



PERSUADING VISUAL ATTENTION THROUGH LOW-LEVEL IMAGE FEATURES

VIKASH KUMAR
CSE, 2nd Year
IIT HYDERABAD
EMAIL: cs11b038@iith.ac.in

Project Guide:
Dr. Rajarshi Pal
Assistant Professor
IDRBT, Hyderabad

INSTITUTE OF DEVELOPMENT AND RESEARCH IN BANKING
TECHNOLOGY (IDRBT)
ROAD No. 1, CASTLE HILLS, MASAB TANK,
HYDERABAD-500057

CERTIFICATE

This is to certify that **Mr. VIKASH KUMAR**, pursuing Computer Science and Engineering course at **Indian Institute of Technology Hyderabad** has undertaken a project as an intern at **IDRBT, Hyderabad** from **May 13, 2013 to July 12, 2013**.

He was assigned the project "**Persuading Visual Attention Through Low-Level Image Features**" under my guidance and has done excellent work.

I wish him All the Best in his endeavours for Knowledge discovery.

Dr. Rajarshi Pal
(Project Guide)
Assistant Professor
IDRBT, Hyderabad

Acknowledgment

I express my deep sense of gratitude to my Guide **Dr. Rajarshi Pal**, Assistant Professor, IDRBT for giving me an opportunity to do this project in the Institute for development and research in Banking Technology and providing all the support and guidance needed which made me complete the project on time.

I am also thankful to **Indian Institute of Technology Hyderabad** for giving me this golden opportunity to work in a high-end research institute like **IDRBT**.

Vikash Kumar

CSE 2nd Year

IIT Hyderabad

Contents

1. Introduction	5
2. Project Description	6
3. Saliency	7
3.1 Visual Saliency	7
3.2 Application of Visual Saliency	8
3.3 Visual Saliency Modelling	9
3.3.1 Bottom-Up Model	9
3.3.2 Top-Down Model	11
4. Computing Saliency Map	12
4.1 Graph Representation of Feature Map	12
4.1.1 Forming Nodes	13
4.1.2 Calculating Edge Weights	14
4.2 Binarizing Graph	15
4.3 Calculating Degree Centrality	15
4.4 Forming Saliency Map	16
5. Changing the Saliency Map	17
6. Conclusion and future scope	21
7. References	22

1. INTRODUCTION

Attention is one of the most important components of primate vision. Selective visual attention is the mechanism by which we can rapidly direct our gaze towards objects of interest in our visual environment. Attention only allows a small part of the incoming sensory information to reach short-term memory and visual awareness, allowing us to break down the problem of scene understanding into rapid series of computationally less demanding, localized visual analysis problems [1].

Visual attention can be bottom-up (image-based) and mostly driven by unconscious mind or top-down (task-dependent) and mostly driven by memory and conscious mind.

Computational visual attention systems aim to detect regions of interest in images and psychovisual experiments have shown that in the absence of any external guidance, attention is directed to visually salient locations in the image.

Now-a-days almost every device let it be a watch, phone or even television have got the ability to capture images on the go. But sometimes, the portion of our interest may fail to draw attention and probably some other object in the image draws the user attention. In this project, a step has been taken to address this problem by proposing an algorithm which will tell by how much the intensity of the desired object has to be increased so that it becomes the most salient part in the whole image.

[2] Proposes a methodology that makes a camera even smarter by converting the objects of interest to the user more salient. It relies on different positions by which the object stands out different from its surrounding and taking saliency at each position and choosing that position in which the saliency of the object is maximum.

2. PROJECT DESCRIPTION

An entirely different approach is used here where required object's saliency has been tried to improve by changing the intensity of the original image. In this project, an attempt has been made to find a method based on Intensity so that the user defined region in the gray-scale image draws immediate attention of the viewer which the original image may fail to do so.

An algorithm has been given which tell the amount of change that has to be made in the intensity so the saliency of the desired object is increased in the whole image.

A network and degree centrality based approach [3] where nodes represent accumulation of similar pixels and the dissimilarity in terms of features between any pair of such accumulations is encoded as edge-weight between corresponding nodes is used to form the saliency map of the image.

Modification of this saliency map is done using the algorithm proposed and results were obtained.

3. SALIENCY

The **salience** (also called **saliency**) of an item – be it an object, a person, a pixel, etc. – is the state or quality by which it stands out relative to its neighbours. Saliency typically arises from contrasts between items and their neighbourhood, such as a red dot surrounded by white dots, a flickering message indicator of an answering machine, or a loud noise in an otherwise quiet environment. Saliency detection is often studied in the context of the visual system. The term is widely used in the study of perception and cognition to refer to any aspect of a stimulus that, for any of many reasons, stands out from the rest. Salience may be the result of emotional, motivational or cognitive factors and is not necessarily associated with physical factors such as intensity, clarity or size.

3.1 VISUAL SALIENCY

Visual salience (or visual saliency) is the distinct subjective perceptual quality which makes some items in the world stand out from their neighbours and immediately grab our attention. Visual saliency is the distinct subjective quality which makes some item stand different from other objects in its vicinity. Our attention is mostly attracted towards most salient object in the field of view.

Visual salience helps your brain achieve reasonably efficient selection. Early stages of visual processing give rise to a distinct subjective perceptual quality which makes some stimuli stand out from among other items or locations. Our brain has evolved to rapidly compute salience in an automatic manner and in real-time over the entire visual field. Visual attention is then attracted towards salient visual locations.

3.1 APPLICATIONS OF VISUAL SALIENCY

Visual saliency provides a relatively inexpensive and rapid mechanism to select a few likely candidates and eliminate obvious clutter and has applications like:

- Automatic target detection (e.g. finding traffic signs along the road or military vehicles in deserts.)
- Robotics (using salient objects in the environment as navigation landmarks.)
- Image and video compression (e.g. giving higher quality to salient objects at the expense of degrading background clutter.)
- Automatic cropping/centering of images for display on small portable screens
- Finding tumors in mammograms.

4. COMPUTING SALIENCY

Visual attention is an indispensable component of complex vision tasks. A network-based approach for determining visual saliency is used as described in [1]. It uses degree centrality over a network of image regions to form a saliency map. Centrality of a vertex within a graph determines the relative importance of a vertex within the graph, examples, of centrality could be: how influential a person is within a social network or how well-used a road is within an urban network.

Degree centrality have also been used in old search engines where a link is given preference if it is linked to more sites and thus, appear above the search results.

Degree of a node in a graph is the number of edges incident upon it.

4.1 GRAPH REPRESENTATION OF FEATURE MAP

Low-level feature that we are using to construct the saliency map is intensity of the pixels in the image and hence saliency map is called intensity saliency map. Experiments have proved that if a location is significantly different from all other locations and specifically from its surround, it is salient, that is, it draws user's attention.

In this method image is represented as a graph consisting of set of nodes and edge-weights and node with the maximum degree is the most salient node in the whole image or the portion represented by the node is the most salient. Nodes in this network represent accumulation of similar pixels and the dissimilarity in terms of features between any pair of such accumulations is encoded as edge-weight between corresponding nodes. This dissimilarity measure is modulated by the distance between the pair of nodes. This ensures that nodes which are closer to each other, the edge between them get higher weightage as compared to nodes which are farther away.

Incorporation of degree centrality analysis with this type of network suggests that a centrality situated node belongs to a salient location.

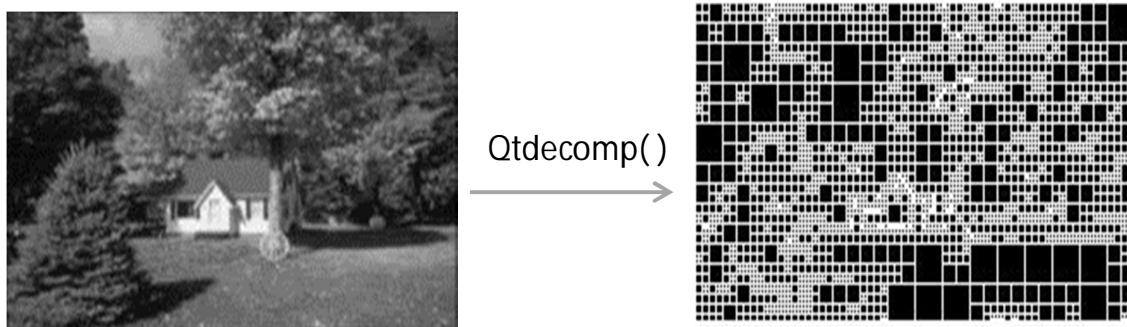
4.1.1 FORMING NODES

Graph is constructed over the intensity feature map and a block of connected and homogenous pixels is represented by a single node or block in the graph. Quadtree decomposition technique is used to divide the image into blocks.

The homogeneity across a block of pixels in terms of a feature is determined by estimating the difference between maximum and minimum feature values within that block. If the difference between the maximum and the minimum feature values in a block is greater than a particular threshold, taken to be 15/255-th of the dynamic range of the feature map, then that block is not homogeneous and is decomposed further using quadtree decomposition technique.

Quadtree decomposition technique divides a square image into four equal-sized square blocks, and then tests each block to see if it meets some criterion of homogeneity. If a block meets the criterion, it is not divided any further. If it does not meet the criterion, it is subdivided again into four blocks, and the test criterion is applied to those blocks. This process is repeated iteratively until each block meets the criterion. The result can have blocks of several different sizes.

An example for an image that has been subjected to quadtree decomposition in Matlab has been shown below:



All the nodes above are homogenous i.e they satisfy the given threshold and form the collection of nodes in the graph.

4.1.2 CALCULATING EDGE-WEIGHTS

After forming the collection of nodes, edge-weights between each pair of nodes is calculated. For edge-weights two things is taken into account:

- Feature distance
- Spatial distance i.e actually distance between the pair of nodes

Edge-weight is directly proportional to feature distance and inversely proportional to spatial distance between the pair of nodes. This is because feature distance is directly related to intensity and nodes which are closer to one another contribute more to salience as compared to nodes which are farther.

Feature distance is estimated as the absolute difference of the mean feature values of these blocks and normalised with respect to the maximum value of feature distance among any pair of blocks.

$$D_{f_{ij}} = |\mu_{f_i} - \mu_{f_j}| / \max_{\forall k, \forall l} (|\mu_{f_k} - \mu_{f_l}|)$$

where μ_{f_i} represent the mean feature value of node i.

Edge-weight is inversely proportional to spatial distance between those blocks and a Gaussian function is used here to stimulate the decay of feature influence with spatial distance. The spatial distance two blocks is computed as the Cartesian distance between the midpoints of these nodes and normalised with respect to the maximum possible distance among any pair of blocks.

So, the final edge-weight formula between two nodes i and j is:

$$F(V_i, V_j) = D_{f_{ij}} e^{-D_{f_{ij}}^2 / 2\sigma^2}$$

Where σ is the standard deviation of the Gaussian function.

4.2 BINARIZING GRAPH

As the edge weights are proportional to feature dissimilarity, edges with higher weight are taken for calculation. Now the graph so formed contains edges whose edge weights are negligible and hence are of less importance. Thus the subset of edges with weight greater than a particular threshold T is thus retained and other edges are discarded.

This threshold T is used to binarize the graph where edge weights which are greater than T is given a value of 1 and those with less value is given a value of 0.

4.3 CALCULATING DEGREE CENTRALITY

Degree of a node, which is also a centrality measure, is a structural attribute of the node in a network. It shows the importance of that node among all other nodes as it is defined by the number of other nodes to which it is directly connected.

Since graph is already formed that means its adjacency matrix is with us. Degree of each node can be calculated by row-wise/column-wise summation of the adjacency matrix of the graph.

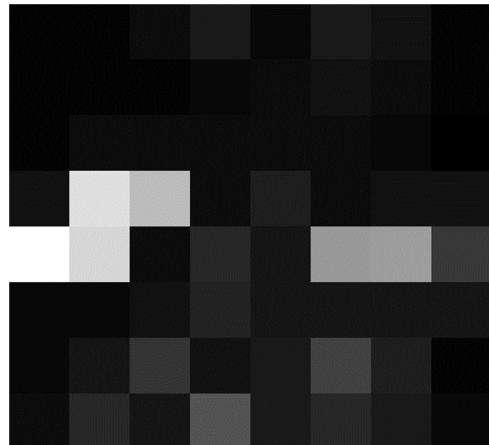
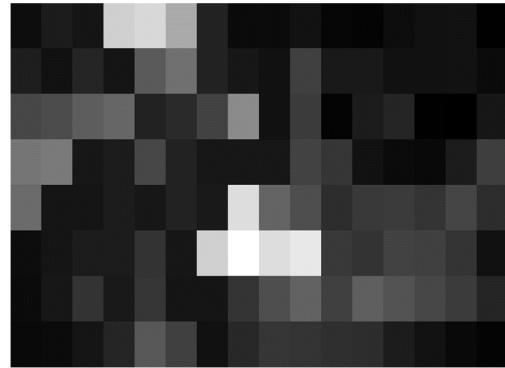
Thus, Degree Centrality of node ‘i’ is given by

$$DC_i = \sum_j A_{i,j}$$

As saliency of certain location is determined by how dissimilar it is from other locations, specially its surroundings, the degree of a node measures the saliency of the location corresponding to that node.

4.4 FORMING SALIENCY MAP

A saliency map is formed by mapping the degree centrality values of all the nodes to the pixels corresponding to their respective locations. Nodes with higher intensity in saliency map are most salient node in the map.



Original Image

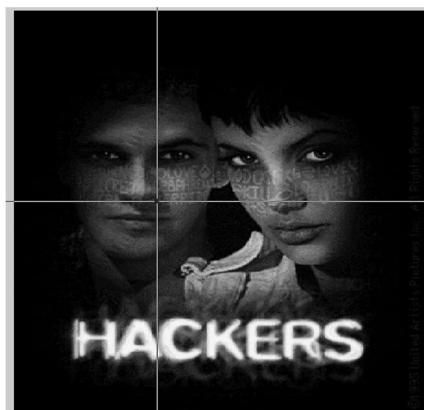
Saliency Map

Above figures represent original image along with their saliency map. Brighter region correspond to salient part in their corresponding part in the original image. These maps were taken for 50 gray-scale images.

5. CHANGING THE SALIENCY MAP

Since we want to make the desired node most salient so we have to either change the feature distance i.e feature value of the node or we have to decrease the spatial distance as spatial distance is inversely proportional to saliency map but changing the spatial distance is not an option because this will change the size of the modified image which is not acceptable. So, only option left is changing the feature value of the node to be made salient.

For the implementation part, user is given the choice to select a point in the region he wants to be salient. User selected point is then used to find the node in which the point lies.



Now, the below algorithm will tell by how much the feature value of the node is too be increased such that it becomes the most salient node in the whole saliency map.

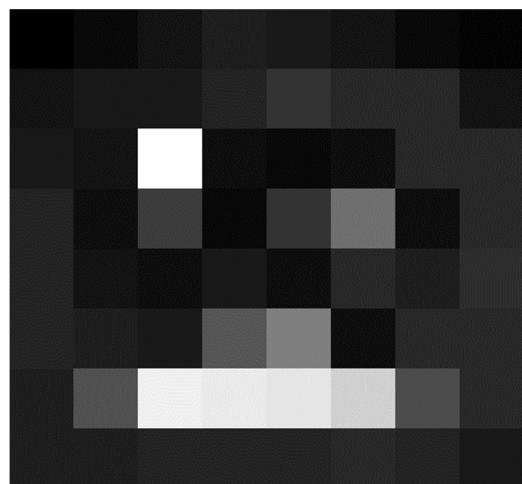
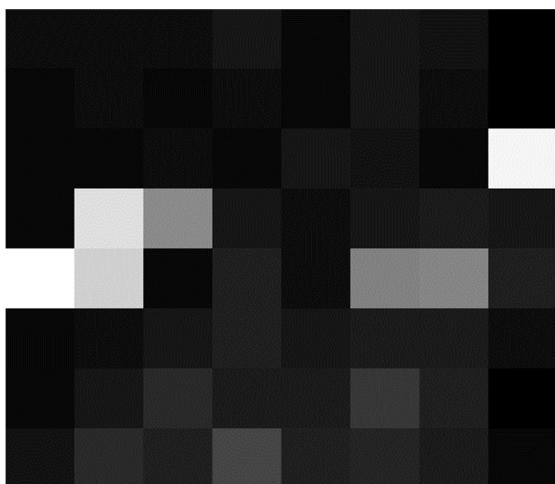
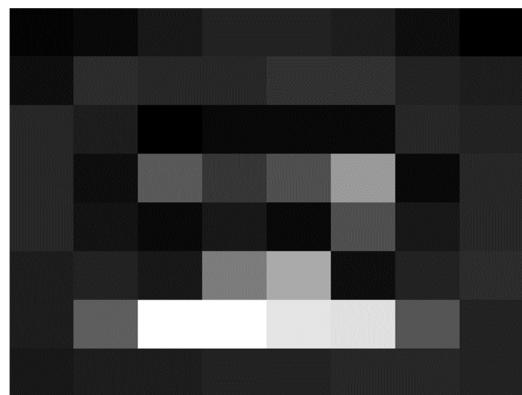
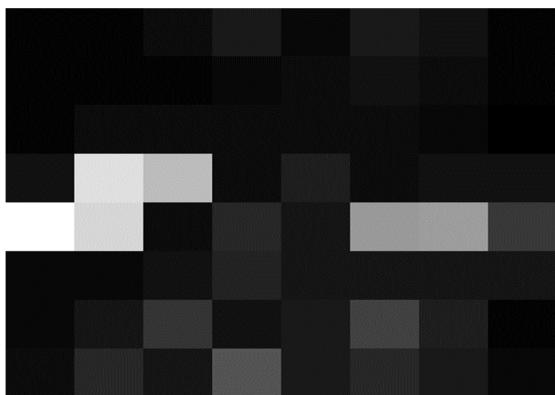
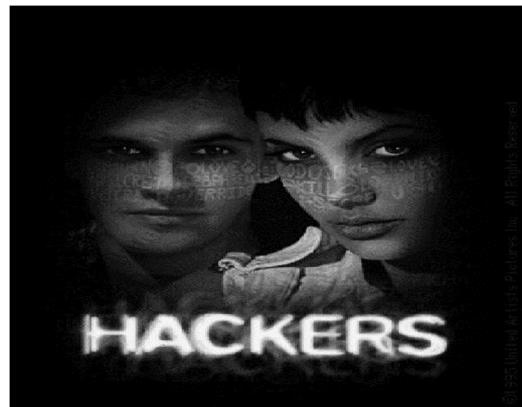
Algorithm:

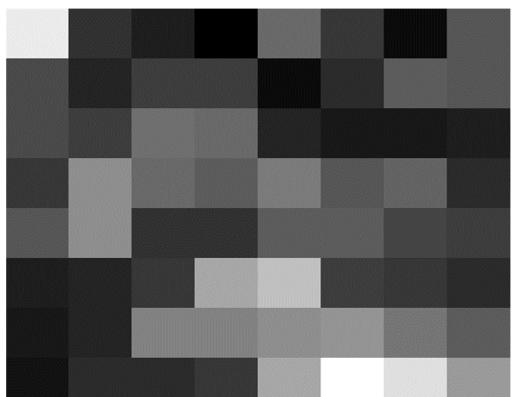
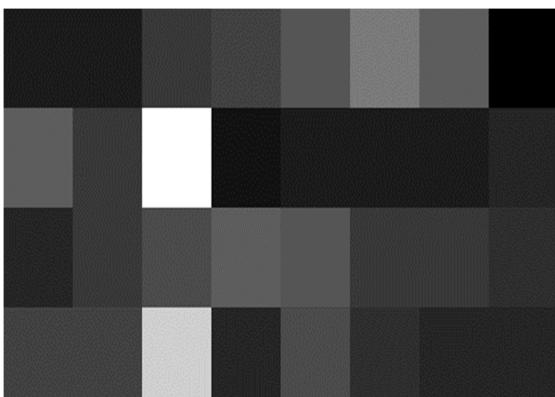
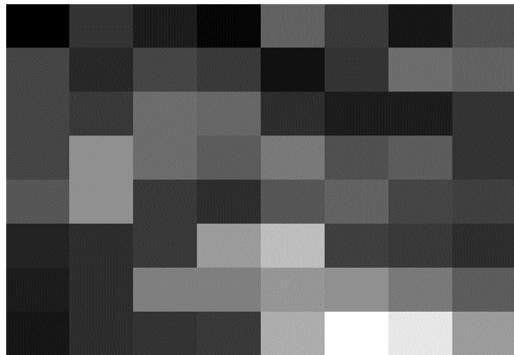
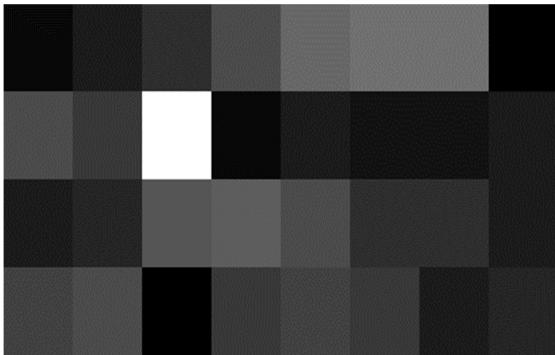
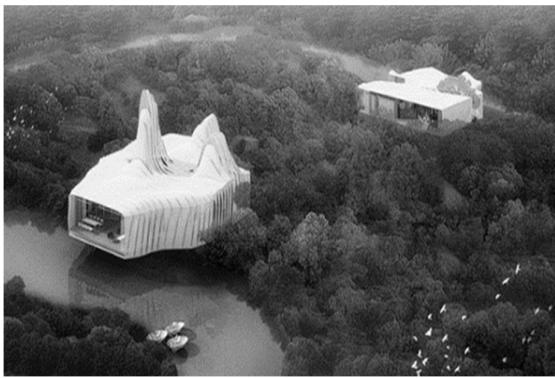
1. Read an image
2. Find the threshold for Qtdecomp
Threshold = min value of intensity in image + (max value – min value)*0.0588
3. Do qtdecomp() of image with above threshold
4. Form the collection of nodes and edge-weights
5. Find the cut-off threshold and binarize the graph
6. Form the degree centrality matrix
7. Find the node with the max. degree and note its coordinate
Index= ind2sub (size (degree), find (degree == max (degree)))
8. Find the coordinate of the node which has to be made salient.
9. Use these coordinates and find their feature value and then the difference

$$\Delta f\text{value} = f\text{value}(\text{salient node}) - f\text{value}(\text{node to be made salient})$$

10. Use this difference to increase the feature value of node to be made salient.
-

RESULTS: Below are images and their modified image by using the above algorithm. First figure in a column is original image, second is its saliency map and third is the modified image followed by the modified image saliency map.





6. CONCLUSION AND FUTURE SCOPE

In this study, we have taken a small step and have considered the problem where we tried to increase the saliency of the required part of the image based on the algorithm proposed and results were obtained as expected.

After comparing the original image feature map and modified image feature map we found that using the proposed technique of changing the feature value i.e the intensity of the image saliency of the selected part was increased.

In this approach we only considered one feature value, that is, intensity but this method can be further modified to inculcate many feature values like orientation, colour, texture etc.

Image segmentation can also be used to select only the object then the block effect will also not be there.

7. References

- [1] Laurent Itti, University of Southern California “Visual Attention”
- [2] Rajarshi Pal, Pabitra Mitra, Jayanta Mukhopadhyay
“ICam:Maximizes Viewers’ Attention on Intended Objects”
- [3] R. Pal, A. Mukherjee, P. Mitra, J.Mukherjee, “Modelling visual saliency using degree centrality”
- [4] Simone Frintrop, Erich Rome, Henrik I. Christensen
“Computational Visual Attention Systems and Their Cognitive Foundations: A Survey”
- [5] www.scholarpedia.org/article/Visual_salience