BLIND PSNR ESTIMATION OF VIDEO SEQUENCES USING QUANTIZED DCT COEFFICIENT DATA

*Tomás Brandão**

IT-Lisbon/ISCTE

Maria Paula Queluz

IT-Lisbon/IST-UTL

ABSTRACT

This paper proposes a no-reference PSNR estimation method for video sequences subject to lossy DCT-based encoding, such as MPEG-2 encoding. The proposed method is based on DCT coefficient statistics, which are modeled by Laplace probability density functions, with parameter λ. The distribution's parameter is computed from the received quantized data, by combining maximum-likelihood with linear prediction estimates. The resulting coefficient distributions are then used for estimating the local error due to lossy encoding. Since no knowledge about the original (reference) sequences is required, the proposed method can be used as a no-reference metric for evaluating the quality of the encoded video sequences.

Index Terms— Image quality, no-reference metric, parameter estimation

1. INTRODUCTION

In the past few years, quality monitoring of multimedia data has become an important matter, especially due to the increasing transmission of digital video contents over broadband and wireless networks. From a quality of service perspective, it would be desirable to evaluate the quality of the received contents at the user's end. This kind of system would have to deal with different distortion sources, such as lossy encoding of media data and transmission errors. Moreover, and since the original signals are not available at the receiver, quality scores must be provided with few knowledge about the original - *reduced reference* metrics - or no knowledge at all - *no-reference* metrics.

This paper suggests a new technique that estimates errors due to lossy compression in block-based DCT (discrete cosine transform) video encoding schemes, the most used in today's panorama. It is assumed that the statistics of the original DCT coefficient data are well modeled by a *Laplace* distribution, with parameter λ. This parameter is estimated from the quantized values, available at the receiver, combining maximum-likelihood (ML) estimation method with a linear prediction scheme, as suggested in [1] for still images. The method proposed here can be seen as a generalization of [1] for DCT coefficients subject to linear quantization, with quantization scales that may differ from block to block. This feature allows to apply the algorithm to video sequences, encoded using current DCT-based standards. The final result is a PSNR estimate that is computed without the need of the original data, thus resembling a no-reference quality metric.

The performance of the proposed algorithm has been evaluated using several video sequences, subject to different coding rates. The resulting PSNR estimates have shown greater accuracy than the ones provided by a state-of-the-art method [2], proposed with the same objective.

The paper is organized as follows: after the introduction, section 2 depicts the framework for the ML estimation; the use of prediction is synthesized in section 3; section 4 shows how to compute a blind PSNR score from the estimated DCT coefficient distribution; results are depicted in section 5, and finally, some remarks and topics for future investigation are given in section 6.

2. ML PARAMETER ESTIMATION

Block-based DCT coefficient data distribution of natural images can be modeled by a zero-mean *Laplace* probability density function (*pdf*) [3](excluding the DC coefficients). More accurate models can be found in literature, such as generalized gaussian [4] or gaussian mixtures [5]. The choice of the laplacian model represents a reasonable tradeoff between accuracy and simplicity.

Using 8×8 blocks, with horizontal and vertical frequency pairs $(i,j) \in \{0,...,7\} \times \{0,...,7\}$ and $(i,j) \neq (0,0)$, the coefficient's distribution is thus described by:

$$f_X(x) = \frac{\lambda(i,j)}{2} \exp\left(-\lambda(i,j)|x|\right), \qquad (1)$$

where $\lambda(i,j)$ is the distribution's parameter for frequency pair (i,j) and x is the coefficient value. For notation simplicity, the indexes (i,j) will be dropped along the text.

2.1. λ estimation using the original coefficient data

An estimate for λ, using the original coefficient data, can be obtained using the maximum-likelihood (ML) method. Representing by x_k the k-th original coefficient value at a given

*Corresponding author: tomas.brandao@lx.it.pt

frequency, an ML estimate for λ is given by:

$$\lambda_{ML} = \arg\max_{\lambda} \left\{ \log \prod_{k=1}^{N} f_X(x_k) \right\} = \frac{N}{\sum_{k=1}^{N} |x_k|}, \quad (2)$$

where N represents the number of DCT coefficients at the given frequency (which is the same as the number of blocks). For the remainder of this paper, λ_{ML} will often be referred as *original* λ.

2.2. λ estimation using quantized coefficient data

Admitting that only quantized data is available for estimating the original coefficient distribution, the ML method can also be used for estimating the value of λ:

$$\hat{\lambda}_{ML} = \arg\max_{\lambda} \left\{ \log \prod_{k=1}^{N} P(X_k) \right\}, \quad (3)$$

where $P(X_k)$ represents the probability of having value X_k at the quantizer's output. Assuming that the quantizer is linear with quantization step q_k, which may differ from block to block, $P(X_k)$ can be written as:

$$\begin{aligned} P(X_k) &= \int_{X_k - \frac{q_k}{2}}^{X_k + \frac{q_k}{2}} \frac{\lambda}{2} e^{-\lambda|x|} dx \\ &= \begin{cases} 1 - e^{-\frac{\lambda q_k}{2}}, & \text{if } X_k = 0; \\ \frac{1}{2} e^{-\lambda|X_k| + \frac{\lambda q_k}{2}} (1 - e^{-\lambda q_k}), & \text{otherwise.} \end{cases} \end{aligned} \quad (4)$$

Substituting (4) in (3) leads to:

$$\begin{aligned} \hat{\lambda}_{ML} = \arg\max_{\lambda} \{ & \sum_{\substack{k=1, \\ X_k=0}}^{N} \log(1 - e^{-\frac{\lambda q_k}{2}}) + \\ & \sum_{\substack{k=1, \\ X_k \neq 0}}^{N} \log \frac{1}{2} (e^{-\lambda|X_k| + \frac{\lambda q_k}{2}})(1 - e^{-\lambda q_k}) \}. \end{aligned} \quad (5)$$

Differentiating with respect to λ and after simple algebraic manipulations, we get:

$$\sum_{\substack{k=1, \\ X_k=0}}^{N} \frac{\frac{q_k}{2} e^{-\frac{\lambda q_k}{2}}}{1 - e^{-\frac{\lambda q_k}{2}}} + \sum_{\substack{k=1, \\ X_k \neq 0}}^{N} \left(\frac{q_k}{2} - |X_k| + \frac{q_k e^{-\lambda q_k}}{1 - e^{-\lambda q_k}} \right) = 0, \quad (6)$$

whose solution can be found by using an iterative root finding algorithm. In this work, a simple implementation of *Newton-Raphson*'s method has been used, taking a small value (i.e. 0.01) as the initial solution guess. Convergence to the correct solution has been achieved in all experiments.

For the particular case where the encoder uses the same quantizer scales for all image blocks, i.e. $\forall_k : q_k = \Delta$, such as in JPEG encoding, (6) can be written as:

$$e^{-\lambda\Delta}(N\Delta + 2S_1) + e^{-\frac{\lambda\Delta}{2}} N_0\Delta + N_1\Delta - 2S_1 = 0, \quad (7)$$

where N_0 and N_1 are the number of coefficients quantized to zero and non-zero values, respectively, and S_1 is the sum of the absolute values of non-zero coefficients. The solution of (7) is given by:

$$\hat{\lambda}_{ML} = -\frac{2}{\Delta} \log \frac{-N_0\Delta + \sqrt{N_0^2\Delta^2 - 4(N\Delta + 2S_1)(N_1\Delta - 2S_1)}}{2N\Delta + 4S_1}, \quad (8)$$

a result already derived in [6].

Let us now take a deeper insight to see what happens to $\hat{\lambda}_{ML}$ as the quantization steps during encoding increase. In lossy DCT-based encoding, important compression gains result from increasing the number of coefficients that are quantized to zero. As the quantization steps increase, more coefficients will fall on this situation. It is common, even at average compression rates, to have all DCT coefficients quantized to 0 at high frequency positions. For such cases, (6) can be approximated by:

$$\sum_{\substack{k=1, \\ X_k=0}}^{N} \frac{\frac{q_k}{2} e^{-\frac{\lambda q_k}{2}}}{1 - e^{-\frac{\lambda q_k}{2}}} = 0. \quad (9)$$

It can be easily concluded that the solution for this equation is $\hat{\lambda}_{ML} = \infty$, which means that the estimated distribution will be a *Dirac's delta* function. Similarly, for the constant quantization scales (eq. (8)), as $N_0 \to N$, then $N_1 \to 0$, $S_1 \to 0$ and, as a consequence, $\hat{\lambda}_{ML} \to \infty$.

Aware of these problems, the authors of [2] propose to model the original DCT coefficients distribution as a mixture of two Laplace *pdf*s: one is estimated based on all quantized coefficient values, while the other is estimated considering non-zero quantized values only. This strategy is more robust when a high percentage of the coefficients is quantized to 0, but still fails if all DCT coefficients at a given frequency are quantized to zero.

3. λ ESTIMATION BY ADDING PREDICTION

The issue described at the end of the previous section can be tackled by exploring the correlation between λ values at neighbouring DCT frequencies, as suggested in [1]. Using matrix notation, a prediction value for λ at a given frequency can be written as:

$$\hat{\lambda}_p = \boldsymbol{\lambda}_{\boldsymbol{\upsilon}}^{\mathbf{T}} \boldsymbol{\beta}, \quad (10)$$

with

$$\boldsymbol{\lambda}_{\boldsymbol{\upsilon}} = \begin{bmatrix} 1 \\ \lambda_{\upsilon_1} \\ \vdots \\ \lambda_{\upsilon_K} \end{bmatrix} \text{ and } \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix},$$

where K is the neighbourhood size, $\boldsymbol{\lambda}_{\boldsymbol{\upsilon}}$ is a vector consisting of the neighbourhood values and $\boldsymbol{\beta}$ is the weight vector.

The weight vector $\boldsymbol{\beta}$ has been computed by minimizing the square error between the original λ and $\hat{\lambda}_p$, in a randomly

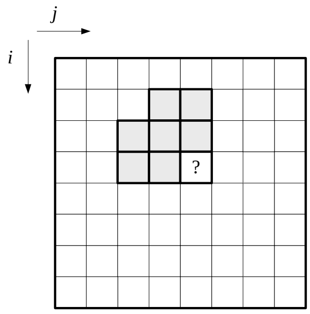


Fig. 1. Neighbourhood configuration.

chosen image set - 15 still images taken from LIVE image database [7]. By following the square error minimization criterion, $\boldsymbol{\beta}$ is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{\Lambda^T \Lambda})^{-1}\mathbf{\Lambda^T}\boldsymbol{\lambda}, \tag{11}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & \lambda_{\upsilon_1}^{(1)} & \dots & \lambda_{\upsilon_K}^{(1)} \\ 1 & \lambda_{\upsilon_1}^{(2)} & \dots & \lambda_{\upsilon_K}^{(2)} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_{\upsilon_1}^{(L)} & \dots & \lambda_{\upsilon_K}^{(L)} \end{bmatrix}, \boldsymbol{\lambda} = \begin{bmatrix} \lambda^{(1)} \\ \lambda^{(2)} \\ \vdots \\ \lambda^{(L)} \end{bmatrix}.$$

$\mathbf{\Lambda}$ is a $L \times (K+1)$ matrix, where L is the number of images in the training set. Each element $\lambda_{\upsilon_k}^{(l)}$ is the k-th neighbour in the l-th image. As for the vector $\boldsymbol{\lambda}$, it consists of the original values of λ, per image, at the position to predict. The training procedure is carried out for each frequency position to predict.

The neighbourhood configuration used in the experiments is illustrated in figure 1. Since low-frequency DCT coefficients are less vulnerable to the effects of lossy compression, its structure has been chosen with the purpose of recursively predict values for λ, starting from those frequency positions.

The prediction $\hat{\lambda}_p$ that results from (10) can then be combined with $\hat{\lambda}_{ML}$ in order to improve the estimation accuracy for the original DCT distribution parameter. Since ML estimates become more inaccurate as the rate of coefficients quantized to zero increases, more trust should be given to the predictor in this case. On the other hand, if the number of coefficients quantized to zero is low, the ML estimator will most likely get accurate results, so there is no real need for the predicted value. Based on these premises, a simple criterion for combining $\hat{\lambda}_p$ with $\hat{\lambda}_{ML}$ is to weight them proportionally to the rate of DCT coefficients quantized to zero:

$$\hat{\lambda}_f = r_0\hat{\lambda}_p + (1 - r_0)\hat{\lambda}_{ML}, \tag{12}$$

where $r_0 = N_0/N$ represents the rate of coefficients quantized to zero and $\hat{\lambda}_f$ is the final estimation for the distribution's parameter.

To estimate $\hat{\lambda}_f$ for each frequency, we start by computing $\hat{\lambda}_{ML}$ and r_0 for all frequency positions. Then $\hat{\lambda}_p$ is computed recursively starting from the lower frequencies, in zigzag scan order. At the start of the recurrence the values of $\hat{\lambda}_{ML}$ are used for prediction but, as the recurrence progresses, predictions will be based on previously estimated values.

4. PSNR ESTIMATION

The parameter estimation method described in the previous section can be used for the purpose of blindly estimate the PSNR of images subject to DCT-based lossy encoding, without requiring the original image data. Assuming pixel values in the range of $[0; 255]$, the image PSNR is usually given by:

$$\text{PSNR}_{[\text{dB}]} = 10\log_{10}\frac{255^2}{\frac{1}{M}\sum_{k=1}^{M}\varepsilon_k^2}, \tag{13}$$

where M is the number of pixels and ε_k^2 is the squared error between the k-th reference and distorted pixel. In the context of this paper, M will be the number of DCT coefficients under analysis and $\varepsilon_k^2 = (X_k - x_k)^2$ will be the squared difference between original and quantized coefficients. Note that, in accordance with Parceval's theorem (and since DCT is an unitary transform), it is indifferent to measure PSNR in the pixel or in the DCT domain.

If the original DCT data distribution was known, the local mean square error $\hat{\varepsilon}_k^2$ at the k-th coefficient could be estimated by observing the quantized values, according to:

$$\hat{\varepsilon}_k^2 = \int_{-\infty}^{+\infty} f_X(x|X_k)(X_k - x)^2 dx. \tag{14}$$

Using *Bayes rule* and knowing that $P(X_k|x) = 1$ if the coefficient x lies in the quantization interval centered in X_k, and $P(X_k|x) = 0$, otherwise, (14) can be rewritten as:

$$\hat{\varepsilon}_k^2 = \frac{1}{P(X_k)}\int_{X_k - \frac{q_k}{2}}^{X_k + \frac{q_k}{2}} f_X(x)(X_k - x)^2 dx, \tag{15}$$

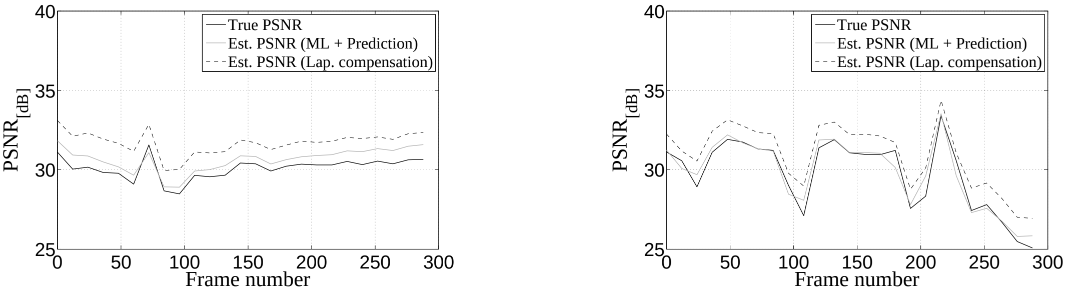
with $f_X(x)$ and $P(X_k)$ given by (1) and (4), respectively. The laplacian parameter to use in these expressions will be the value of $\hat{\lambda}_f$ that results from (12).

5. RESULTS

The effectiveness of the proposed algorithm has been evaluated using the video sequences displayed in figure 2 subject to MPEG-2 encoding. Sequences have been encoded at different rates, between 256 and 4096 kbit/s, using a GOP-12 frame structure with a prediction interval of 3 frames (IBBPBBPBBPBB...).

PSNR of the encoded sequences has been estimated for each I-frame and confronted with its true value. Figure 3 depicts two examples that illustrate the PSNR evolution along

Fig. 2. Video sequences used in the experiments. From left to right: *Akyo*; *Coastguard*; *Tempete*; *Football*; *Foreman*; *Stephan*; *Table-tennis*; *Mobile & Calendar*. All sequences have 352×288 resolution and 25 Hz frame rates.



(a) Coastguard (512kb/s). (b) Stephan (1024kb/s).

Fig. 3. PSNR estimation examples.

	λ **Pred.**	**LC**
Mean error [dB]	0.632	0.968
Root mean square error [dB]	0.806	1.171
Error value at percentile 99 [dB]	2.324	3.257
Correlation (with true value)	0.993	0.988

Table 1. PSNR estimation error statistics.

the video sequences. For comparison purposes, the results computed by the *laplacian compensation* (LC) algorithm proposed in [2] are also displayed.

Table 1 depicts the global experimental results, which account for all the video sequences, encoded at the mentioned rates (these results correspond to 960 PSNR estimates). As can be observed both from the figures an table, PSNR estimates based on the proposed algorithm for λ estimation are quite accurate, and closer to the true values than those resulting from the *laplacian compensation* method.

6. CONCLUSIONS

A new method for estimating the original DCT coefficient distribution, from their quantized values, has been proposed. This method assumes that quantization is linear, but may have different scales throughout each video frame. The proposed frame work has been applied to compute the PSNR of encoded video sequences, without requiring the original ones.

For future work, we are planning on using similar ideas for H.264/AVC encoded sequences, eventually with a different coefficient distribution model and a different predictor structure due to smaller block size in H.264/AVC. Due to its ability to estimate the local error due to encoding, the method proposed in the paper may also be useful for other applications, such as artifact reduction.

7. REFERENCES

[1] T. Brandão and M. P. Queluz, "Estimation of DCT coefficient statistics from their quantized values: application to image quality evaluation", in *proc. of ConfTele 2007, 6th Conference on Telecommunications*, Peniche, Portugal, May 2007.

[2] A. Ichigaya, M. Kurozumi, N. Hara, Y. Nishida, and E. Nakasu, "A method of estimating coding PSNR using quantized DCT coefficients", *IEEE Transactions on Circuits and Systems for Video Technology* 16(2), pp. 251–259, February 2006.

[3] E.Y. Lam and J.W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images", *IEEE Transactions on Image Processing* 9(10), pp. 1661–1666, October 2000.

[4] F. Muller, "Distribution shape of two-dimensional DCT coefficients of natural images", *Electronics Letters* 29(22), pp. 1935–1936, 1993.

[5] T. Eude, R. Grisel, H. Cherifi and R. Debrie, "On the distribution of the DCT coefficients", in *proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, Adelaide, Australia, pp. 365–368, 1994.

[6] J. Price and M. Rabbani, "Biased reconstruction for JPEG decoding", *IEEE Signal Processing Letters* 6(12), December 1999.

[7] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik, "LIVE image quality assessment database, release 2", 2006, http://live.ece.utexas.edu/research/quality.