PCTV: A BIOLOGICALLY- AND PSYCHOLOGICALLY-INSPIRED EDGE AND LINE DETECTION

X. Kang[1], *P.Y.S. Cheung*[2], *W.P. Yau*[1], *Y. Hu*[1]

[1]Department of Orthopaedics & Traumatology, Li Ka Shing Faculty of Medicine,
[2]Department of Electrical & Electronic Engineering, Faculty of Engineering,
The University of Hong Kong

ABSTRACT

A novel method for detecting edges and lines simultaneously and automatically is proposed. This method, based on phase congruency and tensor voting (hence PCTV), makes use of the properties of how edges and lines are built from the Fourier decomposition of an image, and how the primary visual cortex responds to them, instead of making assumptions on the intensity profiles of the regions near a feature. Experiments showed that the detection results were more consistent to the "ground truth" manually drawn by humans. For detecting edges, this method is superior to three commonly used detectors in that it reduces the production of false detections.

Index Terms— Edge detection, line detection, phase congruency, tensor voting, primary visual cortex (V1)

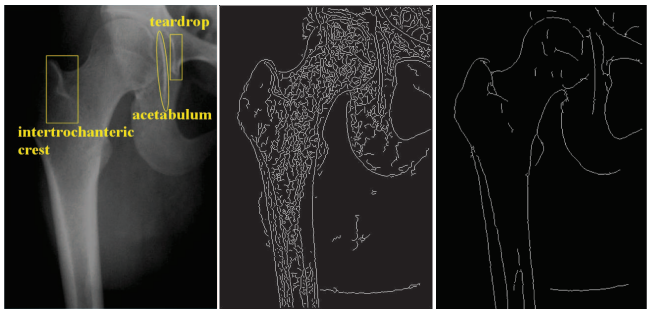
1. INTRODUCTION

Detection of low-level features such as edges and lines (including straight lines and curves) is crucial for image analysis, pattern recognition and computer vision. Edge detection has been studied for a long time and used in various tasks. But features in real images are often composed of combinations of step edges, ramps, roofs and lines.

Fig. 1(a) shows a typical X-ray image in which the proximal femur overlaps with the pelvis. While the outlines of the bones are edges, the acetabulum, the teardrop and the intertrochanteric crest projected from the original 3D structures are line-like features. In addition, the bones are not homogeneous in intensity; there are many trivial details due to the differences in bone density and the image acquisition conditions in clinical practice.

To obtain an anatomically meaningful interpretation of the image, one needs to detect the edges and lines simultaneously. However, most edge detectors (e.g. Canny [1]) always mark two "edges" on either side of a line-like feature (Fig. 1(b)). Neither of them is the feature of interest, and neither is localized at the actual position of the feature. This substantially affects the robustness of subsequent analyses.

One approach may be to detect edges and lines separately and then fuse them together. But in addition to the difficulties in fusion methods, one has to first distinguish which edges in the output of an edge detector are the "real edges," which is a much more complicated recognition problem. Therefore, a method is needed to detect edges and lines simultaneously, without presupposing the feature type.



(a) (b) (c)

Fig. 1. Automatic detections of edges and lines in an X-ray image (a) by Canny detector ($\sigma = 1$) (b) and our method (c).

Trivial details often introduce false detections because their existence violates the fundamental assumptions of most edge or line detectors. Treated as a kind of noise, they may be suppressed by spatial smoothing. However, this will make weak edges/lines harder to be detected. High-degree smoothing will produce significant shift in the localization of the features, which is unacceptable for many biomedical and anatomical images.

It seems clear in the literature that the early stage of vision serves to compress the input image [2], and one form of such compression involves finding information-rich features such as edges and lines. It also has been shown that the responses of primary visual cortex (V1) neurons have a close correlation with the perceptual saliency of contours [3]. On the other hand, both retinal and digital images, as raw data, are totally devoid of structure. Some general laws must be followed so that groups of pixels can be organized into spatially extended features. These laws are stated in the Gestalt theory, which focuses on the characteristics of stimuli cause us to structure or interpret a visual field in a certain way. This motivated

the proposals of edge detector [4] and straight line segment detector [5] using stochastic geometry.

In this paper, we are primarily concerned with the modeling of the simultaneous detection of edges and lines in one unified framework, which few edge or line detectors can do. Our goal is to show that a more adequate modeling that makes use of the response properties of V1 neurons to visual stimuli does translate into better feature detection. We model how edges and lines are built up in terms of the Fourier decomposition of an image, and how V1 population responds to them, rather than modeling the neighborhoods near an edge or a line as constants, as statistic distributions of intensity values [6], or as spatial distributions of pixels [7]. The trivial details are moderately inhibited, without spatial smoothing, because the neuron responses to them are inconsistent with their responses to edges or lines in the sense of some Gestalt laws[1].

This method makes no assumption about the shape of the features it is trying to detect; it is insensitive to the relative contrast and uneven illumination in images; and the denoising and feature detection are performed simultaneously,

2. PROPOSED METHOD

We model the behavior of V1 population, i.e., the visual processes that rely on the synergy of many neurons. It has similar characteristics to the neuron code [8, 9]: spatial localization, bandpass orientation tuning, bandpass spatial frequency tuning, quadrature phase, rotational similarity, similarity across scale, and long-range interaction.

2.1. Modeling Simple and Complex Cells in V1

The majority of neurons in V1 respond to an edge or a line of a certain orientation [8]. There are two types of orientation-selective cells: simple and complex. They are characterized by their classical receptive field (RF) properties.

As for simple cells, the closest approximation to their RF is the two-dimensional (2D) Gabor filter [10], as formulated in (1). The filter (or cell), spatially localized at $(x_0, y_0) \in \mathbb{R}^2$ and oriented along the angle θ, is tuned to the optimal spatial frequency f, and has an RF phase ϕ. The bandwidth and the ellipticity of RF are determined by σ and γ.

$$\begin{aligned} g(x,y) &= e^{-\left(x'^2+\gamma^2 y'^2\right)/\sigma^2} \cos\left(2\pi f x' + \phi\right), \\ x' &= (x-x_0)\cos(\theta) + (y-y_0)\sin(\theta), \\ y' &= -(x-x_0)\sin(\theta) + (y-y_0)\cos(\theta). \end{aligned} \tag{1}$$

Tuning to different phases, simple cells/filters represent images in terms of even-symmetric, odd-symmetric, and intermediate structures.

Complex cells receive input from simple cell subunits of different phases [11], and are modeled by a quadrature pair, which leads to representations of even-symmetric axial structures (lines) and odd-symmetric oblique structures (edges).

[1] see http://www.sapdesignguild.org/resources/optical_illusions/gestalt_laws.html for some illustrations of these laws.

2.2. Modeling the Behavior of V1 Population

2.2.1. Modeling response properties to edges and lines

We model the neuron responses to edges and lines using phase congruency (PC) [12], because phase is a very important cue for feature detection [13, 14]. Recent neurological investigations have also shown that very few neurons were phase insensitive whereas most were tuned to PC [15], and this congruence across scales indicates significant structures in scale-space [16].

A robust measure of PC [17] is adopted. For each given orientation o,

$$PC_o(x,y) = \frac{W_o(x,y)\,\lfloor E_o(x,y) - T_o \rfloor}{\sum_n A_{no}(x,y) + \varepsilon}, \tag{2}$$

where $E_o(x,y)$ is the local energy computed using quadratic Gabor filters across scales, A_{no} the filter response amplitude at scale n, $W_o(x)$ a function weighting the frequency spread, T_o the radius of the estimated noise circle, $\lfloor \cdot \rfloor$ the round operator, and ε a small constant (see [17] for details).

Instead of thresholding the PC measure of an image (as in [17]), we use tensors to represent the PC measure at each position since the tensor field is superior to the vector or scalar field in describing local structures. The tensor is built up as

$$\boldsymbol{T} = \begin{bmatrix} \sum_o PC_o^2 \cos^2\theta_o & \sum_o PC_o^2 \sin 2\theta_o \\ \sum_o PC_o^2 \sin 2\theta_o & \sum_o PC_o^2 \sin^2\theta_o \end{bmatrix}, \tag{3}$$

where θ_o is the angle corresponding to the orientation.

2.2.2. Modeling Long-range Interactions

The response of a neuron in V1 is significantly influenced by stimuli outside its classical RF [9] in a region called its context. Recent neurophysiological experiments have also shown increased information content in synchronous neuronal activity compared with the activity of single neurons [18]. This interaction between the computation of nearby neurons makes a neuron's response clearly reflect global properties of the input image [19].

Though the activities of the neurons show excitatory and inhibitory responses, we do not distinguish between excitatory and inhibitory neurons. Their responses do not have to be classified prematurely or dealt with solely, allowing us to determine whether a response is excitatory or inhibitory after enough information has been collected. This simplifies the model without losing generality of the neural mechanisms.

At this stage, each point is represented by a tensor $\boldsymbol{T}$. Expressed using its eigenvalues $\lambda_{1,2}$, $(\lambda_1 \geq \lambda_2)$ and eigenvectors $\boldsymbol{e}_{1,2}$, it can be written as

$$\boldsymbol{T} = (\lambda_1 - \lambda_2)\,\boldsymbol{e}_1\boldsymbol{e}_1^T + \lambda_2\left(\boldsymbol{e}_1\boldsymbol{e}_1^T + \boldsymbol{e}_2\boldsymbol{e}_2^T\right). \tag{4}$$

In this form, the first term depicts the local structure having one preferred orientation, $\boldsymbol{e}_1$, with certainty $(\lambda_1 - \lambda_2)$ [20].

The contextual influences are formulated using tensor voting [20]. Assume there is a neuron at spatial position $p \in \mathbb{R}^2$. If its response to the stimulus in its classical RF has one preferred orientation, its excitation or inhibition to the neuron at position $q \in \mathbb{R}^2$ is represented as

$$\mathbf{F}_S(\boldsymbol{v}) = e^{-\frac{s^2+c\kappa^2}{\sigma^2}} \boldsymbol{t}\boldsymbol{t}^T, \quad \boldsymbol{t} = [-\sin 2\theta, \cos 2\theta]^T. \qquad (5)$$

Otherwise, its isotropic effect does not influence the local structure and is omitted. In (5), $\boldsymbol{v}$ is the vector pointing from p to q, s the arc-length, κ the curvature, θ the angle between $\boldsymbol{v}$ and the tangent of the osculating circle at p, σ the spatial scale and c a factor determined by σ (see [20] for details).

Denote as $\mathcal{N}_p \subset \mathbb{R}^2$ the context of a neuron at p. The neuron accumulates the excitations and inhibitions from other neurons within its context as

$$\boldsymbol{T}(p) = \boldsymbol{T}(p) + \sum_{q \in \mathcal{N}_p} (\lambda_{1,q} - \lambda_{2,q}) \mathbf{F}_{S,q}(q - p), \qquad (6)$$

where $\mathbf{F}_{S,q}$ is the contextual influences from the neuron at q to the one at p. Afterward, individual neurons refine their responses to the global input.

The response of each neuron after this synergy is still represented by a tensor. Decompose it into the form of (4), from the geometric meaning of the first term, we know that a point that belongs to an edge or a line is locally maximal along the normal $\boldsymbol{e}_1$ of the tensor, i.e., $\mathrm{d}(\lambda_1 - \lambda_2)/\mathrm{d}\boldsymbol{e}_1 = 0$.

Finally, edges and lines can be simultaneously and automatically extracted by finding the zero-crossings in the directional derivatives, without using any thresholds.

3. EXPERIMENTS

The proposed method was applied to X-ray and natural images, and compared with Canny [1], compass [6] and SUSAN [7] operators. Ideally, an algorithm is expected to accurately detect the edges and lines, without false detection, and without the need to manually tune parameters for each image or group of images. To illustrate how far our method goes in this direction, our method's parameters were fixed in all the experiments, while parameters of the other methods were determined by their algorithms or set to their suggested values.

As shown in Fig. 1(c), all the edges and the line-like features were detected using our method, which as a whole depicts the anatomical structure correctly with much less false detections. For another low quality X-ray image, compare to the "ground truth" (GT) drawn by an orthopedics surgeon (the first row, Fig. 2), our method also detected more detail anatomical structures than the others. For the SUSAN detector, we had to set the brightness threshold to 5 to obtain a good result (it detected nothing when the default value was used).

We also show the detection results for several natural images associated with GT in the BSDB [21] in Fig. 2. Though provided for segmentation purposes, they illustrate to a large extent what features are perceived by humans. Our algorithm detects edges and lines closest to the GT as compared against other algorithms, and produces the least false detection (e.g. wrinkles of the old man on the last row).

The quantitative evaluation performed using F-score [22] is shown in Table 1. It can be seen that our method obtained the highest score for all the images in Fig. 2.

Table 1. Quantitative evaluation of different methods.

image	Canny	compass	SUSAN	proposed
1	0.461	0.463	0.158	0.761
2	0.426	0.447	0.508	0.592
3	0.392	0.430	0.375	0.457
4	0.402	0.493	0.325	0.612
5	0.521	0.510	0.408	0.545
6	0.380	0.494	0.405	0.615

We performed such comparison using a large number of X-ray and nature images. In most cases, our method outperforms the other three methods and produced much fewer false detections, especially for X-ray images.

4. CONCLUSION

We have proposed a novel method for detecting edges and lines in an image simultaneously and automatically by mimicing the synergetic responses of V1 neurons to them, inspired by biological and psychological findings. Applied to different types of images, it was shown that this method, which models the nature of the computation according to the mechanisms of the human visual system, produces good detection of edges and lines as it better matches the human perceptions.

5. REFERENCES

[1] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. PAMI*, vol. 8, no. 6, pp. 112–131, 1986.

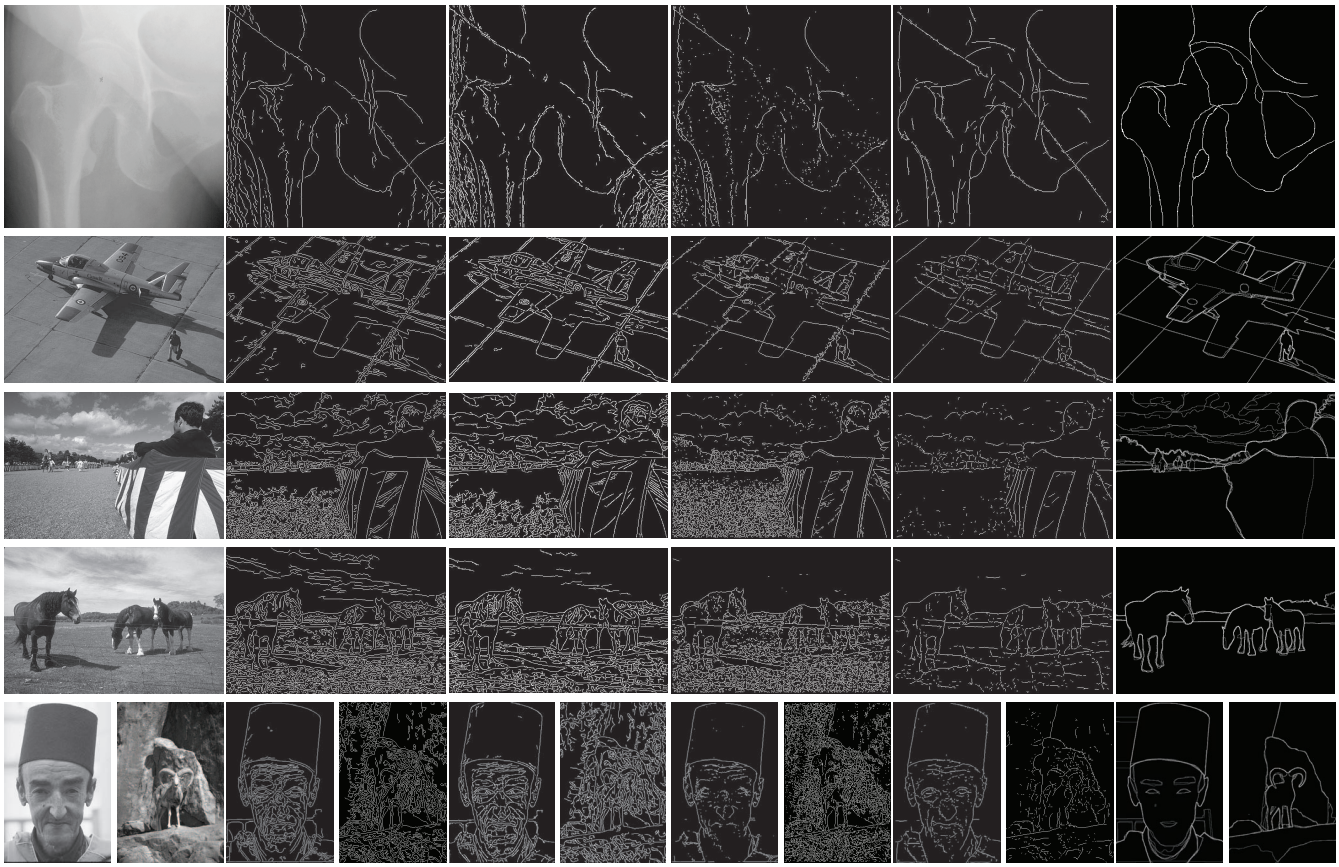
[2] Z. Li, "Theoretical understanding of the early visual processes by data compression and data selection," *Network: Computation in Neural Systems*, vol. 17, no. 4, pp. 301–334, 2006.

[3] W. Li, V. Piëch, and C.D. Gilbert, "Contour saliency in primary visual cortex," *Neuron*, vol. 50, no. 6, pp. 951–962, 2006.

[4] A. Desolneux, L. Moisan, and J.-M. Morel, "Edge detection by Helmholtz principle," *J. Math. Imaging Vis.*, vol. 14, no. 3, pp. 271–284, 2001.

[5] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "On straight line segment detection," *J. Math. Imaging Vis.*, vol. 32, no. 3, pp. 313–347, 2008.

[6] M. A. Ruzon and C. Tomasi, "Edge, junction, and corner detection using color distribution," *IEEE Trans. PAMI*, vol. 23, no. 11, pp. 1281–1295, 2001.



(a) image (b) Canny (c) compass (d) SUSAN (e) proposed (f) ground truth

Fig. 2. Experimental results on natural images using various algorithms.

[7] S.M. Smith and J.M. Brady, "SUSAN – a new approach to low-level image processing," *Int. J. Computer Vision*, vol. 23, no. 1, pp. 45–78, 1997.

[8] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction, and functional architecture in the cat"s visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.

[9] M. K. Kapadia, M. Ito, C. D. Gilbert, and G. Westheimer, "Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in v1 of alert monkeys," *Neuron*, vol. 15, pp. 843–856, 1995.

[10] John G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, no. 10, pp. 847 – 856, 1980.

[11] L. M. Martinez and J. M. Alonso, "Construction of complex receptive fields in cat primary visual cortex," *Neuron*, vol. 32, no. 3, pp. 515–525, 2001.

[12] M. C. Morrone and R. A. Owens, "Feature detection from local energy," *Pattern Recogn. Lett.*, vol. 6, no. 5, pp. 303–313, 1987.

[13] M. C. Morrone, J. Ross, D. C. Burr, and Robyn Owens, "Mach bands are phase dependent," *Nature*, vol. 324, no. 6094, pp. 250–253, Nov. 1986.

[14] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE*, vol. 69, no. 5, pp. 529–541, 1981.

[15] Ferenc Mechler, Daniel S. Reich, and Jonathan D. Victor, "Detection and Discrimination of Relative Spatial Phase by V1 Neurons," *J. Neurosci.*, vol. 22, no. 14, pp. 6129–6157, 2002.

[16] Zhaoping Li and J. J. Atick, "Efficient stereo coding in the multiscale representation," *Network: Computation in Neural Systems*, vol. 5, no. 2, pp. 157–174, 1994.

[17] P. D. Kovesi, "Image features from phase congruency," *Videre: J. Comput. Vis. Res.*, vol. 1, no. 2, pp. 1–26, 1999.

[18] J. M. Samonds, J. D. Allison, H. A. Brown, and A. B. Bonds, "Cooperation between area 17 neuron pairs enhances fine discrimination of orientation," *J. Neurosci.*, vol. 23, no. 6, pp. 2416–2425, 2003.

[19] Z. Li, "Pre-attentive segmentation in the primary visual cortex," *Spatial Vision*, vol. 13, no. 1, pp. 25–50, 2000.

[20] G. Medioni, M.-S. Lee, and C.-K. Tang, *A Computational Framework for Segmentation and Grouping*, Elsevier, Amsterdam, 2000.

[21] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. PAMI*, vol. 26, pp. 530–549, 2004.

[22] C. J. van Rijsbergen, *Information Retrieval*, Butterworth, 1979.