ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

IJARCCE

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 3, Issue 2, February 2014*

# TOP-DOWN MODELS OF HUMAN VISIUAL ATTENTION USING DYNAMIC BAYESIAN NETWORK

**Chanchal Satheesh[1], A.Diana Andrushia[2], R.Thangarajan[3]**

ECE Department, Karunya University, Coimbatore, India[1,2]

CSE Department, Kongu Engineering College, Erode, India[3]

**Abstract:** Visual attention is one of the built in mechanisms in system that quickly selects regions in a visual scene, which are most likely to contain items of interest. Many computational models are developed to predict the behaviour of human visual attention. This paper describes the computational model involving top-down process for most attended region using Dynamic Bayesian Network. By using Dynamic Bayesian Network, Top-down knowledge is predicted. Top down saliency map is found using three different methods namely weighted modulation, weighted combination and joint learning. The performance of each method is compared using Normalized scan-path saliency (NSS).

**Keywords:** Attention, Dynamic Bayesian Network, Top-Down Knowledge

## I. INTRODUCTION

We live in the visual world. Every day human eyes manipulate $10^5$-$10^6$ bits of data for every second. All the information are not going directly to the visual brain. The filters are used to reduce the information. How human brains select the most attended region and how the information's are filtered? Many researches are going in this direction.

To filter the information from visual scene deals with the domains such as Psychophysics, Computational sciences, Cognitive Neurosciences, Computational sciences, Cognitive Psychology. Most of the information are commonly referred in the Computational Neuroscience, defined by Trappenberg [1] as: "the theoretical study of the brain used to discover the principles and mechanisms that guide the development, organization, information processing and mental abilities of the nervous system".
Visual Attention defined by Corbetta [2] as: " the mental ability to select stimuli, responses, memories, or thoughts that are behaviorally relevant among many others that are behaviorally irrelevant" The visual attention is applied in the field [3] of computer vision, pattern recognition such as object detection, object recognition, action recognition, segmentation, background subtraction, video summarization, compression, scene understanding, computer-human interaction, robotics, driver assistance.

## II. RELATED WORKS

Many computational models are reviewed in [3] [4] which are related to bottom-up, top-down and combined of two models. The Bottom-up computational models are integrated with the low level features which are discussed in [5][8].The Top-down models are integrated with high level features which are discussed in the paper [6], [7].This paper is mainly focussed on top-down model with the probabilistic prediction.

## III. TOP DOWN MODEL

Models incorporating additional cues such as prior knowledge about the search targets, human intention and cognitive states generally fall under Top down models.

Mainly three types of top down models and they are weighted modulation, weighted combination and join learning models. Each models deals with finding the saliency map using Dynamic Bayesian Network.

(i) weighted modulation

In weighted modulation, the input image features are extracted, the features can be color (red-green), (blue-yellow), intensity and orientation (0, 45, 90, 135). Each extracted features will be having their own weight. Using Dynamic Bayesian network, a top down knowledge is given such a way maximum probability can be taken as the weight. Then the maximum weight gets added to Top down saliency map as shown in fig 1. With r, g, b as the input colors the intensity is obtained using

$$I = r + g + b/3 \quad (1)$$

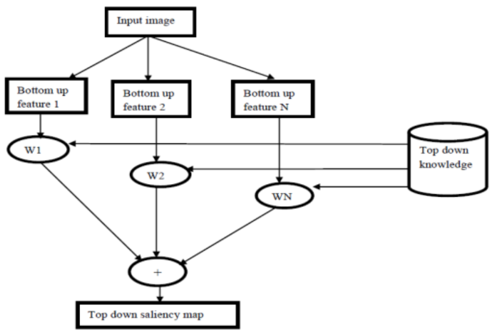



Fig 1.weighted modulation

Dynamic Bayesian network is a probabilistic network like a kalman filter used for prediction purpose. Here the prediction is done so as to get maximum output, for this conditional probability table(CPT) is constructed.CPT table can be constructed for the features namely color, orientation and intensity. While dealing with color CPT then it can be given as CPT (1, 1) = [1 0], CPT (1, 1) represents the color contain red-green and blue-yellow, the value on the right hand side represents the probability for it to become salient and not salient. From the CPT table constructed maximum probability is given as weight to

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

individual feature channel resulting in Top down saliency map. Since the weight is modified according to the designed probability, it names as weighted modulation

*(ii) Weighted Combination*

In weighted combination there are three stage .In first stage which deals with Bottom-up saliency map gaining using itti's approach. Initially the features are extracted from the input image. Features extracted are same as that has been selected for weighted modulation. According to [8], Gaussian pyramid is constructed from the extracted features, which then undergo centre-surround difference to form individual feature map. Feature map with same features are integrated to form the conspicuity map, all these conspicuity map contribute to get the saliency map.

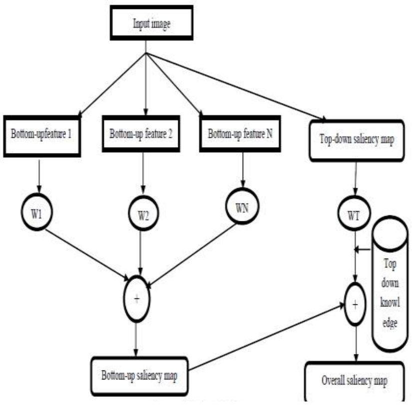



Fig 2.weighted combination

And in second stage the top down saliency map is computed over the weighted modulation was obtain. In last stage, bottom-up feature map from itti's approach and top down saliency map from weighted modulation these two maps are combined to combination third stage ,also dynamic Bayesian network is carry out predictive process. Hence last overall saliency map was obtained.

*(iii) Joint Learning*

Fig 3 shows the joint learning of visual stimuli. This approach is based on guided search theory.

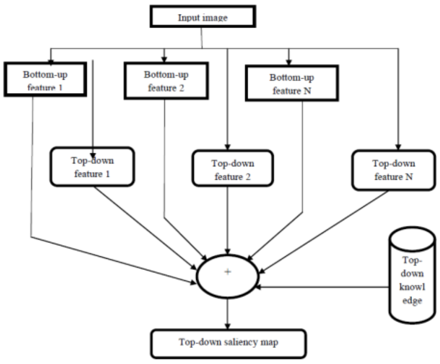



Fig 3.Joint learning

Guided search employs a activation map[3].This activation map is obtained from bottom-up activation based on existing features and top down activation map using characteristics of target stimuli[9][10].The obtained top down features and bottom up features are integrated to get the top down saliency map.

## IV. EVALUATIONS METRIC

To quantify how a model predicts well in modeling eye position, a normalized operator is taken. Let $X_{ij}$=Saliency Map and $Y_{ij}$=Eye Map (binary), Mean of $X_{ij}$=Z, Standard Deviation of $X_{ij}$= A.

$X_{ij}$ (Normalized) = $(X_{ij}-Z)/(A)$ (2)

$[C, D] = \sum^{N}_{i=1}\sum^{N}_{j=1} (Y_{ij} > 0)$ (3)

Normalized Vector = $\sum^{N}_{C=1}\sum^{N}_{D=1} X_{(C, D)}$ (4)

N = 1 indicates eye positions fall in a region whose predicted density is above average. NSS less than zero indicate that the model performs is not better than picking a random position on the map.

## V. RESULTS

Three types of top-down models are computed, six images are given to each of the models to obtain the saliency map.

Fig 4.Input image

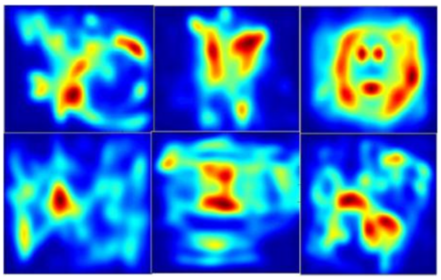
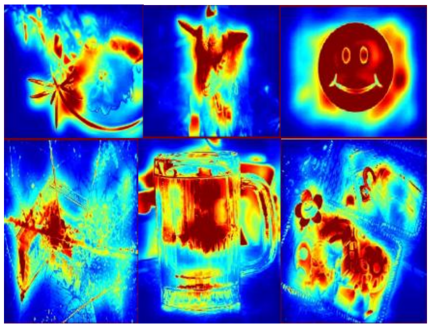


Fig 5.weighted modulation output



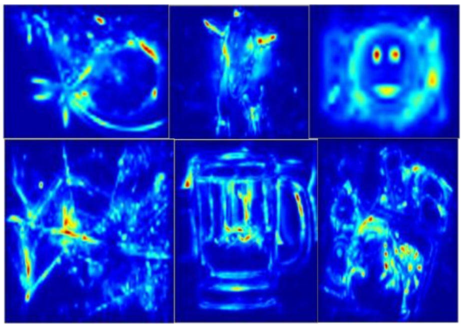


Fig 7.Joint learning output

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

## VI. PERFORMANCES METRIC

Table 1.Normalized vector score

| | M1 | M2 | M3 |
|---|---|---|---|
| Image 1 | 1.1918 | 2.3677 | 1.1312 |
| Image 2 | 1.4382 | 2.2728 | 1.213 |
| Image 3 | .8526 | 1.3471 | .7321 |
| Image 4 | 1.4538 | 2.6613 | 1.112 |
| Image 5 | 1.1792 | 1.6955 | .8756 |
| Image 6 | 1.2235 | 2.0858 | .9176 |

M1-weighted modulation
M2-weighted combination
M3-Joint learning

## VII. CONCLUSION

The work describes the computational model for most attended region using Machine learning Techniques. By using Dynamic Bayesian Network, Top down knowledge is predicted. Top down saliency map for different classes namely weighted modulation, weighted combination and joint learning was done. NSS score for each class is also performed. Weighted combination shows highest score when compared to Weight modulation and Joint learning. Weighted combination is used to predict the most attended region

## REFERENCES

[1] Trappenberg, TP.: Fundamentals of computational neuroscience, 2nd edition Oxford University Press,Oxford(2010)
[2] Corbetta, M.: Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? National Academy of Science U S A 95(3):831–838(1998)
[3] Akisato kimura,Yonetani,R.: Computational Models of Human Visual Attention and Their Implementations: A Survey IEICE Transaction on Information & System Volume E96-D,N0 3 (2013)
[4] Sílvio Filipe, Luis Alexandre.:From the human visual system to the computational models of visual attention: a survey:Journal of Artifical Intelligence Review (2013)
[5] Klein A, Frintrop.:S Center-surround divergence of feature statistics for salient object detection. IEEE International conference on computer vision, Barcelona, Spain (2011)
[6] Aude Oliva.: Top-down Control of Visual Attention in Object Detection Vision Research vol 12(2010)
[7] Borji A, Ahmadabadi MN, Araabi BN, Hamidi M.: Online learning of task-driven object-based visual attention control. Image and Vision Computing 28(7):1130–1145(2010)
[8] L.Itti,C.koch,and E.Neibur , "A model of saliency-based,visual attention for rapid scene analysis,"IEEE Transaction on pattern analysis and machine intelligence,vol-20,no.11,pp.1254-1259(1998)
[9] A.Borji,D.Sihite,and L.Itti, "Probabilistic learning of task specific visual attention," Proceedings of IEEE conference on computer vision and pattern Recognition(CVPR),pp.1-8(2012)
[10] A.Borji, "Boosting bottom-up and top-down visual features for saliency detection," Proceedings of IEEE conference on computer vision and pattern Recognition(CVPR),pp.1-8(2012)