ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 1, January 2014

Cognitive Sensor Fusion Based Environmental Awareness System for Visually Impaired People

Dhivya.G, Jenish ponraj peter.J
M.E –Embedded System Technologies
Sri Shakthi Institute of Engineering and Technology, Coimbatore

Abstract— ***To overcome the difficulty of navigation for visually impaired people in a self assistive manner an image processing based cognitive fusion and sensing technique is adopted. Here a CCD camera acts as a sensor by sensing the real time video images and the obstacle in front of the user is identified with the help of a cognitive algorithm. Here we implement a cognitive technique of visual saliency detection based on a graphical approach to determine a salient feature in the obstacle image. This salient feature is associated with other common features of the image. We already store some predefined real world images in a template such that whenever a salient feature is determined in the input image a corresponding image is matched if it is present in the template and informed to the user through an audio device or else the input image is stored in the template as a new real world object. Along with the obstacle information alternate path is also suggested to the user.***

I. INTRODUCTION

There are many self assisting electronic devices to guide visually impaired people such as: i) Laser Canes: It uses invisible laser beams to detect obstacles and produces a specific audio signal. It has three distinct audio signals, each indicates a specific distance. ii)Sonic Mobility Device: This is a device that is generally mounted on users head. It uses ultrasonic technology to detect obstacles. It uses the musical scale's 8 tones to indicate the distance of the object. Each tone signifies a particular distance from the obstacle .iii)Handheld Mobility Device: This is a small device through which the user points around the surroundings. Once the handheld device detects the obstacle it will vibrate. Depending upon the level of vibration the user can identify the distance of the obstacle. Here for the first time we implement the cognitive concept of visual saliency for guiding visually impaired. Normally visual saliency detection is used in machine vision and robotics.

But here for the first time tried to implement for the benefit of visually challenged. Most vertebrates including humans move their eyes. This ability is used to sample most relevant features of a scene, while spending only limited processing resources elsewhere. The ability to predict, given an image, where a human might fixate in a fixed-time free viewing scenario has long been of interest in the vision community. Apart from other technologies mentioned here we implement a cognitive concept of visual saliency detection is implemented to identify the obstacles in front of the visually impaired. The standard approaches(e.g.,[2],[9]) of visual saliency detection are based on biologically motivated feature selection, followed by center-surround operations which highlight local gradients, and finally a combination step leading to a "master map". Recently, Bruce [5] and others [4] have hypothesized that fundamental quantities such as "self-information" and "surprise" are at the heart of saliency/attention. Thus the leading models of visual saliency may be organized into these three stages:

(s1) extraction**:** extract feature vectors at locations over the image plane
(s2) activation**:** form an "activation map" (or maps) using the feature vectors
(s3) normalization/combination**:** normalize the activation map (or maps, followed by a combination of maps into a single map)

Here in the graph based algorithm we define Markov chains over various image maps, and treat the equilibrium distribution over map locations as activation and saliency values. Here we take a unified approach to steps (s2) and (s3) of saliency computation, by using dissimilarity and saliency to define edge weight on graphs which are interpreted as Markov chains. As in previous cases here there is no attempt made to connect features only to those which are somehow similar. Here this method is also compared with other methods, using power to predict human fixations as a performance metric.
The contributions in this paper are:

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 1, January 2014

1) A complete bottom-up saliency model based on graph computations, including a framework for "activation and "normalization/combination".

2) A comparison of this visual saliency detection technique against existing benchmarks on a data set of gray scale images of natural environments (viz.,foliage)with the eye-movement fixation data of seven human subjects, from a recent study by Einhauser et.al.[1].

II. IMPLEMENTED ALGORITHM OF VISUAL SALIENCY

Given an image I, we wish to ultimately highlight a handful of 'significant' locations where the image is 'informative' according to some criterion, e.g. Human fixation. This process is conditioned on first computing feature maps (s1) ,e.g. by linear filtering followed by some elementary nonlinearity [15]. "Activation" (s2), "Normalization and combination" (s3) steps follow as described below.

A. Formation of Activation Map (s2)

Suppose we are given a feature map M : [n]^2 → R. Our goal is to compute an activation map A:[n]^2 → R, such that intuitively, locations (i,j) [n]^2 where I, or as a proxy, M(i,j), somehow is unusual in its neighbourhood will correspond to high values of activation A.

1. Existing Schemes

Of course "unusual" does not constrain us sufficiently, and so one can choose several operating definitions. "Improbable" would lead one to the formulation of Bruce [5], where a histogram of M(i,j) values is computed in some region around (i,j), subsequently normalized and treated as a probability distribution, so that A(i,j) = -log(p(i,j)) is clearly defined with p(i,j)=Pr{M(i,j)| neighbourhood}. Another approach compares local "center" distributions to broader "surround" distributions and calls the Kullback-Leibler tension between the two "surprises" [4].

Markovian Approach

We propose a more organic approach. Let us define the dissimilarity of M(i,j) and M(p,q) as d((i,j) || (p,q) =|log(M(i,j)/M(p,q)|.This is a natural definition of dissimilarity: simply the distance between the one and the ratio of two quantities, measured on a logarithmic scale. For some of our experiments, we use |M(i,j) – M(p,q)| instead, and we have found that both work well. Consider now the fully connected directed graph Ga, obtained by connecting every node of the lattice M, labeled with two indices (i,j) [n]^2 , with all other n-1 nodes. The directed edge from node (i,j) to node (p,q) will be assigned a weight w1((i,j),(p,q))=d((i,j)||(p,q)).F(i-p,j-q) where F(a,b) = exp(-(a^2+b^2)/2α^2). α is a free parameter of this algorithm.

Thus, the weight of the edge from the node (i,j) to node (p,q) is proportional to their dissimilarity and to their closeness in the domain of M. Note that the edge in the opposite direction has exactly the same weight. We may now define a Markov chain on Ga by normalizing the weights of the outbound edges of each node to 1, and drawing equivalence between nodes & states, and edges weights & transition probabilities. The equilibrium distribution of this chain, reflecting the fraction of time a random walker would spend at each node/state if he were to walk forever, would naturally accumulate mass at nodes that have high dissimilarity with their surrounding nodes, since transitions into such sub graphs is likely, and unlikely if nodes have similar M values. The result is an activation measure which is derived from pair wise contrast. We call this approach "organic" because, biologically, individual "nodes" (neurons)exist in a connected, retinotopically organized, network(the visual cortex), and communicate with each other (synaptic firing) in a way which gives rise to emergent behaviour, including fast decisions about which areas of a scene requires additional processing. Similarly, our approach exposes connected (via F) regions of dissimilarity (via w), in a way which can in principle be computed in a completely parallel fashion. Computations can be carried out independently at each node: in a synchronous environment, at each time step, each node simply sums incoming mass, then passes along measured partitions of this mass to its neighbours according to outbound edge weights. The same simple process happening at all nodes simultaneously gives rise to an equilibrium distribution of mass.

Technical Notes

The equilibrium distribution of this chain exists and is unique because the chain is ergodic, a property which emerges from the fact that our underlying graph Ga is by construction strongly connected. In practice, the

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 1, January 2014

equilibrium distribution is computed using repeated multiplication of the Markov matrix with an initially uniform vector. The process yields the principal eigen-vector of the matrix. The computational complexity is thus O ((n^4)K) where K<<n^2 is some small number of iterations required to meet equilibrium.

B. "Normalizing" an Activation Map (s3)

Armed with the mass concentration definition, we propose another Markovian algorithm as follows: This time, we begin with an activation map A:[n]^2 → R, which we wish to normalize. We construct a graph Gn with n^2 nodes labeled with indices from [n]^2. For each node (i,j) and every node (p,q) to which it is connected, we introduce an edge from (i,j) to (p,q) with weight:w2((i,j),(p,q)) = A(p,q).F(i-p,j-q). Again normalizing the weights of the outbound edges of each node to unity and treating the resulting graph as a Markov chain gives us the opportunity to compute the equilibrium distribution over the nodes. Mass will flow preferentially to those nodes with high activation. It is a mass concentration algorithm by construction, and also one which is parallelizable, as before, having the same natural advantages. Experimentally it seems to behave very favorably compared to the standard approaches such as" DoG" and "NL".

III. EXPERIMENTAL RESULTS

A .Preliminaries and Paradigm

We perform saliency computations on real images of the natural world, and compare the power of the resulting maps to predict human fixations. The experimental paradigm we pursue is the following: for each of a set of images, we compute a set of feature maps using standard techniques. Then, we process each of these feature maps using some activation algorithm, and then some normalization algorithm, and then simply sum over the feature channels. The resulting master saliency map is scored (using an ROC area metric described below) relative to fixation data collected for the corresponding image, and labeled according to the activation and normalization algorithms used to obtain it. We then pool over a corpus of images, and the resulting set of scored and labeled master saliency maps is analyzed in various ways presented below. Some notes follow:

Algorithm Labels

Hereafter, "graph (i)" and "graph (ii)" refer to the activation algorithm previously described. The difference is that in graph (i), the parameter α=2.5, whereas in graph (ii), α=5. "graph (iii)" and "graph (iv)" refer to an iterated repetition of the normalization algorithm previously described. The difference is the termination rule associated with the iterative process: for graph (iii), a complicated termination rule is used which looks for a local maximum in the number of matrix multiplications required to achieve a stable equilibrium distribution, and for graph iv, the termination rule is simply "stop after 4 iterations". The normalization algorithm referred to as "I" corresponds to "Identity", with the most naive normalization rule: it does nothing, leaving activations unchanged prior to subsequent combination. The algorithm "max-ave" and "DoG" were run using the publicly available "saliency toolbox". The parameters of this were checked against the literature [2] and [3], were found to be almost identical, with a few slight alterations that actually improved performance relative to the published parameters. The parameters of "NL" were set according to the better of the two sets of parameters provided in [11].

Performance Metric

We wish to give a reward quantity to a saliency map, given some target locations, e.g., in the case of natural images, a set of locations at which human observers fixated. For any one threshold saliency value , one can treat the saliency map as a classifier, with all points above threshold indicated as "target" and all points below threshold as "background". For any particular value of the threshold, there is some fraction of the actual target points which are labelled as such (true positive rate), and some fraction of points which were not target but labelled as such anyway (false positive rate). Varying over all such thresholds yields an ROC curve [14] and the area beneath it is generally regarded as an indication of the classifying power of the detector. This is the performance metric we use to measure how well a saliency map predicts fixation locations on a given image.

B. Human Eye-Movement Data on Image of Nature

In a study by Einhauser et al. [1], human and primate fixation data was collected on 108 images, each modified in nine ways. In the present study, 749 unique modifications of the 108 original images, and

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 1, January 2014

24149 human fixations from [1] were used. Only pictures for which fixation data from three human subjects were available were used. Each image was cropped to 6 0 0 *400 pixels and was presented to subjects so that it took up 76˚*55˚ of their visual field. In order to facilitate a fair comparison of algorithms, the first step of the saliency algorithm, feature extraction (s1), was the same for every experiment.

Two spatial scales (½,¼) were used, and for each of these, four orientation maps corresponding to orientations {0˚,45˚,90˚,135˚}were compute using Gabour filters, one contrast map was computed using luminance variance in a local neighbourhood of size 80*80, and the last map was simply a luminance map (the grayscale values). Each of these 12 maps was finally down sampled to a 25*37 raw feature map."c-s" (center-surround) activation maps were computed by subtracting, from each raw feature map, a feature map on the same channel originally computed at a scale 4 binary orders of magnitude smaller in overall resolution and then resized smoothly to size 25*37. In [2], this overall scheme would be labeled c={2,3}, for ½ and ¼, and ω={4}, corresponding to a scale change of 4 orders. There are also other activation procedures available. Similarly there are also other normalization procedures available. In the below figure an image is taken from the study of Einhauser et al.[1] and saliency map is calculated and the image for that saliency map is obtained from the original image and finally this saliency map is overlayed on the original image.

(a)Original Image (b)Graph Based Saliency Map

(c)Most Salient Parts (d)Saliency Map Overlayed

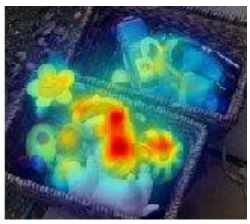


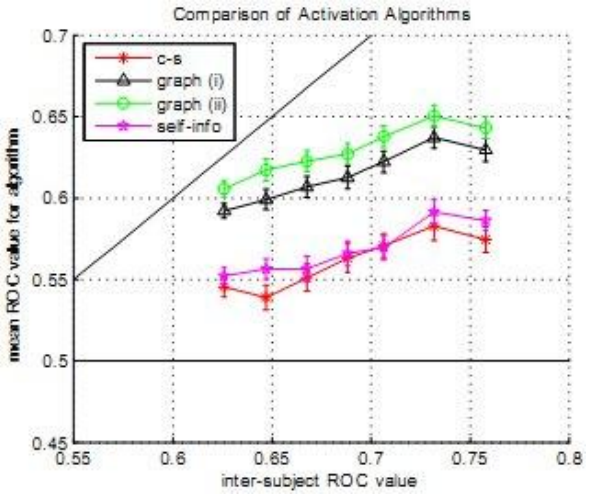
Fig 1:(a)An image taken from the studies of Einhauser et al.[1].(b)The Saliency map formed when using(activation, normalization)=(graph(i),graph(iii)).(c)75% le of Salient parts obtained from the Saliency map.(d)Saliency Map overlayed on the original image.

Einhauser [1] done a study on fixation points on this image by three human subjects. Our result shows a greater similarity to it compared to other traditional algorithms by obtaining the 75% of salient parts over the image. Finally, we show the performance of this algorithm on the corpus of images. For each image, a mean inter-subject ROC area was computed as follows: for each of the three subjects who viewed an image, the fixation points of the remaining two subjects were convolved with a circular, decaying kernel with decay constant matched to the decaying cone density in the retina.

This was treated as a saliency map derived directly from human fixations, and with the target points being set to the fixations of the first subject, an ROC area was computed for a single subject. The mean over the three is termed "inter-subject-ROC value" in the following figures. For each range of this quantity, a mean performance metric was computed for various activation and normalization schemes. For any particular scheme, an ROC area was computed using the resulting saliency map together with the fixations from all 3 human subjects as target points to detect. The results are shown below.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 1, January 2014

(a) Activation Comparison



(b) Normalization Comparison

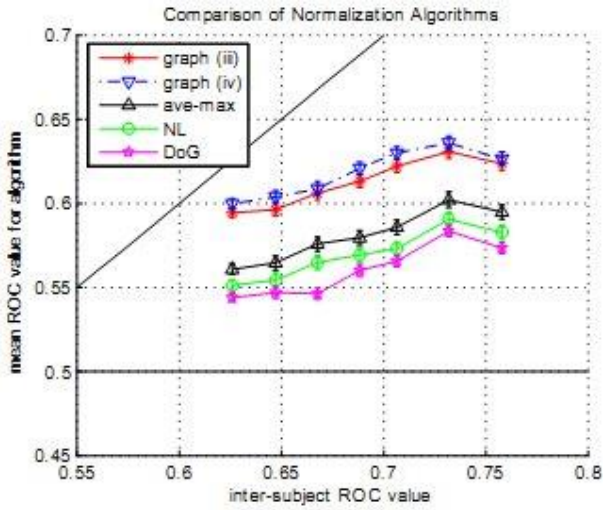


Fig 2:(a) A mean of ROC metric is computed for each range of inter-subject ROC values. Each curve represents a different activation scheme, while averaging over individual image numbers and normalization schemes. (b) A mean ROC metric is similarly computed, instead holding the normalization constant while varying the activation scheme.

In both the figures 2(a) and 2(b), the boundary lines above and below show a rough upper and strict lower bounds on performance (based on a human control and chance performance).These figures clearly demonstrate the tremendous predictive power of the graph-based algorithms over standard approaches.

IV. DISCUSSION AND CONCLUSION

This algorithm can be implemented in the Black fin processor to which the camera is connected and the real-time video images are processed. It guides the visually impaired people in an very efficient manner. Although a novel, simple approach to an older problem is always welcome, we must also seek to answer the scientific question of how it is possible that there are at least two reasons for this observed difference. The first observation is that, because nodes are on average closer to a few center nodes than to any particular point along the image periphery, it is an emergent property that this algorithm promotes higher saliency values in the center of the image plane. We hypothesize that this "center bias" is favourable with respect to predicting fixations due to human experience both with photographs, which

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 1, January 2014

are typically taken with a central subject, and with everyday life in which head motion often results in gazing straight ahead. Notably, the images of foliage used in the present study had no central subject.

One can quantify this algorithm induced center bias by activating, then normalizing, a uniform image using our algorithms. However if we introduce this center bias to the output of the standard algorithms master maps (via point- wise multiplication), we find that the standard algorithms predict fixations better , but still worse than our algorithm. In some cases (e.g.,"DoG"), introducing this center bias only explains 20% of the performance gap to GBVS – in the best case (viz,"max-ave"),it explains 90% of the difference. We conjecture that the other reason for performance difference stems from the robustness of our algorithm with respect to differences in the sizes of salient regions. Experimentally, we find that the "c-s" algorithm has trouble activating salient regions distant from object borders, even if one varies over many choices of scale differences and combinations thereof. Since most of the standard algorithms have "c-s" as a first step, they are weakened ab inito. Similarly the "self-info" algorithm suffers the same weakness, even if one varies over the neighbourhood size parameter. On the other hand, our present algorithm robustly highlights salient regions, even far away from object borders.

We note here that what lacks from this algorithm described as above in any notion of a multiresolution representation of map data. Therefore, because multiresolution representations are so basic, one may extend both the graph-based activation and normalization steps to a multiresolution version as follows: We begin with, instead of a single map A:[n]^2→R, a collection of maps {Ai}, with each Ai:[ni]^2→R representing the same underlying information but at different resolutions. Proceedings as we did before, we instantiate a node for every point on every map, introducing edges again between every pair of nodes, with weights computed same as before with one caveat: the distance penalty function F(a,b) accepts two arguments each of which is a distance between two nodes along a particular dimension. In order to compute F in this case, one must define a distance over points taken from different underlying domains. The authors suggest a definition whereby:(1)each point in each map is assigned a set of locations, (2)this set corresponds to the spatial support of this point in the highest resolution map, and (3) the distance between two sets of locations is given as the mean of the set of pair wise distances. The equilibrium distribution can be computed as before. We find that this extension improves performance with little added computation. Therefore, we have presented a method of computing bottom-up saliency maps which shows a remarkable consistency with the attention deployment of human subjects. The method uses a novel application of ideas from graph theory to concentrate mass on activation maps, and to form activation maps from raw features. We compared our method with established models and found that ours performed favourably, for both of the key steps in our organization of saliency computations. Our model is extensible to multiresolutions for better performance, and it is biologically plausible to the extent that a parallel implementation of the power-law algorithm for Markov chains is trivially accomplished in hardware.

REFERENCES

[1] W.Einhauser, W.kruse,K.P. Hoffmann, & P.Konig "Differences of Monkey and Human Overt Attention under Natural Conditions", Vision Research 2006.

[2] L.Itti, C.Koch, & Niebur "A model of saliency based visual attention for rapid scene analysis", IEEE transactions on Pattern Analysis and Machine 1998.

[3] L.Itti & C.Koch "A Saliency based search mechanism for overt and covert shifts of visual attention", Vision Research, 2000.

[4] L.Itti & P.Baldi "Bayesian Surprise Attracts Human Attention", NIPS*2005.

[5] N.Bruce & J.Tsotsos "Saliency Based on Information Maximization", NIPS*2005.

[6] L.F.Costa "Visual Saliency and Attention as Random Walks on Complex Networks", arXiv preprint 2006.

[7] G.Boccignone, & M.Ferraro "Modelling gaze shift as a constrained random walk", Physica A 331, 207 2004.

[8] D.Brockmann, T.Giesel "Are Human Scan paths Levy flights?",ICANN 1999.

[9] D.Parkhurst, K.Law, & E.Niebur "Modelling the role of salience in the allocation of overt visual attention", Vision Research, 2002.

[10] D.K.Lee, L.Itti, C.Koch, & J.Braun "Attention activates winner-take-all competition among visual features", Natural Neuroscience, 1999.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 1, January 2014

[11] L.Itti, J.Braun, D.K.Lee, & C.Koch "Attention Modulation of Human Pattern Discrimination Psychophysics Reproduced by a Quantitative Model", NIPS*1998.

[12] W.Einhauser & P.Konig, "Does luminance-contrast contribute to saliency map for overt visual attention?", Eur.J.Neurosci. 2003.

[13] U.Rutishauser, D.Walther, C.Koch, & P.Perona "Is bottom-up attention is useful for object recognition?", CVPR 2004.

[14] B.W.Tatler, R.J.Baddeley, & I.D.Gilchrist "Visual correlates of fixation selection: Effects of scale and time."Vision Research 2005.

[15] J.Malik & P.Perona "Preattentive texture discrimination with early vision mechanisms" Journal of the Optical Society of America A 1990.