P. Bodesheim: Spectral Clustering of ROIs for Object Discovery. DAGM 2011, pp. 450–455.

Spectral Clustering of ROIs for Object Discovery

Paul Bodesheim

Chair for Computer Vision
Friedrich Schiller University of Jena
Paul.Bodesheim@uni-jena.de
http://www.inf-cv.uni-jena.de

Abstract. Object discovery is one of the most important applications of unsupervised learning. This paper addresses several spectral clustering techniques to attain a categorization of objects in images without additional information such as class labels or scene descriptions. Due to the fact that background textures bias the performance of image categorization methods, a generic object detector based on some general requirements on objects is applied. The object detector provides rectangular regions of interest (ROIs) as object hypotheses independent of the underlying object class. Feature extraction is simply constrained to these bounding boxes to decrease the influence of background clutter. Another aspect of this work is the utilization of a Gaussian mixture model (GMM) instead of k-means as usually used after feature transformation in spectral clustering. Several experiments have been done and the combination of spectral clustering techniques with the object detector is compared to the standard approach of computing features of the whole image.

1 Introduction and Related Work

Unsupervised image categorization for object discovery is a challenging task in computer vision. Algorithms try to group images according to categories of the pictured objects only using the visual content. This can be done by utilizing similarities between representations of images assuming that images containing objects of the same class provide similar feature vectors. A clustering of all vectors then implies a clustering of the corresponding images.

Commonly used approaches for object discovery include spectral clustering techniques, which are characterized later in this paper. A main part of those methods rely on graph partitioning based on optimizing the *Normalized Cut* [11]. Closely related to *Normalized Cuts Spectral Clustering* is a dimensionality reduction technique called *Laplacian Eigenmaps* [2], where at last, the same eigenvalue problem of the graph Laplacian as for the Normalized Cut optimization needs to be solved. A good overview of spectral clustering and graph Laplacians is provided by von Luxburg [5].

Another alternative to discover objects in images is the usage of *Topic models* [12, 13]. Object categories are determined by estimating the parameters of a

recommended for submission to YRF2011 by Prof. Dr.-Ing. Joachim Denzler

statistical model, which involves hidden (latent) topic variables [12]. Both approaches for object discovery, spectral clustering and topic modeling, are compared in Tuytelaars et al. [14]. In the present paper, we focus on spectral clustering techniques and present their combination with a general object detector.

2 Spectral Clustering Techniques

Spectral clustering techniques are methods that rely on the eigen-decomposition of a modified similarity matrix containing pairwise similarities of feature vectors [14]. Using the eigenvectors and eigenvalues of such matrices, feature vectors can be transformed by projections into a low-dimensional feature space prior to clustering.

In this section, four selected methods, which meet that definition of spectral clustering, are described briefly. They have in common that each of them uses pairwise similarities of feature vectors $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(M)} \in \mathbb{R}^N$ calculated by a kernel function κ and collected in a kernel matrix $\boldsymbol{K}$ with $\boldsymbol{K}_{ij} = \kappa\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right)$. Each method realizes a specific feature transformation and the transformed data points are always clustered using standard techniques. While k-means is usually applied, we use a GMM, which generalizes k-means by estimating arbitrary covariance matrices.

2.1 Nonlinear Component Analysis

Kernel methods treat the kernel matrix $\boldsymbol{K}$ as a matrix containing inner products of the feature vectors in a higher-dimensional space $\mathbb{F}$, which mostly depends on the input space in a nonlinear way. The following two approaches of nonlinear component analysis both project the data points on principal axes in $\mathbb{F}$ without computing vectors in this space, but they differ in the selection of the axes.

Kernel Principal Component Analysis (Kernel-PCA). For *Kernel-PCA*, those principal axes are chosen, which offer largest variance of data points in $\mathbb{F}$. Thus, Kernel-PCA is equal to standard PCA in this higher-dimensional space. As in standard PCA, a centering step is necessary to ensure centered data points in $\mathbb{F}$ [10]. The largest eigenvalues and corresponding eigenvectors of the centered kernel matrix $\bar{\boldsymbol{K}}$ solving $\bar{\boldsymbol{K}}\boldsymbol{v} = \lambda\boldsymbol{v}$ are required to compute transformed feature vectors $\tilde{\boldsymbol{x}}^{(1)}, \ldots, \tilde{\boldsymbol{x}}^{(M)}$ [10].

Kernel Entropy Component Analysis (Kernel-ECA). Using *Kernel-ECA* for feature transformation also results in computing projections of data points on principal axes. Different to Kernel-PCA, the eigenvectors are not chosen according to the largest eigenvalues of the centered kernel matrix, but with respect to their contribution to an approximation of the quadratic Renyi entropy [8] $\mathrm{H}\left(p\right) = -\log \int p^2\left(\boldsymbol{x}\right) \mathrm{d}\boldsymbol{x}$. As stated in [4], the aim is to select principal axes with highest contributions to this entropy. The contribution of the m-th principal axis to an approximation of this entropy is $c_m = \left(\sqrt{\lambda_m}\mathbf{1}^\top \boldsymbol{v}^{(m)}\right)^2$ with eigenvalue λ_m and the corresponding eigenvector $\boldsymbol{v}^{(m)}$ of $\boldsymbol{K}$. Compared to Kernel-PCA, there is no centering step of the kernel matrix involved [4].

2.2 Normalized Cuts Spectral Clustering

For Normalized Cuts Spectral Clustering, a weighted and undirected graph is constructed treating feature vectors as vertices and pairwise similarities as edge weights. Thus, it is possible to use the kernel matrix $\boldsymbol{K}$ to represent a full graph. The two methods of this section optimize the Normalized Cut [11] of the graph determined by $\boldsymbol{K}$. In general, the optimization can be done by minimizing the Rayleigh quotient $\frac{\boldsymbol{y}^\mathsf{T}(\boldsymbol{D}-\boldsymbol{K})\boldsymbol{y}}{\boldsymbol{y}^\mathsf{T}\boldsymbol{D}\boldsymbol{y}}$, which ends in computing eigenvectors according to the smallest eigenvalues of the generalized eigenproblem $(\boldsymbol{D}-\boldsymbol{K})\,\boldsymbol{y} = \lambda\boldsymbol{D}\boldsymbol{y}$, where $\boldsymbol{D}$ is a diagonal matrix containing row sums of $\boldsymbol{K}$ [11]. The eigenvalue λ is equal to the Normalized Cut with respect to $\boldsymbol{y}$, which in theory is a binary vector describing the corresponding bipartition of the graph.

Random Walks Laplacian Eigenmaps (Random Walks LEM). The work of Meila and Shi [6] gives an interpretation of spectral partitioning with random walks using the stochastic matrix $\boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{K}$. Instead of the generalized eigenproblem, they solve $\left(\boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{K}\right)\boldsymbol{y} = \lambda\boldsymbol{y}$, with $\boldsymbol{I}$ the identity matrix, by computing eigenvectors of $\boldsymbol{P}$ according to the largest eigenvalues. Since eigenvectors of $\boldsymbol{P}$ are also solutions for the generalized eigenproblem [5], these eigenvectors minimize the Normalized Cut as well. Forming a matrix $\tilde{\boldsymbol{X}}$ containing the eigenvectors of $\boldsymbol{P}$ in its columns, the rows of $\tilde{\boldsymbol{X}}$ represent the transformed feature vectors $\tilde{\boldsymbol{x}}^{(1)}, \ldots, \tilde{\boldsymbol{x}}^{(M)}$. Because of the strong connection between Normalized Cuts Spectral Clustering and Laplacian Eigenmaps (cf. Sect. 1) as well as the random walks point of view [6], this method is termed *Random Walks Laplacian Eigenmaps (Random Walks LEM)* throughout this paper.

NJW-Algorithm. The *NJW-Algorithm* [7] uses eigenvectors of the normalized Laplacian matrix $\boldsymbol{L} = \boldsymbol{D}^{-\frac{1}{2}}\,(\boldsymbol{D}-\boldsymbol{K})\,\boldsymbol{D}^{-\frac{1}{2}}$ by computing eigenvectors according to the largest eigenvalues of $\tilde{\boldsymbol{L}} = \boldsymbol{I} - \boldsymbol{L}$. Compared to Random Walks LEM, this leads to scaled eigenvectors $\boldsymbol{z} = \boldsymbol{D}^{\frac{1}{2}}\boldsymbol{y}$ [11]. Transformed feature vectors are computed as done in the algorithm called Random Walks LEM, but with an additional normalization of the rows of $\tilde{\boldsymbol{X}}$ having unit length [7].

3 Object Detection and Categorization of ROIs

As in [14], feature extraction is often performed on the whole image. To avoid clusterings based on background textures, it is desirable to compute features only at regions, which are covered by an object. The key idea of this paper is to integrate a general object detector into an unsupervised learning framework for object discovery. For this purpose, the general object detector of Alexe et al. [1] is applied to generate bounding boxes as object hypotheses independent of the object's class and feature extraction can be limited to these rectangular areas. Using this detector, we get an arbitrary number of bounding boxes, each of them having a score between 0 and 1 measuring how likely the rectangle contains an object of any class. The scoring and thus the detector works generic across categories using some object cues such as closed contour and color contrast [1].

At first glance, applying this detector is not possible in an unsupervised framework, because the detector needs to be trained with images and ground-truth-information about ROIs. But if the training images are completely independent of the clustered images, there is no information utilized about the latter. So, when we use the detector with the default parameter setting, which comes with the software of [1] and whose values are obtained using images containing objects of classes different to those that should be discovered, it can be seen as an unsupervised scenario as well.

First Approach: One ROI per Image. In a first approach, we sample a fixed number of ROIs for every image, but using only the ROI of each image with the highest score given by the detector. Feature extraction, transformation and clustering is simply done for those ROIs and the category label of one ROI directly specifies the label of a single image.

Second Approach: Multiple ROIs per Image. The second approach employs the idea of [9] for object discovery, where multiple segmentations of each image are used with the assumption that at least one segment covers one single object in a sufficient way. In the case of ROIs, assuming that at least one ROI is a good bounding box for an object in the image, multiple ROIs per image are sampled at the beginning, e.g. b ROIs with highest score. Subsequent, feature extraction is performed on all ROIs as well as feature transformation and clustering. In the end, there are b labels for each image, one per ROI. To avoid images with multiple labels and to compare the results with the first approach, it is necessary to have one label for each image. Using a GMM for clustering, one can determine a single ROI per image, which has the highest probability for being a member of the specific category and the image is assigned to the label of this ROI.

4 Experimental Results

In experiments, all images of 20 object categories of the *Caltech-256* dataset selected by [14] are grouped. *PHOG* features [3] as well as the χ^2-kernel [14] are applied, and also a kernel particular for *PHOG* similarity [3], which we term PHOG-kernel. As proposed in [14], the conditional entropy is measured to evaluate a clustering. A low conditional entropy corresponds to a high quality of the clustering.

In Fig. 1, the conditional entropy of achieved clusterings is displayed depending on the dimension of the transformed feature vectors, where $\langle \mathit{IMAGE} \rangle$ stands for feature calculation on the whole image, $\langle \mathit{1\ ROI} \rangle$ for applying the first approach proposed in Sect. 3 and $\langle \mathit{1\ of\ 10\ ROIs} \rangle$ for the usage of ten ROIs per image selecting the best one as described in the second approach. It can be seen clearly that only using one ROI per image leads to poor clusterings according to the conditional entropy, whereas multiple ROIs show better performance. As stated before, we also calculated features on the whole image. Indeed this leads

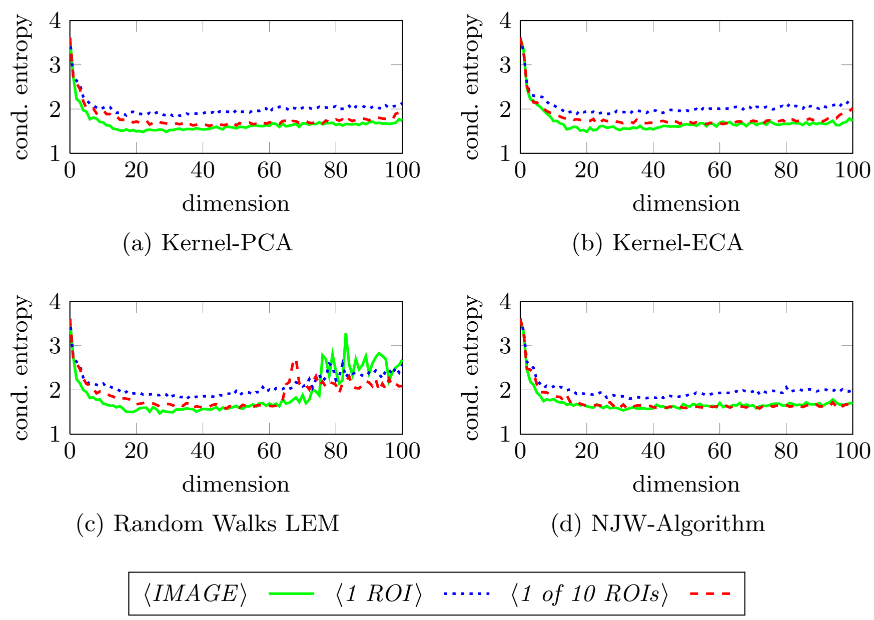

Fig. 1. Conditional entropy of the clusterings depending on the dimension of the transformed feature vectors (number of eigenvectors used for transformation), obtained by four spectral clustering techniques (a)–(d) and three mentioned approaches using PHOG-kernel [3] and a GMM (best viewed in color)

to the best results, but especially the NJW-Algorithm produces nearly equal outputs comparing ⟨*IMAGE*⟩ and ⟨*1 of 10 ROIs*⟩. For clarity and due to the lack of space, Fig. 1 only shows the results obtained by the GMM since in further experiments, clusterings using k-means achieve a higher conditional entropy.

In comparison to the results of [14] using 20 eigenvectors for feature transformation, the clusterings are better for ⟨*IMAGE*⟩, notably obtained by Kernel-PCA with a conditional entropy of 1.55 (cf. Table 1). Also ⟨*1 of 10 ROIs*⟩, using twice the number of eigenvectors, because there is an additional performance gain for a dimension higher than 20, achieves good results near the lower bound given by the intervals of [14].

5 Conclusions

The presented results show the ability of applying a general object detector in an unsupervised object discovery framework, where the usage of multiple ROIs per image leads to better performance. Although the proposed method of spectral clustering of ROIs does not provide a clear quantitative performance benefit, our approach of first detecting an object in general and subsequent discovering the category is promising and improvements should be aspired in further work.

Table 1. Conditional entropy of the clusterings with χ^2-kernel compared to results of [14], where different features are evaluated (that's why there are intervals denoted)

Spectral Clustering Technique	⟨*IMAGE*⟩ (20 Eigenvectors)	⟨*1 of 10 ROIs*⟩ (40 Eigenvectors)
Kernel-PCA & GMM	1.55	1.61
Kernel-ECA & GMM	1.60	1.62
Random Walks LEM & GMM	1.56	1.67
NJW-Algorithm & GMM	1.61	1.66
Kernel-PCA & k-means [14]	1.64 − 2.35	–
NJW-Algorithm & k-means [14]	1.58 − 2.54	–

In our studies, it turned out that a GMM for grouping transformed feature vectors, compared to commonly used k-means, boosts the quality of categorizations obtained by spectral techniques.

Acknowledgements. I want to thank Erik Rodner, advisor of my diploma thesis, for his great support and Michael Kemmler for helpful suggestions.

References

1. Alexe, B., Deselaers, T., Ferrari, V.: What is an object? In: CVPR. pp. 73–80 (2010)
2. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation 15(6), 1373–1396 (2003)
3. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: CIVR. pp. 401–408 (2007)
4. Jenssen, R.: Kernel entropy component analysis. TPAMI 32(5), 847–860 (2010)
5. von Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing 17(4), 395–416 (2007)
6. Meila, M., Shi, J.: Learning segmentation by random walks. In: NIPS. pp. 873–879 (2000)
7. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: NIPS. vol. 14, pp. 849–856 (2002)
8. Renyi, A.: On measures of entropy and information. In: Proceedings of the Berkeley Symposium on Mathematics, Statistics and Probability. vol. 1, pp. 547–561 (1960)
9. Russell, B.C., Freeman, W.T., Efros, A.A., Sivic, J., Zisserman, A.: Using multiple segmentations to discover objects and their extent in image collections. In: CVPR. vol. 2, pp. 1605–1614 (2006)
10. Schölkopf, B., Smola, A., Müller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10(5), 1299–1319 (1998)
11. Shi, J., Malik, J.: Normalized cuts and image segmentation. TPAMI 22(8), 888–905 (2000)
12. Sivic, J., Russell, B.C., Efros, A.A., Zisserman, A., Freeman, W.T.: Discovering objects and their location in images. In: ICCV. vol. 1, pp. 370–377 (2005)
13. Sivic, J., Russell, B.C., Zisserman, A., William T. Freeman, Efros, A.A.: Unsupervised discovery of visual object class hierarchies. In: CVPR. pp. 1–8 (2008)
14. Tuytelaars, T., Lampert, C.H., Blaschko, M.B., Buntine, W.: Unsupervised object discovery: A comparison. IJCV 88(2), 284–302 (2010)