Reconstructive and Discriminative Sparse Representation for Visual Object Categorization

Huanzhang Fu
huanzhang.fu@ec-lyon.fr
Emmanuel Dellandrea
emmanuel.dellandrea@ec-lyon.fr
Liming Chen
liming.chen@ec-lyon.fr

Université de Lyon, CNRS
Ecole Centrale de Lyon, LIRIS
UMR5205, F-69134, France

Abstract

Sparse representation was originally used in signal processing as a powerful tool for acquiring, representing and compressing high-dimensional signals. Recently, motivated by the great successes it has achieved, it has become a hot research topic in the domain of computer vision and pattern recognition. In this paper, we propose to adapt sparse representation to the problem of Visual Object Categorization which aims at predicting whether at least one or several objects of some given categories are present in an image. Thus, we have elaborated a reconstructive and discriminative sparse representation of images, which incorporates a discriminative term, such as Fisher discriminative measure or the output of a SVM classifier, into the standard sparse representation objective function in order to learn a reconstructive and discriminative dictionary.

Experiments carried out on the SIMPLIcity image dataset have clearly revealed that our reconstructive and discriminative approach has gained an obvious improvement of the classification accuracy compared to standard SVM using image features as input. Moreover, the results have shown that our approach is more efficient than a sparse representation being only reconstructive, which indicates that adding a discriminative term for constructing the sparse representation is more suitable for the categorization purpose.

1 Introduction

Generic Visual Object Categorization (VOC) aims at predicting whether at least one or several objects of some given categories are present in an image. In fact, VOC is a fundamental problem in computer vision and pattern recognition, and has become an important research topic due to the wide range of possible applications such as video monitoring, video coding systems, security access control, automobile driving support as well as automatic image and video indexation and retrieval [12] [22]. Until now, many VOC methods have been proposed and applied to the classification of numerous objects categories like, for example, cars, motorbikes, animals, people, furniture etc. Despite many efforts and much progress that have been made during the past years, it remains an open problem and is still considered as one of

BMVC 2011 http://dx.doi.org/10.5244/C.25.39

the most challenging topics in computer vision [7]. In particular, the image representation is a key problem since, from the image visual content presented in the form of image features, it has to be able to model effectively this content in a discriminative way to allow an efficient classification of the image.

In this paper, we propose an approach for the image representation inspired by the principles of sparse representation theory that we have adapted to the problem of VOC. Indeed, sparse representation models of signals have received a lot of attentions and is a very active research area in recent years. It is originally used as a powerful tool for acquiring, representing and compressing high-dimensional signals in the signal processing applications and has achieved great successes. These successes are mainly due to the fact that important classes of signals have naturally sparse representations with respect to fixed bases, or concatenations of such bases. Moreover, a set of efficient and effective algorithms based on convex optimization or greedy pursuit has been proposed for solving the sparse representation problem and computing such representations with high fidelity [3].

The goal of sparse representation is to obtain a compact high-fidelity representation of a given signal, which can be considered as a linear combination of atoms from an overcomplete dictionary [15]. The property of sparsity in the representation of signals has also been approved in human perception by some studies of human vision [16] [17]. In fact, many neurons in the visual pathway are selective for a variety of specific stimuli in the human vision and then can be considered as an overcomplete dictionary. Thus, the firing of the neurons with respect to a given input image is typically highly sparse. Recent research on wavelet, ridgelet, curvelet and contourlet transforms has also greatly accelerated and promoted the development of sparse representation model. Until now, it has been widely used and obtained promising results in many different applications, such as signal separation [23], denoising [6], coding [18], image inpainting and restoration [14] and magnetic resonance spectroscopy quantification [9].

Recently, techniques from sparse signal representation have significantly impacted the domain of computer vision and pattern recognition [30] [31] [13], in which we are often more interested in extracting the visual content of an image rather than a compact high-fidelity representation. Variations and extensions of these representations have been widely used in many vision tasks, including face recognition [31], image super-resolution and classification [32] [13], motion segmentation [21], background modeling [4]. In almost all of these applications, the sparse representation based methods has provided encouraging results which are comparable to the state of the art ones. This has motivated us to propose an approach adapting these principles to the problem of VOC.

The rest of this paper is organized as follows. Sparse representation background is presented in section 2. Our reconstructive and discriminative sparse representation for VOC is then described in section 3. Experiments we have conducted to evaluate this model are detailed in section 4. Finally, the conclusion and perspectives are drawn in section 5.

2 Sparse representation model

Let consider a signal $y \in \mathbb{R}^n$, which will be represented as a linear combination of basic elements from a dictionary $D \in \mathbb{R}^{n \times K}$ composed by atoms in columns $\{d_j\}_{j=1}^K$. We say that a representation of the signal y based on this specific dictionary D is any vector $x \in \mathbb{R}^K$ which satisfies:

$$y = Dx \tag{1}$$

In the case where $n < K$, the dictionary D is said to be overcomplete and this equation is underdetermined thus having many possible solutions. Conventionally, in this case, the minimum ℓ^2 norm solution is chosen:

$$\min_x(||x||_2) \ \ subject \ \ to \ \ Dx = y \tag{2}$$

where $||x||_2$ is the ℓ^2 norm of x. The above problem can easily be solved and it has a unique solution as follow:

$$x = D^+ y = D^T (DD^T)^{-1} y \tag{3}$$

where D^+ is the pseudoinverse of D. However, this solution is generally non sparse with many nonzero elements corresponding to the atoms from the dictionary and consequently does not satisfy our expectation. Indeed, we would rather prefer a sparse solution, that is to say we want to find a linear combination of only a few atoms to approximate the signal y. This problem can be formally described by

$$\min_x(||x||_0) \ \ subject \ \ to \ \ Dx = y \tag{4}$$

where $||x||_0$ is ℓ^0 norm of x and equals the number of nonzero elements in the vector x. Solving the equation (4) is a NP hard problem because of its nature of combinational optimization. Nevertheless, there exist many approximation techniques for this task such as Matching Pursuit (MP) [15] which consists in selecting one atom at each stage based on the minimization of the residue in a greedy way, and Orthogonal Matching Pursuit (OMP) [20] which involves the computation of inner products between the signal and dictionary columns. If the dictionary is an orthogonal vector set and the signal is indeed a sparse combination of atoms, OMP is guaranteed to find this sparse set.

Another crucial aspect for applying sparse representation model successfully on the signals (images) is the design of the dictionary, namely D in the equation (1). One type of approaches consists in using the preconstructed dictionaries which do not change during the problem solving. Such dictionaries based on the transforms mentioned above, i.e. ridgelet, curvelet and contourlet, have been widely used in signal processing. Another possibility consists in using the dictionary composed by the training images themselves, which has also given promising results as in [31] and [8].

However, this conventional setting may not be suitable to be directly employed in the domain of computer vision and pattern recognition as there is no given basis with good property compared to signal processing [30]. In order to address this new situation, another type of approaches has been proposed in order to learn a task-specific dictionary from given samples by updating the dictionary, with the purpose of describing the image content more effectively. We can mention here two appealing and widely used methods: Method of Optimal Directions (MOD) [8] and K-SVD [1]. Both of them are iterative methods, containing a sparse coding stage which finds the corresponding coefficients x of a signal y based on the current dictionary and a dictionary update stage which updates the dictionary using coefficients obtained from previous stage to better fit the data.

3 RDSR_VOC: a Reconstructive and Discriminative Sparse Representation for Visual Object Categorization

Let consider a set of N training signals $\{y_i\}_{i=1}^N$ belonging to M categories. $Y = [y_1, y_2, ..., y_N]$ is a signal matrix with the corresponding sparse coefficients based on the dictionary D as

$X = [x_1, x_2, ..., x_N]$. Moreover, we suppose that N_i signals are in the category M_i, for $1 \leq i \leq M$.

The objective function of the standard reconstructive sparse representation can be expressed as:

$$\min_{D,X}\{||Y - DX||_F^2\} \quad subject \quad to \quad ||x_i||_0 \leq L \quad \forall i \tag{5}$$

If we incorporate the sparsity constraint into the function, it can be reformulated as:

$$\begin{aligned} &\min_{D,X,\Lambda}\{\lambda_1||Y - DX||_F^2 + \lambda_2 \sum_{i=1}^{N} ||x_i||_0\} \\ \Rightarrow &\min_{D,X,\Lambda}\{\lambda_1 \sum_{i=1}^{N} ||y_i - Dx_i||_2^2 + \lambda_2 \sum_{i=1}^{N} ||x_i||_0\} \end{aligned} \tag{6}$$

where $\Lambda = \{\lambda_1, \lambda_2\}$ is a set of regularization parameters which adjust the tradeoff between the reconstruction error and the sparsity.

The main goal of our approach is to learn a reconstructive and discriminative dictionary which helps to increase the discriminative power of the signal sparse representation based on this dictionary, while keeping a relative low reconstruction error, i.e. the reconstructed signal using the obtained sparse coefficients being as close to the original signal as possible. Therefore, inspired by [11], the Fisher discriminative term [2] is introduced to the objective function.

Suppose S_W is the "intra-class scatter" which measures the within-class covariance:

$$S_W = \sum_{i=1}^{M} S_i \tag{7}$$

where

$$S_i = \sum_{x_j \in M_i} (x_j - m_i)(x_j - m_i)^T \tag{8}$$

$$m_i = \frac{1}{N_i} \sum_{x_j \in M_i} x_j \tag{9}$$

m_i is the mean of the signals belonging to category M_i. Let S_B denote the "inter-class scatter" which we identify as a measure of the between-class covariance

$$S_B = \sum_{i=1}^{M} N_i (m_i - m)(m_i - m)^T \tag{10}$$

where m is the mean of all signals

$$m = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{11}$$

Then, the Fisher discriminative score can be expressed as

$$F(X) = \frac{||S_B||_2^2}{||S_W||_2^2} = \frac{||\sum_{i=1}^{M} N_i (m_i - m)(m_i - m)^T||_2^2}{||\sum_{i=1}^{M} \sum_{x_j \in M_i} (x_j - m_i)(x_j - m_i)^T||_2^2} \tag{12}$$

The Fisher score is maximized when the distance between different categories is maximized while that within a category is minimized, thus making the classification task easier.

Incorporating the Fisher discriminative term to (6) gives:

$$\min_{D,X,\Lambda}\{\lambda_1\sum_{i=1}^{N}||y_i-Dx_i||_2^2+\lambda_2\sum_{i=1}^{N}||x_i||_0-\lambda_3 F(X)\} \tag{13}$$

where $\Lambda=\{\lambda_1,\lambda_2,\lambda_3\}$ is, similarly to (6), the set of regularization parameters used to tune the tradeoff between the reconstruction error $\sum_{i=1}^{N}||y_i-Dx_i||_2^2$, the sparsity $\sum_{i=1}^{N}||x_i||_0$ and the discriminative power $F(X)$. The expected reconstructive and discriminative dictionary can be learned by solving properly the previous minimization problem. Thus, the signal sparse representation which gains the discriminative ability while retaining its faithfulness to the original signal can also be obtained through sparse coding based on the learned dictionary.

As mentioned previously, most of works in the literature use an iterative method to solve the dictionary learning problem. They generally contains two stages: sparse coding and dictionary update. We have followed this strategy for solving the minimization problem in (13). The first question that arises is "Given the dictionary, how to do the sparse coding faced with our reconstructive and discriminative objective function?". Since it involves not only a single signal but also all the training signals, the traditional sparse coding methods, such as BP and OMP, can not be directly applied to (13). Therefore we propose here a Sequential Forward Sparse Coding algorithm (SFSC) to do this task.

Let G being the function to be minimized:

$$G=\lambda_1\sum_{i=1}^{N}||y_i-Dx_i||_2^2+\lambda_2\sum_{i=1}^{N}||x_i||_0-\lambda_3 F(X) \tag{14}$$

The first step of SFSC consists in selecting one atom from the dictionary D with the smallest value of function G which is calculated by assuming that only that specific atom has been used for the sparse decomposition to obtain the sparse coefficients of all signals $\{x_i\}_{i=1}^{N}$ as well as X. Indeed, if we know beforehand the subset Γ of indices of atoms which are used for sparse decomposition, the sparse coefficients can easily be obtained using

$$X=D_\Gamma^{+}Y \tag{15}$$

where D_Γ is a reduced dictionary composed only by the atoms whose indices are in Γ. Then in each following step, we continue to select one atom among the remaining ones, which yield the smallest value of G based on the subset of atoms formed by the combination of pre-selected atoms and this new one, until reaching the stopping rule. Here, the stopping rule can consist in achieving the predefined number of atoms used for sparse decomposition or stopping when the value of G begins to increase. The detailed algorithm is as follows:

SFSC algorithm

- Task: Given the dictionary $D\in\mathbb{R}^{n\times K}$, the regularization parameter set Λ and the set of signal $Y=[y_1,y_2,...,y_N]$ to be represented by a linear combination of atoms from D, find the corresponding coefficients $X=[x_1,x_2,...,x_N]$ that minimize G

$$G=\lambda_1\sum_{i=1}^{N}||y_i-Dx_i||_2^2+\lambda_2\sum_{i=1}^{N}||x_i||_0-\lambda_3 F(X).$$

- Initialization: Set the initial index set $\Gamma^0=\emptyset$ and the indicator of iteration $t=1$.

- Repeat until stopping rule:
  - For each $i \notin \Gamma^{t-1}$ and $i \in \{1,2,...,K\}$, let $\Psi = \Gamma^{t-1} \cup i$. Then calculate the sparse coefficients $X = D_{\Psi}^{+} Y$ as well as the value of G_i based on X. D_{Ψ} represents the reduced dictionary composed by the columns in D whose indices are in Ψ
  - $i_{min} = \arg_i \min(G_i)$
  - $\Gamma^t = \Gamma^{t-1} \cup i_{min}$
  - $t = t+1$
- Calculate the sparse coefficients $X = D_{\Gamma^t}^{+} Y$.

Concerning the dictionary update stage, we can employ the method of K-SVD [1] mentioned in section 2. Thus one complete dictionary learning algorithm is formed and ready to be used for generic visual object categorization. The entire classification algorithm is described as follows:

RDSR_VOC algorithm

1. Extract the feature vector representing the image visual content for all the images: $f_{i,j} \in \mathbb{R}^n, i \in \{1,2,...,M\}, j \in \{1,2,...,N_i\}$, where M is the number of categories and N_i is the number of images for i-th category.
2. Normalize all $f_{i,j}$ to have unit ℓ^2 norm.
3. Learn a reconstructive and discriminative dictionary D of sparse representation based on training images, by iteratively running the following two stages with the purpose of minimizing the objective function G. D is initialized by a subset of training image vectors, chosen randomly.
   - *Sparse Coding* using SFSC.
   - *Dictionary Update* similar to the dictionary update stage of K-SVD [1]. $k = 1,2,...,K$ in D^{t-1}, update it by
     - Define the group of signals that use this atom $\omega_k : \{i | 1 \le i \le N, x_T^k(i) \neq 0\}$ where x_T^k is the k-th row of X.
     - Compute the overall representation error matrix, E_k, by
     $$E_k = Y - \sum_{j \neq k} d_j x_T^j.$$
     - Restrict E_k by choosing only the columns corresponding to ω_k, and obtain E_k^R.
     - Apply SVD decomposition $E_k^R = U \Delta V^T$. Choose the updated dictionary column $\tilde{d}_k$ to be the first column of U. Update the coefficient vector x_R^k to be the first column of V multiplied by $\Delta(1,1)$. Here x_R^k is a reduced version of the row vector x_T^k by discarding of the zero entries.
4. Compute the sparse coefficients of all the images based on the learned dictionary D, including the training images and test images.
5. Use a classifier (SVM for example) to accomplish the classification task, using the obtained sparse coefficients as input.

One advantage of our proposed RDSR_VOC is that other discriminative criteria can be easily employed by replacing $F(X)$ into the objective function, without changing the whole classification scheme. For example, we have tested the use a SVM classification accuracy as the discriminative term. All these experiments are presented in the next section.

4 Experimental results

We present in this section the experiments that have been conducted in order to evaluate the discriminative ability of our approach, RDSR_VOC, for the problem of VOC.

For training and testing, we have used the SIMPLIcity image dataset [28] which contains a total of 1000 images from ten categories (100 images per category): African & village (C1), Beach (C2), Building (C3), Bus (C4), Dinosaur (C5), Elephant (C6), Flower (C7), Horse (C8), Mountain & glacier (C9) and Food (C10). Some sample images are presented in Figure 1.

Figure 1: Some sample images from SIMPLIcity dataset. From left to right, from top to bottom, they belong to African & village, Beach, Building, Bus, Dinosaur, Elephant, Flower, Horse, Mountain & glacier and Food respectively.

A total number of 2446 features has been computed to represent each image from SIMPLIcity dataset. The corresponding feature set includes Color Auto-Correlogram (CAC) [10], Color Coherence Vectors (CCV) [19], Color Histogram (CH) [25] , Color Moments (CM) [24], Edge Histogram (EH) [29], Grey Level Co-occurrence Matrix (GLCM) [27], Texture Auto-Correlation (TAC) [27] and Local Binary Pattern (LBP) [26].

Considering different discriminative criteria incorporated in the objective function, three tests have been done to evaluate the performance of our proposed RDSR_VOC: using Fisher discriminative measure (noted as Fisher in the following); using the output of a SVM classifier with RBF kernel (noted as SVM_RBF in the following); using the output of a SVM classifier with linear kernel (noted as SVM_Linear in the following). All the regularization parameters are empirically set to have the same value, meaning that all the three terms, namely the reconstruction error, the sparsity and the discriminative power, in the objective function G in (14) have the same weight. The stopping rule of SFSC is set to use 60 atoms for sparse coding (this point is discussed further in this section).

Moreover, for comparison purpose, two additional experiments have been carried out. In the first one, we have used SVM directly on the feature vectors extracted from images to classify a test image into the corresponding category according to the object it contains (noted SVM in the following). Several configurations have been evaluated for SVM (kernels and parameters), and only the best one is presented below. In the second experiment (noted RSR_VOC in the following), we have modeled the image representation thanks to

a traditional sparse representation that is only reconstructive. In this case, the dictionary consists of the feature vectors computed from the training images and the objective function that is purely reconstructive (excludes the discriminative term) is optimized using OMP algorithm (mentioned in section 2) to obtain sparse coefficients. The image sparse representations made of these sparse coefficients are then used to feed SVM classifiers to perform the classification task.

The results are give in Table 1. Rows C1, C2, ..., C10 corresponds to the 10 image categories and columns correspond to the different methods that are compared. The result values are given in term of average classification accuracy rate (ratio of number of examples correctly classified with respect to the total number of examples classified) which has been obtained using a 4-fold cross-validation.

From Table 1, we can clearly see that using a sparse representation allow to significantly improve the classification results compared to the traditional method using SVM since it presents an improvement of 4% for the purely reconstructive sparse representation (RSR_VOC) and of near 6% for the reconstruction and discriminative sparse representation (RDSR_VOC with Fisher). We can also notice that the powerful SVM has already obtained a relatively high classification rate.

Now, focusing on sparse representation, we can note that the overall classification rate increases from 87.5% with RSR_VOC to 89.1% with RDSR_VOC using Fisher, which means that the classification ability of RDSR_VOC is really reinforced by adding the Fisher discriminative term to the standard sparse representation framework. Indeed, although the improvement may appear relatively low with an increasing value of around 2%, it is in fact significant since the improvement space left is very small. Indeed, the higher the classification rate is, the more difficult it will be to increase it. Now let us concentrate on the different image categories. We can see that the superiority of RDSR_VOC with Fisher among RSR_VOC is mainly due to the large improvement for difficult categories, namely the ones with lower rate such as C2 (Beach), C3 (Buildings) and C9 (Mountains & glaciers). For instance, 9% of augmentation has been observed for C9 using RDSR_VOC compared to RSR_VOC.

The results of SVM_RBF and SVM_Linear are very similar with an overall classification rate of 87.6% for both of them, showing no advantage compared to RSR_VOC and being worse than RDSR_VOC with Fisher. This is probably due to the "overfitting" during the dictionary learning and classifier training, as we have used two independent SVM classifiers in the process, one for the discriminative term and the other for the final classification. However, it did not hurt much the performance either, proving that our proposed algorithm RDSR_VOC can robustly cooperate with different discriminative terms without changing the algorithm itself.

Our proposed sparse coding method SFSC needs a criterion as stopping rule. It can be either the number of atoms used for sparse decomposition or the decrease of the objective function value. We have chosen the first criterion for its simplicity and the fact that it can avoid the case where many atoms have been used but only with very small coefficients thus yielding a non-sparse representation, which may probably happen with the second criterion. However, determining the optimal number of atoms used still remains an open question. In our experimentation, we have tested three typical values (30, 60, 100) for all three experiments, namely Fisher, SVM_RBF and SVM_Linear. Besides the results presented above for 60 atoms used, the results with 30 and 100 atoms used are presented in Table 2.

From this table, we can clearly observe that the classification rates with 60 atoms and 100 atoms are much higher than that with 30 atoms, presenting an improvement of 4% in

Table 1: Average classification rate for the 10 image categories of SIMPLIcity (in rows) using different approaches (in columns).

	SVM	RSR_VOC	RDSR_VOC (Fisher)	RDSR_VOC (SVM_RBF)	RDSR_VOC (SVM_Linear)
C1	80%	90%	88%	86%	85%
C2	82%	73%	78%	76%	74%
C3	62%	78%	82%	79%	77%
C4	84%	97%	96%	95%	97%
C5	100%	100%	100%	100%	100%
C6	86%	86%	83%	85%	82%
C7	84%	95%	97%	97%	97%
C8	98%	98%	94%	94%	95%
C9	72%	73%	82%	77%	78%
C10	86%	85%	91%	87%	91%
Average	83.4%	87.5%	89.1%	87.6%	87.6%

Table 2: Influence of the number of atoms in the learned dictionary on the average classification rate with different configurations of RDSR_VOC.

Number of atoms	RDSR_VOC (Fisher)	RDSR_VOC (SVM_RBF)	RDSR_VOC (SVM_Linear)
30	84.5%	83.5%	84.6%
60	89.1%	87.6%	87.6%
100	89.0%	87.2%	87.4%

average. However, there is not much difference between the results with 60 atoms and 100 atoms, the results with 60 atoms being a little bit better than those with 100 atoms. This suggests that using 60 atoms is a good choice for space coding with SFSC and using more atoms may not be helpful to improve the performance.

5 Conclusion

We have presented in this paper a new approach for Visual Object Categorization via a sparse representation based on a reconstructive and discriminative principle, RDSR_VOC, which includes a discriminative term, such as Fisher discriminative measure or the output of a SVM classifier, to the standard sparse representation objective function in order to learn a reconstructive and discriminative dictionary.

Experiments conducted on the SIMPLIcity dataset have clearly revealed that our reconstructive approach has gained an obvious improvement of the classification accuracy compared to standard SVM using image features as input. Moreover, our reconstructive and discriminative approach has obtained better results than a pure reconstructive one which shows that adding a discriminative term for constructing the sparse representation is more

suitable for the classification task.

Therefore, we are convinced that sparse representation can greatly help for designing efficient approaches for VOC purpose. Thus, we envisage to go further in our future works by investigation different directions including the way to identify optimal regularization parameters for RDSR_VOC, the way different spatial pyramid levels should be fused, the way to combine the results from different features (for example, we can replace SVM by MKL to realize an automatic feature combination in the kernel level) and the design of novel kernels to best fit the properties of our sparse image representation.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[2] C.M. Bishop. *Pattern recognition and machine learning*. Springer, 2007.

[3] A.M. Bruckstein, D.L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *Society for Industrial and Applied Mathematics Review*, 51(1):34–81, 2009.

[4] M. Dikmen and T. Huang. Robust estimation of foreground in surveillance video by sparse error estimation. In *Proceedings of the International Conference on Image Processing*, pages 1–4, 2008.

[5] M. Elad and M. Aharon. Image denoising via learned dictionaries and sparse representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 895–900, 2006.

[6] K. Engan, S.O. Aase, and J.H. Hakon-Husoy. Method of optimal directions for frame design. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 2443–2446, 1999.

[7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results. http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html.

[8] H. Fu, C. Zhu, E. Dellandréa, C.E. Bichot, and L. Chen. Visual object categorization via sparse representation. In *Proceedings of the International Conference on Image and Graphics*, pages 943–948, 2009.

[9] Y. Guo, S. Ruan, J. Landré, and J.M. Constans. A sparse representation method for magnetic resonance spectroscopy quantification. *IEEE Transactions on Biomedical Engineering*, 57(7):1620–1627, 2010.

[10] J. Huang, S.R. Kumar, M. Mitra, W. Zhu, and R. Zabih. Image indexing using color correlograms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 762–768, 1997.

[11] K. Huang and S. Aviyente. Sparse representation for signal classification. In *Proceedings of the Advances in Neural Information Processing Systems*, volume 19, pages 609–616, 2006.

[12] M.S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing Communication and Applications*, 2(1):1–19, 2006.

[13] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[14] J. Mairal, G. Sapiro, and M. Elad. Learning multiscale sparse representations for image and video restoration. *SIAM Multiscale Modeling & Simulation*, 7(1):214–241, 2008.

[15] S.G. Mallat and Z.F. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transaction on Signal Processing*, 41(12):3397–3415, 1993.

[16] B.A. Olshausen and B.J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

[17] B.A. Olshausen and B.J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23):3311–3325, 1997.

[18] B.A. Olshausen, P. Sallee, and M.S. Lewicki. Learning sparse image codes using a wavelet pyramid architecture. In *Proceedings of the Advances in Neural Information Processing Systems*, volume 13, pages 887–893, 2001.

[19] G. Pass and J. Miller R. Zabih. Comparing images using color coherence vectors. In *Proceedings of the ACM international conference on Multimedia*, pages 65–73, 1997.

[20] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 40–44, 1993.

[21] S. Rao, R. Tron, R. Vidal, and Y. Ma. Motion segmentation via robust subspace separation in the presence of outlying, incomplete, and corrupted trajectories. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[22] I. El Sayad, J. Martinet, T. Urruty, and C. Djeraba. Toward a higher-level visual representation for content-based image retrieval. *Journal of Multimedia Tools and Applications*, pages 1–28, 2010.

[23] J. Starck, M. Elad, and D. Donoho. Image decomposition via the combination of sparse representation and a variational approach. *IEEE Transaction on Image Processing*, 14 (10):1570–1582, 2005.

[24] M.A. Stricker and M. Orengo. Similarity of color images. In *Proceedings of the Storage and Retrieval for Image and Video Databases III*, volume 2, pages 381–392, 1995.

[25] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[26] V. Takala, T. Ahonen, and M. Pietikainen. Block-based methods for image retrieval using local binary patterns. In *Proceedings of the Scandinavian Conference on Image Analysis*, pages 882–891, 2005.

[27] M. Tuceryan and A.K. Jain. Texture analysis. In *The Handbook of Pattern Recognition and Computer Vision (2nd Edition)*, pages 207–248, 1993.

[28] J.Z. Wang, J. Li, and G. Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947–963, 2001.

[29] C.S. Won. Feature extraction and evaluation using edge histogram descriptor in mpeg-7. In *Proceedings of the Advances in Multimedia Information Processingï£¡CPCM*, volume 3333, pages 583–590, 2004.

[30] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. In *Proceedings of IEEE*, volume 98, pages 1031–1044, 2009.

[31] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[32] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.