Non-strictly Sparse Source Modelling Using Heavy-tailed Distributions for DCT Coefficients

Masoud Aramideh
Electrical Engineering Department, Engineering Faculty
Shahid Chamran Universiy of Ahvaz
Ahvaz, Iran
m-aramideh@stu.scu.ac.ir

Ehsan Namjoo
Electrical Engineering Department, Engineering Faculty
Shahid Chamran Universiy of Ahvaz
Ahvaz, Iran
e.namjoo@scu.ac.ir

Mahdi Nooshyar
Electrical and Computer Engineering Department, Engineering Faculty
University of Mohaghegh Ardabili
Ardabil, Iran
nooshyar@uma.ac.ir

***Abstract*—In this study heavy-tailed distributions including Pareto, Weibull, Levy, and Log-normal distributions are proposed to model the DCT coefficients of real-world sources. It will be shown that the proposed heavy-tailed distributions are more accurate to model non-strictly sparse sources rather than Laplace and Cauchy distributions. Maximum Likelihood Estimation (MLE) method is utilized to parameter estimation of the proposed distributions. Finally, the Shannon Lower Bound (SLB) on the rate distortion function of each model is calculated, accordingly a framework is provided that could be used to assess the rate distortion performance of any coding scheme that is supposed to compress DCT coefficients as outcome of a sparse source.**

Keywords—lossy source coding, sparse sources, heavy-tailed distribution, maximum likelihood estimation, Pareto distribution, Weibull distribution, Levy distribution, Log-normal distribution

I. INTRODUCTION

Source coding is a fundamental aspect in today's digital communication systems. Depending on the type of the source, both lossy and lossless source coding scenarios are used. However, lossless compression techniques could not be performed on every source, e.g. continuous ones. Also in many applications, the bit budget might be limited. So in such cases, lossy compression is considered as a general solution to compress a source when there are restriction on available rate. Lossy compression implies distortion. Rate distortion theory determines the fundamental theoretical limits[1]. Given a distortion measure and a source distribution, rate distortion function determines the minimum expected distortion to achieve a specific rate. Every data compression scheme includes two main phases: *modeling* and *coding* [2]. *Modeling* focuses on the behavior of the source. Many studies have been done to model DCT coefficients for still and moving images. For instance, in [3] it is argued that DCT coefficients must have Gaussian distribution according to the central limit theorem, while in [4] it is shown that Laplace and Gamma distributions may model DCT coefficients better than Gaussian distribution. In [5] a zero mean Cauchy distribution is used to model DCT coefficients. Also it is stated that a truncated Cauchy Probability Density Function (PDF) presents a proper description of the DCT coefficient [6]. A Symmetric Normal Inverse Gaussian (SNIG) probability distribution function is also proposed in [7] to model the DCT coefficient of still images. In [8] a comparison between generalized Gaussian density function and Laplace distribution in image compression is made. Some other studies focus on estimating models' parameters for DCT coefficients. For example in [9] an approach called *one moment method* is used to estimate the shape parameter of the generalized Gaussian distribution with application to JPEG reconstruction. Recently some new models for DCT coefficients including Transparent Composite Model (TCM) and Synthesized Laplacian Distribution (SynLD) are introduced in [10] and [11] respectively.

After finding the appropriate model, in *coding* phase, figuring out the fundamental limits of lossy compression is the next important step to focus on. Assessing the performance of any lossy compression method depends on finding precise appropriate coding bounds. Also figuring out the rate distortion function of an information source is vital; because according to the source-channel separation theorem, transmitting and reconstructing an information source with rate-distortion function $R(D)$ through a channel with capacity C, is only possible when $C > R(D)$.

due to the fast development of compressive sensing [12], sparse sources have found a special place in both signal processing and compression studies. Sparsity is a key fundamental property of a sparse source that make it well prepared to be compressively coded. Sparse sources are classified into two classes: strictly and non-strictly sparse sources. Strictly sparse sources produces zero with probability close to one, while non-strictly sparse sources produce some zero-mean small values instead of zero [13]. Examples of sparse sources are binary sparse sources, Bernoulli-Gaussian, and zero-mean Gaussian mixtures [13]. Also in [13] it is shown that heavy-tailed PDFs are capable of modelling non-strictly sparse sources.

Wavelet as a well-known sparsifying transform is widely used in a vast areas of studies and applications. Rate distortion behavior of the coefficients of Wavelet transform has been investigated in [13], while not enough attention have been paid to investigate the rate distortion behavior of discrete cosine transform (DCT) coefficients. According to the fact that DCT is efficiently used in image and video coding, it is worth it to pay a special attention to model DCT coefficients. Consequently, this work could shade a light to a wide area of research on compressive sampling and lossy source coding scenarios based on DCT.

In this paper we have tried to model DCT coefficients using heavy-tailed distributions including Pareto, Weibull, Levy, and Log-normal distributions. To this aim plenty of DCT coefficients from real world sources including still images and video sequences have been gathered. After applying DCT on these sources, a suitable heavy-tailed distribution is fitted and the parameters of the distribution is estimated. The chi-square goodness of fit statistic (χ^2) [14] is

978-1-7281-1508-5/19/$31.00 ©2019 IEEE

Authorized licensed use limited to: University of Limerick. Downloaded on August 13,2024 at 20:35:42 UTC from IEEE Xplore. Restrictions apply.

used to assess the performance of the proposed distributions to model DCT coefficients. Finally by achieving proper models, a lower bound on rate distortion function for each model is calculated.

This paper is managed as follows. In section II rate distortion theory is stated. In section III the heavy-tailed distributions are introduced as non-strictly sparse source models and the models' parameters are estimated using maximum likelihood estimation approach. In section IV rate distortion bound (SLB) of each model is calculated. Section V is dedicated to results and conclusions. Finally the paper is concluded in Section VI.

II. RATE DISTORTION THEORY

Rate distortion function of a source X determines the minimum rate R that is needed to describe X when the expected distortion is kept less than or equal to D. rate distortion function is written as below [15].

$$R(D) = \min_{f(\hat{x}|x):\, Ed(x,\hat{x})\leq D} I(X;\, \hat{X}) \tag{1}$$

where $\hat{X}$ is the quantized version of X, $I(X;\, \hat{X})$ is the mutual information between X and $\hat{X}$, $f(\hat{x}|x)$ is the PDF of $\hat{X}$ given X, and $Ed(x,\hat{x})$ is the expected distortion measure. Equivalently the goal is to minimize the distortion introduced in reconstructing $\hat{X}$ given X; i.e. the expected distortion measure have to be bounded by D as stated in (2).

$$Ed(x,\hat{x}) = \sum_{(x,\hat{x})} f(x,\hat{x})d(x,\hat{x}) \leq D \tag{2}$$

Also, (2) could be rewritten as (3).

$$Ed(x,\hat{x}) = \sum_{(x,\hat{x})} f(x)f(\hat{x}|x)d(x,\hat{x}) \leq D \tag{3}$$

So the rate distortion function stated in (1) is written as (4).

$$R(D) = \min_{f(\hat{x}|x):\, \sum_{(x,\hat{x})} f(x)f(\hat{x}|x)d(x,\hat{x})\leq D} I(X;\, \hat{X}) \tag{4}$$

According to the nature of different sources, different distortion measures are used to describe the rate distortion function. The distortion measure utilized in this research is mean square error, as presented in (5).

$$d(x,\hat{x}) = (x - \hat{x})^2 \tag{5}$$

For a continuous source with mean square error as the distortion measure, a famous lower bound is stated as (6).

$$R(D) \geq R_{SLB}(D) = h(X) - \frac{1}{2}\log(2\pi e D) \tag{6}$$

This lower bound is known as the SLB in which $h(X) = -\int_{-\infty}^{+\infty} f(x)\log\big(f(x)\big)dx$ is the differential entropy of the source X with PDF $f(x)$. The SLB on the rate distortion function is used in section V to assess the rate distortion performance of our proposed heavy-tailed models.

III. PROPOSED HEAVY-TAILED MODELS FOR NON-STRICTLY SPARSE SOURCES

In order to find a well-suited model for DCT coefficients as a non-strictly sparse source, a set of real world samples is collected from BiodID dataset which contains 1500 face images in different brightness condition [16] (Fig. 1), and QCIF video sequences including ice, carphone, and foreman (Fig. 2). Performing DCT on images and video frames, forms the observation set that is used to estimate non-strictly sparse models. In this case, after normalizing the observations, the normalized histogram is attained. Finally, parameters of heavy-tailed distributions including Pareto, Weibull, Levy, and Log-normal are estimated according to the normalized histogram. In the following part, the detailed procedure of parameter estimation based on MLE approach is given.

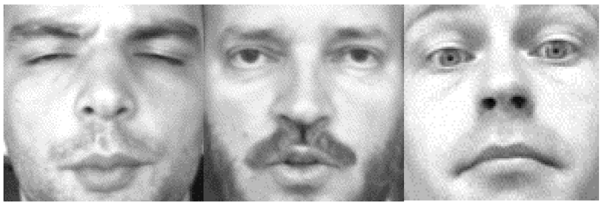

Fig. 1. Example of some frontal view face images from BioID

(a) (b) (c)

Fig. 2. Example of some famous video sequences: a) foreman b) ice c) carphone

In this study, it is assumed that the observations are drawn from a known i.i.d. sequence with unknown parameters; i.e. MLE is actually performed on N i.i.d. observations with joint distribution $f(x_0, x_1, \dots, x_{N-1})$ with unknown parameters.

A. Parameter Estimation of Pareto Distribution

Pareto distribution as a heavy-tailed PDF is our first choice to model DCT coefficients. The PDF of Pareto distribution for $x \geq x_m$ is presented in (7). In this section, the MLE approach is utilized to estimate the parameter α.

$$f(x) = \frac{\alpha\, x_m^{\alpha}}{x^{\alpha+1}} \tag{7}$$

The PDF of N observation is given in (8).

$$f(x_0, x_1, \dots, x_{N-1}) = \prod_{i=0}^{N-1} f(x_i) \tag{8}$$

Then the likelihood function is written as (9).

$$\prod_{i=0}^{N-1} f(x_i) = \prod_{i=0}^{N-1} \frac{\alpha\, x_m^{\alpha}}{x_i^{\alpha+1}} = \frac{\alpha^N\, x_m^{\alpha N}}{\prod_{i=0}^{N-1} x_i^{\alpha+1}} \tag{9}$$

Applying a natural logarithm to (9), log-likelihood function is obtained as (10).

$$ln\big(\prod_{i=0}^{N-1} f(x_i)\big) = \ln(\alpha^N x_m^{\alpha N}) - \ln(\prod_{i=0}^{N-1} x_i^{\alpha+1}) \tag{10}$$

By some simplifications, the log-likelihood function is rewritten as (11).

$$ln\big(\prod_{i=0}^{N-1} f(x_i)\big) = N\ln\alpha + \alpha N \ln x_m - \sum_{i=0}^{N-1} ln x_i \tag{11}$$

Authorized licensed use limited to: University of Limerick. Downloaded on August 13,2024 at 20:35:42 UTC from IEEE Xplore. Restrictions apply.

Now, to find the estimation of α, the log-likelihood function should be maximized with respect to the parameter α. To this aim, the derivative of log-likelihood function is set equal to zero as shown in (12).

$$\frac{\partial ln\left(\prod_{i=0}^{N-1} f(x_i)\right)}{\partial \alpha} = 0 \qquad (12)$$

The result reveals:

$$\frac{N}{\alpha} + N \ln x_m - \sum_{i=0}^{N-1} ln x_i = 0 \qquad (13)$$

Finally, the ML estimation of α is obtained as (14).

$$\hat{\alpha}_{ML} == \frac{N}{\sum_{i=0}^{N-1}(ln x_i) - N ln x_m} \qquad (14)$$

Or equivalently,

$$\hat{\alpha}_{ML} = \frac{N}{\sum_{i=0}^{N-1}(ln x_i - ln x_m)}. \qquad (15).$$

$\hat{\alpha}_{ML}$ is the maximum likelihood estimator for the parameter α.

B. Parameter Estimation of Levy Distribution

The second heavy-tailed distribution that we have utilized to model DCT coefficients is Levy distribution with PDF presented in (17). The Levy distribution is defined for $x > \mu$.

$$f(x) = \begin{cases} \sqrt{\frac{C}{2\pi}} \frac{e^{-\frac{C}{2(x-\mu)}}}{(x-\mu)^{3/2}}, & x \geq \mu \\ 0, & otw \end{cases} \qquad (17)$$

Since DCT coefficients are centralized around zero, it is assumed that $\mu = 0$. So, equation (17) is simplified to (18).

$$f(x;\, \mu = 0) = \begin{cases} \sqrt{\frac{C}{2\pi}} \frac{e^{-\frac{C}{2(x)}}}{(x)^{3/2}}, & x \geq 0 \\ 0, & otw \end{cases} \qquad (18)$$

Again, because of the i.i.d. assumption, the PDF of N observations is given in (19) and the estimating process is given in equations (19)- (24).

$$f(x_0, x_1, \dots, x_{N-1}) = \prod_{i=0}^{N-1} f(x_i) = \prod_{i=0}^{N-1} \sqrt{\frac{C}{2\pi}} \frac{e^{-\frac{C}{2x_i}}}{x_i^{3/2}} = \left(\frac{C}{2\pi}\right)^{N/2} \left(e^{-\sum_{i=0}^{N-1} \frac{C}{2x_i}}\right) \prod_{i=0}^{N-1} \frac{1}{x_i^{3/2}} \qquad (19)$$

$$ln\left(\prod_{i=0}^{N-1} f(x_i)\right) = \ln\left(\frac{C}{2\pi}\right)^{N/2} + \ln\left(e^{-\sum_{i=0}^{N-1} \frac{C}{2x_i}}\right) + \ln \prod_{i=0}^{N-1} \frac{1}{x_i^{3/2}} \qquad (20)$$

$$ln\left(\prod_{i=0}^{N-1} f(x_i)\right) = \frac{N}{2C} \ln\left(\frac{C}{2\pi}\right) - \sum_{i=0}^{N-1} \frac{C}{2x_i} + \sum_{i=0}^{N-1} \frac{1}{x_i^{3/2}} \qquad (21)$$

$$\frac{\partial ln\left(\prod_{i=0}^{N-1} f(x_i)\right)}{\partial C} = 0 \qquad (22)$$

$$\frac{N}{2C} - \sum_{i=0}^{N-1} \frac{1}{2x_i} = 0 \qquad (23)$$

$$\hat{C}_{ML} = \frac{N}{\sum_{i=0}^{N-1} \frac{1}{x_i}} \qquad (24)$$

$\hat{C}_{ML}$ is the maximum likelihood estimator for the parameter C.

C. Parameter Estimation of Weibull Distribution

Weibull distribution is the third heavy-tailed distribution, we have aimed to use as a model to represent DCT coefficients. The distribution is given in (24). Weibull distribution have two parameters λ and k.

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad (24)$$

The likelihood function for this distribution is given in (25).

$$f(x_0, x_1, \dots, x_{N-1}) = \prod_{i=0}^{N-1} f(x_i) = \left(\frac{k}{\lambda^k}\right)^N e^{-\sum_{i=0}^{N-1} \left(\frac{x_i}{\lambda}\right)^k} \prod_{i=0}^{N-1} x_i^{k-1} \qquad (25)$$

The log-likelihood function is obtained as in (26).

$$ln\left(\prod_{i=0}^{N-1} f(x_i)\right) = \ln\left(\frac{k}{\lambda^k}\right)^N + \ln\left(e^{-\sum_{i=0}^{N-1} \left(\frac{x_i}{\lambda}\right)^k}\right) + \ln \prod_{i=0}^{N-1} x_i^{k-1} \qquad (26)$$

The log likelihood function for this distribution is written as (27).

$$ln\left(\prod_{i=0}^{N-1} f(x_i)\right) = N \ln k - Nk \ln \lambda - \sum_{i=0}^{N-1} \left(\frac{x_i}{\lambda}\right)^k + \sum_{i=0}^{N-1} \ln x_i^{k-1} \qquad (27)$$

The estimated value for λ is obtained by putting the derivative of (27) with respect to λ, equal to zero.

$$\frac{\partial ln\left(\prod_{i=0}^{N-1} f(x_i)\right)}{\partial \lambda} = 0 \qquad (28)$$

By some simplification, we have:

$$-\frac{Nk}{\lambda} - (-k)\lambda^{-k-1} \sum_{i=0}^{N-1} x_i^k = 0. \qquad (29)$$

By solving (29) for λ, the *ML* estimator of the parameter λ, ($\hat{\lambda}_{ML}$) is calculated as in (30).

$$\hat{\lambda}_{ML} = \left(\frac{\sum_{i=0}^{N-1} x_i^k}{N}\right)^{1/k} \qquad (30)$$

In order to find the ML estimator of the other parameter, k, derivative of (27) with respect to k is set to zero.

$$\frac{\partial ln\left(\prod_{i=0}^{N-1} f(x_i)\right)}{\partial k} = 0 \qquad (31)$$

The resultant is given in (32).

$$\frac{N}{k} - N \ln \lambda - \sum_{i=0}^{N-1} \left(\frac{x_i}{\lambda}\right)^k \ln\left(\frac{x_i}{\lambda}\right) + \sum_{i=0}^{N-1} \ln x_i = 0 \qquad (32)$$

The recent equation should be solved for the parameter k, But there is no closed form solution and k should be calculated iteratively.

Authorized licensed use limited to: University of Limerick. Downloaded on August 13,2024 at 20:35:42 UTC from IEEE Xplore. Restrictions apply.

D. Parameter Estimation of Log-normal Distribution

The last heavy-tailed distribution we have chosen in order to model DCT coefficients is the Log-normal distribution with PDF presented in (33). Just like the Gaussian distribution, the parameters of Log-normal distribution are μ and σ^2.

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \tag{33}$$

Similar to aforementioned cases, the likelihood function can be written as (34).

$$f(x_0, x_1, \ldots, x_{N-1}) = \prod_{i=0}^{N-1} f(x_i) = \frac{1}{(\sigma^2)^{N/2}\sqrt{2\pi}\prod_{i=0}^{N-1} x_i}\, e^{-\sum_{i=0}^{N-1}\frac{(\ln x_i-\mu)^2}{2\sigma^2}} \tag{34}$$

The log-likelihood function is written as (35)

$$\ln\left(\prod_{i=0}^{N-1} f(x_i)\right) = -\ln\prod_{i=0}^{N-1} x_i - \frac{N}{2}\ln\sigma^2 - N\ln\sqrt{2\pi} - \sum_{i=0}^{N-1}\frac{(\ln x_i-\mu)^2}{2\sigma^2} \tag{35}$$

Similarly, in order to calculate maximum likelihood estimator of μ, log-likelihood function should be maximized by taking derivative of (35) respect to μ and set the result equal to zero as shown in (36).

$$\frac{\partial \ln\left(\prod_{i=0}^{N-1} f(x_i)\right)}{\partial\mu} = 0 \tag{36}$$

Therefore, we have:

$$-\sum_{i=0}^{N-1} - \frac{2(\ln x_i-\mu)}{2\sigma^2} = 0. \tag{37}$$

The maximum likelihood estimator of μ ($\hat{\mu}_{ML}$) is obtained as in (38).

$$\hat{\mu}_{ML} = \frac{1}{N}\sum_{i=0}^{N-1}\ln x_i \tag{38}$$

The same procedure is performed to calculate the ML estimator of σ^2 by maximizing the log-likelihood function respect to σ^2 and set the result equal to zero, as in (39).

$$\frac{\partial \ln\left(\prod_{i=0}^{N-1} f(x_i)\right)}{\partial\sigma^2} = 0 \tag{39}$$

The result is presented in (40).

$$-\frac{N}{2\sigma^2} - \frac{-1}{(\sigma^2)^2}\sum_{i=0}^{N-1}\frac{(\ln x_i-\mu)^2}{2} = 0 \tag{40}$$

Finally, by solving (40) for σ^2, the desired ML estimator, $\widehat{\sigma^2}_{ML}$, is obtained as in (41).

$$\widehat{\sigma^2}_{ML} = \frac{\sum_{i=0}^{N-1}((\ln x_i-\hat{\mu})^2}{N} \tag{41}$$

IV. Lower Bound on Rate Distortion Function of the Heavy-Tailed Models

In this section a lower bound on rate distortion function for each utilized continuous heavy-tailed source is calculated using the SLB formula given in (6). The procedure is straightforward. Actually, in this case, to calculate a lower bound, the only thing needed to be done is to calculate the differential entropy of each model. Let $h_{Pareto}(X), h_{Levy}(X), h_{Weibull}(X)$, and $h_{Log-normal}(X)$ presents the differential entropy of Pareto, Levy, Weibull, and Log-normal distributions respectively [17], then the exact formulas for their differential entropy is stated in (42)- (45).

$$h_{Pareto}(X) = log\left(\left(\frac{x_m}{\alpha}\right)e^{1+\frac{1}{\alpha}}\right) \tag{42}$$

$$h_{Levy}(X) = \frac{1+3\gamma+log(16\,\pi C^2)}{2} \tag{43}$$

$$h_{Weibull}(X) = log\left(\frac{\lambda}{k}e^{1+\frac{k-1}{k}\gamma}\right) \tag{44}$$

$$h_{Log-normal}(X) = log(\sigma\, e^{\mu+\frac{1}{2}}\sqrt{2\,\pi}) \tag{45}$$

The parameter γ in (44) is the Euler's constant.

V. Results

In this section according to calculations presented in sections III and IV, the parameters of the proposed PDF's i.e. Pareto, Weibull, Levy, and Log-normal distributions are estimated for DCT coefficients of real data including samples collected from BiodID dataset, ice, carphone, and foreman QCIF video sequences, respectively. As shown in Fig. 3- Fig. 6, for every dataset composed of normalized histogram of DCT coefficients, the proposed heavy-tailed models are compared with Laplace distribution [4] and a zero mean Cauchy distribution [5]. The PDF of Laplace distribution and zero mean Cauchy distribution is given in (45) and (46) respectively.

$$f(x) = \frac{1}{2b}e^{\left(-\frac{|x-\mu|}{b}\right)} \tag{45}$$

$$f(x) = \frac{1}{\pi}\frac{\mu}{\mu^2+x^2} \tag{46}$$

TABLE I. presents the numerical estimation results for the parameters of the proposed heavy-tailed, Laplace, and Cauchy distributions when they are used to model DCT coefficients for each dataset. A distribution with less χ^2 test value is more accurate to model the data. According to this test, it is seen that all the proposed heavy-tailed distributions provide a better fit for DCT coefficients histogram in comparison to Laplace and Cauchy distributions. In addition, as mentioned in the literature, Cauchy distribution describe DCT coefficients histogram more precisely than Laplace distribution.

DCT coefficients include both positive and negative values, while all proposed models are one-sided distributions. This seems to be a challenge for our proposed models, but in fact, it could be handled easily by extending the models to support both positive and negative values just using the simple trick, $g(x) = 0.5[f(x) + f(-x)]$. In that case any one-sided distribution $f(x)$ could be easily extended to a symmetric two-sided distribution $g(x)$. Also in this study only the DCT coefficients with magnitude in the range [0, 50] are shown just for convenience, while actual range of DCT coefficients is much wider than the mentioned range.

Fig. 7 shows SLB on rate distortion function for estimated models. This figure provides a fundamental information theoretic limits on lossy compression of DCT coefficients.

Authorized licensed use limited to: University of Limerick. Downloaded on August 13,2024 at 20:35:42 UTC from IEEE Xplore. Restrictions apply.

TABLE I. ESTIMATED PROPOSED MODELS VS. CAUCHY AND LAPLACE DISTRIBUTIONS

Dataset	***Distributions***	***Estimated Parameters***	***Chi square goodness of fit statistic (χ^2)***
BioID	**Pareto**	$\hat{\alpha}_{ML} = 0.9558$ $X_m = 0.5$	$\mathbf{2.2511 \times 10^7}$
	Cauchy	$\hat{\mu} = 1.4749$	1.1670×10^8
	Laplace	$\hat{b} = 4.4155$ $\hat{\mu} = 0$	1.2624×10^8
ice	**Weibull**	$\hat{\lambda}_{ML} = 7.7612$ $\hat{k}_{ML} = 0.8651$	$\mathbf{2.1779 \times 10^6}$
	Cauchy	$\hat{\mu} = 4.4810$	4.1931×10^6
	Laplace	$\hat{b} = 8.3812$ $\hat{\mu} = 0.0015$	5.0220×10^6
carphone	**Levy**	$\hat{C}_{ML} = 3.8470$	$\mathbf{7.4596 \times 10^5}$
	Cauchy	$\hat{\mu} = 4.1582$	1.7956×10^6
	Laplace	$\hat{b} = 7.5094$ $\hat{\mu} = -0.0137$	2.1895×10^6
foreman	**Log-normal**	$\hat{\mu}_{ML} = 1.2305$ $\widehat{\sigma^2}_{ML} = 2.0225$	$\mathbf{3.0087 \times 10^6}$
	Cauchy	$\hat{\mu} = 3.7776$	2.1696×10^7
	Laplace	$\hat{b} = 7.3566$ $\hat{\mu} \approx 0$	2.4093×10^7

Any quantization or source coding scheme provides better results when it gets closer to the corresponded SLB. It means that it provides better rate distortion performance. As seen in Fig 7, the histogram of DCT coefficients with faster drop, is more compressible and its corresponding rate distortion curve is placed below the other histograms with slower decline. Here DCT coefficients' histogram of BioID, modeled by Pareto distribution, drops much faster than the other histograms. As a result, its corresponding SLB curve is placed far below the other SLB curves.

To make the comparison fair enough, similar curves (Fig. 8) have been derived for different datasets but the same model is chosen. In Fig.8 the Weibull distribution is fitted to represents DCT coefficients for different datasets. The estimation parameters used to derived curves depicted in Fig. 8, is provided in TABEL. II. Once more, as expected, the SLB curves for BioID dataset is placed far below the other similar curves that emphasis on the fact that histogram of data with faster drop is more compressible than one with slower decline.

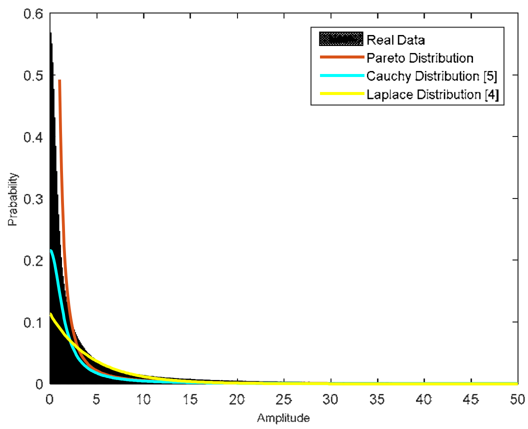

Fig. 3. Normalized DCT coeffients histogram (Real Data) for BioID dataset, and its corresponding fitted Pareto, Cauchy, and Laplace distributions.

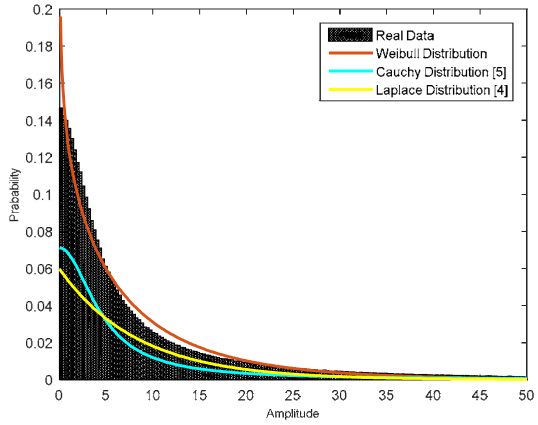

Fig. 4. Normalized DCT coeffients histogram (Real Data) for ice sequence dataset, and its corresponding fitted Weibull, Cauchy, and Laplace distributions

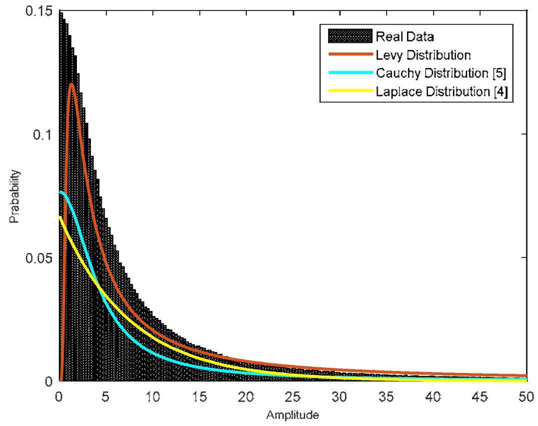

Fig. 5. Normalized DCT coeffients histogram (Real Data) for carphone sequence dataset, and its corresponding fitted Levy, Cauchy, and Laplace distributions

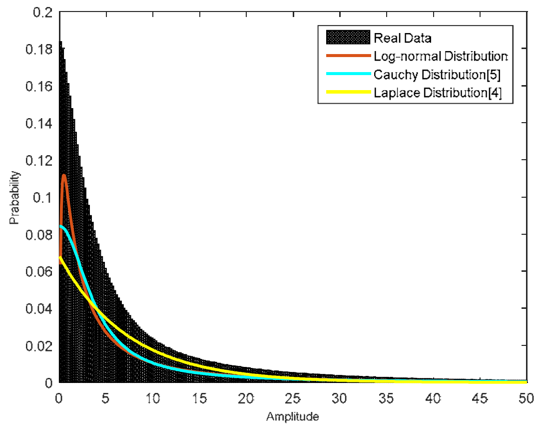

Fig. 6. Normalized DCT coeffients histogram (Real Data) for foreman sequence dataset, and its corresponding fitted Log-normal, Cauchy, and Laplace distributions

Authorized licensed use limited to: University of Limerick. Downloaded on August 13,2024 at 20:35:42 UTC from IEEE Xplore. Restrictions apply.

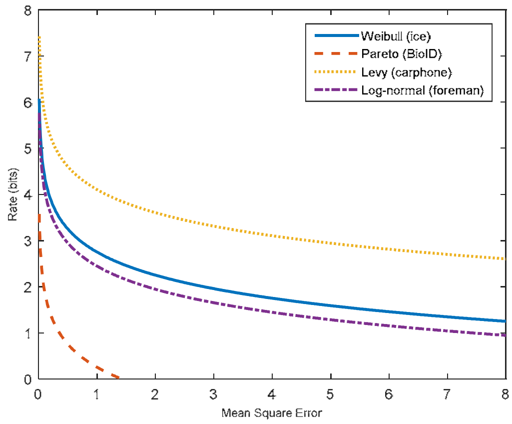

Fig. 7. SLB for differeent datasets and distributions

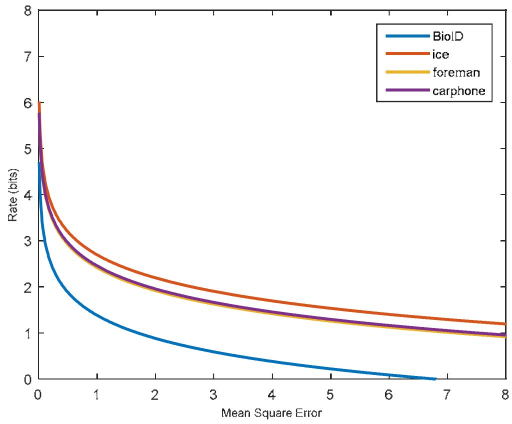

Fig. 8. SLB for differeent datasets but the same distribution (Weibull Distribution)

TABLE II. ESTIMATED PARAMETERS OF WEIBULL DISTRIBUTION FOR DIFFERENT DATA

Dataset	**Distribution**	***Estimated Parameters***
BioID	Weibull	$\hat{\lambda}_{ML} = 3.5025$ $\hat{k}_{ML} = 0.6225$
ice	Weibull	$\hat{\lambda}_{ML} = 9.0337$ $\hat{k}_{ML} = 0.7783$
carphone	Weibull	$\hat{\lambda}_{ML} = 7.7380$ $\hat{k}_{ML} = 0.8116$
foreman	Weibull	$\hat{\lambda}_{ML} = 7.4292$ $\hat{k}_{ML} = 0.7680$

VI. CONCLUSION

In this paper four heavy-tailed distributions including Pareto, Weibull, Levy, and Log-normal distributions, which were not introduced in the literature before, are proposed to model DCT coefficients of still images and video sequences as non-strictly sparse sources. The MLE of the parameters of the proposed models are derived. Numerical results show that the proposed models provide a better fit of DCT coefficients histogram in comparison to the well-known Laplace and Cauchy distributions according to the χ^2 test. Furthermore, fundamental limits of lossy compression is investigated for the DCT coefficients of several datasets of still images and video sequences. Comparing the lower bound on the rate distortion function of studied data, shows that data with faster histogram drop are more compressible rather than one with slower decline.

REFERENCES

[1] C. E. Shannon, “A Mathematical Theory of Communication,” Bell Labs Tech. J., vol. 27, no. April 1928, pp. 379–423, 1948.

[2] K. Sayood, Introduction to data compression. Morgan Kaufmann, 2017.

[3] W.-H. Chen and C. Smith, “Adaptive coding of monochrome and color images,” IEEE Trans. Commun., vol. 25, no. 11, pp. 1285–1292, 1977.

[4] T. Eude, R. Grisel, H. Cherifi, and R. Debrie, “On the distribution of the DCT coefficients,” in Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on, 1994, p. V-365.

[5] N. Kamaci, Y. Altunbasak, and R. M. Mersereau, “Frame bit allocation for the H. 264/AVC video coder via Cauchy-density-based rate and distortion models,” IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 8, pp. 994–1006, 2005.

[6] J. Eggerton and M. Srinath, “A visually weighted quantization scheme for image bandwidth compression at low data rates,” IEEE Trans. Commun., vol. 34, no. 8, pp. 840–847, 1986.

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

[8] R. L. Joshi and T. R. Fischer, “Comparison of generalized Gaussian and Laplacian modeling in DCT image coding,” IEEE Signal Process. Lett., vol. 2, no. 5, pp. 81–82, 1995.

[9] R. Krupiński and J. Purczyński, “Modeling the distribution of DCT coefficients for JPEG reconstruction,” Signal Process. Image Commun., vol. 22, no. 5, pp. 439–447, 2007.

[10] E.-H. Yang, X. Yu, J. Meng, and C. Sun, “Transparent composite model for dct coefficients: Design and analysis,” IEEE Trans. Image Process., vol. 23, no. 3, pp. 1303–1316, 2014.

[11] W. Gao, S. Kwong, H. Yuan, and X. Wang, “DCT Coefficient Distribution Modeling and Quality Dependency Analysis Based Frame-Level Bit Allocation for HEVC.,” IEEE Trans. Circuits Syst. Video Techn., vol. 26, no. 1, pp. 139–153, 2016.

[12] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. theory, vol. 52, no. 4, pp. 1289–1306, 2006.

[13] C. Weidmann and M. Vetterli, “Rate distortion behavior of sparse sources,” IEEE Trans. Inf. theory, vol. 58, no. 8, pp. 4969–4992, 2012.

[14] F. Müller, “Distribution shape of two-dimensional DCT coefficients of natural images,” Electron. Lett., vol. 29, no. 22, pp. 1935–1936, 1993.

[15] T. M. Cover and J. A. Thomas, “Elements of information theory.” Hoboken, NJ: Wiley-Interscience, pp. 301–346, 2006.

[16] A. G. BioID, “BioID face Ddatabase.” 2001.

[17] J. V. Michalowicz, J. M. Nichols, and F. Bucholtz, Handbook of differential entropy. Crc Press, 2013.

Authorized licensed use limited to: University of Limerick. Downloaded on August 13,2024 at 20:35:42 UTC from IEEE Xplore. Restrictions apply.