Learning-based Descriptor for 2-D Face Recognition

CAO, Zhimin

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Information Engineering

The Chinese University of Hong Kong
July 2010

Abstract of thesis entitled:

Learning-based Descriptor for 2-D Face Recognition

Submitted by CAO, Zhimin

for the degree of Master of Philosophy

at The Chinese University of Hong Kong in May 2010

Face recognition (verification) with 2-D images is a traditional and important computer vision problem. It is the key component of many applications, such as bio-information security system, Internet face image search, and electronic photo-album management. Two main steps are involved in a face recognition system, representing the face image with certain image descriptor and then defining a distance (similarity) measure between two face image descriptors. Early stage research work concatenate the image pixels into a vector as the face representation. Recently, some novel face descriptors based on the image micro-structure encoding have been proposed. Most of these micro-structure encoding methods are manually designed, which makes them far from optimal.

In this thesis, we present a novel face image descriptor to address the representation issue in 2-D face recognition (verification). Firstly, our approach encodes the micro-structures of the face by a new learning-based encoding method. Unlike many previous manually designed encoding methods, we use unsupervised learning techniques to learn an encoder from the training examples, which can au-

tomatically achieve very good tradeoff between discriminative power and invariance. Then we apply a (unsupervised) dimension reduction technique, PCA, to get a compact face descriptor. While the previous usage of PCA technique is hindered by performance degradation, we find a simple normalization mechanism after PCA can reverse the degradation and significantly improve the discriminative ability. The resulting face representation, learning-based (LE) descriptor, is compact, highly discriminative, and easy-to-extract/compare.

The proposed novel descriptor is tested on several 2-D face recognition benchmark to demonstrate its good recognition performance. With training on one dataset and testing on the other mode, our method is proven to have excellent generalization ability across different datasets.

摘要

基於二維圖像的人臉識別（認證）是一個計算機視覺中的傳統重要問題。它是很多實際應用系統的關鍵部分，比如生物信息安全系統，因特網的人臉圖片搜索以及電子照片管理系統。一個人臉識別系統中包含兩個主要部分： 一是用某種人臉描述子（特徵）來表示人臉圖片；二是定義兩個人臉描述子之間的距離（相似度）。早期的研究工作一般把人臉圖片的像素拉成一個向量作為表示。最近的研究工作提出了一些基於人臉圖片微結構編碼的新穎的描述子，其中大部分使用了手工設計的微結構編碼方法，這導致它們遠遠不是最優的描述子。

在本論文中，我們提出了一種新的人臉描述子來解決二維人臉圖像的表示問題。首先，我們的方法用一種基於學習得到的編碼方法來編碼人臉圖片的微結構。與之前提出的手工設計的編碼方法相比，我們使用了一種無監督學習方法從訓練集中得到了編碼方法，從而能夠自動地在描述子的鑑別能力和不變性之間取得與一個良好的平衡。隨後我們使用了一種無監督的維數降低算法，主分量分析，來得到一種緊緻的人臉描述子。之前很多研究工作中主分量分方法的使用往往會帶來性能的損失，我們在研究中發現一個簡單的後歸一化處理卻能夠避免性能的降低，並且大幅度提高鑑別力。最終獲得的人臉描述子，稱之為 LE 描述子，不僅具有緊緻的特性，而且擁有很高的鑑別力，也容易抽取和比較計算。

本論文提出的這種新的描述子在幾個二維人臉識別的標準庫上進行了實驗，結果表明確實具有很好的鑑別力。通過在一個數據集上訓練而在另一個數據集上測試，實驗還表明了我們提出方法具有很好的推廣性。

Acknowledgement

In the past two years, I have enjoyed an abundant and fruitful life. Many people give me their warm hands when I came into difficulty. Here I hope to have the opportunity to express my sincere gratitude to them.

Firstly, I wish to give my greatest gratitude to my supervisor Prof. Tang. He gave me great instructions and advices on how to formulate an idea, how to do a good research and how to write a paper. He is always nice to the students, offering us sufficient room to be an independent researcher with our own insights into the research topic. When we encounter difficulty or do not know which direction is worth to try during experiments, he is always ready to stand there to give us his suggestions and remarks. Discussion with him is an enjoy since exciting and interesting ideas keep coming to your mind. I feel so lucky to have a nice and smart professional as my supervisor.

I also learned much from Dr. Jian Sun, my mentor when I did my intern in Microsoft Research Asia. Dr. Sun has deep insights into the computer vision area, and he is an expert in supporting a fresh student to do good research. He never criticizes my "silly" questions or ideas during discussions. Instead, he always gives his praise to each small step I made and told me what he thinks of the

current results and what I might need to explore further.

I would also like to give my thanks to Dr. Jianzhuang Liu. He is an excellent researcher and keeps serious attitude towards academic activities, which gives me deep impressions. Much thanks to Boqing and Hao, and to all the lab-mates, Weige, Ming, Xiaowei, Mo, Ke, Tianfan, Yueming, Xiaotian, Yichen, Kaiming, and Kui. We have happy days together, discussing questions, playing cards, PC games, badminton, *etc*.. It has become a piece of good memory I will never forget in my future life.

Finally, I would like to thank my parents. Though they do not know my research, they told me that I should do what I think right. Without their support and love, I could not finish the two-year graduate life smoothly and happily.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and related work

Face recognition for 2-D images is an extensively studied, yet challenging vision task. Various face recognition systems have been successfully applied in recognition task under the controlled conditions. There are two main kinds of face recognition tasks: face identification (who is who in a probe face set, given a gallery face set) and face verification (same or not, given two faces). In this thesis, we focus on the verification task, which is more widely applicable and is also the foundation of the identification task. For convenience, a pair of face images belong to the same person (different persons) is termed as intra-person (extra-person) pair.

Since face verification is a binary classification problem on an input face pair, there are two major components of a verification approach: face representation and face matching. The extracted feature (descriptor) is required to be not only discriminative but also invariant to apparent changes and noise. The matching should be robust to variations from pose, expression, and occlusion, as shown in Figure 1.1. These requirements render face verification a challenging problem.

Figure 1.1: Images from the same person may look quite different due to pose (upper left), expression (upper right), illumination (lower left), and occlusion (lower right).

Traditional methods concatenate the image pixels into a vector as the face representation, and then apply different vector subspace analysis (learning) algorithms, *e.g.*, Eigen-face [28], Fisher-face [22], Laplacian-face [13], to extract a discriminative holistic face descriptor. Though achieving success under certain well-controlled scenario, these methods used a holistic face representation, which could not utilize the abundant information included in the image micro-structures. A research direction is designing (proposing) effective face representation to encode more face micro-structure information.

Recently, this direction has attracted much research effort [12], [30], [31], [14], [15], [17], [24], [25], [36], [38] due to the progresses of local face descriptors [7], [20], [21], [27], [33], [34], [35], [36] and increasing demands of real-world applications, such

as face tagging on the desktop [6] or the Internet[1]. Currently, these descriptor-based approaches [14], [25], [37] have been proven to be effective face representations producing best performance [11], [23], [16]. Ahonen *et al.* [1] proposed to use the histogram of Local Binary Pattern (LBP) [21] to describe the micro-structures of the face. LBP encodes the relative intensity magnitude between each pixel and its neighboring pixels. It is invariant to monotonic photometric change and can be efficiently extracted. Since LBP is encoded by a handcrafted design, many LBP varieties [26], [36], [40] have been proposed to improve the original LBP. SIFT [20] or Histogram of Oriented Gradients (HOG) [7] are other kinds of effective descriptors using handcrafted encoding. The atomic element in these descriptors can be viewed as the quantized code of the image gradients. Essentially, different encoding methods and descriptors have to balance between the discriminant power and the robustness against data variance.

However, existing handcrafted encoding methods suffer two drawbacks. On one hand, manually getting an optimal encoding method is difficult. Usually, using more contextual pixels (higher dimension vector) can generate a more discriminative code. But it is non-trivial to manually design an encoding method and determine the codebook size to achieve reasonable tradeoff between discrimination and robustness in a high dimension space. In addition, handcrafted codes are usually unevenly distributed as shown in Figure 1.2. Some codes may rarely appear in real-life face images. It means that the resulting code histogram will be less informative and less compact, degrading the discriminant ability of the descriptor.

[1]Picasa Web Albums, *http://picasaweb.google.com/*

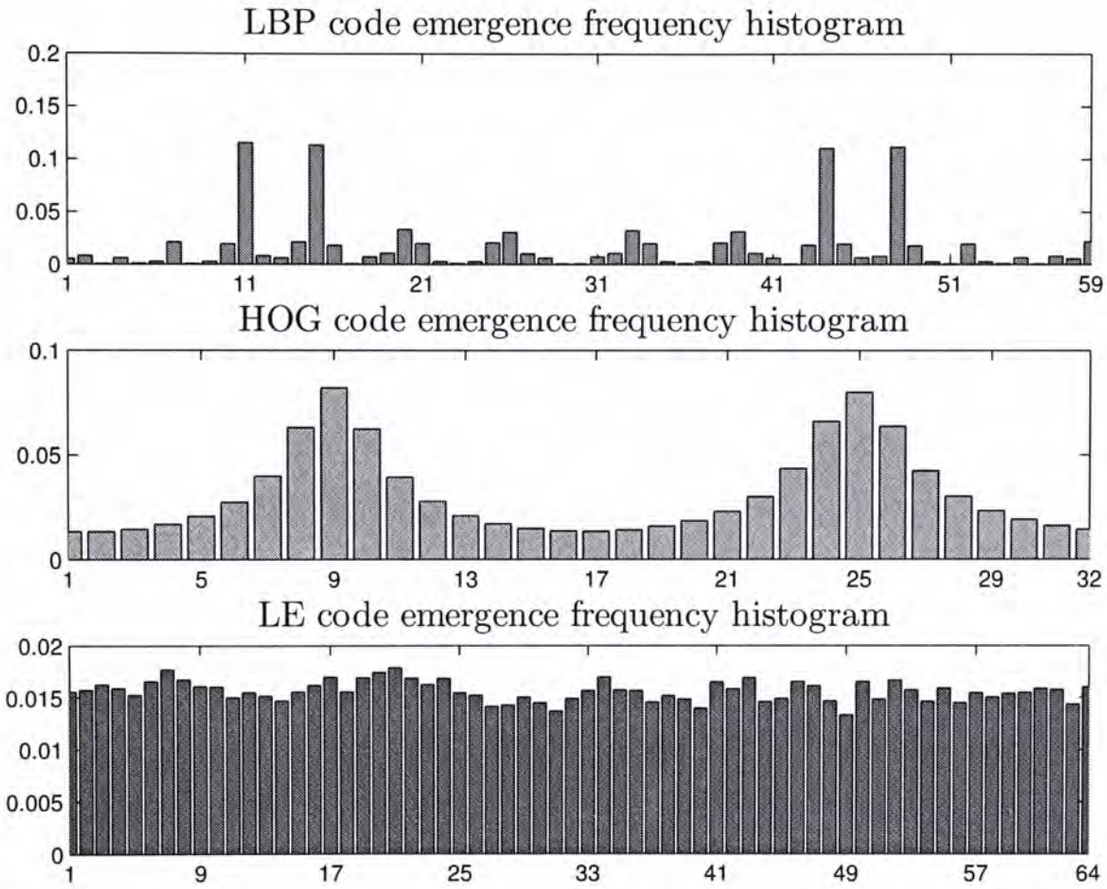

Figure 1.2: The code uniformity comparison of LBP, HOG, and the proposed LE code. We computed the distribution of code emergence frequency for LBP (59 uniform codes), HOG (32 orientation bins) and LE (64 codes) in 1000 face images. Clearly, the histogram distribution is uneven for LBP and HOG while our LE code is close uniform.

In this thesis, to tackle the aforementioned difficulties, we present a learning-based encoding method, which uses unsupervised learning methods to encode the local micro-structures of the face into a set of discrete codes. The learned codes are more uniformly distributed (as shown in Figure 1.2) and the resulting code histogram can achieve much better discriminative power and robustness trade-off than existing handcrafted encoding methods. Furthermore, to pursue the compactness, we apply the dimension reduction technique, PCA, to the code histogram. And we find a proper normalization mechanism after PCA can improve the discriminative ability of the code histogram. Using two simple unsupervised learning methods, we obtain a highly discriminative and compact face representation, the learning-based (LE) descriptor.

Many recent researches also apply learning approaches in face recognition, such as subspace learning [30], [31], metric learning [12], high-level trait learning [17], discriminant model learning [25], [36], [37], but few of these works focus on the issue of local feature encoding [18], [29] and the study of descriptor compactness. Though Ahonen *et al.* [2] tried K-means cluster to build local filter response codebook, they argued manual thresholding is faster and more robust.

Besides the representation, the matching also plays an important role. In most practices, the face is aligned by a similarity or affine transformation using detected face landmarks. Such 2D holistic alignment is not sufficient to handle large pose deviations from the frontal pose. Further, the large localization error of any landmark will result in misalignment of the whole face. 3D alignment [3] is more principled but error-prone and computationally intensive.

Wright *et al.* [38] recently encoded the geometric information into descriptors and used an implicit matching algorithm to deal with the misalignment and pose problem. Gang [14] demonstrated that a simple elastic and partial matching metric can also handle pose change and clutter background.

To explicitly handle large pose-variance, we propose a pose-adaptive matching method. We found that a specific face component contributes differently when the pose combinations of input face pairs are different. Based on this observation, we train a set of pose-specific classifiers, each for one specific pose combination, to make the final decision.

Combining a powerful learning-based descriptor and a pose-adaptive matching scheme, our system achieves the leading performance on both the LFW [16] and the Multi-PIE [11] benchmarks. We will describe our methods in detail in Chapter 2. Experiments will be presented in Chapter 3. We will conclude the thesis and discuss several possible future work in Chapter 4.

□ **End of chapter.**

Chapter 2

Learning-based descriptor for face recognition

2.1 Overview of framework

Pipeline overview. Our system is a two-level pipeline: the upper-level is the learning-based descriptor pipeline while the bottom-level is the pose-adaptive face matching pipeline.

As shown in Figure 2.1, we first use a standard fiducial point detector [19] to extract face landmarks. Nine different components (*e.g.*, nose, mouth) are aligned separately based on detected landmarks. The resulting component images are fed into a DoG filter (with $\sigma_1 = 2.0$ and $\sigma_2 = 4.0$) [14] to remove both low-frequency and high-frequency illumination variations. In each component image, a low-level feature vector is obtained at each pixel and encoded by our learning-based encoder. The final component representation is a compact descriptor (LE descriptor) generated by the concatenated patch histogram of the encoded features after PCA reduction and normalization. The component similarity is measured by L_2 distance between corresponding LE descriptors of the face pair. The

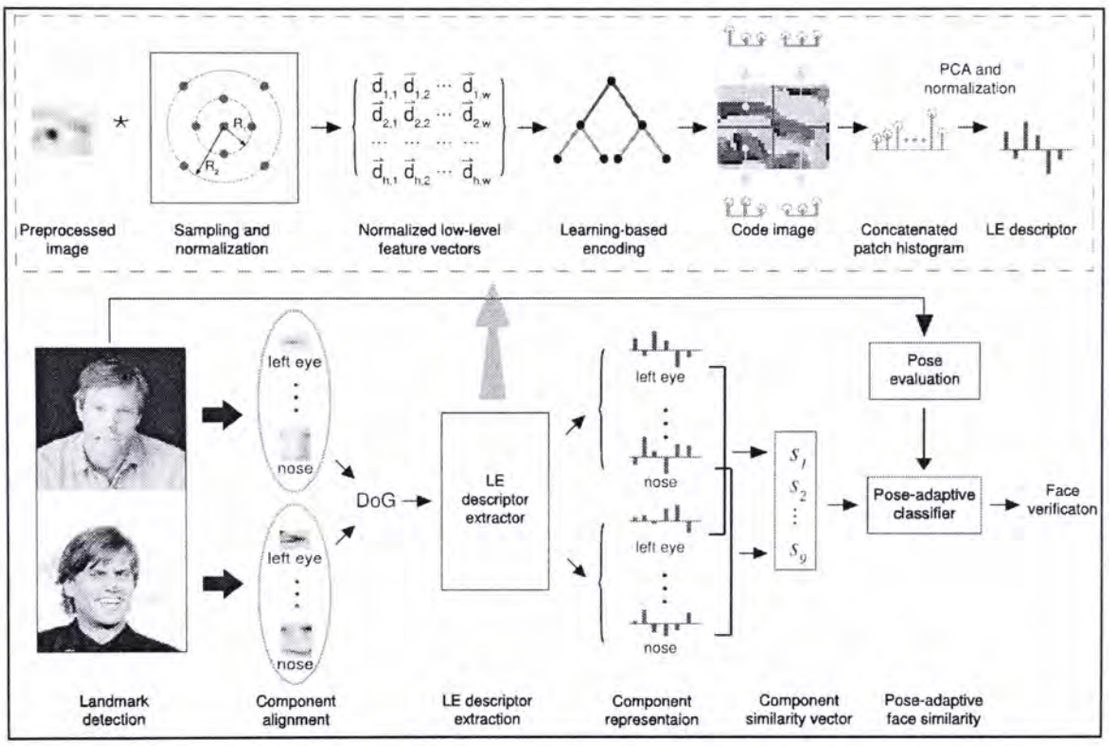

Figure 2.1: The proposed LE descriptor pipeline and the pose-adaptive face matching framework.

resulting 9 component similarity scores are fed into a pose-adaptive classifier, consisting of a set of pose-specific classifiers. The pose-specific classifier optimized to the pose combination of the matching pair gives the final decision.

Experiment overview. We mainly use the LFW benchmark [16] in our experiments and follow their protocol. The LFW standard test set consists of ten subsets and each subset contains 300 intra-personal/extra-personal pairs. The recognition algorithm needs to run ten times for formal evaluation purpose. At each time, one subset is chosen for testing and the other nine are used for training. The final average recognition performance serves as the evaluation criterion.

2.2 Learning-based descriptor extraction

In this section, we describe the critical steps in the learning-based (LE) descriptor extraction. In order to study the LE descriptor's power precisely, all the experiments in this section are conducted in holistic face level, without using component-level pose adaptive matching.

2.2.1 Sampling and normalization

At each pixel, we sample its neighboring pixels in the ring-based pattern to form a low-level feature vector. We sample $r * 8$ pixels at even intervals on the ring of radius r. Figure 2.2 shows four effective sampling patterns we found in an empirical manner. We extensively varied the parameters (*e.g.*, ring number, ring radius, sampling num-

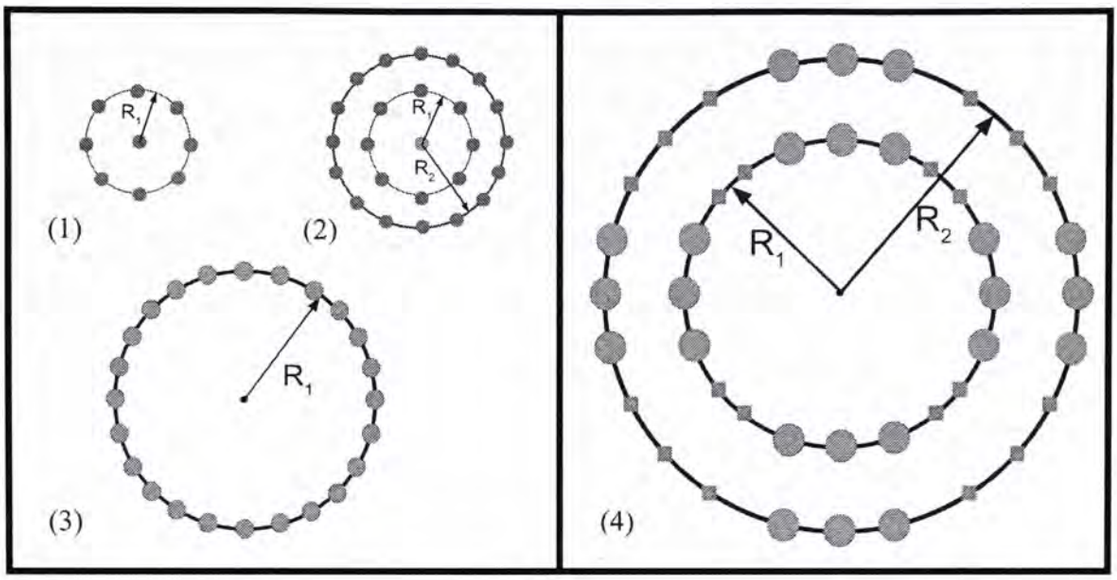

Figure 2.2: Four typical sampling methods used in our experiments: (1) $R_1 = 1$, with center; (2) $R_1 = 1, R_2 = 2$, with center; (3) $R_1 = 3$, no center; (4) $R_1 = 4, R_2 = 7$, no center. (The sampling dots on the green-square labeled arcs are omitted for better visuality).

ber of each ring) but found the differences among good patterns are not significant - no more than 1% on the LFW benchmark. The 2^{nd} pattern in Figure 2.2 is our best single pattern and we use it as our default sampling method.

Although the performances of single patterns are similar, combining them together may give us a chance to exploit the complementary information captured by different sampling methods. We will discuss the use of multiple patterns later in this section.

After the sampling, we normalize the sampled feature vector into unit length. Such normalization combined with DoG preprocessing makes the feature vector invariant to local photometric affine change.

2.2.2 Learning-based encoding and histogram representation

Next, an encoding method is applied to encode the normalized feature vector into discrete codes. Unlike many handcrafted encoders, in our approach, the encoder is specifically trained for the face in an unsupervised manner from a set of training face images. We have tried three unsupervised learning methods: K-means, PCA tree [8], and random-projection tree [8]. While K-means is commonly used to discover data clusters, random-projection tree and PCA tree are recently proved effective for vector quantization. In our implementation, random-projection tree and PCA tree recursively split the data based on uniform criterion, which means each leaf of the tree is hit by the same number of vectors. In other words, all the quantized codes have a similar emergence frequency in the vector space (as shown in Figure 1.2).

After the encoding, the input image is turned into a "code" image (Figure 2.1). Following the method described in Ahone *et al.*'s work [1], the encoded image is divided into a grid of patches (5×7 patches for the holistic face (84×96) used in this section). A histogram of the LE codes is computed in each patch and the patch histogram is concatenated to form the descriptor of the whole face image.

The choice of the learning method and the code number are important for our learning-based encoding. Figure 2.3 shows the performance comparison of the three learning methods under different code number setting. We select 1,000 images from the LFW training set to train our learning-based encoders. On each image, a number of 8,064 (=84×96) feature vectors are sampled as training examples. We varied the code number from 4 to 131,072 (=2^{17}) and plot-

ted the recognition rate (we stopped testing K-means after reaching 2^9 codes since the computation becomes intractable). Notice that random-projection tree slightly outperforms the other two and thus is adopted in the following as default. We compare our LE descriptor with LBP (59-bin), HOG (8-bin), and Gabor [35] on the LFW. Our LE descriptors start to beat existing descriptors (LBP 72.35%, HOG 71.25%, and Gabor 68.53%) when the code number reaches 32. And our LE descriptor achieves 77.78% rate when the code number reaches 2^{15}.

2.2.3 PCA dimension reduction

If we use the concatenated histogram directly as the final descriptor, the resulting face feature may be too large (e.g., 256 codes $\times$ 35 patch = 8,960 dimension). A large feature not only limits the number of faces which can be loaded into memory, but also slows down the recognition speed. This is very important for the applications that need to handle a large number of faces, for example, recognizing all face photos on a desktop. To reduce the feature size, we apply Principle Component Analysis (PCA) [28] to compress the concatenated histogram, and call the compressed descriptor as our final learning-based (LE) descriptor.

Surprisingly, we found that PCA compression substantially improves the performance if a simple normalization is applied after the compression. Figure 2.4 shows the recognition rates of LE descriptors with different normalization methods. Without the normalization, the compressed feature is inferior to the uncompressed one by 6% points. But with L_1 or L_2 normalization, the PCA version can be

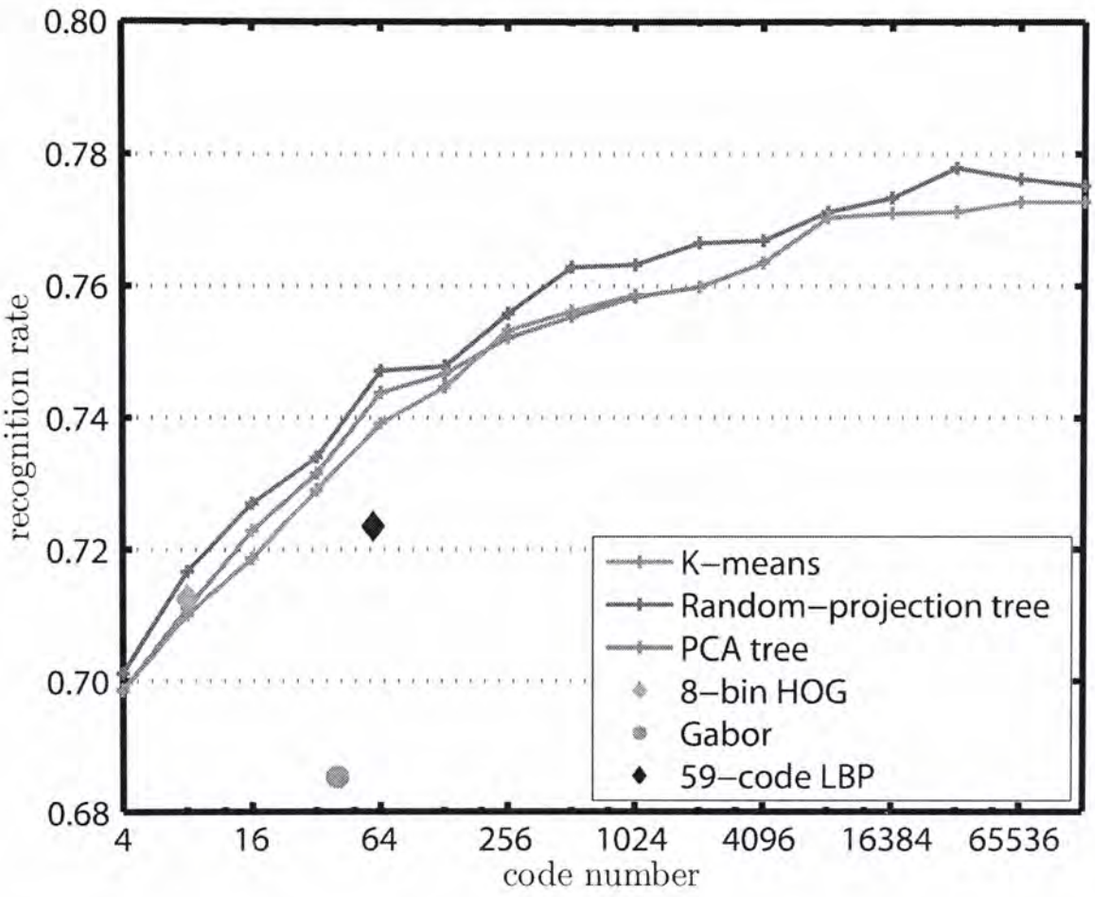

Figure 2.3: Performance comparison vs. learning method. We studied the recognition performance of the LE descriptors using three learning methods (random projection tree, PCA-tree, and K-means) under different code number settings. We also gave several existing descriptors' results for comparison.

5% higher. This result reveals the angle difference between features is most essential for the recognition in the compressed space. To confirm this key observation, we also tried to apply PCA compression to LBP. We repeated the same compression and normalization operations and also found simple normalization can boost uncompressed LBP's performance 3% points while skipping such step will detract it 5% points.

To obtain the optimal setting for the LE descriptor, we extensively studied the parameter combination of code number and PCA dimension. For large code number shows little performance advantage after PCA compression, we choose 256 code and 400 PCA-dimension as our default setting in the following experiments.

Our default LE descriptor achieves recognition rate as high as 81.22%, which significantly outperforms previous descriptors, using only 400-dimension feature vector for the holistic face, about 20% the size of the 59-code LBP descriptor. This demonstrated that our descriptor extraction pipeline (pre-processing, sampling and normalizing, learning-based encoding, and dimension reduction) is very effective for producing a compact and highly discriminative descriptor.

2.2.4 Multiple LE descriptors

As discussed in Section 2.2.1, our flexible sampling method enables us to generate a class of complementary LE descriptors, and the combination of multiple LE descriptors may achieve better performance. In this thesis, we take a simple approach by training a linear SVM [5] to combine the similarity scores generated by different LE

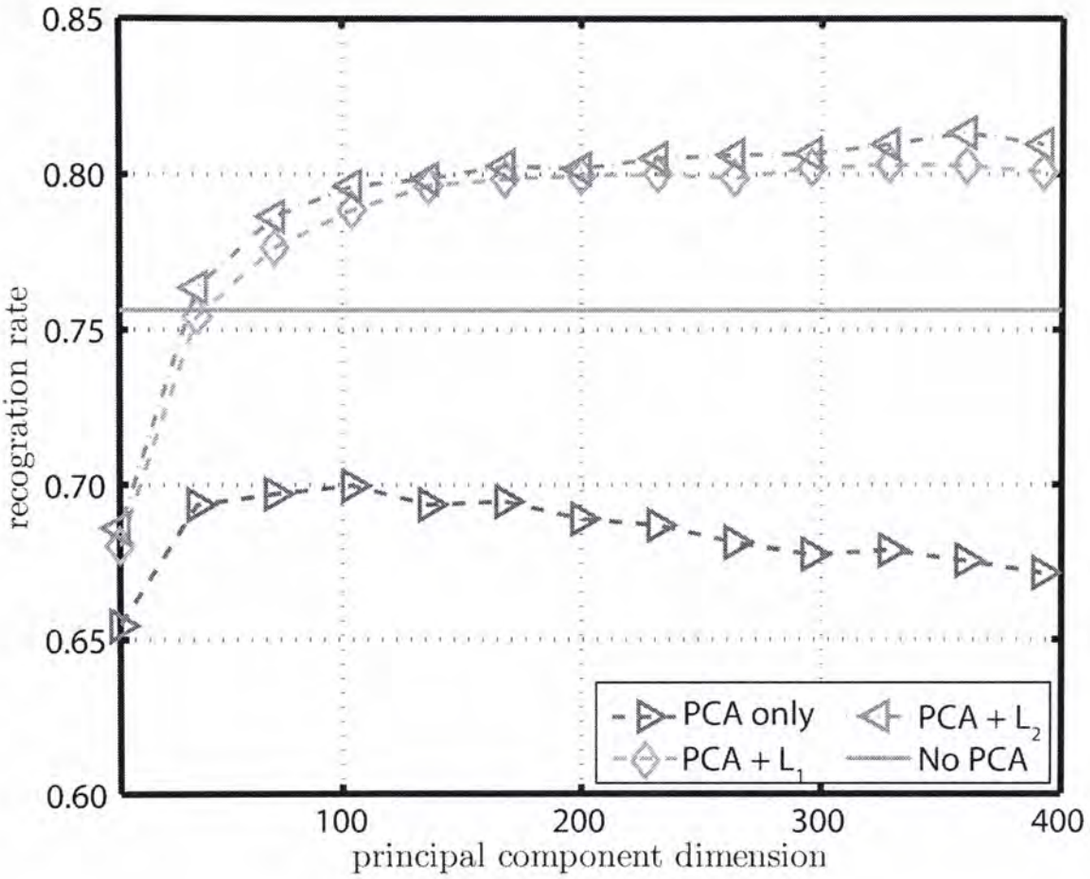

Figure 2.4: Investigate the effects of the PCA dimension with different normalization methods. After applying PCA compression to the concatenated patch histogram vector, we normalize the resulting vector with different normalization methods and then compute the similarity score with L_2 distance.

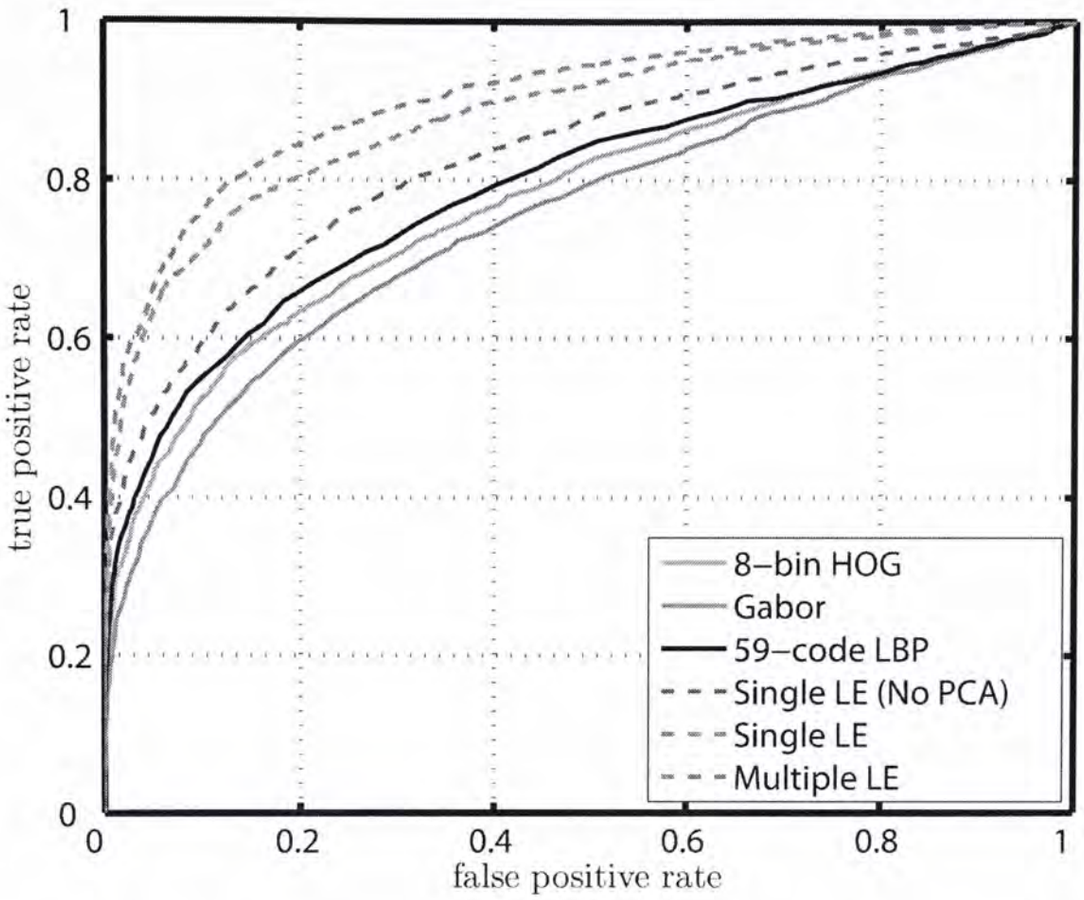

Figure 2.5: ROC curve comparison between our LE descriptors and existing descriptors.

descriptors. Generally, the combination can always achieve better result. In our experiments, the combination of four LE descriptors (shown in Figure 2.2) obtained the best performance on the LFW. Figure 2.5 gives the comparison curves of different descriptors.

2.3 Pose-adaptive matching

In the previous section, we use 2D holistic alignment and matching for the comparison purpose. In this section, we will show that a

pose-adaptive matching at the component-level can effectively handle large pose variation and further boost the recognition accuracy.

2.3.1 Component-level face alignment

Instead of using a 2D holistic (similarity) alignment on the whole face, we align 9 face components (shown in Figure 2.6) separately using similarity transform. For each component, two landmarks are selected from the five detected fiducial landmarks (eyes, nose, and mouth corners) to determine the similarity transformation (details in Table 2.1). Compared with the 2D holistic alignment, the component-level alignment presents advantages in large pose-variant case. The component-level approach can more accurately align each component without balancing across the whole face. And the negative effect of landmark error will also be reduced. Figure 2.6 shows aligned components and Table 2.2 compares the performance of different alignment methods.

2.3.2 Pose-adaptive matching

Using the component-level alignment, the face similarity score is the sum of similarities between corresponding components. We found that each component contributes differently for the recognition when the pose combination of the matching pair is different. For example, the left eye is less effective when we match a frontal face and a left-turned face. Based on this observation, we take a simple pose-adaptive matching method.

Firstly, we categorize the pose of the input face to one of three poses (frontal (F), left (L), and right (R)). To handle this pose cate-

Figure 2.6: Fiducial points and component alignment.

Component	Selected landmarks
Forehead	left eye + right eye
Left eyebrow	left eye + right eye
Right eyebrow	left eye + right eye
Left eye	left eye + right eye
Right eye	left eye + right eye
Nose	nose tip + nose pedal*
Left cheek	left eye + nose tip
Right cheek	right eye + nose tip
Mouth	two mouth corners

Table 2.1: Landmark selection for component alignment. (* means the pedal of the nose tip on the eye line.)

Alignment mode	Recog. rate
2-Point holistic	$79.85\% \pm 0.42\%$
5-Point holistic	$81.22\% \pm 0.53\%$
Component	$82.73\% \pm 0.43\%$

Table 2.2: Recognition rate vs. alignment mode.

gory, three images are selected from the Multi-PIE dataset, one image for each pose, and the other factors in these three images, such as person identity, illumination, expression remain the same. After measuring the similarity between these three gallery images and the probe face, the pose label of the most alike gallery image is assigned to the probe face.

Given the estimated pose of each face, the pose combinations of a face pair could be {FF, LL, RR, LR (RL), LF (FL), RF (FR)}. Our final pose-adaptive classifier consists of a set of linear SVM classifiers, each trained by a subset of training pairs with a specific pose combination. The "best-fit" classifier having the same pose combination with the input matching pair makes the final decision. Through pose-adaptive matching, we explicitly handle the large pose variation by this "divide-and-conquer" method.

2.3.3 Evaluations of pose-adaptive matching

To best evaluate the ability of pose change handling, we constructed a new test set from the LFW dataset by randomly sampling 3,000 intra-personal/extra-personal pairs for each pose combination. The total pair number in our new test set is $3,000 \times 6 = 18,000$. Note that this new test set is more challenging than the standard test data

Component	Image size	Patch division
Forehead	76×24	7×2
Left eyebrow	46×34	4×3
Right eyebrow	46×34	4×3
Left eye	36×24	3×2
Right eye	36×24	3×2
nose	24×76	2×7
Left cheek	34×46	3×4
Right cheek	34×46	3×4
Mouth	76×24	7×2

Table 2.3: Patch division for face components.

in the LFW due to the larger pose difference between the matching pair. We use half of them as the training set and the rest as the test set. Subjects are mutually exclusive in these two sets. And the patch division in component-level setting is shown in Table 2.3. Recognition performances were compared before ($76.20\% \pm 0.41\%$) and after ($78.30\% \pm 0.42\%$) pose-adaptive matching was adopted and results showed that the proposed technique is useful in such large pose-variant case.

□ **End of chapter.**

Chapter 3

Experiment

In this section, we report our final face recognition performance on the LFW benchmark systematically and then validate the excellent generalization ability of our system across different datasets.

3.1 Results on the LFW benchmark

We present our recognition results on the LFW benchmark in the form of ROC curves. Figure 3.1 shows comparison results for the validation of our proposed individual techniques. In Figure 3.1, "single LE + holistic" means that we only use the single best LE to represent the holistic face, and it is the baseline to show the power of LE without other techniques. "single LE + comp" indicates the application of component-level, pose-adaptive matching to the baseline single LE. Multiple LE descriptors are combined to form "multiple LE + holistic". And "multiple LE + comp" is our best performer. The accuracies for these four methods are $81.22\% \pm 0.53\%$, $82.72\% \pm 0.43\%$, $83.43\% \pm 0.55\%$, and $84.45\% \pm 0.46\%$. Despite the strong discriminant ability of the LE descriptor itself, the pose-

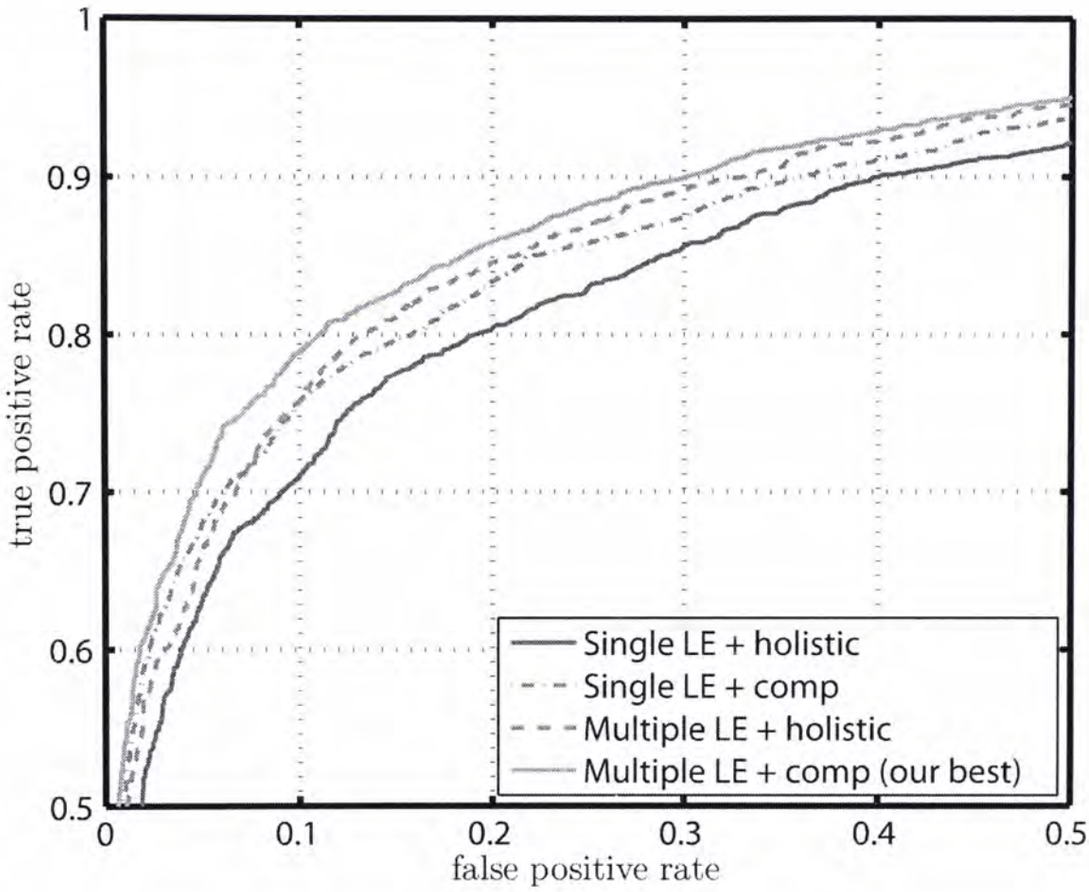

Figure 3.1: Demonstrate the effects of our proposed techniques on the LFW benchmark. Here, "holistic" means using holistic face representation while "comp" means component-level, pose-adaptive matching.

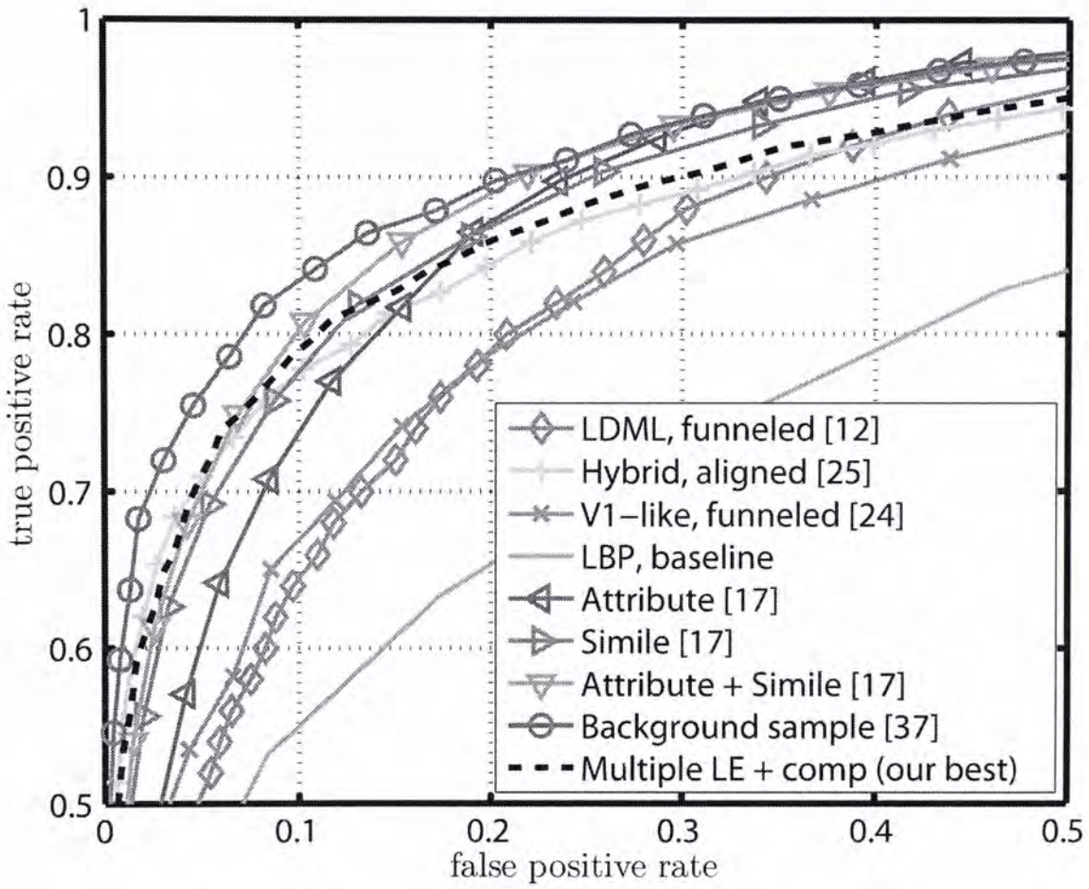

Figure 3.2: Face recognition comparison on the LFW benchmark in restrict protocol.

adaptive matching and multiple descriptor combination further enhance the recognition performance of our system.

Our best ROC curve is comparable with previous state-of-the-art methods, as shown in Figure 3.2. On the LFW benchmark, two new algorithms show the leading performance. Wolf *et al.*'s work [37] adopts the background learning by using the identity information within the training set. Kumar *et al.* [17] used the supervised learning to train high-level classifications through a huge volume of training images outside of the LFW dataset. These two methods [17, 37] both use additional information outside the LFW test protocol. So the comparison with other methods (including ours) in Figure 3.2 is not really fair. Additional training data or information may also improve other approaches.

Our system achieves the best performance when the standard test protocol is strictly respected [16]. More importantly, our work focuses on low-level face representation, which can be easily combined with previous algorithms to produce better performance.

3.2 Results on Multi-PIE

We also perform extensive experiments on the Multi-PIE dataset to verify the generalization ability of our approach. The Multi-PIE dataset contains face images from 337 subjects, imaged under 15 view points and 19 illumination conditions in 4 recording sessions. Large differences exist between LFW and Multi-PIE, considering the pose compositions, illumination variance, and resolution. Moreover, Multi-PIE is collected under a controlled setting systematically simulating the effects of pose, illumination, and expression. On the

Descriptor	Recog. rate on Multi-PIE
59-code LBP	$84.30\% \pm 0.89\%$
8-bin HOG	$84.02\% \pm 0.66\%$
Gabor	$86.42\% \pm 0.85\%$
single LE + holistic	$\mathbf{91.58\% \pm 0.50\%}$
single LE + comp	$\mathbf{92.12\% \pm 0.52\%}$
multiple LE + holistic	$\mathbf{92.20\% \pm 0.49\%}$
multiple LE + comp	$\mathbf{95.19\% \pm 0.46\%}$

Table 3.1: Recognition performance on the Multi-PIE dataset.

other hand, the LFW is more close to the real-life setting since its faces are selected from news images. For these reasons, training on one dataset and testing on the other can better demonstrate the generalization ability of a recognition system.

Similar to the LFW benchmark, we randomly generate 10 subsets of face images with Multi-PIE, each has 300 intra-personal and 300 extra-personal image pairs. The identities of subjects are mutually exclusive among these 10 subsets, and cross-validation mode similar to LFW is applied. The default "single LE" descriptor and "multiple LE" descriptors trained on the LFW benchmark are adopted in the experiments.

As shown in Table 3.1, the single LE with holistic face representation outperforms the commonly used descriptors more than 5 points, and pose-specific classifiers trained on the LFW dataset also perform well on the Multi-PIE dataset. All these results demonstrated the excellent generalization ability of our system.

□ **End of chapter.**

Chapter 4

Conclusion and future work

4.1 Conclusion

We have introduced a new approach for 2-D face recognition using learning-based (LE) descriptor and pose-adaptive matching. Our whole system (algorithm) involves the following main steps:

- Locate the fiducial points of the face image with a detector and then align its nine components separately with the detected landmarks.
- Process the aligned components with a DoG filter and then encode each pixel of the component images with a learned encoder. Each component image is converted into a code image.
- Extract each component code image's concatenated patch histogram and apply the PCA to get a more compact LE descriptor.
- Given a pair of face images, their corresponding component similarity scores (with angle distance metric) are fed into a

pose-adaptive classifier to get the final similarity measure between them.

We validated our recognition system on the LFW benchmark and demonstrated its excellent generalization ability on Multi-PIE.

4.2 Future work

Although good results have been obtained, there are still some problems and interesting directions that have not been explored.

- In this work, the face micro-pattern encoding is learned but the pattern sampling is still manually designed. Automating this step with learning techniques [33] may produce a more powerful descriptor for face recognition.
- We apply unsupervised algorithm to learn the encoding. A heuristic code emergence criterion is used. Since our ultimate purpose is the recognition performance, it is more suitable to adopt a supervised learning algorithm to tune the encoding step to make the final system more discriminative in differentiating persons.
- In this work, Principal Component Analysis (PCA) is used to obtain a compact form LE descriptor. Though simple and easy to implement, unsupervised dimension reduction techniques are generally inferior to supervised ones in boosting the recognition performance. Given various supervised dimension reduction algorithms available [39], [10], [32], it is quite promising to apply them into the descriptor compression.

- Combining multiple LE descriptors is not fully studied in our work. A probable better algorithm is treating different LE descriptors and different feature and then formulate the combining as an ensemble learning [4] [9] problem to solve.

□ **End of chapter.**

Bibliography

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face Recognition with Local Binary Patterns. *Lecture Notes in Computer Science*, pages 469–481, 2004.

[2] T. Ahonen and M. Pietikäinen. Image description using joint distribution of filter bank responses. *Pattern Recognition Letters*, 30(4):368–376, 2009.

[3] V. Blanz and T. Vetter. Face recognition based on fitting a 3 D morphable model. *IEEE Transactions on pattern analysis and machine intelligence*, 25(9):1063–1074, 2003.

[4] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[5] C. Chang and C. Lin. LIBSVM: a library for support vector machines, 2001. *Software available at http://www. csie. ntu. edu. tw/cjlin/libsvm*, 2001.

[6] J. Cui, F. Wen, R. Xiao, Y. Tian, and X. Tang. EasyAlbum: an interactive photo annotation system based on face clustering and re-ranking. In *Proc. of the SIGCHI conference on human factors in computing systems*, page 376, 2007.

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. CVPR*, 2005.

[8] Y. Freund, S. Dasgupta, M. Kabra, and N. Verma. Learning the structure of manifolds using random projections. In *Neural Information Processing System*, 2007.

[9] Y. Freund and R. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory*, pages 23–37, 1995.

[10] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighbourhood components analysis. *Neural Information Processing System*, 17:513–520, 2005.

[11] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *International Conference on Automatic Face and Gesture Recognition*, 2008.

[12] M. Guillaumin, J. Verbeek, C. Schmid, I. LEAR, and L. Kuntzmann. Is that you? Metric learning approaches for face identification. In *Proc. ICCV*, 2009.

[13] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. *IEEE Transactions on pattern analysis and machine intelligence*, 27(3):328–340, 2005.

[14] G. Hua and A. Akbarzadeh. A robust elastic and partial matching metric for face recognition. In *Proc. ICCV*, 2009.

[15] P. Hua, G. Viola and S. Drucker. Face recognition using discriminatively trained orthogonal rank one tensor projections. In *Proc. CVPR*, 2007.

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

[17] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. Attribute and Simile classifiers for face verification. In *Proc. ICCV*, 2009.

[18] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.

[19] L. Liang, R. Xiao, F. Wen, and J. Sun. Face Alignment via Component-based Discriminative Search. In *Proc. ECCV*, 2008.

[20] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[21] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[22] N. Peter, P. João, and J. David. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997.

[23] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation mthodology for face-recognition algorithms. *IEEE*

Transactions on pattern analysis and machine intelligence, 22(10):1090–1104, 2000.

[24] N. Pinto, J. DiCarlo, and D. Cox. How far can you get with a modern face recognition test set using only simple features. In *Proc. CVPR*, 2009.

[25] Y. Taigman, L. Wolf, T. Hassner, and I. Tel-Aviv. Multiple One-Shots for utilizing class label information. In *British Machine Vision Conference*, 2009.

[26] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Lecture Notes in Computer Science*, 4778:168, 2007.

[27] E. Tola, V. Lepetit, and P. Fua. A fast local descriptor for dense matching. In *Proc. CVPR*, 2008.

[28] M. Turk and A. Pentland. Face Recognition using Eigenfaces. In *Proc. CVPR*, 1991.

[29] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *Proc. CVPR*, 2003.

[30] X. Wang and X. Tang. A unified framework for subspace face recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 26(9):1222–1228, 2004.

[31] X. Wang and X. Tang. Random sampling for subspace face recognition. *International Journal of Computer Vision*, 70(1):91–104, 2006.

[32] K. Weinberger and L. Saul. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research*, 10:207–244, 2009.

[33] S. Winder and M. Brown. Learning local image descriptors. In *Proc. CVPR*, 2007.

[34] S. Winder, G. Hua, and M. Brown. Picking the best DAISY. In *Proc. CVPR*, 2009.

[35] L. Wiskott, J. Fellous, N. Krüger, and C. Von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):775–779, 1997.

[36] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Faces in Real-Life Images Workshop in ECCV*, 2008.

[37] L. Wolf, T. Hassner, and Y. Taigman. Similarity scores based on background samples. In *Proc. ACCV*, 2009.

[38] J. Wright and G. Hua. Implicit elastic matching with random projections for pose-variant face recognition. In *Proc. CVPR*, 2009.

[39] E. Xing, A. Ng, M. Jordan, and S. Russell. Distance metric learning with application to clustering with side-information. *Neural Information Processing System*, pages 521–528, 2003.

[40] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Li. Face detection based on multi-block lbp representation. *Lecture Notes in Computer Science*, 4642:11, 2007.