An Improved Neural Network for Segmenting Objects' Boundaries in Real Images*

Wee Kheng Leow and Seet Chong Lua
Department of Information Systems and Computer Science
National University of Singapore
Lower Kent Ridge Road, Singapore 119260, Singapore
leowwk@iscs.nus.sg, sclua@iti.gov.sg

Abstract

An important task in object recognition is to first identify the boundaries of the objects in the input image. Several neural networks have been proposed to perform edge detection and boundary segmentation. Among them, Grossberg and Mingolla's Boundary Contour System (BCS) seems promising because it is able to complete missing object boundaries. Although BCS has been shown to work well on synthetic and silhouette images, we found that it has some shortcomings when applied to real images. This paper presents an improved version of BCS for handling the shortcomings.

1. Introduction

A machine vision system often begins its processing by extracting edges in the input image. Edges, representing sharp changes in image intensity, are indeed important features for object recognition. However, to recognize objects in the image, the vision system has to perform not only edge detection but also *boundary segmentation*, i.e., to identify the edges that form the objects' boundaries. The system must be able to segment object boundaries at various scales, as in human shape perception [1, 7]. Consider, for example, the image of a comb (Fig. 3(a)). At a fine scale, the system should extract the boundaries that define the comb's teeth. These boundaries are important in describing the exact physical structure of the object. At a coarse scale, however, only the overall boundary is important in characterizing the overall shape of the comb; fine details are unessential and should be ignored. Scale-dependent boundary segmentation is therefore an important task in object recognition.

*Supported by NUS Academic Research Grant No. RP950656.

2. Related work

Most of the recent works on boundary segmentation have been focused on the sub-task of *edge grouping* [3, 6, 2, 8], with the aim of bridging small gaps between edges. These works typically apply the *Gestalt principles* of proximity, smoothness, and continuity to join disconnected edges. These principles suggest that two disconnected edge segments can be joined if they are close in proximity and can form a smooth, continuous curve. In other words, two edge segments that are near to each other and have similar orientations can be connected.

The edge grouping methods proposed by Mohan and Nevatia [6], Dolan and Riseman [2], and Tabbone [8] are non-neural network algorithms. These algorithms typically begin by extracting *zero-crossings* of the second derivatives of image intensity. Since the second derivatives are typically computed using isotropic filters, such as the Laplacian of Gaussian, the zero-crossings lack information about the edges' orientations. As a result, additional processing is required to compute the orientations of edges and curves, so that disconnected edges can be connected into smooth curves. However, the computation of edge orientations is often inexact and ambiguous, especially in noisy real images. Without proper disambiguation of edge orientations, edges will be incorrectly connected.

In contrast, the Boundary Contour System (BCS) of Grossberg and Mingolla [3] is a neural network system that lets its neural units cooperate and compete to determine edge orientations. The BCS consists of five layers of units. The first layer is the input layer. The second layer extracts, at each pixel location, the local in-

Appears in *Proceedings of International Conference on Neural Networks*, 1997, vol. III, 1663–1668.

tensity contrasts measured along various orientations. As discussed before, orientations of local contrasts can be ambiguous. Certain locations such as corners and vertices may exhibit strong contrasts along more than one orientations. Disambiguation is performed in the third and fourth layers. The units in these layers compete to determine the dominant orientations of local contrasts. Units in the last layer, which is the output layer, cooperate to activate units that form continuous, smooth curves, thereby completing disconnected edges. The network is run for several iterations for the units' activities to settle. The equilibrium network activities represent completed boundaries whose local orientations have been disambiguated.

So far, BCS has been shown to work well on synthetic and silhouette images [3, 5]. Our application of BCS to real images, however, revealed two major shortcomings of the system:

1. The boundaries extracted by BCS is often very thick. As a result, the locations of the boundaries are inexact.

2. The results obtained by BCS is often too noisy to be useful.

This paper describes an improved version of BCS, called *Boundary Segmentation Network*, that overcomes the above shortcomings.

3. Network architecture and dynamics

The Boundary Segmentation Network (BSN) consists of six layers of neural units (Fig. 1). Except for the first layer, each other layers consists of eight sublayers, one for each orientation from 0° to 180°. The first five layers are similar to those in BCS, except for some modifications. The last layer of BSN, the Sharpening Layer, is absent in the original BCS [3], but present in a later version of BCS [4]. This layer also serves as the BSN's output layer. Let us examine each layer one at a time.

The first layer is the input layer, and the units' activities I_q represent the intensities of the input image at location $q = (x_q, y_q)$. The second layer, the Contrast Extraction Layer, extracts, at each pixel location, the local intensity contrasts measured along eight different orientations. The total input u_{kp} of the unit in orientation sublayer k $(0 \leq k \leq 7)$ and at location p is given by the weighted sum of the input units' activities I_q:

$$u_{kp} = \sum_q W_{kpq} I_q \tag{1}$$

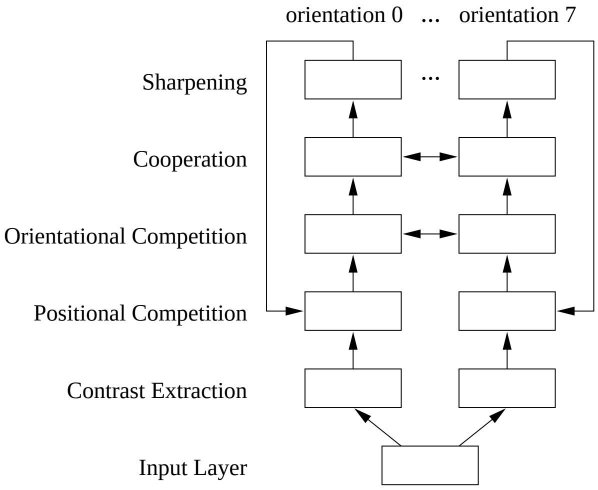

Figure 1. The Boundary Segmentation Network (BSN) consists of 6 layers of units. Except for the input layer, each other layers consists of 8 sublayers, one for each orientation.

where W_{kpq} is the weight of the connection from input unit I_q to unit u_{kp}. Input unit's activity I_q is assumed to range from 0 to 1. The weights W_{kpq} takes the form of an oriented elongated Gaussian, except that half of it is inverted:

$$\begin{aligned} W_{kpq} &= G_k(x_q - x_p, y_q - y_p) \qquad (2)\\ G_k(x,y) &= \begin{cases} g_k(x',y') & \text{if } y' > 0 \\ 0 & \text{if } y' = 0 \\ -g_k(x',y') & \text{otherwise} \end{cases} \qquad (3)\end{aligned}$$

where

$$\begin{aligned} g_k(x',y') &= \frac{1}{2\pi\lambda\sigma^2}\exp\left(-\frac{(x'/\lambda)^2 + y'^2}{2\sigma^2}\right) \qquad (4)\\ x' &= x\cos\theta_k + y\sin\theta_k \\ y' &= -x\sin\theta_k + y\cos\theta_k \ . \end{aligned}$$

The variable $\theta_k = k \cdot 22.5°$ denotes the angle of orientation k, and λ determines the amount of elongation (= 2 in the current BSN). The *scale factor* σ determines the scale of the network. To extract fine details, σ should be small. Conversely, to extract coarse boundaries, σ should be large. In BCS, the connection weights have uniform values. In BSN, Gaussian weights are adopted because, with Gaussian weights, the unit at the location of maximum contrast tend to be more active than its neighbouring units. In other words, the positions of the most active units correspond well with the locations of maximum contrasts. Second layer units' activities a_{kp} is given by the absolute value of the total input, with noise filtering:

$$a_{kp} = f(|u_{kp}|) \tag{5}$$

$$f(u) = \begin{cases} 1 & \text{if } u \geq \beta_u \\ \dfrac{u-\alpha_u}{\beta_u-\alpha_u} & \text{if } \alpha_u < u < \beta_u \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where α_u and β_u are positive lower and upper bounds. Function $f(u)$ has a shape similar to a sigmoid. It regards inputs smaller than α_u as noise and filters them out. It also clips inputs larger than β_u and outputs the maximum value of 1. In BSN, $\alpha_u = 0.05$ and $\beta_u = 1$.

The third and the fourth layers serve to determine the position and the dominant orientations of local contrasts through competition among the units. The third layer, the Positional Competition Layer, determines the positions of local contrasts by imposing a competition between neighbouring units in the same orientation sublayer:

$$\frac{\mathrm{d}}{\mathrm{d}t} b_{kp} = -\alpha_b b_{kp} + \beta_b a_{kp} + \gamma_b s_{kp} - \delta_b b_{kp} \sum_{q \neq p} E_{pq} a_{kq} \tag{7}$$

where α_b, β_b, γ_b and δ_b are positive constant parameters. Every third layer unit's activity decays at a rate of α_b. The unit at orientation sublayer k and location p receives two excitatory inputs: one from a second (Orientation Extraction) layer unit a_{kp} and the other from a sixth (Sharpening) layer unit s_{kp}, both units located at the same position p with the same orientation k. In addition, it also receives inhibitory inputs from second layer units a_{kq} located in the same orientation sublayer and at neighbouring positions. The inhibitory weights E_{pq} is defined by an isotropic Gaussian function, which is just Eq. 3 with $\lambda = 1$ and $\theta_k = 0$. In contrast, BCS adopts uniform weights. The constant parameters determine how strongly the four terms in Eq. 7 affect the units' activations. Experimental studies show that, if γ_b is too large compared to β_b, many suprious edges are created. In the current BSN, $\alpha_b = 0.2$, $\beta_b = 1$, $\gamma_b = 0.25$, and $\delta_b = 1$. Given a fixed input, Eq. 7 will eventually stabilise at equilibrium values, and can, thus, be simplified as:

$$b_{kp} = \frac{\beta_b a_{kp} + \gamma_b s_{kp}}{\alpha_b + \delta_b \sum_{q \neq p} E_{pq} a_{kq}} \; . \tag{8}$$

The fourth layer, the Orientational Competition Layer, serves the complementary role of the Positional Competition Layer. It determines the dominant orientations of local contrasts by imposing a competition between units at different orientation sublayers. A fourth layer unit c_{kp} receives excitatory input from the third layer unit b_{kp} and inhibitory inputs from third layer units b_{lp} at the same location p but different orientation sublayers l:

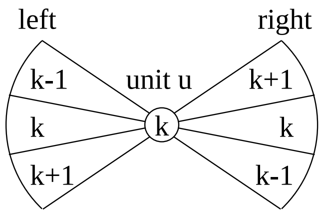

Figure 2. The receptive field of a unit in orientation sublayer k in the Cooperation Layer is divided into the left and the right halves. Each half is further divided into three sections, aligned at orientations $k-1$, k and $k+1$.

$$c_{kp} = \max\left(b_{kp} - \sum_{l \neq k} \Phi_{kl} b_{lp}, 0 \right) \; . \tag{9}$$

The inhibitory weights Φ_{kl} are defined in such a way that units with larger orientation difference inhibit each other more strongly than units with smaller orientation difference:

$$\Phi_{kl} = \frac{\sin|\theta_k - \theta_l|}{\sum_m \sin|\theta_k - \theta_m|} \; . \tag{10}$$

The largest orientation difference is 90°, i.e., the two orientations are perpendicular to each other. In BCS, competition of the form described by Eq. 9 applies only to units with perpendicular orientation differences. Competition between units with smaller orientation differences takes a form similar to Eq. 7. Experimental results show that, compared to BCS, BSN's orientational competition produces object boundaries that have more even strengths.

The fifth layer is the Cooperation Layer, in which pairs of units cooperate to complete disconnected object boundaries. Cooperation is carried out as follows. Consider a unit u at orientation sublayer k (Fig. 2, assume that orientation k is horizontal). Its receptive field is divided into the left and the right halves with respect to orientation k. Each half is further divided into three sections, aligned at orientations $k-1$, k and $k+1$.[1] Sections at orientation k fall in the same sublayer as unit u. Sections $k-1$ and $k+1$ fall respectively in orientation sublayers $k-1$ and $k+1$. That is, unit

[1] The subtraction in $k-1$ and the addition in $k+1$ are modulo 8 operations.

u receives excitatory inputs from neighbouring units in the sections and at the respective orientation sublayers. This way, any unit in the left section may cooperate with any unit in the right section to activate unit u, and the activated units will form a smooth curve. In BCS, the left and the right halves each has only 1 section. BSN's cooperation scheme is, therefore, more elaborate than that of BCS, and it has the advantage of producing smoother boundaries. The activity d_{kp} of a unit in the Cooperation Layer is given by the product of the total excitatory inputs from the two halves:

$$d_{kp} = f\left(\sum_{l\in K}\sum_{q\in L} Z_{pq} c_{lq} \cdot \sum_{l\in K}\sum_{q\in R} Z_{pq} c_{lq}\right) \qquad (11)$$

where $K = \{k-1, k, k+1\}$, L and R denote the sets of locations in the left and the right receptive fields, and $f(u)$ is the noise-filtering function defined in Eq. 6. The weights Z_{pq} of the connections from units c_{lq} to unit d_{kp} are defined by an appropriate Gaussian function:

$$Z_{pq} = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(D_{pq}-D)^2}{2\sigma^2}\right) \qquad (12)$$

where D_{pq} is the distance between locations p and q, and D denotes the optimal distance of cooperation, which is set to 2σ in BSN.

The sixth layer in BSN is the sharpening layer, which reduces the width of the edges, and thus helps to localize the edges more accurately. Sharpening is achieved by the inhibition from flanking units:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} s_{kp} = \; & -\alpha_s s_{kp} + \beta_s(1-s_{kp})g(s_{kp}) \\ & + h(d_{kp}) - \gamma_s \sum_{q\neq p} X_{kpq}\, g(s_{kq})\end{aligned} \qquad (13)$$

where α_s, β_s, and γ_s are positive parameters, $g(s_{kp})$ is a scaling function, and $h(d_{kp})$ is an input function. The function $h(d_{kp})$ outputs the value d_{kp} for a brief moment such that unit s_{kp} receives a brief excitatory input from unit d_{kp}. Unit s_{kp} also receives excitatory feedback from itself and inhibition from its neighbouring units. The inhibitory weights X_{kpq} are given by a circular oriented Gaussian (i.e., Eq. 3 with $\lambda = 1$), except within a narrow band along the major axis at which the weights are 0. Therefore, unit s_{kp} receives inhibition only from neighbouring units flanking the Gaussian's major axis. The constant parameters should be appropriately set to balance the amount of decay, self-feedback and flanking inhibition. In BSN, $\alpha_s = 0.1$, $\beta_s = 0.8$, and $\gamma_s = 0.3$. To achieve strong sharpening effect, $g(s_{kq})$ should be a faster-than-linear function, such as a square function $g(s_{kq}) = s_{kq}^2$, which is used in BSN. The original BCS [3] does not have a sharpening layer. A later version of BCS [4] does have a sharpening layer but without self-feedback, and its inhibitory weights have uniform values. That is, units representing edges receive inhibition not only from the flanking units but also from neighbouring edge units. In comparison, BSN's sharpening layer performs better in enhancing object boundaries.

4. Test results

In BCS as well as BSN, neural units that represent edges at different orientations are located in different orientation sublayers. For the purpose of illustrating test results, activities of the sublayers in the output layer are summed up into a single 2-D representation.

Figure 3 illustrates the results of processing a comb image. Fig. 3(b) shows that the boundaries extracted by BCS are quite thick. Moreover, due to weaker noise filtering and edge sharpening, boundaries that are close to each other interfere at the cooperation stage, producing spurious unit activations. The effect is most pronounced on the left side of the comb where the teeth are more closely packed. In contrast, the boundaries extracted by BSN (Fig. 3c) are thinner and more well defined. Fig. 3 also shows that BSN's boundary segmentation is scale-dependent. At a fine scale (Fig. 3c), BSN can extract boundaries that define the comb's fine structure. At a coarse scale (Fig. 3d), BSN can ignore fine details and extract the boundaries that represent the object's overall shape. The coarse-scaled boundaries tend to be thicker than the fine-scaled boundaries due to the units' larger receptive fields. This is, however, not a serious problem because the boundaries extracted at a coarse scale are well separated and do not interfere with each other.

Figure 4 compares BCS and BSN in the processing of an image with a nosy background. Fig. 4(b) shows that the object boundaries segmented by BCS are thicker, and that many background details are extracted. The strength of these background details are so strong that they are easily confused with the fan's boundaries. In contrasts, BSN can reject noisy details better than BCS, and the boundaries that it extracts are thinner (Fig. 4c).

5. Further enhancements

BSN, as well as BCS, allows the neural units to compete to determine the dominant orientations of local contrasts. This competition is especially strong at corner points. As a result, the strength of the boundaries

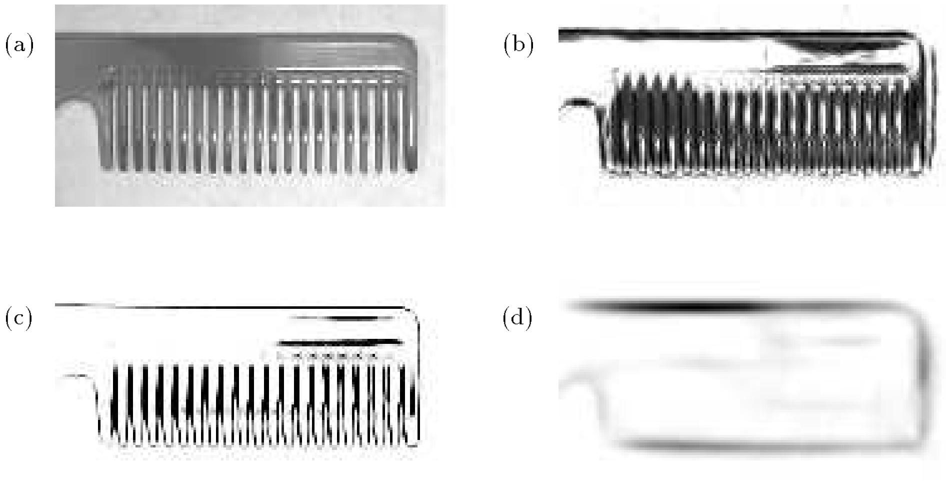

Figure 3. Comparison of BCS and BSN. (a) Input image of a comb. (b) Boundaries extracted by BCS with scale factor $\sigma = 0.5$. (c) Boundaries extracted by BSN with $\sigma = 0.5$. (d) Boundaries extracted by BSN with $\sigma = 6$.

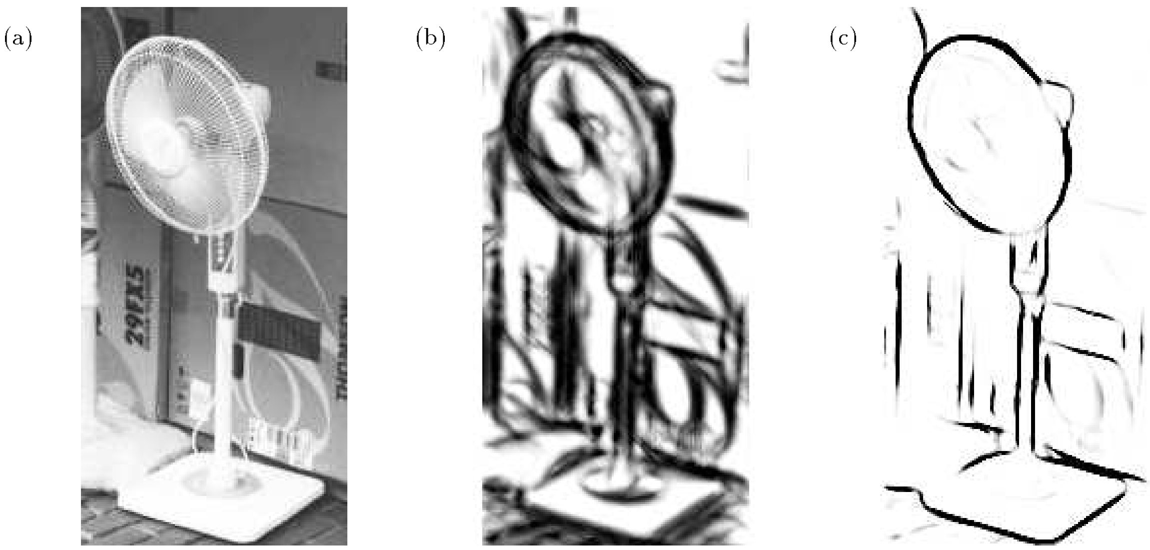

Figure 4. Comparison of BCS and BSN. (a) Input image of a fan. (b) Boundaries extracted by BCS with scale factor $\sigma = 1$. (c) Boundaries extracted by BSN with $\sigma = 1$.

at the corners can be quite weak. A method to overcoming this shortcoming is to use corner detectors to provide strong neural activations at corner points. Currently, BSN incorporates only the Gestalt principles of proximity, smoothness, and continuity in its processing. It would be interesting to see how other principles such as symmetry and parallelism can help to improve BSN's performance in object boundary segmentation.

6. Conclusion

An improved version of BCS, called BSN, is presented in this paper. BSN's boundary segmentation is scale-dependent. At a fine scale, detailed boundaries are extracted, whereas at a coarse scale, only the boundaries that represent objects' overall shapes are segmented. In comparison with BCS, the boundaries extracted by BSN is thinner and more well defined. BSN also performs better at rejecting noisy background details while segmenting objects' overall boundaries. These characteristics make BSN more suitable than BCS for segmenting the boundaries of objects in real images.

References

[1] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987.

[2] J. Dolan and E. Riseman. Computing curvilinear structure by token-based grouping. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1992.

[3] S. Grossberg and E. Mingolla. Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92:173–211, 1985.

[4] S. Grossberg and E. Mingolla. Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentation. *Perception & Psychophysics*, 38:141–171, 1985.

[5] S. Grossberg and L. Wyse. Figure-ground separation of connected scenic figures: Boundaries, filling-in, and opponent processing. In G. A. Carpenter and S. Grossberg, editors, *Neural Networks for Vision and Image Processing*. MIT Press, Cambridge, Massachusetts, 1992.

[6] R. Mohan and R. Nevatia. Segmentation and description based on perceptual organization. Technical report, University of Southern California, Los Angeles, 1989.

[7] I. Rock. *The Logic of Perception*. MIT Press, Cambridge, Massachusetts, 1983.

[8] S. Tabbone. Cooperation between edges and junctions for edge grouping. In *Proceedings of 4th International Conference on Image Processing*, pages 954–957, 1993.