De-Emphasis of Distracting Image Regions Using Texture Power Maps

Sara L. Su, MIT CSAIL, sarasu@mit.edu
Frédo Durand, MIT CSAIL, fredo@mit.edu
Maneesh Agrawala, Microsoft Research, maneesh@microsoft.com

MIT Laboratory for Computer Science
Technical Report
MIT-LCS-TR-987
April 12, 2005

Abstract

A major obstacle in photography is the presence of distracting elements that pull attention away from the main subject and clutter the composition. In this article, we present a new image-processing technique that reduces the salience of distracting regions. It is motivated by computational models of attention that predict that texture variation influences bottom-up attention mechanisms. Our method reduces the spatial variation of texture using power maps, high-order features describing local frequency content in an image. We show how modification of power maps results in powerful image de-emphasis. We validate our results using a user search experiment and eye tracking data.

Keywords: texture manipulation, texture discrimination, image editing, saliency, visual attention

1 Introduction

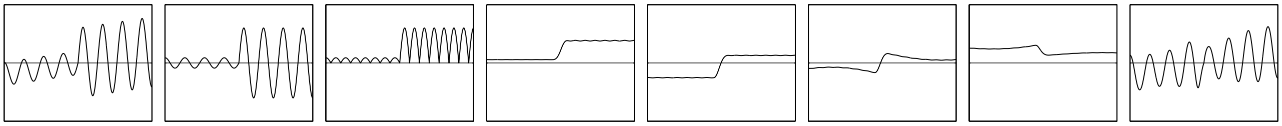
A major obstacle in photography is the presence of distracting elements that pull attention away from the main subject and clutter the composition. Much of the art of photography involves directing the viewer's attention to or away from regions of an image. Photographers have developed a variety of post-processing techniques, both in the darkroom and on the computer, to reduce the salience of distracting elements by altering low-level features to which the human visual system is particularly attuned: sharpness, brightness, chromaticity, or saturation.

Surprisingly, there is one low-level feature that cannot be directly manipulated with existing image-editing software: *texture variation*. From a perceptual standpoint, variations and outliers in texture are salient to the human visual system [14, 5], and the human and computer vision literature show that discontinuities in texture can elicit an edge perception similar to that triggered by color discontinuities [1, 11, 20, 10].

We introduce a new technique for selectively altering texture variation to reduce the salience of an image region. In a nutshell, we modify the region to make it look more like a uniform texture. Our method is based on perceptual models of attention that hypothesize that contrast in texture contributes to salience. We review the filter-based model of texture discrimination and the computational models of visual attention based on it (Section 2) before presenting the following contributions.

Image manipulation with power maps. Higher-order image features have been heavily used in image analysis. For example, power maps encode the local average of the response to oriented filters. We show how power maps provide a powerful representation for *manipulating* frequency content in an image We introduce a perceptually-motivated technique for selective manipulation of texture variation.

Psychophysical study of texture and attention. 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 eye tracker. In addition, a search experiment shows quantitatively that applying our technique to distractors directs viewer attention toward unmodified regions.



(a) Input (b) Band-pass (c) Rectified (d) Low-pass (e) Log (f) High-pass (g) Exponent (h) Scaled

Figure 1: Texture discrimination and manipulation in 1D. Band-pass filtering the input signal (a) produces a response (b) which, averaged over a small neighborhood, is zero. Rectification, in this case full-wave, is necessary (c). Pooling the rectified response captures local frequency content (d). Converting to log-space (e) allows multiplicative rather than additive operations. The high-pass response (f) captures global texture variation and, converting back to linear-space (g), is used to scale the bandpass response Because the goal is to reduce variation, a negative multiple of the high-pass is used as the scale factor. The resulting signal has been 'flattened' (h).

2 Background

2.1 Texture segmentation and discrimination

Texture discrimination and texture edge detection has received much attention in computational and human vision [1, 11, 10, 9]. These approaches compute local variations in frequency content to detect texture edges. Most approaches roughly follow Malik and Perona's biologically-inspired model [11], illustrated with a 1D example in Figure 1. 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 1(b)). Because it is band-limited, the response to such a filter averaged over a small neighborhood is usually zero; the positive and negative lobes of the response cancel each other. The signal must be rectified. Possible non-linearities include full-wave rectification (absolute value) and energy computation (square response); the absolute value is shown in Figure 1(c).

Low-pass filtering (or "pooling") of this rectified response produces the local average of the filter response strength; we call this the *power map* (Figure 1(d)). As suggested by Northdurft [12], an analysis similar to intensity images can then be performed on these power maps at each scale and orientation.

In addition to its applications in edge detection and image segmentation, this approach to texture discrimination has inspired texture synthesis methods that match histograms of filter responses [6]. We show how power maps can be applied to a different problem: image manipulation.

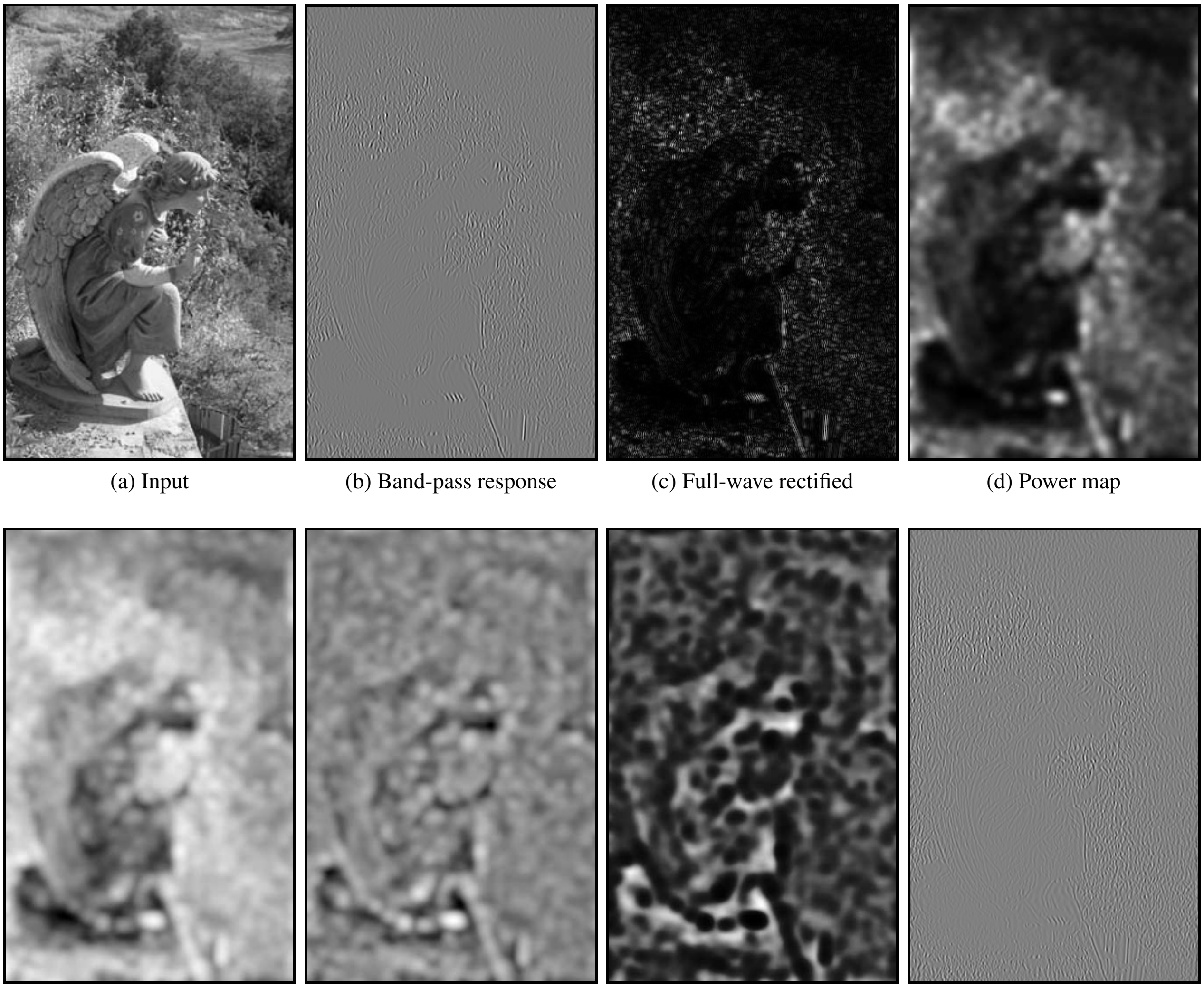
2.2 Computational models of visual attention

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 or the task. Top-down processes are important to understanding attention, however, in this paper, 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. A number of influential computational models of attention have explicitly identified *salient* objects as statistical outliers in low-level feature distributions [17, 15, 16]. Other well-known models implicitly capture the same behavior [7].

Most models focus on the response to filter banks that extract contrast and orientation in the image. Various non-linearities can then be used to extract and combine maxima of the response to each feature. These *first-order* salience models capture low-level features such as contrast, color, and orientation. Increasing or decreasing the presence of outliers or large variations in the feature distribution for a region of the image results in a respective increase or decrease in the salience of the region, as exploited by traditional image editing techniques [19, 13, 22].

Recently, Parkhurst and Niebur [14] presented a saliency model that captures texture variation in order to explain psychophysical experiments by Einhäuser and König who reported salience effects that could not be explained by first-order models [3] . Their *second-order* model performs additional image processing on the response to a first-order filter bank. This effectively performs the same computation as first-order models but on what we call power maps rather than on image intensity. This motivates our strategy of performing image manipulations on power maps in order to modify contrast in texture.



(a) Input (b) Band-pass response (c) Full-wave rectified (d) Power map

(e) Log space (f) High-pass response (g) Scaling factor (h) Scaled coefficients

Figure 2: Texture discrimination and manipulation in 2D. We consider a single band-pass subband (b) of the steerable pyramid of the input image (a). The average of (b) over a small neighborhood, is zero, so full-wave rectification is necessary (c). Pooling the rectified response produces the power map (d), which captures local texture content. Converting to log-space (e) allows multiplicative rather than additive operations. The high-pass response (f) captures global texture variation and, converting back to linear-space (g), is used to scale the subband coefficients (h) to reduce texture variation.

3 Texture equalization

We have reviewed how the response to multiscale oriented filters can be used for texture discrimination. A plethora of such filters has been developed. In our work, we use *steerable pyramids* because they permit straightforward analysis, processing, and reconstruction of images [4, 18].

We have developed a post-processing technique to de-emphasize distracting regions of a photograph by reducing contrast in texture. Informally, our goal is to invert the outlier-based computational model of saliency to perform *texture equalization*. 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 decrease spatial variation of texture as captured by the response to steerable oriented filters.

3.1 Power maps to capture local energy

We compute power maps using the texture-discrimination approach described in Section 2.1. The local frequency content is computed using steerable pyramid, and a power map is computed for each subband s. We illustrate the steps for one subband in Figure 2.

As discussed above, the subbands are band-limited and their local average is zero. We perform a full-wave rectification to correct this, taking the absolute values of the steerable coefficients (Figure 2(c)).

We next apply a low-pass filter with a Gaussian kernel g_l to compute the local average of the response magnitude (Figure 2(d)); we call the resulting image s_l the *power map*.

$$s_l = |s| \otimes g_l \tag{1}$$

We must choose a variance σ_l for the Gaussian kernel that is large enough to blur the response oscillation but small enough to selectively capture response variations. We have found that a value of $\sigma_l = 5$ pixels worked consistently well. Note that the low-pass filter has the same size for each subband, Note that the low-pass filter has the same size or each subband. Thus, for coarser scales the power map averages responses over a larger region of the image.

3.2 Log power manipulation

Because the computation of power maps includes an absolute-value rectifying non-linearity, propagating modifications on the power map to the image is not straightforward. In particular, linear image processing may result in negative values that are invalid power map coefficients, since the power map is computed from absolute values. While these issues are not a concern for analysis, they are crucial to consider for image editing.

This is why we choose to perform multiplicative manipulation, scaling the coefficients rather than using additive arithmetic. For this, we perform all subsequent processing in the natural logarithmic domain of the power map (Figure 2(e)). An additive change to the log power map translates to a multiplicative change to the original steerable pyramid coefficients.

3.3 Reducing global texture variation

The power maps capture local frequency content in the image. We next perform high-pass filtering reveals the spatial variation s_h of frequency content over the image (Figure 2(f)). Recall that this variation is defined for each subband s.

$$s_h = \ln(s_l) - (\ln(s_l) \otimes g_h) \tag{2}$$

We have experimented with different values of σ_h for the Gaussian kernel g_h. In contrast to the low-pass g_l, the high-pass filter must scale with the size of the subband such that if it is translated to image-space, it is the same at each pyramid level. We have found that a value of $\sigma_h = 60$ pixels for the finest subband worked consistently well. We have found that the technique is robust to this choice and that the value of σ_h has a small effect on the final output.

To reduce texture variation in the image, we must remove some portion of the high frequencies of the power maps, which is a trivial image-processing operation. However, we must define how a modification of the power map translates into a modification of the pyramid coefficients. Recall that we are working in the log domain to perform multiplicative modification to the power map and steerable-pyramid coefficients. A subtraction on the log power map corresponds to a division of the linear coefficients (Figures 1(g) and 2(g)):

$$s' = se^{-ks_h} \tag{3}$$

We have found values of $k = 1,2,3$ to work well. In practice, at the boundary between low and high values of the power map, the high-pass of the log power map gos from negative to positive values, which results in scaling up the coefficients on the low side and scaling down on the other side (Figure 1 (g) and (h)).

Clamping. Smooth regions correspond to zero values of the power map. When they are adjacent to highly-textured regions, they result in extremely high values of the high-frequency of the power map s_h. As a result, the applied scaling factor is large and can overly enhance the small amount of noise present in the smooth region.

To avoid enhancing noise present in the original subband, it is necessary to clamp the isolated extreme values in the scaling (high-pass response) map. In particular, it is important to avoid amplifying isolated extreme values in large uniform regions. To prevent such artifacts, we use a simple non-linearity to clamp extreme values to a specified fraction of the maximum:

$$s'_h = \frac{cs_h}{c+s_h} \tag{4}$$

where $c = k_c \max(s_h)$. In practice, we have found that a value of $k_c = 0.5$ works well for most natural images.

3.4 Correcting first-order effects

Our technique smooths the spatial variation of local frequency content. However, we found that the non-linearities involved in clamping and log manipulation can also result in changes in first-order properties such as overall sharpness. We correct for this by re-normalizing each subband to the average of the original:

$$s' = s' \frac{mean(|s|)}{mean(|s'|)} \tag{5}$$

This is similar in spirit to Heeger and Bergen's multiscale texture synthesis [6]. We also perform an histogram match on the pixel values from the input to the reconstructed output. This ensures that the average intensity of the image is not altered by our technique.

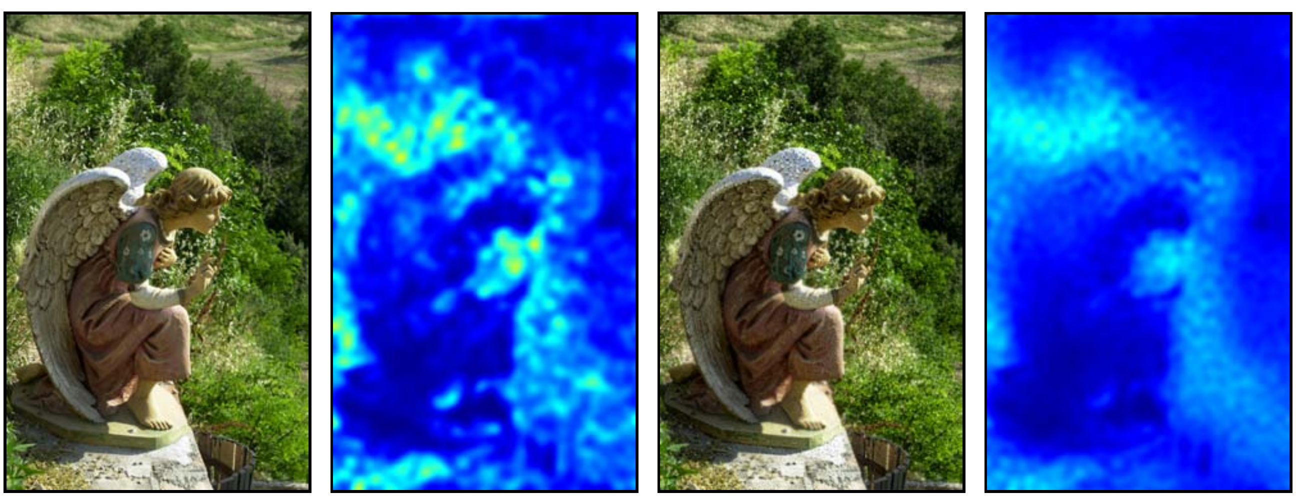
4 Results

We have implemented the technique described in the previous section in Matlab and have used it to reduce the salience of distracting backgrounds in a variety of photographs.

We first show the effect of our technique on an entire image. Figure 3 shows a casual photograph before and after processing with our technique. As can be seen in the false-color visualization of the power maps, texture variation has been reduced across the bottom image. The clear boundaries between regions of high and low texture variation have been softened.

In many natural images, we have also observed that distracting specular highlights in textured regions have been reduced. Note the toned-down highlights in the leaves of Figure 3. Photographers often strive to achieve this effect using polarizers.

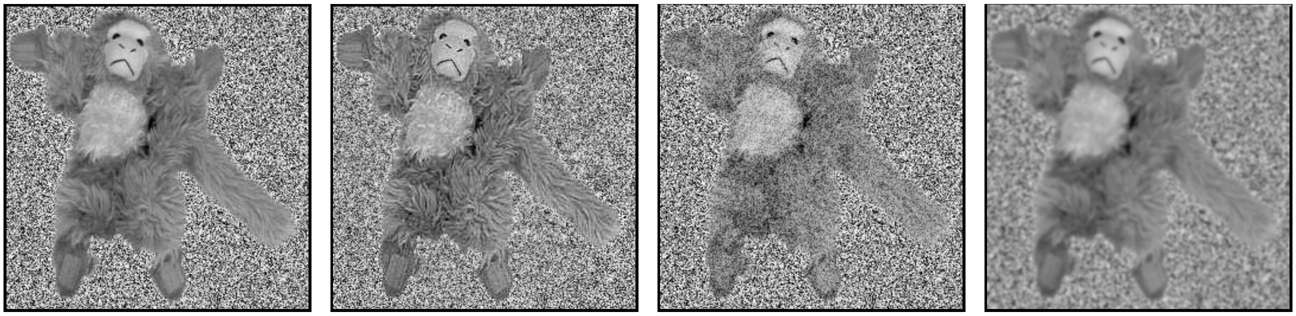
For selective de-emphasis, we use a simple alpha mask and blend processed and unprocessed images. Figure 7 was submitted to us by an amateur photographers who complained that the specular highlights of the leaves are a distraction. Reducing texture variation in this region improves foreground/background separation. Note that in all of these examples, we have applied our technique to only the luminance channel, leaving the chrominance unchanged. This decision is motivated by the low sensitivity of human vision to high frequencies in chrominance.



(a) High frequencies in input

(b) High frequencies after texture equalization

Figure 3: This casual photograph was globally modified to illustrate the effects of reducing texture variation. With high frequencies made more uniform, the texture-equalized image exhibits a "camouflage" effect that masks medium-frequency content. The change in high-frequency distribution can be seen in the corresponding false-color power maps.



(a) Input (b) Texture equalization (c) Additive white noise (d) Gaussian blur

Figure 4: Comparison of de-emphasis techniques.

At first glance, one might guess that our technique simply adds uniform noise. This is not the case. As can be seen in the side-by-side comparison of Figure 4, our technique amplifies existing high frequencies where needed to make texture variation uniform. This strategy preserves key features of the objects in the image whereas simply adding white noise imposes an overall graininess.

Gaussian blur is an alternative de-emphasis technique that can introduce depth-of-field effects. However, the reduced sharpness can be undesirable, particularly if the distracting element is at the same distance as the main subject. In addition, Gaussian blur removes the high-frequency content of an image region, which can emphasize the medium frequencies and result in a more distracting object. (See Figures 4 and 8.) In contrast, our technique makes high-frequencies more uniform, creating a "camouflage" effect that masks medium-frequency content.

Gaussian blur and texture equalization are complementary tools in a photographer's toolkit. Our technique works well when the distracting region is already somewhat textured. Gaussian blur works well when depth-of-field effects are already present and when the medium frequencies are not distracting.

It is not straightforward to do a fair visual comparison between the two techniques because the control parameters are not comparable. In Section 5.1, we discuss the visual search experiment used to determine the size of Gaussian blur kernel comparable to the strength of texture equalization that we have found works well for most natural images.

4.1 Reversing things: "sharpening"

The technique presented so far *reduces* the high frequencies in power maps. It is natural to explore the effect of *increasing* high frequencies, thereby producing a sort of *second-order sharpening*. This can be achieved by simply reversing the sign of k_s in Eq. 3. The technique is expectedly less stable because it increases frequency content that is already high, thereby leading to effects that can appear stylized. Similarly, unsharp mask leads to more artifacts than Gaussian blur. In particular, we have found that it can be dangerous to increase texture variation in the main subject because this is where the beholder's vision is expected to attend and pay more attention to artifacts.

5 Psychophysical validation

Because the goal of our technique is to reduce the salience of distracting regions, we can evaluate its effectiveness using psychophysical experiments. We have conducted two user studies to experimentally validate the effectiveness of our de-emphasis techniques. We have used a visual search task to perform quantitative validation and an eye tracker device for qualitative evaluation.

5.1 Visual search experiment

Saliency is commonly studied through visual search for a target object in the presence of distractors. Subject response time is a reliable indicator of target saliency [8]. We have conducted a controlled search experiment to measure how our de-emphasis technique affects response time. Subjects were shown a series of images and asked to locate a target object as quickly as possible. We compared mean search times for unmodified images and those in which texture variation had been reduced

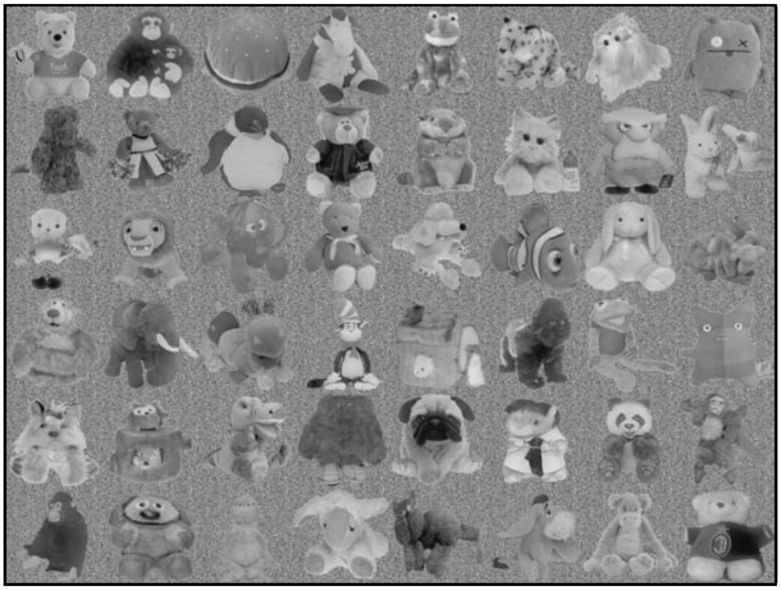


Figure 5: Example search stimuli. After normalization for luminance, the 48 objects in each image are of comparable salience.

everywhere except for the target. The experimental results support our hypothesis that search time is reduced when distractor regions are de-emphasized.

We also use the search task to compare the effect of Gaussian blur and our technique. We point out that this comparison is meaningful only for this type of input images and that generalization requires further experimentation. As discussed above, each technique has its preferred domain of application, and the highly-textured stimuli correspond to our approach's strength. The study however provides an interesting ballpark to quantitatively relate the two approaches.

Stimuli. The stimuli in this experiment were 45 images, each depicting 48 toys on a uniform white noise background (Figure 5). Grayscale images were used to avoid attentional bias for color, and all stimuli have resolution 1600x1200. We refer to the 45 distinct scenes as *layouts*. For each layout, six *conditions* were tested:

- **Original.** The unmodified "flat" image.
- **Texture-equalized.** The background region (everything except for the search target) is texture-equalized. To reduce texture variation, the following parameters were used: low-pass filter $\sigma_l = 5$, high-pass filter maximum $\sigma_h = 60$, high-pass clamping factor $= 0.5$, and final scale factor $k_s = 2$.
- **Gaussian-blurred.** Four conditions were created in which Gaussian blur with $\sigma = \{0.25, 0.50, 1.0, 1.25\}$ pixels was applied to the background.

Images were displayed on a 50 cm LCD screen at a full-screen resolution of 1600×1200. The screen was at a distance of 125 cm to ensure that images appeared fully in the field of view.

Experimental procedure. Data were collected from 12 volunteer subjects. Each subject was shown the series of 45 layouts. For each layout, one of the 6 conditions was randomly displayed. To prevent a learning effect, no subject was shown the same layout twice.

Each subject was shown a search target before viewing a layout and was instructed to locate the target and click twice with the mouse: once immediately upon locating the object and again on the object itself. Time to fixation was approximated by the first-click response time. The second click was used to verify that the target was found. A fixation screen was displayed between consecutive images, and subjects were required to click on the center of the screen to proceed; this ensured that all mouse movements originated at the center of the screen for consistent timing.

Condition	Reaction time (sec)
Unmodified	3.7594
Texture-equalized	2.9160
Gaussian-blurred, $\sigma = 0.25$	4.0446
Gaussian-blurred, $\sigma = 0.50$	3.9288
Gaussian-blurred, $\sigma = 1.00$	3.4382
Gaussian-blurred, $\sigma = 1.25$	3.1234

Table 1: Mean response times for our visual search experiment using stimuli such as Figure 5. De-emphasizing distractors using our texture equalization results in a speedup of more than 20%.

Analysis. The mean response time for the background-equalized images was 2.916 seconds, compared to 3.7594 seconds for unmodified images. This 22.43% speed-up supports our hypothesis that de-emphasizing distractors by reducing texture variation increases salience of target objects.

Three-way ANOVA was used to test the statistical significance of the difference in image condition. A probability $p = 0.0417 < 0.05$ was computed for the condition variable, indicating that the effect of the condition is significant.

Comparing mean response times for the search task, we found that texture equalization of strength $k_s = 2$ produces a change in salience stronger than Gaussian blurring with $\sigma = 1.25$. Extrapolation indicates that $\sigma = 1.5$ would correspond to a similar effect. It may come as a surprise that a small Gaussian blur with $\sigma < 0.5$ increases response time. Our hypothesis is that for the highly-textured images we use, the elimination of high-frequencies removes the "camouflage" effect and enhances the influence of medium frequencies. We can see in Figure 4 that Gaussian blur can enhance the main structure of an object. This does not mean that small Gaussian blur is a bad technique for de-emphasis, but rather that its domain of application is different from that of our technique. Textured image regions are better handled by our approach.

5.2 Fixation experiment

Using an eye tracker, we studied how subjects' gaze paths and fixations changed as they viewed a series of casual photographs before and after modification with our technique.

Experimental procedure. Two versions each of 24 photographs were displayed in random order on a 50 cm CRT screen at a resolution of 1024×768 pixels. Subjects were asked to study each for 5 seconds. Eye movements were recorded by an ISCAN ETL 400 table-mounted eye tracker with an accuracy of 1 visual degree. Subjects' head were secured on an optometric chin-rest to minimize head movement and to maintain a eye-to-screen distance of 75 cm, eye-to-camera distance of 65 cm, and subtended visual angle of 30×20 degrees. The eye tracker output a data file of screen fixations sampled at a rate of 240 Hz.

Discussion. We evaluated the results of the eye tracking experiment by visual inspection of scan paths and fixation maps [21, 2] (Figure 6). This qualitative evaluation supported our expectation that image regions emphasized using our technique would attract and hold fixations. Although this study was less controlled than the search experiment and included fewer subjects, the initial qualitative results are promising. An extended study is future work.

6 Conclusions

We have presented a novel post-processing technique to reduce the salience of distracting regions in an image. 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 reducing variation in texture. We use steerable pyramids to define a set of power maps which capture local frequency content at each scale and orientation and provide a perceptually-meaningful tool for image manipulation. Psychophysical evaluation showed that the technique reduces the salience of modified regions.

Our de-emphasis technique is complementary to existing post-processing methods such as Gaussian blur that increases depth-of-field effects. Our technique is most efficient for textured image regions, while Gaussian blur works best when small depth-of-field effects are already present and when medium-frequency content is not distracting.

(a) Scan path for input

(b) Scan path after texture equalization of distractors

Figure 6: Scan paths for images before and after texture equalization. The photograph was modified to emphasize the leftmost boy and the girl in the upper left. Small green circles indicate saccadic jumps recorded by the eye tracker, while red circles indicate fixations, with the duration indicated by the radius of the circle.

Areas of future work include the application of such image-manipulation methods for the study of bottom-up visual attention. We are planning more extensive experiments to study the variables that influences the effectiveness of de-emphasis and emphasis techniques. The combination of first-order features such as sharpness and brightness with our second-order features is an exciting topic, which raises the challenging task of appropriate calibration. Our search experiment provides a first data-point, but more data is needed to fully understand the effect of image content. Finally, the idea of image processing on texture feature space has potential for image in-painting and restoration.

Acknowledgments

We thank Aude Oliva and Ruth Rosenholtz and their students for use of their eyetracker and advice on experimental design. Paul Green and Eric Chan provided invaluable assistance with data acquisition and analysis for the user study. Thanks to Eric, Paul, Barbara Cutler, Tom Mertens, Sylvain Paris, and Peter Sand for feedback on the text, and to members of the MIT Graphics and Vision groups for volunteering to be test subjects. David Salesin and Patrice Simard hosted SLS at Microsoft Research during the early stages of this project. This material is based on work supported by the National Science Foundation under Grant No. 0429739 and the Graduate Research Fellowship Program, MIT Project Oxygen, and the Royal Dutch/Shell Group.

References

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

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

[3] 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.

[4] 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.

[5] 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.

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

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

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

[9] 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. MIT Press, 2004.

[10] 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.

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

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

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

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

[15] 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.

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

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

[18] 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.

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

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

[21] 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.

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

(a) Input

(b) Texture equalization of background

Figure 7: Specular highlights in the leaves and other distractors prevent clear foreground/background separation in the original image. Reducing texture variation in the background de-emphasizes these distractors, thereby increasing salience of the intended subject, the tiger.

(a) Input

(b) Gaussian blurring of background

(c) Texture equalization of background

Figure 8: Gaussian blur de-emphasizes everything in the image except for the left tiger by introducing depth-of-field effects. The reduced sharpness is undesirable because of the new conflicting depth cues between the two tigers, which should appear at the same distance. Reducing texture variation in the background effectively de-emphasizes without this effect.