Advanced Computer Vision — Review: 1/26

Murat Ambarkutuk {murata@vt.edu}

1 Fisher kernels on visual vocabularies for image categorization

Image categorization is the effort of extracting semantic level information from images and labeling them based on the extracted information. In their paper [1], the authors propose an approach in which Fisher Kernels (FK) and Gaussian Mixture Models (GMM) are utilizied to achieve image categorization problems. By doing so, the authors provide a unified framework for two different kinds of pattern classification methods, namely, generative and discriminative approaches. The intuition is to approximate low-level feature vectors extracted from the images with Gaussian functions, while representing them with FK.

The proposed approach contains two major steps in constructing the vocabulary: the approximation of the low-level feature vectors with a number of Gaussian functions, and the representation of the derived Gaussian mixtures with the gradients of log-likelihood functions. The first contribution of this paper is to employ the pipeline of GMM, and FK in the image categorization problems to provide a combined framework for two seperate approaches of categorization. Each N-dimensional feature vector is formulated as the sum of log-likelihood functions given the a set number of Gaussian functions. One can object the idea of representing any arbitrary distribution with a set number of Gaussian functions, given that one may end up having highly non-Gaussian (multi-modal) distributions. Given the influence of this parameter in the whole process, choosing a constant is not trivial. The second step of the approach is to calculate the gradient vectors of the log-likelihood functions of feature vectors. Along with utilization of FK in image categorization problems, the derivation of the closed-form formulation of the Fisher information matrix is the second contribution of the paper. The images were preprocessed before extracting features by resizing the images approximately to the same size. After that, the dimensionality of the feature vectors are reduced to 50. In the training (unsupervised) step Maximum Likelihood estimation was used to construct "universal vocabulary"[1], from which the class-vocabulary is established by using Maximum a Posteirori estimation.

The proposed method was compherensively tested with two different databases, in house and VOC 2006 [2], under supervised and unsupervised learning schemes with two different low-level feature vectors to validate the derived approach. Both of the dataset has more than 5000 images, for testing and training combined for less than 20 categories. Thus, it can be inferable that FK is a tailored solution for particular problems, given the need of amount of effort and data in the training process. The results provided by the authors validate the the overall performance of FK. Showing at least 10% better performance than median of the VOC 2006 applicants, FK seems to prove its success in categorization problems, which may be resulted from having higher dimensionality in representing the words. Along with its relatively higher succes, the computational efficacy of FK is higher by the factor of 25 than the conventional BOV representation, albeit the aforementioned dimensionality.

References

[1] Florent Perronnin and Christopher Dance. Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[2] M. Everingham, A. Zisserman, C. K. I. Williams, and L. Van Gool. The PAS-CAL Visual Object Classes Challenge 2006 (VOC2006) Results. http://www.pascal-network.org/challenges/VOC/voc2006/results.pdf.