NOVATEUR PUBLICATIONS
International Journal Of Research Publications In Engineering And Technology [IJRPET]
ISSN: 2454-7875
VOLUME 2, ISSUE 8, Aug. -2016

DETECTING SALIENT OBJECTS OF NATURAL SCENE IN A VIDEO'S USING SPATIO-TEMPORAL SALIENCY & COLOUR MAP

SUHAS B. KHADAKE
Department of Electronics Engineering,
Bharatratna Indira Gandhi College of Engineering,
Kegaon, Solapur, India

ABSTRACT:

Detection of spatio temporal saliency is the main aim of this paper. It uses system where it utilizes the human eye perceptive action. This uses the signal coming from human eye, for detection of saliency, for that a large database are used (about 15000 frames). For Saliency detection two pathways are decided. This uses the data on the basis of same thing on images and in accordance similarity will be calculated. According to it visual summary data are used. It gives much amount of data, which gives a lot of research in computer vision and results are used for calculation of saliency.

We have take the view of all method used calculation of saliency. So by using the data of different method we develop a Novel method which will give best results of detection of saliency. Here technique used is calculation of saliency by using spatio temporal saliency and colour map of video or images. In this method number of experiments in progress for saliency. This method has fast and robust result.

Index Terms: HVS, pixel level, phase spectrum, resolution, Gaussian, Particle filter, Spatio-Temporal saliency.

INTRODUCTION:

The human visual system (HVS) does a remarkable job at separating foreground from background clutter even when the objects in a scene have never been encountered before. This capability is often attributed to its ability to quickly locate 'attentive' regions in the scene and subsequently understand the rest of the scene. In computer vision, the phenomenon of identifying attentive regions in image and video is known as salient region detection, where the regions may or may not represent a single object. The two main approaches to salient region detection are top-down and bottom-up. The former approach is an object or task-oriented process and incorporates a fair amount of training, while the bottom-up approach is driven by contrast between features leading to a measure of conspicuity of a region.

The distinguishing aspect between image and video saliency is the temporal motion information in the latter that introduces the notion of motion saliency in addition to spatial saliency of images. Bottom-up saliency for videos helps in foreground region detection by assigning a saliency measure to a pixel to be spatially and temporally salient. The most commonly used technique to easily segment the foreground from the background is by applying a threshold to the saliency map [2]. However, if the saliency map is not accurate this might result in spurious regions being identified as foreground.

In this paper, we propose a method for detecting the most salient object in a video based on a computationally simple approach for the generation of spatio-temporal saliency maps. The most salient object is identified by localizing a set of pixels using particle filters whose weights are computed from the saliency maps and colour features. The location of the particles is influenced by the saliency maps as well as by the colour information in the frame. The spatio-temporal saliency map allows the particle filter to converge faster on the most salient foreground object by refining the weights of the particles while the colour map provides the robustness in matching the reference and target distributions. The spatio-temporal saliency map is generated at the pixel-level in the original resolution as opposed to other methods [2–4] that operate on a reduced resolution or on patches to reduce the computation time. In the latter case, the accuracy of the saliency maps is compromised when resized to the original resolution while our pixel-level computation to generate saliency maps has been shown to generate accurate saliency maps at an average processing time of eight frames per second on a frame resolution of 352 × 288.

The first step of a salient motion detection algorithm is the foreground-background segmentation and subtraction step. To segment out the foreground from background there are lot of techniques available. The pixels intensity values are added with the colour illumination into the algorithm, so it is easy to improve the image subtraction techniques. The background subtraction model in [4] models each pixel as a mixture of Gaussians. Mixture of several Gaussians is preferred to model the background. The model is updated regularly and evaluated to determine which of the Gaussians are sampled from the background process. When changes in the background are too fast then the variance of the Gaussians becomes too large and non parametric approaches are more suited. The

concept of observably from the output pixels is provides a good clue to the salient motion detection in natural videos.

LITERATURE REVIEW:

Salient object detection in videos is challenging because of the competing motion in the background, resulting from camera tracking an object of interest, or motion of objects in the foreground. The authors present a fast method to detect salient video objects using particle filters, which are guided by spatio-temporal saliency maps and colour feature with the ability to quickly recover from false detections. The proposed method for generating spatial and motion saliency maps is based on comparing local features with dominant features present in the frame[1]. A region is marked salient if there is a large difference between local and dominant features. For spatial saliency, hue and saturation features are used, while for motion saliency, optical flow vectors are used as features. Experimental results on standard datasets for video segmentation and for saliency detection show superior performance over state-of-the-art methods.

A new method for automatic Salient object segmentation is an important research area in the field of object recognition, image retrieval, image editing, scene reconstruction, and 2D/3D conversion. In this work, salient object segmentation is performed using saliency map and color segmentation. Edge, color and intensity feature are extracted from mean shift segmentation (MSS) image, and saliency map is created using these features [2]. First average saliency per segment image is calculated using the color information from MSS image and generated saliency map. Then, second average saliency per segment image is calculated by applying same procedure for the first image to the thresholding, labeling, and hole-filling applied image. Above are applied to the mean image of the generated two images to get the final salient object segmentation. The effectiveness of proposed method is proved by showing 80%, 89% and 80% of precision, recall and F-measure values from the generated salient object segmentation image and ground truth image.

A dynamical neural network then selects attended locations in order of decreasing saliency. The system breaks down the complex problem of scene understanding by rapidly selecting, in a computationally efficient manner, conspicuous locations to be analyzed in detail[3]. We propose a principled approach to summarization of visual data (images or video) based on optimization of a well-defined similarity measure. The problem we consider is re-targeting (or summarization) of image/video data into smaller sizes. A good "visual summary" should satisfy two properties: (1) it should contain as much as possible visual information from the input data; (2) it should introduce as few as possible new visual artifacts that were not in the input data (i.e., preserve visual coherence). We propose a bi-directional similarity measure which quantitatively captures these two requirements: Two signals S and T are considered visually similar if all patches of S (at multiple scales) are contained in T, and vice versa. The problem of summarization/re-targeting is posed as an optimization problem of this bi-directional similarity measure. We show summarization results for image and video data. We further show that the same approach can be used to address a variety of other problems, including automatic cropping, completion

However, computational modeling of this basic intelligent behavior still remains a challenge[3]. This paper presents a simple method for visual saliency detection. Our model is independent of features, categories, or other forms of prior knowledge of the objects[4]. By analyzing the log-spectrum of an input image, we extract the spectral residual of an image in spectral domain, and propose a fast method to construct the corresponding saliency map in spatial domain. We test this model on both natural pictures and artificial images such as psychological patterns. The results indicate fast and robust saliency detection of our method [5]. Salient image regions permit non-uniform allocation of computational resources. The selection of a commensurate set of salient regions is often a step taken in the initial stages of many computer vision algorithms, thereby facilitating object recognition, visual search and image matching. In this study, the authors survey the role and advancement of saliency algorithms over the past decade. The authors first offer a concise introduction to saliency. Next, the authors present a summary of saliency literature cast into their respective categories then further differentiated by their domains, computational methods, features, context and use of scale. The authors then discuss the achievements and limitations of the current state of the art. This information is augmented by an outline of the datasets and performance measures utilized as well as the computational techniques pervasive in the literature [6]. A Primate demonstrates unparalleled ability at rapidly orienting towards important events in complex dynamic environments. During rapid guidance of attention and gaze towards potential objects of interest or threats, often there is no time for detailed visual analysis. Thus, heuristic computations are necessary to locate the most interesting events in quasi real-time. We present a new theory of sensory surprise, which provides a principled and computable shortcut to important information. We develop a model that computes instantaneous low-level surprise at every location in video streams. The algorithm significantly correlates with eye movements of two humans watching complex video clips, including television programs (17,936 frames, 2,152 saccadic gaze shifts). The system allows

more sophisticated and time-consuming image analysis to be efficiently focused onto the most surprising subsets of the incoming data [7]. A spatiotemporal saliency algorithm based on a center-surround framework is proposed. The algorithm is inspired by biological mechanisms of motion-based perceptual grouping and extends a discriminant formulation of center-surround saliency previously proposed for static imagery. Under this formulation, the saliency of a location is equated to the power of a predefined set of features to discriminate between the visual stimuli in a center and a surround window, centered at that location. The features are spatiotemporal video patches and are modeled as dynamic textures, to achieve a principled joint characterization of the spatial and temporal components of saliency. The combination of discriminant center-surround saliency with the modeling power of dynamic textures yields a robust, versatile, and fully unsupervised spatiotemporal saliency algorithm, applicable to scenes with highly dynamic backgrounds and moving cameras. The related problem of background subtraction is treated as the complement of saliency detection, by classifying nonsalient (with respect to appearance and motion dynamics) points in the visual field as background. The algorithm is tested for background subtraction on challenging sequences and shown to substantially outperform various state-of-the-art techniques. Quantitatively, its average error rate is almost half that of the closest competitor [8]. Detection of the motion of foreground objects on the backdrop of constantly changing and complex visuals has always been challenging. The motion of foreground objects, which is termed as salient motion, is marked by its predictability compared to the more complex unpredictable motion of the backgrounds like fluttering of leaves, ripples in water, smoke filled environments etc. We introduce a novel approach to detect this salient motion based on the control theory concept of 'observability' from the outputs, when the video sequence is represented as a linear dynamical system. The resulting algorithm is tested on a set of challenging sequences and compared to the state – of - the-art methods to showcase its superior performance on grounds of its computational efficiency and detection capability of the salient motion [9]. Salient motion detection is a challenging task especially when the motion is obscured by dynamic background motion. Salient motion is characterized by its consistency while the non- salient background motion typically consists of dynamic motion such as fog, waves, fire etc. In this paper, we present a novel framework for identifying salient motion by modeling the video sequence as a linear dynamic system and using controllability of states to estimate salient motion. The proposed saliency detection algorithm is tested on a challenging benchmark video dataset and the performance is compared with other state-of-the-art algorithms.a Motion saliency is a key component for video saliency model, and attracts a lot of research interest. However, the existing methods will lose the foreground objects or parts of the foreground objects when they stop moving for a certain period of time, and the interior of objects will be marked un-salient unless the moving objects are sufficiently textured. In this paper, a novel saliency model based on motion history map is proposed. We generate a spatial saliency sub-map by pixel-wise self-information and global contrast, and use it to adaptively adjust the effect period of motion history map, which makes the foreground objects are still marked silent for a long time after they stop moving, while the background noise is soon marked un-silent. Otherwise, a fast region growing method based on the spatial saliency sub-map is applied to mark the interior of moving objects salient.

METHODOLOGY

We employ a particle filter for its ability to approximate the posterior distribution of a system based on a finite set of weighted samples.

1 System Block Diagram:-

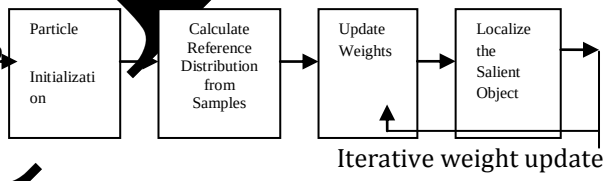

Fig. 1. Proposed System Block Diagram

This System consists of 4 blocks.

i) Particle Initialization ii) Reference distribution from sample iii) Weight Update iv) Localize the salient objects.

In above figure we apply the video input which we have calculates saliency. This is applied to the particle initialization block. Which is the first step of saliency calculation.nxt step calculate reference distribution from the samples. In accordance with weight are updated. Next step is localizing the salient objects are localized.

Particle filters are also highly robust to partial occlusion and computationally inexpensive. The weight of each particle is initialized using a uniform distribution. The first set of particles is initialized around the centre of the first frame of the video. Weights are subsequently calculated as the weighted sum of distance measures of the candidate regions to the reference distribution. The spatio-temporal saliency and the colour maps are used to calculate the weight of the samples, which allows subsequent iterations to move the particles closer to the most salient object. In the proposed framework, we detect only one object of interest. Fig. 1 illustrates a block diagram of how the particle filter framework is used to

detect the salient object. Colour versions of all the Figures used in this are available online [11]; a Video saliency mechanism is crucial in the human visual system and helpful to object detection and recognition. In this paper we propose a novel video saliency model that video saliency should be both consistently salient among consecutive frames and temporally novel due to motion or appearance changes. Based on the model, temporal coherence, in addition to spatial saliency, is fully considered by introducing temporal consistence and temporal difference into sparse feature selections. Features selected spatio-temporally are enhanced and fused together to generate the proposed video saliency maps. Comparisons with several state-of-th-art methods on two public video datasets further demonstrate the effectiveness of our method[12]. Multimedia applications such as image or video re- trivial, copy detection, and so forth can benefit from saliency detection, which is essentially a method to identify areas in images and videos that capture the attention of the human visual system. In this paper, we propose a new spatio-temporal saliency detection framework on the basis of regularized feature reconstruction. Specifically, for video saliency detection, both the temporal and spatial saliency detection is considered. For temporal saliency, we model the movement of the target patch as a reconstruction process using the patches in neighboring frames. A Laplacian smoothing term is introduced to model the coherent motion trajectories. With psychological findings that abrupt stimulus could cause a rapid and involuntary deployment of attention, our temporal model combines the reconstruction error, regularizer, and local trajectory contrast to measure the temporal saliency. Finally, the temporal saliency and spatial saliency are combined together to favor salient regions with high confidence for video saliency detection. We also apply the spatial saliency part of the spatio-temporal model to image saliency detection. Experimental results on a human fixation video dataset and an image saliency detection dataset show that our method achieves the best performance over several state-of-the-art approaches [13]. By the guidance of attention, human visual system is able to locate objects of interest in complex scene. We propose a new visual saliency detection model for both image and video. Inspired by biological vision, saliency is defined locally. Lossy compression is adopted, where the saliency of a location is measured by the Incremental Coding Length (ICL). The ICL is computed by presenting the center patch as the sparsest linear representation of its surroundings. The final saliency map is generated by accumulating the coding length. The model is tested on both images and videos. The proposed model is inspired by the first steps of the human visual system, from the retina cells to the complex cells of the primary visual cortex. The visual information goes through the retina preprocessing to the cortical-like filter decomposition. The retina extracts two signals from each frame that correspond to the two main outputs of the retina [12]. Each signal is then decomposed into elementary features by a bank of cortical-like filters [13]. These filters are used to extract both static and dynamic information, according to their frequency selectivity, providing two saliency maps: a static and a dynamic one. Both saliency maps are combined to obtain a master spatio-temporal saliency map per video frame.

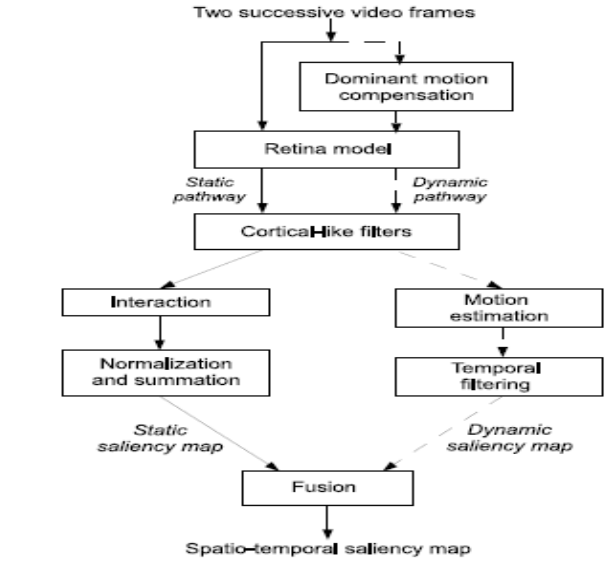

Fig.2.Proposed Spatio-Temporal Scheme

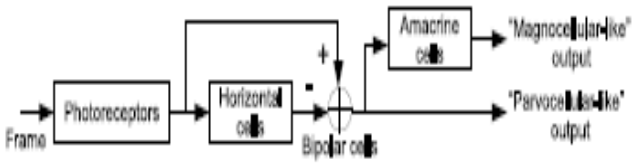

Fig.3. The retina "parvocellular-like" output

(Fig.3). This map predicts the gaze direction to particular areas of the frame.

DETAILED ANALYSIS OF THE TWO PATHWAYS:

The proposed model is a good predictor of the eye movements of subjects when looking freely at clip snippets. However, this model is a better predictor for the first frames of a snippet. It can be interesting to inquire now what is more attractive in the static saliency map and what is more attractive in the dynamic saliency map.

As we said before, the static and the dynamic saliency maps do not have the same appearance. On one side, the static saliency map exhibits a large number of salient areas, corresponding to textured areas, which are

spread over the whole image. On the other side, the dynamic saliency map can exhibit only small and compact areas corresponding to moving objects. Concerning the question "what is more salient in the static and the dynamic pathways?" we can suppose:

• For the static map: a frame would be salient if its static saliency map has a high value and not if its static saliency is spread. The saliency of a frame would be correlated with the maximum of its corresponding static saliency map.

• For the dynamic map: a frame would be salient if its dynamic saliency map has small and compact areas. The saliency of the frame would be linked to the number and the size of the salient areas in its corresponding dynamic saliency map.

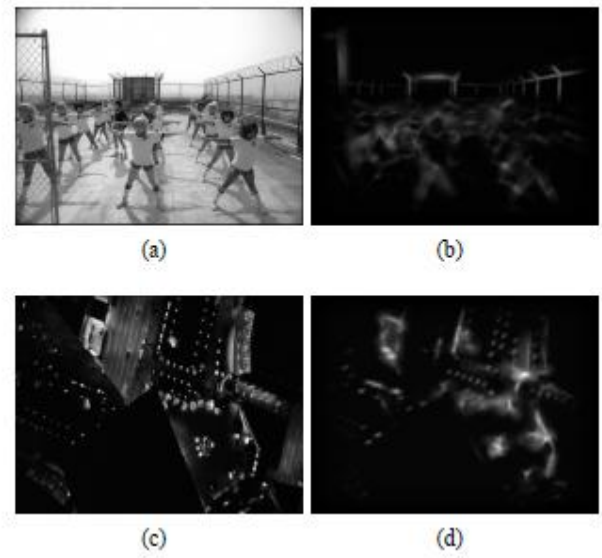

Figure 4.Examples of a natural scene (a) with a static saliency map (b) with low maximum (1.5) and scene (c) with dynamic saliency map (d) with high maximum (2.03).

Figure 5. Example of a natural scene (a) with its saliency Map Mskew−max (b)

RESULTS:

We analyzed the eye positions rather than the fixation points for two reasons. First, we had more data when choosing all the eye positions, and so we could extract one point per frame and per subject. Second, in most of the cases, eye positions and fixations are very close except during smooth pursuit. So by retaining all the eye positions we obtained data even during smooth pursuit. The aim is to compare the salient areas given by the model, with the fixated areas.

Various criteria have been proposed and used for this comparison: the Kullback-Leibler distance [25] or the Receiver Operator Curve (ROC) [26] (others examples can be found in [23], [8], [27]). In this paper, we chose to focus on the correlation coefficient and the Normalized Scanpath Saliency (NSS) [28]. This last criteria was especially designed to study eye movement data and so, the corresponding results can be easily interpreted. The correlation coefficient and the NSS lead to the same conclusion on the data analysis.

However, the correlation coefficient is very dependent on the we analyzed the eye positions rather than the fixation points for two reasons. First, we had more data when choosing all the eye positions, and so we could extract one point per frame and per subject. Second, in most of the cases, eye positions and fixations are very close except during smooth pursuit.

So by retaining all the eye positions we obtained data even during smooth pursuit. The aim is to compare the salient areas given by the model, with the fixated areas.

Various criteria have been proposed and used for this comparison: the Kullback-Leibler distance [20] or the Receiver Operator Curve (ROC) [21]. In this paper, we chose to focus on the correlation coefficient and the Normalized Scan path Saliency (NSS) [22]. This last criteria was especially designed to study eye movement data and so the corresponding results can be easily interpreted. The correlation coefficient and the NSS lead to the same conclusion on the data analysis. However, the correlation coefficient is very dependent on the standard deviation of the Gaussian applied to gaze positions to compute the human eye position density map. The NSS was therefore preferred. The NSS criteria is a Z-score, (also called standard score). This Z-score expresses the divergence of the experimental result from the model mean as a number\ of standard deviations of the model. The larger the value of Z, the less probable it is that the experimental result is due to chance.

CONCLUSION:

This method followed wherein the saliency value that corresponds to a human saccade position is computed as the maximum of the human saccade positions over a circle of diameter 128 pixels, centered at the human saccade position. The saliency values collected over the entire database are discredited into 10 bins which is subsequently normalized to obtain the probability distribution P. The distribution for Q is calculated in a similar manner from spatio-temporal saliency maps. As the

positions are sampled randomly, we repeat the experiment.

REFERENCES:

1) Karthik Muthuswamy, Deepu Rajan," *Particle filter framework for salient object detection in videos".* Centre for Multimedia and Network Technology, School of Computer Engineering, Nanyang Technological University, 50 Nanyang Avenue,N4-02C-92 639798, Singapore
2) Han, S.-H., Jung, G.-D., Lee, S.-Y., Hong, Y.-P., Lee, S.-H.: '*Automatic salient object segmentation using saliency map and color segmentation'*, J. Central South Univ., 2013, 20, (9), pp. 2407–2413.
3) Itti, L., Koch, C., Niebur, E.: '*A model of saliency-based visual attention for rapid scene analysis'*, IEEE Trans Pattern Anal. Mach. Intell., 1998, 20, (11), pp. 1254–1259.
4) Simakov, D., Caspi, Y., Shechtman, E., Irani, and M.: '*Summarizing visual data using bidirectional similarity'*. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008.
5) Hou, X., Zhang, L.: *'Saliency detection: a spectral residual approach'.* Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 2007
6) Duncan, K., Sarkar, S.: *'Saliency in images and video: a brief survey'*, IET Compute. Vis., 2012, 6, (6), pp. 514–523
7) Itti, L., Baldi, P.: '*A principled approach to detecting surprising events in video'*. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, pp. 631–637
8) Mahadevan, V., Vasconcelos, N.: '*Spatiotemporal saliency in dynamic scenes'*,IEEE Trans. Pattern Anal. Mach. Intell., 32, (1), pp. 171–177
9) Gopalakrishnan, V. ,Hu, Y., Rajan, and D.: *'Sustained observability for salient motion detection'.*Proc. 10th Asian Conf. on Computer Vision, Queenstown, New Zealand, 2010, pp. 732–743
10) Muthuswamy, K., Rajan,D.: *'Salient motion detection through state controllability'*. Proc. IEEE Int. Conf on Acoustics, Speech and Signal Processing, Kyoto, Japan, 2012, pp. 1465–1468
11) Xia, Y., Hu, R., Wang, Z.: '*Salient map extraction based on motion history map'*. Proc. 4th Int, Congress on Image and Signal Processing, Shanghai, China, 2011, pp. 427–430
12) Luo, Y., Tian, Q.: '*Spatio-temporal enhanced sparse feature selection for video saliency estimation'*. Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 2012, pp. 33–38
13) Ren, Z., Gao, S., Chia, L., Rajan, D.: '*Regularized feature reconstruction for spatio-temporal saliency detection'*, IEEE T. Image Process., 2013, 22, (8), pp. 3120–3132
14) Li, Y., Zhou, Y., Xu, L., Yang, X., Yang, J.: '*Incremental sparse saliency detection'.* Proc. 16th IEEE Int. Conf. on Image Processing, Cairo, Egypt, 2009, pp. 3093–3096
15) Guo, C., Ma, Q., Zhang, L.: *'Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform'*. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008
16) Zhai, Y., Shah, M.: '*Visual attention detection in video sequences using spatiotemporal cues'*. Proc. 14th Annual Int. Conf. Multimedia, Santa Barbara, CA, USA, 2006, pp. 815–824
17) Seo, H.J., Milanfar, P.: *'Static and space-time visual saliency detection by self- resemblance'*, J. Vis., 2009, 9, (12), pp. 1–27
18) Li, Q., Chen, S., Zhang, B.: '*Predictive video saliency detection'.* Proc. Chinese Conf. on Pattern Recognition, Beijing, China, 2012, pp. 178–185
19) Le Meur, O., Le Phuoc, T.H., Granjon, L., Guyader, N., Pellerin, D., Guérin-Dugué, A.: 'Modelling spatio-temporal saliency to predict gaze direction for short videos', Int. J. Comput. Vis., 2009, 82, (3), pp. 231–243.
20) U. Rajashekar, L. K. Cormack and A. C. Bovik, "*Point of gaze analysis reveals visual search strategies,*" Human vision and electronic imaging IX 2004, Proc. of SPIE, vol. 5292, pp. 296-306, 2004.
21) W. Tatler, R. J. Baddeley and I. D. Gilchrist, "*Visual correlates of fixation selection: effects of scale and time,"*Vision Research, vol. 45, pp. 643-659, 2005.
22) R. J. Peters, A. Iyer, L. Itti and C. Koch "*Components of bottom-up gaze allocation in natural images*," Vision Research, vol. 45, pp. 2397-2416, 2005.