Image Quality Metrics Based on Single and Multi-Channel Models of Visual Processing

Joyce E. Farrell † Xuemei Zhang ‡ Christian J. van den Branden Lambrecht †
D. Amnon Silverstein †
†Hewlett-Packard Laboratories, 1501 Page Mill Road, Palo Alto, CA, 94304
‡Psychology Department, Stanford University, Stanford, CA 94304

Abstract

We review two classes of image analysis tools based on single and multiple channel models of human vision processing, designed to predict the visibility of printed dots and halftone texture, respectively.

1 Introduction

Nearly every imaging peripheral Hewlett-Packard manufactures is designed to optimize subjective image quality. Whether we are capturing, transmitting or rendering visual images, we consider the impact that our design decisions have upon subjective image quality.

Subjective impressions of image quality, however, are difficult, if not impossible, to measure, quantify and predict. Subjective image quality is influenced by personal experience and other high-level factors outside of any simple physical description of the image. Thus we are faced with the challenging task of optimizing a signal that we find difficult to measure.

Some physical attributes of an image can be strongly correlated with perceived quality. Our approach to image quality evaluation is to identify the visual factors that influence these perceptions and then design machine vision tools to measure these factors. For example, we are developing image analysis tools that predict the visibility of halftone texture, image compression artifacts, and printer banding. These visual factors adversely influence our customers' subjective impressions of image quality.

Our goal is to design imaging peripherals and image processing methods that minimize the visibility of these types of distortions. To achieve this goal, we must develop metrics for the detectability of printed dots, halftone texture, printer banding, compression artifacts, and other annoying visual signals.

In this paper we consider the power and the limitations of two classes of image quality metrics that we refer to as single-channel and multi-channel metrics, respectively.

2 Single channel metrics

A large collection of image quality (IQ) metrics are based on single channel models of human visual processing. Typically these metrics quantify the energy in a signal after it has been passed through a visual filter that represents human spatial sensitivity as a function of spatial frequency. This filter is referred to as the human contrast sensitivity function (CSF). The filtered signal is pooled together in a single "channel" of information.

Single channel (SC) metrics have a long evolutionary history, beginning with the application of linear systems theory to the study of optical systems. The image quality of a lens, for example, is determined by its' optical blur which, in turn, can be characterized by an optical blur function (in space) or a modulation transfer function (in the Fourier frequency domain). Unlike the low-pass MTF, the CSF is band-pass. The CSF includes neural as well as optical factors.

SC metrics are typically based on the the summed energy of the perceptual signal. For example, the mean square error metric (MSE), popularly used as a fidelity metric in digital image processing, is based on the sum of the squared differences between corresponding pixels in an original and a distorted image or, in other words, the squared vector length of the difference image.

An improvement of the MSE is the visually-weighted mean square error (VMSE), which is computed by filtering the difference image with a human contrast sensitivity function or CSF (e.g. see [7]). The CSF describes how sensitive a typical, or standard, observer is...
to the spatial frequency components in any spatial image. Filtering an image with the human contrast sensitivity function effectively weights the image by how sensitive the visual system is to the spatial frequencies present in the image. To quantify the difference between two visually filtered or weighted images we compute the vector length of the filtered difference between the two images.

Conceptually similar metrics can be computed in the frequency domain. For example, the image quality metric commonly referred to as "granularity" is computationally equivalent to the root mean square error in the spatial reflectance domain. Granularity [2] measures the power spectrum of image distortion, integrated over the range of a filter function. This filter function is similar to the CSF but it has been adjusted to predict the detectability of film grain noise. This metric is a SC metric because it is based on the integrated output of the distortion signal passed through a single filter.

SC metrics have been extremely useful in characterizing image quality in optics, image processing and analog and electro-photography. They can be used to evaluate the visibility of film grain noise, the visibility of toner particles, and the visibility of printer halftone noise when the noise is fixed and varies only in amplitude. These measures work well when the image distortion signal is stochastic in nature, as is typical in analog imaging.

When the filter function is tuned to predict the detectability of a certain signal structure, such as sinusoidal gratings (in the case of the CSF) or film-grain noise (in the case of granularity), it is very successful. However, when the filter function predicts one form of signal, it is likely to fail on other structures. For example, granularity can not predict the visibility of a thin line.

Although no single filter metric can predict the detectability of all types of distortion, a group of filter metrics can be combined in a systematic way to provide much more robust predictions. Instead of pooling the signal into a single channel, multi-channel metrics are capable of predicting the visibility of patterns based on the combined output of many filters (independent visual channels) sensitive to different spatial frequencies and orientations.

3 Multichannel Metrics

Single channel models have a limited field of application. It is known that they fail to predict the visibility of complex structured signals and they cannot account for observable effects of pattern adaptation [5]. A large body of research [4] supports the hypothesis that several independent mechanisms or visual channels determine our sensitivity to spatial patterns. Each channel is sensitive to a particular band of spatial frequencies and orientations. A collection of independent channels spans the frequency plane, partitioning the plane into frequency- and orientation selective bands. The tuning of the bands is also consistent with measures of the responses of the neurons in the primary visual cortex that proved to be band-limited.

This hypothesis gave birth to a variety of vision models [3,9,10]. An illustration of this concept is presented in Fig. 1, where the magnitude of the channel is plotted over the spatial frequency plane. The channels realize a dyadic (octave band) decomposition along the radial frequency axis and a linear decomposition in orientation.

Multichannel (MC) metrics use at least two different types of summation to predict detection. The image is filtered independently by a collection of filters. Each filter is pooled to provide a separate channel. The channels are then weighted and pooled together with a different function (such as a probability summation function). MC metrics are typically more computationally expensive than SC metrics, but they can predict the visibility of more complex, structured and periodic patterns, as well as account for some effects of visual adaptation.

MC models are also able to incorporate models for visual masking, i.e. of the interference between several...
stimuli in a single visual sensation. More precisely, it is known that the presence of a background stimulus modifies the perception of a foreground stimulus: masking corresponds to a modification of the detection threshold of the foreground according to the local contrast of the background.

To compare how well SC and MC metrics predict the visibility of printed halftone texture, we conducted a visual psychophysical experiment, described briefly below.

4 Comparison of single and multichannel metrics

4.1 Stimuli

We printed a set of halftoned gray and sinusoid patches on a 600 dpi laser printer. The halftone methods used were clustered-dot dither and void-and-cluster dither. Halftone cell sizes of 10 and 12 were used for the clustered-dot dither, and cell sizes of 16, 64, and 256 were used for the void-and-cluster dither. The halftoning was done in MATLAB.

4.2 Method

We collected rank ordering data from 18 subjects on the printed halftone patches. Subjects were asked to sort the 7 halftone patches by how visible the pattern was in comparison to a uniform gray field. This method has several disadvantages. First, although we asked subjects to keep the patterns on a viewing stand as they sorted them and to keep their viewing distance at approximately 12 inches, there was no precise control of viewing distance. To compute both SC and MC metrics, we must assume a fixed viewing distance. Second, subjects are making judgments about halftone texture patterns that are clearly visible (in other words, above their thresholds for visibility). SC and MC metrics are designed to predict contrast thresholds for fixed viewing distances. To predict contrast thresholds for fixed viewing distances. Silverstein and Farrell [8] have shown that for super-threshold patterns, perceptual strength is not a simple function of contrast. Thus, we should not expect to find a strong relationship between ranking and SC and MC metric predictions.

SC and MC metrics are designed to predict the perceptual strength of the patterns near the detection threshold, and the strength of the pattern in terms of contrast threshold is the only thing the described SC and MC can predict. To judge the models in terms of near-threshold performance, we reduced the strength of the patterns to the point of subjective threshold.

We effectively reduced the contrast of the patterns by varying the viewing distance. Contrast is reduced with distance when certain conditions hold. If the pattern has only high frequency signal content and the CSF has a linear fall off in the region where the pattern has detectable signal, then perceptual contrast is directly reduced with distance. If these conditions are approximately true, then contrast will be reduced monotonically with distance. Our patterns mostly fit these criteria (with perhaps one exceptional pattern, as will be discussed), so we expected distance to correlate better with contrast sensitivity than the ranking experiment (see [8,11] for an improved method of contrast adjustment).

The data were collected in a lab room with indoor lighting (fluorescent). Subjects completed the rank ordering task first, and then completed the threshold viewing distance task. The printed halftone patches were 0.8 × 0.8 inches in size, and mounted on 3.5 × 4 inch white cards. The visual angle of the print area of each patch is about 4 degrees at a viewing distance of 12 inches.

We found that threshold viewing distance is monotonically related to rank order within the precision of the distance measurements. This tells us that subjects’ ordering of halftone visibility does to some extent indicate how easy it is to see the texture patterns.

4.3 Analysis

We implemented both SC metrics as Matlab routines that take, as input, calibrated luminance bitmaps and generate, as output, predicted halftone visibility strength. To create the calibrated bitmaps, we scanned the printed patches at 1200 dpi on an AGFA scanner using raw reflectance mode. For each scanned image, we took a 512x512 square to use in all computations. The 512x512 square corresponds to a 256x256 square on the printed page (because of the difference in scanning and printing resolution), which at a viewing distance of D should have a visual angle of \( \text{atan}(\frac{256}{600D}) \times \frac{180}{\pi} \) on each side. Thus the visual angle of the input pattern on paper is 2.04 degrees at 12 inch viewing distance, and 4.07 degrees at 6 inch viewing distance.

Our Matlab implementations of single and multichannel metrics are based on the assumption that the perceptual strength of halftone texture varies directly with the contrast of the pattern. For high contrast patterns, however, factors other than contrast may determine the perceived pattern strength [8].

Since the task involved the detection of a pattern on a uniform gray field, we did not need to model visual
stimuli in a single visual sensation. More precisely, it is known that the presence of a background stimulus modifies the perception of a foreground stimulus: masking corresponds to a modification of the detection threshold of the foreground according to the local contrast of the background.

To compare how well SC and MC metrics predict the visibility of printed halftone texture, we conducted a visual psychophysical experiment, described briefly below.

4 Comparison of single and multi-channel metrics

4.1 Stimuli

We printed a set of halftoned gray and sinusoid patches on a 600 dpi laser printer. The halftone methods used were clustered-dot dither and void-and-cluster dither. Halftone cell sizes of 10 and 12 were used for the clustered-dot dither, and cell sizes of 16, 64, and 256 were used for the void-and-cluster dither. The halftoning was done in MATLAB.

4.2 Method

We collected rank ordering data from 18 subjects on the printed halftone patches. Subjects were asked to sort the 7 halftone patches by how visible the pattern was in comparison to a uniform gray field. This method has several disadvantages. First, although we asked subjects to keep the patterns on a viewing stand as they sorted them and to keep their viewing distance at approximately 12 inches, there was no precise control of viewing distance. To compute both SC and MC metrics, we must assume a fixed viewing distance. Second, subjects are making judgments about halftone texture patterns that are clearly visible (in other words, above their thresholds for visibility). SC and MC metrics are designed to predict contrast thresholds for fixed viewing distances. Silverstein and Farrell [8] have shown that for super-threshold patterns, perceptual strength is not a simple function of contrast. Thus, we should not expect to find a strong relationship between ranking and SC and MC metric predictions.

We effectively reduced the contrast of the patterns by varying the viewing distance. Contrast is reduced with distance when certain conditions hold. If the pattern has only high frequency signal content and the CSF has a linear fall off in the region where the pattern has detectable signal, then perceptual contrast is directly reduced with distance. If these conditions are approximately true, then contrast will be reduced monotonically with distance. Our patterns mostly fit these criteria (with perhaps one exceptional pattern, as will be discussed), so we expected distance to correlate better with contrast sensitivity than the ranking experiment (see [8,11] for an improved method of contrast adjustment).

The data were collected in a lab room with indoor lighting (fluorescent). Subjects completed the rank ordering task first, and then completed the threshold viewing distance task. The printed halftone patches were 0.8 x 0.8 inches in size, and mounted on 3.5 x 4 inch white cards. The visual angle of the print area of each patch is about 4 degrees at a viewing distance of 12 inches.

We found that threshold viewing distance is monotonically related to rank order within the precision of the distance measurements. This tells us that subjects' ordering of halftone visibility does to some extent indicate how easy it is to see the texture patterns.

4.3 Analysis

We implemented both SC metrics as Matlab routines that take, as input, calibrated luminance bitmaps and generate, as output, predicted halftone visibility strength. To create the calibrated bitmaps, we scanned the printed patches at 1200 dpi on an AGFA scanner using raw reflectance mode. For each scanned image, we took a 512x512 square to use in all computations. The 512x512 square corresponds to a 256x256 square on the printed page (because of the difference in scanning and printing resolution), which at a viewing distance of D should have a visual angle of \(\arctan(\frac{256}{D}) \times \frac{180}{\pi}\) on each side. Thus the visual angle of the input pattern on paper is 2.04 degrees at 12 inch viewing distance, and 4.07 degrees at 6 inch viewing distance.

Our Matlab implementations of single and multi-channel metrics are based on the assumption that the perceptual strength of halftone texture varies directly with the contrast of the pattern. For high contrast patterns, however, factors other than contrast may determine the perceived pattern strength [8].

Since the task involved the detection of a pattern on a uniform gray field, we did not need to model visual
masking [6]. Masking occurs when a high contrast pattern is on or near the stimulus. The masking pattern may reduce sensitivity to the pattern. In developing quality metrics for printers, it is usually more useful to know the detectability of distortion on a uniform field, since this is the worst case of distortion. For other imaging applications, such as image compression, visual masking will need to be considered. Masking may also need to be considered for printer band detection. Banding is a periodic thin line that is sometimes introduced on top of a significantly strong background of halftone noise. In some cases the halftone noise can mask the banding.

Our implementation of the SC metric is based on the contrast sensitivity function by Comes [1]. We also use this function to adjust the gain of the visual channels in the MC metric. The single-channel visibility is just the visually weighted power of the input image integrated over all frequencies. We computed the single-channel visibility for the same 7 halftone patches.

We implemented a MC metric with a bank of 17 filters tuned to 5 spatial frequencies (0, 2, 4, 8 and 16 cpd) and four orientations [1]. The filterbank is analytical, which permits a very easy estimation of contrast values. Again, because we wish to predict the visibility of halftone texture without reference to a particular image, masking is not taken into account. Contrast is thresholded within the channels according to the detection threshold of each channel. The visibility metric is obtained by Minkowski summation of the outputs of all channels after thresholding.

The algorithm can be summarized as follows:

**Algorithm 1 Visibility Prediction**

\[
\begin{align*}
 & I \leftarrow \text{considered image} \\
 & \text{for } j = 0 \text{ to } 17 \\
 & \quad CP_j \leftarrow I \ast \text{Filter}_j \\
 & \quad \text{for each pixel } x,y \\
 & \quad \quad \text{if } \text{Contrast}(x,y) < C_{T0} \text{ then} \\
 & \quad \quad \quad CP_j(x,y) \leftarrow 0 \\
 & \quad \quad \text{else} \\
 & \quad \quad \quad CP_j(x,y) \leftarrow CP_j(x,y) \times \left(1 - \frac{C_{T0}}{\text{Contrast}(x,y)}\right) \\
 & \quad \quad \text{end if} \\
 & \quad \text{end for} \\
 & \text{Visibility} = \left(\frac{1}{N_c \times N_p} \sum_{j=1}^{N_c} \sum_{i=1}^{N_p} |\text{Re}(CP_j(x,y))|^p\right)^{1/p} \\
& \text{end for} \\
\end{align*}
\]

where \(C_{T0}\) is the detection threshold of the considered channel, \(N_c\) the number of channels (17 in this case), \(N_p\) the number of pixels in a channel. The exponent \(p\) used for the Minkowski summation is a free parameter and has been set to 2 (i.e. summation of channel outputs are Euclidean).

5 Results

To evaluate the ability of the SC and MC metrics to predict the visibility of halftone texture, we plot the predicted output as a function of both rank orders and threshold viewing distances in Figure 2. The MC values are monotonically related to both the rank order score and to the threshold viewing distance of the halftone patterns. The SC values show one violation of monotonicity. Thus, although we predict some of the variance in the subjects' judgments of halftone texture visibility with the SC metric, we do a better job with a multichannel metric.

In Fig. 2, we identify one outlier data point corresponding to a pattern that was ranked third in texture visibility, had the shortest threshold viewing distance, and yet has relatively small SC and MC predicted values. Careful inspection revealed that this pattern, unlike all the others, had a low frequency component that was described by some subjects as a visible "dark ring" in the pattern. Several subjects reported seeing the "ring" from a rather large viewing distance, but did not report seeing the halftone pattern until they moved much closer to the patch. This "emergent" figure in the pattern must have influenced their criteria for halftone visibility for this particular pattern. We suspect that this was due to a failure of our method to reduce contrast as was described previously.

6 Conclusions

When used properly, SC and MC metrics can be powerful tools in the evaluation of image quality. Whereas SC metrics can be used to predict the visibility of stochastic noise (such as film grain), MC metrics are necessary to predict the visibility of periodic noise, such as printed halftone texture noise.

We are developing a collection of different machine vision metrics to predict the visibility of different types of visual signals. As the complexity of the signal increases, so too does the complexity of the metric designed to measure its visibility. An image quality metric capable of predicting halftone texture visibility requires multiple visual channels sensitive to different frequencies and orientations, but does not require the additional expense of masking operations. A metric capable of predicting the visibility of printer banding may require additional masking operations. And, as you will hear in a subsequent paper in this session [11], mult-
Multiple luminance and chrominance channels are required to predict the visibility of color halftone patterns.

References


