A Computational Model for Boundary Detection

Gopal Datt Joshi and Jayanthi Sivaswamy

Centre for Visual Information Technology,
IIIT Hyderabad, Andhra Pradesh
India, 500032
gopal@research.iiit.ac.in, jsivaswamy@iiit.ac.in

Abstract. Boundary detection in natural images is a fundamental problem in many computer vision tasks. In this paper, we argue that early stages in primary visual cortex provide ample information to address the boundary detection problem. In other words, *global visual primitives* such as object and region boundaries can be extracted using *local features* captured by the receptive fields. The anatomy of visual cortex and psychological evidences are studied to identify some of the important underlying computational principles for the boundary detection task. A scheme for boundary detection based on these principles is developed and presented. Results of testing the scheme on a benchmark set of natural images, with associated human marked boundaries, show the performance to be quantitatively competitive with existing computer vision approaches.

1 Introduction

Boundary detection constitutes a crucial step in many computer vision tasks. A boundary map of an image can provide valuable information for further image analysis and interpretation tasks such as segmentation, object description etc. Fig. 1 shows an image and the associated boundary map as marked by human observers. It can be noted that the map essentially retains gross but important details in the image. It is hence sparse yet rich in information from the point of scene understanding. Extracting a similar boundary map is of interest in computer vision.

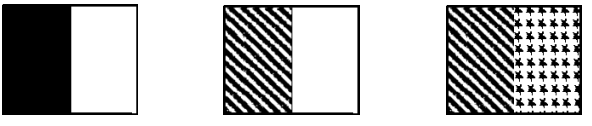
The problem of boundary detection is different from the classical problem of edge detection. A boundary is a contour in the image plane that represents a change in pixel's ownership from one object or surface to another [2]. In contrast, an edge is defined as a significant change in image features such as brightness or color. Edge detection is thus a low-level technique that is commonly applied toward the goal of boundary detection. In general, it is desirable to be able to accurately extract all types of boundaries: for instance those formed between two luminance regions, two textured regions and texture-luminance regions as shown in Fig. 2. There are some attempts in computer vision to address all these attributes completely [2] [3] [4] [5] using complex and computationally intensive schemes. In contrast, humans have an outstanding ability to detect boundaries pre-attentively (fast in nature). This means that the human visual system (HVS)

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 172–183, 2006.

(a) (b)

Fig. 1. (a) Example image (b) Human-marked segment boundaries. Image shows boundaries marked by 4-8 observers. The pixels are darker where more observers marked a boundary [1].

is capable of extracting all important boundary information in its *early* stages of processing. Studying the visual mechanisms underlying these tasks can provide an alternative solution to the boundary detection problem. It may also lead to simple and fast scheme for boundary detection in computer vision.



(a) (b) (c)

Fig. 2. Types of boundaries

Some attempts have been made to model boundary detection in HVS. One such model assumes that saliency of boundaries arise from long-range interaction between orientation-selective cortical cells [6]. This model accounts for a number of experimental findings from psychophysics but its performance is unsatisfactory on natural images and it is computationally intensive. Another model emphasises the role of local information and focuses on cortical cells which are tuned to bar type features [7] [8]. It extracts edge information which is followed by an assessment based on the local context. These models have been shown to perform well on natural images but are incapable of detecting boundaries formed by texture regions (shown in Fig. 2(b), 2(c)). In this paper, we present a computational model for the boundary detection functionality of the HVS which can extract all types of boundaries. We present results of testing this model on a set of benchmarked images where boundaries are marked by human observers.

The presentation in this paper is organised as follows. In the next section, we review the relevant neurophysiological and psychophysical findings in early stages of the HVS and end with a proposal for a computational model for boundary detection. In section 3, a computational scheme is developed based on the proposed computational model. In section 4, the performance of the proposed scheme is compared against human marked boundaries followed by some concluding remarks.

2 Computation Model for Boundary Detection

Real world images are processed in our visual system to produce boundaries. These images are characterised by colour, texture [1] and non-texture (only regular luminance/colour based) regions. Thus, boundaries can arise due to the adjacency of any of these regions in natural images. Some of these that can occur in grey scale images (which is the focus of this paper) are shown in figure 2: luminance-luminance or LL boundary, texture-luminance or TL boundary, and texture-texture or TT boundary.

Any image point can be declared as a boundary only after understanding its local context. By context is meant a characterisation of the local surround in terms of luminance and texture. In the early stages of HVS, there is evidence that the derived representation provides enough texture and non-texture information to address boundary detection effectively. At the retinal level, visual input (image) is filtered by ganglion cells whose local classical receptive field's (CRF's)[2] are a close fit to a Laplacian of Gaussian [9]. Thus the representation derived at the retinal level is an edge map. The results of this processing form direct input to Lateral Geniculate Nucleus (LGN) in the mid-brain. This area has no known filter function but serves mainly to project binocular visual input to various sites, especially to the visual cortex. The cells found in this area have a functional role similar to that of the retinal ganglion cells except that they also perform binocular mapping. In our work, we ignore binocular details associated with the LGN cells.

The Ganglion and LGN cells are classified into two classes, known as P and M-cells [10] [11] [12]. The P-cells have smaller receptive fields and signal high spatial frequencies in the image while the M-cells have (2-3 times) larger receptive fields and they cannot resolve high spatial frequencies [10] [12] [13]. We can infer that P-cells strongly respond to fine and coarse edges whereas M-cells respond to coarse edges and quite poorly to fine edges. At this stage of HVS, there is not much information associated with any detected edge to declare it as a boundary point. When we consider a texture patch for example, the M-cells respond to its contour while the P cells respond to its contour as well as any edges arising from the texture elements within the patch. Hence, there is an ambiguity in determining if an edge belongs to a texture region or not based on the cell responses. Thus, it is difficult to separate out texture and non-texture information effectively at this stage of HVS. Such a situation however, gets resolved in the cortical level which is the next stage of the HVS, called as area V1.

The cortical cells in area V1 are sensitive to some new attributes like orientation. Furthermore, their sensitivity to edge features becomes more specialised compared to the LGN cells. Hubel and Wiesel [14] distinguished between *simple* and *complex cells* in cat primary visual cortex (area $V1$) that are selective to

[1] It is a spatial structure characterising, apart from colour and the gray level, the visual homogeneity of a given zone of an image.

[2] The receptive field is, by definition, the visual area within which one can activate an individual neuron.

intensity changes in specific orientation (oriented edge features). Although complex cells have many properties in common with simple cells, including orientation selectivity, their defining feature is that a suitably oriented pattern will elicit a response no matter where it lies in the receptive field [14] [15]. This property is known as "phase invariance". Although, simple and complex cells bring orientation selectivity in feature detection, their response to texture and non-texture patterns is ambiguous, similar to LGN cells. There are some other cells in area V1 having more specialised behavior like *bar cells*, *grating cells* and *end- stopped cells.* [16] [17] [18]. These cells are more specialised forms of complex cells.

The bar and grating cells play an important role in boundary detection [18] [19]. It is important to know their characteristics and inter-connections with the previous stages. These cells mostly get their input from the M- cells of area LGN [10] [11] [20] [21]. The grating cell responds only to a texture edge and not to any isolated edge or line [18] [19]. On the other hand, a bar cell responds only to an isolated edge or line but does not respond to any texture edge [7]. Hence, it is possible to disambiguate between an edge belonging to a textured region and a non-textured region.

To summarise, the HVS appears to use a principle of increasing functional specialisation to enable certain features of the visual pattern to become more explicit in successive stages of processing [9]. Such functional specialisation serves to resolve the ambiguity present in the previous stages.

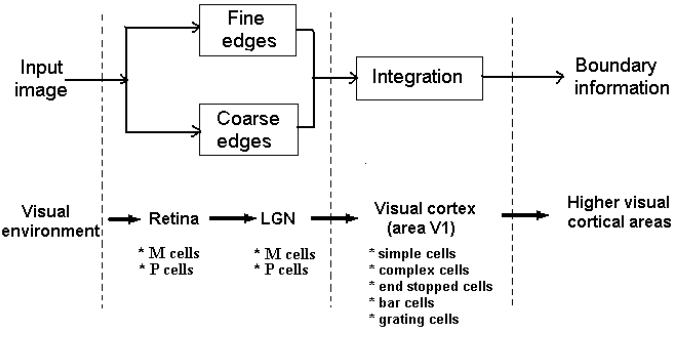


Fig. 3. Computational model for boundary detection and corresponding processing stages in HVS

Based on the above findings, we propose a computational model for boundary detection (given in Fig. 3): The visual input (or an image) is processed by P and M types of ganglion and LGN cells in order to extract redundant subsets of fine and coarse edges. This information is passed to the next stage (area V1) where the bar and grating cell operators help extract texture and non-texture information. The boundaries that are formed by texture and non-texture regions are extracted via an integration process that combines the outputs of the bar and grating cells. The output of integration is usable in any high level task. Next, we present an implementation scheme based on the proposed model.

3 Implementation

3.1 Image Representation at LGN Level

At the retinal level, *ganglion cells* signal the spatial difference in the light intensity falling upon adjacent locations in the retina. At the output of this stage (retina), the visual system provides an efficient representation in terms of fine edge locations. This *fine* edge map can be computed using any standard gradient-based edge detector. Assuming the gradient is computed in two orthogonal directions x and y. The gradient map G be for a given input image I is:

$$G(x, y) = \sqrt{(I_x^2 + I_y^2)} \tag{1}$$

where I_i is partial derivative of image in i direction. The P-type cells produce a response similar to the gradient map and extract fine edges in the image whereas, M-cells are tuned for coarse edge features. We derive such characteristics by using local surround. For every point p in an image, we consider its surround and associate with the point a histogram of the surround which we call as the *Photoreceptor Histogram* (h_p). For computational purpose, the surround is taken to be be a window of fixed size. The Photoreceptor Histogram (h) is a K-long vector where K is the maximum no. of grey levels in the image. The histogram operation ignores spatial details and captures coarse details within a local surround which is actually relevant to get a boundary details. Such details can lead to the detection of coarse edges similar to the M-type cells. Here, we do not present the detection of such edges (as it is not of use) but it can be easily obtained by a sum of gradient values computed at every element of the transformed vector (h_p).

3.2 Image Representation at Area V1

In area V1, cells gain orientation selectivity and exhibit more specialised behavior towards texture and non-texture patterns. In the context of boundary detection, bar and grating cells are more useful as they provide unambiguous information about such patterns.

Bar Cells. A bar cell responds most strongly to a single bar stimulus, such as a line or edge, in its receptive field and it has a reduced response when more bars are present in the surrounding region of the stimulus. In natural images, it is equivalent to a detector which responds only to isolated edges and not to edges which belong to a texture region [7]. Such a characteristic can be achieved by a surround (local) assessment of P-type LGN cell response. This notion is called *surround inhibition* which models intra-cortical interaction among cells.

For a given point in the image, the inhibition term is computed in an annular area around it. Let a filter function $g_\sigma(x, y)$ be defined as follows (inverse of $g_\sigma(x, y)$ is shown in Fig. 4):

$$g_\sigma(x, y) = \frac{1}{||P(DoG)||} P(DoG(x, y)) \tag{2}$$

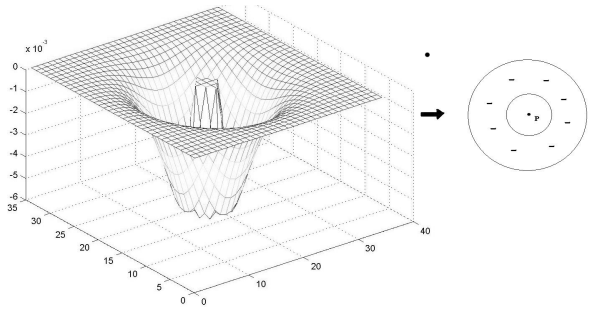


Fig. 4. Inhibition function which models the contribution of the surround (2D and 1D profile)

where $P(x) = \frac{x+|x|}{2}$ is a rectification operator, $|.|$ denotes modulus, $DoG(x, y)$ is a difference of Gaussian functions with standard deviations σ in the ratio of $n : 1$ for some integer n; and $||.||$ denotes the L_1 norm. The surround influence is applied to the gradient image G (obtained from equation 1) as follows:

$$S(x, y) = (G * g_\sigma)(x, y) \tag{3}$$

A bar cell response E_α at a location (x, y) is then obtained as:

$$E_\alpha(G(x, y)) = P(G(x, y) - \alpha.S(x, y)) \tag{4}$$

where the factor α controls the strength of the influence by the surround inhibition. If there is no texture surrounding (i.e., there is an isolated edge) a given image point, the response at that point will be equal to the gradient value as there will be no inhibition. However, if there are other edges in the surrounding region, the inhibition term $S(x, y)$ will become strong enough to cancel completely, the contribution of the gradient term. This model for bar cells provides a contour representation for any given input by discarding irrelevant edges within texture regions. In the later stages of our boundary detection scheme, we will use this functional model by the name *surround inhibition*.

Grating Cells. The grating cells are responsible for texture processing in the early stages of HVS. These cells respond strongly to a grating (periodic pattern of edges) of specific orientation, periodicity and position and not to isolated edges [18]. The role of the grating cells as a texture operator has been established in [19]. Texture regions are distinguishable based on the distribution of the edges within. Using this fact, we can define a similarity measure between two texture regions. Such a measure is useful to determine any boundary between two texture regions. For instance, any point which lies in between two texture regions which are dissimilar can be declared as a boundary point. A measure of such similarity is therefore of interest. Given two photoreceptor histograms h_{p1} and h_{p2}, we use the χ^2-statistic [22] to define a (dis-)similarity measure:

$$\chi^2(h_{p1}, h_{p2}) = \sum_{k=1}^{K} \frac{[h_{p1}(k) - h_{p2}(k)]^2}{[h_{p1}(k) + h_{p2}(k)]} \tag{5}$$

where $k \in [1, K]$ is the intensity level. Two similar texture regions will cause the numerator of the above expression to diminish, and hence the similarity measure to be low. In natural images, the diverse nature of the texture regions results in a wide range of variability in the above measure. To address this problem we transform these values to fit in the range of $0 - 1$ in such a way as to emphasise only low values as follows.

$$R(\chi^2) = e^{-\frac{(\chi^2)^2}{2\tau^2}} \tag{6}$$

where τ is a parameter that controls the level of penalty. The similarity measure can be used to determine if a point lies on TL and TT types of boundaries by considering the histograms of points located to its left and right. Next, we integrate the information extracted up to this point in order to obtain boundaries.

3.3 Integration: A Scheme for Boundary Detection

This stage integrates information gathered from the P and M cells of LGN as well as the bar and grating cells of V1. Let $\chi^2 = [(\chi_x^2)^2 + (\chi_y^2)^2]^{\frac{1}{2}}$ where, χ_x^2 and χ_y^2 are computed along the x and y-axis, respectively. The integration is achieved as follows:

$$\tilde{B}(x, y) = \gamma.G(x, y) + \beta.\chi^2(x, y) \tag{7}$$

where, γ and β are appropriate weights. The integration scheme has two subparts with each part contributing to the extraction of specific types of boundaries: the first part will be a maximum at the location of a LL boundary whereas the second part will be a maximum at TL and TT boundaries. To determine how the weights are to be assigned, let us re-examine the first part. This term will also be significant within the texture regions of TL and TT boundary which needs to be suppressed. This can be partially achieved by choosing the weight γ to be dependent on the texture measure χ^2 as follows: $\gamma = 1.0 - R(\chi^2)$. This choice of weight ensures that the first term nearly vanishes in equation 7 when edges are formed due to sub-patterns in a texture region. The weight β can simply be a scalar.

In principle, equation 7 signals (with a maximum) texture boundaries and edges. Of these, to extract only boundary points due to all types of boundaries, we need to further suppress the response for edges within texture regions. This is accomplished by applying *surround inhibition* (E_α) as found in bar cells.

$$B(x, y) = E_\alpha(\tilde{B}(x, y)) \tag{8}$$

Next, we present the results of testing the proposed scheme on natural images and evaluate the same against human-marked boundaries.

4 Performance Evaluation and Results

Most of the methods for the evaluation of edge and boundary detectors use natural images with associated desired output that is subjectively specified by the human observer [1] [4] [7]. We tested the performance of the proposed scheme by applying a *precision-recall (PR)* framework using human-marked boundaries from the Berkeley segmentation dataset [1] as ground truth. The segmentation dataset contains 5-10 segmentations for each image. The dataset has training images and testing images. The training images were not used as there is no training involved in our scheme and hence the evaluation was done only on the test image set (100 images) which consisted of both indoor and outdoor scenes.

The precision-recall curve is a parametric curve that captures the trade off between accuracy and noise as the detector's threshold varies. *Precision* is the fraction of detections that are true positives rather than the false positives, while *recall* is the fraction of true positives that are detected rather than missed. The PR curves are hence appropriate for quantifying boundary detection. The PR measures are particularly meaningful in the context of boundary detection when we consider applications that make use of boundary maps, such as stereo or object recognition. It is useful to characterise a detector in terms of how much true signal is required to succeed R (recall), and how much noise can be tolerated P (precision). A method to determine the relative cost μ between these quantities for a given application is given in [2]. We follow the same and use the *F-measure* (proposed therein) which is defined as

$$F = PR/(\mu R + (1 - \mu)P) \tag{9}$$

The location of the maximum F-measure along the curve provides the optimal threshold for an application for a desired μ, which we set to be 0.5 in our experiments. When a single performance measure is required or is sufficient, precision and recall can be combined with the F-measure. The F-measure curve is usually unimodal, so the maximal F-measure may be reported as a summary of the detectors performance.

Precision and recall are appealing measures, but to compute them we must determine which true positives are correctly detected, and which detections are false. We have used the correspondence algorithm presented in [2] to compute true and false detection using output boundary map and available ground truths. In summary, given the computed boundary map, we compute the points on the precision-recall curve independently by first thresholding the output image to produce a binary boundary map and then matching this computed boundary map against each of the human boundary maps in the ground truth segmentation data set.

In our scheme, the following parameter values were empirically chosen to obtain best results. Once chosen, they were fixed to remain constant for all 100 test images. The window sizes in bar and grating cells' functional modelling were 7×7 and 15×15, respectively. The value for β was chosen to be 0.6 and value of α was 0.1. In equation 5, the intensity level was quantised from 256 to

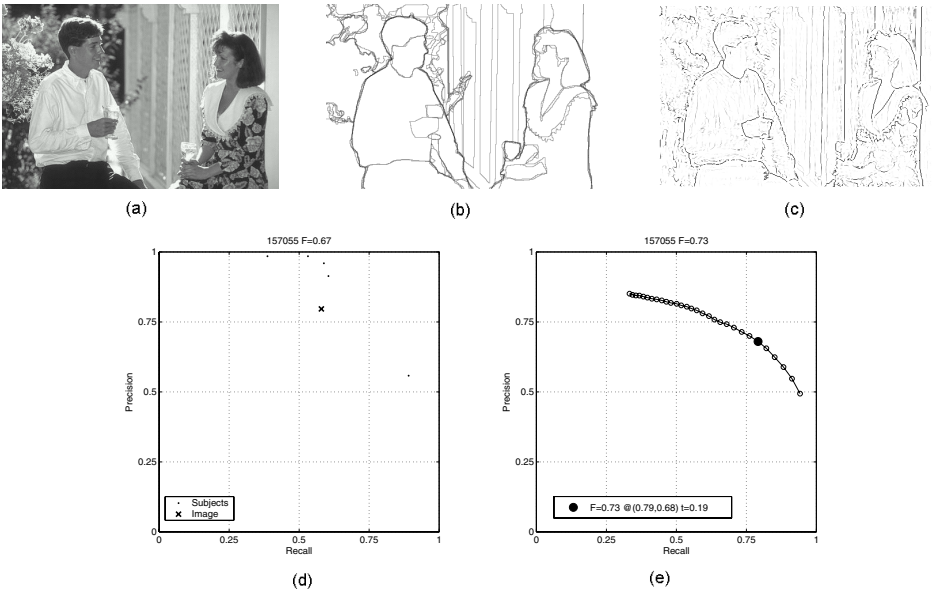


Fig. 5. (a) Sample test image, (b) Associated observer-segmented results, (c) Extracted boundary map, (d) PR curves for each observer relative to the rest.(e) PR curve for the proposed scheme. The curve is scored by its maximal F-measure, the value and location of which are shown in the legend.

Table 1. Comparison of proposed scheme with other schemes

Method	Performance
Brightness and texture gradient	0.63
Brightness gradient	0.60
Proposed scheme	**0.59**
Texture gradient	0.58
Multi-scale gradient magnitude	0.58
Second moment matrix	0.57
Gradient magnitude	0.56

32 as it had no effect on the value of χ^2 while it helped greatly minimise the computation. Fig. 5 shows the performance of the proposed scheme on a sample test image. It provides a comparison of the scheme against human observers. The points marked by a dot on the figure 5(d) show the precision and recall of segmentation by each human observer relative to other observers(a total of five). The median F-measure for the observers is 0.67 while the maximum obtained value using the proposed scheme is 0.73 indicated by a big dot in the PR curve (in fig 5(e)). The scheme was tested on a test dataset of 100 images and the overall performance was computed using a bench-marked algorithm [2] which gives a score based on the obtained results. The obtained score is 0.59 (shown in table. 1). Some of obtained soft boundary maps are shown in the Fig. 6.

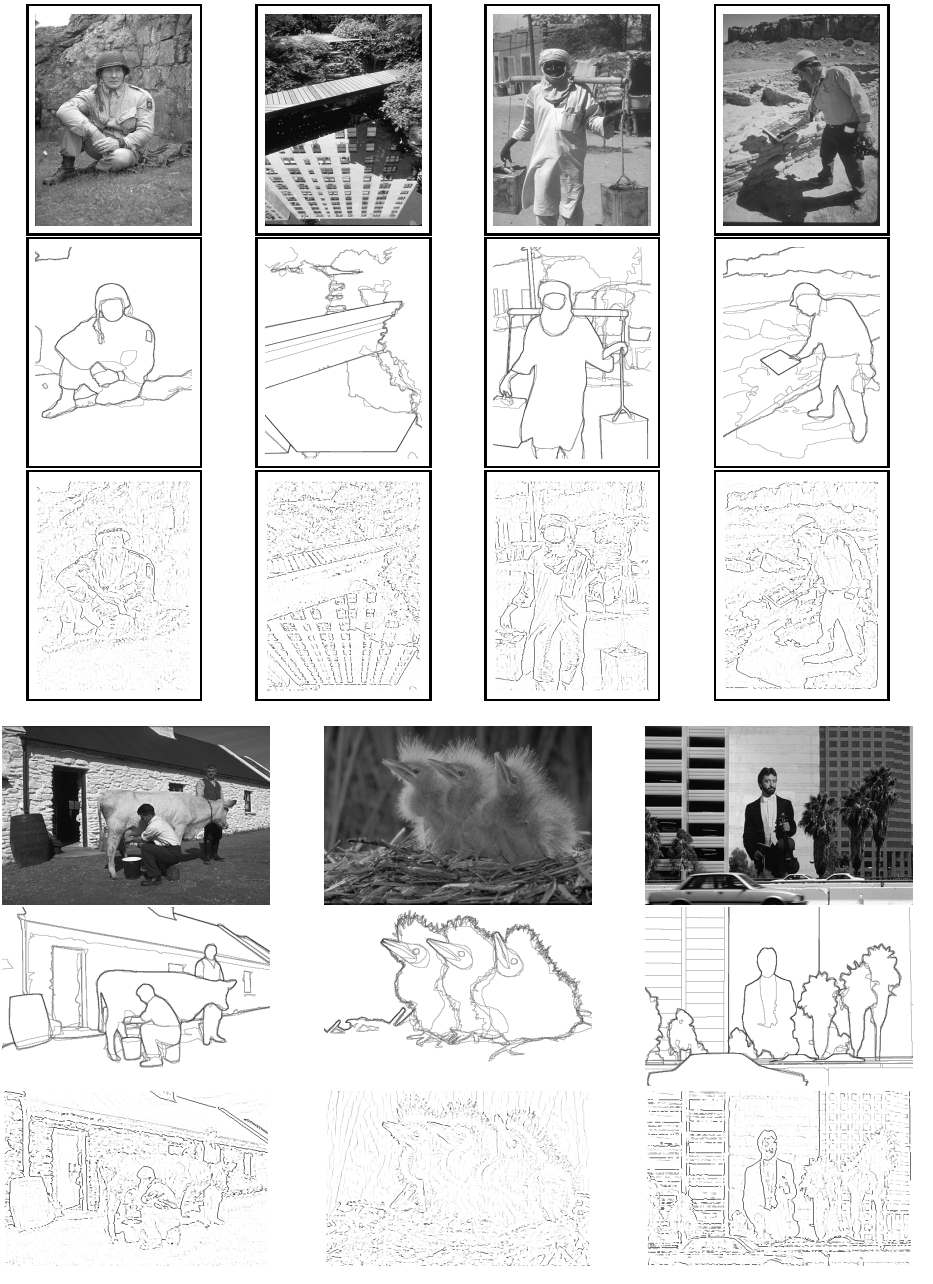


Fig. 6. Sample test images, their corresponding ground truth and obtained results from the presented boundary detection scheme. First row shows the original images; second row shows the corresponding ground truth images; third row shows obtained soft boundary map. In the soft boundary map, intensity of the boundaries varies from 0-1.

Our scheme was also assessed against the existing boundary detection approaches reported in [2] using the same Berkeley dataset. Table 1 shows the relatives scores. The top two methods differ from the proposed scheme in the types of texture and non-texture features used and in the complex manner in which they are processed to compute boundary maps. In general, all the reported methods use training images for tuning parameters to obtain the best boundary map. In contrast, our scheme is simpler and the reported performance was achieved without any training. The latter is an attractive feature. In short, the performance of the proposed scheme is reasonably good.

5 Discussion and Conclusion

Evidence that complex cells receive direct input from the LGN cells in addition to simple cells [10] [11] [20] [21] is significant in terms of understanding the computations performed in V1. However, this has generally not received much attention in the computational modelling literature. It appears that the early stages in primary visual cortex provide ample information to address the boundary detection problem. The richness of information emerges from the capability of the HVS to extract global visual primitives from local features with no top-down influence.

A model for boundary detection based on these principles has been developed and presented. The model is useful for computing boundary points in images with performance which is competitive with existing computer vision approaches. It is also computationally simpler than most of the existing approaches to boundary extraction.

The functions of individual cells found in HVS have been modelled at a fixed single scale. However, evidence for multi-scale processing exists in the form of cortical cells of different sizes. Our initial attempt has been limited to understand the kind of processing and interaction carried out by the cells of fixed size. The model can be enhanced by extending it to a multi-scale framework and by including colour information.

References

1. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Proc. of International Conference on Computer Vision (2001)
2. Martin, D., Fowlkes, C., Malik, J.: Learning to detect natural image boundaries using brightness and texture. IEEE Transactions on Pattern Analysis and Machine Intelligence **26 (5)** (2004) 530–549
3. Ma, W.Y., Manjunath, B.S.: Edgeflow: A technique for boundary detection and segmentation. IEEE Transactions on Image Processing **9 (8)** (2000) 1375–1388
4. Malik, J., Belongie, S., Leung, T., Shi, J.: Contour and texture analysis for image segmentation. International Journal of Computer Vision **42 (1)** (2001) 7–27
5. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22 (8)** (2000) 888–905

6. Yen, S., Finkel, L.: Extraction of perceptually salient contours by striate cortical networks. Vision Research **38 (5)** (1998) 719–741
7. Grigorescu, C., Petkov, N., Westenberg, M.: Contour detection based on nonclassical receptive field inhibition. IEEE Transactions on Image Processing **12 (7)** (2003) 729–739
8. Joshi, G.D., Sivaswamy, J.: A simple scheme for contour detection. Proc. of the Conference on Computer Vision Theory and Applications (2006) 236–242
9. Marr, D., Hildreth, E.: Theory of edge detection. Proceedings of the Royal Society of London, Series B **207** (1980) 187–217
10. Hoffmann, K.P., Stone, J.: Conduction velocity of afferents to cat visual cortex: a correlation with cortical receptive field properties. Brain Research **34** (1971) 460–466
11. Martinez, L., Alonso, J.M.: Complex receptive fields in primary visual cortex. The Neuroscientist **9(5)** (2003) 317–331
12. V.Bruce, Green, P.R., Georgeson, M.A.: Visual Perception: physiology, psychology and ecology. Fourth edition, Psychology Press (2004)
13. Lennie, P., Trevarthen, C., Essen, D.V., Wassle, H.: Parallel processing of visual information. Visual Perception-The Neurophysiological Foundations, Academic Press, San Diego **92** (1990)
14. Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in the cats visual cortex. Journal of Psychology **160** (1962) 106–154
15. Alonso, J.M., Martinez, L.M.: Functional connectivity between simple cells and complex cells in cat striate cortex. Nature Neuroscience **1(5)** (1998) 395–403
16. Baumann, R., van der Zwan, R., Peterhans, E.: Figure-ground segregation at contours: a neural mechanism in the visual cortex of the alert monkey. European Journal of Neuroscience **9 (6)** (1997) 1290–1303
17. Dobbins, A., Zucker, S.W., Cynader, M.S.: Endstopped neurons in the visual cortex as a substrate for calculating curvature. Nature **329 (6138)** (1987) 438–441
18. von der Heydt, R., Peterhans, E., Drsteler, M.R.: Grating cells in monkey visual cortex: coding texture? Channels in the Visual Nervous System: Neurophysiology, Psychophysics and Models (Blum B, ed) (1991) 53–73
19. Kruizinga, P., Petkov, N.: Nonlinear operator for oriented texture. IEEE Transactions on Image Processing **8 (10)** (1999) 1395–1407
20. Alonso, J.M.: The microcircuitry of complex cells in cat striate cortex. Society for Neuroscience **22(198.1)** (1996) 489
21. Mel, B.W., Ruderman, D.L., Archie, K.A.: Translation-invariant orientation tuning in visual *Complex Cells* could derive from intradendritic computations. The Journal of Neuroscience **18(11)** (1998) 4325–4334
22. Liu, X., Wang, D.: A spectral histogram model for textons and texture discrimination. Vision Research **42 (23)** (2002) 2617–2634