Impact of Video Parameters on The DCT Coefficient Distribution for H.264-Like Video Coders

Nejat Kamaci and Ghassan Al-Regib

Georgia Institute of Technology, Atlanta, GA 30332, USA

ABSTRACT

We examine the impact of various encoding parameters on the distribution of the DCT coefficients for H.264-like video coders. We model the distribution of the frame DCT coefficients using the most common Laplacian and Cauchy distributions. We show that the resolution, the quantization levels and the coding type have significant impact on the accuracy of the Laplacian and Cauchy distribution based models. We also show that the transform kernel (4×4 vs 8×8) has little impact. Moreover, we show that for the video sources that have little temporal or spatial detail, such as flat regions, the distribution of the frame DCT coefficients resembles a Laplacian distribution. When the video source exhibits more detail, such as texture and edges, the distribution of the frame DCT coefficients resembles a Cauchy distribution. The correlation between the details of the video source to the two probability distributions can be used to further improve the estimation of the distribution of the frame DCT coefficients, by using a classification based approach.

Keywords: H.264, DCT modeling

1. INTRODUCTION

Compression is a key part of the video processing for which numerous video coding standards have been developed and adopted by the industry[1–3] to date. These standards provide us with the necessary tools to compress and encode video sources to satisfy the needs of the visual information processing and communication applications.

From the video communications perspective, the user experience is affected by numerous factors including but not limited to the network conditions, and the user environment characteristics such as the user interface capabilities, and the physical environment. The network conditions affect the amount of data transmitted between the subjects. The user environment conditions might dictate certain requirements such as video resolution and complexity. The physical environment of the environment also affect the quality of the coded video because the output bit rate and the quality of the coded video depends on the statistical characteristics of the video source.

To improve the video experience in video communications, the video subsystem can be designed to handle the aforementioned variations in an efficient manner. The encoded video can be adapted based on the conditions of the communication environment, and therefore a good understanding of the impact of the communication environment on the video coding performance is crucial. From the video coding point of view, this translates into how the coded video output will be affected by the nature of the source video and the coding constraints.

Most of the video coding algorithms use a block-based spatial transform as part of the coding algorithm. The two-dimensional discrete cosine transform (DCT) is the most common used transform. The statistical properties of the DCT coefficients in a transform-based video coding algorithm has great importance in satisfying the application constraints and controlling the quality of the coded video. 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,[6] or more complex distributions.[7] Among these, the Laplacian distribution has been the most popular because of its simplicity.

Further author information: (Send correspondence to Prof. Al-Regib)
Ghassan Al-Regib : E-mail: alregib@gatech.edu
Nejat Kamaci: E-mail: nejat.kamaci@gmail.com

The Laplacian distribution is characterized with the probability density function (pdf) as follows:

$$p(x) = \frac{\lambda}{2} \exp\{-\lambda |x|\}, \qquad x \in \mathbf{R}, \tag{1}$$

where $\lambda > 0$ is the pdf parameter. The Laplacian pdf has an exponential form, leading to the property that the tail of the density decays very fast.

Recently, the Cauchy probability density function (pdf) was shown to be a better estimate for most video sources than the Laplacian pdf.[6] The Cauchy distribution is characterized with the following pdf:

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

where $\mu > 0$ is the pdf parameter. The tail of the Cauchy pdf decays much slower than the Laplacian. The Laplacian and the Cauchy pdf each has a single pdf parameter. These two distributions are used more commonly in the literature for practical purposes.

To the best of our knowledge, in all of the statistical analysis studies, a single statistical model is used for all kinds of 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.

In this work we analyze the impact of several factors present in video communication environments on the DCT coefficient statistics for video coding purposes. We analyze the impact of the resolution of the video source, the output video bit rate and complexity requirements. The bit rate requirements are addressed through the selection of quantization parameters. The complexity is addressed through the use of different tools such as 4×4 and 8×8 transforms, and intra versus inter coded frames.

The paper is organized as follows. In Section 2 we describe our experimental setup. In Section 3 we analyze the impact of resolution, the quantization parameter (QP), the transform type, and the coding type. In Section 4 we summarize the results and draw conclusions.

2. EXPERIMENTAL SETUP

Res.	Scan	Fps	Name
CIF	P	30	Akiyo, Bus, City, Coastguard, Crew, Flower, Football, Foreman, Ice, Mobile & Calendar, Crew, Paris, Soccer, Stefan, Waterfall
SD	I	30	Ceremony, Concert, Downtown, Fast Food, Festival, Football, Formula 1, Letters, Rugby, Tempete, Waterfall
SD	P	30	City, Ice, Soccer
qHD	P	25	Mobile 2, Park Run, Shields
qHD	P	30	Blue Sky, Pedestrian Area, River Bed, Rush Hour, Station, Stockholm, Sunflower, Tractor
HD	P	50	Mobile 2, Park Run, Shields
HD	P	60	Blue Sky, Pedestrian Area, River Bed, Rush Hour, Station, Stockholm, Sunflower, Tractor

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

To examine the effect of the considered encoding parameters on the DCT statistical distribution, we use the H.264 reference encoder JM 17.0[8] to generate the histograms and consider the following experimental setup:

- We consider the video sequences shown in Table 1. In this table, 'P' stands for *Progressive Scan*, 'I' stands for *Interlaced Scan*, and 'qHD' stands for *quarter-HD*.

- We use these quantization parameters: $QP \in \{6, 12, 18, 24, 30, 36, 42, 48\}$. Note these are the quantization indices. The actual quantization level has an exponential relation to the quantization index.
- The encoder is configured to use an open GOP structure with only I and P frames.
- Intra and skip macroblocks are disabled in P frames. SAD is selected as the distortion metric for motion estimation with a maximum search range of 32. Weighted prediction is not used, and the loop filter is enabled.
- To generate the 4×4 DCT histograms, we use the main profile tools. To generate the 8×8 DCT histograms, we use the high profile (ID 100) tools and enable only 8×8 transform.
- We consider histogram values generated by all coefficients lumped together rather than individual frequency components. This is more suitable for quantizer selection and rate control applications.
- We use the Kolmogorov-Smirnov (K-S) goodness-of-fit criterion to assess the goodness-of-fit when the histograms are fitted using the Laplacian and the Cauchy statistical distributions that are considered for comparison.

3. EXPERIMENTS

In this section, we analyze the impact of the four parameters to the histogram of the frame DCT coefficients and the Laplacian and Cauchy distribution based models.

3.1 Impact of Resolution

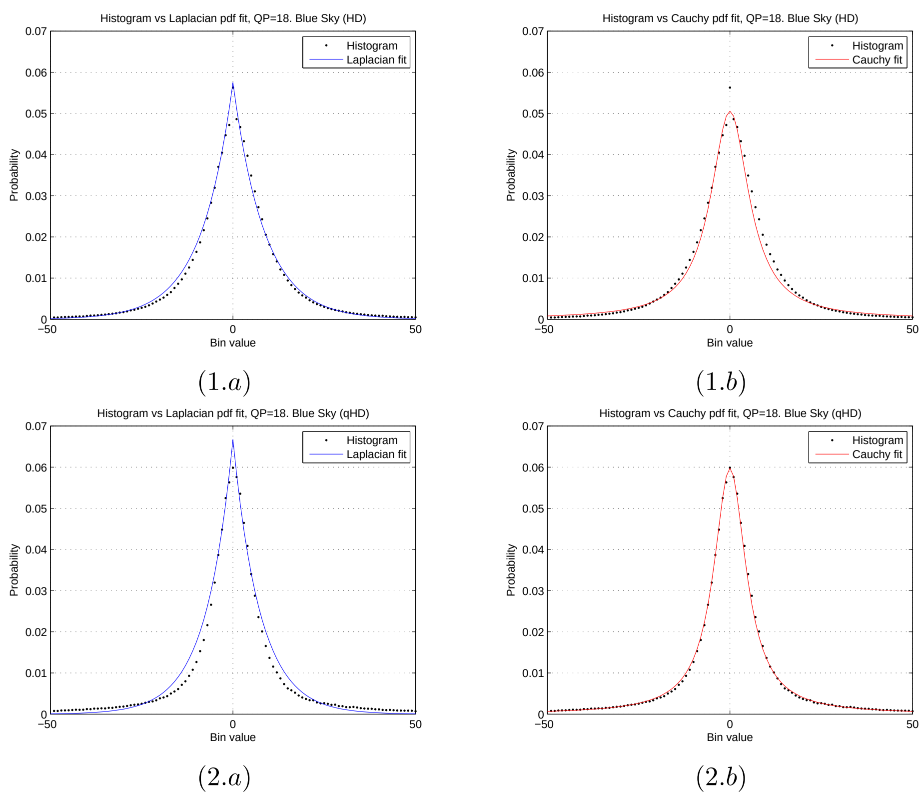


Figure 1. Impact of resolution on the 4×4 transform coefficient histogram and goodness-of-fit for the Blue Sky sequence ($QP = 18$). (1.a) Laplacian and (1.b) Cauchy fit at HD resolution. (2.a) Laplacian and (2.b) Cauchy fit at qHD resolution.

The spatial and temporal correlation among the pixels of video sources will vary depending on the spatial resolution and the frame rate. This in turn will impact the statistical distribution of the DCT coefficients. Figure 1 illustrates how the resolution might impact the shape of the distribution of the frame DCT coefficients and the statistical modelling of them using the Laplacian and the Cauchy distributions. The accuracy of the Cauchy distribution based approximation improves greatly when the resolution is reduced from HD to qHD. On the contrary, the accuracy of the Laplacian distribution based approximation becomes poorer as the resolution reduces.

Obviously, one example is not good enough to make claims. So we test our claim by experimenting with a wide set of video sequences obtained at two different resolutions. The original video sequences have HD resolution. We generated the qHD resolution sequences by downsampling the HD sequences by two and using a 3-tap Lanczos anti-aliasing filter.

Table 2 shows the K-S error values for the two resolutions. According to these results, as the resolution decreases, the statistical distribution of the DCT coefficients gets closer to the Cauchy distribution. At HD resolution, the Cauchy distribution is 12% more accurate than the Laplacian distribution on average. At qHD resolution, the Cauchy distribution is 38% more accurate than the Laplacian distribution on average. At HD resolution, there are four streams out of eleven that are better modelled by a Laplacian pdf. At qHD resolution, there is only one.

	K-S goodness-of-fit errors					
	HD Resolution			qHD Resolution		
Sequence	Laplacian	Cauchy	Improvement%	Laplacian	Cauchy	Improvement%
Blue Sky	0.0335	0.0293	12.56	0.0465	0.0196	57.80
Mobile 2	0.0231	0.0158	31.67	0.0305	0.0225	26.49
Park Run	0.0305	0.0103	66.33	0.0208	0.0110	47.47
Pedestrian	0.0294	0.0368	-25.35	0.0229	0.0155	32.17
River Bed	0.0243	0.0141	41.59	0.0291	0.0088	70.16
Rush Hour	0.0204	0.0264	-29.30	0.0205	0.0220	-7.50
Shields	0.0214	0.0146	31.39	0.0240	0.0169	29.75
Station	0.0220	0.0264	-19.65	0.0216	0.0171	21.23
Stockholm	0.0195	0.0093	52.85	0.0216	0.0171	21.23
Sunflower	0.0268	0.0275	-2.99	0.0231	0.0185	19.97
Tractor	0.0295	0.0233	21.22	0.0319	0.0148	53.82
Ave	0.0255	0.0213	12.02	0.0266	0.0165	38.13

Table 2. Impact of the resolution on the goodness-of-fit results. Comparing HD and qHD resolution results using 4×4 transform.

3.2 Impact of the Quantization Parameter

The quantization parameter selection has an indirect effect on the DCT coefficient distribution via the predictive coding. The reconstructed picture distortion is proportional to the quantization level used while encoding. Thus the motion compensated prediction will in general produce a smaller residue when the reference picture is reconstructed with less distortion (i.e. with a smaller QP), assuming all other coding parameters are identical. In this case, the DCT coefficients should be concentrated around zero statistically, and the tail of the DCT coefficient distribution should be lighter than when the reference is encoded with a larger QP. Intuitively, the Cauchy pdf will characterize a high tail better than a Laplacian pdf. Therefore, we expect that the accuracy of the Cauchy pdf will improve as QP increases and the accuracy of the Laplacian pdf will improve as QP decreases.

Figures 2 and 3 illustrate how the DCT distribution varies based on the quantization levels for and HD resolution (Sunflower) and a CIF resolution (Akiyo) sequence, respectively. Figures show the DCT histogram of inter coded second frame of each sequence when encoded using $QP = 18$ and is fitted with (1.a) a Laplacian pdf and (1.b) a Cauchy pdf, respectively. Notice in this case the Laplacian pdf is a better fit. Figures also show the

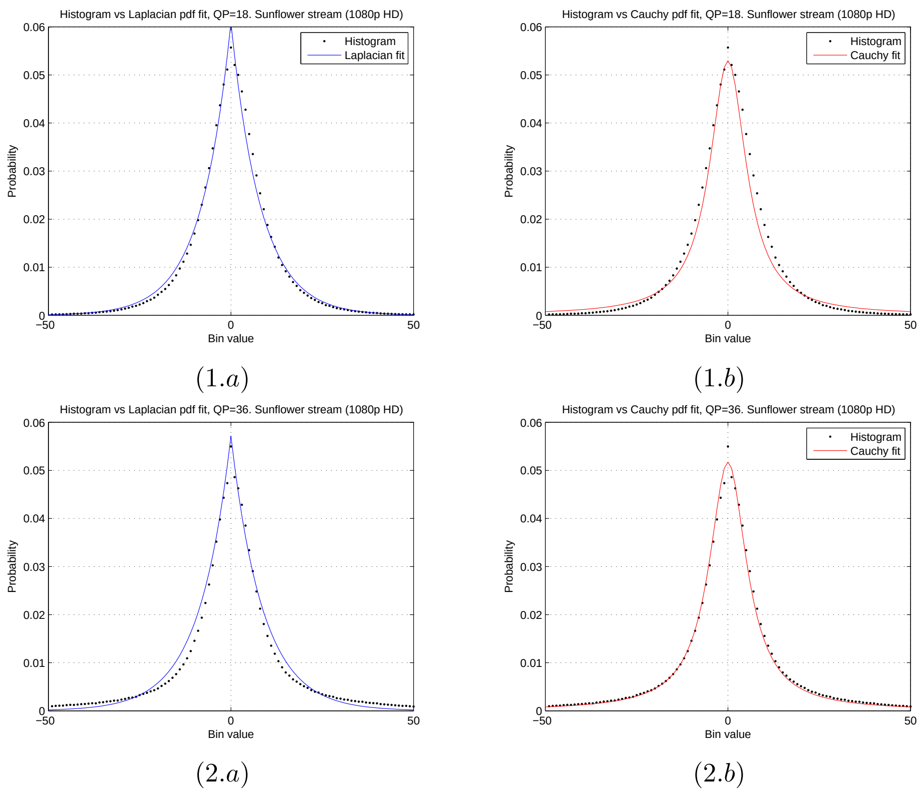


Figure 2. Impact of QP on the 4×4 transform coefficient histogram and goodness-of-fit for the HD resolution Sunflower sequence. (1.a) Laplacian and (1.b) Cauchy fit when $QP = 18$. (2.a) Laplacian and (2.b) Cauchy fit when $QP = 36$.

DCT histogram when encoded using $QP = 36$ and is fitted with (2.a) a Laplacian pdf and (2.b) a Cauchy pdf, respectively. Notice in this case the Cauchy pdf is a better fit for both video sequences.

To test our claims, we experimented with a wide range of video sequences using the experimental setup described in Section 2. In our tests, he first picture is intra coded the rest is inter coded. Figure 4 shows how the accuracy is impacted by QP for the SD resolution sequences in our test setup. The error value for each QP index is calculated as an average over all frames encoded for each video sequence. Results on Figure 4 suggest that the estimation accuracy of the Laplacian distribution is affected by the QP significantly more than the estimation accuracy of the Cauchy distribution. Compared to the Cauchy distribution based estimation, the Laplacian distribution based estimation errors are significantly larger for higher QP values. For lower QP values, the results are mixed. For six out of the total 16 sequences, the Laplacian distribution based estimation performs better. For the rest, the Cauchy distribution based estimation performs equal or better.

Table 3 tabulates the numerical results. In the table, the approximation error for the Laplacian and Cauchy approximations as a function of the QP for CIF and SD resolution sequences are shown. The approximation errors are according to the K-S criterion. The error values are averaged over all sequences.

According to the results, there is a clear relation between QP and the accuracy of the two approximations. For the Laplacian approximation, interestingly, the best accuracy is observed when $QP = 24$. For the Cauchy approximation, the accuracy is in general better when QP is small. Comparing both approximations, the Cauchy approximation is better for all QP values, especially when QP is higher.

The results also suggest that the resolution has an impact on the accuracy for different QP values. When QP is in the lower range, an increase in the resolution affects the estimation accuracy in favor of the Laplacian distribution. However when QP is in the higher range, the resolution increase affects the estimation accuracy in favor of the Cauchy distribution. This is an interesting result.

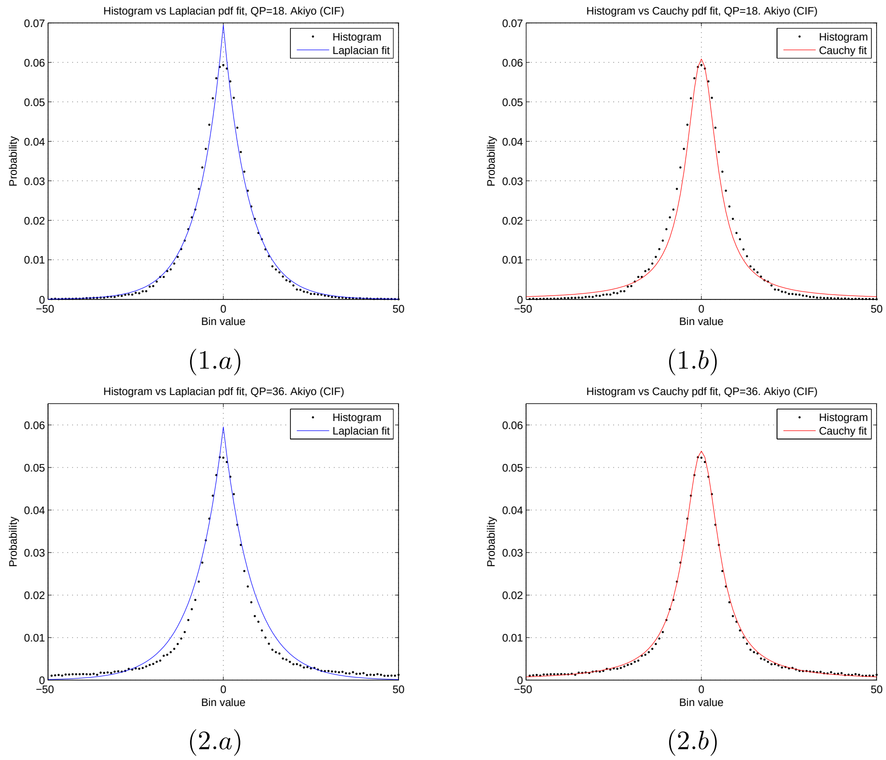


Figure 3. Impact of QP on the 4×4 transform coefficient histogram and goodness-of-fit for the CIF resolution Akiyo

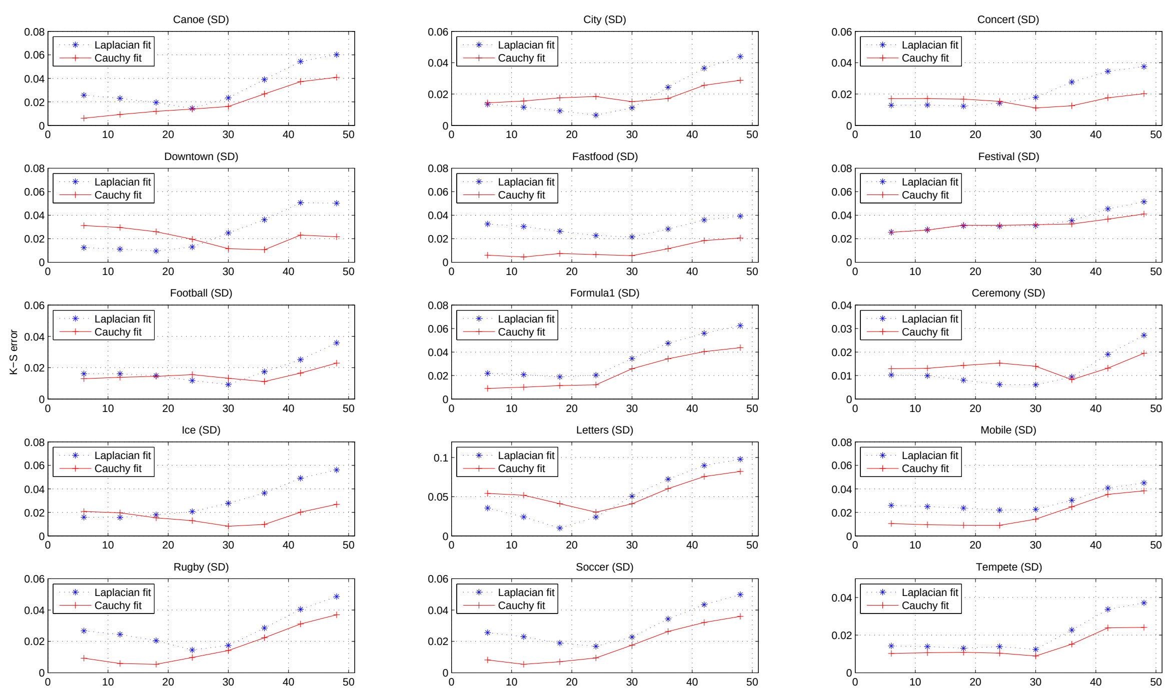


Figure 4. QP versus goodness-of-fit plots for the SD resolution sequences

	K-S goodness-of-fit errors						
	CIF resolution)				SD resolution)		
QP	Laplacian	Cauchy	Improvement%		Laplacian	Cauchy	Improvement%
6	0.0265	0.0170	35.99		0.0219	0.0189	13.55
12	0.0255	0.0179	29.66		0.0212	0.0184	13.16
18	0.0226	0.0182	19.61		0.0201	0.0181	9.89
24	0.0213	0.0192	9.84		0.0190	0.0173	8.97
30	0.0263	0.0206	21.81		0.0230	0.0174	24.05
36	0.0374	0.0256	31.64		0.0437	0.0277	36.65
42	0.0467	0.0324	30.66		0.0437	0.0277	36.65
48	0.0528	0.0367	30.40		0.0487	0.0307	36.92
Ave	0.0324	0.0234	27.78		0.0288	0.0211	26.74

Table 3. Impact of the QP on the goodness-of-fit results for the CIF and SD resolution sequences in terms of the K-S criterion using 4×4 transform.

3.3 Impact of Transform Type

To assess whether and how the selection of the transform type will impact the DCT coefficient distribution, we experimented with 8×8 and 4×4 DCT transforms.

Table 4 summarizes the goodness-of-fit errors combased on the K-S criterion for the CIF resolution sequences. The 4×4 and the 8×8 transform cases are shown separately. Histograms are obtained by encoding with 8 different QP values and the goodness-of-fit errors are averaged. The left-most column shows the goodness-of-fit error values averaged over all quantization parameter values and using the Laplacian pdf. The middle column shows the goodness-of-fit error values averaged over all quantization parameter values and using the Cauchy pdf. The right-most column shows the improvement of goodness of fit of the Cauchy pdf over that of the Laplacian pdf. Comparing 8×8 statistics to that of the 4×4, we observe that the results are slightly different. For the 8×8 DCT, the Cauchy distribution approximates the histogram values about 20% better compared to the 4×4 case.

	K-S goodness-of-fit errors						
	4×4 DCT				8×8 DCT		
Sequence	Laplacian	Cauchy	Improvement%		Laplacian	Cauchy	Improvement%
Akiyo	0.0295	0.0263	10.72		0.0281	0.0236	15.93
Bus	0.0251	0.0143	42.92		0.0293	0.0175	40.52
City	0.0174	0.0170	2.72		0.0215	0.0138	35.95
Coastguard	0.0137	0.0169	-23.65		0.0224	0.0143	36.24
Crew	0.0322	0.0173	46.37		0.0336	0.0135	59.97
Flower	0.0781	0.0616	21.15		0.0754	0.0593	21.35
Football	0.0325	0.0153	52.92		0.0389	0.0180	53.90
Foreman	0.0240	0.0195	18.83		0.0234	0.0163	30.41
Ice	0.0468	0.0208	55.44		0.0462	0.0216	53.18
Mob. & Cal.	0.0398	0.0326	18.04		0.0329	0.0264	19.79
Paris	0.0319	0.0233	26.88		0.0299	0.0229	23.59
Soccer	0.0188	0.0157	16.67		0.0229	0.0146	36.37
Stefan	0.0476	0.0306	35.72		0.0476	0.0300	37.03
Tempete	0.0346	0.0222	35.82		0.0315	0.0202	36.03
Waterfall	0.0138	0.0182	-31.67		0.0226	0.0178	21.33
Ave	0.0324	0.0234	27.62		0.0337	0.0220	34.92

Table 4. Goodness-of-fit errors based on the K-S criterion for the CIF resolution sequences. Comparison of the 4×4 and the 8×8 transforms.

3.4 Impact of Intra vs Inter Coding

In the H.264 video coding standard and its extensions (SVC,[9] MVC[10]), both the intra and the inter coded pictures use predictive coding. However, for intra pictures, only a spatial prediction using neighboring pixels of the previously encoded macroblocks is used for a given macroblock to be encoded e.g. *intra prediction*. The inter prediction uses pixels from previously encoded pictures and has much greater prediction ability compared to the intra prediction. As a result, the DCT coefficient distribution shows different characteristics when a given video source is intra coded versus inter coded. This claim is also supported in the literature.[7] In this paper, we provide additional experiments to analyze the impact of intra and inter coding.

To examine the impact of intra versus inter coding, we encode the sequences in our test bench using intra and inter coding and collect the frame histograms. Then we estimate the generated histograms and calculate K-S goodness-of-fit errors of both the Laplacian and Cauchy distributions. Table 5 shows the results of our experiments. The results show that the Laplacian estimation to the frame DCT distribution is approximately 20% worse when the frame is intra coded. The Cauchy estimation performs similarly for both cases.

	K-S goodness-of-fit errors						
	Intra Coded				Inter Coded		
Sequence	Laplacian	Cauchy	Improvement%		Laplacian	Cauchy	Improvement%
Blue Sky	0.0258	0.0183	28.93		0.0246	0.0213	13.20
Mobile 2	0.0143	0.0109	23.93		0.0166	0.0175	-5.53
Park Run	0.0371	0.0115	69.04		0.0172	0.0068	60.29
Pedestrian	0.0200	0.0243	-21.45		0.0203	0.0262	-28.81
River Bed	0.0180	0.0095	47.17		0.0167	0.0108	35.33
Rush Hour	0.0145	0.0189	-30.94		0.0154	0.0195	-27.07
Shields	0.0226	0.0139	38.52		0.0184	0.0137	25.30
Station 2	0.0143	0.0140	2.66		0.0165	0.0195	-18.15
Stockholm	0.0228	0.0246	-8.03		0.0121	0.0061	49.47
Sunflower	0.0181	0.0167	7.60		0.0194	0.0201	-3.22
Tractor	0.0267	0.0136	49.06		0.0210	0.0170	19.05
Ave	0.0213	0.0160	18.77		0.0180	0.0162	10.89

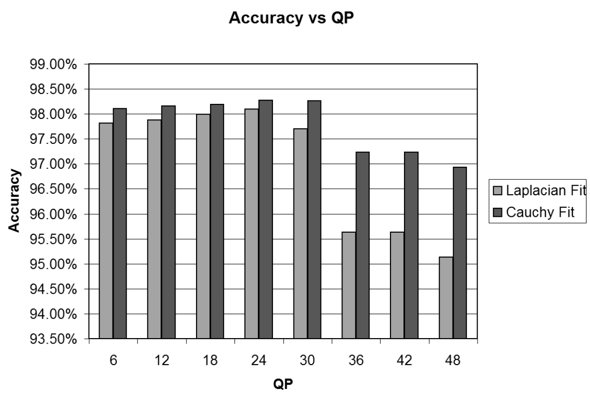
Table 5. Goodness-of-fit errors based on the K-S criterion for the HD resolution sequences. Comparison of the inter and the intra coded pictures using 4×4 transform.

4. SUMMARY AND CONCLUSION

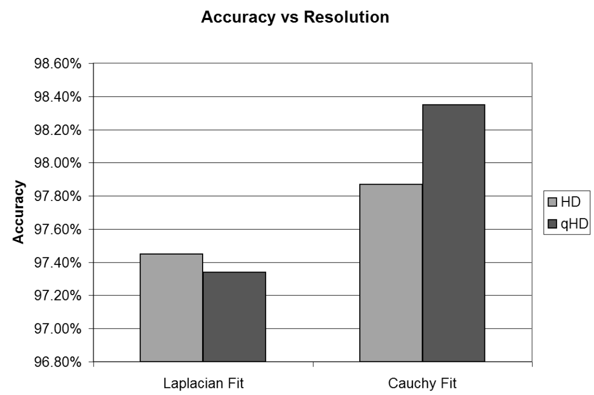
In this work, we presented an experimental analysis of the impact of some of the most fundamental parameters of the coded video on the distribution of the DCT coefficients for H.264-like video coders. We chose the Laplacian and the Cauchy distributions as basis for comparison for approximating the actual DCT coefficient distribution due to their popularity and practicality over other statistical distributions proposed in the literature. We analyzed the impact of the resolution, the QP selection, the transform size, and the coding type. A summary of our analysis is shown on Figure 5. We observed that:

- The resolution has a great impact on the distribution of the frame DCT coefficients. On the average, the accuracy of the Cauchy distribution in estimating the frame DCT coefficients reduces by approximately 30% as the resolution increases.
- The quantization level has the biggest impact on the distribution of the frame DCT coefficients. For the Laplacian estimation, the accuracy can decrease as much as 150% when QP gets higher, whereas for the Cauchy estimation, the impact on the accuracy is a maximum of 120%. Interestingly, the best approximation is obtained when QP is close to the mid-value (24) of its allowed range $[0, 51]$.
- The transform kernel size has small impact on the distribution of the frame DCT coefficients. The estimation errors for the 4×4 and the 8×8 transform differ by about 5% only.

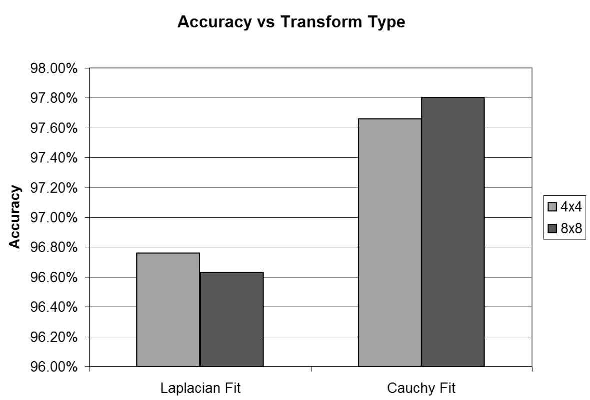
- The coding type (intra vs. inter) has a significant impact on the distribution of the frame DCT coefficients. The Laplacian estimation to the frame DCT coefficients distribution is approximately 20% worse when the frame is intra coded.



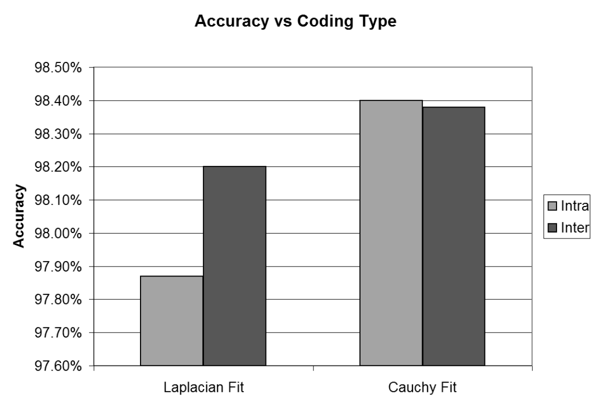
(*a*)



(*b*)



(*c*)



(*d*)

Figure 5. Charts showing the average accuracy of Laplacian and Cauchy pdf fit as a function of (a) QP, (b) resolution, (c) transform size, and (d) coding type.

Our experimental results indicate that overall, the Cauchy distribution based estimation is more accurate than the Laplacian distribution based estimation. The accuracy of the Cauchy distribution is particularly better when the video source has more detail, hence the energy of the residual pixels is high. For the video sources that have little temporal or spatial detail, such as flat regions, the frame DCT coefficients tend to have a Laplacian distribution. This is mainly due to the fact that the Laplacian pdf is heavily concentrated around zero. When the video source exhibits more detail, such as texture and edges, the frame DCT coefficients tend to have a Cauchy distribution. This is mainly due to the fact that the Cauchy pdf has a heavier tail than the Laplacian pdf.

The correlation between the details of the video source to the two probability distributions can be used to further improve the estimation of the distribution of the frame DCT coefficients, by using a classification based approach.

REFERENCES

[1] Wallace, G. K., "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics* **38**(1) (1992).

[2] Haskell, B. G., Puri, A., and Netravali, A. N., [*Digital Video: An introduction to MPEG-2*], Chapman & Hall, Ltd., London, UK, UK, 1st ed. (1996).

[3] Wiegand, T., Sullivan, G. J., Bjontegaard, G., and Luthra, A., "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology* **13**(7), 560–576 (2003).

[4] Netravali, A. N. and Limb, J. O., "Picture coding: A review," *Proceedings of the IEEE* **68**(3), 366–406 (1980).

[5] Reininger, R. and Gibson, J., "Distributions of the two-dimensional DCT coefficients for images," *IEEE Transactions on Communications* **31**(6), 835–839 (1983).

[6] Kamaci, N., Altunbasak, Y., and Mersereau, R. M., "Frame bit allocation for the H.264/AVC video coder via cauchy-density-based rate and distortion models," *IEEE Transactions on Circuits and Systems for Video Technology* **15**(8), 994–1006 (2005).

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

[8] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, "Joint model reference software version 17.0."

[9] Schwarz, H., Marpe, D., and Wiegand, T., "Overview of the Scalable Video Coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology* **17**(9), 1103–1120 (2007).

[10] Vetro, A., Wiegand, T., and Sullivan, G. J., "Overview of the Stereo and Multiview Video Coding extensions of the H.264/MPEG-4 AVC standard," *Proceedings of the IEEE* **99**, 626–642 (2011).