A Case-Based Reasoning Approach for Detection of Salient Regions in Images

Aditi Kapoor*
Amar Nath and Shashi Khosla School of Information Technology
Indian Institute of Technology Delhi
Hauz Khas, New Delhi - 110016
aditi@cse.iitd.ernet.in

K. K. Biswas
Department of Computer Science and Engineering
Indian Institute of Technology Delhi
Hauz Khas, New Delhi - 110016
kkb@cse.iitd.ernet.in

ABSTRACT

Automatic salient region detection in an image is useful for image compression, image cropping for resizing on smaller displays, object recognition and tracking. In this paper case-based reasoning is used to make a system learn the relevant attributes of a salient region, based on global and local color contrast. CIELab colorspace is used as it is perceptually uniform and matches the human visual perception. The parameters used are background colors at the boundary, color distance, spatial variance, size of the connected components of salient color and dominant colors. Intensity values are used to deal with images containing black and white shades. Exemplar images are presented to the system to categorize the cases. The method is tested on a large image database.

Figure 1: Sample images with salient objects marked with yellow rectangles.

1. INTRODUCTION

Identification of salient regions in images and videos helps in object based image cropping, adaptive image resizing for display on smaller devices, image or video compression, advertising design, besides object recognition and tracking [1] - [6], [8] - [10]. Our brains do not register everything that is presented to our visual field. Although a number of objects could be visible to the human eye at any point in time, the attention gets focused on a particular object or a group of objects which are more conspicuous by virtue of their contrast with the surrounding, or by virtue of having a striking textural difference with neighboring objects. Such regions of the image or a video are referred to as salient regions. Figure 1 presents some images with salient regions marked out. The complexity of the problem can be judged from the fact that there is a large variability of color distribution across the salient regions and the background image. Although humans tend to locate faces in images very easily, this aspect is not discussed in the paper.

Detection of salient regions in images has attracted a large number of research workers [1]- [6], [8] - [10]. A computational model of saliency based spatial attention has been presented in [8]. The saliency maps are computed for luminance, color and orientation (at multiple scales) features by aggregating the information about each location in an image in a bottom-up manner. Ma et al. [10] first transform the image to CIELuv space, divide into small blocks, and then grow a saliency map for each block by using local contrast. As the variation of gray levels in saliency map is not consistent, they propose fuzzy growing concept to extract the attended areas. They have also shown how attended views could be extracted from the saliency map. An attention model based on three attributes (region of interest, attention value, and minimal perceptible size) associated with each attention object has been proposed by Chen et al. [6] to detect human faces and text material from an image. Lui et al. [9] use CRF learning to combine multi-scale contrast, center-surround histogram and color spatial distribution information from the images for detection of salient objects in an image. They also provide an extensive database with ground truths carefully labeled by multiple users.

The neural network model proposed by Bruce et al. [2] -[4] is somewhat close to the human visual system which relies on attention based on information maximization. In their work, they hold the premise that the saliency of visual content may be equated to a measure of the information present locally within a scene as defined by its surrounding patches. The control of focus of attention is stimulus driven and corresponds to a bottom up perceptual process. A correlation exists between visual saliency and fixation behavior

*Corresponding author

ICVGIP '10, December 12-15, 2010, Chennai, India

in human observers which has been demonstrated in [2],[3], [4]. Chalmond et al. [5] have used multiscale features as input to a probabilistic mixture mode for detecting salient regions in remote-sensed images. Recently, Achanta et al. [1] have shown that by using luminance and color maps at three levels of image pyramid and applying contrast determination filters at each stage, results similar to Itti's [8] can be obtained. A K-means algorithm is used on the overall map to segment whole objects.

The common approach used by all the researchers has been to propose a set of parameters which help in highlighting the salient region, and then propose a scheme to tune the parameters by training with a large number of example images. In this paper, a case-based reasoning approach is used to learn the parameters. Case-based reasoning has been applied in various image classification, segmentation and recognition problems especially in medical image interpretation [11] - [14] . In [13], a catalogue based image classifier is built to reduce the number of prototypical images required in medical image interpretation systems. In [12], learning using case-based approach has been carried out on ultra sonic images from non-destructive testing showing defects inside a metallic component. Frucci et al. [7] have presented a case-based reasoning approach for watershed segmentation. The approach is used to select the segmentation parameters involved in the segmentation algorithm by taking into account the features characterizing the current image. Parameters producing the best segmentation for a large number of images are recorded. They illustrate their scheme on images containing faces, animals and natural scenes.

In this paper case-based reasoning (CBR) is used to let the system learn from a diversity of cases presented to it. The proposed scheme is tested on images taken from the large dataset made available in [9]. The dataset contains images with single/multiple objects, varying vastly in size and color combinations. CBR aids in learning the fuzzy parameters for detection of salient regions by taking into account the features characterizing the prominent boundary colors, the color distance between two regions (as per CIELab chart) , their relative spatial variances, relative locations, and the relative sizes of connected components of regions having colors different from the background colors of the current image. A number of images are presented to the system and the parameters which provide the best match to the salient regions marked as ground truth are recorded. These images are grouped to form relevant cases, where each case includes all the images having similar image features, under the assumption that the same parameters will discover similarly good salient regions for all the test images. CBR selects the top most relevant cases and thus helps in reducing the rule-set greatly. Thus CBR not only simplifies the problem of variations across a range of sample images but also returns a set of characteristics defining the given dataset. This can be especially useful in making inferences about the datasets containing similar types of images and can also aid in image retrieval.

The rest of the paper is organized as follows. Section 2 presents the pre-processing while the fuzzy features are explained in section 3. Section 4 describes the case-based reasoning approach. Section 5 presents the analysis of results of the proposed approach and section 6 provides the conclusions.

Table 1: Color numbering

1	Dull Green	6	Pink
2	Green	7	Red
3	Bluish Green	8	Brown
4	Blue	9	Orange
5	Purple	10	Yellow
11	Dull Yellow		

2. PRE-PROCESSING

Color is the most prominent feature of any object. One of the most important requirements of salient region detection is, therefore, to categorize the colors correctly. We make use of CIELab space as the perceptual differences are nearly Euclidean and conform to human perception of color.

This colorspace has been designed in such a way it transforms colors to a representation which is approximately perceptually uniform in the sense that Euclidean distances between different colors in this space correspond roughly to perceived color differences. It also enables us to compute color contrast and luminance contrast wherever applicable. We quantize the CIELab color space into 11 categories and number each color depending on the proximity of colors in the CIELab space as shown in Table 1. These categories are fuzzified to take care of overlap between colors in the CIELab colorspace. Further to reduce computational complexity, we divide the image into blocks of size 10X10 and consider the dominant color of each block. This also helps in smoothing the colors.

3. FUZZY FEATURES FOR SALIENT REGION DETECTION

The features used for salient region detection in this paper include both background characteristics like boundary area and spatial variance as well various salient region characteristics like color distance, spatial variance, size of connected components, intensity components and dominant colors. These are detailed in the following subsections.

3.1 Background detection based on location information

In most of the images, the colors at the boundary of the image also turn out to be the background colors. While for some images, there is only a single color (such as green from an image of grass) all around the boundary, for other images, there are more than one boundary colors. To restrict the number of background colors to a small manageable number, we put a restriction that a color qualifies as a "background" color ONLY if it occupies at least 25% of the total number of 10x10 blocks at the boundary of the image. This automatically limits the number of background colors to 4.

The number of blocks occupied by the specific color at the boundary is assigned a fuzzy membership using the linguistic variables : {small, medium and large}, as shown in Figure 2. Various samples of background images are given to the system to allow it to learn what percentage generally constitutes the background. A few examples are presented in Figure 3 to illustrate single and two background color cases. Intuitively, greater the percentage of area of boundary color, the more likely it is going to be part of the background.

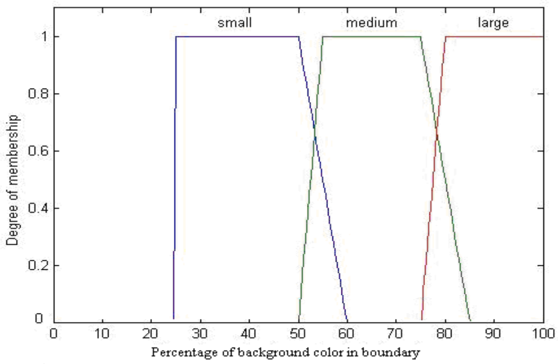

Figure 2: Membership functions for relative area of background

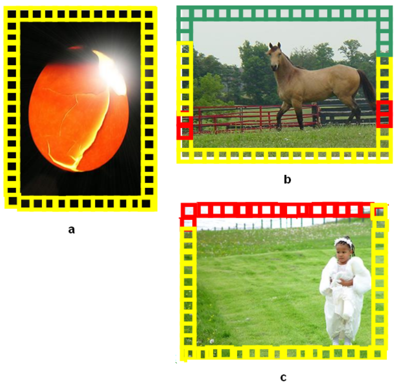

Figure 3: (a)Consists of one boundary color, (b)Consists of two boundary colors (sky , ground), the fence color occupies less than 25% so the color is not included, (c)Consists of two boundary colors.

3.2 Color Distance

The color distance is a measure of nearness of a segment color number C_s to any of the background color numbers C_b in the quantized CIELab space in terms of the color numbers given in Table 1. A particular color could be a potential candidate to be labeled as a salient color if its distance from all background colors is large. The color distance CD is computed as follows:

$$CD = (1/6)|C_s - C_b| \tag{1}$$

Here C_b refers to each of the prominent background colors of the image. The color distance is calculated with respect to each background color of the image. However, one has to take care of the fact that the color numbers are in circular order. For instance, color 1 is nearer to color 2 as well as to color 11 though the CD is different in the two cases. Accordingly, we need to consider two separate cases depending on whether CD 1 or CD >1. For this purpose, CD is fuzzified using fuzzy membership set: {very small, small, medium, large } and separate fuzzy rules are fired for the two cases. The membership functions for the case CD < 1 is shown in Figure 4. It is easy to infer that greater the color distance, better are the chances of this being part of a salient region. For the intensity cases, separate rules are formed on the basis of which white is separated from dark grey or black etc.

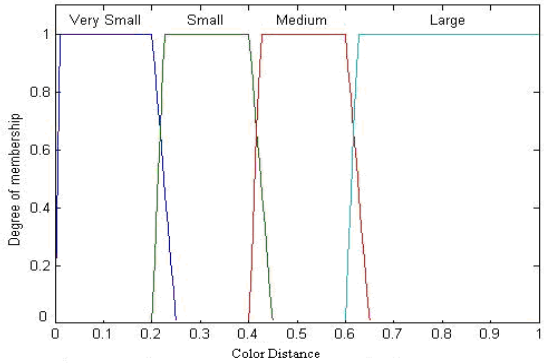

Figure 4: Membership functions for color distance

3.3 Spatial Variance

The image is segmented on the basis of the color assigned to each block. The spatial variance is a good measure of the spatial distribution of colors [9] . Spatial variance is computed both in the vertical and horizontal directions. Assuming region with color c contains N blocks, vertical variance V_{cx} and horizontal variance V_{cy} are computed as shown below:

$$\mathrm{V}_{cx}\text{=}(1/\mathrm{N})\sum_{i=1}^{N}(\mathrm{l}_{ci} - \mu_{cx})^2 \tag{2}$$

$$\mathrm{V}_{cy}\text{=}(1/\mathrm{N})\sum_{i=1}^{N}(\mathrm{l}_{ci} - \mu_{cy})^2 \tag{3}$$

where l_i and t_i are the leftmost and topmost co-ordinates of the i^{th} block and μ_{cx} and μ_{cy} are the means of the color c in the horizontal and vertical direction respectively. The spatial color distribution SV_c is then computed as:

$$\mathrm{SV}_c\text{=} V_{cx} + V_{cy} \tag{4}$$

Once SV_c is computed for all the colors, the average spatial variance percentage AV is obtained as:

$$\mathrm{AV}\text{=}(1/\mathrm{m})\sum_{c=1}^{m}(\mathrm{SV}_c) \tag{5}$$

where m is the number of colors present in the image. This is then used to compute r_c , the relative spatial variance of a particular color using

$r_c = (SV_c \ / \ AV) * 100$ (6)

Spatial variance of any background color is expected to be greater than the average spatial variance of all colors in that image while the spatial variance of any salient color is less than the average spatial variance. This is defined in terms of r_c which is assigned a fuzzy membership value from the set : {very small, small, medium, large } as shown in Figure 6.

3.4 Saliency Color Ordering

Some colors attract the human eye more than other colors. By providing relevant images, the system is made to learn the fuzzy rules which weigh more favorably the red, pink, brown and yellow colors as candidates for salient region over colors such as shades of green and blue when colors from both categories are found to be competing for salient region.

3.5 Using Color Distance and Spatial Variance in CBR

Spatial variance gives the spread of the colors. It is possible that a color has too small a percentage in the background to be a boundary but isn't a salient color either. In Figure 5 (a), yellow is chosen over black before it has smaller spatial variance as compared to black. This is also true for 5 (b) where brown is chosen over the blue stream. Similarly, it is possible that spatial variance is large (because of the object size) but it is chosen because of large color distance 5 (c). 5(d) and (e) show cases where both color distance and spatial variance are used. Various types of such images are fed to the system to allow it to learn cases and fuzzy variables for color and spatial analysis.

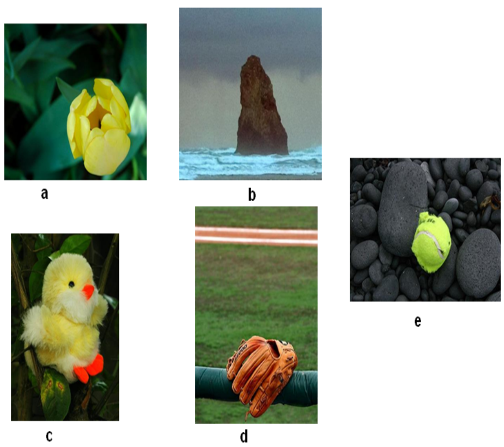

Figure 5: Sample images for color distance and spatial variance

3.6 Intensity Analysis

If the image consists of black, white or grayish shades, a different approach is needed. It is because on the CIELab color chart, the "a" and "b" values for black and white colors are located near the origin and are placed very close to each other. In such cases we need to make use of intensity values (the Luminance component of the CIELab space). We

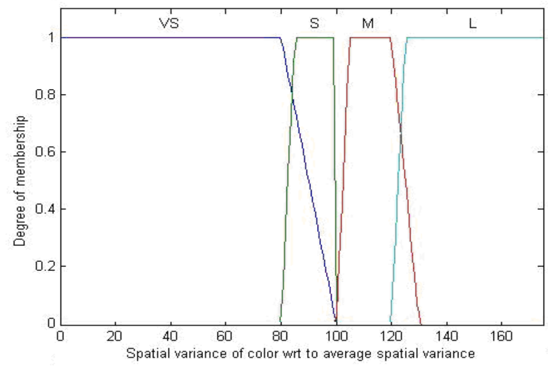

Figure 6: Membership functions for spatial variance

fuzzify the set black, dark gray, light gray, white over the total luminance range of [0...100].

An image containing mostly black and white shades can be handled on the basis of fuzzy intensity rules. If the background contains black/white shade, and the foreground region contains one or more of the 11 colors, then fuzzy rules related to spatial variances as well as rules related to intensity are used. Figure 7 b & c show the use of only intensity rules. Figures 7 a & d show cases where spatial variance rules are also required to segment out the colored portion.

If the background contains one of the 11 colors, while the salient region partially or completely contains black/ white shades, then simply the spatial variance rules could be used to process the image. Figure 8 shows cases where the grey/white parts are selected as salient colors using spatial variance.

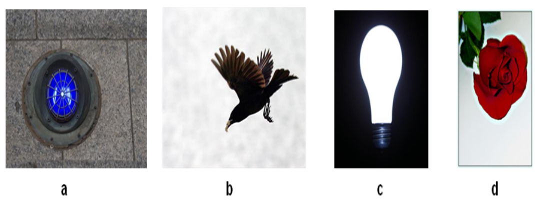

Figure 7: Sample images for cases where background uses intensity rules

Figure 8: Sample images for cases where foreground contains black or white portions

3.7 Connected Component Analysis

One needs to use connected component analysis to connect adjacent segments of the same color. Very often, the salient region is composed of an object having a number of distinct colors. All these colors (which are different from the background colors) will be declared as salient colors. In such a case, it is necessary to aggregate adjacent salient color blocks. Some samples of cases where multiple color components are aggregated to form a single salient object are shown in Figure 9.

Figure 9: Aggregation in connected component analysis examples

The image is binarized into background colors and foreground (salient colors). All the background colors are assigned value 1 and all the salient colors are assigned value 0. A connected component analysis helps in generating the salient region. If the blocks of salient colors are not adjacent to each other, then the identification of salient region is based on the size and location of the connected components. The connected component size relative to the total image size, rcc is fuzzified with the set : {small, medium, large, very large}, as shown in Figure 11. The system is given cases containing multiple objects with the salient one marked as shown in Figure 10, to allow it to learn the properties of the salient components.

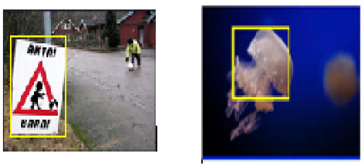

Figure 10: Selection in connected component analysis examples

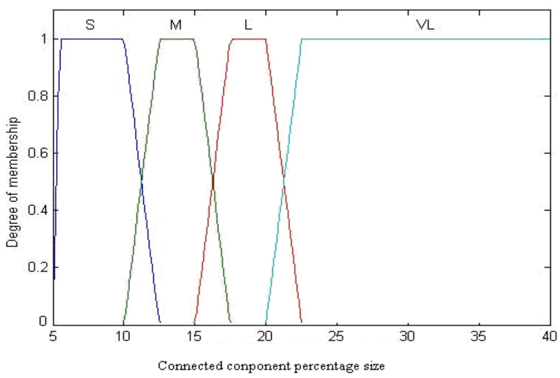

Figure 11: Membership functions for connected component analysis

4. CASE-BASED REASONING

As discussed in the previous sections, all the parameters used for salient region detection typically are fuzzified with 4 linguistic variables. Thus the number of possible combinations of these features is more than 1000. We use case based reasoning to select the topmost combinations. Given a set of training images, with the salient regions marked with a rectangle [9], each case represents a combination derived from an image.

To start with, the number of background colors and the salient region colors are ascertained for each image. The top 2 colors with maximum area inside the background and foreground region each are selected. Then, for each of these combinations, cases containing information on intensity, salient and background color variances, distance between salient and background color and the relative size of the connected components are generated. New cases are created incrementally as more and more images are input. This is done by calculating the similarity of the new case with respect to the already created cases. Initially, the membership of each new case is calculated. This is done by taking the mean of the maximum memberships for each feature.This gives the similarity of this case (c), with respect to any existing case which matches it in terms of the number of features having maximum membership variables as shown below:

$$\text{sim(c)}= \sum_{i=1}^{n} (\forall_j \max(\mu_{cij}))/\text{n} \tag{7}$$

where n gives the number of features and j refers to the linguistic variables.

To increase the stability of the system, a new case is added to the set only after 5 instances of the same have been repeated. Many examples of images of various complex cases are fed to the system to allow it to include as many cases as possible. For simple cases, one feature is predominant over other features, as discussed in the previous sections. However, as the number of colors in the image increases, so does the complexity of the cases. This also happens when one feature is more dominant but not adequate. This includes intensity images, images with large color distances, images with many colors in foreground but each having less variance than background. However, no single feature is more important than the other, instead this depends on the membership values for the various features of that case. As we increase the complexity further, images with multiple objects are analyzed. Some complex cases are shown in Figure 12.

Some examples of compound cases for above images are given below:

IF the boundary area of the background color is “small”
AND the spatial variance of the background color is “medium”
AND the spatial variance of the salient color is “small”
AND the color distance is “large”
THEN the region is salient

IF the boundary area of the background color is “medium”
AND the spatial variance of the background color is “medium”
AND the spatial variance of the salient color is “very small”
AND the color distance is “medium”

Table 2: Result analysis

Datasets for training	Datasets for testing	Overall good	Perfect	Very Good	Partial overlap
1	2,3	371 (74.2%)	294(58.8%)	77(15.4%)	129(21.8%)
2	1,3	375 (75.0%)	282(56.4%)	93(18.6%)	125(19.8%)
3	1,2	379 (75.8%)	286(57.2%)	93(18.6%)	121(24.2%)

AND the size of connected component is "small"
THEN the region is salient

IF all the background colors belong to intensity shades
AND the spatial variance of the salient color is "small"
AND the connected component size is "medium"
THEN the region is salient.

For a set of 500 such images, approximately 35-40 cases are generated. These are then fed as input during testing and similarity analysis is carried out. For testing purposes, background colors are initially selected based on percentage of area. Cases are incrementally added by analyzing the background characteristics of selected colors and adding the foreground characteristics based on maximum similarity with existing case. For each foreground color under consideration, similarity is computed taking into account each of the features. This is done by taking maximum memberships of features for color under consideration.

5. RESULTS AND ANALYSIS

We have used 1500 images from the data set indicated in the paper by Liu [9] for our study. The dataset also provides separately the coordinates of the rectangle R_{gt} [ground truth] bounding the salient object/region for each image. We partitioned the images into 3 subsets 1, 2 and 3 of equal size to carry out the testing. Once the salient region is identified by our approach for a given image, a bounding rectangle R_{our} is drawn to mark out the area. This area of our rectangle A_{our} is compared with the area of ground truth rectangle A_{gt} to compute the accuracy as follows:

$$\text{Accuracy}=(A_{gt}\cap A_{our})/A_{gt} \quad (8)$$

The results were categorized into "Perfect", "Very Good" and "Partial overlap" classes based on the percentage of area of match between intersection of our rectangle with the ground truth and the area of ground truth. For "Perfect" class, the threshold is taken as 0.8 and for "Very good" it is taken as 0.6. It is found that in almost all cases the system is able to correctly identify the salient region in the image. The results are shown in Table 2. The average performance of our proposed approach (considering the "Perfect" and "Very good cases") is 75%. For cases where the area of the salient region is greater than the ground truth, a separate evaluation criteria was employed in which the accuracy is decreased as co-ordinates of the salient region approach the boundaries of the image. This was done to take into account the possibility of our rectangle actually covering the entire image. Thus such cases, though rare come under "partial overlap" as expected.

A small sample of the results of our approach is shown in Figures 13 and 14 to compare the results with those obtained by other researchers. Since the dataset (for training and testing) is different for all authors, so only a goodness of method in terms of quantitative comparisons has been presented. The results of Bruce were obtained by running the MATLAB code given on his website (on the link http://wwwsop.inria.fr/members/Neil.Bruce/SOURCECODE) and selecting the parameter values as resizesize = 0.5 convolve=1, basisname= '21jade950.mat' and output =1.

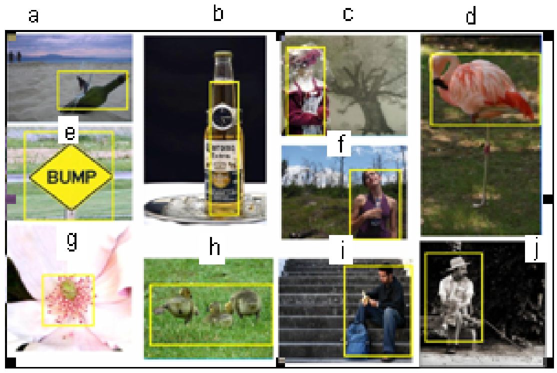

Figure 12: Some complex cases

6. CONCLUSIONS

In this paper we have presented a case-based reasoning approach to the problem of salient region detection in an image.Since color is the most dominant feature in an image, we have focused on that in this paper. However we plan to include texture features also in our future work. A large number of different types of images are fed to the system and various cases are learnt. CBR selects the top most relevant cases and thus also helps in reducing the rule-set greatly. It has been shown that the approach is effective in outlining the salient region in an image, and compares favorably with other published results.

7. REFERENCES

[1] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk. Salient region detection and segmentation. In *Int. conf. Computer Vision Syst. (ICVS '08)*, volume 5008, pages 66–75, 2008.

[2] N. Bruce. Features that draw visual attention: An information theoretic perspective. *Neurocomputing*, 65-66:125–133, 2005.

[3] N. D. B. Bruce and J. K. Tsotsos. Saliency based on information maximization. *Advances in Neural Information Processing Systems*, 18:155–162, 2006.

[4] N. D. B. Bruce and J. K. Tsotsos. Saliency, attention, and visual search, an information theoretic approach. *Journal of Vision*, 9:3:1–24, 2009.

[5] B. Chalmond, B. Francesconi, and S. Herbin. Using hidden scale for salient object detection. *IEEE Transactions on Image Processing*, 15(9):2644–2656, 2006.

[6] L. Chen, X. Fan, W. Ma, H. Zhang, and H. Zhou. A visual attention model for adapting images on small displays. *9th Int. Conference on Multimedia Modeling (MMM03)*, pages 353–364, 2003.

[7] M. Frucci, P. Perner, and G. S. di Baja. Watershed segmentation via case-based reasoning. In *The 2nd Int. Conf. on Advances in Brain, Vision and Artificial Intelligence*, pages 244–253, 2007.

[8] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions PAMI*, 20(11):1254–1259, 1998.

[9] T. Liu, J. Sun, J. Zheng, X. Tang, and H. Shum. Learning to detect a salient object. *Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[10] M. Ma and H. Zhang. Contrast-based image attention analysis by using fuzzy growing. *Proc. of the 11th ACM Int. conference on Multimedia*, 2003.

[11] M. Oliver, S. Naidu, and A. Koronios. Utilizing case-based reasoning and multimedia to enhance clinical decision making of novice practitioners: Product implementation and evaluation. In *The 17th Annual Conf. of the Australian Society for Computers in Learning in Tertiary Education(ASCILITE 2000)*, 2000.

[12] P. Perner. Different learning strategies in a case-based reasoning system for image interpretation. In *ECCBR 98. Lecture Notes in Artificial Intelligence*, 1998.

[13] P. Perner. Prototype-based classification. *Applied Intelligence*, 2008.

[14] P. Perner, B. C. Pierce, and D. N. Turner. Why case-based reasoning is attractive for image interpretation. *Case Basis-Reasoning,Research and Developments, Springer-Verlag*, LNAI 2080, 2001.

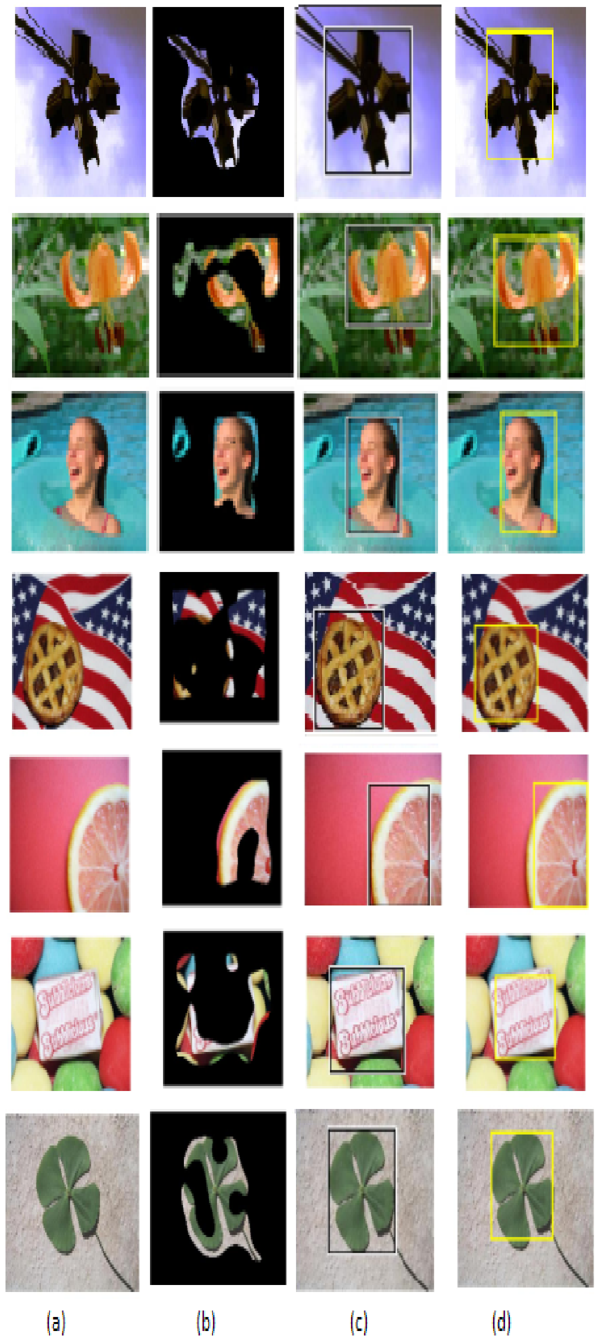

Figure 13: Comparisons with other approaches: (a) Original image (b) Bruce - non salient region is suppressed (c) Liu and (d) our approach.

(a) (b) (c) (d) (e) (f)

Figure 14: Comparisons with other results (a) Liu's Ground Truth (b) Ma (c) Itti (d) Liu (e)Our approach (f) Bruce