Proceedings of the IEEE Conference on Multimedia Computing and Systems, October 1997.

An Objective Approach to Assessing Relative Perceptual Quality of MPEG-Encoded Video Sequences[1]*

Lee Wang[†], Ranga S. Ramanujan[‡], James A. Newhouse[‡], Maher Kaddoura[‡], Atiq Ahamad[‡], Kenneth J. Thurber[‡], and Howard Jay Siegel[†]

[†]School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47907-1285 USA
{lwang, hj}@ecn.purdue.edu

[‡]Architecture Technology Corporation
Minneapolis, MN 55424 USA
{rsr, jnewhouse, mkaddoura, aa, kthurber}@atcorp.com

Abstract

A novel objective approach to assessing the relative perceptual quality of MPEG-encoded video sequences is presented. Quality comparison of the original and filtered MPEG-encoded video sequences is performed directly in the frequency domain. Experimental results show a very good correlation between the objective quality measurement and the human subjective perception. This objective approach provides a sensible and meaningful automated quality measure for MPEG compression. The approach should become a viable tool to aid MPEG encoding optimization.

1. Introduction

The goal of this work is to derive an objective measure that reflects the amount of degradation perceived by the human visual system for compressed digital motion video sequences [5]. Such a computable objective measure can be used to help develop and evaluate lossy techniques that further compress an MPEG-encoded [2] video sequence to reduce the bitrate with minimal viewing quality degradation.

2. Objective quality assessment

Experiments contained two steps: calibration and validation. In the calibration step, a MPEG-1-encoded video clip was further compressed to various degrees. Human subjects viewed these further compressed clips and gave subjective quality numbers. These numbers were used to determine the exponents and other parameters in the objective measurement equations below.

In the validation step, a set of MPEG-encoded video sequences (excluding the one used above) were further compressed to various degrees using multiple lossy techniques together. Then the approach discussed below computed the objective quality using the parameters determined in the calibration step. These numbers were compared with human subjective quality numbers for validation.

As an example of the validation step, a total of 81 further compressed clips were generated from an MPEG-1-encoded 384-frame video sequence named Zoom using two lossy techniques together. The largest difference between the objective and the subjective quality numbers was less than 5.5%. Tests on other video sequences gave similar results in the validation step.

The proposed quantitative quality assessment approach contains two phases: masking and pooling. The masking phase incorporates two techniques: luminance masking and contrast masking [1]. In the following formulas, the indices i, k, and f represent the i-th frequency component, the k-th block, and the f-th frame. The index z stands for a color component. Let the measured detection threshold under normal conditions for the i-th frequency of color z be t_{iz} [4]. Define C_{0kfz} to be the DC coefficient of the k-th block in the f-th frame for color component z and define C_{0fz} to be the average DC of the f-th frame for color component z. Luminance masking can be formulated by a power function [6]. The adjusted detection threshold after luminance masking becomes

$$t_{ikfz} = t_{iz}\,(C_{0kfz}/C_{0fz})^{\alpha} \qquad (1)$$

The next step is contrast masking. A model from [3] is used below

$$m_{ikfz} = t_{ikfz} \times \max[1, (|C_{ikfz}| / t_{ikfz})^{\omega iz}] \qquad (2)$$

Each m_{ikfz} is a just-noticeable-difference (JND). The ω_{iz} values vary for different frequencies, which are determined by experimentation in this research.

Let $Cref_{ikfz}$ and C_{ikfz} be the i-th DCT coefficient of k-th block in the f-th frame for color z from the

1 This work was supported by DARPA under contract number DAAH01-96-C-048. Wang and Siegel are ATC consultants.

original MPEG-encoded and a further compressed video sequences, respectively. The raw error, e_{ikfz} , is defined as

$$e_{ikfz} = | C_{ikfz} - Cref_{ikfz} | \quad (3)$$

which is content-based, i.e., it is the actual absolute difference of the comparing DCT coefficients. The perceptual error, d_{ikfz}, is then

$$d_{ikfz} = e_{ikfz} / m_{ikfz} \quad (4)$$

The pooling phase starts with spatial pooling and frequency pooling [6]. This approach then adds four extra pooling steps for the motion video objective quality assessment. For each frame, spatial pooling combines the perceptual errors for each frequency together across all blocks. Let K_{fz} be the number of blocks in the f-th frame for color z. The pooled error p_{ifz} is given by

$$p_{ifz} = [(1/ K_{fz}) \Sigma_k \mid d_{ikfz} \mid^{\beta z}]^{(1/\beta z)} \quad 5)$$

Let I_{fz} denote the number of frequencies in each block of the f-th frame and color z. Frequency pooling combines these frequency errors together into a single number for each frame and color

$$P_{fz} = [(1/ I_{fz}) \Sigma_i (p_{ifz})^{\gamma z}]^{(1/ \gamma z)} \quad 6)$$

Next, the errors for all color components are pooled to form a perceptual frame error P_f in component pooling, where the ρ_z's and σ are determined experimentally and $\Sigma_z \rho_z = 1$

$$P_f = [\Sigma_z \rho_z (P_{fz})^{\sigma}]^{(1/\sigma)} \quad (7)$$

Temporal pooling combines the perceptual frame errors to form an overall frame error P. Let F denote the number of frames. Temporal pooling is given by

$$P = [(1/F)\Sigma_f (P_f)^{\lambda}]^{(1/\lambda)} \quad (8)$$

The value of λ is obtained experimentally.

Next step is temporal derivative pooling, which is to capture the perceptual frame error variance. Let

$$P'_f = P_{f+1} - P_f \quad (9)$$

be the temporal perceptual frame error derivative. Temporal derivative pooling combines these derivatives to form an overall error derivative P'

$$P' = [(1/ (F-1)) \Sigma_f | P'_f |^{\tau}]^{(1/ \tau)} \quad (10)$$

The value of τ is obtained by experimentation.

The final step is grand pooling, which combines the overall frame error and the overall error derivative to give a total perceptual error E for the entire further compressed video sequence relative to the original sequence. In the following formula, μ_1, μ_2, and ν are determined by experimentation, and $\mu_1 + \mu_2 = 1$.

$$E = [\mu_1 P^{\nu} + \mu_2 (P')^{\nu}]^{(1/\nu)} \quad (11)$$

The relative perceptual quality Q_{rel} is obtained by a formula shown below, where E_w is the perceptual error of the video sequence of the worst perceptual quality (i.e., greatest E).

$$Q_{rel} = 1 - (E / E_w) \quad (12)$$

3. Conclusions

A novel objective approach to assessing the relative perceptual quality of MPEG-encoded video sequences is presented in this paper. Quality comparison of original and further-compressed MPEG-encoded video sequences is performed directly in the frequency domain. Experimental results show a very good correlation between this objective quality measure and subjective perception. This objective approach provides a sensible and meaningful automated quality measure for MPEG compression, which could be a viable tool to aid in MPEG encoding optimization.

References

[1] A. J. Ahumada Jr. and H. A. Peterson, "Luminance model-based DCT quantization for color image compression," *Proc. SPIE, Human Vision, Visual Processing, and Digital Display III,* B. E. Rogowitz, Ed., Vol. 1666, Aug. 1992, pp. 365-374.

[2] ISO/IEC, *MPEG-1 Standard,* ISO/IEC, Geneva, Switzerland, 1993.

[3] G. E. Legge, "A power law for contrast discrimination," *Vision Research,* Vol. 21, No. 3, Mar. 1981, pp. 457-467.

[4] B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard,* Van Nostrand Reinhold, New York, NY, 1992.

[5] L. Wang, R. S. Ramanujan, J. A. Newhouse, M. Kaddoura, A. Ahamad, K. J. Thurber, and H. J. Siegel, *An Objective Approach to Assessing Relative Perceptual Quality of MPEG-encoded Sequences,* Technical Report ATC-RD-97-07, Architecture Technology Corporation, Mar. 1997.

[6] A. B. Watson, "DCT quantization matrices visually optimized for individual images," *Proc. SPIE, Human Vision, Visual Processing, and Digital Display IV,* J. P. Allebach and B. E. Rogowitz, Eds., Vol. 1913, Sep. 1993, pp. 202-216.