Detectability and Annoyance of Artifacts in Watermarked Digital Videos *

Mylène C.Q. Farias[1], *Marco Carli*[3], *John M. Foley*[2], *Sanjit K.Mitra*[1]
[1]Department of Electrical and Computer Engineering, [2]Department of Psychology,
University of California, Santa Barbara, California 93106 USA
[3]Department of Electrical Engineering, University of Roma TRE
Via della Vasca Navale, 84 00146 Roma ITALY
e-mail: (mylene,mitra)@ece.ucsb.edu, carli@ele.uniroma3.it, foley@psych.ucsb.edu

ABSTRACT

This paper presents an evaluation of the annoyance and visibility of the artifacts generated by embedding a watermark into a video. To measure the detection threshold and mean annoyance values, a psychophysical experiment is carried out. The results show that the choice of the image to be embedded into the video does not affect the visibility and annoyance of the artifacts significantly. The mean annoyance curve can vary considerably depending on the physical characteristics of the particular video.

1 INTRODUCTION

In the past few years a rapid diffusion of the Internet along with a proliferation of digital multimedia data has been witnessed. The use of digital techniques make copying, modifying and illegally delivering these data easier. Therefore, it has become important to develop effective methods to identify the owner of digital data and discourage 'piracy'. A common way to do this is to embed some identification data in the file (*watermark embedding*). Recently, embedding techniques have also been used for different purposes: finger printing, multimedia indexing, context base retrieval, channel quality assessment, etc. [2].

Every watermarking system has to satisfy three main constraints:

- invisibility: the watermark should not affect the perceptual quality of the video and should not produce noticeable distortions into the data,
- robustness to image alterations: the watermark cannot be altered by malicious (an attempt to alter the mark) or unintentional (compression, transmission or filtering) operations,
- security: the watermark may not be removed from the video, even if the embedding scheme is known.

The objective of a watermarking system is drastically reduced if the mark is visible when the video is displayed on a computer or on a TV screen. A good watermark is invisible and preserves the quality and content integrity of the video.

In this paper, the first constraint has been analyzed by evaluating the visibility and annoyance of artifacts introduced by watermarking embedding techniques. A psychophysical experiment has been carried out to measure the detection threshold and annoyance values of these artifacts.

The remainder of the paper is organized as follows. Section 2 introduces the watermark embedding procedure. Section 3 describes the psychophysical experiment method. Section 4 discusses the experiment results. Conclusions are presented in Section 5.

2 THE WATERMARKING EMBEDDING PROCEDURE

Several watermarking methods have been proposed in the literature; a first classification of these methods can be done according to the particular domain in which the embedding process is performed. The watermark insertion can be done in the spatial domain [1] or in an *ad hoc* transform domain such as the DCT domain [3], the Fourier domain [4], or the Wavelet domain [6], [8].

In our experiment, a two-dimensional watermark is embedded in the DCT domain. A preprocessing is performed on the mark before the actual embedding into the DCT mid-frequency samples. A spread-spectrum technique is simulated to hide the watermark, by using a set of uncorrelated pseudo-random noise (PN) matrices (one per each frame) which are later multiplied by the reference watermark (the same for the whole video). The result is that the spatial localization of the mark is different frame by frame, avoiding temporal summation.

A high bit rate multimedia communication system has been considered: the video frame size is 720×486, in

Figure 1: *Random* (left), and *Logo* (right) images used as watermarks.

*This work is supported in part by a National Foundation grant No. CCR-0105404 and in part by an Italian National Research Council grant.

YCrCb format, 4:2:2 digital component video. The watermark image is an 88×88 size image. Two kinds of watermarks have been tested: the *Random* image and the *Logo* image shown in Figure 1.

In order to reduce the notational complexity, the lower case bold letters are used to represent 2-tuples of the form $\mathbf{n} \doteq (n_1, n_2)$, where n_1 and n_2 are integers. The embedding procedure can be summarized as follows. First, the watermark image $w(\mathbf{n})$ is normalized, and binarized:

$$\tilde{w}(\mathbf{n}) = sign\left(w(\mathbf{n}) - E\{w(\mathbf{n})\}\right),$$

Second, a pseudo random algorithm is used to generate the pseudo-noise image $p(\mathbf{n})$, which has values in the range [-1,1] and is of zero mean. One pseudo-noise image is created for each frame of the video.

For a generic frame of a video sequence $f_i(\mathbf{n})$, the final watermark w_i is obtained by multiplying the watermark by $p_i(\mathbf{n})$:

$$w_i(\mathbf{n}) = \tilde{w}(\mathbf{n})\, p_i(\mathbf{n})$$

The *log* of the luminance is taken and then the DCT coefficients are computed. The mark is added only to the mid-frequencies DCT coefficients. In our experimental trials, the mark is embedded starting from the (215, 390) DCT coefficient. The range of frequencies where the watermark is inserted is strongly dependent on the application. For the purpose of delivering a high quality video through an ideal channel, the mid-frequencies are a good choice. Inserting the mark in the low-frequencies would cause visible artifacts in the image, while inserting it in the high frequencies would make it easier to remove it.

The final mark is multiplied by a scaling factor α and is added to the DCT coefficients:

$$Y_i(\mathbf{n}) = \mathrm{DCT}\left(log\left(f_i(\mathbf{n})\right)\right) + \alpha \cdot w_i(\mathbf{n})$$

This scaling factor α can be used to vary the strength of the watermark. In various applications and video formats different values for α are desirable. By increasing α, we also decrease the quality of the video.

3 THE PSYCHOPHYSICAL EXPERIMENT

We used twenty test subjects drawn from a pool of students in the introductory psychology class at UCSB. The students are thought to be relatively naive concerning video artifacts and the associated terminology. They are asked to wear any vision correcting devices (glasses or contacts) that they normally wear to watch television. A Sony PVM-1343 monitor is used to display the test video sequences. The experiment is run with one subject at a time. Each subject is seated straight ahead of the monitor, located at or slightly below eye height for most subjects. The subjects are positioned at a distance of four screen heights (80 cm) from the video monitor.

The experimental session consists of four stages. In the first stage, the subject is verbally given instructions. In the second stage, training sequences are shown to the subject. The training sequences represent the impairment extremes for the experiment and are used to establish the annoyance value range. In the third stage, the test subjects runs through several practice trials. The practice trials are identical to the experimental trials and are used to familiarize the test subject with the experiment.

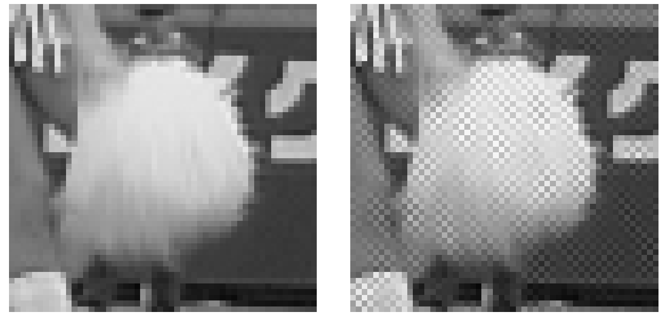

Figure 2: Zoomed version of the 10th frame of Cheerleaders: original (left) and embedded with watermark *Logo* and $\alpha = 0.6$ (right).

Finally, the actual experiment is performed with the complete set of test sequences. After each video is displayed, the subject is asked to enter if he/she saw any defect. If the answer is no, no further questions are asked and the next video is shown. If the answer is yes he/she is asked to enter a value between 0 and 100, representing how annoying the defect is, compared to the worst defect present in the training sequences.

In this experiment, the goal has been to measure the detection threshold and annoyance values for the artifacts introduced by the watermark embedding procedure. To generate the test video sequences, we start with a set of five original video sequences of assumed high quality: Bus, Cheerleader, Flower-garden, Football and Hockey. The video clips are all 5 seconds long and contain scenes that are typical of normal television. Each original is then embedded with the watermark.

To find artifact detection and annoyance values, the contrast of the error patterns must range from nearly imperceptible to highly annoying. This is obtained by varying the scaling factor used in the watermark embedding algorithm. The set of scaling factors used is 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6. Figure 2 displays a detail of 10th frame of the video Cheerleaders with and without watermark. The left picture of Figure 2 corresponds to the original video, while the right picture corresponds to the same video embedded with the watermark *logo* and $\alpha = 0.6$.

The total number of test sequences used in this experiment is 65, which includes 60 test sequences (5 originals times 6 strength factors times 2 watermark images) plus the 5 original sequences. The sequences are shown in a random order during the main experiment. The total squared error (TSE) is used as our objective error measure:

$$\mathrm{TSE} = \Sigma_{\mathbf{n}} \Sigma_i (y_i(\mathbf{n}) - f_i(\mathbf{n}))^2,$$

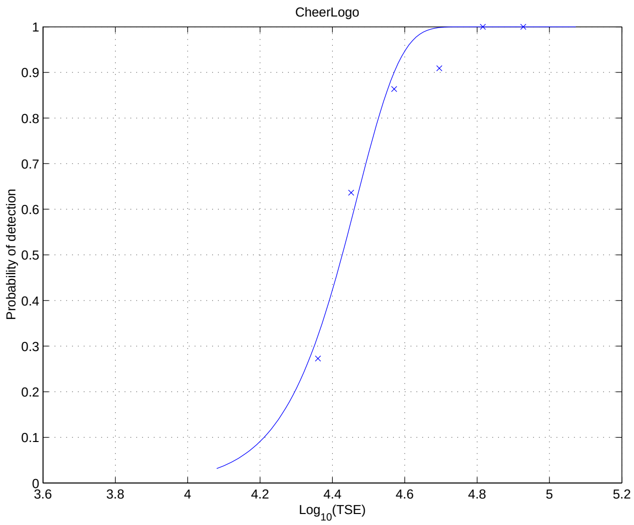

Figure 3: Probability of detection curve for embedding the watermark *Logo* into the video cheerleaders.

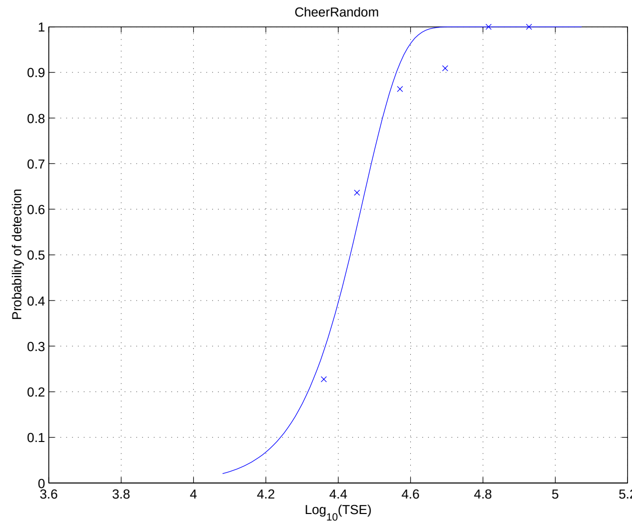

Figure 4: Probability of detection curve for embedding the watermark *Random* into the video cheerleaders.

where y_i is the i-th frame of the watermarked video and f_i is the i-th frame of the original video.

4 EXPERIMENTAL RESULTS

To measure the detection threshold for the artifacts, the probability of detection for each artifact as a function of the error energy is estimated. The threshold is defined as the error energy such that the artifact is seen by 50% of the subjects. The probability of detecting each artifact is estimated by counting the number of people who detected the artifact and dividing by the number of observations. The logarithm of the error energy (TSE) is used for each artifact as the independent variable. The probability as a function of the logarithmic error energy is then fitted using the Weibull function [10], which has an S-shape similar to the experimental data and is defined as

$$P(x) = 1 - 2^{(Sx)^k},$$

where $P(x)$ is the probability of detection, x is the logarithmic error energy, $1/S$ is the 50% detection threshold in logarithmic error energy, and k is a constant that determines the steepness of the function.

Figures 3 and 4 depict the probability of detection curves for the artifacts caused by embedding the watermark image *Logo* and *Random*, respectively, into the video Cheerleaders. The two curves are very similar implying that the choice of a different image as a watermark does not have a significant effect on the visibility of the artifacts.

Table 1: Threshold error energy and detection threshold curve fit parameters for the artifacts.

Test Sequence	Detection Threshold		Curve Fit Parameters	
	TSE	log_{10}(TSE)	S	k
Bus-Logo	15488	4.19	0.2407	27.71
Bus-Random	16218	4.21	0.2378	31.17
Cheer-Logo	27542	4.44	0.2259	37.66
Cheer-Random	28184	4.45	0.2256	42.38
Flower-Random	39811	4.60	0.2179	87.50
Foot-Logo	26303	4.42	0.2270	25.68
Foot-Random	26303	4.42	0.2273	27.38

Table 1 summarizes the 50% detection threshold found in terms of error energy (TSE) and logarithmic error energy (log_{10}(TSE)). Overall, the threshold values do not change considerably over the sequences and remain almost constant when only the embedded image is changed. Table 1 also includes the curve fit parameters found for each sequence and image tested. The parameter S (sensitivity) corresponds to the inverse of the log-threshold and therefore has the same behavior.

The parameter k, which represents the steepness of the probability of detection curve, varies between different videos and may be due to variation in content of the videos. The same artifact at the same strength will vary both in visibility and annoyance depending on where and when it appears, and also on the texture and luminance characteristics of the background of the video. In particular, it is noticed that white and/or smooth backgrounds facilitate the visibility of the artifacts. For the video Hockey, which has very large white smooth areas, it has not been possible to estimate the detection threshold. More than half of all subjects tested have seen the weakest artifact. This particular video has probably a lower threshold than the rest of the videos. It has not also been possible to calculate the threshold for the video Flower embedded with the watermark *Logo* due to artifacts (blurring) previously present in this video.

Standard methods [10] are used to analyze the data provided by the test subjects and to compute the mean annoyance values. The mean annoyance values for each test sequence is fitted with the standard logistic function[10]:

$$y = y_{min} + \frac{(y_{max} - y_{min})}{1 + exp\left(-\frac{(x-\overline{x})}{\beta}\right)}$$

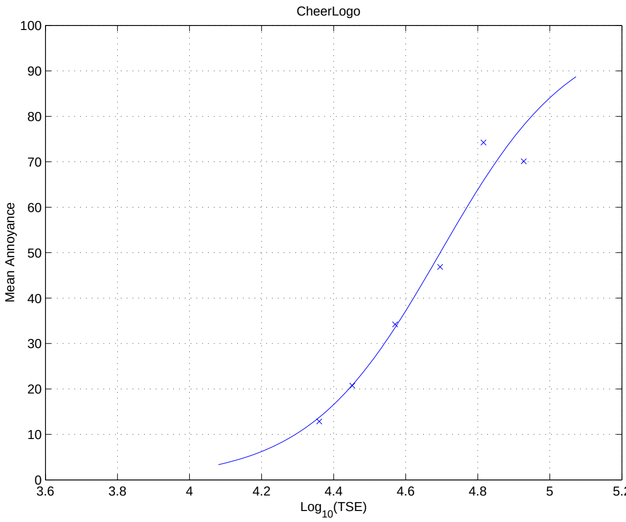

Figure 5: Mean annoyance curve for embedding the watermark *Logo* into video cheerleaders.

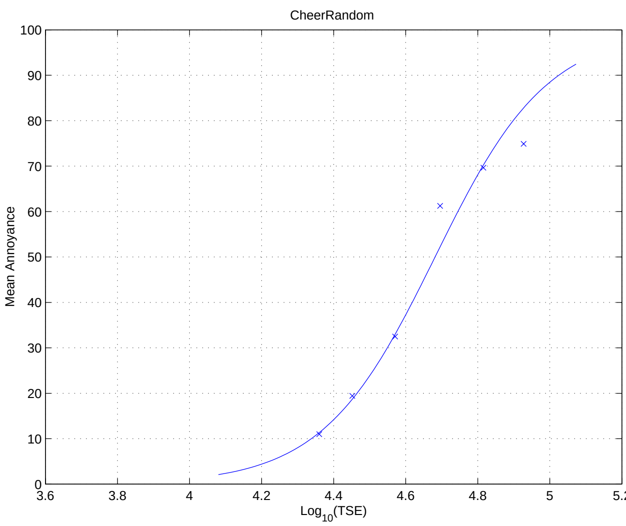

Figure 6: Mean annoyance curve for embedding the watermark *Random* into video cheerleaders.

where y is the predicted annoyance and x is the logarithmic error energy. The parameters y_{max} and y_{min} establish the limits of the annoyance value range. The parameter $\overline{x}$ translates the curve in the x-direction and the parameter β controls the steepness of the curve.

Figures 5 and 6 depict the mean annoyance curves versus the logarithmic error energy for the artifacts caused by embedding the watermark image *Logo* and *Random*, respectively, into the video Cheerleaders. Again, the two curves are very similar. The steepness of the annoyance curve β does not vary significantly for different watermarks, but it does vary between test sequences. The same is true for the parameter $\overline{x}$.

5 CONCLUSION

In this paper an evaluation of the artifacts caused by embedding a watermark into a video has been performed. A psychophysical experiment has been carried out to measure the detectability and annoyance values of these artifacts. The results show that the choice of the image to be embedded into the video does not affect the visibility and annoyance of the artifacts significantly. The mean annoyance curve can vary considerably depending on the physical characteristics of the particular video. The detection thresholds found are indicators of the strength of the watermark which is seen by 50% of the observers. In our experiment, the thresholds found correspond to scaling factors α no greater than 0.2.

References

[1] N. Nikolaidis, I. Pitas, "Robust image watermarking in the spatial domain," *Signal Processing*, vol. 66, no.3, pp. 385-403, May 1998.

[2] Hanjalic, G. C. Langelaar, P.M.B. van Roosmalen, J. Biemond, R.L. Lagendijk, *Image and video databases: restoration, watermarking and retrieval*, New York: Elsevier Science B. V., 2000.

[3] M. Barni, F. Bartolini, V. Cappellini, A. Piva, "A DCT-domain system for robust image watermarking," *Signal Processing*, vol. 66, no. 3, pp. 357-372, May 1998.

[4] I. Cox, J. Kilian, F. Leighton, T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, December 1997.

[5] M. D. Swanson, B. Zhu, A. H. Tewfik, "Transparent robust image watermarking transform," *Proc. IEEE Int. Conf. Image Proc. 1996*, Lausanne, Switzerland, pp. 211-214, 16-19 September, 1996.

[6] H. Inoue, A.Miyazaki, T. Katsura, "An image watermarking method based on the wavelet transform," *Proc. IEEE Int. Conf. Image Proc. 1999*, pp. 296-300, Kobe, Japan, 25-28 October, 1999.

[7] H. J. M. Wang, P. C. Su, C. C. J. Kuo, "Wavelet-based blind watermark retrieval technique," *Proc. SPIE*, vol 3528, Conf. on Multimedia Systems and Applications, Boston, MA, November 1998.

[8] P. Campisi, A. Neri, M. Visconti, "A wavelet based method for high frequency subbands watermark embedding," SPIE Multimedia Systems and Applications III, Boston (MA), November 2000.

[9] M.Yuen, H.R.Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70 (1998), pp. 247-278.

[10] ITU Recommendation BT.500-8, "Methodology for subjective assessment of the quality of television pictures," 1998.