![DiVA logo](http://www.diva-portal.org)
This is the published version of a paper presented at *International Conference on Computer Vision Theory and Applications*.

N.B. When citing this work, cite the original published paper.

Leveraging Gabor Phase for Face Identification in Controlled Scenarios

Yang Zhong, Haibo Li

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

Abstract: Gabor features have been widely employed in solving face recognition problems in controlled scenarios. To construct discriminative face features from the complex Gabor space, the amplitude information is commonly preferred, while the other one — the phase — is not well utilized due to its spatial shift sensitivity. In this paper, we address the problem of face recognition in controlled scenarios. Our focus is on the selection of a suitable signal representation and the development of a better strategy for face feature construction. We demonstrate that through our Block Matching scheme Gabor phase information is powerful enough to improve the performance of face identification. Compared to state of the art Gabor filtering based approaches, the proposed algorithm features much lower algorithmic complexity. This is mainly due to our Block Matching enables the employment of high definition Gabor phase. Thus, a single-scale Gabor frequency band is sufficient for discrimination. Furthermore, learning process is not involved in the facial feature construction, which avoids the risk of building a database-dependent algorithm. Benchmark evaluations show that the proposed learning-free algorithm outperforms state-of-the-art Gabor approaches and is even comparable to Deep Learning solutions.

1   INTRODUCTION

Face recognition (FR) is a well established research area and it has been studied for more than two decades (Turk and Pentland, 1991; Belhumeur et al., 1997; Wiskott et al., 1997; Ahonen et al., 2004; Zou et al., 2007; Wright et al., 2009; Chan et al., 2013; Taigman et al., 2013). Typically, face recognition works in two essentially different modes: face verification or face identification under either in controlled scenarios or in the wild. Face verification performs *1:1* matching and provides a binary decision to the claimed identity. Face verification in controlled scenarios has reached a rather high accuracy (Givens et al., 2013). To tackle face verification in uncontrolled scenarios, many approaches have been proposed for more effective alignment (Cao et al., 2014; Yi et al., 2013; Chen et al., 2012), utilization of different types of feature representations (Lowe, 2004; Dalal and Triggs, 2005; Ahonen et al., 2004) and matching metric for comparing faces (Hua and Akbarzadeh, 2009; Pinto et al., 2009; Li et al., 2013). Driven by innovation in Deep Learning approaches, face verification performance has been greatly advanced in recent years. By learning from big data, end-to-end artificial networks can outperform human on challenging verifica-

tion tasks, e.g., the Labeled Face in the Wild (Huang et al., 2007), (Huang et al., 2012; Taigman et al., 2013; Sun et al., 2013; Schroff et al., 2015).

In contrast, face identification is more difficult. It performs *1:N* matching to sort out the gallery images based on pair wise similarity measurements. Obviously, the operating requirement of face identification is ***vastly more demanding*** than operating merely in verification: an identifier needs to be roughly *N* times better than a verifier to achieve comparable odds against making false matches (Daugman, 2006). This is probably why progress in face identification has been relatively insignificant over the last five years. Though the proposed face identification approaches have become increasingly complex, recognition performance according to the benchmark evaluations remained relatively constant (Xie et al., 2010; Yang et al., 2013; Cament et al., 2014; Chai et al., 2014). To make a breakthrough in face identification, it seems we must revisit the foundation of face recognition, and have a fresh look at the fundamental building blocks of face recognition.

A most fundamental building block of face recognition is construction of features for measuring similarity between two face images. The construction of features consists of two steps: (1) the selection

of a suitable face representation; (2) feature extraction from the representation. There is a large collection of research papers on how to extract stable, local or global discriminative features, e.g., the commonly used SIFT (Lowe, 2004), HOG (Dalal and Triggs, 2005), and LBP (Ahonen et al., 2004). Recently, these features have been criticized as "hand-crafted". It has been claimed that better features can be learned automatically from big face data collections through Deep Learning approaches as in (Huang et al., 2012; Taigman et al., 2013; Sun et al., 2013).

While Deep Learning solutions generally have significant performance advantages over conventional approaches, the dependence on access to very large training datasets, the knowledge of designing delicate hand-crafted artificial neural network and careful engineering is considerable. This indicates that the learned network could bias (or likely overfit) to the training data; to adapt well to novel tasks, sufficient data for fine tuning or even the construction of additional components in the learned architecture is often a must (Parkhi et al., 2015) (we demonstrate this in the last section). It is also worth nothing that Deep Learning solutions are generally computational demanding: typically the Convolutional Neural Networks (CNN) involves millions of parameters. Thus, although we believe a well tuned deep neural network is a most effective FR solution (e.g., for FR in the wild), in many scenarios where users are cooperative and environments are controlled, it is often worth to consider efficient and effective solutions depending on much shallower networks. Considering this, it is tenable to revisit one of the most widely adopted approaches, Gabor transformation, which paved the way for face representation in controlled scenarios.

The Gabor transformation enables the employment of rich low-level multi-scale features by transforming images from pixel domain to the complex Gabor space. In the complex Gabor transformed space, one reasonable option for many state-of-the-art approaches was to utilize the amplitude for face representation and feature construction. This is because the amplitude varies slowly with spatial shift, making it robust to texture variations caused by dynamic expressions and imprecise alignment. By constructing LBP-type features mostly from the amplitude and applying various learning techniques, many Gabor based approaches have shown remarkable advantages over pixel-featured based methods: the identification rate in benchmark evaluations has been found to be improved by more than 20% (reaching around 90%) thanks to the "blessing of dimensionality" (Givens et al., 2013) (but at the high cost of computational efficiency (Mu et al., 2011; Chai et al., 2014)). Now

the question is how to achieve face identification rates in the range from 90% to 95% or even higher. In this paper we argue that leveraging Gabor phase could enable such performance improvement for face identification.

The Gabor phase is robust to light change and indeed has been well-known that phase is more important than amplitude for signal reconstruction (Oppenheim and Lim, 1981). Gabor phase should have played a more important role in face identification. However, use of 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 (Gao et al., 2006; Zhang et al., 2009; Xie et al., 2010; Cament et al., 2014). This is largely due to two challenging issues: (1) Gabor phase is a periodic function and a hard quantization occurs for every period; (2) it is very sensitive to spatial shift (Wiskott et al., 1997; Zhang et al., 2009), which imposes a rigid requirement on face image alignment. The first issue was partly solved by introducing the phase-quadrant demodulation technique (Daugman, 2004), but the second issue is still far from being solved. The state-of-the-art Gabor phase approach (LGXP (Xie et al., 2010)) extracts varied LBP from the phase spectrum. Since the combination of the phase and LBP is also sensitive to spatial shift, the power of Gabor phase was not demonstrated in face identification.

In this work, we propose a method that merely leverages the power of Gabor phase to address the problem of face identification in controlled scenarios. We apply a slim filter bank of only two Gabor filters to extract the Gabor phase information and perform explicit matching on the quantized phase map via our Block Matching scheme (Zhong and Li, 2014). Different from other elastic matching schemes, the Block Matching scheme not only cancels the patch-wise spatial shift in 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 exploitment of high-difinition phase information (4 times higher than (Xie et al., 2010)) from only 2 Gabor filters. Thus, the proposed approach can significantly bring up the algorithmic efficiency without sacrificing the recognition accuracy. Further more, it is totally comparable to those state of the art Gabor solutions and even CNN based solutions.

The disposition of our paper is as follows: we first briefly review the related Gabor based and CNN based solutions in Section 2; our approach is then described in Section 3 followed by comparative experiments presented in Section 4. We further compare the

performance between our approach and CNN architectures in Section 5, where we discuss our work as a whole and offer our conclusions.

2 RELATED WORK

In this section, we first describe the Gabor representation and then review recent Gabor based face recognition methods utilizing the Gabor amplitude or phase in different ways.

In state-of-the-art face identification methods based on Gabor wavelet representations, there are two steps to construct features for identification. The first step is to use Gabor filters for an optimal image representation. A Gabor face is obtained by filtering a face image with the Gabor filter function, which is defined as:

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

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

$$k_{u,v} = k_v e^{i\phi_u},$$

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 (Liu and Wechsler, 2002). 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.

From the Gabor face representation, we then need to form face features for identification. The most popular way is to extract the LBP type patterns from the complex Gabor transformed image. As in (Zhang et al., 2005), the LGBP feature is extracted from the amplitude spectrum. In (Zhang et al., 2007) and (Xie et al., 2010), 4-ray phase-quadrant demodulator is applied to demodulate the phase from each of the complex Gabor coefficients, and local binary phase descriptors are subsequently generated from the demodulated phase spectrum. Dimension reduction can also be used for feature construction. As in (Xie et al., 2010), FLD is applied to form local Gabor features. Besides, face identification can be built on other types of local representations, exemplified as GOM (Chai et al., 2014) and SLF (Yang et al., 2013).

Fusing other features that are independent of the local Gabor features can also lead to better performance: (Tan and Triggs, 2007; Su et al., 2009; Zhang et al., 2007) fuses the global (holistic) features with local ones at feature level; (Xie et al., 2010) proposes fusion of Gabor phase and amplitude on the score and feature levels; (Chai et al., 2014) fuses real, imaginary, amplitude and phase. Alternatively, attaching an illumination normalization step and weighting the local Gabor features is shown to be helpful as well (Cament et al., 2014).

Another face representation trend is to utilize comparatively and/or exclusively trained CNNs to learn discriminative metrics and features. With recently launched hardware platforms (Jia et al., 2014; Vedaldi and Lenc, 2014) and especially public accessible large-scale dataset (Yi et al., 2014), developing deep learning based face recognition approaches becomes feasible with less resources. Many deep network architectures have been proposed as in (Schroff et al., 2015; Parkhi et al., 2015; Yi et al., 2014).

3 THE GABOR PHASE BLOCK MATCHING APPROACH

In this section, we first introduce the philosophy of our proposed approach in Subsection 3.1, and then describe the signal representation selected for our approach and the details of the approach in Subsection 3.2.

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 image into blocks and performs similarity measurements between them.

Even if being different in how to construct local features from either Gabor amplitude or phase representation, state-of-the-art face recognition methods do share a common implication: the spatially corresponding patches/features are the best match (since matching is only performed between spatially corresponding features). This implication is hardly true because of the movement of facial components, head pose variablity and imprecise alignment, the spatially corresponding patches easily become dislocated (see Fig. 2 in (Zou et al., 2007)). In addition, matching the spatially corresponding Gabor phase patches is even worse than comparing Gabor amplitude patches since phase is very sensitive to spatial shift of facial textures. To handle this problem, some solutions deploy elastic matching strategy to allow each segmented patch of one image to search the best matching from spatially neighboring locations on the other

image. This has achieved better robustness to spatial shift of the textures as in (Lades et al., 1993; Wiskott et al., 1997; Hua and Akbarzadeh, 2009; Zhong and Li, 2014).

It is not surprising that different facial area (blocks) have different utilities for identifying people. For instance, patches containing eyes would have higher discrimination power than other patches. But, it is also straightforward that the discriminative patches could locate at any position on the face when matching faces. This is because the discriminative features are totally matching-pair specific. For example, patches from the cheek area with almost no texture would be useful only when matched to a cheek patch has scar or a mole; it won't contribute much when matched to similar texture-less ones. Thus, to improve discrimination, it is necessary to weight the corresponding features based on their utility.

In this work, we show how to use the Blocking Matching method (Zhong and Li, 2014) to construct local face features and perform matching between best-matching features. Our Block Matching scheme explicitly handle spatial shift between two patches, so that high-definition [1] Gabor phase information can be maximally utilized and patch-wise utility is evaluated on-line in the same matching process as well. We demonstrate the combination of Gabor phase with the Blocking Matching is powerful in handling the hard factors, like light, pose, facial expressions, and aging involved in face identification tasks.

3.2 Algorithm

The matching process of our Gabor Phase Block Matching (GPBM) approach is illustrated in Fig.1 and the details of the Block Matching searching scheme is shown in Fig.2.

To match a gallery and a probe image, our GPBM approach consists of two steps. In the first step, both probe and gallery face images are filtered using a Gabor filter bank (in the following we show that a single-scale Gabor filter with two orientations is sufficient in our approach). The filtered images are demodulated by a Gray-coded Phase Shift Keying (PSK) demodulator for smooth phase quantization. In the second step, we divide phase images into non-overlapping blocks. It is a natural choice to employ such a block based approach since it is our way to construct features. As mentioned in (Xie et al., 2010), the block (patch) based strategy is an effective tool to handle the so called "curse of dimensionality" and has been proved to be an effective tool in face recognition.

---

[1]Compared to approaches used quadratic phases, e.g., in (Xie et al., 2010).

In our approach, the Block Matching method is used to form features for identification. The demodulated phase spectra are input to the Block Matching method (Zhong and Li, 2014) to form features and calculate the the pair wise distance between a probe (pb) image and a gallery (gl) image. Specifically (Step 2 of Fig. 1), we first segment the probe phase spectrum into N non-overlapping patches and the patches $\{f_n^{(pb)}\}_0^{N-1}$ are simply formed by the ***raw phase*** codes of the patches. For each probe patch $f_n^{(pb)}$ centered at image coordinate (x_n, y_n) (denoted as $f^{(pb)}(x_n, y_n)$), it searches its best matching block within the corresponding search window and yields a patch-wise distance vector d_n denoted as:

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

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 full search method, R and C stands for the searching offset in vertical and horizontal directions respectively. Each element in d_n is computed by performing an explicit matching over the raw demodulated phase as:

$$d_n^i = \left\| \text{XOR}(f^{(pb)}(x_n, y_n), f^{(gl)}(x_i, y_i))_{decimal} \right\|_2, \tag{2}$$

where the patch-wise distance metric is the $l2$-norm of element wise Humming distance in decimal 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,

$$\begin{cases} x_i = x_n + \Delta x, & \Delta x \in [-C, C] \\ y_i = y_n + \Delta y, & \Delta y \in [-R, R]. \end{cases} \tag{3}$$

We then evaluate the patch-wise utility by applying linear regression to the matching distance values. This is achieved by calculating 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 normalization factor s_n is calculated as:

$$s_n = k_n / d_n^*, \tag{4}$$

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

$$s_n^* = s_n / \sum_{n=0}^{N-1} s_n. \tag{5}$$

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

$$dist^{(pb, gl)} = \sum_{n=0}^{N-1} s_n^* \cdot d_n^*. \tag{6}$$

It is noteworthy that 1) feature extraction is not carried out through the Block Matching process; 2)
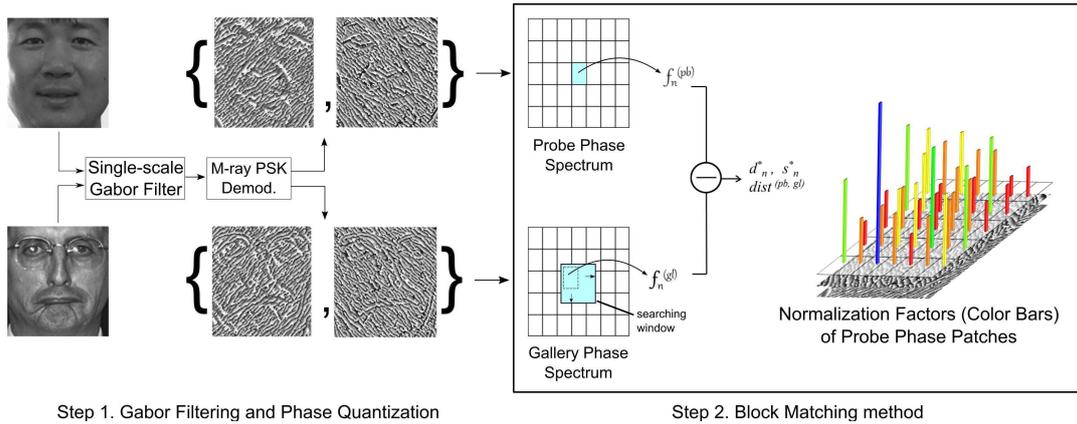
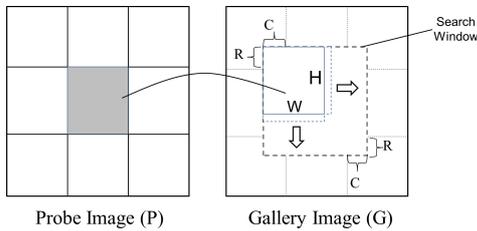Figure 1: Gabor Phase Matching using the Block Matching Method

Figure 2: Block Matching Scheme

the patch-wsie utility evaluation process is realized only based on the two matching images at hand and learner is not involved.

4 EXPERIMENTS AND RESULTS

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 (Phillips et al., 2000), FRGC2.0(Phillips et al., 2005) and the LFW dataset. 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 (Sim et al., 2002) dataset is selected to evaluate our GPBM against variations of pose, expression and illumination.

4.2 EXPERIMENTAL SETUP

Face images are first normalized (aligned) based on the positions of both eyes as in (Xie et al., 2010).

A central facial area of 150×136 , which maintains the same height/width ratio (1.1 : 1) as in (Xie et al., 2010; Zhang et al., 2010), is 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 have 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 is used for phase demodulation and the constellation is shown in Fig.3. Compared to the quadrature phase demodulation used in (Zhang et al., 2007; Xie et al., 2010), 4 times the phase information can be utilized thanks to the employment of our block-matching approach. With Gabor phase information, the block matching approach is used to form features for face identification.

4.3 Evaluations on the CMU-PIE database

The CMU-PIE database contains 41368 images of 68 subjects. Images with Pose Label 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 blocking matching method, the most important parameters are the block size (H and W) and searching offset (R and C). We have con-
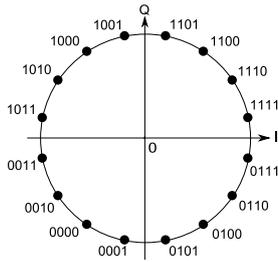
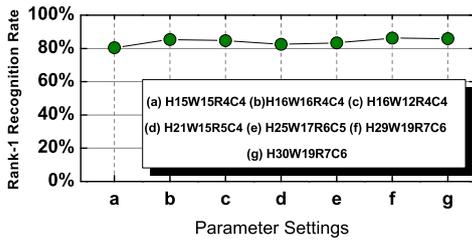Figure 3: Phase codeword generator: 16-PSK demodulator constellation

Figure 4: Recognition rates under different parameters on PIE

ducted a set of empirical tests over other datasets to select suitable parameters. We found that it makes sense to divide a central facial area into 5×7 batches, which correspond to semantic facial macro features, like eyes, nose, etc. Thus, for a facial area of 150×136, a reasonable size of a block is 30×20. In our implementation, we select the block size of 29×19 for convincing block searching (where we prefer the block size with odd numbers). To have good coverage while keeping low computational complexity, the searching offset is chosen as around $\frac{1}{4}$ of the block size and we select search offset of $R = 7$, $C = 6$ pixels in our experiments. To test how sensitive the performance is 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 are shown in Fig. 4. From the test results one can see that the performance is rather insensitive to the selection of parameters. Thus, in our experiments with both CMU-PIE and FERET databases we used the chosen parameters.

We then conduct experiments on the CMU-PIE probe set and compare our GPBM with G_LBP and G_LDP (Zhang et al., 2010). 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

Table 1: Comparative rank-1 recognition rates of GPBM on the CMU-PIE database

Method	Accuracy
G_LBP [*]	71%
G_LDP 2nd-order [*]	72%
G_LDP 3rd-order [*]	79%
G_LDP 4th-order [*]	74%
GPBM	**82%**

[*] The recognition rates are estimated from Fig. 12a in (Zhang et al., 2010).

is a good reference for comparison. The comparative rank-1 recognition rates are listed in Table 1. It can be seen that our method is at least 3% better than the G_LDP, even though LDP extract much more complicated patterns than the LBP from the Gabor amplitude space. Utilizing the Gabor phase in the Block Matching scheme is more effective in dealing with pose and illumination changes than LBP-type patterns extracted from the Gabor amplitude space.

4.4 Evaluations on the FERET database

The FERET database contains 1196 frontal face images in the gallery set, 1195 images with different expressions in the probe set "Fa", 194 images with illumination variations in the probe set "Fc", 722 images taken in later time in the "Dup1" set, and 234 images taken at least 1 year later than the gallery set form the hardest "Dup2" set. We faithfully follow the evaluation protocol of the FERET dataset. The results of our GPBM with other approaches using Gabor-phase are listed in Table 2.

From Table 2 one can see that in a fair comparison, when only Gabor phase is utilized for matching, our GPBM is almost 12% better than LGXP on the hardest "Dup2"; even in unfavorable comparisons, where pre-processing, training, and fusion methods are exploited by LMGEW//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 — is state-of-the-art Gabor phase based method in terms of performance on the hardest FERET "Dup2", and our GPBM is entirely comparable.

We also 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 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

Table 2: Comparative rank-1 recognition rates of **Gabor-phase** based approaches on the FERET database

Method	Fb	Fc	Dup1	Dup2
LGBP_Pha (Zhang et al., 2009)	93%	92%	65%	59%
LGBP_Pha^{weighted} (Zhang et al., 2009)	96%	94%	72%	69%
HGPP^{weighted} (Zhang et al., 2007)	97.5%	99.5%	79.5%	77.8%
LGXP (Xie et al., 2010)	98%	100%	82%	83%
LGXP+BFLD (Xie et al., 2010)	99%	100%	92%	91%
S[LGBP_Mag+LGXP] (Xie et al., 2010)	99%	100%	94%	93%
LMGEW//LN+LGXP (Cament et al., 2014)	99.9%	100%	94.7%	91.9%
GPBM	**99.4%**	**100%**	**95.3%**	**94.9%**

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.

The computational complexity is always a big concern. From Table 4 in (Mu et al., 2011), under the image size of 128×128 with a 5×8 Gabor filter bank, the histogram extraction of LGBP takes around 0.45 seconds, S[LGBP_ Mag+LGXP] takes 0.99 seconds. Extracting GOM feature takes 0.7 seconds (Chai et al., 2014). However, the "feature extraction" time in our method is **0** seconds 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 second in average (Gabor filtering included) on a 3.4GHz 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 and 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.

5 DISCUSSIONS AND CONCLUSIONS

Before we conclude this paper, it would be interesting to investigate "*How good Deep Learning can be in face recognition in controlled scenarios?*". To answer this, we trained several CNNs with well-known architectures of AlexNet (Krizhevsky et al., 2012), VGG-net (Simonyan and Zisserman, 2014) and Google's InceptionNet (Szegedy et al., 2014) and FaceNet (Schroff et al., 2015)), and evaluated them on the most difficult probe set "Dup2" of FERET database. For fair comparisons on 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 . We used WebFace dataset (Yi et al., 2014) to train our networks and carefully fine tuned the trained nets afterwords with FERET gallery images.

To illustrate how architecture choice affects recognition performance, we investigated how the Rank-1 accuracy varies under different sizes of the FC layers. The results are enlisted in Table 4. We can see that the architecture (length of FC) does influence the recognition accuracy. On the one hand, explicitly inherit networks designed for other image classification tasks may not perform well in novel face recognition tasks (compare the left column to the right); investigations on suitable deep feature representations must be made correspondingly (here we found that $FC - 1024$ is a good choice which is also verified by (Parkhi et al., 2015)). On the other hand, the performance strongly correlates to architecture in general: even with $FC - 1024$ the InceptionNet outperformed others. While it is not astonishing that some CNNs outperformed the proposed approach for almost 4%, we can see that such advantage is not statistically significant: the best CNN correctly identified 9 more probe faces than our proposal which made 222 correct answers out of 234 probes on the "Dup2" set. One can expect even higher accuracy from better CNN solutions, but under limited conditions where CNN based solutions are not feasible (e.g., due to lack of training data) the proposed method is still a good alternative due to its comparable effectiveness and high efficiency.

The Gabor based solutions share a common limitation. They extract low level features from image

Table 3: Comparative Summary of Recent State-of-the-art Face Identification Approaches

Methods	Image Size	Photometric Processing	Gabor Filter Scale × Orient.	Gabor Feature Space	Training Data Independent	Rank-1 Accuracy on FERET Dup2
LGXP (Xie et al., 2010)	88×80	No	5×8	Phase	No	83%
LGBP+LGXP (Xie et al., 2010)	88×80	No	5×8	Amplitude + Phase	No	93%
GOM (Chai et al., 2014)	160×128	No	5×8	Amplitude + Phase	No	93.1%
LN+LGXP (Cament et al., 2014)	251×203	Yes	5×8	Phase	No	91.9%
LN+LGBP (Cament et al., 2014)	251×203	Yes	5×8	Amplitude	No	93.6%
SLF-RKR_l2 (Yang et al., 2013)	150×130	No	5×8	Amplitude	No	94.4%
GPBM, ours	150×136	No	$\mathbf{1 \times 2}$	Phase(**explicit matching**)	**Yes**	**94.9%**

Table 4: Rank-1 accuracy of several well-known CNN architectures on FERET Dup2 (Input image size to CNNs is 120×120)

Architectures	Length of the Last 2 FCs	
	FC-4096	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%

texture through a shallow filter bank (typically of 40 filter kernels). When it comes to matching faces with dramatic texture changes, the features are not "deep" enough to construct discriminative representations. This makes Gabor based solutions unsuitable for face recognition in uncontrolled scenarios where the CNN solutions, which build high-level features from low level Gabor-like filters, demonstrated overwhelming advantages.

To conclude, we propose a plain approach to leverage the demodulated Gabor phase for face identification based on the Block Matching method. The proposed approach neither utilizes a large Gabor filter bank nor a training process. It only depends on the signal representation from a single-scale Gabor filter pair to perform explicit matching over the raw Gabor phase spectrum.

Comparative experiments show that: 1) our approach features the highest accuracy utilizing the Gabor phase for face recognition; 2) our approach retains very low computational complexity yet with comparable performance to other state of the art methods including Deep Learning methods when it works on face recognition in controlled scenarios. Our experiments demonstrate that our Block Matching method is a powerful tool that can leverage the power of Gabor phase to boost the face recognition performance.

REFERENCES

Ahonen, T., Hadid, A., and Pietikainen, M. (2004). Face recognition with local binary patterns. In *Computer Vision - ECCV 2004*, volume 3021 of *Lecture Notes in Computer Science*, pages 469–481. Springer Berlin Heidelberg.

Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720.

Cament, L. A., Castillo, L. E., Perez, J. P., Galdames, F. J., and Perez, C. A. (2014). Fusion of local normalization and gabor entropy weighted features for face identification. *Pattern Recognition*, 47(2):568–577.

Cao, X., Wei, Y., Wen, F., and Sun, J. (2014). Face alignment by explicit shape regression. *International Journal of Computer Vision*, 107(2):177–190.

Chai, Z., Sun, Z., Mendez-Vazquez, H., He, R., and Tan, T. (2014). Gabor ordinal measures for face recognition. *Information Forensics and Security, IEEE Transactions on*, 9(1):14–26.

Chan, C. H., Tahir, M. A., Kittler, J., and Pietikainen, M. (2013). Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(5):1164–1177.

Chen, D., Cao, X., Wang, L., Wen, F., and Sun, J. (2012). Bayesian face revisited: A joint formulation. In *Computer Vision–ECCV 2012*, pages 566–579. Springer.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE.

Daugman, J. (2004). How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):21–30.

Daugman, J. (2006). Probing the uniqueness and randomness of iriscodes: Results from 200 billion iris pair comparisons. *Proceedings of the IEEE*, 94(11):1927–1935.

Gao, Y., Wang, Y., Zhu, X., Feng, X., and Zhou, X. (2006). Weighted gabor features in unitary space for face

recognition. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 6–pp. IEEE.

Givens, G. H., Beveridge, J. R., Lui, Y. M., Bolme, D. S., Draper, B. A., and Phillips, P. J. (2013). Biometric face recognition: from classical statistics to future challenges. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(4):288–308.

Hua, G. and Akbarzadeh, A. (2009). A robust elastic and partial matching metric for face recognition. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2082–2089. IEEE.

Huang, G., Lee, H., and Learned-Miller, E. (2012). Learning hierarchical representations for face verification with convolutional deep belief networks. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2518–2525.

Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc.

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., and Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *Computers, IEEE Transactions on*, 42(3):300–311.

Li, H., Hua, G., Lin, Z., Brandt, J., and Yang, J. (2013). Probabilistic elastic matching for pose variant face verification. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3499–3506. IEEE.

Liu, C. and Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *Image processing, IEEE Transactions on*, 11(4):467–476.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

Mu, M., Ruan, Q., and Guo, S. (2011). Shift and gray scale invariant features for palmprint identification using complex directional wavelet and local binary pattern. *Neurocomputing*, 74(17):3351–3360.

Oppenheim, A. V. and Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541.

Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep face recognition. *Proceedings of the British Machine Vision Conference*.

Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., Marques, J., Min, J., and Worek, W. (2005). Overview of the face recognition grand challenge. In *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, volume 1, pages 947–954. IEEE.

Phillips, P. J., Moon, H., Rizvi, S. A., and Rauss, P. J. (2000). The feret evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1090–1104.

Pinto, N., DiCarlo, J. J., and Cox, D. D. (2009). How far can you get with a modern face recognition test set using only simple features? In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2591–2598. IEEE.

Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823.

Sim, T., Baker, S., and Bsat, M. (2002). The cmu pose, illumination, and expression (pie) database. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 46–51. IEEE.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.

Su, Y., Shan, S., Chen, X., and Gao, W. (2009). Hierarchical ensemble of global and local classifiers for face recognition. *Image Processing, IEEE Transactions on*, 18(8):1885–1896.

Sun, Y., Wang, X., and Tang, X. (2013). Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1891–1898.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2014). Going deeper with convolutions. *Computer Vision, 2014 IEEE 12th International Conference on*.

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2013). 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.

Tan, X. and Triggs, B. (2007). Fusing gabor and lbp feature sets for kernel-based face recognition. In *Analysis and Modeling of Faces and Gestures*, pages 235–249. Springer.

Turk, M. A. and Pentland, A. P. (1991). Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE.

Vedaldi, A. and Lenc, K. (2014). Matconvnet-convolutional neural networks for matlab. *arXiv preprint arXiv:1412.4564*.

Wiskott, L., Fellous, J.-M., Kuiger, N., and Von Der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):775–779.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. (2009). Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227.

Xie, S., Shan, S., Chen, X., and Chen, J. (2010). Fusing local patterns of gabor magnitude and phase for face recognition. *Image Processing, IEEE Transactions on*, 19(5):1349–1361.

Yang, M., Zhang, L., Shiu, S.-K., and Zhang, D. (2013). Robust kernel representation with statistical local features for face recognition. *Neural Networks and Learning Systems, IEEE Transactions on*, 24(6):900–912.

Yi, D., Lei, Z., and Li, S. Z. (2013). Towards pose robust face recognition. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3539–3545. IEEE.

Yi, D., Lei, Z., Liao, S., and Li, S. Z. (2014). Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*.

Zhang, B., Gao, Y., Zhao, S., and Liu, J. (2010). Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. *Image Processing, IEEE Transactions on*, 19(2):533–544.

Zhang, B., Shan, S., Chen, X., and Gao, W. (2007). Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition. *Image Processing, IEEE Transactions on*, 16(1):57–68.

Zhang, W., Shan, S., Gao, W., Chen, X., and Zhang, H. (2005). Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 786–791. IEEE.

Zhang, W., Shan, S., Qing, L., Chen, X., and Gao, W. (2009). Are gabor phases really useless for face recognition? *Pattern Analysis and Applications*, 12(3):301–307.

Zhong, Y. and Li, H. (2014). Is block matching an alternative tool to lbp for face recognition? In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 723–727.

Zhou, E., Cao, Z., and Yin, Q. (2015). Naive-deep face recognition: Touching the limit of lfw benchmark or not? *arXiv preprint arXiv:1501.04690*.

Zou, J., Ji, Q., and Nagy, G. (2007). A comparative study of local matching approach for face recognition. *Image Processing, IEEE Transactions on*, 16(10):2617–2628.