Discriminative Dictionary Learning with Local Constraints for Face Recognition with Occlusion

Tao Zhang[1]($\boxtimes$), Zhuoqun Yang[1], Yaqi Xu[1], Bin Yang[2], and Wenjing Jia[3]

[1] Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi, China
taozhang@jiangnan.edu.cn
[2] Department of Design, Jiangnan University, Wuxi, China
[3] Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, Australia

Abstract. Facial occlusion, such as sunglasses, mask etc., is one important factor that affects the accuracy of face recognition. Unfortunately, faces with occlusion are quite common in the real world. In recent years, sparse coding becomes a hotspot of dealing with face recognition problem under different illuminations. The basic idea of sparse representation-based classification is a general classification scheme in which the training samples of all classes were taken as the dictionary to represent the query face image, and classified it by evaluating which class leads to the minimal reconstruction error of it. However, how to balance the shared part and class-specific part in the learned dictionary is not a trivial task. In this paper we make two contributions: (i) we present a new occlusion detection method by introducing sparse representation-based classification model; (ii) we propose a new sparse model which incorporates the representation-constrained term and the coefficients incoherence term. Experiments on benchmark face databases demonstrate the effectiveness and robustness of our method, which outperforms state-of-the-art methods.

Keywords: Face recognition · Occlusion detection · Sparse representation

1 Introduction

Automatic face recognition with occlusion has been a hot topic in the area of computer vision and pattern recognition due to the increasing need for real-world applications. Various approaches have been proposed, including subspace mapping algorithm [1–3], feature extraction [4–6] and kernel models [7–10]. However, all these method used the reconstructed images for classification, considering the fact that the reconstructed images might remove some useful information and introduce some redundant information, therefore, whether the reconstructed images are suitable for occluded face recognition needs study.

To avoid over-fitting, a regularization term is generally imposed upon the LR model. There are two widely-used constraints: the L2-normand the L1-norm. While

L1-norm regularizer is a traditional model for sparse representation. Recently proposed sparse representation classification (SRC) algorithms have obtained promising performance on image classification and image super-resolution tasks, etc. [11–17]. Since sparse representation classification approaches obtained competitive performance in face recognition area [18], it has attracted researchers' attention in image classification. Sparse representation classification model has shown the robust ability to deal with sparse random pixel corruption and block occlusion. Nevertheless, a discriminative dictionary for both sparse data representation and classification is still a difficult learning problem.

Some recent work, on the other hand, began to investigate the role of sparsity in face recognition [10, 19–21]. Liu et al. [21] introduce the dual form of dictionary learning and provided some theoretical proof. They argued that it is L1 constraint together with L2 that makes SRC more effective. To overcome high residual error and instability, Zeng et al. [20] analyzed the main principle of SRC and believed that the collaborative representation strategy can enhance the interpretability. They presented a collaborative representation classifier (CRC) based on Ridge regression. CRC can be thus considered as a special case of the SRC algorithm, however, does not provide a mechanism for noise removal, so it is not robust to detect occluded face.

Face recognition with occlusion algorithm need to be robust against arbitrary occlusions. Despite the emergence of a large number of face detection algorithms, most of the existing algorithm are focused on partial occlusion. In the early years, Wen et al. proposed a face occlusion detection method using Gabor filter function [22]. In order to use the temporal feature, some algorithm based on spatial-temporal information were used to find frontal faces [23–25]. Some skin color-based detection approaches are used to find faces [26, 27]. Interestingly, other researchers focus on the popular "recognition by parts" scheme, whose main aim is to predict the head position by determining the appropriate human body model of other parts, such as a probabilistic weighted retrieval [28], locally salient ICA information [29], a new learning algorithm PRSOM (Probabilistic Self-Organizing Model) [30], discriminative and robust subspace model [31], dynamic similarity function [32], local non-negative matrix factorization [33], holistic PCA model [34], SVM model [35], Markov Random Field method [36], optimal feature selection model [37], confidence weighting model [38], embedded hidden Markov method [39]. These approaches can cope with the partial occlusion cases through extracting other features of the non-occluded parts. However, for the severe occluded cases, the performance of these algorithm will be reduced. Other head detection approaches are also a hot research area in surveillance application, such as color model-based approaches, contour-based approaches and matching-based approaches, and these can also be regarded as a different application of face detection. Color model-based approaches [40, 41] determine face regions by extracting hair and face color information. The computational complexity of these algorithm is very low, but when the region of head is severely covered, these methods will not work. Matching-based approaches [42, 43] detect head by comparing the similarity of training template and the current area. Contour-based approaches [44, 45] adequately use complex geometric curves to depict face contour feature. This kind of algorithm can deal with severe occlusion problem, but the computational cost is high. Also, it is hard

to work in a low-resolution image. In this paper, we try to detect head regions by a novel and robust algorithm.

It is worth mentioning that the convolutional neural networks methods, such as DeepID [46] and WebFace [47], are proved to handle face recognition with various variations. However, they exploit lots of information with very complex image variations to assist the training process, so the main drawbacks contribute to the high computational complexity and complex parameter tuning. Thus, they are not suitable to treat with undersampled face recognition, especially for face occlusion problem.

Note that the residual image (a difference between the raw and reconstructed image) contains most of the occluded information as shown in Fig. 1. It is obviously that the occluded region in the residual image is very intuitive. In this paper, we propose a discriminative sparse coding model to deal with recognition task. With the same setting as [20, 21], we consider the scenario that only one non-occluded training sample is available for each subject of interest, which is close to many real application scenarios such as security, video surveillance et al. Compared with some related methods, the advantages of our proposed model are highlighted as follows:
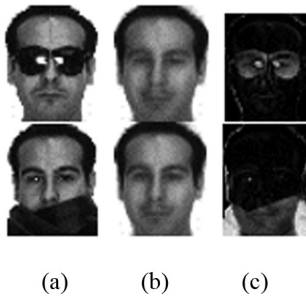

(a)   (b)   (c)

Fig. 1. Examples of raw images, reconstructed images and residual images. (a) The raw images. (b) The reconstructed images. (c) The residual images.

- An occlusion variation dictionary is learned for representing the possible occlusion variations between the training and testing samples. Different from SRC, our proposed model extracts the features from the covariance of occlusion variations based on deep networks to construct the occlusion variation dictionary. Experimental results show that the learned dictionary can efficiently represent the possible occlusion variations.
- Proposing novel measurements strategy to improve sparsity, robustness and discriminative ability. Different from traditional sparse representation which task is to minimize the reconstruction error only, in this proposed model, two terms, the similarity constrain term and the coefficient incoherence term are introduced to ensure that the learned dictionary has the powerful discriminative ability.

The reminder of the paper is organized as follows: Sect. 2 presents some related works. Section 3 describes our proposed model. Section 4 shows experiment results and Sect. 5 draws conclusions.

2   Related Work

In SRC [15], Wright et al. proposed a general classification scheme in which the training samples of all classes were taken as the dictionary to represent the query face image, and classified it by evaluating which class leads to the minimal reconstruction error of it. Since SRC scheme has shown impressive performance in FR, how to design a framework and algorithm to learn a discriminative dictionary for both sparse data representation and classification are attracting a great deal of attention.

Wright et al. [15] proposed the sparse representation based classification (SRC) scheme for robust face recognition (FR). Given K classes of subjects, and let $D = [A_1, A_2, \cdots, A_K]$ be the dictionary formed by the set of training samples, where A_i is the subset of training samples from class i. Let y be a test sample. The algorithm of SRC is summarized as follows.

(a) Normalize each training sample in A_i, $i = 1, 2, \cdots, K$.

(b) Solve l_1-minimization problem: $\hat{x} = \mathrm{argmin}_x \left\{ \|y - Dx\|_2^2 + \gamma \|x\|_1 \right\}$, where γ is scalar constant.

(c) Label a test sample y via: $\mathrm{Label}(y) = \mathrm{argmin}_i\{e_i\}$, where $e_i = \|y - A_i\hat{\alpha}^i\|_2^2$, $\hat{x} = [\hat{\alpha}^1, \hat{\alpha}^2, \cdots, \hat{\alpha}^K]^T$ and $\hat{\alpha}^i$ is the coefficient vector associated with class i.

Obviously, the underlying assumption of this scheme is that a test sample can be represented by a weighted linear combination of just those training samples belonging to the same class. Its impressive performance reported in [15] showed that sparse representation is naturally discriminative.

According to predefined relationship between dictionary atoms and class labels, we can divide current supervised dictionary learning into three categories: shared dictionary learning, class-specific dictionary learning and hybrid dictionary learning. In shared dictionary learning, a dictionary shared by all classes is learned, meanwhile the discriminative power of the representation coefficients is also mined [48, 49]. In generally, in this scheme, a shared dictionary and a classifier over the representation coefficients are together learned. However, there is no relationship between the dictionary atoms and the class labels, and thus no class-specific representation residuals are introduced to perform classification task.

In the class-specific dictionary learning, a dictionary whose atoms are predefined to correspond to subject class labels is learned and thus the class-specific reconstruction error could be used to perform classification [50, 51].

The hybrid dictionary models which combines shared dictionary atoms and class-specific dictionary atoms have been proposed [52, 53]. However, the shared dictionary atoms could encourage learned hybrid dictionary compact to some extent, how to balance the shared part and class-specific part in the hybrid dictionary is not a trivial task.

3   Proposed Model

Machine learning algorithm are often used in computer vision due to their ability to leverage large amounts of training data to improve performance. For face recognition task, the deeply learned features are required to be generalized enough for identifying new unseen classes without label prediction. In order to enhance the discriminative power of the deeply learned features, wen etc. propose new supervision signal, called center loss. Specifically, the center loss simultaneously learns a center for deep features of each class and penalizes the distances between the deep features and their corresponding class centers [54]. It is encouraging to see that their CNNs achieve the state-of-the-art accuracy. Therefore, in this paper, we adapt this deep network model to extract occluded face feature.

Recently proposed sparse representation classification (SRC) algorithms have obtained promising performance on image classification and image super-resolution tasks, etc.. For a detailed introduction to sparse representation, which can be found in [15–17]. Since sparse representation classification approaches obtained competitive performance in face recognition area [18], it has attracted researchers' attention in image classification. Nevertheless, a discriminative dictionary for both sparse data representation and classification is still a pending problem.

To address these difficulties, we propose a modified sparse model for this purpose. Different from traditional sparse representation which task is to minimize the reconstruction error only, in this proposed model, two terms, the representation-constrained term and the coefficient incoherence term are introduced to ensure that the learned dictionary has the powerful discriminative ability.

3.1   Proposed Sparse Classification Model

The representation-constrained term is used to project each descriptor into its local coordinate system which captures the correlations between similar descriptors by sharing dictionary. On the other hand, the coefficients incoherence term ensures that samples from different classes can be built by independent dictionary.

In the class-specific dictionary learning, each dictionary atom $D = [D_1, D_2, L, D_K]$ indicate class label, where D_i is the sub-dictionary of class i. In our experimental settings, corresponding training deep feature samples set $\{a_{ij} | i = 1, 2, L, \ k; j = 1, 2, L, N.\}$, where a_{ij} indicates the j-th sample of class i, K is the number of classes, and N denotes the number of training samples in each class. Let $A = [A_1, A_2, L, A_i] \in R^{n \times N}$, where $A_i = [a_{i1}, a_{i2}, L, a_{iN}]$, n is the deep feature dimension. Our purpose is to contain the classification error as a term in the objective function for dictionary learning for purpose of making the lexicon be optimal for classification. Sparse code Z can be directly utilized as a characteristic for classification. Let $Z = [Z_1, Z_2, L, Z_i]$, denote the learned dictionary by $D = [d_1, d_2, L, d_k] \in R^{n \times k}$ ($k > n$ and $k < N$). We propose the following novel sparse model:

$$\langle D, W, Z \rangle = \arg\min\{\|A - DZ\|_F^2 + \lambda_1\|Z\|_1 + \lambda_2\|Z - m\|_F^2 + \gamma_1\|WZ - B\|_F^2 + \gamma_2\|W\|_F^2\} \quad (1)$$
$$s.t.\|d_c\|_2 \le 1, \forall c \in \{1, 2, \cdots, k\}$$

Where $m = [m_1, m_2, L, m_i] \in R^{k \times N}$, m_i denotes mean vector Z_i of in class i, $\|WZ - B\|_F^2$ denotes the classification error, $B = [0, 0, L, b_N] \in R^{m \times N}$ are the class labels of input feature. $b_i = [0, 0, L1L, 0]^T \in R^m$ is a label vector. $W \in R^{m \times k}$ denotes the matrix of classifier parameters, and λ_1, λ_2, γ_1 and γ_2 are the scale adjustment parameters.

In our proposed model, the representation-constrained term $\|WZ - B\|_F^2$ and coefficients incoherence term $\|W\|_F^2$ are introduced in Eq. 1.

3.2   Optimization Process

Obviously, the function of Eq. 1 is not co-convex to (D; W; Z), when the other two variables are fixed, it is convex for each of D, W, and Z. So, we can optimize D, W and Z respectively to make the Eq. 1 be three sub-problems, i.e., update the Z when D and W are fixed, update D when W and Z are fixed, and update W when D and Z are fixed. Let us explain the details.

Updating Z: When D and W are constant value, it can be regarded as sparse coding problem for solving Z. When Z_i is updated, all $Z_j(j \ne i)$ are also fixed. Thus, for each Z_i, the objective function in Eq. 2 can be replaced by:

$$\langle Z \rangle = \arg\min\{\|A - DZ\|_F^2 + \lambda_1\|Z\|_1 + \lambda_2\|Z - m\|_F^2 + \gamma_1\|WZ - B\|_F^2 \quad (2)$$

By solving Eq. 2, we have:

$$Z_i = \{D^T D + (\lambda_1 + \lambda_2)I + \gamma_1 W^T W\}^{-1}(D^T A_i + \lambda_2 m_i + \gamma_1 W^T b_i) \quad (3)$$

Updating D: When Z and W are constant value, Eq. 1 can be regarded as solving $D = [D_1, D_2, L, D_K]$ sparse coding problem. When D_i is updated, all $D_j(j \ne i)$ are also fixed. Thus, Eq. 1 can be replaced by:

$$\langle D \rangle = \arg\min\left\{\|A - DZ\|_F^2\right\}, s.t.\|d_c\|_2 = 1, \forall c \in \{1, 2, \cdots, k\}. \quad (4)$$

The above problem in Eq. 4 can be solved effectively by the Lagrange dual method. Updating W: When D and Z are fixed, Eq. 1 can be replaced by:

$$\langle W \rangle = \arg\min\left\{\gamma_1\|WZ - B\|_F^2 + \gamma_2\|W\|_F^2\right\} \quad (5)$$

Obviously, Eq. 5 can be solved using the least square method. Thus we can get the following solution:

$$W_i = b_i Z_i^T (Z_i Z_i^T + \frac{\gamma_2}{\gamma_1} I)^{-1} \qquad (6)$$

Therefore, according to the above equations, the optimized values of all parameters in Eq. 1 can be got.

4 Experimental Results

To evaluate the proposed model, we compare it with the state-of-the-art methods for face recognition with occlusion. Sparse representation-based approaches: sparse representation based face classification (SRC) [15], robust sparse coding (RSC) [16], correntropy-based sparse representation (CESR) [17] and extended sparse representation-based classification (ESRC) [19].

4.1 Results on the AR Database with Real-World Occlusion

We evaluate the performance of our proposed model in dealing with real occlusion using the AR face database [18], which consists of 4000 frontal-face images from 126 subjects (70 men and 56 women). Each subject has two separate sessions and 13 images for each session. These images are taken under different variations, including various facial expressions, illumination variations and occlusions (such as sunglasses and scarf).

In the first group, all the rest samples with occlusion from the 80 subjects in session 1 and session 2 are used as testing set, which is divided into three subsets in session 1 or session 2, respectively. (sunglasses subset and scarf subset). The final results with the existing methods are shown in Table 1. Based on the results, we can draw the following conclusions:

- In the same session (session 1 or session 2), the face recognition performances of these existing methods are much better on sunglasses subset than on scarf subset. That is because the sunglasses occlude roughly 20% of the image, while the scarf occlude roughly 40% of the image.

Table 1. Recognition rates for different methods on the AR database.

Method	Session1:		Session2:	
	Sunglass	Scarf	Sunglass	Scarf
SRC	40.83	25.33	22.92	13.33
RSC	71.67	59.58	52.92	35.83
CESR	86.24	83.75	65.42	57.50
ESRC	63.75	16.25	41.67	10.00
Ours	**89.68**	**86.48**	**70.16**	**64.29**

- The performances of some sparse representation-based face recognition approaches such as SRC, RSC and CESR, are bad to address occluded problem for face recognition. For example, the recognition rates of SRC, RSC and CESR are only 13.33%, 35.83% and 10.00% on scarf subset from session 2, respectively. Lack of sufficient training samples to represent the test sample is the main reason.
- Our proposed algorithm obtains significantly higher recognition rates than most of these compared methods, the recognition results is 89.68%, 86.48%, 70.16% and 64.29%, respectively. It indicates that our proposed model is more robust to occlusion variation than these existing methods.

4.2   Results on the CAS-PEAL Database with Real Occlusion

The CAS-PEAL [18] face database consists of 9594 images of 1040 subjects (595 males and 445 females), which are obtained in different variations, including pose, expression, accessory, lighting, time and distance. Each subject is captured under at least two kinds of these variations. Thus, here we take a subset with normal and different accessory variations, which contains 3038 images of 434 subjects and 7 images for each subject. So each subject has 1 neutral image, 3 images with glasses/sunglasses, and 3 images with hats. Finally, all the images are cropped to 120×100 size.

In each recognition process, we select 350 subjects of interest for training and testing, and the remaining 84 subjects are considered as external data for learning the occlusion variation dictionary. In training process, we choose only the neutral image of each of the 350 subjects. While in testing process, we consider three separate test subsets of the 350 subjects. The first test subset constitute with 3 images of each subject wearing glasses/ sunglasses (glass subset). The second test subset constitute with 3 images of each subject wearing hats (hat subset). The third test subset constitute with 6 images of the subject from glass subset and hat subset.

The final recognition rates for all the methods on CAS-PEAL database are given in Table 2. Based on the results, we can get the following conclusions:

Table 2. Recognition rates for different methods On the CAS-PEAL databas.

Method	Sunglass	Hat
SRC	79.05	42.24
RSC	89.62	80.10
CESR	89.33	29.43
ESRC	89.71	63.52
Ours	**92.68**	**87.56**

- The hats occlusion subset is more challenging than the glasses/sunglasses occlusion subset on the database.
- SRC-based face recognition approaches, such as ESRC, RSC and CESR improve the face recognition performance of ordinary SRC for the glasses/sunglasses occlusion, but their performance is not good to the hats occlusion. For example, the

recognition rate of CESR can reach up to 89.33% for the glasses/sunglasses occlusion, but it degrades seriously for the hats occlusion, only 29.43%.

- Our proposed algorithm achieves the best results for all subsets. That is because the representation-constrained and the representation coefficients are more discriminative, and the corresponding classification method is effective to reveal such information. It indicates that proposed model can well learn the occlusion variation, and is also effective to detect occlusion cases.

5 Conclusion

In this paper, we present a novel sparse representation-based classification model and apply the alternating direction method of multipliers to solve it. Different from traditional sparse representation which task is to minimize the reconstruction error only, in this proposed model, two terms, the representation-constrained term and the coefficient incoherence term are introduced to ensure that the learned dictionary has the powerful discriminative ability. Proposed model takes advantage of the structural characteristics of noise and provides a unified framework for integrating error detection and error support into one sparse model. Extensive experiments demonstrate that the proposed model is robust to occlusions.

References

1. Tai, Y., Yang, J., Luo, L., Zhang, F.L., Qian, J.J.: Learning discriminative singular value decomposition representation for face recognition. Pattern Recognit. **50**(C), 1–16 (2016)
2. Zhang, G., Zou, W., Zhang, X., et al.: Singular value decomposition based virtual representation for face recognition. Multim. Tools Appl. **5**(11), 1–16 (2017)
3. Hu, C., Lu, X., Ye, M., et al.: Singular value decomposition and local near neighbors for face recognition under varying illumination. Pattern Recogn. **64**, 60–83 (2017)
4. Lei, Z., Pietikainen, M., Li, S.Z.: Learning discriminant face descriptor. IEEE Trans. Pattern Anal. Mach. Intell. **36**(2), 289–302 (2014)
5. Lei, Z., Yi, D., Li, S.Z.: Learning stacked image descriptor for face recognition. IEEE Trans. Circuits Syst. Video Technol. **26**(9), 1685–1696 (2016)
6. Zhang, T., Yang, Z., Jia, W., et al.: Fast and robust head detection with arbitrary pose and occlusion. Multimed. Tools Appl. **74**(21), 9365–9385 (2015)
7. Wang, D., Lu, H., Yang, M.H.: Kernel collaborative face recognition. Pattern Recogn. **48**(10), 3025–3037 (2015)
8. Wang, M., Hu, Z., Sun, Z., et al.: Kernel collaboration representation-based manifold regularized model for unconstrained face recognition. Signal Image Video Process. (C), **12** (5), 1–8 (2018)

9. Hua, J., Wang, H., Ren, M., et al.: Collaborative representation analysis methods for feature extraction. Neural Comput. Appl. **28**(S1), 1–7 (2016)
10. Yang, J., Luo, L., Qian, J., Tai, Y., Zhang, F., Xu, Y.: Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. IEEE Trans. Pattern Anal. Mach. Intell. **39**(1), 156–171 (2017)
11. Fan, Z., Ni, M., Zhu, Q., Sun, C.: L0-norm sparse representation based on modified genetic algorithm for face recognition. J. Vis. Commun. Image Represent. **28**, 15–20 (2015)
12. Han, B., Wu, D.: Image representation by compressive sensing for visual sensor networks. J. Vis. Commun. Image Represent. **21**, 325–333 (2010)
13. Jorge, S., Javier, R.: Exponential family fisher vector for image classification. Pattern Recognit. **59**, 26–32 (2015)
14. Cheng, H., Liu, Z., Yang, L., Chen, X.: Sparse representations and learning in visual recognition: theory and applications. Signal Process. **93**, 1408–1425 (2013)
15. Wright, J., Yang, A.Y., Ganesh, A., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern Mach. Intel. **31**, 210–227 (2009)
16. Xu, Y., Zhang, B., Zhong, Z.F.: Multiple representations and sparse representations for image classification. Pattern Recognit. **68**, 9–14 (2015)
17. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. IEEE Trans. Image Process. **19**, 2861–2873 (2010)
18. Zhang, Z., Xu, Y., Yang, X., Li, X.: A survey of sparse representations: algorithm and applications. IEEE Access **3**, 490–530 (2015)
19. Lai, J., Jiang, X.: Class-wise sparse and collaborative patch representation for face recognition. IEEE Trans. Image Process. **25**(7), 3261–3272 (2016)
20. Liu, B.D., Shen, B., Gui, L., Wang, Y.X., Li, X., Yan, F., et al.: Face recognition using class specific dictionary learning for sparse representation and collaborative representation. Neurocomputing **204**, 198–210 (2016)
21. Zeng, S., Gou, J., Deng, L.: An antinoise sparse representation method for robust face recognition via joint l 1, and l 2, regularization. Expert Syst. Appl. **82**, 1–9 (2017)
22. Wen, C., Chiu, S., Tseng, Y., Lu, C.: The mask detection technology for occluded face analysis in the surveillance system J. Forensic Sci. **3**, 1–9 (2005)
23. Yoon, S.M., Kee, S.C.: Detection of partially occluded face using support vector machines. In: Proceedings of IAPR Conference on Machine Vision Applications, pp. 546–549 (2002)
24. Kim, J., Sung, Y., Yoon, S.M., Park, B.G.: A new video surveillance system employing occluded face detection. In: Ali, M., Esposito, F. (eds.) IEA/AIE 2005, vol. 3533, pp. 65–68. Springer, Heidelberg (2005). https://doi.org/10.1007/11504894_10
25. Choi, I., Kim, D.: Facial fraud discrimination using detection and classification. In: Bebis, G., et al. (eds.) ISVC 2010. LNCS, vol. 6455, pp. 199–208. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17277-9_21
26. Dong, W., Soh, Y.: Image-based fraud detection in automatic teller machine. Int. J. Comput. Sci. Network Secur. **11**, 13–18 (2006)
27. Kakumanu, P., Makrogiannis, S., Bourbakis, N.: A survey of skin-color modeling and detection methods. Pattern Recognit. **3**, 1106–1122 (2007)
28. Zhang, Y., Martinez, A.M.: A weighted probabilistic approach to face recognition from multiple images and video sequences. Image Vis. Comput. **6**, 626–638 (2006)
29. Kim, J., Choi, J., Yi, J., Turk, M.: Effective representation using ICA for face recognition robust to local distortion and partial occlusion. IEEE Trans. Pattern Anal. Mach. Intell. **12**, 1977–1981 (2005)
30. Tan, X., Chen, S., Zhou, H. Zhang, F.: Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble. IEEE Trans. Neural Networks **4**, 875–886 (2005)

31. Fidler, S., Skocaj, D., Leonardis, A.: Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. IEEE PAMI **3**, 337–350 (2006)
32. Liu, Q., Yan, W., Lu, H., Ma, S.: Occlusion robust face recognition with dynamic similarity features. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006), vol. 3, pp. 544–547 (2006)
33. Oh, H.J., Lee, K.M., Lee, S.U.: Occlusion invariant face recognition using selective local non-negative matrix factorization basis images. Image Vis. Comput. **11**, 1515–1523 (2008)
34. Rama, A., Tarres, F., Goldmann, L., Sikora, T.: More robust face recognition by considering occlusion information. In: Proceedings of the Eighth IEEE International Conference on Automatic Face Gesture Recognition (FG 2008), pp. 1–6 (2008)
35. Jia, H., Martinez, A.: Support vector machines in face recognition with occlusions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009), pp. 136–141 (2009)
36. Zhou, Z., Wagner, A., Mobahi, H., Wright, J., Ma, Y.: Face recognition with contiguous occlusion using Markov random fields. In: Proceedings of the IEEE 12th International Conference on Computer Vision (ICCV 2009), pp. 1050–1057 (2009)
37. Lin, J., Ming, J., Crookes, D.: Robust face recognition with partial occlusion, illumination variation and limited training data by optimal feature selection. IET Comput. Vis. **1**, 23–32 (2011)
38. Struc, V., Dobrisek, S., Pavesic, N.: Confidence weighted subspace projection techniques for robust face recognition in the presence of partial occlusions. In: Proceedings of the 20th International Conference on Pattern Recognition (ICPR 2010), pp. 1334–1338 (2010)
39. Huang, S.-M., Yang, J.-F.: Robust face recognition under different facial expressions, illumination variations and partial occlusions. In: Proceedings of the 17th International Conference on Advances in Multimedia Modeling (MMM 2011), vol. 2, pp. 326–336 (2011)
40. Yang, T., Pan, Q., Li, J., Cheng, Y.M.: Real-time head tracking system with an active camera. In: Proceedings of the 5th World Congress on Intelligent Control and Automation, pp. 1910–1914 (2006)
41. Chen, M.L., Kee, S.: Head tracking with shape modeling and detection. In: Proceedings of the Second Canadian Conference on Computer and Robot Vision (2006)
42. Huang, W.M., Luo, R.J.: Real time head tracking and face and eyes detection. In: Proceedings of IEEE TENCON, pp. 507–510 (2002)
43. Yao, Z.R., Li, H.B.: Tracking a detected face with dynamic programming. Image Vis. Comput. **6**, 573–580 (2006)
44. Finlayson, G.D., Hordley, S.D., Lu, C., Drew, M.S.: On the removal of shadows from images. IEEE PAMI **25**(10), 59–68 (2006)
45. Zou, W., Li, Y., Yuan, K., Xu, D.: Real-time elliptical head contour detection under arbitrary pose and wide distance range. J. Vis. Commun. Image R. **20**, 217–228 (2009)
46. Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. Computer Vision and Pattern Recognition, pp. 1891–1898. IEEE (2014)
47. Wang, D., Otto, C., Jain, A.K.: Face search at scale: 80 million gallery. Comput. Sci. 1–14 (2015)
48. Bach, F., Mairal, J., Ponce, J.: Task-driven dictionary learning. IEEE Trans. Pattern Anal. Mach. Intell. **34**(4), 791–804 (2012)
49. Jiang, Z., Lin, Z., Davis, L.S.: Label consistent k-svd: learning a discriminative dictionary for recognition. IEEE Trans. Pattern Anal. Mach. Intell. **35**(11), 2651–2664 (2013)
50. Yang, M., Zhang, L., Feng, X., Zhang, D.: Fisher discrimination dictionary learning for sparse representation. In: Proceedings, vol. 24(4), pp. 543–550 (2011)

51. Yang, M., Zhang, L., Feng, X., Zhang, D.: Sparse representation based fisher discrimination dictionary learning for image classification. Int. J. Comput. Vis. **109**(3), 209–232 (2014)
52. Kong, S., Wang, D.: A dictionary learning approach for classification: separating the particularity and the commonality. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7572, pp. 186–199. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33718-5_14
53. Zhou, N., Shen, Y., Peng, J., Fan, J.: Learning inter-related visual dictionary for object recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 157, pp. 3490–3497. IEEE Computer Society (2012)
54. Wen, Y.D., Zhang, K.P., Li, Z.F., et al.: A Discriminative Feature Learning Approach for Deep Face Recognition. **47**(9), pp. 499–515 (2016)