

一般論文

JPEG非可逆モードにおける統計モデルの改善

黒 木 祥 光^{*1}

Improvement of the Statistical Models in the JPEG Lossy Mode

Yoshimitsu KUROIKI

The JPEG statistical models in the lossy mode specify the procedures for converting Discrete Cosine Transform (DCT) coefficients into binary strings and the context modeling in the case where the binary arithmetic coder called QM-coder is employed as an entropy coder. The JPEG lossy mode establishes two statistical models, one for prediction residuals of the DC coefficients and the other for the AC coefficients. In this paper, we redesign these two models by taking account of the coefficients' distribution. We confirm that a Laplacian distribution is appropriate for both the DC coefficients and the AC coefficients through Kolmogorov-Smirnov (KS) test; consequently, we propose statistical models that fit Laplacian distributions. By adopting the proposed statistical models in lieu of the conventional models, the number of the states decreases from 294 to 210 and the compression performance on several test images including super high definition images improves by 0.01 to 1.48%.

1 はじめに

静止画像符号化の国際標準JPEG (Joint Photographic Experts Group) では、必須機能としてDCT (Discrete Cosine Transform) を用いた非可逆符号化を採用している^[1]。DCT係数は量子化された後、ハフマン符号もしくは2値算術符号であるQM-coderによって符号化される。QM-coderは動的適応性が高く、符号化しながらダイナミックに符号化パラメータを変化させるため、任意の情報源に対し、常に高効率な符号化が可能である。また、既に符号化済みの情報によって符号化すべき情報源を分離するマルコフモデル符号化も容易に実現され、更なる符号化効率の改善が可能である。従って、QM-coderはJPEGのみならず2値画像符号化の国際標準JBIG (Joint Bi-level Image experts Group)^[2]においても採用されている。しかし、QM-coderを用いて多値情報源を符号化する際、2値系列の信号に変換する必要がある。2値化の手法として、レベルプレーン展開等の手法があるが、一般に各判定を独立した2値情報源として取り扱わない限り、情報量の増加は避けられない。また、マルコフモデル符号化を行う場合、各判定を更に独立した情報源に分離し、符号化パラメータを記憶す

る必要がある。なお、QM-coderを使用する際、独立と見なされる個々の情報源を一般に状態と呼ぶ。QM-coderは2値シンボルの出願確率の初期値をそれぞれ0.5とし、符号化しながら2値シンボルの出現確率を学習していくため、状態において出現するシンボル数が少ない場合、学習が十分に行われず、却って符号化効率の低下が生じる。同時に、状態毎に用意すべき記憶容量の増大も問題となる。この解決策として、2値シンボルの出現確率が近い状態を選出し、統合する手法が考えられる。JPEGの統計モデルは、レベルプレーン展開に準じた2値化と、マルコフモデル条件付けを記述したものであり、高効率なマルコフモデル符号化及び状態統合の手法と言えるが、ハフマン符号との親和性を考慮しているため、その最適性には疑問が残る。本論文はDCT係数の分布を参考とし、JPEGより状態数が少なく、且つ符号量も低減可能な統計モデルを提案する。

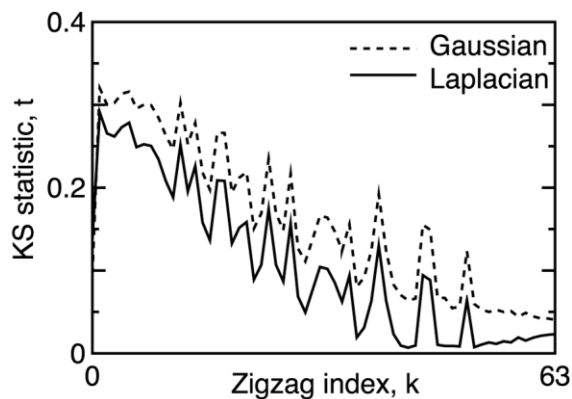
2 DCT係数の分布

JPEGでは、視覚特性を利用するため、周波数毎に量子化幅を可変としている。また、一般にDCT係数の分散は周波数毎に異なると云われており、各周波数を独立した情報源と考えるのが適当と思われる。各周波数成分は8×8画素から成る矩形ブロック内の画素値の線形和であるため、中心極限定理により、

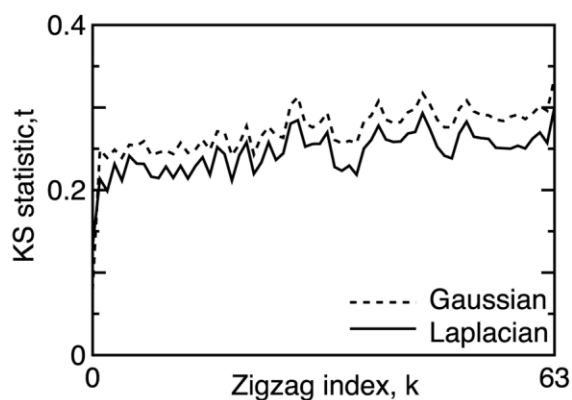
平成17年1月20日 受理

^{*1} 久留米工業高等専門学校制御情報工学科

Copyright 2005 国立久留米工業高等専門学校



(a) 原画像におけるKS統計量



(b) 解像度を下げた場合のKS統計量

図1：画像“PM001”のDCT係数に対するKS統計量

正規分布に従うと考えられる^[3]。しかし、ReiningerとGibsonは256×256画素、256階調の画像数種類に対してKS (Kolmogorov-Smirnov) 統計量を測定した結果、特に画素値の変化が大きな画像を除いた場合、低周波成分の分布は正規分布よりラプラス分布に近いという結果を得ている^[4]。Reiningerらを使用した画像は画素数が少なく、低周波成分のみの結果であったため、ここでは今後様々な分野での使用が期待される超高精細画像、および比較のため、超高精細画像の解像度をNTSC程度まで低下させた画像を用い、全ての周波数成分に対し、KS統計量を測定する。測定に用いる分布は上記の理由により、正規分布とラプラス分布の2種類とする。KS統計量は標本値 $X = (x_1, x_2, \dots, x_M)$ によって定義される標本累積分布関数 $F_X(x)$ と与えられた分布関数 $F(x)$ に対し、仮説 $F_X(x) \equiv F(x)$ を検定する際用いられる指標である。標本累積分布関数 $F_X(x)$ は次式で定義される。

$$F_X(x) = \begin{cases} 0, & x < x_{(1)} \\ n/M, & x_{(n)} \leq x < x_{(n+1)}, \\ & n = 1, 2, \dots, M-1 \\ 1, & x \geq x_{(M)} \end{cases} \quad (1)$$

ここで、 $x_{(n)}$ ($n=1, 2, \dots, M$) は標本 X を小さい順に並べたものである。KS統計量 t は次式で与えられる^[4, 6]。

$$t = \max_{i=1,2,\dots,M} |F_X(x_i) - F(x_i)| \quad (2)$$

従って、 t が小さいほど近似度が高いといえる。超高精細画像“PM001” (4,096×4,096画素、12bit/pel) に対するKS統計量の測定結果を図1に示す。図1(a)は原画像の結果である。図1(b)は原画像の画素数を512×512画素まで低下させた画像に対する結果である。低解像度の画像は原画像を8×8画素のブロックに分割し、各ブロックを平均化して作成した。図1の横軸はジグザグスキャンの順序 k を表し、実線は正規分布の、波線はラプラス分布のKS統計量をそれぞれ示している。DC係数 ($k=0$) はJPEGの符号化手法に従い、直前のブロックとの差信号を測定の対象とした。図1より、(a)、(b)双方において、DC係数を除く全ての係数でラプラス分布の方が正規分布よりKS統計量が小さい。また、原画像では高周波になるほどKS統計量は減少する傾向であるのに対し、低解像度の画像では高周波になるほど増加する結果となった。ここで、検定における有意点 c はサンプル数 M が多い場合、有意水準を α とすると、

$$c = \sqrt{\frac{1}{2M} \ln \frac{1}{\alpha}} \quad (3)$$

で近似されるため^[6]、 $\alpha=0.05$ (両側10%) における有意点は原画像の場合 $c=2.39 \times 10^{-3}$ 、低解像度の場合 $c=2.39 \times 10^{-2}$ である。従って、双方の画像において正規分布、ラプラス分布共にDCT係数の分布であるとは言えない。なお、その他数種類の画像に対し、KS統計量を測定したが、同様の結果であった。

3 JPEGの概略

JPEG非可逆モードでは、DC係数用とAC係数用、2種類の統計モデルを用意している。ここで、DCT係数は予め整数に量子化されていると仮定する。JPEGにおけるDCT係数の2値化はDC係数、AC係数に関わらず、図2に示す整数値 V の2値化に基づいている^[5]。ここで、中心に S_0 、 SS 等を示した菱形の記号は、流れ図の左に示した $V \neq 0$ 、 $V < 0$ 等の2値判定の項目を表し、偽の場合はQM-coderで0を符号化した後右方向へ、真の場合は1を符号化した後下方向の処理に進む。判定の項目が異なる場合、それらは全て独立した情報源と見なされる。以下、DC係数とAC係数それぞれの2値化および状態割り当ての手法について述べる。

3.1 DC係数の統計モデル

DC係数の統計モデルでは、現在のブロックと直前に符号化したブロックのDC係数との差が図2の値 V に相当する。各判定に対する条件付けの方法を表1に示す。DC係数の場合、まず、 V に対し、ゼロ判定 S_0 ($V \neq 0$?)が行われる。もし、判定が偽であれば、 $V = 0$ と決定されるため、符号化を

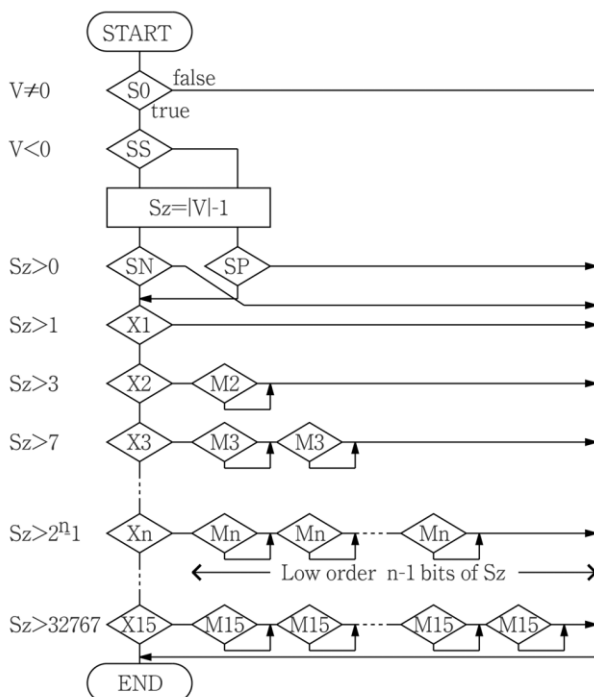


図2 JPEGにおける2判定と条件付け

表1: JPEGのDC係数に対する各2値判定の条件付け

判定項目	S_0	SS	SN	SP	X_1	X_n, M_n
状態数	5	5	5	5	1	それぞれ1

終了する。続いて、サイン判定 SS ($V < 0$?)が行われる。更に、 $S_z = |V| - 1$ に対し、グループ判定が順次行われる。ここで、グループ0の判定 ($S_z < 0$?)では V が負の場合は SN 、正の場合は SP と別個の項目となる。また、グループ2の判定 X_2 ($S_z > 3$?)以降で偽となる場合、グループ判定のみでは V の値が一意に決まらないため、グループ内のどの値であるかを示す付加ビットと呼ばれる2進数(判定 X_n であれば $n-1$ ビット)を最上位ビットから順に符号化する。判定の項目 S_0 、 SS 、 SN および SP は直前ブロックのDC差分により、それぞれ大(正)、小(正)、ゼロ近傍、小(負)、大(負)の5つの条件に分離される。ここで、大と小、および小とゼロ近傍を分離する閾値は正と負に関して対称であり、パラメータ L と U で与えられる。 X_1 以降の判定には、判定毎に1つの条件のみが割り当てられている。従って、DC係数における状態数は49となる。

3.2 AC係数の統計モデル

AC係数に用いられる2値判定の手順を図3に、各判定における条件付けを

表2に示す。図3の κ ($\kappa = 1, 2, \dots, 63$)はジグザグスキャンのインデックスを表す。AC係数の場合、係数の判定に先立ち、EOB (End of Block) か否かの判定 SE が行われる。これは $\kappa = 1$ および直前の係数が0以外の時のみ行われる。続いて係数の値 ZZ (κ)に対し、DC係数と同様の判定が行われる。つまり、図2内の“V decision tree”とは図2に示した判定手順の挿入を表す。

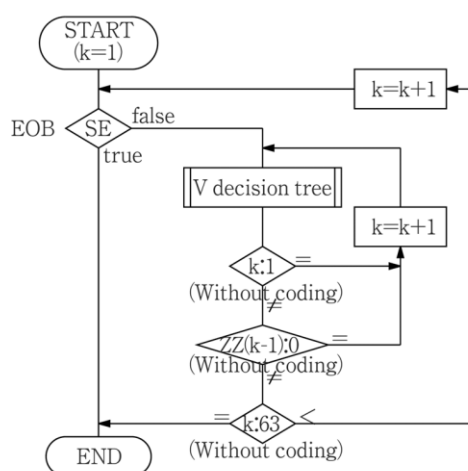


図3: JPEGにおけるAC係数の2判定と条件付け

表2: JPEGのAC係数に対する各2値判定の条件付け

判定項目	SE	S_0	SS	$SN/SP/X_1$	X_n, M_n
状態数	63	63	固定	63	それぞれ2

当然この場合、 $ZZ(\kappa)$ は V に替わる。各2値判定の条件付けは表2に示すように、DC係数と異なる。判定 SE , SO , $SN/SP/X_1$ に対しては、DCT係数の周波数毎にそれぞれ63状態が割り当てられる。ここで、 $SN/SP/X_1$ とは、これら3種類の判定が全て同じ状態に統合されていることを意味する。これは、AC係数の場合、正と負の出現確率がほぼ等しいため、判定 SN と SP の統計的振る舞いは同等である事に起因する。更に判定 X_1 も統合されているが、これは判定 X_1 における0, 1の出現確率も SN あるいは SP と同様であるとの経験則に従っている。判定 SS の符号化パラメータが固定されているが、これも正と負の出現確率がほぼ等しいことに由来する。 X_2 以降の判定に関してはグループ判定、付加ビットそれぞれに対し、低周波と高周波を分離するパラメータ K_x により、各判定は更に2つの状態に分離される。以上より、AC係数における状態数は245 (SS を除く)となる。

4 提案する統計モデル

2で述べたように、正規分布とラプラス分布、双方とも検定の結果、DCT係数の分布を十分に近似するとは言えない。しかし、本研究の目的は分布を正確に近似する関数を求めることではないため、DC係数の差分値およびAC係数は正規分布よりKS統計量の小さいラプラス分布に従うと仮定し、統計モデルの改善を行う。JPEGでは、判定 SO から X_1 まではレベルプレーン展開と同様であり、それぞれ別の状態を割り当てており、これ以上の改善は不可能と考えられる。 X_1 以降の判定は絶対値より1を引いた値 S_z を対象としているが、ラプラス分布はゼロを中心とする対称分布であるため、絶対値の判定および正と負の判定(ゼロ以外の場合)で表しても情報量は増加しない。一方、グループ2以降はグループ判定と付加ビットにより2値化を行っている。ここでは、このグループ判定と付加ビットによる2値化の改善を試みる。DC係数の差分あるいはAC係数の絶対値 y は非負の整数であるため、DCT係数の周波数成分 κ ($\kappa=0, 1, \dots, 63$)における y の出現確率 $P_\kappa(y)$ は次に示す幾何分布で近似される。

$$P_\kappa(y) = (1 - p_\kappa)p_\kappa^y \quad (y = 0, 1, 2, \dots) \quad (4)$$

パラメータ p_κ は $0 < p_\kappa < 1$ の範囲にあり、一般に周波数毎に異なる。ここで、幾何分布となるのは次の2つの理由に因る。即ち、ラプラス分布に従う信号の絶対値は指数分布となること^[7]、幾何分布におけるベルヌーイ試行の時間間隔 Δt に対し、

$\Delta t \rightarrow 0$ とした場合の極限が指数分布になること^[8]である。なお、この場合 Δt は量子化幅に相当する。従って、幾何分布に従う要因として、線形量子化であることも挙げられる。ところで、式(4)の確率変数 y は、出現確率をそれぞれ p_κ および $1 - p_\kappa$ で与えられる2項変数1, 0のベルヌーイ試行において、初めて0を出力するまでに出力した1の個数に他ならない。従って、このベルヌーイ試行はレベルプレーン展開の判定と等価と言える。ここで、幾何分布は記憶の欠如を有する分布であり、確率 p_κ と $1 - p_\kappa$ は同一周波数において常に一定である事を考慮すると、信号 y に対するレベルプレーン展開の各判定は同一の情報源と見なすことが可能である。従って、グループ判定と付加ビットを用いず、1つの状態に統合されたレベルプレーン展開で2値化する手法が適切と考えられるため、図2の代わりに図4の2値判定を提案する。図4において、 SO から SN/SP はJPEGと同じ手法であるが、 S_z に対する判定 Y_n ($n=1, 2, \dots$) は1つの状態に統合される。ここで、 SO , SN , SP に関する統合も考えられるが、JPEGの量子化表はミッドレッド型であり、絶対値に対してはゼロ付近が他の半分の幅であること、また、これらの判定は5つに条件付けされており、各条件下のDCT係数が全てラプラス分布に従うとは考えにくいいため、統合は行わないこととした。従って、状態数はDC係数の場合21である。一方、AC係数の場合、判定 Y_n は判定 SN/SP に統合される。ここで、状態 $SN/SP/Y_n$ は周波数毎に63の条件付けが行われ、状態数は189 (SS を除く)となる。この場合、JPEGで行っている低周波と高周波の分離は必要ない。

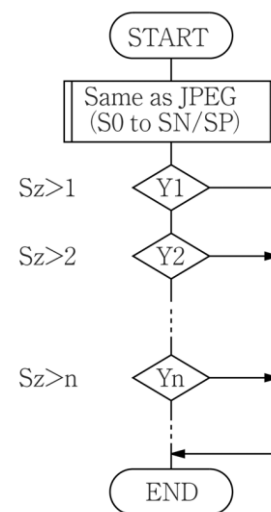


図4：提案する2値判定と条件付け

5 シミュレーション

超高精細画像およびITE標準画像に対する符号化シミュレーションの結果を表3に示す。表3に示した超高精細画像と同名の低解像度の画像は、KS統計量の測定と同様の手法により作成した。その他はITE標準画像であり、その書式は“肌色チャート”と“天気予報”が 512×480 画素、それ以外が 512×472 画素で構成され、全て256階調の輝度信号成分である。表3(a)は量子化表として、JPEG Annex Kに記載してある表を使用した結果である。一方、表3(b)はJPEGの量子化表を全周波数において $1/2$ とした場合の結果である。提案モデルにおいて、最初の幾つかの判定はJPEGと同じであるため、両モデルの差異は量子化幅が小さい場合により現れると考え、表3(b)の結果を示した。なお、各種パラメータはデフォルト値 $L=0$ 、 $U=1$ 、 $Kx=5$ （提案法では使用しない）とした。符号量にヘッダ情報は含まれていない。参考のため、PSNRを掲載しているが、これは何れの統計モデルを用いても同じ結果となる。表3より、提案した統計モデルを使用すると、JPEGに比べ、全ての画像において符号量を0.01%~1.48%低減することが確認される。また、表3の(a)と(b)を比較すると、量子化幅が小さい場合、殆どの画像において、本手法はより有効であることが分かる。

6 まとめ

本研究では、JPEG非可逆モードにおける統計モデルの改善を行った。DCT係数は理論上正規分布に、経験則ではラプラス分布に従うと云われている。双方の分布に対し、DCT係数全周波数成分のKS統計量を測定し結果、何れの分布にも従わないが、正規分布よりラプラス分布に近いとの結論を得た。DCT係数がラプラス分布に従う場合、量子化されたDCT係数の絶対値は幾何分布となるため、レベルプレーン展開で2値化した場合、各判定における2値信号の出現確率は等しい。この点に着目し、JPEGでは294あった状態数を210まで低減した統計モデルを提案した。符号化シミュレーションの結果、提案した統計モデルを使用すると、JPEGに比べ符号量が0.01%~1.48%減少する結果となった。今後はプログレッシブモードへの適用等について検討する予定である。

表3：符号量の比較

(a) JPEGの推奨する量子化表を使用した場合

		データ量 (bytes)		PSNR (dB)
		JPEG	提案法	
超高精細画像	PA022	395,166	393,037	39.06
	PM001	598,239	592,880	36.07
	SA001	654,295	648,479	37.05
低解像度	PA022	15,923	15,823	30.68
	PM001	25,175	25,052	26.29
	SA001	20,625	20,613	26.56
肌色チャート		10,996	10,977	32.36
ヘアーバンドの女性		14,640	14,591	30.98
天気予報		15,025	14,949	30.81
スイスの山村		41,096	40,907	23.27
チューリップ		36,859	36,608	24.34

(b) JPEGの $1/2$ 倍の量子化表を使用した場合

		データ量 (bytes)		PSNR (dB)
		JPEG	提案法	
超高精細画像	PA022	675,366	668,575	42.13
	PM001	990,738	985,117	38.81
	SA001	1,034,365	1,019,101	40.20
低解像度	PA022	25,383	25,174	33.15
	PM001	36,928	36,643	29.89
	SA001	34,698	34,566	29.37
肌色チャート		18,540	18,538	33.95
ヘアーバンドの女性		24,319	24,311	32.88
天気予報		23,109	22,985	33.19
スイスの山村		59,531	58,852	26.38
チューリップ		53,515	52,867	27.17

参考文献

- [1] ISO/IEC 10928-1: “Digital compression and coding of continuous-tone still images : Requirements and guideline”, (Feb. 1994)
- [2] ISO/IEC 11544: “Coded representation of picture and audio information—Progressive bi-level image compression”, (Dec. 1993)
- [3] A. N. Netravali and J. O. Lim: “Picture coding: A review”, Proc. IEEE, 68, 3, pp. 366—406, (Mar. 1980)
- [4] R. C. Reininger and J. D. Gibson: “Distributions of the Two-Dimensional DCT Coefficients for Images”, IEEE Trans. Commun., 6, pp. 835—839, (June 1983)
- [5] W. B. Pennebacker and J. L. Mitchell: “JPEG : still image data compression standard”, Van Nostrand Reinhold, (Sep. 1992)

-
- [6] 竹内 啓：確率分布と統計解析，日本規格協会（1975）
 - [7] 蓑谷 千風彦：すぐに役立つ統計分布，東京図書（1998）
 - [8] 小倉久直：確率過程入門，森北出版（1998）