Evaluation of Gabor filter enhanced Hierarchical Temporal Memory in image classification applications

Dipan Kumar Pal
Dept. of Electronics and Communication Engineering, Birla Institute of Technology, Mesra, Ranchi, India
dipan2912@gmail.com

Abstract - **This paper gives an evaluation of the combination of two biologically inspired tools i.e. hierarchical temporal memories (HTM) and Gabor filters, when applied to a problem that the biological systems they model have solved, namely image classification. Hierarchical temporal memory (HTM) is an under development computational model of the human neo cortex whereas Gabor filters which utilize Gabor functions that can model simple cells in the visual cortex of mammalian brains. Gabor filters when combined with HTMs form a mathematical model of the visual cortex capable of performing robust image classification. We test the complete model over an image classification dataset i.e. the Corel 1000 image and a 1440 image face recognition dataset. The results obtained are very promising giving 99.8 percent and 100 percent accuracy respectively of the two datasets. This was achieved with comparatively very few training examples for each category and face.**

Keywords – **Hierarchical temporal memory, Gabor filter, image classification, face recognition, HTM**

I. INTRODUCTION

Image classification has long been a problem which tests the capability of a system to understand the semantics of visual information within an image and to develop a model which can store such information. The system needs to effectively extract important feature information which when combined with knowledge gained by training allows the system to perceive the object in the image. Thus image classification can be said to also have the potential to make use of efficient knowledge representation techniques. The arguments presented above allow us to split the image classification problem into two sub-problems whose combined solution would enable us to solve the parent problem.

The first one is the problem of effectively extracting feature information such as edges and higher level information such as texture. This problem has been tackled from many different perspectives. But it is logical that when we aim to mimic the image classification capability of mammalian brains, we also might study and implement the processes occurring in them which make the feat possible. Our target area is the visual cortex with its simple and complex cells. The behavior of these cells can be modeled using Gabor functions [6]. It was also found that the real part of the complex Gabor functions appear to be very close to the receptive field weight functions in a cat's striate cortex [5]. Moreover, a bank of Gabor filter functions varying in orientation when convolved with the input signal created what is called a Gabor space. This process appeared to be very similar to processes occurring in the primary visual cortex [7]. These behavioral similarities between the biological systems and the Gabor filter make it a promising candidate for the solution of the first sub-problem mentioned earlier.

The second sub-problem is the representation of the visual knowledge that an image contains. To try and solve this problem, we must understand that when we perceive an image that has rich visual information in it, we are interpreting it using our own knowledge database. Thus to develop a system capable of building the knowledge database we use a hierarchical temporal memory (HTM). The HTM is a machine learning technique which builds such a database by modeling the world it observes.

There has been previous work on the use of Gabor filters along with a supervised classifier using a modified minimum distance classifier for face recognition [8]. The classifier they used was termed as minimum average distance classifier. It did employ a kind of clustering but it was completed in a single stage. Multistage clustering methods like HTMs are shown to extract more information from an image and thus offer more robust classification given smaller training sets.

We present in this paper, the results of classification and recognition experiments conducted on two different datasets. The first dataset is the Corel 1000 image dataset consisting of 10 categories with a 100 images in each [11]. Whereas the second dataset in a face recognition dataset [10] consisting overall of 72 different faces with 20 images per face. The samples have mild changes in expression and lighting. The experiments are explained in a more detailed way in a later section. The HTM and Gabor filter used were part of the Numenta Vision Framework [3]. The very promising results we obtained serve to show us the effectiveness of the two models. It might also be intuitive to expect so, because they both are individually mathematical models of parts of a biological system which has already solved the parent problem.

II. GABOR FILTER

This section gives a brief overview of Gabor filters, there relevance to image processing and explains their compatibility with HTMs. Gabor filters are very popular in face recognition though they have also been used in general image processing applications. These applications also include other domains such as image smoothing, image coding, texture analysis, shape analysis, edge detection, fingerprint and iris recognition [8].The Gabor filter is a band-pass linear filter with its impulse response defined by a harmonic function multiplied by a Gaussian function. Thus, a bidimensional Gabor filter constitutes a complex sinusoidal plane of particular frequency and orientation modulated by a Gaussian envelope [12].

Gabor filters are usually used as banks of multiple filters with varying scales and orientations. When these filters are convolved with the input signal or pattern, they generate a Gabor space. The Gabor space has a useful property that the activations of spatial locations in an image are very distinct for different objects within the same image. This allows for easier information extraction. Also only important activations may be extracted from the Gabor space in order to create a sparse representation of the object. This property is extremely important for the filter to be compatible with HTMs. The reason as we shall see is because the mechanism of learning in an HTM is specialized for sparse representations which help in efficient storage and generalization. It also reduces the input space for HTMs which makes computation more feasible. The above mentioned properties make Gabor filters very compatible with HTMs and this fact is again very intuitive.

The formal definition of a 2D Gabor filter in the spatial domain is

$$\psi_{u,v}(x,y) = \frac{f_u^2}{\pi\gamma\eta} e^{-(\frac{f^2}{\gamma^2}x'^2+\frac{f^2}{\eta^2}y'^2)} e^{j2\pi f_u x'} \quad (1)$$

Where $x' = x\cos\theta_v + y\sin\theta_v$, $y' = -x\sin\theta_v + y\cos\theta_v$, and the parameters f_u and θ_v are defined as $f_u = f_{max}/2^{(u/2)}$ and $\theta_v = v\pi/8$. The Gabor filter's center frequency and orientation are defined by f_u and θ_v. The parameters γ and η determine the ratio between the center frequency and the size of the Gaussian envelope [9].The Gabor filter's popularity may be justified by its computational properties and also its biological relevance as mentioned earlier.

III. HIERARCHICAL TEMPORAL MEMORIES

A. Relevance of HTMs in image classification

An image to a computer is just a sequenced collection pixel intensity values. If say we were given a randomly generated computer image, we would most probably not understand it. To recognize and classify the content of an image, we might argue that we must 'know' the content of the image. But then it raises the question that by what mechanism should the system gather the required knowledge. In [2], Hawkins and George describe such a theory which resulted in the development of such a mechanism i.e. the HTM.

There have been no assumptions in the design of the HTM algorithm about the data except for the fact it must also have been generated by a hierarchical system [1] and [2]. Fortunately this is true of most data especially images. Any image consists of combinations and definite sequences of lesser complex structures. These further can be decomposed into simpler structures such as a line or corner. This hierarchical structure of data is what makes an HTM possible. Hence a way to tackle the knowledge representation problem is to repeatedly break down complex patterns into their constituent simpler patterns and remember sequenced combinations of those. We continue to give a more detailed description of the general structure and functioning of an HTM.

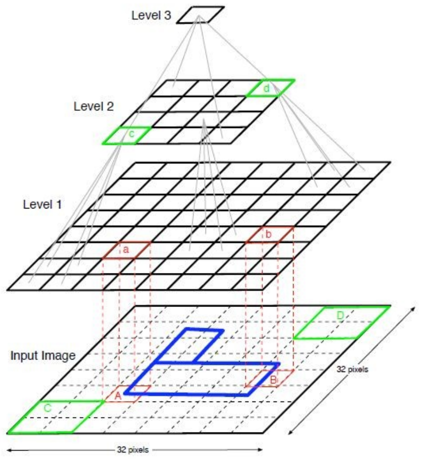

Fig. 1. General structure of a 3 level HTM

B. Description of the HTM

The Hierarchical Temporal Memory (HTM) is an algorithm which tries to capture the data modeling and processing capabilities of the human neocortex. HTM is similar to Bayesian networks which use belief propagation, but they are self-training and are easier to handle. The algorithm essentially uses clustering mechanisms to achieve invariance in output when an input belonging to a particular class is presented to the network. It does this by forming a spatial temporal correlation between low level input patterns which appear to the network. Thus knowledge and understanding about the HTM environment is only gained with what the HTM perceives as input.

HTMs in general are a tree structured multi-leveled hierarchy with each level consisting of a region of nodes. A typical 3 level HTM is shown in Fig. 1 .An HTM can consist of any number of levels, but for most applications a 2 or 3 level network suffices. Each level consists of a fixed number of nodes all of which perform the same algorithm. The bottom most level of the HTM is fed with the raw input data, which in this case is the output of a Gabor filter fed with a RGB color image. Each node performs clustering in overall three dimensions and it does this in two stages. The first stage is called the spatial pooler and the second one is the temporal pooler.

TABLE I
MAJOR PARAMETER CHANGES FOR THE HTM USING THE NUMENTA VISION FRAMEWORK

HTM parameters	Values
numCategories	10
seed	24
midLevelPatches	160
gaborNumOrients	(is varied)
gaborPhaseMode	'single'
gaborCenterSurroud	False
spatialPoolerAlgorithm	kthroot_product
maxDistance	0.3
temporalPoolerAlgorithm	maxProp
spatialPoolerTraining Algorithm	RandomFlash
temporalPoolerTrainingAlgorithm	MultiSweep

As the name suggests, the spatial pooler pools or clusters data in the spatial dimension. Each pattern appearing at the input during learning of the spatial pooler is compared with the database of other patterns, if the distance between the input pattern and each is less than the maxDistance parameter, then the input pattern is considered same as the corresponding existing pattern, termed as a coincidence. If the previous condition does not satisfy, then the input pattern is "memorized" as a new coincidence. Thus the spatial pooler quantizes the input space but only remembers the patterns which appear. The temporal pooler performs clustering over time and forms temporal groups of coincidence patterns. These groups are formed on the basis of the statistical behavior of the input data, which is captured using a Markov graph whose nodes are the coincidence patterns learned previously. Hence, the members of a temporal group are likely to follow one another. After training, a vector of probabilities of membership of the input pattern to each of the temporal groups is the input to the next level of nodes. Therefore, the overall effect of this approach causes the lower level nodes to remember and recognize patterns of lower complexities such as a line or corner. As we ascend the hierarchy, we find that the coincidences represent combinations of patterns of lower complexities. This increases the variance and complexity of data represented at higher levels. But in spite of the seemingly large input space at higher levels, the spatial pooler at higher levels only remembers patterns it encounters thereby improving efficiency.

HTMs can be run in two modes, the learning mode and the inference mode. In the learning mode, a level tries to find new coincidences and keeps updating the Markov graph time progresses. In inference mode, the probability distribution of the membership of the input pattern is outputted to the next higher level. The learning mode provides no such output. During training of a particular level, all levels below it are run in inference mode and it itself is run in the learning mode.

IV. EXPERIMENTAL DETAILS AND RESULTS

All experiments were conducted using the Numenta Vision Framework implemented with NuPIC1.7 [3] and [4]. The experiments were conducted in two major sections; the first was training and testing of the HTM along with the Gabor filter on the 384x256 RGB Corel 1000 dataset. The second one was on a 200x180 RGB face recognition dataset which had images of the faces of 72 individuals with 20 images per face. The dataset along with experimental results are discussed next.

A. Structure of the HTM used

The HTM consisted of 5 levels. The first level consists of a Gabor filter with a receptive field which receives the input image from the image sensor provided with the Numenta Vision Framework. The Gabor filter can be said to work as a spatial pooler since it too clusters the input space and reduces it. Its output is sent to a temporal pooler region. The spatial pooler and temporal pooler are considered to be different levels in the convention of the Vision Framework. The output of the temporal pooler is sent again to a spatial pooler region followed by a temporal pooler region up the hierarchy. This completes 4 levels of the hierarchy. But effectively the hierarchy forms only a 2 level HTM network. The top level is the classifier node which outputs a probability distribution of membership of the image in each of the categories.

B. Image classification on the Corel 1000 dataset

The Corel 1000 image dataset was primarily intended for research on content based image retrieval, but the variations in the image within each individual category make it very favorable for testing with the Gabor filter enhanced HTM network. The dataset as mentioned earlier has 10 categories with 100 images each. The experiments were carried out on the Numenta Vision Framework with major parameters changes to defaults of the HTM are in Table I.

Multi sweep was the training algorithm chosen for the level 2 and level 4 temporal pooler nodes, whereas the default training algorithm was found to give good results for the level 3 spatial pooler nodes.

TABLE II
RESULTS KEEPING NUMBER OF ORIENTATIONS OF THE GABOR FILTER = 2

Number of training images per category	Accuracy in %	Level 2 coincidences (temporal)
3	98.6	936
6	98.9	936
9	99	936
12	97.3	936
15	98.1	936

TABLE III
RESULTS KEEPING NUMBER OF TRAINING IMAGES PER CATEOGRY = 3

Parameter gaborNumOrients	Accuracy in %	Level 2 coincidences (temporal)
2	98.6	936
5	99.6	2340
10	99.2	4680
15	99.8	7018
20	99.9	9358

There were two experiments designed to emphasize on the importance of the Gabor filter along with the HTM. The first experiment measures accuracy in percentage for an increasing amount of training images per category while keeping the number of orientations of the Gabor filter constant. The results are depicted in Table I. In the second experiment, we kept the number of training to just three images per category and increased the number of orientations of the Gabor filter. Table II depicts the results of this experiment. In all experiments, the coincidence counts on level 3 and level 4 were found to be 160 and 640 respectively with 160 temporal groups in level 4. Fig. 2 and Fig. 3 show some of the correctly classified images from the Corel 1000 and faces95 dataset respectively.

We find some random behavior of accuracy upon increasing the number of images. But this is simply because the inherent statistics of the data changed with changing the number of testing images. We must understand that the performance of the HTM essentially remained the same because the number of coincidences observed and groups formed were constant.

When the number of orientations of the Gabor filter was increased the HTM grouped the coincidences together for each of the orientation. Each orientation of the Gabor filter gives a convolved output which has components from the image in only that orientation. This creates high temporal coherence among patterns within each orientation. The temporal pooler groups them together thereby increasing the number of groups. More groups allow the HTM to learn and classify more varied coincidence patterns as allowed by more varied outputs from the Gabor filter, resulting in a positive effect on recognition performance. Hence we can infer that the number of learnt coincidences is directly related to the amount of knowledge stored in the HTM.

Also, at lower levels, the HTM is only able to distinguish between the various low level image feature components. It is only at higher levels that it is able to form spatial and temporal groups that store knowledge about the distinction between the high level categories. Moreover, due to efficient clustering and thorough training algorithms such as MultiSweep and ExhaustiveSweep in the Vision Framework, enough coincidences are generated from the image space to help the Markov graph to model the statistical behavior of the input. This in the end enables us to drastically reduce the number of training images required for good accuracy and also provides a big advantage in scalability. The system when scaled up to larger datasets with many more categories would require only relatively small increase in training images.

TABLE IV
RESULTS ON FACE RECOGNITION DATASET KEEPING NUMBER OF ORIENTATIONS OF GABOR FILTER = 10

Number of training images per face	Accuracy in %	Level 2 coincidences (temporal)
5	99.5	1680
3	99.7	1680
1	100	1680

C. *Face recognition*

The experiments conducted on the face recognition dataset were very similar to the ones conducted on the Corel 1000 dataset. These experiments were primarily conducted to support the qualitative analysis of the results of the Corel 1000 tests and also to show that the HTM can efficiently distinguish between categories even when the nature of the categories are very similar (each category represents a different face, but each face has the same structure).

The face recognition dataset is publicly available online as faces95 [10]. It has samples of 72 individuals with a total of 20 images per face. For the experiments, the platform and parameters were the same as mentioned earlier. Similar experiments were carried out with one increasing the number of training images per face keeping number of orientations of the Gabor filter constant and the other increased the number of orientations keeping the number of training images per face constant at 1 image. The results obtained are shown in Table IV and Table V respectively. The fact that the system showed 100 percent recognition rate for the entire dataset with only 1 training image per face is very encouraging and shows that the higher levels of the HTM are able to find spatial and temporal discrepancies between facial features of different faces even though the there are only subtle low level differences between them.

Another reason that the two datasets chosen for testing are so different is because the first dataset had a more varied high level input space i.e. the categories. Whereas the second dataset had very little variation at higher levels i.e. all categories had the basic structure and enough similarity for face detection algorithms to exploit them. But the HTM was still able to discover classify using low level variations. This shows the versatility of the Gabor filter enhanced HTM system, which can independently and easily discover spatial and temporal

TABLE V
RESULTS ON FACE RECOGNITION DATASET KEEPING NUMBER OF TRAINING IMAGES PER FACE = 1

Parameter gaborNumOrients	Accuracy in %	Level 2 coincidences (temporal)
2	99.4	336
5	99.9	840
10	100	1680
15	100	2520

coherence between coincidence patterns at any level of data abstraction.

V. CONCLUSION AND FUTURE WORK

The results show that the Gabor filter enhanced HTM system offers very robust image classification. The results also support the initial intuition that the system will perform excellently due to their biological relevance. The very high accuracies promise very robust image processing systems incorporating HTMs in the future. Hence further work into the theoretical modelling of the HTM along with optimising its performance based on a given training set might be very useful if successfully undertaken. Another opportunity for research might be on the problem that given a parameterised HTM and a complete dataset, what is the number of training samples to be chosen and which statistic or measure to maximise, such that the ‘knowledge’ contained within the HTM is maximised. These problems deviate from the image processing domain, but recent work has shown that the human brain solves most of the problems with the same algorithm [2]. Hence to solve any classification problem or knowledge representation problem in any domain, it might turn out to be useful to first understand and develop a detailed theory the hierarchical nature of world data itself. This might be a hardcore artificial intelligence problem but it has been long known that image processing and artificial intelligence are deeply connected.

Numenta is also going to release a new version of HTMs employing the new sub-cortical algorithms. Those HTMs will also feature prediction of coincidences. The paradigm that it will provide in data and especially image processing will also be interesting to research upon.

REFERENCES

[1] D. George and B. Jaros, “The HTM Learning Algorithms”, Numenta Incorprated, March 1, 2007.
[2] J. Hawkins and D. George, “Hierarchical Temporal Memory: Concepts, Theory, and Terminology”, Numenta Incorprated , March 27, 2007.

Fig. 2. Correctly classifed examples from the Corel 1000 image database

Fig. 3. Correctly classified samples from the face recognition dataset

[3] “VisionFrameworkGuide”, Numenta Incorprated, August 2009.
[4] “nupic_gettingstarted”, Numenta Incorprated, September 2008.
[5] J. P. Jones and L. A. Palmer. “An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex.” J. Neurophysiol, 58(6):1233-1258, 1987.
[6] J. G. Daugman. “Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters.” Journal of the Optical Society of America A, 2(7):1160–1169, July 1985.
[7] J. G. Daugman. "Two-dimensional spectral analysis of cortical receptive field profiles", Vision Res. 20 (10): 847–56, PMID 7467139, 1980.
[8] T. Barbu. "Gabor Filter-based Face Recognition Technique", Proc. of the Romanian Academy, Series A, Volume 11, Number 3/2010, pp. 277–283.
[9] L. Shen, L. Bai, and M. Fairhurst, “Gabor wavelets and general discriminant analysis for face identification and verification,” Image and Vision Computing, vol. 25, no. 5, pp. 553–563, 2007.
[10] faces95 face recognition dataset.
[11] Corel 1000 image database.
[12] J. R. Movellan, “Tutorial on Gabor filters”.