

# A Survey on Various Saliency Detection Methods

Elsa Sebastian  
M.Tech student,  
Viswajyothi College of  
Engineering and Technology

Neenu Daniel  
Assistant professor,  
Viswajyothi College of  
Engineering and Technology

## ABSTRACT

Visual saliency is an important characteristic of Human Visual System (HVS) that select the visually significant information from scenes. The salient objects stand out relative to their neighbourhood regions. Detecting and segmenting salient objects, also known as salient object detection is used to extract the most interesting object/objects in a scene and has resulted in many applications. There are many different methods to detect saliency known as visual attention models or saliency detection methods. In past few years many saliency detection methods have been proposed. One of the main objectives of the work is to perform a detail study in the field of Saliency detection by keeping focus on the different bottom-up computational models and the methods used to predict saliency. The work aims to analyze various solutions that aid the task of HVSs properties. This paper presents various saliency detection methods.

## Keywords

Saliency detection, Visual attention, Human Visual System

## 1. INTRODUCTION

Visual attention is an important characteristic in the human visual system (HVS) that try to predict the most relevant and important areas of videos or images viewed by the human eye. There are two main mechanisms for directing visual attention: top-down and bottom-up. Top-down is voluntary and task-driven, i.e. it needs an observer to identify salient regions. In the bottom-up process, the saliency is detected through computational models. Visual saliency has been widely used in various applications such as image retrieval, object recognition, image/video compression, image cropping/retargeting, human gaze and actions, automatic target detection, boundary detection, sensation enhancement, image analysis and computer vision.

In this paper, survey of various saliency detection methods is studied. The discussed methods include a model of saliency-based visual attention for rapid scene analysis, low-Level features based on Wavelet Transform, multiresolution spatiotemporal saliency detection, video saliency incorporating spatiotemporal cues and uncertainty weighting, local gradient flow optimization and global refinement, spatiotemporal saliency detection, video saliency detection model in compressed domain, hybrid of local feature-based saliency and global feature-based saliency, region-based saliency detection, superpixel based saliency detection, spatiotemporal saliency for HDR images and video.

## 2. LITERATURE SURVEY

In [1] Itti, Koch and Niebur proposed a visual attention model in which image features such as color, intensity, and orientation for LDR images are combined to form a single saliency map. It makes use of Gaussian based approach in which a Gaussian pyramid is formed in each channel by subsampling the input image. Each feature is computed by a

set of linear centre-surround operations. Centre-surround is implemented in the model as the difference between fine and coarse scales. The across-scale subtraction is obtained by interpolation to the centre scale and point-by-point subtraction. Afterwards, all feature maps are combined into a “conspicuitymap” in each channel. Conspicuity maps from all channels one master saliency map, which topographically represents the local saliency.

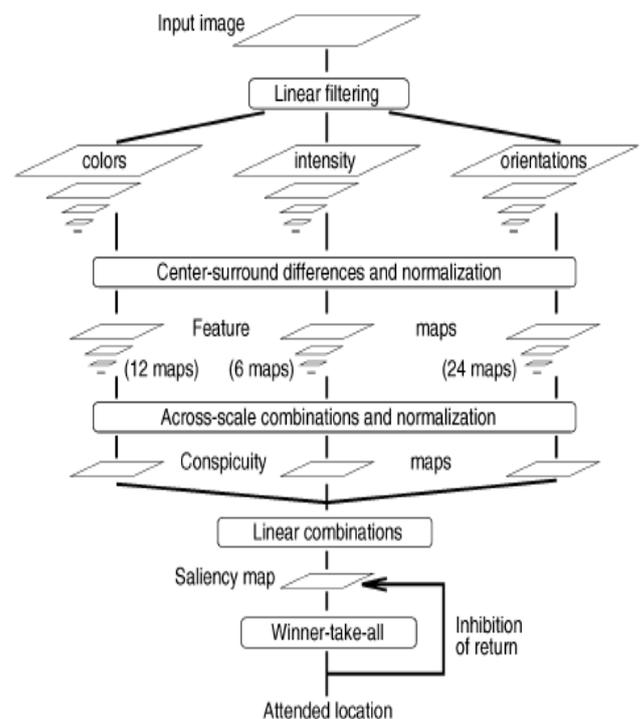
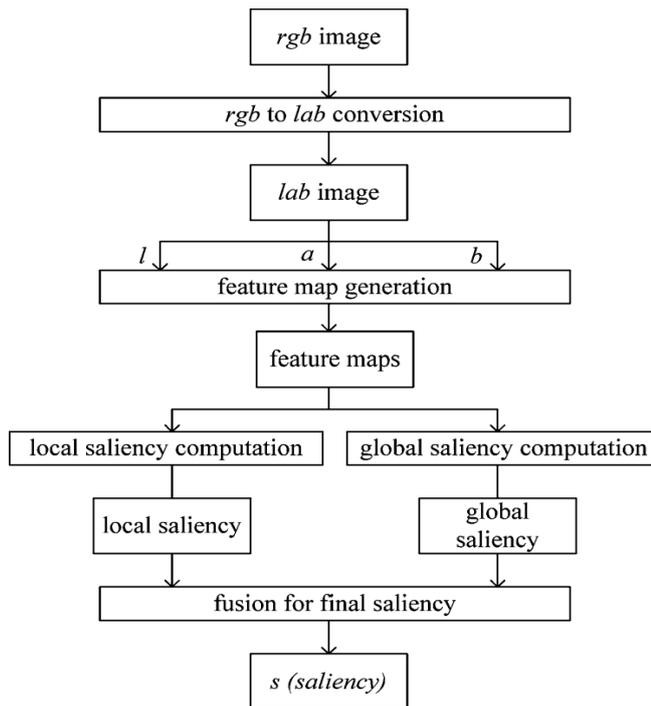


Figure 1: Framework of visual attention for Rapid Scene Analysis [1]

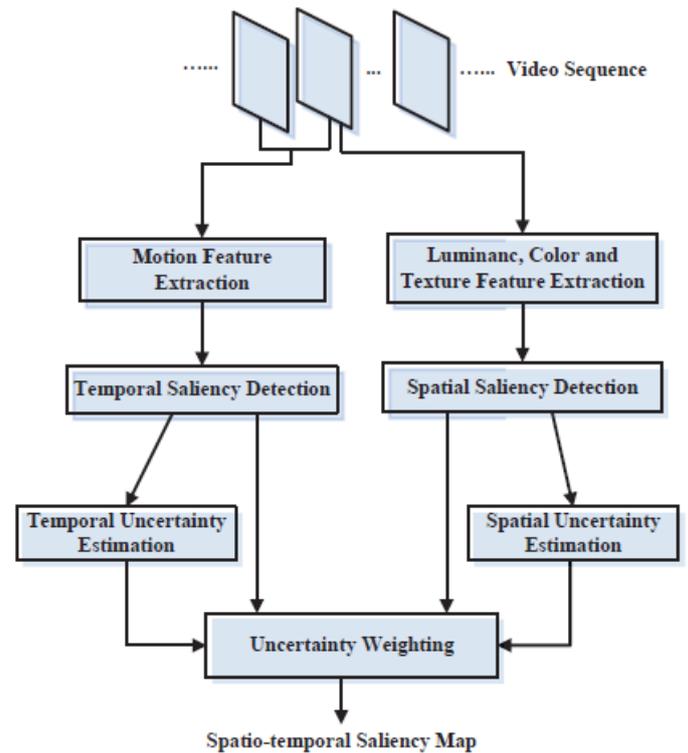
In [2] Chenlei Guo and Liming Zhang presents a saliency detection model based on high-pass coefficients of the wavelet decomposition. The idea is to create the feature maps by IWT on the multi-level decomposition. First, RGB image is converted into LAB color space. Then, apply wavelet transform decomposition on the noise removed version to find scaling coefficients. In addition to it, it focuses on creating the feature maps by IWT on the multi-level decomposition. Two saliency maps are created: local and global saliency maps. Finally, the local and global maps are combined to yield the final saliency map. The final saliency map represents both the local contrast of each location on the scene and the global distribution of the features as an amplifier for local saliency values.



**Figure 2: Framework of saliency detection model based on wavelet transform [2]**

In [3] Nevrez mamolu, Weisi Lin, and Yuming Fang suggested a saliency detection model in which image is represented as quaternionic and developed a multiresolution spatiotemporal saliency detection model called phase spectrum of quaternion Fourier transform (PQFT) to compute the spatiotemporal saliency map from the images quaternion representation. First, each pixel of the image is represented by a quaternion that consists of color, intensity and motion feature. Then, the phase spectrum of QFT is used to calculate the spatiotemporal saliency map, which considers not only salient spatial features like color, orientation and others in a single frame but also temporal feature between frames like motion.

In [4] Yuming Fang, Zhou Wang, and Weisi Lin proposed a method to detect visual saliency from video signals by combing both spatial and temporal information and statistical uncertainty measures. First, separate spatial and temporal saliency maps are generated. Second, the spatial and temporal saliency maps are merged into one using a spatiotemporally adaptive entropy-based uncertainty weighting approach. Low-level spatial and motion features are first extracted from the input video sequence, where the spatial features (including luminance, color and texture) and the motion feature are used to calculate the spatial and temporal saliency maps, respectively. The spatial and temporal uncertainty maps are then calculated to assess the confidence of the corresponding saliency maps. Finally, the spatial and temporal saliency maps are fused using an uncertainty weighting approach, resulting in the final spatiotemporal saliency map.



**Figure 3: Framework of spatiotemporal cues and uncertainty weighting [4]**

In [5] Wenguan Wang, Jianbing Shen, and Ling Shao, proposed a saliency detection method that estimates the salient regions in videos based on the gradient flow field and energy optimization. Firstly, the method utilises a gradient flow field that incorporates two distinctive features: 1) intra-frame boundary information and 2) inter-frame motion information together for indicating the salient regions. Further, local as well as global contrast saliency measures using the foreground and background information is estimated from the gradient flow field. Finally, a spatiotemporal saliency energy function is presented to encourage the spatiotemporal consistency of the output video salience maps.

In [6] Wonjun Kim, Chanho Jung, and Changick Kim suggested a novel unified method for detecting salient regions in both images and videos based on a discriminant center-surround hypothesis that the salient region stands out from its surroundings. First of all, a set of visual features composed of edge and color orientations and temporal gradients are computed. Then, compute the spatiotemporal saliency at each scale in which the spatial saliency is computed as the distances between ordinal signatures of edge and color orientations obtained from the center and the surrounding regions and the temporal saliency, by simply computing the sum of absolute difference (SAD) between temporal gradients of the center and the surrounding regions. Finally, resize saliency map to the same size of input image to obtain the final saliency.

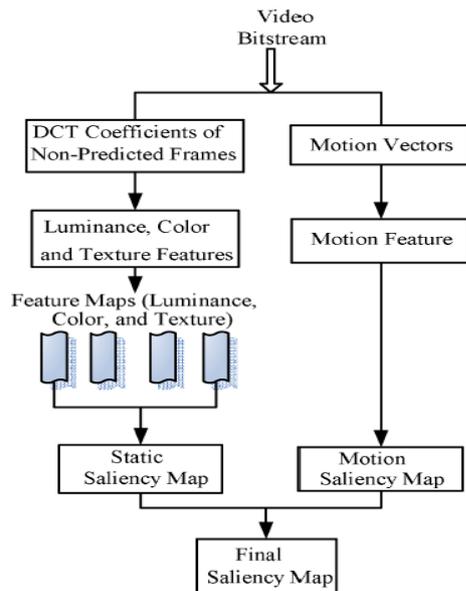


Figure 4: Video Saliency Detection Model [6]

In [7] Yuming Fang, Weisi Lin, Zhenzhong Chen, Chia-Ming Tsai, and Chia-Wen Lin proposed a novel video saliency detection model based on feature contrast in compressed domain. Firstly, three features including luminance, color, and texture are extracted from the discrete cosine transform coefficients for unpredicted frames (I frames), and the motion feature is extracted from the motion vectors in video bit stream for predicted frames (P and B frames). Then, the static saliency map is obtained based on the features of luminance, color, and texture for unpredicted frames, while the motion saliency map is calculated on the basis of motion feature for predicted frames. Finally, the static saliency map and the motion saliency map are combined to get the final saliency map for each video frame.

In [8] Qinmu Peng, Yiu-ming Cheung, Xinge You, and Yuan Yan Tang, presents a visual saliency detection approach, which is a hybrid of local feature-based saliency and global feature-based saliency. First, for a given input image, use an automatic selection of smoothing parameter scheme to make the image region more homogeneous. Then, partition the smoothed image into a set of regions and compute the local saliency by measuring the color and texture dissimilarity in

the smoothed regions and the original regions, respectively. Furthermore, compute the global saliency by utilizing the global color distribution model embedded with color coherence, together with the multiple edge saliency. Finally, combine the local and global saliencies, and utilize the composition information to obtain the final saliency.

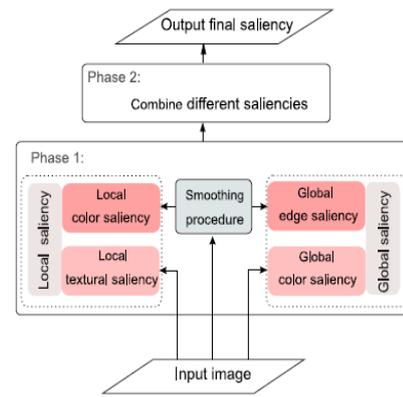


Figure 5: Hybrid of local and global saliency detection [8]

In [9] Zhixiang Ren, Shenghua Gao, Liang-Tien Chia, and Ivor Wai-Hung Tsang presents a region-based solution for saliency detection applicable to better encode the image features for solving object recognition task. First use the adaptive mean shift algorithm to extract superpixels from the input image. Then apply Gaussian mixture model (GMM) to cluster superpixels based on their color similarity. Finally, calculate the saliency value for each cluster using spatial compactness metric together with modified PageRank propagation.

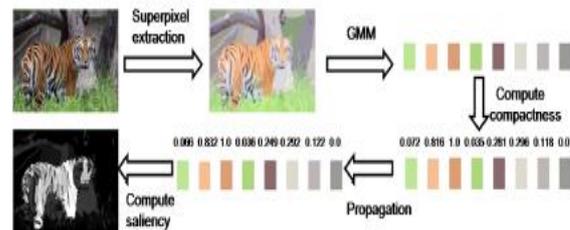
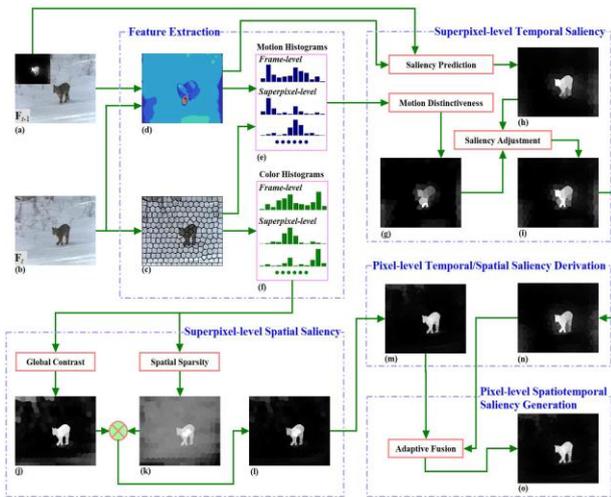


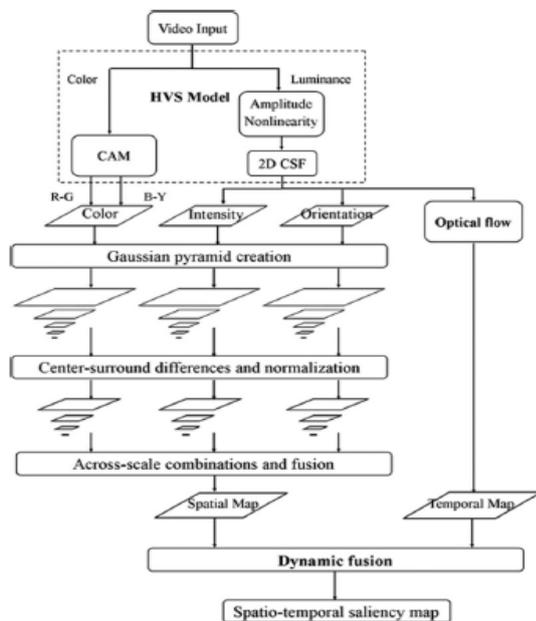
Figure 6: Region-Based Saliency Detection [9]

In [10] Zhi Liu, Xiang Zhang, Shuhua Luo, and Olivier Le Meur proposed a superpixel-based spatiotemporal saliency model for saliency detection in videos. First, based on the superpixel representation of video frames, motion histograms and color histograms are extracted at the superpixel level as local features and frame level as global features. Then, superpixel level temporal saliency is measured by integrating motion distinctiveness of superpixels with a scheme of temporal saliency prediction and adjustment, and superpixel-level spatial saliency is measured by evaluating global contrast and spatial sparsity of superpixels. Finally, a pixel-level saliency derivation method is used to generate pixel-level temporal and spatial saliency maps, and an adaptive fusion method is exploited to integrate them into the spatiotemporal saliency map.



**Figure 7: Superpixel-Based Spatiotemporal Saliency Detection [10]**

In [11] Yuanyuan Dong, Mahsa T. Pourazad, and Panos Nasiopoulos proposed a saliency detection method that detects the saliency of HDR images and HDR video frames. The spatial and temporal cues are taken into account, leading to two saliency maps: the spatial saliency map and the temporal saliency map. First, to obtain the spatial saliency map, use the HVS model to decompose feature channels from an HDR input and then follow the procedure of the classical bottom-up method in [1]. Then to compute the temporal saliency map, an optical flow based method is used to estimate motion. Finally, a dynamic fusion method is proposed to combine both the spatial and temporal saliency maps.



**Figure 8: Superpixel-Based Spatiotemporal Saliency Detection [11]**

### 3. CONCLUSION

Saliency Detection is found to be a broad area wherein many studies and researches have been carried out. This paper presents a brief review of different bottom-up saliency detection methods of images and videos done over the past

decade. The survey outlined how saliency detection methods can be categorized on the basis of feature sets and steps used to detect saliency. From this survey, we were able to determine the advantages of each method. It could be seen that majority of the works focused on detecting saliency of LDR images and videos and only few works on HDR images and videos.

### 4. REFERENCES

- [1] Laurent Itti, Christof Koch, and Ernst Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 20, No. 11, Nov 1998.
- [2] Chenlei Guo and Liming Zhang, "A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression," *IEEE Transactions On Image Processing*, Vol. 19, No. 1, Jan2010.
- [3] Nevrez mamolu, Weisi Lin, and Yuming Fang, "A Saliency Detection Model Using Low-Level Features Based on Wavelet Transform," *IEEE Transactions On Multimedia*, Vol. 15, No. 1, Jan 2013.
- [4] Yuming Fang, Zhou Wang, and Weisi Lin, "Saliency Incorporating Spatiotemporal Cues And Uncertainty Weighting," *IEEE Transactions On Image Processing*, Vol. 23, No. 9, sept 2014.
- [5] Wenguan Wang, Jianbing Shen, and Ling Shao, "Consistent Video Saliency Using Local Gradient Flow Optimization and Global Refinement," *IEEE Transactions On Image Processing*, Vol. 24, No. 11, Nov 2015.
- [6] Wonjun Kim, Chanho Jung, and Changick Kim, "Spatiotemporal Saliency Detection and Its Applications in Static and Dynamic Scenes," *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 21, April 2011.
- [7] Yuming Fang, Weisi Lin, Zhenzhong Chen, Chia-Ming Tsai, and Chia-Wen Lin, "A Video Saliency Detection Model in Compressed Domain," *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 24, No. 1, January 2014.
- [8] Qinmu Peng, Yiu-ming Cheung, Xinge You, and Yuan Yan Tang, "Hybrid of Local and Global Saliencies for Detecting Image Salient Region and Appearance," *IEEE Transactions On Systems, Man, And Cybernetics: Systems*, April 2016.
- [9] Zhixiang Ren, Shenghua Gao, Liang-Tien Chia, and Ivor Wai-Hung Tsang, "Region-Based Saliency Detection and Its Application in Object Recognition," *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 24, No. 5, May 2014.
- [10] Zhi Liu, Xiang Zhang, Shuhua Luo, and Olivier Le Meur, "Superpixel Based Spatiotemporal Saliency Detection," *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 24, No. 9, September 2014.
- [11] Yuanyuan Dong, Mahsa T. Pourazad, and Panos Nasiopoulos, "Human Visual System-based Saliency Detection For HDR Content," *IEEE Transactions On Multimedia*, Vol. 18, No. 4, April 2016.