Combination of DFT as Global Face Descriptor and LBP/LDiP/LDNP as Local Face Descriptor for Face Recognition

Riko Arlando Saragih, Dodi Sudiana and Dadang Gunawan
Department of Electrical Engineering
Universitas Indonesia, Indonesia
riko.arlando@ui.ac.id

***Abstract*—This paper describes the combination of DFT as a global face descriptor and LBP/LDiP/LDNP as a local face descriptor that results in a final feature vector. Each of these face descriptors does not need a complex learner to classify a novel face pattern when operates separately. However, it will not work when they combine together. The main contributions of our work are in determining the final feature vector that discriminatively represents a face image and the optimal classifier (SVM) that efficiently and accurately classify a novel feature pattern. We conduct simulations on ORL face database by varying the number of face images in training and testing sets on two well-known global face descriptors (PCA and LDA), three local face descriptors (LBP, LDiP, and LDNP), and also the combination of DFT and LBP/LDiP/LDNP. Simulation results show that, the more the number of face images in the training phase, the better the recognition rate of the combination face descriptors rather than either each global or local face descriptor.**

***Index Terms*—Face Recognition; Global Face Descriptor; Local Face Descriptor; Recognition Rate.**

I. INTRODUCTION

Face image as visual information is very useful for face analysis, such as facial expression, face recognition and age estimation. Among several factors that affect the determination of a biometric identifier, a face has high universality, high collectability, and also high acceptability, that make it becomes one of biometrics that is commonly used to identify someone [1, 2]. However, due to its low uniqueness and low performance, to design a robust face recognition system especially in an uncontrollable situation like extreme large illumination variation and deep pose variation for the typical practical condition is still an open and a challenging issue [3].

Two crucial issues must be considered to develop a face recognition system: feature representation and classifier design [4]. Feature representation tries to efficiently and discriminately extract a set of compact features. The main aim of facial representation is nothing but to minimise the intra-class variations and also at the same time maximise the extra-class variations [4, 5]. It means that the best facial representation has a good discriminating power and also invariant to any different imaging factors such as scale, orientation, pose, facial expressions and lighting conditions that may affect the recognition accuracy, which is characterised by the range of values for objects in different classes. They should be different and preferably be well separated and non-overlapping, but all objects of the same class should have similar values. Meanwhile, classifier design is an essential process to find an optimal hypothesis learner that best determines a novel pattern into the correct class belonging. It is important to carefully consider an adequate facial representation to enable a classifier works easier [4].

Along with the development of face recognition algorithm, feature representation techniques are commonly grouped as the global face descriptor [6, 7] or local face descriptor [8, 9], or [10, 11]. Global face descriptor works on the whole face image to get its representation. In contrast, local face descriptor usually first divides a face image into several sub-images or patches to extract the prominent information and subsequently encodes in a hand-crafted way every pixel in the sub-images or patches.

Recently, researchers are more interested in developing local face descriptor due to its robustness to the variations of facial expression, illumination, and occlusion [10]. However, there are different roles between the global face-descriptor and local face descriptor. A global face descriptor may accommodate the information, such as facial contours and hairstyles, resulting in a coarse representation. Meanwhile, a local face descriptor may describe more details the local components on a face, such as eyes, mouth, and nose, resulting in a finer representation. Therefore, designing the combination of both as a final image descriptor is very reasonable to achieve a better face recognition rate [10, 11].

In this paper, we propose a combination of Discrete Fourier Transform (DFT) as the global face descriptor and either Local Binary Pattern (LBP) or Local Directional Pattern (LDiP) or Local Directional Number Pattern (LDNP) as the local face descriptor. Unlike [10] that assumed face images as linear data, we consider them as a non-linear one. We choose LBP/LDiP/LDNP as a local face descriptor due to its simplicity. Moreover, we choose LBP due to its popularity as a local face descriptor; LDiP is for its simple technique like LBP, but more discriminative, and LDNP because of its code more succinct (only use 6 bits) than LBP and LDiP. We also use Support Vector Machines (SVM) as the classifier, because it is effective in cases where the number of dimensions is higher than the number of features (LDA usually fails for this). The simulations are conducted on ORL face database by varying the number of face images in the training phase and testing phase. We do not vary the intensity and also do not reduce the resolution all face images in the simulations like our previous work [12]. We also compare the results to two global face descriptors (Principal Component Analysis/PCA and Linear Discriminant Analysis) and three local face descriptors (LBP, LDiP, and LDNP).

II. Face Image Descriptors

As stated earlier, the role of a face image descriptor in a face recognition system is very crucial. Two well-known global face descriptors are used in this work: PCA and LDA. In PCA, mostly, we try to project higher dimensional feature vectors (face images) onto a space of lower dimension in such a way that, in the lower dimension, whatever projected feature vector we get, that best represents the initial/original feature vector that we have projected. Best represent means the best representation regarding the least square error, that is the least square error between our original feature vectors and the minimum reduced feature vectors [6].

Meanwhile, in LDA, our aim is to separate feature vectors (face images) which belong to different classes, by trying to take a projection onto a different feature space. The projection tries to separate the mean vectors of different classes, and at the same time, that projection also tries to make the samples belonging to the same class more compact. What that means is that within class scatter is reduced while between class scatter is increased [7].

DFT, where the global face descriptor that we combine with the local face descriptor, is defined as follows :

$$F(u,v)=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\exp\left[-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)\right] \tag{1}$$

where $f(x,y)$ represents a 2-D face image of size M by N pixels, $0 \leq \mu \leq M-1$ and $0 \leq \nu \leq N-1$ are frequency variables. The resulting $F(u,v)$ are complex numbers and consist of the real parts and the imaginary parts, that is :

$$F(u,v)=\mathrm{Re}(u,v)+\mathrm{Im}(u,v) \tag{2}$$

where $\mathrm{Re}(u,v)$ and $\mathrm{Im}\ (u,v)$ are the real and imaginary components of $F(u,v)$. Though after Fourier transforms, a face image is represented by the real and imaginary components of all the frequencies, in this work, we only keep 12.5% of the coefficients, and resulting in no more than 15% reconstruction energy ([10] used 50% reconstruction energy in their experiments). Moreover, to represent a face image, after Fourier transforms, the selecting real and imaginary components of the DFT coefficients are concatenated into a single feature vector as a global face descriptor.

LBP is a well-known local face descriptor. Many local face descriptors are developed following this idea. LBP is a kind of local face descriptor that extracts features from a face image by firstly dividing the face image into several sub-images /blocks/regions. As a grey-level comparison technique, every pixel of each region from an image is labelled by first thresholding the 3×3-neighbourhood of each pixel with the centre pixel value. The resulting binary number can be considered as its decimal one as expressed in :

$$LBP_{P,R}(f_c)=\sum_{k=1}^{P} s(f_k-f_c)2^{k-1} \tag{3}$$

where $s(u)=1$ if $s \geq 0$ and 0 otherwise; P and R are the number of neighbouring pixels considered and the radius of the neighbourhood, respectively; and f_c is the centre pixel value. We choose $P = 8$ and $R = 1$ in this work, following the work of [8]. Figure 1 illustrates how LBP extracts a face image locally.

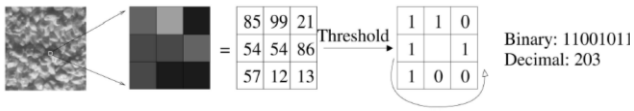

Figure 1: The basic LBP operator [8]

Essentially, LD_iP is an 8-bits binary string that codes each pixel from an image [13]. These binary code patterns are acquired by calculating the relative edge response for several different orientations of each pixel from an image. LDiP uses eight edge response from an image by using a mask (Kirsch mask) for eight different orientations, namely M_0– M_7. The eight Kirsch masks are depicted in Figure 2.

$$\begin{bmatrix}-3&-3&5\\-3&0&5\\-3&-3&5\end{bmatrix}\quad\begin{bmatrix}-3&5&5\\-3&0&5\\-3&-3&-3\end{bmatrix}\quad\begin{bmatrix}5&5&5\\-3&0&-3\\-3&-3&-3\end{bmatrix}\quad\begin{bmatrix}5&5&-3\\5&0&-3\\-3&-3&-3\end{bmatrix}$$

East (M_0) North East (M_1) North (M_2) North West (M_3)

$$\begin{bmatrix}5&-3&-3\\5&0&-3\\5&-3&-3\end{bmatrix}\quad\begin{bmatrix}-3&-3&-3\\5&0&-3\\5&5&-3\end{bmatrix}\quad\begin{bmatrix}-3&-3&-3\\-3&0&-3\\5&5&5\end{bmatrix}\quad\begin{bmatrix}-3&-3&-3\\-3&0&5\\-3&5&5\end{bmatrix}$$

West (M_4) South West (M_5) South (M_6) South East (M_7)

Figure 2: Kirsch edge response masks in eight directions [13]

The first step to get the LDiP for each pixel is by applying eight masks to obtain eight response value $m_0, m_1, ..., m_7$. For the response values which are not equally important in all directions due to the presence of corner or edge that show high response values in particular directions, then we may choose the k prominent directions in order to generate the LDiP code. Therefore, we may get the k top values $|m_j|$ and set them to 1. The other (8-k) bits of 8-bits LDiP are set to 0 [13]. The following equation details the process :

$$C[f(x,y)] := (c_j = 1) \text{if } 0 \leq i \leq 7 \text{ and } m_i \geq \phi \tag{4}$$

where $\phi=k^{th}(M)$ and $M=\{m_0, m_1, ..., m_7\}$[13].

LDNP code is generated by analysing edge response for each mask $\{M^0, M^1, ..., M^7\}$ that represents a significant edge in its own orientation and combining the numbers that have dominant orientations [9]. Not all edge response is equally important; the most negative and positive number show dark and bright, respectively. Therefore, to encode the prominent area, we use three most significant bits to represent the maximum positive number, and three least significant bits to represent the minimum negative number to get the LDNP code. Formally the LDNP code is expressed as the following equations :

$$\mathrm{LDNP}(x,y)=8i_{x,y}+j_{x,y} \tag{5}$$

where (x,y) is the central pixel of the neighbourhood being coded, $i_{x,y}$ is the directional number of the maximum positive response and $j_{x,y}$ is the directional number of the minimum negative response, defined by :

$$i_{x,y} = \arg\max_{i} \left\{ \text{Ɨ}^{i}(x,y) \mid 0 \le i \le 7 \right\} \tag{6}$$

$$j_{x,y} = \arg\max_{j} \left\{ \text{Ɨ}^{j}(x,y) \mid 0 \le j \le 7 \right\} \tag{7}$$

where Ɨ^{i} is the convolution between the original image I and the ith mask, M_i, defined by:

$$\text{Ɨ}^{i} = I * M^{i} \tag{8}$$

We represent all pattern codes resulted in each region for the three local-feature descriptors as a histogram as its feature vector. These histograms are concatenated together to represent the whole image. For classification, we compare the encoded feature vector with all other candidate's feature vector with the chi-square dissimilarity measure. This measure between two feature vectors, S and M, of length N, is defined as :

$$\chi^{2}(S,M) = \sum_{i=1}^{N} \frac{(S_i - M_i)^2}{(S_i + M_i)} \tag{9}$$

where the corresponding image of the feature vector with the lowest measured value indicates the match found.

III. Methodology

In our proposed scheme, each face image, either from the training or testing sets, is processed as follows :

Step 1: Apply DFT as in Equation (1) to get the global face descriptor.

Step 2: Pick only 12.5% of the resulting DFT coefficients (top right corner and bottom right corner).

Step 3: Put the real part of each DFT coefficient in a vector and, likewise, also the imaginary part.

Step 4: Concatenate the real part vector and the imaginary part vector into one single feature vector (global face feature vector).

Step 5: Concatenate this global feature vector with one of the histogram vector (LBP/LdiP/LDNP), that represents the local face feature vector, to get the *final* face descriptor.

Step 6: These final feature vectors (either from the training or the testing sets) are fed together to the SVM as the classifier.

Step 7: The true positive is collected during testing and considered as the recognition rate (%).

IV. Simulation Setup

In our work, we use ORL (Olivetti Research Laboratory) database to evaluate our proposed scheme to combine DFT and one of local face descriptor (LBP/LDiP/LDNP) [14]. We choose this database because there are 40 subjects in this database, of which each subject has ten face images for different pose and expressions. There are also several subjects who use glasses as occlusions that may deteriorate the recognition task. Figure 3 shows several sample face images from this database.

Figure 3: Sample face images from ORL database

We conduct nine kind simulations, of which we pick randomly one face image from each person for the first simulation, two face images for the second one, and so on, until nine face images for the last one. We treat these as the training sets. The rest face images for every simulation are used as the testing sets.

We also repeat ten times for each simulation, of which for each time we pick randomly several face images from each subject as training images and testing images, and take the average the recognition rate, to meet the uncontrollable pose variations for a typical practical face recognition problem. We report the whole simulation results in Table 1.

Unlike for DFT, before extracting the local features, each face image is resized into 100×100 pixels, both for training and testing. We do the same thing (resize into 100×100 pixels) for PCA and LDA to make it appropriate when comparing each other to the face descriptors. All the simulations are run with Intel Core i5-3330 (3 GHz) CPU on the MATLAB platform.

V. Results And Discussion

Table 1 displays all the simulation results. Each row denotes each kind simulation. For example, the first row is the recognition rate for one face image from each subject as the training images and nine other face images as testing images. The second row displays the recognition rate for each descriptor where there are two face images for each subject as the training images set, and eight else face images as testing images set; and so on.

Table 1
Recognition rate for all simulations (%)

PCA	LDA	LBP	DFT+ LBP	LDiP	DFT+ LDiP	LDNP	DFT+ LDNP
55.75	58.03	72.61	65.86	68.75	57.67	71.34	61.17
63.72	75.41	86.57	83.28	81.88	75.13	86.10	77.03
49.82	52.32	60.57	87.96	58.39	81.71	59.18	81.25
42.04	45.87	48.88	93.33	47.54	87.33	48.62	88.50
35.15	42.7	44.1	94.65	43.1	90.55	41.7	90.45
38.01	42.38	39.73	97.81	40.19	94.32	38.63	94.13
28.08	33.5	30.17	97.99	32.33	95.75	31.01	96.33
24	23	25	99	28.75	97.63	26.5	97.25
34.25	28.5	32	100	31	100	30.5	100

As we may see from Table 1, mostly in every simulation the global face descriptor has a lower recognition rate than the local face descriptor. This is because each local face descriptor alone has more discriminative power than each global face descriptor.

LDA as the development of PCA works well along with more number of training images, but LDiP and LDNP as the development of LBP do not perform likewise. It might be that LDiP and LDNP are more robust in illumination variation than LBP, but not in pose variation.

From Table 1, we also may see the recognition rate for each global face descriptor and local face descriptor decreases along with more training images, whereas for all combinations DFT+LBP/LDiP/LDNP, the recognition rate increases. These results might explain the strong combination of global and local face descriptor to make a more robust face recognition system due to their complementary roles.

VI. CONCLUSION

We have proposed a new scheme to combine a global face descriptor using DFT and a local face descriptor (LBP/LDiP/LDNP) to represent a face image. The combinations of global and local face descriptor show an increasing recognition rate along with more number of training images. These results indicate how these combination concepts should be further elaborated to gain insight to get a more robust face recognition system. However, in this paper, we still do not explain how much contribution from each face descriptor to increase the recognition rate.

ACKNOWLEDGEMENT

The first author is on leave from Maranatha Christian University and expresses his sincere thanks to the Ministry of Research, Technology and Higher Education of the Republic of Indonesia for supporting his doctoral study in the form of Higher Education Scholarship (BPPS).

REFERENCES

[1] A.K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no.1, pp. 4-20, Jan.2004.

[2] D. Zhang, Z. Guo, and Y. Gong, *Multispectral Biometrics: Systems and Applications*. Switzerland: Springer-Verlag, 2016, ch. 1.

[3] G. Hua, M.-H. Yang, E. Learned-Miller, Y. Ma, M. Turk, D.J. Kriegman, and T. S. Huang, "Introduction to the Special Section on Real-World Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no.10, pp. 1921-1924, Oct. 2011.

[4] S.Z. Li and A.K. Jain, *Handbook of Face Recognition*. London: Springer-Verlag, 2011, ch.4.

[5] L. Zhen, M. Pietikainen, and S.Z. Li, "Learning discriminant face descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no.2, pp. 289-302, Feb. 2014.

[6] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991.

[7] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, Jul.1997.

[8] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Recognition with Local Binary Patterns : Application to Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no.12, pp. 2037–2041, Dec.2006.

[9] A.R. Rivera, J.R. Castillo and Oksam Chae," Local Directional Number Pattern for Face Analysis : Face and Expression Recognition," *IEEE Transactions on Image Processing*, vol.22, no.5, pp. 1740-1752, May. 2013.

[10] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition," *IEEE Transactions on Image Processing*, vol.18, no.8, pp.1885-1896, Aug. 2009.

[11] L. Wolf, T. Hassner, and Y. Taigman,"Effective Unconstrained face Recognition by Combining Multiple Descriptors and Learned Background Statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no.10, pp. 1978–1990, Oct.2011.

[12] R.A. Saragih, D. Sudiana, and D. Gunawan," Face Recognition using SCAN-based Local Face Descriptor," *ARPN Journal of Engineering and Applied Sciences*, vol. 9, no. 10, pp. 1779-1782, Oct. 2014.

[13] T. Jabid, M.H. Kabir, and O. Chae, "Local directional pattern (LDiP) for face recognition," in *Proc. of the IEEE International Conference on Consumer Electronics*, 2010, pp. 329–330.

[14] F.S. Samaria and A.C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, 1994, pp. 138–142.