Review the Strength of Gabor features for Face Recognition from the Angle of its Robustness to Mis-alignment

Shiguang Shan[1], Wen Gao[1,2], Yizheng Chang[2], Bo Cao[1], Pang Yang[1]
ICT-ISVISION JDL, Institute of Computing Technology, Chinese Academy of Sciences, China
College of Computer, Harbin Institute of Technology, Harbin, China
{sgshan, wgao, yzchang, bcao, pyang}@jdl.ac.cn

Abstract

Gabor feature has been widely recognized as better representation for face recognition in terms of rank-1 recognition rate. In this paper, we review the strength of Gabor feature for face recognition from the new angle of its robustness to mis-alignment using a novel quantificational evaluation method combining both the alignment precision and the recognition accuracy. Our experiments show that, compared with the gray-level intensity, Gabor feature is much more robust to image variation caused by the imprecision of facial feature localization, which further support the feasibility of Gabor representation.

1. Introduction

Motivated by both its scientific values and its wide potential applications, face recognition technologies (FRT) have attracted more and more attention. And much progress has been achieved during the past few years [1]. However, most of the current systems work only under constrained conditions, even requiring the subjects highly cooperative. The essential problems in face recognition area remain unsolved, especially under the practical unconstrained imaging conditions. Clearly, challenges lie in not only the academic level but also the application system designing level.

Fisherface [2] has been recognized as one of the most successful FRTs. Our static tests of Fisherface on many databases also show its impressive performance provided that the faces have been manually aligned accurately. However, it is totally not the same case for our practical system based on Fisherface, which puzzled us a lot. We finally found that most of the incorrect recognition comes from the imprecisely localization of the eye centers with a possible deviation of only one or two pixels from their real positions. Namely, the performance degradation mostly resulted from the incorrect alignment. We then elaborately designed experiments to quantitatively evaluate the Fisherface's robustness to mis-alignment. The results surprised us greatly because abrupt degradation is observed even if only one pixel deviation is introduced (Please refer to Figure 5.). Such abrupt degradation of the performance is unacceptable for a practical face recognition system, in which one or two pixels of mis-alignment are almost unavoidable.

To highlight the mis-alignment problem, we had explicitly defined the "***curse of mis-alignment***" problem as "*the abrupt degradation of the recognition performance when small mis-alignment occurs which is caused by the inaccurate localization of the facial landmarks.*"[3]. Evidently, to solve the curse of mis-alignment problem, one should further develop more accurate face alignment method; on the other hand, the robustness of the face representation and classification method to mis-alignment should be greatly improved.

In addition, the Gabor wavelet representation has been successfully used in many face recognition systems. Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity, and therefore achieve higher top-one recognition rate. Among them, Dynamic Link Architecture [4](later named by Elastic Graph Matching [5]), Liu's Gabor-Fisher Classifier (GFC) [6], and Gabor Wavelet Networks (GWN) [7] are the most famous. This paper further reviews the Gabor wavelet representation from the angle of its robustness to the abovementioned mis-alignment problem, and reveals that the Gabor representation is more robust than the image intensity representation.

The remainder of the paper is organized as follows: in Section 2, we briefly introduce the Gabor-based face recognition method; how we should evaluate a FRT when considering its robustness to mis-alignment is presented in Section 3; The next section details the evaluation and

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

comparison experiments, followed by the conclusion in the section 5.

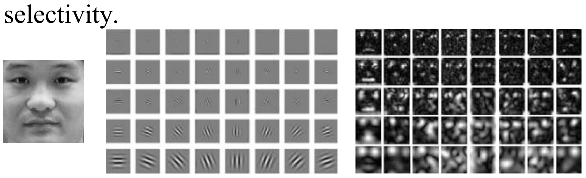
2. Discriminant analysis of Gabor representation

Curse of mis-alignment compels us to seek a representation more robust to mis-alignment. The Gabor wavelet representation seems a natural choice since it can capture the local feature corresponding to spatial frequency (scale), spatial localization, and orientation selectivity. As a result, the Gabor wavelet representation of face images should be robust to variations due to the mis-alignment caused by the imprecisely localized facial landmarks. To verify this point, we implement a face recognition system based on the Discriminant Analysis of Gabor Representation (DAGR), which is very similar to Liu's GFC[] method except that we use the standard Fisher discriminant analysis instead of the so-called enhanced fisher model.

In our DAGR, The Gabor filters are formulated as follows:

$$\psi_{u,v}(z) = \frac{\| k_{u,v} \|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2 \|z\|^2 / 2\sigma^2)} \left[e^{i\vec{k}_{u,v} z} - e^{-\sigma^2/2} \right], \quad (1)$$

where $k_{u,v} = k_v e^{i\phi_u}$, $k_v = \frac{k_{\max}}{f^v}$ gives the frequency, and $\phi_u = \frac{u\pi}{8}, \phi_u \in [0, \pi)$ gives the orientation, and $z = (x, y)$. Note that, in the Equation 1, u controls the scale of the Gabor filters, which mainly determines the center of the Gabor filter in the frequency domain; v controls the orientation of the Gabor filters. This can be observed intuitively from the visualization of the real part of the Gabor filters shown in Figure 1(b). The parameters for the Gabor filters are as follows: $\sigma = 2\pi$, $k_{\max} = \pi/2$, $f = \sqrt{2}$, five scales $v \in \{0,1,2,3,4\}$ and eight orientations $u \in \{0,1,2,3,4,5,6,7\}$. These Gabor kernels form a bank of 40 different filters and exhibit desirable characteristics of spatial frequency, spatial locality, and orientation selectivity.



(a) (b) (c)

Fig. 1 Gabor wavelet representation of face images (a) the 40 Gabor kernels (b) One cropped face image (c) Gabor wavelet representation of the *b* image.

The Gabor representation for a face image is then the convolution of the image with the family of 40 Gabor filters as defined above. Then, the 40 Gabor coefficients (the magnitudes) for each position are computed (as shown in Figure 1) and concatenated pixel by pixel to form a high dimensional feature space. To facilitate the subsequent discriminant and classification, we down-sample the feature space by averaging the magnitude in the 4 by 4 window and throw away the bound points.

In our DAGR system, the normalized face images is 64 by 64 pixels, therefore the dimension of the down-sampled Gabor features g is 225*40=9,000. To further reduce its dimensionality, PCA is applied to g to reduce its dimension to 300. Fisher discriminant analysis is then performed to exact the final features for recognition.

3. Evaluate FRT's robustness to mis-alignment

Distinct algorithms would have different robustness to mis-alignment. Hence, the pure rank-1 recognition rate, when no mis-alignment occurs, would no longer be appropriate for evaluating and comparison. Considering two different algorithms A and B, how their recognition rates vary with the degree of mis-alignment has been drawn in Figure 2. As can be seen, under well-alignment situation, B's recognition rate is as high as 100%, while that of A's is only 92%. Traditionally, we would safely conclude that B outperforms A. However, is it the fact? Our answer is "***NO***". This may seem somewhat anti-intuitive, but we would soon demonstrate its correctness.

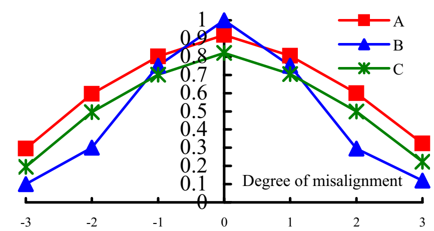


Fig.2 Relationship between the mis-alignment and the recognition rates of three FR methods A, B, and C

Let us consider how it would be if we integrate A and B into a practical face recognition system. We further assume A and B adopt the same frontal-end feature alignment method, which is unavoidably non-perfect, but with a Gaussian distributed mis-alignment from the ground truth, that is, the alignment error satisfies:

$$p(\partial) \sim N(\mu, \sigma^2) \cdot \quad (1)$$

where $\partial = d(P, P^*)$ is the deviation, with P the localized position and P^* the ground truth. We then evaluate the performance of different algorithms as follows:

Definition 1. Overall recognition rate considering mis-alignment robustness is defined as:

$$r^* = \int_{\Omega} P(\partial) r(\partial) d\partial \quad (2)$$

where ∂ *is the degree of mis-alignment;* Ω *restricts the range of possible mis-alignment;* $P(\partial)$ *is the pdf of the* *mis-alignment; and* $r(\partial)$ *represents the recognition rate when mis-alignment* ∂ *occurs.*

R^* is in fact the weighted average of the recognition rate with its corresponding mis-alignment probability. Therefore, it is more appropriate than the pure single recognition rate to evaluate the performance of a practical system integrated by the feature alignment procedure and the recognition procedure.

Nevertheless, it is also necessary to evaluate the robustness of an algorithm to the mis-alignment independent of its recognition rate. For example, consider the algorithm C, whose recognition rates are 10% lower than A's, as shown in Figure 2. Intuitively, C should have the same robustness to mis-alignment as A. To process this case, we further define the following robustness measurement:

Definition 2. Robustness to mis-alignment is defined as:

$$R = \int_{\Omega} P(\partial)\frac{r(\partial)}{r_0}d\partial = \frac{r^*}{r_0}. \qquad (3)$$

where r_0 *is the recognition rate with perfect alignment.*

R, ranging in (0, 1), measures the degradation degree of a recognition method against the mis-alignment. A larger R implies the recognition method be more robust (i.e. less sensitive) to the mis-alignment.

The definition of the r^* and R greatly facilitates the evaluation of different algorithms when considering the mis-alignment. Take the A, B, and C in Figure.2 for example, assuming $p(\partial) \sim N(0,1)$, their r^* and R are shown in Table.1, from which we can evidently conclude that ***A outperforms B*** when mis-alignment is considered. In addition, one can see that C has the same robustness as A, that is, $R_C = R_A$, though its r^* is 10% lower than A's, which completely coincides with the intuition.

Table.1 Performance comparison between A, B, and C with the proposed evaluation measurements

Algorithms	r_0 (%)	r^*(%)	R
A	92	82.3	0.895
B	100	79.5	0.795
C	82	72.3	0.895

4. Robustness comparison of DAGR and Fisherface

To verify the DAGR's robustness, Fisherface [2] is chosen as the comparison benchmark and test them on the FB probe set from the FERET face database. Table.2 shows the structure of the FERET standard face database. Note that the FERET face database has strictly distinguished the testing set (comprised of Gallery and Probe sets) from the training set.

Table.2 Structure of the FERET face database to evaluate FRT's robustness to mis-alignment

Database		#Persons	#Images	Description
Training set (L)		429	1002	Near-frontal faces
Testing Set	Gallery (G)	1196	1196	Near-frontal faces under normal lighting
	Probe Set-FB	1195	1195	Near-frontal faces under Normal lighting with different expressions.

In the FERET face database, the coordinates of the eyes in all the face images have been provided, which can be used as the ground-truth alignment. In both DAGR and Fisherface, faces are normalized as shown in Fig.3. Faces are firstly cropped out, as Fig.3 (b), by placing the two eyes at fixed locations. A mask, as shown in Fig.3 (c), is then covered over the face region to eliminate the background and hairstyle. Eventually, all faces are warped to the size of 64x64 as shown in Fig.3 (d) from its original form as in Fig.3 (a).

(a) (b) (c) (d)

Fig.3 Face normalization in our experiments

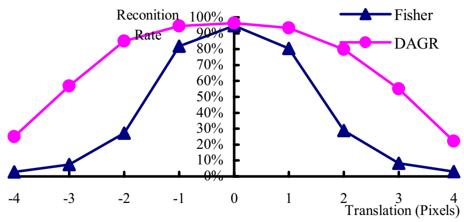
To evaluate the DAGR and Fisherface's robustness to mis-alignment systematically and quantitatively, we test the variance of their recognition rates with the deliberated perturbation of the eye coordinates of the ***probes*** in order to observe the relationship between the recognition rate and the mis-alignment degree. It is not difficult to understand that the mis-alignment of the eyes is equivalent to the variation of the affine parameters such as translation, rotation and scale. Concisely, experiments are conducted to investigate the influence of the variation of translation, rotation and scale separately rather than their combination. Figure 4 illustrates some examples with normalization error due to mis-alignment, from which much appearance variation can be observed. Note that, nevertheless, in our experiments, the alignment of the images from the training set and the gallery is kept precise

Fig. 4 Normalization error due to mis-alignment

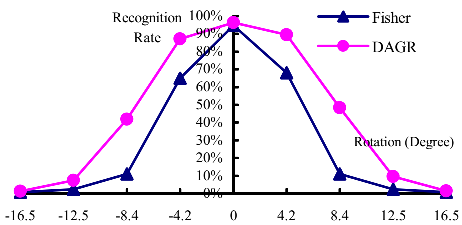
The evaluation results of DAGR and Fisherface are shown in Figure 5 with (a) (b) and (c) representing the translation, rotation and scale case respectively. Note that, in Figure 5 (b), each graduation (about 4.2 degrees) of the horizontal axis is caused by one pixel deviation of the two eyes from their ground-truth position along the opposing vertical direction (that is, one up, the other down).

Similarly, in Figure 5 (c), each graduation (about 0.07 scale change) comes from one pixel deviation of the two eyes from their ground-truth position along the opposing horizontal direction (that is, one left, the other right).

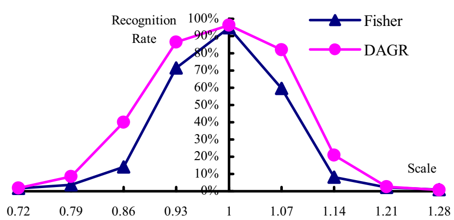
Table 3 illustrates the comparison using the proposed evaluation method. The comparison has evidently indicated that the DAGR has much better overall performance than the Fisherface.



(a) Translation



(b) Rotation



(c) Scale

Fig.5 Comparison between the Fisherface and the DAGR from the angle of robustness to mis-alignment

Table.3. Performance comparison of the Fisherface and the DAGR using the proposed evaluation measurement assuming $p(\partial) \sim N(0,1)$

Mis-align	Methods	r_0 (%)	r^*(%)	R
Translation	Fisherface	94.8	80.2	0.846
	DAGR	96.3	93.2	0.968
Rotation	Fisherface	94.8	71.2	0.751
	DAGR	96.3	86.1	0.894
Scale	Fisherface	94.8	70.8	0.747
	DAGR	96.3	82.5	0.857

5. Conclusions and future work

Gabor feature has been widely recognized as one of the best representation method for face recognition in terms of rank-1 recognition rate. In this paper, we review the strength of Gabor feature for face recognition from the new angle of its robustness to mis-alignment. By modeling the alignment error statistically, a novel quantificational evaluation method combining the alignment accuracy and recognition accuracy is proposed to evaluate different face recognition algorithms. Our experiments have shown that, Gabor feature is much more robust to image variation caused by the imprecision of facial feature localization, which further support the feasibility of Gabor representation.

We have only investigated the mis-alignment problem in terms of affine transform. How Gabor feature be extended to adapt to the pose and illumination would be one of our future efforts. In addition, though Gabor representation is robust than gray-level intensity, its robustness is not as robust as one has expected, especially for scale and rotation cases. Our future work would focus on seeking for better representation more robust.

Acknowledgements

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, National Hi-Tech Program of China (No. 2001AA114190 and 2002AA118010), and ISVISION Technologies Co., Ltd.

Reference

[1] R.Chellappa, C.L.Wilson, S.Sirohey, Human and Machine Recognition of faces: A survey, Proceedings of the IEEE, vol.83, no.5, 1995

[2] P.N.Belhumeur, J.P.Hespanha etc. Eigenfaces vs Fisherfaces:recognition using class specific linear projection. TPAMI, 20(7), pp711-720, 1997

[3] S.Shan,Y.Chang, W.Gao, Curse of Mis-alignment in Face Recognition: Problem and a Novel Mis-alignment Learning Solution, To appear in Proceeding of FG04, Seoul, Korea, May,2004

[4] M.Lades, J.C.Vorbruggen, etc, Distortion Invariant Object Recognition in the Dynamic Link Architecture, IEEE Trans. On Computers, 42(3), pp 300-311, 1993

[5] Laurenz Wiskott, Jean Marc Fellous, Norbert Kruger, Christoph von der Malsburg, Face Recogniton by Elastic Bunch Graph Matching, IEEE Trans. On PAMI, 19(7), pp775-779, 1997

[6] C. Liu and H. Wechsler: "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition", IEEE Trans. Image Processing, 11(4), pp467-476, 2002.

[7] V. Krueger and G. Sommer. Gabor Wavelet Networks for Object Representation, Journal of the Optical Society of America (JOSA), 2002