KERNELS ON BAGS OF FUZZY REGIONS FOR FAST OBJECT RETRIEVAL

Philippe H. Gosselin[⊤], *Matthieu Cord*[⋆] *and Sylvie Philipp-Foliguet*[⊤]

[⊤]ETIS / CNRS, 6 ave du Ponceau, 95014 Cergy, France
[⋆]LIP6 / CNRS, 104 ave Kennedy, 75006 Paris, France

ABSTRACT

We propose in this paper a general kernel framework to deal with database object retrieval embedded in images with heterogeneous background. We use local features computed on fuzzy regions for image representation summarized in bags, and we propose original kernel functions to deal with sets of features and spatial constraints. Combined with SVMs classification and online learning scheme, the resulting algorithm satisfies the robustness requirements for representation and classification of objects. Experiments on a specific database having objects with heterogeneous backgrounds show the performance of our object retrieval technique.

Index Terms— Information retrieval, Learning systems, Interactive systems, Object recognition, Machine vision

1. INTRODUCTION

Significant progress in the performance of object categorization and retrieval systems has been achieved in the last decade. Powerful strategies have been proposed for object recognition in different poses, in the presence of clutter, occlusion and varying lighting conditions. However, the problem remains very hard when considering object categorization and retrieval. Databases may be very large, with many objects embedded in large images with heterogeneous background.

In this difficult context, object retrieval systems must satisfy two main requirements: an effective data representation based on local descriptors in order to catch the object characterization, and an effective classification strategy. Many papers focus on the representation of local features in images, and very efficient techniques are now available, such as points of interest approaches [1] or region-based techniques [2]. SVMs are state-of-the-art large margin classifiers which have demonstrated remarkable performance in object recognition.

We propose in this paper a general kernel framework to deal with object retrieval. We use local features computed on fuzzy regions for image representation summarized in bags, SVMs for classification [3], and we combine these two successful approaches via the introduction of a specific class of kernels on bags of features. The resulting algorithm satisfies the robustness requirements for representation and classification. We present several experiments on a specific database built from the well-known Columbia database when adding heterogeneous background. The RETIN active learning framework previously introduced in [4] is used to investigate the performance of our kernel-based learning algorithm for online object retrieval.

(a) (b) (c)

Fig. 1. An example of fuzzy segmentation. (a) Original image. (b) Defuzzified regions. (c) Fuzzy region 11.

2. REGION-BASED FEATURE REPRESENTATION

In the same way than the human visual system perceives coarse zones with their approximate colors and sizes, we build regions which roughly correspond to the main color parts of the image. Despite the fact that our visual system does not perform an accurate segmentation of the scene, the recognition of a landscape or a painting is instantaneous. Differently from the other systems which use regions [2], we use a segmentation into fuzzy regions. The main advantage is to be able to segment any image, even in difficult cases, when there is no clear limit between some parts of the objects. Fig. 1 shows an example of fuzzy segmentation. Also note that, compared to points of interest (PoI) [1, 5], the region-based representations are able to provide a very compact local representation : where it is usually required to have hundreds to thousands of PoI, only a few dozen of regions are necessary for an image.

2.1. Fuzzy segmentation algorithm

Details of the algorithm of fuzzy segmentation are given in [6]. Extracted regions have the following properties : uniformity in terms of color, contained expansion by high gradient norms, uncertainty where two (or more) regions encounter. The algorithm first performs a watershed algorithm on the

image of gradient norms, the uniform areas (of null color gradient norm) constitute the cores of the fuzzy regions. The membership degrees of pixels to regions are then computed using the topographic distance to these cores, which is defined as the length of the shortest path connecting the pixel to the core, along the surface constituted by the gradient norm in the 3D space. The degrees slowly decrease according to the spatial distance to the core and strongly decrease when meeting an edge, zone of a large gradient norm. Impulse noise is bypassed, because a shorter path is found around it.

2.2. Region indexing

Each image i of the database is represented with a bag $B_i = \{\mathbf{b}_{ri}\}_r$ of regions $\mathbf{b}_{ri} \in \mathbb{R}^p$. Vectors $\mathbf{b}_{ri}$ are the concatenation of one color histogram and 3 texture histograms. Histograms are obtained by adding the membership degrees of the pixels to the region. Thus pixels with small membership degrees belonging to transitions or outliers inside a region have little influence on the histogram shape.

3. KERNEL DESIGN FOR BAGS OF FEATURES

After the region feature computation, each image is represented by sets of unordered vectors. If heterogeneous background is considered, several feature vectors are relevant for object characterization, many are irrelevant or let's say noise. The next step is now to consider similarity functions between them. The major aim is to find the set of local descriptors that discriminate an object from other objects and the background. In other words, we have to detect within the bags B_i which features are relevant.

3.1. Kernel on bags

As bags $B_i = \{\mathbf{b}_{ri}\}_r$ belongs to the set of subsets $\mathcal{P}(\mathbb{B})$, the input space is a non vectorial space. Let us note $\Phi : \mathcal{P}(\mathbb{B}) \rightarrow \mathcal{H}$ the embedding function which maps any bag B_i to a vector $\Phi(B_i)$ in a Hilbert space $\mathcal{H}$. To design kernels over sets, one can find a function K corresponding to a dot product in the induced space:

$$K(B_i, B_j) = \langle \Phi(B_i), \Phi(B_j) \rangle$$

Some authors have recently proposed strategies using explicit mapping Φ [7]. These approaches are not really kernel-oriented and most of the work focuses in that case on preprocessing to map bags into a finite $\mathbb{R}^p$ space.

Contrary to previous techniques, kernel framework deals with building a function without explicit evaluation of the corresponding mapping Φ. Several kernel functions have been proposed [8, 9, 10]. We address this issue for the following class of kernels:

$$K(B_i, B_j) = \sum_{\mathbf{b}_{ri} \in B_i} \sum_{\mathbf{b}_{sj} \in B_j} k(\mathbf{b}_{ri}, \mathbf{b}_{sj}) \qquad (1)$$

Where k is the minor function on $\mathbf{b}_{ri} \in \mathbb{B}$ with ϕ as embedding function into the feature space $\mathcal{H}$. We have in this case: $\Phi(B_i) = \sum_{\mathbf{b}_{ri} \in B_i} \phi(\mathbf{b}_{ri})$. One interesting property of this formalization is that changing ϕ to another does not change the Φ embedding structure.

The function of Eq. (1) satisfy the Mercer's conditions (see [11], Chap. 9 for proof), however it returns the average similarity between all the local descriptors. In order to increase the high matches $k(\mathbf{b}_{ri}, \mathbf{b}_{sj})$, Lyu introduces the following function [12]:

$$K_{lyu}(B_i, B_j) = \frac{1}{|B_i|} \frac{1}{|B_j|} \sum_{\mathbf{b}_{ri} \in B_i} \sum_{\mathbf{b}_{sj} \in B_j} k(\mathbf{b}_{ri}, \mathbf{b}_{sj})^q \qquad (2)$$

Using a high value of q, high matches will be increased much more than low matches.

However, with high values of q, the function of Eq. (2) produces a very discriminative kernel. In other words, function of Eq. (2) tends to be the value of the highest match power q, i.e. $K_{lyu}(B_i, B_j) \propto \max_{r,s} k(\mathbf{b}_{ri}, \mathbf{b}_{sj})^q$. Then, the Hilbert space induced by K_{lyu} tends to be of infinite dimension, just like the Gaussian kernels with a very small σ. In order to get a higher generalization capacity, we propose the following function:

$$K_{single}(B_i, B_j) = \left(\sum_{\mathbf{b}_{ri} \in B_i} \sum_{\mathbf{b}_{sj} \in B_j} k(\mathbf{b}_{ri}, \mathbf{b}_{sj})^q \right)^{\frac{1}{q}} \qquad (3)$$

As Minkowski distance tends to be the L^∞ distance as $q \rightarrow \infty$, this function tends to be the function $\max_{r,s} k(\mathbf{b}_{ri}, \mathbf{b}_{sj})$ as $q \rightarrow \infty$.

3.2. Integration of spatial constraints

Representing an image as an unordered set of regions is assuming that the regions are independent, which means that objects with similar regions laid out differently are indistinguishable. We propose to extend our kernel function in order to improve discrimination using spatial constraints. We show an example in Fig. 2, where the query is the left image that contains a green can. Using only independent regions, the image on the top right with a red can will be more similar than the image on the bottom right with a green can, since red regions 2,5,10 in the left image will match with the bottom of the red can (region 10). However, if we use pairs of regions instead of single regions, the bottom right image with a green can will be the closest one to the query : pair (4,11) in the query only matches with the pair (9,1) in the bottom right image that also contains a green can.

To take into account spatial dependencies between regions, we are considering pairs of adjacency regions : for each region of an image, we build 3 pairs with its 3 closest regions. Each image i is then represented with a set $\mathcal{P}_i$ of pairs $P_{vi} \in \mathbb{B} \times \mathbb{B}$.

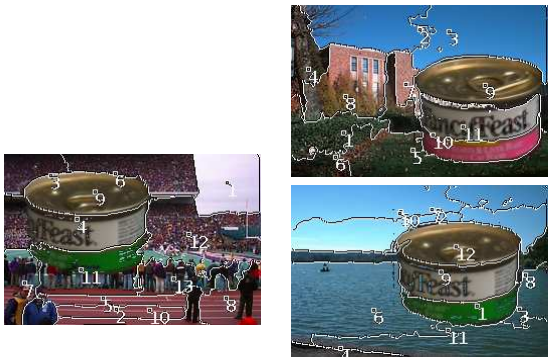

Fig. 2. Integration of spatial constraints.

Next, each pair is considered as a "mini-bag" of two regions, which leads to the following function:

$$K_{pairs}(\mathcal{P}_i, \mathcal{P}_j) = \left(\sum_{P_{vi} \in \mathcal{P}_i} \sum_{P_{wj} \in \mathcal{P}_j} K_{single}(P_{vi}, P_{wj})^q \right)^{\frac{1}{q}} \quad (4)$$

Another way to represent spatial relations on objects concerns graph representations. Kernel on graph design has been proposed by considering matches on paths of the graph [13]. However, their computational needs are intractable for real applications.

4. EXPERIMENTS

4.1. Database setup

Columbia database has been used a lot to evaluate object recognition methods. With 100 objects, and 72 shots from different points of view for each object on a homogeneous black background, it turns out that most of the methods now achieve high performances with very few training data.

In order to retrieve objects in a much more realistic scenario, we built a synthetic database with objects from Columbia and background from Washington database[1]. We randomly selected 12 views of 50 different objects from the Columbia database, removed the background, and embedded them on the images of the Washington database (rescaled to a fixed size). The final database contains 600 images in specific backgrounds. All these backgrounds have very heterogeneous content. Examples are shown in Fig.3. In [12], Lyu also considers an extended database with objects. The one we built for our experiments is comparable to the Lyu's one, except that the objects are much smaller in our final database, making the problem much harder.

[1] http://www.cs.washington.edu/research/imagedatabase/

4.2. Experimental setup

The computation of the fuzzy regions require a main parameter which is the number of regions the algorithm should return. Since the segmentation has to be automatic, the level of detail is set through an interval for the number of regions. For this database the interval was set to $[5, 15]$ fuzzy regions. Each region is represented by 4 histograms, one of 8 chrominances values from $CIEL^\star a^\star b^\star$, and 3 of 8 textures from Gabor filters for 3 different scales. Concerning pairs of regions, we get around 20 pairs after removing repeated pairs.

We tested several minor kernel functions, and found that a Gaussian kernel with a χ^2 distance is the best choice against linear, polynomial, Gaussian L1, Gaussian L2, triangle, and minima kernels. Next, we use the function of Eq. 3 for single regions, and the function of Eq. 4 for pairs of regions. In both cases, we found that the best power value is $q = 5$ on average.

Thanks to the kernel functions on features, we train SVM classifiers to discriminate images that contain or not the query object from the others. This also allow us to use active learning, which has proved its huge interest for interactive search [14]. In the following experiments, we use a precision-oriented active learning technique in conjunction with a correction of the boundary and a fast and efficient batch selection [4]. Each retrieval session is initialized with 1 image containing the object the user is looking for. Next, 5 images chosen by the active learning are labeled with binary annotations (positive: the image contains the query object, negative: the image does not contain the query object). This process is repeated 10 times. At the end of a retrieval session, the training set is made of 51 labeled images. Performances are evaluated with the Mean Average Precision, i.e. the sum of the Precision/Recall curve[2].

4.3. Results

Results are shown on Fig. 4. The two fuzzy region-based methods start at the same position, but next MAP with pairs of regions increases much more than single regions, which shows the interest of this strategy for object retrieval. We compare fuzzy regions to global histograms, which represent an image with a single histogram of colors and textures [4]. We can see on Fig. 4 that local representations clearly outperform the global one. We also compare our fuzzy region indexing technique to the well-known approach combining MSER region detectors and SIFT descriptors[1]. We use them with the function of Eq. 3. With few training samples, MSER/SIFT gives better results, but the performances of our technique increase fast, and equal MSER/SIFT with 35 training samples. This shows the generalization capacity of our fuzzy region approach, which is well suited to semantic learning tasks.

The main drawback of the local methods is the time complexity. In our strategy, we tried to overcome this problem

[2] cf. TRECVid: http://www-nlpir.nist.gov/projects/trecvid/

Fig. 3. Objects of Columbia on random backgrounds.

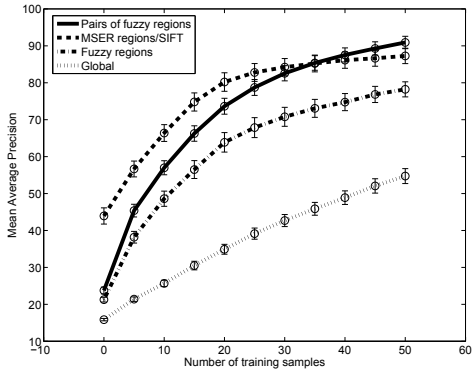

Fig. 4. Average performances in Mean Average Precision(%) for an interactive search of one object using different features and kernels.

using a limited number of region features. It turns out that our method is about 100 times faster than MSER/SIFT. This is due to the very large number of PoI required by the approach. Since the two methods gave approximately the same results, using pairs of fuzzy regions is more interesting for real applications.

5. CONCLUSION

In this paper, a kernel-based method for object retrieval is presented. Local feature computation on fuzzy regions provides a robust local-based image representation. The resulting bags of features are compared thanks to a new class of kernels on sets to deal with images composed of objects with heterogeneous background. These kernels highlight the best local matches between features, so that they efficiently overcome the noise problem introduced by the background features. Results show that our approach is able to perform object retrieval in real-world scenes with heterogeneous background. The method using kernels integrating spatial constraints on the regions achieved results as good as PoI ones, but is 100 times faster !

6. REFERENCES

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, pp. 43–72, November 2005.

[2] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 8, pp. 1026–1038, 2004.

[3] V.N. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, New York, 1999.

[4] P.H. Gosselin and M. Cord, "Precision-oriented active selection for interactive image retrieval," in *IEEE International Conference on Image Processing*, Atlanta, GA, USA, October 2006.

[5] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal on Computer Vision (IJCV)*, vol. 2, no. 60, pp. 91–110, 2004.

[6] S. Philipp-Foliguet and J. Gony, "FReBIR : Fuzzy regions-based image retrieval," in *Image Processing and Multimedia Understanding (IPMU)*, Paris, France, July 2006.

[7] Y. Chen and J.Z. Wang, "Image categorization by learning and reasoning with regions," *International Journal on Machine Learning Research*, vol. 5, pp. 913–939, 2004.

[8] R. Kondor and T. Jebara, "A kernel between sets of vectors," in *International Conference on Machine Learning (ICML)*, 2003.

[9] C. Wallraven, B. Caputo, and A.B.A. Graf, "Recognition with local features: the kernel recipe," in *International Conference on Computer Vision (ICCV)*, 2003, vol. 2, pp. 257–264.

[10] L. Wolf and A. Shashua, "Learning over sets using kernel principal angles," *Journal of Machine Learning Research (JMLR)*, pp. 913–931, 2003.

[11] J. Shawe-Taylor and N. Cristianini, *Kernel methods for Pattern Analysis*, Cambridge University Press, ISBN 0-521-81397-2, 2004.

[12] S. Lyu, "Mercer kernels for object recognition with local features," in *IEEE International Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005.

[13] H. Kashima and Y. Tsuboi, "Kernel-based discriminative learning algorithms for labeling sequences, trees and graphs," in *International Conference on Machine Learning (ICML)*, Banff, Alberta, Canada, 2004.

[14] S Tong and D Koller, "Support vector machine active learning with application to text classification," *Journal of Machine Learning Research*, pp. 2:45–66, November 2001.