How Good Can a Face Identifier Be Without Learning?

Yang Zhong$^{(\boxtimes)}$, Anders Hedman, and Haibo Li

Computer Science and Communication,
KTH Royal Institute of Technology, 100 44 Stockholm, Sweden
{yzhong,ahedman,haiboli}@kth.se

Abstract. Constructing discriminative features is an essential issue in developing face recognition algorithms. There are two schools in how features are constructed: hand-crafted features and learned features from data. A clear trend in the face recognition community is to use learned features to replace hand-crafted ones for face recognition, due to the superb performance achieved by learned features through Deep Learning networks. Given the negative aspects of database-dependent solutions, we consider an alternative and demonstrate that, for good generalization performance, developing face recognition algorithms by using hand-crafted features is surprisingly promising when the training dataset is small or medium sized. We show how to build such a face identifier with our Block Matching method which leverages the power of the Gabor phase in face images. Although no learning process is involved, empirical results show that the performance of this "designed" identifier is comparable (superior) to state-of-the-art identifiers and even close to Deep Learning approaches.

Keywords: Face recognition · Controlled scenario · HD Gabor Phase · Block Matching · Learning-free · Deep Learning

1 Introduction

Automatic human face recognition is a well-defined research problem in the fields of computer vision and pattern recognition. The technical core is to define a distance to measure the similarity between two given face images X and Y. The simplest way to define the distance is using the $l2$ metric on the whole raw images as:

$$d = \|X - Y\|_2. \tag{1}$$

Besides $l2$ form, other forms like $l0$, $l1$, etc., are also widely used. A distance metric is believed good if d is small when both X and Y are from the same person and large when they are from two different persons (so-called small intra-personal variations and large inter-personal variations). Unfortunately, it is hard to directly employ the raw face images for similarity measurement in practice. This is because human face images exhibit significant appearance variations in

scale, pose, lighting, background, hairstyle, clothing, expression, color saturation, image resolution, focus, etc., as they occur in real world applications.

To distinguish persons from their faces, a more effective and efficient way is to represent face images using visual features so face images can be projected into a feature space and classified. Then the similarity between two images X and Y is measured with the following distance metric:

$$d = \sum_i \|x_i - y_i\|_2 , \tag{2}$$

where x_i and y_i are features extracted from two face images X and Y.

The power of using features for face recognition comes from, not only the construction of visual features, but importantly from the flexibility and possibility of **weighting** visual features for classification. With weighting, the similarity is measured by calculating the distance metric:

$$d = \sum_i w_i \|x_i - y_i\|_2 , \tag{3}$$

where w_i is the weight received by feature i. The intuition of giving weights is that for each face image point (in a high-dimension space) such a metric should make face image points from the same person closer than points from different persons.

In face recognition, one of the most technically challenging issues is how to construct suitable facial features for face classification. The facial features constructed by conventional approaches are so-called "hand-crafted features", i.e. features are constructed mathematically or engineered. Commonly used mathematical tools include Wavelet and Gabor filtering. The two most remarkable engineering features used in face recognition are SIFT [28] and LBP [1]. An entirely different way to construct facial features is through learning from face image data, i.e. learning to extract facial representations from training sets. The classical Eigenface approach is about how to extract principal facial components from training data sets for classification. Since the principal components are learned from training sets, the extracted facial features are called learned features. Another well-known algorithm for feature learning is Linear Discriminant Analysis (LDA). Today, with the rise of Deep Learning networks, almost all facial features used in face recognition are learned features.

Although it is easy to see that hand-crafted features and learned features are two different approaches, few realize that they are from two different facial feature constructing schools and there is consequently little debate around this topic. Successful stories of deep neural networks have led us to believe that learning is king! The unspoken assumption is: hand-crafted features are out of date, and only approaches using learned features are viable. The consequence is that we have become blind to their inherent problems. Solutions that (over) learn from training sets (particularly Deep Learning) are becoming increasingly database-dependent, even worse, it is hard to distinguish cases where general progress is made in face recognition from just good solutions to particular problems defined over specific databases.

In this paper, we argue that in the interest of making fundamental progress in face recognition, we ought to adequately study how to develop database-independent face recognition algorithms. We are interested in how good a modern face recognition system can be without learning. We consider face identification mainly due to two reasons: the problem itself is more challenging than face verification; it has been a research topic for quite some time and there are extensive experimental results available for comparison. The scientific methodology we employ here is to construct a face identifier and test and compare with state-of-the-art identifiers to explore empirically the question of how good a face identifier can be without learning.

We propose a method that merely leverages the power of the Gabor phase to address the problem of face identification in controlled scenarios. A slim filter bank of only two Gabor filters is applied to extract the Gabor phase information and explicit phase code matching is performed on the quantized phase map via our Block Matching scheme [55]. Different from other elastic matching schemes, the Block Matching scheme not only cancels the patch-wise spatial shift in the phase map but also simultaneously evaluates the patch-wise utility during the **learning-free matching process**. Combining the matching scheme with phase codewords enables the employment of high-definition phase information (4 times higher than [47]) from the 2 utilized Gabor filters. Thus, the proposed approach can significantly bring up the algorithmic efficiency without sacrificing the recognition accuracy. Furthermore, it is totally comparable to state-of-the-art Gabor solutions and even Deep Learning based solutions.

The disposition of our paper is as follows: we first briefly review related Gabor based approaches in Sect. 2; our approach is then described in Sect. 3 followed by comparative experiments presented in Sect. 4 where we also compare the performance between our approach and Deep Learning solutions. Finally, we discuss our work as a whole and offer our conclusions.

2   Related Work

Gabor filtering enables the employment of rich low-level, multi-scale features by transforming images from the pixel domain to the complex Gabor space. A Gabor face is obtained by filtering a face image with the Gabor filter function, which is defined as:

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2\|z\|^2/2\sigma^2)}[e^{ik_{u,v}z} - e^{-\sigma^2/2}], \qquad (4)$$

where u and v define the orientation and scale of the Gabor kernels respectively, and the wave vector is defined as:

$$k_{u,v} = k_v e^{i\phi_u}, \qquad (5)$$

where $k_v = k_{max}/f^v$, $\phi_u = u\pi/8$; k_{max} is the maximum frequency, σ is the relative width of the Gaussian envelop, and f is the spacing factor between

kernels in the frequency domain [27]. The discrete filter bank of 5 different spatial frequencies ($v \in [0, \cdots, 4]$) and 8 orientations ($u \in [0, \cdots, 7]$) is mostly exploited to filter face images to facilitate multi-scale analysis for face recognition.

In the complex Gabor transformed space, most state-of-the-art face recognition approaches utilize the amplitude of Gabor filtered image for face representation and facial feature construction. As in [51], the LGBP feature is extracted from the amplitude spectrum. One of the motivations is because the amplitude varies slowly with spatial shifts, making it robust to texture variations caused by dynamic expressions and imprecise alignment.

By constructing LBP type features from the amplitude and adopting different learning techniques, many Gabor filtering based approaches have shown remarkable advantages over pixel feature based methods: the identification rate in benchmark evaluations has been improved by more than 20% (reaching around 90%) thanks to the so-called "blessing of dimensionality" [17] (but with a high cost of less computational efficiency [8,29]).

The Gabor phase is robust to light change. It has been well-known that phase is more important than amplitude for signal representation and reconstruction [30]. It is reasonable to believe that the Gabor phase should have played a more important role in face identification. However, the use of the Gabor phase in face recognition is far from common and it has often been unsuccessful with worse or nearly the same performance as the amplitude in comparative experiments [6,16,47,53].

This is largely due to two challenging issues: (1) the Gabor phase is a periodic function and a hard quantization occurs for every period; (2) the Gabor phase is very sensitive to spatial shifts [45,53], which imposes a rigid requirement on face image alignment. The first issue was partly solved by introducing the phase-quadrant demodulation technique [14], but the second issue is still far from being solved. The state-of-the-art Gabor phase approach (LGXP in [47]) extracts varied LBP from the phase spectrum. Since the combination of phase and LBP is also sensitive to spatial shifts, the power of the Gabor phase has not been demonstrated in face identification.

Fusing other features that are independent of the local Gabor features can also lead to better performance: [38,42,52] fuse the global (holistic) features with local ones at feature level; [47] proposes a fusion of the Gabor phase and amplitude on score and feature levels; [8] fuses real, imaginary, amplitude and phase data. Alternatively, attaching an illumination normalization step and weighting the local Gabor features is shown to be helpful as well [6].

3   Our Learning-Free Face Matching Approach

In this section, we first introduce the philosophy of our proposed approach in Sect. 3.1, and then demonstrate how it is used in face identification to achieve competitive performance with respect to the state-of-the-art.

3.1   Overview

Repeatable features extracted from small face portions are known as good discriminative traits for identifying persons. In addition, such local features are less likely to be influenced than the holistic features by pose changes and facial expressions. Thus, it is natural to divide face images into blocks and perform similarity measurements between them. Practically, in most face recognition methods, the matching process compares spatially corresponding patches after face alignment.

But such a matching process implies that the spatially corresponding features are the best match. This implication is hardly true even after face alignment. Because of the movement of facial components, head pose variability and imprecise alignment, the spatially corresponding patches easily become dislocated (see Fig. 2 in [58]). It is nearly impossible to achieve reasonable face alignment by using similarity transformations applied holistically to images.

In our approach, facial components are aligned individually by our Block Matching algorithm. Our Block Matching segments a face image into non-overlapping blocks and treats individual blocks as features explicitly. Given a pair of face images X and Y, the core of the algorithm is to use a given block (feature) x_i of image X to search for the corresponding block y_i in image Y. Then we measure the distance of two blocks as $\|x_i - y_i\|_2$, which is used to form the similarity between two face images as in Eq. 2. This is a direct application of the Elastic Matching concept [45] in face recognition.

Moreover, since not all blocks contribute to face identity equally, it is natural to weight the face blocks during the matching process as shown in Eq. 3. By computing proper weight factors, we can expect larger distance values for patches from different persons and smaller distances for patches from the same persons. The key is how to acquire the weight factor w_i.

Without doubt, we can learn weights from the training sets using metric learning techniques as in [12,18], but the developed algorithm will be database dependent. To have good generalization performance, we developed an efficient on-line learning step to calculate the weight factor w_i during matching the face pair at hand in our Block Matching approach, which is introduced in the following.

3.2   Algorithm

We designed the algorithm based on the observation that a face can be distinguished by its unique feature(s) which is more informative than its surrounding one(s), e.g. scars, moles, nasolabial folds, etc. This means that in an Elastic Matching context, if a segmented patch is discriminative, it gives very small distance when a good match is found and the distance varies dramatically if it is matched to surrounding locations. By considering both the minimum matching distance and the variation of the matching distance, we can evaluate the discrimination power of the local patches.

Specifically, given a face-matching pair, a probe image P (denoted by pb) and a gallery image G (denoted by gl), we first segment the probe image into N non-overlapping patches that are denoted by $\{f_n^{(pb)}\}_0^{N-1}$. (The local features are simply formed by the corresponding patches, e.g. by pixels from a patch for gray-scale images.)

For each probe patch $f_n^{(pb)}$ centered at image coordinate (x_n, y_n) (denoted by $f^{(pb)}(x_n, y_n)$), it searches its best matching block within the corresponding search window and yields a patch-wise distance vector d_n where:

$$d_n = \{d_n^i\}, i \in [0, L-1] \tag{6}$$

where L is the number of candidate gallery patches within the $(2R+1) \times (2C+1)$ search window, i.e. $L = (2R+1) \cdot (2C+1)$ when applying an "exhaustive" search method, R and C stand for the search offset in vertical and horizontal directions respectively. Each element in d_n is computed as:

$$d_n^i = \left\| (f^{(pb)}(x_n, y_n) - f^{(gl)}(x_i, y_i)) \right\|_2, \tag{7}$$

where the patch-wise distance metric is the $l2$-norm of element wise distance of local features (patches) and $f^{(gl)}(x_i, y_i)$ denotes the patch that centered at image coordinate (x_i, y_i) within the search window on the gallery face image so that

$$x_i = x_n + \Delta x, \Delta x \in [-C, C], \ y_i = y_n + \Delta y, \Delta y \in [-R, R]. \tag{8}$$

We then calculate the slope k_n of the linear fitting of the first 5 ascendingly sorted values of d_n for normalization of the patch wise distance for each patch, such that the weight factors for each local feature w_n is calculated as:

$$w_n^* = k_n / d_n^*, \tag{9}$$

where $d_n^* = min(d_n)$. w_n^* is then normalized by its $l1$-norm as:

$$w_n = w_n^* / \sum_{n=0}^{N-1} w_n^*. \tag{10}$$

Finally, the distance between a matching pair of probe and gallery face images is the weighted sum of d_n^* as:

$$dist^{(pb,gl)} = \sum_{n=0}^{N-1} w_n \cdot d_n^*. \tag{11}$$

More details of our Block Matching approach are given in [55].

3.3   Gabor Phase Block Matching (GPBM)

It is known that the features constructed from pixels are vulnerable to lighting and pose variations. To further improve the recognition performance, one way is

to construct more robust features. Another effective way is increasing the dimensionality of features to raise recognition rate dramatically thanks to the "blessing of dimensionality". Traditionally, the most popular way is to exploit the Gabor features via Gabor transformation, which normally increases dimensionality of image representations by 40 times [27].

The Gabor transformation enables the employment of rich low-level multiscale features by transforming images from the pixel domain to the complex Gabor space. Different strategies of using either Gabor magnitude or Gabor phase, or a hybrid of both magnitude and phase have been proposed to construct features. One reasonable option for many state-of-the-art approaches has been to utilize Gabor amplitude for face representation and feature construction.

But high dimensional features lead to high cost and create difficulties for training, computation, and storage (as pointed out in [8,29]). To build a practical solution, patch-based approaches and dimensional reduction techniques, such as PCA or LDA, and rotated sparse regression, are commonly used to learn a subspace to reduce intra-class variation and expand inter-class variations. Since the learning process has to be involved and training datasets are needed (e.g. for LDA, the leading eigenvectors of the covariance matrix are needed to calculate over training image pairs), the advantage of using hand-crafted features to achieve generalization performance is diminished.

Can we remain learning-free (to promise generalization) in our face matching approach and also further improve the recognition performance without suffering from heavy computational load brought about by high-dimension representations? We focus on the **Gabor Phase**, since it better reconstructs signals than amplitude [30]. We combine the Gabor Phase face representation with our Block Matching approach introduced in Sect. 3.2, and demonstrate that increasing the signal dimension is not the only way to boost the recognition performance.

Specifically, we filter faces with only a single-scale Gabor filter pair and calculate the phase of the filtered face. That is, for each face image, only two demodulated Gabor phase spectra are used as in the input of our Block Matching method, see Fig. 1. We first segment the probe phase spectrum into N
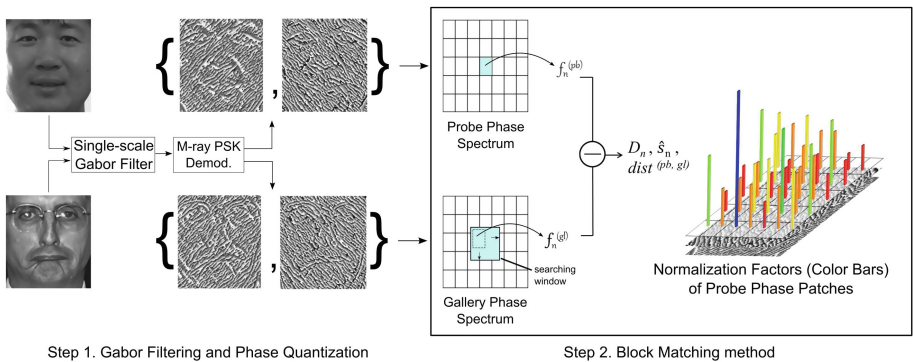

Fig. 1. The matching process of our Gabor Phase Block Matching (GPBM) approach.

non-overlapping patches and the patches $\{f_n^{(pb)}\}_0^{N-1}$ are simply formed by the **raw phase** codes of the patches. Then the Block Matching approach is utilized to calculate the distance of the two faces. The only difference is that when calculating the phase distance, each element in d_n is computed by performing an explicit matching over the raw demodulated phase as:

$$d_n^i = \left\| \mathrm{XOR}(f^{(pb)}(x_n, y_n), f^{(gl)}(x_i, y_i))_{decimal} \right\|_2, \tag{12}$$

where the patch-wise distance metric is the $l2$-norm of element wise Hamming distance in decimal. More technical details are provided in [56].

4   Experiment

4.1   Database Selection

There are a variety of large-scale datasets available for benchmark evaluation of different face recognition approaches, such as the FERET [33], FRGC2.0 [32] and the LFW [21] datasets. Since we focus on face recognition in controlled scenarios in this paper, the FERET database—the most commonly used face identification benchmark—is selected to evaluate and compare our method with state-of-the-art face identification approaches. In addition, the CMU-PIE [36] dataset is selected to evaluate our GPBM against variations of pose, expression and illumination.

4.2   Experimental Setup

Face images were first normalized (aligned) based on the positions of both eyes as in [47]. A central facial area of 150×136, which maintained the same aspect ratio (1.1:1) as in [47,54], was segmented from the face image and used for our experiments.

Due to our Block Matching scheme, the Gabor phase information with a higher definition can be utilized in our approach. We found that a single-scale Gabor filter pair with two orientations is sufficient for face identification. In our implementation, the selected Gabor filters had the following parameters: $v = 0$, $u \in \{2, 6\}$, $f = \sqrt{2}$, $k_{max} = \pi/2$, $\sigma = 2\pi$.

One can see that the chosen Gabor filters have broad high-frequency coverage. These high-frequency components correspond to facial texture variations and are insensitive to the factors of lighting, pose, and aging. Accordingly, to retain high phase definition and to be tolerant to potential phase change caused by texture shift, a Gray-coded 16-PSK demodulator was used for phase demodulation and the constellation is shown in Fig. 2. Compared to the quadrature phase demodulation used in [47,52], 4 times the phase information can be utilized thanks to the employment of our Block Matching approach.

To provide a thorough answer to "How good can a face identifier be without learning", we compare our Block Matching approach [55] and our GPBM to other methods on both image domain and Gabor transformed space in the following.
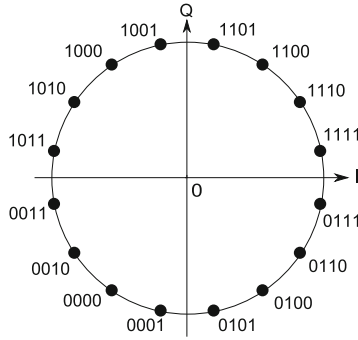
Fig. 2. 16-PSK demodulator constellation.

4.3 Evaluations on the CMU-PIE Database

The CMU-PIE database contains 41368 images of 68 subjects. Images with pose labels 05, 07, 09, 27, and 29 under 21 illuminations (Flash 2 to 22) of all the 68 persons are selected as the probe set.

When applying the Block Matching method, the most important parameters are the block size (H and W) and searching offset (R and C). Our empirical tests on other datasets indicate that it makes sense to divide a central facial area into 5×7 patches, which semantically correspond to components of human faces, like eyes, nose, etc. Thus, for a facial area of 150×136, a reasonable size of a block is 30×20. In our implementation, our block size was 29×19 (we prefer odd block sides) in the block search. To have a good coverage while keeping low computational complexity, the search offset was set to around a quarter of the block size and we selected the search offset of $R = 7$, $C = 6$ pixels in our experiments. To test how sensitive the performance was to the selected parameters, we selected the first 2000 probe images on the CMU-PIE to evaluate the performance with the chosen parameters and other parameters randomly selected around them. The evaluation results in Fig. 3 show that the recognition performance is rather insensitive to parameter selections.
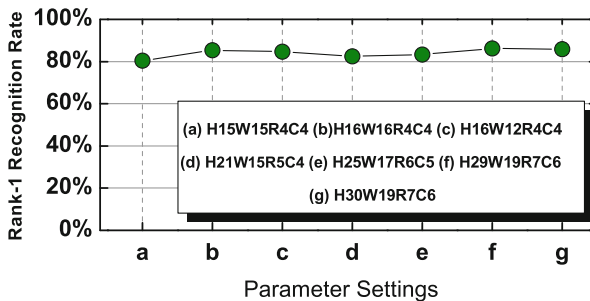

Fig. 3. Recognition rates under different parameters on the CMU-PIE dataset.

We then conducted experiments on the CMU-PIE probe set and compared our GPBM with G_LBP and G_LDP [54]. The G_LBP is the Gabor version LBP and the G_LDP is a type of improved Gabor amplitude Local Binary Pattern. The G_LDP achieved equivalent performance as LGXP (Gabor Phase pattern) on the FERET evaluations so it is a good reference for comparison. Since the Gabor phase is inadequate to handle the non-monotopic illumination variations, we employed the Difference of Gaussian (DoG) image to equalize the illumination on the face. The corresponding results is denoted by $GPBM + DoG$. The comparative rank-1 recognition rates are listed in Table 1. It can be seen that even with use of the raw pixels without any photometric processing for face matching, the Block Matching approach performs comparably to the LBP and LDP approaches where hand-crafted features were employed. Similarly, for our GPBM, it is slightly better than the G_LDP, even though LDP extracts much more complicated Gabor amplitude patterns. The results indicate that with the environment of dramatic illumination and pose variation, our Block Matching approaches have equal recognition power to the hand-crafted features.

Table 1. Comparative rank-1 recognition rates on the CMU-PIE database.

Method	Accuracy
Unweighted LBP*	58%
Best of LDP, 3rd order*	61%
Block Matching [55]	**60%**
G_LBP*	71%
Best of G_LDP, 3rd-order*	79%
GPBM	**82%**
GPBM + DoG	**85%**

* The recognition rates are estimated from Fig. 12a in [54].

4.4   Evaluations on the FERET Database

The FERET database is the most commonly used face identification benchmark. It contains variations in illumination, expression and aging. The gallery set "Fa" contains 1196 frontal face images and the easiest probe set "Fb" contains 1195 images with variations mainly in expression. The probe set "Fc" has 194 images with illumination variations. The "Dup1" set contains 722 images taken later in time than the "Fa". 234 images in the "Dup1" taken at least 1 year after the "Fa" session were selected to form the hardest "Dup2" set. We faithfully followed the evaluation protocol of the FERET dataset and compare with other feature based methods in Table 2.

It is easy to observe that when recognizing faces in the image domain, the Block Matching approach outperforms the LBP and LDP (first 3 rows in

Table 2. Comparative rank-1 recognition rates on the FERET database.

Method	Fb	Fc	Dup1	Dup2
Unweighted LBP [2]	93%	51%	61%	50%
LDP 3rd-order [54]	90%	88%	63%	61%
Block Matching [55]	**93%**	**83%**	**66%**	**64%**
LGBP_Pha [53]	93%	92%	65%	59%
HGPP [52]	97.5%	99.5%	79.5%	77.8%
LGXP [47]	98%	100%	82%	83%
LGXP+BFLD [47]	99%	100%	92%	91%
S[LGBP+LGXP] [47]	99%	100%	94%	93%
LN+LGXP [6]	99.9%	100%	94.7%	91.9%
PCANet-2 [9]	99.6%	100%	95.4%	94%
GPBM	**99.4%**	**100%**	**95.3%**	**94.9%**

Table 2). If Gabor transformation is employed, in a fair comparison (non-learning component was involved and only Gabor phase was utilized for face matching), our GPBM is almost 12% better than LGXP on the hardest "Dup2". Even in unfavorable comparisons, where pre-processing, training, and fusion methods were exploited by LN+LGXP and S[LGBP Mag+LGXP], our GPBM still excels. To our best knowledge, the method S[LGBP_ Mag+LGXP]—aided by the Gabor amplitude and training procedures—was the state-of-the-art Gabor phase based method in terms of performance on the hardest FERET "Dup2", and our GPBM is entirely comparable.

Our approach also has comparable performance to the Deep Learning based PCANet-2 [9]. It firmly confirms again that in a controlled scenario by weighting the image-wise distance via our Block Matching process, we can achieve equally effective face identification as the state-of-the-art. Our results indicate that feature design and high-dimensional signal representation might be less important than commonly believed.

We further compare our GPBM with other state-of-the-art approaches based on other techniques on the FERET in Table 3. From the table one can see that all these approaches are based on Gabor features, which indicates that the Gabor filter is a very effective tool for signal representation. Our GPBM method outperforms all the other approaches on the hardest "Dup2" set and it features three advantages: (1) it enables high definition Gabor phase to be utilized for face identification; (2) a single-scale Gabor filter with two orientations is sufficient to generate an effective face image representation, with 1/20 of the computational complexity of other methods that utilized 40 Gabor filters; (3) further to this, it is not a learning-based face identification method and, therefore, promises good generalization.

We also evaluated our approach under pose variations using the pose probe sets "bd", "be", "bf", "bg" and gallery set "ba" on the FERET dataset. These

Table 3. Comparative summary of recent state-of-the-art face identification approaches.

Methods	Image size	Gabor feature space Gabor filter bank	Training data independent	Rank-1 Rate on FERET Dup2
LGXP[a]	88×80	Phase 5×8	No	83%
LGBP+LGXP[b]	88×80	Amplitude + Phase 5×8	No	93%
GOM[c]	160×128	Amplitude + Phase 5×8	No	93.1%
LN+LGXP[d]	251×203	Phase 5×8	No	91.9%
LN+LGBP[e]	251×203	Amplitude 5×8	No	93.6%
SLF-RKR_$l2$[f]	150×130	Amplitude 5×8	No	94.4%
GPBM, ours	150×136	Phase (explicit matching) 1×2	**Yes**	**94.9%**

[a][47], [b][47], [c][8], [d][6], [e][6], [f][48]

sets correspond to pose angles of $+25°$, $+15°$, $-15°$, $-25°$, and $0°$ to the camera, and each of these sets contains 200 persons. The comparative performance of our GPBM is listed in Table 4. It can be observed that our GPBM approach is significantly more accurate than the non-learning feature-based approaches ($LGBP$ and $LGXP$) when probe faces have relatively large pose angles. In addition, it outperforms the learning-based methods on all probe sets and is even comparable to the recent Deep Learning approach $SPAE$ on "be" and "bf" sets.

Table 4. Comparative rank-1 recognition rates on the FERET dataset with the posed probe sets.

Method	bd	be	bf	bg
Pose angle	$-25°$	$-15°$	$+15°$	$+25°$
LGBP [6]	86.5%	98%	97.5%	88.5%
LGXP [6]	73.5%	95.5%	96%	65.5%
StackFlow [3]	89%	96%	94%	92%
DAE [4]	93%	96%	96%	94%
LN//LGXP [6]	97.5%	99%	99.5%	96%
SPAE [23]	98%	99%	99%	99%
GPBM	**96%**	**100%**	**100%**	**95%**

The computational complexity is always a big concern. As in Table 4 of [29], under the image size of 128×128 with a 5×8 Gabor filter bank, the histogram extraction of LGBP takes around 0.45 s, S[LGBP_ Mag+LGXP] takes 0.99 s. Extracting GOM feature takes 0.7 s [8]. However, the "feature extraction" time in our method is **0** s since only the raw phase is used for matching; the demodulation is the only on-line computation of the probe face, thus, it is extremely fast. Our Matlab implementation executes the matching of a face pair in 0.05 s

on average (Gabor filtering included) on a 3.4 GHz Intel CPU. We can therefore safely conclude that our GPBM outperforms the best Gabor-phase based approach (S[LGBP₋ Mag+LGXP]) in efficiency with a big margin. We can also infer that the other methods in Table 3 could hardly be more efficient than our GPBM due to higher image resolution, Gabor face dimensions, and additional photometric processing. Here we should mention that our GPBM needs to run block matching. Right now, we used an "exhaustive search" strategy. Since we have just a few blocks per probe image, matching is still fast. In future work, we could also incorporate fast-search strategies from the video compression field to speed up face matching.

4.5   Deep Learning for Face Recognition in Controlled Scenarios

Before we conclude this paper, it would be interesting to investigate *how good Deep Learning can be for face recognition in controlled scenarios*. To answer this, we trained several CNNs with well-known architectures of AlexNet [24], VGG-net [37], Google's InceptionNet [40] and FaceNet [35]), and evaluated them on the most difficult probe set "Dup2" of the FERET database.

For fair comparisons between different architectures, layers after the last spatial pooling in our implementation of the InceptionNet and the FaceNet were replaced by two concatenated Fully Connected (FC) layers, and Softmax was selected as the lost function. We used the WebFace dataset [50] to train our networks. WebFace is a face image collection with half a million instances of around 10000 celebrities. Since the FERET dataset was formed with non-celebrity people, all the trained nets were fine-tuned carefully with the FERET gallery images.

To illustrate how architecture choice affects the recognition performance, we investigated how rank-1 accuracy varies under different sizes of the FC layers. The results are listed in Table 5. We can see that the architecture (length of FC) does influence the recognition accuracy. Explicitly inherit network architectures designed for one image classification task (e.g. networks with $FC - 4096$ layers work well on the ImageNet) may not perform well in a novel face recognition task. Investigations on suitable deep feature representations must be made correspondingly and here we found that $FC - 1024$ is a good choice which is also verified in [31]. On the other hand, the performance strongly correlates to architecture in general: even with $FC - 1024$, the InceptionNet outperformed others.

Table 5. Rank-1 accuracy of several well-known CNN architectures on FERET Dup2 (input image size to CNNs was 120×120).

Architectures	Length of the FC-4096	Last 2 FCs FC-1024
AlexNet	91.9%	94.4%
VGG-13 layers	93.6%	94.9%
VGG-16 layers	93.2%	97.0%
InceptionNet	95.3%	98.7%
FaceNet	94.9%	98.3%

Apparently, CNNs can outperform the proposed approach for around 4%, but such advantage is not statistically significant for the test: the best CNN correctly identified 9 more probe faces than our proposed approach which made 222 correct answers out of 234 probes on the "Dup2" set. We can see that even with the most advanced CNNs and trained with a massive face dataset, deep-learning doesn't solve the face identification problem defined over the FERET set significantly better than our non-learning approach. At least, one can conclude that for an unseen face identification task, the approach developed without learning could be a promising solution.

5   Discussion

Although without learning, we have shown that a combination of Block Matching with Gabor phase could work as a very good identifier. Unlike in most state-of-the-art face identifiers where a highly engineered design of facial features or high dimensional features ("blessing of dimensionality") are "must-have", the proposed face identifier has no designed features. It just uses the blocks of raw face images, two orientation channels of a single scale Gabor filter to construct the phase features for face recognition. Our experimental results show that the form and dimensionality of features are of course important, but not the key in building a good identifier. The key lies in how to handle the factors causing unlimited variations of facial textures. The crucial issue in constructing features is knowing how the features affect the relationship between within-class variability and between-class variability. Face recognition can be performed reliably only when the between-class variability is larger than the within-class variability. Why does the proposed identifier work? It is due to that spatial shift caused by camera, pose, expression, scale, and aging between two face images is effectively tackled by our Block Matching approach. In addition, the 2D Gabor filtering family uniquely achieves the theoretical lower bound on joint uncertainty over spatial position and frequency. These properties are particularly useful in characterizing facial textures. Leveraged by our Block Matching technique, the Gabor phase demonstrates its power in handling lighting factors and detailing of local facial texture changes.

Just like others, the performance of our face identifier was also tested with standard database sets. However, there is a clear difference between the evaluations. Since it is not learned from the training data, our identifier is of good generalization. One can expect a similar performance when applying to other databases. Of course, according to the *No Free Lunch Theorem*, if we are interested solely in the generalization performance, there is no reason to prefer one identifier over another. Certain prior knowledge about the problem or a concrete application is always used explicitly or implicitly, for example, choice of operating parameters in our case. In fact, our identifier itself is of a good technical platform where learning can be well integrated. For example, instead of being computed on-line, the weights can be learned from database sets. Thus, our approach is becoming the so-called metric learning:

$$d = (\mathbf{x} - \mathbf{y})^T W (\mathbf{x} - \mathbf{y}), \tag{13}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the feature representations of X and Y, W is a weight matrix, typically a symmetric positive definite matrix. A typical example is the Mahalanobis metric. The weight matrix can be learned from either sets of labeled image pairs or just sets of labeled images with an objective of finding a matrix such that positive pairs have smaller distances than negative pairs. Of course, once learning is involved the developed identifier will be more database-dependent and less apt for generalization.

Face recognition has been developed for over more than three decades and three-order of magnitude improvement in recognition rate has been achieved. One has to realize that such an achieved performance increase was only on the selected databases.

Due to the popularity of machine learning, we don't know how to measure the real progress that has been made in face recognition. Taking the "LFW benchmark" as an example, so many advanced CNN networks have been trained to do face recognition and some of the best can achieve 99.5% recognition accuracy in benchmark evaluations. But when such networks were practically deployed in a real-world application, it was found that they were still far from usable mostly due to the divergence between the training dataset and the real-world data [57]. Though Deep Learning made a big stride in solving challenging face recognition tasks, it is still early to confirm that it is the only right way to go. It is wise to include diverse solutions using "hand-crafted" features and/or features learned from data. This is why in this paper we take on a radical approach to see how far face recognition can go without learning.

6   Conclusions

In this paper we argue strongly that it makes sense to study how good a face identifier can be without learning, particularly today when Deep Learning is very commonly used. We have shown how to construct such an identifier that simply uses Block Matching technique over Gabor phase codes to achieve state-of-the-art performance. We have demonstrated that engineered feature designs or those adhering to the slogan "blessing of dimensionality" are not essential ingredients for building a good identifier. The key issue in constructing features is to achieve between-class variability larger than within-class variability. Since it is not learned from the training data, our identifier lends itself well for generalization. One can expect similar performance when applying it to other databases. This is very important for developing algorithms that constitute real progress in face recognition.

References

1. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24670-1_36

2. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
3. Ashraf, A.B., Lucey, S., Chen, T.: Learning patch correspondences for improved viewpoint invariant face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)
4. Bengio, Y.: Learning deep architectures for AI. Found. Trends Mach. Learn. **2**(1), 1–127 (2009)
5. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
6. Cament, L.A., Castillo, L.E., Perez, J.P., Galdames, F.J., Perez, C.A.: Fusion of local normalization and Gabor entropy weighted features for face identification. Pattern Recogn. **47**(2), 568–577 (2014)
7. Cao, X., Wei, Y., Wen, F., Sun, J.: Face alignment by explicit shape regression. Int. J. Comput. Vis. **107**(2), 177–190 (2014)
8. Chai, Z., Sun, Z., Mendez-Vazquez, H., He, R., Tan, T.: Gabor ordinal measures for face recognition. IEEE Trans. Inf. Forensics Secur. **9**(1), 14–26 (2014)
9. Chan, T.-H., Jia, K., Gao, S., Lu, J., Zeng, Z., Ma, Y.: PCANet: a simple deep learning baseline for image classification? IEEE Trans. Image Process. **24**(12), 5017–5032 (2015)
10. Chen, D., Cao, X., Wang, L., Wen, F., Sun, J.: Bayesian face revisited: a joint formulation. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7574, pp. 566–579. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33712-3_41
11. Chan, C.H., Tahir, M.A., Kittler, J., Pietikainen, M.: Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **35**(5), 1164–1177 (2013)
12. Cui, Z., Li, W., Xu, D., Shan, S., Chen, X.: Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3554–3561. IEEE (2013)
13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
14. Daugman, J.: How iris recognition works. IEEE Trans. Circ. Syst. Video Technol. **14**(1), 21–30 (2004)
15. Daugman, J.: Probing the uniqueness and randomness of iriscodes: results from 200 billion iris pair comparisons. Proc. IEEE **94**(11), 1927–1935 (2006)
16. Gao, Y., Wang, Y., Zhu, X., Feng, X., Zhou, X.: Weighted Gabor features in unitary space for face recognition. In: 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, p. 6. IEEE (2006)
17. Givens, G.H., Beveridge, J.R., Lui, Y.M., Bolme, D.S., Draper, B.A., Phillips, P.J.: Biometric face recognition: from classical statistics to future challenges. Wiley Interdisc. Rev.: Comput. Stat. **5**(4), 288–308 (2013)
18. Hu, J., Lu, J., Tan, Y.-P.: Discriminative deep metric learning for face verification in the wild. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1875–1882, June 2014
19. Hua, G., Akbarzadeh, A.: A robust elastic and partial matching metric for face recognition. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 2082–2089. IEEE (2009)

20. Huang, G., Lee, H., Learned-Miller, E.: Learning hierarchical representations for face verification with convolutional deep belief networks. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2518–2525 (2012)
21. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical report 07-49, University of Massachusetts, Amherst (2007)
22. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. arXiv preprint. arXiv:1408.5093 (2014)
23. Kan, M., Shan, S., Chang, H., Chen, X.: Stacked progressive auto-encoders (spae) for face recognition across poses. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1883–1890. IEEE (2014)
24. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (eds.) Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105. Curran Associates Inc. (2012)
25. Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. IEEE Trans. Comput. **42**(3), 300–311 (1993)
26. Li, H., Hua, G., Lin, Z., Brandt, J., Yang, J.: Probabilistic elastic matching for pose variant face verification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3499–3506. IEEE (2013)
27. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans. Image Process. **11**(4), 467–476 (2002)
28. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
29. Mu, M., Ruan, Q., Guo, S.: Shift and gray scale invariant features for palmprint identification using complex directional wavelet and local binary pattern. Neurocomputing **74**(17), 3351–3360 (2011)
30. Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. Proc. IEEE **69**(5), 529–541 (1981)
31. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: Proceedings of the British Machine Vision Conference (2015)
32. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 947–954. IEEE (2005)
33. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1090–1104 (2000)
34. Pinto, N., DiCarlo, J.J., Cox, D.D.: How far can you get with a modern face recognition test set using only simple features? In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 2591–2598. IEEE (2009)
35. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)
36. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (PIE) database. In: Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Proceedings, pp. 46–51. IEEE (2002)

37. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556 (2014)
38. Su, Y., Shan, S., Chen, X., Gao, W.: Hierarchical ensemble of global and local classifiers for face recognition. IEEE Trans. Image Process. **18**(8), 1885–1896 (2009)
39. Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1891–1898 (2013)
40. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: 2014 IEEE 12th International Conference on Computer Vision (2014)
41. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2013)
42. Tan, X., Triggs, B.: Fusing Gabor and LBP feature sets for kernel-based face recognition. In: Zhou, S.K., Zhao, W., Tang, X., Gong, S. (eds.) AMFG 2007. LNCS, vol. 4778, pp. 235–249. Springer, Heidelberg (2007). doi:10.1007/978-3-540-75690-3_18
43. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings of CVPR 1991, pp. 586–591. IEEE (1991)
44. Vedaldi, A., Lenc, K.: Matconvnet-convolutional neural networks for matlab. arXiv preprint arXiv:1412.4564 (2014)
45. Wiskott, L., Fellous, J.-M., Kuiger, N., Von Der Malsburg, C.: Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 775–779 (1997)
46. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210–227 (2009)
47. Xie, S., Shan, S., Chen, X., Chen, J.: Fusing local patterns of Gabor magnitude and phase for face recognition. IEEE Trans. Image Process. **19**(5), 1349–1361 (2010)
48. Yang, M., Zhang, L., Shiu, S.-K., Zhang, D.: Robust kernel representation with statistical local features for face recognition. IEEE Trans. Neural Netw. Learn. Syst. **24**(6), 900–912 (2013)
49. Yi, D., Lei, Z., Li, S.Z.: Towards pose robust face recognition. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3539–3545. IEEE (2013)
50. Yi, D., Lei, Z., Liao, S., Li, S.Z.; Learning face representation from scratch. arXiv preprint arXiv:1411.7923 (2014)
51. Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: 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 2005, vol. 1, pp. 786–791. IEEE (2005)
52. Zhang, B., Shan, S., Chen, X., Gao, W.: Histogram of Gabor phase patterns (HGPP): a novel object representation approach for face recognition. IEEE Trans. Image Process. **16**(1), 57–68 (2007)
53. Zhang, W., Shan, S., Qing, L., Chen, X., Gao, W.: Are Gabor phases really useless for face recognition? Pattern Anal. Appl. **12**(3), 301–307 (2009)
54. Zhang, B., Gao, Y., Zhao, S., Liu, J.: Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. IEEE Trans. Image Process. **19**(2), 533–544 (2010)

55. Zhong, Y., Li, H.: Is block matching an alternative tool to LBP for face recognition? In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 723–727 (2014)
56. Zhong, Y., Li, H.: Leveraging Gabor phase for face identification in controlled scenarios. In: Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 49–58 (2016)
57. Zhou, E., Cao, Z., Yin, Q.: Naive-deep face recognition: touching the limit of LFW benchmark or not? arXiv preprint. arXiv:1501.04690 (2015)
58. Zou, J., Ji, Q., Nagy, G.: A comparative study of local matching approach for face recognition. IEEE Trans. Image Process. **16**(10), 2617–2628 (2007)