Fusing Gabor and LBP Feature Sets for Kernel-Based Face Recognition

Xiaoyang Tan, Bill Triggs

Fusing Gabor and LBP Feature Sets for Kernel-based Face Recognition

Xiaoyang Tan and Bill Triggs

INRIA & Laboratoire Jean Kuntzmann, 655 avenue de l'Europe, Montbonnot 38330, France
{xiaoyang.tan,bill.triggs}@imag.fr

Abstract. Extending recognition to uncontrolled situations is a key challenge for practical face recognition systems. Finding efficient and discriminative facial appearance descriptors is crucial for this. Most existing approaches use features of just one type. Here we argue that robust recognition requires several different kinds of appearance information to be taken into account, suggesting the use of heterogeneous feature sets. We show that combining two of the most successful local face representations, Gabor wavelets and Local Binary Patterns (LBP), gives considerably better performance than either alone: they are complimentary in the sense that LBP captures small appearance details while Gabor features encode facial shape over a broader range of scales. Both feature sets are high dimensional so it is beneficial to use PCA to reduce the dimensionality prior to normalization and integration. The Kernel Discriminative Common Vector method is then applied to the combined feature vector to extract discriminant nonlinear features for recognition. The method is evaluated on several challenging face datasets including FRGC 1.0.4, FRGC 2.0.4 and FERET, with promising results.

1 Introduction

One of the key challenges for face recognition is finding efficient and discriminative facial appearance descriptors that are resistant to large variations in illumination, pose, facial expression, ageing, partial occlusions and other changes [32]. Most current recognition systems use just one type of features. However for complex tasks such as face recognition, it is often the case that no single feature modality is rich enough to capture all of the classification information available in the image. Finding and combining complementary feature sets has thus become an active research topic in pattern recognition, with successful applications in many challenging tasks including handwritten character recognition [9] and face recognition[16].

In this paper, we show that face recognition performance can be significantly improved by combining two of the most successful local appearance descriptors, Gabor wavelets [12, 28, 15] and Local Binary Patterns (LBP) [18, 19, 2]. LBP is basically a fine-scale descriptor that captures small texture details. Local spatial invariance is achieved by locally pooling (histogramming) the resulting texture codes. Given that it is also very resistant to lighting changes, LBP is a good choice for coding fine details of facial appearance and texture. In contrast, Gabor features [12, 28, 15] encode facial shape and appearance information over a range of coarser scales (although they have also been used as a preprocessing stage for LBP feature extraction [30]). Both representations are

rich in information and computationally efficient. Their complementary nature makes them good candidates for fusion.

Here we evaluate and normalize the two modalities independently before combining them (although some works argue that it can be more effective to fuse modalities at an earlier stage of processing [10]). Both feature sets are high-dimensional (typically at least 10^4-D) and simply concatenating them would tend to exacerbate any 'curse of dimensionality' problems. To counteract this we run dimensionality reduction on each modality before fusion. Many dimensionality reduction techniques could be considered – Principal Component Analysis (PCA) [8], Independent Component Analysis (ICA) [20], Non-negative Matrix Factorization (NMF) [14], and Canonical Correlation Analysis (CCA) [13] to mention only some of the linear ones – but here we find that simple PCA suffices. The reduced feature vectors are separately normalized before being concatenated into a single combined feature vector. Finally the Kernel Discriminative Common Vector (KDCV) [4] method is applied to the combined feature vector to provide effective multi-class recognition from relatively few training examples. To illustrate the effectiveness of our approach we present experimental results on three state-of-the-art face recognition datasets containing large lighting variations similar to those encountered in natural images taken under uncontrolled conditions: the Face Recognition Grand Challenge 1.0.4 and 2.0.4 datasets [21] and FERET [22].

2 Related Work

Information fusion for visual recognition can occur at feature-level or at decision-level [10]. Feature-level methods combine several incoming feature sets into a single fused one that is then used in a conventional classifier, whereas decision-level ones combine several classifiers (e.g. based on distinct features) to make a stronger final classifier [11] (this is also called post-classification fusion or mixture of experts).

Face recognition is an area that is well-suited to the use (and hence fusion) of multiple classes of descriptors owing to its inherent complexity and need for fine distinctions. Much of the past work in this area adopts classifier-level fusion, *e.g*. [6,17]. For example, in [17] PCA, ICA and LDA provide the component subspaces for classifier combination. Each test sample is separately projected into these three subspaces and the resulting distance matrices are then fused to make the final decision using either the sum rule [11] or an RBF network. However motivated by the belief that the original features are a richer representation than distance matrices or individual classifier decisions, several works have studied feature-level fusion. J. Yang *et al*. [29] concatenate different features into a single vector and use Generalized PCA for feature exaction. C. Liu *et al*. [16] concatenate shape and texture information in a reduced subspace then use an enhanced Fisher classifier for recognition: their framework has similarities to ours but the underlying features and recognition methods are different.

Selecting appropriate and complementary component features is crucial for good performance. There is some work on fusing different biometric modalities (*e.g*. face and speech [3], face and fingerprint [23]), but most studies concentrate on fusing different representations of a single underlying modality (e.g. 2D and 3D facial shape in [6]). Our work belongs to this category, studying the effectiveness of fusing local 2D

texture descriptors at both the feature and the decision stages but focusing mainly on the feature based fusion. Our initial experiments selected two of the most successful local appearance descriptors, Gabor wavelets and LBP, as promising candidates for fusion. As both features are strongly normalized and quite local in nature, we also test whether the inclusion of a less-normalized feature set (raw gray levels) further improves the quality of the combined representation.

Finally, we use a kernel discriminant to extract as much information as possible from the resulting combined features. Methods such as Kernel Principal Component Analysis (KPCA) [25] have proven to be effective nonlinear feature extractors and here we use a related discriminative method, Kernel Discriminative Common Vectors (KDCV). Like other kernel methods, KDCV uses a nonlinear (kernel) mapping to implicitly transform the input data into a high dimensional feature space. It then selects and projects out an optimal set of discriminant vectors in this space, using the kernel trick to express the resulting computation in terms of kernel values in the input space. A simple Nearest Neighbour (NN) classifier is applied to the resulting KDCV feature vector. H. Cevikalp *et al*. [4] have shown that the combination of KDCV and NN significantly outperforms several other kernel methods including KPCA+LDA and SVM in related problems.

3 Fusing Gabor and LBP Feature Sets for Kernel-based Face Recognition

This section describes the components of our face recognition system in detail: Gabor and LBP features, PCA dimensionality reduction and feature fusion, Kernel DCV feature extraction and Nearest neighbour recognition. The stages of processing are illustrated in Fig. 1.

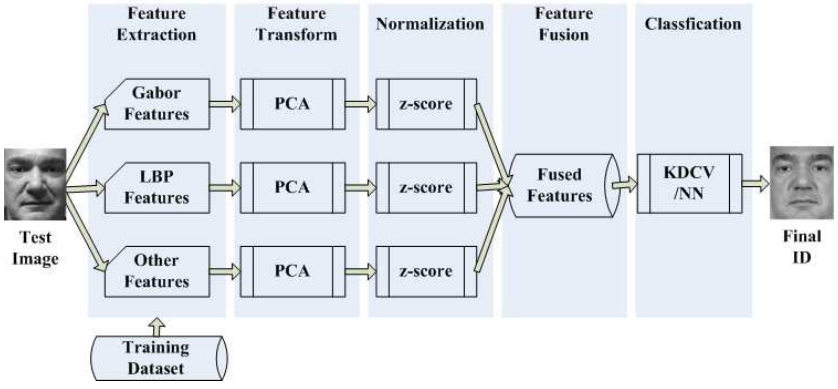

Fig. 1. The overall architecture of our face recognition system.

3.1 Gabor Features Representation

Gabor wavelets were originally developed to model the receptive fields of simple cells in the visual cortex and in practice they capture a number of salient visual properties including spatial localization, orientation selectivity and spatial frequency selectivity quite well. They have been widely used in face recognition since the pioneering work of Lades *et al*. [12]. Computationally, they are the result of convolving the image with a bank of Gabor filters of different scales and orientations and taking the 'energy image' (pixelwise complex modulus) of each resulting output image. The filters most commonly used in face recognition have the form

$$\psi_{\mu,\nu}(\mathbf{z}) = \frac{\|\mathbf{k}_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|\mathbf{k}_{\mu,\nu}\|^2 \|\mathbf{z}\|^2}{2\sigma^2}} [e^{i\mathbf{k}_{\mu,\nu}\mathbf{z}} - e^{\frac{-\sigma^2}{2}}] \tag{1}$$

where μ and ν define the orientation and scale of the Gabor kernels, $\mathbf{z} = (x, y)$, $\|\cdot\|$ denotes the norm operator, and the wave vector is $\mathbf{k}_{\mu,\nu} = k_\nu(\cos\phi_\mu, \sin\phi_\mu)$ where $k_\nu = k_{\text{max}}/f^\nu$ and $\phi_\mu = \pi\mu/8$ with k_{max} being the maximum frequency and f being the spacing factor between kernels in the frequency domain. Many face recognition studies use 40 Gabor wavelets of five different scales, $\nu \in \{0, 1, 2, 3, 4\}$, and eight orientations, $\mu \in \{0, \ldots, 7\}$, with $\sigma = 2\pi$, $k_{\text{max}} = \frac{\pi}{2}$, and $f = \sqrt{2}$. The Gabor wavelet representation is essentially the concatenated pixels of the 40 modulus-of-convolution images obtained by convolving the input image with these 40 Gabor kernels. In practice, before concatenation, each output image is downsampled according to the spatial frequency of its Gabor kernel and normalized to zero mean and unit variance.

3.2 Local Binary Patterns

Ojiala *et al*. [18] introduced the Local Binary Pattern operator in 1996 as a means of summarizing local gray-level structure. The operator takes a local neighbourhood around each pixel, thresholds the pixels of the neighbourhood at the value of the central pixel and uses the resulting binary-valued image patch as a local image descriptor. It was originally defined for 3×3 neighbourhoods, giving 8 bit codes based on the 8 pixels around the central one. Formally, the LBP operator takes the form

$$LBP(x_c, y_c) = \sum_{n=0}^{7} 2^n \, s(i_n - i_c) \tag{2}$$

where in this case n runs over the 8 neighbours of the central pixel c, i_c and i_n are the gray-level values at c and n, and $s(u)$ is 1 if $u \geq 0$ and 0 otherwise. The LBP encoding process is illustrated in fig. 2.

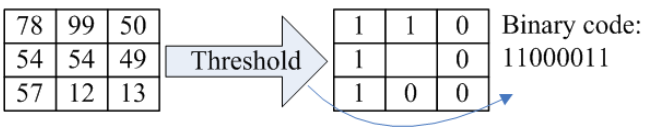

Fig. 2. Illustration of the basic LBP operator.

Two extensions to the original operator were made in [19]. The first defined LBP's for neighbourhoods of different sizes, thus making it feasible to deal with textures at different scales. The second defined the so-called *uniform patterns*: an LBP is 'uniform' if it contains at most one 0-1 and one 1-0 transition when viewed as a circular bit string. For example, the LBP code in fig. 2 is uniform. Uniformity is an important concept in the LBP methodology, representing primitive structural information such as edges and corners. Ojala *et al*. observed that although only 58 of the 256 8-bit patterns are uniform, nearly 90 percent of all observed image neighbourhoods are uniform. In methods that histogram LBP's, the number of bins can be thus significantly reduced by assigning all non-uniform patterns to a single bin, often without losing too much information.

LBP's are resistant to lighting effects in the sense that they are invariant to monotonic gray-level transformations, and they have been shown to have high discriminative power for texture classification [18]. T. Ahonen *et al*. introduced an LBP based method for face recognition [1] that divides the face into a regular grid of cells and histograms the uniform LBP's within each cell. Finally, the cell-level histograms are concatenated to produce a global descriptor vector. Like the Gabor descriptor, the LBP descriptor is usually high dimensional. For example, a 128×128 face image with 8×8 pixel cells produces a 15104-D LBP descriptor vector (256 patches with 59 entries/patch).

3.3 Feature-level Fusion with PCA

Before combining the Gabor and LBP features, we reduce their dimensionality to remove some of the redundancy and noise inherent in them. Given that we will later be feeding the results to a sophisticated nonlinear discriminant feature extractor (KDCV), we do not attempt to select discriminative directions at this stage. Instead we use simple PCA-based dimensionality reduction [8], retaining enough components to give KDCV scope to find good discriminant directions while still significantly reducing the size and redundancy of the data. Other methods could be used (ICA, CCA, NMF, etc), but PCA has the advantage of minimizing reconstruction error without making strong assumptions about the nature or use of the resulting data – we prefer to postpone such assumptions to the classifier stage.

Formally, let faces be represented by n-D vectors $\mathbf{x}$. PCA seeks a set of m orthogonal directions that capture as much as possible of the variability of the face set $\{\mathbf{x}\}$, or equivalently an m-D projection $\mathbf{y}$ of $\mathbf{x}$ from which $\mathbf{x}$ can be reconstructed with as little error as possible. Encoding these directions as an $n \times m$ matrix $\mathbf{U}$ with orthonormal columns, we seek to maximize $\mathrm{tr}(\mathbf{U}^T\mathbf{C}\mathbf{U})$ where $\mathbf{C}$ is the covariance matrix of the face set $\{\mathbf{x}\}$. This leads to an eigenvalue problem $\mathbf{C}\mathbf{U} = \mathbf{U}\boldsymbol{\Lambda}$ where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ is the matrix of eigenvalues of $\mathbf{C}$. For the best reconstruction we need to take the m largest eigenvalues. Then given $\mathbf{x}$, its projection is $\mathbf{y} = \mathbf{U}^T(\mathbf{x} - \boldsymbol{\mu})$ and its reconstruction is $\mathbf{x} \approx \mathbf{U}\mathbf{y} + \boldsymbol{\mu}$ where $\boldsymbol{\mu}$ is the mean of the training set $\{\mathbf{x}\}$. m necessarily satisfies $m \leq \min(n, N-1)$ where N is the number of training samples. In the experiments below $N \ll n$ and we preserve as much discriminant information as we can by taking m to be large enough to include all 'significantly non-zero' eigenvalues, so in practice $m \approx N - 1$.

Letting $\mathbf{x}_1 \in R^{n_1}$ and $\mathbf{x}_2 \in R^{n_2}$ be respectively the Gabor and LBP features of a face image, and $\mathbf{y}_1 = \mathbf{U}_1^T(\mathbf{x}_1 - \boldsymbol{\mu}_1)$, $\mathbf{y}_2 = \mathbf{U}_2^T(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ be the corresponding

centred and PCA-reduced vectors, the combined feature vector $\mathbf{z} \in R^{m_1+m_2}$ is then the '*z-score*' normalized combination

$$\mathbf{z} = (\mathbf{y}_1/\sigma_1, \mathbf{y}_2/\sigma_2)^T \tag{3}$$

where σ_1, σ_2 are the (scalar) standard deviations of $\mathbf{y}_1, \mathbf{y}_2$.

3.4 Seeking Optimal Discriminant Subspace with Kernel Trick

The next stage of the process extracts optimally discriminative nonlinear features from the combined feature vector $\mathbf{z}$. This is the only point at which class label information is used during training. It is based on a kernelized variant of Linear Discriminant Analysis (LDA) [24] called KDCV [4]. Classical LDA seeks a low-dimensional projection matrix $\mathbf{P}$ that maximizes the objective function

$$J(\mathbf{P}) = \frac{\mathbf{P}^T \mathbf{S}_B \mathbf{P}}{\mathbf{P}^T \mathbf{S}_W \mathbf{P}} \tag{4}$$

where $\mathbf{S}_B$ denotes the between-class and $\mathbf{S}_W$ the within-class scatter matrix of the training data. Formally the solution is given by the largest-eigenvalue eigenvectors of $\mathbf{S}_W^{-1}\mathbf{S}_B$. However this is not always stably computable. In particular, if there are more feature dimensions than training examples or if the examples lie in a lower dimensional affine subspace – both of which are true in our case – $\mathbf{S}_W$ is rank deficient and its inverse does not exist. The singularity is intrinsic in the sense that directions in the null space of $\mathbf{S}_W$ have no observed covariance so LDA predicts that they should be infinitely discriminant. In particular, if $\mathbf{S}_B$ has a nontrivial projection along these directions, LDA considers the corresponding classes to be perfectly separable. Techniques proposed to solve this classical problem include the perturbation method [31], two stage PCA+LDA [26], and the *null space* methods pioneered by Chen *et al*. [7]. The latter have dominated research in recent years. They focus *only* on the null space of $\mathbf{S}_W$, so they are really complements to traditional LDA not stabilized variants of it. They optimize the *null space based* LDA criterion

$$J(\mathbf{P}) = \max_{|\mathbf{P}^T \mathbf{S}_W \mathbf{P}|=0} |\mathbf{P}^T \mathbf{S}_T \mathbf{P}| \tag{5}$$

where $\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W$ is the total scatter matrix of the training set. Cevikalp *et al*. [4] proved that the optimal discriminant subspace in the sense of (5) is the intersection of the null space $N(\mathbf{S}_W)$ of $\mathbf{S}_W$ and the range space $R(\mathbf{S}_T)$ of $\mathbf{S}_T$, and to find it one can first project the training set sample onto $N(\mathbf{S}_W)$ and then apply PCA. This method is called Discriminative Common Vectors (DCV) [5] because all of the training samples in a class are projected to a unique vector in $N(\mathbf{S}_W)$ called the class' *common vector*. It can be shown that if the affine spans of the training sets are linearly separable, the corresponding common vectors are distinct resulting in a perfect recognition rate [5].

In many face recognition problems the class distributions are not separable using linear DCV but introducing a nonlinear embedding $\boldsymbol{\phi} : R^d \mapsto F$ into a kernel-induced feature space F allows them to be separated. Kernel DCV [4] finds projection vectors

that optimize the null space LDA criterion (5) in the induced feature space F by applying KPCA to project the training set onto the range space $R(\mathbf{S}_T^\phi)$ of $\mathbf{S}_T^\phi$, the total scatter matrix induced in F, then finding an orthonormal basis for the null space $N(\mathbf{S}_W^\phi)$ of the within-class scatter matrix $\mathbf{S}_W^\phi$ within this range space. The computation is kernalizable (expressible using inner products) precisely because it suffices to work within the span of $\mathbf{S}_T^\phi$: although $N(\mathbf{S}_W^\phi)$ typically contains many directions orthogonal to this, they are irrelevant as far as inter-class discrimination is concerned because test sample components in these directions are identical for all classes and hence not useful for discrimination based on this training set.

We will only summarize KDCV briefly here. See [4] for details. Let $\widetilde{\mathbf{K}}$ be the empirical kernel matrix of the training set, with eigendecomposition $\widetilde{\mathbf{K}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ where $\boldsymbol{\Lambda}$ is the diagonal matrix of nonzero eigenvalues. $\mathbf{U}$, the associated matrix of normalized eigenvectors, doubles as a basis for the span of $\mathbf{S}_T^\phi$. Let $\overline{\boldsymbol{\Phi}}$ be the matrix of the centered training set with respect to the empirical feature space. The matrix that projects the training set onto $R(\mathbf{S}_T^\phi)$ is then $\overline{\boldsymbol{\Phi}}\mathbf{U}\boldsymbol{\Lambda}^{-1/2}$. This is used to obtain the projected within-class scatter matrix $\widetilde{\mathbf{S}}_W^{\boldsymbol{\Phi}}$, from which a basis $\mathbf{V}$ for the null space of $\widetilde{\mathbf{S}}_W^{\boldsymbol{\Phi}}$ is obtained:

$$\mathbf{V}^T\, \widetilde{\mathbf{S}}_W^{\boldsymbol{\Phi}}\, \mathbf{V} = 0 \tag{6}$$

The optimal projection matrix $\mathbf{P}$ is then:

$$\mathbf{P} = \overline{\boldsymbol{\Phi}}\, \mathbf{U}\, \boldsymbol{\Lambda}^{-1/2}\, \mathbf{V} \tag{7}$$

3.5 Face Recognition in the Optimal Discriminant Subspace

When a face image is presented to the system, its Gabor and LBP representations are extracted, projected into their PCA subspaces, normalized separately (3) and integrated into a combined feature vector $\mathbf{z}_{\text{test}}$, which is then projected into the optimal discriminant space by

$$\boldsymbol{\Omega}_{\text{test}} = \mathbf{P}^T\, \boldsymbol{\phi}(\mathbf{z}_{\text{test}}) = (\mathbf{U}\, \boldsymbol{\Lambda}^{-1/2}\, \mathbf{V})^T\, \mathbf{k}_{\text{test}} \tag{8}$$

where $\mathbf{P}$ is the optimal projection matrix given by (7) and $\mathbf{k}_{\text{test}} \in R^M$ is a vector with entries $K(\mathbf{z}_m^i, \mathbf{z}_{\text{test}}) = \langle \boldsymbol{\phi}(\mathbf{z}_m^i), \boldsymbol{\phi}(\mathbf{z}_{\text{test}}) \rangle$, where $\boldsymbol{\phi}(\mathbf{z}_m^i)$ are the mapped training samples. The projected test feature vector $\boldsymbol{\Omega}_{\text{test}}$ is then classified using the nearest neighbour rule and the cosine 'distance'

$$d_{\cos}(\boldsymbol{\Omega}_{\text{test}}, \boldsymbol{\Omega}_{\text{template}}) = -\frac{\boldsymbol{\Omega}_{\text{test}}^T \boldsymbol{\Omega}_{\text{template}}}{\|\boldsymbol{\Omega}_{\text{test}}\| \|\boldsymbol{\Omega}_{\text{template}}\|} \tag{9}$$

where $\boldsymbol{\Omega}_{\text{template}}$ is a face template in the gallery set. Other similarity metrics such as L_1, L_2 or Mahalanobis distances could be used, but [15] found that the cosine distance performed best among the metrics it tested on this database, and our initial experiments confirmed this.

4 Experiments

We now present experiments designed to illustrate the effectiveness of the proposed method. Three publicly available databases containing large illumination variations

(a) (b)

Fig. 3. Examples of images from FRGC-104: (a) target images (upper row) and query images (lower row) without illumination preprocessing; (b) the corresponding illumination normalized images from the proposed preprocessing chain.

were used: Face Recognition Grand Challenge version 1 experiment 1.0.4 ('FRGC-104') and version 2 experiment 2.0.4 ('FRGC-204') [21], and the FERET dataset [22]. We first conducted a series of pilot experiments on the FRGC-104 dataset, then we verified the results on FERET and the challenging FRGC-204 dataset.

4.1 Experimental Settings

Prior to analysis, all images undergo geometric and photometric normalization to counter the effects of pose and illumination variations, local shadowing and highlights. First they are converted to 8 bit gray-scale, rigidly scaled and rotated to place the centers of the two eyes at fixed image positions using the eye coordinates supplied with the original datasets, and cropped to 128×128 pixels. Then they are photometrically normalized using the following sequence of steps: strong gamma compression; Difference of Gaussian (DoG) filtering; robust normalization of the range of output variations; and sigmoid-function based compression of any remaining signal highlights. A detailed description of this simple but very effective normalization procedure can be found in [27]. Some examples of preprocessed images are shown in Fig. 3.

The downsampling factor for the Gabor features is set to 64, resulting a dimensionality of 10 240 ($128 \cdot 128 \cdot 40/64$), while the cell size of the LBP features is set to 8×8 pixels, giving a dimensionality of 15 104. For the kernel method we tried polynomial kernels $k(\mathbf{x},\mathbf{y}) = (\langle \mathbf{x},\mathbf{y} \rangle)^n$ with degrees $n = 2, 3$ and Gaussian kernels $k(\mathbf{x},\mathbf{y}) = e^{-\|\mathbf{x}-\mathbf{y}\|^2/(2\sigma^2)}$ with scale parameter chosen on a validation set and reported the best result.

4.2 Results on FRGC-104

The FRGC-104 dataset [21] is challenging because although the gallery images were obtained under carefully controlled conditions, the probe images were captured in uncontrolled indoor and outdoor settings with large changes in illumination, appearance and expression. Fig. 3 shows some examples. For the experiments reported here the gallery contains 152 people with one image per person, while the probe set contains 608 images of the 152 subjects. For training we chose the 886 images of 198 subjects

with at least two images per subject from the FRGC-104 training set. There is no overlap between the training, gallery and probe sets. Besides the well-normalized LBP and Gabor local texture features, we also test whether the inclusion of raw gray-level image pixels can improve the results.

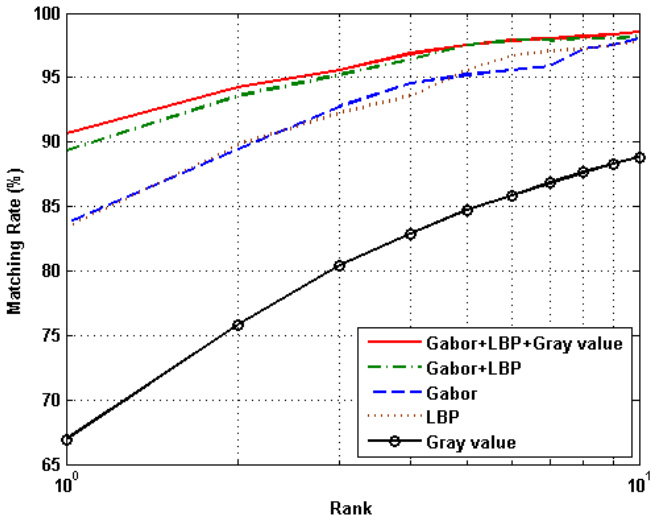

Fig. 4. The comparative recognition performance of KDCV/NN on different feature sets.

Fig. 4 shows the FRGC-104 performance of our Kernel DCV/NN method for several different types of input features. As expected the raw pixel features perform poorly owing to their sensitivity to various common appearance variations, while both Gabor wavelets and LBP features give much better, and here very similar, performance. However, fusing the Gabor and LBP features still provides a significant performance gain – about 6.0% relative to either feature set individually – which suggests that these two feature sets do indeed capture different and complementary information. Incorporating the somewhat unreliable information provided by raw gray-levels provides a modest further improvement, reaching a rank 1 recognition rate of over 90%. This suggests that there is scope for further improvement by including additional higher-quality feature sets.

We also checked the effects of using our combined features in several other popular face recognition frameworks including LDA, DCV and KPCA. The results are shown in Table 1. The recognition performance of every method was improved by using the combined features as input. Among the methods compared, KDCV consistently performs best, particularly on the combined features 'Gabor+LBP+Gray value'.

The influence of different PCA projection dimensions (as represented by the percentage of the total energy retained by the projection) is illustrated in fig. 5. The figure reveals a positive, albeit somewhat irregular, correlation between PCA energy (projec-

Input Features	LDA	DCV	KPCA	KDCV
Gabor	52.3*	82.2*	45.1*	83.7*
LBP	50.8*	78.6*	52.7	83.4*
Gray value	36.2*	63.2*	35.2*	66.9*
Gabor+LBP	56.1	89.1	50.0	89.3
Gabor+LBP+Gray value	**59.7**	**89.8**	**53.5**	**90.6**

Table 1. FRGC-104 recognition rate (%) for different feature sets and different recognition methods. The asterisks indicate performance differences that are statistically significant at the 5% level between the given method and the corresponding result in bold.

tion dimensionality) and recognition rate, underlining the importance of preserving as much as possible of the original information during the projection. In particular, increasing the energy retained from 70% to 85% gives a 30% improvement in recognition rate. However, note that the maximum possible PCA dimension is limited by the number of training samples and that for larger samples than those used here some overfitting may occur if all of the dimensions are used.

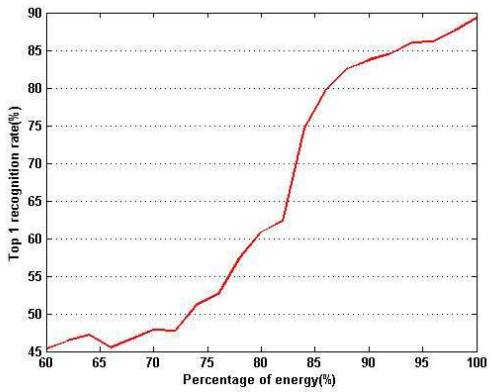

Fig. 5. The influence of PCA dimension (percentage of total energy preserved during the PCA) on FRGC-104 recognition rate.

To compare the relative effectiveness of feature-level and decision-level fusion we conducted some experiments based on a simple decision-level fusion method. We project each test image onto three KDCV discriminant subspaces trained respectively with Gabor wavelets, LBP features and gray-value features, and for each feature class we compute the cosine distances between the test image and the gallery templates. The *z-score* normalization procedure (3) is applied to the three distance matrices, and they are then combined by simple addition. As before, test samples are assigned to the class containing the closest template under the combined distance metric and we considered several different feature combinations.

Features	Gabor	LBP	Gray value	Gabor+LBP	Gabor+LBP+Gray value
CPU Time	73.9	74.2	72.8	60.5	61.96

Table 2. CPU times (s) for FRGC-104 recognition runs on a 2.8 GHZ single processor PC.

The results for 'Gabor + LBP' and 'Gabor + LBP + Gray value' are shown in fig. 6. The decision-level 'Gabor + LBP' method predominates. As a general rule, both decision-level and feature-level fusion benefited from using a mixture of different feature types. The main exception was that for decision-level fusion, the 'Gabor+LBP' scheme worked significantly better than the 'Gabor+LBP+Gray value' one. Decision-level fusion by simple averaging tends to be sensitive to the performance of the worst of its component classifiers, and to perform best when they are both *diverse* and *uniformly accurate*, whereas here the raw pixel based classifier is significantly weaker than the other two, thus decreasing the overall system performance. In contrast, feature-level fusion provided a performance increment for each new feature set included in its pool.

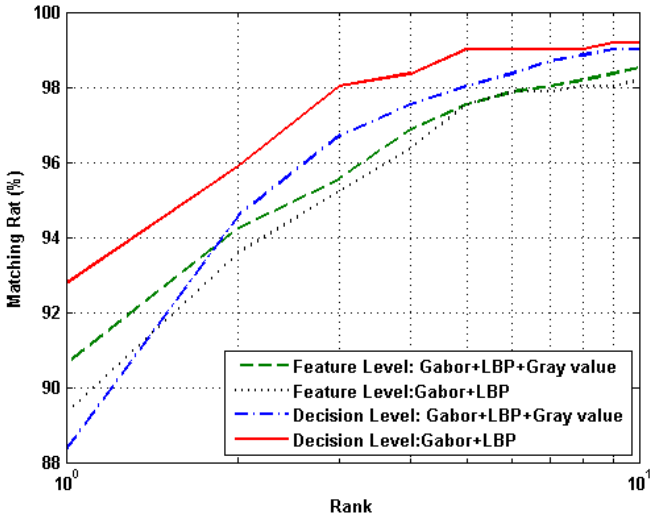

Fig. 6. The comparative face recognition performance of feature level fusion and decision-level averaging on FRGC-104.

Regarding computational cost, average CPU times for complete recognition runs on FRGC-104 on our 2.8 GHz single processor PC are shown in Table 2. Note that the combined feature sets actually have *lower* cost than the individual features: after reduction, the combined features have lower dimensionality than the individual ones and most of the run time is spent doing KDCV and NN search in this reduced space.

Method	fb	fc	dup1	dup2
Fisherfaces[30]	0.94	0.73	0.55	0.31
Best Results of [22]	0.96	0.82	0.59	0.52
Best Results of [1]	0.97	0.79	0.66	0.64
Best Results of [30]	0.98	0.97	0.74	0.71
Our 'Gabor+LBP' method	0.98	0.98	0.90	0.85

Table 3. Comparative recognition rates of various methods on the FERET partitions.

4.3 Results on FERET

Fig. 7. Some sample images from the FERET dataset.

A second series of experiments was conducted on the FERET dataset. This contains five standard partitions: 'fa' is the gallery containing 1196 grayscale images and 'fb', 'fc', 'dup1' and 'dup2' are four sets of probe images. The diversity of the probe images is across gender, ethnicity and illumination ('fc'), expression ('fb') and age/time ('dup1' and 'dup2'). Some examples of FERET images are shown in fig. 7. All of the images were preprocessed as described in section 4.1. The gallery set is always available to a face recognition system so in addition to the distributed training set we used the images in 'fa' to train the Kernel DCV classifier. As there is only one image per person in 'fa', these images do not contribute to the null space of the within-class scatter matrix, but they do help to shape the between-class distribution and to increase the dimensionality of final discriminative subspace.

We compared the proposed 'Gabor+LBP' method to several previously published results on FERET including Fisherfaces, the best result in FERET'97 [22], and the recent results of [1] and [30]. The rank-1 recognition rates of the different methods on the FERET probe sets are shown in table 3. The performance of the proposed method is comparable to or better than existing state-of-the-art results on this dataset, especially on the challenging probe sets 'dup1' and 'dup2'. Besides better performance, our method also requires much less memory than weighted LGBPHS [30], which allows it to to scale efficiently to large datasets such as FRGC version 2.

4.4 Results on FRGC-204

FRGC-204 is the most challenging FRGC experiment [21]. It extends the FRGC-104 dataset, defining a standard tripartite partition into a training set of 12,776 images (in-

cluding both images with controlled lighting and uncontrolled ones), a target set of 16,028 controlled images, and a query set of 8,014 uncontrolled images. Again the pre-processing method described in section 4.1 was used. To allow a better comparison with the state of the art on this dataset we used the training set of [15], which includes 6,388 images selected from the full FRGC-204 training set.

The results of FRGC version 2 experiments are usually reported using the Receiver Operating Characteristic (ROC) curves for Face Verification Rate (FVR) as a function of False Accept Rate (FAR). For a given distance matrix three types of ROC curves can be generated by the Biometric Experimentation Environment (BEE): ROC-I, ROC-II,and ROC-III, corresponding respectively to images collected within a semester, within a year, and between semesters. Owing to space limitations we report results only for ROC-III (the most commonly reported benchmark) – see fig. 8. The figure shows that the proposed 'Gabor+LBP' method increases the FVR over separate Gabor or LBP from 73.5% to 83.6% at 0.1% FAR. The best previous performance that we are aware of on this dataset at 0.1% FAR is 76.0% FVR [15].

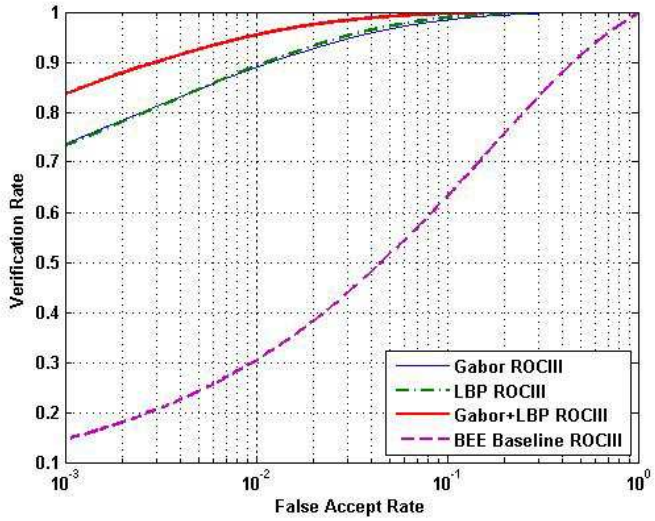

Fig. 8. FRGC-204 face recognition performance (ROC-III curves) for Gabor, LBP and Gabor+LBP methods. The FRGC baseline performance is also included for comparison.

5 Conclusions

This paper investigated the benefits of combining two of the most successful feature sets for robust face recognition under uncontrolled lighting: Gabor wavelets and LBP features. We found that these features are more complementary than might have been expected, with the combination having only around 2/3 of the errors of either feature set

alone. The method was tested in a novel face recognition pipeline that includes: robust photometric image normalization; separate feature extraction, PCA-based dimensionality reduction and scalar variance normalization of each modality; feature concatenation; Kernel DCA based extraction of discriminant nonlinear features; and finally cosine-distance based nearest neighbour classification in the KDCA reduced subspace. The proposed face recognition method is scalable to large numbers of individuals and easy to extend to additional feature sets. We illustrated its performance with a series of comparative experiments on the challenging FRGC version 1 experiment 4, FRGC version 2 experiment 4, and FERET datasets.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive suggestions and H. Cevikalp for helpful discussions. The research used the FERET and FRGC face datasets collected respectively under the FERET and FRGC programs.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *Proc. ECCV'04*, pages 469–481, Prague, 2005.

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE TPAMI*, 28(12), 2006.

[3] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz. Fusion of face and speech data for person identity verification. IDIAP-RR 03, IDIAP, 1999.

[4] H. Cevikalp, M. Neamtu, and M. Wilkes. Discriminative common vector method with kernels. *IEEE TNN*, 17(6):1550–1565, 2006.

[5] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana. Discriminative common vectors for face recognition. *IEEE TPAMI*, 27(1):4–13, 2005.

[6] K. Chang, K. Bowyer, P. Flynn, and X. Chen. Multi-biometrics using facial appearance, shape and temperature. In *Proc. AFGR'04*, pages 43–48, 2004.

[7] L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu. A new lda-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33(10):1713–1726, 2000.

[8] H.Hotelling. Analysis of a complex of statistical variables into principal components. *J.Educational Psychology*, 24:417–441, 1933.

[9] Y. S. Huang and C. Y. Suen. A method of combining multiple experts for the recognition of unconstrained handwritten numerals. *IEEE TPAMI*, 17(1):90–94, 1995.

[10] A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, 2005.

[11] J. Kittler, M. Hatef, R. P. Duin, and J. Matas. On combining classifiers. *IEEE TPAMI*, 20(3):226–239, 1998.

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

[13] P. L. Lai and C. Fyfe. Kernel and nonlinear canonical correlation analysis. *IJCNN*, 04:4614, 2000.

[14] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–56obtainable2, 2000.

[15] C. Liu. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE TPAMI*, 28(5):725–737, 2006.
[16] C. Liu and H. Wechsler. A shape- and texture-based enhanced fisher classifier for face recognition. *IEEE TIP*, 10(4):598–608, 2001.
[17] X. Lu, Y. Wang, and A. Jain. Combining classifiers for face recognition. In *Multimedia and Expo, 2003. ICME '03*, 2003.
[18] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29, 1996.
[19] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invarianat texture classification with local binary patterns. *IEEE TPAMI*, 24(7):971–987, 2002.
[20] P.Comon. Independent component analysis-a new concept? *Signal Processing*, 43:287–314, 1994.
[21] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek. Overview of the face recognition grand challenge. In *Proc. CVPR'05*, pages 947–954, San Diego, CA, 2005.
[22] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE TPAM*, 22(10):1090–1104, 2000.
[23] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recogn. Lett.*, 24(13):2115–2125, 2003.
[24] R.R.Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugen*, 7:179–188, 1936.
[25] B. Scholkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001.
[26] D. L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE TPAMI*, 18(8):831–836, 1996.
[27] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Proc. AMFG'07*, 2007.
[28] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE TPAMI*, 19(7):775–779, 1997.
[29] J. Yang and J.-Y. Yang. Generalized k-l transform based combined feature extraction. *Pattern Recognition*, 35(1):295–297, 2002.
[30] W. Zhang, S. Shan, W. Gao, and H. Zhang. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In *Proc. ICCV'05*, pages 786–791, Beijing, China, 2005.
[31] W. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant analysis of principal components for face recognition. In *FG '98*, page 336, Washington, DC, USA, 1998.
[32] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Survey*, 34(4):399–485, 2003.