International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.6, No.5 (2013), pp.205-216
http://dx.doi.org/10.14257/ijsip.2013.6.5.18

Robust Degraded Face Recognition based on Multi-scale Competition and Novel Face Representation

Guangling Sun and Xiaofei Zhou

School of Communication and Information Engineering, Shanghai University, Shanghai, China
sunguangling@shu.edu.cn

Abstract

Robust degraded face recognition means the recognizer is robust to low resolution and blurry images and as well as other variations such as illumination, expression and et al. Such task is frequently encountered yet a challenging problem. In this paper, we propose appealing solutions to the task without any image reconstruction and without any blur type limitation. Short-Term Fourier Transform (STFT) is first conducted on face image and then two components relying on STFT are proposed: one is related to window size of STFT named scale and the other is face representation construction from STFT. The goal of the first component is to be robust to low resolution and blur. We propose a multi-scale competition strategy that extracts multiple descriptors corresponding to multiple window sizes of STFT and take the identity corresponding to maximum first candidate confidence as the final recognition result. The goal of the second component is to be robust to other variations. We explore the increased discrimination brought by joint coding and using of multiple frequencies. In particular, we propose a novel local descriptor in which information in local areas coming from two frequencies is jointly encoded and further multiple two-frequency-combinations are jointly utilized so as to construct a more discriminative and descriptive face representation. The experiments conducted on AR and Extended Yale B databases demonstrate that state-of-the-art performance has been achieved by multi-scale competition strategy and the proposed novel face representation.

Keywords: *robust face recognition; low resolution and blurry image; multi-scale competition; joint coding; multiple two-frequency-combinations*

1. Introduction

Due to a wide range of potential applications as well as academic challenges, face recognition has attracted much attention during the last decade. Despite great progress has been made in design of scheme robust to expressions, aging of subjects, partial occlusions, illuminations and inaccurate registrations, most of them aimed at recognizing faces in high quality image. Once coping with degraded images caused by such as blur, low resolution, noise *etc.*, the performance will decline dramatically. Or, robust to degradation whereas sensitive to other variations. Hence, in this paper, we focus on face recognition robust to blur, low resolution and other typical variations simultaneously.

There roughly exist four categories of frameworks in literature to handle face recognition from blurry or low resolution image. The first category is to deblur or super resolve an image, then feed the restored image to the recognition engine [1, 2]. While the separated scheme is straight-forward, it is not a good choice since the goal of image restoration is not consistent with that of recognition. And even worse, especially for blurry image, if the blur model is unknown or complex, notable artifacts introduced by deblurring will decline the recognition performance on the contrary. The second category is to do a direct recognition from blurry or low resolution image without deblurring or super re-

ISSN: 2005-4254 IJSIP

solving. Zhang *et al.*, [3] presented a joint blind restoration and recognition framework based on sparse representation. Once blur kernel is estimated, it is used to blur the training set to generate a blur dictionary and the sparse coding of the blurry face using the blur dictionary is determined to give recognition result. Moreover, the kernel is estimated iteratively in a close loop. Sun *et al.*, [4] also explored the blind blurred image recognition in which two frameworks are investigated. One is first to infer the kernel as a separate step, then the kernel is used to generate a data dictionary and an adaptive SIFT feature dictionary is also obtained accordingly. The other is to integrate the kernel estimation and the adaptive SIFT dictionary inference into a common model. The two steps are alternatively executed until stop criterion is reached. The main drawback of works in [3] and [4] is the low efficiency because of the heavy time consumption of blurring operation. In addition, the determination between classes in low degraded image space is limited. The third category is to extract blur invariant or insensitive features. Heikkilä *et al.*, analyzed Local Phase Quantization (LPQ) descriptor robust to centrally symmetric blur [5]. LPQ relied on Short-Term Fourier Transform (STFT). They noticed that the local quantized phase is nearly invariant in low frequency band. Clearly, phase information alone is not appropriate since magnitude is also very useful even more important for recognition demonstrated by work [6]. R.Gopalan et al proposed a subspace as the blur invariants and do recognition over the Grassmann manifold [7]. However, only taking pixel intensity as features is not benefit for further improving robustness. The fourth category is to directly enhance features. Li *et al.*, [8] learned two coupled mapping matrix that mapped a pair of high and low resolution image to a unique feature space. The target of the couple mapping matrix is to make the distance between two points in feature space as close as possible provided that they are corresponding to a pair of high and low resolution version of a same image. The efficiency of the approach is high and super resolving is not necessary, but the mapped feature is global not being benefit for recognition. K. Nguyen *et al.*, also super resolved directly the Gabor wavelet features whereas their method depended on multiple face images [9].

Our idea stems from the STFT related work in [5, 10]. It has been shown that local descriptors based on STFT are effective for recognizing degraded face to a certain extent. We notice that for a given tested degraded image, a suitable scale of STFT being the most favoured for recognition would be selected. Furthermore, the joint LBP-like coding of local areas coming from two frequencies which will bring about more information for recognition would be explored to enhance the robustness of the descriptor. To reach the two purposes, we propose multi-scale competition strategy and a novel face representation. The former selects the winner among multi-scales in terms of corresponding recognition confidence of first candidate and the latter connects the LBP-like coding of local areas of two frequencies into one descriptor and multiple two-frequency-combinations compose the novel face representation.

The paper is organized as follows: Section 2 reviews the STFT and related descriptors. Section 3 presents multi-scale competition and section 4 gives a detail discussion of the proposed face representation. Section 5 demonstrates good empirical results on AR and Extended Yale B databases. Conclusions and future work are provided in Section 6.

2. Review of STFT and Related Descriptors

STFT is calculated over a local area N_{x} centred at $\mathbf{x}$ of an image $f(\mathbf{x})$ as follows:

$$\mathrm{F}_{\mathbf{x}}(\mathbf{u}) = \sum_{\mathbf{y}_i \in N_{\mathbf{x}}} f(\mathbf{y}_i)\omega^*(\mathbf{y}_i - \mathbf{x})e^{-j2\pi\mathbf{u}^T\mathbf{y}_i} \tag{1}$$

Where $\mathbf{u} \in \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L\}$ denote a set of two dimensional frequencies, $\omega(\mathbf{x})$ denote a window function and $\omega^*(\mathbf{x})$ is the conjugate of it. Under the assumption that $\omega(\mathbf{x})$ is real and even, equation (1) could be rewritten as follows:

$$\mathrm{F}_{\mathbf{x}}(\mathbf{u}) = \sum_{\mathbf{y}_i \in N_{\mathbf{x}}} f(\mathbf{y}_i)\omega(\mathbf{x}-\mathbf{y}_i)e^{-j2\pi\mathbf{u}^T\mathbf{y}_i} \tag{2}$$

Obviously, the computation can be separated into two 1-D convolutions for the rows and columns successively. In rest of the paper, the window function will be real and even without particular explanation. Four frequencies $\mathbf{u}_1$=[a,0],$\mathbf{u}_2$=[0,a],$\mathbf{u}_3$=[a,a],$\mathbf{u}_4$=[a,-a] are selected as in work [5] and ‘a’ is just the reciprocal of window size of STFT. After STFT of an image, LBP-like descriptors such as local magnitude descriptor (LMD) and local phase descriptor (LPD) are calculated from magnitude and phase respectively [10]. LMD and LPD are both dependent on binary strings describing relations between value of a position and its eight neighbours. Once a binary string is obtained, it will be encoded into an integer and integers in a local area are pooled into a histogram adopted as recognition feature.

3. Multi-scale Competition

While STFT and LMD/LPD descriptors are demonstrated to be effective for blurry and low resolution face recognition, further problems still deserved to be carefully studied. In following, we will discuss two related topics.

3.1. The Analysis of Magnitude and Phase of STFT for Degradation Face Recognition

In STFT, both magnitude spectra and phase spectra are distributions of harmonic wave bands, namely, the former is energy distribution and the latter is spatial position distribution. Since STFT is spatial-frequency decomposition, the frequency resolution is definitely lower than that of Fourier Transform (FT). Furthermore, compared with magnitude resolution, a higher phase resolution should be favoured since more accurate spatial information is required for better recognition. However, spatial resolution and frequency resolution is a trade-off in STFT so that a higher phase resolution determinedly is accompanied by a lower spatial resolution leading to a larger window size of STFT. Unfortunately, larger window size will lose more detail information helpful for recognition because the extracted frequency is the reciprocal of window size as mentioned above. Moreover, general blur kernel functions except for Gaussian and linear motion will destroy the phase much worse than that of magnitude. Consequently, we argue that the phase of STFT is not appropriate for degraded image recognition. In the sequel, we will only take magnitude into account.

3.2. Multi-scale Competition based on Generalized Confidence

A low resolution or blurry image $g(\mathbf{x})$ could be modelled as a convolution between a high quality image $f(\mathbf{x})$ and a blur kernel function $k(\mathbf{x})$ regardless of noise:

$$g(\mathbf{x}) = f(\mathbf{x}) \otimes k(\mathbf{x}) \tag{3}$$

Assume we focus on two positions $\mathbf{x}_i$ and $\mathbf{x}_j$ and two local regions centred at the two positions. In terms of STFT, two FT of image in which the two local regions are extracted are as follows:

$$\begin{aligned} \mathrm{F}_{\mathbf{x}_i}(\mathbf{u}) &= F[\omega(\mathbf{x}-\mathbf{x}_i)f(\mathbf{x})] \\ \mathrm{F}_{\mathbf{x}_j}(\mathbf{u}) &= F[\omega(\mathbf{x}-\mathbf{x}_j)f(\mathbf{x})] \end{aligned} \tag{4}$$

Now let the two images blurred by $k(\mathbf{x})$, corresponding FT according to Convolution Theorem of FT is follows:

$$\begin{aligned}\tilde{\mathrm{G}}_{\mathbf{x}_i}(\mathbf{u}) &= F\left[k(\mathbf{x})\otimes[\omega(\mathbf{x}-\mathbf{x}_i)f(\mathbf{x})]\right] = \mathrm{K}(\mathbf{u})\bullet \mathrm{F}_{\mathbf{x}_i}(\mathbf{u})\\ \tilde{\mathrm{G}}_{\mathbf{x}_j}(\mathbf{u}) &= F\left[k(\mathbf{x})\otimes[\omega(\mathbf{x}-\mathbf{x}_j)f(\mathbf{x})]\right] = \mathrm{K}(\mathbf{u})\bullet \mathrm{F}_{\mathbf{x}_j}(\mathbf{u})\end{aligned} \tag{5}$$

where $\mathrm{K}(\mathbf{u})$ denotes FT of $k(\mathbf{x})$. For an identical frequency $\mathbf{u}_k$, the blur invariant hold due to $\tilde{\mathrm{G}}_{\mathbf{x}_i}(\mathbf{u}_k)/\tilde{\mathrm{G}}_{\mathbf{x}_j}(\mathbf{u}_k)=\mathrm{F}_{\mathbf{x}_i}(\mathbf{u}_k)/\mathrm{F}_{\mathbf{x}_j}(\mathbf{u}_k)$. However, the blurring operation is first then followed by local area extraction in practice:

$$\begin{aligned}\mathbf{G}_{\mathbf{x}_i}(\mathbf{u}) &= F\left[\omega(\mathbf{x}-\mathbf{x}_i)g(\mathbf{x})\right] = F\left[\omega(\mathbf{x}-\mathbf{x}_i)[k(\mathbf{x})\otimes f(\mathbf{x})]\right]\\ \mathbf{G}_{\mathbf{x}_j}(\mathbf{u}) &= F\left[\omega(\mathbf{x}-\mathbf{x}_j)g(\mathbf{x})\right] = F\left[\omega(\mathbf{x}-\mathbf{x}_j)[k(\mathbf{x})\otimes f(\mathbf{x})]\right]\end{aligned} \tag{6}$$

undoubtedly, the blur invariant has been compromised.

Fortunately, blur or low resolution could be generally regarded as low frequency pass filter implying that the two degradations only destroy the magnitude in high frequency and much less in low and middle frequency to a limitation. As such, once the frequency band corresponding to the window scale of STFT is not beyond the limitation, the blur insensitive will be preserved to a certain extent for that scale. In other words, there are appropriate scales adaptive to the degraded image that make it near as close as possible to the high quality image for magnitude in low and middle frequency band. Here, we will demonstrate twelve degraded cases including two low resolutions(‘LR4’ and ‘LR6’), two parametric blur kernels(‘G’ and ‘M’) and eight nonparametric blur kernels(‘k1’to’k8’). The details about the twelve degradations are provided in experiment section.

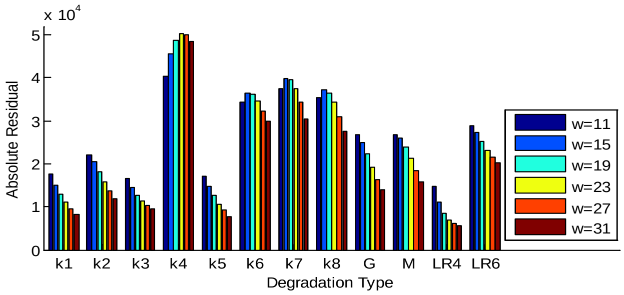


Figure 1.Absolute Residuals between Original Image and Degraded One Covering Six Scales and Twelve Degradations

The statistical result in Figure 1 is obtained by sampling a random subset of AR database [11]. Figure 1 shows the absolute residuals of magnitude of u1 between original and degraded one covering six scales and twelve degradations. The statistics is average values of all randomly selected samples. It can be seen that in a scale range, larger scales actually monotonically decrease the absolute residual between original image and degraded ones with the exception of ‘k4’,’k6’,’k7’and‘k8’. For these four blur kernels, the absolute residuals corresponding to smaller ones of six scales increase on the contrary. In addition, the overall absolute residuals for the four kernels are larger than the other degradations. This problem will be further discussed in experiment section.

Of course, increasing scale is not benefit for obtaining detailed information because of lacking more high frequency just as discussed above. Indeed, degradation insensitive and detail sensitive are two contradictory factors. We propose a straightforward but effective strategy to solve this problem: in a reasonable scale range, confidences of first candidate for all possible scales are calculated and then the identity corresponding to maximum confidence is regarded as the final recognition result. This procedure is intuitively a competition among multiple scales and the scale that obtains the highest confidence would win. We adopt the generalized confidence presented by [12] in our work as this value is a good approximation of Posterior probability which is declared by author. Notice that in the multi-scale framework, features and classifiers corresponding to all scales must be extracted and constructed in advance from original high quality samples.

4. Proposed Face Representation

Now revisit the LMD descriptor, we notice that among the eight coding bits, the four bits corresponding to horizontal or vertical direction is closely related with the four bits corresponding to diagonal or anti-diagonal direction. It means that there are redundancies for the eight bits. However, if the four bits of horizontal or vertical direction is preserved and the other four bits are replaced with the bits of same directions but coming from other frequency, the redundancy will be reduced and information useful for recognition will also be increased. So that the performance obtained by arbitrary two combined frequencies will be higher than that of individual frequency of them. In addition, if more powerful coding pattern such as centre-symmetric pattern [13] shown in Figure 2(a) is adopted, the performance will further be increased compared with the horizontal or vertical pattern. Clearly, from four frequencies $\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{u}_4$,there will be a total of six different two frequency combinations:$(\mathbf{u}_1,\mathbf{u}_2)$,$(\mathbf{u}_1,\mathbf{u}_3)$,$(\mathbf{u}_1,\mathbf{u}_4)$, $(\mathbf{u}_2,\mathbf{u}_3)$, $(\mathbf{u}_2,\mathbf{u}_4)$, and $(\mathbf{u}_3,\mathbf{u}_4)$. The six combination coding results will be pooled into sub-region histogram to feed the classifier. In sum, the proposed novel face representation is calculated as follows:

Based on the magnitude $M(\mathbf{u},\mathbf{x})$ at frequency **u** and spatial position **x**, the binary relation for centre-symmetric pattern is defined as follows:

$$T\left(M\left(\mathbf{u},\mathbf{k}\right),\ M\left(\mathbf{u},\mathbf{m}\right)\right)=\begin{cases}1\text{ ,if } M\left(\mathbf{u},\mathbf{k}\right)\geq M\left(\mathbf{u},\mathbf{m}\right)\\ 0\text{ ,otherwise}\end{cases} \tag{7}$$

Where **k** and **m** denote two centre-symmetric positions to the focused pixel located at **x**. Then, we define centre-symmetric local magnitude descriptor (CSLMD) as follows:

$$\mathrm{CSLMD}_{\mathbf{u},\mathbf{x}}=\sum_{w=1}^{4}T\left(M\left(\mathbf{u},\mathbf{k}\right),\ M\left(\mathbf{u},\mathbf{m}\right)\right)2^{w-1} \tag{8}$$

For an arbitrary two-frequency-combinations and an identical focused pixel, the joint binary relations contain two centre-symmetric coding coming from two frequencies as illustrated in Fig. 2(b), where $\mathbf{u}_a$ and $\mathbf{u}_b$ denote two frequencies. To this end, the novel local descriptor called joint CSLMD(JCSLMD) is encoded as an integer:

$$\begin{aligned}\mathrm{JCSLMD}_{\mathbf{u}_a,\mathbf{u}_b,\mathbf{x}}&=\sum_{w=1}^{4}T\left(M\left(\mathbf{u}_a,\mathbf{k}\right),\ M\left(\mathbf{u}_a,\mathbf{m}\right)\right)2^{w-1}\\&+\sum_{w=5}^{8}T\left(M\left(\mathbf{u}_b,\mathbf{k}\right),\ M\left(\mathbf{u}_b,\mathbf{m}\right)\right)2^{w-1}\end{aligned} \tag{9}$$

Where $\mathbf{u}_a$ denotes one frequency, $\mathbf{u}_b$ denotes the other frequency and **x** denotes position of the focused pixel. After that, the encoded integers of all positions compose a label image and a total of six label images are shown in Figure 3.

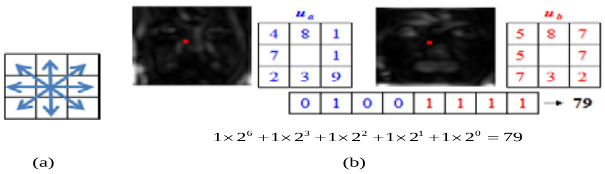


Figure 2.JCSLMD a Location of a Two-Frequency-Combination

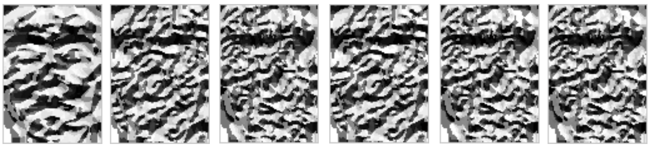


Figure 3.Six Labelled Images of JCSLMD

Because of 8-bits length in coding, the encoded integer is a value between 0 and 255 leading to a histogram with 256 bins. In our experiment, each of the 6 labelled images is divided empirically into 4×4=16 non-overlapping sub-regions and a total of 16×6=96 regional label histograms are generated and concatenated into a long feature vector. Considering the both factors, the dimension of this feature vector will be 96×256=24,576. Clearly, this extremely high dimension will introduce curse of dimensionality and make the feature unstable. We tackle the issue with a learning scheme: depending on training set, a global label histogram is obtained first. In the sequel, percentages of all bins are ordered and two bins of least percentage are combined into a new larger bin and the two percentages are summed as the percentage of the combined bin. Ordering and combinations of bins are executed alternatively and iteratively until a satisfied number of bins are achieved. The final kept and combined bins are called valid bins. During recognition, the original bins of a tested sub-region histogram will be adjusted and combined into a much lower number of bins in terms of the learned valid bins. The specific number of valid bins will be explained in experiment section.

5. Experiment Results and Analysis

The performance of multi-scale competition scheme and the proposed novel face representation are evaluated on two public face databases: AR(rescaled to 110×80, a total of 100 persons, for each person, 7 samples with different expressions and illuminations in the first session as training and corresponding 7 samples in the second session as testing) and Extended Yale B (EYale B) [14] (rescaled to 96×84, great illumination variations). For EYale B, all samples of each person are partitioned randomly and equally into two parts, one part is training and the other part is testing. This partition is done 5 times so as to complete a 5-cross-validation testing. A total of twelve synthesized degradations of an original image are demonstrated in Figure 4. It includes two low resolution degradations with down sampling scale 4 and 6(LR4 and LR6), parametric blur kernels including Gaussian kernel (standard deviation 5 and size 11×11) and linear motion kernel (15 pixel-length with 45 degree), eight complex non-parametric kernels [15]. The non-parametric kernel synthesized ones are blur space invariant. Gaussian window is adopted in STFT.

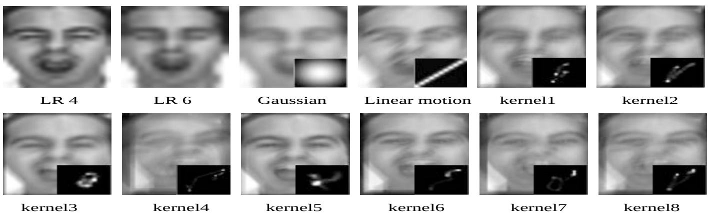


Figure 4. Twelve Degradations of an Original Image

5.1. Face Recognition Classifier and Parameter Setting

Though we do not focus on issue of classifier in this work, the performance of adopted classification approach is rather important. Hence, we implement a classification that slightly different from [16] since we only take the reconstruction errors as recognition distance as follows:

Step 1: Calculate optimal coding $\hat{\boldsymbol{\alpha}}$ for tested sample y upon dictionary D with l_2-norm regularization:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_2^2$$

Where λ is the regularization factor. In all experiments, we set λ=0.01.

Step 2: Classification according to reconstruction error associated with each category:

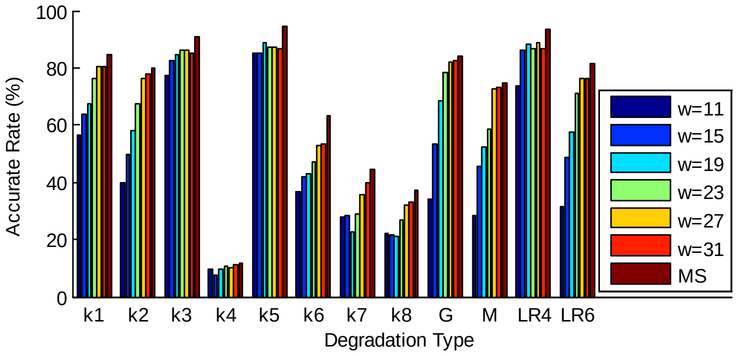
$$\text{identity}(\boldsymbol{y}) = \arg\min_{i} \|\boldsymbol{y} - \boldsymbol{D}_i \hat{\boldsymbol{\alpha}}_i\|_2$$

Where $\hat{\boldsymbol{\alpha}}_i \in \hat{\boldsymbol{\alpha}} = \{\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\alpha}}_2, \cdots \hat{\boldsymbol{\alpha}}_c\}$.

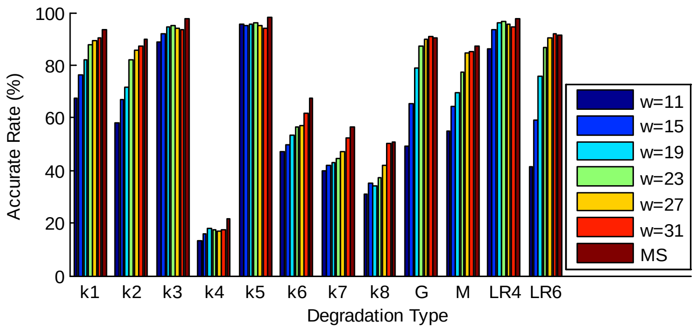
For LMD and CSLMD, four feature planes corresponding to four frequencies are concatenated to compose a complete feature vector and for JCSLMD, six feature planes corresponding to six two-frequency-combinations are concatenated to compose a complete feature vector. All categories of feature vectors are then fed into the classifier to give recognition result. The number of histogram bins for CSLMD is 16 and the optimal valid number of histogram bins for LMD and JCSLMD is 48.

5.2 Multi-scale Competition

The selected multiple scales are in the range of 11×11 to 31×31 with a step 4×4 and their specific values are listed in Figure 5. 'MS' refers to multi-scale competition. The tested descriptor is CSLMD. Accuracies corresponding to six scales and multi-scale competition have been illustrated in Figure 5, where 'G' and 'M' refer to Gaussian kernel and linear motion kernel respectively.



(a)



(b)

Figure 5. Accuracies of Multi-scale competition using CSLMD(a) Accuracies of AR (b) Accuracies of EYale B

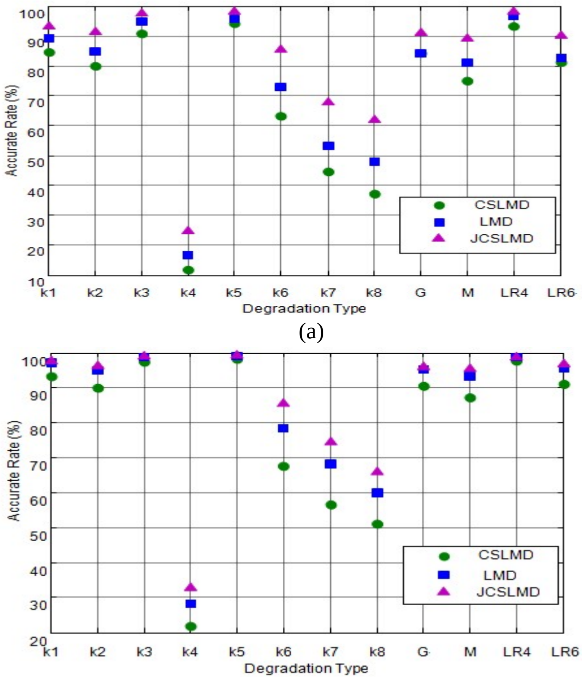
From the result of AR, we can see that the performance of multi-scale competition strategy actually is superior to the highest one among the six scales for nearly all degradation types only except for 'k4' being equal. The result has fully indicated the feasibility and necessity of this strategy. However, we also notice that for EYale B, the improvement is not that remarkable as AR even there is a trivial decrease for 'Gaussian' and 'LR6'. It is probably because that the illumination variation on a face image has destroyed the low frequency to a certain degree whereas a larger scale also utilizes the low frequency band! Despite this case, we still argue that the multi-scale competition has important roles on boosting the recognition performance for degraded face image. Of course, it must be admitted that the obtained performance is at the expense of increased storage and computation because of multiple scales.

5.3. CSLMD and LMD versus JCSLMD

To demonstrate the advantage of the proposed face representation, we compare the performance of three local descriptors including CSLMD, LMD and JCSLMD and corresponding representations on both the degraded and high quality face image. All results have been shown in Figure 6, Table I, Table II and Table III. The suffix 'M' in Figure 6 denotes that the degraded face recognition is depending on the strategy of multi-scale competition. Table I lists the accuracies of high quality face recognition on two databases and the used window size of STFT is 9×9. To make a fair comparison with Zhang's method [16], the images of AR and EYale B are normalized into the same size as that in[16],which are 60×43 and 54×48 respectively. Table II and Table III list the specific accuracies of twelve degradations for two databases using JCSLMD.

For high quality face recognition, the statistics in Table I show that the performance of JCSLMD is superior to that of other related representations and method [16] on the two databases. For degraded face recognition, except the similar performances for 'k1', 'k3', 'k5'and 'LR4' on EYale B, the performance of JCSLMD has been advantageous over that of CSLMD and LMD for all other cases. The exception may be because that the accuracies corresponding to the exceptions have been enough high. Certainly, we have noticed that performances of JCSLMD corresponding to 'k4','k6','k7' and 'k8' yet are insufficient just being consistent with the exceptions in Fig.1. This is absolutely not only a coincidence but closely related. As shown in corresponding four degraded images in Fig.4, there is an overt spatial shift relative to the original image which is one of the major reasons leading to large absolute residuals. With window scale increasing and frequency decreasing accordingly, the magnitude values become larger so that the absolute residuals also become larger if there are more spatial shifts. This is because the absolute residual is calculated point by point. However, if the scale further increases, the absolute residual will again decrease because of losing much detail leading to more similar between images. But this is neither advantage for recognition. In addition, of these four kernels, the spatial shift of 'k4'is the most serious so the accuracy corresponding to it is the lowest. In fact, the spatial shift will only exist in the synthesized scenario and not happen in reality as the face is detected first and then normalized through alignment in real face recognition pipeline. So we argue that the unsatisfied performance should not be emphasized.

To compare with Zhang's approach [3], the results on eight non-parametric kernels are listed along with in Table III. Because there are different parameter settings on 'Gaussian' and 'Linear motion' between our work and Zhang's and there are no testing on 'LR4' and 'LR6', their results corresponding to Zhang's have not been listed in table. From the comparison for the eight non-parametric kernels, the bold numbers emphasize that dramatic improvements have been achieved by the proposed JCSLMD except 'k4' and'k6'. In fact, this also could be owing to spatial shift.



(a)

(b)

Figure 6.Comparison of Three Local Descriptors (a) Accuracies of AR (b) Accuracies of EYale B

Table 1.Accuracies of High Quality Images of AR and EYale B and Comparison with ZHANG'S Method [16] (%)

	CSLMD	LMD	JCSLMD	Zhang's
AR	92.6	96.4	**99.1**	93.7
EYale B	97.6	99.5	**99.6**	97.9

Table 2. Accuracies of Degraded Images of AR (%)

Degradation	k1	k2	k3	k4	k5	k6	k7	k8
Proposed	93.4	91.4	97.7	24.9	98.3	85.4	68.0	61.9
Degradation	G	M	LR4	LR6				
Proposed	91.1	89.4	98.3	90.3				

Table 3. Accuracies of Degraded Images of EYale B and Comparison with Zhang's Method [3] (%)

Degradation	k1	k2	k3	k4	k5	k6	k7	k8
Zhang's	86.2	79.3	85.7	43.1	81.9	86.4	64.7	54.8
Proposed	**97.6**	**96.2**	**98.9**	32.7	**99.3**	85.6	**74.5**	**65.9**
Degradation	G	M	LR4	LR6				
Proposed	96.1	95.4	98.8	96.8				

6. Conclusion and Future Work

We investigate degraded face recognition task and present two contributions. One is multi-scale competition strategy and the other is a novel face representation based on JCSLMD, which are for the purpose of improving robustness to degradation and to variations of expression, illumination and et al simultaneously. A scale adaptive to tested image which obtains the best insensitive-detail trade-off is found by using multi-scale competition and a depth discussion about it is presented. In addition, JCSLMD improves the performance of LMD and CSLMD by utilizing the increased discrimination brought by joint local encoding of two frequencies and joint use of multiple two-frequency-combinations. Encouraging results have been obtained on public AR and EYale B. The proposed approach need not any image reconstruction and even is effective for space variant blue degradation for it does not rely on any blur assumption though we do not evaluate it on space variant blurry samples.

Future work would complement discriminate analysis for the proposed new representation instead of direct using it and develop faster and cleverer multi-scale fusion approach. Further, we will investigate the more difficult scenario of low resolution and blurry image corrupted by noise simultaneously.

References

[1] M. Nishiyama and A. Hadid, "Facial Deblur Inference Using Subspace Analysis for Recognition of Blurred Faces", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 4, **(2011)**, pp. 838-845.

[2] C. Liu, H. Y. Shum and W. T. Freeman, "Face hallucination: Theory and Practice", International Journal of Computer Vision, vol. 75, no. 1, **(2007)**, pp. 115-134.

[3] H. Zhang and J. Yang, "Close the Loop: Joint Blind Image Restoration and Recognition with Sparse Representation Prior", IEEE International Conference on Computer Vision, **(2011)**, pp. 770-777.

[4] G. Sun, J. Yin and X. Zhou, "Blurred Image Classification based on Adaptive Dictionary", (DOI: 10.5121/ijma.2013.5101), The International Journal of Multimedia & Its Applications, vol. 5, no. 1, **(2013)**, pp. 1-9.

[5] J. Heikkilä and V. Ojansivu, "Methods for Local Phase Quantization in Blur-Insensitive Image Analysis", Local and Non-Local Approximation in Image Processing, **(2009)**, pp. 104-111.

[6] S. Xie, S. Shan, X. Chen and J. Chen, "Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition", IEEE Transactions on Image Processing, vol. 19, **(2010)**, pp. 1349-1361.

[7] R. Gopalan, S. Taheri, P. Turaga and R. Chellappa, "A Blur-robust Descriptor with Applications to Face Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, **(2012)**, pp. 1220-1226.

[8] B. Li, H. Chang, S. Shan and X. Chen. "Low-Resolution Face Recognition via Coupled Locality Preserving Mappings", IEEE Signal Processing Letters, vol. 17, no. 1, **(2010)**, pp. 20-23.

[9] K. Nguyen, S. Sridharan, S. Denman and C. Fookes, "Feature-Domain Super-Resolution Framework for Gabor-based Face and Iris Recognition", IEEE Conference on Computer Vision and Pattern Recognition, **(2012)**, pp. 2642-2649.

[10] Z. Lei and T. Ahonen, "Local Frequency Descriptor for Low-Resolution Face Recognition", IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, **(2011)**, pp. 161-166.

[11] A. M. Martinez and R. Benavente, "The AR face database", CVC Technical Report, no. 24, **(1998)**.

[12] X. Lin and X. Ding, "Adaptive Confidence Transform based on Classifier Combination for Chinese Character Recognition", Pattern Recognition Letters, vol. 19, **(1998)**, pp. 975-988.

[13] M. Heikkilä, M. Pietikänen and C. Schmid, "Description of Interest Regions with Local Binary Patterns", Pattern Recognition, vol. 42, no. 3, **(2009)**, pp. 425-436.

[14] A. Georghiades, P. Belhumeur and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, **(2001)**, pp. 643-660.

[15] A. Levin, Y. Weiss, F. Durand and W. Freeman, "Understanding and Evaluating Blind Deconvolution Algorithms", IEEE Conference on Computer Vision and Pattern Recognition, **(2009)**, pp. 1964-1971.

[16] L. Zhang, M. Yang and X. C. Feng, "Sparse Representation or Collaborative Representation: Which Helps Face Recognition?", IEEE Conference on Computer Vision, **(2011)**, pp. 471-478.

Authors

Guangling Sun, she received the B.S. degree in Electronic Engineering from Northeast Forestry University, China, in 1996, and the M.E. and Ph.D. degrees in Computer Application Technology from Harbin Institute of Technology, China, in 1998 and 2003, respectively. Since 2006, she has been with the faculty of the School of Communication and Information Engineering, Shanghai University, where she is currently an Associate Professor. Her research interests include pattern recognition, image and video processing and multimedia security.

Xiaofei Zhou, he received the B.S. degree in Electronics and Communication Engineering from Anhui Polytechnic University, China, in 2012, and now he is working on a master's degree in signal and information processing at the School of Communication and Information Engineering, Shanghai University. His research interests include pattern recognition, image processing and computer vision.