

Colour Object Recognition using Biologically Inspired Model

S. Arivazhagan¹ R. Newlin Shebiah² P. Sophia³, A. Nivetha⁴
^{1,2,3,4} Department of Electronics and Communication Engineering
^{1,2,3,4} Mepco Schlenk Engineering College, Sivakasi - 626 005

Abstract- Human visual system can categorize objects rapidly and effortlessly despite the complexity and objective ambiguities of natural images. Despite the ease with which we see, visual categorization is an extremely difficult task for computers due to the variability of objects, such as scale, rotation, illumination, position and occlusion. Utilization of characteristics of biological systems for solving practical problems inevitably leads towards reducing the gap between manmade machines and live systems. Biologically motivated information processing has been an important area of scientific research for decades. This paper presents a Biologically Inspired Model which gives a promising solution to object categorization in colour space. The features are extracted in YCbCr colour space and classified by using SVM classifier. The framework has been applied to the image dataset taken from the Amsterdam Library of Object Images (ALOI). The proposed framework can successfully detect and classify the object categories with a good accuracy rate of about 91.3% for the Cb plane.

I. INTRODUCTION

Object recognition plays an important role in car number plate recognition[1], face recognition for the purpose of access control [2] and cancer recognition [3], applications related to computer vision such as video surveillance [4], image and video retrieval [5], web content analysis [6], human computer interactions [7] and biometrics[8]. In general, object categorization is a difficult task in computer vision because of the variability in illumination, scales, rotation, deformation and clutter as well as the complexity and variety of backgrounds.

W. Niblack, R. Barber, W. Equitz, et.al proposed traditional appearance-based approaches in the object recognition which mainly use global low-level visual features such as gray value, color, shape, and texture [9]. These methods do not consider local discriminative information and are sensitive to lighting conditions, object poses, clutter, and occlusions.

J. Amores, N. Sebe, and P. Radeva proposed Part-based models [10] that make matches between particular patches and interesting objects through various searching schemes. In this framework, it is challenging to robustly segment and find the meaningful parts, so the spatial relationships of meaningful parts cannot be duly modeled. G. Csurka, C. Bray, C. Dance introduced the original bag-of-features based scheme [11] is efficient for recognition, but it ignores the spatial relationship of features, and thus it is hard to represent the geometric structure of the object class or to distinguish between foreground and background features. D. G. Lowe extracted Distinctive image features

from dcale-invariant key-points. This is a local feature based approach that combines the interest point detectors and local descriptors with spatial information. Representative local features include scale-invariant feature transform (SIFT) [12].

Although these features are effective in describing local discriminative information, they lack higher level information, e.g., relations of local orientations. T. Serre, L. Wolf, and T. Poggio in [13] used a set of complex biologically inspired features obtained by combining the response of local edge-detectors that are slightly position- and scal e-tolerant over neighboring positions and multiple orientations. It produced an admirable results as the classification rate obtained is above 35% correct when using 15 training examples, however, it could not perform well specially when it was applied to images with high clutter or with partially occluded objects.

T. Serre, L. Wolf, S. Bileschi, et.al in [14] proposed a new set of scale and position-tolerant feature detectors that are adaptive to the training set. This approach demonstrates good classification results on a challenging (street) scene understanding application that requires the recognition of both shape-based as well as texture-based objects. The classification rate obtained is above 44% correct when using 15 training examples. Jim Mutch and David G. Lowe in [15] builds on the approach of [14] by incorporating some additional biologically-motivated properties, including sparsification of features, lateral inhibition, and feature localization. These modifications enhance recognition accuracy, but the heavy computational load and the feature selection are still a big problem. The classification rate obtained is 51% correct when using 15 training examples. J. Mutch and D. G. Lowe in [16] updates and extends the approach of [15] by incorporating some additional biologically motivated properties, specifically, sparsity and localized intermediate-level features. These modifications show that each of these changes provides a significant boost in generalization performance. The classification rate obtained is 51% correct when using 15 training examples similar to [15].

This paper is structured as follows: Section II describes about the proposed methodology. Results and discussion is given in section III. Next section gives the concluding remarks.

II. PROPOSED METHODOLOGY

The BIM consists of four layers of computational units: S_1 , C_1 , S_2 , and C_2 .

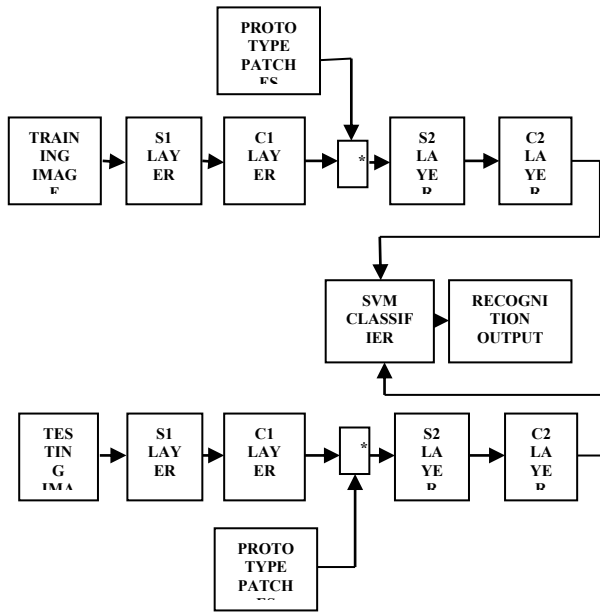


Fig.1: Block Diagram of Proposed Methodology

A. S_1 Units

The units in the S_1 layer correspond to the simple cells in the primates' visual cortex. An initial input image is convolved with different Log-Gabor filters to produce the S_1 layer. The Log-Gabor filters are used here since they are found to be more advantageous than the Gabor filters. Gabor filters are not optimal if one is seeking broad spectral information with maximal spatial location and also the maximum bandwidth of the Gabor filter is limited to approximately one octave.

B. Log-Gabor filters

Log-Gabor filters basically consist in a logarithmic transformation of the Gabor domain [17] which eliminates the annoying DC-component allocated in medium and high-pass filters. Log-Gabor wavelet transforms allow exact reconstruction and strengthen the excellent mathematical properties of the Gabor filters.

C. C_1 units

The C_1 units describe complex cells in the visual cortex. To generate a C_1 unit, BIM pools over S_1 units using a maximum operation. This process can be considered as a down-sampling operation over the S_1 images.

$$C_1(i) = \max(S_1(i, :, :))$$

Where $C_1(i)$ is the afferent C_1 image and $S_1(i)$ is the group of S_1 images.

D. S_2 units

The S_2 units describe the similarity between C_1 images and prototypes via convolution operation. An S_2 image is calculated by

$$S_2(i, j, k) = C_1(j, k) * P_i$$

$C_1(j, k)$ is the afferent C_1 image with a specific (2) and a specific orientation k . P_i is a patch selected from the input image

E. Extraction of patches

The patches are not selected randomly. Threshold is applied to the input image. The pixels in the image having the value lesser than the threshold are assigned with zero value. Threshold value of 50% is chosen here. Then the resulting input image is divided into blocks of uniform sizes. Thus for the input image of size 200x200, hundred 20x20 blocks are obtained. From the divided blocks, the blocks having the pixel value greater than 75% are selected as the essential patches. From reducing the computational complexity only twenty patches from the essential patches are taken for the experiment. These prototype patches are then convolved with the Log-Gabor filters of four scales and four orientations. Finally (20*16=320) three hundred and twenty prototype patches are obtained and stored.

F. C_2 units

A C_2 value is the global maximum response of a group of S_2 image over different scales and orientations. An C_2 image is calculated by

$$C_2(i) = \max(S_2(i, :, :)) \quad (3)$$

where $C_2(i)$ is the afferent C_2 image and $S_2(i, :, :)$ is the group of S_2 images.

G. SVM classifier

The C_2 vectors are classified using a linear classifier called SVM classifier. It separates a set of objects into their respective groups. It constructs a hyper plane or set of hyper planes in an infinite dimensional space which can be used for classification. SVM is widely used in object detection and recognition content-based image retrieval, text recognition, biometrics, speech recognition.

III. RESULTS AND DISCUSSIONS

The proposed methodology was evaluated on the image dataset taken from the Amsterdam Library of Object Images (ALOI). For the object recognition using the biologically inspired model five categories are such as car, bike, basket, bowl and shoe are taken into consideration. Each category contains two classes. For example the car category contains two different types of car such as blue car and white car. The car dataset contains 108 images in the blue car and 108 images in the white car. Similarly in each category 108 images are seen in the first class and 108 images are seen in the second class. These images are of various sizes and for the experiment they were resized to 200x200. The images of the five categories are shown in the fig.2



Fig. 2: Sample images from Amsterdam Library of Object Image. From left to right, the categories are car, basket, bowl, shoe, and ball.

Initially as a sample, the car image is taken as input to the proposed framework. The input image is resized to 200x200. Fig.3 shows the resized input image. Colour transformation on the input image is performed. The Y, Cb, Cr contents of the input image are also shown in the fig.3

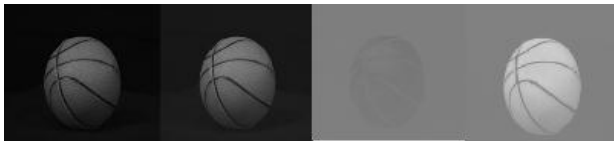


Fig.3: Colour transformed images. From left to right are the input image, Y content, Cb content and Cr content

After the color transformation the pixels having value lesser than the threshold value are replaced with zero. Threshold of about 50% is applied to the transformed image. Then the features are extracted in four layers such as S_1 layer, C_1 layer, S_2 layer and the C_2 layer.

A. S_1 layer

The image after applying threshold is convolved with the Log-Gabor filters. Here Log-Gabor filter of four orientations (0° , 45° , 90° and 135°) and four scales are used. Hence the output of 16 images is obtained for a single input image. The S_1 units obtained for the sample input image 'ball' is shown in the fig.4.

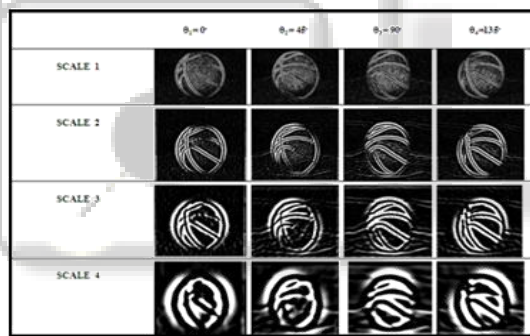


Fig. 4 : Sample Image of ball at S_1 layer

B. C_1 layer

The output values from S_1 layer are used in the C_1 layer. To generate a C_1 unit, BIM pools over S_1 units using a maximum operation. Each C_1 unit is obtained by taking the maximum over two S_1 units of same orientation but different scales. Thus for the S_1 layer of 16 images 8 C_1 images are obtained. This process can be considered as a down-sampling operation over the S_1 images.

The C_1 images obtained for the sample input image 'ball' is shown in the Fig.5

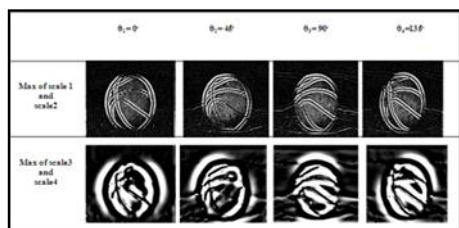


Fig.5 : C_1 units

C. Extraction of patches

For the S_2 layer the patches are selected from the input image. The thresholding at the initial stage is followed by the segmentation for the patch extraction. It is divided into hundred uniform blocks of size 20x20. It is not necessary that all the patches are having more information. So the patches that are having more than 75% of image pixels are selected as essential patches. For reducing the computational complexity only twenty patches from the essential patches are taken for the experiment. These prototype patches are then convolved with the Log-Gabor filters of four scales and four orientations. Finally ($20 \times 16 = 320$) three hundred and twenty prototype patches are obtained for the sample input image.

D. S_2 layer

The values for S_2 units are calculated using the output values of the C_1 layer. The S_2 units describe the similarity between the C_1 images and prototype patches via the convolution operation. The extracted 320 prototype patches are convolved with the 8 C_1 images. Out of these 320 patches, first 80 patches belong to the orientation 0° , next 80 patches to the orientation 45° , next 80 patches to the orientation 90° and the last 80 patches belong to the orientation 135° . Convolution operation is taken along the C_1 images and the prototype patches of same orientation. And thus at last we obtain 640 S_2 images. Fig.6 shows the S_2 layer units obtained when a single prototype patch is convolved with a C_1 unit of same orientation

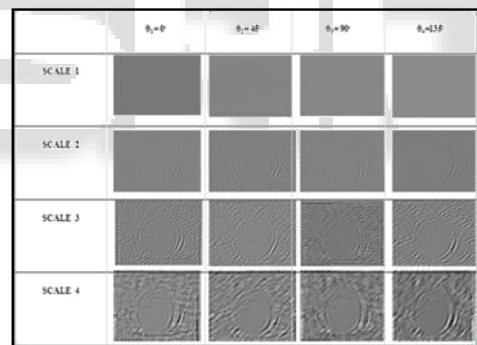


Fig.6 : S_2 units

E. C_2 layer

The output values from S_2 layer are used in the C_2 layer. A C_2 value is the global maximum response of a group of S_2 image over different scales and orientations. This further allows the layer to be more tolerant to shifts and scale changes within its receptive field. Each C_2 unit is obtained by taking the maximum over two S_2 units of same orientation but different scales. Thus for the S_2 layer of 640 images 320 C_2 values are obtained. Finally 320 features are obtained for the sample input image 'car'.

F. Classification

These 320 extracted features are fed into the classifier stage classification. The detection of the objects is done by the SVM classifier. The car images are divided into training set and testing set, where 50% of images are used to train the system and the remaining 50% of the images serves as the testing set. Similarly other images are also divided into two

sets, training set, and testing set. The number of images used for training, testing and classification gain for each category of objects in YCbCr plane is shown in the table.1 The classification gain is improved with 91.3% for Cb plane.

Category	No of Images		Recognition rate		
	Training	Testing	Y	Cb	Cr
Car	108	108	50	82.4	70.4
Ball	108	108	80.5	100	85.2
Shoe	108	108	47.2	89.8	93.5
Bowl	108	108	88.9	100	100
Basket	108	108	79.6	84.3	90.7

Table. 1: Results obtained using SVM classifier in YCbCr colour space

Y represents luma or luminance. Luma is the brightness in an image (black and white). It ranges from 0-100 where 0 denotes the absolute black and 100 denotes absolute white. Whereas chroma is the purity of a colour. It represents the colour information. It ranges from 0-100 where 0 denotes the absence of colour and 100 denotes the maximum of colour. High chroma means rich and full colour. Low chroma means dull and pale colour. Cb represents blue chroma and Cr represents red chroma. Here the recognition rate obtained for Cb plane is higher when compared to Y and Cr plane especially for the categories such as car, ball and bowl due to the presence of colours such as blue, yellow, etc. The recognition rate obtained for Cr plane is good for the categories shoe, bowl and basket due to the presence colours such as red. When comparing plane with CbCr plane, it doesn't yield that much good result. The comparison of Y content, Cb content and Cr content is shown in the fig.7.

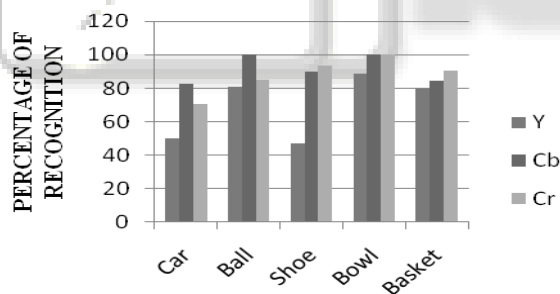


Fig.7 : Comparison of Y, Cb and Cr using SVM classifier

IV. CONCLUSION

Recognizing and classifying various objects is mainly the purpose of the proposed approach. Thus the proposed methodology was tested on the five categories namely Car, Bowl, Basket, Shoe and Ball from the Amsterdam Library of Object Images. The experimental results indicate that the proposed approach is a valuable approach, which can significantly recognize and classify the objects with a little computational effort. From the analysis of results obtained by classifying those object categories, the classification rate using Cb plane is nearly 91.3% which is the highest recognition rate obtained. It gives good results for car, ball and bowl. The classification rate using Cr plane is about 87.96% on an average. It gives better results for the

categories such as shoe, bowl and basket. This model can be used for machine vision applications.

REFERENCES

- [1] Lihong Zheng and Xiangjian He.: Number Plate Recognition Based on Support Vector Machines. In: *IEEE International Conference on Video and Signal Based Surveillance, AVSS '06*. page(s): 13-13.(2006).
- [2] D. Bryliuk and V. Starovoitov.: Access Control by Face Recognition Using Neural Networks, <http://neuroface.narod.ru.>(2000).
- [3] Jihong Liu and Weina Ma.: An Effective Recognition Method of Breast Cancer Based on PCA and SVM Algorithm. *ICMB 2008, LNCS 4901*, pp. 57-64 (2007).
- [4] R. Collins, A. Lipton, T. Kanade, et.al "A system for video surveillance and monitoring," *Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep.: CMU-RI-TR-00-12*, 2000.
- [5] A. Yoshitaka and T. Ichikawa, "A survey on content-based retrieval for multimedia databases," *IEEE Trans. Knowl. Data Eng., vol. 11, no. 1*,pp. 81-93, Jan./Feb. 1999.
- [6] R. Kosala and H. Blockeel, "Web mining research: A survey," *ACMSIGKDD Explorations Newslett., vol. 2, no. 1*, pp. 1-15, Jul. 2000.
- [7] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7*, pp. 677-695, Jul. 1997.
- [8] A. K. Jain, A. Ross, and S. Prabhakar, Eds., *Biometrics: Personal Identification in Networked Society*. Norwell, MA: Kluwer, 1999.
- [9] W. Niblack, et.al, "The QBIC project: Querying images by content using color, texture and shape," in *Proc. SPIE—Storage and Retrieval for Image and Video Databases*, 1993, pp. 173-187.
- [10] J. Amores, et.al, "Context-based object class recognition and retrieval by generalized correlograms," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 10*, pp. 1818-1833, Oct. 2007.
- [11] G. Csurka, et.al, "Visual categorization with bags of keypoints," in *Proc. ECCV*, 2004, pp. 1-16.
- [12] D. G. Lowe. Distinctive image features from dcale-invariant key-points. *International Journal of Computer Vision*, 2(60):91-110, 2004.
- [13] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," in *Proc. CVPR*, 2005, pp. 994-1000.
- [14] T. Serre, L. Wolf, S. Bileschi, et.al, "Robust object recognition with cortex-like mechanisms," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 3*, pp. 411-426, Mar. 2007
- [15] J. Mutch and D. G. Lowe, "Multiclass object recognition with sparse,localized features," in *Proc. CVPR*, 2006, pp. 11-18.
- [16] J. Mutch and D. G. Lowe, "Object class recognition and localization using sparse features with limited receptive fields," in *Proc. IJCV*, 2008, pp. 45-57.
- [17] D.J. Field. Relation between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379_2394,1987