

Developing a Modified HMAX Model Based on Combined with the Visual Featured Model

Yaghoub Poursad

Faculty of Electrical Engineering, Urmia university of Technology, Urmia, Iran
Corresponding author, e-mail: Bakhteeyar@gmail.com

Abstract

Identify objects based on modeling the human visual system, as an effective method in intelligent identification, has attracted the attention of many researchers. Although the machines have high computational speed but are very weak as compared to humans in terms of diagnosis. Experience has shown that in many areas of image processing, algorithms that have biological backing had more simplicity and better performance. The human visual system, first select the main parts of the image which is provided by the visual featured model, then pays to object recognition which is a hierarchical operations according to this, HMAX model is also provided. HMAX object recognition model from the group of hierarchical models without feedback that its structure and parameters selected based on biological characteristics of the visual cortex. This model is a hierarchical model neural network with four layers, is composed of alternating layers that are simple and complex. Due to the high complexity of the human visual system is virtually impossible to replicate it. For each of the above, separate models have been proposed but in the human visual system, this operation is performed seamlessly, thus, by combining the principles of these models is expected to be closer to the human visual system and obtain a higher recognition rate. In this paper, we introduce an architecture to classify images based on a combination of previous work is based on the basic operation of the visual cortex. According to the results presented, the proposed model compared with the main HMAX model has a much higher recognition rate. Simulations was performed on the database of Caltech101.

Keywords: Object recognition, hierarchical, HMAX model, visual featured

Copyright © 2017 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Eyesight as the most important sense among the human senses, has great importance in understanding the world around. Thus, a large number of human neural processes related to this sense. Using the principles of human vision in providing image processing algorithms makes these algorithms become intelligent [1]. Experience has shown that in many areas of image processing algorithms that have modeled the structure or function of the human visual system, have the greater simplicity and efficiency.

There are two general methods for studying the human visual process. In the first method that is called psychological analysis, focus on the study of the relationship between perceived stimuli and thus simulate the input-output mental phenomena vision is based on performance of vision system. In this way, there was no attention the details of how to understand but important issue is the relationship between excitation and understanding. In the second method to study the relationship between perceived and physiological processes of the brain and sensors associated with the human visual system, which it called physiological analysis. In these models, simulation processes of different layers of the primary visual cortex is considered, as a result, hierarchical models have been developed. The physiological analysis carried out, eye receptors convert stimulus energy into an electrical signal. These signals transmitted by the nerve fibers of the brain which leads to different understandings. Vision system model of mammals mainly introduced by Hubel & Wiesel [2]. By studying neurons of vision, a strong theory of human vision with the name of scale-space theory have been proposed [3-7]. According to this theory, from the beginning to the end of the visual cortex with a family of local and scaled Gaussian operators will model. Modeling can be performed with the help of multilayer neural networks which are composed of simple units [8].

In this paper, according to biological models, the lowest multi-layered network operators are local multi-directional filters at multiple scales, for example, Gaussian derivatives or Gabor filters can be used. Good recognition system must give a similar response to objects within a class (immutable property) and give different response to objects between classes (selectivity property). In fact, one of the key aspects of each desirable recognition system is that can establish a proper balance between these two properties. Many developments in the field of computer vision is in order to achieve this goal. Physiological studies show that in object recognition, feedback low activity ahead with fast response and low accuracy but a lot of feedback produces slow and accurate response [8-11]. If only activity ahead modeled, the biggest challenge, create a high-level representation with flexibility and rigidity is acceptable. In fact, if we want a simple feature local and national produce a complex feature, we are faced with the problem of the balance between the two, in this case, we cannot have confidence of the distinction between classes and immutable within each class. Identifying objects is a hierarchical process that HMAX model is designed with inspiration from it and in four layers. The main cases of non-compliance with the human visual system model, the random selection prototypes can be named. However, the human visual system selects the optimal prototypes. To fix the problem random selection prototypes in the HMAX model, the high sensitivity of the human visual system to the lines and edges will be attended. Thus, instead of a random selection prototypes, top prototype, the prototype with the greatest variations (including lines and edges), will be selected. According to the results of the simulation, the proposed amendment increases the recognition rate.

2. Different Stage Detect Objects in the Visual Cortex

Vision process is started from the eye and continues until the last layers of the visual cortex in the human brain. The task of the eye, forming a clear picture of the outside world on the retina, processing visual stimuli begins from the retina (where the electromagnetic radiation in a particular frequency band is transformed into neural code). According to Figure 1 when light reaches the retina after detection by optical receivers, is processed by the bipolar cells, horizontal and Amacrine and finally in Ganglion cells, is transformed into a train of frequency modulated blow.

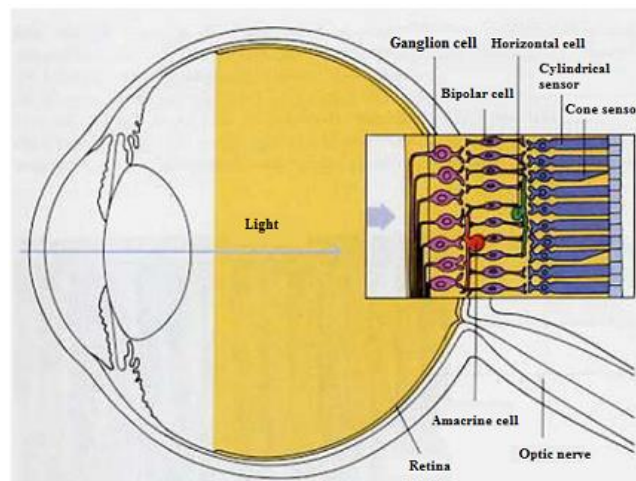


Figure 1. Structure of the retina in the eye [2]

Receivers and cells are combined in such a way in the retina which are divided into two structures of center-on and center-off. According to Figure 2, the structure of the center-turn is stimulated by triggers which is covered their center and is not covered around them, and the center-off vice versa. Due to this feature, Ganglion cells can act as contrast detector (edge or line detector). Such a structure enables visual system to respond to the contrast, in fact, is one

of the most important characteristics of human vision. This feature is very important physiologically because of the contrast, the observed properties of the object is determined. Image contrast is independent of the surrounding light intensity. This type of processing is best done in the retina because the structure's center around small differences in light intensity can be detected by neurons [11]. Based on the above explanation, the retina has circular receptive field.

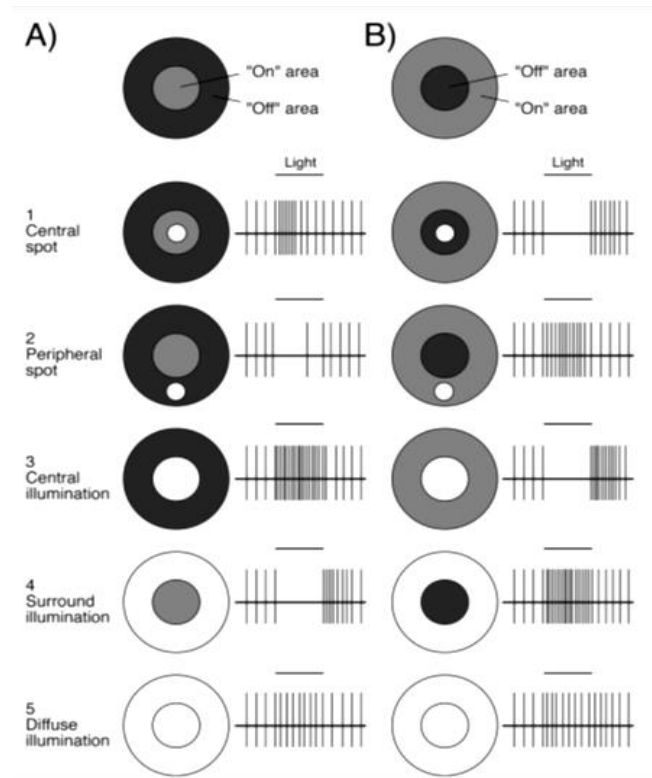


Figure 2. How to respond Ganglion cells of the center-on and centers - off to the edge of external stimuli

Retinal processing system, is so organized that as much as possible done much analysis on the input image. The retina is the unit that there are a lot of data in it and the results appear in the output of Ganglion cell. Axons of Ganglion cells create optic nerves which will lead to a reinforcing structure in the thalamus that it is called Lateral Geniculate Nucleus (LGN). Such as the receptive field of retinal cells of LGN have a circular field which contains the center-around the time axis and to act opposite. However, important differences between the retina and LGN is that the area around the LGN create a stronger deterrent effect than surrounding areas Ganglion cells in the retina. This means that LGN cells increase the difference in brightness mapping of Ganglion cells. In addition, because of their receptive field is circular, like perceptions of the retina respond to the border or contour in any direction. After LGN information is entered into the visual cortex of the brain. The visual cortex in humans is divided into different parts of the cell performance that interact with each other, to create a perception of the image. The primary visual cortex, is a part of the occipital cortex that can take input from LGN. Like other areas of the cerebral cortex, the area of the primary visual cortex contains a layer arrangement of cells which has about two millimeters thick. In total there are approximately 100 million cells per hemisphere of the brain in the area. Figure 3 shows a picture of the primary cortical brain tissue.

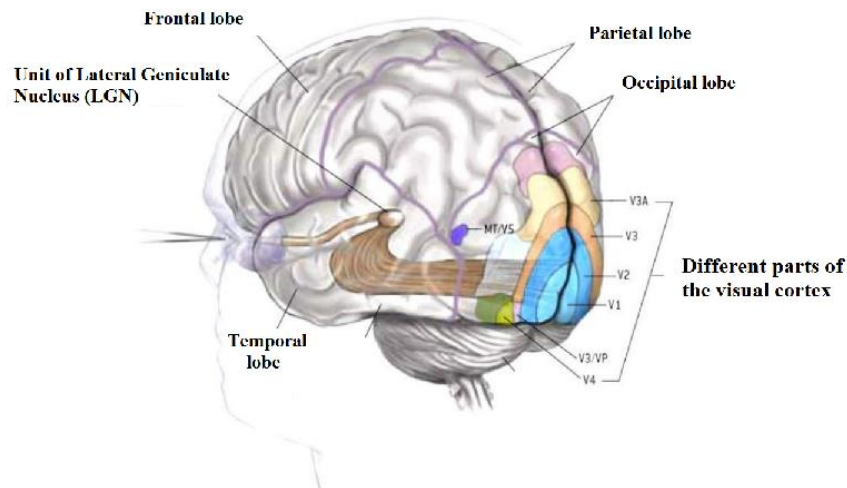


Figure 3. Overall shape of brain structure in the process of vision

- V_1 : The first part of the visual cortex called V_1 that low-level visual processing that takes place in it which represents the image details. Like all parts of the cortex consists of six layers, and it is considered a pillar structure. Each column processes the features of a different part of the field of view (e.g., contrast, color, orientation and movements).
- V_2 : The second main part of the visual cortex, which is the first area in the visual association. Here, V_1 has received strong connection of forward and sends strong connections to the next section, strong connection returns feedback to V_1 . District V_1 provides a mapping of the visual world. This area has three different cell groups in terms of functional which include sensitive cells in order to detect motion and three-dimensional structure of objects, cell shape recognition and wavelength-sensitive cells to distinguish colors.
- V_3 : This part is related to image dynamic processes. Receives input from sensitive cells in regions V_1 and V_2 , and sends to posterior parietal (PP) area.
- V_4 : is related to color processing, studies have shown that selective attention can change the response rate to approximately 20% of the cells in this area. This section receives forward and strong connections of V_2 and sends forward connections to IT. Receives input of V_1 directly, but connections with V_5 is too low. V_4 like V_1 by changing the orientation, spatial coordinates and color change frequency, but unlike V_1 is active for objects with low complexity.
- V_5 : This area has direct inputs from the visual cortex and lower levels of wide connectivity with most areas of the brain. This area is also called MT. This part plays an important role in visual motion processing (for example, in understanding the movement and eye movement). As it mentioned V_1 neurons also respond by changing the direction and speed of movement. But in this area, the local motion signals becomes the general movement of an object. Blob areas in the visual cortex are related to color processing that have circular receptive field and rest areas respond to linear actuators such as edges, rods, lines and animated lines.

The most important feature cortex cells, is their directional selection. This is while in the retina and Lateral Geniculate Nucleus (LGN) due to the circular of the receptive field, the cell's responses are uncertain in all directions. In this sense, the visual cortex cells are classified in two forms: simple and complex cells.

2.1. Simple cell

The cortex cells are sensitive to stimulants, as shown in Figure 4 each cell line will respond only in one direction. In other words, each cell cortex has a preferential direction in which direction is most responsive. If the line deviates slightly from the preferred cell response is reduced a lot. Seeing these different arrangements, we can assume that any significant motivation can be invoked for best answered in a specified cell. For this reason, the cortex cells in the targeted areas, known as simple cells. It means that there is a simple relationship between a field and receive preferential driving in the same direction. This behavior can be summarized by showing a graph of the power cell response as a function of direction, peak value of cone tuning curve shown in Figure 4, shows a preferential orientation [12]. Preferential orientation varies from one cell to another, so that in general, the cells will be present in all directions. In most cells, inclination of about fifteen degrees of preferential orientation is enough to call cell is completely destroyed.

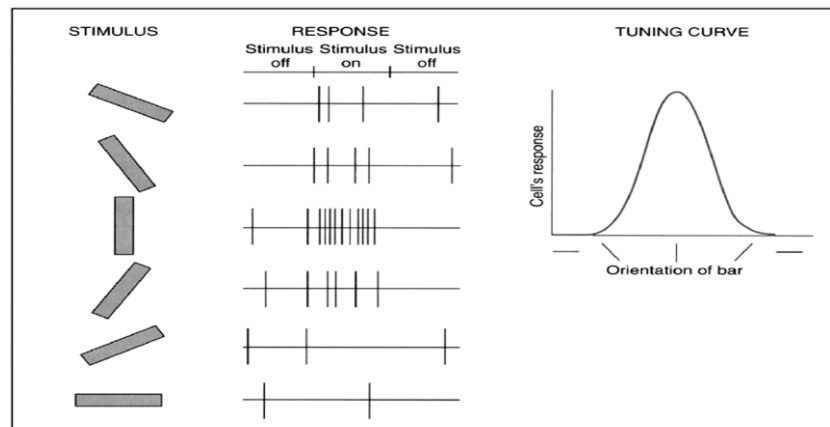


Figure 4. Orientation cell response to input stimuli at different angles

At the same time with discover the selection of directional properties of simple cells, a plan was presented to justify this property. This design is shown in Figure 5, shows that the general principles of the proposed all simple cell models by Hubel & Wiesel [13]. As seen in Figure 5, along with finding the appropriate rows of cells Lateral Geniculate Nucleus (LGN) together and the aggregate responses of center - surrounding cells, directional response relates to a simple cell of model. The proposed model, based on three assumptions. Entrance must be received from a large number of cells of Lateral Geniculate Nucleus (LGN), entrance must be strong enough and have the greatest significance and there must be a very special geometry and precise wiring between the rows of cells Lateral Geniculate Nucleus (LGN) and target cells in the cortex.

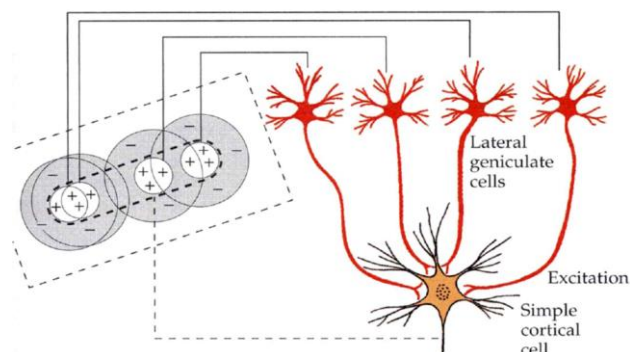


Figure 5. Construction of single cell model

2.2. Complex cell

Complex cells of cortex, unlike simple cells of off and on areas which can be defined precisely with shape, but their response is also related to a specific direction. These cells, called complex cells, such as to predict exactly what the stimulus can respond appropriately to them, a little more complicated. The exact location stimulus for complex cells is not so important. Also, as in complex cells of off and on areas is not clearly specified, these cells respond strongly to the rod or line that moves in the space of receiving it. More complex cells respond to the contours in motion, but simple cells respond only to stationary or slow-moving contours. In addition, complex cells respond to dynamic contours only in one direction. For example, complex cells is shown in Figure 6, may give severe response when a vertical contour move from left-to-right, and have no response in right-to-left direction. Complex cells may have the opposite response and have response for stimulus moves from right to left. This property (which is the direction selection) shows that this type of cortex cells, play an important role in the perception of motion.

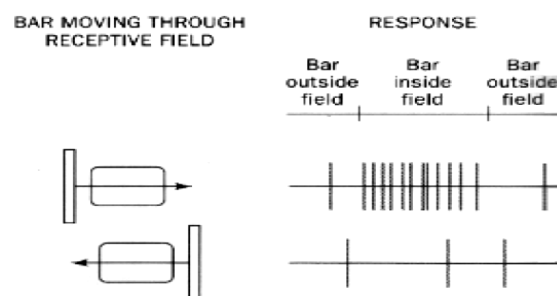


Figure 6. Complex cell responses to stimuli bar

After moving bar, new place must be determined by simple cell that by combining this information, we can discern motion. The conclusion that can be drawn from the above is that the complex cell model is created using the collect call several simple cells on the surface of neurons.

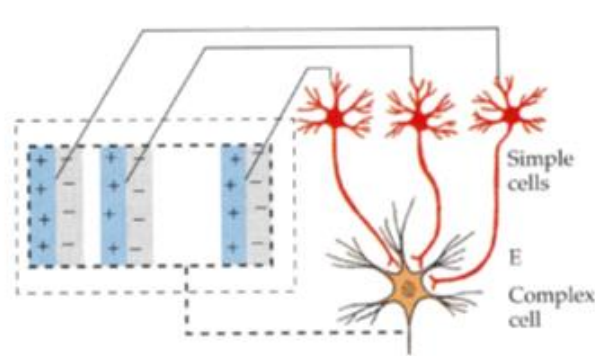


Figure 7. Complex cell model made from simple cells

3. Introduction of the Visual Featured Model

Extraction of image relief map is one way to mimic the human visual system. Much evidence suggests that the human brain image and select an area of the visual system imposes its details. The process leading to this area is not yet fully known [14, 15]. In visual featured, there are two main approaches of bottom-up saliency and top-down saliency. Visual featured based on bottom-up saliency in a scene responds to stimuli faster. For example, a flashing light because of physical properties such as light intensity and more general move to attract attention and somehow be featured. One of the important tasks on bottom-up saliency-based approach is based on a hierarchy of operations that is based on three parameters: intensity, direction and

color, providing a model for the detection of image saliency map. The final model is derived by combining the output models of these three features [16]. In every feature, the pyramid is built with dimensions of half-size images before. Then, using the human visual system performance center around the top six difference images added together. The output of this method is not the size of the main image. This is making it a potential biological models based on actual study neural processes. Maps for each scale for four directions and produced a total of 24 maps for directions made features. The main feature of this model, incorporating all the features maps middle. This is done by scoring the maximum. A normal operator is also used to accomplish this matter. For relief maps, three maps emerge, normalized and added together. In this model explains a map of the bumps. Maximum relief map, the most prominent place in the picture. Another method is based on the frequency spectrum. In this way the input image, taken Fourier spectrum and by selecting high coefficients and highlights of spectrum, prominent image is achieved.

4. Introduction of HMAX Model

Basic HMAX model, in 1999 on the basis of simple cell arrangement (s) and complex (c) in the V1 area of the visual cortex that were repeatedly been proposed. The model is derived and built based on the theory of visual processing hierarchy of Hubel & Wiesel model, belongs to the category hierarchical models and uses convolution design / integration that take place periodically. This model includes the ventricular sight of V_1 (first processing area of the visual cortex) to the higher levels of the visual cortex, such as IT and PFC and follows the theory of object recognition forward path in the cerebral cortex related to vision processing 100 to 200 milliseconds primary ventricular flow. HMAX model structure, is composed of four layers that are hierarchically increase its selectivity and immutable. Simple units to increase selectivity, their input with a series of special functions such as Gabor and the Gaussian convolution, each stage of convolution provide the set of feature maps. Complex units to increase rigidity, as aggregate output of simple units by applying a nonlinear function such as a maximum, until the final version to a different answer to the same position and the same inputs. In the training phase, a large number of random bits of training images with different dimensions of the prototype are called, in the layer C_1 for all angles (0°, 45°, 90°, 135°) extract (ie a prototype with dimensions $n \times n$ of $n \times n \times 4$ is element, from 4 starts n value to 16 pixels with a distance of 4 times). The function of each layer is described in detail in the following:

Layer 1 (S_1): The first layer HMAX model that simulates the activity of the visual cortex simple cells V1 area. In this layer, each map features $L1_{\delta, \theta}$ by convolution of the input image with a set of Gabor filters $g_{\delta, \theta}$ for edge detection filters that are applied to θ and size δ has been achieved. Two-Dimensional Gabor is shown in the following Equations:

$$G(x, y) = s(x, y) \times w(x, y) \quad (1)$$

That $s(x, y)$ is a sine function and called carrier, $w(x, y)$ is a two-dimensional Gaussian function that is capped.

$$w(x, y) = e^{\left(\frac{-X^2 + \gamma^2 Y^2}{2\sigma^2}\right)}, s(x, y) = \cos\left(\frac{2\pi}{\lambda} X\right) \quad (2)$$

$$G(x, y) = e^{\left(\frac{-X^2 + \gamma^2 Y^2}{2\sigma^2}\right)} \times \cos\left(\frac{2\pi}{\lambda} X\right) \quad (3)$$

$$X = x \cos \theta + y \sin \theta, Y = -x \sin \theta + y \cos \theta$$

Where σ during Gaussian function, θ for Gaussian, γ development rate and λ is the wavelength. If the input image I call, the exact output layers (1) to θ and size of δ is obtained by convolution of the following:

$$L1_{\delta,\theta} = |g_{\delta,\theta}(x,y) * I| \quad (4)$$

The input image is analyzed with filters in different dimensions and directions (from 7×7 to 37×37 in steps of 2 pixels), so filters are in 16 different sizes. The filter includes four angles of 0° , 45° , 90° and 135° , thus the total number of filters $16 \times 4 = 64$ (4 to 16 dimensions) is. The filter set is divided into eight bands, parameters related to the eight bands in Table 1.

Layer 2 (C_1): output layer before they are sent to this layer, this layer is similar to the behavior of complex cells in the primary visual cortex. This type of cortical cells are resistant to change position and size of images on the perceptions, this property is run by the operator maximum in this layer. According to the above explanations C_1 output layer is obtained by applying the operator maximum S_1 the output layer. Maximum operator on two adjacent filter with the same directions (such as 7×7 and 9×9) in order to create an immutable characteristics applied to size and position the image thus obtained 32 (4 directions and 8 bands). The next step is to calculate the maximum value of each pixel in networks is to dimension $\mathcal{N}^\Sigma \times \mathcal{N}^\Sigma$, dimensions \mathcal{N}^Σ of eight starts in band 1 and band 8 to 22 continues to overlap Δ_s . Table 1 above shows two layers parameters.

Layer 3 (S_2): This layer acts on the aggregation of local neighborhoods takes place in all four directions C_1 layer, this layer acts as a radial basis function (RBF). The output layer S_2 , Gaussian-shaped Euclidean distance between the new entry and a sample is stored, which means that for a piece of image X which related to prior layer (C_1), in a specific band, call r corresponding to these layers is equal to:

$$r = \exp(-\beta \|X - P_i\|^2)$$

Where sharpness of the tuning and P_i features single center RBF the network has learned during training. The total number of properties RBF center is equal to N that in proportion to the initial samples extracted in the learning phase. At runtime, the all maps S_2 opportunity for every eight band is calculated for each of the prototypes P_i (up to 1000 samples) is calculated as a multiple scale maps.

Layer 4 (C_2): The final output of the network, the answer is resistant to scaling and movement, with the public on all aspects maximum position for each sample obtained in the previous layer. For example, after the implementation of the prototype P_i input image stored in all positions and scale, only the most valuable preserved and the rest are discarded. The final signature that invariance property also has generally achieved. Layer C_2 , is the last stage in the process HMAX model. The total maximum layer S_2 all responses occur on all scales and directions. The output layer C_2 for each input image vector of length of prototypes (Features) is. Then call the C_2 to apply a classification. Figure 8 shows a block diagram the main HMAX model.

Table 1. Parameters of S_1 and C_1 Layers

Scale band S	C_1 Layer		S_1 Layer		
	Spatial pooling grid ($N_s \times N_s$)	Overlap Δ_s	Filter size s	Gabor σ	Gabor λ
Band 1	8 x 8	4	7 x 7	2.8	3.5
Band 2	10 x 10	5	9 x 9	3.6	4.6
			11 x 11	4.5	5.6
Band 3	12 x 12	6	13 x 13	5.4	6.8
			15 x 15	6.3	7.9
Band 4	14 x 14	7	17 x 17	7.3	9.1
			19 x 19	8.2	10.3
Band 5	16 x 16	8	21 x 21	9.2	11.5
			23 x 23	10.2	12.7
Band 6	18 x 18	9	25 x 25	11.3	14.1
			27 x 27	12.3	15.4
Band 7	20 x 20	10	29 x 29	13.4	16.8
			31 x 31	14.6	18.2
Band 8	22 x 22	11	33 x 33	15.8	19.7
			35 x 35	17.0	21.2
			37 x 37	18.2	22.8

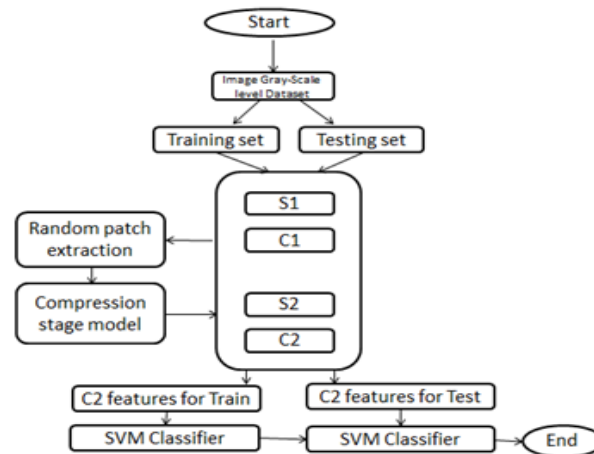


Figure 8. Block-diagram of the main HMAX model

5. Proposed model

In the main HMAX model in recognition of objects similar to the successful model of the human visual cortex, the processing input images would not be taken. However, the human visual system first visual featured done and then recognition is done. In order to model more similar to the human visual system to the principles we combine the two models. In this way, the visual featured done to the main image is selected then we consider the output of the system as input of HMAX model.

HMAX model that was introduced in 1999 based on the makeup simple and complex cells in the visual cortex area V_1 was suggested which were periodically. This model is based inference based on the theory of visual processing hierarchy of Hubel & Wiesel model. In [12] the main function of the four-layer structure, the main HMAX model has been explained. In [8] the first layer is the first layer of simple cells in the visual cortex area V_1 of HMAX model and simulates the activities described. Each map features on this layer $L1_{\sigma,\theta}$ with the convolution of the input image with a set of Gabor filters $g_{\sigma,\theta}$ (filters with the direction θ and size σ) is obtained. Input image apply a set of filters edge detection. These filters are made based on Gabor function. In [14, 15] the second layer is described, the layer model of complex cells in the visual cortex activity is to change the location and size, show resistance. To identify the correct angular lines or edges anywhere in the receptive field (immutability of the position), and more accurate than simple cells (unchanging relative to the size), in the field of modeling for complex cells get bigger (twice the simple cells) considered. In [9, 10, 13] of the project convolution / aggregation that occur periodically been used. In [1, 9] introduced the system follows the theory of object recognition forward path in the cerebral cortex related to vision processing 100 to 200 milliseconds primary ventricular flow. In [16] the output layer 3 (S_2) with convolution filters HL with $L2_{\sigma}$ layer can be obtained.

$$L3_{\sigma}^m = \alpha^m * L2_{\sigma} \quad (5)$$

HL filters represent the visual descriptors of mid-level visual image that in fact, the combination of Gabor filters is a low level for a given scale and in different directions. This layer acts on the aggregation of local neighborhoods where the layer C_1 in all four directions. This layer acts as a radial basis function (RBF). The layer 4 (C_2) the final output response network C_2 resistant to change the scale of displacement, with the public on all aspects maximum

position for each of the previous layer (S_2) achieved on the entire network S_2 . For example after matching the input image with prototype P_i stored in all positions and scale, only the most valuable preserved and the rest are discarded. As mentioned above, a method of extracting map image bumps mimic the human visual system. Much evidence suggests that the human brain selects an area of the image and it imposes the details of the visual system. One of the important tasks based approach is bottom-up saliency method based on hierarchical operations that based on three parameters; intensity, direction and color relief map image provides a model for diagnosis. The final model is derived by combining the output models of these three features [16]. Identification and object recognition can be performed using this model. In [17, 18] bump map will be affected by the characteristics of the target object. This effect is desired objects are masks that are defined for the purpose of weighting coefficients. In [19] face detection is executed with the help of this method. Rutishauser and colleagues have used the model for mining regions and identify objects by examining the surrounding area's most prominent area in an image and then use the techniques of region-growing. Another method mentioned is based on the amplitude of frequency spectrum, in this method, the Fourier spectrum of the input image and by selecting the coefficients and leading the whole, the relief image.

The model presented in [20] based on Gabor wavelet in a multi-scale image is prominent. According to the model presented in [21] is the Fourier transform. In [22] four route calculation is made of the brightness and color that the Fourier transform of all the paths taken quartet and finally combine. In reference models [23, 24] used Fourier transform and Gabor wavelet and then the map bumps is produced by weighting process. A model that bumps the frequency spectrum residual image is obtained by removing the additional fields' frequency. In order to remove the extra parts of the frequency, is reducing the whole picture of the frequency spectrum average. In [25] the method based on the frequency spectrum is used. First, the input image data as a non-linear change is the size of 64×64 and then the Fourier transform of the image is taken and the size and phase separate. The average logarithm of size Fourier transform decreases the size of the Fourier transform of the logarithm and finally collect the logarithm of the amplitude and phase Fourier Transform photos taken multiplied by the imaginary unit. Average stuck with a radius of three convolution filter output and ultimately the size of the original image is resized.

6. Results

The proposed model on data collection Caltech101 which contains 101 groups of objects, for 2, 5, 10, 15, 20 and 25 classes have been tried and recognition rate of the proposed model were compared with the standard HMAX model. At each stage, the simulation was performed 10 times and mean it as recognition rate is considered. Some examples of classes with the number of images per class data sets in Table 1 are given.

Table 2. Some examples of classes Caltech101 and the number of pictures in class

Data collection	Number of Pictures	Data collection	Number of Pictures
Airplanes	800	Chair	62
Cars	123	Chandelier	107
Faces	435	Cougar_face	69
Motorbikes	798	Crab	73
Bonsai	128	Crayfish	70
Bass	54	Crocodile	50
Brain	98	Elephant	64
Buddha	85	Faces	435
Butterfly	91	Motorbike	798
car_side	123	Umbrella	75

In Figure 9, three images of the five classes of the database is given. Each set of images can be divided into two sets of training and testing, which is 80 percent for education and 20 percent of the images chosen for the test images, respectively, 80% of the images are

selected for training and 20% of the images are selected for testing, select the group and also images of each group is randomly. All the images are converted to images with gray levels to match the images, their size is changed to 140×140 . At the beginning of each group a picture taken randomly prototype for extraction, this image is extracted from various prototypes that will be selected from among them the ideal prototype. The desirable prototype, prototype for which the maximum change in the simulation, the variance command is used. The set of images of education, as input into the system by comparing them with the prototype of their features are extracted. Finally set of test images, the system acts as a testing phase extracted features. Finally, these features has been entered to SVM classification and then the class of the input image is determined. In Figure 10, the block-diagram model has been improved.



Figure 9. The number of images of Caltech101 database

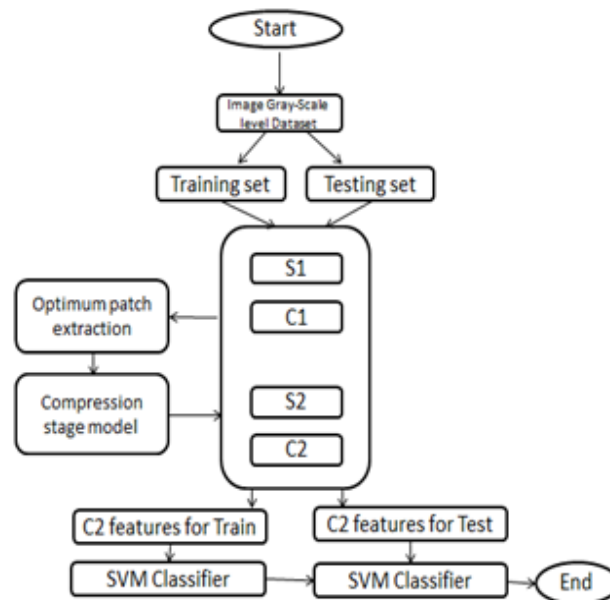


Figure 10. visual model block diagram

According to Table 3 recognition rate model introduced with the original image of Caltech101 set compared, which are significantly improved. In Figure 11, the rate of recognition compared together.

Table 3. Comparison of the results improved model and HMAX model

Number of data groups	Recognition percent of HMAX model	Recognition percent of improved HMAX model
5	97.8	100
10	96.2	99.6
15	91.6	99.2
20	89	99
25	84.4	98.6

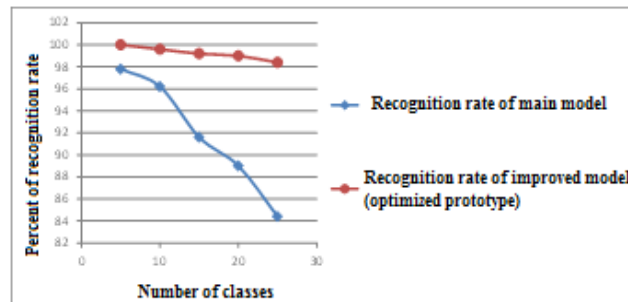


Figure 11. Comparison of recognition rate of standard model and improved model

7. Conclusion

In this paper, the attempt was to improve HMAX model, which is one of the few models similar to the structure of the brain, and recognition percent be upgraded. One of the main objections that enters to the HMAX model is that the initial samples are done randomly whereas the humans do not act in recognizing objects in random order. On the other hand, human is sensitive to the lines, so we choose between the selected samples, a sample that has the greatest variance that in comparison, certainly compared to the other is more likely to include lines.

References

- [1] M Mishkin, L G Ungerleider, "Contribution of Striate Inputs to the Visuospatial Functions of Parieto-Preoccipital Cortex in Monkeys," *Behavioural brain research*, 1982; 6: 57-77.
- [2] D H Hubel, T N Wiesel, "Receptive Fields of Single Neurons in the Cat's Striate Cortex," *The Journal of physiology*, 1959; 148: 574.
- [3] Z Li, Z Hu, X Du, S Fu, Y Ran, "Feature Extraction for Object Recognition Based on Visual Cortex-Like Mechanisms," *Energy Procedia*, 2011; 13: 3499-3506.
- [4] A Andersona, E Brunib, A Lopopolob, M Poesiob, M Baroni, Reading visually embodied meaning from the brain: Visually grounded computational models decode visual-object mental imagery induced by written text, *NeuroImage*, 2015; 120: 309–322.
- [5] J Koenderink, "Operational Significance of Receptive Field Assemblies," *Biological Cybernetics*, 1988; 58: 163-171.
- [6] C Theriault, N Thome, M Cord, "Extended Coding and Pooling in the HMAX Model," *IEEE Transactions on Image Processing*, 2013; 22: 764-777.
- [7] E Hernández, E Clementea, G Olaguea L Briseño, Evolutionary multi-objective visual cortex for object classification in natural images, *Journal of Computational Science*, 2016, In press.
- [8] E T Rolls, G Deco, "Computational Neuroscience of Vision," Oxford university press Oxford, 2002.
- [9] M Rouholamini, M Rashidinejad, A Abdollahi, H Ghasemnejad, "Frequency reserve within unit commitment considering spinning reserve uncertainty", *International Journal of Energy Engineering*, 2012; 2: (4) 177-83.
- [10] J Manikandan, C S Celin, VM Gayathri, A Document Imaging Technique for Implementing Electronic Loan Approval Process, *Indonesian Journal of Electrical Engineering and Informatics (IJEI)*, 2015; 3: (2) 109-114.
- [11] G Paxinos, JK Mai, "THE HUMAN NERVOUS SYSTEM" Chapter 35, Elsevier Academic Press, 2004.
- [12] L Simard, "Fundamentals of Computer Vision, Biological Vision, Student prepared course notes," COMP-558, Winter. 2012.
- [13] D H Hubel, T N Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol. (Lond.)*, 1968; 195: 215–243.

- [14] Anton Yudhana, Sunardi Sunardi, Shoffan Saifullah, Segmentation Comparing Eggs Watermarking Image and Original Image, *Bulletin of Electrical Engineering and Informatics (BEEI)*, 6: (1) 595-560.
- [15] L Itti, C Koch, "A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention," *Vision research*, 2000; 40: (10) 1489-1506.
- [16] L Itti, C Koch, E Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Transactions on pattern analysis and machine intelligence*, 1998; 20: 1254-1259.
- [17] V Navalpakkam, L Itti, "Sharing Resources: Buy Attention, Get Object Recognition," *International Workshop on Attention and Performance in Computer Vision WAPCV*, 2003: 73-79.
- [18] D Walther, L Itti, M Riesenhuber, T Poggio, C Koch, "Attentional Selection for Object Recognition -A Gentle Way," *Biologically Motivated Computer Vision*, 2002: 472-479.
- [19] C Siagian, L Itti, "Biologically-Inspired Face Detection: Non-Brute-Force-Search Approach," *Computer Vision and Pattern Recognition Workshop, CVPRW'04*, 2004: 62.
- [20] D B Walther, C Koch, "Attention in Hierarchical Models of Object Recognition," *Progress in brain research*, 2007; 165: 57-78.
- [21] M Ferni Ukrit, G R Suresh, Super-Spatial Structure Prediction Compression of Medical, *Indonesian Journal of Electrical Engineering and Informatics (IJEI)*, 2016; 4: (2) 126-133.
- [22] C Guo, L Zhang, "A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression," *IEEE Transactions on Image Processing*, 2010; 19: 185-198.
- [23] M Wang, J Li, T Huang, Y Tian, L Duan, G Jia, "Saliency Detection Based on 2d Log-Gabor Wavelets and Center Bias," Proceedings of the international conference on Multimedia, 2010: 979-982.
- [24] X Cui, Q Liu, D Metaxas, "Temporal Spectral Residual: Fast Motion Saliency Detection," Proceedings of the 17th ACM international conference on Multimedia, 2009: 617-620.
- [25] X Hou, L Zhang, "Saliency Detection: A Spectral Residual Approach," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'07), 2007: 1-8.