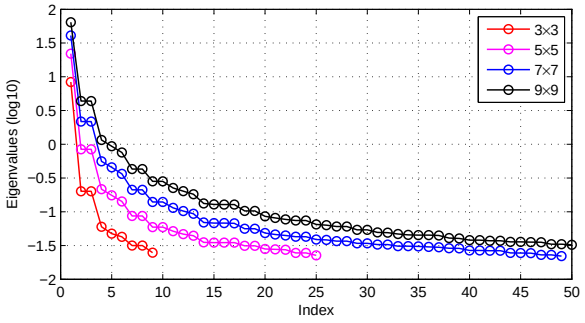


Fig. 2. Eigenvalue spectrum computed from 3×3, 5×5, 7×7, and 9×9 random fields with a neighboring correlation coefficient of 0.95.

neighboring pixels, the pixel covariance matrix can be computed as follows:

$$\Sigma_{ij} = \rho^{\|P_i - P_j\|_2} \tag{13}$$

where $\|\cdot\|$ denotes the L_2 norm. The designed principal components (in image form) of 3×3, 5×5, 7×7, and 9×9 random fields with $\rho = 0.95$ are shown in Fig. 1. These components are ordered by their variance (by column and then by rows). The first few components are composed of a small number of low-frequency components, displaying a certain oriented structure. This result indicates that the lowest spatial frequencies account for the greatest part of the variance in the random field. For decreasing variance, the spatial frequency increases since the spectral power of the natural images decreases with power law of the frequency [46]. Surprisingly, the major eigenfilters exhibit an invariant organization regardless of the filter size: principal component (PC) 1 is a constant-like component. PCs 2 and 3 are rotated versions of the same "edge", PCs 7 and 8 are rotated versions of the same "bar", and PCs 4 and 6 are two versions of the same "wedge". PC 5 is a Gaussian-like "blob". As illustrated in Fig. 2, the corresponding eigenvalue spectra exhibit two plateaus on indices 2–3 and 7–8, where neighboring (orthogonal) PCs with identical eigenvalues span a space modeling the rotation invariance of the random field.

Algorithm 1 Compressive Binary Patterns (CBP)

Input: Input image. The K compressive filters denoted as $F = [f_1, \ldots, f_K] \in \mathbb{R}^{d\times K}$, where f_i is the i-th vectorized compressive filter computed by Eq. (12). The N pre-defined cells by regularly sampling on the image or around landmarks.

Output: The feature vector for the CBP descriptor.

1: **for** every pixel location (u, v) **do**

2: Extract the local patch and normalize the vectorized image patch by

$$\mathbf{X}^{u,v} = \mathbf{X}^{u,v} - m^{u,v}\mathbf{1} \tag{14}$$

where $m^{u,v}$ is the mean value of the vector $\mathbf{X}^{u,v}$ and $\mathbf{1}$ is an all-ones vector.

3: Compute the K responses of the compressive filters on the local patch and convert them to binary code $\mathbf{B}^{u,v}$ by a threshold of zero as

$$\mathbf{B}^{u,v} = \text{sgn}(F^T\mathbf{X}^{u,v}) \tag{15}$$

4: Convert the binary code $\mathbf{B}^{u,v}$ to a decimal number $D^{u,v} \in [0, 2^K - 1]$.

end for

5: **for** $i = 1, \ldots, N$-th cell **do**

6: Count the histogram (denoted as h_i) with 2^K bins of the decimal values within the cell region.

end for

7: Concatenate the histograms of all N cells to form a single output descriptor $H_{CBP} = [h_1, \ldots, h_N]$.

B. Compressive Binary Patterns (CBP)

Motivated by the above optimal design principle, CBP replaces the local derivative filters of LBP with the RF eigenfilters in the image filtering stage and applies the same binary coding and spatial histogram procedure to retain the simplicity and efficiency of the LBP descriptor. *The term "compressive" emphasizes that the RF eigenfilters generate compressive responses to efficiently represent the local image characteristics.* The designed pipeline is illustrated in Fig. 4, and the computational procedure is detailed in Algorithm 1. In this algorithm, the patch mean is subtracted in (14) before filtering to achieve enhanced invariance. The image filtering is conducted by the vector inner product in (15), which aims to detect the patterns in F that are stably preserved during image contamination and degradation. In each preassigned cell, the histogram of binary code, i.e., h_i, approximates the joint distribution of the detected patterns. Finally, a concatenation of these histograms forms the CBP descriptor.

Compared with the commonly used local derivative filters [15], derivative of Gaussian [24], and Gabor-like filters [33], the designed RF eigenfilters have two advantages. First, they are information preserving because they pass most energy through to the following binary coding stage. Hence, CBP encodes sufficient information by a small number of filter responses, resulting in a compact binary code. Second, they are noise resistant because small noise turbulence does not change the sign of the high-amplitude responses. In addition, the degraded images commonly preserve the low-frequency patterns (detected by RF eigenfilters) but lose high-frequency details (detected by derivative filters). In this sense, the CBP descriptor may be robust and invariant to image degradations, although we only justified its optimality under restrictive Gaussian assumptions.

In our implementation, the CBP descriptor adopts $K = 6$ RF eigenfilters of size 7×7, i.e., PCs 2, 3, 4, 6, 7, and 8, as illustrated in Fig. 3. Note that PC 1 and PC 5 are discarded to keep a short coding

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2018.2800008, IEEE Transactions on Pattern Analysis and Machine Intelligence

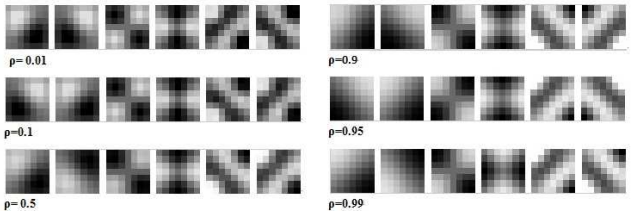


Fig. 3. Visualization of the random-field eigenfilters computed from various correlation coefficient ρ. They consist of nearly identical structures, including two edge filters, two wedge filters, and two bar filters.

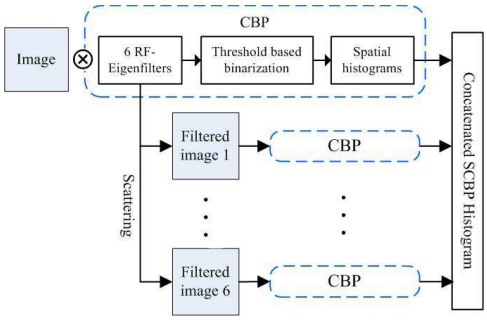


Fig. 4. The designed pipeline of CBP and SCBP descriptors with six primitive filters, where the CBP is a basic module of SCBP.

length because they cannot characterize explicit local structures. The selected eigenfilters are computed in terms of $\rho = 0.95$ in Eq. 13, but note that these leading eigenfilters are (roughly) invariant to ρ, possibly suggesting that they characterize the intrinsic structures of pairwise correlations.

C. Scattering Compressive Binary Patterns (SCBP)

The 6 selected RF eigenfilters are well adapted to detect primitive elements (edge, wedge, and bar), but they may not have sufficient frequency and directional resolution to distinguish fine-grained details of facial structures. A straightforward solution is to apply more eigenfilters with higher frequency and directional resolution, such as the filters shown in the second and third rows of Fig. 1 (b–d). However, the high-frequency eigenfilters generally produce low-amplitude responses, and their signs are easily altered by noise. To avoid introducing these *noise-sensitive eigenfilters*, we apply scattering-like operators [34] to design an enhanced descriptor called SCBP. Using CBP as the basic module, SCBP consists of two layers, where the term "*scattering*" vividly describes the expansion process from a single image (first layer) to a group of feature maps (second layer). The designed pipeline is illustrated in Fig. 4, and the computational procedure is detailed in Algorithm 2.

In our implementation, the first layer convolves $K = 6$ RF eigenfilters on input image, and outputs the family of filtered images (feature maps), as well as the first-layer CBP histogram sequence H_{CBP}. In the second layer, CBP histogram sequences $H_{CBP}^{(1)}, \ldots, H_{CBP}^{(K)}$, are extracted separately from each filtered image using the same filter bank convolutions. In this layer, the filter responses come from the sequential convolution of two filters, and the binary code actually characterizes the co-occurrence of two convolved patterns. As shown in Fig. 5, the second-layer binarized filtered images in (b) characterize richer details than the first-layer ones in (a). Finally, the first-layer and second-layer histogram sequences are concatenated to yield H_{SCBP} that characterizes both the primitive and complex structures of the image.

Note that the two-layer scattering convolutions by K filters of size $L \times L$ are equivalent to convolutions by K^2 filters of larger size of $(2L-1) \times (2L-1)$[1]. However, there are three advantages by using the scattering convolution. Firstly, it effectively controls the complexity of the (equivalently larger-size) filters to reduce the chance of overfitting. Specifically, the second-stage convolution only processes the feature maps passing through the first layer, which has filtered out the high-frequency noise and distortion. As a result, the second-layer encoding can distinguishe the facial details without the risk of matching noisy components. Second, it ensures that the CBPs of second layer are extracted from K uncorrelated filtered images, although the orthogonal property does not hold for the (equivalently larger-size) filters across different CBPs. Finally, the scattering operator reduces the convolution complexity from $(2L-1)^2$ to $\left(1 + \frac{1}{K}\right) L^2$.

In addition to the compact and robust binary code, RF eigenfilter-based binary patterns also benefit from good *code utilization*, making effective use of the available codes to avoid collisions [47]. Because local face patches easily yield conflicted binary code, sufficient code utilization is important to distinguish the fine-grained difference between similar-looking faces. As shown in Fig. 6, handcrafted binary codes are generally unevenly distributed, but SCBP yields evenly distributed binary code. In contrast to the codebook-learning-based method [28], SCBP simply binarizes the responses by a threshold of zero. This indicates that the responses roughly follow a Gaussian-like distribution (or other axis-symmetric distributions) in R^d [47], and each of the K orthogonal RF eigenfilters functions as a hyperplane to divide equally the ensemble of local patches.

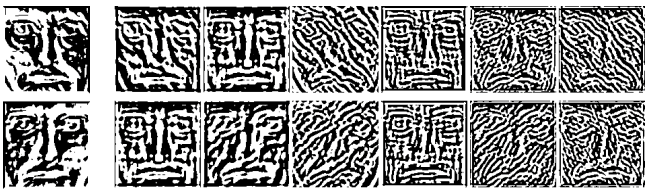
Algorithm 2 Scattering Compressive Binary Patterns (SCBP)

Input: Input image. The K compressive filters. The N pre-defined cells by regular sampling on the image or around landmarks.

Output: The feature vector for the SCBP descriptor.

1: Extract the CBP descriptor of the input images, denoted as H_{CBP}, according to Algorithm 1.
2: Save the K intermediate filtered images before binarization.
3: **for** $i = 1, \ldots, K$-th filtered image **do**
4: Extract the CBP descriptor of the i-th filtered image, denoted as $H_{CBP}^{(i)}$, according to Algorithm 1.
end for
5: Concatenate the $K+1$ CBP descriptors to form a single SCBP descriptor.

$$H_{SCBP} = \left[H_{CBP}, H_{CBP}^{(1)}, \ldots, H_{CBP}^{(K)}\right]. \quad (16)$$



(a) (b)

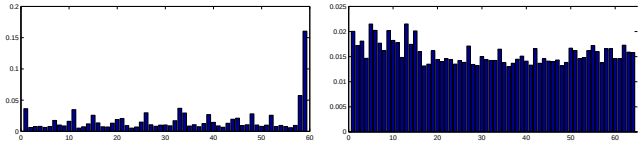
Fig. 5. Binarization of the filtered images of (a) first layer and (b) second layer of SCBP.

IV. Experiments

In this section, we evaluate the effectiveness and robustness of the proposed CBP/SCBP using the FERET [35], LFW [36], and PaSC [37] databases.

[1]We would like to thank the anonymous reviewer who indicates this property.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2018.2800008, IEEE Transactions on Pattern Analysis and Machine Intelligence



(a) Uniform LBP (b) SCBP

Fig. 6. The bin distributions of (a) uniform LBP and (b) SCBP counted in the 1196 FERET gallery images. The 59 bins of uniform LBP are unevenly distributed, but the 64 bins of SCBP are evenly distributed.

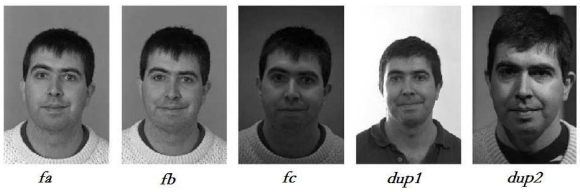


Fig. 7. Example images of different subsets of the FERET database.

A. Comparison with Other Nonstatistical Face Descriptors

The first experiment evaluates the discriminative power of the RF eigenfilters for face description. Our experiment follows the standard FERET data partitions: *fa (gallery)* set, *fb* probe set taken with alternative expressions, *fc* probe set taken under different lighting conditions, *dup1* probe set taken at different times, and *dup2* probe set taken at least a year later. As shown in Fig. 7, fb and fc sets contain only a single-source variation in expression or illumination, but dup1 and dup2 sets are more difficult because they involve blended variations in expression, illumination, makeup, and facial shape as time passes. Facial images are first aligned by two eye centers and then normalized to a size of 128×128 for studying feature descriptors. Following the criterion in [17][48], the block-weighted histogram intersection is applied to measure the distance between facial images.

We first evaluate the effectiveness of RF eigenfilters by replacing them with the same number of PCA-learned and ICA-leaned filters in the CBP descriptor. These filters are learned from 50000 random patches of FERET training images, as detailed in [43]. The results in Table I show that the proposed RF eigenfilters perform better than the PCA-learned filters, followed by the ICA-learned filters. The worse performance may because that the learned filters are easily fit to harmful variance from image noise or intra-class variations. Table II further compares our methods with other handcrafted descriptors. In addition to other previous LBP variants, CBP also outperforms the recently proposed DCP by 3%–4% on average, although its feature size, i.e., 64 histogram bins, is only one eighth of DCP, i.e., 512 bins, which suggests that the RF eigenfilters are more compressive and discriminative than the dual-cross patterns for feature description.

Through the scattering operator on CBP, SCBP notably increases the average accuracy from 88% to 92% using a similar feature dimension (CBP with 16×16 non-overlapped cells and SCBP with 8×8 non-overlapped cells), clearly showing the discriminative power

TABLE I
COMPARISON OF FERET RECOGNITION RATES (%) OF DIFFERENT CBP DESCRIPTORS USING DIFFERENT EIGENFILTERS

Filters	fb	fc	dup1	dup2	avg
PCA learned [38]	96.3	92.3	79.8	77.9	86.6
ICA learned [43]	96.5	91.3	75.8	73.9	84.4
RF eigenfilter	**97.5**	**93.3**	**82.8**	**79.9**	**88.4**

TABLE II
COMPARISON OF FERET RECOGNITION RATE (%) WITH STATE-OF-THE-ART HANDCRAFTED FEATURE DESCRIPTORS USING WEIGHTED HISTOGRAM INTERSECTION

Descriptor	Dims.	fb	fc	dup1	dup2	avg
LBP [15]	2,891	97.0	79.0	66.0	64.0	76.5
LDP [18]	458,752	94.0	83.0	62.0	53.0	73.0
LGBP-M [17]	2,252,800	98.0	97.0	74.0	71.0	85.0
LGBP-P [17]	2,252,800	96.0	94.0	72.0	69.0	82.6
GV-LBP-M [48]	105,600	98.1	98.5	80.9	81.2	89.7
GV-LBP-P [48]	105,600	97.9	**99.0**	81.9	83.8	90.7
DCP [24]	131,072	97.4	79.4	80.3	80.3	84.4
MD-DCP [24]	131,072	98.2	98.5	83.7	83.3	90.9
CBP	16,384	97.5	93.3	82.8	79.9	88.4
SCBP	28,672	**98.9**	**99.0**	**85.2**	**85.0**	**92.0**

of the concatenated second-layer histograms on the fine-grained structures. Note that small cell size is important to CBP because RF eigenfilters cannot capture fine-grained details. Its accuracy reduces to 80% with 8×8 cells. SCBP performs the best on all four probe sets compared to the other descriptors, including the very-high-dimensional descriptors extracted from 4-directional gradient images (MD-DCP) and 40 Gabor-filtered images (LGBP). Note that SCBP encodes the joint distribution of the six responses of the eigenfilter bank, whereas MD-DCP, LGBP and GVLBP encode the filtered images individually. The higher accuracy suggests that the coding scheme of SCBP not only reduces dimensionality but also encodes the co-occurrence of filtering responses, which is more important for recognition.

Due to the blended variations of the duplicate sets, the accuracy of many descriptors drops severely. In contrast, the proposed descriptors obtain relatively stable performance on them. This clearly shows that 1) RF eigenfilters are robust to complex image variations, not only for Gaussian noise. 2) scattering architecture can characterize informative features at a finer scale, and at the same time, retain the robustness of the descriptor. It is possible that the stable performance comes from the large filter size, since the scatter convolution of two 7×7 filters is equivalent to a convolution by a single 13×13 filter. To test this possibility, we enlarge RF eigenfilters from 7×7 to 13×13 for CBP, but the accuracy is severely reduced by more than 10%, which indicates that large-size filters miss some discriminative localized structures. Concatenating two CBPs with 7×7 and 13×13 filters to form a 16,384×2 dimensional feature (with higher dimension than SCBP) only yields an accuracy about 89%. These results suggest the scattering architecture provides a distinctive enhancement for recognition, rather than just benefiting from large filter size.

B. Comparison with the State-of-the-Art Face Descriptors

This experiment evaluates whether the proposed method can generalize well to the web-collected LFW database [36], which contains more than 13000 face images of 5749 subjects with various expressions, ages, illuminations, resolutions, and backgrounds. Our experiment is conducted under an image-restricted setting with label-free outside data [36]. We first crop LFW-a aligned images into 150×130, and then extract CBP by 16×16 and SCBP by 8×8 nonoverlapped cells. We also implement three typical learning-based face descriptors by following the alignment and parameter settings reported in their original papers. Among them, Fisher vector face [49] applies dense SIFT to extract informative features and encodes both the first- and second-order quantities of the GMM codebook. DFD

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2018.2800008, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE III
COMPARATIVE LFW PERFORMANCE OF DIFFERENT FACE DESCRIPTORS UNDER THE IMAGE RESTRICTED SETTING USING THREE WIDELY USED LEARNED METRICS. THE DIMENSION AND RUN TIME (MS) ARE ALSO PRESENTED.

Methods	Dim.	Time	CSML [50]	S-SML [51]	DDML [52]
Fish.Vec [49]	67,584	3533	0.8776	0.8834	0.8897
DFD [27]	50,176	1432	0.8482	0.8464	0.8572
CBFD [28]	32,000	254	0.8634	0.8712	0.8732
CBP	16,384	84	0.8408	0.8412	0.8460
SCBP	28,672	564	**0.8812**	**0.8868**	**0.8932**

TABLE IV
COMPARISONS OF THE MEAN VERIFICATION RATE AND STANDARD ERROR (%) WITH THE STATE-OF-THE-ART RESULTS ON LFW UNDER THE IMAGE RESTRICTED SETTING

Methods	10-Fold Accuracy
V1-like/MKL [53]	0.7935 ± 0.0055
MRF-MLBP [54]	0.7908 ± 0.0014
Fisher vector faces [49]	0.8747 ± 0.0149
Eigen-PEP [55]	0.8897 ± 0.0132
Single LE + holistic [56]	0.8122 ± 0.0053
LBP + CSML [50]	0.8557 ± 0.0052
LARK supervised [57]	0.8510 ± 0.0059
DML-eig SIFT [58]	0.8127 ± 0.0230
Pose Adaptive Filter [59]	0.8777 ± 0.0051
OCBP+TSML [60]	0.8710 ± 0.0043
PCANet [44]	0.8628 ± 0.0110
Gabor-PCA [45]	0.8863 ± 0.0140
DFD [27]	0.8402 ± 0.0140
CBFD [28]	0.8757 ± 0.0143
Supervised-DAE [61]	0.8702 ± 0.0183
Convolutional-DBN [62]	0.8777 ± 0.0062
CBP	0.8460 ± 0.0167
SCBP	**0.8932 ± 0.0134**

[27] and CBFD aim to learn region-specific filters to extract features and learn a k-means codebook to encode the long binary code.

Following common practice, we first apply PCA to reduce these features to 300 dimensions and then use three popular methods [50][51][52] to learn a distance metric to compute the similarity of each face pair. Cosine similarity metric learning (CSML) aims to learn a metric space in which cosine similarity performs well for verification [50]. Sub-SML learns the metric by solving a convex optimization problem [51]. Discriminative deep metric learning (DDML) [52] learns a set of hierarchical nonlinear transformations to project face pairs into the same feature subspace. The final comparative performances are shown in Table III along with the feature dimensions and extraction run times. Under all three tested metrics, the proposed SCBP descriptor obtains test performance using the lowest feature dimension. Although learning-based descriptors are commonly preferred, SCBP demonstrates that handcrafted descriptors can achieve competitive performance by considering the robustness of the designed filter and the distinctiveness of the scattering architecture.

Table IV compares our method with other face verification methods in terms of the performance reported in the original papers, which also shows that SCBP achieves better accuracy than many face descriptors with complicated parameter tuning. Some deep-learning based descriptors have been tested on this restricted protocol (where outside training data is not allowed). For example, a latest auto-encoder based method called class sparsity based supervised encoder [61] obtains 0.87 accuracy (without ensembles), and the local convolutional restricted Boltzmann machines (RBMs) [62] reports 0.8777 accuracy. Their performance is worse than SCBP, although they potentially learn deep representations that capture higher-order statistics than hand-crafted image descriptors. The off-the-shelf VGG face descriptor [63] can yield much higher accuracy, but it violates the restricted protocol by using millions of labeled outside training data. The results clearly suggest that *although the optimality of RF eigenfilter is derived under the constrained Gaussian assumptions, it indeed generalizes well on the real-world complex conditions*. Fig. 8 illustrates typical image pairs from the error cases, and they are mostly caused by large variations such as pose, occlusion, and makeup.

Fig. 8. Examples of the error cases of our method, where 'FP' indicates the false positive pair and 'FN' indicates the false negative pair.

C. Extended Evaluation of the Robustness of Face Descriptors

This experiment evaluates the robustness of the descriptors by extending the FERET evaluation with synthetic noise and degradation. For clarity, we express the interference of face recognition $\eta = \eta_f + \eta_q$ [64], where η_f denotes facial variations such as misalignment, expression, illumination, and age and η_q denotes the image variation due to sensor or coding-related issues, such as Gaussian noise, blur, compression, and low resolution. Most studies on the FERET database focused only on the effect of η_f, whereas our extended experiments study both the pure effect of η_q and the superposed interference of $\eta_f + \eta_q$. For a comprehensive study, we synthesize four types of noise or degradations that are most common in real-world systems but that have not appeared in the standard databases.

Specifically, we generate the following versions of probe sets: 1) five levels of *Gaussian noise*. The images are normalized in the range of $(0, 1)$, and then we apply additive Gaussian noise with zero mean and standard derivations of $\sigma = 0, 0.01, 0.02, 0.03, 0.04, 0.05$; 2) four different *Gaussian blur* sets of gallery and four probes using a Gaussian kernel of size 10×10 with $\sigma = \{2, 4, 6, 8\}$; 3) four different *compressed* images using MATLAB's JPEG codec of quality 60, 45, 30, and 15; and 4) four different *low-resolution* sets of test images by first downsampling the images by ratios of 2, 3, 4, and 5 and then interpolating them to the original resolution by the "nearest" method in MATLAB. Example probe images are shown in Fig. 9, and as shown in this figure, these degraded faces are recognizable by humans and are very common in real-world surveillance scenarios. Therefore, it is important to study how the accuracy of the face descriptor changes under these degradations.

For comparison purposes, we also implement several commonly used local descriptors: LBP [15], DCP [24], MD-DCP [24], NRLBP

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2018.2800008, IEEE Transactions on Pattern Analysis and Machine Intelligence

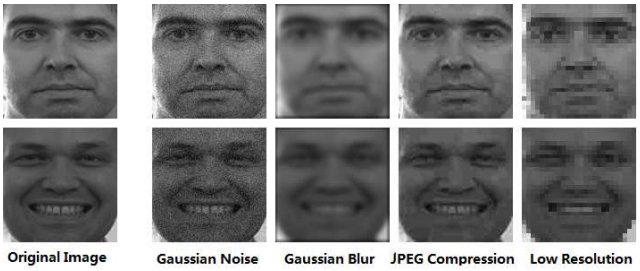


Fig. 9. Examples of original and degraded images used in our extended FERET evaluation. The last four columns correspond to the most severe degrees of Gaussian noise, Gaussian blur, JPEG compression, and reduced resolution applied on the probe images.

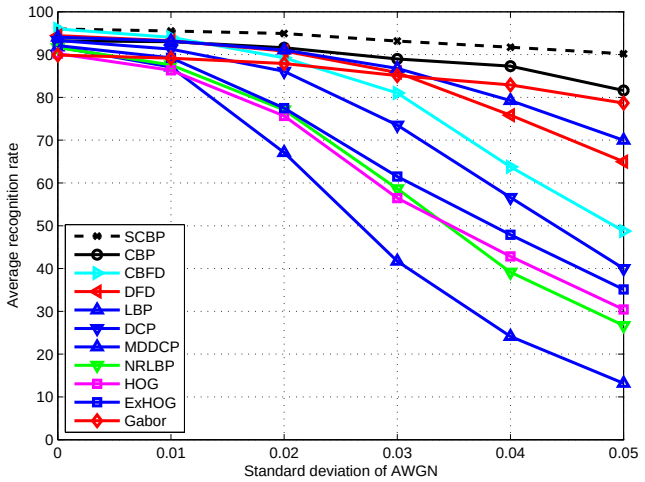


Fig. 10. Comparative FERET performance of face descriptors as a function of the standard deviation of additive Gaussian white noise. The average accuracy across the four probe sets is reported.

[31], HOG [65], ExHOG [66], Gabor [67], DFD [27], and CBFD [28]. Specifically, the LBP descriptor adopts the $LBP^{U2}_{8,2}$ operator [15] in 16×16 cells of 59 bins, resulting in a 15,104 (16×16×59)-dimensional feature vector.The DCP descriptor is 131,072 dimensional with 16×16 cells of 512 bins. The MD-DCP descriptor is 131,072 dimensional with 8×8 cells on the 4 filtered images. The uncertainty threshold of NRLBP is empirically set to $t = 0.5\sigma$ for the images contaminated by the Gaussian noise of standard deviation σ. The HOG [65] descriptor first divides the image into multiple 16×16 cells, and a local histogram of 18 signed gradient directions over the pixels of the cell are accumulated for each cell. "L2-Hys" contrast normalization with a threshold of 0.2 is applied over each block of 2×2 cells. The combined histogram entries form the final 4,608 (16×16×18)-dimensional feature vector. ExHOG doubles the number of bins of the HOG histogram to enhance robustness [66]. The 10,240-dimensional Gabor feature [67], 50,176-dimensional DFD [27] and 32,000-dimensional CBFD [28] are extracted according to their original papers. All the tested descriptors are extracted from the same aligned face images and used for face identification using parameter-free linear regression analysis [68].

Fig. 10 presents the recognition accuracy of ten descriptors as a function of the standard deviation of Gaussian noise. This figure shows severe performance deterioration with increasing noise, which suggests that the descriptors are more sensitive to the superposed noise with real-world variations. As expected, the traditional LBP descriptor performs the worst across various noisy conditions. N-RLBP largely improves LBP by the error-correction encoding with an increase of 10–20% accuracy observed in Fig. 10. DCP indeed enhances the accuracy of LBP by approximately 30% with its local sampling of dual-cross patterns in a large neighborhood, and MD-DCP further improves the noise robustness by the first derivative of the Gaussian operator [24]. The quantized gradient orientation of HOG appears to be less sensitive to noise than the thresholded derivative of LBP, and ExHOG further improves the robustness to some extent by doubling the histogram bins.

Unfortunately, these handcrafted improvements are not sufficient to handle the probe image with severe noise, and their performance begins to decrease when the noise $\sigma > 0.02$. When the noise $\sigma > 0.04$, Although its performance is common on the original image, the downsampled Gabor feature outperforms all other previously proposed handcrafted and learning-based descriptors, which clearly supports the filter based approach for robust face descriptor. In general, CBP achieves much better accuracy than LBP/NRLBP/DCP, clearly validating the robustness of the RF eigenfilters. This robustness is further enhanced by the scattering operator. Compared with Gabor feature, CBP and SCBP are more discriminative to the original image. As the noise level increases, the relative performance gain of the SCBP descriptor over the others become increasingly more significant. It can also be observed that the accuracy loss of SCBP is less than CBP. This observation indicates that the second-layer encoding is very robust to the image noise by focusing only on the low-frequency feature maps generated by RF eigenfilter.

Table V shows that SCBP and CBP exhibit much better robustness than the other descriptors under image blur, compression, and reduced resolution. Although descriptors with Gabor filtering or directional derivative of Gaussian filtering (MD-DCP and LGBP) also exhibit a certain degree of robustness, their absolute accuracy is notably lower than that of SCBP, lacking sufficient distinctiveness. On the original FERET probe sets, the accuracy difference among the four best descriptors, i.e., SCBP, MD-DCP, DFD, and CBFD, is approximately only 1%–2%, which all show high distinctiveness. On the severely blurred and reduced resolution probe sets, however, the accuracy gap dramatically increases to 70%–80%. It is possible that their discriminative objectives result in noise-sensitive filters. For example, CBFD learns as many as 15 local filters for binary coding in each cell. To optimize three joint objective functions, approximately half of the CBFD-learned filters characterize the high-frequency components that are easily *overfitting* the noise and distortion, resulting in noise-sensitive binary codes. In contrast, *although the optimality of RF eigenfilter is derived under the constrained Gaussian noise, it generalizes well on various types of image degradations.* The scattering architecture of SCBP naturally achieves a balance between distinctiveness and robustness. Although the off-the-shelf VGG deep learning descriptor yields the best accuracy, its robustness to image blur and reduced resolution is still worse than our methods.

D. Digital point and shoot camera images

The final experiment evaluates the feasibility of the proposed method on real-world unconstrained degraded conditions using the Point and Shoot Face Recognition Challenge (PaSC) database [37]. The PaSC database contains both still images and videos. The images and videos were taken using digital point and shoot cameras, particularly handheld cameras found in cell phones. The still image portion consists of 9,376 images of 293 people. These still images were taken at nine locations, both inside buildings and outdoors, with five point-and-shoot still cameras. As illustrated in Fig. 11, since the images were taken at a variety of poses and distances from the camera, they show low image quality due to blurring and low resolution.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2018.2800008, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE V
COMPARATIVE RECOGNITION RATES (%) OF EXTENDED FERET EVALUATION ON THE ROBUSTNESS TO THE THREE TYPES OF COMMON DEGRADATIONS. ACCURACY LOSS OF EACH DEGRADATION DEGREE ON EACH PROBE SET IS REPORTED IN DETAIL.

Feature	Basic Accuracy[1]	Gaussian Blur				JPEG Compression				Reduced Resolution				**Summarized Accuracy[2]**
		2	4	6	8	60	45	30	15	1/2	1/3	1/4	1/5	
LBP [15]	91.8	-3.7	-18.0	-36.1	-52.0	-5.0	-8.4	-15.9	-43.9	-2.4	-47.4	-86.3	-90.1	57.7 (-34.1)
DCP [24]	93.3	-2.1	-10.9	-30.5	-48.3	-1.7	-2.4	-6.0	-19.8	-1.5	-4.1	-36.6	-64.5	74.3 (-19.0)
MD-DCP [24]	95.9	-3.4	-8.7	-17.0	-29.4	-1.1	-1.9	-4.0	-13.6	-1.5	-6.0	-20.7	-38.9	83.7 (-12.2)
HOG [65]	90.2	-3.1	-11.8	-26.0	-43.1	-3.8	-5.8	-10.3	-30.5	-6.0	-24.3	-54.5	-69.9	66.1 (-24.1)
ExHOG [66]	92.1	-1.7	-8.8	-22.4	-38.7	-2.1	-4.4	-10.4	-30.4	-4.5	-25.3	-52.1	-68.6	69.7 (-22.4)
Gabor [67]	89.9	-5.2	-12.3	-20.5	-30.1	-1.4	-2.5	-4.6	-13.9	-2.2	-8.7	-24.9	-46.2	75.5 (-14.4)
LGBP [17]	96.1	-2.7	-6.6	-14.7	-28.6	-0.9	-1.7	-3.3	-9.3	-1.3	-5.1	-16.6	-43.4	84.9 (-11.2)
DFD [27]	94.7	-4.3	-17.2	-70.5	-91.3	-1.3	-3.3	-6.5	-27.6	-1.3	-8.7	-59.7	-87.4	63.1 (-31.6)
CBFD [28]	96.0	-0.3	-3.7	-38.0	-71.7	-2.5	-4.3	-8.8	-39.0	-1.9	-15.4	-72.9	-91.0	66.9 (-29.1)
VGG-face [63]	**97.8**	-0.2	-1.3	-8.1	-27.5	-0.3	-0.7	-2.1	-9.2	-0.3	-1.5	-7.5	-23.7	90.9 (-6.9)
CBP	93.2	-0.2	-2.0	-10.2	-33.7	-0.1	-0.8	-2.1	-12.6	-0.6	-1.8	-7.1	-17.1	85.9 (-7.3)
SCBP	96.7	-0.2	-0.7	-5.4	-21.6	-0.4	-0.4	-2.1	-11.0	0.0	-1.3	-7.0	-15.7	**91.2 (-5.5)**

To provide a comprehensive result, the average accuracy across the four types of probe sets is reported.
[1] The average accuracy on the original FERET data set.
[2] The average accuracy across all types and all degrees of the tested degradations.

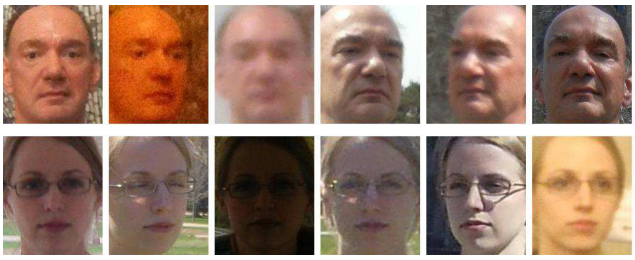


Fig. 11. Example images of PaSC database with real-world degradations by weak lighting, motion blur, poor focus, and low resolution.

TABLE VI
VERIFICATION RATES AT FAR OF 0.01 ON THE PASC STILL-TO-STILL MATCHING DATABASE

Algorithm	Frontal Only	Full Database
CohortLDA	0.22	0.08
LRPCA	0.19	0.10
PittPatt (Commercial)	0.55	0.41
L-CSSE [61]	0.61	0.54
VGG-Face [63]	0.77	0.72
HD-SCBP	0.64	0.56
HD-SCBP+VGG	**0.82**	**0.76**

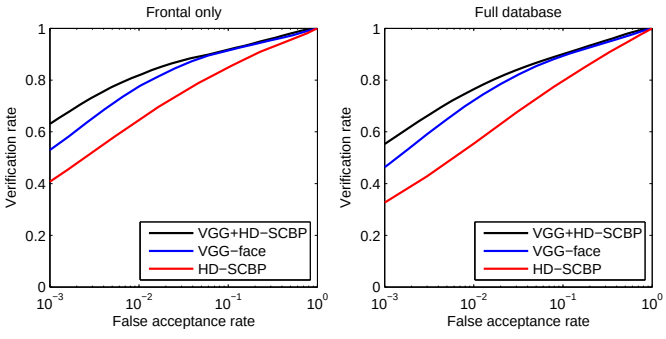


Fig. 12. ROC curves on PaSC still-to-still matching protocol. (a) Frontal only images. (b) Full database.

To design an informative descriptor for this in-the-wild task, we borrow an idea from the work of "bless of dimensionality" [23]. It has been observed that multi-scale facial features extracted both locally (patch based) and holistically (full face) help to jointly encode high-dimensional discriminative features. Specifically, we first preprocess the face as in [69], and then the aligned image are resized to three scales, where the side lengths of the image are 180, 128, and 90. In the 3 scaled images, local patches at 22 facial landmarks predefined in [23] are cropped with a fixed size of 32×32. Each patch is divided into 2×2 non-overlapped cells to characterize local-level features. At the same time, each of the three scaled images is divided into 8×8 non-overlapped cells to characterize holistic-level features. Finally, we concatenate the SCBP descriptors for encoding each cell to form a high-dimensional SCBP (HD-SCBP) for face descriptors. The dimensions of the features are reduced to 500 by PCA for joint Bayesian learning [70], which seeks a metric space where the inter-class and intra-class differences are best separated. Both the PCA and joint Bayesian models were trained on the LFW and FRGC databases.

As shown in Table VI, HD-SCBP yields the second best verification accuracy, which is much better than the two baseline algorithms CohortLDA and LRPCA. On the frontal only images, HD-SCBP yields a 0.64 verification rate at a 0.01 false acceptance rate, whereas the recently proposed fusion of supervised deep auto-encoders (called L-CSSE in [61]) yields 0.61 and the commercial matcher PittPatt yields 0.55. Similar improvements are also observed on the full database, where HD-SCBP yields 0.55 and PittPatt provides a 0.41 verification rate. Our method is also comparable to the off-the-shelf VGG-face descriptor [63], which is based on a deep CNN pre-trained by millions of face images. Moreover, we have also observed that our handcrafted HD-SCBP has a certain complementary effect to the deep-CNN-based VGG feature because simply adding the cosine similarity of these two features improves the verification accuracy by approximately 4–5%. The ROC curves are shown in Fig. 12.

V. CONCLUSIONS

A number of conclusions can be drawn from the experiments:

1. The proposed RF eigenfilters, designed from the neighborhood correlation between image pixels, are efficient and robust for characterizing facial texture, although their optimality is justified only under restrictive Gaussian assumptions. By simply replacing the local derivative filters with the RF eigenfilters, CBP significantly improves the robustness of LBP.

2. The scattering-like architecture provides a simple paradigm to design the descriptor with both distinctiveness and robustness. Although designed by only six predefined RF eigenfilters, the proposed SCBP achieves comparable accuracy on the FERET, LFW, and PaSC databases with other state-of-the-art face descriptors.
3. The negative effects of image noise and degradation may be underestimated for the applicability of face descriptors. Low-level descriptors such as LBP, DCP and HOG tend to break down under a moderate degree of image degradation because high-frequency elements such as local derivative and gradient orientation are highly unstable.
4. The commonly preferred learning-based descriptors, such as DFD and CBFD, tend to derive noise-sensitive filters by adapting to fine-grained structures. In contrast, our designed RF eigenfilters with a scattering structure, which focus on the low-frequency components of images, exhibit considerably better robustness. Additionally, the handcrafted SCBP can outperform learning-based descriptors with an average margin of 20%–30% accuracy on degraded probe images.
5. When training samples are limited, SCBP can outperform the up-to-date deep auto-encoder based descriptors, and obtain better robustness than CNN based features under severe image degradations. On the challenging scenario on PaSC database, SCBP based high dimensional descriptor demonstrates complementary effects to the VGG face descriptor learned from millions of training samples.

By restricting our design to be simple, we have shown that the SCBP descriptor handcrafted by 6 RF eigenfilters is sufficient to achieve accurate and robust performance. Naturally, adopting an increased number of filters with some learning and regularization techniques would probably enhance the performance. The balance of designed robustness and learning-based adaptation is the major issue for our future work on deriving an optimized descriptor that combines distinctiveness, robustness, and compactness.

VI. Acknowledgements

This work was partially supported by the National Natural Science Foundation of China under Grant Nos. 61573068, 61471048, 61375031, and 61532006, and Beijing Nova Program under Grant No. Z161100004916088.

References

[1] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 5, pp. 815–830, 2010.

[2] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[4] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 5, pp. 978–994, 2011.

[5] R. Arandjelović and A. Zisserman, "Three things everyone should know to improve object retrieval," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2911–2918.

[6] R. Arandjelovic and A. Zisserman, "All about vlad," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2013, pp. 1578–1585.

[7] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2. Ieee, 1999, pp. 1150–1157.

[8] K. Van De Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1582–1596, 2010.

[9] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Computer vision–ECCV 2006*, pp. 404–417, 2006.

[10] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "Brief: Computing a local binary descriptor very fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1281–1298, 2012.

[11] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4353–4361.

[12] I. Melekhov, J. Kannala, and E. Rahtu, "Image patch matching using convolutional descriptors with euclidean distance," in *Asian Conference on Computer Vision*. Springer, 2016, pp. 638–653.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[15] T. Ahonen, A. Hadid, and M. Pietikinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, 2006.

[16] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recognition*, vol. 62, pp. 135–160, 2017.

[17] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," in *ICCV*, vol. 1. IEEE, 2005, pp. 786–791.

[18] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor," *Image Processing, IEEE Transactions on*, vol. 19, no. 2, pp. 533–544, 2010.

[19] X. Jiang, "Extracting image orientation feature by using integration operator," *Pattern Recognition*, vol. 40, no. 2, pp. 705–717, 2007.

[20] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li, "Learning multi-scale block local binary patterns for face recognition," in *Advances in Biometrics*. Springer, 2007, pp. 828–837.

[21] L. Wolf, T. Hassner, and Y. Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 33, no. 10, pp. 1978–1990, 2011.

[22] N.-S. Vu and A. Caplier, "Enhanced patterns of oriented edge magnitudes for face recognition and image matching," *Image Processing, IEEE Transactions on*, vol. 21, no. 3, pp. 1352–1365, 2012.

[23] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *CVPR*, 2013, pp. 3025–3032.

[24] C. Ding, J. Choi, D. Tao, and L. Davis, "Multi-directional multi-level dual-cross patterns for robust face recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 38, no. 3, pp. 518–531, 2016.

[25] C. H. Chan, M. A. Tahir, J. Kittler, and M. Pietikainen, "Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 35, no. 5, pp. 1164–1177, 2013.

[26] S. U. Hussain, T. Napoléon, and F. Jurie, "Face recognition using local quantized patterns," in *British Machive Vision Conference*, 2012, pp. 11–pages.

[27] Z. Lei, M. Pietikainen, and S. Z. Li, "Learning discriminant face descriptor," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 36, no. 2, pp. 289–302, 2014.

[28] J. Lu, V. E. Liong, X. Zhou, and J. Zhou, "Learning compact binary face descriptor for face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 10, pp. 2041–2056, 2015.

[29] T. Ahonen and M. Pietikäinen, "Soft histograms for local binary patterns," in *Proceedings of the Finnish signal processing symposium, FINSIG*, vol. 5, 2007, p. 1.

[30] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.

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

[32] J. Ren, X. Jiang, and J. Yuan, "Lbp encoding schemes jointly utilizing the information of current bit and other lbp bits," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2373–2377, 2015.

[33] T. Ahonen and M. Pietikäinen, "Image description using joint distribu-

tion of filter bank responses," *Pattern Recognition Letters*, vol. 30, no. 4, pp. 368–376, 2009.

[34] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 35, no. 8, pp. 1872–1886, 2013.

[35] P. J. Phillips, H. Moon, P. Rizvi, and P. Rauss, "The feret evaluation method for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 0162–8828, 2000.

[36] G. B. Huang and E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," *Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep*, pp. 14–003, 2014.

[37] J. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer *et al.*, "The challenge of face recognition from digital point-and-shoot cameras," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. IEEE, 2013, pp. 1–8.

[38] F. Ade, "Characterization of textures by eigenfilters," *Signal Processing*, vol. 5, no. 5, pp. 451–457, 1983.

[39] M. Unser, "Local linear transforms for texture measurements," *Signal processing*, vol. 11, no. 1, pp. 61–79, 1986.

[40] T. Aach, A. Kaup, and R. Mester, "On texture analysis: Local energy transforms versus quadrature filters," *Signal processing*, vol. 45, no. 2, pp. 173–181, 1995.

[41] F. De la Torre, J. Vitria, P. Radeva, and J. Melenchon, "Eigenfiltering for flexible eigentracking (efe)," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 3. IEEE, 2000, pp. 1106–1109.

[42] T. Randen and J. H. Husoy, "Filtering for texture classification: A comparative study," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 21, no. 4, pp. 291–310, 1999.

[43] J. Kannala and E. Rahtu, "Bsif: Binarized statistical image features," in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 1363–1366.

[44] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, 2015.

[45] C.-Y. Low, A. B.-J. Teoh, and C.-J. Ng, "Multi-fold gabor, pca and ica filter convolution descriptor for face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[46] B. A. Olshausen *et al.*, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[47] M. A. Carreira-Perpinán and R. Raziperchikolaei, "Hashing with binary autoencoders," in *CVPR*, 2015, pp. 557–566.

[48] Z. Lei, S. Liao, M. Pietikäinen, and S. Z. Li, "Face recognition by exploring information jointly in space, scale and orientation," *Image Processing, IEEE Transactions on*, vol. 20, no. 1, pp. 247–256, 2011.

[49] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild." in *BMVC*, vol. 5, no. 6, 2013, p. 11.

[50] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in *Asian Conference on Computer Vision*. Springer, 2010, pp. 709–720.

[51] Q. Cao, Y. Ying, and P. Li, "Similarity metric learning for face recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2408–2415.

[52] J. Hu, J. Lu, and Y.-P. Tan, "Discriminative deep metric learning for face verification in the wild," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[53] N. Pinto, J. J. DiCarlo, and D. D. Cox, "How far can you get with a modern face recognition test set using only simple features?" in *CVPR*. IEEE, 2009, pp. 2591–2598.

[54] S. R. Arashloo and J. Kittler, "Efficient processing of mrfs for unconstrained-pose face recognition," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. IEEE, 2013, pp. 1–8.

[55] H. Li, G. Hua, X. Shen, Z. Lin, and J. Brandt, "Eigen-pep for video face recognition," in *Computer Vision–ACCV 2014*. Springer, 2014, pp. 17–33.

[56] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *CVPR*. IEEE, 2010, pp. 2707–2714.

[57] H. J. Seo and P. Milanfar, "Face verification using the lark representation," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 4, pp. 1275–1286, 2011.

[58] Y. Ying and P. Li, "Distance metric learning with eigenvalue optimization," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 1–26, 2012.

[59] D. Yi, Z. Lei, and S. Li, "Towards pose robust face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3539–3545.

[60] L. Zheng, K. Idrissi, C. Garcia, S. Duffner, and A. Baskurt, "Triangular similarity metric learning for face verification," in *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, vol. 1. IEEE, 2015, pp. 1–7.

[61] A. Majumdar, R. Singh, and M. Vatsa, "Face verification via class sparsity based supervised encoding," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1273–1280, 2017.

[62] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *CVPR*. IEEE, 2012, pp. 2518–2525.

[63] O. M. Parkhi, A. Vedaldi, A. Zisserman *et al.*, "Deep face recognition." in *BMVC*, vol. 1, no. 3, 2015, p. 6.

[64] R. Gopalan, S. Taheri, P. Turaga, and R. Chellappa, "A blur-robust descriptor with applications to face recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 34, no. 6, pp. 1220–1226, 2012.

[65] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, vol. 1. Ieee, 2005, pp. 886–893.

[66] A. Satpathy, X. Jiang, and H.-L. Eng, "Human detection by quadratic classification on subspace of extended histogram of gradients," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 287–297, 2014.

[67] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 467–476, 2002.

[68] W. Deng, J. Hu, X. Zhou, and J. Guo, "Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning," *Pattern Recognition*, vol. 47, no. 12, pp. 3738–3749, 2014.

[69] W. Deng, J. Hu, Z. Wu, and J. Guo, "Lighting-aware face frontalization for unconstrained face recognition," *Pattern Recognition*, vol. 68, pp. 260–271, 2017.

[70] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: A joint formulation," *Computer Vision–ECCV 2012*, pp. 566–579, 2012.