

JPEG の裏側

ゆずねり 著

2025-05-31 版 Neln 発行

はじめに

JPEG の品質は「85 がよい」と、耳にしたことはありませんか？^{*1} そもそも、JPEG の「品質」という数値が何を意味しているのか、深く考えたことはあるでしょうか。

JPEG は 1992 年に誕生し、30 年以上たっています。今でも写真やウェブ画像など、私たちの身の回りで広く使われている画像フォーマットです。しかし、JPEG の圧縮の仕組みや「品質」の意味について、詳しく知っているでしょうか。

JPEG は、人間の視覚特性や画像の特徴、情報を効率的に表現するためのさまざまな符号化技術を、うまい具合に組み合わせてデータを削減します。その裏側には、長年にわたる研究と工夫が積み重ねられており、とてもおもしろい仕組みになっています。

本書では、なるべく予備知識がなくても理解できるように、JPEG の仕組みを解説します。

普段何気なく使っている JPEG 画像の舞台裏を、一緒にのぞいてみましょう。きっと、これまでとは違った視点で画像を見られるようになるはずです。

本書サポートサイトのご案内

本書のサポートサイトは次の URL になります。

<https://neln.net/b/nt03/>

免責事項

本書に記載された内容は、情報の提供のみを目的としています。したがって、本書を用いた開発、製作、運用は、必ずご自身の責任と判断によって行ってください。これらの情報による開発、製作、運用の結果について、著者はいかなる責任も負いません。

^{*1} <https://developers.google.com/speed/docs/insights/OptimizeImages>

目次

はじめに	2
第 1 章 色を視るしくみ	6
1.1 光の正体	6
1.2 光を色と認識するまで	7
1.2.1 桿体細胞の役割	7
1.2.2 錐体細胞の役割	7
1.3 色の正体	8
1.3.1 光を視る場合	8
1.3.2 物体を視る場合	8
1.4 カメラのしくみ	9
1.4.1 標本化	10
1.4.2 量子化	10
1.4.3 イメージセンサー	10
第 2 章 JPEG ファイルフォーマット	13
2.1 JPEG の誕生	13
2.2 JPEG の規格	13
2.2.1 JPEG コア規格 (ISO/IEC 10918-1)	14
2.2.2 JPEG File Interchange Format (ISO/IEC 10918-5)	14
2.2.3 Exchangeable Image File Format	14
2.3 JPEG のデータ構造	14
2.3.1 Start Of Image	15
2.3.2 Application segment	15
2.3.3 Define Quantization Table	16
2.3.4 Define Huffman Table	16
2.3.5 Start Of Frame	16
2.3.6 Start Of Scan	17

2.3.7	End Of Image	17
第 3 章	JPEG 圧縮アルゴリズム	18
3.1	JPEG 圧縮の考え方	18
3.1.1	色の変化を省略	18
3.1.2	細かい情報を省略	19
3.1.3	データを圧縮	19
3.2	圧縮処理の流れ	19
3.3	色空間変換	19
3.3.1	色空間の視覚的効果	20
3.4	ダウンサンプリング	21
3.4.1	クロマサブサンプリング方式	21
3.4.2	視覚的影響	24
3.5	ブロック分割	24
3.6	レベルシフト	26
3.7	離散コサイン変換	26
3.7.1	画像の周波数表現	26
3.7.2	2 次元 DCT	30
3.7.3	逆 DCT	32
3.8	量子化	34
3.8.1	量子化テーブル	34
3.8.2	デコーダーの品質	37
3.8.3	ブロックノイズ	37
3.8.4	モスキートノイズ	38
3.9	エントロピー符号化	38
3.9.1	DC 成分の符号化	39
3.9.2	AC 成分の符号化	40
3.10	全体の計算過程	44
3.10.1	符号化過程	45
3.10.2	復号化過程	48
付録 A	JPEG 以外のフォーマット	51
A.1	JPEG 2000	51
A.2	JPEG XR	51
A.3	JPEG XT	52
A.4	JPEG XL	52
A.5	JPEG XS	52

A.6	TIFF (Tagged Image File Format)	52
A.7	GIF (Graphics Interchange Format)	53
A.8	PNG (Portable Network Graphics)	53
A.9	BMP (Microsoft Windows Bitmap Image)	53
A.10	WebP	54
A.11	RAW	54
A.12	SVG (Scalable Vector Graphics)	54
付録 B	色空間	56
B.1	sRGB	56
B.2	Adobe RGB	56
B.3	Display P3	56
B.4	CMYK	57
B.5	HSV / HSB	57
B.6	HSL	57
B.7	YCbCr	57
付録 C	画質評価	58
C.1	平均二乗誤差	58
C.2	ピーク信号対雑音比	58
あとがき		59

第1章

色を視るしくみ

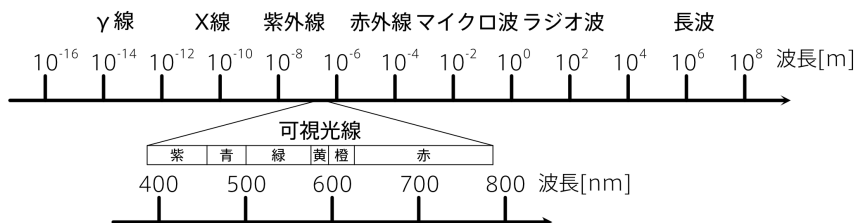
日常生活において色は物体を識別し、周囲の世界を視るのに重要な役割を果たしています。

では、私たち人間の眼は、普段どのようにして色を認識しているのでしょうか。そして、その色をデジタル世界に持って行くには、どうすればいいのでしょうか。

この章では、人間の眼がどのように光を色として認識しているのか、そのメカニズムを解説します。また、デジタルカメラが光から画像に変換する仕組みも、合わせて解説します。

1.1 光の正体

色を視るには、まず太陽をはじめとした光源が必要です。光は電磁波の一種であり、エックス線、紫外線、赤外線などと同じ種類のエネルギーです。



▲図 1.1 電磁スペクトル

人間の眼で視られるのは約 380nm から 780nm までの、可視光線と呼ばれる範囲に限られます。この範囲内で私たちはさまざまな色を知覚し、豊かな色彩の世界を体

験しています。

一見すると、太陽光には色がないように見えるかもしれませんが、しかし、プリズムを通った光や、雨上がりに現れる虹を観察すると、光がさまざまな色に分かれる様子を目にできます。

1.2 光を色と認識するまで

光が眼に入ると、まず角膜が光を屈折させて眼球内に導きます。続いて虹彩が瞳孔の大きさを調整し、取り込む光の量をコントロールします。瞳孔は周囲の明るさに応じて大きさを変え、明るい場所では小さく、暗い場所では大きくなります。その後、水晶体が光を集めて網膜上で焦点を合わせます。

網膜にある桿体（かんたい）細胞と錐体（すいたい）細胞の2種類の視細胞が、光を受け取ります。視細胞は受け取った光を、電気信号（神経インパルス）に変換し、神経線維の束を経由して脳に信号が送られます。脳は受け取った信号を統合し、色として認識します。

1.2.1 桿体細胞の役割

桿体細胞は、薄暗い条件下（暗所視）で主に機能し、光に対する感度が優れています。色の識別はできませんが、わずかな光でも反応し、夜間や薄暗い環境で物体や動きを捉えられます。

約1億個の桿体細胞が存在します。

1.2.2 錐体細胞の役割

錐体細胞は、明るい条件下（明所視）で主に機能し、視覚や色の識別を担当しています。桿体細胞と比べて光への感度は低いながらも、色や形の識別能力に優れています。

錐体は、約560nm付近の黄緑色に反応するL錐体（Long wavelength）、約540nm付近の緑色に反応するM錐体（Medium wavelength）、約440nm付近の青色に反応するS錐体（Short wavelength）の3種類あります。3種類の錐体細胞の強度を組み合わせると、人は数百万色以上を識別することが可能です。

全部で約600万個の錐体細胞が存在します。各錐体細胞の割合には個体差があり、一般的にL錐体が約60%、M錐体が約30%、S錐体が数%とされています。個体差があるにもかかわらず正常な色覚体験を得られているのは、錐体細胞からの相対的な信号に基づいて、脳が色覚を調整していると考えられています。

1.3 色の正体

ここまで、眼に入った光を脳が認識するまでの過程でした。では、そもそも色はどのように決まっているのでしょうか。

1.3.1 光を視る場合

光源自体が発光している場合には、その光が直接視覚細胞を刺激し、色として認識します。光源が放出する光のスペクトルによって決まる色を、光源色と呼びます。たとえば、赤色は主に 700nm 付近の波長を含む光が放出されている状態です。

複数の光源を組み合わせることで、新たな色を作り出せます。この方法を、加法混色と呼びます。

加法混色には、錐体細胞がもっともよく反応する赤、緑、青の三原色の光が用いられます。三原色をさまざまな割合で混ぜ合わせることで、幅広い色を表現できます。

たとえば同じ割合で、緑色と青色を重ね合わせるとシアン（水色）に、青色と赤色を重ね合わせるとマゼンタ（赤紫色）に、赤色と緑色を重ね合わせるとイエロー（黄色）に視えます。3色を重ね合わせると、白色になります。

光を重ね合わせて色を作るため、光を重ね合わせて黒色を作ることはできません。ディスプレイやプロジェクターは、光を遮ることで黒色を表現します。周囲の光や反射などの影響により、完全な黒を表現することは難しくなります。

1.3.2 物体を視る場合

光は自分自身が発光していたため、その色で眼で捉えられます。では、物体の色はどのように決まるのでしょうか。

実は、**物体そのものに色がついているわけではありません。**

物体に光が当たると、物体を構成する物質や色素が特定の波長の光を吸収します。吸収されなかった光は、物体の表面で反射されたり、物体を通り抜けたりします。この反射や透過した光のスペクトルを通じて、色を認識します。この現象を物体色と呼び、物体の色の正体です。

たとえば、赤いりんごが赤く視えるのは、太陽光に含まれる波長の中から赤色の光を強く反射し、それ以外の光を吸収しているためです。

第 2 章

JPEG ファイルフォーマット

JPEG は 30 年以上の歴史を持つ画像フォーマットです。

この章では、JPEG の誕生から現在に至るまでの歴史と、その内部構造に迫ります。1986 年に物語が始まった JPEG、その過程で生まれた JPEG コア規格、JFIF や Exif などの拡張規格、JPEG ファイルのデータ構造について解説します。

2.1 JPEG の誕生

JPEG の物語は、1980 年代に始まります。ストレージ容量が今ほど豊富でなかった当時、ファイルサイズを小さく抑えつつ、専用ソフトウェアを使わずに画像を交換、表示できる共通フォーマットの必要性が高まっていました。

国際標準化機構 (ISO) と国際電気通信連合電気通信標準化部門 (ITU-T) が共同で、1986 年に Joint Photographic Experts Group (JPEG) が設立されました。そこから 6 年後の 1992 年 9 月 18 日、JPEG フォーマットが発表されました。

2.2 JPEG の規格

JPEG は、圧縮アルゴリズムやデータ構造を定義したコア規格と、アプリケーション固有の拡張規格に分かれています。

コア規格は、ISO/IEC 10918-1 として国際標準化されました。拡張規格は、アプリケーション固有の情報を格納するための規格で、JFIF や Exif などがあります。

2.2.1 JPEG コア規格 (ISO/IEC 10918-1)

1992年に発表されたJPEGコア規格では、写真画像を効率的に圧縮するための技術を定義しています。

圧縮以外では、バイナリデータで構造化するための、マーカーやセグメントの仕様を定義しています。

JPEGコア規格は主に、圧縮、復号のアルゴリズムと、データストリームの構造に焦点を当てていました。色空間や解像度といった、アプリケーションレベルでの相互運用に必要な情報は規定されておらず、互換性の問題がありました。

2.2.2 JPEG File Interchange Format (ISO/IEC 10918-5)

C-Cube Microsystems社を中心としたコンソーシアムは、JPEGの草案段階から互換性の問題を認識していました。曖昧さを解消し、シンプルな方法を目的として1991年にJPEG File Interchange Format (JFIF)を開発しました。

JFIFは曖昧だった識別子、バージョン情報、色空間、解像度などの定義を行いました。色空間では、各成分8ビットのYCbCr表色系の使用を義務づけました。

これらの定義により曖昧さが解消され、プラットフォームを問わず画像交換が容易になり、JFIFはデファクトスタンダードとして浸透しました。約20年後の2013年に、ISO/IEC 10918-5として国際標準化されました。

2.2.3 Exchangeable Image File Format

デジタルカメラの普及に伴い、画像の撮影日やカメラの情報、露出設定、位置情報などを記録するニーズが出てきました。

日本電子工業振興協会（現在の電子情報技術産業協会、JEITA）は1995年、JFIFを基にExchangeable Image File Format (Exif)を開発しました。

Exifの目的は、撮影時の状況や使用した機材などのメタデータを、画像ファイル内に埋め込むための標準的な仕様を提供することです。

2.3 JPEG のデータ構造

実際のJPEGファイルは、どのようなデータ構造になっているのでしょうか。

JPEGファイルでは、内部データの管理にマーカーという方式が用いられています。

マーカーは、JPEGファイル内の各セクションの開始や、内容を識別するために使われる目印で、0xFFから始まる2バイトで構成されます。マーカー以外のデータで

0xFF が出現したときは、マーカーと誤認されないよう 0x00 を追加して 0xFF00 とします。

全体のファイル構造は、次のようになります。



▲図 2.1 JPEG ファイルの構造

代表的なマーカーを、いくつか見ていきましょう。

2.3.1 Start Of Image

Start Of Image (SOI, 0xFFD8) は、JPEG ファイルの先頭を示すマーカーです。JPEG ファイルの開始を示します。

2.3.2 Application segment

Application segment (APPn, 0xFFE0～0xFFEF) は、画像データ以外の情報を格納するマーカーです。APP0～APP15 まで予約されています。

APP0 (0xFFE0) は、JFIF 規格のときに利用します。主にバージョン情報、ピクセル密度、解像度、サムネイルが含まれます。

APP1 (0xFFE1) は、Exif 規格のときに利用します。主にカメラの種類、画像の向き、解像度、色空間、露出時間、F 値、ISO 感度、シャッタースピード、絞り値、焦点距離、フラッシュの有無、GPS、サムネイルなどが含まれます。また、Exif では APP2 (0xFFE2) に ICC プロファイルを格納することで、色の再現性が高められます。

JFIF 規格では APP0 が、Exif 規格では APP1 が、それぞれ SOI の直後に配置されることが前提となっています。そのため、両者は規格上互換性がありません。現実的には、SOI の後に APP0 (JFIF)、続いて APP1 (Exif) の構造になっている JPEG ファイルが多く存在します。ほとんどのデコーダーは両方のセグメントを含んでいても、正しく処理できるようになっています。

第3章

JPEG 圧縮アルゴリズム

画質を保ちながらファイルサイズを小さくする、その秘密は人間の視覚特性をうまく利用した仕組みにあります。

この章では、JPEG 圧縮アルゴリズムの核心に迫ります。人間の眼の特性とは何なのか、その特性をどう圧縮に活かしているのか。知恵と工夫が詰まった JPEG 圧縮の秘密を解き明かします。

3.1 JPEG 圧縮の考え方

JPEG は、自然写真のように色や明るさが滑らかに変化する画像の圧縮を得意としています。これは、人間の色覚特徴を上手に利用しているためです。

人間の眼は、すべての情報を均等に認識しているわけではありません。JPEG は、人間が気づきにくい、あるいは気にならない情報を選択的に減らすことで、見た目の印象を大きく損なわずにデータ量を削減しています。さらに、削減したデータの記録方法にも工夫があります。

3.1.1 色の変化を省略

人間の眼は、色のわずかな変化には鈍感で、明るさのわずかな変化には敏感に反応します。モニターや画像編集の際、色の濃淡をわずかに調整しても変化には気づきにくいですが、明るさを調整するとすぐに変化を感じられます。

この特性を活かし、明るさの情報はできるだけ残しつつ、色の情報を削減します。

(「3.3 色空間変換」、「3.4 ダウンサンプリング」)

3.1.2 細かい情報を省略

人間の眼は、細かい縞模様や急激な色の変化には鈍感で、広くなだらかな範囲の色の変化には敏感に反応します。非常に細かい白黒の縞模様は、細くなるにつれて灰色に見えてしまい、模様として認識できなくなります。

この特性を活かし、目立つ大きな形や色の変化はできるだけ残しつつ、細かな部分の情報を削減します。

（「3.7 離散コサイン変換」、「3.8 量子化」）

3.1.3 データを圧縮

JPEG では、データが特定の値に集中するように設計されています。

このデータの出現傾向を解析し、圧縮処理を行います。よく現れるデータには短い符号を、あまり現れないデータには長めの符号を割り当てることで、効率的に圧縮します。

（「3.6 レベルシフト」、「3.9 エントロピー符号化」）

3.2 圧縮処理の流れ

JFIF の Baseline DCT では、次のような順番で圧縮します。復号は、逆順で行います。

1. 色空間変換
2. ダウンサンプリング
3. ブロック分割
4. レベルシフト
5. 離散コサイン変換
6. 量子化
7. エントロピー符号化

3.3 色空間変換

JPEG 圧縮は、画像の色情報を、人間の眼にとって分かりやすい形に整理し直すことから始まります。

第3章 JPEG 圧縮アルゴリズム

赤 (R)、緑 (G)、青 (B) の光の組み合わせで表現する RGB 表色系は、ディスプレイで表示しやすい形式です。

人間の眼は、色の違いよりも明るさの違いに敏感であるという特性があります。そのため、画像を明るさ (Y) と、青みの違い (Cb)、赤みの違い (Cr) の組み合わせで表現する、YCbCr 表色系に変換します。こうすることで、明るさ (Y) の輝度成分をできるだけ残しつつ、色 (Cb、Cr) の色差成分が削減できます。

変換には、ITU-R BT.601 の変換式を利用します。この変換は可逆であり、情報の損失はありません。

式 3.1: RGB 表色系から YCbCr 表色系へ変換

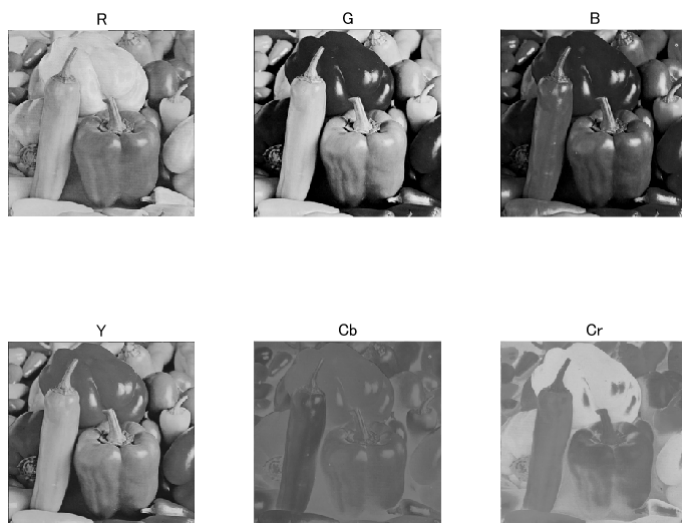
$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ C_b &= -0.168736R - 0.331264G + 0.5B + 128 \\ C_r &= 0.5R - 0.418688G - 0.081312B + 128 \end{aligned}$$

式 3.2: YCbCr 表色系から RGB 表色系へ変換

$$\begin{aligned} R &= Y + 1.402(C_r - 128) \\ G &= Y - 0.344136(C_b - 128) - 0.714136(C_r - 128) \\ B &= Y + 1.772(C_b - 128) \end{aligned}$$

3.3.1 色空間の視覚的效果

RGB 表色系と YCbCr 表色系の視覚的特徴を見るため、それぞれの成分に分離した画像を見てみましょう。



▲図 3.1 成分ごとに分離した画像

RGB 表色系は、どの成分もグレースケール画像のように見え、輪郭や構造がはっきりしています。

YCbCr 表色系は、Y 成分に関してはグレースケール画像のように見え、輪郭や構造がはっきりしています。Cb 成分や Cr 成分は、ぼんやりとしていて、細かい形はあまり分かりません。

このように、YCbCr 表色系で輝度成分と色差成分を分離することで、人間の視覚特性を活かした圧縮が可能になります。

3.4 ダウンサンプリング

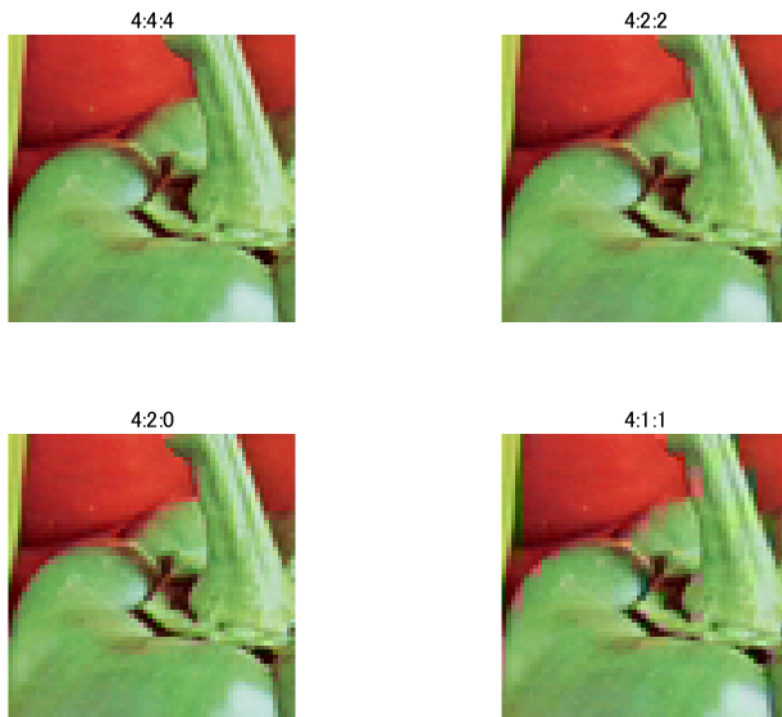
人間の眼が色の違いに鈍感であるという特性を利用し、色差成分（クロマ）の解像度を、輝度成分（ルミナンス）よりも、少なくすることでデータ量を削減します。この処理を、クロマサブサンプリングと呼びます。

3.4.1 クロマサブサンプリング方式

クロマサブサンプリングの方式は、一般的に J:a:b という 3 つの数字の比率で表記されます。水平方向に対して、J は 1 行あたりのピクセル数、a と b はそれぞれ 1 行目と 2 行目の、色差成分のピクセル数を示します。

3.4.2 視覚的影響

ほとんどの場合、色差成分を間引いても、目に見える劣化はあまり感じられませんが、色の境界がはっきりしている部分を極端に拡大すると、わずかに色がにじんで見える場合があります。



▲図 3.6 クロマサブサンプリング方式の拡大比較^{*1}

3.5 ブロック分割

JPEG では、画像を効率的に処理するため、 8×8 ピクセルのブロックに分割します。このサイズは、圧縮効率と画質のバランスを考慮しつつ、1990 年代のコンピュー

^{*1} 紙媒体でご覧の方は、サポートサイトにてカラー画像をご覧いただけます

レベルシフト

各画素値から 128 を足し、小数点を丸めて輝度成分を得ます。

▼表 3.15 復号化した輝度成分

-70	-97	-101	-100	-45	2	10	12
-93	-99	-100	-73	-11	11	9	13
-100	-100	-89	-31	6	3	-7	-32
-102	-100	-60	-2	-3	-23	-34	-55
-102	-88	-15	-5	-34	-44	-51	-55
-89	-15	-9	-39	-50	-49	-53	-43
-12	-8	-40	-46	-53	-54	-47	-30
-10	-44	-43	-48	-51	-52	-41	-23

復号化結果

入力画像と復元画像を比較すると次のようになります。



▲図 3.23 入力画像と復元画像の比較

差分は、圧縮によって変動した画素値の情報を表しています。量子化によって DCT 係数の多くを 0 にしたにもかかわらず、多くの箇所です ± 10 程度に収まっています。このように誤差を小さく抑えられることが、JPEG 圧縮の効率性の高さにつながっています。

付録 A

JPEG 以外のフォーマット

この付録では、JPEG 以外のファイルフォーマットを紹介します。

A.1 JPEG 2000

JPEG 2000 は、2000 年に JPEG 委員会によって規格化された画像フォーマットです。JPEG と比較してより高い色深度への対応、エラー耐性の向上、低ビットレートでの圧縮性能向上、多機能化、そして可逆圧縮と非可逆圧縮の両方に対応している点が特徴です。

圧縮には、周波数変換に離散ウェーブレット変換 (DWT)、エントロピー符号化に算術符号が用いられます。JPEG よりも高い圧縮率を実現し、可逆圧縮と非可逆圧縮の両方を扱えます。

主に、医療画像やデジタルシネマなど、高画質・高機能性が求められる分野で利用されています。

A.2 JPEG XR

JPEG XR は、Microsoft 社が 2006 年に開発した技術を、JPEG 委員会が採用した画像フォーマットです。JPEG と比較して高い圧縮率、高い色深度、アルファチャンネルによる透過処理、そして可逆圧縮と非可逆圧縮の両方に対応している点が特徴です。

圧縮には、 4×4 ブロックを基本とする Photo Core Transform (PCT) と、ブロックノイズを低減する Photo Overlap Transform (POT) が用いられます。

付録 B

色空間

この付録では、色を数値で表現する色空間を紹介します。

B.1 sRGB

国際電気標準会議（IEC）によって定められた、国際標準規格の色空間です。赤、緑、青の三原色の光を混合して色を表現する加法混色に基づいています。

一般的な PC モニターやウェブコンテンツで広く使用されており、多くのデジタルデバイスが対応しています。デバイス間での色再現性の一貫性を保つ基準となっています。

B.2 Adobe RGB

Adobe 社によって策定された色空間です。sRGB と比較してより広い色域を持ち、とくに緑から青にかけての鮮やかな色の再現性に優れています。

プロフェッショナル向けの写真編集や、高品質な印刷物制作に適しています。色域を正確に表示、印刷するためには、Adobe RGB に対応したモニターやプリンターなどのデバイスが必要です。

B.3 Display P3

Apple 社によって提唱されている色空間で、デジタルシネマ規格である DCI-P3 の色域をベースに、白色点を D65 に設定するなど、ウェブコンテンツ向けに調整したものです。sRGB と比較して色域が広く、とくに緑から赤にかけての鮮やかな色の再現性に優れています。

Apple 製品を中心に搭載が進み、ウェブコンテンツでの対応も広がっています。

付録 C

画質評価

この付録では、画像の劣化具合を定量的に評価する方法を紹介します。

C.1 平均二乗誤差

平均二乗誤差 (MSE) は、元画像の画素値 $I(i, j)$ と、比較対象画像の画素値 $K(i, j)$ の差を二乗し、それらの総和を全要素数 $H \times W$ で割った値です。

式 C.1: MSE の計算式

$$MSE = \frac{1}{HW} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} [I(i, j) - K(i, j)]^2$$

MSE の値が 0 に近いほど、2 つの画像の誤差は少なく、元画像に忠実であることを意味します。

C.2 ピーク信号対雑音比

ピーク信号対雑音比 (PSNR) は、信号が取りうる最大の画素値とノイズの比率を、対数スケールで示したものです。

式 C.2: PSNR の計算式

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

PSNR の値が大きいほどノイズの割合が相対的に小さく、画質が良いと評価されます。