DERIVATIVE VARIATION PATTERN FOR ILLUMINATION-INVARIANT IMAGE REPRESENTATION

Mohammad Tavakolian[1], *Farshid Hajati*[1,2], *Ajmal S. Mian*[3], *Yongsheng Gao*[2]

[1]Electrical Engineering Department, Tafresh University, Tafresh, Iran
[2]School of Engineering, Griffith University, QLD 4111, Australia
[3]Computer Science and Software Engineering, The University of Western Australia, WA 6009, Australia
m_tavakolian@tafreshu.ac.ir, f.hajati@griffith.edu.au, ajmal@csse.uwa.edu.au, yongsheng.gao@griffith.edu.au

ABSTRACT

This paper presents a novel image descriptor called Derivative Variation Pattern (DVP) and its application to face and palmprint recognition. DVP captures image variations in both the frequency and the spatial domains. The effects of uncontrolled illumination are compensated in the frequency domain by discarding the illumination affected frequencies. Image pixels are encoded as binary patterns based on the higher-order spatial derivatives computed in the spatial domain. The proposed descriptor was evaluated on the Extended Yale-B and FERET face databases, and the PolyU palmprint database. Experimental results demonstrate the effectiveness of the DVP descriptor in both the face and the palmprint recognition tasks under uncontrolled illuminations.

Index Terms— Derivative pattern, spatial domain, frequency domain, descriptor, recognition.

1. INTRODUCTION

Image representation and recognition has attracted much attention during the past few years and the effectiveness of local descriptors has been established in this area. Most image representation methods use only the spatial domain to compute features. The Local Binary Pattern (LBP) operator [1] encodes the local structures of images by comparing each pixel within a neighborhood. LBP was originally proposed for texture analysis however, it has been extended to many applications such as facial image analysis [2], image and video retrieval [3], and motion analysis [4]. Guo et al. [5] proposed Completed LBP (CLBP) for texture classification by representing local regions using the center pixels and a local deference sign-magnitude transform. LBP Variance (LBPV) [6] is a rotation-invariant operator that aligns the LBP histogram of the orientation estimation of the image and uses the histogram's local information for texture classification.

Zhang et al. [7] proposed an object descriptor called Local Derivative Pattern (LDP) that models the high-order local derivative variations. The LDP templates extract high-order local information by encoding various distinctive partial relationships contained in the local regions. Vu et al. [8] introduced a descriptor for object representation called Patterns of Oriented Edge Magnitudes (POEM). POEM is an oriented feature that computes the gradient orientation of the pixels incorporating the gradient information of the pixel's neighborhood. Radon Representation-based Feature Descriptor (RRFD) was proposed by Liu et al. [9] for texture classification. RRFD converts pixels into Radon-pixels using the Radon transform [10] and then projects the Radon-pixels onto a feature space that is invariant to geometric affine transformations. RRFD extracts Radon features that are invariant against geometric affine transformations as well as illumination variations.

While most researchers have used the spatial domain for image representation, others have considered only the frequency domain for image description. Celik et al. [11] proposed a texture classifier which exploits both the magnitude and the phase of the Dual-Tree Complex Wavelet Transform (DT-CWT) sub-bands by extracting a texture feature vector from the magnitude and the phase of DT-CWT sub-bands. Xu et al. [12] proposed a texture descriptor based on Multi-fractal spectrum (MFS) [13] defined on a multi-orientation wavelet pyramid including low-frequency wavelet components [14] and wavelet leaders [15]. MFS is estimated for each individual component in the wavelet pyramid (low-frequency component and high-frequency component and wavelet leaders).

In this paper, we present a novel image descriptor called Derivate Variation Pattern (DVP). DVP describes the image by characterizing pixels with a binary code extracted from both the spatial and frequency variations making an invariant feature to illumination conditions. We evaluate the proposed method under uncontrolled illuminations for the tasks of face and palmprint recognition using the standard Extended Yale-B [16] and FERET [17] face databases and the PolyU palmprint database [18].

2. DERIVAITIVE VARIATION PATTERN (DVP)

External conditions, especially lighting, affect the appearance of an object in an image and consequently the image's frequency coefficients [19]. However, when the light source is far from the object, it is possible to represent the lighting condition by defining the intensity of the light as a function of its direction. According to Lambert's law [20], if an object with the reflectance $r(x, y)$ and the normal vector $\bar{n}(x, y)$ is under the lighting condition $l(x, y)$, the intensity of the captured image $f(x, y)$ can be modeled using the Lambertian reflectance property [19] as

$$f(x, y) = r(x, y) \cdot \max[l(x, y) \cdot \bar{n}(x, y), 0] \quad (1)$$

where *max*$(\cdot, \cdot)$ denotes the maximum operator.

Assuming that the light source is far from the object's surface, we can approximate Equation (1) as

$$f(x, y) \cong r(x, y) \cdot l(x, y) \quad (2)$$

where $r(x, y)$ and $l(x, y)$ denote the reflectance and the illumination parts of the object's surface, respectively.

In Equation (2), $l(x, y)$ is dependent on the illumination changes; while the reflectance part $r(x, y)$ is invariant to illumination changes. We can separate these two parts to obtain an illumination-invariant image of the given object. Since only $f(x, y)$ is known in an image, it is not possible to separate $r(x, y)$ from $l(x, y)$. Therefore, we change the product in Equation (2) to summation by taking the logarithmic transform as

$$\underbrace{\log\left(f(x, y)\right)}_{F(x,y)} = \underbrace{\log\left(r(x, y)\right)}_{R(x,y)} + \underbrace{\log\left(l(x, y)\right)}_{L(x,y)} \quad (3)$$

where *F(x,y)*, *R(x,y)*, and *L(x,y)* are the logarithm of the captured image, the reflectance, and the illumination condition, respectively.

Given the logarithm of the captured image, $F(x, y)$, we compute and remove the effect of the illumination condition in the frequency domain. Using the Total Variation (TV-L^1) model [21], the illumination variation of the captured image, $L^*(x, y)$, can be approximated as

$$L^*(x, y) = \arg\min_L \iint (\|\nabla L(x, y)\| + \lambda \|F(x, y) - L(x, y)\|_1) dxdy \quad (4)$$

where λ is a parameter that balances the first and the second terms of Equation (4) and "1" subscript denotes the L-1 norm [22].

Substituting $R(x, y) = F(x, y) - L(x, y)$ from Equation (3) in Equation (4) leads us to an estimation of the reflectance part, $R^*(x, y)$, (neglecting the $F(x, y)$ term as a fixed term in the left side of the equation) as

$$R^*(x, y) = \arg\min_R \iint (\|\nabla(F(x, y) - R(x, y))\| + \lambda \|R(x, y)\|) dxdy \quad (5)$$

Since the illumination condition, $L(x, y)$, varies slower than the reflectance, $R(x, y)$ [23], we estimate $R^*(x, y)$ from $F(x, y)$ by adjusting the low-frequency components and keeping the high-frequency components. Here, we define the low-frequency components which are affected by the illumination changes with a novel method. For this purpose, we use Discrete Cosine Transform (DCT) [24] to transform the image $F(x, y)$ from the spatial domain to the frequency domain and eliminate the frequency variation by discarding the low-frequency sub-bands that are most affected by the illumination changes.

Given the logarithm of the captured image, $F(x, y)$, the Discrete Cosine Transform (DCT) of the logarithmic captured image is defined as

$$\hat{F}(u, v) = DCT\{F(x, y)\} \quad (6)$$

where $\hat{F}(u, v)$ is the DCT transform of the logarithmic captured image.

Using the DCT transform, we define $2N - 1$ frequency sub-bands for the given $N \times N$ image. For an $N \times N$ image, the sub-bands are numbered from 1 to $2N - 1$ (see Figure 1).

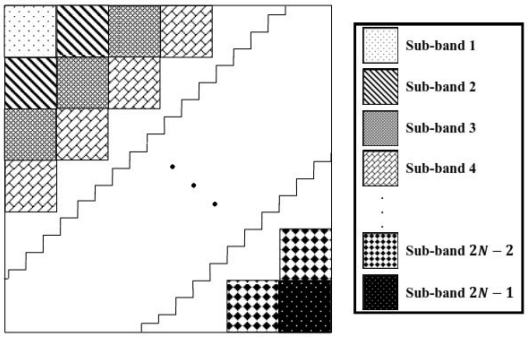

Fig. 1: Definition of the frequency sub-bands in DCT space.

Discarding the most affected frequency sub-bands in the DCT space and using the inverse DCT function, we compute an illumination compensated image $\tilde{R}_n(x, y)$ as

$$\tilde{R}_n(x, y) = IDCT\{\hat{F}(u, v) \cdot Q_n(u, v)\} \quad (7)$$

where $IDCT\{\cdot\}$ denotes the Inverse Discrete Cosine Transform and $Q_n(u, v)$ is a mask matrix in which the most affected frequency sub-bands are zeros and others are ones.

In order to determine the best mask matrix $Q_n(u, v)$, we define a cost function using Equation (5) as

$$J(F(x, y), \tilde{R}_n(x, y)) = \arg\min_n \iint (\|\nabla\big(F(x, y) - \tilde{R}_n(x, y)\big)\| + \lambda \|\tilde{R}_n(x, y)\|) dxdy \quad (8)$$

where $F(x, y)$ is the original logarithmic input image and $\tilde{R}_n(x, y)$ is an image in which the most affected frequency sub-bands are discarded using Equation (7).

By removing the frequency sub-bands, sequentially, and computing the cost function in Equation (8), we determine the frequency sub-band in which the cost function is minimum. The minimum of the cost function corresponds to

the least frequency variation in the image caused by illumination changes. In this way, we discard the affected low-frequency sub-bands and obtain an illumination compensated image $\tilde{R}_n(x,y)$ which is used for the spatial image description.

In order to define the spatial descriptor, we consider each pixel as the center of a circle with radius r (see the bold circle in Figure 2). This circle specifies a neighborhood called the primary circle. We draw another circle, centered on the primary circle's pixels, with the same radius r (see the dashed circle in Figure 2) which specifies another neighborhood called the secondary circle.

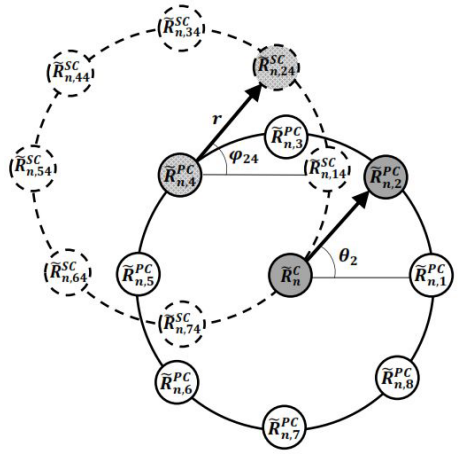

Fig. 2: Defining the primary and the secondary circles.

Assuming that the gray level of the central pixel is $\tilde{R}_n^C(x,y)$, the gray level of the neighbors on the primary circle are defined as

$$\tilde{R}_{n,i}^{PC}(x_i,y_i) = \tilde{R}_n(x - r\sin(\theta_i), y + r\cos(\theta_i)) \quad (9)$$

where $\tilde{R}_{n,i}^{PC}(x_i,y_i)$, $i = 1,\ldots,8r$, denotes the gray level of the primary circle's pixels. r is the radius of the circles and θ_i is the angle between the x axis and the i-th neighbor in the primary circle.

Each point on the primary circle is surrounded by $8r$ neighbors on the secondary circle that their gray levels are defined as

$$\tilde{R}_{n,ji}^{SC}(x_j,y_j) = \tilde{R}_n(x_i - r\sin(\varphi_{ji}), y_i + r\cos(\varphi_{ji})) \quad (10)$$

where $\tilde{R}_{n,ji}^{SC}(x_j,y_j)$, $j = 1,\ldots,8r$, denotes the j-th neighbor of the i-th pixel on the primary circle. φ_{ji} is the angle of the j-th neighbor of the i-th pixel on the primary circle with respect to the x axis.

Using Equations (9) and (10), we compute the first-order spatial derivatives as

$$\tilde{R}_{n,i,\theta}^{(1)}(x,y) = \tilde{R}_{n,i}^{PC}(x_i,y_i) - \tilde{R}_n^C(x,y) \quad (11)$$

$$\tilde{R}_{n,ji,\varphi}^{(1)}(x,y) = \tilde{R}_{n,ji}^{SC}(x_j,y_j) - \tilde{R}_{n,i}^{PC}(x_i,y_i) \quad (12)$$

where $\tilde{R}_{n,i,\theta}^{(1)}(x,y)$ and $\tilde{R}_{n,ji,\varphi}^{(1)}(x,y)$ denote the first-order derivative within the primary circle and the first-order derivative between the primary and the secondary circle, respectively.

The second-order derivative in the primary circle, $\tilde{R}_{n,i,\theta}^{(2)}(x,y)$, and the second-order derivative between the primary and the secondary circles, $\tilde{R}_{n,ji,\varphi}^{(2)}(x,y)$, are defined as

$$\tilde{R}_{n,i,\theta}^{(2)}(x,y) = \tilde{R}_{n,i}^{PC^{(1)}}(x_i,y_i) - \tilde{R}_n^{C^{(1)}}(x,y) \quad (13)$$

$$\tilde{R}_{n,ji,\varphi}^{(2)}(x,y) = \tilde{R}_{n,ji}^{SC^{(1)}}(x_j,y_j) - \tilde{R}_{n,i}^{PC^{(1)}}(x_i,y_i) \quad (14)$$

Generally, the m^{th}-order derivatives are calculated based on the $(m-1)^{th}$-order as

$$\tilde{R}_{n,i,\theta}^{(m)}(x,y) = \tilde{R}_{n,i}^{PC^{(m-1)}}(x_i,y_i) - \tilde{R}_n^{C^{(m-1)}}(x,y) \quad (15)$$

$$\tilde{R}_{n,ji,\varphi}^{(m)}(x,y) = \tilde{R}_{n,ji}^{SC^{(m-1)}}(x_j,y_j) - \tilde{R}_{n,i}^{PC^{(m-1)}}(x_i,y_i) \quad (16)$$

Using the m^{th}-order derivative within the primary circle and between the primary and the secondary circles, we define the m^{th}-order Derivative Variation (DV) for the image $\tilde{R}_n(x,y)$ as

$$DV_r^{(m)}\big(\tilde{R}_n(x,y)\big) = \{u\big(\tilde{R}_{n,i,\theta}^{(m)}(x,y)\cdot\tilde{R}_{n,ji,\varphi}^{(m)}(x,y)\big)| \quad (17)$$
$$i = 1,\cdots,8r \;\;\&\;\; j = 1,\cdots,4r\}$$

where $DV_r^{(m)}(\tilde{R}_n(x,y))$ returns the m^{th}-order derivative variation using a neighborhood with radius r, and $u(\cdot)$ is the unit step function.

Using the m^{th}-order derivative variation, we compute the m^{th}-order Derivative Variation Pattern (DVP) of the image as

$$DVP_r^{(m)}\big(\tilde{R}_n(x,y)\big) = \sum_{k=1}^{32r^2} DV_{r,k}^{(m)}\big(\tilde{R}_n(x,y)\big)\; 2^{32r^2-k} \quad (18)$$

where $DV_{r,k}^{(m)}$ is the k-th component of the m^{th}-order derivative variation computed using Equation (17).

Figure 3 illustrates the extraction of DVP on a face image. Notice that as the order of the operator increases, details are extracted from the image even in adverse lighting conditions.

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

Fig. 3: Extracted DVPs. (a) original face image, (b) the illumination compensated image ($\tilde{R}_n(x,y)$), (c) first-order DVP, (d) second-order DVP, (e) third order DVP.

3. EXPERIMENTAL RESULTS

We evaluate the performance of the proposed method using Extended Yale-B [16] and FERET [17] face databases and the

PolyU palmprint database [18]. The Extended Yale-B face database contains a total of 38 subjects with severe illumination variations. Each subject was imaged under nine poses and 64 different illumination conditions. The FERET face database consists of a total of $14,051$ gray-scale images representing $1,199$ individuals. The images contain variations in lighting, facial expression, pose angle, etc. In this work, only the frontal face images are considered. The PolyU palmprint database contains $7,752$ gray-scale images corresponding to 386 different palms (around twenty samples per subject). All the frontal face images were normalized by aligning the location of the eyes and cropped to 160×160 pixels, while the palmprint images were cropped to 128×128 regions of interest. The first face/palmprint image from each subject was selected to make a gallery set and all the remaining frontal images were used as the probe.

First, the parameter of the proposed algorithm (i.e. r) is determined using a randomly selected dataset from the Extended Yale-B database. Our results show that the optimum values for the primary and the secondary circles' radius is $r = 1$. Hence, we select $r = 1$ in the all experiments.

Second, the rank-1 recognition rate of the proposed algorithm is computed using DVPs with different orders for the three databases (see Figure 4). As can be seen, the recognition accuracy is significantly improved when the order of local pattern is increased from the first-order DVP to the second-order DVP in all databases. Then, the performance drops when the DVP's order increases. These results prove the effectiveness of the second-order DVP in extracting more discriminative information even with severe illumination variation. Therefore, we use second-order DVP in the remaining experiments.

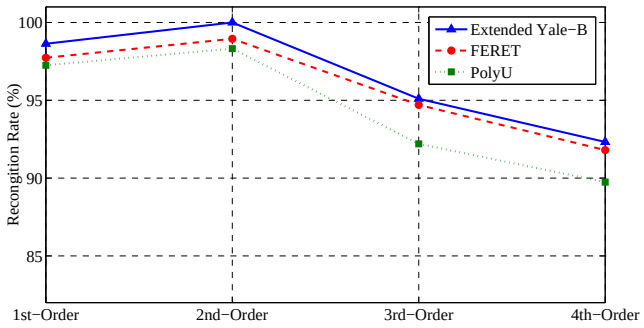

Fig. 4: Recognition rate versus DVP's order.

Finally, the rank-1 recognition rates of DVP on the face and palmprint databases are compared to the state-of-the-art approaches (see Table 1 and Table 2). As can be seen from Table 1, DVP achieves 6.25% and 7.15% performance improvement over LBP and LDP on the FERET face database, respectively. It also improves the recognition accuracy by 24.10%, 7.10%, 1.90%, and 1.04% over LBP, LDP, Sparse Representation-based Classification (SRC) [25], and Gradientfaces [26] methods respectively on the Extended Yale-B face database showing that the DVP significantly improves the performance of the face recognition system even in adverse lighting conditions.

In Table 2, the rank-1 recognition rate of DVP and other benchmark palmprint recognition approaches are reported. Here, DVP achieves 98.32% identification rate compared to 84.67%, 93.33%, and 95.71% identification rates for 2D Locality Preserving Projections (2DLPP) [27], 2D Local Graph Embedding Discriminant Analysis (2DLGEDA) [28], and Discriminant Projection Embedding (DPE) [29], respectively. Hence, DVP outperforms current benchmark techniques for palmprint identification.

Table 1: Comparison the rank-1 recognition rate (%) of DVP and the benchmarks in the face recognition task.

Method	Extended Yale-B	FERET
*LBP [2]	75.90	92.70
*LDP [7]	92.90	91.80
SRC [25]	98.10	N/A
Gradientfaces [26]	98.96	N/A
DVP	**100**	**98.95**

*The results are from [7]

Table 2: Comparison the ranke-1 recognition rate (%) of DVP and the benchmarks in the palmprint recognition task.

Method	PolyU
2DLPP [27]	84.67
2DLGEDA [28]	93.33
DPE [29]	95.71
DVP	**98.32**

4. CONCLUSION

A novel image descriptor called Derivative Variation Pattern (DVP) is proposed which simultaneously captures the spatial and the frequency domain features of an image. An image is first transformed to the frequency domain using Discrete Cosine Transform, and the low-frequency sub-bands that are most affected by the illumination variation are discarded. This makes the algorithm invariant to illumination changes. In the spatial domain, a binary code is obtained by taking radial derivatives along two circular paths. The proposed DVP descriptor was tested on three standard databases and compared to benchmark techniques for the task of person identification. Our results show the effectiveness of the proposed method.

ACKNOWLEDGEMENT

Ajmal S. Mian was supported by ARC Discovery grant DP110102399.

REFERENCES

[1] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. on PAMI*, vol. 24, pp. 971–987, 2002.

[2] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. on PAMI*, vol. 28, pp. 2037–2041, 2006.

[3] D. Grangier and S. Bengio, "A discriminative kernel-based approach to rank images from text queries," *IEEE Trans. on PAMI*, vol. 30, pp. 1371–1384, 2008.

[4] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. on PAMI*, vol. 28, pp. 657–662, 2006.

[5] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. on Image Processing*, vol. 19, pp. 1657–1663, 2010.

[6] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP Variance (LBPV) with global matching," *Pattern Recognition*, vol. 43, pp. 706–719, 2010.

[7] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Trans. on Image Processing*, vol. 19, pp. 533–544, 2010.

[8] N. S. Vu, H. M. Dee, and A. Caplier, "Face recognition using the POEM descriptor," *Pattern Recognition*, vol. 45, pp. 2478–2488, 2012.

[9] G. Liu, Z. Lin, and Y. Yu, "Radon representation-based feature descriptor for texture classification," *IEEE Trans. on Image Processing*, vol. 18, pp. 921–928, 2009.

[10] S. R. Deans, *The Radon Transform and Some of Its Applications*, Wiley, 1983.

[11] T. Celik and T. Tjahjadi, "Bayesian texture classification and retrieval based on multiscale feature vector," *Pattern Recognition Letters*, vol. 32, pp. 159–167, 2011.

[12] Y. Xu, X. Yong, H. Ling, and H. Ji, "A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid," in *IEEE CVPR*, 2010, pp. 161–168.

[13] Y. Xu, S. B. Huang, H. Ji, and C. Fermuller, "Combining powerful local and global statistics for texture description," in *IEEE CVPR*, 2009, pp. 573–580.

[14] S. G. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, CA: San Diego: Academic Press, 2008.

[15] H. Wendt, S. G. Roux, S. Jaffard, and P. Abry, "Wavelet leaders and bootstrap for multifractal analysis of images," *Signal Processing*, vol. 89, pp. 1100–1114, 2009.

[16] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. on PAMI*, vol. 23, pp. 643–660, 2001.

[17] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. on PAMI*, vol. 22, pp. 1090–1104, 2000.

[18] The Hong Kong Polytechnic University (PolyU) Palmprint Database (The Second Version) www.comp.polyu.edu.hk/∼biometrics, ," .

[19] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. on PAMI*, vol. 25, pp. 218–233, 2003.

[20] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible lighting conditions?," in *IEEE CVPR*, 1996, pp. 270–277.

[21] T. Chen, W. T. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," *IEEE Trans. on PAMI*, vol. 28, pp. 1519–1524, 2006.

[22] W. Yin, D. Goldfarb, and S. Osher, "The total variation regularized L1 model for multiscale decomposition," *Multiscale Modeling and Simulation*, vol. 6, pp. 190–211, 2007.

[23] T. Zhang, B. Fang, Y. Yuan, Y. Y. Tang, Z. Shang, D. Li, and F. Lang, "Multiscale facial structure representation for face recognition under varying illumination," *Pattern Recognition*, vol. 42, pp. 251–258, 2009.

[24] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Processing*, vol. 9, pp. 1661–1666, 2000.

[25] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and M. Yi, "Robust face recognition via sparse representation," *IEEE Trans. on PAMI*, vol. 31, pp. 210–227, 2009.

[26] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Trans. on Image Processing*, vol. 18, pp. 2599–2606, 2009.

[27] D. Hu, G. Feng, and Z. Zhou, "Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition," *Pattern Recognition*, vol. 40, pp. 339–342, 2007.

[28] M. Wan, Z. Lai, J. Shao, and Z. Jin, "Two-dimensional local graph embedding discriminant analysis (2DLGEDA) with its application to face and palm biometrics," *Neurocomputing*, vol. 73, pp. 197–203, 2009.

[29] Y. Yan and Y. J. Zhang, "Discriminant projection embedding for face and palmprint recognition," *Neurocomputing*, vol. 71, pp. 3534–3543, 2008.