Biologically-inspired Edge Detection Through Surround Modulation

Arash Akbarinia
www.cvc.uab.es/people/sakbarinia/
C. Alejandro Parraga
www.cvc.uab.es/people/aparraga/

Centre de Visió per Computador
Universitat Autònoma de Barcelona
Barcelona, Spain

Abstract

Edges are key components of any visual scene to the extent that we can recognise objects merely by their silhouettes. The human visual system captures edge information through neurons in the visual cortex that are sensitive to both intensity discontinuities and particular orientations. The "classical approach" assumes that these cells are only responsive to the stimulus present within their receptive fields, however, recent studies demonstrate that surrounding regions and inter-areal feedback connections influence their responses significantly. In this work we propose a biologically-inspired edge detection model in which orientation selective neurons are represented through the first derivative of a Gaussian function resembling double-opponent cells in the primary visual cortex (V1). In our model we account for four kinds of surround, *i.e.* full, far, iso- and orthogonal-orientation, whose contributions are contrast-dependant. The output signal from V1 is pooled in its perpendicular direction by larger V2 neurons employing a contrast-variant centre-surround kernel. We further introduce a feedback connection from higher-level visual areas to the lower ones. The results of our model on two benchmark datasets show a big improvement compared to the current non-learning and biologically-inspired state-of-the-art algorithms while being competitive to the learning-based methods.

1 Introduction

Our ability to recognise objects is completely entangled with our ability to perceive contours [24, 35]. The primary and secondary visual cortices – *i.e.* V1 and V2 – play a crucial role in the process of detecting lines, edges, contours, and boundaries [22], to such extent that an injury to these areas can impair a person's ability to recognise objects [42]. Furthermore, edges (a form of image gradient and sometime also referred to as "boundary" or "contour") are indispensable components of computer vision algorithms in a wide range of applications, such as colour constancy [34], image segmentation [2], document recognition [21] and human detection [8].

Given its importance, many computational models have been proposed to detect edges (for a comprehensive review refer to [24]). In its earliest form Prewitt [26] proposed a convolutional-based image gradient to capture local changes. Marr [23] suggested a correspondence between edges and zero-crossing points. Canny [5] improved on previous algorithms by incorporating non-maximum suppression and hysteresis thresholding. The greatest

Pages 5.1-5.13
DOI: https://dx.doi.org/10.5244/C.30.5

challenge faced by these classical methods is the distinction between authentic boundaries and undesired image textures. This issue was partially addressed by local smoothing techniques, such as bilateral filtering [33] and mean shift [6]. Thereafter, graph-based models emerged, *e.g.* [7, 12]. More recent frameworks extract relevant cues (*e.g.* colour, brightness and texture) feeding them to machine learning algorithms, such as probabilistic boosting tree [11], gradient descent [2] and structured forest [10]. Currently, state-of-the-art algorithms [3, 4, 19, 28, 38] rely heavily on deep-learning techniques.

Despite their success, learning methods suffer from two major drawbacks: (a) their performance might be dataset dependant and (b) they are computationally demanding since for every single pixel a decision must be made (in both training and testing steps) on whether it corresponds to an edge or not. In addition to these, there is no evidence from a biological point of view supporting the notion that edge detection is the result of such laborious learning process. On the contrary, edges have long been associated with low-level features that are modulated by feedback from higher-level visual areas, *e.g.* those responsible for global shape [22].

In line with this, a number of biologically-inspired edge detection models have been recently proposed with promising results. Spratling [31] proposed a predictive coding and biased competition based on the sparsity coding of neurons in V1. Wei *et al.* [36] presented a butterfly-shaped inhibition model based on non-classical receptive fields operating at multiple spatial scales. Further improvement came from Yang *et al.* [41] who explored imbalanced colour opponency to detect luminance boundaries. The same authors demonstrated employing the spatial sparseness constraint, typical to V1 neurons, helps to reserve desired fine boundaries while suppressing unwanted textures [40]. Another improvement in contour detection originated from introducing multiple features to the classical centre-surround inhibition common to most cortical neurons [39]. The introduction of feedback connections has also been beneficial. Díaz-Pernas [9] extracted edges through oriented Gabor filters accompanied with top-down and region enhancement feedback layers. In this work we propose a new edge detection model that incorporates recent knowledge of the physiological properties of cortical neurons. Our work is novel compared to the methods mentioned above in four main aspects: (i) we incorporate a more sophisticated set of cortical interactions which includes four types of surround, *i.e.*, full, far, iso- and orthogonal-orientation, (ii) we account for contrast variation of surround modulation, (iii) we model V2 neurons that pool signals from V1 responses over a larger region corresponding to the centre and neighbouring spatial locations, and (iv) we consider a fast-conducting feedback connection from higher visual areas to the lower ones.

Figure 1 illustrates the flowchart of our framework. At first, the original image is convolved by single opponent cells in the retina and sent though the lateral geniculate nucleolus (LGN) in the form of colour opponent channels [27]. These channels are processed by double-opponent cells in V1 (known to be responsive to colour edges [27]), whose receptive field (RF) are modelled through the first derivative of a Gaussian function [5]. To consider the RF surround: we define a short range circular (isotropic) region corresponding to full surround [22], long range iso- and orthogonal-orientation surrounds along the primary and secondary axes of the RF [13], and we model far surround via feedback connections to enhance the saliency of edge features. All these interactions are inversely dependant on the contrast of the RF centre [29]. The output signal from V1 is pooled at V2 by a contrast-variant centre-surround mechanism applied orthogonally to the preferred direction of the V1 RF [25]. Finally, to account for the impact of global shapes on local contours [22], we feed the output of V2 back into V1.

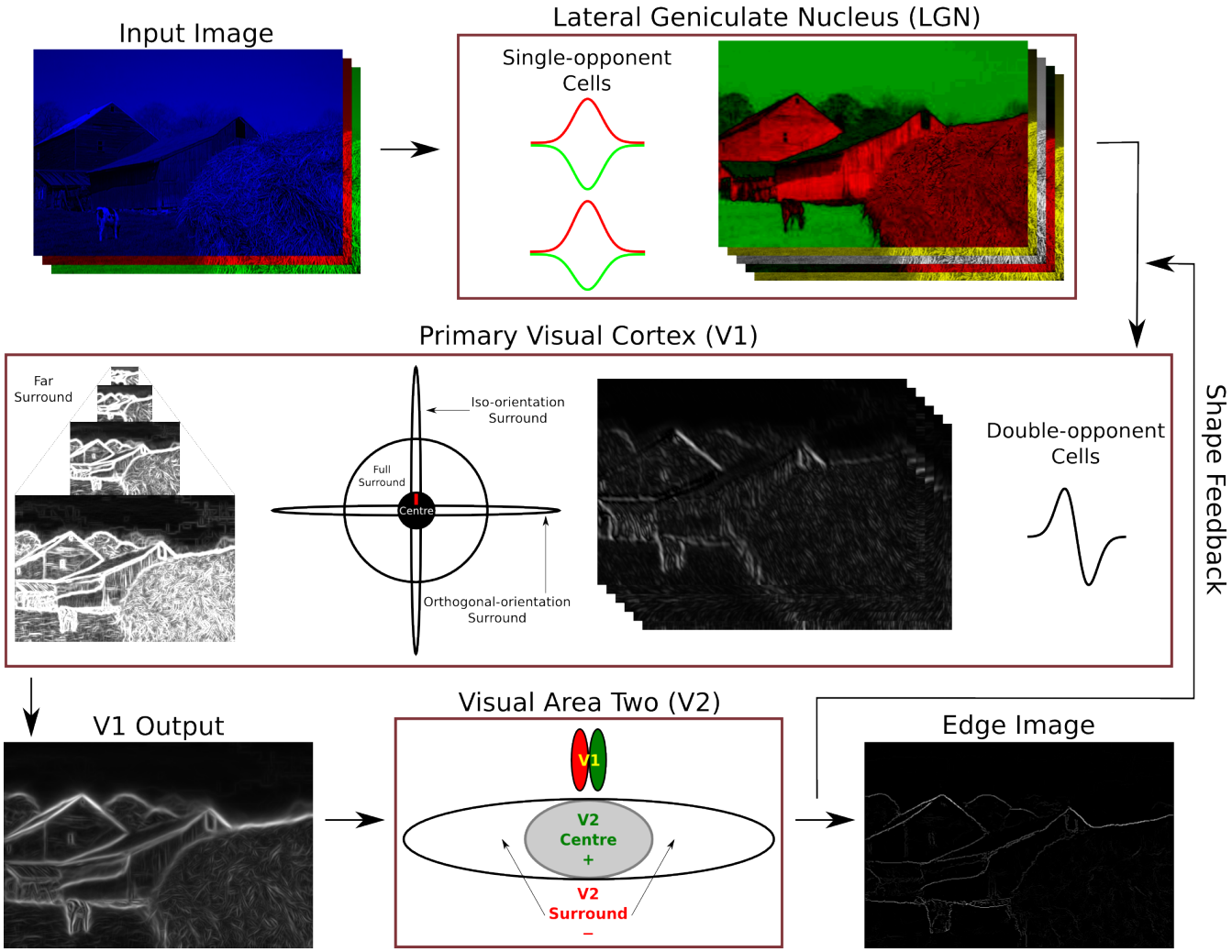

Figure 1: The flowchart of our model. Balanced and imbalanced colour opponent channels are created in the retina and sent through the LGN. Orientation information is obtained in V1 by convolving the signal with a derivative of Gaussian at twelve different angles. We model four types of orientation-specific surround: full, far, iso- and orthogonal-orientation. In V2 the signal is further modified by input from surrounding areas in a directional orthogonal to that of the original RF. Shape feedback is sent to V1 as an extra channel.

2 Surround Modulation Edge Detection

2.1 Retina and lateral geniculate nucleus (LGN)

Cone photoreceptor cells located at the back of the retina absorb photons at every spatial location. Their output is processed in an antagonistic manner by further layers of single-opponent cells (ganglion cells) and sent to the cortex through the LGN in the form of a luminance and two chromatically-opponent channels [27], usually modelled as

$$\begin{aligned} SO_{lu}(x,y) &= S_r(x,y) + S_g(x,y) + S_b(x,y), \\ SO_{rg}(x,y) &= \kappa_r S_r(x,y) - \kappa_g S_g(x,y), \\ SO_{yb}(x,y) &= \kappa_b S_b(x,y) - \kappa_{rg} \left(\frac{S_r(x,y) + S_g(x,y)}{2} \right), \end{aligned} \tag{1}$$

where SO represents the response of single-opponent cells, $\{lu, rg, yb\}$ represent the luminance, red-green and yellow-blue opponent-channels, (x,y) are the spatial coordinates, and

$\{r,g,b\}$ are the original red, green and blue cone signals. S is the spectral response function of each cone and can be approximated by a two dimensional Gaussian function as follows

$$S_h(x,y) = I_h(x,y) * G(x,y,\sigma), \tag{2}$$

where I is the input image, $h \in \{r,g,b\}$ is the index of colour channels, $*$ denotes the convolution operator and G is the Gaussian kernel, defined as

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}, \tag{3}$$

with variance σ. This Gaussian convolution is equivalent of a smoothing preprocessing stage in computer vision which has been demonstrated to play an important role in the successive edge detection [24].

When the chromatically-opponent input to single-opponent cells in Eq. 1 is equilibrium, parameter κ is equal to one for all channels. However, there is physiological evidence showing that some types of single-opponent cells combine chromatic information in an imbalanced fashion [27]. The significance of these cells has also been shown in practice through many computer vision algorithms, *e.g.* edge detection [40, 41] and colour constancy [14]. Following this insight, we included two imbalanced opponent-channels: $SO_{rg'}$ with $\kappa_g = 0.7$ and $SO_{yb'}$ with $\kappa_{rg} = 0.7$.

2.2 Primary visual cortex (V1)

SO channels are processed by a number of double-opponent cells in V1 that are responsive to boundaries [27]. The response of these cells are modulated by regions beyond their RF centres, with facilitation predominantly at low contrast and inhibition at high contrast [1, 18, 29]. Thus, we defined our orientation-tuned double-opponent cells *DO* as

$$DO_c(x,y,\theta) = CR_c(x,y,\theta) + \zeta_c^{-1}(x,y) SR_c(x,y,\theta), \tag{4}$$

where c is the index of *SO* channels, θ is the preferred orientation of the RF (set to twelve evenly distributed angles in the range $[0,2\pi)$ [25]), *CR* and *SR* are the centre and surround responses respectively, and ζ is the contrast of the RF centre approximated by the local standard deviation of its constituent pixels. Double-opponent cells are typically modelled in biologically-inspired algorithms by Gabor filters, [9, 31, 39], or the first derivative of a Gaussian function, [40, 41]. We settled for the later one originally proposed by Canny [5]. Therefore, we defined the *DO* centre response, *CR*, as

$$CR(x,y,\theta) = SO * \left| \frac{\vartheta G(x,y,\sigma)}{\vartheta \theta} \right|, \tag{5}$$

where σ is the RF size (set to 1.5 in our model corresponding to the typical RF size of foveally-connected neurons in V1 or 0.25° of visual angle [1], which is equivalent to approximately 13 pixels when viewed from $100cm$ in a standard monitor).

2.2.1 Surround modulation

We defined the surround response, *SR*, as follows

$$SR(x,y,\theta) = LS(x,y,\theta) + IS(x,y,\theta) + OS(x,y,\theta) + FS(x,y,\theta), \tag{6}$$

where *LS* is full surround referring to the isotropic region around the RF; *IS* denotes iso-orientation surround along the RF preferred axis; *OS* is orthogonal-orientation surround in the direction perpendicular to the RF preferred axis; and *FS* denotes far surround.

Because the full surround is an isotropic region (*i.e.* stimulus occupying the entire surrounding region rather than isolated flanking lines [22]) it can be modelled as the average response of a circular window around the cell's RF centre. This surround is inhibitory when it shares the same orientation as the centre and strongly facilitatory when its orientation is perpendicular to the centre [22]. Thus, we defined the full surround *LS* as

$$LS(x,y,\theta) = \lambda\,\zeta^{-1}(x,y)\Big(CR(x,y,\theta_\perp) * \mu\Big) - \zeta(x,y)\Big(CR(x,y,\theta) * \mu\Big), \tag{7}$$

where $\theta_\perp = \theta + \frac{\pi}{2}$, μ is the circular average kernel and λ determines the strength of orthogonal facilitation in comparison to the iso inhibition. The former facilitation is reported to be stronger than the later inhibition [22], therefore λ must be larger than one.

The iso-orientation surround, *IS*, extends to a distance two to four times larger than the RF size [13]. Within this region elements parallel to the RF preferred orientation are facilitatory while orthogonal ones are inhibitory [13, 22], therefore, we modelled *IS* as

$$IS(x,y,\theta) = \zeta^{-1}(x,y)\Big(CR(x,y,\theta) * E(\sigma_x,\theta)\Big) - \zeta(x,y)\Big(CR(x,y,\theta_\perp) * E(\sigma_x,\theta)\Big), \tag{8}$$

where E is an elliptical Gaussian function elongated in the direction θ, defined as

$$E(x,y,\sigma_x,\sigma_y,\theta) = e^{-\left(ax^2-2bxy+cy^2\right)},$$
$$a = \frac{\cos^2\theta}{2\sigma_x^2} + \frac{\sin^2\theta}{2\sigma_y^2}, \quad b = -\frac{\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2}, \quad c = \frac{\sin\theta^2}{2\sigma_x^2}\frac{\cos\theta^2}{2\sigma_y^2}$$

We set $\sigma_y = 0.1\sigma_x$ and $\sigma_x = 3\sigma$ corresponding to physiological measurements [13].

The orthogonal-orientation surround, *OS*, projects to a distance half of the iso-orientation surround [13]. In the orthogonal-surround elements parallel to the RF preferred orientation are inhibitory while perpendicular ones are facilitatory [13, 22], thus, we modelled *OS* as

$$OS(x,y,\theta) = \zeta^{-1}(x,y)\Big(CR(x,y,\theta_\perp) * E(\sigma_x,\theta_\perp)\Big) - \zeta(x,y)\Big(CR(x,y,\theta) * E(\sigma_x,\theta_\perp)\Big). \tag{9}$$

The far surround could extend to regions up to 12.5° of visual angle [29] which is approximately equivalent to 673 pixels when viewed from 100*cm* in a standard monitor. Consequently the feedforward and horizontal connections in V1 that mediate interactions between the RF and its near surround are too slow to account for the fast onset of far surround. Due to this, it has been suggested that far surround is operated through a different mechanism via inter-areal feedback connections [1, 30]. We speculate that parts of these inter-areal connections come from spatial scale layers in V1 [16], and assume their influence to be facilitatory when image elements in this region share the same orientation as the centre [17]. Therefore, we defined *FS* as

$$FS(x,y,\theta) = \zeta^{-1}(x,y)\sum_{s=2}^{4}\frac{CR_s(x,y,\theta)}{s} \tag{10}$$

where s is the index of the corresponding spatial frequency scale. This processing is analogous to the multi-scale processing common to both visual sciences and computer vision, with the distinction that we account for both contrast and distance, since surround modulation has been reported to be stronger in the near than in the far regions [1].

2.3 Visual area two (V2)

Visual processing becomes more global along the brain's ventral pathway. In line with this, many V2 neurons have been reported to respond to curvatures and extended arcs [37]. It has been proposed that V2 RFs extract curvature information by pooling signals from V1 using a centre-surround mechanism in the direction orthogonal to the V1 orientations [25, 37]. In order to model this, first, we defined the V1 response, $V1R$, as the most activated DO orientation. This operation is assumed to be realised by complex cells pooling the maximum value of DO cells [32], modelled as

$$V1R_c(x,y) = \underset{\theta\in[0,2\pi)}{\arg\max}\left(DO_c(x,y,\theta)\right). \tag{11}$$

The V2 RFs show similar contrast-variant surround modulation as those of V1 [29]. Therefore, we modelled the V2 response, $V2R$, through a Difference-of-Gaussians (DoG) as

$$V2R_c(x,y) = V1R_{c,\theta}(x,y) * E(\sigma_x,\theta_\perp) - \upsilon_c(x,y)V1R_{c,\theta}(x,y) * E(5\sigma_x,\theta_\perp) \tag{12}$$

where υ is the contrast of $V1R$ computed by its local standard deviation, the index θ at $V1R_\theta$ shows the preferred orientation of that RF. Cortical RFs increase their diameters systematically by approximately a factor of three from lower to higher areas [37]. Therefore, we set the size of V2 RF, σ_x, to three times the size of a V1 RF. In Eq. 12 surround is five times larger than the centre according to physiological findings [29].

There are massive feedback connections from higher visual areas to the lower ones [1]. In our model we accounted for only a fraction of those from V2 to V1 corresponding to the well established fact that global shape influences local contours [22]. We simulated this global shape by averaging the V2 outputs of all channels and sending it as feedback to V1. This feedback is processed as all other inputs to V1. The final edge map is created as a sum of all V2 output channels

$$edge(x,y) = \sum_c V2R_c(x,y), \quad \text{with} \quad c \in \{lu, rg, yb, rg', yb', feedback\}. \tag{13}$$

3 Experiments and results

We tested our model – termed Surround-modulation Edge Detection (SED) – on two datasets[1], the Berkeley Segmentation Dataset and Benchmark (BSDS) [2] and the Contour Image Database (CID) [15]. The former includes two sets of colour images BSDS300 (100 test images) and BSDS500 (200 test images). The later contains forty grey-scale images. The ground truth of both datasets are edges manually-drawn by human subjects. We evaluated our algorithm in the standard precision-recall curve based on their harmonic mean (referred to as F-measure) on three criteria: optimal scale for the entire dataset (ODS) or per image (OIS) and average precision (AP). The results we report in this paper were obtained with a fixed set of parameters (see details in Section 2) for all datasets much in the same way as the human visual system. Table 1 compares the results of our model to several other state-of-the-art algorithms that have also reported theirs results on the BSDS dataset. Table 2 compares the results of SED to four algorithms driven by low-level features on the CID dataset.

[1]The source code and all the experimental materials are available at https://github.com/ArashAkbarinia/BoundaryDetection.

In our proposal, we have modelled different areas and aspects of the visual cortex. In order to investigate the contribution of each of the components of our model we conducted four experiments on the BSDS500 dataset.

- **Gaussian Derivative** – In this scenario, we accounted neither for the surround modulation in V1, nor for the V2 pooling and feedback. Essentially only convolving the single-opponent cells with the first derivative of Gaussian function similar to CO [41].
- **Only V1 Surround** – In this case, we excluded V2 pooling and feedback. We only included full, far, iso- and orthogonal-orientation surround modulation for V1 RFs.
- **No V2 Feedback** – In this scenario, we excluded the feedback sent from V2 to V1, *i.e.* $c \in \{lu, rg, yb, rg', yb'\}$ in Eq. 13.
- **No Far surround** – In this case, we did not account for far surround modulation, *i.e.* $FS = 0$ in Eq. 6.

The result of these experiments are reported in Figure 2.

		Method		BSDS300 ODS	BSDS300 OIS	BSDS300 AP	BSDS500 ODS	BSDS500 OIS	BSDS500 AP
		Human		0.79	0.79	–	0.80	0.80	–
Biological	Low-level features	Canny	(1986) [5]	0.58	0.62	0.58	0.60	0.63	0.58
		Mean Shift	(2002) [6]	0.63	0.66	0.54	0.64	0.68	0.56
		Felz-Hutt	(2004) [12]	0.58	0.62	0.53	0.61	0.64	0.56
		Normalised Cuts	(2005) [7]	0.62	0.66	0.43	0.64	0.68	0.45
		PC/BC	(2013) [31]	0.61	–	–	–	–	–
		CO	(2013) [41]	0.64	0.66	0.65	0.64	0.68	0.64
		MCI	(2014) [39]	0.62	–	–	0.64	–	–
		dPREEN	(2014) [9]	0.65	–	–	–	–	–
		SCO	(2015) [40]	0.66	0.68	0.70	0.67	0.71	0.71
Deep-learning	Machine-learning	BEL	(2006) [11]	0.65	–	–	0.61	–	–
		gPb	(2011) [2]	0.70	0.72	0.66	0.71	0.74	0.65
		DeepNets	(2014) [19]	–	–	–	0.74	0.76	0.76
		DeepEdge	(2015) [3]	–	–	–	0.75	0.75	0.80
		DeepContour	(2015) [28]	–	–	–	0.76	0.77	0.80
		HFL	(2015) [4]	–	–	–	0.77	0.79	0.80
		HED	(2015) [38]	–	–	–	0.78	0.80	0.83
		SED (Proposed)		**0.69**	**0.71**	**0.71**	**0.71**	**0.74**	**0.74**

Table 1: Results of several edge detection algorithms on the BSDS300 and BSDS500 [2].

4 Discussion

The results of our model on BSDS and CID datasets demonstrate a systematic quantitative improvement of about 4% comparing to the state-of-the-art biologically-inspired algorithms in all three criteria of ODS, OIS and AP (see Tables 1 and 2). This improvement is also qualitatively evident in Figure 3. On the one hand, our model shows greater robustness in textural areas in comparison to CO [41] as it is evident in the rocks image (third row) and the grass image (fourth row). On the other hand, thanks to its surround modulation, SED does a

Method	CID		
	ODS	OIS	AP
Canny [5]	0.56	0.64	0.57
CO [41]	0.55	0.63	0.57
MCI [39]	0.60	0.63	0.53
SCO [40]	0.58	0.64	0.61
SED	**0.65**	**0.69**	**0.68**

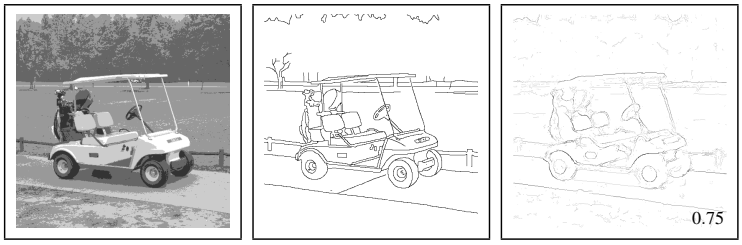

Table 2: Results of five edge detection algorithms on the CID dataset [15]. On the right: one example result of our model is shown with the F-measure on the right bottom corner.

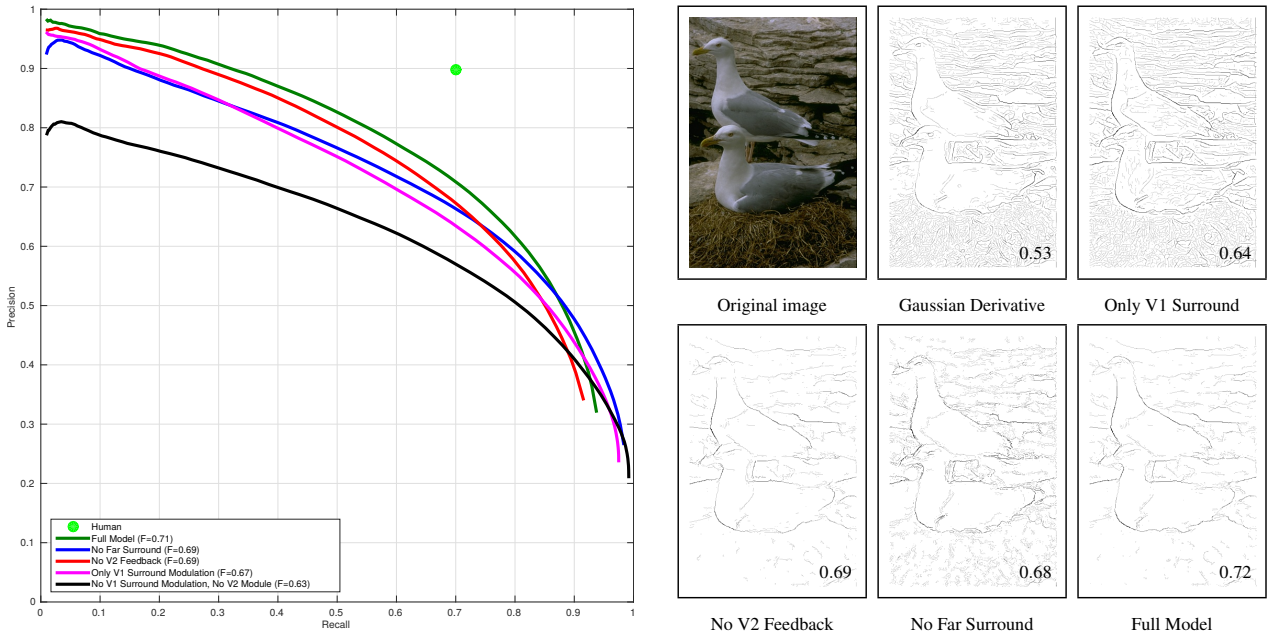

Figure 2: Evaluation of the different components of SED. The graph on the left is the precision-recall curve on the BSDS500. In the legends the ODS F-measures are indicated. The images on the right show the result of our full model on one exemplary image along with the four experiments we conducted. F-measures are on the right bottom corner of images.

better job at detecting continuous lines, compared to SCO [40] which is evident for images in the second and fourth rows where the land and sea boundaries are more clearly defined.

Our improvements over the state-of-the-art originates from the combination of different components of our model. The precision-recall curve in Figure 2 shows that excluding surround modulation and the V2 module all together drops the ODS F-measure to 0.63 (black curve). This is in line with the results of CO [41] which essentially is same as our model in the absence of V1 surround modulation and the V2 module. Including surround modulation (*i.e.* full, far, iso- and orthogonal-orientation regions) raises the F-measure to 0.67 (pink curve). This clearly shows that surrounding regions play a crucial role in edge detection in agreement with previous psychophysical findings [22].

Comparison of *"Only V1 Surround"* and *"No V2 Feedback"* pictures in Figure 2 reveals that the V2 module strongly assist the process of eliminating textures. This is consistent with physiological findings that suggest texture statistics might be encoded in V2 [20] and could be explained by the fact that V2 RFs are large and therefore suppress tiny textural pixels. Although V2 centre-surround suppression is beneficial in general (2% improvement in F-measure of the red curve in comparison to the pink one), it causes occasional over-smoothing and consequently in high recalls the precision of the pink curve is higher than the blue one. We postulate that this problem can be addressed by a mechanism similar to the

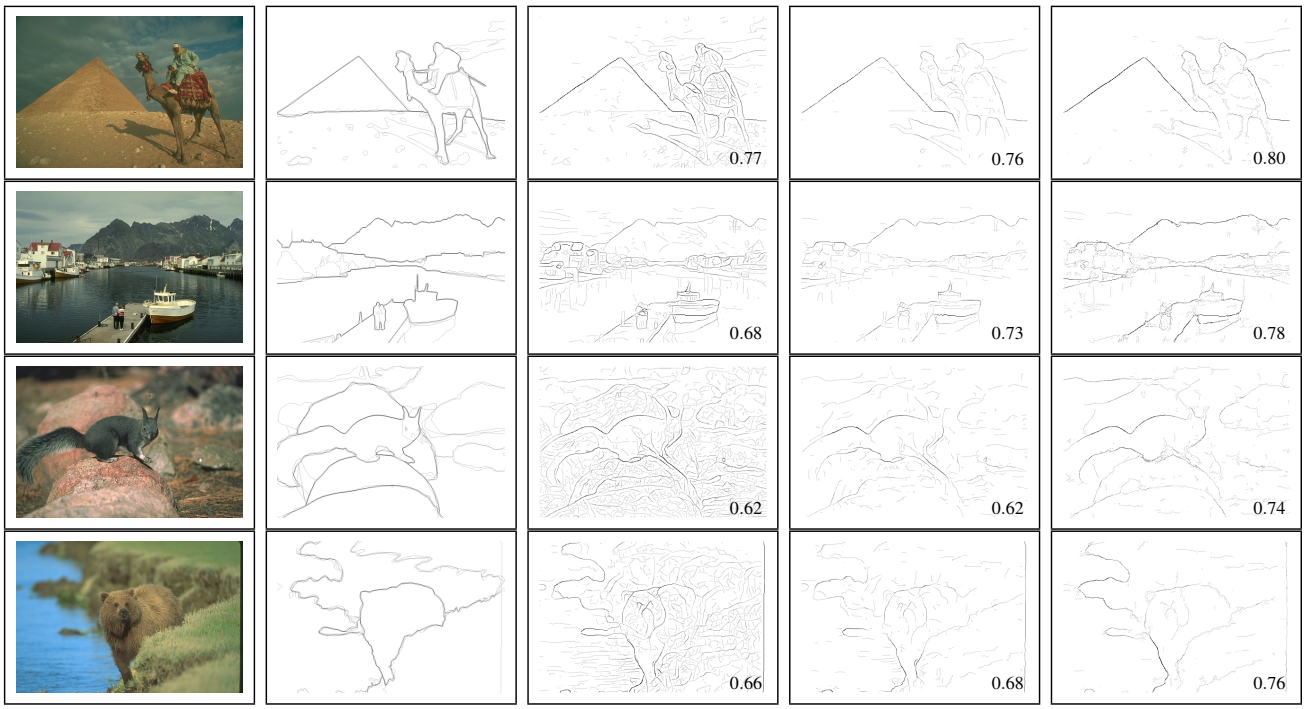

Figure 3: Edge detection results of three biologically-inspired methods. The F-measure is indicated on the right bottom corner. The first two rows show results for a picture from the BSDS300 dataset and the last two rows from BSDS500 dataset.

visual cortex where suppression can change to facilitation at low contrast levels [1]. Modelling this phenomenon is onerous since the threshold between suppression and facilitation is cell specific and there is no universal contrast level or surround stimulus size that triggers facilitation across the entire cell population [1]. We propose this as a line of future work.

The precision-recall curve in Figure 2 shows that excluding far surround modulation reduces the ODS F-measure to 0.69 (blue curve), which still is better than other non-learning state-of-the-art algorithms. A qualitative comparison of *"No Far Surround"* and *"Full Model"* results reveals that far surround appears to play an important role in enhancing continuous edges while suppressing broken ones. We observe similar patterns with V2 surround modulation. In high recall *"No Far Surround"* has a higher precision than *"Full Model"* (blue versus green curves). Resolving this is a subject for further investigation.

5 Conclusion

In this paper, we presented a biologically-inspired edge detection model grounded on physiological findings of the visual cortex. Our main contributions can be summarised as follows: (i) modelling a contrast-dependant surround modulation of V1 receptive fields by accounting for full, far, iso- and orthogonal-orientation surround; (ii) introducing a V2 centre-surround mechanism to pool V1 signals in their perpendicular orientations; and (iii) accounting for a shape feedback connection from V2 to V1. We quantitatively compared the proposed model to current state-of-the-art algorithms on two benchmark datasets. Our results show a significant improvement compared to the other non-learning and biologically-inspired models while being competitive to the learning ones.

The results of our experiments suggest that V1 surround modulation strengthens edges while V2 contributes to the suppression of undesired textural elements. We can further

improve our model by accounting for the complex shape processing occurring in V4, for example, by concentric summation of V2 signals [37]. Moreover, within our framework we treat different surrounds as individual entities with no interactions between them. This simplification does not consider psychophysical studies showing the non-linear interactions that occur between surround and central regions which depend on the configurations of both inducers and targets [22]. This will be the subject of a future line of work to account for these configurational settings by relating full, iso- and orthogonal-orientation surrounds.

Biologically-inspired solutions such as the one presented here make two contributions, one technological (pushing forward state-of-the-art) and the other scientific (understanding the relationship between the human visual system and the visual environment). The more we learn about the properties of the human visual system the better we can explain visual behaviour. Within our limitations (both in knowledge and resources) we have tried to keep our modelling decisions as close as possible to what we know about the physiology, in two main respects: (a) our architecture reflects the low-level features that are common to mammalian cortical architecture and emerged after millions of years of evolution (*i.e.* are not ad-hoc or dataset-dependant) and (b) our model parameters are the same in all experiments, which is a feature of how the human visual system operates.

Acknowledgements

This work was funded by the Spanish Secretary of Research and Innovation (TIN2013-41751-P and TIN2013-49982-EXP).

References

[1] A Angelucci and S Shushruth. Beyond the classical receptive field: Surround modulation in primary visual cortex. *The new visual neurosciences*, pages 425–444, 2013.

[2] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):898–916, 2011.

[3] Gedas Bertasius, Jianbo Shi, and Lorenzo Torresani. Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4380–4389, 2015.

[4] Gedas Bertasius, Jianbo Shi, and Lorenzo Torresani. High-for-low and low-for-high: Efficient boundary detection from deep object features and its applications to high-level vision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 504–512, 2015.

[5] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698, 1986.

[6] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5): 603–619, 2002.

[7] Timothee Cour, Florence Benezit, and Jianbo Shi. Spectral segmentation with multiscale graph decomposition. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 1124–1131. IEEE, 2005.

[8] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[9] Francisco J Díaz-Pernas, Mario Martínez-Zarzuela, Míriam Antón-Rodríguez, and David González-Ortega. Double recurrent interaction v1–v2–v4 based neural architecture for color natural scene boundary detection and surface perception. *Applied Soft Computing*, 21:250–264, 2014.

[10] Piotr Dollár and C Lawrence Zitnick. Fast edge detection using structured forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(8):1558–1570, 2015.

[11] Piotr Dollar, Zhuowen Tu, and Serge Belongie. Supervised learning of edges and object boundaries. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1964–1971. IEEE, 2006.

[12] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[13] David J Field, James R Golden, and Anthony Hayes. Contour integration and the association field. *The new visual neurosciences*, pages 627–638, 2013.

[14] Shaobing Gao, Kaifu Yang, Chaoyi Li, and Yongjie Li. A color constancy model with double-opponency mechanisms. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 929–936, 2013.

[15] Cosmin Grigorescu, Nicolai Petkov, and Michel A Westenberg. Contour detection based on nonclassical receptive field inhibition. *Image Processing, IEEE Transactions on*, 12(7):729–739, 2003.

[16] Robert F Hess. Spatial scale in visual processing. *The new visual neurosciences*, pages 595–615, 2013.

[17] Jennifer M Ichida, Lars Schwabe, Paul C Bressloff, and Alessandra Angelucci. Response facilitation from the âĂIJsuppressiveâĂİ receptive field surround of macaque v1 neurons. *Journal of Neurophysiology*, 98(4):2168–2181, 2007.

[18] Mitesh K Kapadia, Gerald Westheimer, and Charles D Gilbert. Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proceedings of the National Academy of Sciences*, 96(21):12073–12078, 1999.

[19] Jyri J Kivinen, Christopher KI Williams, Nicolas Heess, and DeepMind Technologies. Visual boundary prediction: A deep neural prediction network and quality dissection. In *AISTATS*, volume 1, page 9, 2014.

[20] Michael S Landy. Texture analysis and perception. *The new visual neurosciences*, pages 639–652, 2013.

[21] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[22] Gunter Loffler. Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision Research*, 48(20):2106–2127, 2008.

[23] David Marr and Ellen Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London B: Biological Sciences*, 207(1167):187–217, 1980.

[24] Giuseppe Papari and Nicolai Petkov. Edge and line oriented contour detection: State of the art. *Image and Vision Computing*, 29(2):79–103, 2011.

[25] Frédéric JAM Poirier and Hugh R Wilson. A biologically plausible model of human radial frequency perception. *Vision research*, 46(15):2443–2455, 2006.

[26] Judith MS Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1):15–19, 1970.

[27] Robert Shapley and Michael J Hawken. Color in the cortex: single-and double-opponent cells. *Vision research*, 51(7):701–717, 2011.

[28] Wei Shen, Xinggang Wang, Yan Wang, Xiang Bai, and Zhijiang Zhang. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3982–3991, 2015.

[29] S Shushruth, Jennifer M Ichida, Jonathan B Levitt, and Alessandra Angelucci. Comparison of spatial summation properties of neurons in macaque v1 and v2. *Journal of neurophysiology*, 102(4):2069–2083, 2009.

[30] S Shushruth, Lauri Nurminen, Maryam Bijanzadeh, Jennifer M Ichida, Simo Vanni, and Alessandra Angelucci. Different orientation tuning of near-and far-surround suppression in macaque primary visual cortex mirrors their tuning in human perception. *The Journal of Neuroscience*, 33(1):106–119, 2013.

[31] Michael W Spratling. Image segmentation using a sparse coding model of cortical area v1. *Image Processing, IEEE Transactions on*, 22(4):1631–1643, 2013.

[32] Christian Thériault, Nicolas Thome, and Matthieu Cord. Cortical networks of visual recognition. *Biologically Inspired Computer Vision: Fundamentals and Applications*, 2015.

[33] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, 1998.

[34] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214, 2007.

[35] Dirk B Walther, Barry Chai, Eamon Caddigan, Diane M Beck, and Li Fei-Fei. Simple line drawings suffice for functional mri decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, 108(23):9661–9666, 2011.

[36] Hui Wei, Bo Lang, and Qingsong Zuo. Contour detection model with multi-scale integration based on non-classical receptive field. *Neurocomputing*, 103:247–262, 2013.

[37] HR Wilson and F Wilkinson. Configural pooling in the ventral pathway. *Werner JS Chalupa L.(Eds.) The new visual neurosciences*, pages 617–626, 2014.

[38] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1395–1403, 2015.

[39] Kai-Fu Yang, Chao-Yi Li, and Yong-Jie Li. Multifeature-based surround inhibition improves contour detection in natural images. *Image Processing, IEEE Transactions on*, 23(12):5020–5032, 2014.

[40] Kai-Fu Yang, Shao-Bing Gao, Ce-Feng Guo, Chao-Yi Li, and Yong-Jie Li. Boundary detection using double-opponency and spatial sparseness constraint. *Image Processing, IEEE Transactions on*, 24(8):2565–2578, 2015.

[41] Kaifu Yang, Shaobing Gao, Chaoyi Li, and Yongjie Li. Efficient color boundary detection with color-opponent mechanisms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2810–2817, 2013.

[42] S Zeki. A vision of the brainblackwell scientific publications, 1993.