SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

January 2018, Vol. 61 012101:1–012101:13
doi: 10.1007/s11432-016-9009-6

Robust sparse representation based face recognition in an adaptive weighted spatial pyramid structure

Xiao MA[1*], Fandong ZHANG[1], Yuelong LI[2,3] & Jufu FENG[1]

[1] *School of Electronics Engineering and Computer Science, Peking University, Beijing* 100871*, China;*
[2] *School of Computer Science and Software Engineering, Tianjin Polytechnic University, Tianjin* 300387*, China;*
[3] *Department of Computer Science, University of York, York* YO10 5GH*, UK*

Received 13 September 2016/Revised 15 November 2016/Accepted 10 January 2017/Published online 22 May 2017

Abstract The sparse representation based classification methods has achieved significant performance in recent years. To fully exploit both the holistic and locality information of face samples, a series of sparse representation based methods in spatial pyramid structure have been proposed. However, there are still some limitations for these sparse representation methods in spatial pyramid structure. Firstly, all the spatial patches in these methods are directly aggregated with same weights, ignoring the differences of patches' reliability. Secondly, all these methods are not quite robust to poses, expression and misalignment variations, especially in under-sampled cases. In this paper, a novel method named robust sparse representation based classification in an adaptive weighted spatial pyramid structure (RSRC-ASP) is proposed. RSRC-ASP builds a spatial pyramid structure for sparse representation based classification with a self-adaptive weighting strategy for residuals' aggregation. In addition, three strategies, local-neighbourhood representation, local intra-class Bayesian residual criterion, and local auxiliary dictionary, are exploited to enhance the robustness of RSRC-ASP. Experiments on various data sets show that RSRC-ASP outperforms the classical sparse representation based classification methods especially for under-sampled face recognition problems.

Keywords face recognition, sparse representation, self-adaptive weighted aggregating, spatial pyramid structure, local robust strategies

Citation Ma X, Zhang F D, Li Y L, et al. Robust sparse representation based face recognition in an adaptive weighted spatial pyramid structure. Sci China Inf Sci, 2018, 61(1): 012101, doi: 10.1007/s11432-016-9009-6

1 Introduction

Face recognition has constantly been an important research subject in the area of computer vision and pattern recognition. Research activities on the subspace' methods, which are based on the hypothesis that all the face images are contained in a linear lower dimensional manifold [1, 2], have increased significantly over the last two decades. Inspired by the sparse coding methods [3, 4], Wright et al. [5] proposed a classical subspace method named sparse representation based classification (SRC), which codes a query sample as sparse linear combination of all training samples and then classifies it by evaluating which class has the minimal coding residual. When the training dictionary is overcomplete, the SRC method can achieve much higher recognition accuracy comparing to other subspace methods. In addition, SRC is also proven robust to occlusion and corruption.

Inspired by SRC, a series of sparse representation based pattern recognition methods have been proposed. Zhang et al. [6] emphasized the collaborative representation mechanism's importance and proposed

* Corresponding author (email: maxiao2012@pku.edu.cn)

© Science China Press and Springer-Verlag Berlin Heidelberg 2017

a more general 'collaborative representation based classification' (CRC) model. Yang et al. [7] discussed how to choose features in CR mechanism and proposed multi-task joint sparse representation based classification (MTJSRC) fusing multiple local features. Ma et al. [8] discussed the role of dictionaries in CR mechanism and learned low-rank dictionaries to enhance CRC based pattern classification. Focussing on separating intra-class variations components, Refs. [9–12] exploited the auxiliary dictionary in SRC's framework, which is more robust to large intra-class variations caused by illumination, poses and occlusions. Feng et al. [13] combined the advantages of collaborative representation and Bayesian residual criterion and proposed collaborative Bayesian representation (CRB), significantly improving CRC's performance in under-sampled cases. In addition, a series of fast algorithms, such as [14, 15], were also proposed to improve the efficiency. Recently, quite a few research studies [16–19] have been proposed to enhance the sparsity based pattern classification's robustness in variety of challenging practical cases.

While the above-mentioned sparse representation based classification methods are only conducted in a single holistic scale, local facial information of multi-scales is rather ignored. Inspired by the idea of spatial pyramid matching (SPM) [20,21], various studies such as ScSPM [22], LLC [23], LCRC [24], SLF-RKR [25], LGR [26], RLR [27] exploit a spatial pyramid (SP) mechanism in sparse representation based methods. These methods work by partitioning the image into several fine sub-patches to conduct local sparse representations, and then aggregating the local coefficients or residuals in a pyramid structure for classification. The sparse representation based classification methods in spatial pyramid structure incorporates facial samples' local and holistic information. Recognition accuracies are largely improved especially for cases with large occlusions.

However, there are still some limitations for the sparse representation methods in spatial pyramid structure. Firstly, all the spatial patches in these methods are directly aggregated with the same weights. All facial sub-patches contribute equally in the representation stage. While for practical face recognitions, intra-class interferences such as illumination, occlusion and noise, may have different influences on different sub-patches. Sub-patches of low quality could hardly be represented with their corresponding local dictionaries and are less reliable for recognition. The differences between the reliable sub-patches and unreliable sub-patches are rather ignored in the original mechanisms. Secondly, all these above-mentioned methods are not quite robust to interferences caused by poses, expressions and misalignment, especially in under-sampled cases. These drawbacks limit the usages of these methods in practical face recognition applications.

In this paper, a novel method named robust sparse representation based classification in an adaptive weighted spatial pyramid structure (RSRC-ASP) is proposed. There are two main contributions for RSRC-ASP: Firstly, RSRC-ASP exploits a sparsity evaluation based self-adaptive weighting strategy for residuals' aggregation, which fully considers the differences among sub-patches' qualities. In addition, three strategies, local-neighbourhood representation, local intra-class Bayesian residual criterion, local auxiliary dictionary are exploited to enhance the robustness of RSRC-ASP. Experiments on various data sets show that the proposed method outperforms the classic sparse representation based methods by about 2%–15%'s improvement especially in under-sampled cases.

The rest of the paper is organized as follows. In Section 2, we introduce related work. Section 3 presents the framework of our proposed algorithm RSRC-ASP and discusses its advantages. Experiments on various databases are presented in Section 4. Finally, Section 5 concludes our paper.

2 Related work

2.1 SRC

As the discussion in Section 1, SRC consists of two major parts: the collaborative representation model and the residual model. In SRC, query face sample is firstly encoded as sparse linear combination of all training samples and then is classified by evaluating which class has the minimal coding residual. The algorithm of SRC [5] is summarized as

$$\hat{\alpha} = \min_{\alpha} \|y - X\alpha\|_2^2 + \lambda\|\alpha\|_1, \tag{1}$$

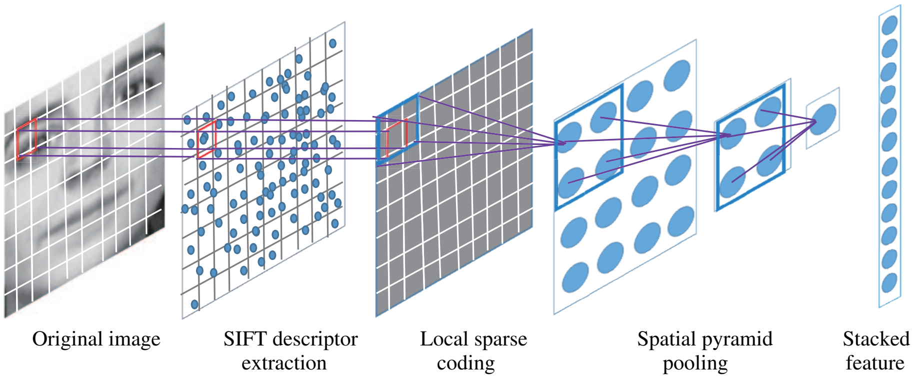

Figure 1 (Color online) The illustration of ScSPM's framework.

$$\text{identity}(y) = \arg\min_i \|y - X_i \hat{\alpha}_i\|_2^2, \tag{2}$$

where, y is the query sample, $X = [X_1, X_2, \ldots, X_C]$ is the training dictionary set of C subjects, and $\alpha = [\alpha_1^{\mathrm{T}}, \alpha_2^{\mathrm{T}}, \ldots, \alpha_C^{\mathrm{T}}]^{\mathrm{T}}$ is the corresponding representation coefficients.

Following Wright's work, a series of sparse representation based methods have been proposed. While most of these studies just use the holistic pixel as feature, the spatial information in different scales is rather ignored. Focussing on this issues, various sptial pyramid based sparse representation methods has been proposed, as shown in the following subsection.

2.2 Local structure based sparse representation methods

Inspired by SPM, Yang et al. [22] proposed the sparse coding based spatial pyramid matching (ScSPM) method. Samples are firstly partitioned into several local patches. Then sparse representations are computed on all sub-patches to learn a shared local dictionary. Afterwards, a spatial pyramid structure is built by maximum pooling of local representations in increasing sub-regions. Finally, the coding coefficients of all sub-regions are directly stacked as the final facial feature for classification. The architecture of ScSPM is illustrated in Figure 1.

ScSPM has made a remarkable success on a range of image classification benchmarks like Caltech-101 and Caltech-256, and is widely used in varieties of practical problems such as face recognition and scene recognition. Recently a few of improvements [23, 28, 29] have been worked out to enhance feature robustness. However, the patch-shared ScSPM does not consider the differences of local features in different locations and scales, ignoring sub-patches' specific information.

Another respective spatial pyramid based sparse representation method is locality constrained representation based classification (LCRC), proposed by Shen et al. [24]. Different from the patch-shared ScSPM, LCRC exploits a patch-specific structure. Images are firstly built a pyramid structure and partitioned into fine sub-regions in each pyramid level. Then, sparse coding is conducted with a specific local dictionary in each sub-region independently. Finally, residuals rather than the coding coefficients are aggregated as the final facial feature. Besides, a simple locality based concentration index (LCI) is exploited to reject the heavily corrupted patches. The final residual criterion of subject i is defined as

$$C_i = \sum_k C_i^{(k)} = \sum_k \|y^{(k)} - X_i^{(k)} \hat{\alpha}_i^{(k)}\|_2^2, \tag{3}$$

where, $C_i^{(k)}$ is the ith subject's representation residual criterion over sub-patch k. $y^{(k)}$, $X_i^{(k)}$ and $\alpha_i^{(k)}$ are its corresponding query sample, ith class's dictionary and coefficients respectively.

Both ScSPM and LCRC incorporate the locality and holistic facial information by conducting local sparse representation in a spatial pyramid local structure. Recently, a series of local structure based sparse representation methods have been proposed: SLF-RKR [25] combines the statical local features and kernel methods in SRC's mechanism. LGR [26] and its improving work RLR [27] exploit local robust representation model in face recognition with single sample per person (SSPP). The patch-specific LCRC

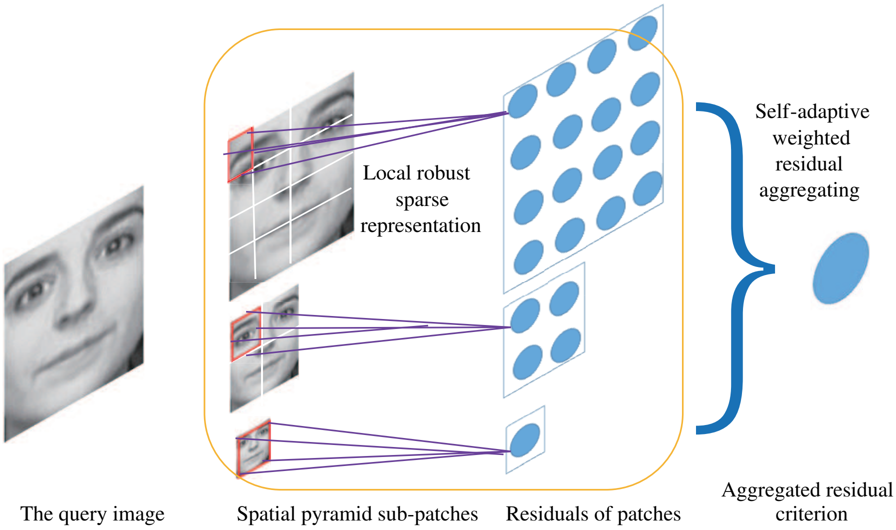

Figure 2 (Color online) The illustration of the RSRC-ASP's framework.

fully considers the differences among features in different sub-regions. It is proven that the aggregated residual criterions are more effective for classification comparing with the aggregated coefficient feature in ScSPM.

However, the residuals of sub-patches are directly aggregated by summation with the same weights in the above-mentioned state-of-art local structure based methods. The patches' residuals (though reserved by LCI in some methods), still contribute equally to the final residual. In addition, there is a pixel-to-pixel correlation between the query sample and training samples in representation stage. The pixel-to-pixel correlation mechanism requires face images being well-aligned, which is quite difficult to guarantee in practical systems. These methods' performance drops sharply when facing large poses variations. Focussing on these problems, a novel method in a spatial pyramid structure is proposed in the following section.

3 RSRC-ASP method

3.1 The framework of RSRC-ASP method

In RSRC-ASP, face samples are firstly built a pyramid structure and partitioned into fine several sub-patches in each pyramid level. Then, local robust sparse representation with three strategies: local-neighbourhood representation, local intra-class Bayesian residual criterion, local auxiliary dictionary, is conducted on each sub-patch. Finally the residuals over local patches are aggregated by an adaptive weighting strategy for classification. The framework of RSRC-ASP's algorithm is illustrated in Figure 2.

3.2 Self-adaptive weighting strategy for residuals' aggregating

For face recognition problems with large interferences such as illumination, occlusion and misalignment, facial sub-patches are influenced in different degrees. In this paper, we define the less corrupted patches as reliable patches, which are of better quality and more effective for classification. In addition, we define the heavily corrupted patches as unreliable patches, which are of bad quality and usually mislead the final classification. The reliable patches should contribute more to the final classifier while the unreliable patches contribute less.

The sparsity of the representation coefficients is exploited in RSRC-ASP as an evaluation index for each patch's reliability. The unreliable patches could hardly be reconstructed by their corresponding local training dictionaries especially in under-sampled cases. For unreliable patches, more dictionary items are

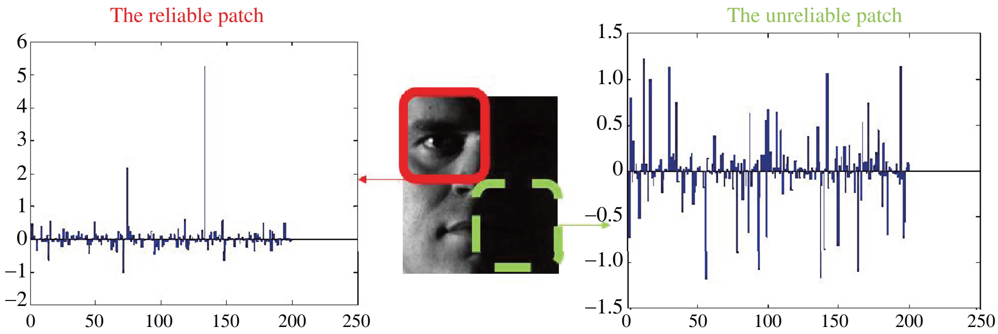

Figure 3 (Color online) The representation coefficients over different patches.

required to fully reconstruct the query sample. Therefore, the representation coefficients of the reliable patches tend to be sparser than those of the unreliable patches as shown in Figure 3.

In RSRC-ASP, the final aggregated residual criterion C_i is defined as

$$C_i(y) = \sum_k w^{(k)} C_i^{(k)} = \sum_k \frac{\|y^{(k)} - X_i^{(k)} \alpha_i^{(k)}\|_2^2}{\|\alpha^{(k)}\|_1}, \tag{4}$$

where, $w^{(k)} = \frac{1}{\|\alpha^{(k)}\|_1}$ is the self-adaptive weight corresponding to sub-patch k, $\alpha^{(k)}$ is the representation coefficient of patch k over all subjects's training dictionary, and $\alpha_i^{(k)}$ is the sub-coefficient vector over the ith subject's training dictionary of patch k.

Then, the final classification is conducted on the aggregated residual criterion, defined as

$$\text{identity}(y) = \arg\min_i C_i(y). \tag{5}$$

Remarkably, there are various inherent differences between CRC-RLS [6] and the proposed RSRC-ASP method, though their residual criterion's forms are very similar. Firstly, CRC-RLS just conduct representation on the holistic face. Secondly, the coefficient on the denominator term corresponds to one specific subject in CRC-RLS. Whereas, the representation coefficient $\alpha^{(k)}$ on the denominator corresponds to sub-patch k's all subjects in RSRC-ASP. In addition, their denominator terms' meanings are also quite different. For CRC-RLS, the denominator term describes one specific subject's contribution in the collaborative representation, which brings some discrimination information for final classification. While for RSRC-ASP, the denominator term mainly describes one specific patch's reliability.

Define y as the query sample vector. The sub-patches $y^{(m)}$ and $y^{(n)}$ can be viewed as projections of y on its corresponding subspaces. When face samples are corrupted by illumination, occlusions or misalignment, the query sample y is transformed to $\tilde{y}$. Sub-patches $\tilde{y}^{(m)}$ and $\tilde{y}^{(n)}$ are projections of the corrupted sample $\tilde{y}$. As illustrated in Figure 4, the projections of $\tilde{y}$ vary differently in subspace m and n. Since sub-patch n is less influenced by interferences, the reliable sub-patch vector $\tilde{y}^{(n)}$ is more closer to the original projection $y^{(n)}$ comparing with unreliable sub-patch $\tilde{y}^{(m)}$. Therefore, higher weights should be added to the more reliable sub-patches in RSRC-ASP, which largely enhances the robustness of under-sampled face recognition.

3.3 Local robust sparse representation for patches

3.3.1 *Strategy A: local-neighbourhood representation*

In RSRC-ASP, we exploit a local-neighbourhood representation (LNR) strategy on the representation stage to improve RSRC-ASP's performance with large poses and misalignment variations. The LNR strategy is based on the assumption that local patches lay in the same linear subspace with their neighbourhood patch samples. For each patch of the training dictionary, we sample a series of patches as the augmentation patches, with a sliding window sampling on its nearby locations. The augmentation

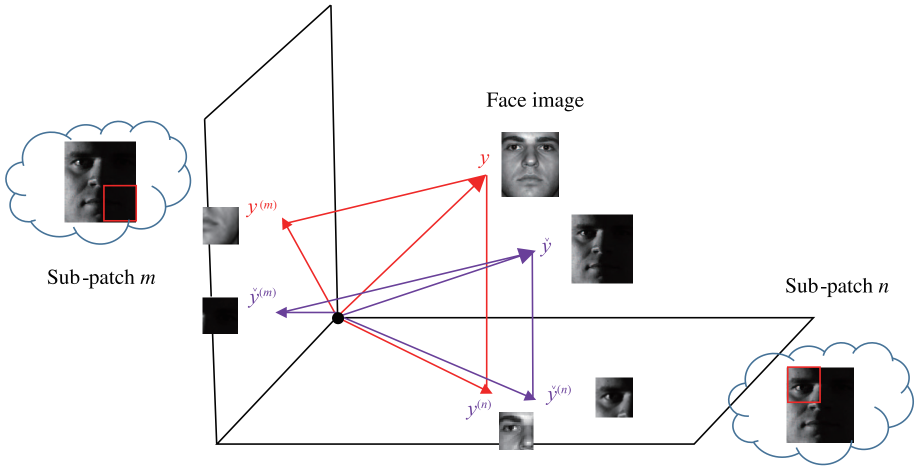

Figure 4 (Color online) The geometric analysis of RSRC-ASP.

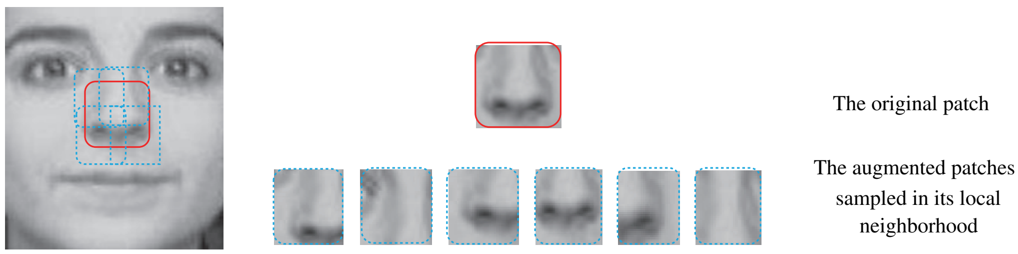

Figure 5 (Color online) The illustration of LNR strategy.

patches are then added to the training dictionary to collaboratively represent its corresponding patches of query image. Figure 5 illustrates the sampling process of LNR. The LNR strategy enhances the representation ability of patches' local training dictionaries, especially for the samples with misalignment and poses variations.

3.3.2 *Strategy B: local intra-class Bayesian residual criterion*

The original sparse representation based classification methods utilizes an Euclidean similarity residual criterion for classification. While for the under-sampled face recognition, where the training dictionary is not over-complete, the intra-class and inter-class variations can not be easily distinguished by the Euclidean similarity criterion. Following the collaborative representation Bayesian (CRB) [13, 30, 31] recognition method's idea, we exploit a better residual criterion, named local intra-class Bayesian residual criterion (LIBRC), in RSRC-ASP.

Assume all the facial intra-class variations lay in a linear subspace. If the query's kth patch $y^{(k)}$ belongs to class i, residual of patch $y^{(k)}$, defined as $r_i^{(k)} = y^{(k)} - X_i^{(k)}\hat{\alpha}_i{}^{(k)}$, is characteristic of the intra-class variation $\Omega_{\mathrm{I}}^{(k)}$, where $\Omega_{\mathrm{I}}^{(k)}$ corresponds to intra-class variations, caused by illumination, poses and expressions. Suppose the intra-class residuals of Ω_{I} are subject to a Gaussian-distribution with covariance Σ_k, the ML local intra-class Bayesian residual criterion of the k patch $P(\Omega_{\mathrm{I}}^{(k)}|r_i^{(k)})$ can be simplified as

$$C_i^{(k)} = {r_i^{(k)}}^{\mathrm{T}} \Sigma_k^{-1} r_i^{(k)}. \tag{6}$$

Σ_k can be estimated by the intra-class intensity differences of generic data's patch samples. Augmentation patches in the local neighborhood can also be added to estimate $\Sigma_{r^{(k)}}$. The final aggregated residual criterion is defined as

$$C_i = \sum_k \frac{C_i^{(k)}}{\|\hat{\alpha}^{(k)}\|_1}. \tag{7}$$

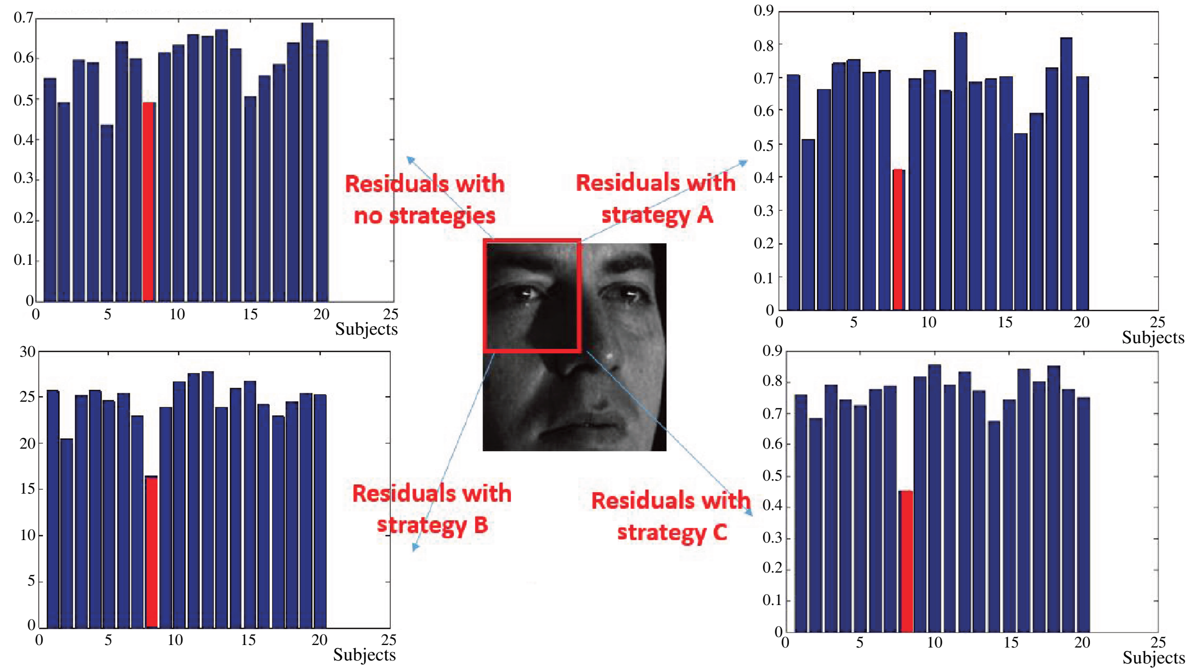

Figure 6 (Color online) Different local residual criterions over an unreliable patch (each of the stripe bars represents the residual criterion of the query image's corresponding subject).

3.3.3 *Strategy C: local auxiliary dictionary*

Inspired by the ideas of ESRC [9], SSRC [10], LAD [11], a series of patch-specific local auxiliary dictionaries, such as LGR [26] and RLR [27], are also exploit to the local sparse representations. In this strategy, local auxiliary dictionaries are trained from corresponding sub-patches' intra-class differences, captured from generic face samples. Auxiliary dictionaries over k patches are defined as $A = [A^{(1)}, A^{(2)}, \ldots, A^{(n_s)}]$, where n_s is the number of sub-patches of a query sample. Components comparing with the training dictionaries X, auxiliary dictionaries are of better ability to reconstruct the intra-class variation components. For the query image's kth sub-patch, the sparse representation process with auxiliary dictionary is summarized as

$$\min_{\alpha^{(k)},\beta^{(k)}} \|y^{(k)} - X^{(k)}\alpha^{(k)} - A^{(k)}\beta^{(k)}\|_2^2 + \lambda\|[\alpha^{(k)};\beta^{(k)}]\|_1, \tag{8}$$

where, $\beta^{(k)}$ is the representation coefficient over auxiliary dictionary $A^{(k)}$. Then the aggregated residual criterion of subject i can be defined as

$$C_i = \sum_k \frac{C_i^{(k)}}{\|\hat{\alpha}^{(k)}\|_1} = \sum_k \frac{\|y^{(k)} - X_i^{(k)}\alpha_i^{(k)} - A^{(k)}\beta^{(k)}\|_2^2}{\|\hat{\alpha}^{(k)}\|_1}. \tag{9}$$

All these three above-mentioned strategies are proved more effective and robust for recognition tasks with large intra-class variations. As shown in Figure 6, the reliability and discrimination of the corrupted patches are dominantly enhanced by these strategies, which leads to a better recognition performance of RSRC-ASP.

4 Experiments and results

4.1 Recognition results

In this section, face recognition experiments are conducted on several benchmark databases: extended Yale B, AR, CMU PIE and LFW. 131465 face images are collected from Internet and external databases as the training set for the local intra-class Bayesian residual criterions and local auxiliary dictionaries.

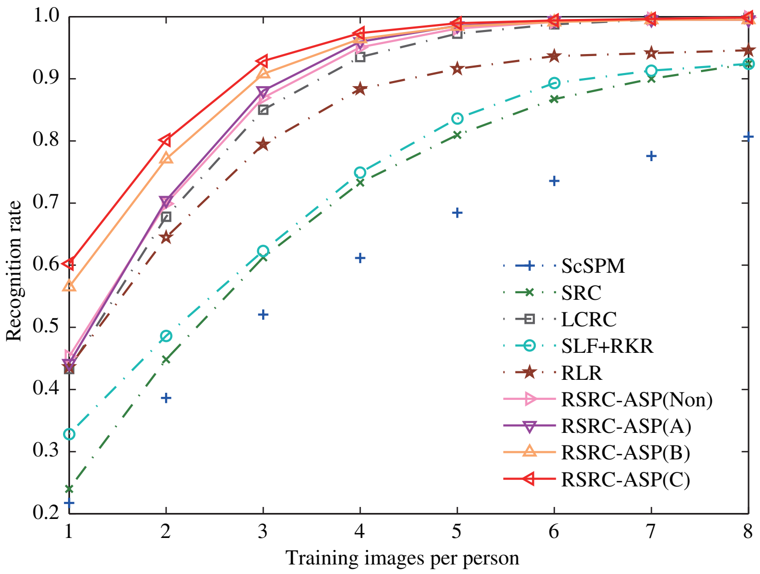

Figure 7 (Color online) The recognition results in extended Yale B.

Table 1 The recognition results (%) in extended Yale B

Gallery	1	2	4	6	8
ScSPM	21.73 ± 2.78	41.31 ± 1.43	61.98 ± 1.21	74.78 ± 0.83	80.69 ± 0.72
SRC	24.00 ± 1.98	46.50 ± 1.21	73.30 ± 0.93	87.00 ± 0.84	92.40 ± 0.43
LCRC	43.30 ± 1.42	72.10 ± 1.14	95.00 ± 0.55	99.30 ± 0.16	99.80 ± 0.03
SLF+RKR	32.83 ± 1.93	48.60 ± 1.62	73.95 ± 0.93	90.56 ± 0.81	92.40 ± 0.44
RLR	43.61 ± 1.87	66.32 ± 1.74	88.52 ± 1.77	93.24 ± 1.66	94.60 ± 1.39
RSRC-ASP(Non)	45.30 ± 0.97	74.10 ± 0.89	96.50 ± 0.32	99.40 ± 0.05	99.90 ± 0.02
RSRC-ASP(A)	44.20 ± 1.59	75.35 ± 1.16	97.20 ± 0.43	**99.40 ± 0.06**	**99.90 ± 0.05**
RSRC-ASP(B)	56.50 ± 0.83	81.50 ± 0.81	97.60 ± 0.38	99.30 ± 0.07	99.50 ± 0.04
RSRC-ASP(C)	**60.24 ± 1.23**	**85.67 ± 1.35**	**98.34 ± 0.55**	99.30 ± 0.18	99.90 ± 0.02

Comparisons are made in the experiments with several sparse representation based methods: ScSPM [22], SRC [5], LCRC [24], SLF-RKR [25], RLR [27] and the proposed RSRC-ASP methods (with different local robust strategies, defined as RSRC-ASP (A)/(B)/(C)/(Non)). All the methods share the same pyramid structure, $\{1, 2, 3, 6\}$. Hyper-parameters are chosen by cross-validation.

4.1.1 *Extended Yale B*

The extended Yale B database [32] consists of 38 individuals' frontal face images with laboratory-controlled illumination conditions. We randomly choose 1–8 face samples per subject as the training dictionary in under-sampled cases, and use all the others for testing. Recognition results are shown in Figure 7 and Table 1. The proposed RSRC-ASP methods with robust strategies outperform the classic sparse representation based methods such as SRC, ScSPM or LCRC, by 2%–17% in nearly all training set sizes.

4.1.2 *AR*

The AR database [33] contains 100 individuals' color images (about 50 images per person) with variations of illumination, expression and occlusion. Recognition results are shown in Figure 8 and Table 2. Different from the experiments in [25, 27], we use the whole AR database (including images with occlusions) for recognition. The proposed RSRC-ASP methods outperform the others by 2%–12%.

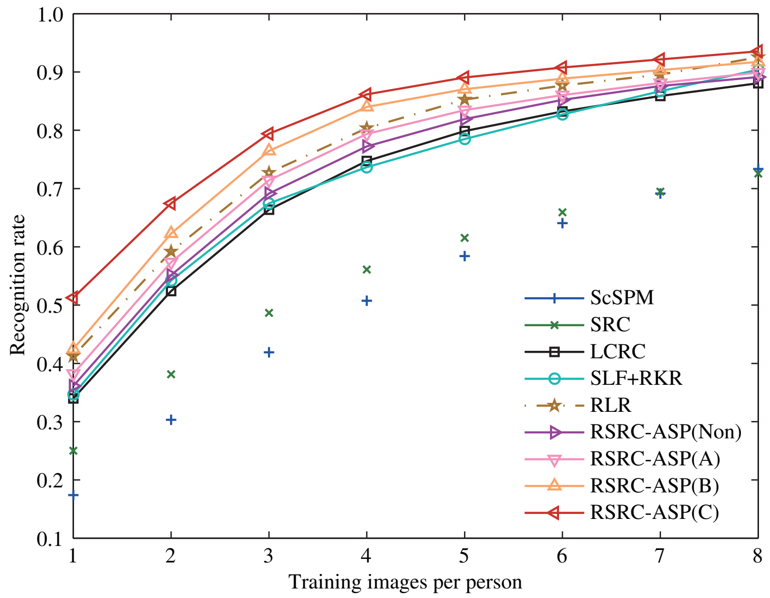

Figure 8 (Color online) The recognition results in AR.

Table 2 The recognition results (%) in AR

Gallery	1	2	4	6	8
ScSPM	17.40 ± 2.83	31.56 ± 1.42	52.13 ± 0.94	64.97 ± 0.97	73.37 ± 0.82
SRC	25.01 ± 2.31	39.64 ± 1.38	56.59 ± 1.21	66.08 ± 0.84	72.56 ± 0.75
LCRC	34.00 ± 1.46	55.00 ± 1.55	75.93 ± 1.09	85.53 ± 0.76	88.07 ± 0.32
SLF+RKR	34.67 ± 1.35	59.76 ± 1.22	74.37 ± 0.96	82.70 ± 0.87	90.39 ± 0.44
RLR	41.25 ± 2.03	63.89 ± 1.99	81.83 ± 1.41	87.04 ± 1.23	92.53 ± 0.83
RSRC-ASP(Non)	36.13 ± 1.96	57.73 ± 1.41	78.20 ± 1.38	85.60 ± 0.86	89.13 ± 0.79
RSRC-ASP(A)	38.23 ± 2.58	59.87 ± 2.12	80.50 ± 1.93	86.13 ± 1.48	89.90 ± 1.12
RSRC-ASP(B)	42.47 ± 1.35	64.73 ± 1.21	85.07 ± 1.08	88.73 ± 0.83	91.73 ± 0.49
RSRC-ASP(C)	**51.27 ± 1.82**	**69.07 ± 1.91**	**87.13 ± 1.50**	**90.73 ± 0.94**	**93.53 ± 0.66**

Table 3 The recognition results (%) in CMU PIE

Gallery	1	2	4	6	8
ScSPM	36.00 ± 2.01	60.10 ± 2.92	68.82 ± 2.66	74.15 ± 1.89	76.61 ± 2.03
SRC	38.35 ± 2.64	59.80 ± 2.03	72.25 ± 1.71	78.20 ± 1.96	81.05 ± 1.80
LCRC	48.50 ± 2.15	67.70 ± 3.03	77.65 ± 1.92	82.60 ± 1.77	84.25 ± 1.55
SLF+RKR	42.31 ± 3.22	61.49 ± 2.56	71.27 ± 1.95	80.54 ± 1.96	83.95 ± 1.41
RLR	60.87 ± 1.46	73.93 ± 1.25	83.37 ± 1.10	86.20 ± 0.85	88.25 ± 0.66
RSRC-ASP(Non)	50.45 ± 2.59	69.30 ± 2.11	79.90 ± 1.90	83.70 ± 1.49	86.50 ± 1.52
RSRC-ASP(A)	58.14 ± 1.85	**76.33 ± 1.71**	**84.60 ± 1.41**	**88.26 ± 0.96**	**89.90 ± 0.77**
RSRC-ASP(B)	56.50 ± 1.41	74.40 ± 1.34	83.10 ± 0.90	86.55 ± 0.65	88.45 ± 0.49
RSRC-ASP(C)	**65.48 ± 1.31**	75.85 ± 0.96	81.50 ± 0.85	85.13 ± 0.54	87.48 ± 0.48

4.1.3 *CMU PIE*

The CMU PIE database [34] contains 68 individuals' frontal color images with variations of illumination, expression and poses. All the experimental environments remain the same with in the former subsections. Recognition results are shown in Figure 9 and Table 3. The proposed RSRC-ASP methods achieve 2%–8%'s improvements.

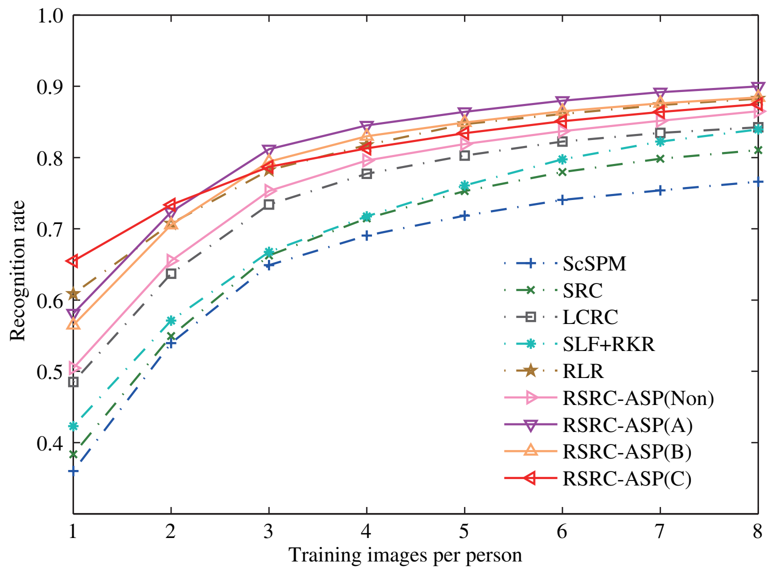

Figure 9 (Color online) The recognition results in CMU PIE.

Table 4 Recognition results (%) in LFW

Method	Recognition result	Method	Recognition result
SRC	72.7 ± 2.25	ScSPM	56.3 ± 3.62
LCRC	74.2 ± 1.46	SLF+RKR	71.9 ± 1.35
RLR	74.3 ± 0.63	RSRC-ASP(Non)	74.3 ± 0.41
RSRC-ASP(A)	74.1 ± 0.22	RSRC-ASP(B)	75.6 ± 0.33
RSRC-ASP(C)	**76.2 ± 0.18**		

4.1.4 *LFW*

LFW [35] is a large-scale database, which contains variations of pose, illumination, expression, misalignment and occlusion. In our experiments, we follow the experimental design in [36]. 143 subjects with no less than 11 samples per subject are chosen (4174 images in total) for experiments. For each person the first 10 samples are used for training data while the rest for testing. 10436 generic face samples collected on the Internet and other data sets are utilized for training auxiliary dictionary or residual criterion. Table 4 illustrates the recognition results comparing with a few of the state-of-the-art methods.

As the recognition results illustrated above, the proposed RARC-ASP method could achieve significant improvements comparing with the original spatial pyramid methods in various data sets. The experimental results verify the efficacy and stability of the proposed RSRC-ASP methods, which are more robust to variations caused by poses and misalignment. Meanwhile, dominant improvements are achieved with the robust local strategies. Additionally, it proves LNR strategy is more effective to misalignment and pose variations, but it is quite sensitive to illumination and occlusion variations as shown in Subsections 4.1.1 and 4.1.2. LIBRC and local auxiliary dictionary strategies own superior representation ability to represent local intra-class variations and are proven more robust to illumination and occlusion variations.

4.2 Efficiency

Experiments for comparing the efficiency of different spatial pyramid method are also conducted in extended Yale B, AR and CMU PIE. In each representation, 20 samples per subject are exploited as training dictionary. Experiments were implemented using MATLAB on a 2.5 GHz PC with 8 GB RAM. The proposed RSRC-ASP methods are proven to be competitively efficient as the state-of-the-art spatial pyramid based sparse representation method, as illustrated in Table 5.

Table 5 Efficiency results (averaged running time per sample)

Method	Efficiency (s)	Method	Efficiency (s)
SRC	1.094	ScSPM	1.191
SLF-RKR(L1)	0.672	RLR	0.665
RSRC-ASP(Non)	**0.423**	RSRC-ASP(A)	1.342
RSRC-ASP(B)	0.581	RSRC-ASP(C)	0.695

Table 6 The recognition results (%) via local auxiliary dictionary's sizes

Method	Auxiliary dictionary's size				
	50	100	200	300	500
RSRC-ASP(C)	85.44 ± 0.96	86.53 ± 0.73	86.92 ± 0.61	$\mathbf{87.18 \pm 0.55}$	87.10 ± 0.38

Table 7 The recognition results (%) via different spatial pyramid structure

Method	Pyramid structure				
	$\{1\}$	$\{1,2,3\}$	$\{1,2,3,6\}$	$\{6\}$	$\{1,2,3,6,36\}$
RSRC-ASP(Non)	72.69 ± 1.89	86.90 ± 0.62	96.50 ± 0.32	94.86 ± 0.21	96.13 ± 0.11
RSRC-ASP(A)	72.69 ± 1.89	87.04 ± 0.88	97.20 ± 0.43	94.44 ± 0.18	96.27 ± 0.21
RSRC-ASP(B)	78.58 ± 0.65	89.00 ± 0.71	97.60 ± 0.38	95.81 ± 0.06	97.69 ± 0.18
RSRC-ASP(C)	82.85 ± 0.72	93.21 ± 0.21	98.34 ± 0.55	97.52 ± 0.13	98.43 ± 0.08

4.3 Further discussion

In this subsection, experiments are conducted on extended Yale B database for checking some hype-parameters' influences in RSRC-ASP. For fair comparison, we fix the training dictionary size as 4 samples per subject and then conduct recognition experiments with other hype-parameters varying.

4.3.1 *Size of local auxiliary dictionary*

RSRC-ASP is a subspace based method, which requires dictionary being over-complete to represent the query facial samples. While for strategy C, the influence of local auxiliary dictionary is rather ignored in the above experiments. Focussing on this issue, firstly, we learn a series of local auxiliary dictionaries with various dictionary sizes. Then we conduct RSRC-ASP method with different local auxiliary dictionaries. Recognition results are illustrated in Table 6. As the local auxiliary dictionary's size increasing, the proposed RSRC-ASP performs much better. However, when the local auxiliary dictionary is over-complete (nearly 200–500), the recognition results of RSRC-ASP become stable.

4.3.2 *Spatial pyramid structure*

The performance of RSRC-ASP via different spatial pyramid structure is shown in Table 7. RSRC-ASP methods achieve higher performances with more sub-patches and deeper pyramid structure. It proves the spatial pyramid structure could better describe face samples' information from multiple perspectives. In addition, the multi-patches structure is more capable of valuing face samples' local reliability in RSRC-ASP's framework, which highly enhances the proposed methods' performances.

5 Conclusion

In this paper, for considering the differences of patches' reliability and enhancing spatial pyramid based methods' robustness, we propose a novel method of robust sparse representation based classification in an adaptive weighted spatial pyramid structure (RSRC-ASP). A sparsity evaluation based weighting strategy is proposed for residual aggravating, emphasizing the reliability of different patches. In addition, three local robust strategies: local-neighbourhood representation, local intra-class Bayesian residual criterion, and local auxiliary dictionary, are exploited to enhance the robustness of RSRC-ASP. Experiments on various data sets show that RSRC-ASP outperforms the classical sparse representation based classification methods especially for under-sampled face recognition problems. Our future work involves building deep RSRC-ASP framework and enhancing RSRC-ASP's computational efficiency.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61333015, 61302127, 11326198), China Postdoctoral Science Foundation (Grant No. 2015M570228), Opening Foundation of Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin Key Projects in the National Science and Technology Pillar Program (Grant No. 14ZCZDGX00033), and International Fostering Plan of Selected Excellent Postdoctors Subsidized by Tianjin City (2015). Thanks for the academic visiting support of the China Scholarship Council.

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Basri R, Jacobs D. Lambertian reflectance and linear subspaces. Pattern Anal Mach Intell, 2003, 25: 218–233
2 Qiao L, Chen S, Tan X. Sparsity preserving discriminant analysis for single training image face recognition. Pattern Recogn Lett, 2010, 31: 422–429
3 Candès E, Romberg J, Tao T. Stable signal recovery from incomplete and inaccurate measurements. Commun Pure Appl Math, 2006, 59: 1207–1223
4 Donoho D. For most large underdetermined systems of linear equations the minimal 1 -norm solution is also the sparsest solution. Commun Pure Appl Math, 2006, 59: 797–829
5 Wright J, Yang A, Ganesh A, et al. Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell, 2009, 31: 210–227
6 Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: which helps face recognition? In: Proceedings of International Conference on Computer Vision (ICCV), Barcelona, 2011. 471–478
7 Yang M, Zhang L. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In: Proceedings of the 11th European Conference on Computer Vision (ECCV), Heraklion, 2010. 448–461
8 Ma L, Wang C, Xiao B, et al. Sparse representation for face recognition based on discriminative low-rank dictionary learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, 2012. 2586–2593
9 Deng W, Hu J, Guo J. Extended src: undersampled face recognition via intraclass variant dictionary. Pattern Anal Mach Intell, 2012, 34: 1864–1870
10 Deng W, Hu J, Guo J. In defense of sparsity based face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, 2013. 399–406
11 Wei C, Wang Y. Learning auxiliary dictionaries for undersampled face recognition. In: Proceedings of International Conference on Multimedia and Expo (ICME), San Jose, 2013. 1–6
12 Yang M, Zhang L, Feng X, et al. Fisher discrimination dictionary learning for sparse representation. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), Barcelona, 2011. 543–550
13 Feng J, Ma X, Zhuang W. Collaborative representation bayesian face recogntion. Sci China Inf Sci, 2017, 60: 048101
14 Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. Siam J Imag Sci, 2009, 2: 183–202
15 Yang J, Zhang Y. Alternating direction algorithms for ℓ_1-problems in compressive sensing. Siam J Sci Comput, 2009, 33: 250–278
16 Li J, Lu C. A new decision rule for sparse representation based classification for face recognition. Neurocomputing, 2013, 116: 265–271
17 Liu B, Shen B, Gui L, et al. Face recognition using class specific dictionary learning for sparse representation and collaborative representation. Neurocomputing, 2016, 204: 198–210
18 Lu Z, Zhang L. Face recognition algorithm based on discriminative dictionary learning and sparse representation. Neurocomputing, 2016, 174: 749–755
19 Ouyang Y, Sang N, Huang R. Accurate and robust facial expressions recognition by fusing multiple sparse representation based classifiers. Neurocomputing, 2015, 149: 71–78
20 Grauman K, Darrell T. The pyramid match kernel: discriminative classification with sets of image features. In: Proceedings of International Conference on Computer Vision (ICCV), Beijing, 2005. 1458–1465
21 Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), New York, 2006. 2169–2178
22 Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification. In: Prcoeedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, 2009. 1794–1801
23 Yu K, Lv F, Huang T, et al. Locality-constrained linear coding for image classification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, 2010. 3360–3367
24 Shen F, Tang Z, Xu J. Locality constrained representation based classification with spatial pyramid patches. Neurocomputing, 2013, 101: 104–115

25 Yang M, Zhang L, Shiu S C, et al. Robust kernel representation with statistical local features for face recognition. IEEE Trans Neural Netw Learn Syst, 2013, 24: 900–912

26 Zhu P, Yang M, Zhang L, et al. Local generic representation for face recognition with single sample per person. In: Proceedings of Asian Conference on Computer Vision (ACCV), Singapore, 2014. 34–50

27 Wang J, Yang J, Yu K, et al. Robust local representation for face recognition with single sample per person. In: Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, 2015

28 Boureau Y, Bach F, Lecun Y, et al. Learning mid-level features for recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, 2010. 2559–2566

29 Zhou X, Yu K, Zhang T, et al. Image classification using super-vector coding of local image descriptors. In: Proceedings of the 11th European Conference on Computer vision (ECCV): Part V, Heraklion, 2010. 141–154

30 Moghaddam B, Nastar C, Pentland A. Bayesian face recognition using deformable intensity surfaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, 1996. 638–645

31 Moghaddam B, Jebara T, Pentland A. Bayesian face recognition. Pattern Recogn, 2000, 33: 1771–1782

32 Georghiades A, Belhumeur P, Kriegman D. From few to many: illumination cone models for face recognition under variable lighting and pose. Pattern Anal Mach Intell, 2001, 23: 643–660

33 Martinez A. The AR Face Database. Cvc Technical Report, 24. 1998

34 Gross R, Matthews I, Cohn J, et al. Multi-PIE. Image Vision Comput, 2010, 28: 807–813

35 Wolf L, Hassner T, Taigman Y. Similarity scores based on background samples. In: Proceedings of the 9th Asian Conference on Computer Vision (ACCV), Xi'an, 2009. 88–97

36 Yang M, Dai D, Shen L, et al. Latent dictionary learning for sparse representation based classification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, 2014. 4138–4145