

Client Specific Image Gradient Orientation for Unimodal and Multimodal Face Representation

He-Feng Yin, Xiao-Jun Wu^(✉), and Xiao-Qi Sun

School of IoT Engineering, Jiangnan University, Wuxi 214122, China
yinhefeng@126.com, xiaojun_wu_jnu@163.com,
704382190@qq.com

Abstract. Multimodal face recognition systems usually provide better recognition performance compared to systems based on a single modality. To exploit this advantage, in this paper, an image fusion method which integrates region segmentation and pulse coupled neural network (PCNN) is used to obtain fused images by using visible (VIS) images and infrared (IR) images. Then, client specific image gradient orientation (CSIGO) is proposed which is inspired by the successful application of client specific technique and image gradient orientations technique. As most of the traditional appearance-based subspace learning algorithms are not robust to illumination changes, to remedy this problem to some extent, we adopt the image gradient orientations method. Moreover, to better describe the discrepancies between different classes, client specific is introduced to derive one dimensional Fisher face per client. Thus CS-IGO-LDA and improved CS-IGO-LDA are proposed in this paper, which combine the merits of IGO and client specific technique. Experimental results obtained on publicly available databases indicate the effectiveness of the proposed methods on unimodal and multimodal face recognition.

Keywords: Image fusion · PCNN · Client specific · Image gradient orientations · Multimodal face recognition

1 Introduction

Face recognition (FR) remains one of the most challenging research topics in computer vision, machine learning and biometrics. In face recognition, face representation plays a vital part. The most widely investigated methods for face representation are linear subspace learning approaches. Principal Component Analysis (PCA) is a typical feature extraction technique widely used in the field of pattern recognition. Based on PCA, the well-known Eigenfaces [1] method for face recognition was developed. In order to use the discriminatory information between different classes, approaches based on Linear Discriminant Analysis (LDA) were studied [2–4]. However, both PCA and LDA fail to discover the nonlinear structure in facial images. Since the important information may be contained in higher order relationships among image pixels of the face pattern [5, 6], the study of kernel principal component analysis (KPCA) [7] and kernel discriminant analysis (KDA) [8] have been attractive topics in pattern recognition.

Visible face image is easily affected by illumination changes which can introduce adverse effect to the performance of face recognition. Thermal infrared image is robust to changes in illumination; however, it is sensitive to temperature changes in the surrounding environment. To exploit the complementary information in the visible and infrared images, we present a method to fuse visible and infrared images based on region segmentation and PCNN. Experimental results show that recognition accuracy on the fused images is better than that on visible images or infrared images.

Conventional subspace learning approaches and their variants are based on a global representation of all the training samples. Departing from this paradigm, Kittler et al. [9] proposed a client specific Fisher face representation in which the client enrollment is insulated from the enrollment of other clients and there is only one Fisher face per client. An improved model of client specific linear discriminant analysis (CSLDA) method was developed by Wu et al. [10]. To incorporate various dimensionality reduction methods with client specific subspace, Sun et al. [11] proposed a framework called the hybrid client specific discriminant analysis.

In addition, most traditional subspace learning methods are based on pixel intensities. However, in practical applications, since the distribution of noise in images is substantially different from Gaussian, classical feature extraction methods from pixel intensities always fail to reliably estimate the low-dimensional subspace of a given data population. To tackle this problem, subspace learning from image gradient orientations (IGO) is proposed in [12], which replaces pixel intensities with gradient orientations.

In order to obtain a more robust face representation and better describe the discrepancies between different classes, in this paper, we present two methods, CS-IGO-LDA and improved CS-IGO-LDA. Experimental results obtained on three publicly available databases demonstrate the effectiveness of our proposed methods for unimodal and multimodal face recognition.

The rest of the paper is organized as follows. IGO and image fusion method are reviewed in Sect. 2. Section 3 presents the proposed CS-IGO-LDA and improved CS-IGO-LDA methods respectively. Experiments on publicly available databases are conducted in Sect. 4. Finally, conclusions are drawn in Sect. 5.

2 Related Works

2.1 An Overview of Image Gradient Orientations

A set of face image samples $\{Z_i\}(i = 1, 2, \dots, N)$, where N is the number of training samples and $Z_i \in \mathbb{R}^{m \times n}$. It is assumed that each image belongs to one of the C classes denoted by $w_j, j = 1, 2, \dots, C$. And the number of samples of the j th class is N_j .

Let $I(x, y)$ be the image intensities at pixel coordinates (x, y) of sample Z_i , we can denote the corresponding horizontal and vertical gradients respectively as:

$$\begin{aligned} G_{i,x} &= h_x * I(x, y) \\ G_{i,y} &= h_y * I(x, y) \end{aligned} \tag{1}$$

where $*$ is convolution operator, h_x and h_y are filters used to approximate the ideal differentiation operator along the image horizontal and vertical directions respectively.

Then the gradient orientation of the pixel location (x, y) is:

$$\Phi_i(x, y) = \arctan \frac{G_{i,y}}{G_{i,x}}, i = 1, 2, \dots, N \quad (2)$$

Let ϕ_i be the 1D representation of Φ_i . Referring to [13], we also define the mapping from $[0, 2\pi)^K$ ($K = m \times n$) onto a subset of complex sphere with radius \sqrt{K}

$$t_i(\phi_i) = e^{j\phi_i} \quad (3)$$

where $e^{j\phi_i} = [e^{j\phi_i(1)}, e^{j\phi_i(2)}, \dots, e^{j\phi_i(K)}]^T$, and $e^{j\theta}$ is Euler form, i.e. $e^{j\theta} = \cos\theta + j\sin\theta$.

By applying the procedure of conventional PCA and LDA to t_i , we can obtain projection matrices U_P for IGO-PCA and U_L for IGO-LDA, respectively.

2.2 Image Fusion

The aim of image fusion is to integrate complementary and redundant information from multiple images to create a composite that contains a better description of the scene than any of the individual source images [14]. This fused image should increase the performance of the subsequent processing tasks. Conventional image fusion methods usually employ pixel-based approaches which is sensitive to misregistration. Furthermore, we are not interested in individual pixels but rather in the objects and regions they represent. It therefore seems reasonable to incorporate object and region information into the fusion process. In this paper, we adopt a new feature level image fusion method which integrates region segmentation and PCNN [15]. First, the fuzzy c-means clustering algorithm is used to segment the image in the feature space formed by multi-channel Gabor filters and then multi-scale wavelet decomposition is performed on the image. Second, the low frequency coefficients are fused with edge intensity and the high frequency coefficients are fused with PCNN for all regions. Finally, the fused image is obtained by taking the inverse wavelet transform. Figure 1 shows some visible and infrared images from the OTCBVS database [16] and their corresponding fused images obtained by our method.



Fig. 1. Visible images (the left two images), infrared images (the middle two images) and the corresponding fused images (the right two images)

3 Client Specific Image Gradient Orientations (CSIGO)

Traditional face representation approaches are based on a global representation of all the training samples. Client specific Fisherfaces can derive only one Fisher face per client. In this paper, to further improve the performance of client specific methods, we combine the advantages of client specific technique and image gradient orientations. Therefore, CS-IGO-LDA and improved CS-IGO-LDA are presented. The proposed methods for multimodal face recognition consist of four steps. Firstly, visible images and infrared images are fused to obtain the corresponding fused images. Secondly, IGO-methods are utilized to reduce the dimensionality of facial images. Then a single Fisher face per subject is obtained via client specific approach. Finally, classification is conducted accordingly. Flowchart of the proposed methods is illustrated in Fig. 2.

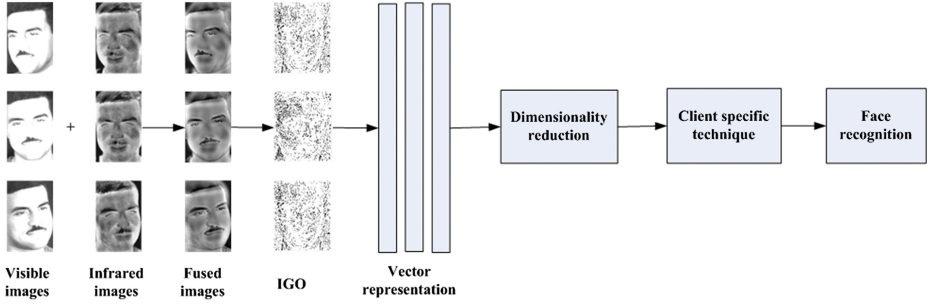


Fig. 2. Flowchart of the proposed methods on multimodal face recognition

3.1 CS-IGO-LDA

Let t_i represent the transformed data in Sect. 2 (i.e. vector representation of training samples in Fig. 2), the total mean of t_i is denoted by m' ,

$$m' = \frac{1}{N} \sum_{i=1}^N t_i \quad (4)$$

In Fig. 2, when we obtain the orthonormal bases U_P by using IGO-PCA, then the projected samples are:

$$p_i = U_P^H (t_i - m') \quad (5)$$

Let us denote the mixture covariance matrix of the projected vectors by S , i.e.

$$S = \frac{1}{N} \sum_{i=1}^N p_i p_i^H \quad (6)$$

The mean of the j th class w_j can be obtained as

$$\mu_j = \frac{1}{N_j} \sum_{i=1}^{N_j} p_i \quad p_i \in w_j \quad (7)$$

In [10], Wu utilized an equivalent Fisher criterion function defined as

$$J(v) = \frac{v^H B_j v}{v^H S v} \quad (8)$$

where B_j is the between class scatter matrix.

The optimal solution to the client specific discriminant problem (8) can be found as

$$v_j = S^{-1} \mu_j \quad (9)$$

After client specific technique is performed to the projected samples in Fig. 2, client specific Fisher face for each class can be derived as

$$a_j = U_p v_j \quad (10)$$

Readers can refer to [10] for more details about the improved client specific discriminant analysis algorithm.

3.2 Improved CS-IGO-LDA

In the previous section, orthonormal bases of IGO-PCA are used for dimensionality reduction prior to client specific technique. To fully utilize the label information in the training samples, orthonormal bases of IGO-LDA can also be exploited to reduce the dimensionality of transformed data t_i . Procedures of the improved CS-IGO-LDA are summarized as follows.

Input: A set of training sample $Z_i, i = 1, 2, \dots, N$ for C classes, a test sample y .

Step 1. Compute orientation images for Z_i and y , denoted by Φ_i and Φ_y , respectively.

Step 2. Transform Φ_i and Φ_y into 1D representation, expressed by ϕ_i and ϕ_y .

Step 3. Compute $t_i(\phi_i) = e^{i\phi_i}$ and $t_y(\phi_y) = e^{i\phi_y}$.

Step 4. Perform IGO-LDA on the mapped data $X = [t_1, t_2, \dots, t_N]$, obtain the set of orthonormal bases U_L .

Step 5. Project mapped data X onto U_L , then obtain the client specific Fisher face $a_k, k = 1, 2, \dots, C$.

Step 6. Project X and t_y onto a_k , then classify y according to the minimum distance classifier.

Output: Identity of y .

4 Experimental Results and Analysis

In order to evaluate the performance of the proposed methods in this paper, first Yale and Extended Yale B [17] are used for unimodal face recognition, then the OTCBVS database is chosen for unimodal and multimodal face recognition. For Yale and Extended Yale B databases, we used images of size 32×32 . In order to tackle the singular problem of LDA-based approaches, PCA-based methods were utilized for dimensionality reduction. For example, for Fisherfaces and IGO-LDA we used PCA to preserve $N - C$ dimensions, and for CS-IGO-LDA we used PCA to preserve 45 dimensions. Experiments on all three databases were repeated 10 times and the average recognition accuracy was recorded. CS-IGO-LDA1 represents the CS-IGO-LDA algorithm and CS-IGO-LDA2 denotes the improved CS-IGO-LDA approach.

4.1 Face Recognition on the Yale Database

The Yale database contains images from 15 individuals, each providing 11 different images. All images are gray-scale and normalized to a resolution of 160×121 pixels. We randomly selected 5 images from each subject for training, whereas the remaining were used for testing. It should be noted that LDA-based methods are different from the other features because the maximal number of valid features is $C - 1$, C is the number of classes. Recognition accuracy of different methods are shown in Table 1 and the corresponding curves are illustrated in Fig. 3.

Table 1. Recognition accuracy on the Yale database.

| Dimension | 3 | 7 | 11 | 14 | 30 | 40 |
|-------------|-------|-------|-------|-------|-------|-------|
| Eigenfaces | 60.44 | 72.89 | 78.22 | 78.00 | 83.33 | 83.56 |
| Fisherfaces | 69.11 | 80.44 | 84.44 | 88.22 | N/A | N/A |
| CS-LDA | 60.89 | 83.11 | 88.89 | 92.00 | N/A | N/A |
| IGO-PCA | 60.89 | 74.44 | 78.07 | 80.44 | 86.44 | 87.56 |
| IGO-LDA | 78.00 | 86.67 | 90.22 | 91.56 | N/A | N/A |
| CS-IGO-LDA1 | 67.56 | 78.00 | 84.44 | 86.44 | 92.22 | 93.33 |
| CS-IGO-LDA2 | 79.33 | 89.33 | 92.22 | 92.98 | N/A | N/A |

4.2 Face Recognition on the Extended Yale B Database

The Extended Yale B database contains 16,128 images of 38 subjects under nine poses and 64 illumination conditions. We used a subset which consists of 64 near frontal images for each subject. For training, we randomly selected a subset with 31 images per subject. For testing, we used the remaining images. Finally, we performed 10 different random realizations of the training/test sets. Table 2 shows the obtained results.

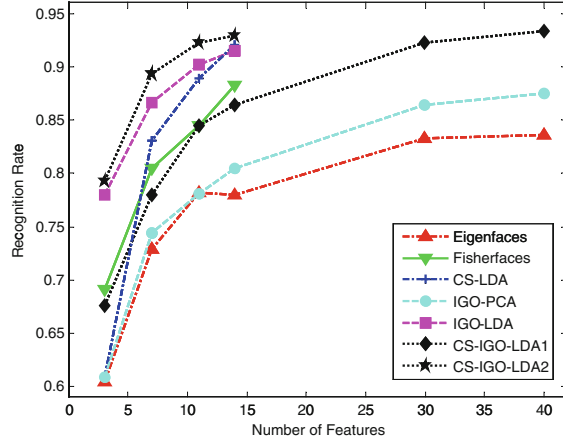


Fig. 3. Curves of recognition accuracy by CS-IGO-LDA and improved CS-IGO-LDA and their competing methods versus feature dimensionality on the Yale database

Table 2. Recognition accuracy on the Extended Yale B database.

| Dimension | 10 | 30 | 37 | 60 | 100 | 140 |
|-------------|-------|-------|-------|-------|-------|-------|
| Eigenfaces | 22.11 | 50.54 | 54.75 | 62.32 | 66.58 | 68.57 |
| Fisherfaces | 61.53 | 74.57 | 76.84 | N/A | N/A | N/A |
| CS-LDA | 73.77 | 89.56 | 90.12 | N/A | N/A | N/A |
| IGO-PCA | 81.73 | 95.81 | 96.13 | 96.71 | 96.83 | 96.91 |
| IGO-LDA | 96.12 | 97.01 | 97.02 | N/A | N/A | N/A |
| CS-IGO-LDA1 | 90.61 | 96.94 | 97.11 | 97.12 | 97.19 | 97.19 |
| CS-IGO-LDA2 | 96.96 | 97.29 | 97.25 | N/A | N/A | N/A |

4.3 Face Recognition on the OTCBVS Database

OTCBVS database contains 4228 pairs of visible and IR thermal images under variable illuminations, expressions, and poses. Image size is 320×240 pixels. There are 30 subjects in the database, with 176–250 images per person. Illumination conditions are Lon (left light on), Ron (right light on), 2on (both lights on), dark (dark room), off (left and right lights off). In our experiments, we selected face images from 14 different subjects and each subject provides 8 pairs of thermal infrared and visible images, which contain 4 types of illumination: Lon, Ron, 2on and off. There are 2 pose images for each illumination condition.

Out of the 8 images per subject, we randomly selected 4 images per person for training and the remaining for test. All the images were cropped and resized to 64×64 pixels. Nearest neighbor classifier (NNC) with cosine distance was employed to classify test samples. Recognition accuracy on the visible images, infrared images and fused images are shown in Tables 3, 4 and 5, respectively. In order to give a vivid illustration of our proposed methods on the fused images, Fig. 4 is plotted which shows

Table 3. Recognition accuracy on the visible images from the OTCBVS database.

| Dimension | 5 | 10 | 13 | 20 | 30 |
|-------------|-------|-------|-------|-------|-------|
| Eigenfaces | 65.54 | 81.25 | 83.93 | 86.07 | 87.86 |
| Fisherfaces | 59.64 | 64.29 | 64.64 | N/A | N/A |
| CS-LDA | 79.46 | 84.46 | 85.71 | N/A | N/A |
| IGO-PCA | 69.64 | 80.18 | 80.89 | 83.57 | 84.29 |
| IGO-LDA | 75.18 | 82.50 | 84.82 | N/A | N/A |
| CS-IGO-LDA1 | 67.14 | 81.36 | 83.50 | 84.50 | 85.64 |
| CS-IGO-LDA2 | 71.25 | 83.21 | 84.64 | N/A | N/A |

Table 4. Recognition accuracy on the infrared images from the OTCBVS database.

| Dimension | 5 | 10 | 13 | 20 | 30 |
|-------------|-------|-------|-------|-------|-------|
| Eigenfaces | 51.61 | 71.96 | 76.61 | 81.79 | 83.21 |
| Fisherfaces | 69.11 | 75.00 | 75.18 | N/A | N/A |
| CS-LDA | 65.00 | 81.61 | 83.93 | N/A | N/A |
| IGO-PCA | 64.11 | 78.93 | 81.96 | 82.50 | 85.00 |
| IGO-LDA | 75.18 | 81.07 | 83.75 | N/A | N/A |
| CS-IGO-LDA1 | 66.43 | 78.57 | 83.29 | 84.54 | 85.25 |
| CS-IGO-LDA2 | 74.82 | 82.82 | 84.25 | N/A | N/A |

Table 5. Recognition accuracy on the fused images.

| Dimension | 5 | 10 | 13 | 20 | 30 |
|-------------|-------|-------|-------|-------|-------|
| Eigenfaces | 56.61 | 76.07 | 79.82 | 85.71 | 87.86 |
| Fisherfaces | 69.11 | 75.00 | 75.54 | N/A | N/A |
| CS-LDA | 70.18 | 83.93 | 85.89 | N/A | N/A |
| IGO-PCA | 67.14 | 83.93 | 84.64 | 86.79 | 87.50 |
| IGO-LDA | 78.04 | 84.46 | 86.07 | N/A | N/A |
| CS-IGO-LDA1 | 72.86 | 85.04 | 86.46 | 88.39 | 88.29 |
| CS-IGO-LDA2 | 82.86 | 85.93 | 87.29 | N/A | N/A |

the recognition accuracy of CS-IGO-LDA and the improved CS-IGO-LDA obtained on visible images, infrared images and fused images, respectively.

Based on the above experimental results on the Yale database, Extended Yale B database and the OTCBVS database, we can have the following observations:

1. In most cases, recognition accuracy of CSLDA is higher than that of Fisherfaces, this indicates Fisher faces obtained by client specific method have more discriminative information than traditional LDA method whose feature space is shared by all classes.

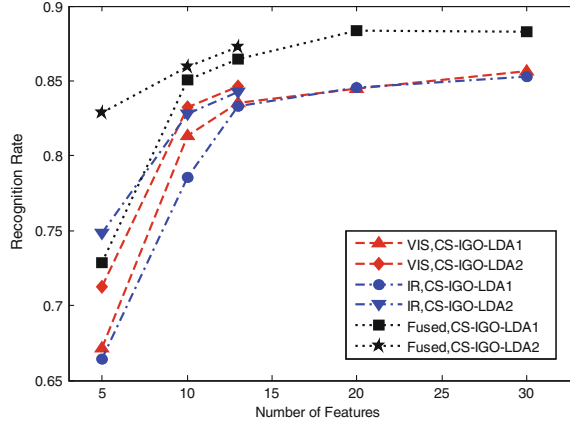


Fig. 4. Curves of recognition accuracy by CS-IGO-LDA and improved CS-IGO-LDA on visible images, infrared images and fused images, respectively

2. Recognition accuracy of PCA and LDA based on IGO are better than that of their intensity-based counterparts, i.e. Eigenfaces and Fisherfaces, this demonstrates that features extracted from image gradient orientations can handle illumination changes to some degree.
3. As expected, by combining the merits of client specific and image gradient orientations, performance of the IGO-based methods is improved. In addition, using the fused images is superior to the visible and infrared images, this indicates that our proposed CS-IGO-LDA and its improved version are feasible and effective for multimodal face recognition.

5 Conclusions

Face recognition on visible image is easily affected by illumination changes. In contrast to visible image, infrared image is robust to illumination changes but it is sensitive to temperature changes in the surrounding environment. To fully exploit the complementary information in visible image and infrared image, an image fusion method which integrates region segmentation and PCNN is adopted. Moreover, in order to improve the performance of IGO methods, CS-IGO-based methods are proposed by combining the advantages of client specific technique and IGO. Experimental results obtained on three publicly available databases not only verify the effectiveness of our proposed methods for unimodal face recognition, but also demonstrate that our proposed methods can achieve better performance on the fused images than that on visible images or infrared images.

Compared with conventional subspace learning approaches and face recognition techniques, our contributions are as follows: (1) image fusion is adopted to integrate information from different domains. (2) IGO and CS technique are combined with LDA method, leading to two new algorithms of face representation.

Although our proposed methods obtain superior results to previous approaches, some additional works are necessary to meet the demands of real-word applications. Thus in future, we will investigate a more robust method for face recognition and authentication.

Acknowledgments. This work was supported by National Natural Science Foundation (NNSF) of China (61373055), Key Grant Project of Chinese Ministry of Education (311024), Innovation Project of Graduate Education of Jiangsu Province (KYLX_1123) and Project of Jiangsu Provincial Department of Science and Technology (Grant No. BY2012059).

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991)
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.: Eigenfaces vs fisherfaces recognition using class specific linear projection. *IEEE PAMI* **9**(7), 711–720 (1997)
3. Chen, X., Yang, J., Zhang, D., Liang, J.: Complete large margin linear discriminant analysis using mathematical programming approach. *Pattern Recognit.* **46**(6), 1579–1594 (2013)
4. Yao, C., Lu, Z., Li, J., Xu, Y., Han, J.: A subset method for improving linear discriminant analysis. *Neurocomputing* **138**, 310–315 (2014)
5. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**(5500), 2323–2326 (2000)
6. Zhang, C., Wang, J., Zhao, N., Zhang, D.: Reconstruction and analysis of multi-pose face images based on nonlinear dimensionality reduction. *Pattern Recognit.* **37**(2), 325–336 (2004)
7. Wen, Y., He, L., Shi, P.: Face recognition using difference vector plus KPCA. *Digit. Signal Proc.* **22**(1), 140–146 (2012)
8. Li, H.M., Zhou, D.M., Nie, R.C., Li, X., Deng, H.Y.: Face recognition using KPCA and KFDA. *Appl. Mech. Mater.* **380**, 3850–3853 (2013)
9. Kittler, J., Li, Y.P., Matas, J.: Face authentication using client specific fisherfaces. In: *The Statistics of Directions, Shapes and Images*, pp. 63–66 (1999)
10. Wu, X.-j., Josef, K., Yang, J.-y., Kieron, M., Wang, S., Lu, J.: On dimensionality reduction for client specific discriminant analysis with application to face verification. In: Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) *SINOBIOMETRICS 2004*. LNCS, vol. 3338, pp. 305–312. Springer, Heidelberg (2004)
11. Sun, X., Wu, X.J., Sun, J., Montesinos, P.: Hybrid client specific discriminant analysis and its application to face verification. In: Hatzilygeroudis, I., Palade, V. (eds.) *Combinations of Intelligent Methods and Applications*, vol. 23, pp. 137–156. Springer, Heidelberg (2013)
12. Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Principal component analysis of image gradient orientations for face recognition. In: *Proceedings of International Conference on Automatic Face & Gesture Recognition and Workshops*, pp. 553–558 (2011)
13. Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Subspace learning from image gradient orientations. *IEEE PAMI* **34**(12), 2454–2466 (2012)
14. Piella, G.: A general framework for multiresolution image fusion: from pixels to regions. *Inf. Fusion* **4**(4), 259–280 (2003)
15. Wu, T., Wu, X.J., Liu, X., Luo, Q.: New method using feature level image fusion and entropy component analysis for multimodal human face recognition. *Procedia Eng.* **29**, 3991–3995 (2012)

16. Oklahoma State University. IRIS Thermal/Visible Face Database. <http://www.vcipl.okstate.edu/otcbvs/bench/>. Accessed 2014
17. Lee, K.C., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. *IEEE PAMI* **27**(5), 684–698 (2005)

Multimodal Pattern Recognition of Social Signals in
Human-Computer-Interaction

Third IAPR TC3 Workshop, MPRSS 2014, Stockholm,
Sweden, August 24, 2014, Revised Selected Papers

Schwenker, F.; Scherer, S.; Morency, L.-P. (Eds.)

2015, VIII, 145 p. 75 illus., Softcover

ISBN: 978-3-319-14898-4