

# Perceptual Quality Assessment for Video Watermarking

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

Genista Corporation  
EPFL – PSE – Genimedia  
1015 Lausanne, Switzerland  
<http://www.genista.com/>  
[swinkler@genimedia.com](mailto:swinkler@genimedia.com)

## Abstract

*The reliable evaluation of the performance of watermarking algorithms is difficult. An important aspect in this process is the assessment of the visibility of the watermark. In this paper, we address this issue and propose a methodology for evaluating the visual quality of watermarked video. Using a software tool that measures different types of perceptual video artifacts, we determine the most relevant impairments and design the corresponding objective metrics. We demonstrate their performance through subjective experiments on several different watermarking algorithms and video sequences.*

## 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;

\* E. Drelie Gelasca and T. Ebrahimi are with the Signal Processing Lab at EPFL, Switzerland.

- 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 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<sup>†</sup> or the Content ID Forum<sup>‡</sup> are 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 [9] or StirMark [10], much less attention has been directed at the evaluation of the visual effects of the watermarking process. In this paper, we propose methods for the objective evaluation of watermarked video quality.

## 2 Perceptual Quality Assessment

### 2.1 Background

Unfortunately, the accurate measurement of quality as perceived by the typical user is a big 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 Recommendation BT.500 [4] or ITU-T Recommendation P.910 [6], which suggest standard viewing conditions, criteria for the selection of observers and test material, assessment procedures, and data analysis methods. However,

<sup>†</sup> <http://www.certimark.org>

<sup>‡</sup> <http://www.cidf.org>

subjective experiments require careful setup and are 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), suggesting that they would be as valid as subjective experiments. 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 more advanced perceptual quality metrics in recent years. An up-to-date review of such metrics can be found in [18]. 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 [16]. Examples of such metrics are described in [8, 12, 17] among others. 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. Examples of such specialized metrics include [11, 15]. 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.

## 2.2 Metrics

We have chosen the VideoQoS\* software for the evaluation of the visual impact of the watermarks. VideoQoS is an application that provides for the measurement of artifacts affecting the perceptual quality of digital video. It does this through full reference quality metrics, i.e. it compares a reference video with a processed one to measure the quality of the degraded video in relation to the reference. VideoQoS provides separate metrics for different types of visual artifacts, which are divided into three categories:

- ANSI Metrics, which rely on algorithms defined by ANSI [1]. This document represents an attempt by a standards body to define objective measures that serve as a basis for the measurement of video quality.
- Perceptual Metrics, which measure specific visual artifacts introduced into the video in a way that is correlated with human perception. These artifacts are intuitive and well known, and are easily recognized even by inexperienced viewers.
- Fidelity Metrics, which are rather wide-spread and represent mathematical fidelity measures of the video, e.g. MSE/PSNR. They do not take into account human perception.

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

- Flicker, which results from visible changes of the watermark pattern between consecutive frames;
- High-frequency (HF) noise, which is the fundamental fingerprint of most watermarks.

Based on these observations, we have designed objective metrics that measure the perceptual impact of these two impairments, which we refer to as Flicker metric and Noise metric in the following.

## 3 Experiments

### 3.1 Watermarking Algorithms

Our interest here is mainly video watermarking. Most video watermarking techniques today are derived from algorithms for still images. Therefore, we adopt a number of watermarking schemes for still images<sup>†</sup> and apply them to each frame of a video sequence. We chose four algorithms from the literature (Cox, Dugad, Wang and Xia). To enhance the test set, we also included a genuine video watermarking algorithm for videos developed by AlpVision. A brief description of each of these algorithms is given in the following.

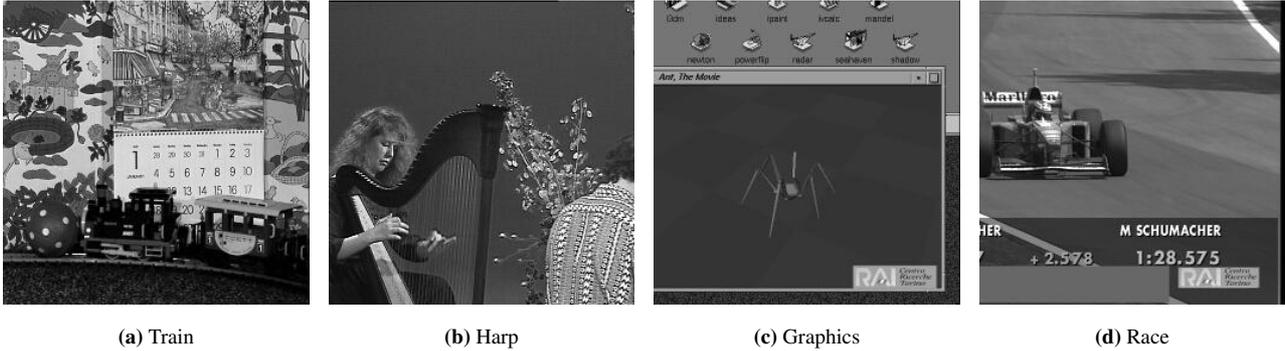
The scheme of Cox et al. [2] is based on the discrete cosine transform (DCT). The DCT of the entire image is computed, and a sequence of  $n$  real numbers is generated from a uniform distribution of zero mean and unit variance, which is then placed into the  $n$  highest magnitude coefficients of the transform matrix. Additionally, a scaling parameter  $\alpha$  can be specified to determine the amplitude of the watermark.

---

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

---

\* VideoQoS version 1.4.1, July 2001, Genista Corp.



**Figure 1:** Frames from test clips.

Dugad et al. [3] 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 subbands except the low-pass band. The watermark amplitude can again be controlled by a scaling parameter.

Wang et al. [14] 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  $\alpha$  and  $\beta$ , which are adjustable by the user.

Xia et al. [19] 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  $\alpha$  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 by Kutter [7]. 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.

### 3.2 Test Clips

We watermarked four different test clips for our analysis. These clips were selected from the set of scenes in the VQEG test [13] to include spatial detail, saturated colors, motion, and synthetic content. They are 8 seconds long with a frame rate of 25 Hz. They were de-interlaced and subsampled from the interlaced ITU-R Rec. BT.601 format [5] to a resolution of  $360 \times 288$  pixels per frame for progressive

display. The implementations of four watermarking algorithms are limited to frame sizes of powers of 2, therefore we have cropped a  $256 \times 256$  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 1.

### 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 flicker and noise 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. Six observers participated in the flicker trial, and five in the noise trial. The data obtained from the subjective ratings were combined to an average rank for comparison with the objective metrics.

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 distortions.

### 3.4 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 flicker for the above-mentioned watermarking algorithms and test

\* <http://www.alpvision.com>

clips is shown in Figure 2(a). For comparison, the scatter plot of perceived flicker versus PSNR is shown in Figure 2(b). The respective correlation coefficients are reported in the table in Figure 2(c).

Figure 3 shows the same data for perceived noise and the HF-Noise 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.

For further improvement and testing of our metrics, more genuine video watermarking algorithms should be used. Also, an extension of the set of test clips could give better indications of how well these metrics generalize to different types of content.

## 4 Conclusions

We have discussed the importance of perceptual quality assessment in watermarking. While this remains a difficult problem, we have presented a possible solution. We have introduced a measurement tool that analyzes video impairments by looking for certain types of artifacts. Using this tool, we have demonstrated that typical video watermarks suffer mostly from added high-frequency noise and/or flicker. The watermarking artifacts, which may be hardly noticeable in still images, become emphasized through the motion effects in video.

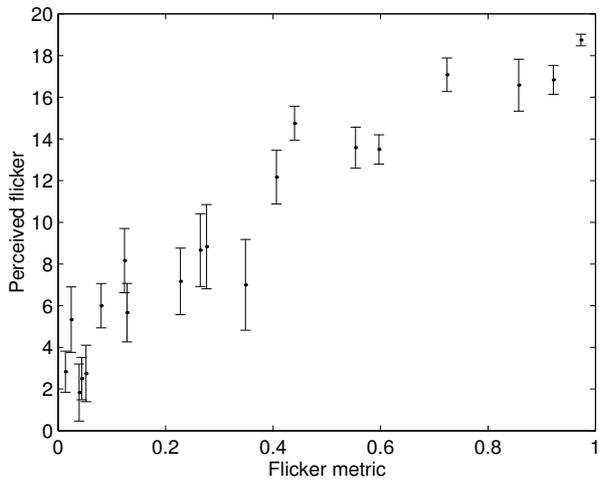
We have proposed a Flicker metric and an HF-Noise metric to measure the perceptual impact of these specific distortions. Through subjective experiments we have demonstrated that the proposed metrics are reliable predictors of perceived flicker and perceived noise and clearly outperform PSNR in terms of prediction accuracy.

## 5 Acknowledgments

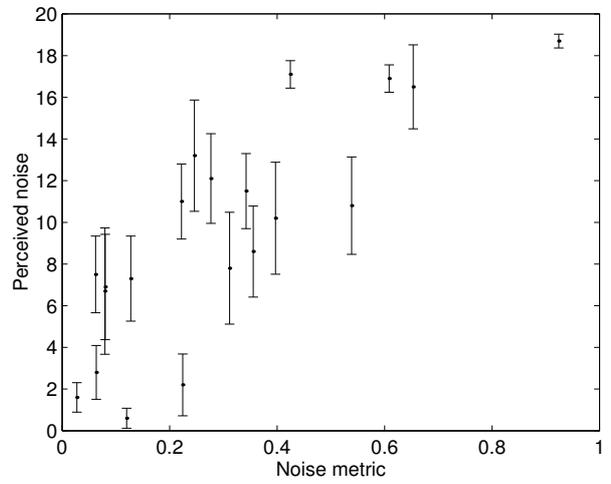
The authors wish to thank Frederic Jordan and Martin Kutter of AlpVision, Switzerland, for watermarking the test clips with their algorithm.

## References

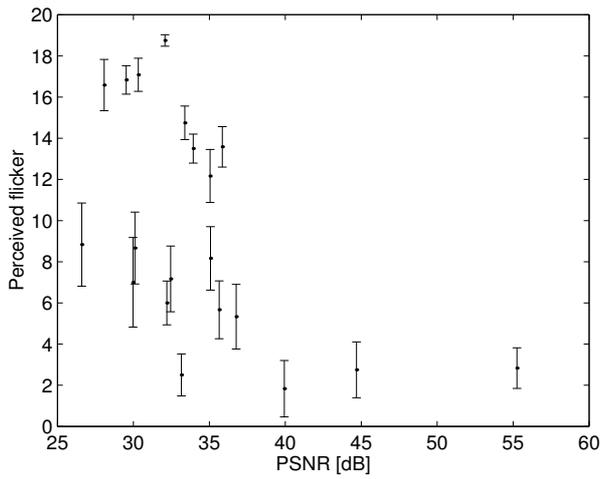
- [1] ANSI T1.801.03. Digital transport of one-way video signals – parameters for objective performance assessment. American National Standards Institute, New York, NY, 1996.
- [2] I. J. Cox, J. Kilian, T. Leighton, and T. G. Shamoan. Secure spread spectrum watermarking for multimedia. In *Proc. ICIP*, volume 6, pages 1673–1687, Santa Barbara, CA, USA, Oct. 26–29, 1997.
- [3] R. Dugad, K. Ratakonda, and N. Ahuja. A new wavelet-based scheme for watermarking for multimedia. In *Proc. ICIP*, Chicago, IL, USA, Oct. 4–7, 1998.
- [4] ITU-R Recommendation BT.500-10. Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union, Geneva, Switzerland, 2000.
- [5] 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.
- [6] ITU-T Recommendation P.910. Subjective video quality assessment methods for multimedia applications. International Telecommunication Union, Geneva, Switzerland, 1996.
- [7] M. Kutter. *Digital Image Watermarking: Hiding Information in Images*. PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 1999.
- [8] J. Lubin and D. Fibush. Sarnoff JND vision model. T1A1.5 Working Group Document #97-612, ANSI T1 Standards Committee, 1997.
- [9] S. Pereira, S. Voloshynovskiy, M. Madueño, S. Marchand-Maillet, and T. Pun. Second generation benchmarking and application oriented evaluation. In *Proc. Information Hiding Workshop*, Pittsburgh, PA, April 2001.
- [10] F. A. P. Petitcolas and R. J. Anderson. Evaluation of copyright marking systems. In *Proc. IEEE Multimedia Systems (ICMCS)*, Florence, Italy, June 7–11 1999.
- [11] K. T. Tan, M. Ghanbari, and D. E. Pearson. An objective measurement tool for MPEG video quality. *Signal Processing*, 70(3):279–294, Nov. 1998.
- [12] C. J. van den Branden Lambrecht. *Perceptual Models and Architectures for Video Coding Applications*. PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 1996.
- [13] VQEG. Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, 2000. available at <http://www.vqeg.org/>.
- [14] H.-J. Wang, P.-C. Su, and C.-C. J. Kuo. Wavelet-based digital image watermarking. *Opt. Express*, 3(12):491–498, Dec. 1998.
- [15] A. B. Watson, J. Hu, J. F. McGowan III, and J. B. Mulligan. Design and performance of a digital video quality metric. In *Proc. SPIE*, volume 3644, pages 168–174, San Jose, CA, Jan. 23–29, 1999.
- [16] S. Winkler. Issues in vision modeling for perceptual video quality assessment. *Signal Processing*, 78(2):231–252, Oct. 1999.
- [17] S. Winkler. A perceptual distortion metric for digital color video. In *Proc. SPIE*, volume 3644, pages 175–184, San Jose, CA, Jan. 23–29, 1999.
- [18] S. Winkler. *Vision Models and Quality Metrics for Image Processing Applications*. PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 2000.
- [19] X.-G. Xia, C. G. Boncelet, and G. R. Arce. Wavelet transform based watermark for digital images. *Opt. Express*, 3(12):497–512, Dec. 1998.



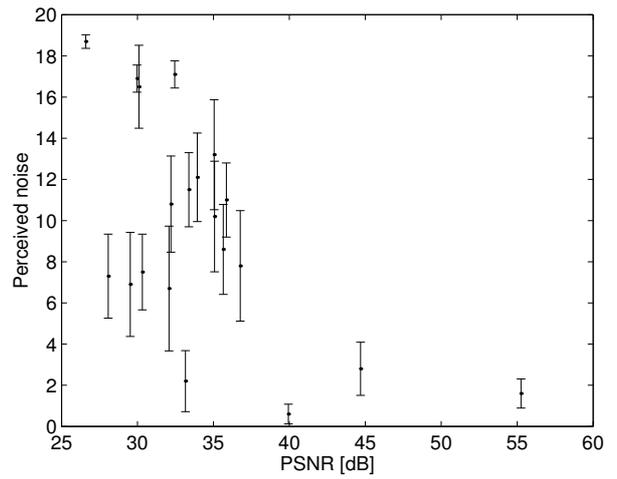
(a) Subjective flicker ratings vs. Flicker metric.



(a) Subjective noise ratings vs. HF-Noise metric.



(b) Subjective flicker ratings vs. PSNR.



(b) Subjective noise ratings vs. PSNR.

Flicker	Metric	PSNR
Pearson	0.95	-0.54
Spearman	0.94	-0.58

(c) Correlations.

Noise	Metric	PSNR
Pearson	0.81	-0.60
Spearman	0.81	-0.41

(c) Correlations.

**Figure 2:** Perceived flicker vs. Flicker metric and PSNR (subjective data are shown with 90%-confidence intervals).

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