BIOLOGICALLY INSPIRED METHODS OF OBJECT RECOGNITION

Tomáš Vaľko

Master Degree Programme (2.), FIT BUT

E-mail: xvalko00@stud.fit.vutbr.cz

Supervised by: Roman Juránek

E-mail: ijuranek@fit.vutbr.cz

Abstract: Object recognition is one of many tasks in which the computer is still behind the human. Therefore, development in this area takes inspiration from nature and especially from the function of the human brain. This work focuses on object recognition based on extracting relevant information from images, features. Features are obtained in a similar way as the human brain processes visual stimuli. Subsequently, these features are used to train classifiers for object recognition. This work examines the feature extraction stage. Its aim is to improve the feature extraction and thereby increase performance of object recognition by computer.

Keywords: Support vector machine, object recognition, visual cortex, feature extraction, Gabor filter

1. INTRODUCTION

Object recognition in images is a broad term which includes detection, localization and categorization of objects. Last mentioned option, the categorization of objects, is aim of this work. Object detection works with one object category of the interest at the time (e.g. detect a human face in the image), object localization takes previous action further and tries to determine exact position of a given object in the image. Finally, object categorization, is a process when we know set of possible categories and our goal is to assign the image into the correct category.

Categories represent objects of the real world (e.g. cars, pedestrians, faces etc.). Images can contain background, other objects (e.g. trees) or main object partly covered (distractors). In addition we have a set of different objects of interest. Overall, it makes object categorization a challenging task. Ideal situation for object categorization is high intra category similarity and high inter category difference. The image 1 shows how similar can two objects from different categories be.

Image 1: An example of low inter category difference from Caltech 101 dataset. Left image from chair category, right image from Windsor chair category.

2. MODEL

Feature extraction from an image used in this work is inspired from [2]. The grounds of the model are simple and complex cells presented by Hubel and Wiesel [3] ordered into hierarchical structure.

This structure is mostly feedforward, because all known facts indicate that this is how the brain works in first few hundreds milliseconds of visual processing [1].

The model consists of 5 layers: **image layer** (input), **S1 layer** (responses on oriented bars and edges just like simple cells in visual cortex), **C1 layer** (corresponds to striate complex cells which hallmark is invariance to the location and scale), **S2 layer** and **C2 layer**.

S1 Layer – Gabor filters. S1 Layer is convolution of image layer with 2D Gabor filters. Important parameter is the number of orientations. Image layer is represented as 3D pyramid of pixels and the S1 layer is a 4D structure. Extra dimension is added for the orientations of Gabor filters. This layer corresponds to V1 simple cells in the brain.

C1 Layer – Local invariance. This layer polls local maximas from S1 layer responses with the same orientation over different scales. Thereby, location and scale invariance is achieved. The result is C1 layer of pyramids (the same number as in S1 layer) with smaller sizes. This layer corresponds to V1 complex cells in the brain.

S2 Layer – Intermediate feature. At each position in the image in all its scales, layer C1 is compared with template by template matching. This is a comparison between the image area from layer C1 and each of *d* templates (*d* – number of templates). These templates represent intermediate features of the model. In the training stage the templates are randomly selected from images in C1 layer. A template is small patch of the size 4×4, 8×8, 12×12 and 16×16 at random position and random scale in the image pyramid.

Smaller patches (4×4) can be seen as encoding of shape, while larger patches (16×16) are more useful for textures (a description of the surface structure or appearance). Considering the random selection at template learning stage, many of these templates do not cover the area of interest or are not suitable for classification. The weighting of templates is left for the classifier (e.g. SVM). It should be noted that the resulting intermediate feature is composed from templates selected from images of different categories.

After template learning stage S2 layer is computed. This gives us *d* pyramids (remind that *d* is the number of templates) representing responses of individual images to templates. S2 layer stands for V4 cortical area or posterior IT (Inferotemporal cortex).

C2 Layer – Global invariance. Building this layer is the last step of computation in this model and leads to a vector of features. Feature vector contains *d* values and is result of choosing global maximum from each pyramid from S2 layer. Thus we loose information about position and scale.

3. RESULTS

The experiments in this work are oriented on tuning parameters of the model which are following:

- number of orientations of Gabor filters (4, 8, 12, 16 orientations from 0 to 180 degrees)
- number of templates/features in S2 Layer (100, 200, 500, 1000, 2000, 5000)

Testing is done on Caltech 101 dataset which contains over 9000 images divided into 101 categories (typically 50 images in one category) and a background category. Most objects are centered and in the foreground with little or no clutter.

For classification is used multi-class SVM, type C–SVC with parameter *cost* set to 32.0. Kernel type is radial basis function with *gamma* (the radius of the gaussian kernel) set to 0.03125. The parameter selection is done experimentally.

The figure 2 presents first results of the work. For four orientations, 1000 features (patches size are from 4×4 to 16×16 randomly chosen) and 15 training images per category is accuracy 32.8%. For the same number of orientations, 2000 features and 30 training images per category is accuracy 44.3%.

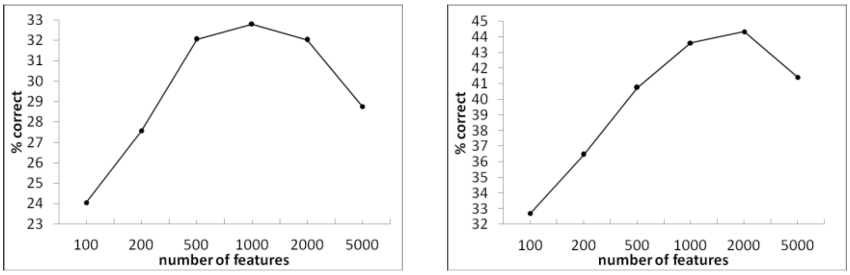

Figure 2: Tuning results for number of features. Left image 15 training images/ cat., right image 30 training images/ cat.

From the shapes of the both curves in the figure 2 we can see relatively quick growth of classification accuracy from beginning (100 features) until peak point (1000 respectively 2000 features). More features mean more information for SVM to find a discrimination hypersurface between the classes. After peak point with growing number of features is classification accuracy relatively slowly falling down. This may be caused by increase of irrelevant information in templates which come from random selection in template learning stage.

The table 1 compares results of this work with other approaches to object categorization on the Caltech 101 dataset [2].

Model	**15 training images/ cat.**	**30 training images/cat.**
this model	32.8	44.3
Serre et al.	35	42
Holub et al.	37	43
Grauman & Darell	49.5	**58.2**
Mutch & Lowe [2]	**51**	56

Table 1: Published results of classification accuracy for Caltech 101 in %.

4. CONCLUSION

Results of this work are positive but are left behind the other works. It can be caused by not optimal setting of SVM classifier. Next steps will be comparison of another classifiers (ANN, k-NN), tuning parameters of Gabor filter or using different filter (e.g. Haar wavelets). In addition to Caltech 101 dataset are planed experiments on Caltech 256 dataset and Letter dataset (own dataset with 26 categories, capital letters from A to Z).

REFERENCES

[1] Serre, T.; Wolf, L.; Poggio, T.: Object Recognition with Features Inspired by Visual Cortex. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'05, Washington, DC, USA: IEEE Computer Society, 2005, ISBN 0-7695-2372-2, s. 994–1000.

[2] Mutch, J.; Lowe, D. G.: Multiclass Object Recognition with Sparse, Localized Features. In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, 2006, ISBN 0-7695-2597-0, s.11–18.

[3] Hubel, D.; Wiesel, T.: Receptive fields of single neurons in the cat's striate cortex. Journal of Physiology, 1959, s. 574–591.