iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** III **Month of publication:** March 2016

DOI:

www.ijraset.com
IC Value: 13.98
Volume 4 Issue III, March 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Detection and Enhancement of 2d Images

Keerthana.C[1], Shanthalakshmi. K[2]
[1]PG scholar,Adhiyamaan College of Engineering, Hosur
[2]Assistant Professor, Adhiyamaan College of Engineering, Hosur

Abstract: Many saliency detection models of stereoscopic display require multi resolution representation of map data in new models for salient region extraction that is not possible in spectral residual. This provides less accuracy with low performance.so,we use enhanced saliency detection framework for binocular images,depending on the feature contrast of luminance,color,texture and depth that are extracted from DCT coefficients for contrast calculation. In this, the feature are extracted and calculated using Gaussian kernel function where the image is transformed and fusion method is processed to obtain the accurate image.We introduce a novel global saliency metric based on sparse representation. We determine the features that are most dissimilar with respect to the entire image as salient, and thereby discriminate the most informative regions in the image from the rest of the image. The high performance provides better accuracy.
Keywords--- Stereoscopic images, Visual attention, Saliency attention, Human visual acuity

I. INTRODUCTION

Human perception can be thought of as responses to visual cues presented in the image. In other words, there are regions in an image that attract visual attention.The parts that attract visual attention are unclear and often subjective. Some of us could be interested in the snow-clad mountains in the background whereas some could be interested in the group of people that is relaxing. Saliency can be defined as the state or quality of an object to stand out relative to its neighbors. The visual attention reduces the complexity of the scene. This provides visual information from the salient region.
Hue is one of the main properties (called color appearance parameters) of a color defined technically, as a degree to which a stimulus can be described as similar to or different from stimuli that are described as RGB and yellow(unique hues). The other color parameters are colorfulness, chroma, saturation, lightness and brightness.

A. Approaches

1) *Bottom-Up Approach:* Alternatively, saliency can be inferred in a bottom-up manner completely agnostic to the objects present in the image. In this approach, the processing is carried out in one direction from the retina to visual cortex. It is a type of information processing based on incoming data from the environment to form perception. In a bottom-up approach the individual base elements of the system are first specified. These elements are then linked together to form larger subsystems, which then in turn linked, in many levels, until a complete top-level system is formed. This is the piecing together of systems to give rise to more complex systems, thus making the original systems sub-systems of the emergent system.

2) *Process in Saliency Identification*: saliency can act as a feature space in which such an observation can be made— i.e. the uniqueness can be inferred. Although not in the context of images specifically, Smith *et al*. talk about the human performance dependence on design features of the performance environment and provide empirical support for the context specificity of important features. We are interested in saliencies that exist in image patches (small regions of pre-determined sizes) and can be computed at low level and then any additional inferences could be made at higher levels. We can then demonstrate how saliency is used to create visual models for various object classes in order to carry out some computer vision tasks. The relationship between visual attention and the image saliency scores we determine is interesting to investigate. In the field of computer vision, saliency has been relevant for feature detection, foreground prediction and compression.

One way to go about detecting saliencies is to use a classification approach. If we attempt to explicitly classify salient vs. non-salient patches, the problem becomes dependent on training information and less bottom-up in nature. It also becomes a highly supervised problem. If we divert our attention to only the salient regions in the image, then we can argue that any addition to this set of salient regions or removal from this set can be striking.

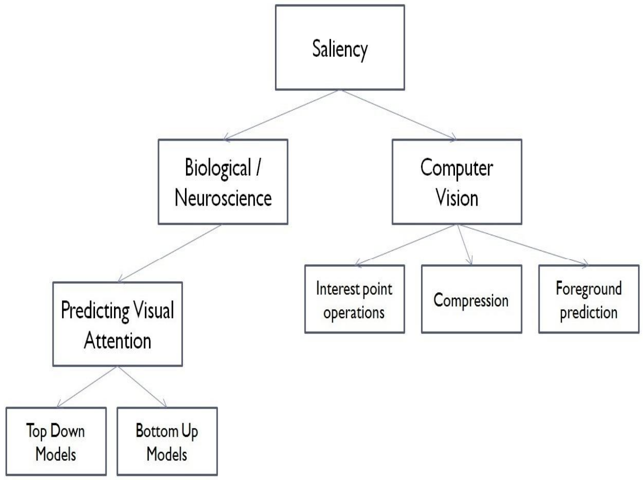


Figure 1: Applications of Saliency in Neuroscience and Computer Vision

II. PROPOSED WORK

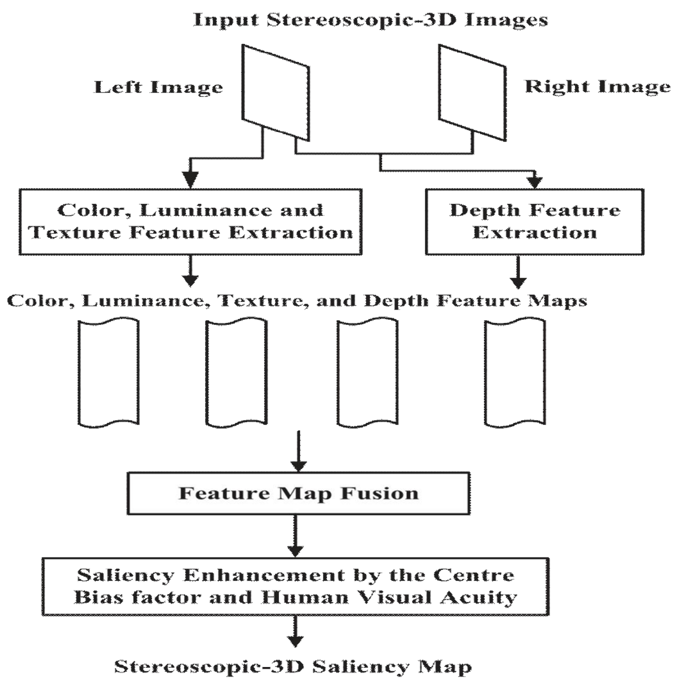


Fig 2: The framework of the proposed model

The transformation of an image from its original form to the saliency feature is processed based on the framework model. firstly,the color,luminance,depth and texture features are extracted from the input stereoscopic image. Based on these features,the feature contrast is calculated for the feature map calculation. A fusion map is designed to fuse all the extracted features into the saliency map. Each step in detail is described as follows.

B. Image Acquisition

Based on human visual acuity, the input image can be of stereoscopic image or binocular images. After the acquisition of the input

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

image, the features such as color, luminance, texture and depth are extracted from DCT coefficients.

C. Feature Detection

The basic feature from an image is the pixel intensity or the RGB tuple. There are still many algorithms that produce state-of-the-art performance with only RGB or gray scale values. For specific applications it is often useful to investigate different color spaces.
The HSV color space is considered to be close to the human perception model and it is considered to be more useful in computer graphics. There are other color spaces like HSV (hue, saturation and value), CMYK (cyan, magenta, yellow and key), CIELAB (International commission on lumination, Luminace a & b color space) etc., which approximate human perception model closely and have found uses in computer vision for classification applications.

D. Edge Detection

The next level of feature detection is probably edge detection (first and second order differences across adjacent pixels). Edges are boundaries or discontinuities in image intensity values. There are many edge detectors like Sobel, Prewitt, Roberts' Cross, Canny etc., which have varying levels of complexity and performance. Corners are points in the image which are intersection of two edges and are the next level of features. Like edge detectors there are many corner detectors. Harris corner detector is perhaps the first robust corner detector and is based on the weighted sum of squared differences (SSD) between image patches along the different directions. Shi and Tomasi provided an improved corner detector based on the Harris corner detector by measure invariance to affine transformations. In an extended discussion of advanced scale and affine adapted corner detectors is presented.

E. Shape Context

Shapes can be considered as a collection of contours which are in turn a collection of edges and therefore make the next level of feature description. Other local edge and corner based features can also be considered in this level. Shape contexts have been demonstrated to be a viable choice for matching and recognizing digits, letters and 3D objects.

F. Texture

The entropy is measured by the probability density function (pdf) of local patch variations. They provide a parametric approach to filter patches with low entropy and to measure local support. In contrast to this approach, our notion of saliency differs in definition. For instance the local entropy of a patch being high may not make it salient by itself. But the pattern can be unique or less like other patterns in the entire image providing a different and more global definition to saliency. Saliency detection for the purpose of feature computation has been approached predominantly as a local phenomenon. In this thesis we propose detecting saliencies globally with respect to the entire image and we determine its usefulness in various applications. The saliency detection algorithm presented in this thesis extracts distinct patterns which are least like other patterns in the image.

G. Feature Extraction

The features such as color, luminance,texture and depth are extracted based on Gaussian kernel function. The kernel function is obtained from K-SVD algorithm.

III. ALGORITHM

KSVD is the algorithm which involves feature extraction from the input images. It consists of Gaussian kernel function which is mainly used to calculate the depth of the pixel.
We observe in the KSVD procedure that the dictionary accommodates new data by changing the dictionary atoms that have the most relevance to the current data sample, by modifying the pertinent atoms. Over multiple iterations (typically 20 to 30) , the dictionary converges to a truly representative dictionary given a large number of training samples minimizing the average training error.
We typically generate 3 scaled versions of the image by repeatedly convolving the original image I with a Gaussian kernel $g(\sigma)$. Where σ is the standard deviation of the kernel.
DImultiscale = $[DI\sigma|DI\sigma1|DI\sigma2|DI\sigma3]$ (3.1)
While performing orthogonal matching pursuit with the combined multiscale dictionary, the different scales compete with each

other in trying to represent the patch of interest. In this way, the effect of varying patch sizes is also reduced.

A. Bottom-Up Approach

The saliency detection method as a feature detector, or a foreground detection method for various applications are described with more relevant work specific to the application. Saliencies are based on euclidean distances in the LAB color space. In this paper we use bottom-up approach as shown in figure 3.3. In a bottom-up approach the individual base elements of the system are first specified. These elements are then linked together to form larger subsystems, which then in turn linked, in many levels, until a complete top-level system is formed.

IV. METHODOLOGY

The features such as color, luminance,texture and depth are extracted from DCT coefficients. The transformation is based on DCT coefficients because it provides finite values. The background of the images are as follows.,

Color – hue is one of the main properties called color appearance parameters technically defined as the degree to which a stimulus can be described as RGB.
Luminance – It is the amount of light that the image contains.
Texture – Texture is the pattern or segmentation of the image. The analysis consists of structured approaches and statistical approaches.
Depth – It is the distance information corresponding to rows and columns of array as in the order.
Our context-aware saliency follows four basic principles of human visual attention, which are supported by psychological evidence
Local low-level considerations, including factors such as contrast and color.
Global considerations, which suppress frequently occurring features, while maintaining features that deviate from the norm.
Visual organization rules, which state that visual forms may possess one or several centers of gravity about which the form is organized.
High-level factors, such as human faces.

A. Feature Calculation

In bottom-up approach, We draw motivation from the theory of Kolmogorov complexity and entropy. We aim to determine the most informative regions in an image or video sequence as defined by its description length. The more complex (or longer) the description length of a region (patch) is the more informative or complex is the patch. This is divided into three parts: theoretical motivation, problem formulation and the algorithmic approach.

B. Algorithmic Approach

There are two parts to this problem. Learning this reconstructive dictionary D, and given the dictionary, how to retrieve the optimal set of sparse coefficients $\alpha^*$to represent the signal x, which is an NP hard problem. The dictionary learning procedure can be carried out using the K-SVD (K-Singular Value Decompositions)algorithm, or the MOD (method of optimal directions) algorithm. A greedy orthogonal matching pursuit algorithm can be used for retrieving the α^*efficiently , despite being suboptimal.
The OMP algorithm selects greedily at each step the atom with the highest correlation (maximum dot product) to the current residual. Once the atom is selected, the signal is orthogonally projected back to the span of the selected atoms, the residual is recomputed, and the process is repeated. It is orthogonal in that the current residual is eliminated for computing the projection in the next step.

Volume 4 Issue III, March 2016
ISSN: 2321-9653

Input: $\mathrm{D} \in \mathbb{R}^{n \times k}$ (dictionary from the previous iteration); M vectors $\mathrm{x}_l \in \mathbb{R}^n$ (input data); $\alpha \in \mathbb{R}^{k \times M}$ (coefficients from the previous iteration).
Output: D and α.
Loop: For $j = 1 \ldots k$, update d, j-th column of D,
• Select the set of patches that use d:

$$\omega \leftarrow \{l \in 1, \ldots, M \text{ s.t. } \alpha_l[j] \neq 0\}. \quad (5)$$

• For each patch $l \in \omega$, compute the residual of the decomposition of x_l: $\mathrm{r}_l = \mathrm{x}_l - \mathrm{D}\alpha_l$.
• Compute a new atom $\mathrm{d}' \in \mathbb{R}^n$ and the associated coefficients $\beta \in \mathbb{R}^{|\omega|}$ that minimize the residual error on the selected set ω, using a truncated SVD:

$$\min_{||\mathbf{d}'||_2=1, \beta \in \mathbb{R}^{|\omega|}} \sum_{l \in \omega} ||\mathrm{r}_l + \alpha_l[j]\mathrm{d} - \beta_l \mathrm{d}'||_2^2. \quad (6)$$

• Update D using the new atom d', and replace the scalars $\alpha_l[j] \neq 0$ from Eq. (5) in α using β.

Figure 3: The KSVD dictionary update procedure

C. *Detection Of Context Aware Saliency*
In this section we propose an algorithm for realizing principles .In accordance with principle

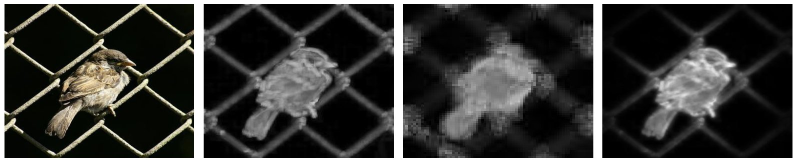
1) Areas that have distinctive colors or patterns should obtain high saliency. Conversely, homogeneous or blurred areas should obtain low saliency values. In agreement with principle
2) Frequently-occurring features should be suppressed. According to principle
3) The salient pixels should be grouped together, and not spread all over the image.
4) Is implemented as post-processing. This implies that a patch p_i is salient when the patches similar to it are nearby, and it is less salient when the resembling patches are far away.

D. *Multi-Scale Saliency Enhancement*
Background pixels (patches) are likely to have similar patches at multiple scales, e.g., in large homogeneous or blurred regions. This is in contrast to more salient pixels that could have similar patches at a few scales but not at all of them. Therefore, we incorporate multiple scales to further decrease the saliency of background pixels, improving the contrast between salient and non-salient regions.
The larger $\bar{S}_i$ is, the more salient pixel i is and the larger is its dissimilarity (in various levels) to the other patches.

E. *Fusion Method*
The extracted features are combined to form the saliency image. The extractions are fused based on the best matching approach. The fusion is based on greedy orthogonal pursuit algorithm.

V. SALIENCY ENHANCEMENT
Finally, the saliency map should be further enhanced using some high-level factors, such as recognized objects or face detection. In our implementation, we incorporated the face detection algorithm of, which generates 1 for face pixels and 0 otherwise. The saliency map of Equation is modified by taking the maximum value of the saliency map and the face map.Let $d_{position}(p_i, p_j)$ be the Euclidean distance between the positions of patches p_i and p_j, normalized by the larger image dimension.

Volume 4 Issue III, March 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)



(a) Input (b) Scale 1 (c) Scale 4 (d) Final

Figure 4: The steps of our saliency estimation algorithm

A. Conspicuity Maps

In 2nd step of the Saliency attention model, conspicuity maps are computed. To achieve transforming each feature map into its conspicuity map, conspicuity maps highlight the parts of the scene which strongly differ according to a specific feature, from their surroundings. In practice, we can implement this mechanism with a *difference-of-Gaussians (DoG)* filter. We can apply *DoG* on feature maps and can extract local activities for each feature map. $M_{j,k} = \left| P_j(c_k) \ominus P_j(s_k) \right|$ -------1

Where $\ominus$ is called cross-scaled difference operator, which first interpolates the coarser scale to the finer one and then it carries out a point by point subtraction.

B. Saliency Maps

In 3rd step of the Saliency attention model, the integration of cue-related conspicuity maps $\hat{C}_{cue}$ with a *saliency map S* is done in a competitive manner.

$$S = \sum_{cue=1}^{l} N\left(\hat{C}_{cue}\right)$$

Where *l* denotes the number of considered cues. As we have mentioned above, then operator *N*(.) We has used seven features (*n=7*) in his implementation in [93], which belongs to the three different cues (*I* = *3*). He computed six (*K* = *6*) multi-scale conspicuity maps $M_{j,k}$ regarding to the conspicuity operator, which means that he computed 62 multi-scale maps for seven features.

VI. EXPERIMENTS AND RESULTS

A. Foreground Detection

For this analysis we used the MSRA salient object detection dataset. Achanta *et al.*provide ground truth for 1000 images from this dataset. Some sample saliency maps for this dataset are shown in Figure 5.3. The ROC areas were compared for different methods. A value of 0.5 indicates a random worthless predictor, whereas a value of 1 indicates that the method is a perfect predictor. The ROC curve is obtained by plotting the true positive rates ($\frac{True\ Positives}{All\ positives}$ also known as hit rate or recall) versus the false positive rate ($\frac{False\ Positives}{All\ negatives}$ also known as fall out) . One such ROC curve is shown in Figure 5. We can obtain this by thresholding the saliency map with different threshold values and computing the false positive and true positive rates. The ROC area is considered a better performance metric than accuracy alone as it is invariant to the choice of the threshold.

The average ROC areas for the different methods are shown in Table 1.

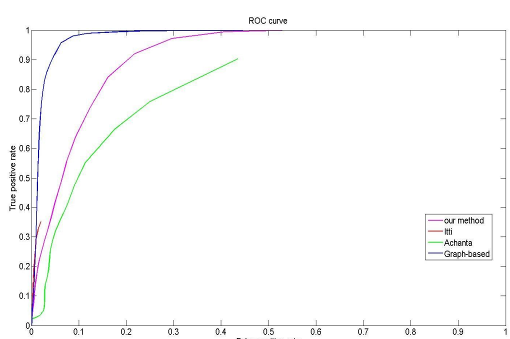


Figure 5: ROC plots corresponding to one image for all methods compared.

www.ijraset.com Volume 4 Issue III, March 2016
IC Value: 13.98 ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

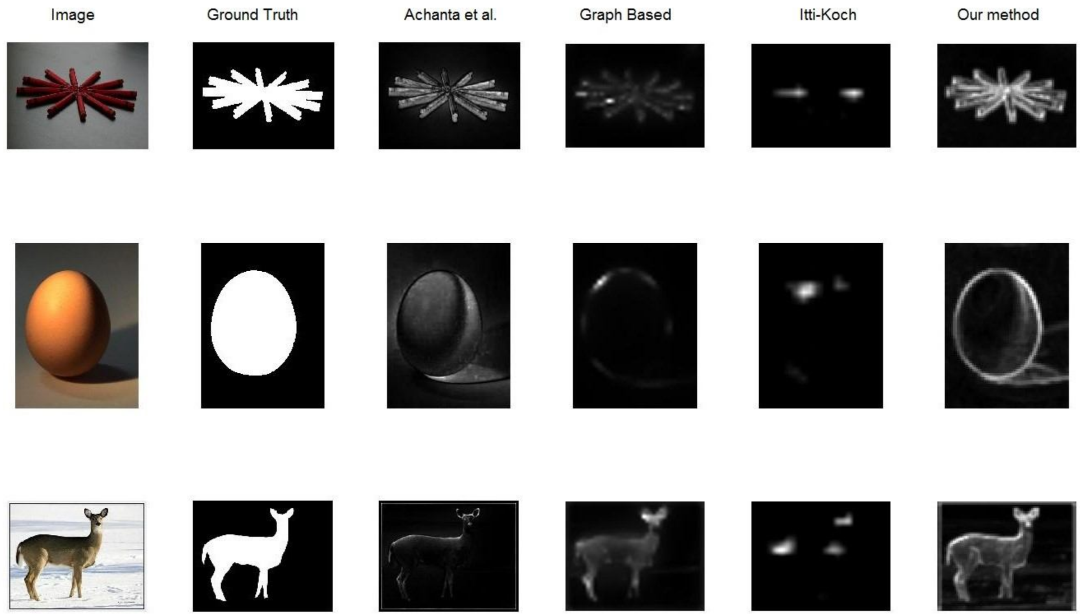


Figure 6: Comparison of Saliency maps for some images from the MSRA dataset.

Method	Average ROC area
Itti	0.58
GBVS	0.61
Frequency Tuned	0.82
Our method	0.75

Table 1: Average ROC areas for different methods on the MSRA salient object dataset

A. Predicting Visual Attention

For this experiment we used the MIT eye fixation dataset . We used their blurred saliency map obtained from eye fixation points as ground truth. With a ground truth G and a saliency map A we can compute the precision and recall of the object segmentation as mentioned before. We used 100 randomly sampled images for training to learn the best value of M for selecting the top salient regions which yielded the best average precision and recall using the ground truth. Since eye gaze track points are sparse, choosing the top 10% salient regions yielded best results. Even though our recall was sometimes poor, we obtained the best average results overall. Our method performs better than the frequency-tuned saliency estimation approach and the Itti-Koch saliency

Volume 4 Issue III, March 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

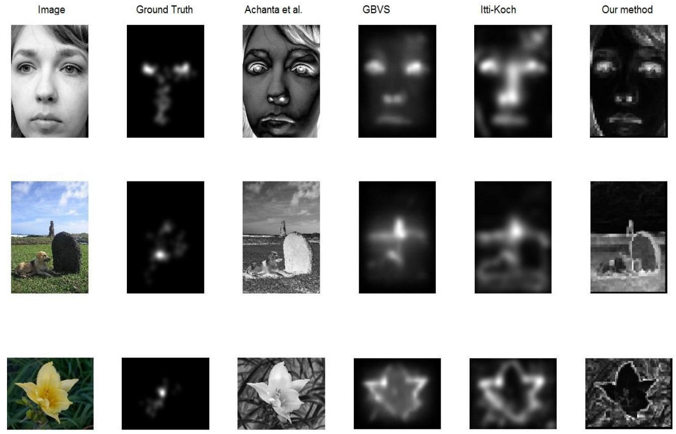


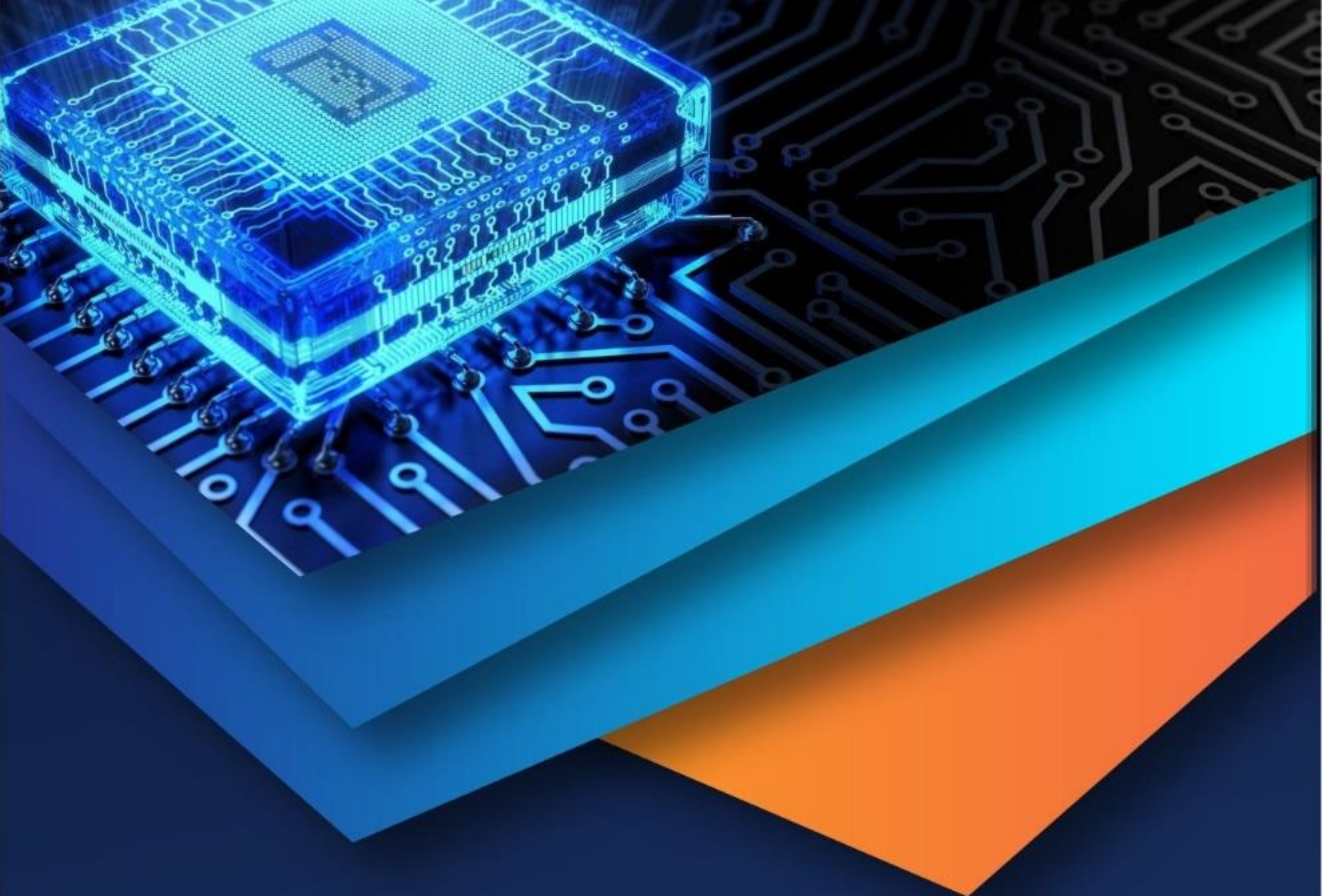
Figure 7: Comparison of saliency maps of different methods on the MIT eye fixation database.

VII.CONCLUSION

In this work saliency detection and its use in multimedia applications were studied. Two novel methods of salient region detection were proposed. The initial idea of region-based processing has been proven to be working. Another novel idea of exploiting depth-spatial relation of objects has also shown to be a possible solution to saliency detection. The evaluation and comparison to state-of-the-art methods has shown high performance of the proposed approaches. In addition, an application-based evaluation has been done through finding a new niche for saliency detection. A novel approach to diversification of visual content based on saliency detection has been presented. Its evaluation has shown the rightfulness of this idea and became a good test-bench for proposed saliency detection algorithm.Although the initial goals of the work were reached there is a room for further development. The proposed models of saliency detection are to large extend based on low-level properties of objects found in images. However, studies on human visual attention has shown that there are two stimulus driving our attention bottom-up and top-down. Addressing top down driven attention via exploiting semantic properties and relations of objects in images may lead to much higher performance. Though this research is possible only through solving tough tasks from natural language processing, semantics understanding and object categorization domains.

REFERENCES

[1] JU/’arcin Marsza lek and Cordelia Schmid. Accurate object localization with shape masks. In IEEE Conference on Computer Vision & Pattern Recognition, jun 2007.
[2] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. European Conference on Computer Vision–ECCV 2006, pages 404–417, 2006.
[3] L. Itti and C. Koch. Computational modelling of visual attention. Nature Reviews Neuroscience, 2(3):194–203, 2001.
[4] T.J. Smith, R.A. Henning, and K.U. Smith. Performance of hybrid automated systemsa social cybernetic analysis. International journal of human factors in manufacturing, 5(1):29–51, 1995.
[5] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 2, pages 2169 – 2178. Ieee, 2006.
[6] 115
[7] Anna Bosch, Andrew Zisserman, and Xavier Munoz. Representing shape with a spatial pyramid kernel. In Proceedings of the 6th ACM international conference on Image and video retrieval, CIVR ’07, pages 401–408, New York, NY, USA, 2007. ACM.
[8] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. European Conference on Computer Vision–ECCV 2006, pages 589–600, 2006.
[9] G. Somasundaram, V. Morellas, and N. Papanikolopoulos. Object classification with efficient global self-similarity descriptors based on sparse representations. In IEEE Internation Conference on Image Processing (ICIP). IEEE, 2012.
[10] G. Somasundaram, Morellas. V, and Papanikolopoulos.N. Action recognition using global spatio-temporal features derived from sparse representation. Submitted to Computer Vision and Image Understanding.
[11] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, June 2008.
[12] Maureen C. Stone, Maureen Stone, and Maureen C. Stone. A survey of color for computer graphics siggraph 2001 description presenter, 2001.
[13] G. Hoffmann. Cie color space, 2002.



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY