2015 IEEE 2015 International Conference on Computer, Communication, and Control Technology (I4CT 2015), April 21 - 23 in Imperial Kuching Hotel, Kuching, Sarawak, Malaysia

# Feature Extraction using Neocognitron Learning in Hierarchical Temporary Memory

Aseel Mousa
School of Computing
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 Sintok, Kedah, MALAYSIA
assiso1970@yahoo.com

Yuhanis Yusof
School of Computing
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 Sintok, Kedah, MALAYSIA
yuhanis@uum.edu.my

*Abstract*— **Hierarchical Temporal Memory (HTM) serves as a practical implementation of the memory prediction theory. In order to obtain the optimum accuracy in pattern recognition, it is crucial to apply an appropriate learning algorithm for the feature extraction step of the HTM. This study proposes the use of neocognitron learning in extracting features of the pattern for image recognition. The integration of neocognitron into HTM addresses both the scale and time issues of the HTM. As for evaluation, a comparison is made against the original HTM and principal component analysis (PCA). The results show that more features are extracted as a function of input patterns than the original HTM and PCA.**

***Keywords*- Hierarchical temporal memory, Neocognitron neural network, Pattern recognition.**

## I. INTRODUCTION

In the modern day, machine learning techniques are found in many applications, such as object and face recognition, videos and images, speech recognition, or medical images, etc. [1]. Pattern recognition approaches to an image analysis, is as a reflecting of how the brain recognize visual patterns [1]. A prior knowledge or the statistical information extracted from the image are used for recognizing an image, which called feature [2].

Feature extraction is used for reduce the dimensions of an image that efficiently represents interesting parts of an image as a compact feature vector [3]. Feature extraction techniques are powerful tools which can significantly increase the classification accuracy [4]. In [4] the author has demonstrated that feature extraction increase classification accuracy, by applying different feature extraction techniques as method of dimensionality reduction, and constructive induction, analyzed with respect to the performance of a classifier. In [5], the authors consider the feature extracting process the most important phase in any speech recognition system. As its objective is to find the features that may help the system to differentiate between voices. They evaluate different feature extraction techniques to find the most suitable one, and then enhance the result by using principal component analysis (PCA). Neural network is used in [6] as feature extracting technique. It has the problem of overlapped data. Other techniques like Principal Component Analysis (PCA), which is used in [7] to recover a specific optimal low-dimensional subspace, still suffer from noise in the image. The Gabor filter [8] which is widely used in feature extraction is also sensitive to noise and distortion. On the other hand, the Lip Geometry Estimation (LGE) method as employed in [9] for speech recognition, generates low performance when the signal to noise ratio is high.

Memory prediction theory and its practical implementation, Hierarchical Temporal Memory (HTM), on the other hand, is a machine learning model innovated by Jeff Hawkins and Dileep George of Numenta, Inc [10]. This models the algorithmic and structural properties of the neocortex, hence suitable to reflect the human brain structure [11]. HTMs organized as a tree hierarchy of nodes [12]. The nodes train data that serves for recognizing the objects, by its hierarchical structure and temporal relations among data [13]. It differs substantially from traditional neural networks [14]. In HTM, feature extraction depends on the temporal analysis of pattern sequences while in neural network, spatial analysis is the basis of recognizing a pattern. Nevertheless, both are simple in comparison to the massive complexity of the human brain that contains about one hundred billion neurons [15]. Despite of such strength, there are issues in HTM that need to be addressed which includes the shift in position of the input image [16]. Such a drawback leads to unknown pattern being recognized as known object. To overcome the problem, it requires effort in identifying suitable feature extraction method to employ in HTM. The neocognitron neural network, which has been proposed by Fukushima in 1980, can recognize the visual patterns robustly through learning [17], so it is suitable to use for this research.

Based on above literature, the choice of feature extraction technique is a major factor affecting the accuracy of the recognition process. Neural network is suitable for feature extracting, and it represents the hierarchy feature of the human brain. However, it is a spatial technique. Hence, it does not have the sense of time [12] which is the record of past events [14]. So, there is a need to combine computational approach that is capable of processing noise and has a hierarchical architecture for pattern recognition This study proposes the use of an enhanced neocognitron [18] as feature extraction step to improve the HTM performance..

## II. AN OVERVIEW OF HIERARCHAL TEMPORAL MEMORY

Hierarchical Temporal Memory (HTM) as a practical implementation model of the memory prediction theory is a machine learning technology that aims to explain the structural and algorithmic properties of the neocortex [2]. This refers to a neocortical bio inspired algorithm for pattern recognition, the control consisting on a network of spatiotemporal pooling nodes and time series prediction arranged in a hierarchy. HTM is a good classifier for cases in which the spatial representation of an input is present at one point in time, such as image recognition. However, HTMs suffers from that when the inputs to the model are composed of a spatial structure evolving over time.

There are several application areas which the researchers use the HTM model to solve the recognition problems. The Hierarchical Bayesian Reservoir Memory proposed in 2009 [3] uses a simple stochastic gradient descent learning algorithm to organize and learn multi-scale spatiotemporal features of the input pattern in an unsupervised manner in a hierarchical structure to provide robust and real-time prediction of future inputs. There is also the Handwritten Digit Recognition [4], Invariant Pattern recognition in the Visual Cortex [2], Pattern Recognition by Hierarchical Temporal Memory[5], Object Identification in Dynamic Images Based on the Memory-Prediction Theory of the Function of Brain [6], and the optimization of Hierarchical Temporal Memory [7]. On the other hand, the Neocognitron convolutional neural networks [8] are examples of models that have pooling operations and network architectures similar to the HTM model. However, the pooling techniques in these models were hard coded and not learned. The structure of HTM that is tree-shaped and multi-level reflects the nested hierarchical structures found in the world, also these hierarchies exist in both spatial and temporal dimensions as in Fig. 1. It consists of three operations [9]:

- Feature extraction
- Learning transition probabilities
- Temporal grouping

In related work based on improving feature extraction step of HTM, every node performs a dot product between the each vector of the already stored and the new one that applied in [10]. Also, reservoir computing methods have been presented in [3] to overcome the problem of local minima, The reservoir enhance the separability by put the input into a high-dimensional space. The quantization centers are the stored input vectors that encode the knowledge of the network in [7] to get performance enhancement in recognition, but still suffer from with noisy images.

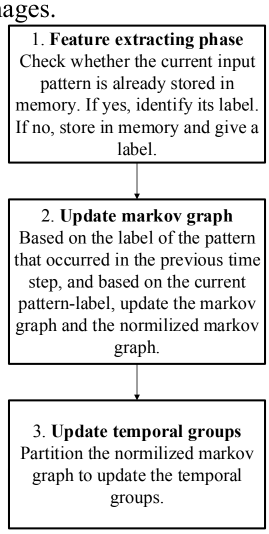


Figure1. Operations in a node during learning

## III. NEOCOGNITRON

Neocognitron is a multi-layer hierarchal neural network proposed by Fukushima [11] where its architecture was initially suggested by neurophysiological findings on the visual systems of mammals [12]. It have the ability to recognize visual patterns strongly through learning. The neocognitron consists of two types of layers: S-cells layer, which contain the simple cells of the visual cortex, they are feature-extracting cells, whose input connections are variable and are modified through learning. The second type is C-cells, which contain complex cells, whose input connections are fixed and unmodified [13]. The C-cells in the highest stage serves as recognition cells, which contain the result of the pattern recognition. These layers of are arranged in a hierarchical manner in an alternate manner [8]. Each S-cell competes with the other cells in its vicinity when every time a training pattern is presented to the input layer. If the output of the cell is larger than the other cells in the competition area [24], the cell is selected as the seed cell. The output of S-cell is as follows: Let $a_i$ be the strength of the excitatory variable connection to an S-cell from the $i^{th}$ C-cell, whose output is $x_i$; and b be the inhibitory variable connection from the V-cell Whose output is $v$, Also let $c_i w_i$ be the strength of the excitatory connection to the inhibitory V-cell from the $i^{th}$ C-cell. Whose output is $v$, the output of s-cell is given by:

$$u_{Sl}(k,\mathbf{n}) = \frac{\theta l}{1-\theta l}\cdot\varphi\left[\frac{1+\sum_{k=1}^{Kc_{l-1}}\sum_{v<AS_l} as_l(v,k,K)\cdot u_{Cl}(k\ ,\mathbf{n}+v)}{1+\theta\cdot bs_l(k\ )\cdot v_l(\mathbf{n},\mathbf{k})}-1\right] \tag{1}$$

Where φ[] is a function defined by φ[x]=max(x,0), the value of *asl(v,k,K)* is the strength of excitatory variable connection coming from the preceding stage. For $l$=1; however, $u_{cl-1\ (n,k)}$ stands for , $u_G\ (n,k)$; It is important to mention here that all cells in a cell-plane share the same set of input connections, hence $as_l(v,k)$ is independent of n. $as_l$ denotes the radius of summation range of n; that is, the size of spatial spread of input connections to a particular S-cell. The positive constant $\theta l$ is the threshold of the S-cell and determines the selectivity in extracting features.

## IV. METHODOLOGY

This study focuses on improving the first step of the standard HTM model, which is the feature extraction. In this section, description on the utilized dataset and experimental setup are described.

### A. Data Collection

The images used are blood type images represented as grayscale and saved as JPEG of size 32x32 pixels. The data set is of 10 blood type images, each image is divided into portion of 4 x 4 pixels. Each level node's input corresponds to one of 4x4 portion of a vector of 16 components and has 64 node arranged in an 8x8 grid, i.e. each image constructed from

64 portion, each of 4x4 pixel size. The S-layer neocognitron network is the same parameters as the one reported by Fukushima (2013), for the extracting level in neocognitron, which is the first layer ($U_{s1}$); chose θ= 0.55. The size of S-layer is 8x8x16 i.e. the S-layer is constructed from 16 cell-plane, each of 8x8 cells (64 nodes). Fig. 3 shows an example of input image, while Fig.4 illustrates example of representation of 4x4 portions of the input image. The nodes are arranged in 32x32, and each level-1 node receives a 4x4 pixel portion of input image. With such representation, the utilized dataset includes a total of 640 portions, each considered as an input pattern.

### *B. Improved Feature Extraction Phase*

The learning process first step, the node extracts the features of the input patterns that were in its receptive field. In general, this can be considered as a vector quantization process of the input data. This process is called the spatial pooling in the HTM [23]. Feature extraction phase in the proposed method contains the implementation of the equation of S-layer of neocognitron neural network with creating of time factor, for further processing in HTM. The input stimuli presented to the input layer of HTM (S-cell layer) is of 4x4 image part. All cells in a cell-plane receive the same set of input connections. When the input stimuli presented to the input layer, a time matrix is created. Each element of the time matrix corresponds to one input vector. Then the input stimuli will be compete to its neighbor cells in its plane. The largest value of competition will be the winner cell that extract the feature. The S-cell becomes a winner cell if the same input pattern is presented to it. Ones the winner-cell is selected from a cell-plane, the input connections of that cell modified. The connections are modified in a way that make the cell responds more strongly to the training patern to which the cell becomes a winner. Otherwise, a new plane is created in the S-layer, and the input pattern is stored in a new cell in this plane as a training pattern. Then the output of winner cell is computed (*Us*). The largest output *Us* of these winner cells will train the input stimuli and send the feature that extracted by this cell to the next stage of HTM. If the patterns are stored in a node as a rows of a matrix, let call it A, each row of the A matrix is a specific pattern. The specific patterns will be named as $a_1$; $a_2$ etc., depending on the location of the pattern in the matrix (rows in the matrix that patterns are stored). The elements of time matrix, T, corresponds to each pattern are $t_1$, $t_2$ etc., $t_1$ corresponds to $a_1$ and so on.
The input matrix and time matrix is as below:

$$A = \begin{bmatrix} 10110010 \\ 00100110 \end{bmatrix} \qquad T = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Output of s-cells will become:

$$U_w(t) = \frac{1}{1-\theta} . \varphi[\Sigma a_n(t).X_n(t) - \theta_v] \qquad (2)$$

Where φ[] is a function defined by max max(x,0) [11], *n*, represents the location of the receptive field center of the cells, $a_n$ is the strength of the excitatory connection to the S-cell. The positive constant $\theta$ is the threshold that controls the S-cell and the strength of variable inhibitory connection coming from the V-cell. The outputs of V-cells are given by:

$$V = \sqrt{\sum c_n x_n^2} \qquad (3)$$

The value of $c_n$ is used as a kind of weight

Matrix A is presented to the input layer in parallel with time associated with each pattern, the flow of improved feature extraction method illustrated in Fig. 2.

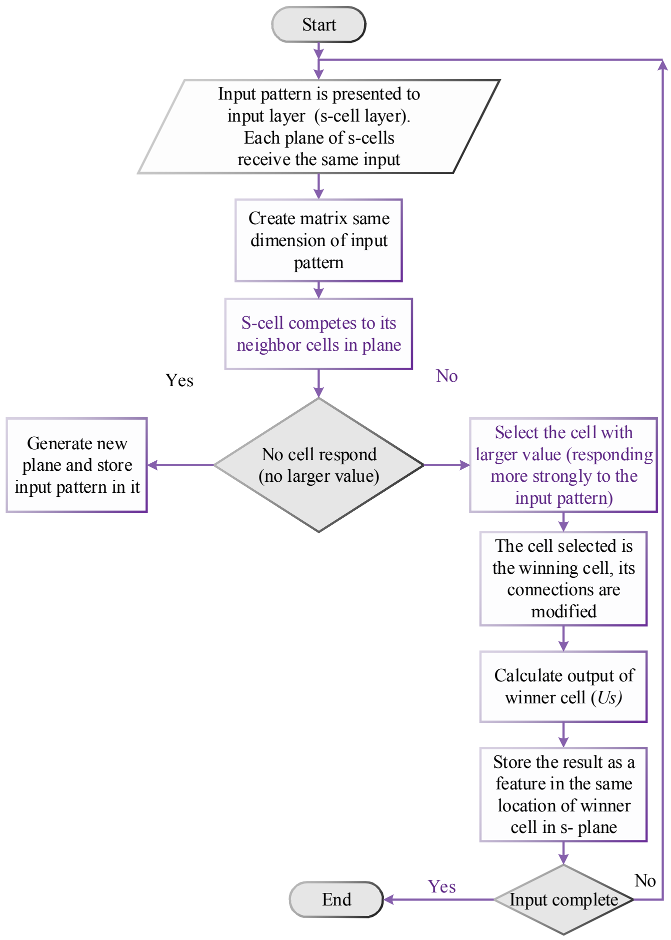


Figure 2. Feature extraction in HTM

### *C. Experimental results* `

In this section, results of the undertaken experiments are presented. The illustration in Fig.5 display the response of S-layer for the input image. The leftmost side of the figure is 4x4 portion of the input image, each square in the s-layer of the rightmost side of figure represent a cell-plane thats respond to a specific feature of input stimuli. So, as mentioned earlier, we have 16 cell-plane each of 4x4 pixels.

Fig. 6 illustrate a plot of the number of feature extracted from input portions. The number of input portion included in this research is 640, the number of extracted input vectors (features) is 550 for the proposed model. The proposed feature extraction method is compared against the one employed in the standard HTM [22], which achieve 266 feature from 640 and in PCA [28], which achieve 150. An illustrative result is presented in Fig.6.

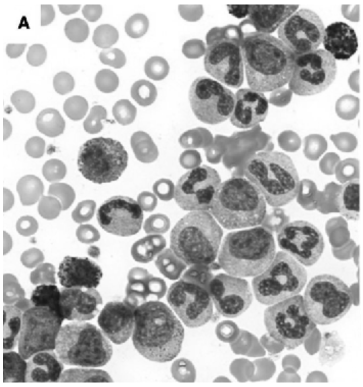


Figure 3. Sample of an input image

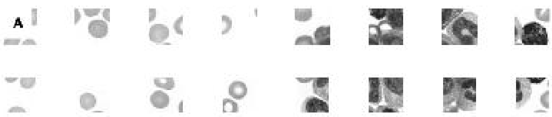

Figure 4. Portions of an input image

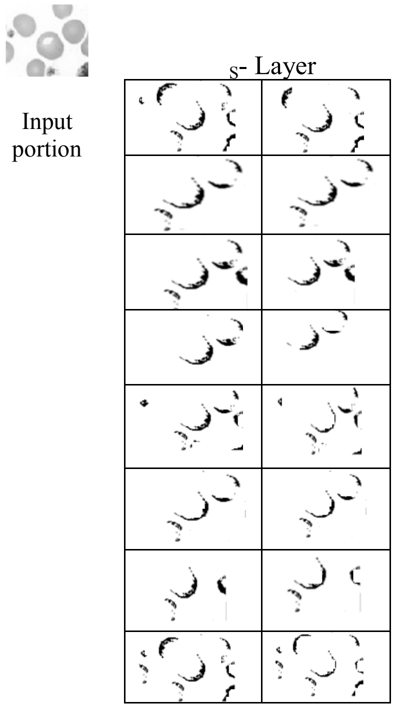


Figure 5. S-layer response to one input portion of an image.

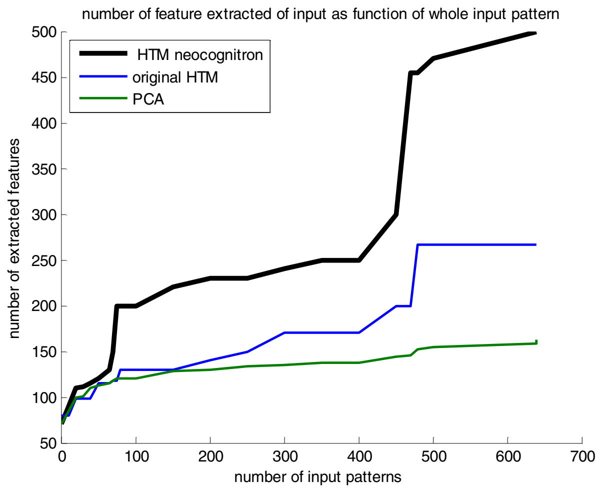


Figure 6. Results: Improved HTM vs. original HTM vs. PCA

From Fig.6, it is learned that the proposed HTM with neocognitron extracted double the features recognized by the original HTM, and 40 percent more than the work of PCA. Such a result may contribute to a better recognition as accuracy is improved when a recognition model can extract more features from the input image [14] [9].

## V. Conclusion

In this study, a combination of HTM and neocognitron neural S-layer has been proposed to recognize blood type image. The S-layer cells of neocognitron were used to extract feature in the first step of HTM. Experiments result shows that the proposed HTM extracts 86% of the dataset, while the original HTM and PCA only recognize 41.5% and 23% respectively. It is hoped that such an approach will improve the accuracy of pattern recognition.

## References


[1] L. Šajn and M. Kukar, "Image Processing and Machine learning for Fully Automated Probabilistic Evaluation of Medical Images," Computer Methods and Programs in Biomedicine, vol. 104, pp. 75-86, 2011.

[2] E. Alpaydin, Introduction to Machine Learning (Adaptive Computation and Machine Learning series), 2010.

[3] R. T. Olszewski, "Generalized Feature Extraction for Structural Pattern Recognition in TimeSeries Data," PHd, School of Computer Science, Carnegie Mellon University, 2001.

[4] M. Pechenizkiy, "The Impact of Feature Extraction on the Performance of a Classifier: kNN, Naïve Bayes and C4.5," in Advances in Artificial Intelligence. vol. 3501, B. Kégl and G. Lapalme, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 268-279.

[5] H. K. Elminir, M. A. ElSoud, and L. M. A. El-Maged, "Evaluation of Different Feature Extraction Techniques for Continuous Speech Recognition," International Journal of Information and Communication Technology Research, vol. 2, p. 8, 2012.

[6] M. F. Othman and M. A. M. Basri, "Probabilistic Neural Network for Brain Tumor Classification," in Intelligent Systems, Modelling and Simulation (ISMS), 2011 Second International Conference 2011, pp. 136-138.

[7] X. Huan, C. Caramanis, and S. Sanghavi, "Robust PCA via Outlier Pursuit," Information Theory, IEEE Transactions, vol. 58, pp. 3047-3064, 2012.

[8] D. p. Tian and S. Baoji, "A Review on Image Feature Extraction and Representation Techniques," International Journal of Multimedia and Ubiquitous Engineering, vol. 8, p. 12, 2013.

[9] A. G. Chitu, L. J. M. Rothkrantz, J. C. Wojdel, and P. Wiggers, "Comparison Between Different Feature Extraction Techniques for Audio-Visual Speech Recognition," Journal on Multimodal User Interfaces manuscript, vol. 8, p. 16, 2006.

[10] D. George and J. Hawkins, Invariant Pattern Recognition using Bayesian Inference on Hierarchical Sequences, 2005.

[11] B. Bobier, Handwritten Digit Recognition using Hierarchical Temporal Memory, 2007.

[12] J. Hawkins and D. George, "Hierarchical Temporal Memory--Concepts, Theory, and Terminology," Whitepaper, Numenta Inc, May, 2006.

[13] S. Štolc and I. Bajla, "On the Optimum Architecture of the Biologically Inspired Hierarchical Temporal Memory Model Applied to the Hand-Written Digit Recognition," Measurement Science Review, vol. 10, pp. 28-49, 2010.

[14] R. William, Hierarchical Temporal Memory Cortical Learning Algorithm for Pattern Recoginition: ProQuest, UMI Dissertation Publishing (October 17, 2012), 2012.

[15] I. Kostavelis and A. Gasteratos, "On the optimization of Hierarchical Temporal Memory," Pattern Recognition Letters, vol. 33, pp. 670-676, 4/1/ 2012.

[16] S. Garalevicius, "Memory–Prediction Framework for Pattern Recognition: Performance and Suitability of the Bayesian Model of Visual Cortex," presented at the Int. Florida Artificial Intelligence Research Society Conference, florida, 2007.

[17] K. Fukushima, "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position," Biological Cybernetics, vol. 36, pp. 193-202, 1980.

[18] K. Fukushima, "Neocognitron: A hierarchical Neural Network Capable of Visual Pattern Recognition," Neural Networks, vol. 1, pp. 119-130, 1988.

[19] A. Nouri and H. Nikmehr, "Hierarchical Bayesian Reservoir Memory," in Computer Conference, 2009. CSICC 2009. 14th International CSI, 2009, pp. 582-587.

[20] D. Maltoni. (2011, Pattern Recognition by Hierarchical Temporal Memory. 45.

[21] M. Bundzel and S. Hashimoto, "Object Identification in Dynamic Images Based on the Memory-Prediction Theory of Brain Function," Journal of Intelligent Learning Systems and Applications, vol. 02, pp. 212-220, 2010.

[22] D. George, "How the Brain Might Work A Hierarchical and Temporal Model for Learning and Recognition," PhD, ElectricalL Engineering, Stanford, 2008.

[23] D. Rozado, "Analysis and Extension of Hierarchical Temporal Memory for Multivariable Time Series," PhD, Universidad Autonoma de Madrid 2011.

[24] K. Fukushima, "Artificial Vision by Multi-Layered Neural Networks: Neocognitron and its Advances," Neural Networks, vol. 37, pp. 103-119, 2013.

[25] T. N. Wiesel and D. H. Hubel, "Single-cellresponses in Striate Cortex of Kittens Deprived of Vision in One Eye," Neurophysiology, vol. 26, 1963.

[26] G. P. M. Priyankara, O. W. Seneviratne, R. K. O. H. Silva, W. V. D. Soysa, and C. R. D. Silva. (2007, An Extensible Computer Vision Application for Blood cell Recognition and Analysis.

[27] G. M. Abdulfattah and M. N. Ahmad, "Face Localization Based Template Matching Approach Using New Similarity easurements," Journal of Theoretical and Applied Information Technology, vol. 57, 2013.

[28] J. Novakovica, M. Minica, and A. Veljovicb, "Classification Accuracy of Neural Networks with PCA in Emotion Recognition," Theory and Applications of Mathematics & Computer Science, vol. 1, pp. 11-16, 2011.