Analysis of features vectors of images computed on the basis of hierarchical models of object recognition in cortex

Ankit Agrawal(Y7057); Jyoti Meena(Y7184)

November 22, 2010

SE 367 : Introduction to Cognitive Sciences

Supervisor: Dr. Amitabha Mukerjee

Department of Computer Science Engineering

IIT Kanpur

Abstract

We tried to analyse the process through which we put objects into different categories.For example when we think of an apple, the characteries of apple like it's shape, color, and taste comes into our mind . Likewise spcific objects defined by specific properties.

Properties of the object goes and store into our mind in the form of Features Vectors. Basis on these specific properties mind can precisely differentiate the object in a image .Clearly the overall process is not known yet, but partially a lot of work [1],[2] has been done on the process of visual categarization in human mind.

Contents

1 Introduction

Here we tried to follow the computational model of object categorization given by Jim Mutch and David G. Lowe [1], in which they try to refine the biologically-inspired model of visual object classification by adding the sparisity and localized features concept in the model of Serre, Wolf, and Poggio [2].

In this model first Gabor filters are applied at all positions and scales of an image; feature complexity and position/scale invariance are then built up by alternating template matching and max pooling operations. Sparsity is increased by constraining the number of feature inputs, lateral inhibition, and feature selection. This refined approach was applied on the database of Caltech 101 object categories[3].

On the basis of MRI data, different kind of computational models has been derived which follows the properties and response of different units in visual cortex(Brain area in which the visual processing happens).We first compute the visual information (features extraction) of a particular object in the image. After that we investigate the response (response matrix given by the model) of these features vectors on images of the same type of object , a different object and a totally different object.

2 Various areas in visual Cortex and their function during the vision

2.1 Primary visual cortex (V1)

The primary visual cortex is the well known visual area in the brain. It is the simplest, earliest cortical visual area. V1 is very sensitive to the pattern recognition by spatial information in vision. During the immediate recognition (40 ms and further) individual V1 neurons have strong tuning to a small set of stimuli(small changes in visual orientations, spatial frequencies and colors).

Function of V1 can be thought of as similar to many spatially local, complex Fourier transforms, or more accurately, Gabor transforms. These filters together can carry out neuronal processing of spatial frequency, orientation, motion, direction, speed (thus temporal frequency), and many other spatiotemporal features. Experiments of neurons substantiate these theories.

2.2 Visual area V2

Visual area V2, receives strong feedforward connections from V1 and sends strong connections to V3, V4, and V5. Functionally, the responses of many V2 neurons are defined by more complex properties, such as the orientation of contours (Qiu and von der Heydt, 2005).

2.3 V3 , V4 & IT

The term third visual complex (V3) refers to the region of cortex located immediately in front of V2. V3 is the third cortical area in the ventral stream, receiving strong feedforward input from V2 and sending strong connections to the PIT(posterior infertemporal area). It also receives direct inputs from V1, especially for central space. V4 is the first area in the ventral stream to show strong attentional modulation.

3 Model

This model based on properties of the ventral visual pathway in an "immediate recognition"(first 200 ms) mode. Within this immediate recognition framework, recognition of object classes from different 3D viewpoints is thought to be based on the learning of multiple 2D representations, rather than a single 3D representation .So, we applied the model on images.

Images are reduced to feature vectors. The dictionary of features is shared across all categories – all images "live" in the same feature space. Our aim is see how the computed features vectors give response change to the same categaries object and to the different categories object.

Features are computed hierarchically in five layers: an initial image layer and four subsequent layers.

3.1 Image layer

The model first convert the image to grayscale and scale the shorter edge to 140 pixels while maintaining the aspect ratio.Then we create 10 scales(factor of between two scales $2^{0.25}$) of the same image using bicubic interpolation and arranged them into pyramid shape.Figure1

3.2 Bicubic interpolation

Suppose the function values f and the derivatives fx, fy and fxy are known at the four corners (0,0), (1,0), (0,1), and (1,1) of the unit square. The interpolated surface can then be written .P(x, y) = $\sum \sum a_i x^i y^j$ for x E {0,3)},y E {0,3}

It preserves fine detail better.

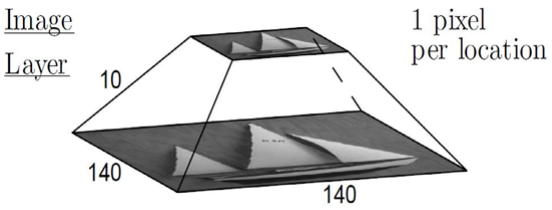

Figure1: Image Layer[1]

3.3 Gabor filter (S1) layer

The S1 layer is computed from the image layer by centering 2D Gabor filters with 4 orientations at each possible position and scale. Computed S1 layer is a 4D (position(2D x,y),scale,orientation). Each unit represents the activation of a particular Gabor filter centered at that position/ scale. Figure 2 This layer corresponds to V1 simple cells.

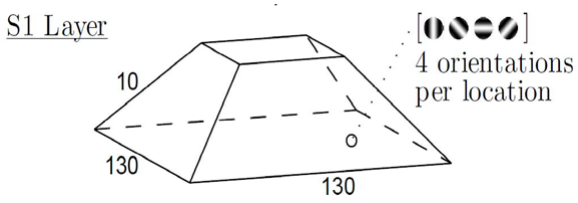

Figure2: Gabor filter S1 layer[1]

The Gabor filters are 11x11 in size, and are described by:

$G(x, y) = \exp\left[-\left(X^2 + \gamma^2 Y^2\right)/2\sigma^2\right] \cos(2\pi X/\lambda)$

where $X = x\cos\theta - y\sin\theta$ and $Y = x\sin\theta + y\cos\theta$. X and Y vary between -5 and 5, and θ varies between 0 and π. The parameters γ(aspect ratio=0.3), σ(effective width=4.5), and λ(wavelength=5.6) are all taken from [7]. We use the same size filters for all scales.

The response of a patch of pixels X to a particular S1 filter G is given by:

$R(X,G) = \left\| \sum X_i G_i / \sqrt{\sum X_i^2} \right\|$

3.4 Local invariance (C1) layer

For each orientation, the S1 pyramid is convolved with a 3D max filter, 10x10 units across in position(x,y) and 2 units deep in scale. A C1 unit's value is simply the value of the maximum S1 unit (of that orientation) that falls within the max filter. The resulting C1 layer is smaller in spatial extent and has the same number of feature types (orientations) as S1. Figure3 This layer provides a model for V1 complex cells. The C1 units within for a 4x4 patch, this means 16 different positions, but for each position,there are units representing each of 4 orientations.So, 4x4 patch means 4x4x4 = 64 C1 unit values.

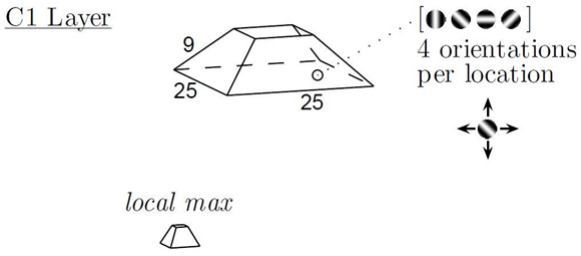

Figure 3 : C1 layer[1]

3.5 Intermediate feature (S2) layer

The S2 layer is intended to correspond to cortical area V4 or posterior IT. The response of a patch of C1 units X to a particular S2 feature/prototype P, of size n × n, is given by a Gaussian radial basis function:

$R(X, P) = \exp\left(-\|X - P\|^2 / 2\sigma^2\alpha\right)$

Both X and P have dimensionality n × n × 4, where n E {4, 8, 12, 16}. The standard deviation $\sigma$is set to 1 in all experiments.Figure 4

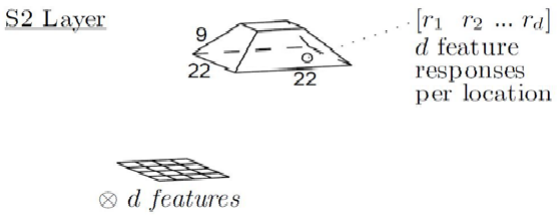

Figure 4 :S2 layer [1]

3.5.1 Sparsify S2 inputs

In brain real neurons are more selective among potential inputs.To increase sparsity among an S2 unit's inputs, we reduce the number of inputs to an S2 feature to one per C1 position.Here we choose the dominant identity and magnitude orientation for each N*N positions in patch.So every resulting 4x4 prototype patch now contains only 16 C1 unit values, not 64.Figure 5

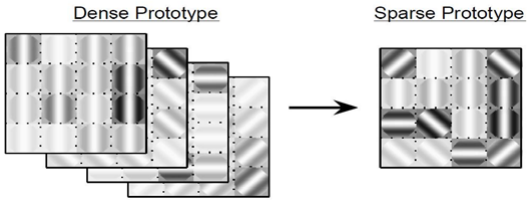

Figure 5 :Sparsify S2 inputs [1]

In conjunction with this we increase the number of Gabor filter orientations in S1 and C1 from 4 to 12. Since we're now looking at particular orientations, rather than combinations of responses to all orientations, it becomes more important to represent orientation accurately.

3.5.2 Inhibit S1/C1 outputs

we ignore non-dominant orientations, we suppressing S1 and C1 unit outputs. In cortex, lateral inhibition(the process whereby nerves can retard or prevent the functioning of an organ or part) refers to units suppressing their less-active

units. At each location, computed the minimum and maximum responses, Rmin and Rmax, over all orientations. Any unit having R $<$ Rmin +h(Rmax $-$Rmin) has its response set to zero.Figure 6

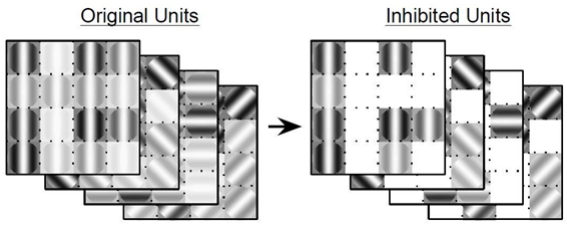

Figure 6 : Inhibited Outputs [1]

3.6 Global invariance (C2) layer

Finally we create a d- dimensional vector, each element of which is the maximum response (anywhere in the image) to one of the model's d prototype patches. At this point, all position and scale information has been removed, i.e., we have a "bag of features".

3.6.1 Limit position/scale invariance in C2

The C2 layer simply takes the maximum response to each S2 feature over all positions and scales. But there could be a false matching of tesing image with training image due to the chance co-occurrence of features from different objects and/or background clutter. So, we want to know some geometric information above the S2 level. In fact, receptive fields of neurons in V4 and IT known are limited to only a portion of the visual field and range of scales [4].

To model this, visual field of given S2 features field is restricted within the region of ,relative to its location in the image from which it was originally sampled, to $\pm$tp% of image size and $\pm$ts scales, where tp and ts are global parameters.This approach assumes the system is "attending" close to the center of the object. This is appropriate for datasets such as the Caltech 101, in which most objects of interest are central and dominant.

4 Multiclass experiments (Caltech 101)

1. We chose training images at random from each category, placing 3 images(First of which is in the same catagories,second one is different catagories and Third one is blank page) in the test set,

2. Learn features vectors(C2) at random positions and scales from the training images (an equal number from each image),

3. Build C2 vectors for the 3 different test images test and classify the test Images basis on response matrix.

5 Results

We have derived some response matrices for 10 different categories of objects with 8 training images and 3 different test images for each category. Response matrix implies that how close the test set is to the training set. It represent the similarity on the scale of 1 and gives 1 for exact matching and 0 for not matching at all.

The response matrix curve for one category(Revolver) is given as follows:-

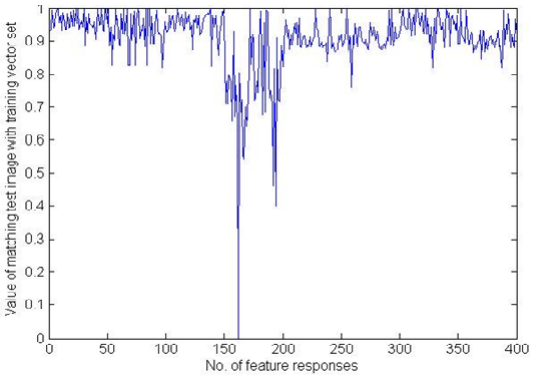

Figure 7 : Response matrix of revolver with revolver.

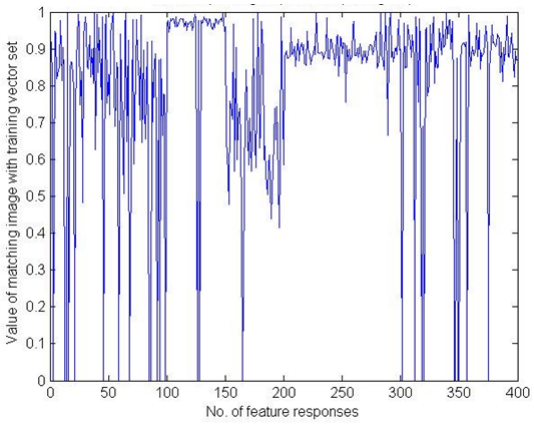

Figure 8 : Response matrix of a revolver with boat.

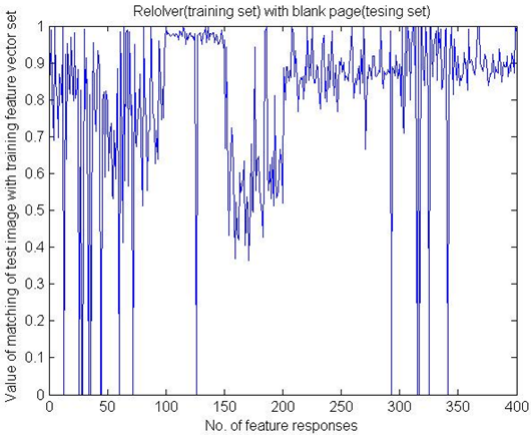

Figure 9 : Response matrix of a revolver with blank image.

Likewise we have done it for 10 catagories of objects and find out their response matrix in tabular form with clearly shows that same categories object shows more correlation.

Response matrix of different objects in tabular form clearly shows that same categories objects are more correlated. Then, we computed the difference(D) between the test set and training set by adding the difference in response matrix at every position (i.e. for every feature) from 1. On the basis of observed data We can digitalize the result as:-

If $D > 50$, Object test image does not belonges to that categories (no)

If $D <= 50$, Object test image belonges to that categories (yes)

Tabular form of different object categories compare with 3 different kind of test images.

Training Images	Same Object test image	Boat test image	Blank page test image
1)Dollar	35.6939 (yes)	102.2775 (no)	125.3542 (no)
2)Menorah	48.3456 (yes)	76.6608 (no)	90.1453 (no)
3)Revolver	34.4904 (yes)	66.9993 (no)	74.0945 (no)
4)Airplane	35.7314	128.2040 (no)	138.8277 (no)
5)Buddha	43.9205 (yes)	50.8058 (no)	76.1193 (no)
6)Camera	55.4419 (no)	77.4703 (no)	113.3936 (no)
7)Chair	58.9212 (no)	75.2972 (no)	100.2698 (no)
8)Cup	48.3616 (yes)	54.2534 (no)	66.5087 (no)
9)Barrel	39.2537 (yes)	47.5643 (yes)	64.0353 (no)
10)Lamp	23.0590 (yes)	41.3912 (yes)	46.1661 (yes)

As we can see from the above results that images containing more specific features (airplanes, chair ,dollar ,menorah ,revolver) can be easily distinguished from the other object images. Similar trend can be observed for the biological visual system. In biological visual system also it is easier to categories the object with more specific features. While the features of more plane objects like barrel,

lamp and camera are not clearly define the object in vector space.Figure 10

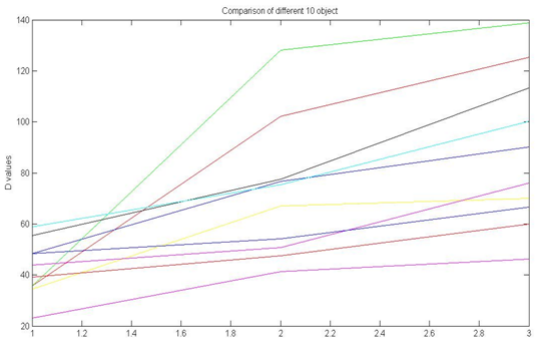

Figure 10 : Graph showing comparison of Different object's D values.

6 Conclusion

Both biological and computer vision systems face the same computational constraints .So, We can expect computer vision research to benefit from the use of similar basis function for describing images. In other words,this approach strengthening the case for investigating biologically-motivated approaches to object recognition.

7 References

- [1] Jim Mutch and David G. Lowe. *Object class recognition and localization using sparse features with limited receptive fields.* International Journal of Computer Vision (IJCV), 80(1), pp. 45-57, October 2008.
- [2] Serre, T., L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio. *Object Recognition with Cortex-like Mechanisms*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29, 3, 411-426, 2007.
- [3] Image Dataset, Caltech, *Computational Vision.*
- [4] E. T. Rolls and G. Deco. *The Computational Neuroscience of Vision.* Oxford University Press, 2001.