A SIMPLE SCHEME FOR CONTOUR DETECTION

Gopal Datt Joshi
Center for Visual Information Technology, IIIT Hyderabad
Gachibowli, Hyderabad, India-500032

Jayanthi Sivaswamy
Center for Visual Information Technology, IIIT Hyderabad
Gachibowli, Hyderabad, India-500032

Keywords: Contour detection, Surround suppression, Primary visual cortex, Human visual system.

Abstract: We present a computationally simple and general purpose scheme for the detection of all salient object contours in real images. The scheme is inspired by the mechanism of surround influence that is exhibited in 80% of neurons in the primary visual cortex of primates. It is based on the observation that the local context of a contour significantly affects the global saliency of the contour. The proposed scheme consists of two steps: first find the edge response at all points in an image using gradient computation and in the second step modulate the edge response at a point by the response in its surround. In this paper, we present the results of implementing this scheme using a Sobel edge operator followed by a mask operation for the surround influence. The proposed scheme has been tested successfully on a large set of images. The performance of the proposed detector compares favourably both computationally and qualitatively, in comparison with another contour detector which is also based on surround influence. Hence, the proposed scheme can serve as a low cost preprocessing step for high level tasks such shape based recognition and image retrieval.

1 INTRODUCTION

Contour detection in real images is a fundamental problem in many computer vision tasks. Contours are distinguished from edges as follows. Edges are variations in intensity level in a gray level image whereas contours are salient coarse edges that belong to objects and region boundaries in the image. By salient is meant that the contour map drawn by human observers include these edges as they are considered to be salient. However, the contours produced by different humans for a given image are not identical when the images are of complex, natural scenes. In such images, multiple cues are available for the human visual system (HVS) - low level cues such as coherence of brightness, texture or continuity of edges, intermediate level cues such as symmetry and convexity, as well as high level cues based on recognition of familiar objects. Even if two observers have exactly the same set cues, they may choose contours at varying levels of granularity. Thus saliency of an edge is a subjective matter and varies accordingly. Nevertheless, the fact remains that a contour map drawn by human observers is sparser than an edge map derived by processing the digital image. This can be seen from Fig 1(a) which shows a test image and the corresponding ground truth data (Fig. 1(b))indicating the contours considered relevant by a human observer (dat, 2003). If we compare this with the edge maps in Fig. 1(c), (d) extracted by a Canny detector we can observe that the contour map is sparse. This is despite selecting a low scale (to capture gross information) and using two different thresholds for the edge detection. In general, a contour map is an efficient representation of an image since it retains only salient information and hence is more valuable for high level computer vision tasks. The design of a detector that can extract all contours from a wide range of images is therefore of interest.

The key to extracting contours appears, from the ground truth, to be the ability to assess what is relevant and what is not in a local neighbourhood. For instance, the grassy texture has been rejected in the ground truth while the edges defining the elephants' feet have been retained. An assessment-based strategy has been attempted to contour detection using local information around an edge such as image statistics, topology, texture, colors, edge continuity, density, etc. Specifically, these approaches have used statistical analysis of gradient field (Meer

Datt Joshi G. and Sivaswamy J. (2006).
A SIMPLE SCHEME FOR CONTOUR DETECTION.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 236-242
DOI: 10.5220/0001374702360242

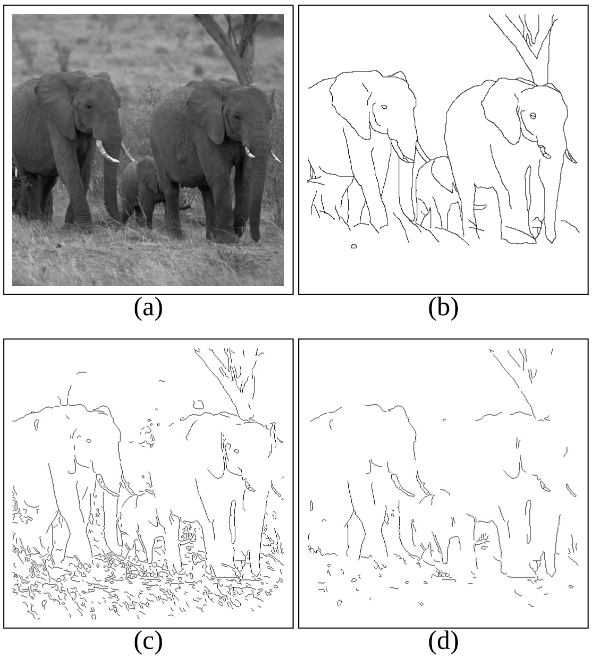

Figure 1: Demonstration of texture as a problem in contour detection process. (a) Image of elephants (b) Ground truth image. Canny edge map using low scale and low threshold, (c) low scale and high threshold.

and Georgescu, 2001), anisotropic diffusion (Perona and Malik, 1990; Black et al., 1998), complementary analysis of boundaries and regions (Ma and Manjunath, 2000) and edge density information (Dubuc and Zucker, 2001). These approaches, by design, are very extensive in computation.

The HVS is capable of extracting all important contour information in its early stages of processing. Some attempts have also been made to model contour detection in the HVS. One such model assumes that saliency of contours arise from long-range interaction between orientation-selective cortical cells (Yen and Finkel, 1998). This model accounts for a number of experimental findings from psychophysics but is computationally intensive and its performance is unsatisfactory on real images. Another model also emphasises the role of local information and focuses on cortical cells which are tuned to bar type features (Grigorescu et al., 2003). This scheme computes oriented Gabor energy at a single scale and over twelve different orientations followed by a non-classical receptive field (non-CRF)[1] inhibition. Results of testing of this scheme on images of animals in their natural habitat, are reasonably good. However, the scheme is computationally expensive and produces a contour map which is quite sparser than an edge map though not as sparse as the ground truth (contour map). In this paper, we seek to find a solution to contour detection which is computationally low in cost as well as effective on a wide range of images including natural images. The paper is organized as follows: section 2 presents the development and details of the scheme; section 3 proposes an implementation of the scheme; section 4 summarises the results and section 5 draws some conclusions based on the performance of the proposed scheme.

2 PROPOSED SCHEME

The human visual system, in its early stages of processing, differentiates between isolated boundaries such as object contours and region boundaries, on the one hand, and edge in group, such as those in texture, on the other hand. This is accomplished in a series of processing stages. At the retinal level, the ganglion cells process the visual input from rods and cones to produce an image similar to that of an edge detector used in computer vision (Marr and Hildreth, 1980). The ganglion cells signal the spatial differences in the light intensity falling upon the retina. Their receptive field is organized into a center-surround fashion, in which the excitatory and inhibitory subfields are integrated into circularly symmetric regions. The classical work in (Marr and Hildreth, 1980) modelled this receptive field with a Laplacian of Gaussian function. At the output of this stage (retina), the visual system provides an efficient representation of the image by removing redundant information such as uniform light intensity on adjacent retinal locations. The axons of the ganglion cells project to an area in the brain called the Lateral Geniculate Nucleus (LGN). This area has no known filter function but serves mainly to project binocular visual input to various sites, especially to the visual cortex.

In the visual cortex, Hubel and Wiesel (Hubel and Wiesel, 1962) found *simple* and *complex cells* in cat primary visual cortex (area V1) that are selective to intensity changes in specific orientation. These orientation-selective cells are organized in columns, in which all cells in a column have the same preferred orientation, and adjacent columns have incremental change in orientations (Hubel and Wiesel, 1962). The orientation selectivity of cells is accomplished through a spatial summation of the inputs from LGN. This leads to interesting features. For example, complex cells differ from simple cells by showing less specificity regarding the position of the

[1]The classical receptive field (CRF) is, by definition, the area within which one can activate an individual neuron. The region beyond this area which can modulate the response of the concerned neuron is called a non-classical receptive field.

stimulus. Computationally, this functionality can be modeled using input from the simple cells.

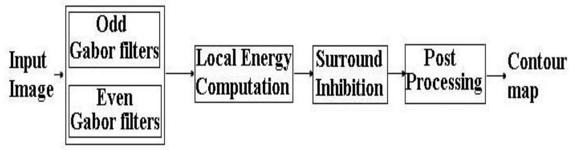

Figure 2: Contour detection scheme in (Grigorescu et al., 2003).

Other classes of cells besides the simple and complex cells also exist in V1. These are the end stopped cells (Dobbins et al., 1987), bar (contour) cells (Baumann et al., 1997), and grating cells (von der Heydt et al., 1991). Neurophysiological measurements on cells have showed that the response of an orientation-selective neuron to an optimal stimulus in its receptive field is reduced if the stimulus extends to the surround. This effect is termed as *surround inhibition* and it is exhibited in a majority (80%) of the orientation-selective cells in the visual cortex of primates(Knierim and van Essen, 1992). In general, an orientation-selective cell with surround- inhibition will respond most strongly to a single bar, line, or edge in its receptive field and will show reduced response when more bars are added to the surroundings like sine wave gratings. These type of cells was called the *bar cell*, referring to the preference of the cell for bars versus gratings. These cells were the source of inspiration for the contour detector reported in (Grigorescu et al., 2003). The detection scheme is based on fitering the input image with a Gabor filter bank; summing the Gabor energy output at 12 different orientations and then applying surround inhibition. This scheme is shown in Fig. 2. The Gabor filtering stage is essentially a local energy computation stage (Morrone and Burr, 1988). Edge and line features are signalled by local maximas in the local energy map. Hence, the input to the surround inhibition stage in Fig. 2 is edge information.

Edge information can also be derived using a gradient computation. The difference between edge detection using gradients as opposed to local energy is that the latter is capabile of detecting step/impulse discontinuities as well as ramps and other luminance profiles in the image. Most of the edges in natural images have step profiles which can be effectively picked by the gradient computation. A drawback of Gabor based commutation is that it leads to a poor localization since the window over which it is computed needs to be wide enough to attain a good fit to the Gabor profile. Furthermore, local energy computation is far too expensive (12×2=24 filters) compared to gradient computation which needs to be done only in two orthogonal directions (2 filters) to determine edges at various orientation. Hence, we argue that a simpler scheme for contour detection would be to derive the edge information from a gradient computation followed by an assessment based on the local context.

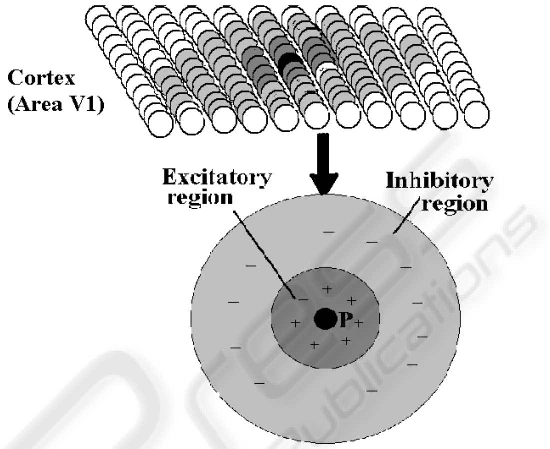

Figure 3: Neighboring cells profile of a cortical cell in area V1.

Next, we turn our attention to the second part of the contour detection scheme, namely assessment of the edge information. This assessment needs to be done based on the local context and a surround inhibition mechanism has been used for this purpose in (Grigorescu et al., 2003). A recent neurophysiological study (Cavanaugh et al., 2002) of cells in V1 has however shown that the surround region actually can have an excitatory influence in addition to an inhibitory influence. Specifically, the findings about neuronal behaviour can be summarised as follows: (i) Every cell responds to a stimulus if it falls on its central region (CRF). This is shown by the circle P in Figure 3; (ii) Besides the CRF, a neuron has a surround region which is made up of two parts, namely an inner annular shaped region (shown in dark gray in Figure 3) which is excitatory and an outer annular shaped region (shown in light grey in Figure 3) which is inhibitory; (iii) The surround region which can influence the response of a neuron is limited in extent. The behavior of these neurons give us a clue about the role of local context in the visual perception of a stimulus which is obtained by combining excitatory and inhibitory influences. We draw inspiration from this study and propose a surround mechanism with excitatory *and* inhibitory components. An contour detection scheme which computes the gradient information first and then employs this second step will not respond (sufficiently) to edges which belong to texture regions. Such a scheme is easy to implement as well.

3 IMPLEMENTATION OF SCHEME

The proposed scheme for contour detection can be implemented in a simple manner as follows. First, compute a gradient map. Next, incorporate the surround influence on the gradient map using a mask operation. Finally, binarise the output of the second stage using a standard procedure.

3.1 Gradient Estimation in the Discrete Domain

As mentioned earlier , it suffice to perform the gradient computation in two orthogonal directions using Sobel or Prewitt masks.

3.2 Surround Interaction

The surround influence can be implemented by either taking into account the direction of the gradient or ignoring the same. The former will lead to less amount of surround supression in natural images because they generally contain texture edges at arbitrary orientations. In other words, derive contour map will be noisy. An surround influence operation which ignore the edge orientation can on the other hand, improve the level of supression. Hence, the edge assesment based on the surround influence can be implemented as a convolution operation with an appropriate isotropic mask. We will now explain how this mask can be designed. As explained in the previous section, this surround consists of two annular regions, the inner being excitatory and the outer being inhibitory in nature. Hence, in the mask, we will define two annular regions surrounding the central pixel. The inner one is assigned positive weights (excitatory) while the outer one is assigned negative weights (inhibitory). The magnitude of the weights is chosen based on the following neurophysiological findings: the influence of a point in the surround on the cell response is dependent on its distance from the centre and this is roughly Gaussian in profile. Furthermore, the excitatory influence is weaker than the inhibitory influence (Series et al., 2003). The size of the required mask was determined by varying it from 7×7 to 15×15. It was found that a mask of size 7×7 is optimal to achieve the best results. Hence, we propose a 7×7 mask as shown in Figure. 4. The weights in the outer inhibitory region of the optimal mask was found by sampling a Gaussian profile of standard deviation 1.4. The sum of the weights inside this mask is set at 0.52 to help enhance a contour pixel.

$\frac{1}{150}$

-5	-6	-5.5	-5	-5.5	-6	-5
-6	-5	-2	0.5	-2	-5	-6
-5.5	-2	0.4	0.4	0.4	-2	-5.5
-5	0.5	0.4	225	0.4	0.5	-5
-5.5	-2	0.4	0.4	0.4	-2	-5.5
-6	-5	-2	0.5	-2	-5	-6
-5	-6	-5.5	-5	-5.5	-6	-5

Figure 4: A 7×7 surround mask.

3.3 Binarisation

A binary contour map can be constructed by using a standard procedure such as nonmaxima supression followed by hysteresis thresholding (Canny, 1986). Let the gradient magnitude $M(x, y)$ and orientation map $\Theta(x, y)$ specify the local strength and local edge direction respectively. Nonmaxima suppression seeks to thin regions where $M(x, y)$ is non-zero, to generate candidate contours as follows: two virtual neighbors are defined at the intersections of the gradient direction with a 3×3 sampling grid and the gradient magnitude for these neighbors is interpolated from adjacent pixels. The central pixel is retained for further processing only if its gradient magnitude is the largest of the three values. The final contour map is computed from the candidates by hysteresis thresholding. This process involves two threshold values t_l and t_h, $t_l < t_h$. All the pixels with $M(x, y) \geq t_h$ are retained for the final contour map, while all the pixels with $M(x, y) \leq t_l$ are discarded. The pixels with $t_l < M(x, y) < t_h$ are retained only if they already have at least one neighbor in the final contour map.

4 EXPERIMENTAL RESULTS

4.1 Ground Truth Image Data

Most of the methods for the evaluation of edge and contour detectors use natural images with associated desired output that is subjectively specified by the observer [(Martin et al., 2004), (Grigorescu et al., 2003)]. Some recent studies (Shin et al., 2001) show that the performance of such an operator must be considered task dependent. For object recognition, for example, some operators may perform better than others despite similar performance on synthetic images. The proposed surround interaction mechanisms is aimed at a better detection of objects contours in natural scenes.

We tested the performance of the proposed scheme on 40 natural images from a database designed to evaluate the performance of contour detection (dat, 2003). For each test image, an associated desired output binary contour map that was drawn by human is given. It should be noted that the ground truth data includes more than one type of pixels: (i) pixels which are parts of a contour of an object (ii) pixels which are part of a boundary between two (textured) regions. Our proposed scheme on the other hand, is designed to extract only the first type of contour pixels.

Table 1: Performance of proposed scheme and reported by Cosmin (Grigorescu et al., 2003) on 3 natural images(elephant, goat and hyena).

Image	Method	Performance
Goat	(Grigorescu et al., 2003) *Proposed Scheme*	0.34 **0.51**
Elephant	(Grigorescu et al., 2003) *Proposed Scheme*	0.42 **0.61**
Hyena	(Grigorescu et al., 2003) *Proposed Scheme*	0.55 **0.76**

4.2 Performance Measure

In order to have a quantitative comparison between the contour detector proposed in (Grigorescu et al., 2003) we use the performance measure introduced in the same. Let f_p and f_n are number of false positive and false negative pixels detected in the final contour map, respectively. The performance measure is defined as follows:

$$P = \frac{t_p}{t_p + f_p + f_n} \tag{1}$$

where, t_p is the number of correctly detected contour pixels (True positive). The performance measure P is a scalar taking values in the interval $[0, 1]$. If the desired output pixels are correctly detected and no background pixels are falsely detected as contour pixels, then $P = 1$. For all other cases, P takes values smaller than one, being closer to zero as more contour pixels are falsely detected and/or missed by the operator.

For computing the performance measure, we must determine which true positives are correctly detected, and which detection is false. The binary contour map in the ground truth data can be used for this purpose. Let us consider how to compute P of a output contour image given a ground truth contour map. One could simply correspond coincident contour pixels and declare all unmatched pixels either as false positives or misses. However, this approach would not tolerate any localization error and result in a poor performance measure. For robustness, it is desirable that the correspondence of output contour pixels to ground truth tolerate localization errors since ground truth data is not accurately localized. The approach proposed in (Grigorescu et al., 2003) considers a contour pixel as correctly detected if a corresponding ground truth contour pixel is present in a 5×5 (empirically find) square neighborhood (window) centered at the respective pixel coordinate. A window based approach leads to a less robust performance measure, as different sizes of the window can be shown to affect the performance value significantly, which is a undesirable. A large window will boost the number of true positive. However, an *explicit* correspondence of the detected contour and ground truth contour pixels is the only way to robustly count the hits, misses and false positives that we need to compute P. We have used the algorithm presented in (Martin et al., 2004) for the correspondence between output contour map and a ground truth contour map. The algorithm converts the corresponding problem into a minimum cost bipartite assignment problem, where the weight between a output contour pixel and ground truth contour pixel is proportional to their relative distance in the image plane. One can then declare all contour pixels matched beyond some threshold to be non-hits. The correspondence computation is detailed in (Martin et al., 2004).

4.3 Results

The proposed contour detection scheme was tested on 40 images from a database reported in (Grigorescu et al., 2003). Of these, we present results on 3 images in Fig 5 for illustrative purposes. A qualitative comparison between the results of our contour detection scheme and the contour reported in (Grigorescu et al., 2003) can be made by observing the results in this figure. The Canny edge detector outputs are also included for reference. The first and second columns show the input images and the corresponding ground truth images, respectively; the third column shows the best results of the Canny edge detector; the fourth column shows the results of the contour detection reported by (Grigorescu et al., 2003); and finally, the fifth column shows the results of the proposed

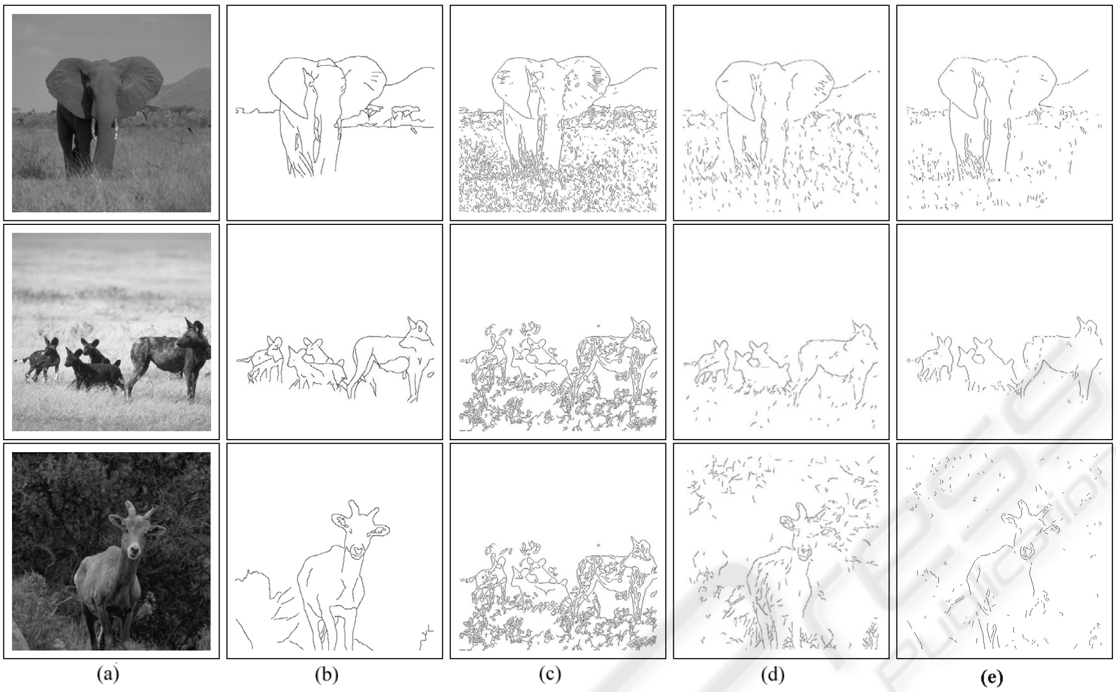

Figure 5: Results of contour detection on test images. (a) & (b) show input images and associated ground truth images respectively (c) Canny edge map (d) Best contour map reported in (Grigorescu et al., 2003) (e) Best results obtained by the proposed contour detection scheme.

scheme. The first point to note is that the obtained contours in the fourth and fifth columns are closer to the ground truth than the Canny output which is justifiably very 'noisy'. Furthermore, it can be seen that the results of the proposed scheme are closer to the ground truth. For instance, the grassy region in the elephant image (top row) is suppressed well. The best result of the contour map can be seen from the hyena images (in the bottom row) where the output is very close to the ground truth image.

A quantitative comparison of the contour and edge detectors is shown in Table 1. The figures in the table show that the proposed scheme outperforms the scheme in (Grigorescu et al., 2003) in all 3 images shown in Fig 5. This is consistent with our visual analysis of the results.

Fig. 6 shows statistical Box- and - Whisker plots for ten of the images used in our experiments. These plots are helpful in interpreting the distribution of performance value over ten different images. The average obtained value is above 0.5 which is encouraging since the measure is computed using ground truth images which included contour pixels belonging to object as well as texture region boundaries, whereas our scheme is designed for extracting only the former.

5 CONCLUSION

Though centre-surround receptive fields have been explored as a possible solution for good edge detection, a centre-surround mechanism is also applicable at a wider level (summation across neighbouring cells) to achieve contour detection. The proposed scheme for contour detection is motivated from such center-surround influence in the *cortical* cells of primates. It contributes to better contour detection not by enhancing responses to contours, but by selectively suppressing edges based on the surround. Specifically, an edge (signalled by a strong gradient) qualifies to be a contour only if it is salient in a local context where saliency implies that either the surround has no edges or the surround has weaker edges. Thus, the proposed scheme is not dependent on any prior knowledge which makes it a general preprocessing step for high order tasks. A further attractive feature of the scheme is that it is also computationally low in cost compared with many of the earlier approaches to general purpose contour detection.

In practice, contour detection is an intermediate level operation in computer vision with its output often used as input for further stages performing higher level processing. It is hence of interest to know the

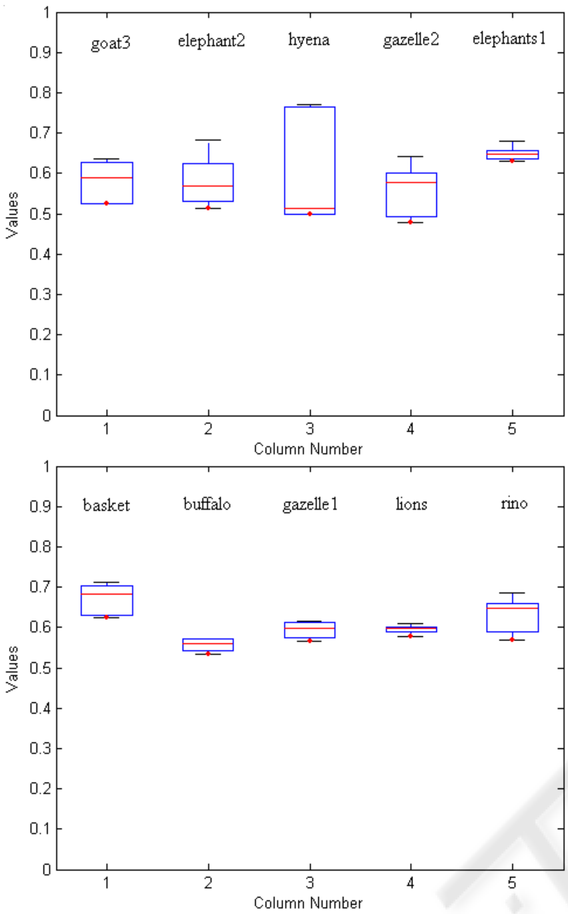

Figure 6: The distribution of performance value over ten different images.

appropriateness of its use given a specific high level task. As can be seen from the results, the proposed contour scheme largely suppresses the local background information and hence it is not appropriate to deploy it in tasks where the background information is essential, e.g. texture classification or region based segmentation. In other high-level tasks such as shape-based recognition and image retrieval, the proposed scheme can play a very useful role in their performance improvement.

REFERENCES

(2003). http://www.cs.rug.nl/∼imaging/databases/contour_database/contour_database.html.

Baumann, R., van der Zwan, R., and Peterhans, E. (1997). Figure-ground segregation at contours: a neural mechanism in the visual cortex of the alert monkey. In *European Journal of Neuroscience*.

Black, M., Sapiro, G., Marimont, D., and Heeger, D. (1998). Robust anisotropic diffusion. In *IEEE Transaction on Image Processing*.

Canny, J. (1986). A computational approach to edge detection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Cavanaugh, J., Bair, W., and Movshon, J. (2002). Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. In *Journal of Neurophysiology*.

Dobbins, A., Zucker, S. W., and Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. In *Nature*.

Dubuc, B. and Zucker, S. (2001). Complexity, confusion and perceptual grouping. part ii: mapping complexity. In *International Journal on Computer Vision*.

Grigorescu, C., Petkov, N., and Westenberg, M. (2003). Contour detection based on nonclassical receptive field inhibition. In *IEEE Transactions on Image Processing*.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cats visual cortex. In *Journal of Psychology*.

Knierim, J. and van Essen, D. (1992). Neuronal responses to static texture patterns in area v1 of the alert macaque monkey. In *Journal of Neurophysiology*.

Ma, W.-Y. and Manjunath, B. (2000). Edgeflow: A technique for boundary detection and image segmentation. In *IEEE Transactions on Image Processing*.

Marr, D. and Hildreth, E. (1980). Theory of edge detection. In *Proceedings of the Royal Society*.

Martin, D. R., Fowlkes, C. C., and Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Meer, P. and Georgescu, B. (2001). Edge detection with embedded confidence. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Morrone, M. C. and Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. In *Proceedings of the Royal Society, London Series B*.

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Series, P., Lorenceau, J., and Fregnac, Y. (2003). The silent surround of v1 receptive fields: theory and experiments. In *Journal of Physiology Paris*.

Shin, M. C., Glodgof, D. B., and Bowyer, K. (2001). Comparision of edge detectors using an object recognition task. In *Computer Vision and Image Understanding*.

von der Heydt, R., Peterhans, E., and Drsteler, M. R. (1991). Grating cells in monkey visual cortex: coding texture? In *Channels in the Visual Nervous System: Neurophysiology, Psychophysics and Models (Blum B, ed)*.

Yen, S. and Finkel, L. (1998). Extraction of perceptually salient contours by striate cortical networks. In *Vision Research*.