Texture Feature Extraction Inspired by Natural Vision System and HMAX Algorithm

Maede Madanian, Abbas Vafaei, S.Amirhassan Monadjemi

Department of Computer Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran

Abstract

In this paper, a new and effective method called HMAX is used for image texture. feature extraction. This method is inspired by the biological system of brain and human vision in order to create feature vectors for image recognition. A set of C2 features obtained from HMAX algorithm that are stable against changes in angle and size, are extracted from all image datasets firstly. Then using artificial neural networks and K-nearest neighbor classifiers, eight different types of natural texture images from VISTEX dataset are classified. In order to evaluate the HMAX feature extraction method, the classification results are compared with Gabor filter banks. Since HMAX model is consistent with natural vision system, it is expected to obtain a better accuracy compared to Gabor filter banks. Experimental results with artificial neural network and K-nearest neighbor classifier show that the accuracy of 90.12% and 84.50% respectively for HMAX features. They have significant improvements compared to Gabor filter banks which obtained 78.62% and 72% accuracy.

Keywords: texture feature extraction, HMAX model, texture classification, artificial neural network, K-nearest neighbor

I.   Introduction

In recent years, great attention has been paid to the texture features and their extraction in the field of image processing and machine vision and Feature vectors extraction from images using texture, can be the base of many other processes such as classification, segmentation and identification of objects. Texture classification itself is an important issue in the field of image analysis as well. Medical X-erography, surface inspection, and documentation are only a few examples of applications in which texture classification plays a key role. Usually in the texture classification, firstly the features have been extracted and then the features would be classified.

Typical ways to extract the features of image texture can be classified into four categories of(Bian,2006):

1.Model-based methods: This scheme is basically established on the structure of a visual model that not only describes the texture but also synthesizes it. Models are capable of simulating the regional texture information by means of multi-variable and dependent probability distribution functions(Krishnamachari and Chellappa, 1997).

2.Syntactic or structural methods: In this approach, texture, as an adjusted array, consists of a number of constant categorized pixels called the primate components. These models suppose that the primate components are located in a place with a rather adjusted connection and texture can

be described by these components together with a set of alternation and connection rules. Then, through adjusting the textures with the given rules and components, images can be partitioned and classified.

3.Statistical methods: These methods are performed on the basis of locational (spatial) distribution of gray levels in a texture, including several categories itself:

- First-order statistical methods using gray surfaces' histograms in a texture.

- Second-order statistical methods, such as Co-occurrence matrices, which is one of the primary methods in extracting texture features and indicates the second order statistical properties of an image(Cooper, 2004).

- Auto correlation function is described by means of auto correlation coefficient indicating the linear locational correlation among the pixels.

- Spectrum of a texture unit, which firstly alternates texture pixels with texture units (functions of rather small neighbors around a pixel), and then, measures texture units distribution (so-called texture spectrum) on the imagograph. This method, in fact, describes the locational relation among image pixels in a given neighborhood.

4.Signal processing methods: As for these methods, texture is modeled as a two-dimensional digital signal. These methods have always been attractive and capable of extracting features properly, on both natural and Artificial textures. One of these methods, is the Multi-Scale,Multi-Directional methods (MSMD). The texture-analysis MSMD methods through emulating human visual systems partition the input image by some filters into partial images with variant frequencies in different orientations. Each of these partial images includes part of the input image features. Among MSMD methods are Gabor filters and wavelet transform(Signolle et al,. 2009),(Li and Staunton, 2008).

As for the popularity of these extracting features and their correspondence with natural visual content, the method utilized in this article, which is also inspired from object recognition structure in natural visual layers, called HMAX, can be categorized as one of these methods. The common tasks performed by Poggio and Marr caused the existence of HMAX model finally by Riesenhuber and Poggio, aimed at emulating the human visual system behavior in ventral channel(Riesenhuber and Poggio, 1999). Similar to Marr theory, in the initial stages of visual system, the cells, respond to simple and primitive forms like gratings, bars, edges, etc(Marr, 1982). Yang models these cells behaviors through Gaussian functions(Young, 1987). In HMAX model, S1 layer emulates the fist-stage visual system by means of Gabor filtering. Following the studies conducted by Poggio team, in 2007, a new model was introduced by Thomas Sir to emulate the recognition mechanism in brain(Serre et al., 2005),(Serre et al., 2007). The main difference between this model and the original HMAX model is the addition of a learning step. The applications of this model for machine vision issues, including object identification and recognition, have proved to be highly efficient (Serre et al., 2005). This model was examined by Mirz and Wolf in face recognition(Meyers and Wolf, 2008). As proposed by Mirz, after performing certain pre-processings on the image, the output C1 layer in new model is measured and then through using RCA analysis, features called S2FF are produced to describe the face. Being tolerant of angle and size changes, these features outperform all other methods. Since the HAMX model is consistent with natural visual systems, we expect its efficiency and accuracy in diagnosis of objects and texture classification. In this paper, we studied the performance of this

algorithm in extraction of the features of natural texture images with random textures, and then compare its performance with the Gabor filter bank algorithm. For images' classification, we used the K-Nearest Neighbors (KNN) and Artificial Neural Networks (ANN) classification and finally we compared their performances too.

II.   Image feature-extracting HMAX method

In this section, the object recognition structure of human and animals visual system is firstly introduced and analyzed. Then, the quality of extracting image texture features by the proposed HMAX model, to classify and segment the image, is described.

A. Object recognition structure

Since human and animals visual systems operate well, it has been long desired to build a model simulating that. During the first milliseconds of human and some animals' visual operation, a hierarchy system in brain is used for object recognition. It is believed that information in visual layer is flown into two Ventral and Dorsal pathways. Object recognition in visual layer is performed through Ventral pathway. Figure 1 illustrates a monkey ventrally oriented visual pathway. As shown, received information from retina, after being transferred to LGN unit, is transferred to V1 layer, V4 layer, and finally the IT. Eventually, IT contributes to vision adjustment and object sighting, and therefore, the intended object is recognized(Gawne and Martin, 2002).
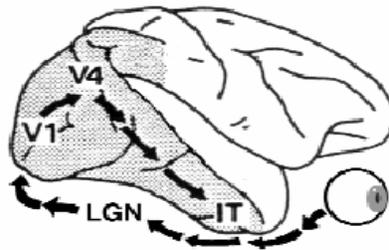
Figure1. The ventral visual stream of the monkey

In recent decade, Poggio and his colleagues tried to perceive and quantavely duplicate the structure and ventral-oriented visual processings of human neural system during the object recognition process. These attempts lead to the introduction of a new method called HMAX. In general, HMAX model, consists of 4 layers including alternative simple S and complex C layers (Riesenhuber and Poggio, 1999). C1 and S1 layers duplicate (model) simple and complex neurons behavior in V1 area of human's vision. C2 and S2 layers behave similarly to neurons in areas above the human V1 visual area. S layers compound out inputs using a Gaussian operation, which results in responding to more specific objects. C layers, also, are more complex units compounding S layer output as the input with maximum operation. This regularly creates

tolerance to size and scale changes. This model can quantitatively duplicate the general properties of neurons in inferotemporal monkey cortex(Hubel and Wiesel. 1965). Figure 2 outlines the HMAX model.
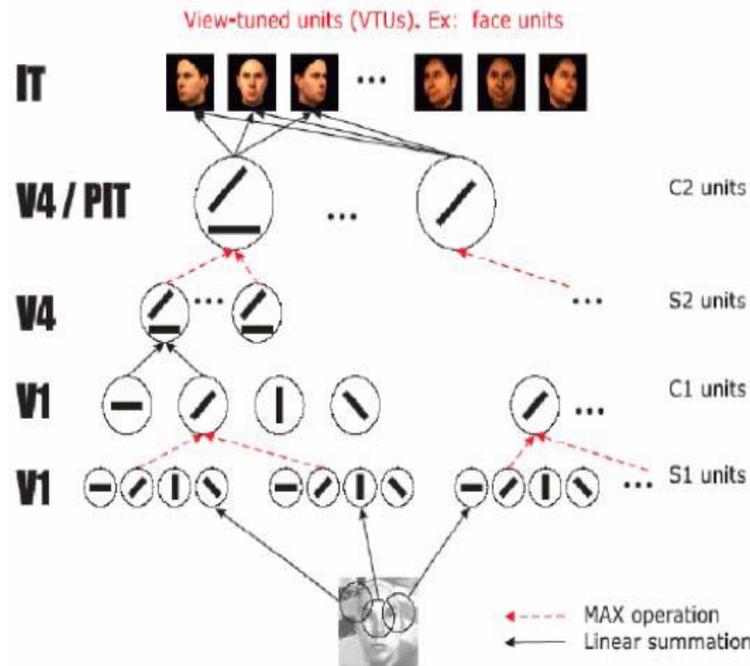
Figuer2. HMAX Model

First, the input images in simple units (S1), with the contribution of 2-dimensional filters arrayed in 4 variant orientation, are sampled concisely. Within compound band area, S1 cells information, a group of cells in the same orientation but rather variant scales and positions, are transferred to complex cells (C1) through MAX operation. Then, in S2 level, and for each frequency band a group of S1 filters with similar frequencies limit, a square-shaped area consisting of four C1 units are entered into a single S2 unit. As a result, all filters responses become tolerant to scale changes. The entire input image is, also, invariant to position changes. Finally, the S2 units are re-compound through the maximum operation to pass the necessary input to C2 units. When the C2 stage being passed, the extracted image features are obtained.

   B. C2-features extraction

   In this stage, through the utilization of HMAX method, the C2 features of texture images used in this article are extracted. As stated in the previous section, the algorithm to extract the above-mentioned features consists of four layers of alternative S1, C1, S2, C2. C2 through the HMAX algorithm is as follows(Serre et al., 2005)

S1 layer:The input image with gray surfaces levels enters as a one-dimensional array into S1 layer. In S1 layer, Gabor filters with variant orientations (θ) and widths (σ) applied to the input image. Equation 1, shows how this filter is applied to the input image(Serre et al., 2005)

In this relation, $X=x\cos\theta+y\sin\theta$ and $Y=-x\sin\theta+y\sin\theta$. (1)

In visual cortical V1 area, there are some neurons, each being sensitive to a particular edge, in terms of size and angle. Thus, in S1 layer, Gabor filter is used to simulate V1 neurons and each filter is applied to each pixel of input image. Gabor filter is applied to the input image in 4 orientations (θ) and 6 scales (S). therefore, 16×4 maps are obtained that are arranged in 8 bands. For each two scales, one band is used, and totally 8 bands are applied. For instance, according to the adjusted parameters in Table 1, band 1 consists of filter output in two scales (sizes 7, 9) with

Band	1	2	3	4	5	6	7	8
Filt.size	7&9	11&13	15&17	19&21	23&25	27&29	31&33	35&37
σ	2.8&3.6	4.5&5.4	6.3&7.3	8.2&9.2	10.2&11.3	12.3&13.4	14.6&15.8	17.0&18.2
Λ	3.5&4.6	5.6&6.8	7.9&9.1	10.3&11.5	12.7&14.1	15.4&16.8	18.2&19.7	21.2&22.8
Grid Size	8	10	12	14	16	18	20	22

all orientations and band 2, consists of filter output in two scales (size 11,13). Following that, Gabor filter is applied to input images. Gabor filter adjustments on S1 layer is illustrated in Table 1(Serre et al., 2005). According to Table 1, the permissible scale limits and pooling through Gabor filters are determined in relations 2 and 3:

scale range= {7-9 ‹11-13‹15-17‹19-21‹23-25‹27-29‹31-33‹35-37}   (2)
pool range ={8‹10‹12‹14‹16‹18‹20‹22}   (3)

Tabel I.
adjusted parameters used in our implementation(Serre et al., 2005)

Orient	$0, \dfrac{\pi}{4}, \dfrac{\pi}{2}, \dfrac{3\pi}{4}$
Patch Size	4×4،8×8،12×12،16×16

C1 layer:This layer consists of complex neurons tending to cover larger receptive fields. Accordingly, this layer responds to all edges and orientations anywhere within their receptive fields. This layer is tolerant to scale and image movement shifts. To obtain C1 through the obtained data of S1 layer, a good number of algorithms are offered. Hubel and Wiesel proposed a structure and method to extract C1 from S1(Hubel and Wiesel, 1965). Poggio and Huber also presented another method for layers structures leading to C1 extraction from S1(Lampl et al., 2004). These methods mostly applied max operator to obtain the required data in C1 layer, considering the obtained data from S1 layer. For each band (1 to 8) the maximum of scales and positions are obtained. To be more precise, in each 4 orientations of two available band scales, a grid cell of size $N^{\Sigma} \times N^{\Sigma}$ is used and then, a maximum of two scales in each orientation is taken. For instance, (as shown in Table 1) in band 1, one grid cell of size 8×8 is applied on each orientation of scales 7 and 9. Finally, for each particular orientation, a max is taken. It must be noted that the maximum is exactly taken between two equal orientations and this is repeated for each of the four orientations. Thus, in C1, each band consists of four maps $(C1)^{\Sigma}$ maps.

S2 layer:A large pool of patches of various sizes at random positions are extracted from a target set of images at the C1 level and for all orientations, i.e. a patch *Pi* of size $n_i \times n_i$ contains $n_i \times n_i \times 4$ elements, where the 4 factor corresponds to the four possible S1 and C1 orientations. In this simulation, patches of size $n_i =4,8,12,16$ are used. But, in practice, any size can be considered. The training process ends by setting each of those patches as prototypes or centers of the S2 units which behave as radial basis function (RBF) units during recognition. In other word, each S2 unit response depends in a Gaussian-like way on the Euclidean distance between an input patch (at a particular location and scale) and the stored prototype. This phenomenon is consistent with well-known neuron response properties in primate inferotemporal cortex and seems to be the key property for learning to generalize in the visual systems(Riesenhuber and Poggio, 2002). When a new input is presented, each stored S2 unit is convolved with the new $(C1)^{\Sigma}$ input image at all scales, This leads to $K \times 8(S2)_i^{\Sigma}$ images, where the K corresponds to the patches extracted during learning and the 8 factor, to the 8 scale bands. As for the conducted experiments in this field, the more the number of extracted patches are, the higher the classification accuracy will be.
**C2 layer**: after taking a final max for each S2 feature vectors, the K final C2 features are obtained. these features are independent of movements and scale changes.

III.   **Results and experiments**

In this section, firstly, the data set applied in this article is introduced. Then, the experiments procedure and the results obtained from applying the proposed methods are surveyed.

A. Data set

in the present article, the [1]VISTEX standard texture data set is used in order to investigate the proposed-methods operation. Figure 1 illustrates eight texture types of natural texture images used for classification. The twin images shown of each class, are samples of training and test stage, respectively. Figure 3 also indicates images samples applied for texture images segmentation.


WATER   FABRIC2   WHER-WALDO   BRICK

FOOD   FOOD2   BARK   TERRAIN
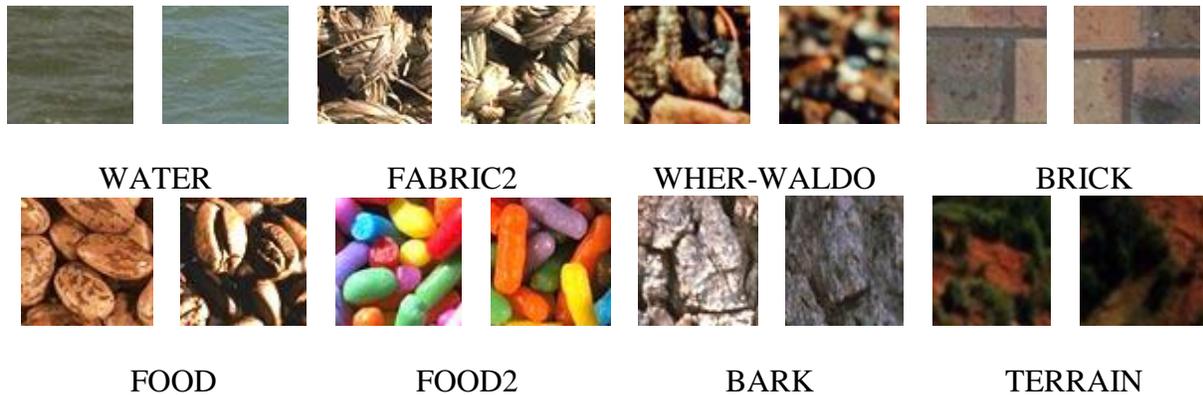
Figure 3: Samples of group of texture images applied in classification process

### B. Natural texture images classification

After extracting texture features through the proposed HMAX model, image classification will be done. This phase consists of 2 stages. First, classifiers are trained through extracted features of training images. Then, classification accuracy is tested and determined through unknown images which are not tolerant to rotation and scale shifts compared to training images. In this article, two classifiers of feedforward neural network and the K-nearest neighbourhood are applied, and finally, the results obtained thereof are compared. According to the experimental findings, utilization of the feedforward neural network classifier increases the system accuracy, compared to the K-nearest neighbourhood. Figure 4 is offered as an illustration.

---

[1] http://wwwwhite.media.mit.edu/vismod/imagery/VisionTexture/vistex.html
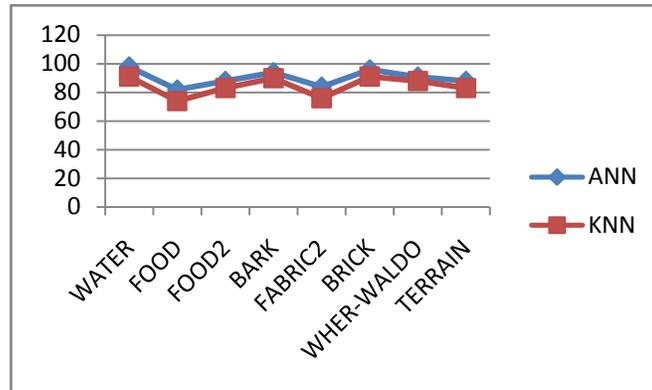
Figure 4. Comparing classification performance of 8-type textures, based on HMAX feature extraction algorithm using K-nearest neighbourhood and feedforward neural network

In order to investigate the proposed algorithm accuracy, the result of classification obtained from this algorithm was compared to the Gabor filter bank. The results show that HMAX method is much more precise and efficient. Figure 5 is a representative of such supremacy.
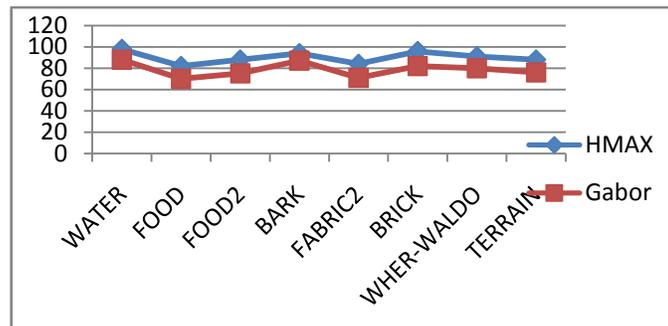
Figure 5. Comparing classification performance of 8-type texture using HMAX feature-extracting algorithm and Gabor, and classifier of feedforward neural network

In Table 2, the general performance obtained from classifying 8-type natural textures is offered. As expected, the general efficiency and accuracy of the proposed method is desirable in classifying random natural textures. As that table shows, HMAX on neural network classifier is 11.5% more accurate than Gabor, while again outperforms it on KNN with 12.5% higher accuracy

TableII.

overall classification accuracy using 8 texture classes

Feature extraction method	classifier	accuracy(%)

HMAX	ANN	90.12
Gabor	ANN	78.62
HMAX	KNN	84.50
Gabor	KNN	72

IV. **Conclusions**

In this paper, the effective and appropriate method, named HMAX, is used to extract texture features. The structure of this method has been inspired form the visual cortex of the animal brain function in detecting objects. With the implementation of this method on VISTEX data set, we achieved the considerable spatial accuracy compared to the performance obtained from the known Gabor filter bank algorithm. After a promosing feature extraction from natural texture images, two classifiers of Artificial Neural networks and K-Nearest neighbor are used to classify the features extracted from textures' images.By comparison of the results we found that the ANN classifier performance is higher than the K-Nearest Neighbor.Regarding this fact that the proposed classification method is associated with high accuracy and efficiency in the context of natural texture images, we intend to implement texture segmentation system based on that successful accomplished classification, as a future work.

References

i. Bian N(2006),"Evaluation of Texture Features For Analysis of Ovarian Follicular Development,"MS Thesis, Department of Computer Science University Saskatchewan, Saskatchewan, Saskatoon

ii. Cooper G(2004)." The textural analysis of gravity data using co-occurrence matrices," Elsevier Journal of Computers & geosciences, Vol(30),pp. 107-115,

iii. Gawne T and Martin J(2002)."Response of primate visual cortical V4 neurons to simultaneously presented stimuli," J.Neurophysiol, Vol(88),pp. 1128– 1135,.

iv. Hubel D, and Wiesel T(1965),"Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) Of the cat," J.Neurophys, Vol(28), pp. 229– 289,.

v. Krishnamachari S and Chellappa R(1997)."Multiresolution Gauss Markov Random Field Models for Texture Segmentation,"IEEE Transactions on Image Processing, Vol(6),pp.251-267,.

vi. Lampl I, Ferster D, Poggio T, and Riesenhuber M(2004). "Intracellular measurements of spatial integration and the max operation in complex cells of the cat primary visual cortex". J. Neurophysiol., 92: 2704– 2713,.

vii. Li M and Staunton R(2008)."Optimum Gabor filter   design and local binary patterns for texture segmentation," Elsevier Journal of Pattern   Recognition Letters,Vol(29), pp. 664 – 672,.

viii. Marr D(1982).Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. San Francisco: WH Freeman and Co,

ix. Meyers E and Wolf L(2008)." Using Biologically Inspired   Features for Face Processing," Springer Journal of Computer Vision, Vol(76), pp. 93-104,.

x. MIT Media Lab. VisTex: Vision Texture database. Retrieved 1 Feb 2010 from the World Wide Web: http://wwwhite.media.mit.edu/vismod/imagery/VisionTexture/vistex.html, 2010.

xi. Riesenhuber M and Poggio T(2002)." Neural mechanisms of object recognition". Curr. Op. Neurobiol.,12:162– 68,.

xii. Riesenhuber M, and Poggio T(1999)." Hierarchical models of object recognition in cortex," Nat. Neurosci, Vol(2), pp. 1019–1025,.

xiii. Serre T, Wolf L , Bileschi S and Riesenhuber M(2007). ,"Object recognition with cortex-like mechanisms," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol(29), pp.411-426,.

xiv. Serre T, Wolf L and Poggio T(2005)."Object recognition with features inspired by visual cortex," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol(2), pp.994-1000,.

xv. Signolle N, Revenu M, Plancoulaine B and Herlin P(2009). ,"Wavelet-based multiscale texture segmentation: Application to stromal compartment characterization on virtual slides," Elsevier Journal of Signal Processing,Vol(90), pp. 2412-2422,.

xvi. Young R(1987)." The Gaussian derivative model for spatial vision: I. Retinal mechanisms,"Spatial Vision, Vol(2), pp.273–293,.