Supervised Spectral Subspace Clustering for Visual Dictionary Creation in the Context of Image Classification

Imtiaz Masud Ziko, Elisa Fromont, Damien Muselet, Marc Sebban

To cite this version:
Imtiaz Masud Ziko, Elisa Fromont, Damien Muselet, Marc Sebban. Supervised Spectral Subspace Clustering for Visual Dictionary Creation in the Context of Image Classification. ACPR 2015: 3rd IAPR Asian Conference on Pattern Recognition, Nov 2015, Kuala Lumpur, Malaysia. hal-01224466v1

HAL Id: hal-01224466
https://hal.archives-ouvertes.fr/hal-01224466v1

Submitted on 9 Nov 2015 (v1), last revised 27 Nov 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Supervised Spectral Subspace Clustering for Visual Dictionary Creation in the Context of Image Classification

Imtiaz Masud Ziko, Elisa Fromont, Damien Muselet and Marc Sebban
Laboratoire Hubert Curien, Saint Etienne, France
ziko.iut@aiub.edu, {elisa.fromont, damien.muselet, marc.sebban}@univ-st-etienne.fr

Abstract

When building traditional Bag of Visual Words (BOW) for image classification, the K-means algorithm is usually used on a large set of high dimensional local descriptors to build the visual dictionary. However, it is very likely that, to find a good visual vocabulary, only a sub-part of the descriptor space of each visual word is truly relevant. We propose a novel framework for creating the visual dictionary based on a spectral subspace clustering method instead of the traditional K-means algorithm. A strategy for adding supervised information during the subspace clustering process is formulated to obtain more discriminative visual words. Experimental results on real world image dataset show that the proposed framework for dictionary creation improves the classification accuracy compared to using traditionally built BOW.

1. Introduction

Bag of visual words (BOW) method [7] has marked significant advancements in the image classification applications. The method involves extracting a large set of discriminative local descriptors (such as SIFT, Color-SIFT [1] etc.) from the training images and learning the visual dictionary by applying clustering algorithm, usually the K-means in the descriptor space. Finally, the cluster representatives are called the visual words of the visual dictionary. However, the K-means algorithm considers all the feature dimensions while calculating the distance during estimating clusters. As a consequence, actual nearest neighbor estimation can get affected by having noisy information from irrelevant feature dimensions [4, 13]. Image descriptor domain is a complex structure having high dimensionality where feature descriptors are assumed to lie in a union of linear or nonlinear subspaces. Thus, applying the K-means algorithm often fails to capture the underlying structure of the descriptor space.

Alternatively, we have experimented to extract the visual words according to the relevant subspaces by using suitable subspace clustering methods. Subspace clustering methods were not explored very well for this specific problem related to the BOW model for image classification. There has been a recent work [5] where they worked to get the nonlinear subspaces from the raw image space by using Restricted Boltzmann Machine (RBM) [11]. In the comparison of results, they incorporate spatial information using the SPM method [14] during the classification. In contrast, we aim to have a fair comparison between simple BOW model using K-means and using subspace clustering approach without using additional tweak of adding spatial information in the built dictionary.

Spectral clustering methods [19] have shown a tendency to perform better in comparison to other existing unsupervised methods for object discovery as reported in [17]. In general, the spectral clustering methods construct a similarity matrix \( W \in \mathbb{R}^{N \times N} \) by estimating pairwise similarity among all the \( N \) data points. Let \( G(V, E) \) be an undirected graph where \( V \) is the set of \( N \) vertices and \( E \) is the set of weighted edges. Ideally, \( W_{ij} = 1 \) if points \( i \) and \( j \) are connected and \( W_{ij} = 0 \) otherwise. Finally, the \( K \) clusters are achieved by applying the K-means to the subset of \( K \) eigenvectors of the Laplacian matrix \( L \in \mathbb{R}^{N \times N} \) formed from \( W \). There are several ways to get the Laplacian matrix such as: the unnormalized Laplacian \( L = \text{diag}(W) - W \), the normalized Laplacian \( L_{sym} = \text{diag}(W)^{-0.5}L\text{diag}(W)^{-0.5} \) and \( L_{rw} = \text{diag}(W)^{-1}L \) where \( 1 \) is the vector of all 1’s.

Spectral subspace clustering methods attempt to build the similarity matrix \( W \) such that only the data points coming from the same subspace will be connected. The rest of the steps remain the same as the spectral clustering method. We compare – SSC [8], SMCE [9], SLBF [20] and SCC [6] for our experiments due to their promising performances reported in the literature for different subspace segmentation problems [18]. Initially we assess these methods on synthetic dataset and based on their performances, pick three of them – SSC, SMCE and SLBF for our proposed framework. The main contribution of this paper is that we propose a
null
eigenvectors which can show the subspace consistency for the underlying cluster [12]. This is why, we need to define a maximum limit on selecting eigenvectors for each subspace cluster such that the number of selected eigenvectors will be within the defined limit. For an example, we define maximum limit as 5 and there are two clusters with data matrix \(20 \times 128\) and \(3 \times 128\) where 128 is the feature dimension, we select only the first 5 eigenvectors for the first case and retain all the 3 eigenvectors for the later case. The flowchart of the whole procedure is shown in fig. 3.

2.2. Supervised Spectral Subspace Clustering

Our purpose of adding supervision for dictionary creation is to get visual words coming from relevant subspaces and from the same categorical images to retain the class purity. The first step involves building the similarity matrix \(W\) as discussed earlier which can be considered as a block diagonal matrix in terms of subspace similarity (step 1 in fig. 4). Each subspace contains descriptors from different categorical images (step 2 in fig 4). We simply disconnect the edge between the points coming from different image classes and retain other edge weights the same as before. In this way, we ensure to have clusters having descriptors from the same subspace and from the same categorical images. To remove the connection, the respective weights are made very close to 0. We chose 0.005 so that the graph does not get too sparse. Finally, the similarity matrix will have connected components in terms of the same subspace and the same image class. Note that, there can be clusters having the same class information but being in different subspaces (step 4 in fig 4) which also implies to have different distinc-
3. Experimental Results

Two datasets have been used for the experiments – synthetic dataset with known ground truth and the 15 Scene Categories dataset [14] for image classification. The synthetic dataset has been generated with total 300 data points lying in $\mathbb{R}^3$. There are two axis aligned subspaces with each of them having 150 points lying in $\mathbb{R}^2$. One of the dimensions is shared by both subspaces. The dataset has been created for both linear (D1) and nonlinear (D2) subspaces (see figure 5). The performance of the subspace clustering methods are determined by the rate of misclassified subspace points calculated as (3). The 15 Scene Categories dataset consists of total 4885 images taken from 15 different scene categories. We select 100 images per class for training and the rest for testing. In total 28,000 descriptors were randomly sampled from all the classes to build the dictionary.

$$\text{Error} = \frac{\text{no. of misclassified subspace points}}{\text{no. of total points}} \times 100\% \quad (3)$$

SIFT descriptors are used [15] along with the dense sampling method to have image patches. For dense sampling, 6 pixel shift in grid spacing is used and the size of the patch is kept $16 \times 16$. We used intersection kernel [16] along with linear SVM classifier by using LIBLINEAR package [10]. We cluster the descriptors into 1000 visual words for the image classification on the 15 Scene Categories dataset. All the experiments are run 10 times by reshuffling the test and training set. Finally, the average classification accuracy is reported for the performance measure.

3.1. Result discussion on Synthetic dataset

The experimental results on both synthetic dataset D1 and D2 are shown in Table 1. From the results, we can observe that the K-means clearly breaks the subspace structures while data points lie in a union of subspaces. On the contrary, the missclassification rates are quite lower for the subspace clustering methods than the K-means. For detecting clusters in linear subspaces, SLBF and SCC algorithm show the best performance having 0% missclassification rate. In case of nonlinear dataset (D2), all of the subspace clustering methods show very good results having missclassification rate within 2%. Interestingly, SLBF method exhibits better consistency for detecting both linear and nonlinear subspaces. Despite of having better results in synthetic data experiments, it was infeasible to run SCC algorithm in our further experiments with real world dataset due to its very high computational time and memory consumption. Finally, three methods – SSC, SMCE and SLBF are selected for further experiments on the real world dataset.

3.2. Result discussion on 15 scene categories

The result of applying spectral subspace clustering with PCA approach is shown in Table 2. We report maximum limits – 3, 5 and $n_i$ on selecting the number of eigenvectors to show the varying influence on the results. For a solid comparison, The results of applying K-means along with PCA approach are also reported. It can be seen that SLBF method is dominating the result when first few eigenvectors (3 or 5) are selected. Note that, if we select number of eigenvectors equal to the number of data points per cluster ($n_i$), the performance becomes very poor due to the bad approximation of the subspaces. In case of using supervised approach (see results in the Table 3), all of the three methods outperform the baseline classification accuracy by selecting maximum 5 eigenvectors empirically.

4. Conclusions and Future Work

We propose a novel framework for using spectral subspace clustering for visual dictionary creation. From the experimental results, it can be observed that the SLBF algorithm works consistently better with our proposed framework and outperforms the baseline classification accuracy of BOW model. However, PCA approach for estimating the subspace basis dimensions has the limitation of tuning the maximum limit on selecting eigenvectors empirically. As future work, we plan to investigate a good solution to estimate the original subspace dimensions during subspace clustering in a coupled manner. Another experiment can be performed on using bi-level approach by at first employing subspace clustering to get clusters with good density and
<table>
<thead>
<tr>
<th>No. of clusters</th>
<th>Max selected eigenvectors</th>
<th>Results (Spectral Subspace Clustering with PCA)</th>
<th>Kmeans with PCA</th>
<th>Baseline using K-means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SMCE</td>
<td>SSC</td>
<td>SLBF</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>71.05</td>
<td>70.38</td>
<td>73.76</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>71.52</td>
<td>69.47</td>
<td>73.90</td>
</tr>
<tr>
<td></td>
<td>$\alpha_i$</td>
<td>59.00</td>
<td>16.18</td>
<td>32.29</td>
</tr>
</tbody>
</table>

Table 2. Results on 15 scene dataset: Spectral Subspace Clustering with PCA approach (in %).

<table>
<thead>
<tr>
<th>No. of clusters</th>
<th>Max selected eigenvectors</th>
<th>Results (Supervised Spectral Subspace Clustering with PCA)</th>
<th>Kmeans with PCA</th>
<th>Baseline using K-means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SMCE</td>
<td>SSC</td>
<td>SLBF</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>72.80</td>
<td>72.76</td>
<td>73.34</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>73.13</td>
<td>73.06</td>
<td>73.91</td>
</tr>
<tr>
<td></td>
<td>$\alpha_i$</td>
<td>56.04</td>
<td>70.81</td>
<td>45.41</td>
</tr>
</tbody>
</table>

Table 3. Results on 15 scene dataset: Supervised Spectral Subspace Clustering with PCA approach (in %).

then using the K-means per cluster with the assumption of having multiple clusters within each subspace.

References