Multiple Object Localization Using Saliency Detection

Mentor: Dr. Shanmuganathan Raman

Group Members: Arun Balajee Vasudevan

Srikanth Muralidharan

Shiva Pratheek Ch

MULTIPLE OBJECT EXTRACTION USING SALIENCY DETECTION

ABSTRACT

This paper describes a novel approach for extraction of multiple salient objects from a set of images as well as video scenes. In the proposed approach, the repetitive and the non-repetitive object from the scene are initially extracted using the computed feature vector value in these regions. We then use two independent approaches for object extraction. One using superpixels and the other way is through use of active contours. The approach is unique in its way of approach for extracting objects from scenes containing multiple objects by using saliency map.

INTRODUCTION

In today's world there are large numbers of digital pictures available which contain both redundant and useful information. The useful information in most cases is in form of objects which may be static, repetitive or non-repetitive. The human eye is a powerful entity which can easily classify the useful information from the redundant information based on spatial and temporal features. Many researchers throughout the world are trying to implement algorithms which can imitate the human eye's perception of objects effectively. These algorithms if implemented efficiently find uses in various sectors of society.

This approach involves the use of visual saliency, active contours and superpixel algorithm to achieve the desired outcome. The reason for choosing visual saliency for object extraction is that it closely resembles the perception of human eye. There are many saliency algorithms existing and most of them use centre versus surround approach. The primary role of visual saliency algorithm is to detect the objects which stand out with respect to background, which includes the interesting information the user is searching. Therefore, many algorithms apart from visual saliency have been developed which can extract interesting objects from a scene.

RELATED WORK

Since the development of image processing especially the field of scene classification, many approaches have been developed for selective extraction of objects. One of the latest automated aproach for object extraction was given by Yu, LI, Tian and Huang in their paper of object extraction using complimentary saliency maps. Similarly, an attempt is made to classify events in static images by integrating scene and object categorizations. [Li Fei-Fei CVPR-2009]. The algorithm classifies the events in the input image and also provides semantic labelling for the objects and the scenes present within the image. Apart from object extraction, detection of salient regions finds usefulness in object segmentation, adaptive compression and object recognition. Various methods are developed for saliency detection in an image.
Superpixels are obtained by over-segmenting an image. The result obtained through over-segmentation clearly demarcates the object edges from the background or other objects present in the scene. This approach has proven to be a major development in the field of image and video analysis, which can be seen in many state of the art algorithms in different fields which have superpixel approach as one of the primary factors for its efficiency. Active contour is an efficient way of outlining an object from a noisy image.

PROPOSED APPROACH

We formulate an approach to extract multiple stationary object regions separately from the scene. The saliency map of an image discriminates salient foreground objects present in the non-interesting static background. The regions in saliency map corresponding to the salient object will be having higher intensity value than the regions corresponding to background. We used context aware saliency which aims at detecting the image regions that represent scene rather than identifying fixation points or dominant objects. We use two separate methods for sub-image extraction from

saliency image. In case, the saliency map is having a different size from the original image, we resize it to the size of original image.

SUPERPIXEL BASED SEGMENTATION

The first method involves corresponding superpixel map extraction from saliency map. We initialize number of distinct segments, typically to a value between 10 and 20. After segmenting into this number of regions, we take the average intensity value pixels of each superpixel. We further refine the number of interesting superpixels or sub-regions by forcing the accumulative superpixel intensity values to zero if the average intensity values are below certain threshold. If the average intensity of the whole saliency is higher as in the case of dense scene, we keep this threshold to be the average intensity value of saliency map. Having a higher multiple (1.5X) of average saliency map intensity as threshold will result in elimination of detection of certain salient object regions. In the cases where average intensity value is less (a sparse scene), we keep the threshold to be a higher multiple (2X) of average saliency map intensity value. Lower threshold value leads to getting many redundant sub-images in case of sparse scenes. Generally, natural objects will have sparse distribution of salient objects and urban, artificial scenes will have dense salient objects.

To illustrate the effect of Intensity threshold based redundancy removal in multiple salient objects extraction, Consider a dense scene as shown in the Fig 1.a. The saliency map corresponding to the scene was obtained using Context aware Saliency algorithm [cite]. The average intensity of the saliency map is 146. The segmented superpixel map of Saliency Map with 20 Superpixels is shown in Fig 1.c. The superpixel map is thresholded at two values: First at 0.75 and the second at 1.50 times the average intensity value, keeping other parameters constant. The 80X80 bounding box is then used to extract superpixels with non-zero intensity values. Fig 1.d and Fig 1.e shows the set of salient regions extracted using the first and second threshold values. Here we note that the scene contains different salient objects like Buildings with distinct salient regions, cars, Vans and a distant pillar. Setting a lower threshold value extracts these salient regions separately without any redundancy as shown in Fig. 1.d. As shown in Fig 1.e, when the Thresholding value is high, the number of salient objects extracted will be less than actual number of salient objects present in the scene. Therefore, we lose out on potentially important salient region by keeping the threshold intensity value high in case of dense scenes.

Fig 1 (a) An urban street scene, Saliency map corresponding to the scene, Superpixels Image of Saliency map

Fig 1(c) Thresholding the saliency map at 0.75 times average saliency map intensity value

Fig 1(d) Thresholding the saliency map at 1.5 times the average saliency map intensity value

Consider another scene which is sparse as shown in the Fig 2.a. The saliency map corresponding to the scene was obtained using Context aware Saliency algorithm [cite]. The average intensity of the saliency map is 58. The segmented superpixel map of Saliency Map with 20 Superpixels is shown in Fig 2.c. The superpixel map is thresholded at two intensity values: First at 0.5(= 29) and the second at 2.5(= 145) times the average intensity value, keeping other parameters constant. The 80X80 bounding box is then used to extract superpixels with non-zero intensity values. Fig 2.d and Fig 2.e shows the set of salient regions extracted using the first and second threshold values. Here we note that the scene contains sparsely distributed salient objects. Setting a lower threshold value extracts these salient regions separately as shown in Fig. 2.d. As shown in Fig 2.e, the salient sub-images obtained at higher threshold of intensity value. As shown in the figures, we observe that salient objects extracted with higher and lower intensity threshold values are same. We can also observe that there is redundancy extracting the salient objects. For example, as shown in Fig. 2.d, the sub-image is extracted around the same man in two of the images extracted. There is no difference in redundancy removal in case of a higher value of intensity threshold.

Fig 2(a) A sparse coast scene, Saliency map corresponding to the scene, Superpixels Image of Saliency map

Fig 2(d) Result of Thresholding the saliency map at 0.5 times the average saliency map intensity value

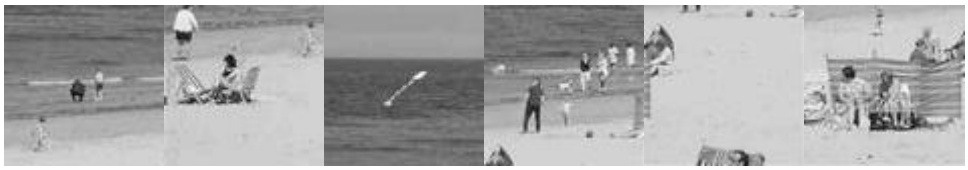

Fig 2(e) Result of Thresholding the saliency map at 2.5 times the average saliency map intensity value

We also use a distance constraint to decrease number of redundant images. In this case, we use half perimeter distance between centres of each superpixel as the distance measure. Let C_x, C_y denote x and y co-ordinates of the centre of superpixels and d_{ij} denote the distance between superpixels i, j. The distance d_{ij} is given by the equation

$$d_{ij} = |C_x(i) - C_x(j)| + |C_y(i) - C_y(j)|$$

$$C_x(k) = \frac{1}{N(k)} \sum_{m \in k} x(m) \quad C_y(k) = \frac{1}{N(k)} \sum_{m \in k} y(m)$$

Here $N(i)$ denotes number of pixels inside superpixel k. Value of d_{ij} if higher than a threshold makes the two sub-regions to consider as distinct salient objects. The value of threshold is decided by the average of intensity value of saliency map. For a high value of average of intensity value of

saliency map, threshold will be set a lesser value (typically less than 10 pixels, sometimes we relax the constraint), and for a low value, we set the threshold value higher (50 pixels or more, depending on the average value). This results in a single sub-image from this region and avoiding the redundancy. A sub-image is finally extracted centred at each superpixel which follows the constraints.

The main advantage with the Superpixel based approach is that we remove redundancy in extracting the objects. Also, we can use this method to classify the scene into numerous categories (dense, sparse or Natural/Urban etc.).

ACTIVE CONTOUR

The other method involves extracting multiple objects using active contours using saliency map. Frequency tuned saliency region detection begets saliency map with well-defined boundaries of salient objects. These boundaries are well used to segment the objects. The skeleton of the objects are obtained by application of active contours on the saliency map as shown in the figure.

Objects can occupy any part of the scene for which we choose the initial contour for active contour to cover the entire scene. We choose small circles placed in the entire scene to be the contour for salient objects are probably small in a natural scene. Albeit the image boundaries are smooth and noisy, location of boundaries are well detected by active contours. The active contour gives a binary image of the scene highlighting the objects. We use the connected component labelling where subsets of connected components are uniquely labelled which helps in object extraction. We make use of the bounding box to extract out object parts from the scene frame.

Active contour has the advantage over the superpixel in the recognition of large number of objects in the scene while we make a bound for the superpixel with the initialisation of the number of divisions. It also gives the appropriate shape of the object.

RESULTS

In order to see the performance of proposed approach, we took a collection of images of varied complex scenes such as streets, kitchen, forest, coast, etc.

CHALLENGES

Active contour on the saliency map fail in the case of the occluded objects as it operates on the intensity. In the case of dense objects scene, algorithm fails to extract multiple objects separately though it manages to acquire the group as a single object.

CONCLUSION

This paper proposes two different approaches in the extraction of objects from a scene. Multiple distributed objects from the complex scenes are extracted using context aware saliency detection and superpixel. In this method, both the minimum distance threshold and the average intensity threshold depend on the scene. For a dense scene with close objects, both the thresholds should be lower. For a scene with locally concentrated salient objects, the only constraint is that the Minimum distance threshold should be lesser and performs equally well in major ranges of intensity threshold constraints. For a sparse scene with scattered objects, both the thresholds should be set high. Setting the appropriate values result in efficient extraction of salient sub images. The active contour techniques on the frequency tuned saliency map also give multiple objects with no bounds on the number of objects in the scene.

FUTURE WORK

An automated framework for estimation of intensity threshold and minimum distance thresholds for different scenes has to be developed. The proposed approach should be applied in classification algorithms for labelling the scenes with multiple objects present.