

# Feature Selection in Top-Down Visual Attention Model using WEKA

Amudha.J  
Research Scholar  
Department of CSE  
Amrita School of Engineering  
Karnataka, India

Soman.K.P  
Professor & Head  
Department of CEN  
Amrita Vishwa Vidyapeetham  
Tamilnadu, India

Kiran.Y  
PG Scholar  
Department of CSE  
Amrita School of Engineering  
Karnataka, India

## ABSTRACT

A feature selection in Top down visual attention model for sign board recognition has been incorporated to reduce the computational complexity and to enhance the quality of recognition. The approach is based on a biologically motivated attention system which is able to detect regions of interest in images based on the concepts of the human visual system. A top-down guided visual search module of the system identifies the most discriminate feature from the previously learned target object and uses to recognize the object. This enables a significantly faster classification and is illustrated in identifying signboards in a road scene environment.

## General Terms

Computer Vision, Pattern Recognition, Biological Vision

## Keywords

Visual Attention, Saliency, Human Perception, Computational Attention system, Decision tree.

## 1. INTRODUCTION

The simple tasks of human vision system open your eyes, see it and interpret it, are indeed a complex processing which has been remarkably well evolved in recent times. The most important aspect of our vision system is based on the attention models [2-3],[7],[8],[12] which focus the brain to restrict to the salient information in the scene. Many evolutionary achievements enable an efficient exploitation of the brain's capacity, e.g. by restriction of high resolution sight to a small area of the retina, the fovea. While this reduces the required processing capacity for recognition, it also disables human to grasp an entire scene at once. However, human perform sequence of eye movements, saccades, for grasping a scene. As goal for a saccade, a subset of the visual input is selected by visual attention which is another evolutionary achievement. Bottom up attention directs the gaze to salient regions, while top-down attention enables goal directed visual search. Computational vision systems are far from achieving human general performance, but good specialized systems exist. For robot vision tasks, real-time performance is essential, and fairly general recognition capabilities are often desired. A biologically inspired approach for reducing the demand for image processing capacity is the computational simulation of visual attention, i.e., the detection of salient regions and a restriction of classification to these fractions of the input image.

Detecting regions of interest with visual attention is an important mechanism in human visual perception. However,

what is of interest depends on the situation. Top-down influences also play an important role in human visual attention: knowledge, motivations, emotions also play an important role in human visual attention. For example, drivers consider signboard in great detail rather than other parts of the road scene. In human behavior, bottom-up and top-down attention are always intertwined and may not be considered separately, although one may outweigh the other in certain situations. Even in a pure exploration mode, each person has own preferences resulting in individual scan-paths for the same scene.

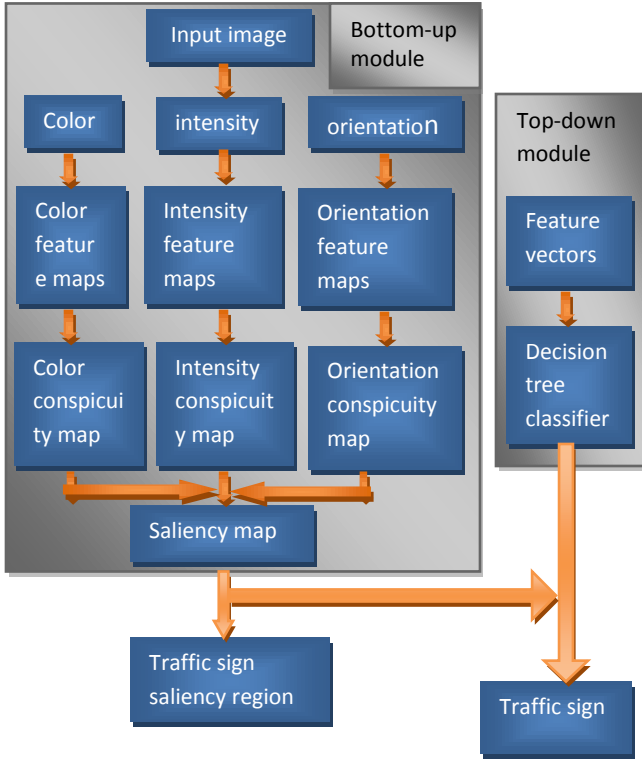
On the other hand, even if searching highly concentrated for a target, the bottom-up pop-out effect is not suppressible, an effect called attention capture [17]. Despite its importance in the human visual system, top-down influences are rarely considered in computational attention systems. One of the reasons is that the neuro-biological foundations are not yet completely understood. Nevertheless, the extension of an attention system with top-down mechanisms is unavoidable if regions of interest shall be detected depending on a task. Moreover, the evaluation of the system is much easier with this extension since ground truth is available. An important model is VOCUS [12-15] that is able to regard top-down cues. In a learning phase, the system learns target-relevant features from a training image considering the properties of the target as well as the surrounding. In search mode, the system considers the information to excite or inhibit features and computes a target dependent top-down saliency map.

In this paper, a top down model using visual attention for the detection of signboard objects in real-world images has been incorporated. First, regions of interest are focused by an attention module, using either pure [2] bottom-up attention (exploration) or the combination with goal-dependent top-down cues (visual search). Secondly, the region of interest is fed into a classifier detecting learned objects. The performance of the bottom up visual attention model has been tested on a video annotation application [1].The performance of the top down visual attention system was tested for signboard detection in a road scene environment. In the proposed model a decision tree (WEKA) is used in learning phase to identify the most discriminate feature and those features are used to identify the target which in turn reduces the computational complexity and increases the efficiency of the system.

We briefly review the details of the model in section II with extensions in the same formal framework. In section III the analysis of the performance of the model has been tested with

various cases and compared. In section IV conclusions and further enhancements are discussed.

## 2. SIGN BOARD DETECTION SYSTEM



**Figure.1 Visual Attention Model with decision tree classifier**

The block diagram in Figure.1 describes the flow of the system. 1) The Visual Attention model (VAM) identifies the most attended region 2) Feature Extraction from the VAM for signboard detection 3) The feature selection identifies the relevant features required for the signboard identification and uses for further classification. The following sections present the algorithm in detail.

### 1.1 Visual attention model

The input image  $I$  is sub-sampled into a Gaussian pyramid on 4 different scales, and each pyramid level is decomposed into channels for red(R), green (G), blue (B), yellow (Y), intensity (I) using (1), (2), (3), (4) and (5) The orientation channels are obtained from  $I$  using oriented Gabor pyramids  $O(\sigma, \theta)$ , where  $\sigma \in [0..5]$  is the scale, and  $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$  is the preferred orientation.

$$I = \frac{r+b+g}{3} \quad (1)$$

$$R = r - \frac{(g+b)}{2} \quad (2)$$

$$G = g - \frac{(r+b)}{2} \quad (3)$$

$$B = b - \frac{(r+g)}{2} \quad (4)$$

$$Y = r + g - 2(|r - g| + b) \quad (5)$$

The quantity corresponding to the double-opponency cells in primary visual context are then computed by center surround differences across the normalized color channels. Each of the three-red /green Feature map is created by first computing (red-green) at the center, then subtracting (green-red) from the surround and finally outputting the absolute value. Accordingly maps  $RG(c,s)$  are created in the model to simultaneously account for red/green and green/red double opponency and  $BY(c,s)$  for blue/yellow and yellow/blue double opponency and orientation using (6), (7) and (8).

$$F_{RG,CS} = N(|(R(c) - G(s)) \ominus (R(s) - G(s))|) \quad (6)$$

$$F_{BY,CS} = N(|(B(c) - Y(s)) \ominus (B(s) - Y(s))|) \quad (7)$$

$$F_{\theta,c,s} = N|O(c, \theta) \ominus O(s, \theta)| \quad (8)$$

The feature maps are then combined into three conspicuity maps, intensity  $\bar{I}$  (9), color  $\bar{C}$  (10), and orientation  $\bar{O}$  (11), at the saliency map's scale ( $\sigma = 4$ ). These maps are computed through across-scale addition ( $\oplus$ ), where each map is reduced to scale four and added point-by-point:

$$\bar{I} = \bigoplus_{c=2}^4 \bigoplus_{s=c+2}^{c+4} N(I(c, s)) \quad (9)$$

$$\bar{C} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [N(RG(c, s)) + N(BY(c, s))] \quad (10)$$

To compute the orientation conspicuity map, four intermediary maps are created by combining the six feature maps. These intermediary maps are then combined into a single orientation conspicuity map.

$$\bar{O} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} N(\bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} N(O(c, s, \theta))) \quad (11)$$

The three conspicuity maps are then normalized and summed into the input  $S$  to the saliency map:

$$S = N(\bar{I}) + N(\bar{C}) + N(\bar{O}) \quad (12)$$

The  $N(\cdot)$  represents the non-linear Normalization operator. The three conspicuity maps are normalized and summed into the final input  $S$  to the Saliency map (13). From the saliency map the most attention regions are identified in the order of decreasing saliency based on the selective tuning model [2].

### 2.2 Feature Extraction

Since we are only interested in the saliency of the target object, here a signboard, the most salient region which is the Focus of Attention (FOA) is determined. From the FOA, 46 feature vectors are determined by extracting the values of the 42 feature maps (RG, BY, Intensity, Orientation  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  with six spatial levels each) and 3 conspicuity maps (Color, Intensity, Orientation) and 1 saliency map as shown in Table 1. Same procedure is repeated for the next five attended locations. It

describes how much each feature contributes to the FOA. The classifier identifies whether the obtained feature vector from the target region matches with the actual target.

**Table 1: Total number of features (FM: Feature map  
 CM: Conspicuity map, x: 1-6)**

NO	TYPE OF FEATURE	TOTAL
1	RG FM x	06
2	BY FM x	06
3	INT FM x	06
4	0 FM x	06
5	45 FM x	06
6	90 FM x	06
7	135 FM x	06
8	COLOUR CM	01
9	INTENSITY CM	01
10	ORIENTATION CM	01
11	SALIENCY MAP	01
12	TOTAL	46

### 1.3 Feature Selection & Classification

To identify whether the salient region is the actual target, a decision tree classifier is used. For the first step of the process we trained multiple machine learning models using the provided training set. The algorithms for these models came from the Waikato Environment for Knowledge Analysis (WEKA), a suite of machine learning algorithms developed at the University of Waikato. The following models were used.

#### 2.3.1. REP Tree:

To identify the class information a quick decision tree generator that uses information gain and variance is used [5].

#### 2.3.2. BF Tree:

A decision tree that uses a best first method for determining its branches [7]

#### 2.3.3. J48 Tree:

It builds the decision tree from labeled training data set using information gain and it examines the same that results from

choosing an attribute for splitting the data. To make the decision the attribute with highest normalized information gain is used. Then the algorithm recurs on smaller subsets. The splitting procedure stops if all instances in a subset belong to the same class. Then the leaf node is created in a decision tree telling to choose that class.

#### 2.3.4. SIMPLE CART

CART (Classification and regression trees) was introduced by Breiman[4]. It builds both classifications and regressions trees. The classification tree construction by CART is based on binary splitting of the attributes. It is also based on Hunt's model of decision tree construction and can be implemented serially [5]. It uses gini index splitting measure in selecting the splitting attribute. Pruning is done in CART by using a portion of the training data set [10]. CART uses both numeric and categorical attributes for building the decision tree and has in-built features that deal with missing attributes [9].

#### 2.3.5. LAD tree

Logical Analysis of Data is the method for classification proposed in optimization literature. It builds a classifier for binary target variable based on learning a logical expression that can distinguish between positive and negative samples in a data set. The basic assumption of LAD model is that a binary point covered by some positive patterns, but not covered by any negative pattern is positive, and similarly, a binary point covered by some negative patterns, but not covered by positive pattern is negative. The construction of Lad model for a given data set typically involves the generation of large set patterns and the selection of a subset of them that satisfies the above assumption such that each pattern in the model satisfies certain requirements in terms of prevalence and homogeneity.

#### 2.3.6. RANDOM tree

It is the randomized process used for split attribute selection and provides the expression for the probability of choosing an attribute/a set of attributes. The attribute selection method we use is as follows. We assume a uniform probability distribution in selecting the attribute variables, that is, attributes which have already not been chosen in a particular branch, have an equal chance of being chosen for the next level.

## 3. RESULTS AND ANALYSIS

Figure 2 shows a sample of the signboards considered to test the system. It includes signboards of pedestrian, crossing and bike.





**Figure 2- Test Images for signboards**

The analysis is done in three levels 1) Survey on signboard images using Visual Attention Model 2) Identification of discriminate features.

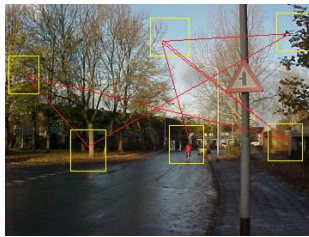
### 3.1 Survey using Visual Attention Model

First, a survey was done to identify whether the signboard is identified by the visual attention model for the signboard dataset. From figure 2, it is shown that the VAM identifies the signboard in some cases at the first level of iteration, in some at later levels and in few it failed. Table 2 gives a statistics of the number of cases that were able to identify the signboard as the first attention, in cases it was able to identify only in the 2, 3 and later levels of iteration and the number of failed cases.



(a)

(b)



(c)

**Figure 3 Attention area identified by VAM in various cases**

(a) First level (b) Fourth level (c) Failed

**Table 2: Identification of signboard as the attention location by the Bottom up Visual attention Model at various levels**

Image Type	Total Number of Images	Attention model		Failed to Identify
		First Level	Other Levels	
Pedestrian	16	7	9	0
Bike	16	5	9	2

Crossing	16	2	13	1
----------	----	---	----	---

### 3.2 Feature Extraction

Hence in this case the features obtained in the bottom up model are used as a training set for the classifier. To train the decision tree classifier, a set of two eighty eight samples are considered as shown in Table 3. The data set contains the best selected features from 42 feature maps (RG, BY, Intensity, Orientation 0°, 45°, 90°, 135° with six spatial levels each) and 3 conspicuity maps (Color, Intensity, Orientation) and 1 saliency map were extracted and the class to which it belongs are taken as input data to train the classifier. The feature vectors Table 3 has the information of class labeling and total number of samples used and is labeled as sb (sign board) and nsb (non sign board).

**Table 3: Training samples for Signboard detection**

Sign board	class	Number of samples
Pedestrian	sb	21
Bike	sb	18
Crossing	sb	19
Others	nsb	230
Total		288

### 3.3 WEKA for feature selection

The data set with all features along with class labels is given as an input to WEKA. The table 4 gives number of features selected for each classifier that are required to recognize the sign board. BF tree gives the least number of selected features when compared to other classifier models.

**Table 4: Different Classifiers and number of features Selected for classification**

Classifier model	Number of features Selected
J48	12
Simple cart	07
Random tree	21
LAD tree	10
REP tree	06
BF tree	05

There are some parameters that evaluates different classifier models. They are as follows:

### 3.4 Testing paramaters

#### 3.4.1 Confusion matrix:

Each row of the matrix represents the instances in a predicted class, while each column represents the instances in an actual class. One benefit of a confusion matrix is that it is easy to see if the system is confusing two classes (i.e. commonly mislabeling one as another)

### 3.4.2 True Positive (TP):

Proportion classified as class x / Actual total in class x . Equivalent to Recall.

### 3.4.3 False Positive (FP):

Proportion incorrectly classified as class x / Actual total of all classes, except x.

### 3.4.4 Precision:

Proportion of the examples which truly have class x / total classified as class x.

### 3.4.5 F-measure:

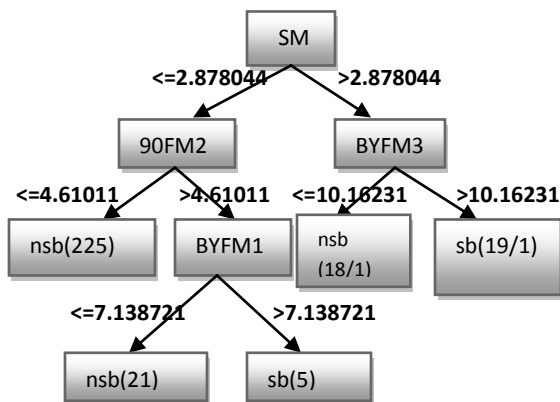
It is a combined measure of precision and recall. It is equivalent to  $2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$ .

## 3.5 Comparison of decision tree classifiers:

Table 5 gives the number of features used by the classifier at different levels. So it is explicit that some features are sufficient to bring out the maximum efficiency for the application considered.

**Table 5: Number of features used by the classifier at different depth of the tree.**

Depth	J48	Rep Tree	BF Tree	Random tree	Lad Tree
0	SM	BYFM <sub>1</sub>	SM	SM	SM, BYFM <sub>1</sub> , 90 FM <sub>2</sub> ,
1	90 FM <sub>2</sub> BY FM <sub>3</sub>	SM	BY FM <sub>3</sub>	BY FM <sub>1</sub> , BY FM <sub>3</sub>	BY FM <sub>3</sub> , INT FM <sub>5</sub> , 90 FM <sub>2</sub> ,
2	BY FM <sub>1</sub>	90FM <sub>2</sub> , INT FM <sub>5</sub>	90 FM <sub>3</sub> .	90FM <sub>2</sub> , BYFM <sub>1</sub> , SM	INTFM <sub>5</sub> ,
3	-----	-----	-----	BYFM <sub>1</sub> , INTFM <sub>5</sub>	----



**Figure 4: J48 tree**

The efficiency of the system does not improve by increasing the number of features. We have certain common features selected by the classifiers at any particular level of the tree. The feature SM is the root of the tree used by all classifiers. INTFM<sub>5</sub>, BYFM<sub>3</sub>, 90FM<sub>2</sub>, BYFM<sub>3</sub>, are the common features at levels 1 to 3.

**Table 6: Different classifiers and their accuracies**

Classifier model	Accuracy
J48	98.26%
Simple cart	97.56%
Random tree	97.22%
LAD tree	98.26%
REP tree	95.48%
BF tree	96.87%

**Table 7: Comparison of confusion matrix parameters of different classifier models**

Classifier model	TP (true positive)	FP (false positive)	Precision	Recall	Fmeasure
J48	0.983	0.039	0.984	0.983	0.983
Simple cart	0.965	0.306	0.964	0.965	0.963
Random tree	0.972	0.192	0.971	0.972	0.972
LAD tree	0.983	0.039	0.984	0.983	0.983
REP tree	0.955	0.383	0.951	0.955	0.951
BF tree	0.969	0.306	0.968	0.969	0.966

Table 6 gives the accuracies of different classifiers. J48 and LAD tree are the classifiers having highest accuracy when compared with other models. This is because of having the highest True positive rate as shown in Table 7 which gives the comparison of confusion matrix parameters of different classifier models. The selected features used by the classifier give accuracy in the classification up to 98.26% whereas the use of the Visual attention model for classification

## 3.6 Advantages of the system

The feature selection adopted here helps to reduce the search time involved in identifying the target object. The feature selection used before recognizing reduces the search time as well give better recognition rate. Hence the usage of feature selection module before applying all the features used in visual attention model for the classifier improves both time and quality of recognition.

## 4. CONCLUSION

The paper proposes an approach to extend the usage of features extracted from the attention model to detection model hence reduces the computational overhead default existing in combining two different systems. This model decreases the computational complexity and increases the quality of the search for the objects. The model has been tested on a vehicle tracking

application also and the results are encouraging [3]. In the future work, the proposed model can be extended to localize the salient area to detect the traffic sign regions, a lower false positive rate can be achieved by increasing the number of training samples and then can be extended for a real time traffic sign detection system.

## 5. REFERENCE

- [1] Amudha J, Soman K.P, Vasanth. K, “Video Annotation Using Saliency”, International conference on Image processing, Computer vision and Pattern Recognition” Vol 1, 191-195, 2008.
- [2] Amudha J and Soman K.P., “Selective tuning Visual Attention Model” International Journal of Recent Trends in Engineering, November 2009, Finland.
- [3] Amudha J and Soman K.P., “Saliency based Visual Tracking of vehicles” International Journal of Recent Trends in Engineering, November 2009 Issue, Finland.
- [4] Breiman, L., Random Forests, Machine Learning, 45(1), 5-32, 2001
- [5] Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. , Classification and Regression trees, Wadsworth, Belmont, CA(1984)
- [6] Ian H. Witten; Eibe Frank, Data Mining: Practical machine learning tools and techniques, 2nd Edition. Morgan Kaufmann, San Francisco. 2005
- [7] L.Itti, C. Koch, and E. Niebur, ‘A model of saliency-based visual attention for rapid scene analysis’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11), 1254–1259, 1998
- [8] Koch.C and Ullman.S, “Shifts in selective visual attention: towards the underlying neural circuitry”, Human Neurobiology, vol. 4, pp. 219-227, 1985.
- [9] Lewis, R.J. An Introduction to Classification and Regression Tree (CART) Analysis. 2000 Annual Meeting of the Society for Academic Emergency Medicine, Francisco, California 2000.
- [10] Podgorelec, V., Kokol, P., Stiglic, B., Rozman, I. Decision trees: an overview and their use in medicine, Journal of Medical Systems Kluwer cademic/Plenum Press, vol.26, Num. 5, pp. 445-463, 2002.
- [11] Shi.H. Best-first decision tree learning. M.Sc. Thesis, University of Waikato, Hamilton, NZ.
- [12] Simone Frintrop and Erich Rome, “Simulating Visual Attention for Object Recognition” Fraunhofer Institute for Autonomous Intelligent Systems (AIS) Schloss Birlinghoven, D-53754 Sankt Augustin, Germany
- [13] Simone Frintrop, “VOCUS: a visual attention system for object detection and goal-directed search,” Ph.D. dissertation, Rheinische Friedrich- Wilhelms-University, Bonn, Germany, July 2005, published 2006 in Lecture Notes in Artificial Intelligence (LNAI), Vol. 3899, Springer Verlag Berlin/Heidelberg.
- [14] Simone Frintrop, Patric Jensfelt, and Henrik Christensen. Pay attention when selecting features. In Proc. Int’l Conf. on Pattern Recognition (ICPR 2006), 2006.
- [15] Simone Frintrop, Patric Jensfelt, and Henrik Christensen. Attentional robot localization and mapping. In ICVS Workshop on Computational Attention & Applications (WCAA), 2007.
- [16] Simone Frintrop, Maria Klodt, and Erich Rome. A real-time visual attention system using integral images. In Proc. of Int’l Conf. on Computer Vision Systems (ICVS), 2007.
- [17] Theeuwes.J, Top Down Search Strategies Cannot Override Attentional Capture. Psychonomic Bulletin and Review, 11:65-70, 2004