RELATIONSHIP BETWEEN SUSPICIOUS COINCIDENCE IN NATURAL IMAGES AND CONTOUR-SALIENCE IN ORIENTED FILTER RESPONSES

A Thesis

by

SUBRAMONIA P. SARMA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2003

Major Subject: Computer Science

RELATIONSHIP BETWEEN SUSPICIOUS COINCIDENCE IN NATURAL IMAGES AND CONTOUR-SALIENCE IN ORIENTED FILTER RESPONSES

A Thesis

by

SUBRAMONIA P. SARMA

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Yoonsuck Choe
(Chair of Committee)

Thomas R. Ioerger
(Member)

Reza Langari
(Member)

Valerie E. Taylor
(Head of Department)

December 2003

Major Subject: Computer Science

ABSTRACT

Relationship between Suspicious Coincidence in Natural Images and
Contour-Salience in Oriented Filter Responses. (December 2003)
Subramonia P Sarma, B.Tech, University of Kerala, Trivandrum, India;
Chair of Advisory Committee: Dr.Yoonsuck Choe

Salient contour detection is an important low-level visual process in the human visual system, and has significance towards understanding higher visual and cognitive processes. Salience detection can be investigated by examining the visual cortical response to visual input. Visual response activity in the early stages of visual processing can be approximated by a sequence of convolutions of the input scene with the difference-of-Gaussian (DoG) and the oriented Gabor filters. The filtered responses are unusually high for prominent edge locations in the image, and is uniformly similar across different natural image inputs. Furthermore, such a response follows a power law distribution. The aim of this thesis is to examine how these response properties could be utilized to the problem of salience detection. First, I identify a method to find the best threshold on the response activity (orientation energy) toward the detection of salient contours: compare the response distribution to a Gaussian distribution of equal variance. Second, I justify this comparison by providing an explanation under the framework of *Suspicious Coincidence* proposed by Barlow [1]. A connection is provided between perceived salience of contours and the neuronal goal of detecting suspiciousness, where salient contours are seen as affording suspicious coincidences by the visual system. Finally, the neural plausibility of such a salience detection mechanism is investigated, and the representational efficiency is shown which could potentially explain why the human visual system can effortlessly detect salience.

To my parents Padmanabhan and Vijayalakshmy

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Yoonsuck Choe, for his constant encouragement and guidance right from the conceptual stage to the completion of this thesis and for the support I received throughout my research. His method of guiding by providing examples and attention to detail have been really motivating for my research. I have learned a lot about research work from him through several interesting discussions.

I would also like to express my gratitude to my committee members, Dr. Tom Ioerger and Dr. Reza Langari, for their valuable and insightful comments on the draft and during the presentation of my ideas.

I would like to acknowledge the contributions of my fellow research group members, S. Kumar Bhamidipati and Yingwei Yu, who provided useful feedback. Thanks are also due to a former research group member, Dr. Hyeon-Cheol Lee, for contributing to some basic parts of the thesis.

Finally, I would like to thank my beloved family for the constant encouragement and belief in me throughout my graduate studies.

TABLE OF CONTENTS

CHAPTER Page

I INTRODUCTION 1

- A. A Visual System Primer 2
- B. Motivation 4
- C. The Problem 5
- D. Approach 5
- E. Outline of the Thesis 6

II BACKGROUND 8

III APPROACH 12

- A. Inputs 12
- B. Orientation Energy Calculation 13
- C. Orientation Energy Distributions (OED) 18
- D. Non-Gaussianity of OED 19

IV EXPERIMENTS AND RESULTS 23

- A. Experiment 1: Comparison with Psychophysical Data . . . 23
  - 1. Methods 23
  - 2. Results 24
  - 3. Discussion 25
- B. Experiment 2: Application to OE Thresholding 27
  - 1. Methods 27
  - 2. Results 29
  - 3. Discussion 32

V ANALYSIS 34

- A. Analysis of the Gaussian Baseline 34
- B. Quantitative Analysis with Synthetic Images 38
  - 1. Synthetic Images and their OEDs 38
  - 2. Variation in Noise and Performance 40
  - 3. Variation in Input Count and Performance 43
  - 4. Summary 45

VI DISCUSSION AND FUTURE WORK 48

CHAPTER Page

A. Potential Neural Basis of Salience Detection 48
B. Computational Efficiency of Orientation Energy Responses 49
C. Extensions to Other Modalities 51
D. Local vs. Global Thresholding 53
E. Comparison of Local Thresholding with Psychophysical Data 54

VII CONCLUSION 55

REFERENCES 56

VITA 61

LIST OF TABLES

TABLE Page

I SS Error means when noise is varied 44
II SS Error Means when input count is varied 47

LIST OF FIGURES

FIGURE Page

1 An illustration of the main visual pathway in primates.(Adapted from [16, 18].) 2

2 Some typical receptive fields of the neurons in the early visual pathway. Positive signs denote excitation and negative signs denote inhibition. (a) the RFs of retinal ganglion cells and LGN cells show center-surround property. (b) The RFs of V1 neurons show orientation selectivity. (Adapted from [16].) 3

3 Some representative natural images depicting a variety of natural terrain obtained from the Kodak website. 14

4 An image before and after windowing with a circular aperture. 15

5 Comparison of gray-scale intensity histograms of three natural images with their orientation energy histograms. (a) to (c) show the gray-scale histograms, and (d) to (f) show the corresponding orientation energy histograms which model V1 response. It can be seen that the gray-scale histograms are quite different, while the orientation energy histograms are remarkably similar. 16

6 The OED derived from the E histograms for six natural images are shown in log-log plot, from a to f. The same images as in Fig. 4 were used. For easier comparison, the curves have been scaled by a factor of 10. We can see that all the curves are mostly straight with a similar slope, which indicates a power law. It may also be noted that the high energy area toward the right of the curves has a lot of noise and empty bins, probably due to the scarcity of samples at high orientation energy. 18

FIGURE Page

7 The figure shows the comparison of a distribution $h(E)$ that follows the power law (solid curve) with a normal distribution $g(E)$ with the same variance (dashed curve), in both the linear and log-log scales. (a) and (b) illustrate a power law distribution, and (c) and (d) show the actual distributions obtained from a natural image. The x-axis represents the orientation energy and the y-axis the probability. It is of interest to look for orientation energy values that have high probabilities, i.e., where $h(E)$ is greater than $g(E)$. Only positive values for E are considered. 20

8 Comparison of OED $h(E)$ of six natural images with their corresponding normal distributions $g(E)$ of the same variance. The second intersection ($L2$) of the two curves are marked by a vertical line in each plot. We can see that beyond this point, $g(E)$ plummets, while $h(E)$ remains steady, thus anything beyond $L2$ may be seen as salient. 21

9 A sample natural image and fixed-percentile thresholded orientation energy images that were shown to a human observer: (a) shows the natural image, (b) - (f) show the results from 55-,65-,75-, 85-, and 95-percentile thresholding. 24

10 Orientation Energy threshold selected by humans vs. L2. The manually chosen thresholds are compared to the $L2$ values for each image. Each point in the plot corresponds to one of 31 natural images used in the calculation. The straight line in the figure shows a linear fit to the data. 25

11 $L2$ vs. σ of the OED. The $L2$ values obtained by comparing $h(E)$ and $g(E)$ are compared to the square root of the raw second moment (σ) of $g(E)$. Each point in the plot corresponds to one of 31 natural images. The straight line in the figure shows a linear fit to the data. 26

12 Orientation Energy thresholds selected by humans vs. σ. The orientation energy thresholds manually selected are compared to the square root of the raw second moment (σ) of $g(E)$. Each point in the plot corresponds to one of 31 natural images. The straight line in the figure shows a linear fit to the data. 26

FIGURE Page

13 Global Thresholding Results. (a) The original image. (b) The threshold result selected by a human. (c) The OED-derived thresholding result where pixels are replaced by oriented lines indicating the local orientation. The result of thresholding is close to our perceived salience of edges in (a). 30

14 Comparison of global OED-derived thresholding with fixed-percentile thresholding. (a) The original image. (b) The threshold result selected by a human. (c) The threshold at the 85-percentile of $g(E)$. (d) The OED-derived thresholding result where pixels are replaced by oriented lines indicating the local orientation. 31

15 Comparison of global thresholding with local thresholding. (a) The original image. (b) The threshold selected by a human which is globally applied. (c) The threshold at the 85-percentile of $g(E)$, also globally applied. (d) The OED-derived global thresholding result where pixels are replaced by oriented lines indicating the local orientation. (e) The result of local OED-derived thresholding using a sliding window of size 21 × 21. 32

16 A white-noise image and its orientation energy distribution. (a) The original image. (b) The orientation energy E. No clear structure can be discerned. 36

17 The OED of a white-noise image (solid)and its matching normal distribution (dotted). The log-log plot shows that the two distributions are quite similar. 36

18 Comparison of white-noise based $L2$ against human-chosen thresholds and σ_h. (a) The new $L2$ values derived from the white-noise based distribution are plotted against the human-chosen thresholds for a set of 42 natural images. The correlation coefficient was 0.98, indicating a strong linearity. (b) The new $L2$ values are compared against the raw standard deviation σ_h of $h(E)$. The correlation coefficient was 0.91, and a strong linearity can be observed between the two sets of values. 37

FIGURE Page

19 A synthetic image consisting of a set of overlapping squares with embedded noise and its corresponding orientation energy distribution shown against the normal distribution of the same variance. We can see that the synthetic image has a distribution almost similar to that of a natural image. 39

20 (a) A synthetic image consisting of 300 overlapping squares with embedded noise. (b) The orientation energy matrix for (a). (c) The orientation energy matrix for the synthetic image without the noise that is used as the reference. Results of thresholding by the four different methods of (d) global OED, (e) global 85-percentile, (f) local OED, and (g) local 85-percentile respectively are shown from the left to the right. The global OED-derived adaptive thresholding method seems to offer the best performance for this input. 41

21 Bar Plots showing the average SS Error values for the thresholding results for a sample noise level for the synthetic images. The global and local OED-derived method have significantly smaller SS Error values than the fixed 85-percentile methods. 43

22 (a) A synthetic image consisting of 500 overlapping squares with embedded noise. (b) The orientation energy matrix for (a). (c) The orientation energy matrix for the synthetic image without the noise that is used as the reference. Results of thresholding by the four different methods of (d) global OED, (e) global 85-percentile, (f) local OED, and (g) local-85 percentile respectively are shown from the left to the right. The local OED-derived adaptive thresholding method seems to offer the best performance for this kind of input. 45

23 Bar plots showing the average SS Error values for the thresholding results for a sample noise level for the synthetic images. The global and local OED-derived methods have significantly smaller SS Error values than the fixed 85 percentile methods. 46

24 Flowchart showing the calculation of sigma by a simple neural mechanism. 49

FIGURE Page

25 Graphs showing the learning efficiency for two training instances with different inputs, with both the plain orientation energy input and its squared form. The graphs clearly show that the efficiency is higher for the squared orientation energy input to learn the threshold value. 52

CHAPTER I

INTRODUCTION

Vision is one of the most important sensory modalities in organisms because it plays a crucial part in helping organisms adapt to and thrive in their environments. Although we often take it for granted, vision is actually a complex process. Understanding vision and explaining it through the use of suitable models is thus an important and worthwhile endeavor. Much of the early successes in vision research have been in low-level vision, where the physical properties of various processing elements in the visual system were clearly identified and explained [5, 18, 22, 23, 28]. These properties have been found to be quite useful in explaining various processes in the early visual system, such as edge detection and contour grouping [13]. Theories based on such low-level processes have in turn formed the foundation for understanding in high-level vision, such as object recognition and perception.

Much of the early visual processes in the human visual system have evolved over time to cope with the natural visual environment. Also, the natural environment is not random but structured and orderly [33, 17]. Thus it can be assumed that the structure and functionality of the human visual system are largely influenced by the statistical properties of the natural scene input it continuously receives from the environment [13]. Since not all input present in the visual scene has immediate relevance to the organism, the visual system could be assumed to concentrate on those areas in the input which are most significant, or in other words, *salient*. One important salient property in visual inputs is edges (or contours), where local contrast abruptly changes along a long stretch of visual space. In this thesis, I will thus focus

---

The journal model is *IEEE Transactions on Neural Networks.*

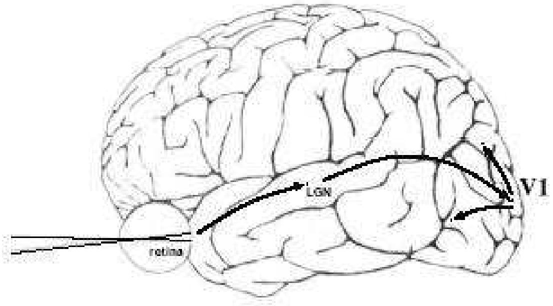


Fig. 1. An illustration of the main visual pathway in primates.(Adapted from [16, 18].)

on understanding how such salient edge features are processed and derived by the visual system. Studying the biological visual system can provide a clue towards this objective.

A. A Visual System Primer

A lot of knowledge has been gained about the primate visual system from various psychophysical and electro-physiological experiments and more recently from advances in brain imaging methods. It has also been widely known that the general structure is quite similar for the human visual system and the primate visual system. In this section, I will attempt to provide a little insight onto the general structure of the visual system and the properties of its neurons. This section is largely based on [16].

Fig. 1 shows an illustration of the early stages of the main visual pathway in primates (adapted from [16, 18]). The light collected by the retinal photo receptors is transmitted by the optic nerve to the retinal ganglion cells. From there the information is sent to the lateral geniculate nucleus (LGN) in the thalamus and is further sent to the primary visual cortex (V1) located at the back of the brain. The V1 is believed to be one of the first locations where the visual information is processed by the cerebral cortex. Information after it is processed in V1 is then sent to other

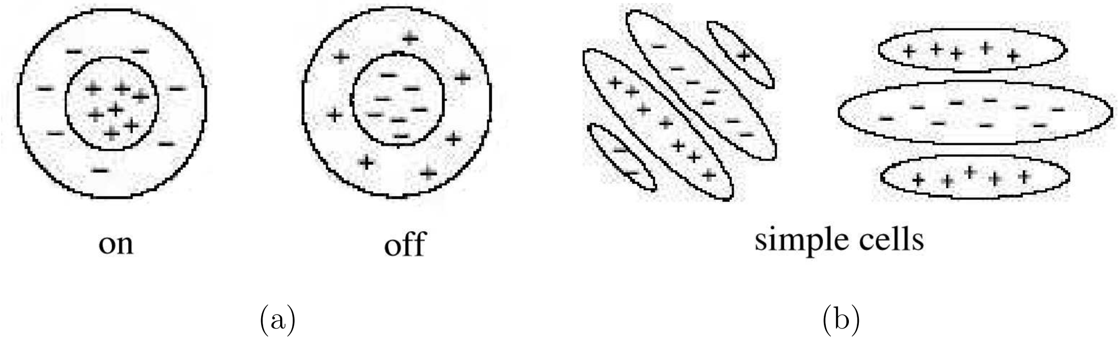


Fig. 2. Some typical receptive fields of the neurons in the early visual pathway. Positive signs denote excitation and negative signs denote inhibition. (a) the RFs of retinal ganglion cells and LGN cells show center-surround property. (b) The RFs of V1 neurons show orientation selectivity. (Adapted from [16].)

locations in extra-striate cortex through several other pathways.

Neurons represent the primary information processing unit in the brain. Neurons communicate within themselves via spikes or action potentials. The response of a typical neuron in the early visual pathways depends on the pattern of input of a small area of the visual field, called the receptive field (RF). Thus changes in the input stimulus in the receptive field will lead to changes in the firing of the corresponding neuron. The receptive fields in different areas of the visual system are known to exhibit different properties. For example, the receptive fields at the retinal ganglion cells and the LGN show a center-surround property, whereby they provide excitatory output by light in a small central circular region, and inhibitory output by light in a surrounding circular region [5, 23]. The opposite effect of inhibition in the center and excitation at the periphery is also shown by other cells. Further downstream in the primary visual cortex (V1), the receptive fields exhibit orientation, phase and frequency tuned properties [22, 28]. This is depicted in Fig. 2 (adapted from [16]).

It is also known that neurons in the visual cortex (as in other cortical areas) show *graded* response to specific stimuli. Also, nearby locations in the visual field are found to be mapped to nearby neurons in the visual cortex. A consequence of the finding about the receptive fields depicted in Fig. 2 is that the early visual processing can be modeled as a sequence of filter convolutions. The center-surround receptive fields can be modeled as the difference of two Gaussian kernels, a classic model of which is given by the Difference-of-Gaussian (DoG) filter. The orientation selective receptive fields can be modeled by Gabor filters which are products of sinusoidal gratings and Gaussian envelopes. Although such a kind of model is quite simplistic, it has been found to be quite effective as a model for preprocessing of visual input to study visual responses, as in [13]. Such a model is also used in this thesis, and will be discussed in detail in Chapter III.

B. Motivation

The response of neurons in the visual system, which can be modeled using the DoG and Gabor filters, have the property that they are quite similar across different natural image input. It is also known that the response distribution of neurons in the primary visual cortex (V1) shows a power law. Thus, it has a heavy tail, or in other words, extreme values are not uncommon. We can then speculate that such a response property is exhibited by the visual system because it is useful in some way. It then becomes a motivating problem to investigate if such properties could be used toward salient contour detection. Studying salience detection can also provide a clue toward understanding higher level visual and cognitive processes such as object recognition and perception.

C. The Problem

The pertinent problem addressed in this thesis can then be stated as follows:

1. Identifying an effective method for detecting salience of contours in natural images, by utilizing the filter response properties.
2. Justification of the use of the method.
3. Investigation of the neural basis and representational efficiency of the method.

D. Approach

In this thesis I propose to answer the first part of the problem by utilizing the properties the filter response, called the orientation energy. I will conduct experiments to see how effectively such properties could be utilized. Since the filter responses are fairly uniform and are usually high for locations in an image where there are prominent edge elements, thresholding the responses can lead to the detection of salient contours. I show that comparing the response distribution from a natural image to a normal distribution with the same variance gives a good thresholding criterion for detecting salient levels of edginess, through comparison with human-chosen thresholds. To precisely measure the effectiveness of this method, I compare the performance of this method on synthetic images having similar response properties as natural images, with human performance.

More significantly, for the second part of the problem, I attempt to interpret the salience detected using orientation energy thresholding under the concept of *Suspicious Coincidence* proposed by Barlow [1]. The central idea is that an image where each pixel is independent from each other (such as a white-noise image) could be defined as having no suspicious feature in it, which is exactly what humans perceive.

Thus salience can be understood as a deviation from the unsuspicious baseline of a Gaussian distribution. If this has to be the case, the orientation energy distribution for a white-noise image should be at least near-Gaussian, and this turns out to be true. Thus the white-noise experiment provides justification for the use of the computationally simpler Gaussian distribution as the baseline for thresholding.

Finally, I suggest a neural basis for the salience detection method by showing that the appropriate threshold can be easily extracted using a simple neural mechanism that utilizes a weighted sum of the squared orientation energy response. I further test the representational effectiveness of squared responses by evaluating the efficiency of a backpropagation network that learns the orientation energy threshold. Such efficiency may possibly be one of the important considerations if it is used in the visual system.

E. Outline of the Thesis

This thesis is organized as follows. Chapter II will provide an overview of related research in this area. I will show how other researchers utilize natural image statistics toward a number of tasks such as denoising and compression, and why my approach differs from theirs. Chapter III details the approach I have used for my experiments, and provides details about the input preparation, calculation of orientation energy and its distribution, and how its properties can be utilized. I examine the relationship between high response levels and perceived salience by humans in Chapter IV, through comparison to human performance. I then derive a thresholding criterion to find salient levels of response by comparison with a Gaussian distribution. Chapter V examines why this comparison is reasonable, and an explanation is given under the framework of Suspicious Coincidence. I also conduct a quantitative analysis to precisely measure the effectiveness of such a thresholding approach. The thesis concludes with a discussion of the neural plausibility of my approach, with experimental results showing its representational efficiency. Some suggestions for future work are also presented.

CHAPTER II

BACKGROUND

There has been extensive research on natural image statistics and their utility for various practical image processing problems. Natural images are structured and orderly and provide unique statistical properties that were found to be useful to develop efficient methods for tasks such as edge detection, segmentation, denoising, and compression. In another research direction, there has been attempts to model visual system processes by studying neuronal responses to visual input. However, a major research path forms a link between these two by studying how natural image statistics influence the development of visual system processes. The basic idea behind such an approach is that the human visual system has evolved over time to adapt to the natural scene input from their environment, and thus has been influenced by natural scene statistics in its development. This research direction is the motivation behind my thesis.

Ruderman [30] investigated the robust scale invariance property of natural images and proposed that this property could be explained by the presence of regions corresponding to statistically independent objects that showed a power-law distribution of sizes. Research done by Zhu, Wu and Mumford [34] showed that exponential models could provide a general framework for natural image modeling. Other researchers such as Huang and Mumford [17] systematically investigated various statistical properties of natural images using a very large calibrated image database and fitted mathematical models to some of the properties such as scale invariance.

The use of natural image statistics to image compression was researched by Buccigrossi and Simoncelli [4]. They developed a statistical model of natural images in the wavelet transform domain that described joint statistics between pairs wavelet coefficients. Simoncelli and Adelson [32] independently in their research on Bayesian wavelet coring used non-gaussian properties to develop suitable thresholding methods for denoising in natural images. They also pointed out that such models could be useful for other kind of tasks such as segmentation and contour identification. In a related work, the thresholding of wavelet responses was researched by Hansen and Yu [14] toward denoising and compression of images.

A recent research direction has been the use of response histograms. For example, spectral histograms of image were used by Liu and Wang [27] for texture discrimination and segmentation. They showed that the spectral histogram model avoided problems such as rectification and spatial pooling which are commonly assumed stages in texture discrimination models. They also showed that by matching spectral histograms, an arbitrary image could be transformed into another image with similar features via statistical sampling. Such use of histograms was found to help in the proper segmentation and synthesizing of textures.

Another direction for research has been to examine the organization and properties of various elements in the visual system to understand various lower-level vision tasks. A detailed study of the receptive fields and maps in the mammalian visual cortex with special regard to the properties of ocular dominance and orientation selectivity was done by Miller [29]. Various models were reviewed that explained the structure of the receptive fields and cortical maps.

A model for early visual processing in primates was proposed by Itti et al. [20] that consisted of a population of linear spatial filters and their interactions. Human psychophysical thresholds were then derived from the population responses. Itti et al. used such a model to predict human thresholds for orientation and contrast discrimination tasks. The detection of salient objects in natural scenes was also researched by Itti and Koch [21] where they studied the importance of selective visual attention to form saliency maps of visual scenes.

There has been considerable research into the study of natural image statistics to understand various visual system processes. Barlow [2] suggested that a role of early sensory neurons is to remove statistical redundancy in the sensory input they receive, which in the early stages of evolution and development would be the natural scene input.

The statistics of natural images toward efficient coding of neural responses was reviewed in detail by Simoncelli and Olshausen [33]. The statistical redundancies present in natural images were considered for various features such as intensity, color and spatial correlations. The non-Gaussian nature of neural responses to natural scene input also was of interest toward understanding efficient coding principles employed by the visual system neurons.

Other researchers such as Geisler et al. [13] were more interested in applying natural image statistics to study performance in contour grouping. They proposed a quantitative method for analyzing grouping performance for natural images, using both absolute and Bayesian edge co-occurrence statistics. This is a unique approach due to its quantitative nature of analysis since most of the research in the area of contour grouping used qualitative analysis methods. Geisler et al. were also able to successfully derive a contour grouping rule from the edge co-occurrence statistics.

Contour integration in low-level vision was also researched by Choe and Miikkulainen [6] and they proposed a model that suggested that lateral interactions between neurons with similar orientation tuning can be learned through input-driven self-organization.

The relation between natural image statistics and visual cortical cell responses was also researched by Field [11]. The representation of images by the visual system in mammals was of interest to Field and various coding schemes were compared and analyzed. One of his findings was that the orientation and spatial-frequency tuning of mammalian simple cells were well-suited for coding the information from natural images. He also proposed that a goal of such coding schemes was to convert higher-order redundancy to first-order redundancy and also ensure a high signal-to-noise ratio. His results thus helped support Barlow's view that the goal of vision was to represent information with minimal redundancy.

More recently, the temporal statistical features of natural video scenes have been studied by some researchers [19] to find the correlations between the temporal responses of complex cells in the visual cortex. The recent advances in statistical and computational modeling have really helped to increase the interest in this area.

My approach in this thesis has been essentially motivated by some of the basic ideas from the research work presented above. Since the primary problem of interest to me is understanding salient contour detection, it was natural to consider the statistics of natural images. It could be believed that the visual system is quite good in detecting salient contours and natural image inputs are of interest since such a capability could be believed to have an evolutionary and developmental background. The non-Gaussian nature of orientation energy (or wavelet response) histograms has been utilized previously, but in different contexts such as denoising and compression. Notable here are the work done by Simoncelli and Adelson [32] for denoising, and by Hansen and Yu [14] for denoising and compression. It was also noted by Barlow [3] that comparing peaked distributions with high kurtosis with distributions derived from an unsuspicious baseline would be quite useful. However, to my knowledge, my approach is the first systematic study of the relationship between perceived salience in humans and the orientation energy distribution under the framework of suspicious coincidence.

CHAPTER III

APPROACH

The early visual processes can be suitably modeled by the responses of filters that mimic the receptive field properties of neurons in the visual system. Such a response, called the orientation energy (OE), can be interpreted as visual cortical activity in response to visual input. The oriented nature of filters that model the orientation-selectivity property of the visual system suggests that the filtered response would be high for locations in an image where there are strong edge elements.

The initial step is then to generate the orientation energy responses for the natural image input. In this chapter, I describe the inputs used and how the orientation energy distribution is calculated. I also show an important property of the orientation energy distribution of natural images,i.e., non-Gaussianity, and consider how it may be useful. The results presented in this chapter are largely based on [25], which describes the research project I took part in.

A. Inputs

The early visual system processes could be believed to have evolved over time by exposure to the natural environment, and thus their development could have been influenced by the natural scene statistics. Thus one of the primary requirements for the inputs for my experiments was that they capture aspects of the natural environment as much as possible, without the presence of artificial (man-made) structures or objects. Thus the images used would have to be from the natural terrain containing mostly terrestrial stimuli, such as images of woods, mountains, rivers, birds, and animals. The natural images selected also needed to be from a wide range of terrain, which would help support the generality of any conclusion that followed out of the experiment. This meant that there were not only natural images containing high-contrast features such as densely populated woods, but also those with low-contrast background such as clouds, river water, etc.

For all the experiments described in this thesis, I utilized a collection of digital stock images obtained from the Kodak website[1], the same source as in [13]. Some sample images are shown in Fig. 3. The images were from a wide variety of natural terrain, and were all obtained in the JPEG image format. All the images were in 24 bit color and had dimensions of 256×256 pixels.

The images were first windowed with a circular aperture of diameter 256 pixels to prevent the addition of artefactual orientation bias due to the horizontal and vertical border regions. An example image and its circular aperture are shown in Fig. 4.

To model visual cortical response, we have to first look at certain properties of the natural scene input. Unlike gray-scale intensity histograms which could be significantly different from each other, it was found that the orientation energy histograms of natural images are quite similar to each other. This is shown in Fig. 5 where the gray-scale and orientation energy histograms of three representative natural images are shown.

B. Orientation Energy Calculation

The calculation of orientation energy is based on that in [25], for which I was a contributor. According to research by Geisler, et al. [13], the orientation energy can be measured as the response obtained by convolving the natural image input with a combination of oriented and non-oriented filters similar to those found in the receptive fields of the primary visual cortex (V1). To calculate the orientation energy response

---

[1] The URL is http://www.kodak.com/digitalImaging/samples/imageIntro.shtml

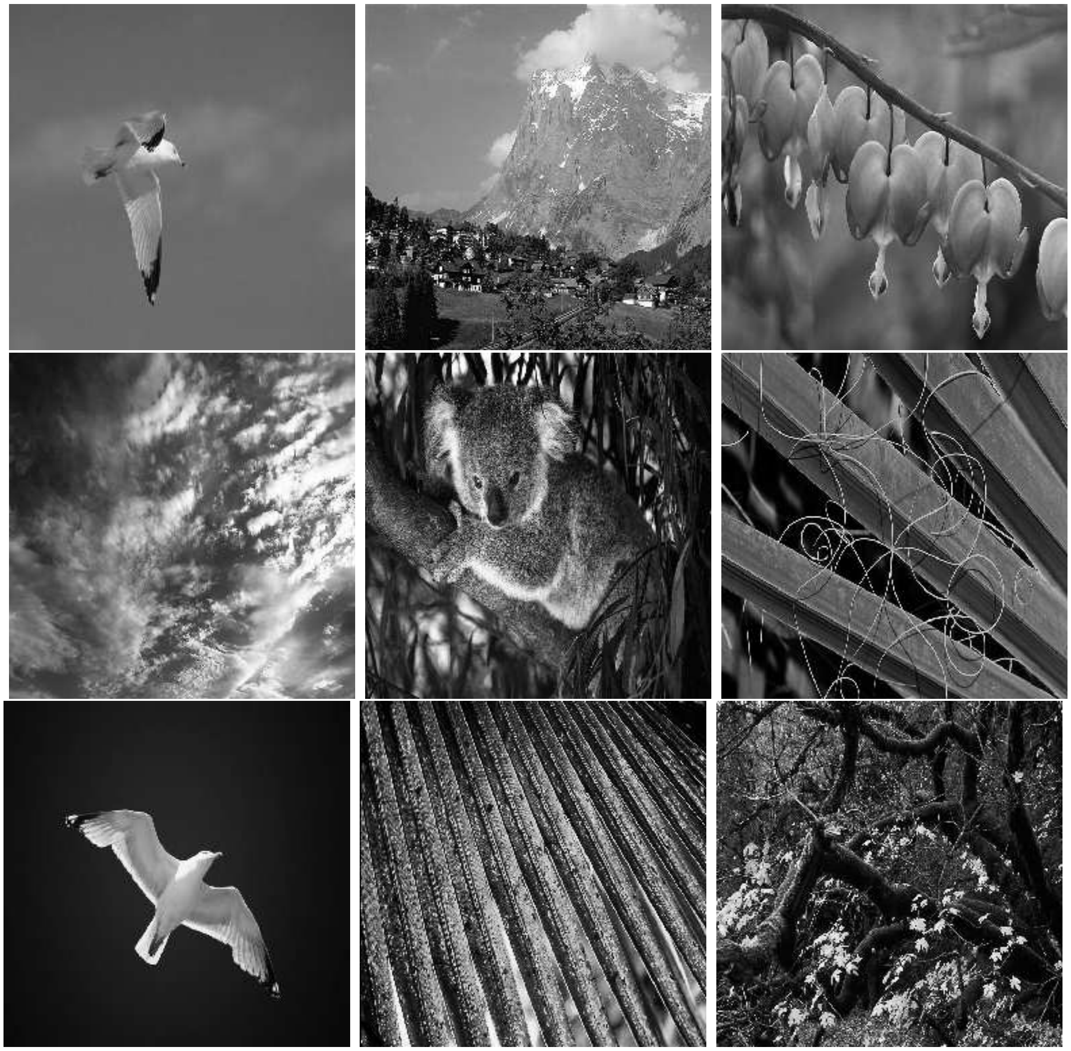


Fig. 3. Some representative natural images depicting a variety of natural terrain obtained from the Kodak website.

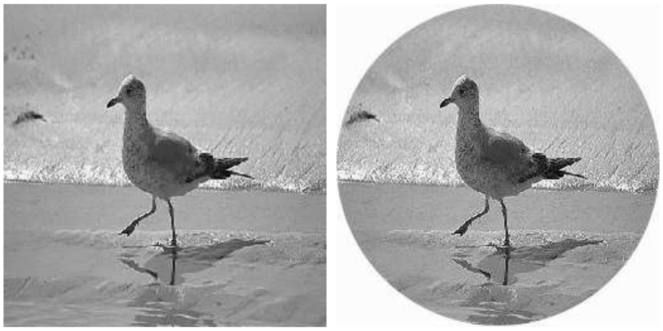


Fig. 4. An image before and after windowing with a circular aperture.

and its distribution, a procedure similar to [13] is used. The procedure involves a sequence of convolutions, first with difference-of-Gaussian (DoG) filters and then with oriented Gabor filters, to give the orientation filter response. The DoG filter is essentially composed of two Gaussian functions whose widths differ by a factor of 0.5, as

$$F(x,y) = G_{(\sigma/2)^2}(x,y) - G_{\sigma^2}(x,y), \tag{3.1}$$

where $G_{\sigma^2}(\cdot)$ is a Gaussian function with variance σ^2, defined as follows:

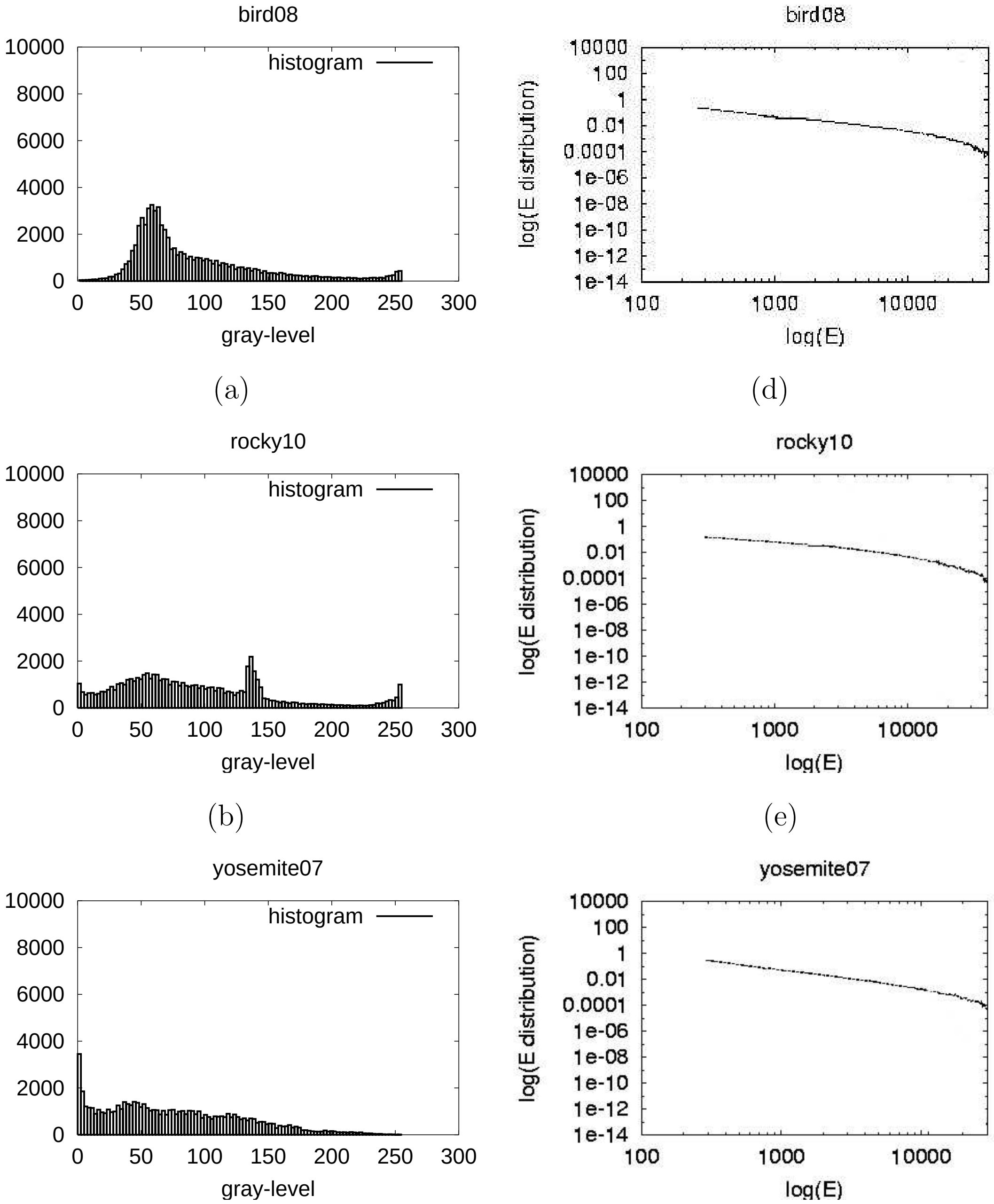
$$G_{\sigma^2}(x,y) = \frac{1}{2\pi\sigma^2}.\exp^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{3.2}$$

where the pixel location is denoted by (x,y).

Other kinds of filter models could be used, notably the Laplacian of Gaussian filter (LoG), instead of the DoG model. However, I used of the DoG filter for simplicity.

For the initial convolution, the gray-level intensity matrix I was obtained from the input image. It was then convolved with the DoG filter to obtain the intermediate matrix I_f, as

$$I_f = I * F, \tag{3.3}$$



(a) (d)

(b) (e)

(c) (f)

Fig. 5. Comparison of gray-scale intensity histograms of three natural images with their orientation energy histograms. (a) to (c) show the gray-scale histograms, and (d) to (f) show the corresponding orientation energy histograms which model V1 response. It can be seen that the gray-scale histograms are quite different, while the orientation energy histograms are remarkably similar.

where the operator * indicates the convolution operation. I used convolution kernels of size 7 × 7 for all the experiments.

Next, the intermediate filtered image I_f is convolved with oriented Gabor filters $R_{\theta,\phi,\sigma}(x,y)$ [10].

$$R_{\theta,\phi,\sigma}(x,y) = \exp^{-\frac{x'^2+y'^2}{2\sigma^2}} . \cos(2\pi x' + \phi), \tag{3.4}$$

where θ is the orientation, ϕ is the phase, σ is the width, and (x,y) represents the pixel location. The Gabor filters have both even and odd phases. The other parameters of the Gabor filters such as spatial frequency and aspect ratio were set to 1 each. Again, the convolution kernels were sized 7 × 7 as above.

Convolving the intermediate filter output from the DoG convolution I_f with this filter gives the orientation energy matrix E_θ.

$$E_\theta = (R_{\theta,0,\sigma} * I_f)^2 + (R_{\theta,\frac{\phi}{2},\sigma} * I_f)^2, \tag{3.5}$$

where

$$x' = x\cos(\theta) + y\sin(\theta), y' = -x\sin(\theta) + y\cos(\theta), \tag{3.6}$$

and (x,y) denotes the pixel location as previously.

The combined orientation energy $E(x,y)$ for each pixel location (x,y) was obtained using a vector sum of six $(\theta, E_\theta(x,y))$ pairs in polar co-ordinates where θ had values of $0, \frac{\pi}{6}, \frac{2\pi}{6}, \frac{3\pi}{6}, \frac{4\pi}{6}, \frac{5\pi}{6}$. This then gives the estimated orientation $\theta^*(x,y)$ and the associated orientation energy value $E^*(x,y)$ at that location. (For simplicity I will refer to the orientation energy at any point (x,y) as just E, instead of $E(x,y)$.)

In the next section, the properties of the E distribution (also referred to as the Orientation Energy Distribution or OED) will be analyzed.

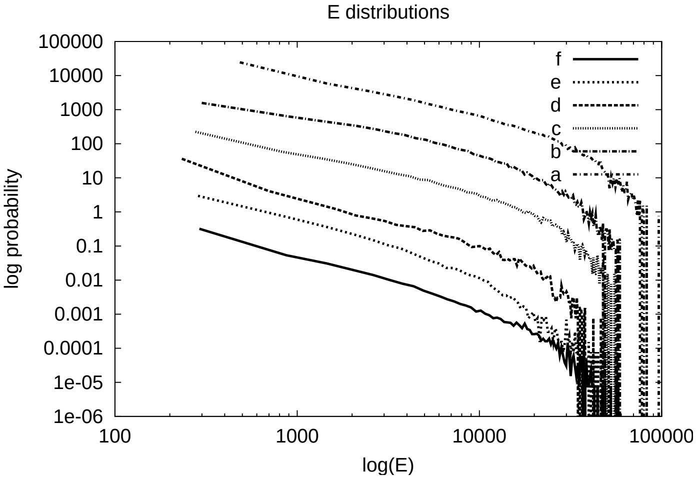


Fig. 6. The OED derived from the E histograms for six natural images are shown in log-log plot, from a to f. The same images as in Fig. 4 were used. For easier comparison, the curves have been scaled by a factor of 10. We can see that all the curves are mostly straight with a similar slope, which indicates a power law. It may also be noted that the high energy area toward the right of the curves has a lot of noise and empty bins, probably due to the scarcity of samples at high orientation energy.

C. Orientation Energy Distributions (OED)

The orientation energy distribution was estimated for several representative natural images by constructing a histogram of 100 bins from the E responses, followed by normalization,

$$h(E) = \frac{f(E)}{\sum_{x \in B_h} f(x)}, \tag{3.7}$$

where $f(E)$ is the frequency of energy level E in the histogram, B_h is the set of all histogram bin locations, and $h(E)$ is the resulting probability mass function. The orientation energy distribution for 6 sample distributions plotted in the log-log scale is shown in Fig. 6. The orientation energy distributions show a strong similarity to each other, and also share a unique feature, i.e., a power law ($p(x) = 1/x^a$, where a is the fractal exponent). When the orientation energy distribution is plotted in the log-log scale, it shows up as a straight line. An interesting property of such power law distributions is that extreme values are not uncommon. In other words, the distributions may have a *heavy tail*. For example, when a power law distribution is compared to a normal distribution with the same variance, it has greatly higher probability for extreme values. Fig. 7 illustrates the theoretical power law distribution in both linear and log scales, and the actual distribution obtained from a natural image.

D. Non-Gaussianity of OED

To detect orientation energy values with unusually high probability, I compared the OED to *discretized half-normal* distributions. Also, to make the two distributions have the same width, the *raw second moment* σ^2 of the OED was calculated as

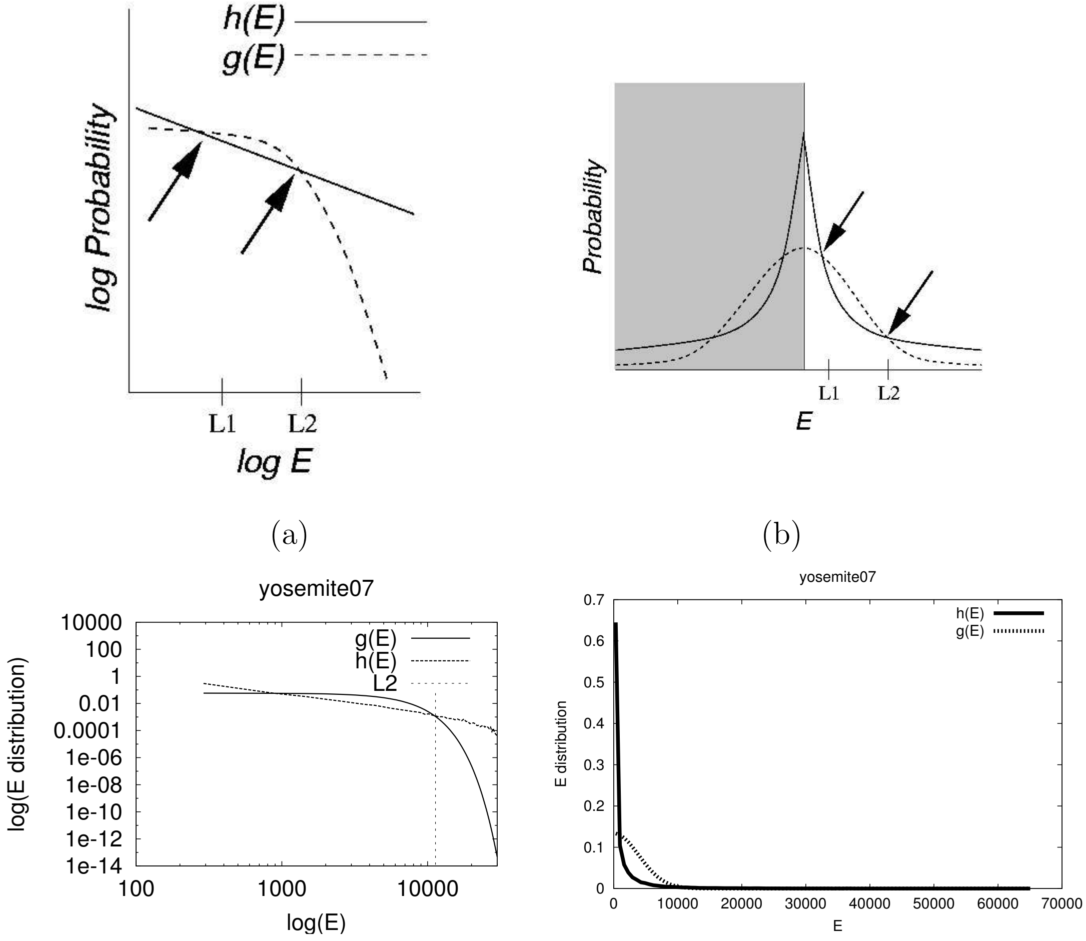
$$\sigma^2 = \sum_{E \in B_E} E^2 h(E), \tag{3.8}$$

where B_E is the set of E values of the histogram bins, and $h(E)$ is the probability of the orientation energy level E derived from the E-histogram.

I then proceeded to calculate the continuous normal probability density function values $N(x; 0, \sigma^2)$ with mean 0 and variance σ^2 for all E values, and normalized as

$$g(E) = \frac{N(E; 0, \sigma^2)}{\sum_{E \in B_E} N(E; 0, \sigma^2)}, \text{ for } E \in B_E, \tag{3.9}$$

where B_E is the set of E values of the histogram bins, and $g(E)$ is the resulting discretized half-normal probability mass function of orientation energy level E. The comparison of $h(E)$ and $g(E)$ of 6 natural images in log-log scale are shown in Fig. 8. The OED distributions from natural images shows a significant deviation from a Gaussian (Normal) distribution. From Fig.8. we can see that after the second



(a) (b)

(c) (d)

Fig. 7. The figure shows the comparison of a distribution $h(E)$ that follows the power law (solid curve) with a normal distribution $g(E)$ with the same variance (dashed curve), in both the linear and log-log scales. (a) and (b) illustrate a power law distribution, and (c) and (d) show the actual distributions obtained from a natural image. The x-axis represents the orientation energy and the y-axis the probability. It is of interest to look for orientation energy values that have high probabilities, i.e., where $h(E)$ is greater than $g(E)$. Only positive values for E are considered.

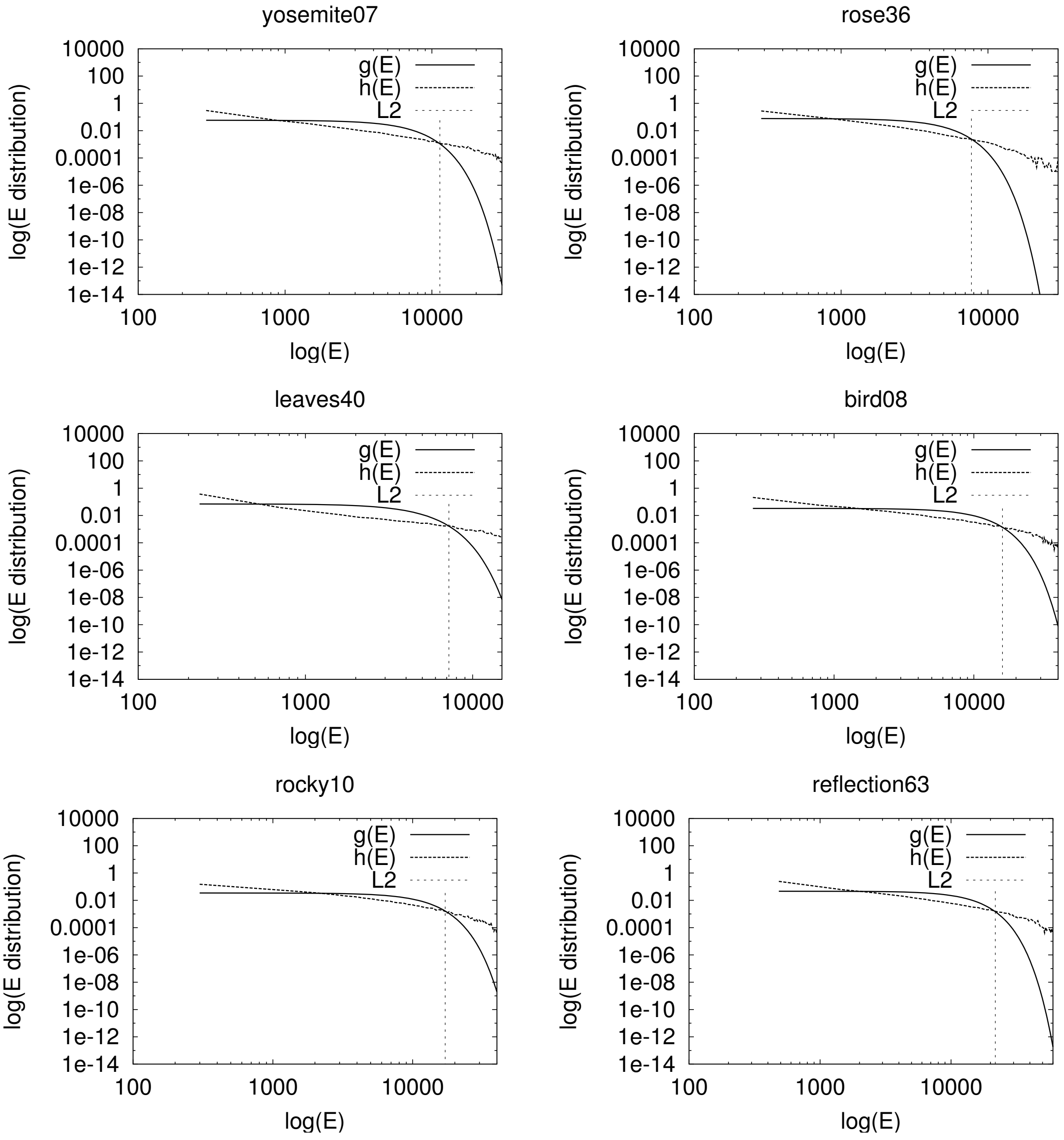


Fig. 8. Comparison of OED $h(E)$ of six natural images with their corresponding normal distributions $g(E)$ of the same variance. The second intersection ($L2$) of the two curves are marked by a vertical line in each plot. We can see that beyond this point, $g(E)$ plummets, while $h(E)$ remains steady, thus anything beyond $L2$ may be seen as salient.

intersection point of $L2$, $g(E)$ significantly drops in comparison to $h(E)$. This point may indicate where salient levels of orientation energy begin to occur. To test if the choice of a Gaussian baseline is reasonable in terms of salience, I compared the $L2$ values with human preference for thresholding. The details and results of this experiment will be described in Chapter IV.

CHAPTER IV

EXPERIMENTS AND RESULTS

It was proposed in the previous chapter that the non-Gaussian nature of orientation energy responses could be utilized to understand the perceived salience by humans. To test this hypothesis, it becomes necessary to compare thresholds obtained by utilizing the non-Gaussian property to thresholds selected by humans. Such a psychophysical analysis is described in this chapter. The results presented in this chapter are largely based on [25], which describes the research project I took part in.

A. Experiment 1: Comparison with Psychophysical Data

1. Methods

Since the $L2$ values obtained by comparing the OED of natural images to a Gaussian distribution could serve as a reasonable measure of salience, they were compared with human data. For this comparison, a total of 31 natural images from a wide variety of natural terrain were used. The $L2$ values were computationally identified based on equations 3.7 and 3.9. The human chosen thresholds for all the 31 natural images were determined by a single person (SB) in my research group. The thresholded E matrices at 55th to the 95th percentile of $h(E)$ at an interval of 5 percentiles were shown to SB for each image. The n-percentile point was found by locating the E value corresponding to the nth percentage point in the cumulative histogram range of the orientation energy. Fig. 9 shows an example image and some fixed-percentile threshold results shown to the human observer. The best threshold was determined by SB by making use of the criteria : (1) contour objects should be presented as much as possible, and (2) noisy background edges should be eliminated as much as possible.

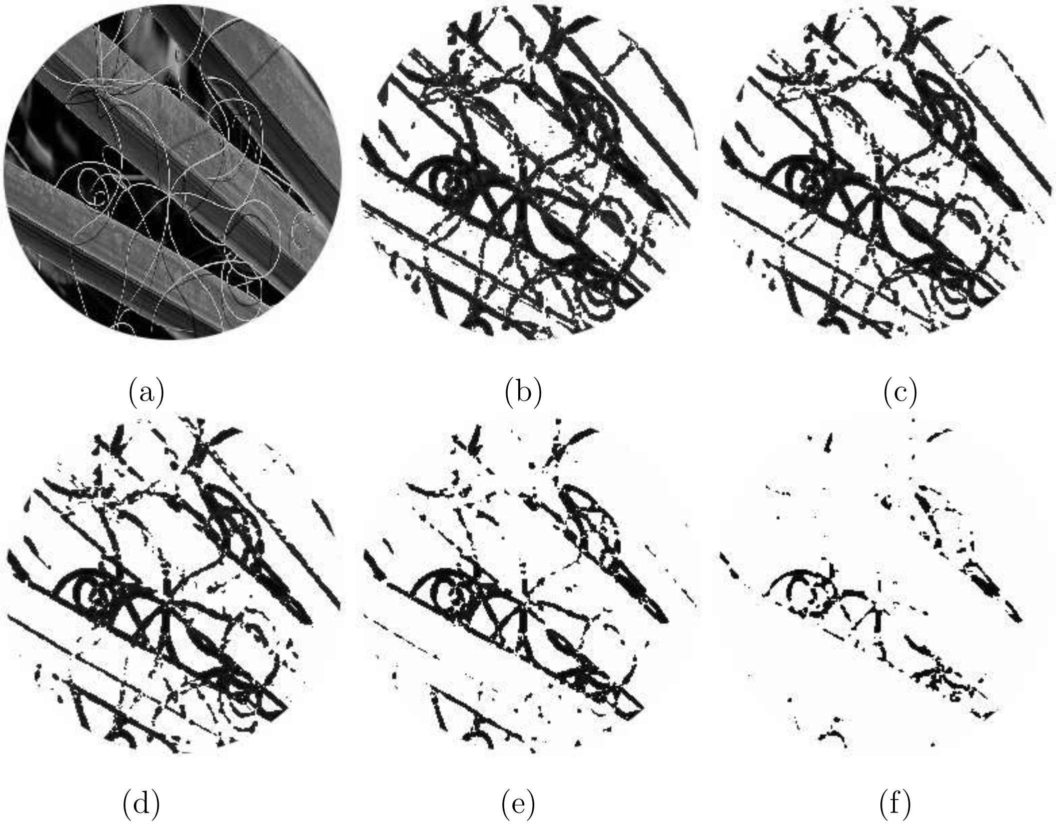


Fig. 9. A sample natural image and fixed-percentile thresholded orientation energy images that were shown to a human observer: (a) shows the natural image, (b) - (f) show the results from 55-,65-,75-, 85-, and 95-percentile thresholding.

Such criteria broadly correspond to our perceived levels of salience in images. For example, in Fig. 9, the human observer chose the threshold result corresponding to the 75-percentile threshold as the one that complies best to the criteria. For each image, the $L2$ value and the human-selected threshold was compared.

2. Results

Fig. 10 shows the results of comparison of the L2 values with the human-chosen threshold values for the 31 natural images. A clear linearity between $L2$ and the threshold selected by humans can be observed from Fig. 10. Furthermore, the $L2$

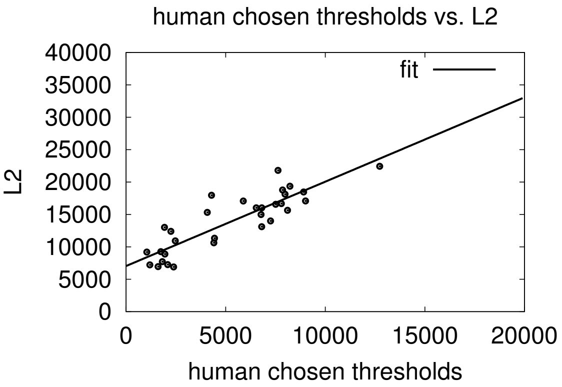


Fig. 10. Orientation Energy threshold selected by humans vs. L2. The manually chosen thresholds are compared to the $L2$ values for each image. Each point in the plot corresponds to one of 31 natural images used in the calculation. The straight line in the figure shows a linear fit to the data.

value has a close linearity with the square root of the raw second moment (σ) of the E values, as is shown in Fig. 11.

To see whether the σ values also linearly correlate with human-chosen thresholds, the two sets of values were compared. As expected, the σ of the orientation energy distribution $g(E)$ showed a clear linearity to the threshold selected by humans, as shown in Fig. 12.

3. Discussion

The linearity of the $L2$ values with the human-chosen thresholds shows that we can use those values to estimate the threshold of E preferred by humans. However, this is not always feasible because of the high time and space requirements for constructing $h(E)$ and $g(E)$, and also for locating $L2$. On the other hand, the linearity of $L2$ values with the square root of the raw second moment (σ) of the OED, and the subsequent result of good linearity between σ and human-chosen thresholds show that a good threshold criterion can be obtained using just the raw second moment. Since the

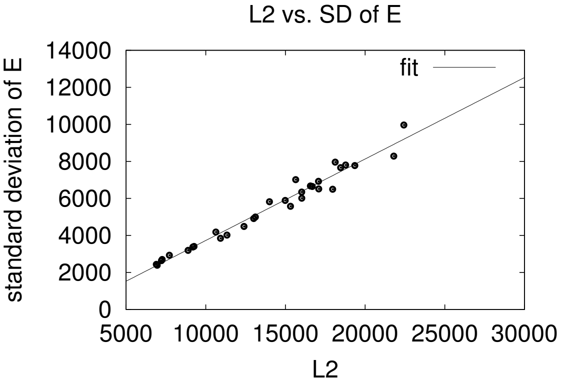


Fig. 11. $L2$ vs. σ of the OED. The $L2$ values obtained by comparing $h(E)$ and $g(E)$ are compared to the square root of the raw second moment (σ) of $g(E)$. Each point in the plot corresponds to one of 31 natural images. The straight line in the figure shows a linear fit to the data.

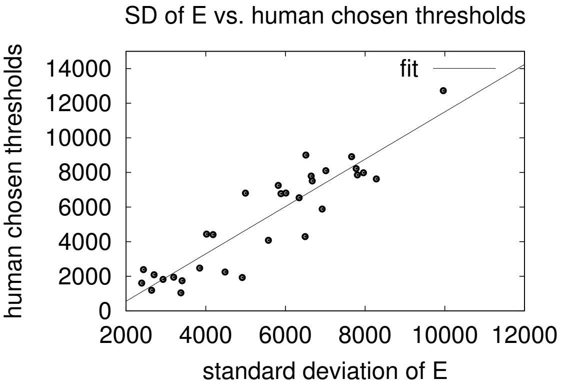


Fig. 12. Orientation Energy thresholds selected by humans vs. σ. The orientation energy thresholds manually selected are compared to the square root of the raw second moment (σ) of $g(E)$. Each point in the plot corresponds to one of 31 natural images. The straight line in the figure shows a linear fit to the data.

threshold calculation depends only on the value of σ, such a criterion could be more computationally efficient.

It may be noted that the effectiveness of the thresholding results using the methods described in this section is tested psychophysically by comparison with human preference. The human observer was shown the original image and was also supplied with the criteria for the selection of the best threshold. It could be argued that since such criteria concern higher-level visual processes in the observer's brain, they could come into play in the selection of the best threshold. However salience detection is mostly a low-level visual process and there is no reason to believe that changes in higher level cognitive factors could influence changes in lower level visual processes. Altering low-level visual processes may, however, alter higher level processes such as perception, but such a situation is not applicable here since the instructions to the human observer are high-level instructions, and do not alter the low level visual processes in any way. Thus comparison with human-chosen thresholds can be justified to give us a clue toward which kind of thresholding mechanism is closest to the salience detection mechanisms in the human visual system.

B. Experiment 2: Application to OE Thresholding

In this section, I show thresholding results using a method that makes use of the σ of the OED, and compare with threshold results selected by humans and those obtained by a fixed percentile thresholding of the OED.

1. Methods

The raw second moment of the E distribution (the expected value of E^2) for the input image can be calculated as in Equation. 3.8. Then a linear equation of the preferred orientation energy threshold T_σ is expressed as a function of the square root of the raw second moment (σ) of the orientation energy, :

$$T_\sigma = 1.37\sigma - 2176.59, \tag{4.1}$$

which is the linear fit shown in Fig. 12. Such a kind of thresholding utilizing the OED of natural images will be called the OED-derived adaptive thresholding elsewhere in the thesis.

For the thresholding experiments, 3 sample natural images belonging to the same set in Fig. 4 were used. A single threshold value computed from Eqn. 4.1 was first applied to the orientation energy matrix of two sample images to see if it comes close to our perceived level of edginess in the images. To analyze the effectiveness of this method, the results were compared both with the human-chosen threshold result and with a fixed 85-percentile thresholding on the orientation energy values, which is a simple way of obtaining salient contours. For the 85-percentile thresholding, the orientation energy value corresponding to the 85th percentile of the orientation energy is used as a cut-off point. The salient edge elements would then be defined by those pixels that have orientation energy values greater than this fixed cut-off point. Experiments with fixed percentile thresholding showed that the 85-percentile cutoff is a reasonable thresholding criterion for many natural images. This percentile threshold value provides the best average fixed-percentile threshold on most natural images, preserving edge contours intact while suppressing background noise as much as possible.

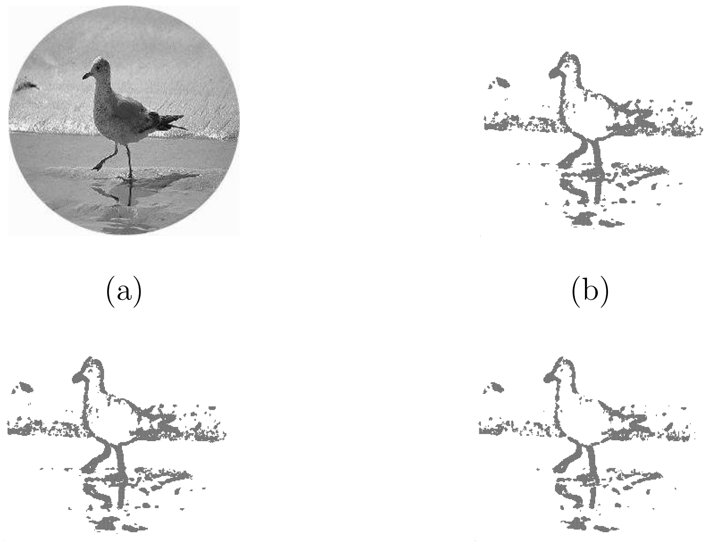
The OED-derived thresholding method offers an advantage because it depends only on the raw second moment of the OED, and this value need not always be collected from the entire orientation energy matrix. Thus the OED-derived thresholding method could be used to calculate locally optimal thresholds for local patches in the image. To test this, a sliding window of of size 21 × 21 pixels was used across different local patches of the orientation energy matrix of another sample image having variations of features to gather thresholds by computing the σ values of the local patches. The threshold at the center of the sliding window was determined by the σ value calculated from within that window. The efficiency of the thresholding result was then tested by comparing with the human-chosen thresholding result, the fixed 85-percentile thresholding result and the global OED-derived thresholding result.

2. Results

Fig. 13 shows the result of using the OED-derived adaptive thresholding globally on a natural image, and comparison with both the human-chosen threshold result and the fixed 85-percentile threshold result. The OED of the image has the property that is not too broad or too narrow, and we can see that all the threshold results are similar.

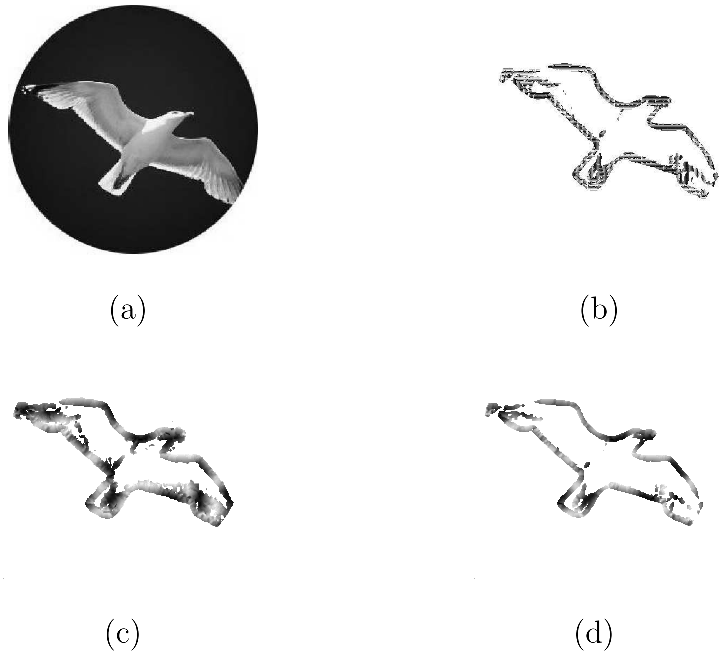
Fig. 14 shows global thresholding results for another image where the OED-derived adaptive thresholding result is compared with the human preference and a fixed percentile threshold result that corresponds to the 85-percentile of the OED. The input image has the property of huge open space in the background. From the figure it can be seen that the fixed 85-percentile thresholding method gives sub-optimal results due to the concentration of the energy values around zero. This has the effect of reducing the 85-percentile point to a lower value than the effective threshold. For example, a lot of edges around the wing tips of the bird are not thresholded in this method. However, the OED-derived thresholding result deals effectively with such a situation because the spread of the distribution is taken into account.

Global thresholding could still be inadequate for cases where there is a large variation of σ across different local patches in an orientation energy matrix. Fig. 15. shows such an example where there are both bold features (the blades of the leaves)



(a) (b) (c) (d)

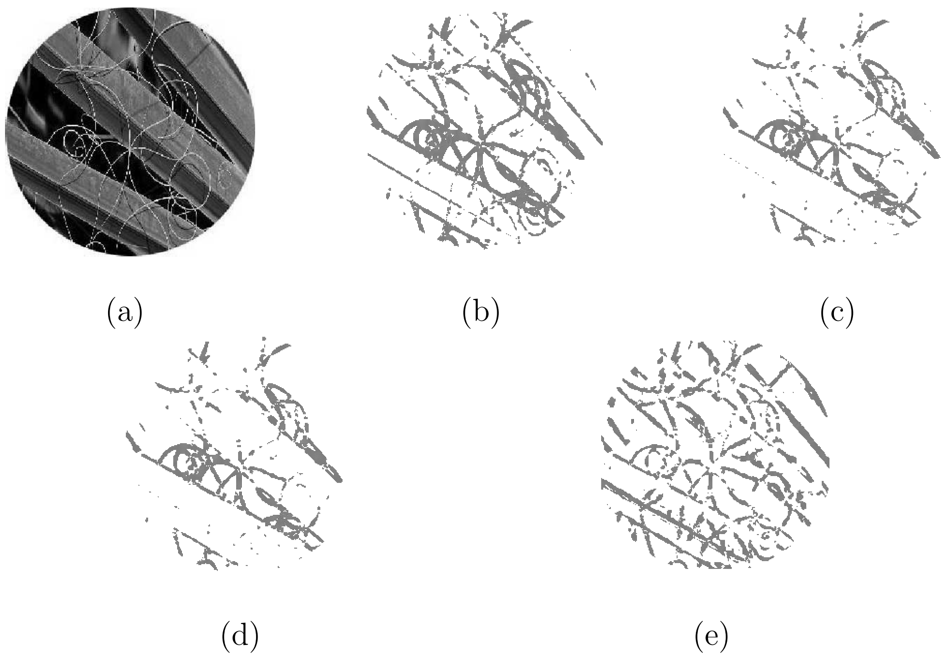
Fig. 13. Global Thresholding Results. (a) The original image. (b) The threshold result selected by a human. (c) The OED-derived thresholding result where pixels are replaced by oriented lines indicating the local orientation. The result of thresholding is close to our perceived salience of edges in (a).



(a) (b)

(c) (d)

Fig. 14. Comparison of global OED-derived thresholding with fixed-percentile thresholding. (a) The original image. (b) The threshold result selected by a human. (c) The threshold at the 85-percentile of $g(E)$. (d) The OED-derived thresholding result where pixels are replaced by oriented lines indicating the local orientation.



(a) (b) (c)

(d) (e)

Fig. 15. Comparison of global thresholding with local thresholding. (a) The original image. (b) The threshold selected by a human which is globally applied. (c) The threshold at the 85-percentile of $g(E)$, also globally applied. (d) The OED-derived global thresholding result where pixels are replaced by oriented lines indicating the local orientation. (e) The result of local OED-derived thresholding using a sliding window of size 21 × 21.

and thin thread like features interspersed together. Fig. 15.(b)-(d) show that the global thresholding methods are not effective in detecting these features. Fig. 15. (e) shows the result of a *local thresholding* method using a sliding window of size 21 × 21 pixels. We can see that the local thresholding method preserves both fine and bold contour features in the image.

3. Discussion

Results of the global and local thresholding methods using the orientation energy distribution show that relative advantages are provided by the methods for different inputs. Both the global OED-derived and the fixed 85-percentile thresholding methods was found to be effective when the OED of the natural image was not too broad or too narrow. The global OED-derived thresholding method seems to be more advantageous if there are not much variations of σ across several regions of the image, since a single value can be effectively and globally applied. For images with wide variations of σ across local patches, the local thresholding method offers the advantage of providing locally optimal threshold values.

CHAPTER V

ANALYSIS

The previous chapter has shown that comparison of the OED from natural images to a Gaussian distribution can lead to orientation energy thresholding mechanisms that can predict the perceived salience of contours in humans. This chapter will analyze the justification behind such an approach, and also provide a quantitative analysis to precisely measure the efficiency of the approach.

A. Analysis of the Gaussian Baseline

It has been shown that the Gaussian distribution serves as an effective baseline for detecting salient levels of orientation energy. It would be useful to consider why it could be such a good baseline. If we consider orientation energy values with super-Gaussian probabilities as being salient, a potential link may be formed between the concept of salience and the concept of *Suspicious Coincidence* proposed by Barlow [1].

The concept of Suspicious Coincidence is basically statistical non-independence applied to brain theory. The view held by Barlow was that the goal of the perceptual system was to effectively detect suspicious coincidence events in the environment. The co-occurrence of statistical events A and B could be said to be a suspicious coincidence if they occur more often together than can be expected from their individual probabilities. This means that the joint probability of the two events should exceed the product of their individual probabilities if they were to be considered suspicious:

$$P(A,B) > P(A)P(B). \tag{5.1}$$

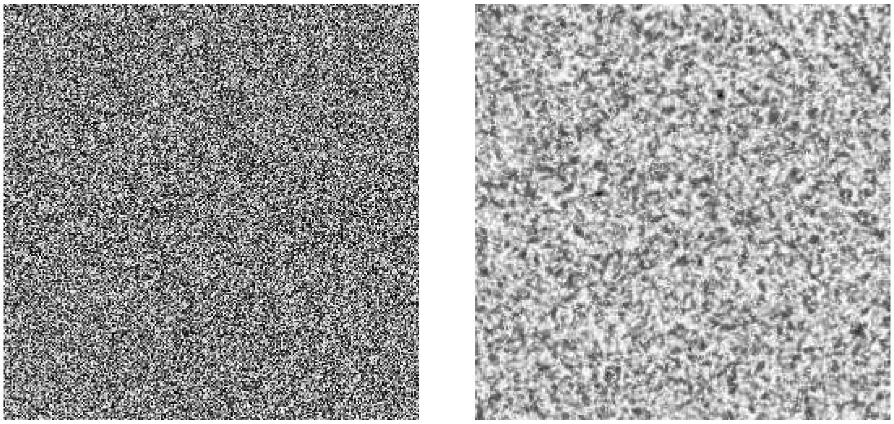
This criterion is based on the comparison to the baseline case when A and B are statistically independent events where $P(A, B) = P(A)P(B)$.

In the domain of image analysis, the concept of suspicious coincidence can be interpreted by treating each pixel as a random variable. We can then consider whether the association of a pair of pixel intensities is suspicious or not. For example, pixel pairs taken from oriented features may be seen as suspicious. We can then test if the orientation energy value in that area is high as well. On the other hand, when we consider an image where all pixels are independent, such as a uniformly distributed white-noise image, we would expect the salience (i.e., the OE) to be low.

If suspicious coincidence is indeed related to salience as in this thesis, the OED of a white-noise image should show up as non-salient. Since salience is defined by a deviation from a Gaussian baseline, the OED for a white-noise image should thus be near-Gaussian, since it does not show any suspicious coincidence.

To test if this is indeed the case, the OED from a white-noise image was calculated and compared to a normal distribution. The white-noise image was of size 256 × 256 consisting of uniformly randomly distributed intensity values between 0 and 255. The OED for this image was calculated using the method outlined previously in Chapter III. Fig. 16 shows the white-noise image and its orientation energy. The OED was then compared with the matching normal distribution of the same variance to see if they were similar. The two distributions were indeed closely overlapped, as expected (Fig. 17; cf. [23]).

A direct consequence of the above results is that the white-noise based OED could also be used to find the threshold for salient contour detection. I conducted another experiment in which new $L2$ values were generated by comparing the OED from natural images to the white-noise based distribution. The variance of the two distributions was made equal by the following procedure. The standard deviation of a random variable which is scaled by a factor of a is $a \times \sigma$ where σ is the standard



(a)

(b)

Fig. 16. A white-noise image and its orientation energy distribution. (a) The original image. (b) The orientation energy E. No clear structure can be discerned.

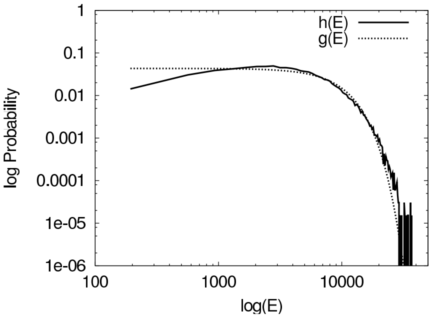
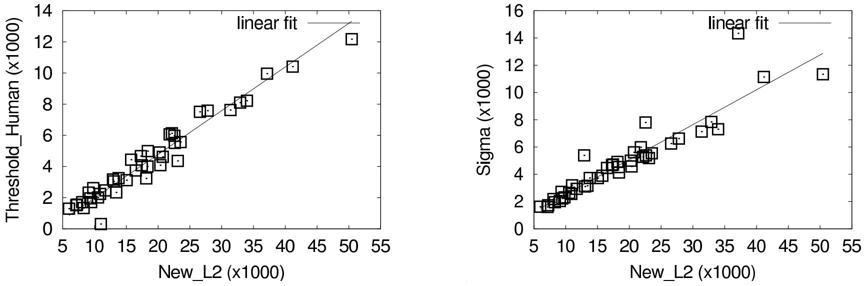


Fig. 17. The OED of a white-noise image (solid)and its matching normal distribution (dotted). The log-log plot shows that the two distributions are quite similar.



(a) (b)

Fig. 18. Comparison of white-noise based $L2$ against human-chosen thresholds and σ_h. (a) The new $L2$ values derived from the white-noise based distribution are plotted against the human-chosen thresholds for a set of 42 natural images. The correlation coefficient was 0.98, indicating a strong linearity. (b) The new $L2$ values are compared against the raw standard deviation σ_h of $h(E)$. The correlation coefficient was 0.91, and a strong linearity can be observed between the two sets of values.

deviation before scaling. Thus to make the two distributions have the same standard deviation, we can simply multiply the orientation energy matrix of the white-noise image with the constant σ_h/σ_r, where σ_h and σ_r are the standard deviations from the natural image and white-noise image respectively. The OED was then calculated and the new $L2$ values obtained by comparing the two distributions. The next step was to compare these values to the orientation energy thresholds selected by humans, according to the criteria outlined in Section A of Chapter IV. The result of the comparison is shown in Fig. 18. (a). The correlation coefficient value for the comparison was 0.98. The result shows that the white-noise based $L2$ values also have a strong linear relationship with the human-chosen thresholds. The new L_2 values also have a strong linear relationship with the σ of the OED of the natural images, with a value of 0.91 for the correlation coefficient, as shown in Fig. 18. (b).

The white-noise distribution is a little more accurate as a baseline; however the important point here is that Gaussian distributions are also close approximations, and this analysis provides justification and theoretical insight for the computationally simpler Gaussian approach to be used as the baseline. Chapter VI will examine the neural plausibility of a thresholding mechanism that uses the Gaussian distribution as the baseline.

B. Quantitative Analysis with Synthetic Images

To establish precisely the effectiveness of the thresholding mechanism by comparison with a Gaussian distribution, I carried out a quantitative analysis. For such an analysis, the natural image input is not the best choice because of the lack of controllability of its features. It then becomes necessary to consider other kinds of input - artificially generated images - that closely mimic the characteristics of the natural image input and yet provide easy access to manipulation and control. In this chapter, I will describe such a type of artificial input, one containing sets of overlapping squares and embedded noise, and the results of thresholding for both the fixed percentile and OED-derived adaptive thresholding methods. Such kinds of synthetic images have the advantage that the error can be precisely measured, by comparison with the orientation energy matrix of the "clean" image without the noise as a reference.

1. Synthetic Images and their OEDs

The synthetic image input was created by generating random squares of various sizes where the gray-level was uniform within each square but different from adjacent squares. Uniform background noise was also embedded among these squares. A large number of such squares were generated and were subjected to a circular aperture to discount any artefactual orientation bias. The orientation energy distribution for the

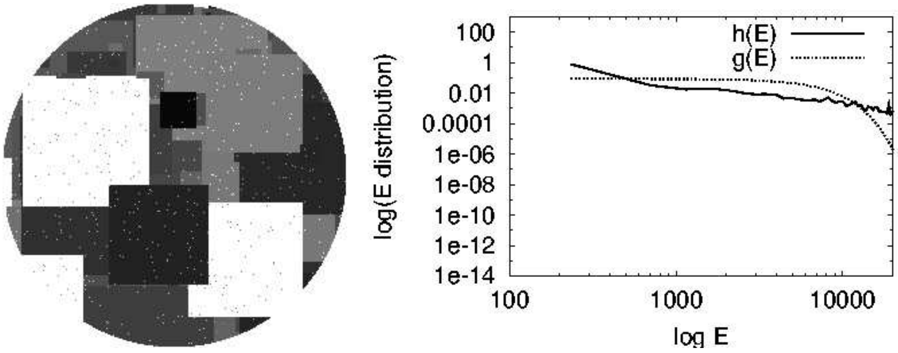


Fig. 19. A synthetic image consisting of a set of overlapping squares with embedded noise and its corresponding orientation energy distribution shown against the normal distribution of the same variance. We can see that the synthetic image has a distribution almost similar to that of a natural image.

resulting image was then calculated using the same procedure as for natural images. The OED for the synthetic images were similar to that for the natural images, and had a characteristic high peak and heavy tail. Fig. 19 shows a sample synthetic image and its OED distribution plotted against the normal distribution in log-log scale. We can see that the OED of the synthetic image closely mimics that of natural images. In fact Lee, Mumford and Huang [24] showed that similarly generated synthetic images have statistical properties very similar to natural images.

A quantitative analysis using synthetic images can then be focused on suitable thresholding of the images to remove the background noise and detect salient features given by the edges of the squares. A good thresholding method should perform two objectives : (1) detect as many salient edges as possible, and (2) remove as much background noise as possible. In the following sections, I describe two sets of experiments for the quantitative analysis of the thresholding methods where, (1) the number of overlapping squares was kept constant but the background noise was varied; (2) and the background noise was kept constant and the number of squares (the input count) was varied. For each of the types, I generated five different image configurations for a more thorough analysis.

2. Variation in Noise and Performance

For this experiment, I generated synthetic images by keeping the input count (the number of overlapping square elements) constant and varied the background noise. The embedded noise was uniformly distributed. Three density levels of noise were used corresponding to 10%, 5% and 1% noise. The combined resultant images were then subjected to a circular aperture, and the performance of the fixed and OED-derived adaptive thresholding methods were measured.

To better investigate the relative merits of the global and local thresholding methods, both the variations were used for the fixed percentile and OED-derived methods on all the images. For each noise density level, 5 representative synthetic images were generated with different configurations of the square patterns in them, with the total number of squares fixed to 300. The OEDs for each image input were calculated according to the method outlined in Chapter III. A fixed threshold of 85-percentile was used for the fixed percentile thresholding. For the local thresholding, a sliding window of size 21 × 21 was used, as previously for the natural image experiments.

There were four different thresholding methods that were compared : (1) the global fixed percentile thresholding, (2) the local fixed percentile thresholding, (3) the global OED-derived adaptive thresholding, and (4) the local OED-derived adaptive thresholding methods.

Thresholding results for a sample image with 300 overlapping squares and with 95-percentile and above embedded noise level are shown in Fig. 20. We can see that the global OED-derived thresholding method provides the best performance for this kind of input.

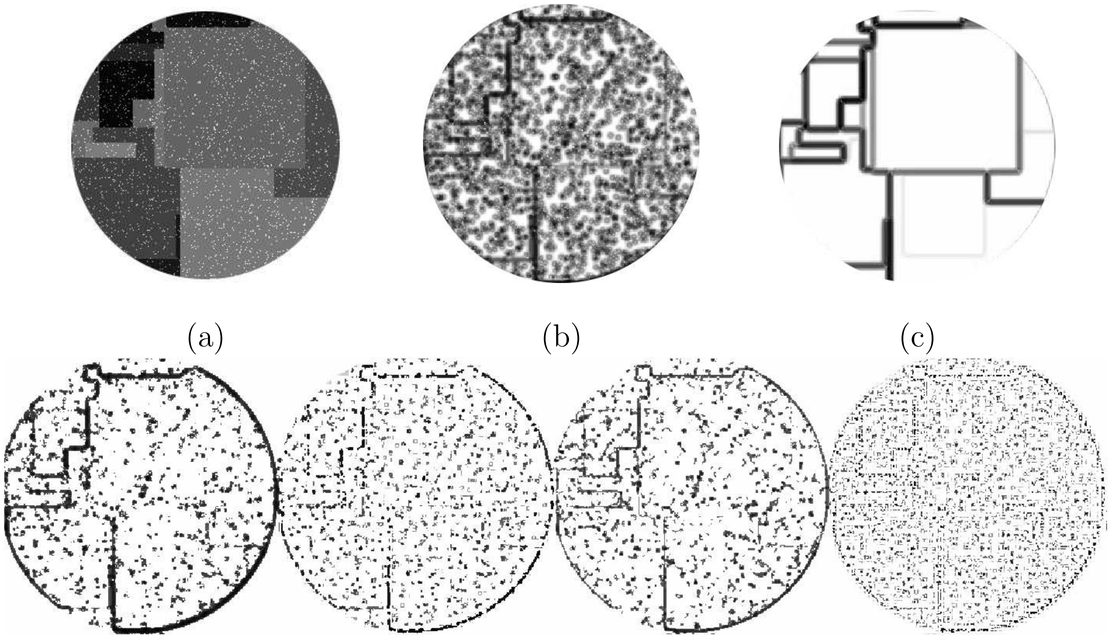

(a) (b) (c)
(d) (e) (f) (g)

Fig. 20. (a) A synthetic image consisting of 300 overlapping squares with embedded noise. (b) The orientation energy matrix for (a). (c) The orientation energy matrix for the synthetic image without the noise that is used as the reference. Results of thresholding by the four different methods of (d) global OED, (e) global 85-percentile, (f) local OED, and (g) local 85-percentile respectively are shown from the left to the right. The global OED-derived adaptive thresholding method seems to offer the best performance for this input.

Such qualitative results are also backed by values for the quantitative measure of *sum of squared error* (SS Error). The SS Error for a thresholding result could be defined as the sum of the square of the difference in the orientation energy values of the thresholded image (containing background noise) from the un-thresholded image representing the ideal output of zero noise. The average SS Error values for a sample noise level is shown in the plot of Fig. 21. We can see that the sum of squared error is the lowest for the global OED derived method, followed closely by the local OED derived method. The SS Error values for the fixed 85-percentile methods were found to be significantly higher. The significance in this difference was evaluated the paired t-test. The p-values for the paired t-test of the difference in the mean SS Error values are given in Table I. (The '<' and '>' symbols in each table cell indicate the relation between the mean SS Error values of the two thresholding methods.) The p-values of all the probable pairs of comparisons except the global OED vs. local OED comparison were found to be less than 0.025, which indicates that the mean SS Error values for the methods could indeed be significantly different. The p-value for the global OED vs. local OED comparison was found to be much higher, leading to inconclusive evidence that the means could be significantly different.

The results for this experiment show that the OED-derived adaptive methods overwhelmingly outperformed the fixed percentile based methods in terms of detection efficiency of edges in the image input and suppression of noise. The global fixed However, the global OED-derived method offered a slight improvement in performance compared to the local method. This could probably be attributed to specific properties of the input images considered, such as the low density of the squares in the image. In fact, a later result with larger input count hints at the possibility that this might indeed be the case.

The variation of noise seemed to have little effect on the relative performance

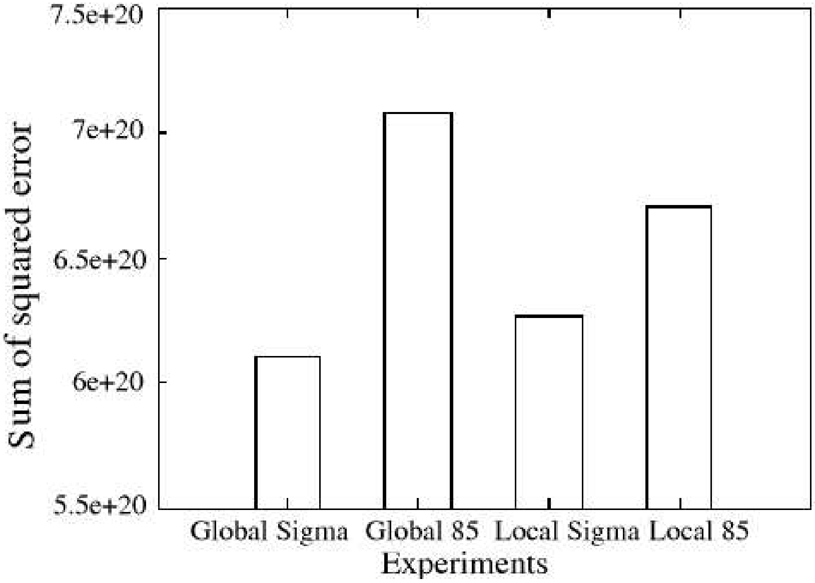


Fig. 21. Bar Plots showing the average SS Error values for the thresholding results for a sample noise level for the synthetic images. The global and local OED-derived method have significantly smaller SS Error values than the fixed 85-percentile methods.

of the thresholding methods. The global OED thresholding method offered the best performance closely followed by the local OED thresholding method. The global and local 85-percentile based thresholding methods performed differently for different inputs but were always weaker than the OED-derived methods.

3. Variation in Input Count and Performance

The second experiment that was carried out was to keep the background noise constant while varying the number of input features. Three different image configurations were used where the number of overlapping synthetic squares were 100, 300 and 500 respectively. The constant background noise level was kept as 1%. For each input configuration, 5 different image samples were generated as for the first experiment. A fixed threshold of 85-percentile was used for the fixed percentile thresholding meth-

Table I. SS Error means when noise is varied

	Global OED	Global 85%	Local OED	Local 85%
Global OED	X	< (p=0.018)	< (p=0.35)	< (p=0.021)
Global 85%	X	X	> (p=0.019)	> (p=0.014)
Local OED	X	X	X	< (p=0.025)
Local 85%	X	X	X	X

ods, and a sliding window of size 21 × 21 was employed for the local thresholding method as previously.

Thresholding results for a sample image with 500 overlapping squares and with 1% embedded noise level are shown in Fig. 22. The OED-derived thresholding methods again outperformed the fixed 85-percentile based methods for all the input configurations. Among the OED-derived methods, the global method offered better performance for smaller input count but fell behind the local method for larger input count. Thus the local thresholding result for the 500-square configuration was better than the corresponding global thresholding result. The average SS Error value is the lowest for the local OED-derived thresholding method, but quite high for the fixed-percentile methods. Fig. 23 shows this comparison. The paired t-test statistic for the average SS Errors for the different thresholding methods had p-values less than 0.15 for all the pairs of comparisons, except for the Global OED vs. Local OED comparison. Table II shows the p-values for the paired t-test of the difference in the mean SS Error values for all the thresholding methods. (The '<' and '>' symbols in each table cell indicate the relation between the mean SS Error values of the two thresholding methods.)

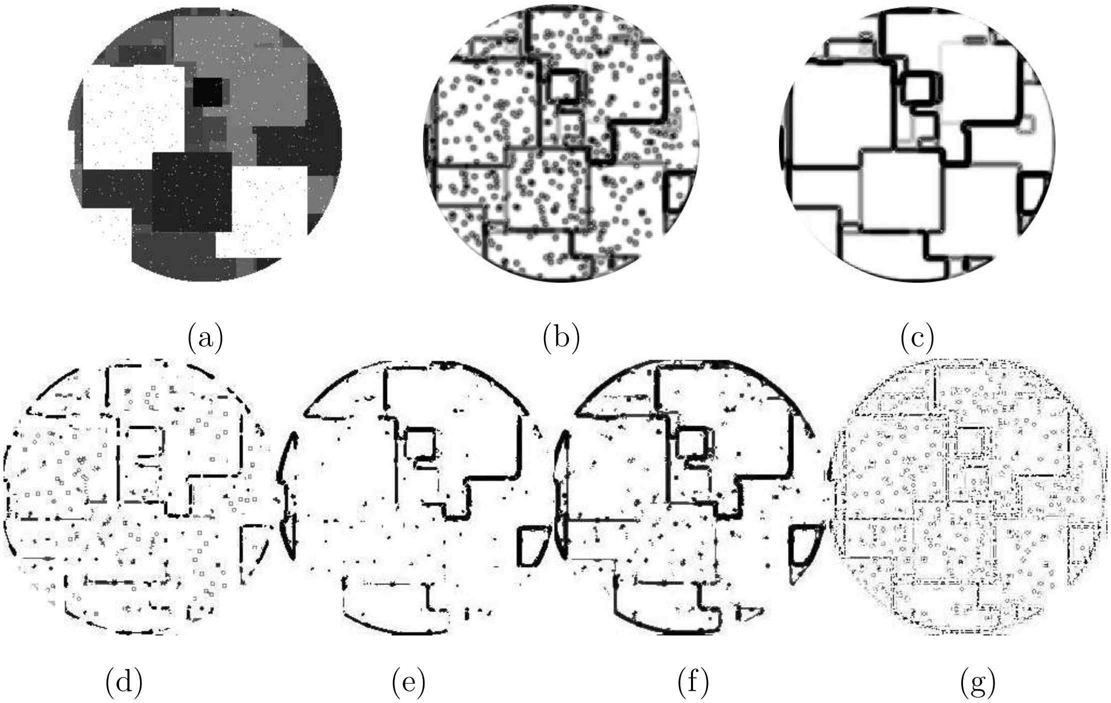


Fig. 22. (a) A synthetic image consisting of 500 overlapping squares with embedded noise. (b) The orientation energy matrix for (a). (c) The orientation energy matrix for the synthetic image without the noise that is used as the reference. Results of thresholding by the four different methods of (d) global OED, (e) global 85-percentile, (f) local OED, and (g) local-85 percentile respectively are shown from the left to the right. The local OED-derived adaptive thresholding method seems to offer the best performance for this kind of input.

4. Summary

The experiments and the quantitative analysis presented in this chapter help to reinforce the effectiveness of the OED-derived adaptive thresholding methods over the fixed-percentile methods. Even though qualitatively it could be seen easily that this is the case, objective measures such as the sum of squared error measure described here help compare the thresholding methods effectively.

One of the properties of the synthetic images used in this analysis was that they contained noise embedded within in. It would be interesting to speculate what such

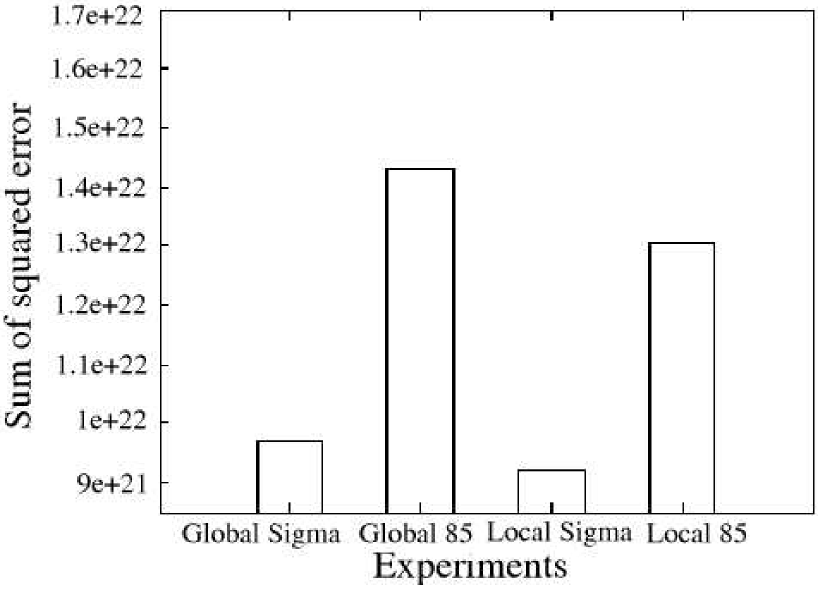


Fig. 23. Bar plots showing the average SS Error values for the thresholding results for a sample noise level for the synthetic images. The global and local OED-derived methods have significantly smaller SS Error values than the fixed 85 percentile methods.

a noise would correspond to for natural images. Since the primary interest is the understanding of visual processes and since the visual system does not need to be concerned with noise from the natural scenes it receives from the environment, it becomes necessary to identify what the noise would mean for natural images. One answer is that the noise could correspond to textures within the natural image. For example, the noise could correspond to certain non-edge features in the image which will normally be excluded from a thresholding mechanism for contour detection. In another context, the noise could be attributed to the properties for scenes which are viewed in dim or low light conditions. Noise could also be introduced into visual scene input if there are defects in the retinal cells leading to poor reception.

One of the results that came up from the above analysis was that there was no clear difference between the global and local OED-derived adaptive thresholding

Table II. SS Error Means when input count is varied

	Global OED	Global 85%	Local OED	Local 85%
Global OED	X	$<$ (p=0.0107)	$>$ (p=0.258)	$<$ (p=0.014)
Global 85%	X	X	$>$ (p=0.011)	$>$ (p=0.006)
Local OED	X	X	X	$<$ (p=0.015)
Local 85%	X	X	X	X

methods. For certain kinds of input where the number of input features was less, the global thresholding method seemed more effective than the local thresholding method. When the number of input features was high, the local thresholding method performed better. The key here is the *variation* of the input features at a local scale, and it could be expected that the synthetic input image with 300 squares has less local variation in input features compared to the one with 500 squares. As was seen with the experimental results on natural images in the previous chapter, the local thresholding method thrives on instances where there is a wide variation of features in the input image.

CHAPTER VI

DISCUSSION AND FUTURE WORK

A. Potential Neural Basis of Salience Detection

The OED-derived adaptive thresholding method described in the previous chapters only depends on the raw second moment σ^2 of the orientation energy matrix. It can be seen that the σ^2 value can be directly calculated from a given orientation energy matrix using a simple neural network. For example, if E is the response matrix in the visual cortex V1 (similar to orientation energy), a suitable threshold on the V1 response could be easily calculated as,

$$\sigma^2 = \sum_{i,j} w_{ij} g(E_{ij}) \tag{6.1}$$

where i, j represent the indices, w_{ij} represents the connection weight which also serves as the normalization constant, $g(x) = x^2$ is the square activation function, and E_{ij} is the V1 response at location i, j. The value σ can be obtained by passing the resultant value through an activation function of the form $f(x) = \sqrt{x}$. This means that the raw second moment can be easily obtained by using a weighted sum of V1 response passed through a quadratic non-linearity and then through a square root activation function. This is depicted in the flowchart shown in Fig. 24.

The relative effectiveness of such a method could be understood given the fact that there is no need to construct a histogram of the responses or its matching normal distribution. It could be speculated that such a mechanism could actually exist in the visual cortex or in higher visual areas. The usefulness of such quadratic non-linearity has been previously suggested by Heeger [15] where it was suggested that the neural circuitry in the complex cells of the primary visual cortex (V1) could

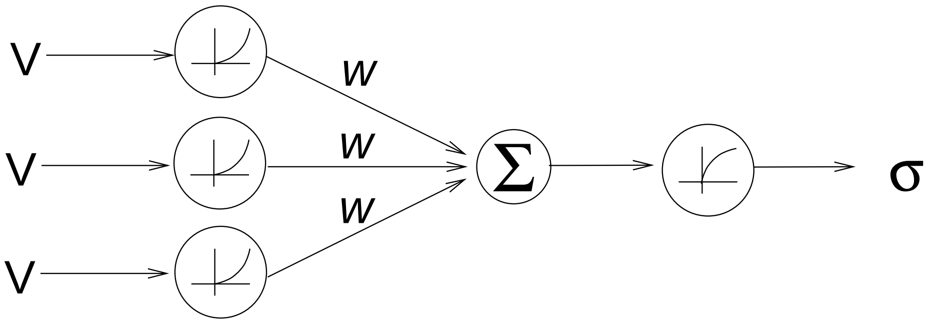


Fig. 24. Flowchart showing the calculation of sigma by a simple neural mechanism.

probably implement squared response functions similar to the squared orientation energy response. Such squared response functions have also been known to improve performance in other learning tasks. For example, Gastaldo et al. [12] use such functions to model a neural network to qualitatively assess the performance of image enhancement algorithms.

If a similar thresholding mechanism utilizing a quadratic non-linear input pattern indeed exists in the visual system, we can speculate about the reason behind it. It is reasonable to assume that if the human visual system has evolved to learn such kinds of thresholding mechanisms, efficiency of the representation could be one of the considerations for it to implemented. Since the quadratic non-linear input pattern corresponds to the squared V1 response, it would be worthwhile to examine if squared V1 response is indeed efficient as a representation.

B. Computational Efficiency of Orientation Energy Responses

A simple experiment was done to see whether the squared orientation energy (which can model the squared V1 response) functions as an efficient input compared with the plain orientation energy. The objective of the experiment is to compare the learning performance of the plain orientation energy with its squared form. I used a backpropagation network which consists of three layers - an input layer, one hidden layer, and an output layer. The units are connected in a feed-forward manner with the input units fully connected to the hidden units and the hidden units fully connected to the output units. (See [31] for details about learning in a backpropagation network.)

For the experiment, I used the backpropagation network with 49 units in the input layer, 4 units in the hidden layer, and 1 unit in the output layer. The output was supposed to give the raw second moment of the orientation energy, which as we have seen, is the parameter to be learned to find the threshold. I tested the learning performance of the plain orientation energy with the squared orientation energy input where the objective was to learn the raw second moment of the plain orientation energy. For the plain orientation energy experiment, the orientation energy values of local patches of size 7×7 pixels from many different images to a total of 1296 patterns were used as the training samples. For the squared orientation energy experiment, the square of the orientation energy values from the previous experiment were used, but the output was expected to be the same as previously (raw second moment of orientation energy). In both cases the inputs were normalized independently to reduce the effects of scaling and learning rate reduction due to one set of inputs being the square of the other set of inputs.

For both the experiments, the learning rate of the network was set at a value of 0.000001. The initial weights for all the units in the network were randomly assigned for ensuring correctness. Other parameters such as the momentum and bias were set as 0 and 1 respectively for both the experiments. The network was then trained independently with both types of input for a large number of training steps (epochs) until the error value at the output (the difference between the actual output and the desired output) was less than the threshold value of 0.0001. The whole experiment was repeated with another set of input values to give a more conclusive result.

Fig. 25 shows the results of the training where the learning efficiency of the network for two repetitions of training is plotted. For the first instance of training, the network took approximately 140,000 learning epochs to learn the raw second moment with the acceptable error value with the plain orientation energy input. In contrast, the network only took approximately 30,000 learning epochs to learn the required output with the squared orientation energy input. For the second instance of training with another set of inputs, the network took close to 135,000 epochs to learn the output for the plain orientation energy input compared to 50,000 epochs for the squared input. Since fewer epochs imply faster learning, the squared orientation energy input seems to be much more effective than the plain orientation energy for the backpropagation network to learn the raw second moment value.

Previously, I suggested that the squared orientation energy input, which models squared activation functions, can be easily implemented in neural hardware. The results from the backpropagation training experiment demonstrates that the squared orientation energy offers a computationally efficient alternative to the plain orientation energy for any network that learns the salience threshold.

C. Extensions to Other Modalities

We can apply this method of utilizing representations of response histograms to other visual and sensory modalities as well. For example, Liu and Wang [27] use spectral histograms to segment and synthesize texture images. A spectral histogram is a combination of many different kinds of filter response histograms. It would be interesting to analyze the response distribution in the different spectral histograms to gain insight into how salient features can be detected in the input scene. For example, the spatial frequency power spectrum also shows a power law. Thus a similar approach could

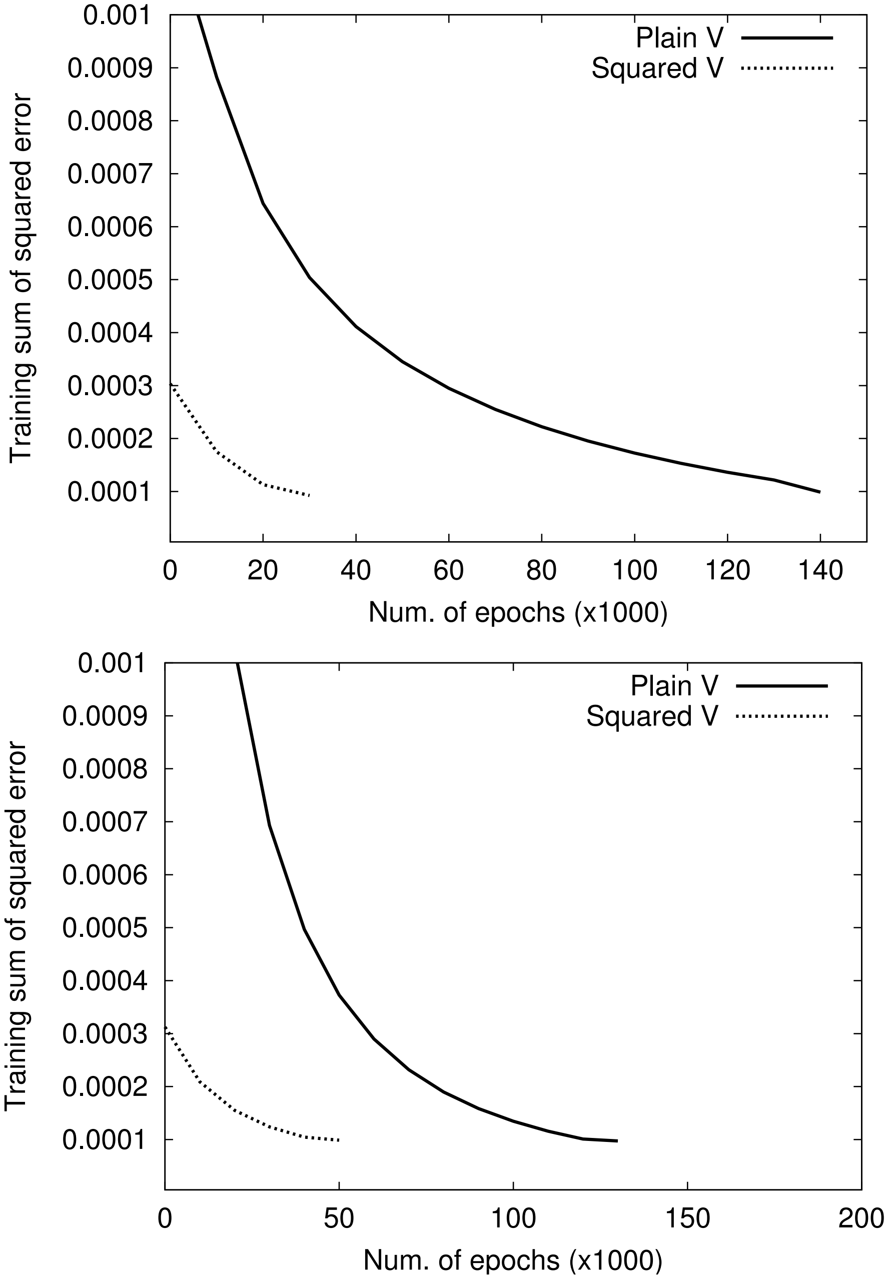


Fig. 25. Graphs showing the learning efficiency for two training instances with different inputs, with both the plain orientation energy input and its squared form. The graphs clearly show that the efficiency is higher for the squared orientation energy input to learn the threshold value.

be employed in the spatial frequency domain as well. In general, such a method of detection of salient features could be extended to any other modality where a similar response distribution can be found.

D. Local vs. Global Thresholding

A conclusion that came out of the quantitative analysis experiments conducted in the previous chapter to compare the different thresholding methods using synthetic images was that there was not much to choose between the local and global OED-derived thresholding methods. The statistical paired t-test also failed to show a clear difference between the effectiveness of the two methods. As was discussed toward the later part of that section, after analyzing the input, it was found that the two methods depended on certain properties of the input. The global thresholding method works best when there are not many variations of input features, while the local thresholding method works well for the opposite case. This was demonstrated both quantitatively and qualitatively, through the experiments with synthetic images and natural images respectively. If there are large variations in the input, the σ values of local patches would also vary widely, and thus appropriate salient edges could be extracted more efficiently using the local thresholding method.

It would be interesting to see if there are any other factors involved other than input scene variations that would help explain the inconclusive results between the local and global OED-derived thresholding methods. More experiments could be conducted using synthetic images in which part of the images could have wide variations and the other parts could have constant variation, to see if the global and local thresholding methods both give similar results.

E. Comparison of Local Thresholding with Psychophysical Data

Another interesting line of work that could be considered is to extend the psychophysical experiments for the local OED-derived thresholding method. In particular, local thresholds could be calculated for all local patches in images using the linear equation for the threshold. The set of human-chosen thresholds could also be found out by obtaining preferences from a human observer as to the best thresholds for each local input patch. These two sets of data could then be compared against each other similar to that for the global method to see if there is a linear association between them. This could then establish the generality and basis for the OED-derived methods.

CHAPTER VII

CONCLUSION

The main contribution of the thesis is the investigation of how salient edge features are processed and derived by the biological visual system. I have explored the utilization of visual system response properties through efficient methods for salient contour detection in natural images. Through various psychophysical and quantitative experiments, I have shown that an adaptive thresholding mechanism that compares the response distribution to the normal distribution, serves to be quite effective in detecting salient contours. I have also suggested a possible justification for selecting the normal distribution as the baseline for the comparison, by the relationship with the concept of Suspicious Coincidence. The results of experiments with white-noise based distributions also help reinforce the validity of the normal distribution as the baseline. Furthermore, the linear thresholds obtained by comparing the orientation energy distribution to the Gaussian baseline were found to have a strong linear relationship with the square root of the raw second moment of the orientation energy. I have then suggested a neural implementation that utilizes the squared response to calculate the raw second moment of the orientation energy and showed that such a kind of representation is quite efficient to learn the salience thresholds as well.

Such a method of utilizing representations of response histograms can be extended to other sensory modalities, where a similar response distribution is found. The psychophysical experiments conducted in this thesis for the global thresholding methods could be extended by gathering human preference for local thresholds in images. I am confident that the ideas in this thesis would be of help to researchers to conduct more insightful research into understanding salient contour detection and other low-level visual processes.

REFERENCES

[1] H. Barlow, “Unsupervised learning,” *Neural Computation*, vol. 1, pp. 295–311, 1989.

[2] ———, “Possible principles underlying the transformation of sensory messages,” in *Sensory Communication* . W.A. Rosenblith, Ed., Cambridge, MA: MIT Press, 1961, pp. 217–34.

[3] ———, “What is the computational goal of the neocortex?,” in *Large Scale Neuronal Theories of the Brain* . C. Koch and J.L. Davis, Eds., Cambridge, MA: MIT Press, 1994, pp. 1–22.

[4] R.W. Buccigrossi and E.P. Simoncelli, “Image compression via joint statistical characterization in the wavelet domain,” *IEEE Transactions on Image Processing*, vol. 8, pp. 1688–1701, 1999.

[5] V.A. Casagrande and T.T. Norton, “Lateral geniculate nucleus: A review of its physiology and function,” in *The Neural Basis of Visual Function* . A.G. Leventhal, Ed., Boca Raton, FL: CRC Press, 1989, pp. 41–84.

[6] Y. Choe and R. Miikkulainen, “Contour integration and segmentation in a self-organizing map of spiking neurons,” *Biological Cybernetics*, 2003, in press.

[7] Y. Choe and S. Sarma, “Relationship between suspicious coincidence in natural images and oriented filter response distributions,” Technical Report, Department of Computer Science, Texas A&M University, College Station, Texas, August 2003.

[8] C.Y. Chou and H.R. Liu, “Properties of the half-normal distribution and its application to quality control,” *Journal of Industrial Technology*, vol. 14, pp. 4–7, 1998.

[9] J.G. Daugman, “Two-dimensional spectral analysis of cortical receptive field profiles,” *Vision Research*, vol. 20, pp. 847–856, 1980.

[10] ——, “Entropy reduction and decorrelation in visual coding by oriented neural receptive fields,” *IEEE Transactions on Biomedical Engineering*, vol. 36, pp. 107–114, 1989.

[11] D.J. Field, “Relations between the statistics of natural images and the response properties of cortical cells,” *Journal of the Optical Society of America A*, vol. 4, pp. 2379–2394, 1987.

[12] P. Gastaldo, R. Zunino, E. Vicario and I. Heynderickx, “CBP neural network for objective assessment of image quality,” in *Proceedings of the International Joint Conference on Neural Networks*, IEEE, 2003, in press.

[13] W.S. Geisler, J.S. Perry, B.J. Super, and D.P. Gallogly, “Edge Co-occurrence in natural images predicts contour grouping performance,” *Vision Research*, vol. 41, pp. 711–724, 2001.

[14] M.H. Hansen and B. Yu, “Wavelet thresholding via MDL: Simultaneous denoising and compression,” *IEEE Transactions on Information Theory*, vol. 45, pp. 1778–1788, 2000.

[15] D.J. Heeger, “Half squaring in responses of cat simple cells,” *Visual Neuroscience*, vol. 9, pp. 427–443, 1992.

[16] P.O. Hoyer, “Probabilistic models of early vision,” PhD dissertation, Helsinki University of Technology, Helsinki, Finland, 2002.

[17] J. Huang and D. Mumford, “Statistics of natural images and models,” *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1541–1547, 1999.

[18] D.H. Hubel, *Eye, brain and vision*. New York: W.H. Freeman and Co., Scientific American Library Series, 22nd edition, 1988.

[19] J. Hurri and A. Hyvarinen. “Simple-cell-like receptive fields maximize temporal coherence in natural video,” *Neural Computation*, vol. 15, pp. 663–691, 2003.

[20] L. Itti, J. Braun, D.K. Lee, and C. Koch. “A model of early visual processing,” *Advances in Neural Information Processing Systems*, vol. 10, pp. 173–179, 1998.

[21] L. Itti and C. Koch, “A saliency-based search mechanism for overt and covert shifts of visual attention,” *Vision Research*, Vol. 40, pp. 1489–1506, 2000.

[22] J.P. Jones and L.A. Palmer, “An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex,” *Neurophysiol.*, vol. 58(6), pp. 1233–1258, 1987.

[23] E. Kaplan, “The receptive field structure of retinal ganglion cells in cat and monkey,” in *The Neural Basis of Visual Function*. A.G. Leventhal, Ed., Boca Raton, FL: CRC Press, 1989, pp. 10–40.

[24] A.B. Lee, D. Mumford, and J. Huang, “Occlusion models for natural images: a statistical study of a scale-invariant dead leaves model ,” *International Journal of Computer Vision* , vol. 41, pp. 35–59, 2001.

[25] H.-C. Lee and Y. Choe, "Detecting salient contours using orientation energy distribution," in *Proceedings of the International Joint Conference on Neural Networks*, IEEE, 2003, in press.

[26] X. Liu, "Computational Investigation of Feature Extraction and Image Organization," PhD dissertation, Ohio State University, Columbus, Ohio, 2000.

[27] X. Liu and D. Wang, "A spectral histogram model for texton modeling and texture discrimination," *Vision Research*, vol. 42, pp. 2617–2634, 2002.

[28] S. Marcelja, "Mathematical description of the response of simple cortical cells," *Journal of Optical Society of America A*, vol. 70(11), pp. 1297–1300, 1980.

[29] K.D. Miller, "Receptive fields and maps in the visual cortex: models of ocular dominance and orientation columns," in *Models of Neural Networks III*. E. Domany, L.L. van Hemmen, and K. Schulten, Eds., New York, NY: Springer, 1996, pp. 55–78.

[30] D.L. Ruderman, "Origins of scaling in natural images," *Vision Research*, vol. 37, pp. 814–817, 1996.

[31] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*. D.E. Rumelhart and J.L. McClelland, Eds., Cambridge, MA: MIT Press, 1986, pp. 318–362.

[32] E.P. Simoncelli and E.H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proceedings of IEEE International Conference on Image Processing*, vol. I, pp. 379–382, 1996.

[33] E.P. Simoncelli and B.A. Olshausen, “Natural image statistics and neural representation,” *Annual Reviews of Neuroscience*, vol. 24, pp. 1193–1216, 2001.

[34] S. Zhu, Y. Wu, and D. Mumford, “Filters, random fields and maximum entropy(FRAME) -towards the unified theory for texture modeling,” *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 686–693, 1996.

VITA

Subramonia Sarma was born in Trivandrum, the capital city of Kerala State, in India on December 1st, 1978, the son of A. Padmanabhan and M. Vijayalakshmy. After graduating from a prominent high school in Trivandrum, India, he went on to pursue his undergraduate studies at the College of Engineering, Trivandrum, India which is affiliated with the University of Kerala. After graduating with a Bachelor of Technology in computer science in May 2000, he worked in the software industry for a year. In the fall of 2001 he entered the Department of Computer Science at Texas A&M University, College Station, Texas to pursue the Master of Science degree in computer science. He worked as an intern at Advanced Micro Devices, Austin, TX for two semesters during his graduate studies.

Permanent Address:

H:No: 109, Sankar Nagar,
Neeramankara,
Kaimanam P.O,
Trivandrum,
Kerala State, India
PIN 695 040

The typist for this thesis was Subramonia Sarma.