Appearance of images

Karen K. De Valois[a], Tatsuto Takeuchi[b], Thomas D. Wickens[a]

[a]Psychology Department and Vision Science Group, University of California, Berkeley, CA, USA
[b]NTT Communication Science Laboratories, NTT Corporation, Atsugi, Kanagawa, JAPAN

ABSTRACT

What makes an image appear to be a veridical representation of a real scene? Knowing what is necessary to produce a "good" image also aids in the design of more efficient compression algorithms. We review our earlier work on video compression and demonstrate the substantial savings and excellent image quality produced by spatial low-pass filtering of most (but not all) of the individual frames. Currently, we work with still images. An example will show that simple filtering can produce unexpected changes in the perceptual interpretation of a complex scene. I will describe and demonstrate a new compression method we are developing based on the assumption that the fine structure in the amplitude domain (and perhaps in phase, as well) can be of minimal importance in conveying the essence of a scene. We find that a complex image can be reproduced surprisingly well by compressing the entire spatial frequency amplitude spectrum to a very small number of terms.

Keywords: image appearance, motion sharpening, spatial frequency filtering, amplitude spectra, color

1. INTRODUCTION

The appearance of images, whether moving or still, is affected by many factors. Some of these are obvious, such as the difference between colored and grey-scale images. Some are not. Even when a particular modification is immediately apparent, some of its effects may not be. We are interested in the effects produced by adding to, subtracting from, or otherwise modifying images. We are especially interested in determining how the appearance of an image can be improved. Of course, improving an image presupposes that it is possible to specify the criteria that make an image good, but those are not fixed and consistent across all situations. A radiographic image used for medical diagnosis, say, should meet standards quite different from those pertaining to a casual vacation photograph. For purposes of this discussion, we shall consider a good image to be one that appears to be a veridical representation of a real scene, but even this loose definition has problems. An image of an actual scene could be modified in such a way that though it appears to be complete and natural, when compared directly to the scene it purports to represent, the differences would be strikingly obvious. Given the complexities of the general problem, we have begun by examining several of the factors that contribute to the appearance of an image.

We began our studies of image appearance by working with moving images to study the process of motion sharpening. First we shall review some of our work on moving images. Much of this work was first presented at an earlier Human Vision and Electronic Imaging meeting. Then we shall describe briefly three of the factors we are currently studying in terms of the ways in which they affect the appearance of still images.

2. MOTION SHARPENING

Although moving images are often blurred on the retina, they sometimes appear to be more sharply focused than they actually are, a phenomenon known as motion sharpening[1]. Several models have been proposed to account for motion sharpening (see, for example, Burr and Morgan, 1997[2], Hammett, Georgeson and Gorea, 1998[3], Pääkkönen and Morgan, 2001[4], Georgeson and Hammett, 2002[5].We wondered in what way a moving image actually appears to differ from the

Human Vision and Electronic Imaging XIII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6806, 680605, © 2008 SPIE-IS&T · 0277-786X/08/$18

same image when it is stationary. In research reported earlier[6], we considered three ways in which a moving image exhibiting motion sharpening might appear to differ from its individual (stationary) frames. One possibility is that the amplitude spectrum of the image appears to extend to higher spatial frequencies. Sharp edges that are not actually present in the image might then be perceived. Another possibility is that the overall contrast of the image appears greater. Increasing the overall contrast is equivalent to multiplying the amplitude at every spatial frequency by a factor greater than 1. A third possibility is that the moving image might look as though the amplitude of its component spatial frequencies had been increased as a function of spatial frequency, with high frequencies being amplified more than low frequencies. This change is equivalent to decreasing the slope of the amplitude function in the log spatial frequency domain. These three possibilities are illustrated in Fig. 1.

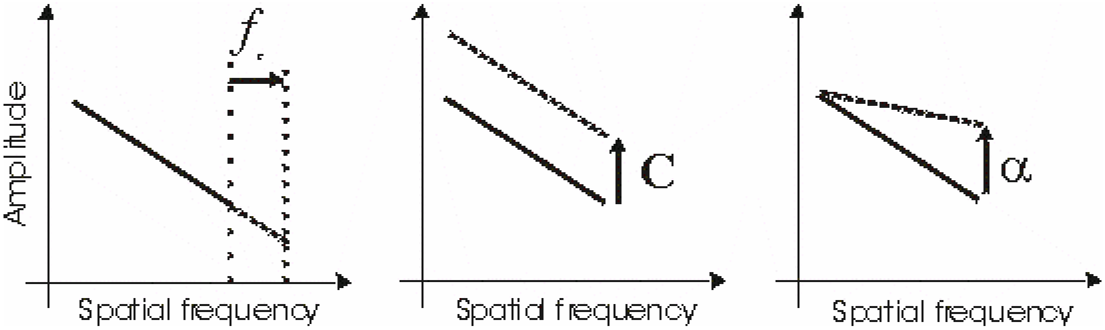


Fig. 1.The three parameters (f_c, C, α) that observers were allowed to manipulate in the still images. Both axes are on a logarithmic scale. (left) The cutoff frequency could be extended to high spatial frequencies; (center) the overall contrast could be increased by multiplying the amplitude at every spatial frequency by the same factor; (right) the slope of the amplitude function could be changed by increasing the amplitude at high spatial frequencies more than at low frequencies.

To evaluate these three models, we asked naïve subjects to compare a moving videotaped image to a set of still frames drawn from the same image sequence. Each still image was presented for 1 sec, then replaced by another. The moving image moved continuously. Each frame of the movie (and the still images taken from it) had previously been low-pass filtered. Observers were given control of the cutoff spatial frequency f_c, the overall luminance contrast, the slope of the amplitude function in the spatial frequency domain, or some combination of those parameters in the still images. They were instructed to make the still images appear as similar as possible to the moving images. When an observer increased f_c, spatial frequency components that before were not present in the image now appeared. When a subject increased the overall contrast C, the amplitude at each spatial frequency was multiplied by the same factor. With this, the overall luminance contrast appeared greater. When an observer changed the slope of the amplitude function, the amplitude at each spatial frequency was multiplied by f^{α}, where f is spatial frequency.

We assessed the amount of motion sharpening by comparing the total power in the adjusted still images to that in the moving images, using a modified motion energy model[7,8]. If the two were equal, no sharpening would have been seen. When the ratio was greater than 1, motion sharpening occurred. When the ratio was less than 1, there was motion blurring. The results were unambiguous. Although observers could reliably set a best-matching high-spatial-frequency cut-off[9], they often remarked that the match was unsatisfactory. When allowed to adjust all three parameters, observers frequently manipulated only the slope of the amplitude function. When each parameter was manipulated independently in separate trials, the amount of motion sharpening reflected by the change in the slope of the amplitude function was markedly greater than that seen by adjusting either of the other two parameters. It appears, thus, that motion sharpening is best characterized as an apparent change in the slope of the amplitude-spatial frequency function, with the higher frequencies being increased in amplitude more than the lower ones.

One particularly interesting observation is that motion sharpening depends on the spatial frequency content of the images. We examined the role of different spatial frequency regions by first filtering each frame of a movie, then assessing the amount of motion sharpening using the same matching procedure we had used before. Fig. 2 shows the results. In the original images, motion sharpening was substantial. When the images were spatially low-pass filtered, an intermediate degree of motion sharpening occurred. However, when the images were high-pass filtered, not only was there no motion sharpening, there was substantial motion blur. The moving images appeared to be of lower quality than the stationary ones. This is not surprising in light of the frequent blurring of moving images on the retina.

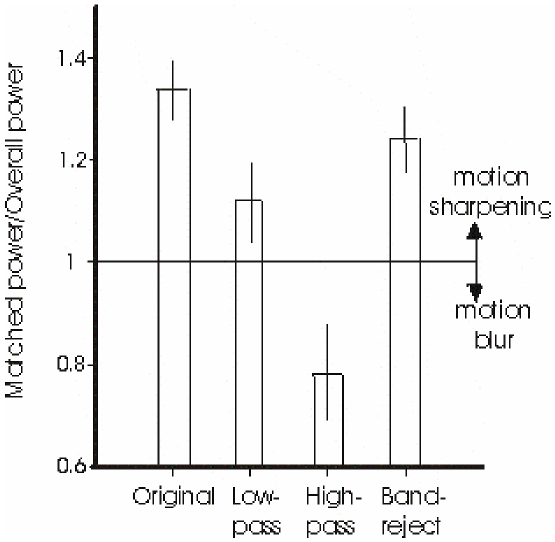


Fig. 2. Motion sharpening in movies to which various filters have been applied. Note that a movie with only high spatial frequency information (“High-pass”) shows motion blur, not sharpening. When only low spatial frequencies are present (“Low-pass”) there is a small amount of motion sharpening, but when both high and low frequencies are present (Band-reject”), motion sharpening is increased.

It is tempting to conclude from this result that the presence of high spatial frequencies actually counteracts motion sharpening and only contributes blur to the appearance of the moving image. The observations made using a band-reject filter, however, demonstrate that this interpretation is incorrect. Though the high frequencies alone apparently do not produce sharpening, the combination of high and low spatial frequencies substantially increases the amount of motion sharpening. With only intermediate spatial frequencies removed, we find more motion sharpening than with the low frequencies alone. Apparently there is an interaction between the high and low frequencies that is important in motion sharpening.

In the course of this research, we noticed that when we were working with a low-pass filtered movie, the addition of a small number of unfiltered frames greatly improved the apparent sharpness of the movie. Accordingly, we examined this phenomenon more closely. We varied both the cut-off spatial frequency of the low-pass filtered images and the proportion of unblurred frames inserted into the movie. We found that when 50% of the frames were significantly blurred (by applying a low-pass filter with a cut-off of 8 c/deg) and 50% were not, the modified movie appears virtually identical to the original. Using greater degrees of blur and varying the percentage of unblurred frames demonstrated that this phenomenon is robust. It is possible to substitute low-pass filtered images for at least 70% of the frames without substantially compromising the appearance of the movie. We note that by occasionally introducing an unfiltered frame with its high spatial frequency components, we are reproducing the stimulus conditions that gave the strongest motion sharpening—namely, the presence of both high and low spatial frequencies. The low sensitivity of the visual system to temporal change with high spatial frequency targets (see, for example, Watson and Nachmias[10]) may account, in part, for the fact that the high spatial frequency information does not need to be refreshed very often. We are also aware that masking may play a significant role in this effect.

Success in compressing a movie by using a large proportion of low-pass-filtered frames raised the question of whether temporal filtering would be equally interesting. We treated each pixel of a movie separately, taking the input to a single pixel over time as a function for analysis and filtering. As with spatial filtering, we found that low-pass filtering the signal to a large proportion of the pixels did not significantly degrade the appearance of the movie, as long as a proportion of pixels continue to receive and display the high temporal frequency components of the signal. Again, as in the space domain, the most significant motion sharpening occurred when both high and low temporal frequency components were present in some pixels. Thus, low spatial frequency information must be supplied at high temporal frequency, while high spatial frequency information is acceptable at low temporal frequency. Our work suggests that substantial compression can be realized by sampling high frequency information much more sparsely than low, either in space or time.

3. SPATIAL FREQUENCY FILTERING

The perceptual conclusions an observer draws about an image can sometimes be altered in significant and surprising ways by simple operations. We are all familiar with the perceptual consequences of either high-pass or low-pass spatial frequency filtering. In the former, edges are emphasized, and the scene will look much like a line drawing if the low-frequency cut-off is sufficiently high. Low-pass spatial filtering, on the other hand, reduces or eliminates edges and blurs the entire scene. It is less common to see a scene that has been modified with a band-reject filter. We wondered what roles the intermediate spatial frequencies play in determining the appearance of an image. We did not use a traditional band-reject filter. Rather, we defined a filter function that is sinusoidal in log spatial frequency, with its frequency, amplitude and phase variable. For the example we show here, the frequency of the sinusoid was 1 c/image, its amplitude was 1, and its phase was set such that f0 was transmitted without attenuation. Thus, the most strongly attenuated components were in the intermediate frequencies.

Consider the photograph of primroses in Fig. 3 (a). The pattern of shadows and highlights is consistent with a single source of illumination above and to the left. However, note what happens when the image is filtered by selectively reducing the amplitude of the intermediate spatial frequencies (b). Although the shadows remain visible, the scene itself appears different in an interesting way. Now, rather than having a single illuminant coming from above and to the left, there appear to be several light sources in the scene. Some of the flowers themselves seem to glow. The regions that appear to emit light are brighter than their immediate surrounds, but they are not necessarily the brightest objects in the scene. In fact, when a scene contains local areas of specular reflection, those regions frequently do not appear to glow.

(a) Original image of primroses

(b) Band-reject filtered image of primroses

(c) Blurred (i.e., low-pass) image of primroses

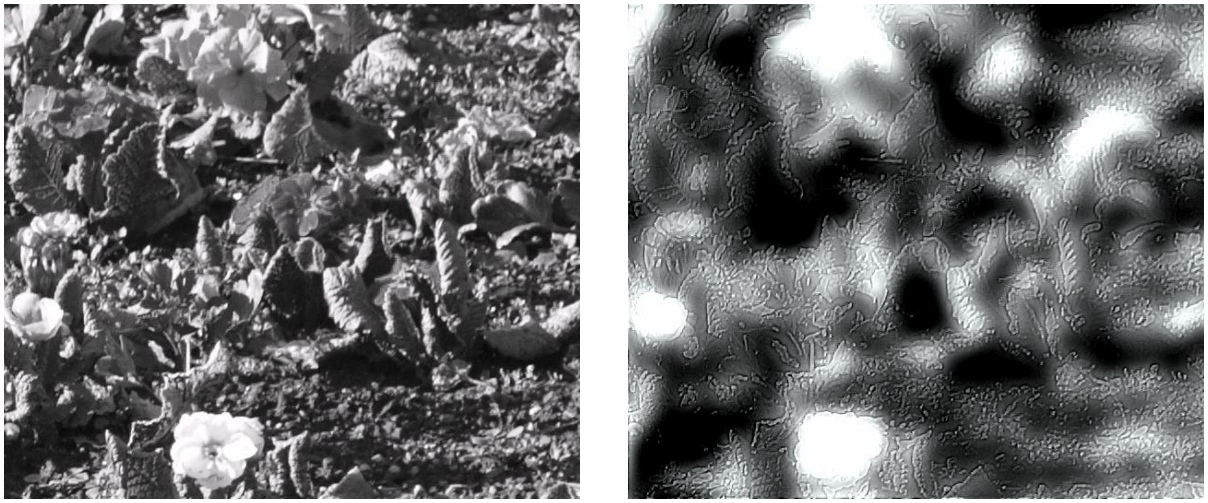
Fig. 3. The original image (a) has obvious shadows and highlights, indicating that a single illuminant is above and to the left. The band-reject filtered image (b) retains the shadows and highlights, but some local areas appear to be emitting light. Notice particularly the white flower at the bottom of the image. See text for filter description. The third image (c) has been deliberately blurred, as by a low-pass filter. Note that it does not appear to have self-luminous areas.

A possible explanation of this effect might be a reduction in the masking of low spatial frequencies by intermediate frequencies. The low frequencies alone could produce the appearance of glowing when they are released from masking. If so, a low-pass filtered version of the same image should also seem to emit light. Note in Fig. 3 (c), however, that no glowing appears in a blurred image in which both high and medium spatial frequencies have been attenuated. We conclude that in this phenomenon, as in the motion sharpening discussed above, both high and low spatial frequency components are required to produce the maximum effect. Apparent light emission seems to occur when the edges of an object are visible, but the light associated with it spreads beyond the edges.

The appearance of self-luminosity is not confined to colored images. Fig. 4 shows the same filtered primrose-garden image in a luminance-only version. Here we display both filtered and original versions of the photograph in gray-scale images. Recall that the filter selectively attenuates intermediate spatial frequencies. In the filtered, luminance-only image, the objects themselves (primroses) cannot readily be identified, but there are still local regions that seem to glow.

The difficulty in recognizing the objects in the images in which the middle spatial frequencies have been attenuated is surprising. Both low-pass-filtered and high-pass-filtered images are usually recognizable. The particular filter we used attenuates the high spatial frequencies somewhat, and this may contribute to the difficulty in recognizing the objects, but visible edges do remain. In a much larger-scale image, we do not notice as much difficulty in recognition. These deleterious effects of severely attenuating the middle spatial frequencies were unexpected. Their role in normal object recognition may be more important than we have previously thought.

Finally, though it is not obvious in the examples we have reproduced here, we have observed another curious phenomenon that often appears when intermediate spatial frequency components are selectively attenuated. Some objects appear to become transparent. The edges of the object are still perfectly visible (because of the remaining high spatial frequencies), but nearby objects or surfaces appear to extend visibly beneath the transparent region. We believe that the frequency characteristics of the filter must be well matched to the spatial characteristics of the image for the transparency to become obvious.



(a) Luminance-only primrose image

(b) Band-reject filtered luminance primrose image

Fig. 4. A luminance-only version of the image shown in Fig. 3. Note that local regions in the band-reject filtered version (b) appear to be self-luminous, but there are no such regions in the unfiltered version (a).

4. COLOR

In asking what characteristics make an image good, we wondered how significant color might be. There is a substantial literature (to which we have contributed; for example, De Valois, 2003[11]) on the question of what roles color plays in normal vision. Color is useful in distinguishing between luminance contrast produced by shadows and luminance contrast that indicates object borders. Color is also useful in the perceptual linking of different parts of an object that are separated in an image because of interposition. Of course, displaying an image in shades of gray eliminates the information about the hues associated with different objects in the scene. Hue, however, is sometimes considered to be a secondary characteristic, one not necessarily required for segmenting a natural scene (except in the special case mentioned above) or the identification of objects. If a scene were truly isoluminant, without color it would be invisible. Isoluminance rarely occurs in nature, however, and gray-scale versions of images usually seem to convey most of the spatial information in a scene.

When we examined full-color and gray-scale versions of the same images, though, we were interested to note that the addition of color often revealed objects or even major segments of a scene that were completely missed in the grayscale image. The apparent loss of scene elements can occur even when the objects or segments have abundant luminance cues. Fig. 5 illustrates this effect. Inspect the gray-scale image. Carefully notice the various parts of the scene. Now compare it to the full-color version of the same scene. The row of red flowers in the background is rarely noticed or identified in the gray-scale image. Those flowers are not isoluminant with their background, and their internal structure produces a considerable amount of luminance contrast, but the luminance cues that are present do not readily lead to the segregation of that part of the scene of the identification of the flowers. Here, then, although the gray-scale image appears to be a good representation of the scene, it fails to convey all of the important information. Color is required to make the image a good representation of the scene.

(a) Gray-scale image of an outdoor scene

(b) Full-color version of the same image

Fig. 5. Luminance-only (a) and full-color (b) versions of the same outdoor scene. Notice the very obvious row of flowers in the background. Despite the fact that the luminance-only version contains luminance contrast corresponding to the flowers, they are not readily seen without color.

5. SPATIAL FREQUENCY AMPLITUDE VARIATIONS

Field[12] (1987) brought to the attention of the vision community the striking similarities in the spatial frequency amplitude spectra of natural scenes. He emphasized the fact that amplitude falls approximately as 1/f **.** When they are plotted on log-log coordinates, the amplitude spectra can generally be well fit by straight lines. The strong dependence of amplitude on spatial frequency made us wonder how significant the small variations around the best-fitting straight line were. Earlier work (Oppenheim and Lim[13], 1981; Piotrowski and Campbell[14], 1982) had shown that when the amplitude spectrum of one complex natural image was combined with the phase spectrum of another, the result more closely resembled the image from which the phase spectrum was taken than the one from which the amplitude spectrum came. We wanted to see whether we could replace the entire Fourier amplitude spectrum with a simple function that would convey just the slope and intercept of the best-fitting straight line.

One option would have been to average the amplitude spectrum across all orientations, as Field had done, then use the slope and intercept of the line that best fit the resulting function. However, Torralba and Oliva[15] (2003) have reported significant asymmetries in the amplitude spectra at different orientations, associated with the character of the scene (e.g., landscape, cityscape, etc.). We have also found appreciable orientation asymmetries, depending upon the particular scene. We decided to treat the amplitude spectrum as a surface and fit it with a simple function that took these asymmetries into account. We treated the f0 amplitude separately from the rest of the spectrum. We fitted the spectrum less the f0 amplitude with a smooth 2-dimensional function for which each radius was a straight line in the logarithm with a slope that varied with orientation according to a simple rule—rather like a conical tent of varying radius. For the slopes we tried low-order polynomials and harmonic functions. The four to six terms of these functions, plus the f0 term, now completely specified the amplitude spectrum.

The new artificial amplitude spectrum was combined with the phase spectrum from the original scene. Fig. 6 below shows (a) an original photograph of a complex natural scene, and (b) the image created by combining the simplified amplitude spectrum with the phase spectrum of the original. Although there are clearly visible deviations—in particular, a general loss of contrast—the result is promising. We have not captured everything of significance, but we believe that what is missing may be relatively simple. In future work, we plan to compare the computed amplitude spectrum to the actual spectrum and modify any point that deviates by more than some specified amount.

(a) Squirrel, original image

(b) Squirrel, modified image

We do not believe that a similar smoothing operation will work for phase, even though there are some consistencies across images (Thomson 1999a[16], 1999b[17]). Therefore, we plan to explore the possibility of simplifying the phase spectrum by binning it into a small number of levels. This binned phase spectrum could be combined with a simplified amplitude spectrum. We hope in this way to encode efficiently the information necessary to produce a good image. Please note that our analysis is based on the entire image, not on arbitrary local patches as in the JPEG transform.

6. SUMMARY

Our studies of motion sharpening have demonstrated that the actual change in appearance is most similar to a decrease in the slope of the amplitude function in the spatial frequency domain. We find that motion sharpening is greater when both high and low spatial frequencies are present, indicating an interaction between them. The high frequency information, whether spatial or temporal, does not need to be refreshed as often as the low frequency information in order to produce a moving display of high quality.

In still images, we have described some perceptual oddities that occur when intermediate spatial frequencies are strongly attenuated. Parts of a complex scene appear to become self-luminous, while parts of some objects may appear transparent. Self-luminosity or transparency occurs only when both high and low spatial frequencies remain in the image. Object recognition can also be compromised when the intermediate spatial frequencies are removed. These observations suggest that the intermediate spatial frequencies play an important role in the most basic perceptual tasks, such as determining the source of illumination and pattern recognition.

We have demonstrated that a grey-scale image that appears to be a good representation of a scene can lead to perceptual errors. Color, by reducing luminance masking and adding an additional source of contrast, can overcome some of these errors.

Finally, we have shown that a greatly simplified, smoothed amplitude function can be substituted for the actual amplitude function in a natural scene. Though the image that results when the simplified function is combined with the phase spectrum is not perfect, it is highly similar to the original.

Although we understand much about the visual system's ability to detect and discriminate between patterns, our ability to predict their appearance is not highly developed. The study of the factors governing the appearance of images holds considerable promise as a tool to increase our understanding of the complex tasks of vision.

REFERENCES

---

[1]Ramachandran, V. S., Rao, V. M. and Vidyasagar, T. R., “Sharpness constancy during movement perception)Short note),” Perception, 3, pp. 97-98 (1974).

[2]Burr, D. .C. and Morgan, M. J., “Motion deblurring in human vision,” Proc. R. Soc. Lond. B, 264, pp. 431-446 (1997).

[3]S. T., Georgeson, M. A. and Gorea, A., “Motion blur and motion sharpening: temporal smear and local contrast non-linearity,” Vision Res., 38, pp. 2099-2108 (1998).

[4]Pääkkönen. K. and Morgan, M. J., “Linear mechanisms can produce motion sharpening,” Vision Res., 41, pp. 2771-2777 (2001).

[5]Georgeson, M. A. and Hammett, S. T., “Seeing blur: ‘motion sharpening’ without motion,” Proc. R. Soc. Lond. B, 269, pp. 1429-1434 (2002).

[6]Takeuchi, T. and De Valois, K. K., “Sharpening image motion based on the spatio-temporal characteristics of human vision,” Human Vision and Electronic Imaging X, Proc. SPIE-IS&T Electronic Imaging, SPIE Vol. 5666, 83-94 (2005).

[7]Adelson E. H. and Bergen, J. R., “Spatiotemporal energy models for the perception of motion,” J. Opt. Soc. Am. A, 2, pp. 284-299 (1985).

[8]Simoncelli, E. P. and Heeger, D. J., “A model of neuronal responses in visual area MT,” Vision Res., 38, pp. 743-761 (1998).

[9]Takeuchi, T. and De Valois, K. K., “Motion sharpening in moving natural images.” J. Vision, 2, Abstract 377, p. 377a, (2002).

[10]Watson, A. B. and Nachmias, J., “Patterns of temporal interaction in the detection of gratings,” Vis. Res. 17, pp. 893-902 (1977).

[11]De Valois, K. K., [The Visual Neurosciences], MIT Press, Cambridge, MA, pp. 924-935 (2003).

[12]Field, D. J., “Relations between the statistics of natural images and the response properties of cortical cells,” J. Opt. Soc. Am. A, 4, pp. 2379-2394 (1987).

[13]Oppenheim, A. V. and Lim, J. S., “The importance of phase in signals,” Proc IEEE, 69, pp. 529-541 (1981).

[14]Piotrowski, L. N. and Campbell, F. W., “A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase,” Perception, 11,337-346 (1982).

[15]Torralba, A. and Oliva, A., “Statistics of natural image categories,” Network: Computation in Neural Systems, 14, pp.391-412 (2003).

[16]Thomson, M. G. A., “Visual coding and the phase structure of natural scenes,” Network: Computation in Neural Systems, 10, pp.123-132 (1999a).

[17]Thomson, M. G. A., “Higher-order structure in natural scenes,” J. Opt. Soc. Am. A, 16, pp. 1549-1553 (1999b).