

Multi-scale Approach to Salient Contour Extraction

Gopal Datt Joshi and Jayanthi Sivaswamy

Centre for Visual Information Technology, IIT Hyderabad, Gachibowli,
Andhra Pradesh, INDIA - 500 019
gopal@research.iit.ac.in, jsivaswamy@iit.ac.in

Abstract

Contour detection is an important and difficult task for object segmentation in computer vision mainly because contours often occur in the presence of background texture. A previously reported scheme, based on the model of cortical cells in primates, analyses the local energy output derived at a single selected scale [7]. Selection of appropriate scale for image analysis is a difficult problem for which the biological visual system has evolved some solutions. We examine the role that scale plays in image analysis and present a scheme for contour detection based on combination of excitatory and inhibitory responses of a multi-scale, log-Gabor filter bank. The proposed scheme is thus a generalization of the approach in [7]. The resulting operator has been tested on a wide range of images and found to successfully extract salient contours. The results indicate that the proposed operator responds strongly to isolated lines/edges and weakly to edges that belong to a background texture.

1. INTRODUCTION

In computer vision and image processing, detection of edges and contours are important tasks which arise in many applications. Edges are defined as points of significant variations in intensity level in a gray level image whereas contours are perceptually salient edges that belong to the boundaries of regions and objects in the image. By perceptually salient we mean that human observers consider these edges to be salient. We focus on detecting contours of interest in a given image. Many techniques have been proposed for contour detection which are mostly developed for specific types of applications. Examples are detecting outlines of tissues in medical images, object contours in natural image scenes, boundaries between different texture regions, etc. These generally start with edge detection and use additional information, such as local image statistics [1, 2], image topology, texture, colors, edge continuity and density, etc [3, 4, 5].

A general-purpose contour detector has been proposed in [7]. This detector is biologically inspired and is based on excitatory and inhibitory responses of the primate cortical cells. The detector employs local energy detection followed by inhibition. Conjugate symmetric Gabor filter pairs are used to calculate the local energy at a specific spatial frequency (scale) whose value is chosen experimentally to detect the desired contours. Every point in the local energy map, which signals the presence of an edge feature, is evaluated according to its local context since for best contour detection; it is desirable to suppress points which are part of texture regions while retaining only those which belong to the object's contours. This suppression is implemented via an inhibition factor.

The main drawback of this approach is that the quality of results is very much scale dependent. Scale plays a major role in visual information processing. The features of interest in natural images are generally present at various scales. Furthermore, textural energy is also typically spread over a wide range of scales and hence any scale selection will result in insufficient inhibition for achieving the best contour detection. Ideally, for a contour detector which analyses an image at a single scale, one needs to select a scale such that detector will equally respond to both texture (fine) and objects (coarse) edges. However, since the characteristics of the texture and the objects are generally unknown, this is not possible. Multi-scale analysis can address this problem. Biological vision appears to employ this type of analysis since there are cells found at the low levels of the visual system tuned to different orientations as well as scales [6, 11]. Studies in computer vision also have established that multi-scale image analysis can help to overcome the problem of appropriate scale selection [12]. We propose a contour detection scheme which uses multi-scale analysis and the notion of inhibition. The scheme removes the scale dependency and achieves improved contour detection performance on a wide range of images.

The paper is organized as follows. Section 2 presents our contour detection scheme which describes the computation of local energy at particular scale, models for two inhibition mechanisms, namely, anisotropic and isotropic and the contour binarisation scheme. Results obtained from the contour detection scheme are presented in Section 3 along with some concluding remarks in section 4.

2. MULTISCALE CONTOUR DETECTION SCHEME

The contour information is extracted in the HVS in its early stages of processing. Various psychophysical studies have shown that the perception of an oriented stimulus, e.g. a line segment, can be influenced by the presence of other such stimuli (destructors) in its neighbourhood [9]. These perceptual effects are supported by neurophysiological measurements on cells in the primary visual cortex of primates [10]. These studies show 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. Neurophysiologists refer to this effect as *surround inhibition* and it is exhibited by a majority (80%) of the orientation-selective cells in the visual cortex of primates [10]. Hence, it is worthwhile to study the underlying mechanism of these cells and develop a scheme for contour detection incorporating surround influence.

From the point of view of the problem at hand, an oriented edge (or line) feature is only a candidate for a contour. For an edge feature to be successfully declared as a contour, its local context, in the form of centre-surround interaction, must be taken into account. Computationally, the surround mechanism can be separated from the centre mechanism which is responsible for edge detection. Hence, we propose a scheme for contour detection made up of two processing steps: 1) Edge detection based on local energy computation at multiple scales [14] 2) Surround inhibition.

2.1 Local Energy Model

Typically, a bank of conjugate symmetric filter pairs which are tuned to different frequencies and orientations is required for the local energy computation. It is desirable to obtain a reasonably broad and uniform coverage of the entire spectrum without having to use too many filters. Neurophysiological studies provide a clue to how this is implemented in mammalian visual systems where families of neurons which span different ranges of locations and scales are present. Specifically, these neurons, viewed as spatial filters, have bandwidth ranging from about 0.5 octaves up to about 3.0 octaves [6] and their transfer functions are roughly symmetric when viewed on a logarithmic frequency scale [8].

A traditional choice for local energy computation is to use Gabor filters. However, the maximum bandwidth obtainable from a Gabor filter is only about 1 octave [15] and a Gabor filter has a Gaussian profile only on the linear frequency scale. A log-Gabor filter on the other hand, allows large bandwidths and has a symmetric (such as Gaussian) profile on the logarithmic scale. This is a desirable feature since, it has been suggested that natural images would be better coded by filters that have Gaussian transfer functions when viewed on the logarithmic frequency scale [13]. Due to these characteristics, a log-Gabor filter bank uniformly covers the frequency spectrum by using a fewer number of filters unlike a Gabor filter bank.

Due to the singularity in the log-Gabor function at the origin, one cannot construct an analytic expression for the shape of log-Gabor function in the spatial domain. Hence, one has to design the filter in the frequency domain. On a linear scale, the transfer function of a log-Gabor filter is expressed as

$$\mathbf{f}_{(r_o, q_o)} = \exp \left\{ \left[\log \left(\frac{r}{r_o} \right) \right] / 2 \left(\log \left(\frac{\mathbf{s}_r}{r_o} \right) \right)^2 \right\} \exp \left\{ - \frac{(\mathbf{q} - \mathbf{q}_o)}{2\mathbf{s}_q^2} \right\} \quad (1)$$

where r_o is the central radial frequency, q_o is the orientation of the filter, \mathbf{s}_q and \mathbf{s}_r represent the angular and radial bandwidths, respectively. The local energy for an image is defined as

$$E_{r_o, q_o}(x, y) = \sqrt{\left(O_{r_o, q_o}^{even}(x, y) \right)^2 + \left(O_{r_o, q_o}^{odd}(x, y) \right)^2} \quad (2)$$

where $O_{r_o, q_o}^{even}(x, y)$, $O_{r_o, q_o}^{odd}(x, y)$ are the responses of the even and odd symmetric log-Gabor filters, respectively. The real-

valued function given in (1) can be multiplied by the frequency representation of the image and, transform the result back to the spatial domain, the responses of the oriented energy filter pair are extracted as simply the real component for the even-symmetric filter and the imaginary component for the odd-symmetric filter. Let $Z_{(r_o, q_o)}$ be the transformed filtered output. The responses of even and odd symmetric log-Gabor filters are expressed as:

$$O_{r_o, q_o}^{even}(x, y) = \text{Re}(Z_{(r_o, q_o)}); \quad O_{r_o, q_o}^{odd}(x, y) = \text{Im}(Z_{(r_o, q_o)}); \quad (3)$$

2.2 Inhibition Models

Following the local energy computation, an inhibition term is to be calculated based on the local energies at different scales and orientations. There are two types of inhibition of interest, one of which depends on orientation of the stimuli in the central (anisotropic) and surround regions while the other does not (isotropic). For a given point in the image, the inhibition term is computed in an circular area around it. In the following, the inhibition terms are constructed in the same way as in [7].

Let $G_{\mathbf{s}}(x, y)$ be a zero-mean and 2-D Gaussian function with standard deviation of \mathbf{s} . Here we define a filter function $h_{\mathbf{s}}(x, y)$ as follows:

$$h_{\mathbf{s}}(x, y) = \frac{1}{\|R(DoG)\|} R(DoG(x, y)) \quad (4)$$

where $R(z) = (z + |z|)/2$, ensures that the operator has only positive response and $\|\cdot\|$ denotes the L_1 norm; $|\cdot|$ denotes modulus; and $DoG(x, y)$ is a difference of Gaussian functions with the ratio of 4:1 in their standard deviation.

(I) Anisotropic inhibition: The anisotropic inhibition is computed in two steps. First step calculates the inhibited contour response and orientation map for a particular scale using local energies in all orientations and the second step integrates information at all the scales and computes the final contour response and orientation map. An inhibition map $S_{r_i, q_j}^A(x, y)$ for each scale r_i and for a particular orientation q_j is found as the response of $h_{\mathbf{s}}(x, y)$ to the local energy $E_{r_i, q_j}(x, y)$ as

$$S_{r_i, q_j}^A(x, y) = (E_{r_i, q_j} * h_{\mathbf{s}})(x, y) \quad (5)$$

The inhibited contour response $\tilde{E}_{r_i, q_j}^{A, a}(x, y)$ is computed as follows:

$$\tilde{E}_{r_i, q_j}^{A, a}(x, y) = R(E_{r_i, q_j}(x, y) - \mathbf{a} * S_{r_i, q_j}^A(x, y)) \quad (6)$$

where factor \mathbf{a} controls the strength of the inhibition by the surround. If there is no texture surrounding (i.e., there is an isolated an edge) a given point, the response will be equal to the local energy term. However, if there are other edges in the surrounding region, the inhibition term $S_{r_i, q_j}^A(x, y)$ can become strong enough to cancel completely, the contribution of the local energy term.

In order to take the orientation into account, an inhibited contour response $N_{r_i}^{A, a}$ is defined as

$$N_{r_i}^{A, a} = \max \{ \tilde{E}_{r_i, q_j}^{A, a}(x, y) \mid i = 1, \dots, N_q \} \quad (7)$$

An associated orientation map $\Theta_{r_{ij}}^A(x, y)$ records the orientation for which this maximum response is achieved

$$\Theta_{r_i}^A(x, y) = \mathbf{q}_k, \quad k = \arg \max \left\{ \tilde{E}_{r_i, \mathbf{q}_j}^{A, \mathbf{a}} \mid j = 1, \dots, N_q \right\} \quad (8)$$

The final contour response and the associated orientation map are computed as

$$N^{A, \mathbf{a}} = \sum_{i=1}^{N_r} N_{r_i}^{A, \mathbf{a}}; \quad \Theta^A = \frac{1}{N_r} \sum_{i=1}^{N_r} \Theta_{r_i}^A \quad (9)$$

where N_r is the total number of used scale.

(2) **Isotropic inhibition:** In isotropic inhibition the orientation of the stimulus in the surround is immaterial. Hence, we construct a local energy map $\tilde{N}_{r_i}^I(x, y)$ with values of maximum local energy over all orientations.

$$\tilde{N}_{r_i}^I(x, y) = \max \left\{ E_{r_i, \mathbf{q}_j}(x, y) \mid j = 1, \dots, N_q \right\} \quad (10)$$

and the orientation map $\Theta_{r_i}^I(x, y)$ records once again the orientation of maximal response as

$$\Theta_{r_i}^I(x, y) = \mathbf{q}_k, \quad k = \arg \max \left\{ E_{r_i, \mathbf{q}_j}(x, y) \mid j = 1, \dots, N_q \right\} \quad (11)$$

The isotropic inhibition map $S_{r_i}^I(x, y)$ is found by filtering the maximum oriented energy map $\tilde{N}_{r_i}^I(x, y)$ with the kernel $h_s(x, y)$

$$S_{r_i}^I(x, y) = \left(\tilde{N}_{r_i}^I * h_s \right)(x, y) \quad (12)$$

The inhibited contour response at scale r_i is computed as follows:

$$N_{r_i}^{I, \mathbf{a}}(x, y) = R \left(\tilde{N}_{r_i}^I(x, y) - \mathbf{a} * S_{r_i}^I(x, y) \right) \quad (13)$$

As before, the factor \mathbf{a} controls the strength of the inhibition. The final contour response and the associated contour map are found as

$$N^{I, \mathbf{a}} = \sum_{i=1}^{N_r} N_{r_i}^{I, \mathbf{a}}; \quad \Theta^I = \frac{1}{N_r} \sum_{i=1}^{N_r} \Theta_{r_i}^I \quad (14)$$

where N_r is the total number of used scale.

(3) **Binary Contour Map Construction:** The response $N(x, y)$ ($N^{I, \mathbf{a}}$ or $N^{A, \mathbf{a}}$) specifies the local edge strength and the orientation map $\Theta(x, y)$ (Θ^I or Θ^A) specifies the associated edge direction at a point (x, y) . These are used to construct binary maps by using a standard procedure of nonmaxima suppression followed by hysteresis thresholding [16]. Nonmaxima suppression seeks to thin regions where $N(x, y)$ 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 candidate contours by hysteresis thresholding. This process involves two threshold values t_l and t_h , $t_l < t_h$. All the pixels with $N(x, y) \geq t_h$ are retained for the final

contour map, while all the pixels with $N(x, y) \leq t_l$ are discarded. The pixels with $t_l < N(x, y) < t_h$ are retained only if they already have at least one neighbor in the final contour map. The values of t_l and t_h are tuned to get the best result.

3. RESULTS

The proposed multi-scale contour detection scheme was implemented with a bank of the filters such that it covers the frequency spectrum uniformly. The length to width ratio of a filter controls its directional selectivity. This ratio can be varied in conjunction with the number of orientations used in order to achieve coverage of the 2-D spectrum. Furthermore, as the degree of blurring introduced by the filters increases with their orientation selectivity, they must be carefully chosen to minimize the blurring. Hence, we consider a filter bank with the following features. (a) The spatial frequency plane is divided into 12 equally spaced orientations. (b) The radial axis is divided into three equal octave bands with a band of width two octaves. The filter with highest central frequency (in each direction) is positioned near the Nyquist frequency to avoid ringing and noise. (c) The angular bandwidth is chosen to be 15 degrees.

We have tested our scheme on different sets of images e.g. synthetic images, natural scenery and medical images. These sets differ in the kind of organisation of image contours and background. We first present a comparison between the performance of our proposed contour detector with Canny edge detector. The best results of the contour detector are shown in Fig. 3. From these we can observe several points. Firstly, as compared with the Canny output, the contour map is sparse, retaining only salient edges such as boundaries and outlines of various regions. Secondly, the extracted contours are closer to perceived ones. This can be seen in the contour map of the stone image where the contours of two of the stones do not include the shadow regions on the left whereas the Canny output includes this region. It is noteworthy that the Canny detector does not distinguish between the contours (signal) and textures (noise) and hence raising the threshold will decrease both the signal and noise. This is in contrast to the proposed contour detector which can be tuned to produce a sparse output which retains mostly the signal while rejecting the noise.

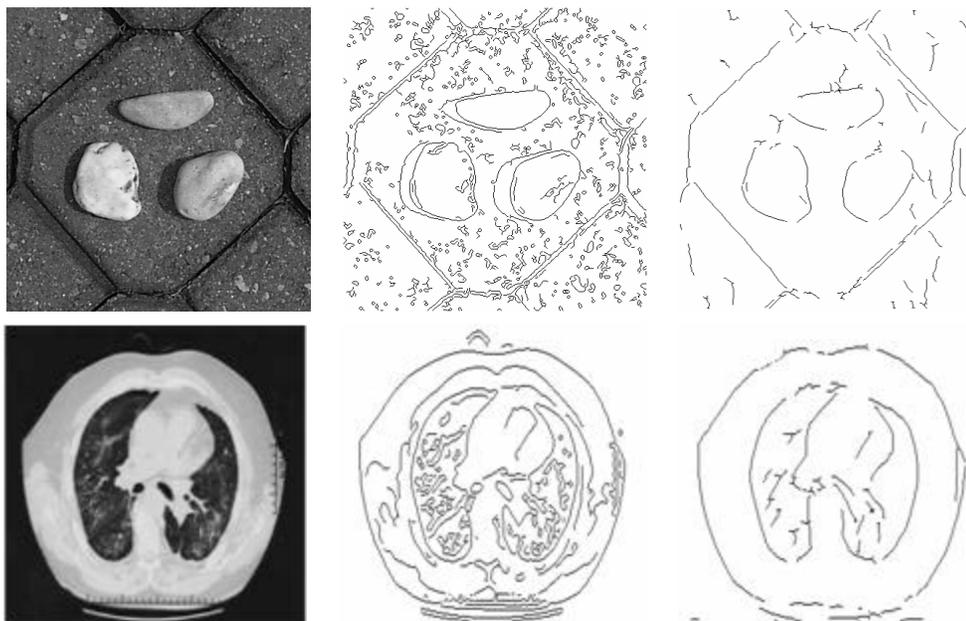


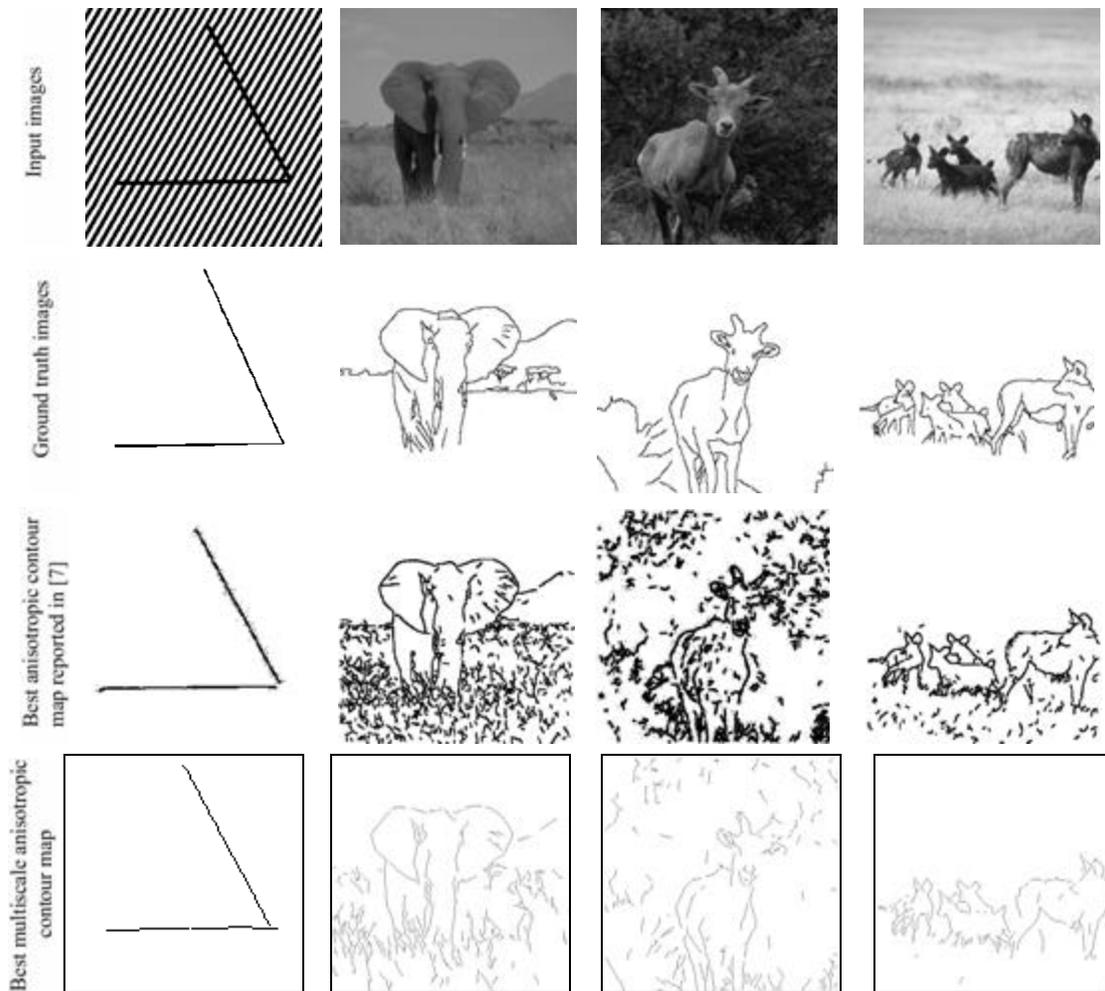
Fig. 3: Results of testing on medical image (row 1) and image of stone on pavement (row 2). First column shows the input images, second column shows the best Canny edge map and third column shows the best contour map obtained from the proposed scheme based on the isotropic inhibition model

We next compare the results of our proposed scheme with the results reported in [7] which are also based on surround inhibition, however at one empirically chosen scale. The results are shown in Fig. 4. These results show that a multi-scale contour detector suppresses surround texture more effectively. For instance, in the result obtained using this operator for the

elephant image, the grassy texture region is suppressed well in both the isotropic as well as anisotropic inhibition models. The best results can be seen on hyena image where the contour map is more close to the ground truth image. These results show that the notion of scale is important for contour detection. Results obtained from testing on synthetic images (in the first column of Fig. 4) show the multi-scale contour operator to be consistent with both the human perception and the results reported in [10].

4. DISCUSSION AND CONCLUSIONS

Based on the premise that scale is important, we have presented a scheme for contour detection that can be used on a wide range of images. The proposed scheme contributes to better contour detection not by enhancing responses to contours, but by selectively suppressing edges based on the surround region. In comparison with the scheme proposed in [10] the proposed scheme produces contour maps much closer to perceived contours. However, the improved performance does come at a higher computational cost. Even though we have tried to reduce this cost by choosing log-Gabor filters, other solutions can be investigated for a further reduction in cost.



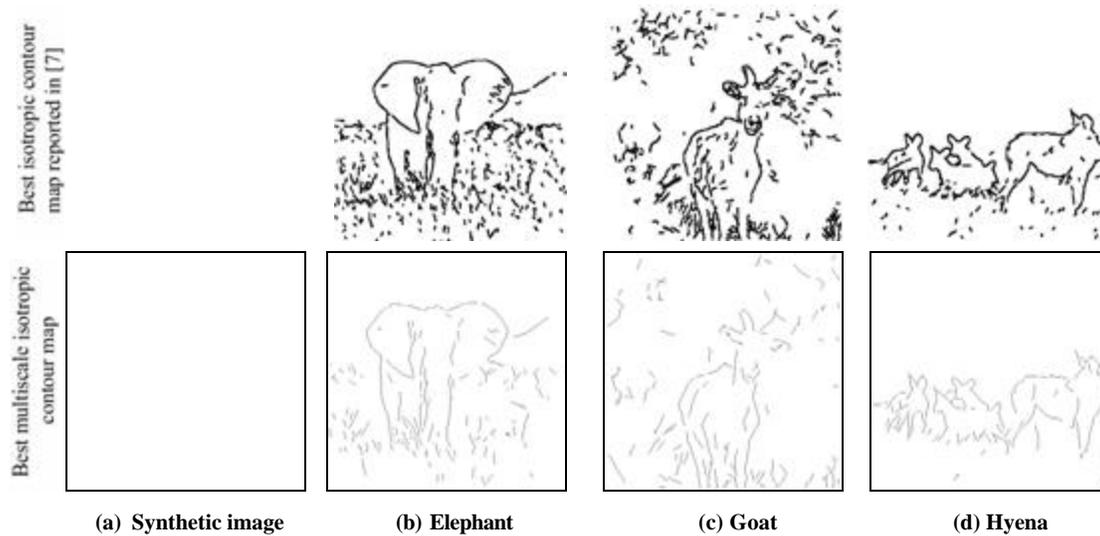


Fig 4: One synthetic image and three natural scenes with objects on textured backgrounds (first row); their corresponding ground truth contour maps (second row); the best anisotropic contour maps reported in [7] (third row); the best contour maps obtained with multi-scale anisotropic contour operator (fourth row); the best isotropic contour maps reported in [7] (fifth row); and the best contour maps obtained with the multi-scale isotropic contour operator (last row)

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 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.

5. REFERENCES

- [1] P. Meer and B. Georgescu, (2001), "Edge detection with embedded confidence", IEEE Trans. Pattern Anal. Machine Intell., Vol. 23, pp. 1351-1365.
- [2] S. Ando, (2000), "Image field categorization and edge/corner detection from gradient covariance", IEEE Trans. Pattern Anal. Machine Intell., Vol. 22(2), pp. 179-190.
- [3] M. J. Black, G. Sapiro, D. Marimont, and D. Heeger, (1998), "Robust anisotropic diffusion", IEEE Trans. Image Processing, Vol. 7, pp. 421-432.
- [4] P. Perona, J. Malik, (1990), "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Pattern Anal. Machine Intell., Vol. 12 (7), pp. 629-639.
- [5] W. Y. Ma, B. S. Manjunath, (2000) "Edgeflow: A technique for boundary detection and image segmentation," IEEE Trans. Image Processing, Vol. 9 (8), pp. 1375-1388.
- [6] J. G. Daugman, (1985), "Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," Journal Optical Society of America, Vol. A 2, pp. 1160-1169.
- [7] C. Grigorescu, N. Petkov, M. A. Westenberg, (2003), "Contour detection based on nonclassical receptive field inhibition," IEEE Trans. Image Processing, Vol. 12 (7), pp. 729-739.
- [8] S. J. Anderson, D. C. Burr, (1985), "Spatial and temporal selectivity of the human motion detection system," Vision Research, Vol. 25 (8), pp. 1147-1154.
- [9] D.J. Field, A. Hayes, R.F. Hess, (1993), "Contour integration by the human visual system: Evidence for a local association field," Vision Research, Vol. 33 (2), pp. 173-193.
- [10] J.J. Knierim, D.C. van Essen, (1992), "Neuronal responses to static texture patterns in area V1 of the alert macaque monkey," Journal of Neurophysiology, Vol. 67, pp. 961-980.

- [11] P. H. Schiller, B. L. Finlay, S. F. Volman, (1976), "Quantitative studies of single-cell properties in monkey striate cortex. III. Spatial frequencies," *Journal of Neurophysiology*, Vol. 39, pp. 1334-1351.
- [12] T. Lindeberg, (1994), "Scale-space theory: A basic tool for analysing structures at different scales," *Journal of Applied Statistics*, Vol. 21(2), pp. 224-270.
- [13] D. J. Field, (1987), "Relations between the statistics of natural images and the response properties of cortical cells." *Journal of the Optical Society of America*, Vol. A 4(12), pp. 2379-2394.
- [14] M. C. Morrone, D. C. Burr, (1988), "Feature detection in human vision: A phase-dependent energy model," *Proceedings of the Royal Society, London Series*, Vol. B 235, pp. 221-245.
- [15] P. Kovess, (1999), "Image features from phase congruency," *Videre: Journal of Computer Vision Research*, Vol. 1(3), pp. 127.
- [16] J. Canny, (1986), "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 8, pp. 679-698.