A Perceptual Distortion Metric for Digital Color Video

Stefan Winkler

Signal Processing Laboratory
Swiss Federal Institute of Technology
1015 Lausanne, Switzerland
http://ltswww.epfl.ch/~winkler/
Stefan.Winkler@epfl.ch

ABSTRACT

In this paper I present a distortion metric for color video sequences. It is based on a contrast gain control model of the human visual system that incorporates spatial and temporal aspects of vision as well as color perception. The model achieves a close fit to contrast sensitivity and contrast masking data from several different psychophysical experiments for both luminance and color stimuli. The metric is used to assess the quality of MPEG-coded sequences.

Keywords: Spatio-temporal vision model, color perception, contrast gain control, video quality assessment

1. INTRODUCTION

With the advent of digital coding of visual information and particularly compression, the evaluation of the actual perceptibility of distortions in images and video has become difficult. Most of the time, the designers of compression algorithms have had to rely on subjective viewing tests in order to obtain reliable ratings for the quality of compressed images or video. While these tests — if executed properly — undoubtedly give valid indications about the visibility of artifacts, they are complex, time-consuming and expensive, and therefore often highly impractical or not feasible at all. Consequently, researchers have computed simple error measures such as mean squared error (MSE) or peak signal-to-noise ratio (PSNR) as alternatives, suggesting that they would be equally valid. However, these simple error measures operate solely on a pixel-by-pixel basis and neglect the much more complex behavior of the human visual system and its dependence on viewing conditions.

These problems prompted the development of objective distortion metrics in recent years. Some exploit *a priori* knowledge about specific artifacts that are introduced by certain compression methods (e.g. block artifacts in JPEG and MPEG), while others are more generally based on models of the human visual system. The metric presented here belongs to the latter category. As such it is independent of the particular compression algorithm applied to the video material.

However, the human visual system is extremely complex, and many of its inner mechanisms are barely understood even today. Therefore, vision models are necessarily limited in scope and incorporate only small parts of the entire system. The perceptual distortion metric presented in this paper is based on a metric developed previously at our lab.[17,27] The underlying vision model — an extension of a model for still images presented by the author[30] — focuses on the following aspects:

- color perception, in particular the theory of opponent colors;
- temporal and spatial mechanisms;
- spatio-temporal contrast sensitivity and pattern masking;
- the response properties of neurons in the primary visual cortex.

Each of these aspects and their implementation in the context of a distortion metric will be explained in detail in the next section.

2. VISION MODEL

The structure of the vision model is shown in Figure 1. After conversion of the input to opponent-colors space, each of the resulting three components is subjected to a spatio-temporal perceptual decomposition, yielding a number of perceptual channels. They are weighted according to contrast sensitivity data and subsequently undergo a contrast gain control stage. Both the reference sequence and the processed sequence are used as input to the model and go through identical stages. Finally, all the sensor differences are combined into a distortion measure. Each of these stages is explained in more detail below.

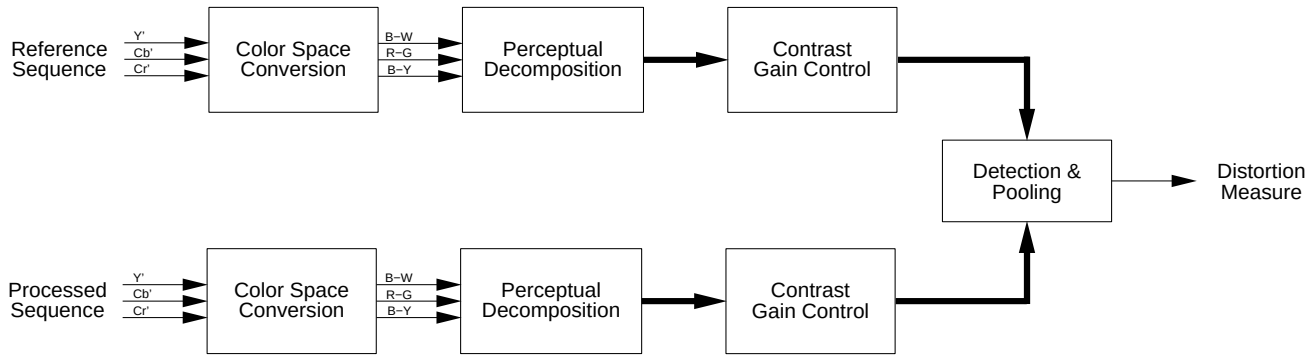

Figure 1. Block diagram of the perceptual distortion metric. After conversion to opponent-colors space, each of the resulting three components is subjected to a perceptual decomposition, yielding several perceptual channels. Subsequently they undergo weighting and contrast gain control, after which all the sensor differences are combined into a distortion measure.

2.1. Color Space Conversion

Hering[10] was the first to point out that some pairs of hues can coexist in a single color sensation (e.g. a reddish yellow is perceived as orange), while others cannot (we never perceive a reddish green, for instance). This led him to the conclusion that the sensations of red and green as well as blue and yellow are encoded in separate visual pathways, which is commonly referred to as the theory of *opponent colors*. Both psychological and physiological experiments in the 1950s[12,14] yielded further evidence to support this theory, and many models have been presented since then. The principal components of opponent-colors space are black-white (B-W), red-green (R-G), and blue-yellow (B-Y). Their precise directions are still subject to debate, however.

The color spaces used in many standards for coding visual information, e.g. PAL, NTSC, JPEG, MPEG and others, already take into account certain properties of the human visual system by coding gamma-corrected color difference components instead of RGB color primaries. A variety of color difference spaces exist; a well-known one for digital video is $\mathrm{Y'C'_BC'_R}$ as defined in ITU-R Recommendation 601.[13] Y' encodes luminance, $\mathrm{C'_B}$ the difference between the blue primary and luminance, and $\mathrm{C'_R}$ the difference between the red primary and luminance. Conversion from $\mathrm{Y'C'_BC'_R}$ to opponent-colors space requires a number of transformations as illustrated in Figure 2.

First, $\mathrm{Y'C'_BC'_R}$ must be converted to R'G'B'. According to ITU-R Recommendation 601,[13] R'G'B' values in the range $0 \ldots 1$ are computed from 8-bit $\mathrm{Y'C'_BC'_R}$ as follows:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \frac{1}{219} \begin{bmatrix} 1 & 0 & 1.3707 \\ 1 & -0.3365 & -0.6982 \\ 1 & 1.7324 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} Y' \\ C'_B \\ C'_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right). \quad (1)$$

Each of the resulting three components undergoes a nonlinearity of the form $f(x) = x^\gamma$ with $\gamma \approx 2.5$ to produce linear RGB values. This is an approximation of the behavior of a conventional CRT display. RGB space further assumes a particular display device, or to be more exact, a particular spectral power distribution of the light emitted from the display phosphors. Once the phosphor spectra of the monitor of interest have been determined, the responses of the L-, M-, and S-cones on the human retina, whose sensitivities are known,[24] can be computed. Their responses are decorrelated very early in the visual system by forming new signals, and a variety of combinations of L-, M-, and

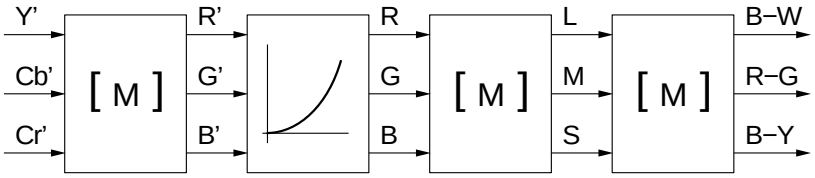

Figure 2. Color space conversion from component video $Y'C'_BC'_R$ to opponent-colors space.

S-cone responses have been proposed. Some authors also suggest applying a nonlinearity to the LMS values before further processing. This distortion metric relies on a recent opponent-colors model by Poirson and Wandell.[20,21] This particular opponent-colors space has the advantage of being pattern-color separable, i.e. color perception and pattern sensitivity can be decoupled and treated in separate stages. The B-W, R-G, and B-Y components are computed from LMS values via the following linear transform:

$$\begin{bmatrix} BW \\ RG \\ BY \end{bmatrix} = \begin{bmatrix} 0.990 & -0.106 & -0.094 \\ -0.669 & 0.742 & -0.027 \\ -0.212 & -0.354 & 0.911 \end{bmatrix} \cdot \begin{bmatrix} L \\ M \\ S \end{bmatrix}. \tag{2}$$

2.2. Perceptual Decomposition

The perceptual decomposition is performed first in the temporal and then in the spatial domain. While this separation is not entirely unproblematic, these two domains can be consolidated in the fitting process as described below.

2.2.1. Temporal Mechanisms

The characteristics of the temporal mechanisms in the human visual system are still under discussion in the vision community. It is believed that there are one temporal low-pass and one, possibly two, temporal band-pass mechanisms,[6,11,19,28] which are generally referred to as sustained and transient channels, respectively. Recent studies also indicate that the peak frequency and bandwidth of the mechanisms change considerably with stimulus energy.[5] The existence of an actual third mechanism is questionable, however.[6,7]

The temporal filters used in this distortion metric are based on the recent work by Fredericksen and Hess,[5,6] who model the temporal mechanisms with derivatives of the following impulse response function:

$$h(t) = e^{-\left(\frac{\ln(t/\tau)}{\sigma}\right)^2}. \tag{3}$$

They achieve a very good fit to their data using only this function and its second derivative, corresponding to one sustained and one transient mechanism, respectively. For a typical choice of parameters $\tau = 0.16$ and $\sigma = 0.2$, the impulse responses of the two mechanisms are shown in Figure 3, and the corresponding frequency responses are shown in Figure 4.

For use in the distortion metric, the filters have been designed so as to keep the delay to a minimum, because in some applications of distortion metrics such as monitoring and control, a short response time is crucial. This fact together with limitations of memory and computing power favor time-domain implementations of the temporal filters over frequency-domain implementations.[17] A trade-off has to be found between an acceptable delay and the accuracy with which the temporal mechanisms ought to be approximated. Recursive infinite impulse response (IIR) filters fare better in this respect than (nonrecursive) finite impulse response (FIR) filters.

In this metric, the temporal mechanisms are modeled by two IIR filters. They were computed by means of a least-square fit to the frequency magnitude responses of the respective mechanisms as given by the Fourier transforms of $h(t)$ and $h''(t)$ from Equation 3. A filter with 2 poles and 2 zeros was fitted to the sustained mechanism, and a filter with 4 poles and 4 zeros was fitted to the transient mechanism. This has been found to yield the shortest delay while still maintaining a good approximation of the frequency responses, which is shown in Figure 4. In the present implementation, the low-pass filters are applied to all three color channels, but the band-pass filter is applied only to the luminance channel in order to reduce computing time. This simplification is based on the fact that color contrast sensitivity is rather low for higher frequencies.

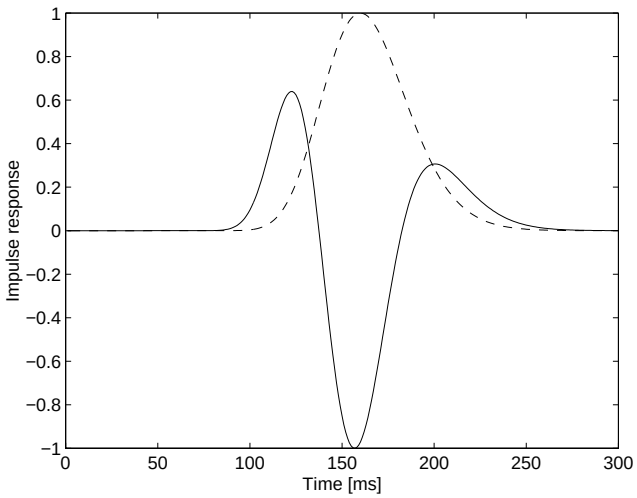

Figure 3. Impulse response function $h(t)$ from equation 3 (dashed) and its second derivative $h''(t)$ (solid) with $\tau = 0.16$ and $\sigma = 0.2$.

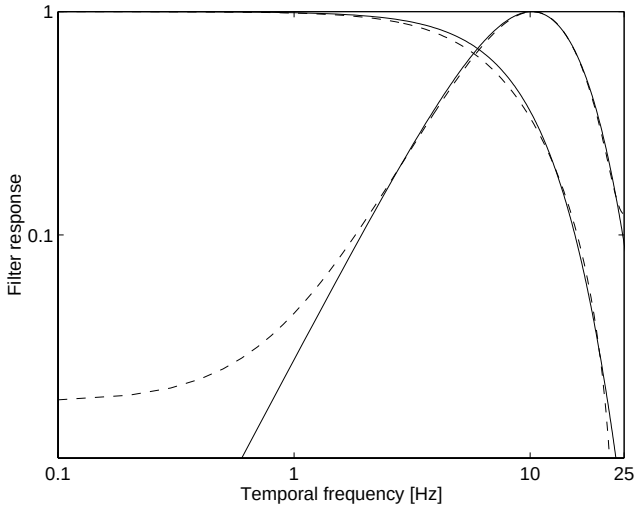

Figure 4. Frequency responses of sustained (low-pass) and transient (band-pass) mechanisms of vision[5,6] (solid) and their IIR filter approximations for a sampling frequency of 50 Hz (dashed).

2.2.2. Spatial Mechanisms

The decomposition in the spatial domain is carried out by means of the steerable pyramid transform proposed by Simoncelli et al.[23] This transform* decomposes an image into a number of spatial frequency and orientation bands; its basis functions are directional derivative operators. For use within a vision model, it has the advantage of being rotation-invariant and self-inverting, and it minimizes the amount of aliasing in the subbands. In the present implementation, the basis filters have octave bandwidth and octave spacing; five subband levels with four orientation bands each plus one low-pass band are computed (cf. Figure 5). The same decomposition is used for all channels.

2.2.3. Contrast Sensitivity

After the temporal and spatial decomposition, each channel is weighted such that the sum of all channels approximates the spatio-temporal contrast sensitivity of the human visual system. While this approach is inferior to pre-filtering the B-W, R-G and B-Y channels with their respective contrast sensitivity functions in terms of accuracy, it is easier to implement and saves computing time. As a matter of fact, the approximation accuracy is still acceptable, as the results in section 3 show.

2.3. Contrast Gain Control Stage

Modeling pattern sensitivity is one of the most critical components of video quality assessment, because the visibility of distortions is highly dependent on the local background. To investigate this phenomenon, also known as masking, vision scientists conduct experiments where a simple stimulus such as a sinusoid or Gabor patch (also referred to as the target) is superimposed on a background stimulus (the masker). With some exceptions, the target is harder to detect as the contrast of the masker increases. In the context of digital video compression, the coding noise plays the role of the target, and the original becomes the masker.

Masking is strongest between stimuli located in the same perceptual channel, and many vision models are limited to this intra-channel masking. However, psychophysical experiments show that masking also occurs between channels of different orientations,[4] between channels of different spatial frequency, and between chrominance and luminance channels,[2,18,25] albeit to a lesser extent.

Models have been proposed which explain a wide variety of empirical masking data with a process of contrast gain control. These models were inspired by analyses of the responses of single neurons in the visual cortex of the

* Source code and filter kernels for the steerable pyramid transform are available at http://www.cis.upenn.edu/~eero/steerpyr.html.

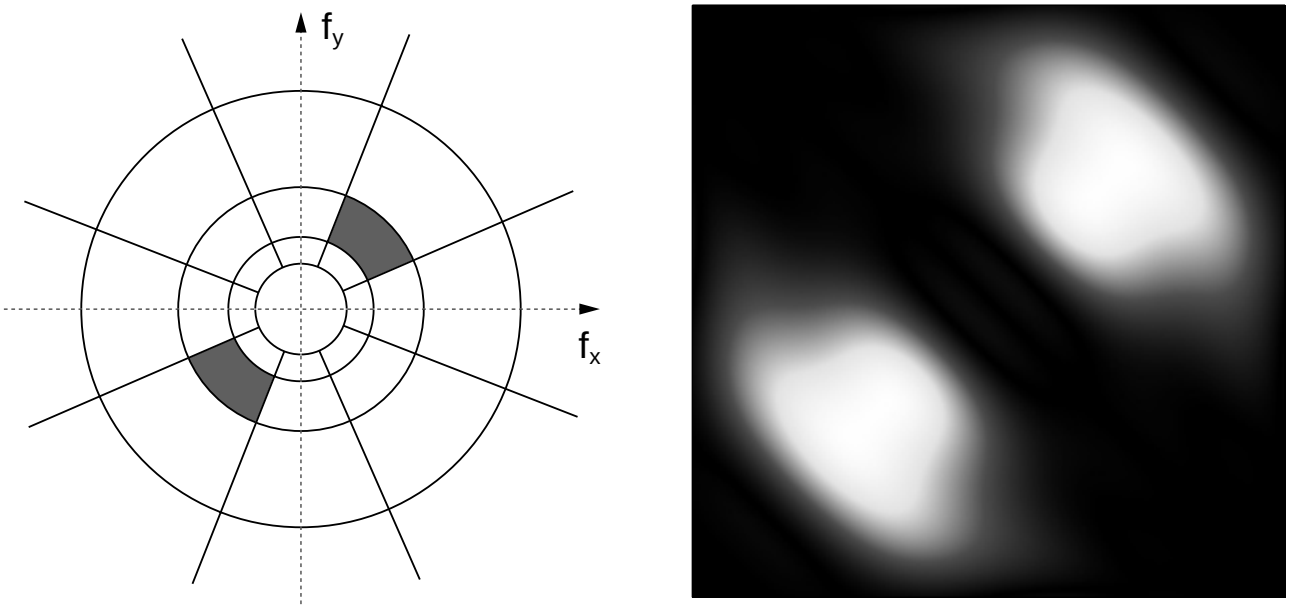

Figure 5. Illustration of the partitioning of the spatial frequency plane by the steerable pyramid transform.[23] Three levels plus one (isotropic) low-pass filter are shown. The bands at each level are tuned to orientations of 0, 45, 90 and 135 degrees. The shaded region indicates the spectral support of a single subband, whose actual frequency response is shown on the right.

cat,[1,8,9] where contrast gain control serves as a mechanism to keep neural responses within the permissible dynamic range while at the same time retaining global pattern information.

Contrast gain control can be realized by an excitatory nonlinearity that is inhibited divisively by a pool of responses from other neurons. Masking occurs through the inhibitory effect of the normalizing pool.[4,26] Watson and Solomon[29] recently presented an elegant generalization of these models, which facilitates the integration of many kinds of channel interactions and spatial pooling. Introduced for luminance images, this contrast gain control model can be extended to color and to sequences as follows: Let $a = a(t, c, f, \theta, x, y)$ be a coefficient of the perceptual decomposition in temporal channel t, color channel c, frequency band f, orientation band θ, at location x, y. Then the corresponding sensor output $s = s(t, c, f, \theta, x, y)$ is computed as

$$s = k \frac{a^p}{b^2 + a^q * h}. \tag{4}$$

The excitatory path in the numerator consists of a power-law nonlinearity with exponent p. The inhibitory path in the denominator controls the gain of the excitatory path. In addition to a nonlinearity with a possibly different exponent q, filter responses are pooled over different channels by means of a convolution with the pooling function $h = h(t, c, f, \theta, x, y)$. In its most general form, this pooling operation may combine coefficients from the dimensions of time, color, temporal frequency, spatial frequency, orientation, space, and phase. In the present implementation of the distortion metric, it is limited to phase and orientation; complete summation over phase is assumed, and a Gaussian pooling kernel is used for the orientation dimension.

The saturation constant b is added to prevent division by zero; k is used to adjust the overall gain of the mechanism. Heeger's model of neural cell responses[9] fixes both $p = q = 2$ so as to be able to work with local energy measures. However, this procedure rapidly saturates the sensor outputs, which necessitates multiple contrast bands (i.e. several different k's and b's) for all coefficients in order to cover the full range of contrasts. Watson and Solomon[29] showed that the same effect can be achieved with a single contrast band when $p > q$. This approach reduces the number of model parameters considerably and simplifies the fitting process, which is why it is used here as well. The fitting procedure and its results are explained in more detail in section 3.1.

2.4. Detection and Pooling

The processes described so far take place before or in the primary visual cortex, also referred to as area V1. It is believed that the information represented there in various channels is integrated in the following brain areas, beginning with area V2. This process can be simulated by gathering the data from these channels according to rules of probability or vector summation, also known as pooling.[22]

The pooling stage combines the elementary differences between the sensor outputs $\mathbf{s} = \mathbf{s}(t, c, f, \theta, x, y)$ for the reference ($\mathbf{s}_0$) and the processed sequence ($\mathbf{s}_1$) over several dimensions:

$$\Delta s = \sqrt[\beta]{\sum |\mathbf{s}_0 - \mathbf{s}_1|^\beta}. \tag{5}$$

In principle, any subset of dimensions can be used for this summation, depending on what kind of result is desired. For example, pooling over pixel locations may be omitted to produce a distortion map for every frame of the sequence. The combination may be nested as well; pooling can be limited to single frames first to determine the variation of distortions over time, and the total distortion can be computed from the values for each frame.

3. RESULTS

3.1. Model Fitting

The model contains several parameters that have to be adjusted in order to accurately represent the human visual system. For this procedure I use threshold data from contrast sensitivity and contrast masking experiments. In the fitting process, the input of the metric imitates the stimuli used in these experiments, and the free model parameters are adjusted in such a way that the output (the distortion measure) approximates these threshold curves.

Contrast sensitivity is modeled by setting the gains of the spatial and temporal filters in such a way that the model predictions match empirical threshold data from spatio-temporal contrast sensitivity experiments for both color and luminance stimuli. For the B-W channels, the weights are chosen so as to match contrast sensitivity threshold data from Kelly.[15] Figure 6 shows the result. For the R-G and B-Y channels, similar data from Kelly[16] is used.

The parameters k, p and b of the contrast gain control stage from equation 4 are determined by fitting the model's responses to masked gratings; q, the inhibitory exponent, is fixed at $q = 2$ in this implementation, as it is mainly the difference $p - q$ which matters.[29] For the B-W channel, empirical data from several intra- and inter-channel contrast masking experiments conducted by Foley[4] is used. For the R-G and B-Y channels, the parameters are adjusted to fit similar data presented by Switkes et al.,[25] as shown in Figure 7 for the R-G channel.

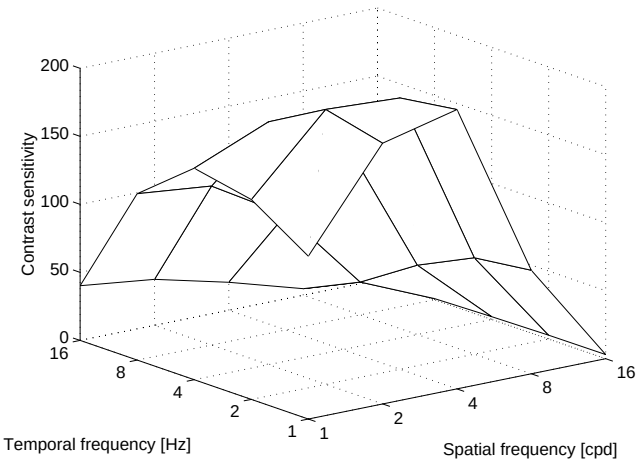

Figure 6. Spatio-temporal contrast sensitivity data for luminance gratings[15] as approximated by the model.

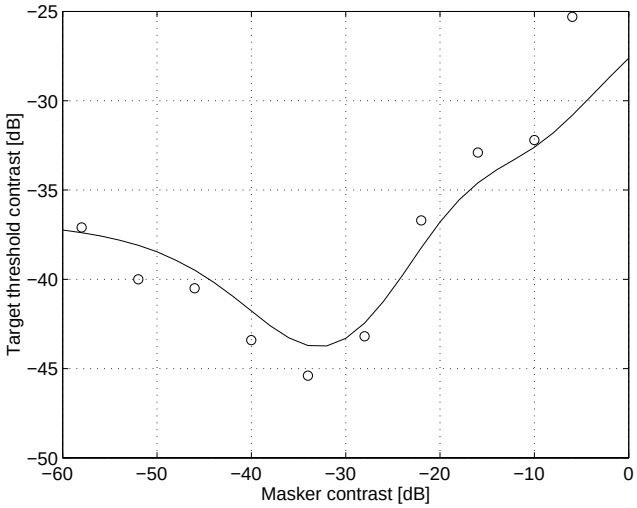

Figure 7. Contrast masking data for red-green gratings[25] as target threshold contrast vs. masker contrast; the solid curve shows the model fit.

Different exponents β have been found to yield good results for different experiments and implementations. $\beta = 2$ corresponds to the ideal observer formalism under independent Gaussian noise which assumes that the observer has complete knowledge of the stimuli and uses a matched filter for detection.[26,30] In a study of subjective experiments with coding artifacts, $\beta \approx 2$ was found to give good results.[3] Intuitively, a few high distortions may draw the viewer's attention more than many lower ones. This behavior can be emphasized with larger exponents. In this distortion metric, pooling over channels and over pixels is carried out with $\beta = 2$, whereas $\beta = 4$ is used for pooling over frames.

The simulation results shown in the figures demonstrate that the overall quality of the fits to the above-mentioned empirical data is quite good and close to the difference between measurements from different observers. Most of the effects found in the psychophysical experiments are captured by the model. However, two drawbacks of this modeling approach should be noted. Because of the nonlinear nature of the model, the parameters can only be determined by means of an iterative least-squares fitting process, which requires a lot of computation. Furthermore, the model lacks flexibility: once a good set of parameters has been found, it is only valid for a particular viewing setup. Future work will deal with how to eliminate these limitations.

3.2. Demonstration

I use the "Basketball" sequence to demonstrate how the proposed distortion metric works. This sequence contains a lot of spatial detail, a considerable amount of fast motion (the players in the foreground), and slow camera panning, which makes it an interesting sequence for a spatio-temporal model. Note that all of the sample images below are normalized in order to show as much spatial detail as possible.

The frame size of the sequence is 704×576 pixels. It was encoded at a bitrate of 4 Mbit/s with the MPEG-2 encoder of the MPEG Software Simulation Group.[†] A sample frame, its encoded counterpart, and the pixel-wise difference between them are shown in Figure 8. The B-W, R-G and B-Y components resulting from the conversion to opponent-colors space are shown in Figure 9. The B-W component looks different from the luminance image in Figure 8 because the transform coefficients differ and because of the gamma-correcting nonlinearity that has been applied as part of the color space conversion. Note the emphasis of the ball in the R-G channel and the blue and orange shirts of the players and the yellow curved line on the floor in the B-Y channel.

Figure 8. A sample frame from the "Basketball" sequence. The reference, its encoded counterpart, and the pixel-wise difference between them are shown (from left to right).

The color space conversions are followed by the perceptual decomposition. The results of applying the temporal low-pass and band filters to the B-W channel are shown in Figure 10. As can be seen, the ball virtually disappears in the low-pass channel, while it is clearly visible in the high-pass channel. As mentioned before, the R-G and B-Y channels are subjected only to the low-pass filter. The decomposition in the spatial domain increases the total number of channels even further; only a small selection is shown in Figure 11, namely the first, third and fifth level of the pyramid at an orientation of 45° constructed from the low-pass filtered B-W channel. The images are downsampled in the pyramid transform and have been upsampled to their original size in the figure. They show very well how different features are emphasized in the different subbands, e.g. the lines on the floor in the highest-frequency channel, the players leaning to the left in the medium-range frequency channel, and the barricades of the field in the lowest-frequency channel.

[†] This software is available at http://www.mpeg.org/~tristan/MPEG/MSSG/.

Figure 9. The B-W, R-G and B-Y components resulting from the conversion to opponent-colors space (from left to right).

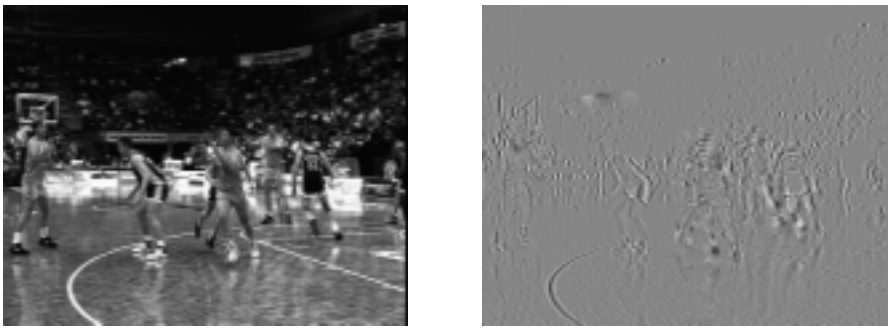

Figure 10. The low-pass (left) and band-pass (right) filtered B-W channel.

Figure 12 shows separate distortion maps for each color and temporal channel. Note that the distortion maps also include temporal aspects of the distortions, i.e. they depend on the previous frames. It is also evident that all four distortion maps are very different from the simple pixel-wise difference between the reference frame and the encoded frame shown in Figure 8. As can be seen, most of the visible artifacts appear in the B-W band-pass channel around the silhouettes of the players currently in motion. The distortions in the color channels are small compared to the other channels, but they have been normalized in the figures to reveal more spatial detail. Note that the distortions in the R-G and B-Y channels show a distinct block structure. This is due to the subsampling in the pyramid transform and shows that the model correctly emphasizes low-frequency distortions in the color channels.

4. CONCLUSIONS

A distortion metric for digital color video based on a model of the human visual system has been presented. The results show that the model can fit psychophysical contrast sensitivity and contrast masking data accurately. The output of the distortion metric is consistent with human observation. At present, the metric undergoes extensive testing with subjectively rated sequences by the Video Quality Experts Group (VQEG) and the ITU in an effort to evaluate and standardize video quality assessment systems. Future work includes comparing prediction accuracy of the color metric over a simple luminance-only metric, an extension of inhibitory pooling to study the interactions between a wider variety of channels, and an investigation of alternative color spaces.

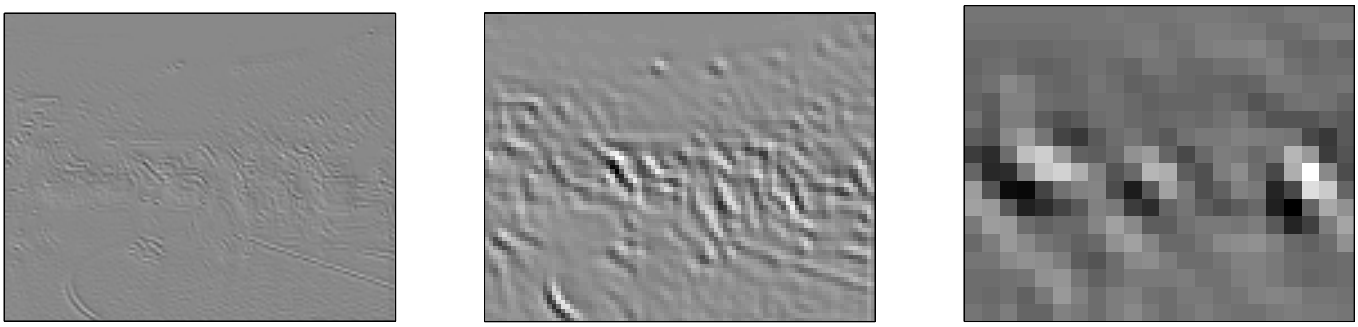

Figure 11. The first, third and fifth level at an orientation of 45° of the pyramid constructed from the low-pass filtered B-W channel.

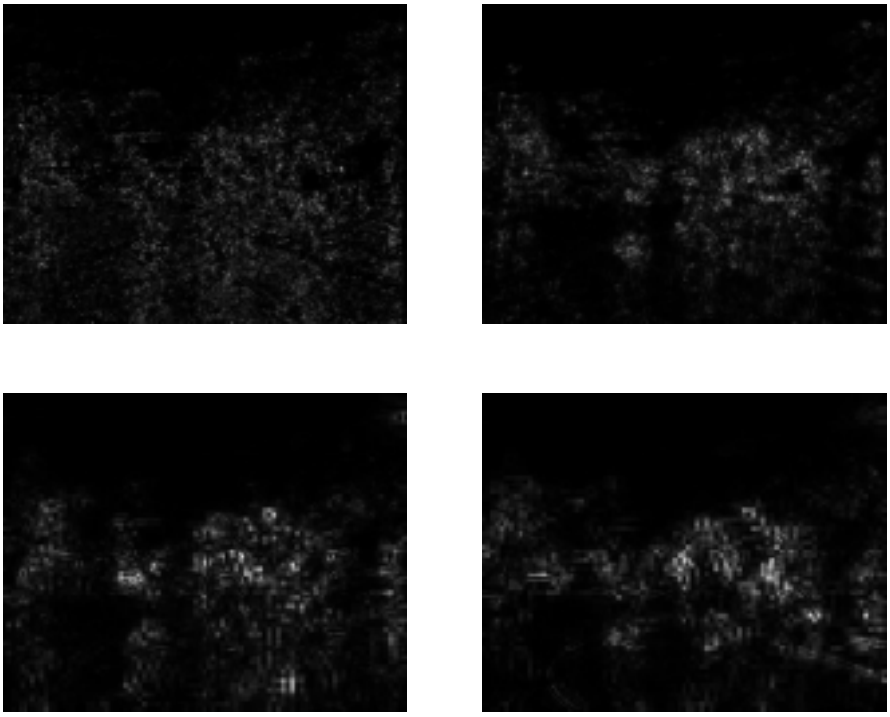

Figure 12. Distortion maps of the sample frame for the low-pass and band-pass B-W channels (top row), the R-G channel (bottom left) and the B-Y channel (bottom right). Note that these four images are not directly comparable, because each has been normalized to better show the spatial structure; the distortions in the color channels are smaller than in the B-W channels.

REFERENCES

1. Duane G. Albrecht, Wilson S. Geisler: "Motion selectivity and the contrast-response function of simple cells in the visual cortex." in *Visual Neuroscience*, vol. 7, pp. 531–546, 1991.
2. G. R. Cole, C. F. Stromeyer III, R. E. Kronauer: "Visual interactions with luminance and chromatic stimuli." in *Journal of the Optical Society of America A*, vol. 7, no. 1, pp. 128–140, January 1990.
3. Huib de Ridder: "Minkowski-metrics as a combination rule for digital-image-coding impairments." in *Proceedings of SPIE Human Vision, Visual Processing and Digital Display*, vol. 1666, pp. 16–26, San Jose, CA, February 9–14, 1992, ISBN 0-8194-0820-4.
4. John M. Foley: "Human luminance pattern-vision mechanisms: Masking experiments require a new model." in *Journal of the Optical Society of America A*, vol. 11, no. 6, pp. 1710–1719, June 1994.

5. R. E. Fredericksen, R. F. Hess: "Temporal detection in human vision: Dependence on stimulus energy." in *Journal of the Optical Society of America A*, vol. 14, no. 10, pp. 2557–2569, October 1997.
6. R. E. Fredericksen, R. F. Hess: "Estimating multiple temporal mechanisms in human vision." in *Vision Research*, vol. 38, no. 7, pp. 1023–1040, 1998.
7. S. T. Hammett, A. T. Smith: "Two temporal channels or three? A re-evaluation." in *Vision Research*, vol. 32, no. 2, pp. 285–291, 1992.
8. David J. Heeger: "Half-squaring in responses of cat striate cells." in *Visual Neuroscience*, vol. 9, pp. 427–443, 1992.
9. David J. Heeger: "Normalization of cell responses in cat striate cortex." in *Visual Neuroscience*, vol. 9, pp. 181–197, 1992.
10. Ewald Hering: *Zur Lehre vom Lichtsinne*. Carl Gerolds & Sohn, Vienna, Austria, 1878.
11. R. F. Hess, R. J. Snowden: "Temporal properties of human visual filters: Number, shapes and spatial covariation." in *Vision Research*, vol. 32, no. 1, pp. 47–59, 1992.
12. Leo M. Hurvich, Dorothea Jameson: "An opponent-process theory of color vision." in *Psychological Review*, vol. 64, pp. 384–404, 1957.
13. ITU-R Recommendation BT.601-5: "Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios." ITU, Geneva, Switzerland, 1995.
14. Dorothea Jameson, Leo M. Hurvich: "Some quantitative aspects of an opponent-colors theory. I. Chromatic responses and spectral saturation." in *Journal of the Optical Society of America*, vol. 45, no. 7, pp. 546–552, July 1955.
15. D. H. Kelly: "Motion and vision. II. Stabilized spatio-temporal threshold surface." in *Journal of the Optical Society of America*, vol. 69, no. 10, pp. 1340–1349, October 1979.
16. D. H. Kelly: "Spatiotemporal variation of chromatic and achromatic contrast thresholds." in *Journal of the Optical Society of America*, vol. 73, no. 6, pp. 742–750, June 1983.
17. Pär [J.] Lindh, Christian J. van den Branden Lambrecht: "Efficient spatio-temporal decomposition for perceptual processing of video sequences." in *Proceedings of the International Conference on Image Processing*, vol. 3, pp. 331–334, Lausanne, Switzerland, September 16–19, 1996.
18. M. A. Losada, K. T. Mullen: "The spatial tuning of chromatic mechanisms identified by simultaneous masking." in *Vision Research*, vol. 34, no. 3, pp. 331–341, 1994.
19. Marc B. Mandler, Walter Makous: "A three-channel model of temporal frequency perception." in *Vision Research*, vol. 24, no. 12, pp. 1881–1887, 1984.
20. Allen B. Poirson, Brian A. Wandell: "Appearance of colored patterns: Pattern-color separability." in *Journal of the Optical Society of America A*, vol. 10, no. 12, pp. 2458–2470, December 1993.
21. Allen B. Poirson, Brian A. Wandell: "Pattern-color separable pathways predict sensitivity to simple colored patterns." in *Vision Research*, vol. 36, no. 4, pp. 515–526, 1996.
22. R. F. Quick, Jr.: "A vector-magnitude model of contrast detection." in *Kybernetik*, vol. 16, pp. 65–67, 1974.
23. Eero P. Simoncelli, William T. Freeman, Edward H. Adelson, David J. Heeger: "Shiftable multi-scale transforms." in *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992.
24. Andrew Stockman, Donald I. A. MacLeod, Nancy E. Johnson: "Spectral sensitivities of the human cones." in *Journal of the Optical Society of America A*, vol. 10, no. 12, pp. 2491–2521, December 1993.
25. Eugene Switkes, Arthur Bradley, Karen K. De Valois: "Contrast dependence and mechanisms of masking interactions among chromatic and luminance gratings." in *Journal of the Optical Society of America A*, vol. 5, no. 7, pp. 1149–1162, July 1988.
26. Patrick C. Teo, David J. Heeger: "Perceptual image distortion." in *Proceedings of SPIE Human Vision, Visual Processing and Digital Display*, vol. 2179, pp. 127–141, San Jose, CA, February 8–10, 1994, ISBN 0-8194-1474-3.
27. Christian J. van den Branden Lambrecht: *Perceptual Models and Architectures for Video Coding Applications*. Ph.D. thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 1996.
28. Andrew B. Watson: "Temporal sensitivity." in *Handbook of Perception and Human Performance*, eds. Kenneth R. Boff, Lloyd Kaufman, James P. Thomas, vol. 1, chap. 6, John Wiley & Sons, 1986, ISBN 0-471-88544-4.
29. Andrew B. Watson, Joshua A. Solomon: "Model of visual contrast gain control and pattern masking." in *Journal of the Optical Society of America A*, vol. 14, no. 9, pp. 2379–2391, September 1997.
30. Stefan Winkler: "A perceptual distortion metric for digital color images." in *Proceedings of the International Conference on Image Processing*, vol. 3, pp. 399–403, Chicago, IL, October 4–7, 1998.