Toward Perceptual Metrics for Video Watermark Evaluation

Stefan Winkler,* Elisa Drelie Gelasca,† Touradj Ebrahimi†

Genista Corporation, Rue du Théâtre 5, 1820 Montreux, Switzerland

ABSTRACT

Assessing and comparing the performance of watermarking algorithms is difficult. The visibility of the watermark is an important aspect in this process. In this paper, we propose two metrics for evaluating the visual impact of video watermarks. Based on several different watermarking algorithms and video sequences, we identify the most prominent impairments as spatial noise and temporal flicker. We design the corresponding measurement algorithms and corroborate their performance through subjective experiments.

1. INTRODUCTION

The rapid spread of digital media (audio, images and video) and the ease of their reproduction and distribution has created a need for copyright enforcement schemes in order to protect content creators and owners. In recent years, digital watermarking has emerged as an effective way to prevent users from violating copyrights. This concept is based on the insertion of information into the data in such a way that the added information is not perceptible yet resistant to (intentional or unintentional) alterations of the watermarked data.

Three factors must be considered in image or video watermarking:

- Capacity, i.e. the amount of information that can be put into the watermark and recovered without errors;
- Robustness, i.e. the resistance of the watermark to alterations of the original content such as compression, filtering or cropping;
- Visibility, i.e. how easily the watermark can be discerned by the user.

These factors are inter-dependent; for example, increasing the capacity will decrease the robustness and/or increase the visibility. Therefore, it is essential to consider all three factors for a fair evaluation or comparison of watermarking algorithms. Organizations such as Certimark (http://www.certimark.org) or the Content ID Forum (http://www.cidf.org) have been working on the definition of procedures for such evaluations. While benchmark tests have already been proposed for the robustness of watermarking algorithms, such as CheckMark[1] or StirMark,[2] much less attention has been directed at measuring the visual effects of the watermarking process. In this paper, we propose two metrics for the objective quality evaluation of watermarked video. They were briefly introduced elsewhere;[3] here we present the measurement algorithms and their evaluation in detail.

The paper is organized as follows: In Section 2 we describe perceptual quality assessment methods in general and our video watermarking metrics in particular. Section 3 outlines the algorithms and video sequences used to test these metrics as well as the design of the subjective experiments. The results are reported and discussed in Section 4. Section 5 offers some conclusions and directions for future work.

E-mail of corresponding author: stefan.winkler@genista.com

[*] S. Winkler is now with the Audiovisual Communications Laboratory at the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

[†] E. Drelie Gelasca and T. Ebrahimi are with the Signal Processing Laboratory at the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

2. PERCEPTUAL QUALITY ASSESSMENT

2.1. Background

The accurate measurement of quality as perceived by a human observer is a great challenge in image or video processing in general. The reason for this is that the amount and visibility of distortions such as those introduced by watermarking strongly depend on the actual image/video content.

The benchmark for any kind of visual quality assessment are subjective experiments, where a number of people are asked to watch test clips and to rate their quality. Procedures for such experiments have been formalized in ITU-R Rec. BT.500[4] or ITU-T Rec. P.910,[5] which suggest viewing conditions, criteria for the selection of observers and test material, assessment procedures, and data analysis methods. However, subjective experiments are intrinsically time-consuming, hence expensive and often impractical. Furthermore, for many applications – such as online quality monitoring and control – subjective experiments cannot be used at all.

Given these limitations, engineers have turned to simple error measures such as mean squared error (MSE) or peak signal-to-noise ratio (PSNR), assuming that they would yield quality indications comparable to human perception. However, these simple measures operate solely on the basis of pixel-wise differences and neglect the important influence of video content and viewing conditions on the actual visibility of artifacts. Therefore, they cannot be expected to be reliable predictors of perceived quality.

The shortcomings of these methods have led to the intensive study of advanced perceptual quality metrics in recent years.[6, 7] Essentially two different approaches can be distinguished:

- Approaches based on models of the human visual system. These are the most general and potentially most accurate ones.[8–10] However, the human visual system is extremely complex, and many of its properties are not well understood even today. Besides, implementing these models is very expensive from a computational point of view due to their complexity.
- Approaches based on a priori knowledge about the compression methods as well as the pertinent types of artifacts.[11–13] While such metrics are not as versatile, they normally perform well in a given application area. Their main advantage lies in the fact that they often permit a computationally more efficient implementation.

In this paper, we take the latter approach, i.e. we first identify the artifacts caused by watermarking, and then we try to find metrics to measure their perceptual severity.

2.2. Metrics

From our experience with the numerous video watermarking algorithms that we tested, we have seen mainly two kinds of impairments:

- Spatial noise, which is the fundamental fingerprint of most watermarks;
- Temporal flicker, which results from visible changes of the watermark pattern between consecutive frames.

Based on these observations, we have designed objective metrics that measure the perceptual impact of these two impairments, which we refer to as Noise metric and Flicker metric in the following. These metrics are part of Genista's *Video PQoS*^{TM} software,[*] which was used for analyzing the videos in our experiments. Video PQoS is an application for the measurement of artifacts affecting the perceptual quality of digital video. It does this through full reference quality metrics, i.e. it compares the video under test with a reference video to measure the quality of the degraded video with respect to the reference. In addition to the watermarking metrics, Video PQoS also provides perceptual metrics for compression artifacts, metrics as defined by ANSI T1.801.03,[14] and fidelity metrics such as PSNR.

[*] See http://www.genista.com for more information.

2.2.1. Noise Metric

For the computation of the noise metric, the watermark is first extracted as the difference d between a frame in the processed sequence and the corresponding frame in the reference sequence: $d(x, y) = Y_{\text{prc}}(x, y) - Y_{\text{ref}}(x, y)$.

Let $D(u, v)$ be the coefficients of the two-dimensional discrete Fourier transform of $d(x, y)$. Based on the vector feature proposed in section 6.1.2 of ANSI T1.801.03,[14] the radial average r_d of the 2D-DFT coefficients is computed as the absolute sum over the Fourier coefficients inside a ring $\mathcal{R}_k$ with radius $k - 1 < \sqrt{u^2 + v^2} < k$ for each k:

$$r_d(k) = \frac{1}{N_{\mathcal{R}_k}} \sum_{(u,v)\in\mathcal{R}_k} |D(u, v)|, \tag{1}$$

where $N_{\mathcal{R}_k}$ denotes the number of coefficients within ring $\mathcal{R}_k$.

Finally, the sum over the higher frequency range $(f_M \dots f_H)$ of this radial spectrum $r_d(k)$ is computed to yield the Noise metric:

$$\text{Noise} = \frac{1}{f_H - f_M} \sum_{k=f_M}^{f_H} r_d(k). \tag{2}$$

We choose the frequency limits to be $f_M = 16\%$ and $f_H = 80\%$ of the maximum spatial frequency.

2.2.2. Flicker Metric

As before, the watermark is extracted as the difference $d(x, y)$ between a frame in the processed sequence and the corresponding frame in the reference sequence. This is done for two consecutive frames, giving d_n and d_{n+1}. The change of this watermark from one frame to the next is computed as $c(x, y) = d_{n+1}(x, y) - d_n(x, y)$.

We again compute the radial frequency spectrum as described above, but this time using the 2D-DFT of $c(x, y)$:

$$r_c(k) = \frac{1}{N_{\mathcal{R}_k}} \sum_{(u,v)\in\mathcal{R}_k} |C(u, v)|. \tag{3}$$

The sum over the low frequencies $(f_L \dots f_M)$ of r_c,

$$s_L = \frac{1}{f_M - f_L} \sum_{k=f_L}^{f_M} r_c(k), \tag{4}$$

as well as the sum over the high frequencies $(f_M \dots f_H)$ of r_c,

$$s_H = \frac{1}{f_H - f_M} \sum_{k=f_M}^{f_H} r_c(k), \tag{5}$$

are calculated. We choose the frequency limits to be $f_L = 1\%$, $f_M = 16\%$, and $f_H = 80\%$ of the maximum spatial frequency.

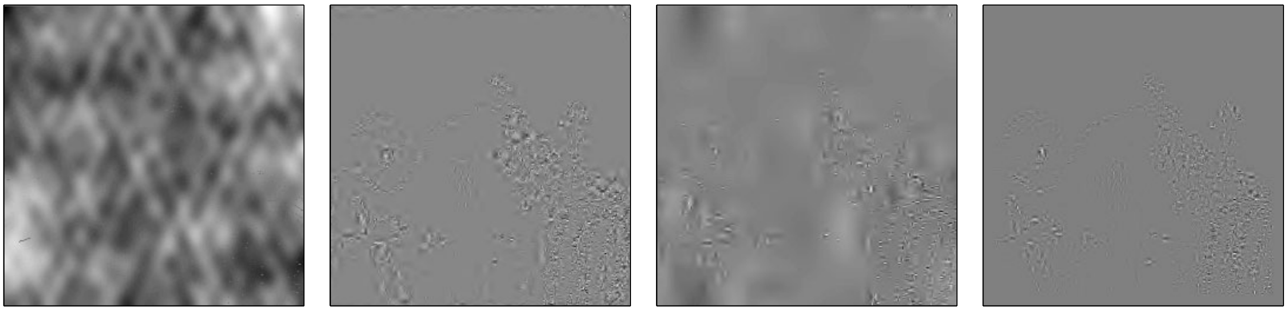
To take into account spatial and temporal masking by the reference sequence, we estimate the spatial and and temporal activity in the reference sequence. This is again based on simple features defined in ANSI T1.801.03,[14] namely the spatial information SI, which is the gradient computed from the horizontally and vertically Sobel-filtered image, as well as the temporal information TI, which is simply the pixel-wise difference between two consecutive frames. We normalize both SI and TI with respect to the maximum possible values. Using the average spatial information (the average gradient magnitude, to be precise) of the reference frame, $\overline{SI} = \sum |SI_r(x, y)|$, and the average temporal information of the reference frame, $\overline{TI} = \sum |TI(x, y)|$, a scalar weight m is computed:

$$m = \max\left(\overline{SI} \cdot \overline{TI}, t\right), \tag{6}$$

where t is a threshold to avoid extreme values of masking. We choose $t = 0.007$.

From the above, the Flicker metric is computed as:

$$\text{Flicker} = \frac{s_L + s_H}{m}. \tag{7}$$



(a) Cox et al.[15] (b) Dugad et al.[16] (c) Wang et al.[17] (d) Xia et al.[18]

Figure 1. Watermark produced by four different algorithms for the frame shown in Figure 2(b). Dark pixels denote negative values, bright pixels denote positive values, medium gray denotes no change. The images were normalized individually to enhance the visibility of the watermarks.

3. EXPERIMENTS

3.1. Watermarking Algorithms

Most video watermarking techniques today are derived from algorithms for still images. Therefore, we adopt a number of watermarking schemes for still images and apply them to each frame of a video sequence. We chose four algorithms from the literature,* as well as a genuine video watermarking algorithm for videos developed by AlpVision. A brief description of each of these algorithms is given in the following.

The scheme of Cox et al.[15] is based on the discrete cosine transform (DCT). The DCT of the entire image is computed, and a sequence of n real numbers is generated from a uniform distribution of zero mean and unit variance, which is then placed into the n highest magnitude coefficients of the transform matrix. Additionally, a scaling parameter α can be specified to determine the amplitude of the watermark.

Dugad et al.[16] use a three level discrete wavelet transform (DWT) with an eight-tap Daubechies filter. The watermark is generated by a sequence of n real numbers and is added to the coefficients above a given threshold in all sub-bands except the low-pass band. The watermark amplitude can again be controlled by a scaling parameter.

Wang et al.[17] adopt a successive subband quantization scheme in the multi-threshold wavelet codec to choose perceptually significant coefficients for watermark embedding. The watermark is inserted in the coefficients above a certain threshold in the current subband while taking into account the scaling factors α and β, which are adjustable by the user.

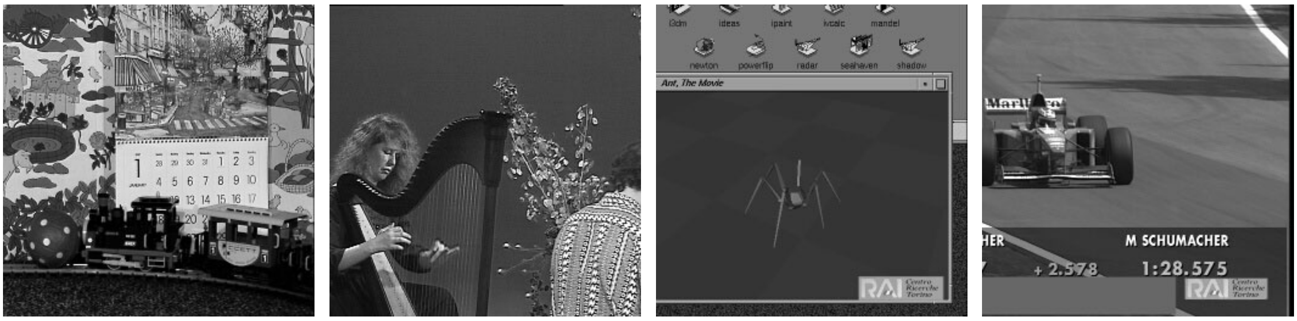
Xia et al.[18] decompose an image into several bands. The watermark is added to the largest coefficients in the high- and middle-frequency bands of the DWT. A parameter α is tuned to control the level of the watermark. The output of the inverse DWT is modified such that the resulting image has the same dynamic range as the original.

The video watermarking scheme developed by AlpVision† is based on a technique initially proposed for still images.[19, 20] It uses spread-spectrum modulation to insert a watermark with variable amplitude and density in the spatial domain. In contrast to the other four algorithms, it considers the temporal content changes in the video.

The default settings of each algorithm were used for all parameters. The resulting watermarks are shown in Figure 1 for a sample frame from one of our test clips. As can be seen, some watermarking algorithms take into account masking phenomena in the human visual system to a certain extent and insert their watermarks mainly in image regions with high spatial activity (bottom right part of the frame).

* The source code for these algorithms can be downloaded from http://www.cosy.sbg.ac.at/~pmeerw/Watermarking/source/.

† See http://www.alpvision.com for more information.



(a) Train (b) Harp (c) Graphics (d) Race

Figure 2. Sample frames from the test clips.

3.2. Test Clips

We watermarked four different test clips for our analysis. These clips were selected from the set of scenes in the first VQEG test[21] to include spatial detail, motion, and synthetic content. They are 8 seconds long with a frame rate of 25 fps. They were de-interlaced and subsampled from the interlaced ITU-R Rec. BT.601 format[22] to a resolution of 360×288 pixels for progressive display. The implementations of some watermarking algorithms mentioned in the previous section are limited to frame sizes of powers of 2, therefore we cropped a 256×256 pixel region from each frame in the video for watermarking and subsequent quality evaluation. A sample frame from each of the four scenes is shown in Figure 2.

3.3. Subjective Experiments

For the evaluation of our metrics, subjective experiments were performed. Non-expert observers were asked to rank a total of 20 watermarked test clips from best to worst according to perceived noise and flicker in two separate trials. The viewing order of the clips was not fixed; observers could freely choose between clips and play them as often as they liked. They could also watch the original clips for comparison. Five observers participated in the noise trial, and six in the flicker trial. For comparison with the objective metrics, the data obtained from the subjective ratings were combined to an average rank.

According to the subjective experiments, the most annoying artifacts in video are produced by watermarking algorithms that add noise patterns with relatively low spatial frequencies, which change from frame to frame and thus create clearly visible flicker. Algorithms that add mainly high-frequency noise or temporally unchanging patterns to the video exhibit much less visible distortion.

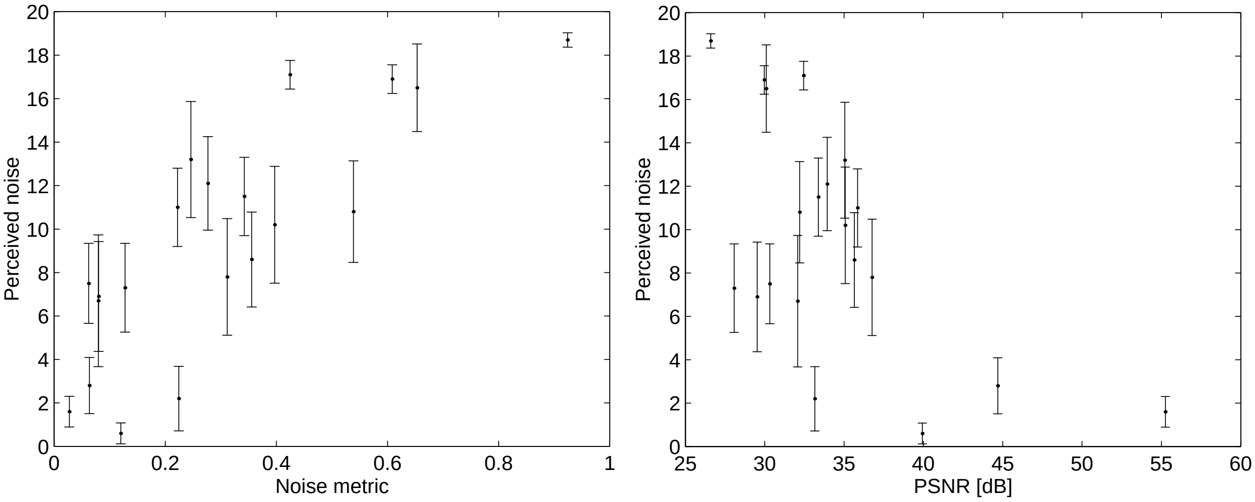
4. RESULTS

A statistical analysis of the data was carried out to evaluate the two proposed metrics with respect to the subjective ratings. Two correlation coefficients are used here to quantify and compare the metrics' performance, namely the (linear) Pearson correlation coefficient as well as the (non-parametric) Spearman rank-order correlation coefficient.

The scatter plot of perceived versus measured noise for the above-mentioned watermarking algorithms and test clips is shown in Figure 3(a). For comparison, the scatter plot of perceived noise versus PSNR is shown in Figure 3(b). The respective correlation coefficients are reported in Figure 3(c).

Figure 4 shows the same data for perceived flicker, PSNR and the Flicker metric.

The proposed metrics clearly outperform PSNR in both cases. The plots show that adding a temporal component such as flicker to the measurements is essential for the evaluation of video watermarks, because PSNR is unable to take this into account. More surprisingly perhaps, PSNR is not well correlated with perceived noise either. This shows the importance of more discriminatory metrics for the perceptual quality evaluation of watermarks.



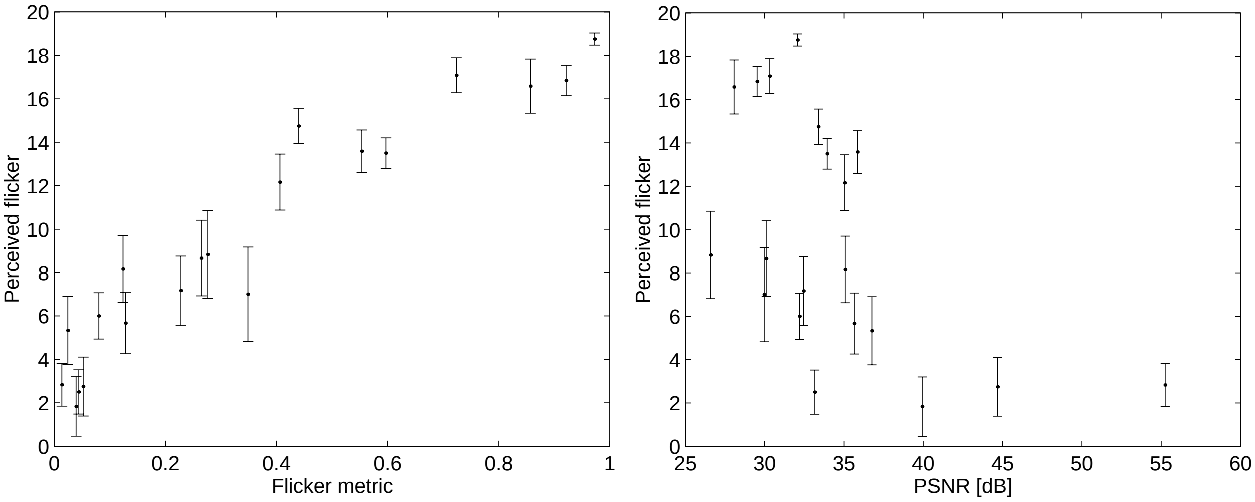
(a) Subjective noise ratings vs. Noise metric.

(b) Subjective noise ratings vs. PSNR.

Noise	Metric	PSNR
Pearson	0.81	−0.60
Spearman	0.81	−0.41

(c) Correlations.

Figure 3. Perceived noise vs. Noise metric and PSNR (subjective data are shown with 90%-confidence intervals).



(a) Subjective flicker ratings vs. Flicker metric.

(b) Subjective flicker ratings vs. PSNR.

Flicker	Metric	PSNR
Pearson	0.95	−0.54
Spearman	0.94	−0.58

(c) Correlations.

Figure 4. Perceived flicker vs. Flicker metric and PSNR (subjective data are shown with 90%-confidence intervals).

5. CONCLUSIONS

We have discussed the importance of perceptual quality assessment in watermarking. While this remains a difficult problem, we presented a possible solution path. We found that watermarked video suffered mostly from added high-frequency noise and/or flicker in our tests. The watermarking artifacts, which may be hardly noticeable in still images, become emphasized through the motion effects in video.

We introduced two measurement algorithms that analyze the video by specifically looking for watermarking impairments, namely a Noise metric and a Flicker metric, which measure the perceptual impact of these specific distortions. Through subjective experiments we have demonstrated that the proposed metrics are reliable predictors of perceived noise and perceived flicker and clearly outperform PSNR in terms of prediction accuracy.

Further improvements of the metrics could be achieved using a local masking model for both metrics; our test set turned out to be too small to effectively see any gains in prediction performance with such an improved model. Additional test sequences could give better indications of how well they generalize to different types of content, and more genuine video watermarking schemes should be used to evaluate the metrics in realistic conditions. Finally, an extension of the metrics to color is necessary for a reliable evaluation of watermarking algorithms that use all three color channels.

6. ACKNOWLEDGMENTS

The authors would like to thank Frederic Jordan and Martin Kutter of AlpVision, Switzerland, for watermarking the test clips with their algorithm.

REFERENCES

1. S. Pereira, S. Voloshynovskiy, M. Madueño, S. Marchand-Maillet, and T. Pun, "Second generation benchmarking and application oriented evaluation," in *Proceedings of the Information Hiding Workshop*, pp. 340–353, (Pittsburgh, PA), April 2001.
2. F. A. P. Petitcolas and R. J. Anderson, "Evaluation of copyright marking systems," in *Proceedings of the International Conference on Multimedia Computing and Systems*, pp. 574–579, (Florence, Italy), June 7–11 1999.
3. S. Winkler, E. Drelie Gelasca, and T. Ebrahimi, "Perceptual quality assessment for video watermarking," in *Proceedings of the International Conference on Information Technology: Coding and Computing*, pp. 90–94, (Las Vegas, NV), April 8–10, 2002.
4. ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures." International Telecommunication Union, Geneva, Switzerland, 2002.
5. ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications." International Telecommunication Union, Geneva, Switzerland, 1996.
6. S. Winkler, *Vision Models and Quality Metrics for Image Processing Applications.* PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 2000.
7. A. M. Rohaly, P. Corriveau, J. Libert, A. Webster, V. Baroncini, J. Beerends, J.-L. Blin, L. Contin, T. Hamada, D. Harrison, A. Hekstra, J. Lubin, Y. Nishida, R. Nishihara, J. Pearson, A. F. Pessoa, N. Pickford, A. Schertz, M. Visca, A. Watson, and S. Winkler, "Video Quality Experts Group: Current results and future directions," in *Proceedings of SPIE Visual Communications and Image Processing*, **4067**, pp. 742–753, (Perth, Australia), June 21–23, 2000.
8. S. Winkler, "Issues in vision modeling for perceptual video quality assessment," *Signal Processing* **78**, pp. 231–252, October 1999.
9. J. Lubin and D. Fibush, "Sarnoff JND vision model." T1A1.5 Working Group Document #97-612, ANSI T1 Standards Committee, 1997.
10. S. Winkler, "A perceptual distortion metric for digital color video," in *Proceedings of SPIE Human Vision and Electronic Imaging*, **3644**, pp. 175–184, (San Jose, CA), January 23–29, 1999.
11. K. T. Tan, M. Ghanbari, and D. E. Pearson, "An objective measurement tool for MPEG video quality," *Signal Processing* **70**, pp. 279–294, November 1998.

12. A. B. Watson, J. Hu, J. F. McGowan III, and J. B. Mulligan, "Design and performance of a digital video quality metric," in *Proceedings of SPIE Human Vision and Electronic Imaging*, **3644**, pp. 168–174, (San Jose, CA), January 23–29, 1999.
13. S. Winkler and R. Campos, "Video quality evaluation for Internet streaming applications," in *Proceedings of SPIE Human Vision and Electronic Imaging*, **5007**, pp. 104–115, (Santa Clara, CA), January 21–24, 2003.
14. ANSI T1.801.03, "Digital transport of one-way video signals – parameters for objective performance assessment." American National Standards Institute, New York, NY, 1996.
15. I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing* **6**, pp. 1673–1687, December 1997.
16. R. Dugad, K. Ratakonda, and N. Ahuja, "A new wavelet-based scheme for watermarking images," in *Proceedings of the International Conference on Image Processing*, **2**, pp. 419–423, (Chicago, IL), October 4–7, 1998.
17. H.-J. M. Wang, P.-C. Su, and C.-C. J. Kuo, "Wavelet-based digital image watermarking," *Optics Express* **3**, pp. 491–496, December 1998.
18. X.-G. Xia, C. G. Boncelet, and G. R. Arce, "Wavelet transform based watermark for digital images," *Optics Express* **3**, pp. 497–511, December 1998.
19. M. Kutter, *Digital Image Watermarking: Hiding Information in Images.* PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 1999.
20. M. Kutter and S. Winkler, "A vision-based masking model for spread-spectrum image watermarking," *IEEE Transactions on Image Processing* **11**, pp. 16–25, January 2002.
21. VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment," April 2000. Available at http://www.vqeg.org/.
22. ITU-R Recommendation BT.601-5, "Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios." International Telecommunication Union, Geneva, Switzerland, 1995.