# Local Frequency Descriptor for Low-Resolution Face Recognition

Zhen Lei[1] Timo Ahonen[2] Matti Pietikäinen[2] Stan Z. Li[1*]

[1]Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu, Beijing 100190, China.
[2]Machine Vision Group, University of Oulu, PL4500, FI-90014 Oulun yliopisto, Finland.

## Abstract

*Face recognition from low-resolution images is a common yet challenging case in real applications. Since the high-frequency information is lost in low-resolution images, it is necessary to explore robust information in the low frequency domain. In this paper, we propose an effective local frequency descriptor (LFD) for low resolution face recognition, by building upon the ideas behind local phase quantization (LPQ) and exploring both blur-invariant magnitude and phase information in the low frequency domain. The proposed descriptor is more descriptive than LPQ with more comprehensive information. In addition, a statistical uniform pattern definition method is introduced to improve the efficiency of the proposed descriptor. Experimental results on FERET and a real video database show that LFD is effective and robust for low-resolution face recognition.*

## 1. Introduction

Face recognition, as one of the primary biometrics technologies, has been widely studied in the recent decades. Although recognition rate in a controlled environment is satisfactory, its performance in real applications is still an unsolved problem partially due to inadequate image quality. The report of MBGC [13] also indicates the face recognition problem with good quality images has been well solved, while the face recognition performance with bad or low-quality face images is still far from satisfaction and hence encourages researchers to pay more attention on these cases. Recently, more and more face recognition techniques have been applied to surveillance or intelligent mobile phone applications. Traditional methods based on high/middle resolution face images could not perform well when the face image captured by a video or web camera is of relatively low-resolution and/or with out-of-focus blur. It is therefore necessary to explore effective descriptors for low-resolution face recognition.

The low-resolution image discussed here roughly includes two cases. One is the traditional low-resolution image where the object size in original image is small. Another one is a normal size object image with the out-of-focus or motion blur whose underlying resolution is therefore comparatively low [21].

In both low-resolution cases, high-frequency information is usually lost. Therefore, some traditional methods [22, 1] involving detail information of face image may not be suitable. To deal with this problem, there are usually two straightforward ways. One is to try to synthesize the high-resolution images from the low-resolution ones and the traditional high/middle resolution face recognition methods could be applied consequently. Lee et al. [7] used an extended support vector data description (SVDD) method to synthesize the high-resolution images with the help of a high-resolution image training set. Dedeoglu et al. [5] exploited spatio-temporal information from video to help hallucinate high-resolution video. Arandjelovic and Cipolla [3] proposed an extended generic shape-illumination manifold (gSIM) framework to derive the high-resolution result implicitly. Nishiyama et al. [14] tried to recover the original facial images by inferring the point spread function (PSF) representing the process of blur. There is also a series of work [4, 19, 12, 11] introducing how to learn the mapping information from low-resolution image to high-resolution image in training set which is then extended into arbitrary images. However, in these methods, the super-resolution image could be well synthesized in training set but the quality of the synthesized images of people outside the training set is often not adequate for face recognition algorithms.

Another way to deal with the low-resolution problem is to extract the discriminant information from the low-resolution images directly. These methods could be divided into two categories further. One is called holistic method

*Corresponding author.

in which the entire structure of the face image corresponding to the low-frequency domain is explored. Su et al. [17] utilized FFT technique to extract the global structure information of face image for recognition. Li et al. [9] proposed coupled metric learning to transform the high and low resolution images into a feature space to be classified. These holistic methods usually require the images to be well aligned and their performance may be degraded by misalignment often present in real applications. Another category is called local feature based methods in which the local low-frequency discriminant information is extracted to represent faces and is proved to be more robust to misalignment problem than the holistic ones. One of the recent methods is named local phase quantization (LPQ) [15] proposed in the context of face recognition by Ahonen et al [2]. It indicates the local phase quantization information in low-frequency domain is nearly invariant to blur operation so that it is suitable for low-resolution face recognition. However, at least two aspects could be improved to deal with issues with low-resolution. First, the magnitude information is removed in LPQ, but many exitsting works [20, 23] show the magnitude information is important for face recognition. Second, LPQ requires the point spread function (PSF) which causes the blur influence to be positive. Although the assumption is supposed to be valid in the low frequency domain, it might not be guaranteed in the real case.

In this paper, we propose a novel Local Frequency Descriptor (LFD) for face recognition. Like LPQ, the proposed method is based on using local frequency information in a way that makes it robust to blur or low resolution. Unlike LPQ, this descriptor uses both magnitude and phase information, thus carrying more information. Furthermore, the LFD is computed so that the positive PSF assumption is not needed for the blur kernel. The representation is therefore expected to be more robust, and more information on wider frequency band could be exploited in real applications. In addition, we also adopt a statistical uniform pattern definition mechanism to improve the effectiveness and efficiency of the proposed method.

The remainder of this paper is organized as follows. Section 2 briefly reviews the methodology of the LPQ method. Section 3 details the method of LFD utilizing both magnitude and phase information. Experimental results are described and analyzed in Section 4 and in Section 5, we conclude the paper.

## 2. Review of local phase quantization (LPQ)

Generally speaking, a low resolution image can be considered to be obtained as a down-sampling operation after a blurring process on a high-resolution image [6]. In the frequency domain, the blurring process can be represented as $G = F \cdot H$, where $G$ is the Fourier transform of the blurred image, $F$ the original image, and $H$ the point spread function (PSF). The magnitude and phase components therefore satisfy

$$\begin{aligned} |G| &= |F| \cdot |H| \\ \angle G &= \angle F + \angle H \end{aligned} \tag{1}$$

Assuming that the PSF is centrally symmetric, the transform $H$ will be real valued and the phase of $H$ will equal 0 or $\pi$. In the LPQ method, it is assumed that in the very-low frequency band, the value of $H$ is positive with $\angle H = 0$, so the phase information of $G$ and $F$ is the same and therefore a blur invariant representation can be obtained from the phase.

In a realization, the local frequency could be computed using a short-term Fourier transform on local $M \times M$ neighborhoods $N_{\mathbf{x}}$ at each pixel position $\mathbf{x}$ of the image $f(\mathbf{x})$ defined by

$$F(\mathbf{u}, \mathbf{x}) = \sum_{\mathbf{y} \in \mathbf{N_x}} \mathbf{f}(\mathbf{x} - \mathbf{y}) \mathbf{e}^{-\mathbf{j2}\pi \mathbf{u^T y}} \tag{2}$$

The transform can be efficiently evaluated for all image positions $\mathbf{x} \in \{\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_N}\}$ using simply 1-D convolutions for the rows and columns successively.

## 3. Face description with local frequency analysis

### 3.1. Local frequency descriptor

In LPQ method, the magnitude information of images is removed. However, in face recognition domain, many existing works [20, 23] have shown that magnitude information is very critical for face recognition, even more robust and discriminant than phase information. Simply removing the magnitude part is likely to lose also meaningful information. Therefore, in this work, we try to extract discriminant information in both magnitude and phase domain to improve the low-resolution face recognition performance.

From Eq. 1, one can see that though the absolute magnitude values of $G$ and $F$ are different, their relative relationship is still preserved in the blurring process. Given a fixed frequency $u$, for image patches $i$ and $j$, if the magnitude response of patch $i$ at frequency $u$ is larger than that of patch $j$ in $F$, this relationship will still hold in $G$ because $|G(u)| = |F(u)| \cdot |H(u)|$, where $|H(u)|$ is a constant at a fixed $u$. That is to say, if $F^i(u) > F^j(u)$, then we have $G^i(u) > G^j(u)$ and vice versa. Based on this, we can formulate the magnitude information by describing the relative relationship among different patches which is also blur invariant. Particularly, we adopt a method similar to LBP (local binary pattern) to describe the neighboring magnitude pattern. For position $i$, denoting the magnitude response at frequency $u$ in local patch as $M(u, i)$, and the counterparts of its neighboring patch centered at position $k$ as $M(u,k)$, we quantize the relative relationship between them as

$$S(M(u,k),M(u,i)) = \begin{cases} 1 & \text{if } M(u,k) \geq M(u,i) \\ 0 & \text{if } M(u,k) < M(u,i) \end{cases}$$

The local magnitude descriptor (LMD) code at position $i$ and frequency $u$ can then be represented as an integer between 0-255 by combining the eight neighboring relationships as

$$f_{LMD}(u,i) = \sum_{k=1}^{8} S(M(u,k),M(u,i))2^{k-1} \tag{3}$$

It is worth noting that the difference of LBP and LMD is that the original LBP is conducted in image space whereas in this work, we consider the magnitude response change in the frequency domain.

Regarding the phase information, LPQ utilizes the absolute phase information to represent faces, and thus the Fourier transform of the blur kernel must be positive to obtain blur invariance. In real scenario, the assumption that $H$ is positive may not always hold. In order to derive more robust face representation, in this work, we propose to model the phase information in a similar way of magnitude. The relative relationship of phase information instead of the absolute value is utilized to describe the face images. For position $i$, denoting the phase information at frequency $u$ in local patch as $P(i,u)$, and the counterparts of its neighboring as $P(k,u)$, we quantize the relative relationship between them as

$$S(P(u,k),P(u,i)) = \begin{cases} 1 & \text{if } P(u,k),P(u,i) \\ & \text{are in the same quadrant} \\ 0 & \text{otherwise} \end{cases}$$

If two phases lie in the same quadrant, the corresponding bit is set to 1, otherwise it is set to 0. In this formulation, we do not need to assume $H$ to be positive any more, since the sign of $H$ does not affect the relative relationship of phases defined in our way. We even don't assume $H$ to be real value, because the phase difference between $P(u,k)$ and $P(u,i)$ is the same as the difference between $P(u,k)+\angle H$ and $P(u,i)+\angle H$. Thus the relative relationship of phase information with different patches is irrelevant to the phase $H$ and this phase description is expected to be more robust in real applications. The local phase descriptor (LPD) code of position $i$ at frequency $u$ can then be represented as

$$f_{LPD}(u,i) = \sum_{k=1}^{8} S(P(u,k),P(u,i))2^{k-1} \tag{4}$$

An example of LFD (LMD and LPD) coding process is illustrated in Fig 1, and Fig. 2 shows an example of LFD for two different blurred images. It shows that the LFD histogram is almost blur-invariant and it is suitable for low-resolution face recognition.

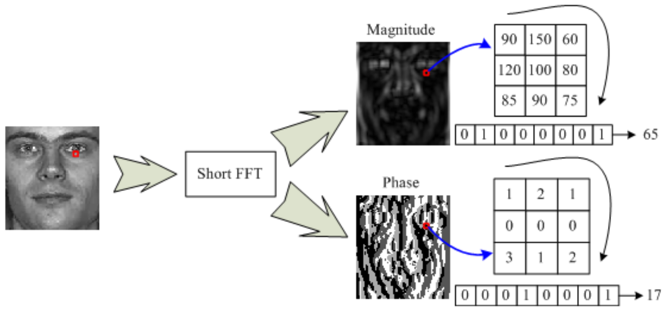



Figure 1. Illustration of the local frequency descriptor.

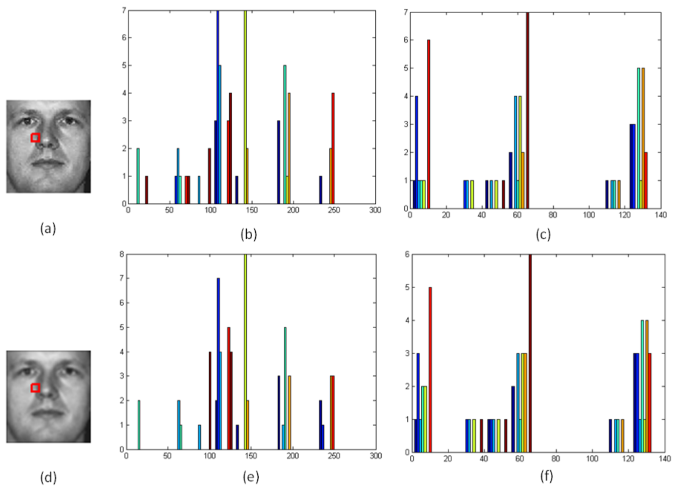
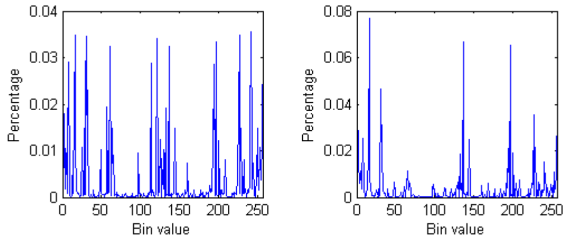



Figure 2. Blur-invariant property of LFD histograms. (a)-(c) original image and its corresponding LFD histograms. (d)-(f) Blurred low-resolution image and its corresponding LFD histograms.

### 3.2. Statistical uniform pattern

In [1], researchers propose uniform pattern mechanism for LBP code which is robust to noise and improves the recognition performance. In LBP code, the uniform patterns are defined as such code that at most two bitwise transitions from 0 to 1 or vice versa occur when the binary string is considered circular. It is based on the observation that there are a limited number of transitions or discontinuities in the circular presentation of the $3 \times 3$ texture patterns. Therefore, the uniform patterns occupy a vast majority proportion of all LBP patterns in local image texture.

In this paper, we adopt a more general strategy as in [10, 8] and define the uniform pattern via statistical analysis, according to the occurrence percentage instead of the number of 0-1 and 1-0 transitions for different codings.




(a) (b)

Figure 3. Distributions of LMD (a) and LPD (b) codes on FERET training set.

Denote a LMD or LPD face image by $f(x, y)$ to indicate the coding value in position $(x, y)$ of the $i$-th image. The occurrence distribution histogram for $n$ face images is computed as

$$H(l) = \sum_{i=1}^{n} \sum_{x,y \in f} I(f(x, y) = l), l = 0, 1, \cdots, 255 \quad (5)$$

where $I(\cdot) \in \{0, 1\}$ is an indication function of a boolean condition. The LMD and LFD codes distributions, calculated on the training set of FERET database are shown in Fig. 3.

The histogram is then sorted according to the occurrence percentage. In this paper, we define the uniform patterns in an iterative way inspired by Huffman coding source reductions [6]. In each step, the patterns corresponding to the two smallest occurrence percentage collapse into a single one and then the histogram is resorted. Suppose we originally have $K$ bins, after $T$ iterations, there are $K-T$ labels left. In this work, $K$ equals to 256 and $T$ can be assigned arbitrarily from 0 to 255. Large $K-T$ value will result in huge feature dimension while small value may lead to the loss of discriminative information for recognition. In this work, we finally select 16 uniform patterns for LMD and LPD respectively, considering the trade-off between the recognition accuracy and the computational cost.

Like the procedure of [1], after the LMD or LPD labeling, the face image is divided into non-overlapped rectangular regions each of which is used to compute a histogram of labels independently, and finally, these histograms are concatenated together to build a global face description.

## 4. Experiments

The effectiveness of the proposed LFD is compared with two state-of-the-art descriptors (LBP and LPQ) for low-resolution face recognition. Experiments with simulated and real low-resolution data are designed to evaluate different methods.

### 4.1. Simulated experiment

We conduct some simulated low-resolution experiments on FERET database [16]. The FERET database consists of one gallery set and four probe sets (fb, fc, dup1, dup2). There are 1196 images of 1196 persons in the gallery set and the four probe sets contain 1195, 194, 722 and 224 images respectively. Two experiments are designed to simulate different low-resolution factors. The resolution of images in gallery set is kept to be $88 \times 80$ in following experiments. In the first one, for the probe images, four resolutions, $88 \times 80$, $66 \times 60$, $44 \times 40$, and $33 \times 30$, are used to evaluate the different methods. For resolutions below $88 \times 80$, the images are first down-sampled into certain resolution and up-sampled to $88 \times 80$ to be recognized. Fig. 4 shows some cropped face examples with different resolutions. In real application such as surveillance scenario, there is also motion blur that make the image low-resolution. Therefore, in the second experiment, we simulate the motion blur problem in probe set by using the shift-invariant linear blur PSF as $H(u, v) = 1/Z$ if $||(u, v)||_2 < b$ and $v = utan\theta$, otherwise $H(u, v) = 0$, where $b$ is the length of camera motion, $\theta$ is the angle and $Z$ is a normalization term. In this experiment, $b$ is set to be 3, three angles ($\theta = 0, 0.25\pi, 0.75\pi$) are selected to simulate different directions of motion. Fig. 5 shows the synthesized motion blurred images with different directions.

In all experiments, the proposed LFD was compared with $\mathrm{LBP}_{8,1}^{u2}$ [1] and LPQ [2]. For LFD and LPQ, four frequency points $\mathbf{u_1} = [\mathbf{a}, \mathbf{0}]^\mathbf{T}, \mathbf{u_2} = [\mathbf{0}, \mathbf{a}]^\mathbf{T}, \mathbf{u_3} = [\mathbf{a}, \mathbf{a}]^\mathbf{T}, \mathbf{u_4} = [\mathbf{a}, -\mathbf{a}]^\mathbf{T}$ are used and the parameters were set to be $M = 7$, $a = 1/7$ which were used in [2]. The LBP, LPQ and LFD histogram features were extracted from the non-overlapping rectangular regions of the size $8 \times 8$.

Fig. 6 and 7 illustrates the face recognition rates on four probe sets in two simulated cases respectively. From the result, one can see that

1. In fb set, where the images are captured under good condition, the performance of LPQ and LFD is very similar in all cases and the LPQ achieved the highest recognition rate.

2. In fc, dup1 and dup2 probe sets, the performance of LFD is mucher better than that of LBP or LPQ. Especially in fc, considering that illumination variation mainly lies in low frequency band, the LFD is proved to be able to alleviate the affect of lighting change and extract the discriminant information effectively, which is very encouraging for low-resolution face recognition.

3. As we know, the face images are always affected by various factors such as illumination, expression, aging etc. The good performance of LFD in all four probe sets with different low-resolution cases shows that the

Figure 4. Cropped FERET face example images. From top to down, the resolution is 88x80, 66x60, 44x40, 33x30 respectively.

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

Figure 5. Examples of face images with motion blur. (a) Original image, (b) $\theta = 0$, (c) $\theta = 0.25\pi$, (d) $\theta = 0.75\pi$.

proposed LFD method is robust, effective and practical for low-resolution face recognition in real world.

### 4.2. Experiment with real low-resolution data

The video database used in this experiment was collected by our group. Two image sets (noted here as set A and set B) were collected by using different video cameras. Set A contains 2400 images from 300 subjects with 8 images per person and set B contains 1016 images from 127 subjects with 8 images per person. The subjects in set B are all included in set A. All images are cropped into $88 \times 80$ according to the automatically detected eye coordinates by an AdaBoost based eye detector [18]. Some cropped example images are shown in Fig. 8. This database is very challenging. There are blur, accessory, lighting and pose variations and the quality of images captured from different cameras is different. All of these increase the difficulties of recognition.

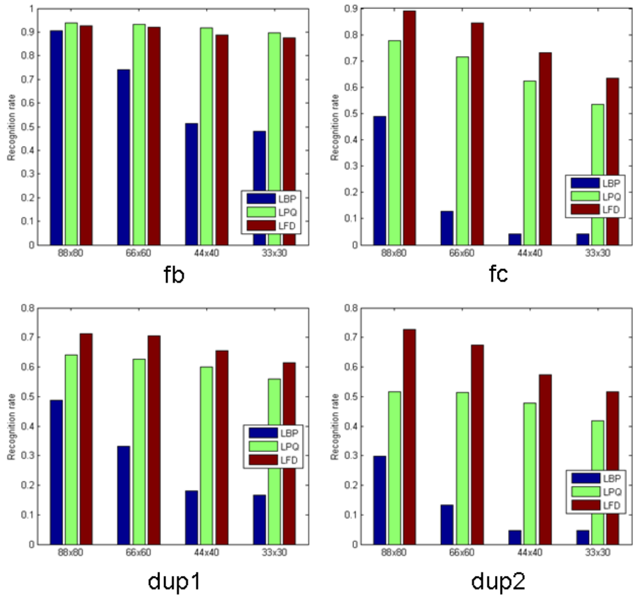



Figure 6. Recognition rates on four probe sets with four resolutions on FERET.

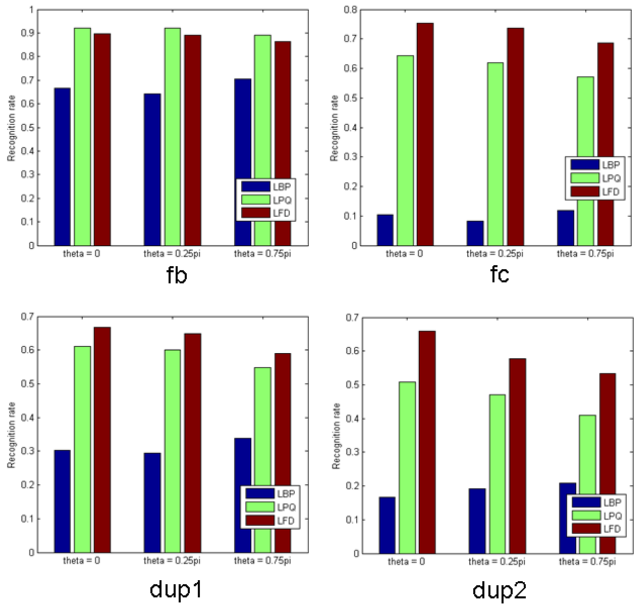



Figure 7. Recognition rates on four probe sets with different motion blur on FERET.

In this experiment, we randomly select 2 images per person from set A to consist the gallery set and use the set B as the probe one. Therefore, it is a cross-matching between two data captured by different cameras.

Table 1 shows the recognition results for different descriptors on the real video database. Both the new magnitude (LMD) and phase (LPD) descriptor show very good performance. The proposed combined LFD descriptor,

Figure 8. Cropped face example images from video database. The first row is the image from set A and the second is from set B.

Table 1. Recognition rates on video database.

| Method | Rec. rate |
|---|---|
| LBP | 0.4117 |
| LPQ | 0.5754 |
| LMD | 0.6538 |
| LPD | 0.6796 |
| LFD (LMD+LPD) | **0.6895** |

which encodes more information in magnitude and phase domains, significantly outperforms the LBP and LPQ methods by improving the recognition rate by 67.5% and 19.8%, respectively. There is great potential for LFD to be applied in low-resolution face recognition in real applications.

## 5. Conclusions

In this paper, we have proposed a new approach to use local phase and magnitude information for robust face recognition. The representation is blur-invariant and is suitable for low-resolution face recognition. Experiments on simulated and real databases show the effectiveness of the LFD in low-resolution face recognition. The future work is to combine the LFD with discriminant learning method to reduce the feature dimension and improve the performance further.


## Acknowledgements

This work was supported by the Chinese National Natural Science Foundation Project #61070146, National Science and Technology Support Program Project #2009BAK43B26, and AuthenMetric R&D Funds.


## References


[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns:application to face recognition. *IEEE T-PAMI*, 28:2037–2041, 2006.

[2] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkilä. Recognition of blurred faces using local phase quantization. In *ICPR*, 2008.

[3] O. Arandjelovic and R. Cipolla. A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution. In *ICCV*, 2007.

[4] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE T-PAMI*, 24(9):1167–1183, September 2002.

[5] G. Dedeoglu, T. Kanade, and J. August. High-zoom video hallucination by exploiting spatio-temporal regularities. In *CVPR*, pages 151–158, 2004.

[6] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, Boston, MA, second edition, 2001.

[7] S.-W. Lee, J. Park, and S.-W. Lee. Low resolution face recognition based on support vector data description. *Pattern Recognition*, 39:1809–1812, 2006.

[8] Z. Lei, S. Liao, M. Pietikäinen, and S. Z. Li. Face recognition by exploring information jointly in space, scale and orientation. *IEEE T-IP*, 20(1):247–256, Jan. 2011.

[9] B. Li, H. Chang, S. Shan, and X. Chen. Coupled metric learning for face recognition with degraded images. In *ACML*, pages 220–233, 2009.

[10] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *ICB*, pages 828–837, 2007.

[11] C. Liu, H.-Y. Shum, and W. T. Freeman. Face hallucination: Theory and practice. *IJCV*, 75(1):115–134, 2007.

[12] W. Liu, D. Lin, and X. Tang. Hallucinating faces: Tensorpatch super-resolution and coupled residue compensation. In *CVPR*, 2005.

[13] MBGC 2008. http://face.nist.gov/mbgc/.

[14] M. Nishiyama, H. Takeshima, J. Shotton, T. Kozakaya, and O. Yamaguchi. Facial deblur inference to improve recognition of blurred faces. In *CVPR*, pages 1115–1122, 2009.

[15] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. In *ICISP*, 2008.

[16] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE T-PAMI*, 22(10):1090–1104, 2000.

[17] Y. Su, S. Shan, X. Chen, and W. Gao. Hierarchical ensemble of global and local classifiers for face recognition. In *ICCV*, 2007.

[18] P. Viola and M. Jones. Robust real-time face detection. *IJCV*, 57:137–154, 2004.

[19] X. Wang and X. Tang. Hallucinating face by eigentransformation. *IEEE T-SMC (Part C)*, 35(3):425–434, August 2005.

[20] L. Wiskott, J. Fellous, N. Kruger, and C. v. d. Malsburg. Face recognition by elastic bunch graph matching. *IEEE T-PAMI*, 19(7):775–779, 1997.

[21] Y. Wong, C. Sanderson, S. Mau, and B. C. Lovell. Dynamic amelioration of resolution mismatches for local feature based identity inference. In *ICPR*, 2010.

[22] D. Yi, S. Liao, Z. Lei, J. Sang, and S. Z. Li. Partial face matching between near infrared and visual images in mbgc portal challenge. In *ICB*, 2009.

[23] W. C. Zhang, S. G. Shan, W. Gao, and H. M. Zhang. Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In *ICCV*, pages 786–791, 2005.