Perlustration of Detecting Saliency in Digital Images

Arvinda Sharma
M.Tech Scholar: Computer Science & Engineering Dept
Global Institutes of Management& Emerging Technology
Amritsar,Punjab, India.

Dr. Rajiv Mahajan
Professor: Computer Science & Engineering Dept
Global Institutes of Management& Emerging Technology
Amritsar, Punjab, India

Abstract:-This paper presents the detection of salient objects in digital images. Saliency is focus of attention in a scene and Salient Object Detection will confined to discover important objects in the digital image or in the video. Discovering Objects in images is valuable for face detection, detecting objects in CCTV camera etc. This paper uses patch based approach ***to detect the salient object through its approximate contour. It also use the concept of Random forest to measure patch rarities and compute similarities among patches. Then on basis of rarity map we get approximate contour. Then apply active contour models and do further the segmentation to refine the results.***

Keywords- object detection, random forest, and active contour model.

I. INTRODUCTION

Visual saliency deals [2] with low-level responses to the human vision system that attracts the attention of person who is viewing it .It is widely used in variety of applications such as range of computer vision, multimedia, and graphics applications, such as[2] automatic object detection , image extraction, video summarization, adaptive image compression , and content-aware image or video resizing. For discovering the objects [3] there are on the least three ways of doing it as: parts-based an object is described as a special spatial arrangement of parts, patch-based in which we classify each rectangular image region as object or background, and region-based where a region of the image is segmented from the background and is explained as set of features .Here we use a patch-based approach. For objects with well-defined shapes a patch usually contains the full object and a small portion of the background.

Object detection [1] is a process of discovering the important objects in the digital images and videos. Detecting such an object is preprocessing stair of segmentation. There are large numbers of methods that segment the multiple objects the scene. Each method has its own benefits and drawbacks. It will discount the effects of background from image and detect the salient objects and saliency is focus of attention in a scene. Saliency maps [9] compute saliency value for each pixel. The salient object is always different from its neighboring perspective, placed near the center of the image and has a sharp closed margin.

Problems in discovering the important entities arise [1] hasty object motion, changing appearance patterns of an object and a background, object-to-object occlusions and object-to-background occlusions. Image segmentation has been the subject of active research in computer vision and image processing. A large body of work on geometric active contours (i.e.) active contours implemented via level set methods, has been proposed to address a wide range of image segmentation problems.

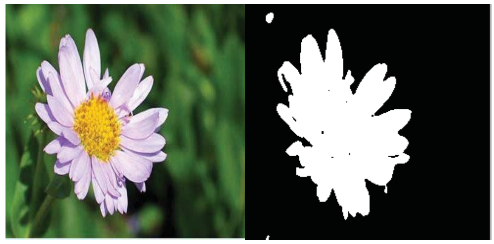


Fig1: Salient Object Detection

II. DETECTING CONTOUR OF SALIENT OBJECT

It consists of two steps: Random Forest Construction and Active Contour models. Contours are the borders of an image. These reflect the shape of an object and help an individual to recognize the object. Contour detection forms the major part of object detection. Object Detection is significant preprocessing step of digital image processing. It [6] is desired to extract regions of interest before further processing and analyzing. Salient region extraction is an important method to detect the region of interest. It is operated to select a subset of the available sensory information before further processing and sample the most relevant features, most likely to reduce the complexity of scene analysis. Due to such superiority, salient region extraction has been adopted in widely applications, such as image compression, distinctive objects detection, image segmentation etc.

A. RANDOM FOREST

It is a first step that is used in processing the digital images. These are based on the use of patches. Patch is a small squared image that is retrieved from the image. Patches are uniquely identified by the certain set of characteristics [10] such as: horizontal location specified by x-coordinate in digital image, vertical location as specified by y-coordinate in image, and size. For a given location and size, patch can be extracted by simply formative which image pixels are located within that image area.

Patches [11] fit in to the class of limited character which means that they portray properties of a particular region of an image. In disparity to that, global description provides information about the image as whole. Classic global features used in object recognition or in the field of image retrieval are for example texture information, the colour distribution, or simply all pixels of the image .As global features account for whole images, they tend to become inadequate if only a small portion of the image is relevant. In the domain of object recognition, it is often the case that images have to be classified based on objects which make up only a very limited part of the image. Local features like patches are better suited for complex images, because they represent restricted regions of the image.

Advantageous properties of local [11] description are:

(a) Inherent translation invariance: using patches it does not matter where in the image ascertain object is shown. An object may be successfully detected even if it uses to be located at different locations in different images.
(b) Robustness to object variance and occlusion: unless the object is very small, patches don't capture it as a whole but capture its parts. Therefore, an object may be successfully.

Generally, a patch is extracted [11] around each pixel of an image at a given scale. Although this seems to be a good idea, as we can be sure not to discard any parts of the image, it turns out that the resulting amount of patches is not feasible for the methods we apply. Therefore, to use a moderate number of patches per image, a subset of points needs to be chosen around which patches are extracted. In the following, different methods to determine such a subset of extraction points are presented: Grid points: choosing grid points is a trivial way of determining extraction points. A regular grid is projected onto the image which immediately gives the extraction points.
With this method the extraction points are distributed uniformly over the image, which means that they are contained in homogeneous regions and in regions with high variance.

The size of the gaps between the patches or whether they overlap depends on the patch size and on the distance between the extraction points which is determined by the size of the grid. In particular, the grid can be chosen such that the patches are aligned.

i. Random points: As the name describes, here points are chosen randomly. These points behave similar to the grid extraction points, as they prefer neither homogeneous nor high-variance regions, but are distributed over the whole image.
ii. Interest points: Finding interest points requires more complicated methods; the degree of interest goes along with the variance within the image, therefore interest points are usually found in regions of high variance.

The type of pre-processing that can be applied to patches is brightness normalization. As long as not artificial images or images which have already been normalized beforehand are used, it is quite normal that the objects appear in different images under different lighting conditions. Possible methods for normalizing complete images are:

a) Minimum or maximum spreading: the brightest and darkest pixels in an image are taken into account. Such spreading shows that full range of brightness values are used in the work.
b) Histogram Equalizations: In this all possible values of the image pixel are seen.

The problem with both the above methodology is that they normalize the entire image without focussing on the objects of interest. Therefore, the brightness divergence of these objects may be retained. We assume that the recognition performance is best if the objects are normalized rather than the whole images. So instead of the whole images, the patches have to be normalized. When the image is split into the patches, these are arranged to form the tree, and collection of trees is known as random forest.

Random forest [9] is an ensemble classifier consisted of a number of decision trees. These can be used for classification and regression .Accuracy and variable important information is also provided with the results. Random forest run efficiently on larger database .It computes proximities between in pair of cases that can be used in clustering, locating outliers. It offers an experimental method for detecting variable interactions. Since the performance of the trees is highly related to the connection among each model in it, the trees are often constructed with some randomization and randomization comes from two points: Sub sampling the training data and each tree are grown with different data, for each internal node, selecting some attributes for split. Besides, each internal node contains a best split of training data.

In the work, first rarity maps are applied to the image using random forest. These will classify the patches of an image belonging to one class. Rarity map measures the extent to which image contents are related to surroundings. It will result into the saliency map, which means measuring the saliency value for each pixel. The equation used is[4]:

$$t_n(s_n : h_1, h_2) = \begin{cases} p_i \epsilon s_l, & if \;\; d_i(h_1, h_2) \leq \theta_{h1,h2} \\ p_i \epsilon s_r, & otherwise \end{cases}$$

Where s_l and s_r are the patch sets contained in node's left and right child, $\theta_{h1,h2} = \frac{1}{|s_n|} \sum \forall p_i \epsilon s_n d_i(h_1, h_2)$, $|s_n|$ is the cardinality of s_n. After the forest is built, we use it to measure the rarities of patches and compute similarities among them, and then detect the salient region through it as [4]:

$$cS(p_i) = \frac{1}{\sum_{i=0}^{n} |L_k|} \cdot w(x_i, x_c)$$

The above equation is used to evaluate the rarity of p_i, where $|L_k|$ is the number of patches contained in L_k. $w(x_i, x_c)$.

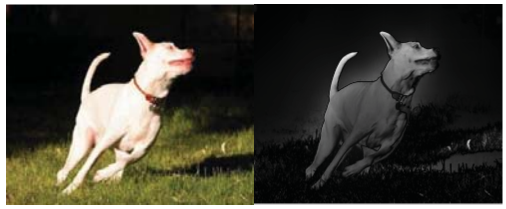


Fig2: Saliency detection

B. ACTIVE CONTOUR MODELS

In computer vision, recognizing objects often depends on identifying particular shapes in an image. This is a difficult task and a central problem in vision. The area of shape representation is concerned with finding ways of describing shape that are sufficiently general to be useful for a range of objects, whilst at the same time allowing their computation from image data, and facilitating comparisons of similar shapes. Both 2-D and 3-D shapes need to be modeled, and a very wide range of ideas can be found in the literature. One of the best known of 3-D representations is Marr's hierarchical approach based on generalized cylinders.

The snakes [12] are the active contour models. These have a very simple form. They consist of a set of control points, effectively connected by straight lines. Each control point has a position, given by x, y coordinates in the image, and a snake is entirely specified by the number and coordinates of its control points. Adjustments to the snakes are made by moving the control points. The control points were not shown separately in the example above, for in future examples will be picked out in a different colour. All the snakes used here form closed loops in the image, though this is not necessarily true of snakes in general.

One of the chief virtues of snake representations [12] is that it is possible to specify a wide range of snake properties, through a function called the energy by analogy with physical systems. A program controlling a snake causes it to evolve so as to reduce its energy, so by specifying an appropriate energy function, we can make a snake that will evolve to have particular properties, such as smoothness. The energy function for a snake is in two parts, the internal and external energies. Thus energy of a snake=internal energy + external energy, internal energy is the part that depends on intrinsic properties of the snake, such as its length or curvature. The external energy depends on factors such as image structure, and particular constraints the user has imposed.

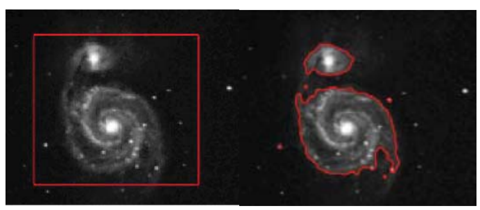
A snake being used for image analysis attempts to minimise its total energy, which is the sum of the internal and external energies. When energies are added their associated forces add too. In the first snake example in this file, finding the clock face, the snake had an elastic energy and image energy, but no bending energy. The parameter gamma was negative, so the image energy drove the snake towards darker parts of the image. The snake contracted under the elastic force until this was balanced everywhere by the image force, and it came to rest shrink-wrapped round the clock face. To explore this balance of forces, it is convenient to be able to initialise a snake on any part of the image. These are connectivity preserving methods that are applied to image segmentation problems. These have been used for image segmentation and in boundary tracking. It is a rubber band [10] of any shape that is deform with time and try to get as close as possible to object contour. These are the energy minimizations [10] that will consider only the pixels or patches that lie on the boundary of object. Active contour models (ACMs, also called snakes or deformable models) have been widely used in image processing and computer vision applications, especially for image segmentation.

The ACM [5] moves the precise parametric curves to pull out objects in images. However, the parametric active contour models has some intrinsic drawbacks, such as its difficulty in handling topological changes and its dependency of parameterization .The level set method later proposed implicitly represents the curve by the zero level of a high dimensional function, and it significantly improves ACM by being free of these drawbacks. The level set methods can be categorized into partial differential equation based ones and variation ones.The level set evolution of partial differential equation -based level set methods is directly derived from the geometric consideration of the motion equations, which can be used to implement most of the parametric. Moreover, the variation LSM can be easily converted into partial differential equation -based LSM by changing slightly the LSE equation while keeping the final steady state solution unchanged.

Active Contour model [1] can be categorized into two types: edge-based models and region-based models. Region-based Active Contour model have many advantages over edge-based ones. First, region-based models utilize the statistical information inside and outside the contour to control the evolution, which are less sensitive to noise and have better performance for images with weak edges or without edges. Second, they are significantly less sensitive to the location of initial contour and then can efficiently detect the exterior and interior boundaries simultaneously. Through level set Active Contour method, we discover the salient object according to the detected, inside and outside, respectively. The formal definition can be written as [5]:

$$\mathbf{p}_i \in \begin{cases} S_{in}, & \text{if } \frac{|\mathbf{p}_i \cap C|}{r^2} > \lambda, \\ S_{out}, & \text{otherwise,} \end{cases}$$

Where $|\mathbf{p}_i \cap C|$ denotes the number of pixels of contained inside the extracted contour λ is a constant, and its value varies between 0 and 1 according to the dataset. The p_i *and* p_j are similar patches if they belong to leaf node and then we measure the contrasts between the inner patches and the outer patches, aiming to suppress the patches of the inner part similar to the outer patches while highlight the outer patches similar to the inner patches. Through the contrasts among patches, our method highlights the whole object uniformly. Its effect over the image is as shown:



(a) Initial contour (b) after ACM moves

Fig3: Active Contour model

C. SEGMENTATION

The function of segmentation [8] is to dividing an image into its basic and disjoint sub-regions, which are identical according to their property, e.g. intensity, color, and quality. Segmentation algorithms are usually based on either discontinuity with sub regions, i.e. edges, or equality within a sub-region, though there are a few segmentation algorithms depends on both discontinuity and equality. Finally we refine the local map using graph-cut based segmentation as the patch-based map is the rough estimation of saliency

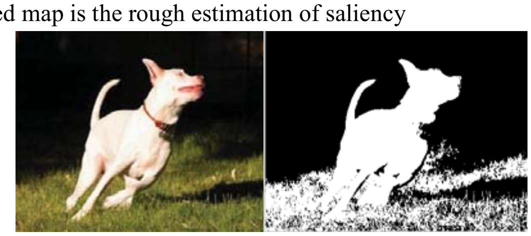


Fig4: Segmentation

III. RESULTS

This section contains the experimental results.

Fig5: Input Image

Fig5 illustrates the Input Image over which the various operations are applied to detect the salient object via random forest.

Fig6 shows the saliency value if inputted image

Fig6: Saliency value of image

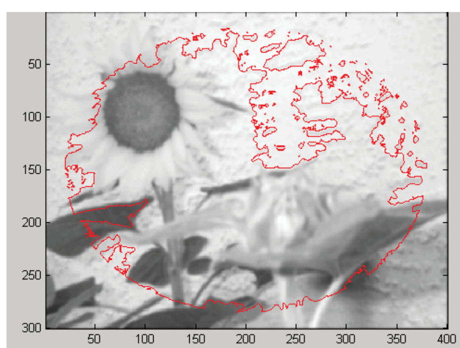


Fig7: Active Contour Models

Fig8: finally Segmented image

IV. PERFORMANCE ANALYSIS

a) F-measure: It is the harmonic connote of precision and recall, while Precision is the percentage of selected items that are correct and recall is percentage of correct items that are selected.
b) Precision and Recall: Precision also called positive predictive value. It is the fraction of retrieved instances that are relevant, Recall is also known as sensitivity. It is the fraction of relevant instances that are retrieved.
c) Geometric accuracy: It is defined as the extent to which we are close to the actual value.

S. no	Precision	Recall	f-measure	Geometric accuracy
1.jpg	0.9999	0.9999	99.9945	1.0000
2.jpg	1.0000	1.0000	99.9960	1.0000
3.jpg	1.0000	1.0000	99.9979	1.0000
4.jpg	1.0000	1.0000	99.9974	1.0000
5.jpg	1.0000	1.0000	99.9983	1.0000
6.jpg	0.9999	0.9999	99.9923	1.0000
7.jpg	1.0000	1.0000	99.9979	1.0000

Table1: Evaluate Precision Recall, F-measure and geometric accuracy over multiple images

V. CONCLUSION

Here we introduce a method to evaluate the salient objects from their contours. This technique is simple and does not use any edge detector. It uses random forest and active contour model for their working. It works well not only for single object but for multiple objects also. It performs well then any other methods. In future we can apply level set free active contour model to level set based model.

REFERENCES

[1] Arvinda Sharma ”A scrutiny of detecting salient object in digital images”. International Journal of advanced research in computer science and software engineering,2014 IJARCSSE in Volume 4,Issue 5.

[2] Mai, Long, Yuzhen Niu, and Feng Liu. "Saliency aggregation: A data-driven approach." Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013.

[3] Murphy, Kevin, Antonio Torralba, and William Freeman. "Using the forest to see the trees: a graphical model relating features, objects and scenes." Advances in neural information processing systems 16 (2003): 1499-1506.

[4] Du, Shuze, and Shifeng Chen. "Salient Object Detection via Random Forest." (2013): 1-1

[5] Zhang, Kaihua, et al. "Reinitialization-free level set evolution via reaction diffusion." Image Processing, IEEE Transactions on 22.1 (2013): 258-271.

[6] Madaan, Er Tanvi, and Er Himanshu Sharma. "Object Detection in Remote Sensing Images: A Review." International Journal of Scientific and Research Publications 2.6 (2012).

[7] Wang, Peng, et al. "Salient object detection for searched web images via global saliency." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.

[8] Baswaraj, Dr A. Govardhan, and D. P Premchand. "Active contours and image segmentation: The current state of the art." Global Journal of Computer Science and Technology 12.11-F (2012).

[9] Jiang, Huaizu, et al. "Automatic salient object segmentation based on context and shape prior." BMVC.Vol. 3.No. 4. 2011.

[10] Cheng, Ming-Ming, et al. "Global contrast based salient region detection."Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on.IEEE, 2011.

[11] Hegerath, Andre, Ing H. Ney, and T. Seidl. Patch-based object recognition. Diss. Diploma thesis, Human Language Technology and Pattern Recognition Group, RWTH Aachen University, Aachen, Germany, 2006.

[12] http://www.sussex.ac.uk/Users/davidy/teachvision/vision7.html

[13] Kass, Michael, Andrew Wit kin, and DemetriTerzopoulos. "Snakes: Active contour models." International journal of computer vision 1.4 (1988): 321-331

[14] Xie, Xianghua, and MajidMirmehdi. "Initialization-Free Active Contour Segmentation." Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, 2010.

[15] Yan, Qiong, et al. "Hierarchical Saliency Detection." CVPR, 2013.