Perceptual Picture Emphasis Using Texture Power Maps

by

Sara Lee Su

B.S. Computer Engineering
University of Washington, 2002

Submitted to the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
[February 2005]
January 2005

Author .....
Department of Electrical Engineering and Computer Science
January 31, 2005

Certified by ....
Frédo Durand
Assistant Professor
Thesis Supervisor

Accepted by.....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Perceptual Picture Emphasis Using Texture Power Maps

by

Sara Lee Su

Submitted to the Department of Electrical Engineering and Computer Science
on January 31, 2005, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

Applying selective emphasis to photographs is a critical aspect of the visual design process. There is evidence from psychophysics that contrast in texture is a key contributor to saliency in an image, yet unlike other low-level perceptual features, texture cannot be directly modified with existing image-processing software. We present a post-processing technique to subtly change the salience of regions of an image by modifying spatial variation in texture. Our method is inspired by computational models of visual attention that capture sensitivity to outliers in local feature distributions. We use the steerable pyramid, which encodes multiscale oriented image features and compute a set of power maps which capture the local texture content at each scale and orientation. With this representation, texture variation can be modified to selectively add or remove emphasis in the image. Two user studies provide qualitative and quantitative psychophysical validation of our approach.

Thesis Supervisor: Frédo Durand
Title: Assistant Professor

Acknowledgments

Many people have helped make this thesis possible.

First and foremost, I thank Professor Frédo Durand for encouraging my interest in computer graphics and for teaching me how to choose and solve worthwhile problems in this field. He and Maneesh Agrawala of Microsoft Research were instrumental in forumlating key ideas in this thesis and preparing the conference submission version of the text. Their guidance has been invaluable, and without it this thesis would not exist.

Thanks to Doctors Agrawala, David Salesin, and Patrice Simard for hosting me at Microsoft Research during the early stages of this project, and to Professors Durand, Jovan Popović, Seth Teller, and Victor Zue for making me feel welcome at MIT.

This work was generously funded by an NSF Graduate Research Fellowship and an MIT Presidential Fellowship.

Professor Aude Oliva kindly allowed me the use of her eyetracker, without which the user study would not have been possible. Thanks to Michelle Greene, Barbara Hidalgo-Sotelo, and Naomi Kenner for teaching me how to use it.

Paul Green provided invaluable assistance with data acquisition and analysis for the user study. Thanks to Paul, Eric Chan, Barbara Cutler, Sylvain Paris, and Peter Sand for proofreading various incarnations of this text.

Thanks to members of the Graphics and Vision groups for volunteering to be test subjects and for making the lab a fun place to work. They and my other terrific friends have helped me ride out the ups and downs of grad school.

Finally, I thank my parents Bernard and Patricia, brother Jonathan, grandparents, and extended family for their unconditional love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Visual attention plays an important role in the understanding of scenes and sequential concepts. An important task of the human visual system is to sift through complex visual stimuli to find objects or regions of interest for further processing and attention. The ability to direct attention to important subsets of that stimuli is a key component of cognition.

Important steps in the visual design process are creating emphasis and directing the viewer's attention to the relevant parts of an image in order to create a mood, tell a story and make information easier to process [Tuf90, Zek99, TMB02]. This aspect is particularly important for educational materials. The order in which viewers scan an image determines the mental representation formed from it [BTss]. Viewers of images often don't know where to look, in contrast to reading, where convention dictates left-to-right, top-to-bottom scanning. The eye is guided by what is visually salient, for example, areas of high contrast such as edges and corners.

Enhancing images using emphasis devices can provide paths for viewers to "read" them. Cognitive studies have shown that image comprehension is significantly improved when a subject is provided with visual cues to guide the gaze path to semantically important regions [TMB02]. In technical illustrations, highlighting and colorization are commonly used to guide viewers through a sequence of instructions or to visually explain the operation of a complex device [Mar89, BRT95]. In paintings and educational diagrams, dramatic techniques such as circling or strong color variations can be used to put create focal points. However, photorealistic styles and photography require more subtle means that maintain

the "naturalness" of the image while appropriately directing the gaze of the viewer.

Much of the art of photography involves creating emphasis and directing the viewer's attention to the relevant parts of an image. When possible, a professional photographer will carefully compose and light a scene before taking a picture to ensure that the subject stands out from the background. However, in many real-world settings, such control over composition and lighting is not possible or not sufficient. Over the years, photographers have developed a variety of post-processing techniques, first in the darkroom and now using a computer, to emphasize important image elements.

Many common post-processing emphasis techniques involve selectively altering low-level perceptual features such as sharpness, brightness, chromaticity, or saturation [War00, Eis04]. The human visual system is particularly attuned to differences in the low-level features manipulated by these techniques.

Human visual attention is driven by a combination of top-down and bottom-up processes. Top-down mechanisms describe how attention is influenced by scene semantics (such as the presence of human faces) and by the task presented to the viewer. Top-down processes are important to understanding attention, however, in this thesis, we focus on image processing techniques that are independent of content.

Bottom-up processes describe the effect of low-level properties of visual stimuli on attention. In a nutshell, the human eye is attracted to change and contrast. Computational models of visual attention shed important light on these processes and provide the inspiration for our work. Several influential models have been developed for predicting the likelihood of a given region of an image to attract visual attention by identifying large differences and outliers in these bottom-up feature distributions [Ros99, IK01, PN04]. By selectively increasing or decreasing one or more of these features in post-processing, photographers adjust the *salience* of chosen regions and guide viewers' gaze through their images [Sol96, Zek99].

Surprisingly, there is one low-level feature that cannot be directly manipulated with existing image-editing software: *texture variation*. Contrast in texture is cited by many visual design and photography manuals as a key element of emphasis in a picture. From a perceptual standpoint, variations and outliers in texture are salient to the human visual

system, and the human and computer vision literature show that discontinuities in texture can elicit an edge perception similar to that triggered by color discontinuities [BPR81, MP90, War00, MBLS01]. We have developed a new technique for selectively altering texture variation to redirect attention in an image.

We make the following contributions in this thesis.

Image manipulation with power maps The texture discrimination tools we call power maps have been heavily used in image analysis. Our work introduces their application to image editing. We show how power maps capture local texture content in an image and provide a powerful representation for manipulating frequency content.

Image emphasis through texture variation. We introduce a perceptually-motivated technique for selective manipulation of texture variation (local frequency content). Our method is complementary to traditional post-processing image emphasis techniques (sharpening and brightening, for example), and we show how it can be applied in instances where traditional techniques may look unnatural.

User study. We conduct two user studies as experimental validation of our technique's effectiveness. Qualitative changes in user fixations on original and modified images are provided using an eyetracker. In addition to showing quantitatively that our image manipulation technique is effective for emphasis and de-emphasis, a visual search experiment verifies the hypothesis that texture is a salient image feature.

1.1 Related work

Human visual attention is driven by a combination of top-down and bottom-up processes. Top-down mechanisms describe how attention is influenced by scene semantics (such as the presence of human faces) and by the task presented to the viewer. Top-down processes are important to understanding attention, however, in this thesis, we focus on image processing techniques that are independent of content.

Bottom-up processes describe the effect of *low-level features* of visual stimuli on attention. The human visual system is neuronally turned to changes and contrast and contrast

in color, contrast, and orientation [Pal99]. Computational models of visual attention shed important light on these processes and provide the inspiration for our work. Several influential models have been developed for predicting the likelihood of a given region of an image to attract visual attention by identifying large differences and outliers in these bottom-up feature distributions [Ros99, IK01, PN04]. By increasing or decreasing the the presence of outliers or large variations in the feature distribution in post-processing, photographers adjust the *salience* of chosen regions and guide viewers' gaze through their images [Sol96, Zek99]. In Section 2.2, we discuss traditional emphasis techniques fitting this model.

1.1.1 Texture

The effects that the low-level visual features listed above have on attention are straightforward. Ware [War00], Parkhurst and Niebur [PN04], Healey [HTER04], and others have shown that contrasts in the second-order feature of texture can also be effectively exploited to guide visual attention.

In computer graphics, Interrante [Int00] has used oriented textures to convey shape, and Healey [HTER04] has explored the use of texture in information visualization by modifying its perceptually-significant characteristics of scale, orientation, and contrast. We recommend the latter paper for its excellent survey of using first- and second-order features to direct attention.

1.1.2 Applied visual attention in computer graphics

The variable resolution of the retina can be exploited to simplify parts of an image outside of the fovea. This idea has been explored using computational prediction of salient regions and has been applied to a number of computer graphics problems including global illumination [YPG01], automatic image cropping [CXF+02, SLBJ03], and animation [PO03]. These approaches combine bottom-up saliency metrics with additional top-level information such as face detection.

Santella et al. [SD04] have evaluated the effect of non-photorealistic image abstrac-

tion on visual attention. They used physical eye-tracking measurements to determine regions of interest in photographs. This information then modulates such properties as stroke density in the painterly filtering of the photograph [DS02, SD02]. They found that their eye-tracking-driven technique performed more effective abstraction than non-perceptually-based methods.

Our work is related to these approaches, however we focus on photographs and photo-realistic styles. In addition, while previous techniques exploited the natural saliency present in images and reinforced it through abstraction [DS02, SD02], our method can also be used to re-direct attention to regions of the image that are not salient in the original.

1.2 Overview

Our method is based on perceptual models of attention that hypothesize that salience can be created by contrast in texture. We first review the filter-based model of texture discrimination that forms the theoretical basis of our work and the computational models of visual attention based on it (Sections 2.3 and 2.4). We review the steerable pyramid representation (Section 2.5) before discussing how it allows us to to analyze and manipulate multiscale oriented features and texture variation using power maps (Chapter 3). Images enhanced using our technique are presented in Chatper 4. We present experimental validation of our technique's effectiveness in Chapter 5 and conclude in Chapter 6 with a discussion of our findings and directions for future work.

Chapter 2

Background

2.1 Perception and saliency

In order to attract or redirect attention in an image, we must consider how the human visual system works. We first review the basics of bottom-up visual attention and the computational models that have been developed to capture these processes. This background provides a unifying framework for the traditional post-processing techniques for modifying saliency and our new texture-manipulation emphasis technique (e.g. sharpening, brightening).

Experiments have shown that visual neurons are tuned to respond to specific classes of stimuli ranging from low-level attributes (intensity contrast, color, direction of motion) to specialized, high-level cues (corners, shape-from-shading) [IK01]. Visual attention is driven by a combination of these top-down and bottom-up processes. Top-down mechanisms describe how attention is influenced by semantic content (such as human faces) and by the task presented to the viewer.

In contrast, bottom-up processes describe the effect of low-level properties of visual stimuli on visual attention. In an early paper on sensory messages, Barlow noted that frogs' eyes contain specialized "fly detector" cells that are sensitive to small, dark objects moving against a light background [Bar61]. Recent studies have suggested that human visual attention is similarly based on *neuronal tuning* for such *low-level features* in the field of view [IK01]. In a nutshell, the human eye is attracted to changes and contrast.

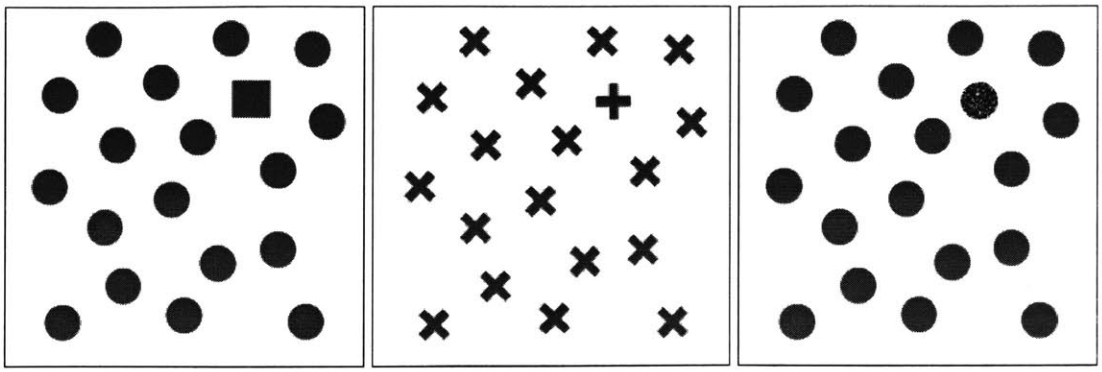


Figure 2-1: Three simple images containing salient, low-level features triggering pop-out phenomena.

The bottom-up view of visual attention theorizes that the human eye first processes an entire scene in parallel at low resolution. This is the first stage at which simple visual features (an area of high intensity contrast, for example) may "pop out" [Ros99]. Figure 2-1 illustrates three types of pop-out phenomena. As attention is disengaged from the foveal region, the eyes move to focus on the first region of interest. The fovea finally focuses on the region of interest for further inspection at high resolution [Duc03]. Computational models of visual attention shed important light on these processes and provide the inspiration for our work.

2.2 Traditional emphasis techniques

As part of an ongoing study of computer depiction and non-photorealistic rendering, we have identified a number of pictorial techniques employed by artists to add emphasis in photographs and illustrations.

Contrast and brightness.

Color highlighting. Color variation, desaturation.

Blurring and sharpening.

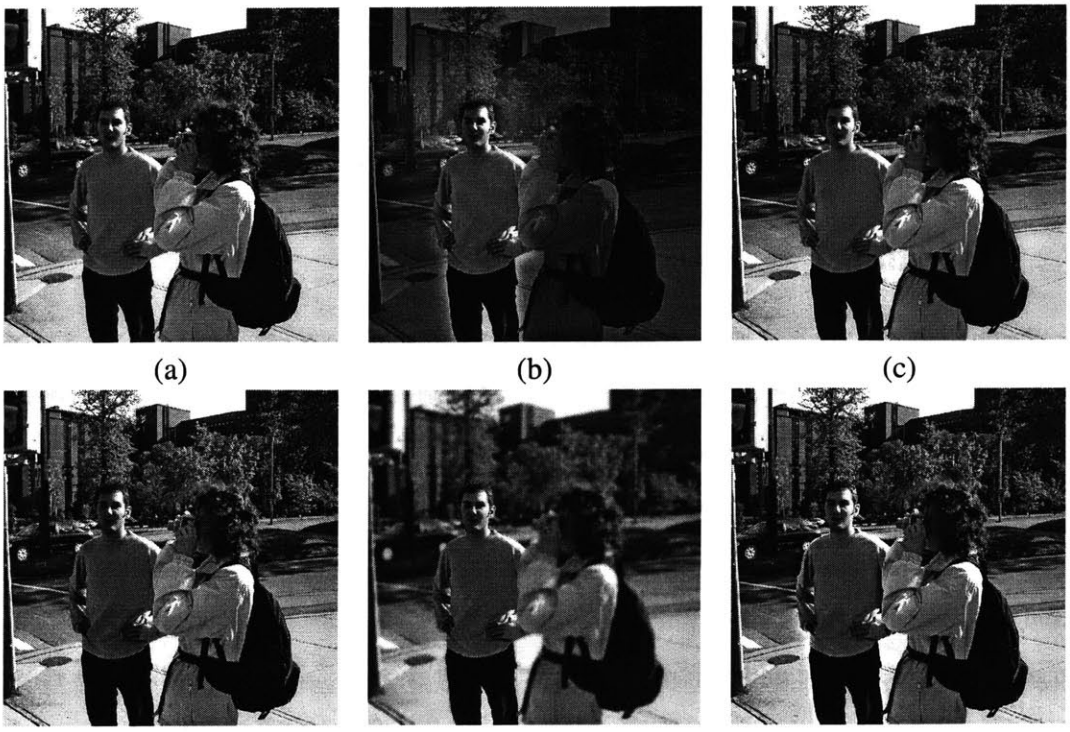


Figure 2-2: Limitations of traditional emphasis techniques. (a) Original image. (b) Vignetting. (c) Adjusting saturation. (d) Adjusting chromaticity. (e) Gaussian blur. (f) Adjusting edge contrast.

Haloing and outlining. Contrast with color and saturation of surrounding objects, contrasting shell edges, fake drop shadows.

Gradient. Brightness, transparency, ghosting, shape and sharpness of edges.

Non-linear scaling and exaggeration.

Levels of detail, simplification. Line density, Tonal modeling.

Metagraphic elements. Arrows, labels, guidelines.

Insets and callouts. Zooming, cropping.

Viewpoint choice. Occluding irrelevant details.

Lines. Style, converging to a point, for physical reference, conveying shape, conveying shading.

These pictorial devices guide the eye by artificially bringing important regions to the foreground and pushing others to the background. Our long-term goal is a more complete formal analysis of a taxonomy of photorealistic and non-photorealistic emphasis techniques.

2.2.1 Limitation of traditional techniques

Traditional post-processing techniques for manipulating saliency increase or decrease sharpness, brightness, chromaticity, or saturation. These techniques directly alter low-level bottom-up features that affect visual attention.

The technique of vignetting, for example, uses a smooth gradient to make the center of attention brighter than the rest of the image, thereby increasing the brightness feature and local contrast. In some cases, however, vignetting has the undesirable effect of darkening large parts of the image and lowering overall contrast (Figure 2-2(b)). Similarly, increasing saturation in a target region draws attention to it, but the effect may look unnatural when it conflicts with viewer expectations about the appearance of familiar objects such as people (Figure 2-2(c)). Conflict with viewer expectations also makes changing chromaticity an undesirable choice when photorealism is important (Figure 2-2(d)).

Another common technique is selective sharpening and blurring of an image to simulate depth-of-field effects This method works well when the emphasized object lies at a different distance than the rest of the scene. However, the resulting image looks unnatural when only one object is modified in a group at a given distance (Figure 2-2(e)); the depth cue introduced by the selective blurring conflicts with the viewer's understanding of the scene.

Similarly, if the object to be emphasized is isolated in front of the background, contrast at the occluding edges can be increased to reinforce figure-ground separation. However, this approach works only when the part to emphasized is separated from a background, and it does not work if part of a continuous object is to be emphasized (Figure 2-2(f)).

Though these techniques all have their limitations, experienced photographers know how to use them in combination to achieve the desired spatial saliency in their images. The techniques draw their effectiveness from the direct manipulation of low-level features and the increase of feature-response variation in regions of an image. Our work shows

how a complementary higher-order feature, texture, can be altered to selectively increase or decrease salience.

2.3 Texture segmentation and discrimination

The ability of humans to perceive texture edges and discriminate between two different textures has received much attention in computational and human vision [BPR81, MP90, MBLS01, LG04]. Precisely defining and representing texture are open problems in vision, but most researchers agree that texture is related to local frequency content. Computational approaches to texture segmentation and texture edge detection have computed local variations in frequency content to detect texture edges. Most approaches roughly follow Malik and Perona's biologically-inspired model of texture discrimination [MP90] which represents texture as a response to a set of multi-scale oriented filters and predicts the salience of the boundary between two different textures. We illustrate this technique with a 1D example (Figure 2-3). The overall principle follows that of edge detection but is applied to local averages of the responses to multiscale-oriented filters rather than to the image intensity.

The first stage of most texture discrimination models is linear filtering with multi-scale oriented Gabor-like functions (Figure 2-3(b)). Note how the response to such filters captures differences in texture frequency and amplitude. However, because the response contains both positive and negative lobes, the response to such a filter averaged over a small neighborhood is usually zero. The signal must be rectified to the unsigned magnitude of the response in order to be a meaningful measure of per-pixel power for each filter. Possible non-linearities include full-wave rectification (absolute value) and energy computation (square response); the absolute value is shown in Figure 2-3(c).

The rectified filter responses encode the per-pixel power of each oriented multi-scale filter (Figure 2-3(c)). These signals still contain many oscillations. Applying a low-pass filter produces the local average of the filter response strength (Figure 2-3(d)). We call this result the *power map*.

As suggested by Northdurft [Nor85], an analysis similar to intensity images can then be performed on these power maps at each scale and orientation. While most work on

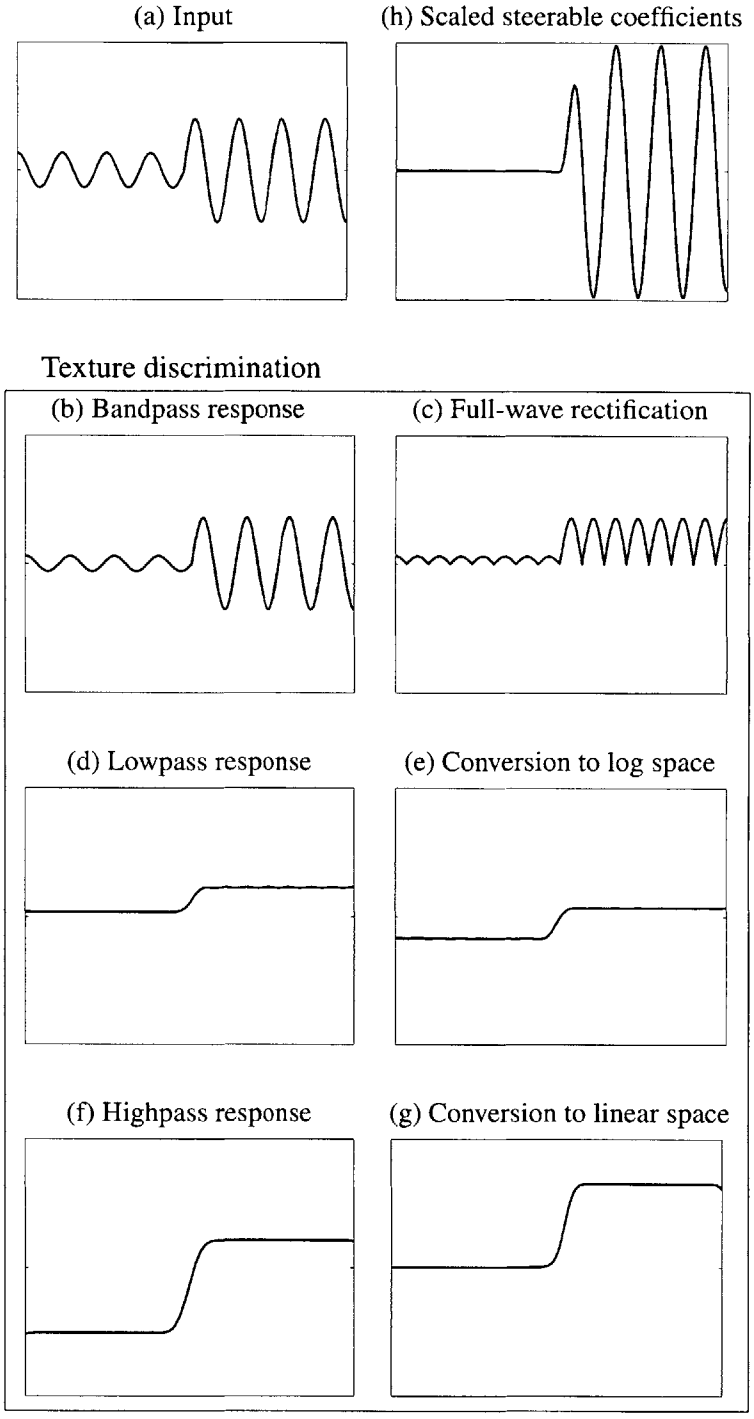


Figure 2-3: Texture discrimination and manipulation in 1D.

computational and machine vision has focused on edge detection and image segmentation, in this paper, we show how power maps can be applied to image manipulation and picture emphasis.

This approach to texture discrimination has also inspired texture synthesis methods that match histograms of filter responses [HB95], but to the best of our knowledge, it has not been applied to photo editing. While more elaborate texture representations based on it have been presented (e.g. [PS00a]), we can rely on this simpler representation because our goal is not to synthesize or recognize texture, but to subtly alter texture variation. This filter-based texture representation naturally affords image manipulation. We will use the ideas of filtering, non-linearity, and pooling to represent texture and alter its spatial variation in an image.

2.4 Computational models of visual attention

Regions of an image are characterized by a set of low-level features such as color, contrast, and orientation at multiple scales that correspond to the early stages of the human visual system [Pal99, IK01]. A number of influential models of attention have identified *salient* objects as statistical outliers in low-level feature distributions [KU85, IKN98, PS00b, RZ99]. Intuitively, a region is salient if it differs greatly from surrounding regions in one or more feature channels. Most computational models focus on the response to filter banks such as Laplacian pyramids or Gabor wavelets that extract contrast and orientation in the image. Various non-linearities are then be used to extract and combine maxima of the response to each feature. These *first-order* salience models deal directly with low-level features such as contrast, color, and orientation. Sharpening or blurring an image or editing its local contrast directly modifies these features, thereby modifying local saliency.

2.4.1 Itti-Koch model

Based on a biologically plausible model proposed by Koch and Ullman [KU85], the bottom-up computational attention model of Itti et al. attempts to simulate feature-specific neuronal responses with *center-surround* filtering of an input image [IKN98]. Their physiological

experiments suggest that the brain processes visual information at multiple levels of detail in feature-specific channels. They focus on three early visual features that, biologically, are computed in parallel across the entire visual field: intensity contrast, color double-opponency [EZW97], and orientation contrast.

To simulate these early visual processes, the Itti model takes as input an image and outputs a *saliency map*, a spatial representation of the saliency of every location in the visual field as a combination of the *conspicuity measures* for each channel (Figure 2-4). We review the basics of the model here and refer the reader to the above references for further detail.[1]

The input image is first decomposed into a nine-scale dyadic Gaussian pyramid [GBP+94] for each of the three conspicuity channels: intensity, color, and orientation. A set of *feature maps* $\mathcal{M}_i$ is generated for each channel, each computed as the *center-surround difference* between two levels of the corresponding pyramid.

The center-surround computation, a difference-of-Gaussians filter, takes the across-scale difference between two scales: the finer, excitatory center I_c and the coarser, inhibitory surround I_s, where scale $c \in \{2,3,4\}$ and $s = c + \delta, \delta \in \{3,4\}$. The across-scale difference is simply a pixel-by-pixel difference between I_c and the interpolation of I_s to scale c.

In practice, 42 feature maps are computed for an image: 6 for intensity, 12 for color (6 each for red/green and blue/yellow chromatic opponency), and 24 for orientation (6 each for the preferred orientations 0°, 45°, 90°, and 135°). The combination results in the saliency map for the image. The architecture of the model is illustrated in Figure 2-4.

2.4.2 Parkhurst-Nieber model

Recently, Parkhurst and Niebur [PN04] presented a model of saliency that captures texture variation in order to explain psychophysical experiments by Einhäuser and König [EK03] who reported salience effects that could not be explained by first-order models. Their *second-order* model performs additional image processing on the response to a first-order

[1] Itti et al. have made a C++ implementation of their model of bottom-up visual attention available for non-commercial use at `http://ilab.usc.edu/toolkit`.

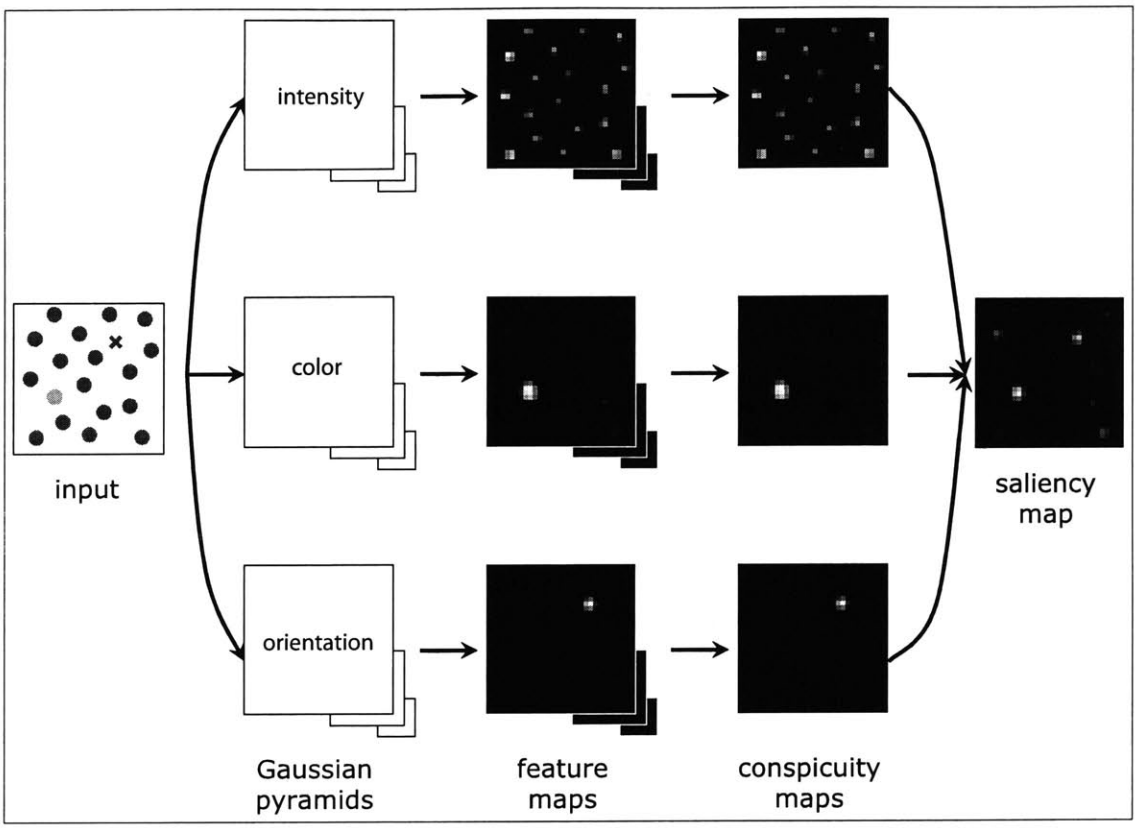


Figure 2-4: Architecture of Itti-Koch bottom-up computational model of human visual attention.

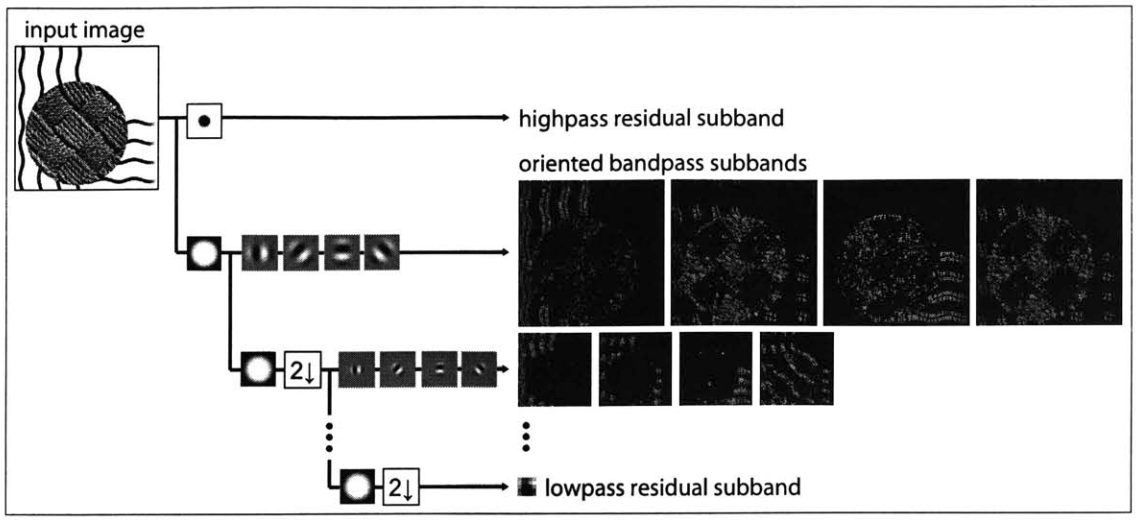


Figure 2-5: System diagram illustrating the process of building a steerable pyramid. The input image is first filtered into a highpass residual subband and a lowpass subband. The lowpass subband is filtered into a set of 4 oriented bandpass subbands and a lower-pass subband which is then bilinearly downsampled. The filtering and downsampling process is repeated. For visualization purposes, the bandpass subbands are displayed as false-color images in which positive coefficients are red, negative blue, and zero black. Note the clear oriented responses to the input image's weave pattern and wavy lines.

filter bank effectively performing the same computation as first-order models but on what we call power maps (Chapter 3) rather than on image intensity[2]. This motivates our strategy of performing image manipulations on power maps in order to modify contrast in texture.

Although it is beyond of the scope of this paper, we recommend the discussion in Parkhurst and Niebur's article [PN04] because it highlights the challenges of designing psychophysics experiments, as well as the opportunities provided by computer graphics and image processing to validate perceptual hypotheses through image manipulation.

2.5 Steerable pyramids

In Section 2.3, we reviewed how the response to multiscale oriented filters can be used for texture discrimination. A plethora of such filters has been developed, and in our work we

[2]Unfortunately, the term '*n*th-order' has a number of definitions related to texture models, including *n*th-order statistics and the one described by Landy [LG04].

use *steerable pyramids* [FA91, SF95] because they permit not only the analysis of images, but also reconstruction and processing.

The multi-scale, multi-orientation steerable pyramid, which uses an overcomplete wavelet transform, is a perceptually meaningful image representation. Because real-world scenes contain features of varying size and distance from the viewer, it is insufficient to consider an image as simply an array of pixel intensities to be processed at a single scale. Psychophysical experiments suggest that the visual cortex processes scenes using orientation- and spatial-frequency-selective filters [WB79]. The low-level features that we use as the basis of our texture model correspond to oriented multi-scale filters encoded in an image pyramid [AAB+84, OABB85].

The steerable pyramid can be seen as an extension of the Laplacian pyramid that also encodes orientation. The corresponding filters are similar to Gabor wavelets and model the response of cells in the early stages of the human visual system. We review the basics of steerable pyramids here and refer the reader to the above references for further detail[3].

The steerable pyramid has the desirable properties of near-perfect image reconstruction and encoded orientation information about the original image and has been used for image denoising [PSWS03, SA96], texture synthesis, [HB95, PS00a, BJEYLW01] and orientation analysis [SF96], among other applications.

The steerable pyramid transform decomposes an image into a set of oriented bandpass images. Each level, or *subband*, of the pyramid is computed from the previous level by convolving with a bank of linear filters and subsampling. For each scale, a set of subbands is constructed, one corresponding to each orientation.

Figure 2-5 shows the decomposition of a simple image into three-scale, four-orientation steerable pyramid. The image is first filtered into a highpass residual subband and a low-pass subband. The lowpass subband is further filtered into a set of four oriented bandpass subbands and one lowpass subband. The lower-pass subband is then bilinearly subsampled by a factor of two. This filtering and subsampling is repeated on successively lower-pass

[3]In addition, Simoncelli et al. have made a steerable pyramid software package available for non-commercial use at `http://www.cns.nyu.edu/~eero/STEERPYR`. We have used this Matlab code in the implementation of our algorithms.

subbands to generate the multi-scale pyramid.

The signed coefficients of each bandpass subband correspond to the response to the scaled oriented filters shown in the figure. For visualization purposes, the bandpass subbands are displayed as false-color images in which positive coefficients are colored red, negative blue, and zero black.

Orientation is a key feature encoded in the steerable pyramid. Because we use Malik and Perona's notion of texture edges [MP90], we consider orientation to be an important discriminant in our analysis. Consider the weave pattern in the input image of Figure 2-5. Although the textures of the individual woven pieces are identical except for orientation, we clearly perceive texture edges between them. Two non-oriented residual subbands are also computed for the pyramid to encode the lowest and highest frequencies of the image for which oriented information cannot be accurately derived.

The steerable pyramid is *self-inverting*; the analysis filters are also used for synthesis. This property of symmetry is important for image manipulation because it guarantees consistency between the analysis and synthesis stages, in contrast to Laplacian pyramids for which the analysis filters are sharper than the synthesis filters. However, the steerable pyramid is overcomplete; that is, it contains more coefficients than the original number of pixels. While this means that the decomposition is non-orthogonal, overcompleteness prevents aliasing within subbands.

Chapter 3

Texture variation with power maps

We have developed a post-processing technique to emphasize or de-emphasize regions of a photograph by modifying contrast in texture. Informally, our goal is to invert the outlier-based computational model of saliency. Recall that this model defines salient regions as outliers from the local feature distribution. Our technique modifies the power maps described in the previous section to increase or decrease spatial variation of texture, as captured by the response to steerable oriented filters. The following pseudocode summarizes our approach to capturing and manipulating texture variation in 2D.

3.1 Power maps to capture local energy

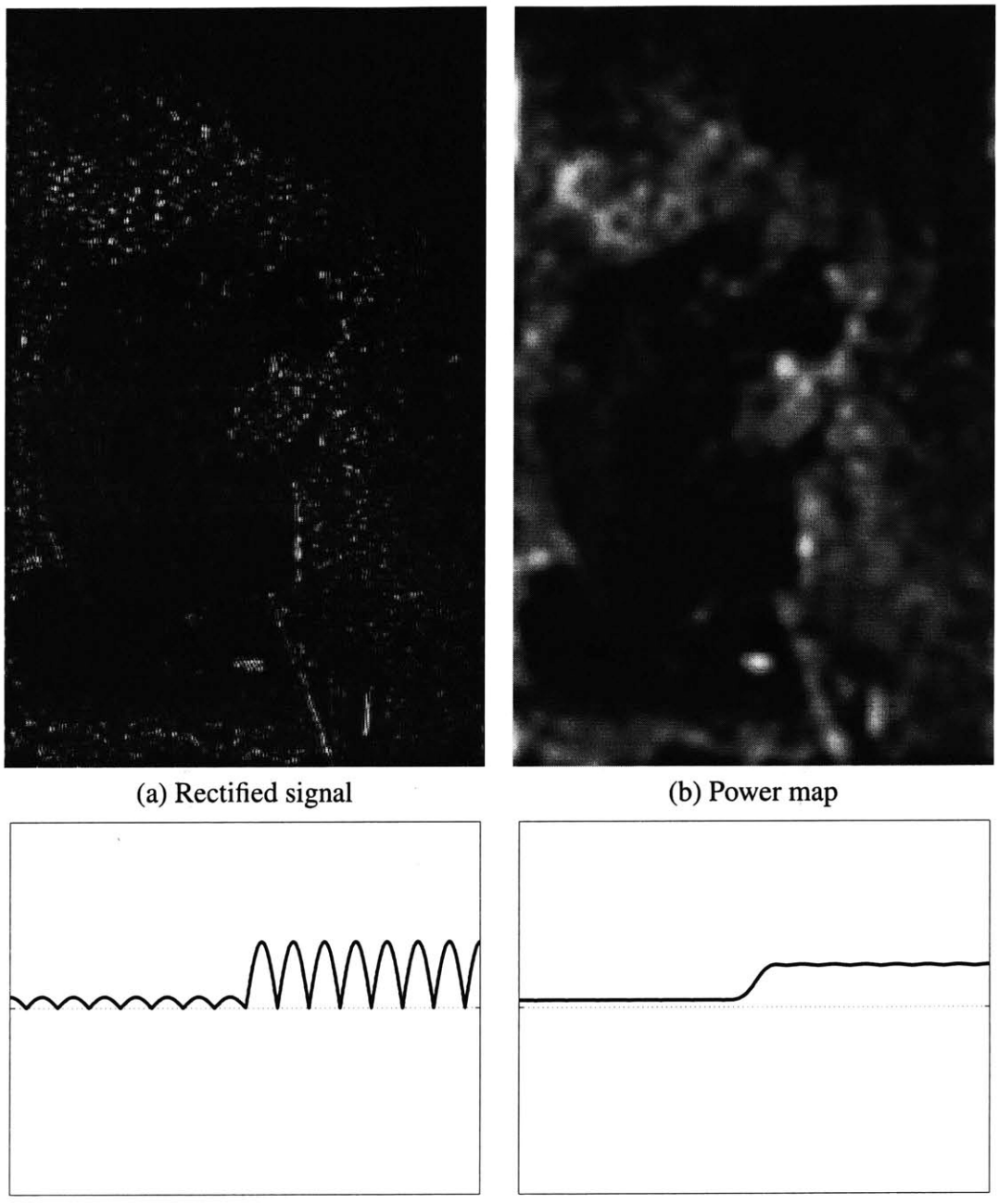
In Section 2.3, we illustrated a filter-based texture-discrimination approach using a 1D example. Now considering the approach in 2D, we compute the local energy content at every scale and orientation of the steerable pyramid. We illustrate the steps for one subband in the figures below. Because the subbands are bandlimited and contain oscillations between positive and negative values, the local average of steerable coefficients for each is zero. As in the 1D case, we perform a full-wave rectification to correct this, taking the absolute values of the steerable coefficients (Figure 3-2(a)).

We next perform a low-pass filtering with a Gaussian kernel to compute the local average of the response magnitude (Figure 3-2(b)); we call the resulting image the power map.

```
for each subband p

  //Full-wave rectification.
  pabs = abs(p)

  //Low pass filter to capture local texture => power map.
  filterlow = lowpassfilter(gaussian, sigma)
  plow = filter(pabs, filterlow)

  //Conversion to log space.
  plog = ln(plow)

  //High pass filter to capture global texture variation.
  filterlow = lowpassfilter(gaussian, sigma)
  phigh = plog - filter(plog, filterlow)

  //Clamp the high pass map to eliminate strong edges.
  kh = clampamt * max(phigh)
  phigh = (kh*phigh) / (kh+phigh)

  //Use these values to scale (blur or sharpen) subband.
  pnew = p * exp(phigh * ks)

end
```

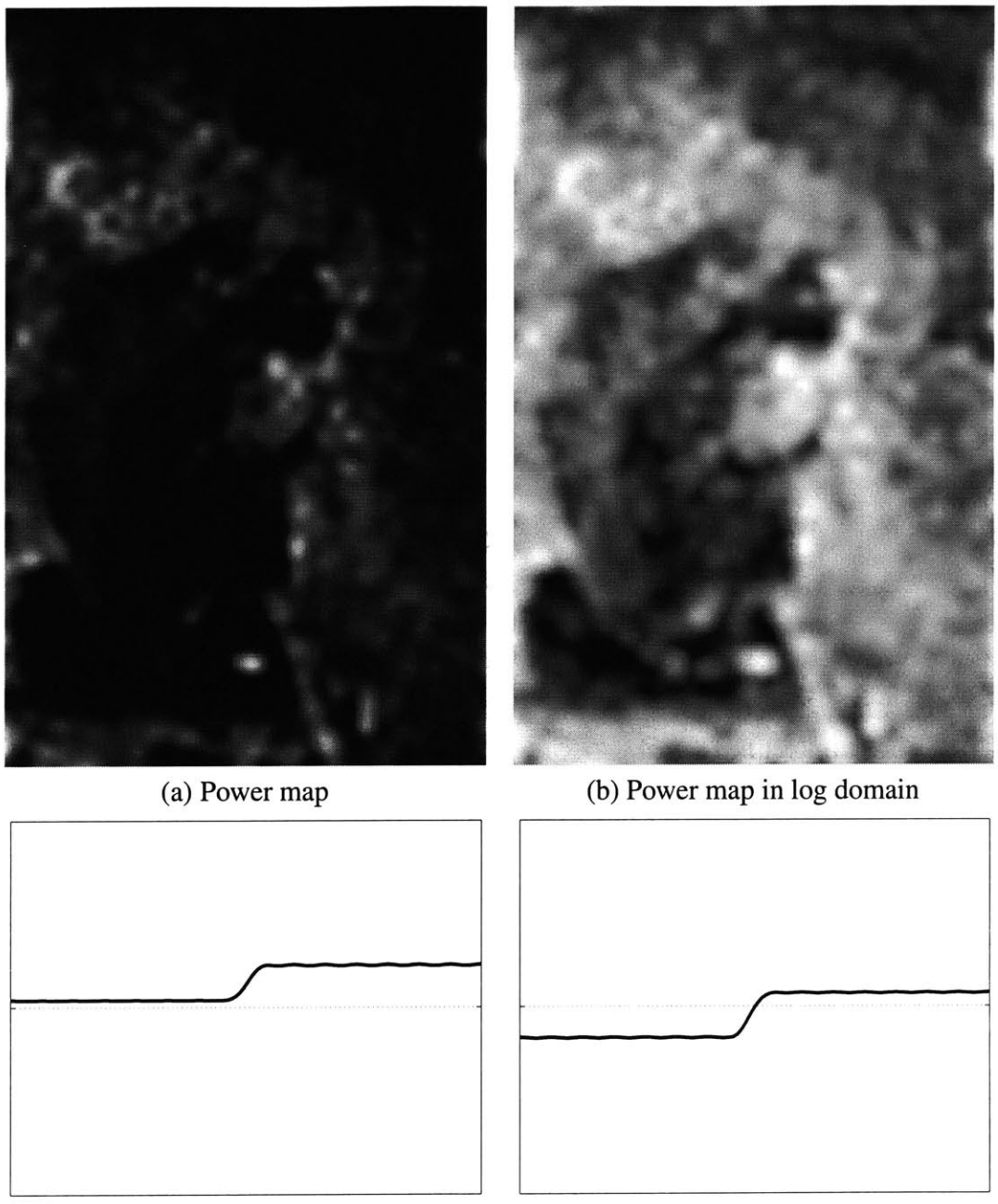
Figure 3-1: Pseudocode for capturing and manipulating texture variation.

We must choose a value of σ for the Gaussian kernel that is large enough to blur the response oscillation but small enough to selectively capture response variations. In practice, we have found that a value of $\sigma = 5$ pixels worked consistently well. Note that the low-pass filter has the same size for each subband, meaning that if it is translated to image-space, it varies at each pyramid level. For coarser scales, the power map averages responses over a larger region of the image. This follows the intuition that local low-frequency content cannot be defined for regions that are smaller than the wavelength.



(a) Rectified signal (b) Power map

Figure 3-2: Power maps to capture local texture content. A low-pass filter with Gaussian kernel $\sigma = 5$ is applied to the rectified signal (a) to compute the local average of the response magnitude (b).



(a) Power map (b) Power map in log domain

Figure 3-3: Log power manipulation.

3.2 Log power manipulation

Because the computation of power maps includes a rectifying non-linearity, propagating modifications on the power map to the image is not straightforward. In particular, applying the final scaling factors directly in the original linear domain may result in negative values that are invalid power map coefficients. While these issues are not a concern for analysis, they are crucial for our image editing context.

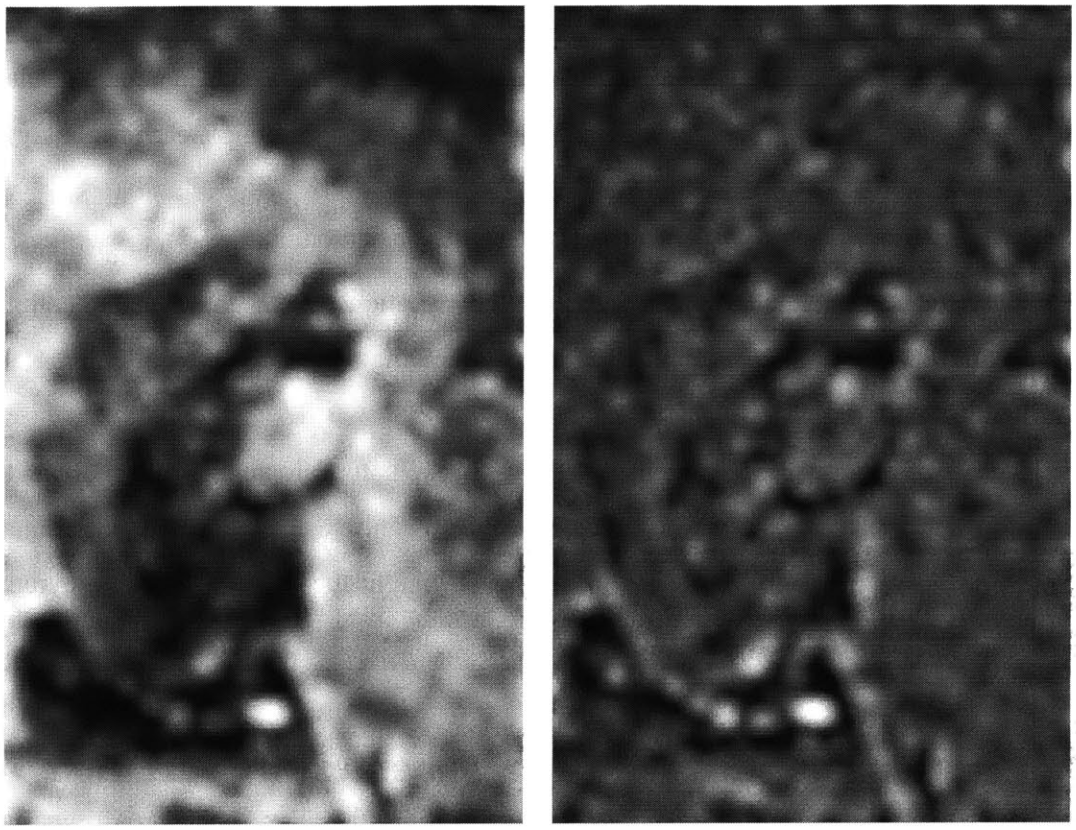
Consider the 1D example of Figure 2-3, where the power map indicates a large disparity between two textures. It is clear that increasing this variation using high-pass sharpening in the linear domain would result in a power map containing negative values, which violates our definition of power maps.

Thus, we perform all subsequent processing in the natural logarithmic domain of the power map, which maps all positive values onto $\mathcal{R}$ (Figure 3-3). An additive change to the log power map translates to a multiplicative change to the original steerable pyramid coefficients.

3.3 Capturing and modifying global texture variation

The power maps capture local texture content in the image. Figure 3-4 shows how high-pass filtering reveals the texture variation over the image. We have experimented with different values of σ for the Gaussian kernel for the high-pass filter. The high-pass filter should scale with the size of the subband such that if it is translated to image-space, it is the same at each pyramid level. In practice, we have found that a maximum value of $\sigma = 60$ pixels for the finest subband worked consistently well. Interestingly, we found that the maximum σ for the high-pass Gaussian had only a small effect on the final result.

Our goal now is to selectively increase or decrease texture variation across an image. Intuitively, to reduce the variation, we want to subtract the high frequencies of the power maps. To increase texture variation, we want to amplify them. Both of these are trivial image-processing operations, however we need to define how a modification of the power map translates into a modification of the pyramid coefficients.



(a) Power map in log domain (b) High-pass filter response

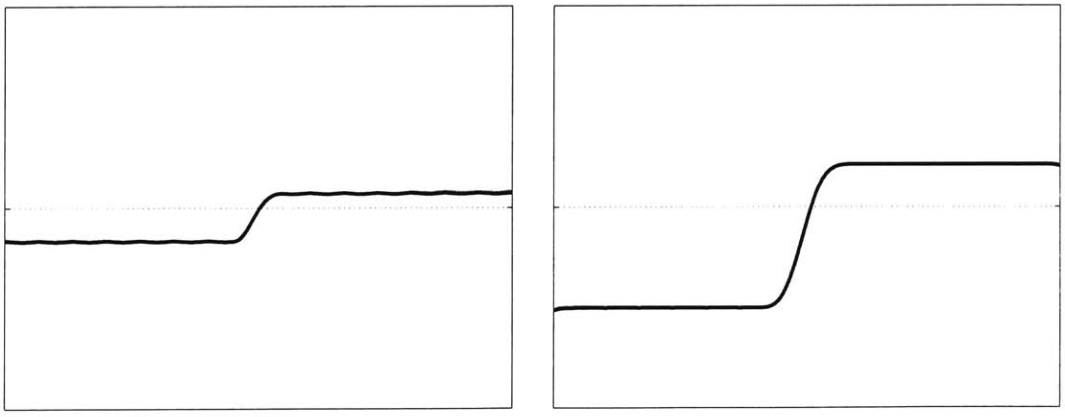


Figure 3-4: High-pass filtering to capture global texture variation. A high-pass filter with Gaussian kernel $\sigma = 60$ is applied to the power map in the log domain.

This translation is clear if we refer back to the 1D example (Figure 2-3). Overlaying the high-pass response signal with the band-pass response (the 1D analogy to the steerable pyramid subband), we see that the former describes the change we want to apply to the latter if our goal is to amplify high frequencies to increase variation (Figure 2-3(h)). Simply using the high-pass response (or some multiple of it) to scale coefficients achieves this because the original high frequencies are multiplied by a factor greater than 1 and low frequencies are multiplied by a factor less than 1. This pushes high frequency and low frequency regions further apart (Figure 3-6). Decreasing texture variation works similarly; we simply use the negative of the high-pass response as the scale. This reduces high frequencies and amplifies low frequencies, pushing the two regions closer together (Figure 3-7).

We must convert back to linear space from log space before the final scaling of the steerable coefficients. As given in the pseudocode above, the final scaling of the subband is as follows.

$pnew = p * e^{(phigh * ks)}$

In practice, we have found a value of $ks = 1$ or 2 to work well as the multiple of the high-pass response to use as the final coefficient scale factor for increasing texture variation; $k = -1$ or -2 work well for decreasing texture variation.

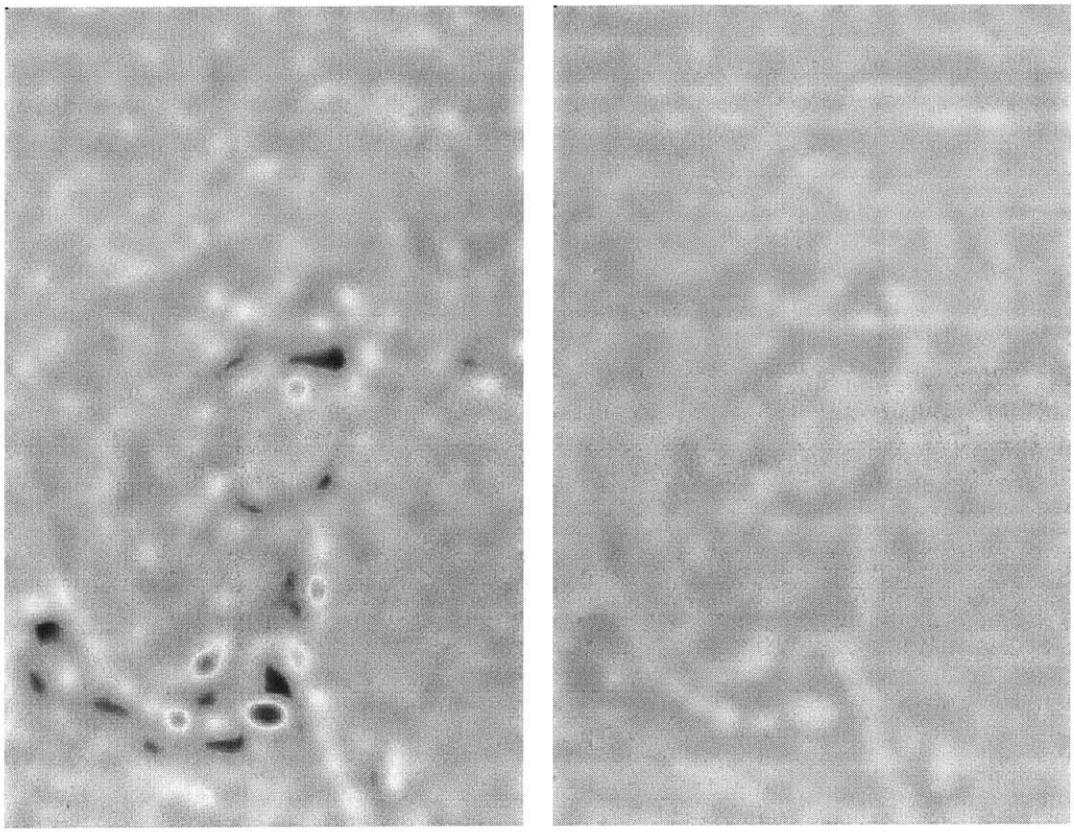
3.3.1 Clamping

It is necessary to clamp the isolated extreme values in the scaling (high-pass response) map to avoid amplifying noise present in the original subband. We use a simple non-linearity to clamp the values to a specified fraction of the maximum, as shown in the pseudocode.

$clamp = c * max(phigh)$

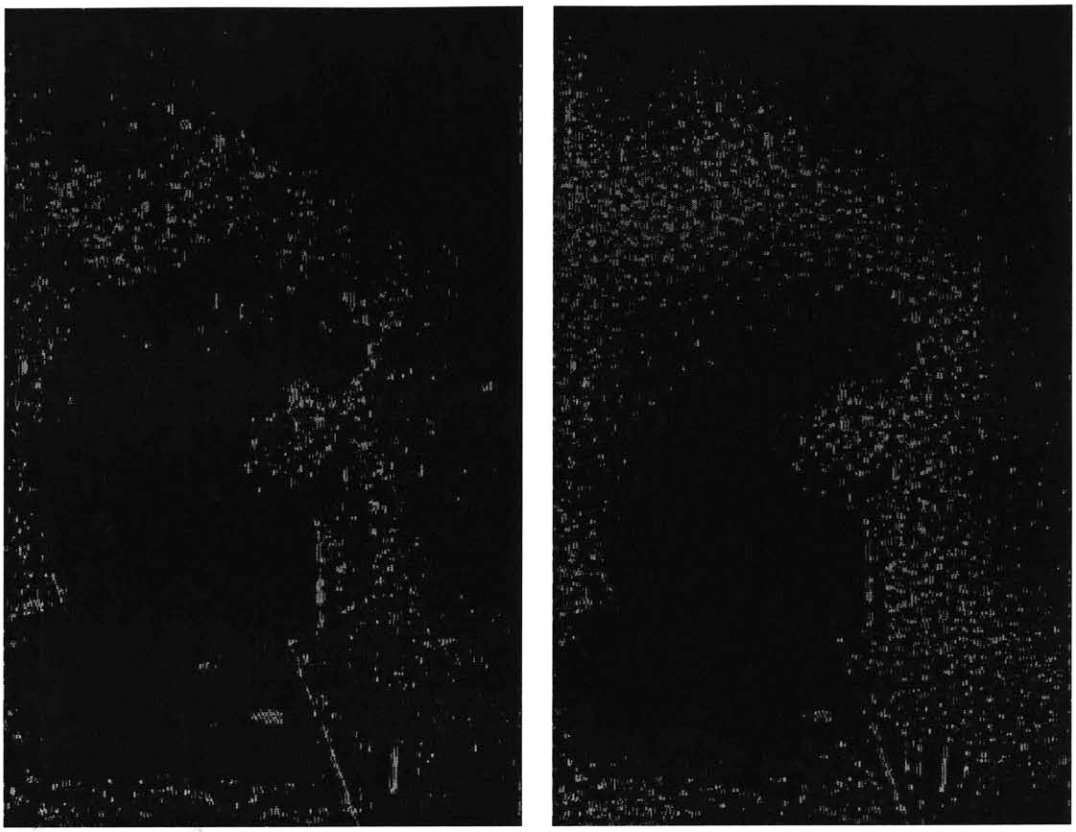
$phigh = \frac{clamp * phigh}{clamp + phigh}$

In practice, we have found that a value of $c = 0.5$ works well for most images.



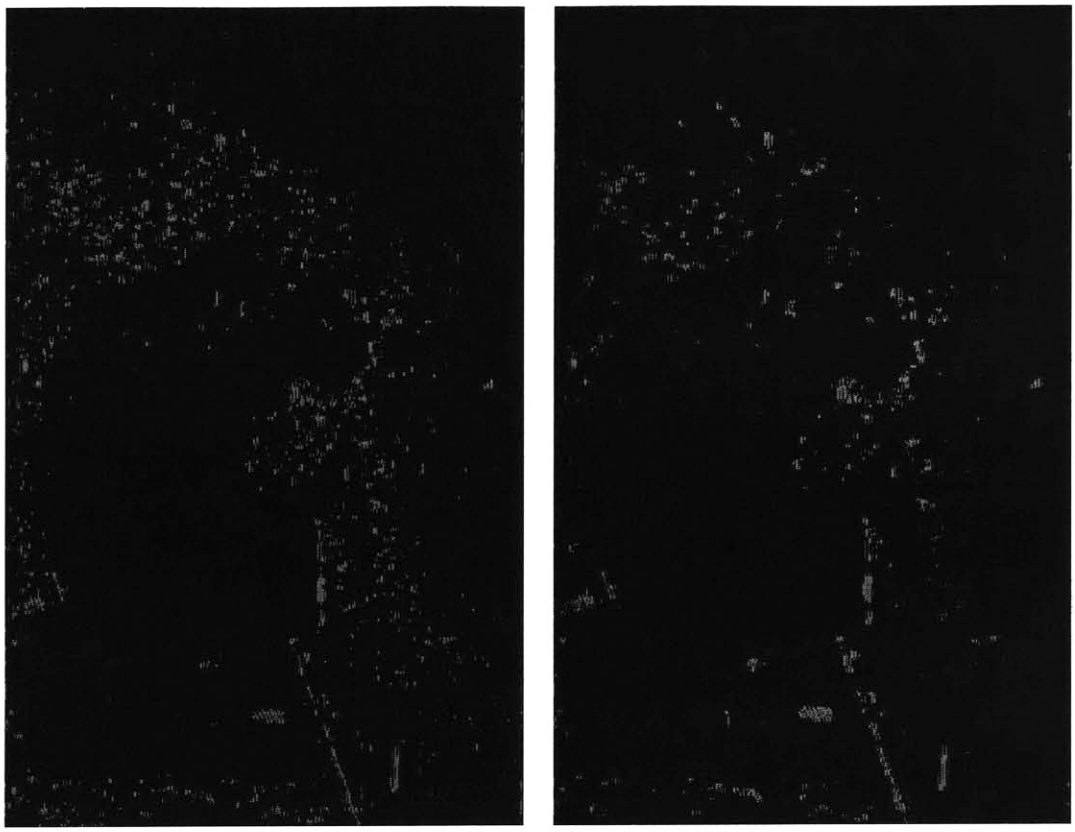
(a) Before clamping (b) After clamping

Figure 3-5: High-pass response before (a) and after (b) clamping. A non-linearity is applied to clamp the values in (a) to half of the maximum. Both before and after maps are visualized with a colormap ranging from maximum negative blue to maximum positive red. It is clear that these isolated extreme values are removed by clamping.



(a) Original subband (b) Texture 'blurred' subband

Figure 3-6: Decreasing texture variation in steerable subband. For visualization purposes, the bandpass subbands are displayed as false-color images in which positive coefficients are red, negative blue, and zero black.



(a) Original subband (b) Texture 'sharpened' subband

Figure 3-7: Increasing texture variation in steerable subband. For visualization purposes, the bandpass subbands are displayed as false-color images in which positive coefficients are red, negative blue, and zero black.

Chapter 4

Results

As discussed in the previous chapter, our goal is to increase or decrease texture variation in image. Intuitively, to reduce variation, we suppress the high frequencies of the power maps, and to increase variation, we amplify them. We refer to these techniques as second-order 'blurring' and 'sharpening' of texture, although their effects may not appear intuitive: to achieve their goals of increasing or decreasing texture uniformity, the two techniques we introduce may make use of either or both of pixel-level blurring and sharpening.

Figure 4-1 shows our texture variation technique used to enhance an image taken by an amateur photographer. Texture variation has been made globally uniform in the middle image, 'blurring' texture edges while preserving true edges. Applied globally, this technique de-emphasizes previously salient objects to produce an image of relatively uniform saliency.

A common post-processing technique for de-emphasis is selective Gaussian blurring to simulate depth-of-field effects. This method works well when the modified object lies at a different distance than the rest of the scene. However, the resulting image may appear unnatural when only one object is modified in a group at a given distance (Figure 4-7). The depth cue introduced by the selective blurring conflicts with the viewer's understanding of the scene. Figure 4-8 shows that a reduction in texture variation is a less obtrusive effect.

Figure 4-2 shows our technique applied to increase texture variation globally. This has the effect of further emphasizing salient objects by increasing contrast at texture boundaries. Note in Figures 4-6 and 4-4 how this method does not globally sharpen all pixels in

the image; rather, it sharpens texture edges by amplifying regions of high frequency and damping regions of low frequency.

Figure 4 shows how increasing global texture variation 'sharpens' texture boundaries while removing noise. Unsharp mask, a standard pixel-space method, sharpens globally and amplifies existing noise and artifacts in an image. In contrast, our technique sharpens texture boundaries by strengthening highly-textured regions and denoising less-textured regions. Sharpening at every scale and orientation prevents haloing artifacts that often result from direct image-space modifications.

Figure 4-1: Decreasing texture variation: `angel`. This photograph taken by an amateur photographer (left) was globally modified to illustrate the effects of decreasing texture variation (right).

Figure 4-2: Increasing texture variation: `angel`. This photograph taken by an amateur photographer (left) was globally modified to illustrate the effects of increasing texture variation (right).

Figure 4-3: Decreasing texture variation: `trunk`. This photograph taken by an amateur photographer (left) was globally modified to illustrate the effects of decreasing texture variation (right).

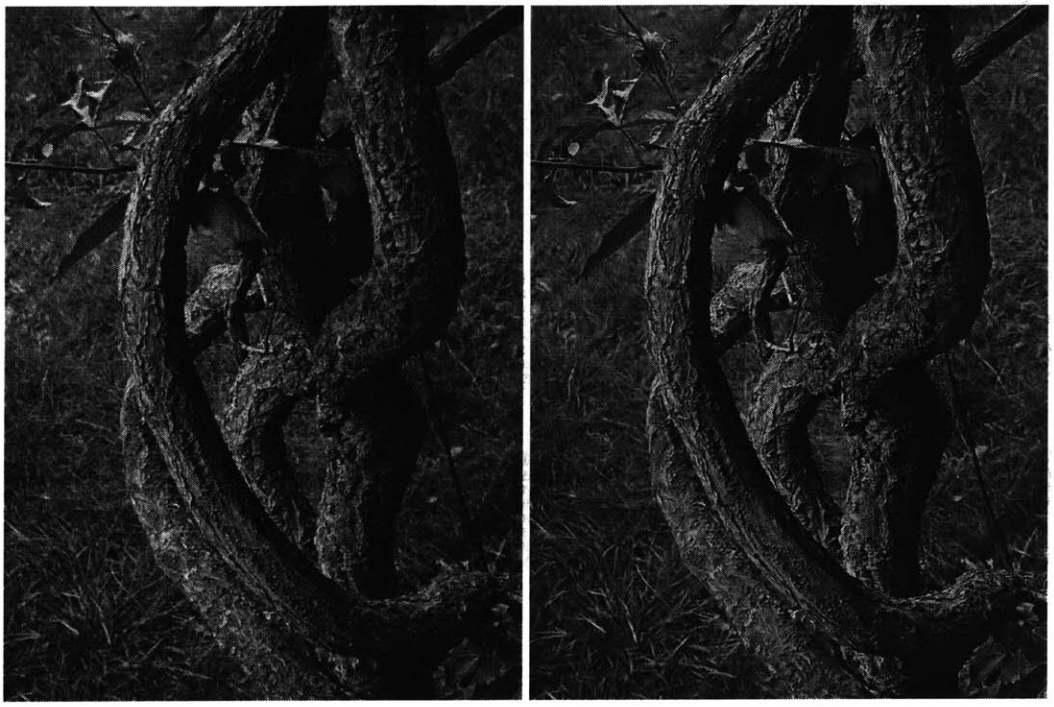


Figure 4-4: Increasing texture variation: `trunk`. This photograph taken by an amateur photographer (left) was globally modified to illustrate the effects of increasing texture variation (right).

Figure 4-5: Decreasing texture variation: `tree`. This photograph taken by an amateur photographer (left) was globally modified to illustrate the effects of decreasing texture variation (right).

Figure 4-6: Increasing texture variation: `tree`. This photograph taken by an amateur photographer (left) was globally modified to illustrate the effects of increasing texture variation (right).

Figure 4-7: Comparison of Gaussian and texture blur: `crosswalk`. This photograph taken by an amateur photographer (top) was selectively modified with Gaussian blur of radius 1.5 (bottom) to de-emphasize the woman. Notice how the Gaussian blurring (even with a relatively small radius) causes a conflicting depth cue in this case because the viewer expects the two objects to be at the same distance. Compare this with the texture blurring shown in Figure 4-8.

Figure 4-8: Comparison of Gaussian and texture blur: `crosswalk`. This photograph taken by an amateur photographer (top) was selectively modified to de-emphasized the woman by reducing texture variation in that region. Unlike with the Gaussian blurring shown in the previous figure, this technique does not cause a depth-of-field effect.

Figure 4-9: Texture 'sharpening' for denoising. The input image (left) was sharpened with an unsharp mask with radius 2 pixels (middle) and our texture sharpening technique (right). Note that while the unsharp mask sharpens globally and amplifies existing noise and JPEG artifacts, our technique sharpens strong texture edges and smoothes the facial region.

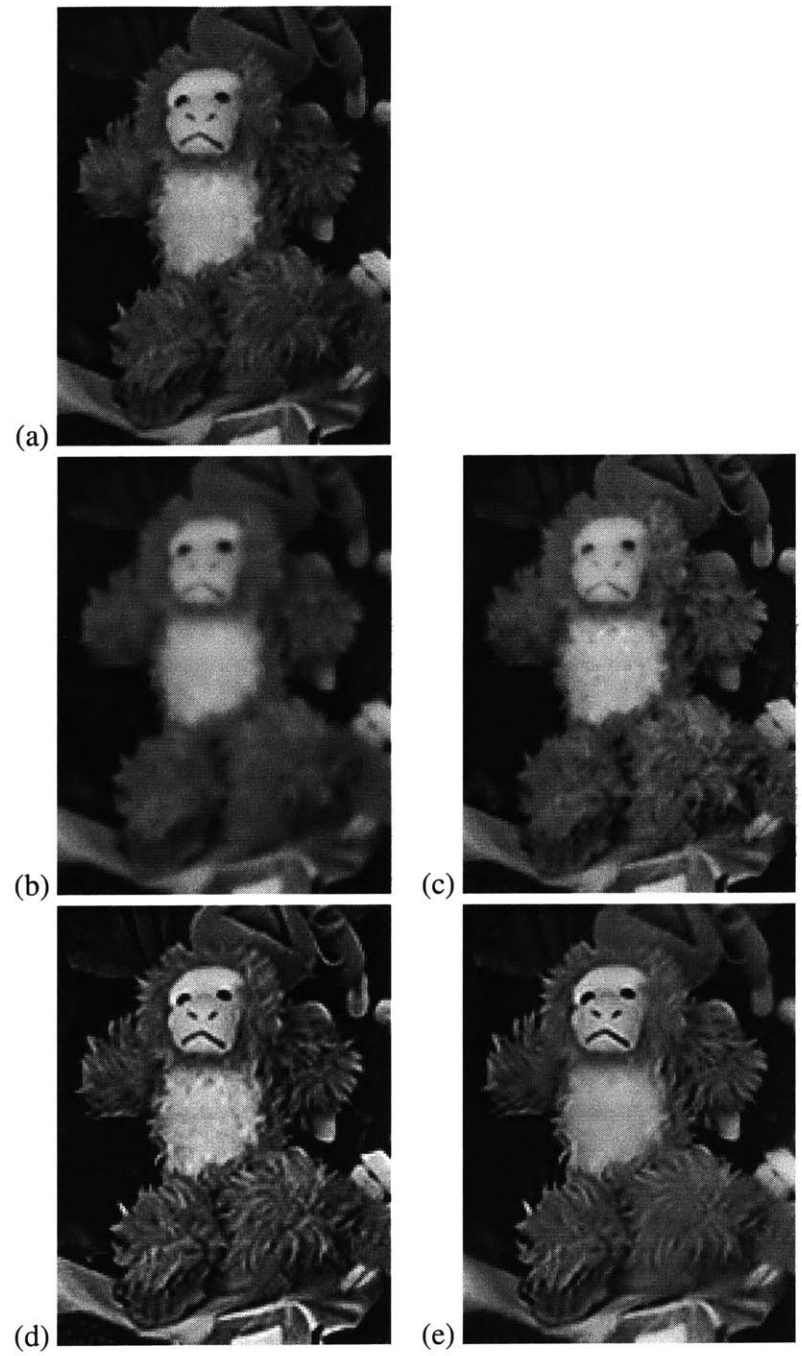


Figure 4-10: Comparison with Gaussian blur and unsharp mask. (a) Original image, (b) with Gaussian blur of radius 1.0 applied, (c) with texture variation reduced to exhibit 'blur' in texture space, (d) with unsharp mask with radius 1.0 applied, (e) and with texture variation increased to exhibit texture 'sharpness'.

Chapter 5

Psychophysical validation

In addition to evaluating our results through visual inspection, we have conducted two user studies to experimentally validate, qualitatively and quantitatively, the effectiveness of our emphasis and de-emphasis techniques.

We have conducted a standard visual search experiment to show that our texture manipulation technique succesfully alters image saliency. Subjects were shown images containing many equally salient objects as well as the same images with our technique applied to selectively emphasize search targets and de-emphasize distractors. We hypothesized that it would take the subjects less time to find the target in the enhanced images.

We showed subjects a series of photographs in their original state and after processing with our techniques. Using an eyetracker, we recorded subjects' eye movements as they viewed the images [Duc03]. The changes in gaze paths confirmed our hypothesis that selectively emphasized image regions attract visual fixation earlier and for greater duration.

These experiments also provided important experimental validation of the outlier-based computational attention models of visual attention that form the theoretical basis of our work [IKN98, Ros99]. Analysis of variance (ANOVA) confirmed that our results are statistically signficant.

5.1 Visual search experiment

Saliency is commonly studied through visual search for a target object in the presence of distractors in a scene. Time to fixation has been shown to be an accurate and reliable indicator of target saliency [JH01]. Psychophysical studies have shown that a target is easier to find if its low-level feature values (chromaticity, intensity, orientation) are more distant from the statistical distribution of the distractors. Our work is based on the insight that texture is an additional low-level feature whose spatial distribution contributes to saliency [PN04].

We conducted a controlled visual search experiment to show that our texture variation technique affects saliency. Subjects were shown a series of images and asked to locate a target object as quickly as possible. We confirmed our hypothesis that search time is reduced when search targets have been emphasized and/or distractors de-emphasized using our technique.

5.1.1 Stimuli

The stimuli used in this experiment were 20 photographs of "messy" scenes, each depicting approximately 50 objects of comparable scale (Figure 5-2). Five versions of each photograph were displayed:

Original. The unmodified photograph.

Blurry. Texture variation of the target is reduced to 'blur' it in texture space. Selected, already-salient objects in the scene are 'sharpened' in texture space to act as distractors. For both 'blurring' and 'sharpening', the following parameters were used: low-pass filter $\sigma = 5$, high-pass filter maximum $\sigma = 60$, high-pass clamping factor $= 0.5$, and final scale factor $= 1$.

Blurrier. The modifications of the previous case are applied with final scale factor $= 2$.

Sharp. Texture variation of the target is increased to 'sharpen' it in texture space. The rest of the image is 'blurred' in texture space. The texture variation parameters used were the same as in the Blurry case above.

Sharper. The modifications of the previous case are applied with final scale factor $= 2$.

Condition	Search time (sec)
Sharper	2.12
Sharp	2.28
Original	2.34
Blurry	2.36
Blurrier	2.76

Table 5.1: Mean time to fixation for visual search experiment.

We refer to the 20 distinct scenes as layouts, each of which has 5 conditions, for a total of 100 distinct images. Grayscale versions of all images were used to isolate the effect of texture variation on saliency without interference from color. Images were displayed on a 36 x 29 centimeter LCD screen at a resolution of 1024 x 768 pixels.

5.1.2 Experimental procedure

Data were collected from 21 volunteer subjects. Each subject was shown the series of 20 layouts on the computer screen. For each layout, one of the 5 conditions was randomly displayed. To prevent a learning effect, no subject was shown the same layout twice.

Subjects were asked to locate the target object and click on it with a mouse. Time to fixation was approximated by the time required for a subject to click on the found search target. A calibration screen was displayed between consecutive layout images, and subjects were required to click on a point at the center of the screen to view the next layout; this was to ensure that all mouse movements begin at the center of the screen.

The target (figure 5.1.2) was selected for its several desirable properties. Without any modification, it is of about average salience in the search scene. Its responses to both our and traditional blurring and sharpening are clear and distinct enough to illustrate the differences between these emphasis techniques.

5.1.3 Analysis

Analysis of variance (ANOVA) was used to test the statistical significance of the difference in conditions. A three-way ANOVA produces a value p for each variable (subject, image,

(a) (b) (c)

Figure 5-1: Search target (a) cropped from an unmodified 'mess' photograph, (b) with texture variation reduced to exhibit 'blur' in texture space, (c) and with texture variation increased to exhibit texture 'sharpness'.

Figure 5-2: Search stimuli: Original.

Figure 5-3: Search stimuli: Blurry.

Figure 5-4: Search stimuli: Blurrier.

Figure 5-5: Search stimuli: Sharp.

Figure 5-6: Search stimuli: Sharper.

condition) estimating the probability that the results occurred by chance. In general, if $p < 0.05$, it is concluded that the variable does have an effect on the measured outcome. A probability $p = 0.0346$ was computed for the condition variable, indicating that the effect of the condition is significant.

5.2 Fixation experiment

Subjects were shown a series of amateur photographs in their original state and after processing with our technique. Using an eyetracker, we recorded subjects' eye movements as they viewed the images [Duc03], hypothesizing that selectively emphasized image regions would attract fixation earlier and for longer.

5.2.1 Experimental procedure

The subject's eye movements were recorded by an ISCAN ETL 400 table-mounted eyetracker with an accuracy of 1 visual degree. The subject's head was secured on an optometric chin-rest to minimize head movement and to ensure a distance of eye to screen 75 centimeters, a distance of eye to camera of 65 centimeters, and a subtended visual angle of 30 x 20 degrees. The eyetracker outputs a data file of screen fixations sampled at a rate of 240 Hz.

Each subject was shown a series of 24 natural images on a 40 x 30 centimeter CRT screen at a resolution of 1024 x 768 pixels. Two versions of each image were displayed: the original and one in which texture variation had been selectively increased or decreased. The subject was shown each image, in random order, for 5 seconds and was asked to study it; no specific task was provided.

5.2.2 Discussion

We evaluated the results of the eyetracking experiment by visual inspection of scan paths (Figure 5-7) and fixation maps [Woo02] (Figure 5-8). This qualitative evaluation supported our expectation that image regions emphasized using our technique would attract and hold

fixations. Although this experiment was not as controlled as the visual search and did not include as many subjects, the initial qualitative results are promising. We intend to conduct an extended study in the future.

(a)

(b)

Figure 5-7: Scan paths: `brick`. The change in scan paths recorded by the eyetracker shows that salient objects in the original image (a) have been successfully de-emphasized by reducing overall texture variation (b). Blue circles show saccadic jumps while red circles represent fixations, with the duration indicated by the radius of the circle.

(a)

(b)

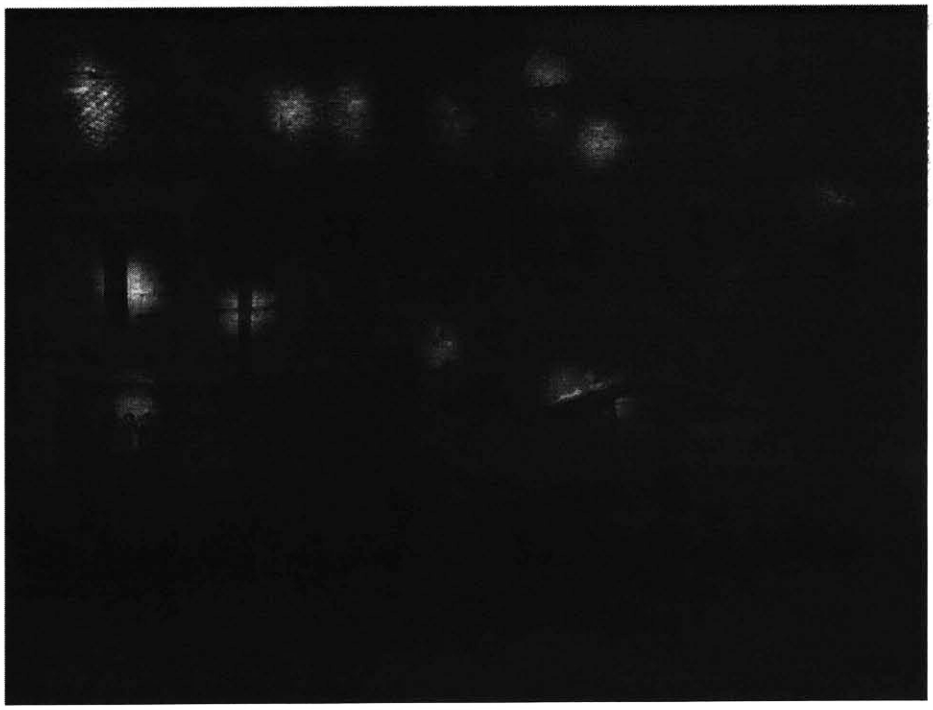


Figure 5-8: Fixation maps: `brick`. This alternative visualization of Figure 5-7 shows the saliency of the before and after images as variations in intensity. Brighter regions are those on which subjects fixated for longer.

Chapter 6

Conclusions and future work

We have presented a novel post-processing technique for modifying texture variation in images. Our method is inspired by bottom-up models of visual attention that predict a strong response to statistical outliers in low-level feature distributions. We have exploited this behavior to alter saliency in an image by adding or removing outliers in the feature distribution to increase or decrease variations in local frequency content.

We use the steerable pyramid decomposition to define for an image a set of power maps which capture local texture content at each scale and orientation and provide a perceptually-meaningful framework for image manipulation.

A visual search experiment verified our hypotheses that texture is a salient image feature and that modifying texture of the search target directly affects time to fixation. ANOVA showed these results to be statistically significant.

We also recorded subjects' eye movements as they viewed original and modified images without a specific search task. Although this experiment was not as controlled as the visual search, the qualitative results supported our hypothesis that objects emphasized with our technique are effective at attracting and holding fixation.

Our texture variation technique is complementary to existing post-processing emphasis methods, such as sharpening and brightening, that modify pixels of the image directly. In instances where these techniques result in objectionable artifacts, our multi-scale method may have more success. We have shown that reduction of texture variation (texture 'blurring') can be used in cases where traditional blurring creates an unnatural depth-of-field

effect. Increasing texture variation (texture 'sharpening') effectively adds emphasis to images and, in many cases, has the desirable effect of noise removal. Traditional unsharp mask tends to amplify existing noise in images.

Our technique also has its limitations. The large scale effect of decreasing texture variation across an image is to add oriented noise in less-textured regions. This is achieved by effectively amplifying existing texture at multiple scales. If large regions of the image are too smooth, this has the effect of only amplifying artifacts.

It is difficult to do a fair comparison between image processing techniques because the various control parameters are not typically comparable. One strategy is to use eyetracking and automatic saliency metrics to calibrate different techniques, e.g. to determine which size Gaussian blur kernel is comparable to the strength of texture blurring we use. This study is future work.

Our technique is not general-purpose, but neither are the existing image processing techniques discussed in this paper. Our contribution is another tool in the toolbox for image emphasis and enhancement.

Appendix A

Consent form for user study

CONSENT TO PARTICIPATE IN
NON-BIOMEDICAL RESEARCH

Effects of image enhancement algorithms on eye movement and fixation

You are asked to participate in a research study conducted by Sara Su and Frédo Durand from the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology (M.I.T.). The results of the study will contribute to the Master's thesis of Sara Su. You were selected as a possible participant in this study because of your affiliation with the computer graphics and/or vision community. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

PARTICIPATION AND WITHDRAWAL

Your participation in this study is completely voluntary and you are free to choose whether to be in it or not. If you choose to be in this study, you may subsequently withdraw from it at any time without penalty or consequences of any kind. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

PURPOSE OF THE STUDY

This study is designed to assess the effectiveness of techniques we have developed for adding visual emphasis to images. This research is applicable to the fields of computer graphics and image processing. Our goal is to provide a tool for novice photographers to subtly enhance images in order to redirect viewers' attention. You will be shown a series of images on a computer screen, and as you view each one, your eye movements will be tracked by a camera. Studying the change in eye movements and fixations when subjects view unaltered and altered photographs will allow us to assess the effectiveness of our image processing algorithms.

PROCEDURES

If you volunteer to participate in this study, we will ask you to do the following things:

You will be shown a series of approximately 20 images on a computer screen. 4 to 6 versions of each image will be shown: an original photograph taken by a novice plus versions subtly altered by our image processing algorithm or existing techniques. You will see one image at a time on the screen in a random order, and you will be asked to study each for approximately 15 seconds. You will be asked to perform a simple task for each image such as search for an object in the scene or judge its aesthetic quality. You will not be asked to make explicit comparisons between different versions of an image. Your eye movements will be recorded by a non-invasive camera system made by ISCAN, Inc. To prepare the camera system, you will be shown a standard calibration image before the test begins; this step should take less than five minutes to complete. The complete test, including informed consent, should taken approximately 45 to 60 minutes.

All testing will take place in an office in the Computational Visual Cognition Laboratory in MIT Building NE20.

POTENTIAL RISKS AND DISCOMFORTS

Each participant in the study will be viewing images on a computer screen for approximately 30 minutes and will be asked to remain relatively still for the duration. If you have a prior history of eye strain, you may want to reconsider your participation in the study.

POTENTIAL BENEFITS

It is not expected that you will receive any direct, personal benefits as a result of your participation in this study.

This research will be used to evaluate the relative performance of image processing algorithms for adding subtle emphasis to photographs. The results of the study will help to contribute to the general knowledge of the computer graphics and human perception communities.

PAYMENT FOR PARTICIPATION

No financial compensation will be offered in exchange for participation in this study.

CONFIDENTIALITY

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law.

The only identifiable information that will be included in this study are the participant's name and e-mail address. This information will be stored electronically and will be accessible only to the researchers who are directly involved in administering the study. Data will be electronically archived following the study. If other researchers use the data in future

projects, personal identifiable information will be excluded.

IDENTIFICATION OF INVESTIGATORS

If you have any questions or concerns about the research, please feel free to contact one of the following investigators:

Frédo Durand, Principal Investigator
Daytime phone: 617-253-7223
MIT Office: 32-D426
E-mail address: fredo@mit.edu

Sara Su, Associated Investigator
Daytime phone: 617-253-8835
MIT Office: 32-D416
E-mail address: sarasu@mit.edu

EMERGENCY CARE AND COMPENSATION FOR INJURY

"In the unlikely event of physical injury resulting from participation in this research you may receive medical treatment from the M.I.T. Medical Department, including emergency treatment and follow-up care as needed. Your insurance carrier may be billed for the cost of such treatment. M.I.T. does not provide any other form of compensation for injury. Moreover, in either providing or making such medical care available it does not imply the injury is the fault of the investigator. Further information may be obtained by calling the MIT Insurance and Legal Affairs Office at 1-617-253-2822."

RIGHTS OF RESEARCH SUBJECTS

You are not waiving any legal claims, rights or remedies because of your participation

in this research study. If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E32-335, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253-6787.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

Name of Subject .....

Name of Legal Representative (if applicable) .....

Signature .....

Date .....

SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

Signature of Investigator .....

Date .....

Bibliography

[AAB+84] Edward H. Adelson, Charles H. Anderson, James R. Bergen, Peter J. Burt, and Joan M. Ogden. Pyramid methods in image processing. *RCA Engineer*, 29(6), 1984.

[Bar61] Horace B. Barlow. Possible principles underlying the transformations of sensory messages. In W. A. Rosenblith, editor, *Sensory Communication*, pages 217–234. MIT Press, 1961.

[BJEYLW01] Ziv Bar-Joseph, Ran El-Yaniv, Dani Lischinski, and Mike Werman. Texture mixing and texture movie synthesis using statistical learning. *IEEE Transactions on Visualization and Computer Graphics*, 7(2):120–135, 2001.

[BPR81] Jacob Beck, K. Prazdny, and Azriel Rosenfeld. A theory of textural segmentation. *Human and Machine Vision*, pages 1–38, 1981.

[BRT95] Lawrence D. Bergman, Bernice E. Rogowitz, and Lloyd A. Treinish. A rule-based tool for assisting colormap selection. In *Proceedings of the 6th conference on Visualization '95*, page 118. IEEE Computer Society, 1995.

[BTss] Mireille Bétrancourt and Barbara Tversky. Simple animations for organizing diagrams. *International Journal of Human-Computer Studies*, In press.

[CXF+02] Liqun Chen, Xing Xie, Xin Fan, Wei-Ying Ma, Hong-Jiang Zhang, and Heqin Zhou. A visual attention model for adapting images on small displays. Technical Report MSR-TR-2002-125, Microsoft Research, 2002.

[DS02] Doug DeCarlo and Anthony Santella. Stylization and abstraction of photographs. In *Proceedings of SIGGRAPH 2002*, pages 769–776, 2002.

[Duc03] Andrew T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, 2003.

[Eis04] Katrin Eismann. *Photoshop Restoration and Retouching*. Pearson Education, 2nd edition, 2004.

[EK03] Wolfgang Einhäuser and Peter König. Does luminance-contrast contribute to a salience map for overt visual attention? *European Journal of Neuroscience*, 17:1089–1097, 2003.

[EZW97] Stephen Engel, Xuemei Zhang, and Brian Wandell. Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, pages 68–71, July 1997.

[FA91] William T. Freeman and Edward H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, September 1991.

[GBP+94] Hayit Greenspan, Serge Belongie, Pietro Perona, Rodney Goodman, Subrata Rakshit, and Charles Anderson. Overcomplete steerable pyramid filters and rotation invariance. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 222–228, June 1994.

[HB95] David J. Heeger and James R. Bergen. Pyramid-based texture analysis/synthesis. In *Proceedings of SIGGRAPH 95*, Computer Graphics Proceedings, Annual Conference Series, pages 229–238, August 1995.

[HTER04] Christopher G. Healey, Laura Tateosian, James T. Enns, and Mark Remple. Perceptually based brush strokes for nonphotorealistic visualization. *ACM Transactions on Graphics*, 23(1):64–96, 2004.

[IK01] Laurent Itti and Christof Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, March 2001.

[IKN98] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, November 1998.

[Int00] Victoria Interrante. Harnessing natural textures for multivariate visualization. *IEEE Computer Graphics and Applications*, 20(6):6–11, 2000.

[JH01] Michael Jenkin and Laurence Harris, editors. *Vision and Attention.* Springer-Verlag, 2001.

[KU85] Christof Koch and Shimon Ullman. Shifts in selective visual attention: Towards the underlying neural circuity. *Human Neurobiology*, 4(2):279–283, 1985.

[LG04] Michael S. Landy and Norma Graham. Visual perception of texture. In L. M. Chalupa and J. S. Werner, editors, *The Visual Neurosciences*, pages 1106–1118. Cambridge, MA: MIT Press, 2004.

[Mar89] Judy Martin. *Technical illustration: materials, methods and techniques.* Macdonald Orbis, 1989.

[MBLS01] Jitendra Malik, Serge Belongie, Thomas Leung, and Jianbo Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, 43(1):7–27, 2001.

[MP90] Jitendra Malik and Pietro Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of Optical Society of America A*, 7(5):923–932, 1990.

[Nor85] H.Christoph Northdurft. Sensitivity for structure gradient in texture discrimination tasks. *Vision Research*, 25:1957–1968, 1985.

[OABB85] Joan M. Ogden, Edward H. Adelson, James R. Bergen, and Peter J. Burt. Pyramid-based computer graphics. *RCA Engineer*, 30:4–15, 1985.

[Pal99] Stephen E. Palmer. *Vision Science: Photons to Phenomenology*. Bradford Books, 1999.

[PN04] Derrick J. Parkhurst and Ernst Niebur. Texture contrast attracts overt visual attention in natural scenes. *European Journal of Neuroscience*, 19(3):783–789, 2004.

[PO03] Christopher Peters and Carol O'Sullivan. Bottom-up visual attention for virtual human animation. In *Proceedings of Computer Animation for Social Agents (CASA) 2003*, May 2003.

[PS00a] Javier Portilla and Eero P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–70, October 2000.

[PS00b] Claudio M. Privitera and Lawrence W. Stark. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(9):970–982, 2000.

[PSWS03] Javier Portilla, Vasily Strela, Martin Wainwright, and Eero P. Simoncelli. Image denoising using a scale mixture of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, November 2003.

[Ros99] Ruth Rosenholtz. A simple saliency model predicts a number of motion popout phenomena. *Vision Research*, 39:3157–3163, 1999.

[RZ99] Pamela Reinagel and Anthony M. Zador. Natural scene statistics at the centre of gaze. *Network: Comput. Neural. Syst*, 10:1–10, 1999.

[SA96] Eero P. Simoncelli and Edward H. Adelson. Noise removal via bayesian wavelet coring. In *IEEE Third International Conference on Image Processing*, September 1996.

[SD02] Anthony Santella and Doug DeCarlo. Abstracted painterly renderings using eye-tracking data. In *Proceedings of NPAR 2002*, pages 75–82, 2002.

[SD04] Anthony Santella and Doug DeCarlo. Visual interest and NPR: An evaluation and manifesto. In *Proceedings of NPAR 2004*, pages 71–78, 2004.

[SF95] Eero P. Simoncelli and William T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *2nd Annual IEEE International Conference on Image Processing*, pages 444–447, 1995.

[SF96] Eero P. Simoncelli and Hany Farid. Steerable wedge filters for local orientation analysis. *IEEE Transactions on Image Processing*, 5(9):1377–1382, 1996.

[SLBJ03] Bongwon Suh, Haibin Ling, Benjamin B. Bederson, and David W. Jacobs. Automatic thumbnail cropping and its effectiveness. Technical Report CS-TR-4469, University of Maryland, April 2003.

[Sol96] Robert L. Solso. *Cognition and the Visual Arts*. Bradford Books, 1996.

[TMB02] Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, 2002.

[Tuf90] Edward R. Tufte. *Envisioning Information*. Graphics Press, 1990.

[War00] Colin Ware. *Information Visualization: Design for Perception*. Academic Press, 2000.

[WB79] Hugh R. Wilson and James R. Bergen. A four mechanism model for threshold spatial vision. *Vision Research*, 28:611–628, 1979.

[Woo02] David S. Wooding. Fixation maps: quantifying eye-movement traces. In *ETRA '02: Proceedings of the symposium on Eye tracking research & applications*, pages 31–36, 2002.

[YPG01] Hector Yee, Sumanita Pattanaik, and Donald P. Greenberg. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics*, pages 39–65, 2001.

[Zek99] Semir Zeki. *Inner Vision: An Exploration of Art and the Brain*. Oxford University Press, 1999.