Empirical Study of Multi-scale Filter Banks for Object Categorization

Manuel J. Marín-Jiménez and Nicolás Pérez de la Blanca
Dpt. of Computer Science and Artificial Intelligence
University of Granada. Spain
mjmarin@decsai.ugr.es, nicolas@ugr.es

Abstract

The aim of this work is the evaluation of different multi-scale filter banks, mainly based on oriented Gaussian derivatives and Gabor functions, to be used in the generation of robust features for visual object categorization. In order to combine the responses obtained from several spatial scales, we use the biologically inspired HMAX model [13]. We have tested the different sets of features on the challenging Caltech-101 database, and we have performed the categorizarion procedure with AdaBoost, Support Vector Machines and JointBoosting classifiers, achieving remarkable results.

1. Introduction

The Marr‘s theory [11] supports that in the early stages of the vision process, there are cells that respond to stimulus of primitive shapes, such as corners, edges, bars, etc. Young [20] models these cells by using Gaussian derivative functions. Riesenhuber & Poggio [13] propose a model for simulating the behavior of the Human Visual System (HVS), at the early stages of vision process. This model is named HMAX. [1] It generates features that exhibit interesting invariance properties. More recently, Serre *et al.* [14], propose a new model for image categorization adding to the HMAX model a learning step and changing the original Gaussian derivative filter bank by a Gabor filter bank. They argue that the Gabor filter is much more suitable in order to detect local features. Nevertheless no experimental support has been given.

Different local feature based approaches are used in the field of object categorization in images. Serre *et al.* [14] use local features based on filter responses to describe objects, achieving a high performance in the problem of object categorization. On the other hand, different approaches using grey-scale image patches, extracted from regions of interest, to represent parts of objects has been suggested, Fei-Fei *et al.* [9], Agarwal *et al.* [1], Leibe [8]. Nevertheless, at the moment, there is not a clear advantage from any of these approaches. However, the non-parametric and simple approach followed by Serre *et al.* [14] in his learning step, suggests that a lot of discriminative information can be learned from the output of filter banks. Computing anisotropic Gabor features is a heavy task that only is justified if the experimental results show a clear advantage on any other type of filter bank.

The aim of this work is to carry out an experimental study in order to propose a new set of simpler filter banks, comparing the local features based on a Gabor filter banks with the Gaussian derivative filter banks. These features will be applied to the object categorization problem.

In section 2 of this paper, we review the use of Gaussian functions as local descriptors. In section 3, we introduce the proposed filter banks for object categorization. In section 4, we describe the experiments and present the experimental results. And finally, in section 5, we present the summary and our conclusions.

2. Using filters to describe images

Koenderink *et al.* [7] propose a methodology to analyze the local geometry of the images, based on the Gaussian function and its derivatives. Several optimization methods are available to perform efficient filtering with those functions [16]. Furthermore, steerable filters [4, 12] (oriented filters whose response can be computed as linear combination of other responses) can be defined in terms of Gaussian functions.

Yokono & Poggio [19] show, empirically, the excellent performance achieved by features created with filters based on Gaussian functions, applied to the problem of object recognition. In other published works, as Varma et al. [17], Gaussian filter banks are used to describe textures.

Our goal is to evaluate the capability of different filter

[1] HMAX consists of 4 types of features: S1, C1, S2, C2. S1 features are the lowest level features, and they are computed as filter responses, grouped into scales. C1 features are obtained by combining pairs of S1 scales with the maximum operation.

banks, based on Gaussian functions, for encoding information usable for object categorization. We will use the biologically inspired HMAX model to generate features.

3. Our proposed multi-scale filter banks

Due to the existence of a large amount of works based on Gaussian filters, we propose to use filter banks compound by the Gaussian function and its oriented derivatives as local descriptors, including them in the first level of HMAX.

In order to improve the information provided by the features, we propose to include the responses of the Forstner operator [3], used to detect regions of interest. For each image point, we can compute a q value, in the range $[0, 1]$, by using equation 2.

$$N(x, y) = \int_W M(x, y)dxdy \approx \Sigma M_{i,j} \tag{1}$$

$$q = 1 - \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2 = \frac{4detN}{(trN)^2} \tag{2}$$

Where M is the moments matrix, W is the neighborhood of the considered point (x, y) and λ_1, λ_2 are the eigenvalues of matrix N.

We will compare our proposed filter banks with the Gabor filter bank (as [14]). On the other hand, Viola and Jones, in their fast object detector [18], use filters which are simplified versions of first and second order gaussian derivative filters, to extract local features. Since those filters achieve very good results and are computable in a very efficient way, we will include them in our comparison.

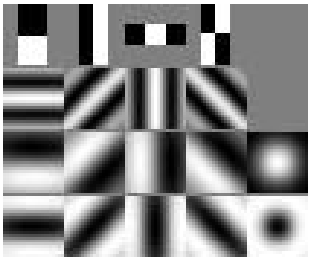

Figure 1. Sample filter banks. From top to bottom: Viola, Gabor, 1oGD with G0 and 2oGD with LoG.

Figure 1 shows sample filter banks: Viola, Gabor, first order Gaussian derivatives (1oGD) with G0, and second order Gaussian derivatives (2oGD) with LoG.

4. Experiments

We have chosen the Caltech 101-object categories [2] to perform the experiments. This database has become, nearly, the standard database for object categorization. It contains images of objects grouped into 101 categories, plus a background category, commonly used as the negative set. This is a very challenging database because the objects are embedded in cluttered backgrounds and have different scales and poses. In order to make a robust comparison, we have discarded the 15 categories that contains less than 40 samples. All the images were normalized in size, so that the longer side had 140 pixels and the other side was proportional, to preserve the aspect ratio.

An extended version of this paper [10] can be found at `http://decsai.ugr.es/vip/publications.php`, where more details are provided.

4.1 Multi-scale filter banks evaluation

We will compute features based on different filter banks. For each feature set, we will train binary classifiers for testing the presence or absence of objects in images from a particular category. The set of the negative samples is compound by images of all categories but the current one, plus images from the background category. This strategy differs from the classic one, where the negative set is compound only by background images, because we are interested in studying the capability of the features to distinguish between different categories, and not only in distinguishing foreground from background.

The eight filter banks defined for this experiment are the following: (1): Viola (2 edge filters, 1 bar filter and 1 special diagonal filter); (2): Gabor (as [14]); (3): anisotropic 1oGD; (4): anisotropic 2oGD; (5): (3) + isotropic zero-order gaussian (G0); (6): (3) + isotropic Laplacian of Gaussian (LoG) and Forstner operator; (7): (3) + (4) + G0 + LoG + Forstner operator; and, (8): (4) + Forstner operator.

The Gabor filter and the anisotropic first and second order Gaussian derivatives (with aspect-ratio equals 0.25) are oriented at 0, 45, 90 and 135. All the filter banks contain 16 scales (as [14]). Table 1 shows the parameters of the filter banks.

Table 1. Filter mask size (*FS*) and filter width (σ) for Gaussian-based filter banks.

FS	7	9	11	13	15	17	19	21
σ	1.75	2.25	2.75	3.25	3.75	4.25	4.75	5.25
FS	23	25	27	29	31	33	35	37
σ	5.75	6.25	6.75	7.25	7.75	8.25	8.75	9.25

[2] The database is available at *http://www.vision.caltech.edu/*

In these filter banks we have combined linear filters (Gaussian derivatives of different orders) and non-linear filters (Forstner operator) to study if the mixture of information of diverse nature enhances the quality of the features.

We will generate features (named C2) following the HMAX method and using the same empirical tuned parameters proposed by Serre *et al.* in [14]. The evaluation of the filters will be done following a strategy similar to the one used in [9]. From one single category, we draw 30 random samples for training, and 50 different samples for test, or less (the remaining ones) if there are not enough in the set. The training and test negative set are both compound by 50 samples, randomly chosen following the strategy previously explained. For each category and for each filter bank we will repeat 10 times the experiment.

Results In order to avoid a possible dependence between the features and the type of classifier used, we have trained and tested, for each repetition, two different classifiers: AdaBoost (with decision stumps) and Support Vector Machines (with linear kernel). The stop conditions used to finish the training of Adaboost have been a fixed error threshold and a maximum of iterations (as many as features). On the other hand, we have choosen *libsvm* library (`http://www.csie.ntu.edu.tw/~cjlin/libsvm/`) for SVM classifiers. For training SVM, a grid search over the parameters space has been performed.

During the patch [3] extraction process, we have always taken the patches from a set of prefixed positions in the images. Thereby, the comparison is straightforward for all filter banks. We have decided, empirically , to use 300 patches (features) per category and filter bank. If those 300 patches were selected (from a huge pool) for each individual case, the individual performances would be better but the comparison would be unfair.

The results obtained from the classification process, are summarized in table 2. For each filter bank and for each classifier, we have computed the average of all classification ratios (first row) and the average of their confidence intervals (second row). The performance is measured at *equilibrium-point* (when the miss-ratio equals the false positive ratio). If we focus on table 2, we see that the averaged performances are very similar. Also, the averaged confidence intervals are overlapped. If we pay attention only at the averaged performance, the filter bank based on second order gaussian derivatives stands out slightly from the others. Hence, our conclusion for this experiment is that Gaussian filter banks represent a clear advantage in comparison to Gabor. It is much better in terms of computational burden and is slightly better in terms of categorization efficacy.

[3] In this context, a *patch* is a piece of a filtered image, extracted from a particular scale. It is three dimensional: for each point of the patch, it contains the responses of all the different filters, for a single scale.

Table 2. Results of classification using different filter banks: averaged performance and confidence intervals.

	Viola	*Gabor*	*FB-3*	*FB-4*	*FB-5*	*FB-6*	*FB-7*	*FB-8*
AdaB	78.4	**81.4**	81.2	**81.4**	**81.9**	77.9	80.3	78.1
	4.3	3.9	3.9	4.2	3.3	4.5	4.2	4.0
SVM	84.2	**85.5**	84.1	**86.0**	84.1	82.6	82.8	82.7
	2.3	2.5	3.6	3.2	3.0	2.7	2.4	2.6

4.2. Multicategorization

In this experiment we deal with the problem of multicategorization on the full Caltech 101-object categories, included the background category. The training set is compound by the mixture of 30 random samples drawn from each category, and the test set is compound by the mixture of 50 different samples drawn from each category (or the remaining, if it is less than 50). Each sample is enconded by using 4075 patches [14], randomly extracted from the full training set. These features are computed by using the oriented second order Gaussian derivatives filter bank. We will use a Joint Boosting classifier, proposed recently by Torralba *et al.*[15].

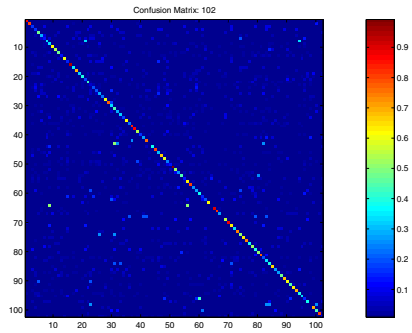

Figure 2. Confusion matrix for 101-objects database. Global performance over 46%.

Under these conditions, we have achieved an average 46.3% of global correct categorization (chance is below 1% for this database), where more than 40 categories are over 50% of correct categorization. With only 2500 patches, we achieve a 44% rate. If we use 15 samples per category for training, we achieve a 39.5% rate. Figure 2 shows the confusion matrix for the 101 categories plus background (by using 4075 features and 30 samples per category). For each row, the highest value should belong to the diagonal.

Other results on this database, using diverse technics, are: Serre 42% [14], Holub 40.1% [6], Grauman 43% [5], and, the best result up to our knowledge, Berg 48% [2].

Caltech selected categories These are: *motorbikes, faces, planes* and *leopards*. We have trained JointBoosting classifiers (using 2000 patches based on 2oGD) with an increasing number of samples, and tested with all the remaining ones. Figure 3 shows how the mean test performance, after 10 random repetitions, evolves according to the number of samples (per category) used for training. With only 50 samples, these results are comparable to the ones shown by Holub *et al.* in [6].

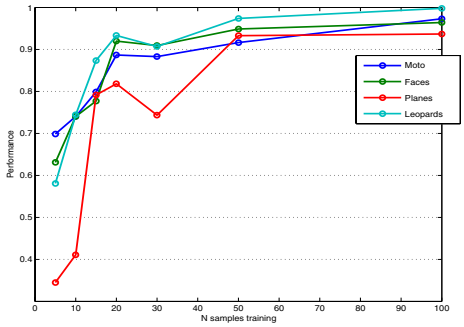

Figure 3. Performance VS number of training samples per category. Multiclass environment.

5. Summary and conclusions

An experimental study has been carried out in order to compare the performance of different filter banks for the object categorization problem. We have generated multi-scale features with eight proposed filter banks, which have been used to learn the object categories included in the challenging Caltech-101 database. The results show that local features generated with filter banks based on Gaussian derivatives, achieve an excellent performance in the object categorization problem, compared to the Gabor-based features. In fact, the results provided in the task of multicategorization on the Caltech-101, combined with JointBoosting classifiers, are very competitive compared to the state-of-the-art. However, we think that including more filter banks is not enough to improve the achieved performance. For this reason, it is necessary to study alternative options to obtain better results.

6. Acknowledgments

Thanks to Dr. Jordi Vitrià and Dr. Javier Portilla for their helpful comments. This work was supported by the Spanish Ministry of Education and Science (FPU grant AP2003-2405) and project TIN2005-01665.

References

[1] S. Agarwal, A. Awan, and D. Roth. Learning to detect objects in images via a sparse, part-based representation. *IEEE PAMI*, 26(11):1475–1490, Nov. 2004.

[2] A. Berg, T. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *CVPR*, 2005.

[3] W. Forstner and E. Gulch. A fast operador for detection and precise location of distinct points, corners and centres of circular features. *ISPRS Interc. Workshop*, June 1987.

[4] W. Freeman and E. Adelson. Steerable filters for early vision, image analysis and wavelet decomposition. In IEEE, editor, *3rd ICCV*, pages 406–415, Dec 1990.

[5] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In *Proceedings of the IEEE ICCV*, October 2005.

[6] A. D. Holub, M. Welling, and P. Perona. Combining generative models and fisher kernels for object recognition. In *ICCV05*, 2005.

[7] J. Koenderink and A. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

[8] B. Leibe. *Interleaved Object Categorization and Segmentation.* PhD thesis, ETH Zurich, October 2004.

[9] F. Li, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. 2004.

[10] M. J. Marín-Jiménez and N. P. de la Blanca. Empirical study of multi-scale filter banks for object categorization. TechReport 121505, VIP, December 2005.

[11] D. Marr. *Vision*. W. H. Freeman and Co., 1982.

[12] P. Perona. Deformable kernels for early vision. *IEEE PAMI*, 17(5):488–499, May 1995.

[13] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025, 1999.

[14] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In *IEEE CSC on CVPR*, June 2005.

[15] A. B. Torralba, K. P. Murphy, and W. T. Freeman. Sharing features: Efficient boosting procedures for multiclass object detection. In *CVPR (2)*, pages 762–769, 2004.

[16] L. van Vliet, I. Young, and P. Verbeek. Recursive gaussian derivative filters. In *l4th ICPR*, volume 1, pages 509–514. IEEE, August 1998.

[17] M. Varma and A. Zisserman. Unifying statistical texture classification frameworks. *Image and Vision Computing*, 22(14):1175–1183, 2005.

[18] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE CVPR*, volume 1, pages 511–518, 2001.

[19] J. J. Yokono and T. Poggio. Oriented filters for object recognition: an empirical study. In *Proc. of the Sixth IEEE FGR*, May 2004.

[20] R. A. Young. The gaussian derivative model for spatial vision: I. Retinal mechanisms. *Spatial Vision*, 2(4):273–293, 1987.