FUZZY ENCODING FOR IMAGE CLASSIFICATION USING GUSTAFSON-KESSEL ALGORITHM

Ashish Gupta and Richard Bowden

Centre for Vision, Speech and Signal Processing
University of Surrey
Guildford, United Kingdom

ABSTRACT

This paper presents a novel adaptation of fuzzy clustering and feature encoding for image classification. Visual word ambiguity has recently been successfully modeled by kernel codebooks to provide improvement in classification performance over the standard 'Bag-of-Features'(BoF) approach, which uses hard partitioning and crisp logic for assignment of features to visual words. Motivated by this progress we utilize fuzzy logic to model the ambiguity and combine it with clustering to discover fuzzy visual words. The feature descriptors of an image are encoded using the learned fuzzy membership function associated with each word. The codebook built using this fuzzy encoding technique is demonstrated to provide superior performance over BoF. We use the Gustafson-Kessel algorithm which is an improvement over Fuzzy C-Means clustering and can adapt to local distributions. We evaluate our approach on several popular datasets and demonstrate that it consistently provides superior performance to the BoF approach.

Index Terms— Fuzzy Clustering, Image Classification, Gustafson-Kessel algorithm

1. INTRODUCTION

The traditional approach to image classification is, use of 'Bag-of-Features' (BoF) method to model occurrence statistics of visual words amongst the feature descriptors in an image. The visual words are representative feature descriptors and commonly equal to the centroid of clusters acquired by a vector quantization technique like k-means clustering. Subsequent developments pursued a search for an optimal set of visual words to model a visual category in images. At present, better results are obtained by increasing the number of visual words at the expense of high sparsity and computational cost. An alternative approach to building huge codebooks is the notion of 'visual word ambiguity' which has been recently introduced by Gemert et al. in [1]. The authors suggest that visual feature space is continuous and consequently a continuum exists between visual words. Therefore,
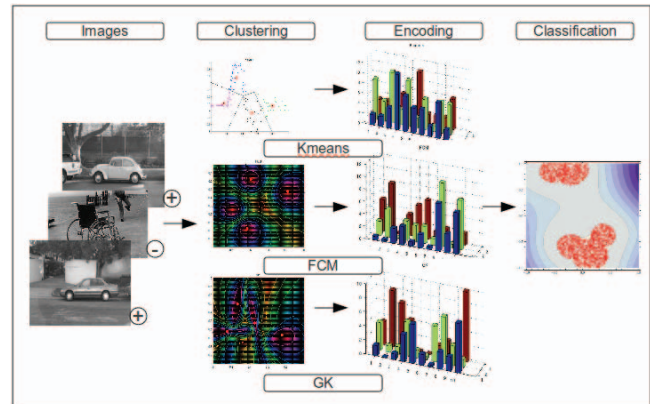
Fig. 1. **Fuzzy Encoding Schematic**: feature vectors from images are clustered to find cluster centroid and fuzzy membership function. An image is encoded based on the cumulative membership score of its descriptors. The encoded feature vector of the image with label is used to train a classifier.

a hard partitioning scheme utilized by BoF is not the best method to model this data. In BoF the assignment of feature descriptors to visual words follows a 'winner-takes-all' scheme which can also be considered as crisp-logic. Gemert et al. assign feature descriptors to visual words 'smoothly' to model the visual continuum using a kernel associated with each visual word. This work motivates our research on utilizing fuzzy logic based methods to model this proposed 'ambiguity' about visual words. In our approach, see figure 1, the feature descriptors are assigned to a visual word based on a fuzzy membership function associated with each visual word. We compute visual words using a fuzzy clustering algorithm, which were developed as a synthesis of fuzzy logic and clustering algorithms. The feature descriptors from an image are encoded using the fuzzy membership of each fuzzy visual word.

In this paper, we present a novel adaptation of the Gustafson-Kessel (GK) algorithm [2] for building a fuzzy visual codebook. GK algorithm belongs to the family of fuzzy clustering algorithms and can be considered an exten-

| (a) Kmeans | (b) Fuzzy C-Means | (c) Gustafson-Kessel |
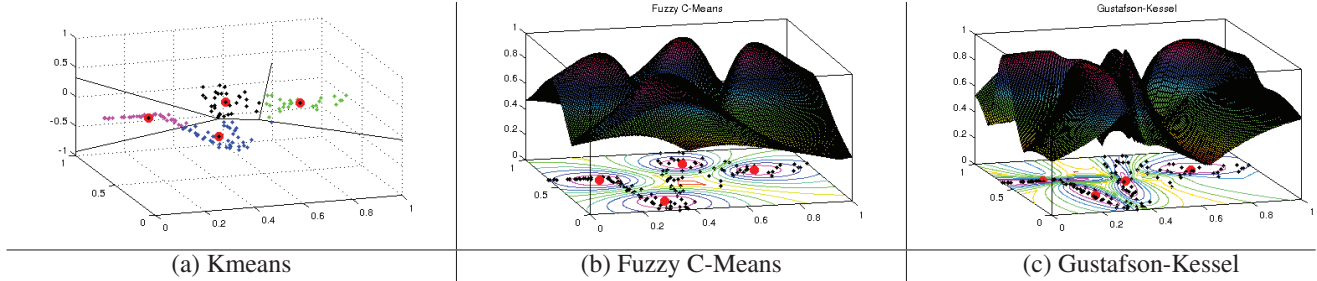
Table 1. Fuzzy Encoding

sion of the Fuzzy C-Means (FCM) algorithm [3]. We analyze the classification performance or our approach against the the standard BoF model on several popular datasets. In addition, we analyze the variation in performance for different sizes of visual codebook.

Our principal contribution is adapting the nascent work on visual word ambiguity, which used intuitively selected kernels, to the well developed field of fuzzy logic. We show that FCM based feature encoding provides improved performance. We build upon this by using GK algorithm which can adapt to local topology of the feature space.

2. SYSTEM

The typical dataset $\mathcal{D}$ utilized for image classification consists of multiple visual categories $\mathcal{C}$, each containing several positive labeled ($\gamma : \mathcal{I} \to 1$) training images $\mathcal{I}^{\oplus}$. The negative labeled ($\gamma : \mathcal{I} \to -1$) training images $\mathcal{I}^{\ominus}$ are culled from remaining categories in the dataset. Local affine invariant feature descriptor vectors x (ignoring image label), collated from a balanced sample of $\mathcal{I}^{\oplus}$ and $\mathcal{I}^{\ominus}$, are clustered using each of Kmeans, FCM, and GK clustering. For each image $\mathcal{I}_i \in \mathcal{D}$ in the dataset we encode the feature vectors of that image. The encoded feature vector ν_i and associated label γ_i are utilized to train a binary SVM classifier using an RBF kernel. The difference amongst the methods lies in the computed cluster centroids and the membership of a feature vector to each cluster.

3. FUZZY ENCODING

The Kmeans algorithm computes cluster centroids $\mathbf{V} = \{v_1, v_2, \ldots, v_K\}$ and assignment of feature vectors $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$ to these clusters by minimizing an objective function:

$$J(\mathbf{X}; \mathbf{V}) = \sum_{i=1}^{K} \sum_{k=1}^{N} \mathbb{1}_k^i \parallel x_k - v_i \parallel^2 \ , \mathbb{1}_k^i = \left\{ \begin{array}{ll} 1 & \text{if } x_k \in v_i \\ 0 & \text{otherwise} \end{array} \right. \tag{1}$$

Each feature vector x_k is assigned to exactly one cluster only, see table 1(a). The encoded feature for an image ν is a discrete valued histogram. In comparison, the FCM algorithm

attempts to minimize the objective function:

$$J(\mathbf{X}; \mathbf{U}, \mathbf{V}) = \sum_{i=1}^{K} \sum_{k=1}^{N} \mu_{ik}^m \parallel x_k - v_i \parallel^2 \ , 1 \le m < \infty \tag{2}$$

where $\mathbf{U} = [\mu_{ik}]$, and μ_{ik} is degree of membership of x_k to cluster i. m is a measure of fuzzification, where for $m = 1$, FCM is equivalent to kmeans. Using FCM, ν reflects a degree of membership of feature vectors to various clusters and consequently is able to model the ambiguity associated with visual words, as compared to Kmeans. However, FCM results in hyper-spherical clusters of equal volume, see table 1(b).

Algorithm 1 Gustafson-Kessel Fuzzy Encoding

$\tau \leftarrow 1$
repeat
$\quad v_i^{(\tau)} \leftarrow \frac{\sum_{k=1}^{N} (\mu_{ik}^{(\tau-1)})^m x_k}{\sum_{k=1}^{N} (\mu_{ik}^{(\tau-1)})^m}$
$\quad F_i \leftarrow \frac{\sum_{k=1}^{N} (\mu_{ik}^{(\tau-1)})^m (x_k - v_i^{(\tau)})(x_k - v_i^{(\tau)})^T}{\sum_{k=1}^{N} (\mu_{ik}^{\tau-1})^m}; 1 \le i \le K$
$\quad D_{ikA_i}^2 = (x_k - v_i^{(\tau)})^T [\rho_i \, det(F_i)^{\frac{1}{n}} F_i^{-1}](x_k - v_i^{(\tau)})$
$\quad \phi_k \leftarrow \{i \mid D_{ik} = 0\}$
$\quad$**for** $k \leftarrow 1, N$ **do**
$\quad\quad$**if** $\phi_k = \emptyset$ **then**
$\quad\quad\quad \mu_{ik}^{(\tau)} \leftarrow (\sum_{j=1}^{K} (\frac{D_{ikA_i}}{D_{jkA_j}})^{\frac{2}{m-1}})^{-1}$
$\quad\quad$**else**
$\quad\quad\quad \mu_{ik}^{(\tau)} \leftarrow \left\{ \begin{array}{ll} 0 & \text{if } D_{ikA_i} > 0 \\ \frac{1}{|\phi_k|} & \text{if } D_{ikA_i} = 0 \end{array} \right.$
$\quad\quad$**end if**
$\quad$**end for**
$\quad \tau \leftarrow \tau + 1$
until $\parallel U^{(\tau)} - U^{(\tau-1)} \parallel < \epsilon$
for $j \leftarrow 1, M$ **do**
$\quad \nu_j \leftarrow \sum \mu_{\mathbb{1}_j k}$
$\quad \gamma_j \leftarrow \left\{ \begin{array}{ll} 1 & \text{if } \mathcal{I}_j \in \mathcal{C} \\ -1 & \text{if } \mathcal{I}_j \notin \mathcal{C} \end{array} \right.$
end for

This simplistic model is unable to represent the complicated topology of visual data, where each cluster roughly corresponds to visually distinct object parts that have a unique

distribution in feature space. Therefore, we employ the GK algorithm with extends FCM, by replacing the Euclidean distance by another metric induced by a positive definite matrix A so that hyper-ellipsoidal clusters can be built instead of simple hyper-spherical clusters only, see table 1(c) [4]. The clusters have an associated norm inducing matrix $\mathbf{A} = \{A_1, A_2, \ldots, A_K\}$, which provides the inner product norm:

$$D_{ikA_i}^2 = (x_k - v_i)^T A_i (x_k - v_i) \, , \, 1 \le i \le K \, , \, 1 \le k \le N \tag{3}$$

The objective function minimized in GK is:

$$J(\mathbf{X}; \mathbf{U}, \mathbf{V}, \mathbf{A}) = \sum_{i=1}^{K} \sum_{k=1}^{N} \mu_{ik}^m D_{ikA_i}^2 \tag{4}$$

Due to the linear dependency between J and A, we constrain A to obtain a feasible solution to J. The determinant $\rho_i = det(A_i)$ is kept constant, which corresponds to optimizing cluster shape whilst keeping its volume constant. A_i can now be expressed as: $A_i = [\rho_i det(\mathbf{F}_i)]^{\frac{1}{n}} \mathbf{F}_i^{-1}$, where $\mathbf{F}_i$ is the fuzzy covariance matrix of the ith cluster.

Our approach to Gustafson-Kessel fuzzy encoding is provided in algorithm 1. The encoded features for each image and associated label $\{\nu, \gamma\}$ is used to train a classifier.

4. EXPERIMENTS

To demonstrate the efficacy of GK encoding technique we comparatively evaluate the classification performance of GK encoding against FCM encoding and the BoF technique. Since, the focus of the experiments is comparison of clustering and encoding techniques, the choice of the feature descriptor and the classifier utilized in the experiments was based on popularity rather then performance. The classifier utilized is SVM with an RBF kernel. We used the SIFT descriptor, which is the most popular member of the family of local affine invariant descriptors. While some other descriptors like LBP, SURF might provide a marginally better classification performance, the relative performance across the techniques being evaluated will remain unchanged. We utilize several popular datasets in the vision community for a comprehensive evaluation of our approach and compare it to the performance of the standard BoF technique. The datasets utilized are Caltech-101 [5]; Caltech-256 [6]; Pascal VOC 2006; Pascal VOC 2010 [7]; and Scene-15. These datasets vary in terms of number of categories; number of images within each category; visual domain of categories; inherent difficulty and together render a comprehensive evaluation of the techniques.

4.1. Performance across datasets

We compare the BoF, FCM, and GK approaches in terms of their classification performance across 5 datasets. The results
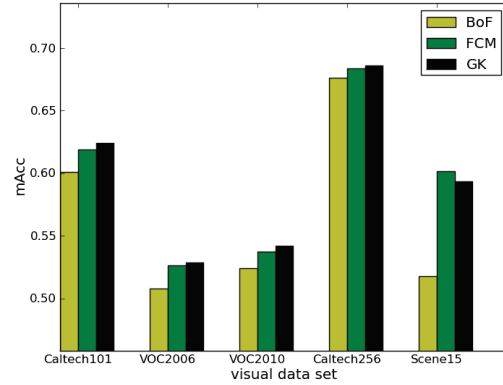
Fig. 2. Classification performance of GK,FCM-encoding and BoF for datasets {*Scene15*, *Caltech101*, *Caltech256*, *VOC2006*, *VOC2010* }.

in figure 2 show the mean accuracy of a cross-validated experiment averaged across all categories of the dataset. The performance varies across datasets due to the inherent difficulty of each dataset. The fuzzy encoding methods, FCM and GK, consistently outperform BoF on all datasets. This is sound empirical support for the efficacy of the fuzzy encoding approach. Note that GK performs marginally better than FCM on almost all the datasets. With optimization of the covariance matrix for each cluster in GK, we expect the margin of improvement provided by GK to improve further.

4.2. Performance across categories

The graphs in table 2 show the comparative performance of BoF and GK approaches for each visual category in the datasets: VOC2006, VOC2010, and Scene15. The absolute and relative performance of both BoF and GK varies across the categories due to the variation in content and complexity of each category. GK is observed to consistently out perform BoF in almost all categories.

4.3. Visual dictionary size

In this experiment, we analyze the relation between the visual dictionary size or the number of fuzzy visual words and classification performance of the encoding techniques considered. The set of codebook sizes considered is {16, 32, 64, 128, 256, 512 }, and the dataset utilized is Caltech101. The graph present the average accuracy across all categories in the dataset for each codebook size. GK performs better than BoF for all dictionary sizes. In addition, there is a gradual improvement in performance as dictionary size increases.
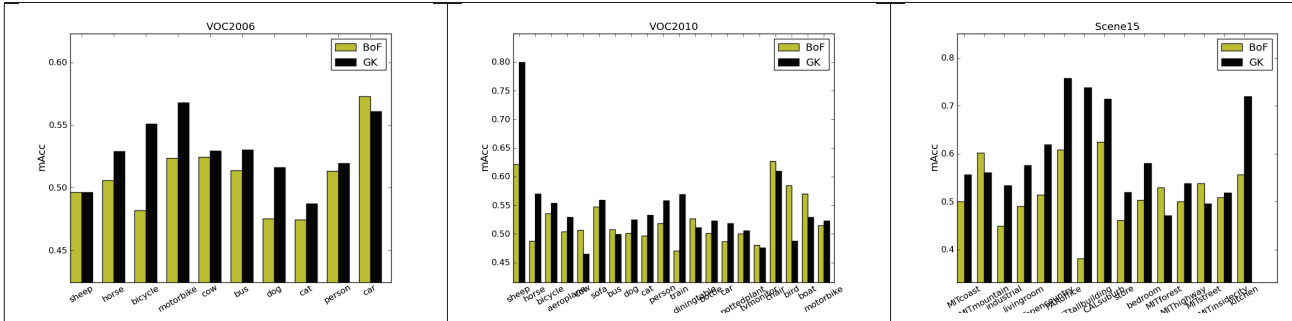
Table 2. The classification performance of BoF, FCM, and GK approaches across the visual categories in a dataset for the VOC2006, VOC2010, and Scene15 datasets. The graphs shows the mean accuracy using a cross validation scheme. The categories are shown on the x-axis. The yellow bar corresponds to the BoF approach while the black bars correspond to the GK approach.
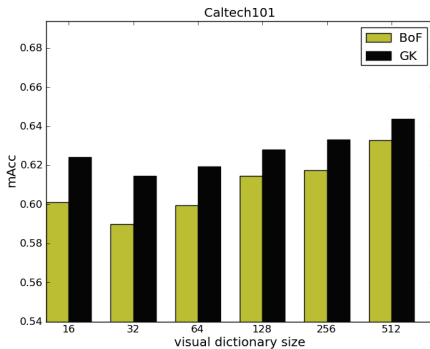
Fig. 3. Analysis of variation in classification performance with codebook size for GK,FCM-encoding and BoF techniques. The dataset utilized is Caltech101. The codebook sizes considered are {16, 32, 64, 128, 256, 512 }.

5. CONCLUSION

We have introduced fuzzy encoding technique for visual categorization using Fuzzy C-Means to compute a fuzzy membership function. We extended this work to the Gustafson-Kessel fuzzy clustering algorithm, which was shown to adapt to local distributions. We demonstrated empirically that our fuzzy encoding approach is consistently better than the BoF model, using several popular datasets. GK algorithm was shown to provide a marginal improvement over FCM, which we expect to improve with optimization of covariance matrices in future.

6. REFERENCES

[1] J.C. van Gemert, C.J. Veenman, A.W.M. Smeulders, and J.-M. Geusebroek, "Visual word ambiguity," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 7, pp. 1271 –1283, july 2010.

[2] Donald E. Gustafson and William C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on*, jan. 1978, vol. 17, pp. 761 –766.

[3] A. Baraldi and P. Blonda, "A survey of fuzzy clustering algorithms for pattern recognition. i," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 29, no. 6, pp. 778 –785, dec 1999.

[4] R. Babuka, P.J. van der Veen, and U. Kaymak, "Improved covariance estimation for gustafson-kessel clustering," in *Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on*, 2002, vol. 2, pp. 1081 –1085.

[5] L. Fei-Fei, R. Fergus, and Pietro Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," 2004.

[6] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," Tech. Rep. 7694, California Institute of Technology, 2007.

[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, 2010.