

FACE RECOGNITION USING FEATURE OF INTEGRAL GABOR-HAAR TRANSFORMATION

Jianguo Li, Tao Wang, Yimin Zhang

Intel China Research Center
{jianguo.li,tao.wang,yimin.zhang}@intel.com

ABSTRACT

Gabor filters are widely known as one of the best representation for face recognition. Since raw Gabor representation is of very high dimensionality, feature reduction is usually required in practice. This paper proposes the feature of integral Gabor-Haar transformation (FIGHT), which is a compact Gabor feature representation while still keeps high recognition performance. This paper also studies fusion strategies for groups of FIGHT feature, and present a discriminative learning scheme to combine group-wise results. Experiments show that FIGHT feature is effective, and the discriminative fusion over FIGHT feature group achieves the state-of-the-art performance on FERET database.

Index Terms— Gabor filters, face recognition, Haar features; discriminative fusion

1. INTRODUCTION

Face recognition has widely potential applications such as biometric authentication and surveillance. It has received significant attention in the research community of computer vision and pattern recognition. In the past two decades, much progress has been made as surveyed in [1].

Many algorithms have been proposed for face recognition. Among those, appearance based algorithm has been dominant techniques in the last decade, which includes three well known algorithms: Eigenface, Fisherface, Bayesian face. These three methods are in fact learning some kind of subspace representation from face appearance. Their effectiveness was thoroughly evaluated by FERET test [2]. However, appearance based algorithm suffers much from small facial variations, such as expression, illumination, and pose. Hence, face representation becomes key issues for a successful recognition system. Many researches have been focused on this topic recently. Among them, two techniques are very successful: Gabor filters and local binary pattern (LBP) [3].

Gabor filters based representation has achieved great success in face recognition. As a multi-scale and multi-orientation representation, raw Gabor features are of very high dimensionality. Several techniques are proposed to alleviate this problem. Elastic bunch graph matching (EBGM) [4] extracts

Gabor features from several local landmarks, and then constructs a graph over landmarks. The recognition is based on the elastic matching over graphs between probe and reference samples. However, the high complexity of the matching prevented it from widely applied. Besides, the landmark misplacement problem will also impact the recognition performance. Other technique is more straightforward. Gabor Fisher classifiers (GFC) [5] adopts Fisher discriminant analysis (FDA) to reduce the dimensionality of raw Gabor features.

LBP is another feature representation, which adopts binary pattern to characterize local variance. The binary pattern can be naturally characterized by histogram. LBP histogram feature is reported performance improvement over FERET 97 best results [3].

To further improve face recognition performance, Zhang et al present to combine LBP with Gabor features, which leads to local Gabor binary pattern histogram (LGBP) [6]. It really rocks. However, the dimensionality of LGBP is in the order of 10^5 , which will lead to great memory burden for a real recognition system. Zhang further adopts FDA over blocks of LGBP to both obtain discriminative features and reduce dimensionality, and then combine the piecewise classifiers together, which yields the ensemble piecewise FDA LGBP algorithm (EPFDA-LGBP) [7]. Experiments show that EPFDA-LGBP achieves the state-of-the-art performance on FERET database. However, it still suffers from the high dimensionality problem (in the order of 10^4 now).

The high dimensionality problem in LGBP may due to the following facts. It requires dividing the original Gabor response image into blocks. To achieve good performance, the block size is usually very small (4x8 in [6]), which will yield many blocks. Since each block is encoded by LBP histogram with 256 bins, this leads to extremely high feature dimensions. Some techniques, such as quantization or uniform operator in LBP, must be used to reduce the histogram bin number. However, the quantization may lose information and uniform LBP still keeps high dimensionality (59 bins in uniform LBP). From the above analysis, it is obvious that LBP is not a compact descriptor.

Motivated by the shortcoming of LGBP, this paper presents the feature of integral Gabor-Haar transformation (FIGHT), which is essentially Gabor transformation followed by inte-

gral image based Haar-like feature extraction. FIGHT can not only extract compact features but also keep high recognition performance. This paper also studies fusion strategies when segmenting FIGHT features into groups, and present a discriminative learning based fusion scheme for further improving recognition performance.

The rest of this paper is organized as follows. Section 2 briefly review Gabor filters. Section3 presents how to extract FIGHT features. Section 4 presents a discriminative learning scheme for fusion over groups of FIGHT feature. Experimental evaluation on FERET database is presented in Section 5. Conclusions are drawn in the final section.

2. GABOR FILTERS

Gabor filters offer the best simultaneous localization of spatial and frequency information. It has been widely applied in image processing tasks such as edge detection, invariant object recognition, and compression [8]. The 2D Gabor filters are defined as follows when assuming $\sigma_x = \sigma_y = \sigma$ [8]:

$$\psi(z, \sigma_s, \theta_o) = \frac{1}{2\pi\sigma_s^2} \exp\left\{-\frac{\|z\|^2}{2\sigma_s^2}\right\} \left[e^{jx'\kappa/\sigma_s} - e^{-\kappa^2/2} \right]$$

$$z = (x', y'), \begin{cases} x' = x \cos \theta_o + y \sin \theta_o \\ y' = -x \sin \theta_o + y \cos \theta_o \end{cases}$$

where x, y are pixel position in spatial domain, κ is a parameter for filter bandwidth, θ_o is the filter angle for o -th orientation, and σ_s are the Gaussian deviation for s -th scale, which is proportional to the wavelength of the filters¹.

The Gabor representation of face images is derived by convolving the image with the Gabor filters:

$$G_{s,o}(x, y) = I(x, y) \otimes \psi(x, y, \sigma_s, \theta_o),$$

where \otimes is the symbol for convolution, and $I(x, y)$ is an input image. This convolution can be fast computed by FFT.

3. FEATURE OF INTEGRAL GABOR-HAAR TRANSFORMATION

Gabor transformation usually yields a very high dimensional representation. Given an image of size K , 5-scale and 8-orientation Gabor filters will yield $40 \times K$ dimensional feature vector (only the magnitude), which is usually at least in 10^4 order. There are much redundant information in the raw Gabor representation. Hence, successive feature selection is required. However, the extremely high feature dimension makes the feature selection step very time-consuming or even makes some algorithm intractable. In this case, people

¹There are some other Gabor-like filters which extend traditional Gabor filters with much better spatial localization, for example the Log-Gabor filter proposed in [9]. They can also be put in our FIGHT framework, but due to space limitation, their descriptions are omitted.

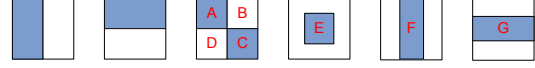


Fig. 1. Illustration of 6 Haar-like features.

usually adopt a grid sub-sampling technique to reduce feature size before feature selection, in which the grid granular becomes a critical problem. This paper presents an alternative solution — we extract the so-called FIGHT features from raw Gabor representation, which is much compact while still keeps high recognition performance.

For face images, Gabor magnitudes do not have great global variations, so that they can be further encoded. Haar-like features have shown great success in face detection due to its efficiency. This paper considers adopting Haar feature to encode Gabor magnitudes. Six Haar-like features are taken into account, which are illustrated in Fig 1. For a given region, they are formally defined as follows:

$$f_1(S) = (|A| + |B|)/|S|, \quad f_4(S) = |E|/|S|,$$

$$f_2(S) = (|A| + |D|)/|S|, \quad f_5(S) = |F|/|S|,$$

$$f_3(S) = (|A| + |C|)/|S|, \quad f_6(S) = |G|/|S|,$$

where $|A| = \sum_{(x,y) \in A} |G(x, y)|$, and S indicates the whole region. The definition has two advantages over widely used subtraction based Haar features. First, it can reduce illumination impact (especially contrast) to some extent. Second, the feature value is normalized to the range $[0, 1]$ for consistent comparison.

Given a local region r of image I , FIGHT is defined as

$$FIGHT_{s,o,r}(I) = \mathbf{f}_r[G_{s,o}(x, y)], \quad (x, y) \in r, \quad (2)$$

where $\mathbf{f}(G) = [f_1(G), \dots, f_6(G)]^T$.

Generally, FIGHT feature may be extracted from many local regions. To speed up feature extraction, this paper adopts the integral image technique, which has been successfully used in face detection [10]. The integral Gabor image is defined as

$$IG(x, y) = \sum_{x', y'} |G(x', y')|, \quad x' < x, \quad y' < y. \quad (3)$$

With integral Gabor image, the operation $|A|$ in \mathbf{f} can be efficiently obtained with only 4 memory access and 4 additions.

The framework of FIGHT feature extraction is illustrated in Fig 2. Since the feature combines integral image, Gabor transformation, and Haar-like features, it is named feature of integral Gabor-Haar transformation (FIGHT).

Specially, a holistic FIGHT feature representation can be defined by a wavefront form over the whole image as follows:

$$FIGHT_{s,o,j,k}(I) = \mathbf{f}_{j,k}[G_{s,o}(x, y)], \quad x \leq \frac{j}{J}W, \quad y \leq \frac{k}{K}H \quad (4)$$

where W and H are the width and height of the region, and $j \in \{1, \dots, J\}$ and $k \in \{1, \dots, K\}$ are the wavefront scale. To make operation $|A|$ meaningful, $J \leq W/2$ and $K \leq H/2$ should be guaranteed.

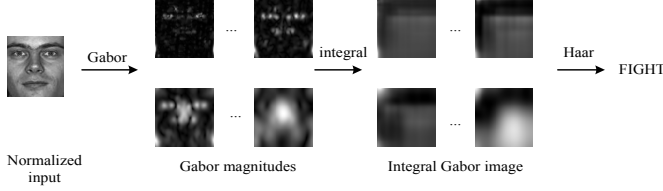


Fig. 2. The framework of FIGHT feature extraction.



Fig. 3. Top 5 local regions learned.

4. LEARNING TO COMBINE GROUPS OF FIGHT FEATURE

FIGHT feature can be used to describe face image by the wavefront form $\mathcal{F} = \{FIGHT_{s,o,j,k}\}$, or by the local region form $\mathcal{F} = \{FIGHT_{s,o,r}\}$. Feature sub-set $FIGHT_{s,o,\dots}$ can be concatenated together for a holistic representation. Besides, they can be also organized into groups for a complementary representation. It is obvious that FIGHT features can be naturally organized by groups of orientation and local regions. This part studies the problem of simultaneously selecting and ensembling groups of FIGHT features.

The fusion of feature group can be done in two typical ways: early fusion or later fusion. In early fusion, people may use the intra-class difference and extra-class difference techniques to generate positive patches and negative patches, and then adopt AdaBoost to select the best patches [11]. The early fusion in feature side can not show how important a patch is for the final recognition, and may be very complex for large patches. This paper studies the later fusion strategy, which considers discriminative fusion over the output of each patch or region based classifier.

Given a feature group organization $\mathcal{F} = \{G_1, \dots, G_n\}$, our late fusion requires the output of the first layer, which are generated in the following way:

- Build dual-space LDA (DSLDA) [12] classifier h_i for each feature group G_i ;
- Adopt classifier h_i to match faces between a left validation set and gallery set. Suppose \mathbf{x} is a validation sample and \mathbf{x}' is a gallery sample, h_i will output a DSLDA distance z_i between them. Define a new label t for $\mathbf{z} = [z_1, \dots, z_n]^T$, let $t=1$ if \mathbf{x} and \mathbf{x}' are in the same class (called positive match), otherwise $t=0$ (called negative match);
- Traverse all the validation samples and gallery samples, we obtain a matching result set $Z = \{(\mathbf{z}, t)\}$.

The goal of later fusion is to find the optimal weight which combines group-wise results. This can be achieved by a probabilistic discriminative learning scheme. Given the first layer output \mathbf{z} , define null hypothesis as $H_0: t = 1$, and the corresponding probability as $P(t = 1|\mathbf{z})$. We use log-odds to characterize how significant the null hypothesis is true:

$$\eta(\mathbf{z}) = \log \frac{P(t = 1|\mathbf{z})}{P(t = 0|\mathbf{z})} = \log \frac{P(t = 1|\mathbf{z})}{1 - P(t = 1|\mathbf{z})}.$$

The log-odds is directly characterized by a linear function, i.e., $\eta(\mathbf{z}) = \beta^T \mathbf{z}$. Now the goal turns to find the optimal weight β . The objective function can be defined as maximally separating the positive matches and negative matches:

$$\max (\mathcal{L}(\mathbf{z}) = \sum_{t(\mathbf{z}_i)=1} \eta(\mathbf{z}_i) - \sum_{t(\mathbf{z}_j)=0} \eta(\mathbf{z}_j)). \quad (5)$$

Note that $\mathcal{L} = t\mathcal{L} + (1-t)\mathcal{L}$, it is easy to show that Eq(5) is equivalent to maximize the following binomial likelihood

$$\max \prod_i P(t = 1|\mathbf{z})^t P(t = 0|\mathbf{z})^{1-t}.$$

This problem can be solved by the iterative re-weighting least square (IRLS) algorithm same as that in logistical regression models. Please refer to [13] for details of the algorithm.

A combined distance measure $\beta^T \mathbf{z}$ is obtained now. The nearest neighbor rule is used for the final decision. Note that the learned weight β can also be used to select feature groups. Groups with larger weight are more important for recognition than those with smaller weight. Some groups with negative weight are in fact negatively correlated to the recognition purpose and thus useless. In our practice, groups with negative weight were discarded.

5. EXPERIMENTS

To evaluate the performance of the proposed FIGHT feature, we conducted experiments on FERET database, which contains a gallery set (1196 images of 1196 subjects), and four probe sets: *fb* (1195 images with expression variation), *fc* (194 images with illumination variation), *dup1* (722 images with short-term aging variation), and *dup2* (234 images with long-term aging variation). In our experiments, we strictly followed the FERET protocol including the pre-processing method provided by CSU [14], and adopted the standard training set in the training CD. One exception is that we resized the CSU cropped images to size 80x90.

We employed FIGHT feature in three ways ².

- (1) Holistic FIGHT: A wavefront form FIGHT was applied for holistic feature representation. In practice, we set $J=K=6$, and started the wavefront from top-left vertex, top-right vertex and mid-point of the image. For 5-scale and 8-orientation Gabor filters, the holistic feature dimension is 4,320.

²Fusion by groups of scales lead to high correlation between groups, and bad performance in experiments. Hence it is omitted.

- (2) Fusion of FIGHT grouped by orientations (FIGHT-OF): In each orientation group, a dual-space LDA (DSLDA) classifier was built [12]. DSLDA reduces the dimension of each feature group to 300 (including dimension from both the principal subspace and the complement subspace). The discriminative learning shows that the first orientation has negative weight, and was neglected according to previous analysis. Hence, this method yields a total of $300 \times 7 = 2,100$ features.
- (3) Fusion of FIGHT grouped by regions (FIGHT-RF): We defined three different size of local regions³: 16×18 , 16×36 , 32×18 . Each local region was sliding in the image with half overlap of the neighbor one. This yields a total of 153 local regions. After discriminative learning, the top 50 regions with largest weight were kept. Fig 3 illustrates the top five regions. Thereafter, each feature group was reduced to 100 dimension by DSLDA. Hence, this method yields 5,000 features in total.

Table 1 shows the rank-1 recognition performance of the proposed features. Since FERET offers a well-defined evaluation protocol so that different algorithms can be directly compared, some known results on FERET are listed in the table for comparison, such as best results in FERET 97 test, results by LBP, LGBP, EPFDA-LGBP, and results by patch based adaptive GFC (PGFC). The feature dimension of different algorithm is also listed in the last column as a comparison.

Three conclusions can be drawn from the results: (1) The holistic FIGHT feature outperforms FERET 97, LBP, non-weighted LGBP, and is comparable to weighted LGBP. However, LGBP has a much higher feature dimension (about 16 times higher than holistic FIGHT). (2) The discriminative fusion over groups of FIGHT feature is helpful. (3) FIGHT-RF achieves the state-of-the-art performance with relatively fewer features than EPFDA-LGBP (5,000 vs 11,000 features).⁴

6. CONCLUSION

This paper proposes the feature of integral Gabor-Haar transformation (FIGHT), which can reduce the dimension of raw Gabor representation significantly while still keep high recognition performance. This paper also studies fusion strategies for group of FIGHT feature, and presents a discriminative learning scheme to combine group-wise results. Experiments show that FIGHT feature is effective, and the discriminant fusion over groups of FIGHT feature achieves the state-of-the-art performance on FERET database.

³This configuration is not necessary the best, can be further optimized.

⁴Specially note that results by LGBP related work adopt a different pre-processing method from that in standard FERET test. LBP and our work accommodate the default pre-processing method in CSU software [14], which keeps pace with the FERET standard. Hence, results by LGBP related work are only for limited comparison. The proposed feature may work much better when adopting better pre-processing method.

Table 1. The rank-1 recognition performance of different algorithms on four FERET probes (%)

Method	fb	fc	dup1	dup2	dimension
FERET97-best [2]	96.2	82.0	59.1	52.1	NA
LBP [3]	97	79	66	64	2,891
PGFC [11]	99	97	87	82	~3,000
LGBP [6]	94	97	68	53	74,000
LGBP-Weight [6]	98	97	74	71	74,000
EPFDA-LGBP [7]	99.6	99.0	92.0	88.9	11,000
FIGHT-Holistic	98.7	95.6	71.3	74.8	4,320
FIGHT-OF	98.9	96.9	76.8	81.7	2,100
FIGHT-RF	99.6	99.0	91.9	88.9	5,000

7. REFERENCES

- [1] W.Y. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Survey*, vol. 2003, 399-458.
- [2] P.J. Phillips, H. Moon, and et al, "The feret evaluation methodology for face-recognition algorithms," *IEEE Trans. PAMI*, vol. 22, no. 10, pp. 1090-1104, 2000.
- [3] T. Ahonen, A. Hadid, and M. Pietikinen, "Face recognition with local binary patterns," in *Proc. ECCV*, 2004, pp. 469-481.
- [4] L. Wiskott, J.M. Fellous, N. Kruger, and et al, "Face recognition by elastic bunch graph matching," *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 775-779, 1997.
- [5] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. PAMI*, vol. 11, no. 4, pp. 467-476, 2002.
- [6] W. Zhang, S. Shan, and et al, "Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," in *Proc. ICCV*, 2005, pp. 786-791.
- [7] S. Shan, W. Zhang, and et al, "Ensemble of piecewise FDA based on spatial histograms of local gabor binary patterns for face recognition," in *Proc. ICPR*, 2006, vol. 3, pp. 590-593.
- [8] T. S. Lee, "Image representation using 2d gabor wavelets," *IEEE Trans. PAMI*, vol. 18, no. 10, pp. 959-971, 1996.
- [9] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of Optical Society of America A*, vol. 4, no. 12, pp. 2379-2394, 1987.
- [10] Paul Viola and Michael Jones, "Robust real-time object detection," *IJCV*, vol. 57, no. 2, pp. 137-154, 2004.
- [11] Yu Su, S. Shan, and et al, "Patch-based gabor fisher classifier for face recognition," in *Proc. ICPR*, 2006, vol. 2, pp. 528-531.
- [12] X. Wang and X. Tang, "Dual-space linear discriminant analysis for face recognition," in *Proc. CVPR*, 2004, vol. 2, pp. 564-569.
- [13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer-Verlag, 2001.
- [14] Ross Beveridge, "The CSU face identification evaluation system," Tech. Rep., <http://www.cs.colostate.edu/evalfacerec/>.