Face Recognition Using Sparse Representation with Illumination Normalization and Component Features

Gee-Sern Hsu*, Ding-Yu Lin
National Taiwan University of Science and Technology
No. 43, Sec.4, Keelung Rd., Taipei, 106, Taiwan
*jison@mail.ntust.edu.tw

***Abstract*—We merge illumination normalization and component features into the framework of Sparse Representation-based Classification (SRC) for face recognition across illumination. Unlike most SRC-based face recognition which constructs a dictionary from a training set with sufficient illumination variation, the proposed method adopts a dictionary with illumination-normalized training set. This can be the first attempt to show that illumination normalization can upgrade the performance of SRC-based face recognition. To further improve the performance, we add in schemes exploiting local features, and prove its effectiveness. Experiments on FERET and Multi-PIE databases show that the performance of the proposed method can be competitive to the state of the art.**

I. Introduction

Many methods proposed for face recognition across illumination focus on illumination normalization that removes unbalnaced lighting and strong shadows cast on the face and makes the face good to be recognized [1], [2], [3]. However, some research considers illumination variation in the training phase, and proposes algorithms able to learn the illumination variation and recognize faces under various lighting conditions. Sparse Representation-based Classification (SRC) is one such method, and it is proven effective handling not just illumination, but also variations caused by expression and occlusion as well [4], [5], [6], [7]. It is shown in [7] that the SRC can yield a superb performance in recognition across illumination given a training set with sufficient illumination variation. However, the requirement of *sufficient* illumination variation in the training set may not be easy to meet in many practical applications. For example, in forensics and law enforcement, there can be as few as one facial image available for learning, and faces are to be recognized under various illumination conditions.

The ESRC (Extended Sparse Representation-based Classification) [6] improves the regular SRC approach so that the cases with as few as one training sample per subject can be handled in the SRC framework. Assuming that the basis of intraclass variation, caused by different expressions and illumination conditions, can be considered similar across different subjects, ESRC constructs an intraclass variant dictionary to describe the variation between training and testing images. The recognition problem is cast as finding a sparse representation of the test subject in terms of the training set and the intraclass variant basis, and the nonzero coefficients are considered contributed by the same subject in the training set and the combination of related intraclass variant basis.

Few, if any, works report the impacts made by illumination normalization on SRC-based face recognition, and it is one of the key issues discussed in this paper. We compared several illumination normalization methods considered highly effective in recent research, and point out which are good to be included in the SRC framework for improving the performance. In addition, we also studied the contribution made by facial components to the recognition performance as such a study is rarely explored in the SRC framework. Extensive experiments on the FERET and Multi-PIE databases [8] show that Illumination Normalization and Component Oriented (INCO) SRC can lead to as much as 10% improvement on the recognition rate.

We first review the selected illumination normalization methods in Sec. II, followed by the proposed combination of holistic and component-oriented SRC in Sec. III. Sec. IV presents experimental setup and results. A conclusion of this study is given in Sec. V.

Fig. 1. Input and output of each phase in the illumination normalization proposed by Tan and Triggs [3]. From left to right, original, gamma-corrected, DOG filtered and the contrast equalized.

II. Review on Illumination Normalization

A few illumination normlization methods which are proven effective in illumination-robust face recognition are selected, including the TT (named after the authors Tan and Triggs) [3], GWLD (Gaussian-smoothed Weber Local Descriptor) [2] and AR (Adaptive Retinex) [1]. TT is composed of several processing phases, including gamma correction, DOG (Difference of Gaussian) filtering, optional region masking and contrast equalization. The optional mask is to block out the regions inappropriate for normalization, such as beard and hair. The core part of the TT algorithm lies on the design of the DOG filter, which must be able to remove shading while maintaining sufficient details of the face. Compared with various methods, including Multiscale Retinex and Logarithmic Total Variation, TT delivers the most satisfying results [3]. An example is given in Fig. 1, which shows the input and output of each phase.

Following the Weber's law [2], the GWLD implements a Gaussian smoothed Weber local descriptor as follows,

$$I_G = \arctan\left(\frac{\mathbf{h} * I_g(x,y)}{I_g(x,y)}\right) \quad (1)$$

where $I_g(x,y)$ is the Gaussian smoothed image $I_(x,y)$ at coordinates (x,y) and $\mathbf{h}$ is a high-pass spatial filter. $\mathbf{h}$ can be, for example, a 3×3 filter with 8 in the center and -1 in its eight neighbors. An example with originals under 21 illumination conditions versus the GWLD processed is shown in Fig. 2.

Fig. 2. An example from PIE database with 21 illumination conditions versus the GWLD [2] processed. The upper three rows are originals.

Extended from the regular retinex that consists of illumination estimation and normalization, the AR (Adaptive Retinex) [1] exploits adaptive smoothing at the estimation phase. The original image is iteratively convolved with an averaging spatial filter with coefficients able to reflect the discontinuity at each pixel on the original. Fig. 3 shows an example processed by the AR scheme [1]. The above three state-of-the-art illumination normalization methods are compared in terms of the performance in the SRC-based face recognition when exploiting them in the sample preprocessing phase. The comparison is reported in Sec.IV

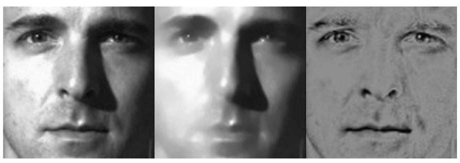

Fig. 3. Left: the original $I(x,y)$; Middle: $I_s(x,y)$ the smoothed using the conduction function; Right: the result obtained by $I(x,y)/I_s(x,y)$.

III. Component-Oriented Sparse Representation

Two schemes that combine the holistic and component features are proposed. One adopts a Dense-on-Features (DOF) grid that has more nodes on local features than on other part of the face, so that the details of the local features can be better considered when forming the coding basis. Unlike most grid based feature extraction with nodes uniformly distributed across the face, the DOF grid has denser nodes on local features, leading to sparse coding basis weighted more on these features. The other adopts a bilayer structure with a regular holistic coding on the first layer and a component feature coding on the second layer, and the outcomes of the two layers are combined using the Bayesian rule. A sample of the DOF grid, compared with a regular grid, is shown in Fig. 4, along with the bilayer structures considered in this study. The one on the left shows dense nodes on three common features, namely eyes, nose and mouth. The one in the middle shows the component layers with eyebrows only, eyes only and both eyes and eyebrows. We have considered nose and mouth for the component layers as well, as shown on the right in Fig. 4.

Fig. 4. Left: Uniform grid (in black) versus DOF (Dense-on-Features) grids, eyes in red grid, nose in yellow and mouth in blue; Middle: Component layers with eyebrows only (in orange), eyes only (in dark blue) and both eyes and eyebrows (the union); Right: holistic versus three component layers.

To apply SRC, we first form a matrix $A = [A_1, A_2, ... A_k]$ from the training set, where A_i denotes the subset formed by all training samples of Subject-i and k is the number of subjects. Each column in A_i is a normalized downsampled feature vector extracted from a training image, and the features can be pixel intensities or others. We have considered the features extracted by Local Binary Pattern (LBP) and Gabor transform in the experiments for comparison purpose. An extensive experimental study on these features is presented in Sec. IV.

Given a probe q^*, the core part of SRC considers the linear representation of q^* in the span of A, i.e.,

$$q^* = Ar^* + \mu^* \quad (2)$$

where r^* is a sparse vector and μ^* is a noise with bounded energy, i.e., $||\mu^*||_2 < \epsilon$. Following the rules in compressing sensing [4], r^* can be obtained by solving the following l_1-minimization:

$$\hat{r^*} = \text{argmin}\, ||r||_1, \quad \text{subject to } ||q - Ar||_2 \leq \epsilon \quad (3)$$

A comprehensive discussion on the solutions for the above l_1-minimization is given in [9], where five fast algorithms

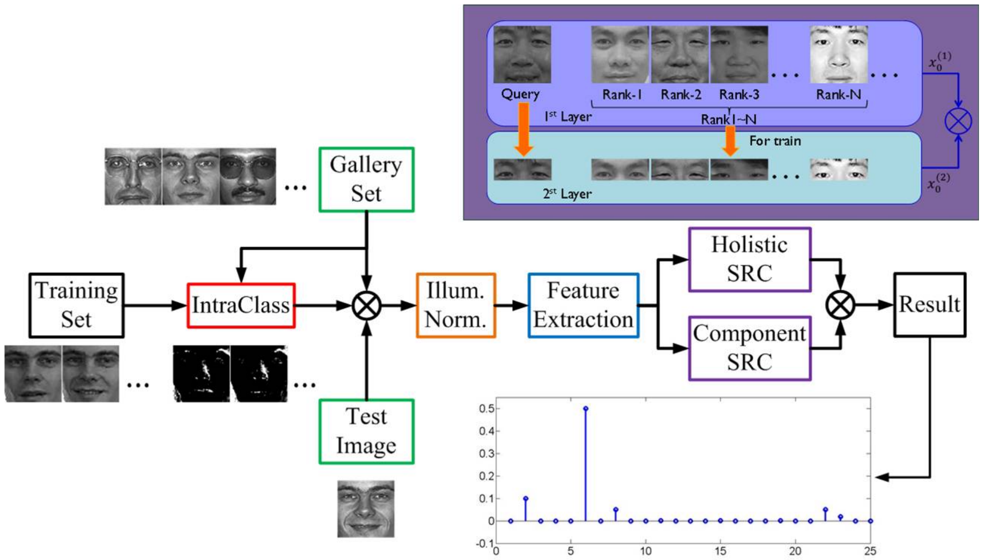

Fig. 5. Workflow of the proposed scheme.

were evaluated on the face recognition performance under illumination variations.

Assuming that the intraclass variation in each gallery face can be approximated by a linear combination of the intraclass differences from a sufficient number of generic faces, the ESRC [6] extends the coding basis to include the basis that spans the intraclass differences in the training set, and change (2) to the following form,

$$q^* = Ar^* + D_i\beta^* + \mu^* \tag{4}$$

where D_i^* is the matrix with its columns assumed able to span the associated intraclass variation. According to [6], D_i can represent the variation caused by unbalanced illumination, different expressions, or occlusions that can not be captured by the noise term μ. If there are redundant and over complete variations in DI, the combination coefficients in β^* would also be sparse. The sparse representation r^* and β^* can thus be recovered simultaneously by l_1-minimization. The merit of the ESRC is the addition of D_i for accounting for intraclass variation, on top of the interclass variation captured by the basis in A.

The framework proposed in this paper follows the formulation in (4), but with the following characteristics:

- A and D_i are obtained from the illumination normalized samples. A performance comparison of this setting with the common setting that uses un-normalized, i.e., original, samples is reported in Sec. IV to highlight its advantages.
- The features extracted to form A and D_i are based on the aforementioned DOF grid, instead of the common uniform grid. A comparison of the two is also given in Sec. IV.

We call the proposed method Illumination Normalized and Component Oriented (INCO) SRC. Its workflow is shown in Fig. 5, where the gallery set contributes to the interclass variant basis, A, the training set contributes to the intraclass variant basis, D_i, and both A and D_i are obtained by illumination-normalized samples. Depending on different features considered, A and D_i can be pixel intensities or the features extracted using LBP or Gabor filter. Given a test image, its illumination is first normalized, the holistic and component oriented features are extracted using the DOF grid or bilayer processing with Bayesian rule for decision making.

IV. Experiments

The following issues are studied in the experiments:

1) Determine the features and related settings good for SRC-based face recognition. Although many use LBP and Gabor features [4], [5], [6], [7], what parameter settings, for example the LBP cell size, can lead to better performance is yet to be answered.
2) Determine the illumination normalization method good to be combined with the SRC framework. A comparison of the three aforementioned methods is carried out to determine the best one for upgrading the performance.
3) Determine whether the holistic or component regions or their combinations are good for the performance.
4) Determine how much improvement the DOF grid scheme can make to the performance when using with different features.
5) The performance comparison of the proposed INCO-SRC with other state-of-the-art methods.

In response to the above issues, we ran experiments on two benchmark databases, FERET and Multi-PIE. The FERET database was collected in 15 sessions over three year. It contains 1564 sets of images for a total of 14,126 images that includes 1199 individuals and 365 duplicate sets of images. A duplicate set is a second set of images of a person already in the database and was usually taken on a different day. For some individuals, over two years had elapsed between their first and last sittings, with some subjects being photographed multiple times. Multi-PIE database contains 337 subjects, captured under 15 view points and 19 illumination conditions in four recording sessions for a total of more than 750,000 images. Only the subset with frontal pose and neutral expression is considered in our experiments. For a fair comparison, we followed the same settings used in the previous works [4], [5], [6], i.e., applying Homotopy to solve the l^1-minimization with error tolerance $\epsilon = 0.05$.

The results are shown in Tables I to VI. Considering the original 128×128 image, the partition into $16 \times 16\times$ cells and each cell transformed into LBP histogram with 59 bins gives the best performance among other LBP cell partitions. It outperforms a regular setup with downsampled 24×24 gray-scaled intensities. However, it is outperformed by the one with Gabor features, which were obtained by applying Gabor filter on the original with 5 scale and 8 orientations, and downsampled to 16×16 in size for better computational efficiency. It must be noted that in all cases the ESRC performs better than the regular SRC.

TABLE I
PERFORMANCE COMPARISON ACROSS DIFFERENT FEATURES WITH VARIOUS SETTINGS (RANK-1 TEST).

FERET dup2	SRC	ESRC
Intensity 24×24	60.7	64.5(+3.8)
LBP $4 \times 4 \times 59$	26.5	30.3(+3.8)
LBP $8 \times 8 \times 59$	56.4	64.1(+7.7)
LBP $16 \times 16 \times 59$	67.9	76.0(+8.1)
LBP $32 \times 32 \times 59$	61.1	73.5(+12.4)
Gabor $16 \times 16 \times 40$	69.6	81.6(+12.0)

Because of the comparison in Table I, we selected the LBP($16\times16\times59$) and Gabor($16\times16\times40$) for the study on the determination of illumination normalization. Table II shows that AR appears slightly better than TT, and both are much better than GWLD when using with LBP features. However, when using with Gabor features, both TT and GWLD perform ideally, and much better than AR. Because the performance with TT appears the most consistent and satisfactory, it is selected as the main tool for normalizing illumination in the rest of experiments.

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT METHODS FOR ILLUMINATION NORMALIZATION.

IN-SRC $\epsilon = 0.05$ FR-rank1					
Illum. Norm.		None	AR	TT	GWLD
FERET	LBP	72.1	89.2	86.0	74.2
Fc	Gabor	97.4	93.3	98.4	98.4

Table III shows the performance with holistic and different facial regions on the most challeging FERET subset Dup-2. Due to the poor performance using mouth and nose, we only consider eyes as the second layer classifier in the bilayer setup, and the DOF grid scheme also covers the eyes region only. Table IV shows the performance improved by the proposed DOF grid. The improvements on LBP($8 \times 8 \times 59$) are more obvious than on LBP($8 \times 8 \times 59$). However, not shown in the table, the processing speed is 3 to 4 times faster.

TABLE III
PERFORMANCE ON DIFFERENT FACIAL REGIONS, USING FERET DUP-2, THE MOST CHALLENGING SUBSET.

FR-rank1	Gray	LBP	Gabor
Holistic	60.7	67.9	69.6
Eyes	48.7	55.5	66.2
Nose	33.8	37.6	39.7
Mouth	26.5	28.2	31.2

TABLE IV
PERFORMANCE COMPARISON OF SRC, ESRC, GRID-SRC AND INCO-SRC (NO-GRID/WITH-GRID).

	SRC	ESRC	Grid-SRC	INCO-SRC
LBP $8 \times 8 \times 59$	56.4	64.1	64.5	67.9/68.4
LBP $16 \times 16 \times 59$	67.9	76.0	67.1	80.8/79.9
Gabor	69.6	81.6	82.0	93.1/-

The comparison of the proposed INCO-SRC with other state of the art is given in Table V, on FERET database, and in Table VI, on Multi-PIE database. The top two are marked in boldface. The INCO-SRC appears to outperform many others, and performs the best when tested on the most challenging subset Dup-2. Similar performance was also observed in Table VI, where the proposed INCO-SRC outperforms others in the subsets Sessions-3 and 4.

TABLE V
PERFORMANCE ON FERET PROTOCOL

Recognition Rate %	Fb	Fc	Dup-1	Dup-2
FERET97 Best result [10]	96.0	82.0	59.0	52.0
ESRC with LBP [6]	**97.3**	95.4	**93.8**	**92.3**
ESRC with Gabor [6]	**97.3**	**98.9**	85.0	84.7
WPCA-POEM [11]	**99.6**	**99.5**	88.8	85.0
INCO-SRC	**97.3**	**98.9**	**90.7**	**93.1**

TABLE VI
THE CONSEQUENCE IN MULTI-PIE

FR-rank1	Session2	Session3	Session4
LBP[7]	**95.2**	**94.7**	93.5
SRC[7]	93.9	93.8	92.3
IN-SRC	93.0	93.9	**94.1**
INCO-SRC	**94.2**	**95.2**	**95.2**

V. CONCLUSION

We apply illumination normalization and component orientation to the SRC framework for face recognition across illumination. This may be the first attempt of altering the SRC dictionary design so that the variation caused by different illumination is substantially reduced, and the remaining variant in the dictionary can better capture the interclass variation. The component orientation emphasizes the features extracted from local regions able to upgrade the performance when combining with holistic features. Experiments show that both schemes can effectively improve the performance.

REFERENCES

[1] J. K. Y.K. Park, S.L. Park, “Retinex method based on adaptive smoothing for illumination invariant face recognition,” *Signal Processing*, vol. 88, p. 19291945, 2008.

[2] W. Y. B. Wang, W. Li and Q. Liao, “Illumination normalization based on webers law with application to face recognition,” *Signal Processing Letters*, vol. 18, pp. 462–465, Aug. 2011.

[3] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *Image Processing, IEEE Trans. on*, vol. 19, pp. 1635–1650, June 2010.

[4] A. G. J. Wright, A.Y. Yang, S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *PAMI, IEEE Trans. on*, vol. 31, pp. 210–227, Feb. 2009.

[5] A. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Fast l1-minimization algorithms and an application in robust face recognition: A review,” *EECS Department, University of California, Berkeley*, Feb 2010.

[6] J. G. W. Deng, J. Hu, “Extended src: Undersampled face recognition via intraclass variant dictionary,” *PAMI, IEEE Trans. on*, vol. 34, pp. 1864–1870, Sept. 2012.

[7] A. Wagner, J. Wright, A. Ganesh, H. Mobahi, Z. Zhou, and Y. Ma, “Towards a practical face recognition system: Robust registration and illumination by sparse representation,” *PAMI, IEEE Trans. on*, vol. 34, pp. 372–386, Feb. 2012.

[8] T. Sim, S. Baker, and M. Bsat, “The cmu pose, illumination, and expression database,” *PAMI, IEEE Trans. on*, vol. 25, pp. 1615–1618, Dec. 2003.

[9] A. Y. Yang, S. S. Sastry, A. Ganesh, and Y. Ma, “Fast $l1$-minimization algorithms and an application in robust face recognition: A review,” in *ICIP*, 2010, pp. 1849–1852.

[10] P. Phillips, H. Moon, P. Rizvi, and P. Rauss, “The feret evaluation method for face recognition algorithms,” *PAMI, IEEE Trans. on*, vol. 22, pp. 1090–1104, Oct. 2000.

[11] N. Vu and A. Caplier, “Enhanced patterns of oriented edge magnitudes for face recognition and image matching,” *Image Processing, IEEE Trans. on*, vol. 21, pp. 1352–1365, Mar. 2012.