Detection Of Optic Disc & Cup Of Digital Fundus Image Using Salient Object Detection Method For Glaucoma Detection

Vyankatesh D. Gutte^a, Prof Yogita M. Vaidya^b

^aDept. of Electronics and Telecommunication, College Of Engineering, Pune, India

^bDept. of Electronics and Telecommunication, College Of Engineering, Pune, India

Vyankatesh D. Gutte: vyankidgutte@gmail.com

ABSTRACT

Automated retinal image analysis is promising screening tool for early detection of eye disease. In this automated analysis some factors need to be considered in order to get better analytical results. We present in this paper methodology to extract exact boundary of OD & Cup in digital retinal fundus images. The method starts with preprocessing of digital fundus images by contrast normalization throughout the image, and removal of blood vessels, which is major reason for distraction of finding OD candidate, using Bottom Hat Transform. Using this OD candidate we find area of interest i.e. area surrounding OD & cup. Then we apply salient object detection algorithm. In this paper, we deal with the salient object detection problem for images. We formulate salient object detection as a binary labeling task that separates a salient object from the background since; one pays more attention to salient object in an image as compared to the background of the image. Feature extraction methods are like edge detection, thresholding, multi scale contrast, Center surround histogram.

Feature maps are prepared for each of the method mentioned above and saliency is computed by a center-surround operation, self-information, or graph-based random walk using multiple features. After normalization and linear/non-linear combination, a master map or a saliency map is computed to represent the saliency of each image pixel. Last, a few key locations on the saliency map are identified by winner-take-all, or inhibition-of-return, or other non-linear operations. Then we highlight the boundaries.

Keywords:- Bottom Hat Transform (BHT), Feature Extraction, Saliency Computation, Visual Attention.

1. INTRODUCTION

Diabetic retinopathy is a chronic disease. One of the most common cause of blindness is Glaucoma with

which about 79 million world population is likely to be affected by the year 2020. The benefits, that a system to automatically detect early signs of this disease would provide, have been widely studied and assessed positively by the experts [2, 3]. Thus Optic Disc and Cup plays an important role for developing automated diagnosis expert system for glaucoma as its segmentation and analysis is a key preprocessing component in many algorithms designed for identifying other fundus features like fovea, vascular tree.

Automated diagnosis of other ophthalmic pathologies can be done using OD and cup segmentation [4, 5]. To detect presence of ophthalmic disease such as diabetic retinopathy, OD and cup segmentation can be used..

Slightly elliptical most bright region in an eye fundus image is OD which contains Cup area which is also slightly elliptical in shape. Size of both may vary significantly, lots of estimations by researchers. Sinthanayothin et. al [6] stated about one-seventh of the entire image is occupied by OD, some researchers pointed out that OD and Cup size depends on each individual occupying one- tenth to one-fifth of the image.

OD segmentation job made harder mainly due to blood vessels presence, ill defined boundaries, variations near the disc boundaries due to pathological changes and variable imaging conditions. Specifically, similar regions near the boundary of optic disc, shape irregularity and boundary irregularity are the most important aspects to be countered by a OD and Cup segmentation method. For better results we remove blood vessels using BHT.

The human brain & visual system pays more attention to some parts of an image which seems more important than others. The salient object can be defined as the objects in an image which draws most visual attention. The object or an area of the image which is of most interest or is more important is labeled as salient objects, or foreground objects that we are familiar with. The salient object detection problem for images is formulated as a binary labeling task, where the salient object is separated from the background. This technique is used for segmenting optic disc & cup from background. For that firstly we found area of interest and then we applied this method of salient object segmentation.

2. PREPROCESSING

Original image enhancement is done by applying bottom hat transform to it. The fig. 1 (b) significantly shows major blood vessels lies on the nasal side and small vessels in temporal side present a good amount of OD and Cup occlusion. We remove or suppress the blood vessels which are causing distraction as follows. The Bottom Hat Transform [1] is applied on red channel of original image as blood vessels are red in color using line as linear structural element. Bottom hat transform is the difference between a closing of the image and original image I_0 defined as:

$$\rho_s^\theta(I_0) = \varphi_s^\theta(I_0) - I_0$$

Where φ_s^θ denotes morphological closing operation with a linear structuring element s of orientation θ .

Morphological closing is nothing but dilation followed by erosion. The erosion of f using b as structuring element at any location (x, y) is defined as the minima of that image in the region projected by b when the start of b is at (x, y) . Therefore, the erosion of an image f at (x, y) with b as structuring element is given by

$$[f \ominus b](x, y) = \min_{(s,t) \in b} \{f(x + s, y + t)\}$$

The dilation of f with b as structuring element at any location (x, y) is defined as the maxima of the image in the region projected by b^- when the start of b^- is (x, y)

$$[f \oplus b](x, y) = \max_{(s,t) \in b} \{f(x - s, y - t)\}$$

The fig. 1 (c) shows result when we apply morphological closing to original image. The fig1(d) shows result after BHT.

Fig. 1. Result of Preprocessing (a) Input image (b) A sample cropped CFI (c) Image obtained by Morphological Closing (d) BHT result

3. REGION OF INTEREST

Region of interest for OD & cup segmentation is area surrounding OD which is of 50*50 pixel size at maximum. For finding area of interest, we have to first find OD candidate i.e. any point in optic disc. It can be found by several methods like maximum difference method, maximum variance method or low pass filtering method, of which low pass filtering method is precise as it gives mostly pixel near to center of optic disc.

Low pass filtering method: The OD pixel of this method is the maximum gray level pixel in a low pass filtered image. Major distraction for finding this pixel, blood vessels, is removed by BHT. In a retinography, usually OD is the brightest and most distinct region but the brightest pixel could not be located as OD candidate. In many cases, due to artifacts in image this pixel may be in some different smaller region outside OD. For removing these distracters, the image I is transformed to the frequency domain and I is filtered by the Gaussian low-pass filter defined as follows

$$H(u, v) = \exp\left(\frac{-D^2(u, v)}{2D_0^2}\right)$$

Where $D(u, v)$ is the Euclidean distance between the point (u, v) and the origin of the frequency plane, and D_0 is the cut-off frequency which is equal to 25 Hz. The pixel with highest gray-level in the filtered image, in the spatial domain, is the result of this method. In major cases, due to preprocessing, the OD pixel found by this method is center of OD.

After finding this pixel, we find area of interest by cutting image 50*50 pixel area surrounding this pixel as center..

Fig. 2.(a) Result of Low Pass Filtering, (b) Region Of Interest

4. FEATURE EXTRACTION:

4.1 SMOOTHENING AND EDGE DETECTION

The Gaussian filter is used for smoothening of an image. The Gaussian blur is a type of image-blurring filter that uses a Gaussian function for calculating the transformation to apply to each pixel in the image. In two dimensions, it is the product of two such Gaussians, one in each dimension:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

where x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution.

Edges characterize boundaries and are therefore a problem of fundamental importance in image processing. Edges in images are areas with strong intensity contrasts – a jump in intensity from one pixel to the next. Edge detecting an image significantly reduces the amount of data and filters out useless information, while preserving the important structural properties in an image. There are many ways to perform edge detection. However we will use the gradient method. The gradient method detects the edges by looking for the maximum and minimum in the first derivative of the image. An edge has the one-dimensional shape of a ramp and calculating the derivative of the image can highlight its location. A pixel location is declared an edge location if the value of the gradient exceeds some threshold. As mentioned before, edges will have higher pixel intensity values than those surrounding it. So once a threshold is set, you can compare the gradient value to the threshold value and detect an edge whenever the threshold is exceeded.

Based on this one-dimensional analysis, the theory can be carried over to two-dimensions as long as there is an accurate approximation to calculate the derivative of a two-dimensional image. The Sobel operator performs a 2-D spatial gradient measurement on an image. Typically it is used to find the approximate absolute gradient magnitude at each point in an input grayscale image. The Sobel edge detector uses a pair of 3x3 convolution masks, one estimating the gradient in the x-direction (columns) and the other estimating the gradient in the y-direction (rows). A convolution mask is usually much smaller than the actual image. As a result, the mask is slid over the image, manipulating a square of pixels at a time. The actual Sobel masks are shown by

-1	0	+1
-2	0	+2
-1	0	+1

G_x

+1	+2	+1
0	0	0
-1	-2	-1

G_y

Fig1: Sobel masks for edge detection

The magnitude of the gradient is then calculated using the formula:

$$|G| = \sqrt{G_x^2 + G_y^2}$$

An approximate magnitude can be calculated using:

$$|G| = |G_x| + |G_y|$$

Fig 3. Gaussian Averaging

4.2 ADAPTIVE THRESHOLDING

Thresholding is used to segment an image by setting all pixels whose intensity values are above a threshold to a foreground value and all the remaining pixels to a background value. Whereas the conventional thresholding operator uses a global threshold for all pixels, adaptive thresholding changes the threshold dynamically over the image. This more sophisticated version of thresholding can accommodate changing lighting conditions in the image, e.g. those occurring as a result of a strong illumination gradient or shadows.

Adaptive thresholding typically takes a grayscale or color image as input and, in the simplest implementation, outputs a binary image representing the segmentation. For each pixel in the image, a threshold has to be calculated. If the pixel value is below the threshold it is set to the background value, otherwise it assumes the foreground value. The main approach to find the threshold is: local thresholding. The assumption behind method is that smaller image regions are more likely to have approximately uniform illumination, thus being more suitable for thresholding. To find the local threshold is to statistically examine the intensity values of the local neighborhood of each pixel. The statistic which is most appropriate depends largely on the input image. the local intensity distribution, The size of the neighborhood has to be large enough to cover sufficient foreground and background pixels, otherwise a poor threshold is chosen. On the other hand, choosing regions which are too large can violate the assumption of approximately uniform illumination. This method is less computationally intensive than the Chow and Kaneko approach and produces good results for some applications.

4.3 MULTISCALE CONTRAST:

Contrast is the most commonly used local feature for attention detection because the contrast operator simulates the human visual receptive fields. Without knowing the size of the salient object, contrast is usually computed at multiple scales.

Visual data is decomposed into multi-scale sub-images which contain multi-scale details on the basis of Gaussian Pyramid, and contrast features are extracted from these multi-scale images for saliency map generation. The multi-scale contrast feature $f_c(x, l)$ is defined as a linear combination of contrasts in the Gaussian image pyramid:

$$f_c(x, l) = \sum_{l=1}^L \sum_{x' \in N(x)} \|I^l(x) - I^l(x')\|^2$$

where I^l is the l th-level image in the pyramid and the number of pyramid levels L is 6. $N(x)$ is a 9×9 window. The feature map $f_c(l)$ is normalized to a fixed range $[0, 1]$.

A Gaussian pyramid is a technique used in image processing, especially in texture synthesis. The technique involves creating a series of images which are weighted down using a Gaussian average (Gaussian blur) and scaled down. When this technique is used multiple times, it creates a stack of successively smaller images, with each pixel containing a local average that corresponds to a pixel neighborhood on a lower level of the pyramid.

Multiscale contrast highlights the high contrast boundaries by giving low scores to the homogenous regions inside the salient object.

Fig 4.(a),(b),(c),(d) Multi Scale Contrast

4.4 CENTER SURROUND HISTOGRAM:

Histogram-based methods are very efficient when

compared to other image segmentation methods because they typically require only one pass through the pixels. In this technique, a histogram is computed from all of the pixels in the image, and the peaks and valleys in the histogram are used to locate the clusters in the image. Color or intensity can be used as the measure. The salient object can always be distinguished by the difference of it and its context.

They are insensitive to small changes in size, shape, and viewpoint. Another reason is that the histogram of a rectangle with any location and size can be very quickly computed by means of an integral histogram.

Suppose the salient object is enclosed by a rectangle A surrounding contour RS with the same area of R is constructed. Here the χ^2 distance between histograms of RGB color is used. The most distinct rectangle $R^*(x)$ centered at each pixel x is found out by varying the size and aspect ratio and is given by:

$$R^*(x) = \arg \max R(x) X^2(R(x), R_s(x))$$

Then, the center-surround histogram feature $fh(x, l)$ is defined as a sum of spatially weighted distances:

$$f_h(x, l) \propto \sum_{\{x' | x \in R^*(x')\}} w_{xx'} X^2(R^*(x'), R_s^*(x'))$$

where $R^*(x')$ is the rectangle centered at x' and containing the pixel x . The weight,

$$w_{xx'} = \exp(-0.5 \cdot \sigma_x^{-2} \|x - x'\|^2)$$

is a Gaussian falloff weight. Finally, the feature map $fh(l)$ is also normalized. Thus the salient object has a large center surround histogram distance.

5. SALIENCY MAP COMPUTATION

The purpose of the saliency map is to represent the conspicuity—or “saliency”—at every location in the visual field by a scalar quantity and to guide the selection of attended locations, based on the spatial distribution of saliency. A combination of the feature maps provides bottom-up input to the saliency map, modelled as a dynamical neural network.

One difficulty in combining different feature maps is that they represent a priori not comparable modalities, with different dynamic ranges and extraction mechanisms. Also, because all feature maps are combined, salient objects appearing strongly in only a few maps may be masked by noise or by less-salient objects present in a larger number of maps. In the absence of top-down supervision, we propose a map normalization operator, $N(\cdot)$, which globally promotes maps in which a small number of strong peaks of activity

(conspicuous locations) is present, while globally suppressing maps which contain numerous comparable peak responses. $N(.)$ consists of:

- 1) Normalizing the values in the map to a fixed range $[0...M]$, in order to eliminate modality-dependent amplitude differences;
- 2) Finding the location of the map's global maximum M and computing the average m of all its other local maxima; and
- 3) Globally multiplying the map by $(M - m)^2$.

Only local maxima of activity are considered, such that $N(.)$ compares responses associated with meaningful "activation spots" in the map and ignores homogeneous areas. Comparing the maximum activity in the entire map to the average overall activation measures how different the most active location is from the average. When this difference is large, the most active location stands out, and the map is strongly promoted. When the difference is small, the map contains nothing unique and is suppressed.

Feature maps are combined into three "conspicuity maps", for intensity, & for color, and for orientation, at the scale ($s = 4$) of the saliency map. They are obtained through across-scale addition, " Θ " which consists of reduction of each map to scale four and point-by-point addition:

$$\bar{I} = U_2^4 U_{s=c+3}^{c+4} N(I(c, s))$$

$$\bar{C} = U_{c=2}^4 U_{s=c+3}^{c+4} [N(RG(c, s)) + N(BY(c, s))]$$

For orientation, four intermediary maps are first created by combination of the six feature maps for a given q and are then combined into a single orientation conspicuity map:

$$\bar{O} = \sum_{\theta=(0,45,90,135)} N(U_{c=2}^4 U_{s=c+3}^{c+4} N(O(c, s, \theta)))$$

**U corresponds to a cross-scale addition

The motivation for the creation of three separate channels I, C, O and their individual normalization is the hypothesis that similar features compete strongly for saliency, while different modalities contribute independently to the saliency map. The three conspicuity maps are normalized and summed into the final input

$$S = \frac{1}{3} (N(I) + N(C) + N(O))$$

At any given time, the maximum of the saliency map (SM) defines the most salient image location, to which the focus of attention (FOA) should be directed. We could now simply select the most active location as defining the point where the model should next attend. However, in a neurally plausible implementation, we model the SM as a 2D layer of leaky *integrate-and-fire* neurons at scale four. These

model neurons consist of a single capacitance which integrates the charge delivered by synaptic input, of a leakage conductance, and of a voltage threshold. When the threshold I reached, a prototypical spike is generated, and the capacitive charge is shunted to zero. The SM feeds into a biologically plausible 2D "winner-take-all" (WTA) neural network, at scale $s = 4$, in which synaptic interactions among units ensure that only the most active location remains, while all other locations are suppressed. The neurons in the SM receive excitatory inputs from S and are all independent. The potential of SM neurons at more salient locations hence increases faster (these neurons are used as pure integrators and do not fire). Each SM neuron excites its corresponding WTA neuron. All WTA neurons also evolve independently of each other, until one (the "winner") first reaches threshold and fires.

This triggers three simultaneous Mechanisms:

- 1) The FOA is shifted to the location of the winner neuron;
- 2) The global inhibition of the WTA is triggered and completely inhibits (resets) all WTA neurons;
- 3) Local inhibition is transiently activated in the SM, in an area with the size and new location of the FOA; this not only yields dynamical shifts of the FOA, by allowing the next most salient location to subsequently become the winner, but it also prevents the FOA from immediately returning to a previously-attended location.

Such an "inhibition of return" has been demonstrated in human visual psychophysics. In order to slightly bias the model to subsequently jump to salient locations spatially close to the currently attended location, a small excitation is transiently activated in the SM, in a near surround of the FOA.

Since we do not model any top-down attentional component, the FOA is a simple disk whose radius is fixed to one sixth of the smaller of the input image width or height.

- This is Itti's computational model for salient object detection
- This is bottom up computational approach.

Fig 4. Saliency Map

6. EVALUTION:

6.1 GROUND TRUTH CONSTRUCTION:

An image can have many salient objects and according to the user their idea of a salient object in an image may change. Hence for training of algorithm over a data set of number of images, the user is asked to specify the salient object in the image according to him. Henceforth, our algorithm puts a rectangle over the salient object. Then the ground truth of the salient image in the image is founded. The saliency probability map is,

$$g_x = \frac{1}{M} \sum_{m=1}^M (a)^m$$

Where M is the number of users.

$A^m = \{a_x^m\}$ the binary mask labeled by the m^{th} user. With this saliency probability map the masked salient object is evaluated using three parameters viz. Precision, Recall and F- measure. These are defined as,

6.1.1 PRECISION

Precision is defined as the ratio of correctly detected salient region to the detected salient region.

$$\text{Precision} = \frac{\text{correctly detected salient region}}{\text{detected salient region}}$$

$$\text{Precision} = \frac{\sum_x g_x a_x}{\sum_x a_x}$$

6.1.2 RECALL

Recall is defined as the ratio of correctly detected salient region to the ground truth of salient region.

$$\text{Recall} = \frac{\text{correctly detected salient region}}{\text{Ground Truth salient region}}$$

$$\text{Recall} = \frac{\sum_x g_x a_x}{\sum_x a_x}$$

Fig 5: Itti's computational model for salient object detection

6.1.3 F-MEASURE

F-measure is the weighted harmonic mean of precision and recall with a non- negative α . F-measure is the overall performance measurement.

$$F_\alpha = \frac{(1+\alpha) \cdot \text{Precision} \cdot \text{Recall}}{\alpha \cdot \text{Precision} + \text{Recall}}$$

Where,

$$\alpha = 0.5$$

To obtain the salient object in an image we have to resolve the rectangles by maximizing the values for the expression:

$$f(x) = \sum_{x \in R} (1 - F(x)) + \sum_{x \notin R} F(x)$$

Where, R is resolve rectangle;

$F(x) \in [0, 1]$ is the normalized saliency map.

7. BOUNDARY

After all this processing, we get saliency map, which shows only two regions cup & optic disc. Now circular boundaries can be traced by finding diameter & center of both regions. Diameter is nothing but maximum distance between two pixels of each region. Fig. 5(b) shows highlighted boundaries of OD & cup.

Fig 5.(a)Original Image

(b) Highlighted OD & Cup

Table 1. Statistics

Sr No.	Algorithm	No. Of Images With Good Results	Execution Timing/ image (Avg) In secs
1	Pre-processing BHT	529	1.0596605
2	LPF	509	0.686599
3	ROI	509	0.042135
4	Optic Disk Segmentation	451	0.482463
5	Cup Segmentation	442	0.432613

8. CONCLUSION

Our project is based on bottom up computational approach only. We have completed basic image processing techniques related to image enhancement, feature extraction, segmentation and object recognition. We have obtained feature maps by thresholding and boundary determination. We are looking forward to combine them to give saliency map and determine salient object in an Image for optic disc and cup segmentation. We are currently working with database from MESIDOR, STARE, and DRIVE and from doctor. Collective results of each database is tabulated in Table 1. We have applied this algorithm on 529 different images. Table 1 shows results step by step and execution timing for each step in algorithm. This algorithm have success percentage as around 85. The algorithm developed in this paper can be extended to sequential images (video) also. We have defined the evaluation criteria to calculate the efficiency of our algorithm. After completion of the implementation we can evaluate our results through techniques mentioned.

REFERENCES

- (1) Gopal Datt Joshi, Jayanthi Sivaswamy, Kundan Karan, S. R. Krishnadas, "Optic Disk And Cup Boundary Detection Using Regional Information".
- (2) N.Patton,T.M.Aslam,T.MacGillivray,I.J.Dear y,B.Dhillon,R.H.Eikelboom, K.Yogesana, and I.J.Constable, " Retinal image analysis: Concepts, applications and potential," *Prog. Retin. Eye Res.*,vol.25,pp. 99-127,2006.
- (3) A.Singalavanja,J.Supoavej,P.Bamroongsuk,C .Sinthanayethin,S.Phoojaruenchamachai,and V.Kongbunkiat,"Feasibility study on computer-aided screening for diabetic retinopathy," *Jpn. J.Ophthalmol.*,vol.50.,pp. 361-366,2006.
- (4) Shijian Lu," Accurate and Efficient Optic Disc Detection and Segmentation by a Circular Transformation" *IEEE Transactions on medical imaging*, vol. 30, no.12, November 2011
- (5) T.Ashok Kumar, S.Priya, M.G.Mini" Optic Disc Localization in Ocular Fundus Images", International Conference on VLSI, Communication & Instrumentation (ICVCI) 2011 Proceedings published by International Journal of Computer Applications® (IJCA)
- (6) *J.Ophthalmol.*,vol.50.,pp. 361-366,2006.
- (7) C.Sinthanayothin,J.F.Boyce,H.L.Cook,and T. H.Williamson,"Automated localization of the optic disc, "Automated localization of optic disc, fovea, and retinal blood vessels from color fundus images," *Br.J.Ophthalmol.*,vol.83,pp.902-910,1999.
- (8) Tie Lie, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, "Learning to detect Salient Object"
- (9) L. Itti, C. Koch, and E. Niebur, "A model of saliency based visual attention for rapid scene analysis," *IEEE Trans.on PAMI*, vol. 20, no. 11, pp. 1254–1259, 1998.attention for rapid scene analysis," *IEEE Trans. on PAMI*,vol. no. 11, pp. 1254–1259, 1998.
- (10) Visual attention home page "www. ilab.usc.edu/bu/"
- (11) Saliency toolbox by Dirk Bernhardt-Walther "www.saliencytoolbox.net"