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

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 [4]. 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 the most challenging topics in computer vision [2]. 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 to adapt the principles of sparse representation theory to the problem of VOC. 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.

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:

$$\min_{D,X}\{||Y-DX||_F^2\} \quad subject \quad to \quad ||x_i||_0\leq L \quad \forall i \tag{1}$$

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

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

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 [3], the Fisher discriminative term [1] is introduced to the objective function. The Fisher discriminative score can be expressed as:

$$F(X)=\frac{||\sum_{i=1}^M N_i(m_i-m)(m_i-m)^T||_2^2}{||\sum_{i=1}^M\sum_{x_j\in M_i}(x_j-m_i)(x_j-m_i)^T||_2^2} \tag{3}$$

where m_i is the mean of the signals belonging to category M_i and m is the mean of all signals. 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 (2) gives:

$$\min_{D,X,\Lambda}\{\lambda_1\sum_{i=1}^N||y_i-Dx_i||_2^2+\lambda_2\sum_{i=1}^N||x_i||_0-\lambda_3F(X)\} \tag{4}$$

where $\Lambda=\{\lambda_1,\lambda_2,\lambda_3\}$ is, similarly to (2), 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.

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 (4). 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 (4). Therefore we propose a Sequential Forward Sparse Coding algorithm (SFSC) to do this task.

Let G being the function to be minimized:

$$G=\lambda_1\sum_{i=1}^N||y_i-Dx_i||_2^2+\lambda_2\sum_{i=1}^N||x_i||_0-\lambda_3F(X) \tag{5}$$

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 $X=D_\Gamma^+Y$ 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 described in the paper, as well as the experiments. They have been conducted on the SIMPLIcity dataset and 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.

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

[2] 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.

[3] 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.

[4] 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.