Hierarchical Classification of Images by Sparse Approximation

Byung-soo Kim,Jae Young Park,Anna C. Gilbert,Silvio Savarese University of Michigan, Ann Arbor, USA
bsookim@umich.edu

Using image hierarchies for visual categorization has been shown to have a number of important benefits [4]). However, a critical question still remains unanswered: would structuring data in a hierarchical sense also help classification accuracy? In this paper we address this question and show that the hierarchical structure of a database can indeed be used successfully to enhance classification accuracy using a sparse approximation framework.

The key idea is to introduce a distance metric that encode the hierarchy of the visual categories and define two images to be similar if they share a similar path in the hierarchy. This allows to cast the categorization problem as the one of discovering the category in the tree structure that has the smallest distance from the query category label. To solve this problem, we introduce a new method called *hierarchical* sparse approximation which enforces that the solution must be few paths out of all possible paths on a given hierarchy of object classes (training set).

Classification Problem. We assume that the query image contains few dominant object(s). The classification problem can be solved by seeking the one in the database that is closest to the query object(s). If the query image contains multiple objects, the classifier must return multiple category labels associated to all of them.

Object representation and distance function. We describe an image as x using a normalized histogram of codewords (i.e., the bag of words representation, also named BOW) [3], where the size of the dictionary is denoted by K; thus, x is a column vector of size K. The similarity between two images represented by x_i and x_j can be measured in l_n norm.

Model matrix. Let us stack all the images in the database in a matrix H. Columns of H will correspond to column vectors x. Thus, H will be $K \times N$, where N is the number of images in the dataset. We call this matrix H the *flat model matrix*. Any query image can be represented as a superposition of images in the training data with small error e such that $x = Hm + e$, where $N \times 1$ vector m is called the *mixing vector* and consists of a few non-zero entries associated to the images in the database that contribute to represent the query image by superposition. The error e captures background clutter and the intra-class variability.

Classification Framework. The classification problem (*what is the object class?*) is recasted into the problem of estimating the vector m (*where is a nonzero entry?*). This formulation also allows us to discover multiple dominant object categories in the image. Solving m is challenging because the system is under-determined ($N \gg K$). Because we postulate or seek a s-sparse mixing vector m, we find the sparsest solution that best approximates (in ℓ_0 error) the observed instance.

Problem 0. $\min \|m\|_0 \quad \text{subject to} \quad \|Hm - x\|_2 \leq \varepsilon$.

Unfortunately, the above problem is an NP-hard problem in general. We can, however, solve this problem in polynomial time with appropriate geometric assumptions on H. As [6], one method is to observe that Problem 0 is an optimization problem with a non-convex objective function and that a convex relaxation of this problem yields a problem which can be solved efficiently with standard optimization techniques [2],

Problem 1. $\min \|m\|_1 \quad \text{subject to} \quad \|Hm - x\|_2 \leq \varepsilon$.

Hierarchical Classification with Sparse Approximation However, the model $x = Hm + e$ fails to take into account any hierarchical information amongst the classes. Furthermore, the error metrics for typical sparse approximation algorithms [2] do not take into account structural relationships amongst the columns of H. Indeed, a small error in the mixing vector $\|\widehat{m} - m\|_2$ or in the reconstruction of the observation x does not necessarily guarantee hierarchical similarity between $\widehat{m}$ and m.

Hierarchical embedding. Assume that object categories are structured in a (rooted, labeled, recursive) tree $\mathscr{T}'$. The encoding matrix E is constructed so as to map the mixing vector m into an embedded mixing vector $\ell = Em$, whose non-zero entries correspond to the paths in $\mathscr{T}'$ from the image to the root of the tree. Note that $x = Hm + e = \Phi\ell$, if we define $\Phi = HE^{\dagger}$. Details to construct E and Φ can be found from the paper.

Hierarchical sparse approximation. The hierarchical embedding allows us to reformulate Problem 1 as a hierarchical sparse approximation problem and find a solution for ℓ given x:

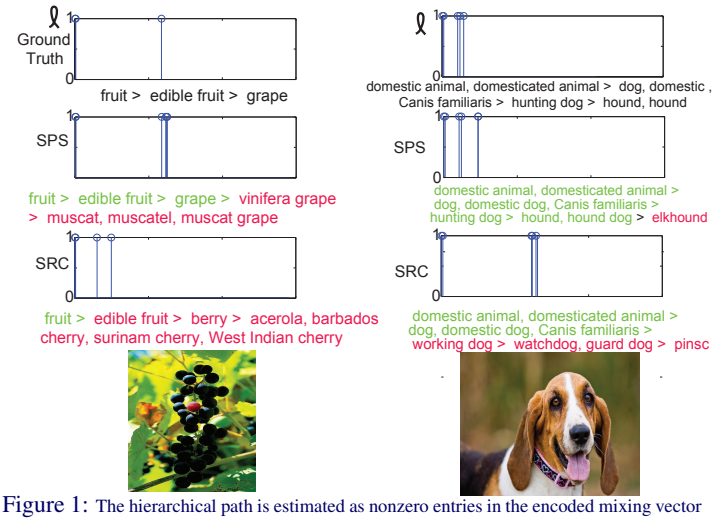

Figure 1: The hierarchical path is estimated as nonzero entries in the encoded mixing vector ℓ. Note that the path estimated by SPS (ours) is closer to the ground truth path than that of SRC. Green (red) indicates correct (incorrect) classification.

Problem 2. $\min \|\ell\|_1 \quad \text{subject to} \quad \|\Phi\ell - x\|_2 \leq \varepsilon$

The sparsity pattern of the vector ℓ is constrained to lie on a single path (or subtree) of the tree $\mathscr{T}'$. It also enforces a model that these non-zero entries must follow; they must lie on paths from individual columns of H to the root of the tree $\mathscr{T}'$. This problem has the structure of a model-based compressive sensing problem [1], and can be solved efficiently by a greedy algorithm called TREE-OMP [5] or MODEL-COSAMP [1].

Theorem 1 *Given a normalized test image x ($\|x\|_2 = 1$) which is sd-sparse with background "noise" n, we can solve $\Phi\ell = x + n$ for the embedded mixing vector ℓ with* TREE-OMP. *After $T > \log(sd)$ iterations, the output vector $\widehat{\ell}$ has at most Td non-zero entries and satisfies*

$$\|\ell - \widehat{\ell}\|_2 \leq 2^{-T} + C\|n\|_2.$$

Sparse Path Selection Algorithm (SPS). Suppose we obtain an estimate of the path ℓ associated to the query image. However, ℓ cannot be used as is for image classification because. Ideally, the sparsest solution of Problem 2 should return a vector of "1" and "0" where the non-zero elements in ℓ allows to estimate the category labels of the query object as well as its parents. Unfortunately, this is not always the case and values between "0" and "1" can be also found because of the estimation noise. To solve this issue, we introduced MAP based post processing step, and called this algorithm SPS. Fig. 1 shows anecdotal examples from SPS and compared with SRC [6].

Conclusion In this work, we introduced a novel framework for hierarchical classification using a new formulation of the sparse approximation problem. We demonstrated that the hierarchical structure of a large and complex database can indeed be used successfully to enhance classification accuracy. Experimental results on several large scale dataset were used to support our claims.

[1] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde. Model-based compressive sensing. *Information Theory*, 2010.

[2] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. In *SIAM review*, 2001.

[3] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop in ECCV*, 2004.

[4] G. Griffin and P. Perona. Learning and using taxonomies for fast visual categorization. In *CVPR*, 2008.

[5] C. La and M.N. Do. Tree-based orthogonal matching pursuit algorithm for signal reconstruction. In *ICIP*, 2006.

[6] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma. Robust face recognition via sparse representation. 2009.