S. Su, F. Durand and M. Agrawala: De-emphasis of distracting image regions using texture power maps. In *Texture 2005: Proceedings of the 4th International Workshop on Texture Analysis and Synthesis*, pp. 119–124, 2005.

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

Abstract

We present a post-processing technique that selectively reduces the salience of distracting regions in an image. Computational models of attention 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. Modification of power maps results in effective regional de-emphasis. We validate our results quantitatively via a human subject search experiment and qualitatively with eye tracking data.

1. Introduction

Much of the art of photography involves directing viewers' 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, one low-level feature that cannot be directly manipulated with existing image-editing software is *texture variation*. Variations and outliers in texture are salient to the human visual system [13, 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, 19, 10].

We introduce a technique for selectively altering texture variation to reduce the salience of an image region. 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 (Sec. 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 flexible, effective representation for *manipulating* frequency content in an image. We introduce a perceptually-motivated technique for selective manipulation of texture variation (see Fig. 1).

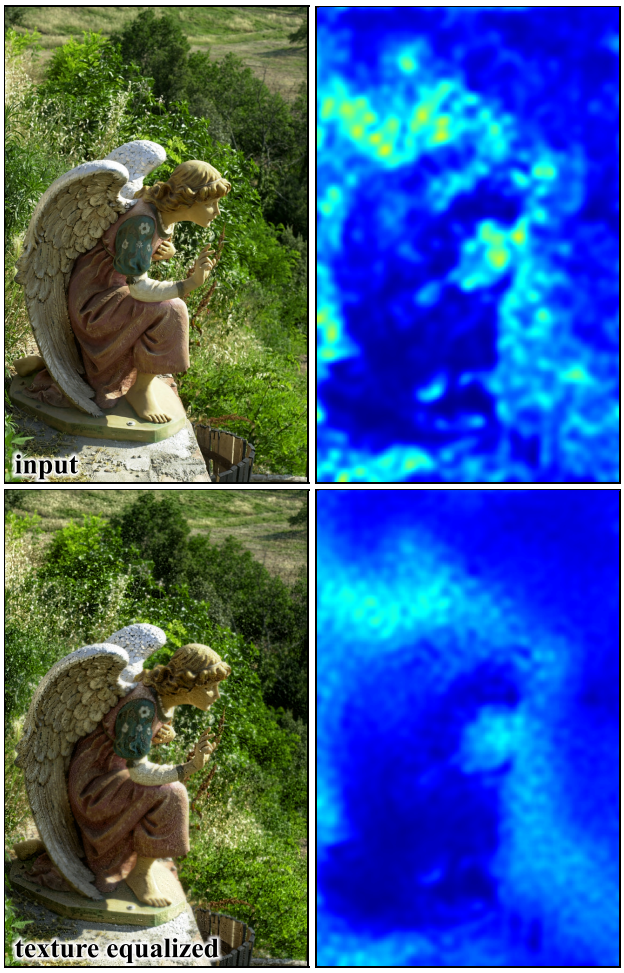
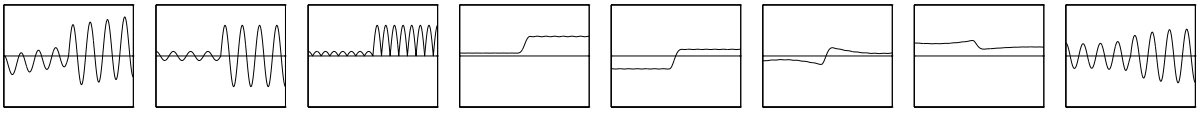


Figure 1. High frequencies have been made more uniform in this texture equalized image. False-color power maps show the change in high-frequency distribution.

Psychophysical study of texture and attention. We conduct two user studies as experimental validation of our technique's effectiveness: A search experiment to measure quantitatively the effectiveness of our technique at directing attention in an image and an eye tracking experiment to record qualitative changes in fixations and scan paths.



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

Figure 2. Texture discrimination and manipulation in 1D. Please see the detailed description in Sec. 2.1.

2. Background

2.1. Texture segmentation and discrimination

Texture discrimination and texture edge detection have 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 roughly follow Malik and Perona's biologically-inspired model [11], illustrated with a 1D example in Fig. 2. The first stage of most texture discrimination models is linear filtering with multi-scale oriented Gabor-like functions (Fig. 2(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 to obtain a meaningful measure of the filter response in a neighborhood. Possible solutions include full-wave rectification (absolute value) and energy computation (square response); the absolute value is shown in Fig. 2(c). Low-pass filtering (pooling) of this response produces the local average of the filter response strength; we call this the *power map* (Fig. 2(d)).

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 top-down and bottom-up processes. Top-down mechanisms, which describe how attention is influenced by scene semantics or the task, are important to understanding attention. However, in this paper, we focus on image processing 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 [16, 14, 15]. 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 nonlinearities 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 [18, 12, 21].

In psychophysical experiments, Einhäuser and König [3] observed salience effects due to texture variation that could not be explained by first-order models. The *second-order* model recently introduced by Parkhurst and Niebur [13] captures these effects by performing the computation of first-order models on the responses to a first-order filter bank (what we call power maps) rather than on image intensity. This strategy motivates our method of manipulating power maps to alter contrast in texture.

3. Texture equalization

We introduce 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 multiscale oriented filters. A plethora of such filters have been developed for texture discrimination. We use *steerable pyramids* because they permit straightforward analysis, processing, and near-perfect reconstruction of images [4, 17].

3.1. Power maps to capture local energy

We compute power maps using the texture discrimination approach of Sec. 2.1. Local frequency content is computed using steerable pyramids, and a power map is computed for each subband s. Because s is band-limited and has local average is zero, we perform a full-wave rectification, taking the absolute values of the steerable coefficients. We apply a low-pass filter with a Gaussian kernel g_l to compute the local average of the response magnitude; we call the resulting image s_l the *power map*.

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

We 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 works consistently well. Note that because the low-pass filter has the same size for

each subband, 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 results in negative values that are invalid power map coefficients; the power map is computed from absolute values. While these invalid coefficients do not interfere with analysis, for image editing they must me scaled rather than summed. We perform all subsequent processing in the natural logarithmic domain of the power map. 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. High-pass filtering of the power maps reveals the spatial variation s_h of frequency content over the image. 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 works 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 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:

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

Values of $k = 1,2,3$ to work well. At the boundary between low and high values of the power map, the high-pass of the log power map goes from negative to positive, resulting in a scaling up the coefficients on the low side and scaling down on the other side (Fig. 2 (g) and (h)).

Clamping. Uniform regions correspond to zero values of the power map. When adjacent to highly-textured regions, they result in extreme high values of the high-frequency of the power map s_h, resulting in a large applied scaling factor that can amplify the small amount of noise present in uniform regions of the original subbands. To prevent such artifacts, we use a simple non-linearity to clamp isolated extreme values in the scaling (high-pass response) map to a 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.

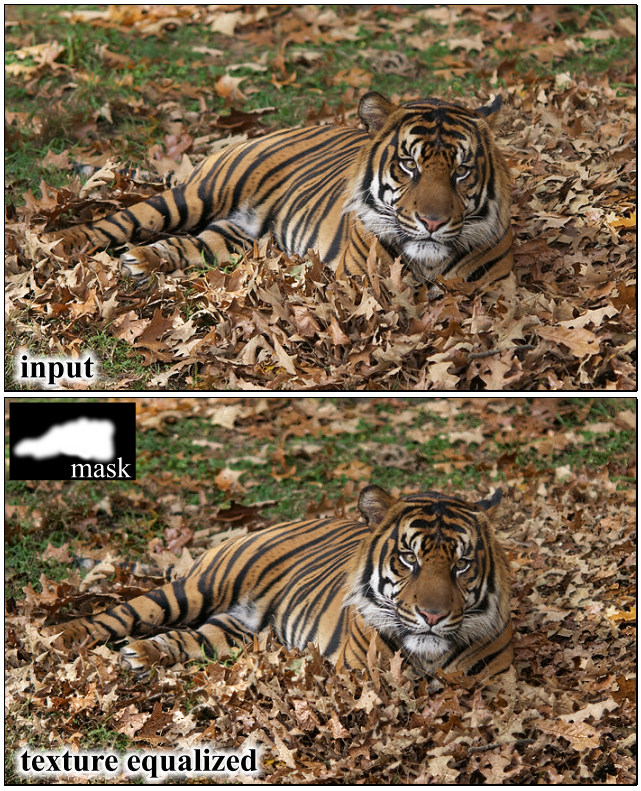


Figure 3. Highlights in the leaves and other distractors prevent clear foreground/background separation in the original photograph. Texture equalization de-emphasizes these distractors, increasing salience of the tiger.

3.4. Correcting first-order effects

Our technique smoothes 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 first-order change by re-normalizing each subband to the average of the original:

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

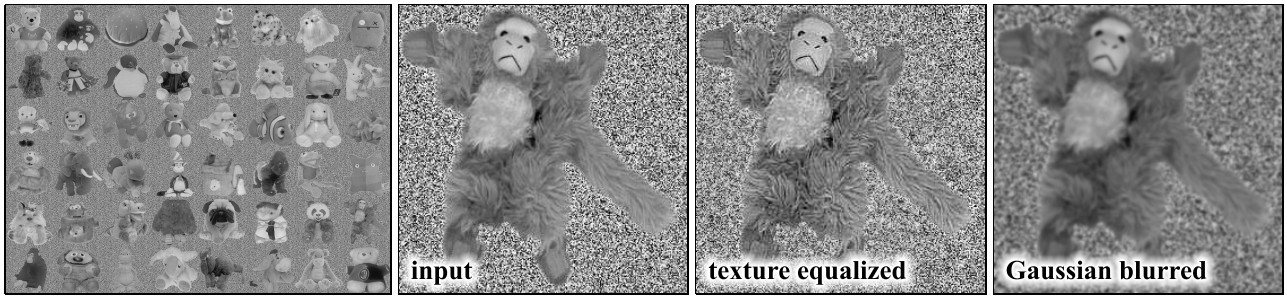


Figure 4. Example search stimulus and close-ups of de-emphasis techniques applied to a stimulus object.

This is similar in spirit to Heeger and Bergen's multiscale texture synthesis [6]. We also perform a 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 our texture equalization method in Matlab and have applied it to a variety of images. Fig. 1 shows a texture equalized photograph. The false-color visualization of the power maps shows how texture variation has been reduced and boundaries between regions of high and low texture variation softened.

For selective de-emphasis, we use an alpha mask and blend processed and unprocessed images. In Fig. 3, we reduce texture variation in the leaves surrounding the tiger, improving foreground/background separation. Note that 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.

At first glance, one might guess that texture equalization simply adds uniform noise. Our technique amplifies existing high frequencies to make texture variation uniform. This strategy preserves key features of the objects in the image while adding white noise imposes an overall graininess.

Gaussian blur is an alternative de-emphasis technique that can introduce depth-of-field effects. The reduced sharpness can be undesirable, particularly if the distracting element is at the same distance as the main subject. In addition, blur removes the high-frequency content of an image region, which can emphasize the medium frequencies and result in a more distracting object. (see Figs. 4 and 6.) In contrast, equalization makes high-frequencies more uniform, creating a "camouflage" effect that masks medium-frequency content. Gaussian blur and texture equalization are complementary tools in an artist's toolkit. Our technique works well when the distracting region is already somewhat textured. Blur works well when depth-of-field effects are already present and medium frequencies are not distracting.

Please see the full-resolution, color images at `http://csail.mit.edu/~sarasu/pub/texture05`.

5. Psychophysical validation

We have conducted two psychophysical experiments to evaluate the effectiveness of our de-emphasis technique: a visual search task for quantitative validation and eye tracking 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 recorded subject responses to unmodified images and those in which texture had been equalized everywhere except for the search target, finding that search time is reduced when distractors are de-emphasized. We also used the search task to compare Gaussian blur and texture equalization.

Experimental procedure. Data were collected from 12 volunteers. Each subject was shown a series of 45 stimulus images at 1600×1200 resolution. Each image depicted a collection of objects arranged in a distinct *layout* on a uniform white noise background (Fig. 4). Grayscale images were used to remove attentional bias for color. For each layout, one of six *conditions* was randomly displayed:

Original. The unmodified image.

Texture-equalized. All parts of the image, except for the search target, are 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. Blur of $\sigma = \{0.25, 0.50, 1.0, 1.25\}$ pixels is applied to all parts of the images except the target.

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

Condition	Mean response time	Std. Dev.
Unmodified	3.7594 s	0.8422
Texture-equalized	2.9160 s	0.6698
Blurred, $\sigma = 0.25$	4.0446 s	0.9519
Blurred, $\sigma = 0.50$	3.9288 s	1.0339
Blurred, $\sigma = 1.00$	3.4382 s	0.6171
Blurred, $\sigma = 1.25$	3.1234 s	0.6193

Table 1. Mean response times for search experiment. Texture equalization results in a speed-up of more than 20%.

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. Trials in which the users second click did not match the search target were discarded from our timing analysis. To prevent a learning effect, no subject was shown the same layout twice.

Analysis. The mean response time for the texture 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.

Two-way ANOVA tested the statistical significance of variables *layout* and *condition*. For *layout*, $p \leq 0.1985$; as expected, this does not achieve the level of significance. For *condition*, $p \leq 0.0487$, indicating that it is a statistically significant variable. A two-sample t-test comparing the data collected in the unmodified and texture-equalized conditions indicated that the null hypothesis can be rejected at the 5% significance level; the difference in timings was not due to chance.

The experiment shows that texture equalization of strength $k_s = 2$ produces a change in salience stronger than Gaussian blurring with $\sigma = 1.25$. It may come as a surprise that a Gaussian blur with $\sigma < 0.5$ increases response time. We hypothesize that for highly-textured images, the elimination of high frequencies removes the "camouflage" effect and enhances the influence of medium frequencies, object structures (see Fig. 4).

5.2. Fixation experiment

Experimental procedure. Using an eye tracker, we studied how 4 subjects' gaze paths and fixations changed as they viewed a series of photographs before and after modification with our technique. Two versions each of 24 photographs were displayed in random order at a resolution of 1024×768 pixels. Subjects were asked to study each for 5 seconds while their eye movements were recorded with an ISCAN ETL 400 table-mounted eye tracker.

Discussion. We analyzed the eye tracking data by visual inspection of scan paths [20, 2] Fig. 5 shows how the

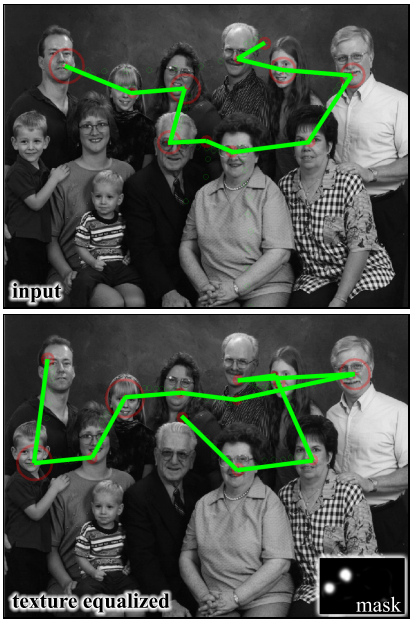


Figure 5. Change in scan paths after texture equalization. Red circles mark fixation points; duration is indicated by the circle radius. See webpage for full-resolution images.

salience of regions can be increased by equalizing the surrounding texture. These emphasized regions attract and hold subjects' fixations. Although this study included fewer subjects, the qualitative results are promising and support our hypothesis that texture variation is a salient feature. An extended study is future work.

6. Conclusions

Inspired by bottom-up models of visual attention, our texture equalization technique reduces the salience of distracting image regions by reducing variation in texture. We use steerable pyramids to define a set of power maps capturing local frequency content and provide a perceptually-meaningful tool for image manipulation that complements other post-processing methods such as Gaussian blur. Our technique is effective for textured image regions, while blur works best when small depth-of-field effects are already present and medium-frequency content is not distracting.

Future work includes the application of such image-manipulation methods to the study of bottom-up visual attention. Our search experiment provides a first data point, but more are needed. We plan more extensive experiments to study the variables that contribute to a technique's effectiveness. The combination of first-order features (e.g. sharpness and brightness) with our second-order features raises the challenging task of appropriate calibration. Finally, image processing in the texture feature space has potential applications in image in-painting and restoration.

7. Acknowledgments

We thank the MIT Graphics Group and anonymous reviewers for feedback; P. Green and E. Chan for invaluable assistance with data acquisition and analysis; and A. Oliva, R. Rosenholtz, and their students for use of their eye-tracker. This work was supported by NSF under Grant No. 0429739 and the Graduate Research Fellowship Program, MIT Project Oxygen, and the Royal Dutch/Shell Group.

References

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

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

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

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

[5] C. G. Healey, L. Tateosian, J. T. Enns, and M. Remple. Perceptually based brush strokes for nonphotorealistic visualization. *ACM Transactions on Graphics*, 23(1):64–96, 2004.

[6] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *Proceedings of ACM SIGGRAPH 95*, 1995.

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

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

[9] M. S. Landy and N. Graham. Visual perception of texture. In L. M. Chalupa and J. S. Werner, editors, *The Visual Neurosciences*, pages 1106–1118. MIT Press, 2004.

[10] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, 43(1), 2001.

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

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

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

[14] C. M. Privitera and L. 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.

[15] P. Reinagel and A. M. Zador. Natural scene statistics at the centre of gaze. *Network: Compututational Neural Systems*, 10:1–10, 1999.

[16] R. Rosenholtz. A simple saliency model predicts a number of motion popout phenomena. *Vision Research*, 39, 1999.

[17] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Proceedings of IEEE International Conference on Image Processing*, pages 444–447, 1995.

[18] R. L. Solso. *Cognition & Visual Arts*. Bradford Books, 1996.

Figure 6. Gaussian blur de-emphasizes everything in the image except for the left tiger by introducing depth-of-field effects. Texture equalization de-emphasizes without the conflicting depth cues introduced by blur.

[19] C. Ware. *Information Visualization: Design for Perception*. Academic Press, 2000.

[20] D. S. Wooding. Fixation maps: quantifying eye-movement traces. In *Proceedings of Symposium on Eye Tracking Research & Applications*, pages 31–36, 2002.

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