BIRD: Learning Binary and Illumination Robust Descriptor for Face Recognition

Zhuo Su[1]
zhuo.su@oulu.fi
Pietikäinen Matti[1]
matti.pietikainen@oulu.fi
Li Liu[*,2,1]
li.liu@oulu.fi

[1] Center for Machine Vision and Signal Analysis
University of Oulu
Oulu, Finland

[2] College of System Engineering
National University of Defense Technology, China

Abstract

Recently face recognition has made significantly progress due to the advancement of large scale Deep Convolutional Neural Network (DeepCNNs). Despite the great success, the known deficiencies of DeepCNNs have not been addressed, such as the need for too much labeled training data, energy hungry, lack of theoretical interpretability, lack of robustness to image transformations and degradations, and vulnerable to attacks, which limit DeepCNNs to be used in many real world applications. Therefore, these factors make previous predominating Local Binary Patterns (LBP) based face recognition methods still irreplaceable.

In this paper we propose a novel approach called BIRD (learning Binary and Illumination Robust Descriptor) for face representation, which nicely balances the three criteria: distinctiveness, robustness, and computationally inexpensive cost. We propose to learn discriminative and compact binary codes directly from six types of Pixel Difference Vectors (PDVs). For each type of binary codes, we cluster and pool these compact binary codes to obtain a histogram representation of each face image. Six global histograms derived from six types of learned compact binary codes are fused for the final face recognition. Experimental results on the CAS_PERL_R1 and LFW databases indicate the performance of our BIRD surpasses all previous binary based face recognition methods on the two evaluated datasets. More impressively, the proposed BIRD is shown to be highly robust to illumination changes, and produces 89.5% on the CAS_PEAL_R1 illumination subset, which, we believe, is so far the best reported results on this dataset. Our code is made available [1].

1 Introduction

As a longstanding, fundamental and challenging problem in computer vision and pattern recognition, face recognition has been one of the most extensively studied problem, and numerous approaches have been proposed in the literature. However, there are still major

* indicates the corresponding author.

[1] https://github.com/zhuogege1943/bird-descriptor

unsolved challenges in making face recognition algorithms work efficiently and reliably in real world environments, especially on wearable and embedded devices. As a typical pattern recognition problem, face recognition primarily consists of two critical subproblems: feature representation and classifier designation [16]. It is generally agreed that the extraction of powerful features plays the central role, and consequently numerous methods have been proposed, with excellent surveys [21, 33].

Various face representation methods can be mainly classified into two categories: traditional method [1, 2, 12, 32, 34], and deep learning based methods [25, 27, 28, 29, 30]. Traditional methods consists of two main classes: holistic features (such as Eigenface [32], Fisherface [2]) and local features (such as Gabor [34], Local Binary Pattern (LBP) [1], SIFT [12]). Recently, deep learning techniques, especially the advances in Deep Convolutional Neural Networks (DeepCNNs) have brought extraordinary progress to face recognition. Representative methods include DeepFace [30], DeepID [28, 29], VGGFace [25] and FaceNet [27].

It seems that the popularity of traditional features have been overtaken by DeepCNN features. Surprisingly however, due to their disadvantages such as data hungry and energy hungry, the adoption of DeepCNN features has been limited in many real world applications. Therefore, due to the outstanding advantages such as good discriminative power, invariance to monotonic illumination changes, low computational complexity and not relying on large amount of training data, LBP features for face recognition remain active, as evidenced by a number of recent excellent binary features [7, 8, 9, 10, 22, 23].

Since the seminar work in [1], the LBP methodology has emerged as one of the most prominent technique for face recognition. Many variants of LBP have been proposed to improve robustness and discriminative power, with recent surveys [13, 21]. Nevertheless, most LBP based facial feature representations still suffer from some of the following drawbacks.

(1) Limited representation power of uniform LBPs. Most approaches such as the original LBP [1], Local Ternary Patterns (LTP) [24], and the method in [6] only exploit the uniform LBP patterns for face representation. However, the proportions of uniform patterns may be too small and inadequate to capture the texture characteristics of human faces.

(2) Limited diversity and spatial support in sampling pattern shapes. It is generally prohibitive for handcrafted LBP descriptors to adopt a large sampling size as increasing the size of local neighborhood increases number of LBP patterns exponentially.

(3) Limited representation power of using only one type of binary feature. To improve the representation power of face descriptors, one usual approach is to fuse the information extracted by different features such as using Gabor filter as preprocessing [34]. It achieved improved performance but at the cost of increased computational complexity.

(4) Limited robustness to real world environments (such as serious illumination variations, pose changes, random noise, image blur, etc). There are some efforts to improve the robustness such as NRLBP [26] and MSLPQ [4], the performance is still unsatisfactory.

(5) Requirement of strong prior knowledge. Most existing LBP based face descriptors are handcrafted, which require strong prior knowledge to engineer them by hand.

Noticeably, there has been a series of work [9, 10, 22, 23] aiming at learning a number of hashing functions to obtain compact binary codes for face representation. This series of work overcome some of the shortcomings such as (1) (2) and (5) listed above; however they still have the disadvantages (3) and (4) because they are derived from the difference between each pixel and its neighboring pixels only. This limits the overall information encoding capability of these learned binary codes and prevent them from leveraging all the texture information existing in a local facial patch, thus affects the overall performance of face recognition systems.

In this paper, we propose a novel approach called BIRD, *i.e.*, learning Binary and Illumination Robust Descriptor (BIRD) for face representation, in order to build features that can inherit the advantages of LBP without suffering the above shortcomings. Specifically, our BIRD based face recognition pipeline consists of the following stages.

Firstly, two kinds of local features are considered, one based on intensities and the other on gradients encoding second order discriminative information along the radial and angular directions. Hence, three kinds of PDVs, *i.e.*, Center PDVs (CPDV), Radial PDVs (RPDV) and Angular PDVs (APDV) are obtained. In order to preserve more texture information, for each type of PDV, its magnitude component is further used to obtain an additional PDV, thus resulting in two complementary features. Therefore, six type of local features are obtained which will be separately fed into the next steps for learning compact binary codes. Secondly, for each of the six local features, a feature mapping is learned to project it into a lower dimensional binary vector based on some criteria. Therefore, six types of compact binary codes can be learned. Then, for each type of binary codes, we cluster and pool these compact binary codes to obtain a histogram representation of each face image. Finally, six global histograms derived from six types of learned compact binary codes are fused for the final face recognition, leading to the final BIRD approach.

Our proposed BIRD descriptor, we argue, nicely balances the following three criteria: distinctiveness, robustness, and computationally inexpensive cost. Each individual learned binary feature representation and the final BIRD approach are extensively evaluated on two popular face datasets: CAS_PEAL_R1 [11] and LFW [14, 15]. We find that the representation power of binary codes can be significantly enhanced through learning, and the final BIRD descriptor achieves consistently superior performance than each individual feature. *The performance of our BIRD surpasses all previous binary based face recognition methods on the two evaluated datasets.* More impressively, the proposed BIRD is shown to be highly robust to illumination changes, and produces **89.5%** on the CAS_PEAL_R1 illumination subset, which, we believe, is so far the best reported results on this dataset. Note that the recent CBFD approach published in IEEE TPAMI [22] reports only 67.4% on this dataset.

2 Proposed BIRD Approach

2.1 Proposed Pixel Difference Vectors

Limitations of Existing Binary Features. As we discussed in the Introduction Section clearly, existing binary features for face recognition have serious limitations such as those from (1) to (5) listed in the Introduction Section. Let's take a close look at three recent representative binary features: the original LBP [1], the Dual Cross Pattern (DCP) [8] and the CBFD [22], whose pattern sampling shapes are contrasted in Fig. 1 (a1), (a2) and (a3). LBP have all the drawbacks from (1) to (5), and most LBP based face representation usually adopt the pattern shape shown in Fig. 1 (a1), *i.e.*, only considering the PDVs between each central pixel and its 8 neighboring pixels on a single scale. CBFD [22] overcomes the drawbacks (1) (2) (5) by learning a feature mapping to project each high dimension PDV (a square layout around a central pixel as shown in Fig. 1 (a2)) into a lower dimensional binary vector and obtained improved face recognition performance. In DCP [8], second order discriminative information in the radial directions are exploited, and the resulting PDV is grouped into two subgroups, as shown in Fig. 1 (a3), to derive two LBP features. To address the limitations of existing binary features, we propose to learning binary features from a extended set of

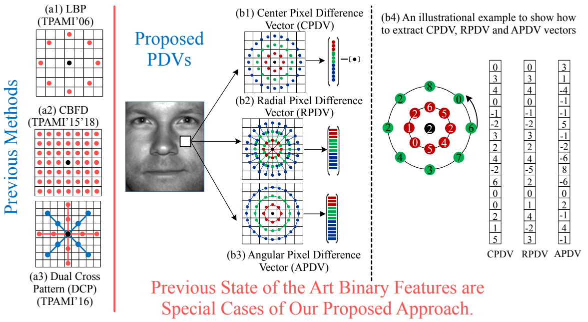

Figure 1: (a): Pattern sampling shapes of existing representative binary features. (b): An illustrational example to show how to extract our proposed PDVs (CPDV, RPDV and APDV) from a local patch.

PDVs, as shown in Fig. 1 (b1), (b2) and (b3). As can be observed from Fig. 1 clearly, existing representative binary features are special cases of our proposed approach.

CPDV. As shown in Figure 1 (b1), for any pixel x_c in the image, we consider a local patch of size $(2R+1)\times(2R+1)$, centered at x_c. For a radius r in this patch, we subtract the gray value of the center pixel x_c from the gray values of its circularly symmetric neighborhood $\{x_{r,p_r,n}\}_{n=0}^{p_r-1}$ (as shown in Equation 1), and thus obtain a CPDV of dimension p_r. $\{x_{r,p_r,n}\}_{n=0}^{p_r-1}$ are the p_r neighboring pixels evenly distributed on a circle of radius r centered at x_c. If the coordinates of the center pixel x_c are $(0,0)$, then the coordinates of $x_{r,p_r,n}$ are given by $(-rsin(2\pi n/p_r), rcos(2\pi n/p_r))$. The gray values of neighbors which do not fall exactly in the center of pixels are estimated by interpolation. CPDV is then formed by concatenating all the PDVs from different radii. In our experiments, we consider local patches of size 11×11, and sample 8, 16 and 24 pixels from radius 1, 3 and 5 respectively, so that each CPDV is a 48-dimensional feature vector.

$$CPDV_{r,p_r,n} = x_{r,p_r,n} - x_c,\ \ n = 0,1,2,...,p_r-1. \tag{1}$$

APDV and RPDV. However, CPDV only encodes the differences between the center pixel and its neighborhood only. In order to explore the second order discriminative information contained in a local patch, we further propose RPDV and APDV, encoding radial and angular pixel differences respectively. In particular, RPDV encodes the between-circumference structure, and preserves the relationship between pixels of different rings. APDV encodes the relationship between neighboring pixels on the same radius and preserves intra-radius information. Formally, RPDV and APDV can be computed as follows:

$$APDV_{r,p_r,n} = \Delta^{Ang}_{r,p_r,n} + \Delta^{Ang}_{r-1,p_r,n},\ \ n = 0,1,2,...,p_r-1\,, \tag{2}$$

$$RPDV_{r,p_r,n} = x_{r,p_r,n} - x_{r-1,p_r,n},\ \ n = 0,1,2,...,p_r-1\,. \tag{3}$$

where, $\Delta^{Ang}_{r,p_r,n}$ $(= x_{r,p_r,n} - x_{r,p_r,mod(n+1,p_r)})$ is the difference between the neighboring pixels on radius r. Note that here we sum angular differences computed from a radii pair (r and $r-1$) to enhance their robustness.

Like CPDV, we also concatenate the calculated PDVs on different radii to form a high-dimensional vector for APDV and RPDV.

To further improve the discriminative power of each type of PDV, we derive a complementary counterpart for each of the above three PDVs by incorporating the magnitude of the calculated differences. As a result, we obtain three additional PDV types and name them as CPDV_M, APDV_M and RPDV_M. Specifically. PDV_M can be computed via:

$$PDV\_M_{r,p_r,n} = |PDV_{r,p_r,n}| - \mu_{r,p_r},\ \ n = 0,1,2,...,p_r-1,\ \ (PDV \in \{CPDV, APDV, RPDV\}), \tag{4}$$

where μ_{r,p_r} $(= \frac{1}{p_r}\sum_{n=0}^{p_r-1}|PDV_{r,p_r,n}|)$ is the mean.

2.2 Compact Binary Pattern Learning

For each PDV type, we aim to learn a feature mapping to project the PDVs to lower dimensional binary vectors. Several unsupervised filter learning techniques, including random projection (RP) [5, 18], PCA, and the recent CBFD [22] method are used for learning. A set of PDVs of dimension d are extracted from the training images and then input to a learning algorithm to learn compact binary vectors of dimension k. For RP, there is no need to learn the projection matrix, and it can preserve the salient information in the signal [18]. For PCA, we only need to learn the principle feature vectors. For CBFD, the learning process is a little complicated. Due to space limitation, details of the learning approaches used can be found in [18, 22].

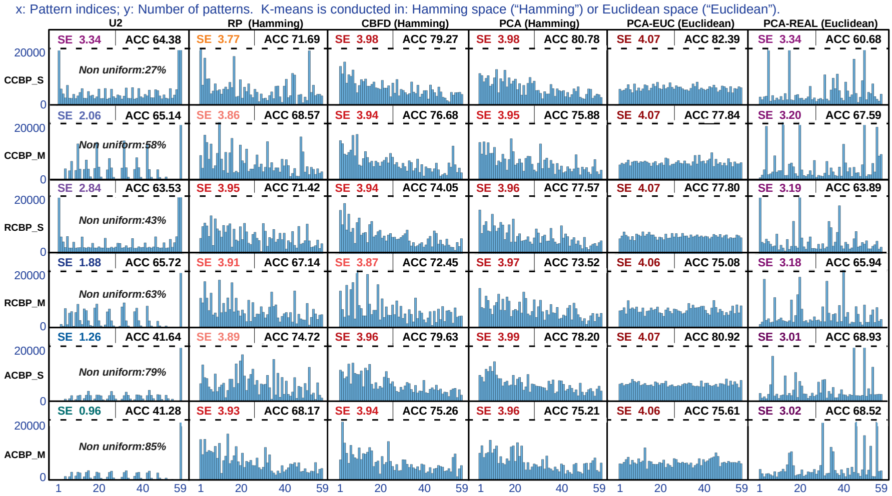

Figure 2: Pattern distributions on different methods. For a fair comparison, all of them adopted the same number of bins for feature representation. We also count the percentages of the non uniform patterns for U2. Generally, the higher Shannon entropy (SE) is, the CBP would be more informative and representative, thus resulting in a better performance.

Having learned the lower dimensional codes, we aim to learn a representative binary codebook from the training set. Here, we obtain six types of Compact Binary Patterns (CBPs), and denote them as CCBP_S, CCBP_M, RCBP_S, RCBP_M, ACBP_S and ACBP_M in response to CPDV, CPDV_M, RPDV, RPDV_M, APDV, APDV_M respectively. We expect the binary codebook to be informative and representative, and look for a binary pattern distribution with a higher Shannon Entropy (SE), which is defined as:

$$H(b_1, b_2, ..., b_N) = -\sum_{i=1}^{N} P(b_i)\log_2 P(b_i),$$

where $P(b_i)$ is the probability of pattern b_i and N is the number of patterns.

Obviously, the maximum of SE is achieved when all the patterns are evenly distributed. For efficiency reasons, we adopt the $k$means clustering on the binary codes to form a codebook. Different from the previous work, we directly apply the $k$means clustering in the Hamming space, which can speed up the learning process considerably. Therefore, we obtain six binary codebooks. Regarding the performance of the learning methods, we have the following findings, as illustrated in Figure 2. 1) Compared with the handcrafted descriptor U2 [1], the learning process achieves a higher representation power; 2) Learning methods PCA and CBFD produce a more evenly distributed pattern than their counterpart RP [18];

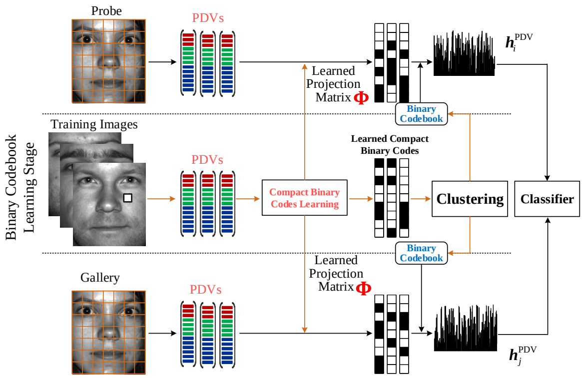

Figure 3: Face recognition pipeline based on learned individual binary features proposed by us. Our final BIRD approach is to fuse all six learned individual binary features, *i.e.* fusing six global feature histograms $\boldsymbol{h}^{\text{CPDV}}$, $\boldsymbol{h}^{\text{RPDV}}$, $\boldsymbol{h}^{\text{APDV}}$, $\boldsymbol{h}^{\text{CPDV\_M}}$, $\boldsymbol{h}^{\text{RPDV\_M}}$, and $\boldsymbol{h}^{\text{APDV\_M}}$. (PDVs in the figure can be one of our proposed PDV types, *i.e.*, CPDVs, RPDVs, APDVs, CPDV_Ms, RPDV_Ms and APDV_Ms.

3) Using binary codes is superior to the one that directly uses real number codes without binarization ("PCA-REAL"). We conjecture from 3) that real number codes tend to be information redundant and contain more noise component. By contrast, the binary codes are more robust and discriminative.

The best distributed patterns are obtained by *k*means in the Euclidean space ("PCA-EUC"). However, the performance of clustering in the Euclidean space is almost the same as that of clustering in the Hamming space with the combination of more CBPs. Due to the efficiency, Hamming space clustering will be used in our experiments.

2.3 BIRD based Face Representation

We first divide the images into $M \times M$ non-overlapped local regions to capture large-scale relations [1, 22]. Then the six CBPs are extracted as described above in the training step for each local region individually. During testing, for each CBP, the learned projection functions and codebooks are used to build histogram representations for the corresponding local regions, which are then concatenated to form a global image description. Therefore, as shown in Fig. 3, each image has six representations in response to the six CBPs respectively. For face recognition, as adopted in [8, 17, 22], we use Whitened PCA (WPCA) to further reduce the feature dimension, and apply cosine similarity as the matching score between two images. To fuse the six CBP features, we average the six similarity scores and use the nearest neighbor classifier in the recognition task.

3 Experiments

3.1 Evaluation on CAS_PEAL_R1

We followed the basic protocol for evaluation on the CAS_PEAL_R1 database, which contains 1200 images of 300 subjects in the Training set, 1040 frontal images of 1040 subjects in the Gallery set. In this experiment, we used the Expression, Accessory and Lighting probe sets, which contains 1570, 2285 and 2243 images respectively under variations in expression, accessory and illumination. All face images were cropped into 128×128 pixels by setting the eyes positions at $(40,48)$ and $(89,48)$ and illumination normalization was used as in [31]. Next, we extracted CPDV from three (r, p_r) pairs: $(1,8)$, $(3,16)$ and $(5,24)$, and

APDV/RPDV from $(2,8)$, $(3,8)$ and $(5,8)$, since we hoped $r \geq 2$ in order to get a $(r, r-1)$ pair. The parameters k, M and the number of patterns were empirically set as 15, as 8 and 500 respectively, leading to a 32000 ($= 8 \times 8 \times 500$) dimensional feature vector for each image. As adopted in [17, 22], WPCA was only conducted on the Gallery set. Noted that the feature dimension after WPCA would be automatically determined according to the rank of the input matrix. The remaining parts followed the framework described in Section 2.3.

Multi-scale vs. Single Scale. We started our analysis by extracting high dimensional PDVs covering multi radii with the following considerations. 1) We assume a big local patch can simulate a multi-scale analysis, to capture both micro-structure and macro-structure information [20, 21, 24]; 2) The crude quantization of angular space caused by a small set of neighboring pixels (*i.e.*, p_r pixels on radius r) could be alleviated by using a larger p_r, thus to increase the robustness against image variants [24]; 3) Inspired by texture analysis in [18] and compressed sensing [3], which demonstrated that most natural images are compressible, it is feasible to learn the low dimensional compact binary codes from a higher dimension space, while with little information loss. As shown in Fig. 4 (a), the recognition rate increases with larger patch size. While we still avoid to apply very large patch size for the subsequent redundancy that degrades the performance.

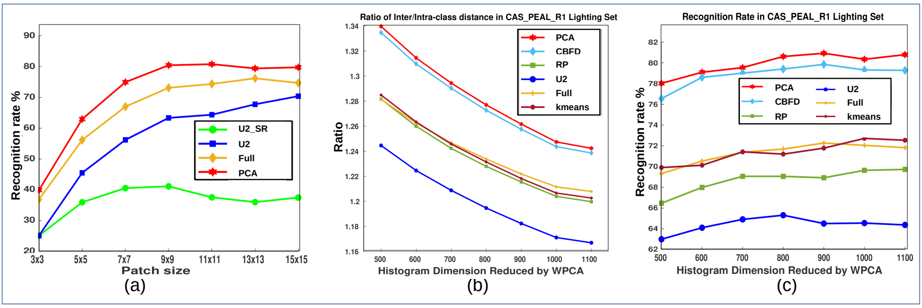

Figure 4: (a) Performances with changing patch sizes. All the method are evaluated with multi radii except U2_SR which only considers a single radius in the the U2 method. (b) The ratio of inter-class discrepancy and intra-class discrepancy w.r.t. Gallery and Probe set. (c) Corresponding results for methods in (b).

Learning based vs. Hand crafted. Our next consideration was what mapping functions should be learned to get discriminative and compact binary codes. Therefore, we further analyzed PCA, PCA-EUC, CBFD, RP and U2 as well as the all patterns encoding method ("Full") [19] by visualizing their distinctive abilities. In addition, the method ("*k*means") directly using *k*means clustering on the raw high-dimension PDV without projection was also considered. The face identification task on the dataset is to find the most similar image in the Gallery set for each image in the Probe set. Therefore, to visualize the differences of all these methods, we can formulate their distinctiveness as the ratio of inter-class discrepancy and intra-class discrepancy on the basis of Gallery and Probe set:

$$\sigma_{intra} = \sum_{i=1}^{C} \frac{1}{N(G_i) \times N(P_i)} \sum_{g \in G_i} \sum_{p \in P_i} \| f(g), f(p) \|, \ \sigma_{inter} = \sum_{i=1}^{C} \frac{1}{N(G_{\sim i}) \times N(P_i)} \sum_{g \in G_{\sim i}} \sum_{p \in P_i} \| f(g), f(p) \|,$$

where, C is the number of subjects on the Gallery set, G_i and P_i are the set of images belong to subject i in the Gallery and Probe set, with number $N(G_i)$ and $N(P_i)$ respectively, while $G_{\sim i}$ is set of Gallery images of subjects other than i, and $f(\cdot)$ is the extracted image feature.

From Fig. 4 (b) and (c), we can see that compared with handcrafted descriptors (Full and U_2), learning based methods achieve better feature description, leading to a higher recognition rate. We may conjecture that on one hand, the handcrafted patterns are not optimal for feature representation, while this is remedied in cases of learning, where patterns are adjusted by *k*means under the supervision of training data distribution. On the other hand, most of noise contained in the raw PDV can be eliminated by PCA or CBFD, which removes the feature dimensions with small variances that usually correspond to redundancy. This is also proved by comparing with *k*means which also uses the raw PDV without projection. We may further discuss the performance difference between PCA and CBFD below.

Individual vs. Extended. As aforementioned, our BIRD descriptor capture textural information with different directions and components, achieving high representation ability. This is emphasized in Fig. 5 (a) by exploring all the 63 combination possibilities over the six CBPs, although there is no guarantee for a monotonically increase in the curves, from which we can see that the overall performance would be improved when more types of CBP descriptors are fused. Another interesting thing is that clustering in Hamming space would not lead to loss of discriminative power, as shown in Fig 5 (a), the performance gap between PCA and PCA-EUC is gradually reduced and eliminated with the CBP combination. This is worthwhile because all the CBPs in BIRD are in the same formulation structure and can be implemented efficiently.

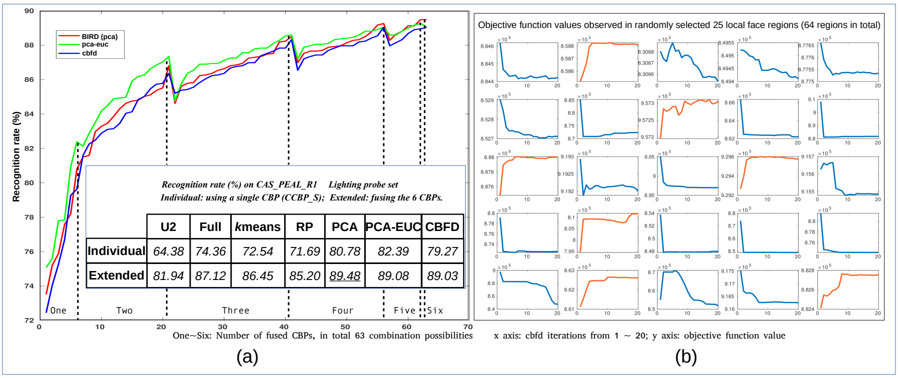

Figure 5: (a) Evaluation on 63 combinations for the 6 CBP descriptors. The six parts on the x axis indicate the the number of fused CBP types, where the results are sorted respectively in each part. (b) Objective function values w.r.t. pattern distribution in CBFD algorithm.

The reason why CBFD is slightly worse than PCA from the above experiments can be explained by tracking the objective function values [22] w.r.t. pattern distributions during the iteration of CBFD algorithm, shown in Fig. 5 (b). In other words, it is hard to ensure the objective function values for all the local face regions are minimized once the parameters are determined. However, this can be improved by selecting better parameters for CBFD.

Comparison with State of the Arts. Table 1 compares different methods, including the original CBFD algorithm [22], DFD [17] and DCP [8] on the CAS_PEAL_R1 dataset. The proposed BIRD method shows high invariance to illumination changes in the Lighting probe set and achieves the best performance with a recognition rate of **89.48%**, which, we believe, is so far the best reported results on this dataset. Furthermore, the proposed CBPs can all produce very competitive performance individually, since the binarization operation in our framework tends to obtain more robustness to variations, while being computationally efficient, as also proved in the Expression and Accessory probe sets, where the results of our framework are superior to or comparable with the recent proposed methods.

Dataset	CAS_PEAL_R1 Expression						CAS_PEAL_R1 Accessory						CAS_PEAL_R1 Lighting					
Method	U2	RP	Full	*k*means	CBFD	PCA	U2	RP	Full	*k*means	CBFD	PCA	U2	RP	Full	*k*means	CBFD	PCA
CCBP_S	99.43	99.81	**99.87**	99.68	99.49	99.75	96.02	96.89	96.76	96.46	97.37	97.29	64.38	71.69	74.36	72.54	79.27	80.78
CCBP_M	98.22	99.04	99.24	99.24	99.30	99.11	88.32	92.43	92.87	93.87	93.22	93.26	65.14	68.57	78.47	72.80	76.68	75.88
RCBP_S	99.49	99.62	99.68	99.62	99.43	99.55	96.54	95.84	97.29	96.67	95.97	96.37	63.53	71.42	74.90	65.23	74.05	77.57
RCBP_M	97.71	98.28	99.11	98.92	98.66	98.66	84.46	89.54	91.33	91.03	92.21	91.55	65.72	67.14	77.89	72.00	72.45	73.52
ACBP_S	99.24	99.62	99.81	99.43	99.62	99.75	94.09	96.76	97.51	97.11	96.85	96.76	41.64	74.72	73.21	70.35	79.63	78.20
ACBP_M	97.96	99.17	99.43	99.04	99.11	99.17	88.53	93.13	92.56	93.09	94.22	94.49	41.28	68.17	79.71	75.17	75.26	75.21
BIRD	99.62	99.81	99.68	99.68	99.55	99.55	95.97	96.46	96.46	96.76	96.72	96.59	81.94	85.20	87.12	86.45	89.03	**89.48**
CBFD* [22]	99.7						97.2						67.4					
DCP [8]	99.62						**99.21**						82.92					
DFD [17]	99.6						96.9						63.9					

Table 1: Recognition rate (%) on CAS_PEAL_R1 database. The results for CBFD* [22] (TPAMI 2015), DCP [8] (TPAMI 2016) and DFD [17] (TPAMI 2014) are from the original papers, others are from the projection-binarization-clustering framework

3.2 Evaluation on LFW

LFW [14, 15] is a widely used database for face verification, containing 13233 images of 5749 subjects created under unconstrained conditions, varying in pose, lighting, focus, resolution, expression, etc. We followed the standard protocol on the "View 2" dataset with the *unsupervised* setting for evaluation, since the proposed method is label free, thus no identification or the *same/different* label information is needed. The "View 2" dataset contains 10 subsets, each of which has 300 matched pairs and 300 unmatched pairs. During the experiments, all the parameters and the cropping methods were the same as those used in the CAS_PEAL_R1 dataset and WPCA was conducted on the training data. Furthermore, similar to [6, 8], we extracted additional 21 local regions based on 21 facial landmarks to gain the pose invariance, so 85 local regions were used for each image ($85 = 8\times 8 + 21$).

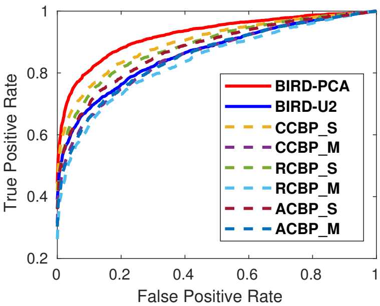

Figure 6: ROC curves for BIRD on LFW with the unsupervised setting.

Methods	AUC
BIRD-U2	0.8626
BIRD-PCA	**0.9238**
CBFD* [22] TPAMI 2015	0.9091
SLBFLE* [23] TPAMI 2018	0.9200
CA-LBFL* [10] TPAMI 2018	0.9166

Table 2: Evaluation on LFW. The results of methods with * are from the original papers

Not surprisingly, the proposed BIRD descriptor achieves better performance than any of the individual components as illustrated in Fig. 6. Table 2 shows the AUC of BIRD and other state of the art methods, including CBFD [22], SLBFLE [23] and CA-LBFL [10]. We can see that our approach achieves the best result among all the methods, demonstrating the high dsicriminative power of the proposed descriptor. Last but not least, there is few deep learning based method recorded on LFW with *unsupervised* setting, due to their need for too much labeled training data. While our approach is label free and compared with deep learning method, it needs much less computation resources, thus can be easily developed for using in embedded or wearable devices.

4 Conclusion and Future Work

This paper presented a novel face descriptor BIRD for face recognition. It is more discriminative and efficient than the previous state of the art binary face descriptors and demonstrated to be highly robust to illumination changes. As future work, we wish to investigate more efficient features (*e.g.*, more powerful learned binary features or compact and efficient binary networks) to enhance the features in terms of discriminative power and robustness, and to reduce the reliance on large scale labelled data.

Acknowledgement. This work was supported by the Business Finland project (Grant No. 3116/31/2017) and Academy ICT 2023 project (Grant No. 313600).

References

[1] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12):2037–2041, 2006.

[2] Peter N Belhumeur, João P Hespanha, and David J Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7):711–720, 1997.

[3] Ingram J Brown. A wavelet tour of signal processing: the sparse way. *Investigacion Operacional*, 30(1):85–87, 2009.

[4] Chi Ho Chan, Muhammad Atif Tahir, Josef Kittler, and Matti Pietikäinen. Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1164–1177, 2013.

[5] Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, and Yi Ma. Pcanet: A simple deep learning baseline for image classification? *IEEE transactions on image processing*, 24(12):5017–5032, 2015.

[6] Dong Chen, Xudong Cao, Fang Wen, and Jian Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3025–3032, 2013.

[7] Weihong Deng, Jiani Hu, and Jun Guo. Compressive binary patterns: Designing a robust binary face descriptor with random-field eigenfilters. *IEEE transactions on pattern analysis and machine intelligence*, 41(3):758–767, 2019.

[8] Changxing Ding, Jonghyun Choi, Dacheng Tao, and Larry S Davis. Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 38(3):518–531, 2016.

[9] Yueqi Duan, Jiwen Lu, Jianjiang Feng, and Jie Zhou. Learning rotation-invariant local binary descriptor. *IEEE Transactions on Image Processing*, 26(8):3636–3651, 2017.

[10] Yueqi Duan, Jiwen Lu, Jianjiang Feng, and Jie Zhou. Context-aware local binary feature learning for face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(5):1139–1153, 2018.

[11] Wen Gao, Bo Cao, Shiguang Shan, Xilin Chen, Delong Zhou, Xiaohua Zhang, and Debin Zhao. The cas-peal large-scale chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(1):149–161, 2008.

[12] Cong Geng and Xudong Jiang. Face recognition using sift features. In *2009 16th IEEE international conference on image processing (ICIP)*, pages 3313–3316. IEEE, 2009.

[13] Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen. Local binary patterns and its application to facial image analysis: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(6):765–781, 2011.

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

[15] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[16] Anil K Jain, Robert P. W. Duin, and Jianchang Mao. Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, 22(1):4–37, 2000.

[17] Zhen Lei, Matti Pietikäinen, and Stan Z Li. Learning discriminant face descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(2):289–302, 2014.

[18] Li Liu and Paul Fieguth. Texture classification from random features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):574–586, 2012.

[19] Li Liu, Paul Fieguth, Guoying Zhao, Matti Pietikäinen, and Dewen Hu. Extended local binary patterns for face recognition. *Information Sciences*, 358:56–72, 2016.

[20] Li Liu, Songyang Lao, Paul W Fieguth, Yulan Guo, Xiaogang Wang, and Matti Pietikäinen. Median robust extended local binary pattern for texture classification. *IEEE Transactions on Image Processing*, 25(3):1368–1381, 2016.

[21] Li Liu, Paul Fieguth, Yulan Guo, Xiaogang Wang, and Matti Pietikäinen. Local binary features for texture classification: Taxonomy and experimental study. *Pattern Recognition*, 62:135–160, 2017.

[22] Jiwen Lu, Venice Erin Liong, Xiuzhuang Zhou, and Jie Zhou. Learning compact binary face descriptor for face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(10):2041–2056, 2015.

[23] Jiwen Lu, Venice Erin Liong, and Jie Zhou. Simultaneous local binary feature learning and encoding for homogeneous and heterogeneous face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(8):1979–1993, 2018.

[24] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7):971–987, 2002.

[25] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. Deep face recognition. In *bmvc*, volume 1, page 6, 2015.

[26] Jianfeng Ren, Xudong Jiang, and Junsong Yuan. Noise-resistant local binary pattern with an embedded error-correction mechanism. *IEEE Transactions on Image Processing*, 22(10):4049–4060, 2013.

[27] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[28] Yi Sun, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation by joint identification-verification. In *Advances in neural information processing systems*, pages 1988–1996, 2014.

[29] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1891–1898, 2014.

[30] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.

[31] Xiaoyang Tan and William Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing*, 19 (6):1635–1650, 2010.

[32] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[33] Mei Wang and Weihong Deng. Deep face recognition: A survey. *CoRR*, abs/1804.06655, 2018.

[34] Wenchao Zhang, Shiguang Shan, Wen Gao, Xilin Chen, and Hongming Zhang. Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 786–791. IEEE, 2005.