Sparse Multiscale Patches (SMP) for Image Categorization

Paolo Piro, Sandrine Anthoine, Eric Debreuve, Michel Barlaud

Sparse Multiscale Patches (*SMP*) for image categorization

Paolo Piro, Sandrine Anthoine, Eric Debreuve, and Michel Barlaud

University of Nice-Sophia Antipolis / CNRS

Abstract. In this paper we address the task of image categorization using a new similarity measure on the space of Sparse Multiscale Patches (*SMP*). *SMP*s are based on a multiscale transform of the image and provide a global representation of its content. At each scale, the probability density function (*pdf*) of the *SMP*s is used as a description of the relevant information. The closeness between two images is defined as a combination of Kullback-Leibler divergences between the *pdfs* of their *SMP*s.
In the context of image categorization, we represent semantic categories by prototype images, which are defined as the centroids of the training clusters. Therefore any unlabeled image is classified by giving it the same label as the nearest prototype. Results obtained on ten categories from the Corel collection show the categorization accuracy of the *SMP* method.

1 Introduction

Image categorization is still one of the most challenging tasks in computer vision. It consists in labeling an image according to its semantic category. The main difficulty lies in using low-level information provided by digital images to retrieve semantic-level classes, which are generally characterized by high intra-variability.

Most approaches address the task of image categorization as a supervised learning problem. They use a set of annotated images to learn the categories and then assign to an unlabeled image one of these categories. These methods rely on extracting visual descriptors, which are to be highly specific, i.e. able to highlight visual patterns that characterize a category. Providing suitable sets of visual descriptors has been a topic of active research in the recent years and several approaches have been proposed.

1.1 Related works

Visual descriptors for image categorization generally consist of either global or local features. The former ones represent global information of images and are based on global image statistics such as color histograms [1] or edge directions histograms [2]. Global feature-based methods were mostly designed to separate very general classes of images, such as indoor vs. outdoor scenes or city vs. landscape images. On the contrary, local descriptors extract information at specific

image locations that are relevant to characterize the visual content. Local approaches are more adapted to cope with the intra-class variability of real image categories. Indeed these techniques are able to emphasize local patterns, which images of the same category are expected to share.

Early approaches based on local features work on image blocks. In [3] color and texture features are extracted from image blocks to train a statistical model; this model takes into account spatial relations among blocks and across image resolutions. Then several region-based approaches have been proposed, which require segmentation of images into relevant regions. E.g. in [4], an algorithm for learning region prototypes is proposed as well as a classification of regions based on Support Vector Machines (SVMs). More recent techniques have successfully used bags-of-features, which collect local features into variable length vectors. In [5], the bags-of-features representing an image are spatial pyramid aggregating statistics of local features (e.g. "SIFT" descriptors); this approach takes into account approximate global geometric correspondences between local features. Other methods using bags-of-features are based on explicitly modeling the distribution of these vector sets. In fact, measuring the similarity between the bags-of-features' distributions is the main difficulty for this kind of categorization methods. For example, Gaussian Mixture Models (GMMs) have been used to model the distribution of bags of low-level features [6]. This approach requires both to estimate the model parameters and to compute a similarity measure to match the distributions.

Apart from what descriptors are used, computing a similarity measure between feature sets is crucial for most approaches. Measuring similarities is particularly adapted to categorization when one prototype of each category is defined in the feature space. In this context the similarity measure is used to find the prototype that best matches an unlabeled image. As pointed out in [7], this framework has provided the best results in image categorization. The main reason is that categorization based on similarity corresponds well to the way human beings recognize visual classes. Indeed human visual categories are mostly defined by similarity to prototype examples, as it results from research on cognitive psychology [8]. Taking advantage of these results, we propose a new categorization technique that is based on category prototypes and uses a similarity measure between feature set distributions.

1.2 Proposed statistical approach

The categorization method that we propose consists of two steps: the training step and the classification step. The training consists in selecting one prototype per category among a set of labeled images (training set). The classification step assigns to an unlabeled image (query image) the label of its closest prototype. Both training and query images are represented by their sets of Sparse Multiscale Patches (*SMP*s).

We have designed the *SMP* descriptors in order to exploit local multiscale properties of images, which generally convey relevant information about their category. Indeed *SMP*s describe local spatial structures of images at different

scales by taking into account dependencies between multiscale coefficients across scale, space and color channels. An image is represented by the set of its *SMP*s at each scale and we measure the similarity between two images as the closeness between their *SMP* probability density functions (*pdf*). Namely we combine their Kullback-Leibler divergence, which has already shown good performances in the context of image retrieval [9].

We use this *SMP* similarity measure in both steps of the categorization, as depicted in the block diagram of Figure 1. During the training step we compute all pairwise *SMP*-based "distances" between images of the same training category C, thus obtaining a distance matrix for this category (see left column of Fig. 1). The entry $s_{i,j}$ of this matrix represents the distance between the image I_i (query image) and the image I_j (reference image); we denote this distance as $S(I_i|I_j)$. The distance matrix is used to select the prototype J_C, which we define as the image minimizing the distance to all other images of the same category:

$$J_C = \underset{j|I_j \in C}{\operatorname{argmin}} \sum_{i|I_i \in C} s_{i,j} \tag{1}$$

Applying the same method to all training categories yields one prototype for each category in terms of *SMP* feature sets.[1]

In the second phase, which is the categorization, we compute the *SMP* similarity of an unlabeled image Q to all category prototypes (see right column of Fig. 1). The query image is given the same label as the nearest prototype.

1.3 Organization of the paper

In the rest of this paper we explain in more details how the *SMP* similarity measure is defined and give some experimental results of categorization. The *SMP* similarity measure is described in two steps. Firstly we define the proposed feature set of Sparse Multiscale Patches in Section 2. Secondly we define the similarity between *SMP* probability densities in Section 3. We also propose a method for estimating this measure non-parametrically, thus avoiding to model the underlying *pdf*s. Finally, in Section 4, we present some results of performing our categorization method on a subset of the Corel database.

2 Feature space: Sparse Multiscale Patches

Les us now define our feature space, which is based on a multiresolution decomposition of the images. Throughout the paper, we will denote by $w(I)_{j,k}$ the coefficient for image I at scale j and location in space k for a general multiresolution decomposition.

[1] Note that the proximity matrix is not symmetric here $s_{i,j} \neq s_{j,i}$.

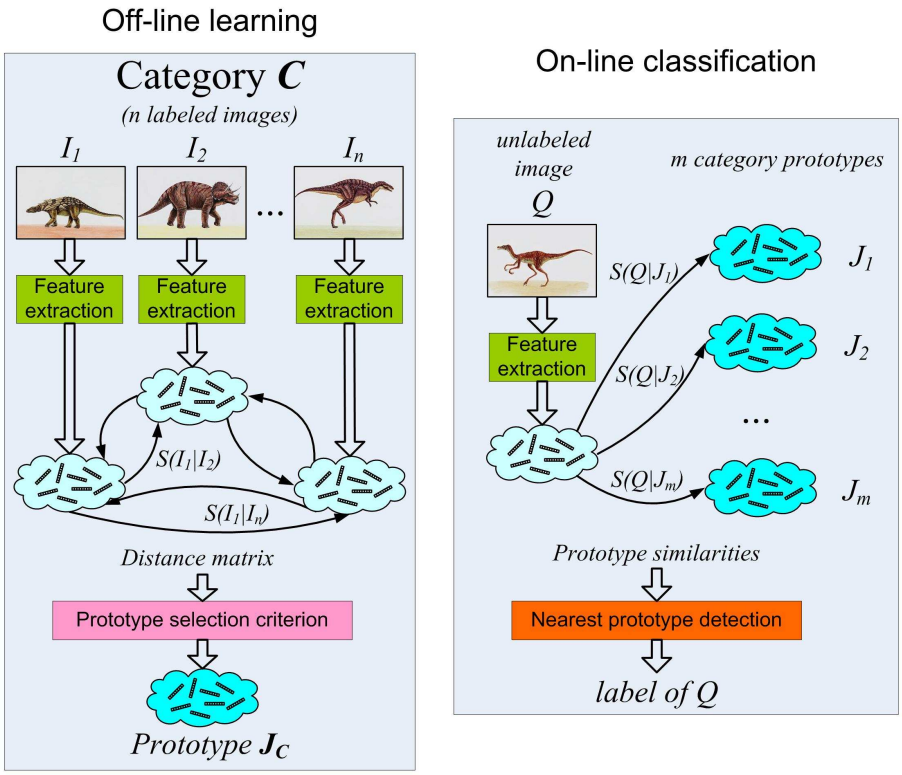

Fig. 1. Overview of the proposed method for image categorization.

2.1 Multiscale patches

The wavelet transform enjoys several properties that have made it quite successful in signal processing and it naturally yields good candidates for image descriptors. Indeed, it provides a sparse representation of images, meaning that it concentrates the informational content of an image into few coefficients of large amplitude while the rest of the coefficients are small. Classical wavelet methods focus on these large coefficients and treat them separately, relying on their decorrelation, to efficiently process images. However, wavelet coefficients are not independent and these dependencies are the signature of structures present in the image. These dependencies have then been exploited in image enhancement (e.g [10, 11]). In particular, the authors of [10] introduced the concept of patches of wavelet coefficients (called "neighborhoods of wavelet coefficients") to represent efficiently fine spatial structures in images.

Following these ideas, we define a feature space based on a sparse description of the image content by a multiresolution decomposition. More precisely, we group the Laplacian pyramid coefficients of the three color channels of image

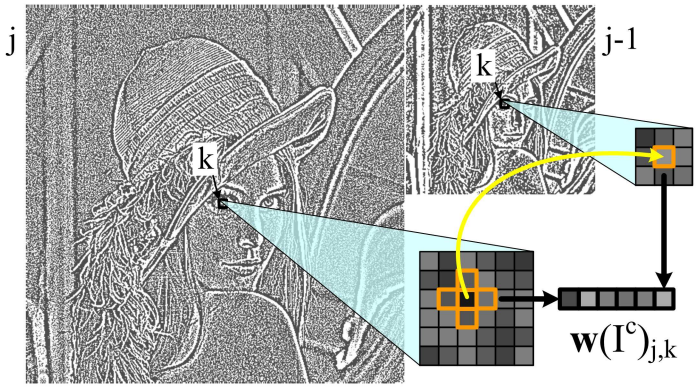

Fig. 2. Building a patch of multiscale coefficients, for a single color channel image.

I into coherent sets called patches. Here the coherence is sought by grouping coefficients linked to a particular scale j and location k in the image.

In fact, the most significant dependencies are seen between a coefficient $w(I)_{j,k}$ and its closest neighbors in space: $w(I)_{j,k\pm(0,1)}$, $w(I)_{j,k\pm(1,0)}$ and in scale: $w(I)_{j-1,k}$, where scale $j-1$ is coarser than scale j. Grouping the closest neighbors in scale and space of the coefficient $w(I)_{j,k}$ in a vector, we obtain the patch $\overrightarrow{w}(I)_{j,k}$ (see Fig. 2):

$$\overrightarrow{w}(I)_{j,k} = \left(w(I)_{j,k}, w(I)_{j,k\pm(1,0)}, w(I)_{j,k\pm(0,1)}, w(I)_{j-1,k}\right) \tag{2}$$

which describes the structure of the grayscale image I at scale j and location k. The probability density functions of such patches at each scale j has proved to characterize fine spatial structures in grayscale images [10, 12].

We consider color images in the luminance/chrominance space: I=(I^Y,I^U,I^V). Since the coefficients are correlated through channels, we aggregate in the patch the coefficients of the three channels:

$$\mathbf{w}(I)_{j,k} = \left(\overrightarrow{w}(I^Y)_{j,k}, \overrightarrow{w}(I^U)_{j,k}, \overrightarrow{w}(I^V)_{j,k}\right) \tag{3}$$

The low-frequency approximation that results from the Laplacian pyramid is also used to build additional feature vectors. The 3×3 pixel neighborhoods along the three channels are joined to form patches of dimension 27 (whereas patches from the higher-frequency subbands defined in Eq.(3) are of dimension 18). The union of the higher-frequency and low-frequency patches forms our feature space. For convenience, the patches of this augmented feature space will still be denoted by $\mathbf{w}(I)_{j,k}$.

2.2 Sparse Multiscale Patches

The coefficients are obtained by a Laplacian pyramid decomposition [13]. Indeed, critically sampled tensor wavelet transforms lack rotation and translation

invariance and so would the patches made of such coefficients. Hence we prefer to use the Laplacian pyramid which shares the sparsity and inter/intrascale dependency properties with the wavelet transform while being more robust to rotations.

Multiscale coefficients provide a sparse representation of images and, similarly, patches of multiscale coefficients of large overall energy (sum of the square of all coefficients in a patch) also concentrate the information. Since the total number of patches in an image decomposition is 4/3N with N the number of pixels in the image, the number of samples we have in the feature space is quite large as far as measuring a similarity is concerned. The possibility of selecting a small number of patches which represent the whole set well is therefore highly desirable. In practice, we selected a fixed proportion of patches at each scale of the decomposition and proved that the resulting similarity measure (defined in Section 3) remains consistent (see [14] for details). This is exploited to speed up our computations. We now define a similarity on this feature space.

3 Similarity measure

3.1 Definition

Our goal is to define a similarity measure between two images I_1 and I_2 from their feature space i.e. from their respective set of patches $\{\mathbf{w}(I_1)_{j,k}\}_{j,k}$ and $\{\mathbf{w}(I_2)_{j,k}\}_{j,k}$. When images are clearly similar (e.g. different views of the same scene, images containing similar objects...), their patches $\mathbf{w}(I_1)_{j_l,k_l}$ and $\mathbf{w}(I_2)_{j_l,k_l}$ do not necessarily correspond. Hence a measure comparing geometrically corresponding patches would not be robust to geometric transformations. Thus, we propose to compare the pdfs of these patches. Specifically, for an image I, we consider for each scale j the pdf $p_j(I)$ of the set of patches $\{\mathbf{w}(I)_{j,k}\}_k$.

To compare two pdfs, we use the Kullback-Leibler (KL) divergence which derives from the Shannon differential entropy (quantifies the amount of information in a random variable through its pdf). The KL divergence (D_{kl}) is the quantity [9]:

$$D_{kl}(p_1||p_2) = \int p_1 \log(p_1/p_2). \tag{4}$$

This divergence has been successfully used for other applications in image processing in the pixel domain [15, 16], as well as for evaluating the similarity between images using the marginal pdf of the wavelet coefficients [17, 18]. We propose to measure the similarity $S(I_1|I_2)$ between two images I_1 and I_2 by summing over scales the divergences between the pdfs $p_j(I_1)$ and $p_j(I_2)$ (with weights α_j that may normalize the contribution of the different scales):

$$S(I_1|I_2) = \sum_j \alpha_j D_{kl}(p_j(I_1)||p_j(I_2)) \tag{5}$$

3.2 Parametric approach to the estimation: pros and cons

The estimation of the similarity measure S consists of the evaluation of divergences between pdfs $p_j(I_i)$ of high dimension. This raises two problems. Firstly, estimating the KL divergence, even with a good estimate of the pdfs, is hard because this is an integral in high dimension involving unstable logarithm terms. Secondly, the accurate estimation of a pdf itself is difficult due to the lack of samples in high dimension (curse of dimensionality). The two problems should be embraced together to avoid cumulating both kinds of errors.

Parametrizing the shape of the pdf is not ideal. The KL divergence is easy to compute when it is an analytic function of the pdf parameters. This is the case for generalized Gaussians models, which fit well the marginal pdf of multiscale coefficients [17, 18]. However, this model cannot be extended to account for the correlations we want to exploit in multiscale patches. Mixture of Gaussians on contrary are efficient multidimensional models accounting for these correlations [12] but the KL divergence is not an analytic function of the pdf parameters.

Thus, we propose to make no hypothesis on the pdf at hand. We therefore spare the cost of fitting model parameters but we have to estimate the divergences in this non-parametric context. Conceptually, we combine the Ahmad-Lin approximation of the entropies necessary to compute the divergences with "balloon estimate" of the pdfs using the kNN approach.

3.3 Proposed non-parametric estimation of the similarity measure

The KL divergence can be written as the difference between a cross-entropy $\mathrm{H_x}$ and an entropy H (see Eq.(4)):

$$\mathrm{H_x}(p_1, p_2) = -\int p_1 \log p_2, \qquad \mathrm{H}(p_1) = -\int p_1 \log p_1 \tag{6}$$

Let us explain how to estimate these terms from i.i.d sample sets $\mathcal{W}^i = \{\mathbf{w}_1^i, \mathbf{w}_2^i, .., \mathbf{w}_{N_i}^i\}$ of p_i for $i = 1$ or 2 (The samples are in $\mathbb{R}^d$.) Assuming we have estimates $\widehat{p_1}$, $\widehat{p_2}$ of the pdfs p_1, p_2, we use the Ahmad-Lin entropy estimators [19]:

$$\mathrm{H_x^{al}}(\widehat{p_1}, \widehat{p_2}) = -\tfrac{1}{N_1} \sum_{n=1}^{N_1} \log[\widehat{p_2}(\mathbf{w}_n^1)], \qquad \mathrm{H^{al}}(\widehat{p_1}) = -\tfrac{1}{N_1} \sum_{n=1}^{N_1} \log[\widehat{p_1}(\mathbf{w}_n^1)] \tag{7}$$

to obtain a first estimator of the KL divergence:

$$\mathrm{D_{kl}^{al}}(\widehat{p_1}, \widehat{p_2}) = \tfrac{1}{N_1} \sum_{n=1}^{N_1} \log[\widehat{p_1}(\mathbf{w}_n^1)] - \log[\widehat{p_2}(\mathbf{w}_n^1)]. \tag{8}$$

General non-parametric pdf estimators from samples can be written as a sum of kernels K with possibly varying bandwidth h (see [20] for a review):

$$\widehat{p_1}(x) = \tfrac{1}{N_1} \sum\nolimits_{n=1}^{N_1} K_{h(\mathcal{W}^1, x)}(x - \mathbf{w}_n^1) \tag{9}$$

We use a balloon estimator i.e. the bandwidth $h(\mathcal{W}^1, x) = h_{\mathcal{W}^1}(x)$ adapts to the point of estimation x given the sample set $\mathcal{W}^1$. The kernel is binary and the bandwidth is computed in the k-th nearest neighbor (kNN) framework [20]:

$$K_{h(\mathcal{W}^1,x)}(x - \mathbf{w}_n^1) = \frac{1}{v_d \ \rho_{k,\mathcal{W}^1}^d(x)} \ \delta\Big[||x - \mathbf{w}_n^1|| < \rho_{k,\mathcal{W}^1}(x)\Big] \tag{10}$$

with v_d the volume of the unit sphere in $\mathbb{R}^d$ and $\rho_{k,\mathcal{W}}(x)$ the distance of x to its k-th nearest neighbor in $\mathcal{W}$. This is a dual approach to the fixed size kernel methods and was firstly proposed in [21]: the bandwidth adapts to the local sample density by letting the kernel contain exactly k neighbors of x among a given sample set.

Although this is a biased pdf estimator (it does not integrate to one), it has proved to be efficient for high-dimensional data [20]. Plugging Eq.(10) in Eq.(8), we obtain the following estimator of the KL divergence, which is valid in any dimension d and robust to the choice of k: [2]

$$D_{kl}(p_1||p_2) = \log\Big[\tfrac{N_2}{N_1-1}\Big] + \tfrac{d}{N_1}\sum_{n=1}^{N_1}\log[\rho_{k,\mathcal{W}^2}(\mathbf{w}_n^1)] - \tfrac{d}{N_1}\sum_{n=1}^{N_1}\log[\rho_{k,\mathcal{W}^1}(\mathbf{w}_n^1)] \tag{11}$$

4 Experiments

4.1 Database and parameter settings

An experimental evaluation of the *SMP* method has been made on a subset of the Corel database. It includes 1,000 images of size 384×256 or 256×384 which are classified in ten semantic categories (*Africa, Beach, Buildings, Buses, Dinosaurs, Flowers, Elephants, Horses, Food, Mountains*). This dataset is well known in the domain of content-based image retrieval and particularly it has been widely used to evaluate several methods of image categorization, like in [22], [23], [4] and [24].

The *SMP* descriptors of images were extracted as described in Section 2. In particular, to build the patches, the Laplacian pyramid was computed for each channel of the image (in the YUV color space) with a 5-point binomial filter $w5 = [1\ 4\ 6\ 4\ 1]/16$, which is a computationally efficient approximation of the Gaussian filter classically used to build Laplacian pyramids. Two high-frequency subbands and the low-frequency approximation were used. In the following experiments, 1/16 (resp. 1/8 and all) of the patches were selected in the first high-frequency (resp. second high-frequency and low-frequency) subband to describe an image (see Section 2.1). At each scale, the KL divergence was estimated in the kNN framework, with k = 10. The contributions to the similarity measure of the divergences in all subbands were equally weighted ($\alpha_j = 1$ in Eq. (5)).

[2] Note that in the log term, N_1 has been replaced by $N_1 - 1$, which corresponds to ommitting the current sample $\mathbf{w}_i^1$ in the set $\mathcal{W}^1$ when estimating the entropy (Eq. (7) and (9)).

Since the computation of KL divergences is a time-consuming task, we developed a parallel implementation of the kNN search on a Graphic Processing Unit (GPU) [25]. This implementation is based on a brute-force approach and was run on a NVIDIA GeForce 8800 GTX GPU (with 768 MB of internal memory). The computation of one similarity measure between two images required 0.1 s on average.

4.2 Categorization results

In order to allow the prototype learning, we have randomly split the whole image set into a training set and a test set. We have tested our method for different sizes of the training set (15, 30 and 50 images per category), while using all the remaining images as a test set. For each value of the training set size we have repeated the same experiment ten times; hence all reported results are averaged over ten experiments with identical parameter values.

These results are shown in Fig. 3 in terms of the average recognition rate per category. This rate corresponds to the average proportion of correct classification as a function of the number of training images. We can see that the accuracy of categorization is quite stable with the size of the training set. In the same figure we can see an example of which images are selected as prototypes from the training set.

We have also measured the accuracy of our method by using a criterion for rejecting unreliable categorization results, so as to reduce the probability of misclassifications. Namely we have empirically fixed a threshold value for the query-to-prototype distance. Hence, whenever the *SMP*-based "distance" between the query image and its nearest prototype is larger than the threshold, we reject the result, thus giving no label to the image. Table 1 contains the confusion matrix which summarizes the categorization results for the case of 50 training images per category. The generic element $a_{i,j}$ of this matrix represents the average percentage of the test images belonging to category i which have been classified into category j. Therefore, the entries on the main diagonal show the categorization accuracy for each category. On the contrary, off-diagonal entries show classification errors. Interestingly note that big values of some of these entries are related to a certain semantic "intersection" between two different categories. This is particularly clear for misclassifications between categories 2 (*Beaches*) and 9 (*Mountains*), since several images belonging to these categories share similar visual patterns (e.g. rocky seaside pictures are visually similar to mountain sceneries containing lakes or rivers).

Our method shows good overall categorization performances, typically of the same order of magnitude as those of state-of-the-art methods developed by Chen and colleagues [22, 4], although not as accurate on some categories". This can be explained by the difficulty of choosing a representative prototype amongst the training images. Indeed, for diverse categories, the classification rate may vary a lot according to the randomly chosen training set which yields to choosing different prototypes. For example, in the category "Africa", some training sets lead to a prototype being a picture of a face (as displayed in Fig. 3)

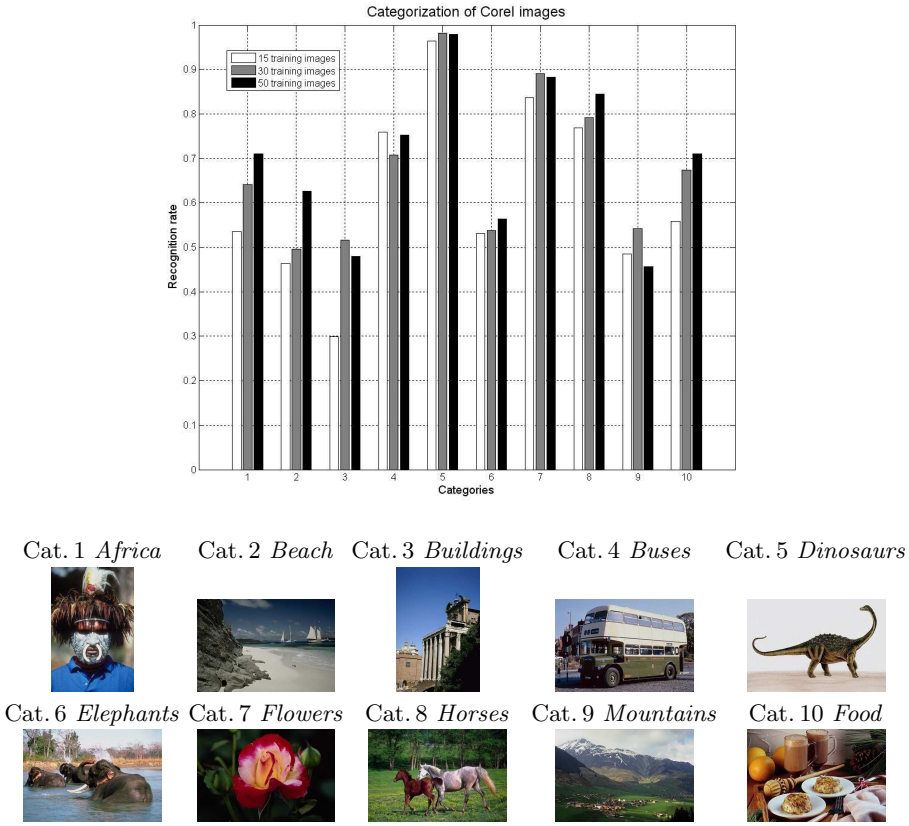

Fig. 3. Top: average recognition rate per category for 15, 30 and 50 training images per category. All results are averaged over 10 experiments with different randomly chosen training sets. Bottom: prototype images for one particular choice of the training sets.

while others lead to village picture prototypes. Different methods are under study to tackle the problem of prototype selection, one of which being to create a poll of representative *SMP*s to define a *pdf* prototype. The patch selection strategy presented here is also basic: one retains a predefined percentage of the patches at each scale. The selection does not adapt to the specificity of the image considered and thus probably results in taking into account outliers of the underlying *pdf*s. Therefore we also study adaptive patch selection strategies. Considering the large prototype variations and the simplistic patch selection process used, the reasonably good performances obtained here show that the *SMP* similarity measure is promising regarding the classification step.

	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Cat. 8	Cat. 9	Cat. 10	reject
Cat. 1	**74.8**	4.2	6.8	0.6	0.0	3.2	0.4	0	1.6	0.4	8.0
Cat. 2	2.0	**56.4**	5.4	0.4	0.0	4.4	0.0	0.0	21.4	0.0	10.0
Cat. 3	7.8	4.0	**74.0**	1.4	0.0	1.6	0.6	0.0	3.8	0.0	6.8
Cat. 4	0.6	1.4	9.4	**73.4**	0.0	0.0	0.0	0.0	1.2	0.2	13.8
Cat. 5	0.0	0.0	0.2	0.0	**88.8**	0.0	0.0	0.0	0.0	0.4	10.6
Cat. 6	4.2	0.6	0.4	0.0	0.0	**87.8**	0.0	0.8	1.8	0.6	3.8
Cat. 7	3.0	0.0	0.0	0.0	0.0	0.0	**89.0**	0.6	0.0	0.0	7.4
Cat. 8	1.2	0.4	0.0	0.0	0.0	1.2	0.4	**91.4**	0.6	0.0	4.8
Cat. 9	0.2	14.0	5.0	0.4	0.0	5.8	1.0	0.0	**62.0**	0.2	11.4
Cat. 10	12.8	1.8	8.2	1.2	0.4	2.4	0.0	0.0	1.2	**55.6**	16.4

Table 1. The confusion matrix resulting from our image categorization experiments (over 10 randomly generated training sets containing 50 images per category). Entry on the row i and column j is the average percentage (over the 10 experiments) of images belonging to the category i which have been classified into the category j (see Fig. 3 for category names). The last column lists the percentage of non-classified images for each category.

5 Conclusion

In this paper we tackled the task of image categorization by using a new image similarity framework. It is based on high-dimensional probability distributions of patches of multiscale coefficients which we call *Sparse Multiscale Patches or SMP*s. Image signatures are represented by sets of patches of subband coefficients, that take into account their intrascale, interscale and interchannel dependencies. The similarity between images is defined as the "closeness" between the distributions of their signatures, measured by the Kullback-Leibler divergence. The latter is estimated in a non-parametric framework, via a *k-th nearest neighbor* or kNN approach.

Our approach to image classification is to represent each category by an image prototype. The latter is defined as the image minimizing the *SMP*-based measure with all other images in the category's training set.

The experiments we made on a subset of the Corel collection show that the *SMP* similarity measure is a promising tool for the categorization problem. The prototype selection as well as the patch selection are to be improved and are among the subjects of ongoing work.

References

1. Szummer, M., Picard, R.W.: Indoor-outdoor image classification. In: CAIVD '98: Proceedings of the 1998 International Workshop on Content-Based Access of Image and Video Databases (CAIVD '98), Washington, DC, USA, IEEE Computer Society (1998) 42
2. Vailaya, A., Figueiredo, M., Jain, A., Zhang, H.J.: Image classification for content-based indexing. Image Processing, IEEE Transactions on **10** (2001) 117–130

3. Li, J., Wang, J.Z.: Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Trans. Pattern Anal. Mach. Intell. **25** (2003) 1075–1088
4. Bi, J., Chen, Y., Wang, J.Z.: A sparse support vector machine approach to region-based image categorization. In: CVPR '05. (2005) 1121–1128
5. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: CVPR (2). (2006) 2169–2178
6. Liu, Y., Perronnin, F.: A similarity measure between unordered vector sets with application to image categorization. In: CVPR. (2008)
7. Zhang, H., Berg, A.C., Maire, M., Malik, J.: Svm-knn: Discriminative nearest neighbor classification for visual category recognition. In: CVPR (2). (2006) 2126–2136
8. Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Braem, P.B.: Basic objects in natural categories. Cognitive Psychology **8** (1976) 382–439
9. Puzicha, J., Rubner, Y., Tomasi, C., Buhmann, J.M.: Empirical evaluation of dissimilarity measures for color and texture. In: ICCV. (1999) 1165–1172
10. Portilla, J., Strela, V., Wainwright, M., Simoncelli, E.P.: Image denoising using a scale mixture of Gaussians in the wavelet domain. TIP **12** (2003) 1338–1351
11. Romberg, J.K., Choi, H., Baraniuk, R.G.: Bayesian tree-structured image modeling using wavelet-domain hidden markov models. TIP **10** (2001) 1056–1068
12. Pierpaoli, E., Anthoine, S., Huffenberger, K., Daubechies, I.: Reconstructing sunyaev-zeldovich clusters in future cmb experiments. Mon. Not. Roy. Astron. Soc. **359** (2005) 261–271
13. Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. IEEE Trans. Communications **31** (1983) 532–540
14. Piro, P., Anthoine, S., Debreuve, E., Barlaud, M.: Image retrieval via kullback-leibler divergence of patches of multiscale coefficients in the knn framework. In: CBMI, London, UK (2008)
15. Boltz, S., Debreuve, E., Barlaud, M.: High-dimensional kullback-leibler distance for region-of-interest tracking: Application to combining a soft geometric constraint with radiometry. In: CVPR, Minneapolis, USA (2007)
16. Angelino, C.V., Debreuve, E., Barlaud, M.: Image restoration using a knn-variant of the mean-shift. In: ICIP, San Diego, USA (2008)
17. Do, M., Vetterli, M.: Wavelet based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. TIP **11** (2002) 146–158
18. Wang, Z., Wu, G., Sheikh, H.R., Simoncelli, E.P., Yang, E.H., Bovik, A.C.: Quality-aware images. TIP **15** (2006) 1680–1689
19. Ahmad, I., Lin, P.E.: A nonparametricestimationof the entropy absolutely continuousdistributions. IEEE Trans.Inform. Theory **22** (1976) 372–375
20. Terrell, George R.and Scott, D.W.: Variable kernel density estimation. The Annals of Statistics **20** (1992) 1236–1265
21. Loftsgaarden, D., Quesenberry, C.: A nonparametric estimate of a multivariate density function. AMS **36** (1965) 1049–1051
22. Chen, Y., Wang, J.Z.: Image categorization by learning and reasoning with regions. J. Mach. Learn. Res. **5** (2004) 913–939
23. Li, T., Kweon, I.S.: A semantic region descriptor for local feature based image categorization. In: ICASSP '08. (2008) 1333–1336
24. Zhu, Y., Liu, X., Mio, W.: Content-based image categorization and retrieval using neural networks. In: ICME '07. (2007) 528–531
25. Garcia, V., Debreuve, E., Barlaud, M.: Fast k nearest neighbor search using gpu. In: CVPR Workshop on Computer Vision on GPU. (2008)