Applying Scattering Operators for Face Recognition: A Comparative Study

Kuang-Yu Chang [1,3], Cheng-Fu Lin [2], Chu-Song Chen [1,2], and Yi-Ping Hung [1,3]
[1] *Institute of Information Science, Academia Sinica, Taipei, Taiwan.*
[2] *Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan.*
[3] *Dept. of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan.*
{kuangyu, song}@iis.sinica.edu.tw, chengfulin@citi.sinica.edu.tw, hung@csie.ntu.edu.tw

Abstract

Face identification is the problem of determining whether two face images depict the same person or not. This is difficult due to variations in scale, pose, lighting, background, expression, hairstyle, and glasses. Thus, a powerful feature descriptor with local-deformation tolerance ability and discriminating capability is essential to fulfill all these variations. In this paper, we present a local descriptor, scattering operator, which includes multi-scale and multi-direction co-occurrence information. It is computed with a cascade of wavelet decompositions and complex modulus. This scattering representation is locally translation invariant and can linearize deformations. We evaluate the abilities of this Gabor-based scattering operator by an effective face recognition paradigm and show that this descriptor outperforms the compared descriptors.

1. Introduction

Gabor transform has been known as one of the most powerful feature representations for face recognition [1][2]. A Gabor basis is constructed by a Gaussian kernel and a Fourier basis, which is a wavelet capturing local time-frequency property of an image. Two-dimensional Gabor textons can be obtained by a number of dilations and rotations of a mother wavelet. In human vision, it has also been found that Gabor filter approximates well the response of the simple cells in the visual cortex in early human visual system [3]. As reported in [1] and [2], Gabor wavelets achieved notable performance for face recognition in well-designed face recognition databases such as FERET.

However, applying the Gabor wavelets of every rotation and scale to an image patch (usually referred to as a Gabor jet) could not tolerate local deformation well. Recent features that are successful in object recognition and image matching, such as Scale-Invariant Feature Transform (SIFT) [4] and Histogram of Oriented Gradients (HOG) [5], can allow reasonable local deformation of a patch while still make distinct patches identifiable. A main issue in designing a powerful feature for a local image patch thus lies in the balance between the local-deformation tolerance ability and the discriminating capability. Recently, this problem has been studied in the design of scattering operators in [6][7]. In particular, a SIFT-like operator called DAISY[8] can approximate the SIFT feature by averaged wavelet coefficients with a partial derivative wavelet.

Unfortunately, tolerance of deformation and discrimination between patches might be two conflict goals, where the enhancement of one could reduce the other. The idea of scattering operators is to employ a series of operators that can tolerate different scales of local changes; the tolerance (or invariance) degrees are increased without losing discriminating abilities. These operators are then cascaded into a single descriptor.

Scattering operators have been used in handwritten digit recognition, texture [6] and audio classification [9]. In this paper, we apply scattering operators in unconstraint face recognition and matching. This descriptor could tolerate deformation caused by expression and head pose. It is because that some detailed characteristics are described by finer-scale descriptors and thus distinctiveness can be preserved.

The rest of this paper is organized as follows. First, we review the scattering operators in Section 2. The recognition method is described in Section 3. Experimental results are presented in Section 4, and Section 5 gives the conclusions.

2. Gabor-based Scattering Operators

Locally invariant feature descriptors (e.g., SIFT and HOG) provide robust representation for image classification and object recognition. These descriptors can be

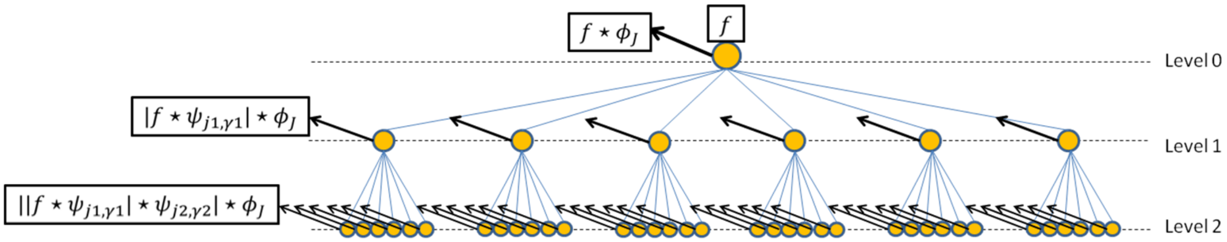

$f \star \phi_J$
f
Level 0
$|f \star \psi_{j1,\gamma1}| \star \phi_J$
Level 1
$||f \star \psi_{j1,\gamma1}| \star \psi_{j2,\gamma2}| \star \phi_J$
Level 2

Figure 1. Scattering operators are a cascade of wavelet decompositions.

seen as averaging the gradient orientations. Some local deformation could be tolerated by the process of averaging since the feature variability is reduced. Lower variability contributes to represent an image with invariant property. However, such process has some drawbacks of information reduction since the high frequencies are removed. Discarding high frequencies information may sacrifice the discriminating capability.

Brunna and Mallat [6] proposed a local descriptor called scattering operators. The descriptor reserves the invariant property of SIFT and HOG. Furthermore, it can recover the lost part in high frequencies by using multiple scales and orientation descriptors. The scattering operators are locally translation invariant including global translation invariant and can linearize small deformation. It retains many complex structures without losing high frequency information.

Let the filter $\psi_{j,\gamma}$ be constructed by scaling a filter ψ with 2^j and rotating $x \in \mathbb{R}^2$ by an angle γ.

$$\psi_{j,\gamma}(x) = 2^{-2j}\psi(2^{-j}R_\gamma x), \tag{1}$$

where $R_\gamma x$ stands for the rotation of x. In this paper, we use complex Gabor function for ψ that has been proven to be effective for texture analysis [10].

The resulted wavelet transform of f at a position x is a filter bank defined by

$$W_J f(x) = \begin{pmatrix} f \star \psi_{j,r}(x) \\ f \star \phi_J(x) \end{pmatrix}_{j<J,\ \gamma\in\Gamma}. \tag{2}$$

$\phi_J(x)$ is a low-pass filter that carries the low frequency information. In this paper, Gaussian filter is adopted as the low-pass filter $\phi_J(x)$.

Some descriptors achieve invariance property by averaging. The wavelet coefficient amplitude of f are averaged by $\phi_J(x)$ as follows:

$$|f \star \psi_{j,r}| \star \phi_J(x). \tag{3}$$

Convoluting with low pass filter can increase the invariant ability but sacrifice high frequencies. High frequency information in Equation 3 can be restored in different scales and multiple orientation coefficients as

$$|f \star \psi_{j_1,r_1}| \star \psi_{j_2,r_2}, \text{where } j_1 < j_2, r_1, r_2 \in \Gamma. \tag{4}$$

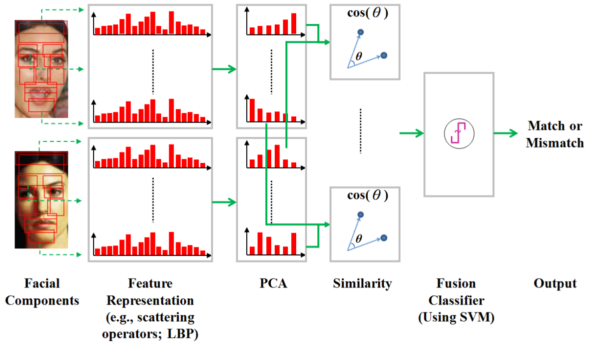


Figure 2. Flowchart of our method.

In order to preserve invariance and keep discriminating information of Equation 4, it is required to reduce the variability of there coefficients. The high frequencies are averaged with low-pass filtering:

$$||f \star \psi_{j_1,r_1}| \star \psi_{j_2,r_2}| \star \phi_J(x). \tag{5}$$

As depicted in Figure 1, the scattering coefficients are computed with a cascade of wavelet decompositions. The input signal can be decomposed into finer scales and wavelet coefficients of different orientations iteratively. In real application, the number of levels can be set to a constant, because the distinctiveness decreases with finer scales. According to [6], three levels are enough for most cases.

In summary, high frequency information can be restored by finer scales and multiple-orientation low-pass wavelet coefficients which are insensitive to local deformation is obtained by averaging their amplitudes with ϕ_J. Scattering coefficients are computed with a cascading structure which is analogous to a convolution network architecture.

3. Method

We divide the faces into several local regions and use these facial components for recognition. We choose 15 components in a face image, including forehead, eyebrows, eyes, nose, cheeks, lips, mouth, and chin. These components are automatically selected after alignment without manual labeling. Dividing face images into local regions can mitigate some problems, such as facial expression and head pose.

Figure 2 describes the schema of our method. A pair of aligned images is cropped, and features are extracted by different descriptors and projected into lower dimensional spaces by PCA. Then, the cosine similarity scores are calculated for each facial component in the same position. We get the output by jointing these similarity scores and classifying them with SVM. In comparison with the component-based approach, holistic-face approach is also evaluated in our experiments.

4. Experimental Results

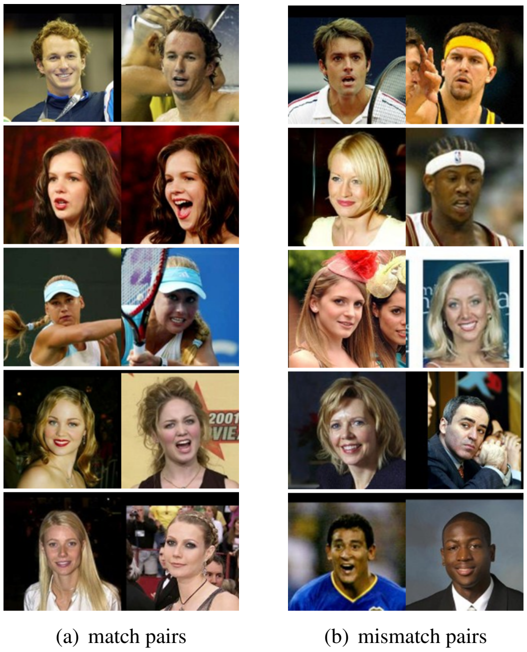
4.1 Database

We perform our experiments on the Labeled Faces in the Wild (LFW) database [11]. One purpose of this database is to help answer the question: “whether the two input faces are from the same person?” and hope to get the answer even when we have not seen these persons before. The database contains 13,233 face images of 250×250 pixel from 5,749 individuals. Some people may contain more than one images. There are totally 9,200 image pairs that are labeled with “match” and “mismatch”. Figure 3 is an example. There are two subsets, view 1 and view 2, in LFW database. View 1 contains 2,200 face pairs for training and 1,000 face pairs for testing, and is used for training and validation. View 2 contains 6,000 face pairs and is used for comparison based on 10-fold cross validation with the parameters of the learning algorithm chosen by using view 1. A holistic image are cropped into 150×80 pixel that are aligned with commercial face alignment software [11].

4.2 Features and Experiment Setup

We evaluated the performance of four feature descriptors of different characteristics. In Local Binary Patterns (LBP), the holistic and component images are divided into non-overlapping 5×5 blocks and histograms with 59 bins are extracted for each block to form the feature vector whose length is 1,475 (= 5× 5×59).

Gabor Wavelets are set with 5 scales and 9 orientations. In the holistic experiment, images with the down-sampling rate of 10×10 are convoluted with Gabor wavelets to form a feature vector whose length is 4,800 (= 5×8×15×8). The component images are convoluted without down-sampling.


(a) match pairs

(b) mismatch pairs

Figure 3. Example image pairs of view 1 in LFW database. (a) and (b) are the match and mismatch image pairs respectively.

HOG descriptor is similar to SIFT that counts the occurrence of gradient orientations in a local region of images, and each image is partitioned into 16-cell grids. Every image represented by HOG is a 128-dimensional histogram of gradient orientations.

Scattering operators use 3 scattering scales (levels) and 6 orientations complex Gabor function for ψ. The holistic and component images are represented with scattering operators descriptor after being normalized to [-1, 1]. The length of scattering operators of a holistic image is 96,520 (=760×127).

4.3 Results

Table 1 shows the accuracy. The results demonstrate that scattering operators outperform other conventional feature descriptors for both holistic and facial component approaches. We try to combine and evaluate different feature descriptors by jointing two of them. The performance of combined features can be improved when scattering operators are included. As can be seen, scattering operators combined with LBP, Gabor, and HOG perform better than the individual LBP, Gabor, and HOG. However, the performance of Gabor plus scattering operators is slightly decreased than the scattering operators only. This could be due to both of them

Descriptor		LBP	Gabor	HOG	Scattering Operators	LBP+ Gabor	LBP+ Scattering	Gabor+ Scattering	HOG+ Scattering
View 1	Holistic	73.4	64.9	66.2	77.9	74.7	78.2	77.6	77.9
	Component	78.4	70.5	75	80.7	78.5	81.9	79.9	82
View 2	Holistic	72.8	63.8	68.2	77.3	73.5	76.3	72.7	74.2
	Component	78.2	71	73.8	**81.5**	78	**83.3**	79.7	81.3

Table 1. Accuracy of different descriptors using holistic and facial regions approaches.

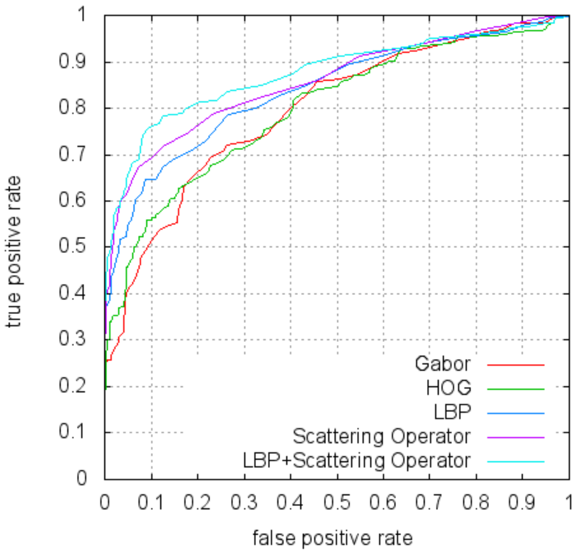


Figure 4. ROC curves averaged over 10 folds of view 2 using facial component approach.

are based on Gabor functions and thus critical information are overlapped.

Figure 4 shows the ROC curves of view 2 using the component approach. From Table 1 and Figure 4, we find that LBP+scattering operators achieves the best performance.

5. Conclusion

In this paper, we introduce scattering operators in LFW database for unconstrained pair matching. The descriptor provides a locally translation invariant representation that could tolerate deformation caused by facial expression and head pose. Furthermore, it can restore the high frequency and preserve the detail information. Our experimental results demonstrate that scattering operators outperform conventional feature descriptors in both holistic and facial component approaches.

Acknowledgment

This work was supported in part by Ministry of Economic Affairs, Taiwan, under Grant No. 100-EC-17-A-02-S1-032.

References

[1] L. Shen and L. Bai, “Information theory for gabor feature selection for face recognition,” *EURASIP JASP*, 2006.

[2] J. Zou, Q. Ji, and G. Nagy, “A comparative study of local matching approach for face recognition,” *IEEE TIP*, 2007.

[3] H. Feichtinger and T. Strohmer, *Gabor analysis and algorithms: Theory and applications*, 1998.

[4] D. Lowe, “Object recognition from local scale-invariant features,” in *ICCV*, 1999.

[5] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *CVPR*, 2005.

[6] J. Bruna and S. Mallat, “Classification with scattering operators,” in *CVPR*, 2011.

[7] S. Mallat, “Group invariant scattering,” *http://arxiv.org/abs/1101.2286*.

[8] E. Tola, V. Lepetit, and P. Fua, “Daisy: An efficient dense descriptor applied to wide-baseline stereo,” *IEEE TPAMI*, 2010.

[9] J. Andén and S. Mallat, “Multiscale scattering for audio classification,” in *ISMIR*, 2011.

[10] T. Leung and J. Malik, “Representing and recognizing the visual appearance of materials using three-dimensional textons,” *IJCV*, 2001.

[11] G. B. 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, Tech. Rep., 2007.