IMPROVED DCT COEFFICIENT DISTRIBUTION MODELING FOR H.264-LIKE VIDEO CODERS BASED ON BLOCK CLASSIFICATION

Nejat Kamaci and Ghassan AlRegib

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332
nejat.kamaci@gmail.com
alregib@gatech.edu

ABSTRACT

Through extensive experimentation with a large set of video sequences, we show that modeling the statistical distribution of the transform coefficients in H.264-like video coders can be improved significantly in terms of accuracy by classifying the video source into multiple classes and modeling each class with a different statistical distribution. In this paper, we present a simple yet effective classification method and best practical models for each class and show that it is possible to improve the statistical modeling significantly without a significant complexity increase. We propose a two-class based approach in which one class is composed of very low detail blocks, and the other class is composed of high texture blocks and blocks with edges. Our two-class based statistical modeling reduces the approximation error up to 70% over the existing single-class modeling approaches for majority of the video sequences experimented. Furthermore, this approach also fits very well with the context of rate control with human visual system considerations, in which distortion in low detail regions of an image is more noticeable than in high detailed regions. In this work, we consider modeling the transform coefficients all lumped together.

1. INTRODUCTION

Most image and video coding standards use a block-based spatial transform as part of the coding algorithm [1, 2, 3]. The two-dimensional discrete cosine transform (DCT) is the most common used transform. In the H.264 video coding standard, an approximation to the DCT is used. Knowledge of the transform coefficients' statistical properties is in the design of the other video processing algorithms used in these video coders such as the design and optimization of the quantizer and the entropy coder. In particular, knowledge of the statistical properties of the transform coefficients is important in the rate control algorithms, since the problems of optimal bit allocation and quantization scale selection require knowledge of the rate-distortion relation as a function of the encoding parameters.

In the literature, several studies on the statistical distribution of the transform coefficients have been proposed. The AC coefficients were conjectured to have Gaussian [4], Laplacian [5], Cauchy [8], or more complex distributions [6, 7]. Among these, the Laplacian distribution has been the most popular because of its simplicity. Recently, the Cauchy probability density function (pdf) was also shown to be a better estimate for most video sources than the Laplacian pdf [8]. In all these studies, a single statistical model is used for all video sources. However, video sources exhibit a wide variety of statistical properties, making it impractical to use a single statistical model in most scenarios. As a result, rate and distortion models based on a single statistical distribution sometimes fail to estimate the actual rate-distortion-coding parameter relations accurately.

One way to alleviate this problem is to use more complex statistical distributions such as Gamma, Generalized Gaussian, or Student-t distributions which have more than one parameter that can be tuned to fit the empirical data. However, these complex models are not practical for mathematical analysis such as calculation of entropy and rate-distortion functions.

Instead, it is more efficient to use the existing mathematically simpler statistical distributions together with a set of visual classes. This method depends on the validity of the existence of these classes. Seeking the answer to the question whether such classification might exist and are practical, we analyzed a huge set of video sources with distinct visual characteristics such as spatial and temporal details, and at different resolutions. In this work, we claim that by using a simple yet effective classification method and using different statistical distributions for each class, it is possible to improve the estimation accuracy of the actual statistical distribution of the transform coefficients.

In this paper, we propose a two-class approach in which we categorize the coding blocks of a frame of a video sequence into two classes based on a simple pixel-domain measure. We also propose using two different probability distribution models for each class. Effectively, the video frame is divided into two classes of video sources with separate statistical properties.

The classification method we propose also fits well in the context of rate distortion optimization with human visual system considerations. The two classes that we propose also have distinct characteristics for the purpose of human eye perception of the distortion. As a result, the approach is expected to fit well with an advanced rate control algorithm.

2. CLASSIFICATION OF FRAME BLOCKS OF A VIDEO SEQUENCE

Intuitively, blocks with minor spatial variations, or so-called flat blocks should have most of their energy concentrated on the top-left transform coefficient, also i.e. the DC coefficient. For these blocks, we expect the distribution to be concentrated around a value of zero since the other coefficient, the AC coefficient, will be small. Also, the tail of the distribution will most likely decay rapidly. Based on this intuition, a class with little spatial variation might have a Laplacian distribution rather than a Cauchy distribution.

Furthermore, blocks with major variations, such as edges, texture, should have a significant portion of their energy concentrated in the AC coefficients. For these blocks, a Cauchy distribution might be a better choice since the tail of the actual distribution might decay slower. In the light of this, we propose these two classes for modeling purposes. As the classification criterion, we propose the following measure:

$$\phi = \frac{\sum_{y=0}^{N-1}\sum_{x=1}^{N-1}(blk[x,y]-blk[x-1,y])^2}{2N(N-1)\times\sigma^2_{frame}} + \frac{\sum_{x=0}^{N-1}\sum_{y=1}^{N-1}(blk[x,y]-blk[x,y-1])^2}{2N(N-1)\times\sigma^2_{frame}}, \quad (1)$$

where $blk\,[x,y]$ is the pixel value at location $[x,y]$ and σ^2_{frame} is the frame variance.

As the classification rule, we propose a simple approach that uses a threshold. For a given block, if the classification measure ϕ is below a threshold, the block is classified as a low detail block, enumerated as Class 0. Otherwise, the block is classified as Class 1:

$$Class = \begin{cases} 0, & if\,\phi < \tau, \\ 1, & otherwise. \end{cases} \quad (2)$$

3. TWO-CLASS BASED DCT COEFFICIENTS STATISTICAL MODELING

Most video coders use a two dimensional block-based 8×8 or 4×4 DCT for compression[1]. The DCT is applied to the residual image obtained by performing a prediction. Fig. 1 shows a typical plot of the histogram of the transform coefficients for a 4×4 block based DCT of a video frame from the TEMPETE sequence.

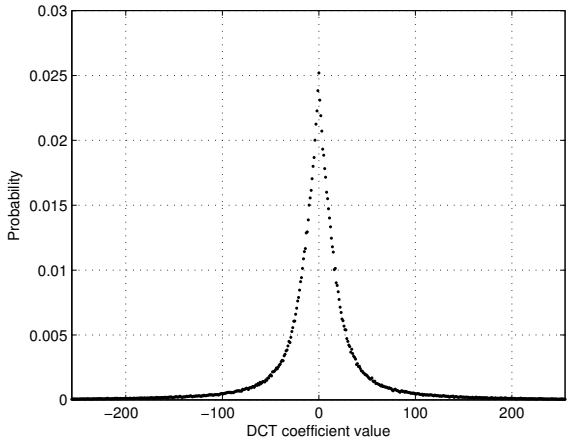

Fig. 1. Distribution of the cumulative DCT coefficients - for a video frame from the Tempete Sequence (CIF format).

Using the mean-square error criterion, we can find the best Laplacian and Cauchy pdf that would fit this histogram. Figs. 2 and 3 illustrate how accurate each fit would be. Notice that the Laplacian pdf does not fit the empirical distribution for larger coefficient values accurately. The Cauchy pdf however, fits better except around smaller values.

[1]The recent coders such as H.264, SVC, and MVC use a 4×4 integer transform that is an approximation to the two dimensional 4×4 DCT.

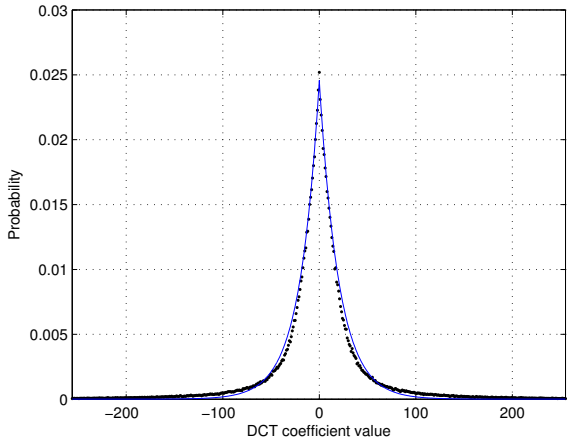

Fig. 2. Solid curve is the Laplacian fit to the distribution of the cumulative DCT coefficients - for the video frame from the Tempete Sequence (SD format).

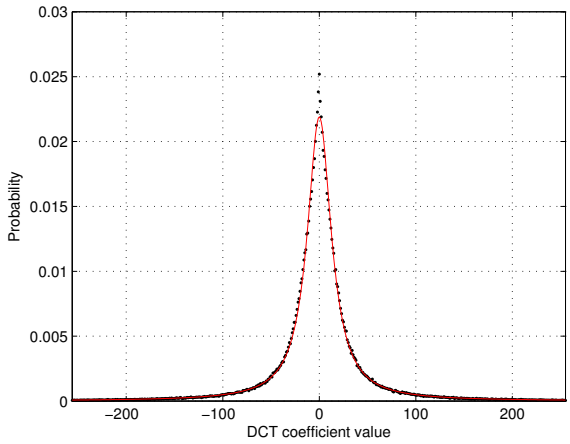

Fig. 3. Solid curve is the Cauchy fit to the distribution of the cumulative DCT coefficients - for the video frame from the Tempete Sequence (SD format).

Using the classification method described in Section 2, we consider separating the cumulative histogram of the frame into two histograms: Class 0, and Class 1, respectively. For Class 0, we consider fitting a Laplacian pdf with parameter λ:

$$p\,(x) = \frac{\lambda}{2}\exp\{-\lambda\,|x|\}, \quad x \in \mathbf{R}. \quad (3)$$

The Laplacian pdf has an exponential form, leading to the property that the tail of the density decays very fast.

For Class 1, we consider fitting a zero-mean Cauchy distribution with parameter μ, having the pdf

$$p\,(x) = \frac{1}{\pi}\frac{\mu}{\mu^2+x^2}, \quad x \in \mathbf{R}, \quad (4)$$

for which the tail of the density decays slow.

The model parameters λ and μ will be estimated using the empirical data and using the Kolmogorov-Smirnov measure. As illustration, Fig. 4 and Fig. 5 show the histogram of the same transform coefficients for separated for Class 0 and Class 1, respectively.

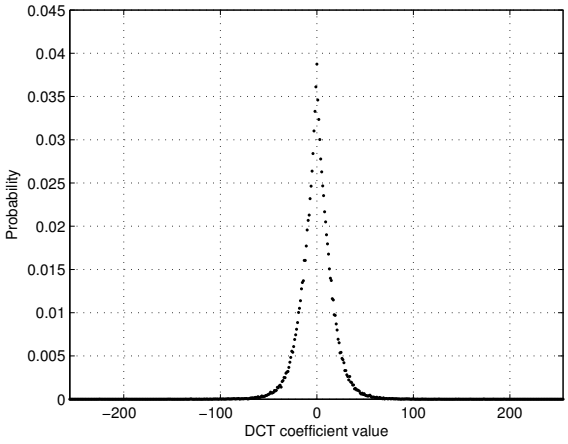

Fig. 4. The empirical distribution of the class 0 DCT coefficients - for the video frame from the Tempete Sequence (SD format).

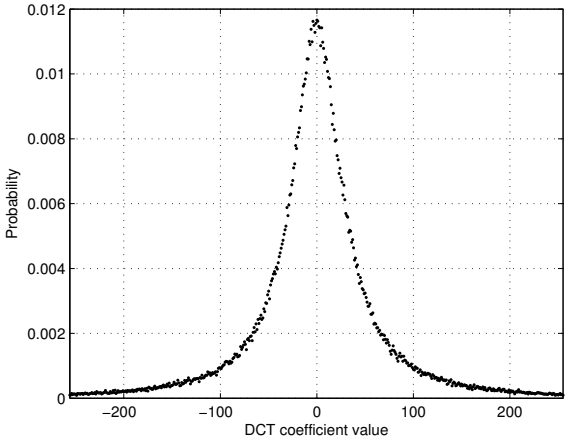

Fig. 5. The empirical distribution of the class 1 DCT coefficients - for the video frame from the Tempete Sequence (SD format).

Applying classification method to the same histogram shown on Fig. 1, we obtain two separate histograms, Class 0 and Class 1, respectively. Fig. 6 shows the Class 0 histogram with a Laplacian pdf fit to it. The Laplacian pdf parameter λ is estimated by using a mean-square error measure, and using an exhaustive search method for all possible values. The convergence criterion is the change in $\lambda \leq 0.01$. Fig. 7 shows the Class 1 histogram with a Cauchy pdf fit to it. The Cauchy pdf parameter μ is estimated similarly.

Comparing these two histogram fits, it is clear that by using a classification such as in Section 2 the empirical distribution of the DCT coefficients of this particular frame can be modelled more accurately.

To support our claims, we will present a set of experiments in which we generate empirical histograms using a wide range of video sequences, and measure the goodness-of-fit for both cases. We will compare the goodness-of-fit measures using a single class versus using the two classes described in Section 2. The experimentation setup and results are presented in Section 4.

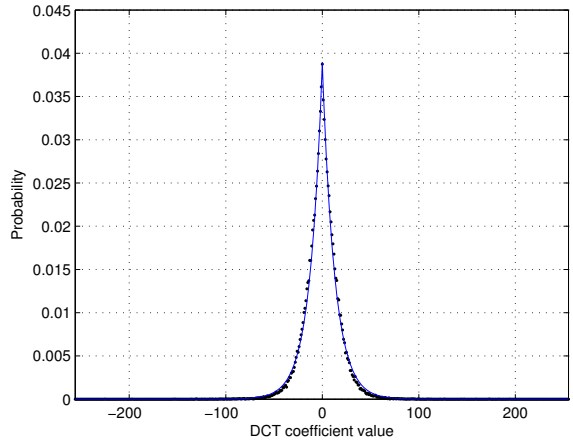

Fig. 6. Laplacian fit to the distribution of the class 0 DCT coefficients - for the video frame from the Tempete Sequence (SD format).

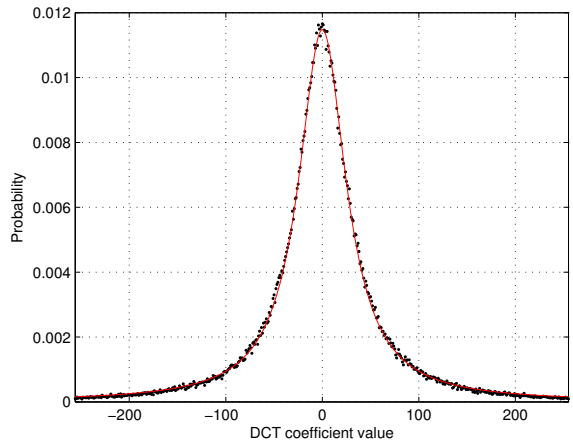

Fig. 7. Cauchy fit to the distribution of the class 1 DCT coefficients - for the video frame from the Tempete Sequence (SD format).

4. EXPERIMENTS

To evaluate the goodness-of-fit of the proposed classification based modeling approach, we generated the empirical distributions of the DCT coefficients of the video sequences shown in Table 1.

For these video sequences, we generated the empirical distributions of the first 20 frames encoded using the H.264 reference software, version JM-17.0 [9]. First frame/field was intra coded and the rest of the frames were inter coded. For the inter coded frames, we disabled intra macroblock modes. We used a constant quantization parameter for each frame. For the results, the quantization parameter is selected as 28 for I-frames, and 29 for P-frames.

Tables 2, 3 and 4 summarize the fit accuracy of both distributions. Tables show both the single histogram fit errors and the proposed two-class histogram approach. Left two columns show the goodness-of-fit error values using the cumulative histogram and Laplacian-vs-Cauchy pdf's. The right-most two columns show the goodness-of-fit error values for the proposed classification based modeling. Class 0 is fit with Laplacian pdf and Class 1 is fit with Cauchy pdf. Overall, the two-class based fit provides up to 70%,

Res.	Scan	Fps	Name
CIF	Progressive	30	BUS, CITY, COASTGUARD, FLOWER, FOOTBALL, ICE, MOBILE & CALENDAR, CREW, FOREMAN, PARIS, STEFAN, WATERFALL
SD	Interlaced	30	CEREMONY, CONCERT, DOWNTOWN, FAST FOOD, FESTIVAL,FOOTBALL, LETTERS, RUGBY, TEMPETE, WATERFALL
SD	Progressive	30	CITY, ICE, SOCCER
HD	Progressive	60	BLUE SKY, PEDESTRIAN, RIVER BED, RUSH HOUR, STATION 2, SUNFLOWER,

Table 1. The list of the test streams and their properties.

and an average of 22% less estimation error.

	K-S goodness-of-fit errors			
	Cumulative		Class 0	Class 1
CIF Sequences	Lapl.	Cauc.	Lapl.	Cauc.
BUS	0.085	0.102	0.039	0.082
CITY	0.092	0.094	0.063	0.079
COASTGUARD	0.101	0.124	0.077	0.082
CREW	0.106	0.065	0.085	0.083
FLOWER	0.199	0.190	0.060	0.065
FOOTBALL	0.082	0.056	0.065	0.077
FOREMAN	0.057	0.061	0.067	0.100
ICE	0.152	0.136	0.050	0.076
MOB. & CAL.	0.133	0.140	0.055	0.063
PARIS	0.097	0.082	0.065	0.073
STEFAN	0.133	0.105	0.099	0.099
TEMPETE	0.114	0.086	0.055	0.067

Table 2. Goodness-of-fit errors based on the K-S criterion for the CIF resolution sequences.

5. SUMMARY AND CONCLUSION

In this paper, we claim that the DCT coefficient distribution is different for different classes of video content. We propose a classification method and a separate analytical model for each class to improve the accuracy of approximating the DCT coefficient distribution over a single class. Our experiments prove that there might be more than one class of content in each video frame, and each of these classes have a different statistical distribution. Although our simple classification metric is successful for majority of the video sources that we experiment, further research is possible to better the classification method, with possibly more than two classes.

6. REFERENCES

[1] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM,* vol. 34, pp. 30-44, Apr. 1991.

[2] "The MPEG-2 international standard," ISO/IEC, Reference number ISO/IEC 13818-2, 1996.

	K-S goodness-of-fit errors			
	Cumulative		Class 0	Class 1
SD Sequences	Lapl.	Cauc.	Lapl.	Cauc.
WATERFALL	0.081	0.120	0.105	0.085
RUGBY	0.082	0.080	0.049	0.062
DOWNTOWN	0.084	0.085	0.044	0.059
GAME CEREMONY	0.063	0.112	0.060	0.059
CONCERT	0.046	0.083	0.065	0.078
FAST FOOD	0.100	0.134	0.109	0.052
FESTIVAL	0.068	0.108	0.094	0.071
FOOTBALL	0.048	0.099	0.051	0.054
LETTERS	0.194	0.210	0.099	0.087
TEMPETE	0.055	0.090	0.050	0.038
CITY	0.079	0.104	0.056	0.044
ICE	0.107	0.120	0.047	0.051
SOCCER	0.089	0.100	0.072	0.076

Table 3. Goodness-of-fit errors based on the K-S criterion for the SD resolution sequences.

	K-S goodness-of-fit errors			
	Cumulative		Class 0	Class 1
HD Sequences	Lapl.	Cauc.	Lapl.	Cauc.
SUNFLOWER	0.070	0.066	0.057	0.072
STATION 2	0.051	0.083	0.075	0.034
RUSH HOUR	0.055	0.077	0.052	0.063
RIVER BED	0.061	0.082	0.069	0.036
PEDESTRIAN AREA	0.052	0.084	0.066	0.052
BLUE SKY	0.095	0.134	0.097	0.041

Table 4. Goodness-of-fit errors based on the K-S criterion for the HD resolution sequences.

[3] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Image Proc.,*, vol. 13, no. 7, July 2003, pp. 560-576.

[4] A. N. Netravali and J. O. Limb, "Picture coding: A review," *Proc. IEEE,* vol. 68, pp. 7-12, Mar 1960.

[5] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Trans. on Commun.,* vol. COM-31, pp 835-839, June 1983.

[6] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. on Image Proc.,* vol. 9, no. 10, Oct. 2000, pp. 1661-1666.

[7] Bhuiyan, M. I. H. and Ahmad, M. O. and Swamy, M. N. S., "Modeling of the DCT coefficients of images," *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on,* May 2008, pp. 272-275.

[8] N. Kamaci, Y. Altunbasak, and R. M. Mersereau, "Frame bit allocation for the H.264/AVC video coder via Cauchy-density based rate and models," *IEEE Trans. on CSVT,* vol. 15, Issue 8, Aug. 2005, pp. 994-1006.

[9] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Joint Model Reference Software Version 17.0.