INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN: 2278 – 0211 (Online)

Saliency Based Human Detection Using The Guided Search System

Malavika H
Department Of Computer Science
Amrita School Of Engineering, Coimbatore, India
Mythili Sukumaran
Department Of Computer Science
Amrita School Of Engineering, Coimbatore, India

Abstract:
Since the necessity to differentiate between objects and human becomes prominent in fields like object tracking, CCTV, and pedestrian detection. This work aims to implement a method for detecting human in static medium. In this paper, a bottom up approach has been adopted for detecting salient regions which are the most prominent regions in an image. This approach uses dominancy feature properties of colour and orientation to detect the salient regions. The guided search system which makes use of the top down approach is employed on salient regions to assure human presence. The top down approach has a learning phase in which weights are calculated for images with human presence and a detection phase that's assures human presence for provided salient regions.

Key words: *Most Salient Regions, Guided search system, Bottom-up approach.*

1.Introduction

The ability of human visual system to detect visual saliency is extraordinarily fast and reliable. However, computational modelling of this basic intelligent behaviour still remains a challenge. Visual salience is the distinct subjective perceptual quality which makes some items in the world stand out from their neighbours and immediately grab one's attention.

Saliency typically arises from contrasts between items and their neighborhood. A "salient object" is the primary component in an image. For instance in an image there is a person standing in a forest surrounded by trees; the trees might be dark brown and the leaves might be green, but in all likelihood, the salient object is human. However, if the salient regions comprise more than half the pixels of the image, or if the background is complex, the background gets highlighted instead of the salient object. High saliency regions correspond to objects or places they are most likely to be found, while lower saliency is associated to background.

The core of visual salience is a bottom-up, stimulus-driven signal that announces "this location is sufficiently different from its surroundings to be worthy of your attention". This bottom-up deployment of attention towards salient locations can be strongly modulated or even sometimes overridden by top-down, user-driven factors. Thus, a lone red object in a green field will be salient and will attract attention in a bottom-up manner. In addition, if you are looking through a child's toy bin for a red plastic dog, and if the bin contains plastic objects of many vivid colors, no one color may be especially salient until your top-down desire to find the red object renders all red objects, whether dog or not, more salient.

In recent years, human detection from images and videos has been one of the active research topics in computer vision and machine learning due to many potential applications like image and video content management, video surveillance and driving assistance systems. Henceforth, despite numerous research efforts, the performance of the current human detection algorithm is still far from what can be reliably used under most realistic environments. This lower performance of the available algorithms can be improvised. So, here we propose the idea of including saliency to the process of human detection.

There are two ways of approaching a problem; bottom-up and top-down. Bottom-up approach is a task independent method to achieve a goal. While top-down approach more task dependent, bottom-up approach aims to develop small subsystems, which are verified and then merged to form bigger sub systems. The incoming data from the lower levels are processed and taken as input by the higher levels, which are then further carried above to form a grander system. In this approach, the systems base elements are specified in detail, these elements are then pieced together at various levels forming more subsystems, ultimately leading to top-level system.

2.Bottom-Up Approach

2.1.Bottom-Up Approach For Saliency Detection

As mentioned previously, saliency detection is identifying the most prominent region of an image. So bottom-up approach utilizes the factors that are solely conspicuous. Bottom-up attention mechanism can be considered highly reflexive or automatic. On seeing an image, the object that pops out will be the one with, high contrast, bright color or the shape of the object. All these factors will tend to be more prominent for the salient object, than the surrounding or background objects. Salient objects can be a flower in a grass bed, man in a desert, bird in the sky, but our work is narrowed down to finding humans as the salient object of an image.

2.2.Bottom-Up Architecture

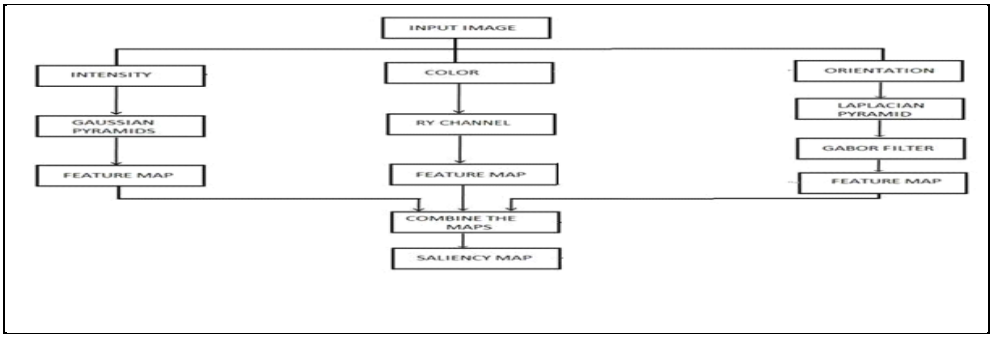

Figure 1 : Bottom Up Architecture

Feature computations are done on the intensity, color and orientation to compute saliency. Since the prominent object can be of varying sizes, saliency is computed on various scales. The image pyramid obtained for each of the feature channel; this is a standard approach to bring out saliency in computer vision. Saliency map is obtained by the combination of intensity map, color map and orientation map. Intensity map is taken into consideration in order to find the contrasting regions. The method we have adopted wraps both type of contrast, bright center with a dark background and vice versa. Color maps are used to find the most salient color of the image. This is because; the salient object has the most eye catching color of the picture.

2.2.1.Colour

Color is a primitive factor that helps narrow down to our region of interest. Color models are used to facilitate the specification of colors. Considering the fact that red, blue and green are the primary colors and they tend to catch the human eye easily, so specifically the RGB color model is used in bottom up search to identify the salient object.

The previous methods described in Itti [1] and Frintrop [2], the RGB channels along with the yellow was extracted; the yellow channel was extracted owing to the yellow color inclination of the human skin color.

Since we are aiming at detecting humans, and most humans have skin color that's more towards the red and yellow color, we combined the RY and BG channels, and the results to this was more towards our interest. Further considering only the RY channel and omitting the BG channel, our results were improved.

The RY color channel was computed as mentioned below:

$$R = r-(g + b)/2; \quad (3.1)$$

$$G = g-(r + b)/2; \quad (3.2)$$

$$B = b-(r + g)/2 \quad \text{and} \quad (3.3)$$

$$Y = (r + g - (2\,|(r-g) + b|) \quad (3.4)$$

Where r, g, b are the red, green, blue color channels of the image.

The Gaussian pyramid is constructed for the red and yellow channels up to level nine (0-8); level zero is of size 480x640. This is followed by the center surround difference.

$$RY(c, s) = |\{R(c)-Y(c)\} - \{(Y(s) - R(s)\}| \quad (3.5)$$

Where c = (2, 3, 4) and s = (3, 4).This results in six feature maps which are combined to form a single color map as shown in Figure 2

Figure 2: (a) Original Image, (b) Color Feature Map

2.2.2.Intensity

The intensity is related to the strength of the light beam. It is the relative lightness or darkness of a particular color; from black (no brightness) to white (full brightness).Intensity is a low level feature as like the color and orientation. The intensity image is obtained by removing the hue and saturation from the image. The intensity (value) factor is dependent on illumination level. Illumination conditions, contrast between face and background and orientation of face plays a vital role in detecting human faces. Since the saturation is removed, it's a gray scale image. With r, g, b being the red, green and blue channels of the input image, an intensity image (I) is obtained as (I) is obtained by I=(r+ g+ b)/3. This is followed by the center surround method. Here the Gaussian levels go up to nine (0 – 8); level zero is of size 480x640. The c = (2, 3, 4) and delta (3, 4). This results in six feature maps which is added to form one map(Figure 3) by across scale addition.

Figure 3: Intensity Feature Map

2.2.3.Orientation

The orientation maps are computed from the oriented pyramids which consists of four pyramids ,one for each of the orientation 0° , 45°,90°,135°.The pyramid for each orientation highlights the edges having this orientation on different scales .The orientations are computed by Gabor filter detecting bar-like features according to a specified orientation.

Gabor filter is a linear filter used for edge detection .Frequency and orientation representations of Gabor filter are similar to those of the human visual system. Their behaviour is similar to the responses of orientation sensitive cells in the human cortex. The Gabor Filters are self similar: all filters can be generated from one wavelet to another by dilation and rotation .It is the product of symmetric components in a particular direction, the other recover anti-symmetric components.

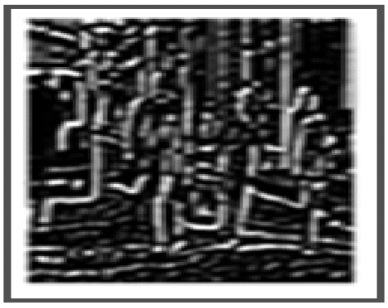

Figure 4: Orientation Feature Map

The parameters wavelength, aspect ratio and effective width are set to standard threshold value that produces the desirable output with edge of the human. The orientations 0°, 45°, 90° and 135° chosen because these orientations are similar to the receptive field profiles in the mammalian cortical field. We take orientation scale maps $O''_{\theta, s}$, for orientations $\theta \in \{0°, 45°, 90°, \text{and } 135°\}$ and scales $s \in \{s_2, s_3, s_4 \ldots\}$. The orientation scale maps are summed up by across scale addition for each orientation, yielding four orientation feature maps O'_{θ} of scale s, one for each orientation.

2.2.4.Final Amalgamation

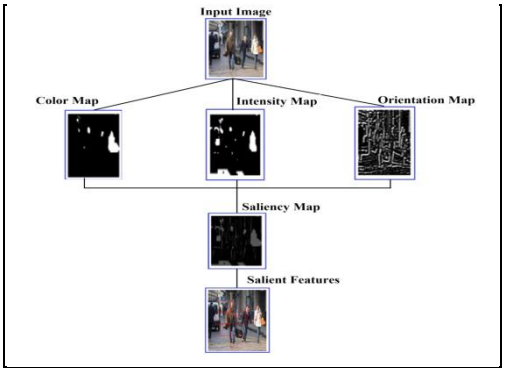

Figure 5: Region Detection Using Bottom-Up Approach

3.Guided Search System

In human visual perception, detecting the region of interest is the most important mechanism, but each time the region of interest changes depending on the situation. If we want to pick a yellow pen from the table among different color pens then the yellow pen is of region of interest, but again if we want to pick a red pen then this situation red pen becomes the region of interest. These ROI are regions of interest with high contrast to the surrounding and are unique. Even bottom-up techniques are considered to detect the region of interest. Even though emotions and motivations of human are considered out of scope, Goal directed search for target objects can be considered.

Normally in human behavior both top-down and bottom-up do not have much difference. They both inter-relate to each other. When we make use of top-down approach and search for a target with high uniqueness then attentional captururing happens. Since neuro-biological factors are not yet understood properly the goal directed search is not used for computational purpose but it can be combined with the top-down procedure to detect the region of interest using the basic training sessions.

Guided search system consists of two phases: Training and learning phase where the different images are trained accordingly to find the target human. It considers the properties of the target human along with its surrounding factors. The weights for different features are considered. The target to be searched will have a range of weights. Next phase is the search phase where it considers all the details from the previous phase of the different images and computes the basic knowledge to detect the target.

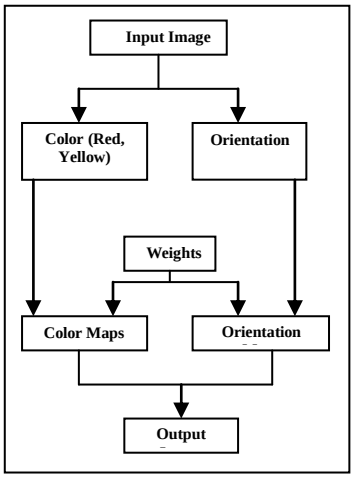

Figure 6: Architecture Of Guided Search System

3.1.Learning Phase

In this phase, the region of interest and a most salient region comprises of the human presence. For every salient region, weights are calculated for the different features such as color and orientation with respect to the human. Range of weights are obtained and normalized and the average weight ‘w’ would be used in the search phase to detect the human.

3.2.Detecting The Most Salient Region

The most salient regions can be detected using the bottom-up approach. First a Region of interest is chosen and within that the MSR is chosen. Region of interest many be the objects surrounding the human while the most salient region is the human. Using the bottom-up approach, saliency map is obtained and within which search for the most salient objects is conducted. The most salient regions are highlighted using red rectangles.

Figure 7: ROI And MSR Of An Image

In Figure 7 the Blue rectangle indicates the Region of Interest and Red rectangle indicates the Most Salient Region. For the region within the Red Rectangle the weight will be calculated to find the most prominent feature.

3.3.Calculation Of The Weight

Weight calculation is the most important part of the guided search system because the search phase fully depends on the weight to determine the human. Based on the weights calculated in the learning phase, the target object will be detected in the later phase. The ratio of mean salient region and the mean of the region around it will give the weight (Eqn 3.1) for that particular salient region.

$W_i = mean_{i(MSR)}/mean_{i(ROI-MSR)}$ (4.1)

$mean_{i(MSR)}$ is the mean average intensity values present in the pixel positions within the most salient region.

The weights will be calculated for the different features namely color (Red, green, blue) and orientation (0 degree,45 degree,90 degree,135 degree).

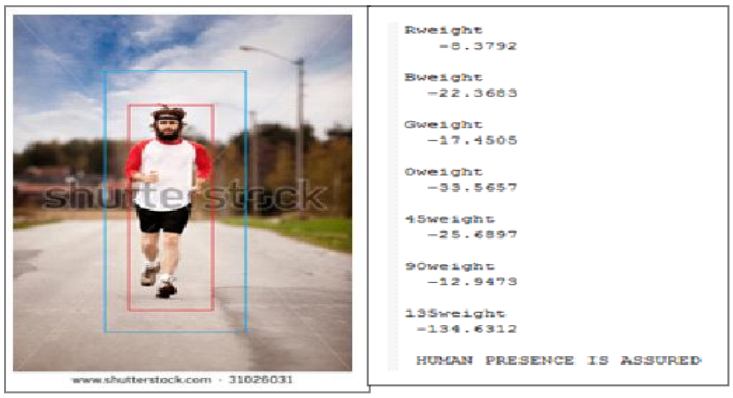

Figure 8: (a) Image, (b) Corresponding Weight For The Image

3.4.Choosing The Right Training Image

Choosing the right kind of image is a challenging task. Images of uniform background are considered rather than a mixture of surrounding objects. For example human walking on a road will have the road as the background so similar pictures with the road as the background should be chosen so that accurate results can be obtained. Images with full visibility should be considered.
Images with full visibility should be considered.

Figure: 9(a) & Figure: 9(b)

Figure 9: (a) image for learning process. The human is of particular distance and the whole structure is visible while as in (b) unsuitable image for training with no orientation visible.

This differentiation between full visibility and partial visibility can be easily found out by a human but for a system it should be trained in a particular way so that the system can understand and be able to predict the result.

3.5.Search Phase

The objective of search phase is to use the weights from learning phase and detect the target object which satisfies the condition. In search mode, all the salient regions are searched based on the weights derived from the learning phase and detect the presence of the human in that image.

3.5.1..Search Phase Using The Bottom Up Results

The bottom up search gives an output image with all the salient regions highlighted in the form of rectangles. This search phase will consider the highlighted regions and search for human presence. Since the salient regions are already detected by the bottom up search, this search phase only searches within those regions instead of searching the whole image. This makes the processing time less and more efficient. In Figure 10(a) shows the output of the bottom up approach with highlighted regions. The regions showed in Figure 10(a) are searched for human presence and the Figure 10(b) shows the region with the target human.


Figure 10: (a) Bottom-Up Result & (b) Detection Of Human Detection

4.Scope for further work

The algorithm we have deduced detects humans in a standing posture and specific skin colour range and can be used in applications such as pedestrian diction system. Project can be extended to detect various postures of human .The current algorithm detects humans having skin regions whitish and so it can be extended for darker skin colour as well.

5.Conclusion

In this paper a computational system for detection of human is presented. This system is influenced by a two step process: Bottom-up approach and a guided search approach. An improvement on bottom-up approach is done, such that the channels of color, orientation and intensity are selectively tuned to detect the human regions in the image. From the saliency map the most salient regions are extracted, that acts as the region of interest. The search is performed in the regions of interest in order to find the target region (human) by the guided search system. It can be noted that a few training samples are sufficient to learn the features of the target object.

Since the channels are fine tuned for bottom-up approach and the search region is minimized for the guided search system, the running time is reduced. The results show that the combination of bottom-up approach and guided search system gives an accurate result of about 91 percent.

6.References

[1] Laurent Itti, Christof Koch and Ernst Niebur, A Model of Saliency-based Visual Attention for Rapid Scene Analysis, IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI),1998

[2] Tie Liu, Jian Sun, Nan-Ning Zheng, Xiaoou Tang and Heung-Yeung Shum Learning to detect a salient object, IEEE, 2007

[3] Zheshen Wang and Baoxin Li, A two stage approach to saliency detection in images, IEEE, 2008

[4] Nikhil Rao, Joseph Harrison Tyler Karrels, Robert Nowak and Timothy T. Rogers," Using Machines to Improve Human Saliency Detection"

[5] Xiaodi Hou and Liqing Zhang, Saliency Detection: A Spectral Residual Approach,

[6] B. Heisele and C. Wohler, "Motion-based recognition of pedestrians", in Proc. International Conference on Pattern Recognition, pp. 1325-1330, vol. 2, August 1998.

[7] T. Haga, K. Sumi, and Y. Yagi. Human detection in outdoor scene using spatio-temporal motion analysis. International Conference on Pattern Recognition, 4:331–334, 2004.

[8]Sang Min Yoon and Hyunwoo Kim. Real-time multiple people detection using skin color, motion and appearance information. International Workshop on Robot and Human Interactive Communication, pages 331–334, 2004.

[9]Lijun Jiang, Feng Tian, Lim Ee Shen, Shiqian Wu, Susu Yao, Zhongkang Lu, and Lijun Xu. Perceptual-based fusion of ir and visual images for human detection. International Symposium on Intelligent Multimedia, Video and Speech Processing, pages 514– 517, 2004.

[10] Vladimir Vezhnevets Vassili, Sazonov Alla Andreeva, "A Survey on Pixel-Based Skin Color Detection Techniques", Graphics and Media Laboratory, Faculty of Computational, Mathematics and Cybernetics, Moscow State University, Moscow, Russia, In Proceedings of the GraphiCon 2003, pp. 85-92.

[11] P Burt and E. Adelson, "The Laplacian Pyramid as a Compact Image Code," IEEE Transactions o n Communication, 1983.