

MPEG の魔法

ゆずねり 著

2025-11-15 版 Neln 発行

はじめに

フル HD 映像を非圧縮で保存すると、1 秒間で約 190MB、1 時間では約 680GB のデータ量になります。

ところが現代では、スマートフォンから SNS への動画アップロードは数秒で完了し、Netflix での映画ストリーミングは途切れることなく再生され、ビデオ会議で世界中の人々と顔を合わせて会話できます。この当たり前の光景の裏には、映像データを劇的に小さくする「魔法」のような技術があります。

その魔法の名前は MPEG。

1980 年代後半、約 650MB の CD に映画を収録するという目標が掲げられました。映像データを大幅に圧縮しながら視覚的な品質を保つ。当時としては夢物語でした。

そんな夢物語に世界中の技術者が挑み、数々の発明と工夫を重ねて生まれたのが MPEG です。

本書は、MPEG という技術の誕生から現代までの物語です。35mm フィルムとブラウン管から始まったアナログ時代、映像のデジタル化、そしてインターネット時代の到来。その進化の過程をたどります。

MPEG には、驚くべき工夫と発想の転換が詰まっています。

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/nt04/>

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。

目次

はじめに	2
第1章 映像圧縮の歴史	8
1.1 映画のはじまり	8
1.1.1 35mm フィルムと映写機の誕生	8
1.1.2 サイレント映画からトーキー映画へ	10
1.1.3 白黒映画からカラー映画へ	10
1.1.4 ワイドスクリーンの登場	10
1.1.5 サラウンド音響の発展	10
1.2 テレビのはじまり	11
1.2.1 世界のアナログ放送規格	11
1.2.2 ビデオ録画技術の誕生	13
1.2.3 家庭用ビデオ規格戦争	13
1.3 MPEG のはじまり	13
1.3.1 テレビ会議から始まった映像圧縮	13
1.3.2 CD 規格がもたらした MPEG の誕生	14
1.3.3 MPEG-1 の登場	14
1.3.4 MP3 が開いた音楽配信	14
1.4 デジタル放送と高画質メディア	15
1.4.1 MPEG-2 の登場	15
1.4.2 DVD の登場	15
1.4.3 デジタル放送への移行	16
1.4.4 HDTV の普及	17
1.5 インターネット映像時代の幕開け	17
1.5.1 Flash Video と Web 映像配信	17
1.5.2 MPEG-4 の登場	18
1.5.3 H.264/AVC の登場	18
1.5.4 次世代光ディスクの規格競争	18

1.5.5	コーデック戦争の終焉	19
1.6	4K 時代と次世代コーデック	19
1.6.1	H.265/HEVC の登場と特許問題	19
1.6.2	オープンコーデックの台頭	19
1.7	映像技術の進化と継承	20
第 2 章	映像と音声のデジタル表現	21
2.1	映像を構成する三つの次元	21
2.1.1	空間次元	21
2.1.2	時間次元	22
2.2	アナログ信号の仕組み	22
2.2.1	走査線	22
2.2.2	インターレース方式	23
2.2.3	プログレッシブ方式	24
2.3	デジタル信号の仕組み	25
2.3.1	標本化	25
2.3.2	量子化	26
2.4	データ量とビットレート	26
2.4.1	非圧縮映像のビットレート	26
2.4.2	非圧縮音声のビットレート	27
2.4.3	ビットレート制御	27
第 3 章	映像圧縮の原理	29
3.1	映像が持つ冗長性	29
3.1.1	空間的冗長性	29
3.1.2	時間的冗長性	29
3.1.3	統計的冗長性	31
3.2	空間的冗長性の削減方針	31
3.3	色空間の変換	31
3.4	クロマサブサンプリング	32
3.5	離散コサイン変換	32
3.6	量子化	33
3.6.1	量子化テーブル	33
3.6.2	空間的マスキング効果	34
3.7	ランレングス符号化	34
3.8	可変長符号化	35
3.9	時間的冗長性の削減方針	35

3.10	フレーム間予測技術	35
3.10.1	I フレーム	36
3.10.2	P フレーム	36
3.10.3	B フレーム	36
3.10.4	GOP	36
3.11	動き補償	37
3.11.1	エンコーダー側の例	37
3.11.2	デコーダー側の例	38
3.11.3	マクロブロック	40
3.11.4	動きベクトルの算出	40
3.11.5	残差の計算と符号化	40
3.12	コンテナとコーデック	43
3.12.1	コンテナフォーマット	43
3.12.2	コーデック	43
3.12.3	コンテナとコーデックの関係	43
第4章	MPEG-1	44
4.1	開発の背景	44
4.2	MPEG-1 Part 1 (Systems)	44
4.2.1	Program Stream	45
4.2.2	タイムスタンプによる同期	45
4.3	MPEG-1 Part 2 (Video)	45
4.3.1	解像度とフレームレート	45
4.3.2	マクロブロックとブロック構造	45
4.3.3	動き補償の仕様	46
4.4	MPEG-1 Part 3 (Audio)	46
4.4.1	Layer I	46
4.4.2	Layer II	46
4.4.3	Layer III	47
4.5	MPEG-1 の限界	47
4.5.1	解像度の制約	47
4.5.2	インターレース映像の非対応	47
4.5.3	ビットレート制御の制約	48
4.6	MPEG-1 の功績	48
第5章	MPEG-2	49
5.1	開発の背景	49

目次

5.2	MPEG-2 Part 1 (Systems)	49
5.3	MPEG-2 Part 2 (Video)	50
5.3.1	インターレース映像対応	50
5.3.2	フレーム DCT とフィールド DCT	50
5.3.3	動き補償の拡張	51
5.3.4	デュアルプライム予測	51
5.3.5	量子化マトリクスと適応量子化	51
5.3.6	プロファイルとレベル	52
5.4	MPEG-2 Part 3 (Audio)	53
5.5	MPEG-2 Part 7 (Advanced Audio Coding)	53
5.6	幻の MPEG-3	53
5.7	MPEG-2 の限界	53
5.7.1	イントラ符号化の制約	54
5.7.2	動き補償精度の制約	54
5.7.3	エントロピー符号化の制約	54
5.7.4	HD 解像度での効率低下	54
5.7.5	インターネット配信への制約	55
5.8	MPEG-2 の功績	55
第 6 章	MPEG-4	56
6.1	開発の背景	56
6.2	MPEG-4 Part 2 (Visual)	57
6.2.1	動き補償の高度化	57
6.2.2	Global Motion Compensation	57
6.2.3	符号化ツールの改良	57
6.2.4	MPEG-4 Visual の実装例	58
6.2.5	技術的成果と課題	58
6.3	MPEG-4 Part 3 (Audio)	58
6.4	MPEG-4 Part 14 (MP4 file format)	59
6.5	MPEG-4 Part 10 (Advanced Video Coding)	59
6.5.1	予測技術の高度化	59
6.5.2	デブロッキングフィルター	60
6.5.3	CABAC/CAVLC エントロピー符号化	60
6.5.4	プロファイルとレベル	60
6.6	H.264/AVC の限界	61
6.6.1	並列処理の制約	61
6.6.2	マクロブロックサイズの制約	61

6.6.3	エントロピー符号化の制約	62
6.7	MPEG-4 の功績	62
第7章 次世代映像圧縮技術		63
7.1	4K 時代の標準コーデック H.265/HEVC	63
7.1.1	H.265/HEVC の技術的革新	63
7.1.2	H.265/HEVC の課題	64
7.2	ロイヤリティフリーに挑戦する AV1	64
7.2.1	AV1 の技術的革新	64
7.2.2	AV1 の課題	65
7.3	究極の圧縮効率を追求する H.266/VVC	65
7.3.1	H.266/VVC の技術的革新	65
7.3.2	H.266/VVC の課題	66
7.4	映像圧縮技術の未来	66
あとがき		67

第1章

映像圧縮の歴史

映像技術の歴史は、映画とテレビの発展とともに歩んできました。35mmフィルムから始まった映像技術は、テレビの普及、デジタル化、そしてインターネット配信へと進化し、それぞれの時代が新しい課題をもたらし、新しい圧縮技術を生み出しました。

本章では、映画の誕生からストリーミング時代まで、映像技術がどのように進化してきたかをたどります。

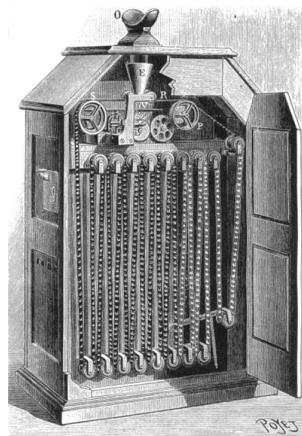
1.1 映画のはじまり

映像の歴史は、映画の誕生と共に始まります。

1.1.1 35mm フィルムと映写機の誕生

1891年にトマス・エジソンと助手のウィリアム・ケネディ・ローリー・ディクソンが、キネトスコープを発明しました。キネトスコープは35mmフィルムをループ状にして高速回転させることで、連続した静止画を映像として見せる装置です。

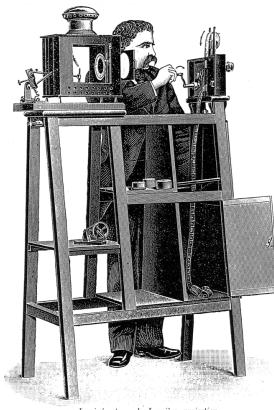
キネトスコープは映写機ではなく、1人ずつのぞき込む形式でした。



▲図 1.1 キネトスコープ

1895 年にフランスのリュミエール兄弟が、シネマトグラフを発明しました。シネマトグラフはスクリーンへの投影を可能にし、多人数での映像鑑賞を実現した、映写機の原型でした。

シネマトグラフでも 35mm フィルムが使用されたことにより、35mm フィルムは事実上の標準として確立されました。



▲図 1.2 シネマトグラフ（右）と幻灯機（左）

1.1.2 サイレント映画からトーキー映画へ

初期の映画は音声がありませんでした。字幕の挿入や、映像に合わせて生演奏やナレーターが補っていました。

映像と音声を同期させるために、多くの試行錯誤が行われました。1926年にワーナー・ブラザース社とウェスタン・エレクトリック社によって、ヴァイタフォン方式が開発されました。ヴァイタフォン方式は、映写機とレコードプレーヤーを同期させて音声を再生する方式です。翌年『ジャズ・シンガー』がこの方式で制作され、トーキー映画の先駆けとなりました。

トーキー映画が登場するまで、フィルムの再生速度は統一されず手動で調整していました。音声と映像の同期には一定のフレームレートが必須となり、音質とフィルムコストのバランスから 24FPS が採用されました。現代でも動きに適度なブラーが乗り、映画らしい印象を与えるため、この伝統が受け継がれています。

レコードと映写機の同期は非常に難しく頻繁に音ズレが発生することから、やがてフィルム自体に音声を記録するサウンド・オン・フィルム方式が主流となります。

1.1.3 白黒映画からカラー映画へ

トーキー化と並行して、カラー化も発展します。

1932年、テクニカラー方式が登場します。テクニカラー方式は光の三原色である赤、緑、青を別々のフィルムに記録し、合成する技術です。テクニカラー方式により、鮮やかなカラー映像が実現し、映画の表現力が大きく向上しました。

1.1.4 ワイドスクリーンの登場

1950年代になると、映画はより大きなスクリーンで映像を楽しむ方向に進化しました。

1953年に20世紀フォックス社がシネマスコープを開発しました。アナモルフィックレンズという特殊なレンズを用いて、撮影時に映像を水平方向に圧縮して 35mm フィルムに記録します。上映時に同じレンズで元に戻すことで、横長のワイド映像を実現するものでした。これにより、2.35:1 のアスペクト比を持つワイドスクリーン映画が登場しました。

1.1.5 サラウンド音響の発展

映画の音響はモノラルからステレオ、そしてサラウンドへと進化してきました。

1970年代にドルビー研究所により左、中央、右、後方の 4 チャンネル音響を実現

するドルビーステレオが開発されました。フィルムの光学トラックに複数チャンネルを記録し、映画館の複数のスピーカーで立体的に再生することで、映画の没入感を大幅に向上させました。

1.2 テレビのはじまり

映画と並行して、テレビも発展を遂げます。

1920 年代から 1930 年代にかけて、テレビの実験放送が各国で始まりました。初期のテレビは機械式テレビと呼ばれ、回転する円盤に開けた穴を通して光を当て、その光の明るさを変えることで、点を並べるように画像を映し出す仕組みです。



▲図 1.3 機械式テレビ

1930 年代に入ると、ブラウン管 (Cathode Ray Tube、CRT) を使用した電子式テレビが登場します。その後、1936 年にイギリスの BBC が世界初の定時テレビ放送を開始し、1939 年にはアメリカでも NBC が定時放送を開始しました。

1953 年にはアメリカで NTSC 方式によるカラー放送が本格的に開始され、1970 年代後半には白黒テレビからカラーテレビへの移行がほぼ完了しました。

1.2.1 世界のアナログ放送規格

アナログテレビの放送規格は NTSC、PAL、SECAM の三つが主要な方式として存在します。なぜ世界で統一された規格が生まれず、複数の方式が並立することになったのでしょうか。

大きな理由に、各国の電力周波数の違いが影響しています。電力周波数は、地域によって 60Hz または 50Hz に分かれています。アナログテレビの初期設計では、電源周波数とテレビの垂直同期周波数を一致させることで、電源由来のノイズによる画面のちらつきを抑えるという技術的な工夫がなされていました。そのため、電力事情に

第1章 映像圧縮の歴史

応じて 60Hz の地域では 30FPS、50Hz の地域では 25FPS が発展しました。

NTSC

National Television System Committee (NTSC) は北米、日本、韓国、中南米などで採用されていた規格です。

1953 年にアメリカで開発された、世界初のカラー放送の規格です。1 秒間に 30 フレーム、525 本の走査線で、 720×480 の解像度で構成されます。

カラー化にあたり、白黒テレビとの互換性を保つため、輝度信号 (Y) と色差信号 (I、Q) に分離する方式が採用されました。しかし、映像信号の帯域幅が拡大し、音声信号との干渉が問題となりました。

この問題に対応するため、色情報を伝送するための副搬送波の周波数を、音声搬送波と干渉しないよう調整する必要がありました。その結果、フレームレートは 30FPS を $1,001/1,000$ 倍した約 29.97FPS^{*1}へと変更されました。この奇妙な仕様は、デジタル時代になっても引き継がれることになります。

また、色再現性に課題があり、撮影機材や放送設備の品質によって色味が変化しやすいという欠点がありました。

PAL

Phase Alternating Line (PAL) は西ヨーロッパ、中国、インド、オーストラリア、南アフリカ、ブラジルなどで採用されていた規格です。

1963 年にドイツで開発された、NTSC の色相ズレを改善した規格です。1 秒間に 25 フレーム、625 本の走査線で、 720×576 の解像度で構成されます。

最大の特徴は、走査線ごとに色差信号の位相を反転させることで、伝送中の位相ズレを自動的に補正する点です。これにより、NTSC よりも安定した色再現性を実現しました。

SECAM

Sequential Couleur avec Memoire (SECAM) はフランス、旧ソ連、東欧諸国、イラン、フランス語圏諸国などで採用されていた規格です。

1967 年にフランスで開発された規格です。PAL と同様に、1 秒間に 25 フレーム、625 本の走査線で、 720×576 の解像度で構成されます。

NTSC や PAL と異なり、二つの色差信号 (R-Y と B-Y) を走査線ごとに交互に伝送し、受信側でメモリに保存して組み合わせる方式を採用しました。周波数変調 (FM) を使用するため位相ズレの影響を受けにくく、三つの規格の中でもっとも安定した色再現性を実現しました。

^{*1} 正確には $30 \times (1000/1001) = 29.970029$ FPS

1.2.2 ビデオ録画技術の誕生

1950 年代後半になると、テレビ放送を録画・再生する技術が登場しました。

1956 年にアメリカのアンペックス社が、世界初の実用的なビデオテープレコーダー (VTR) VRX-1000 を開発しました。これは放送局向けの業務用機器で、2 インチ幅のビデオテープを使用し、テレビ番組の録画と再生が可能になりました。これにより、生放送だけでなく、事前に録画した番組を放送できるようになり、放送業界に大きな変革をもたらしました。

1.2.3 家庭用ビデオ規格戦争

1970 年代に入ると、ソニーの Betamax とビクターの VHS という二つの家庭用ビデオ規格が登場します。両規格は互換性がなく、いわゆるビデオ戦争と呼ばれる規格競争が繰り広げられました。多くのメーカーが VHS に参入したことで、最終的に VHS が市場を制しました。

1980 年代には、VHS でテレビ番組の録画や映画の視聴が一般家庭で広く行われるようになりました。ビデオレンタル店も普及し、映像コンテンツを自由に選んで視聴する文化が定着しました。

1981 年、映画館の音響体験を家庭でも実現するドルビーサラウンドが登場しました。VHS デッキと対応テレビの組み合わせで、家庭でもサラウンド音響を楽しめるようになりました。ホームシアターの概念が生まれました。

1.3 MPEG のはじまり

1980 年代に入ると、コンピューター技術の急速な進歩により、映像のデジタル化が本格的に始まりました。

1.3.1 テレビ会議から始まった映像圧縮

デジタル映像圧縮の研究は、まずテレビ会議の分野から始まりました。

1984 年、国際電気通信連合 (ITU-T) がテレビ会議用の映像圧縮規格 H.120 を策定しました。しかし、圧縮効率が低く実用的ではありませんでした。

1990 年、H.120 をより実用的にした H.261 が標準化されました。H.261 は離散コサイン変換と動き補償を組み合わせた技術で、テレビ会議での映像伝送を実現しました。

第2章

映像と音声のデジタル表現

アナログの世界で捉えられた映像と音声は、デジタルデータとして保存、伝送されます。この変換プロセスにおいて、連続的な信号をどのように離散的なデータとして表現し、効率的に扱うかという原理が確立されました。

本章では、映像と音声をデジタルデータとして表現するための基礎的な仕組みと、その情報量を制御する技術について解説します。

2.1 映像を構成する三つの次元

映像を理解するためには、まず映像が持つ次元について把握する必要があります。映像は「2次元の静止画が時間と共に変化するもの」です。つまり、幅と高さの二つの空間次元に、時間次元を加えた3次元データです。

これら三つの次元を組み合わせることで、私たちは動きのある映像を認識しています。

2.1.1 空間次元

空間次元は、1枚の静止画の中での情報を表現します。

静止画には、幅（横方向）と高さ（縦方向）の二つの次元があります。この空間次元を、細かい格子状に分割したものが画素（ピクセル）です。

空間次元の解像度は、このピクセルの数で決まります。1080p（フル HD）であれば、 1920×1080 ピクセル、4K であれば 3840×2160 ピクセルです。

解像度が高いほど、より詳細で鮮明な映像を表現できます。

2.1.2 時間次元

時間次元は、静止画を時系列に並べることで動きを表現する次元です。

映像は連続した静止画（フレーム）の集まりで構成されており、それらを高速で切り替えることで、動いているように見えます。1秒間に表示されるフレーム数をフレームレートと呼び、単位はFPS（Frames Per Second）で表されます。

フレームレートが高いほど、より滑らかな動きを表現できます。映画では24FPS、テレビ放送では30FPS（NTSC）や25FPS（PAL）が一般的です。

2.2 アナログ信号の仕組み

私たちの眼が見ている風景は、無数の光の点が集まったものです。カメラはこの光を電気信号に変換し、テレビはその電気信号を光の像として再現します。

カメラはレンズを通して入ってきた光を、撮像管という光電変換素子で電気信号に変換します。撮像管は明るい光を強い電気信号に、暗い光を弱い電気信号に変換します。しかし、一度にすべての光を捉えることはできません。

そのため、映像を左上から右下に向かって順番に、各点の明るさの情報を時系列の電気信号（映像信号）に変換します。映像信号には、各点の明るさの情報だけでなく、走査のタイミングを制御するための同期信号も含まれています。

同期信号は、カメラ（送信側）とテレビ（受信側）の走査のタイミングを完全に一致させるために不可欠な信号です。水平同期信号は各走査線の開始を、垂直同期信号は各画面（フレーム）の開始を示します。これにより、受信側のテレビは、送られてきた映像信号を正しく再現できます。

2.2.1 走査線

ブラウン管（CRT）テレビの画面は1枚の絵のように見えますが、実際は左上から右下に向かって順番に線を引くようにして描画されています。この線を走査線（スキャンライン）と呼びます。

CRTテレビでは、電子銃から発射された電子ビームが、蛍光体が塗られた画面の内側に衝突して発光することで、点が描かれます。この点を水平方向に連続して動かすことで線（走査線）となり、さらにその線を上から下へと少しずつずらしながら描くことで、画面全体に映像が映し出されます。

走査線の数は、映像の垂直解像度を決定します。走査線の本数が多ければ多いほど、より高精細で詳細な映像を表現できます。

走査線を使って画面を描画する方法には、大きく分けてインターレース方式とプロ

グレッシップ方式の二つの方式があります。これらは走査線をどのような順序で描画するかという、物理的な実装方法の違いです。

2.2.2 インターレース方式

インターレース方式（飛び越し走査）では、まず奇数番目の走査線（1、3、5、...）だけを上から下へと描画します。これが終わると、次に偶数番目の走査線（2、4、6、...）を同様に描画します。この奇数番目の走査線の集まりを奇数フィールド（または第1フィールド）、偶数番目の走査線の集まりを偶数フィールド（または第2フィールド）と呼びます。

この二つのフィールドが組み合わさることで、はじめて1枚の完全な画像（フレーム）が完成します。

1秒間に30フレーム描画する場合、実際には奇数フィールドと偶数フィールドそれぞれ30回ずつ、合計60回の画面更新が行われます。これにより、人間の目には1秒間に60枚の更新があるように感じられ、動きが滑らかに見えます。

インターレース方式が開発された当時、利用できる電波の周波数帯域は非常に限られていました。1秒間に30枚のフレームを送信する帯域幅はなく、フレーム数を減らせば映像のカクつきや画面のちらつきが発生してしまいます。

インターレース方式は、1回の描画で送る情報量を半分にすることで帯域幅を節約し、同時にフィールドの更新頻度を2倍にすることでちらつきを抑える折衷案として考案されました。また、当時の技術限界により、一度にすべて描画することが困難だったことも、インターレース方式が採用された理由のひとつです。

一方、奇数フィールドと偶数フィールドの描画タイミングのずれが原因で、映像に不自然な動きが生じます。動きが速い時に、映像の輪郭がギザギザに見えるコーミングと呼ばれる現象や、静止画を表示した際にちらつきが発生します。



▲図2.1 インターレース方式

2.2.3 プログレッシブ方式

プログレッシブ方式（順次走査）はインターレース方式とは対照的に、すべての走査線を一度に描画するシンプルな方式です。

インターレース方式のようなフィールドの概念ではなく、1フレームが1枚の完全な静止画に対応します。コーミングのような動きに関するノイズが発生せず、静止画としても非常にクリアな画像を得られます。

現代では半導体技術と信号処理能力の向上により、大容量のデータを効率的に圧縮、伝送できるようになりました。そのため、プログレッシブ方式が標準的な方式となっています。



▲図2.2 プログレッシブ方式

第3章

映像圧縮の原理

フル HD 映像を非圧縮で保存すると、1 秒間で約 190MB ものデータ量になります。この膨大なデータを $1/100 \sim 1/1,000$ に圧縮しながら、視覚的な品質を保つにはどうすればよいのでしょうか。

本章では、人間の視覚特性と映像特有の冗長性を活用し、驚異的な圧縮率と高い画質を両立させる原理に迫ります。

3.1 映像が持つ冗長性

映像データには、主に 3 種類の冗長性が含まれています。空間的冗長性（静止画内の相関）、時間的冗長性（フレーム間の相関）、統計的冗長性（データの偏り）です。これらの冗長性を効果的に削減することで、高い圧縮率を実現します。

3.1.1 空間的冗長性

空間的冗長性とは、1 枚の静止画の中に含まれる冗長性のことです。

自然画像の基本的な特性として、隣接するピクセルは似た色を持つことが多いという点があります。たとえば、空の部分はどこも似たような青い色が続き、草原の部分は緑色が広がっています。

静止画圧縮では、隣接するピクセルの色が似ていることを利用してデータ量を削減します。JPEG も、この原理を利用してデータ量を削減しています。

3.1.2 時間的冗長性

時間的冗長性とは、連続するフレーム間に含まれる冗長性のことです。

30FPS の場合、フレーム間隔は約 33 ミリ秒です。この短い時間では前後のフレー

ムで内容はほとんど変化しません。

$\frac{1}{30}$ 秒



$\frac{2}{30}$ 秒



$\frac{3}{30}$ 秒



⋮

⋮

$\frac{30}{30}$ 秒

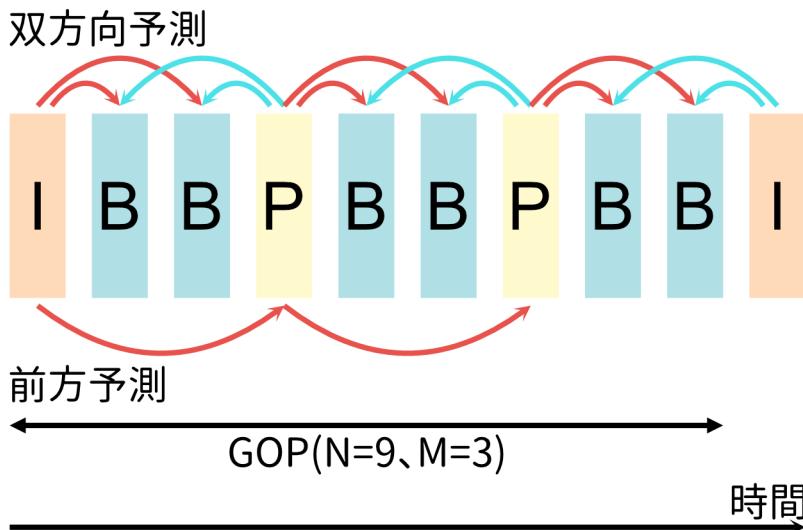


▲図3.1 時速300kmでも1/30秒での変化は小さい

時間的冗長性の削減効果は、映像の内容によって大きく異なります。

ニュース番組のようなキャスターが話す静的なシーンや固定カメラで撮影された映像は、フレーム間の差分が非常に小さいため、大きく削減できます。一方、アクション映画やスポーツ中継のような、動きの激しいシーンではフレーム間で大きな変化が生じるため、圧縮効率が低下します。さらに、シーンチェンジ（場面転換）では、前後のフレームにほとんど関連性がないため、時間的冗長性を利用できません。

ムが 3 フレームごとに配置) となります。



▲図 3.3 フレーム構成

3.11 動き補償

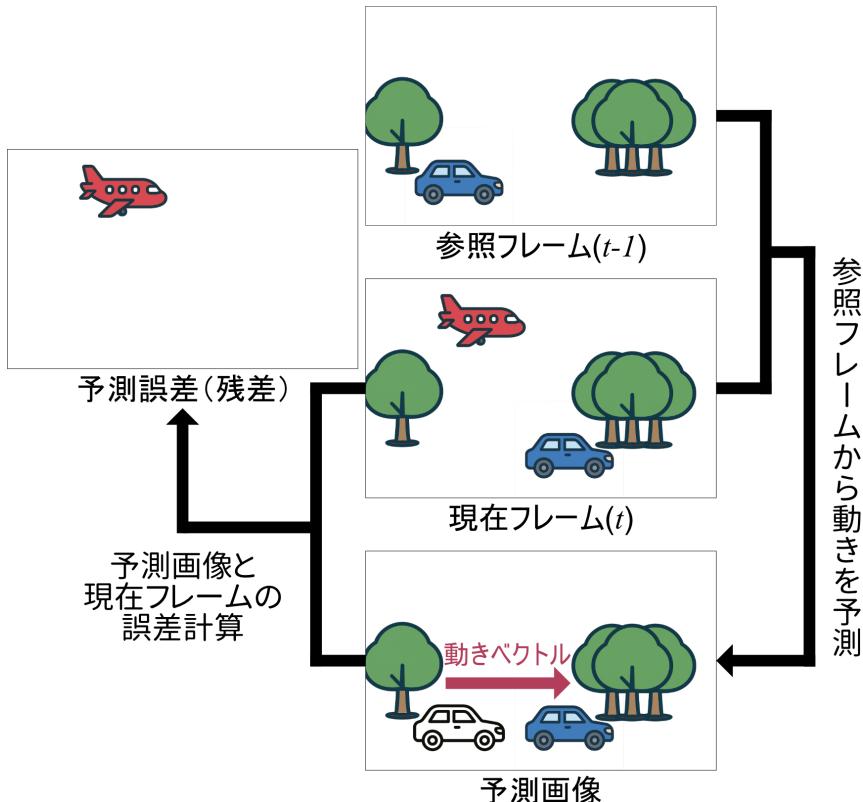
動き補償 (Motion Compensation) は、連続するフレーム間の類似性を利用して、データ量を削減します。

映像は連続したフレームで構成されており、隣接するフレーム同士の内容はほとんど変化しません。

動き補償では、現在のフレームをそのまま符号化するのではなく、参照フレームからの動きと差分だけを記録します。これにより、フレーム全体を保存する場合と比べて、データ量を大幅に削減できます。

3.11.1 エンコーダー側の例

図 3.4 は、エンコーダー側の動き補償の基本的な流れを示しています。



▲図 3.4 エンコーダー側動き補償の概念

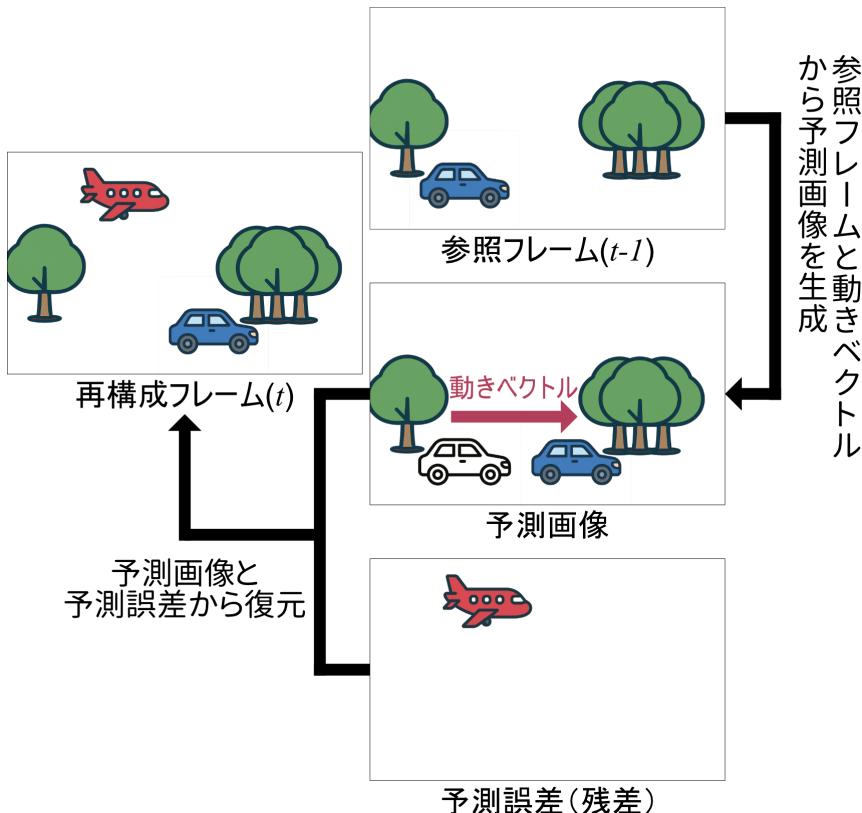
エンコーダーは、参照フレームを使って現在のフレームを予測し、動きベクトルと残差情報を計算します。

背景の木は動いていないため、動きベクトルはゼロになります。車は参照フレームから右に移動しているため、その移動量を動きベクトルとして記録します。

新しく出現した飛行機は、参照フレームに存在しないため動きベクトルだけでは表現できません。そのため、飛行機の部分は予測誤差(残差情報)として記録されます。

3.11.2 デコーダー側の例

図 3.5 は、デコーダー側の動き補償の基本的な流れを示しています。



▲図 3.5 デコーダー側動き補償の概念

デコーダーは、参照フレームと動きベクトル、および残差情報から現在のフレームを再現します。

背景の木は動きベクトルがゼロのため、参照フレームからそのままコピーします。車は動きベクトルが示す分だけ、参照フレームから右方向にずらして配置します。

飛行機は参照フレームに存在しないため、残差情報から飛行機の部分を復元して現在のフレームに追加します。

第4章

MPEG-1

650MB の CD に、VHS 相当の画質で約 74 分の映像を記録する、当時としては夢物語だった課題に挑んだのが MPEG-1 です。

MPEG-1 は、I、P、B フレーム構造、DCT ベースの圧縮、動き補償など、現代のデジタル映像圧縮の基本概念をほぼすべて確立しました。

本章では、デジタル映像圧縮の礎を築いた MPEG-1 の技術的特徴と実用化、そしてその限界について解説します。

4.1 開発の背景

1980 年代後半、コンピューター業界は大きな転換期を迎えていました。

1982 年に CD が登場し、約 650MB という当時としては大容量のデータを、安価なメディアに記録できるようになりました。

この新しいメディアを使って、音楽だけでなく映像も記録できなかつた。この発想が、MPEG-1 開発の出発点となりました。

1988 年、ISO と IEC の合同技術委員会の下に、MPEG が設立されました。

MPEG の最初の目標は、約 74 分の映像を CD に収める圧縮技術の開発でした。この目標を達成するために、映像と音声を合わせて約 1.5Mbit/s のビットレートで VHS 相当の画質を実現する必要がありました。

4 年間の開発期間を経て、1993 年、MPEG-1 規格が正式に国際規格 ISO/IEC 11172 として承認されました。この規格は、MPEG-1 Systems (Part 1)、MPEG-1 Video (Part 2)、MPEG-1 Audio (Part 3) など複数のパートで構成されています。

4.2 MPEG-1 Part 1 (Systems)

映像と音声を多重化し、同期再生するための仕組みを定義しています。

4.2.1 Program Stream

MPEG-1 では、Program Stream (PS) という多重化方式が定義されています。PS は、比較的エラーの少ない環境 (CD やハードディスクなど) での使用を想定しています。

映像ストリームと音声ストリームはそれぞれパケット化され、時刻情報とともに多重化されます。パケットサイズは可変長で、効率的なデータ格納が可能です。

4.2.2 タイムスタンプによる同期

B フレームの存在により、フレームのデコード順序と表示順序が異なります。このため、MPEG-1 では各フレームに Presentation Time Stamp (PTS) と Decoding Time Stamp (DTS) を付加します。

PTS は表示時刻を示し、デコーダーはこの時刻にしたがってフレームを表示します。DTS はデコード時刻を示し、デコーダーはこの時刻にしたがってフレームをデコードします。

この仕組みにより、映像と音声の同期、およびフレームの正しい表示順序が保証されます。

4.3 MPEG-1 Part 2 (Video)

映像圧縮は、DCT ベースの圧縮方式を採用し、I、P、B フレーム構造と動き補償を組み合わせています。すべての仕様は、CD の転送速度で安定再生できることを最優先に決定されました。

4.3.1 解像度とフレームレート

標準的な仕様は、NTSC 方式で 352×240 ピクセルの 30FPS、PAL 方式で 352×288 ピクセルの 25FPS です。

GOP 構造は比較的短く、N=12、M=3 が一般的です。これは、CD の限られた帯域でもランダムアクセスを実現するために必要な設計でした。

4.3.2 マクロブロックとブロック構造

映像は 16×16 ピクセルのマクロブロック単位で処理されます。

16×16 というサイズは、動き補償の探索範囲と計算量のバランスを考慮して決定されました。1990 年代初頭のハードウェア性能でもリアルタイム処理が可能な最大

第 5 章

MPEG-2

MPEG-1 がビデオ CD で成功を収める一方、放送業界はアナログからデジタルへの移行を模索していました。しかし、MPEG-1 にはインターレース映像への非対応、高解像度のサポート不足、動き補償の精度不足といった課題がありました。

MPEG-2 は、これらの課題を解決するために開発されました。

本章では、放送とホームビデオを変革した MPEG-2 の技術的特徴と実用化、そしてその限界について解説します。

5.1 開発の背景

1990 年代初頭、デジタル技術は放送業界とホームビデオ市場の両方で、新たな可能性を切り開こうとしていました。

放送業界では、アナログ放送からデジタル放送への移行が検討されていました。デジタル放送はノイズに強く、高画質を実現し、さらに電子番組ガイド（EPG）などの付加サービスを提供できるという利点がありました。しかし、放送で使われるインターレース映像や高解像度に、MPEG-1 は正式に対応していませんでした。

また、1996 年に容量 4.7GB の DVD が登場し、映画を高画質で収録できるメディアとして期待されました。

これらの新しい用途に対応するため、MPEG-1 を大幅に拡張した MPEG-2 が 1995 年に標準化されました。

5.2 MPEG-2 Part 1 (Systems)

MPEG-1 Systems の Program Stream に加えて、MPEG-2 Systems では Transport Stream (TS) が導入されました。

TS は固定長パケット（188 バイト）を採用し、放送やネットワーク配信などエラーが発生しやすい環境に適しています。固定長パケットによりパケット境界の検出が容易で、パケットロスが発生しても後続のパケットから復帰できます。

また、複数の番組（映像、音声、データ）を一つのストリームに多重化でき、デジタル放送での複数チャンネル配信を可能にします。

5.3 MPEG-2 Part 2 (Video)

映像圧縮は MPEG-1 Video の基本的なアーキテクチャを維持しながら、いくつかの重要な改良を加えています。SD 解像度から HD 解像度までの幅広い解像度に対応し、インターレース映像のサポート、動き補償の拡張など、多くの新機能が追加されました。

5.3.1 インターレース映像対応

MPEG-1 はプログレッシブ映像のみを想定していましたが、放送業界で標準的に使われるインターレース映像を効率的に圧縮するため、MPEG-2 では複数の予測モードが導入されました。

インターレース映像では、1 フレームが時間差のある二つのフィールドで構成されます。この特性に対応するため、MPEG-2 では映像の動きに応じて三つの予測方式を使い分けます。

動きの少ない領域では、二つのフィールドをまとめて処理するフレーム予測を使用します。フィールド間の相関が高いため、まとめて処理することで圧縮効率を高められます。

一方、動きの激しい領域では、時間差による位置ズレが問題となるためフィールド予測を使用します。各フィールドを独立した画像として扱い、参照フレームの同じ時相のフィールドから予測を行います。

MPEG-2 では、マクロブロックごとにフレーム予測とフィールド予測を適応的に切り替えることで、インターレース映像でも高い圧縮効率と画質を両立しました。

5.3.2 フレーム DCT とフィールド DCT

インターレース映像では、DCT 変換の方法も 2 種類用意されています。

フレーム DCT は、マクロブロック内の走査線を順番に並べて 8×8 ブロックを構成します。動きの少ない静止画的な領域では、隣接する走査線の相関が高いため、DCT 変換後の係数が低周波成分に集中し、圧縮効率が高くなります。

フィールド DCT は、奇数フィールドと偶数フィールドの走査線を分離して、それ

第 6 章

MPEG-4

2000 年代になるとインターネットが急速に普及し、映像配信という新たな時代が幕を開けました。また、高解像度の映像コンテンツの需要も高まり、従来の MPEG-2 では対応しきれない課題が浮上しました。

MPEG-4 は、これらの課題を解決するために開発されました。

本章では、現代の映像配信社会の基盤を築いた MPEG-4 の技術的特徴と実用化、そしてその限界について解説します。

6.1 開発の背景

1990 年代後半、インターネットの爆発的普及により、映像配信の世界は根本的な変革を迎えていました。

2000 年前後から、日本では光ファイバー (FTTH)、欧米では ADSL やケーブルインターネットといったブロードバンド回線が普及し始めました。また、NTT ドコモの i モード開始により、携帯電話でのインターネット接続が一般化しました。2001 年には第 3 世代移動通信システム (3G) が開始され、携帯電話でも映像を視聴できる時代が到来しました。

その結果、低速なモバイル回線から高速なブロードバンドまで、多様なネットワーク環境で映像を配信する必要が生まれました。また、HD 解像度での高画質配信や、限られた帯域でのストリーミングにも、より高効率な圧縮技術が求められました。

これらの課題に対応するため、MPEG-4 規格が開発されました。

MPEG-4 は、1998 年から段階的に標準化された広範な規格です。従来の MPEG-1 や MPEG-2 が映像と音声の圧縮に焦点を当てていたのに対し、MPEG-4 はマルチメディア統合規格として設計されました。30 以上のパートからなる包括的な構成を持ち、映像、音声だけでなく、インタラクティブなコンテンツや 3D オブジェクトの扱いにも対応しています。

6.2 MPEG-4 Part 2 (Visual)

MPEG-4 Visual は、1999 年に標準化された映像圧縮規格で、MPEG-2 からさらに圧縮効率の向上を目標に開発されました。インターネットでの映像配信、モバイル通信、ビデオ会議など、多様な用途に対応するため、スケーラブルで柔軟な設計が採用されました。

6.2.1 動き補償の高度化

MPEG-4 Visual では、MPEG-2 の固定的な動き補償から大きく進化しました。

1/4 ピクセル精度の動き補償を導入し、さらに細かい動きの追跡が可能になりました。また、双線形補間フィルターによる中間画素の生成により、予測誤差を削減します。

可変ブロックサイズ (16×16 と 8×8) の採用により、映像の内容に応じて適切なブロックサイズを選択できます。MPEG-2 の固定ブロックサイズから進化し、動きの複雑さに応じた柔軟な符号化が可能になりました。

6.2.2 Global Motion Compensation

Global Motion Compensation (GMC) はカメラのパン、ズーム、回転といった画面全体の動きを、効率的に表現する技術です。

アフィン変換や透視投影変換を使用して、画面全体の動きを数個のパラメーターで記述します。個別のマクロブロックごとに動きベクトルを送信する代わりに、画面全体の変換パラメーターを送信することで、大幅なビット削減が可能です。

6.2.3 符号化ツールの改良

AC 係数予測 (ACpred) により、隣接ブロックの DC 係数と AC 係数から現在のブロックの係数を予測します。これにより、空間的な冗長性をより効果的に削減できます。

量子化方式は、MPEG 量子化と H.263 量子化から選択可能で、コンテンツの特性に応じて最適な方式を使用できます。また、4MV モード (4-Motion Vector mode) により、一つのマクロブロックに対して四つの動きベクトルを使用でき、より複雑な動きに対応します。

第 7 章

次世代映像圧縮技術

2010 年代に入り、映像圧縮技術は 4K、8K 時代という新たな転換期を迎えました。

H.265/HEVC が 4K 映像の実用化を実現し、AV1 がロイヤリティフリーという革新的な選択肢を提供しました。そして、H.266/VVC が究極の圧縮効率を追求しています。

本章では、これら次世代コーデックの技術革新と課題、そして映像圧縮技術の未来について解説します。

7.1 4K 時代の標準コーデック H.265/HEVC

2010 年代に入ると、映像配信の世界は新たな転換期を迎えていました。

4K テレビの開発が本格化し、HD 解像度の 4 倍にあたる画素数を持つ超高精細映像が現実的になりました。しかし、H.264 で 4K 映像を配信するには膨大なビットレートが必要となり、より効率的な圧縮技術が求められました。また、マルチコアプロセッサーの普及により、並列処理に対応したコーデック設計も重要になりました。

これらの課題に対応するため、H.265/HEVC (MPEG-H Part 2, High Efficiency Video Coding) が 2013 年に標準化されました。

7.1.1 H.265/HEVC の技術的革新

H.265 は H.264 の基本構造を維持しながら、最大 64×64 ピクセルの Coding Tree Unit (CTU) を導入しました。この CTU は Quad-Tree 構造による再帰的分割により、映像特性に応じた柔軟な符号化を実現しています。イントラ予測は従来の 9 方向から 35 方向 (33 の角度予測 + DC 予測 + Planar 予測) へと、大幅に拡張されました。動き情報の符号化では、Advanced Motion Vector Prediction (AMVP) と

Merge モードが、隣接ブロックの動き情報を効率的に活用します。

量子化歪みの補正には、Sample Adaptive Offset (SAO) フィルターを導入しました。並列処理では、Tiles 機能で画面を矩形領域に分け、Wavefront Parallel Processing (WPP) で CTU 行単位の並列処理を可能にすることで、マルチコアプロセッサーの性能を引き出せるようになりました。これらの革新により、4K 映像のリアルタイムエンコード、デコードが現実的になりました。

7.1.2 H.265/HEVC の課題

H.265 は技術的には H.264 の約 2 倍の圧縮効率を実現し、4K 映像の実用化に貢献しました。

最大の課題は特許ライセンスの複雑さでした。MPEG LA、HEVC Advance、Velos Media という三つのパテントプールが存在し、それぞれ個別交渉が必要でライセンスコストの予測が困難でした。

技術面でも、エンコード計算量が H.264 と比較して 2~10 倍に増加しました。さらに、WebRTC などのリアルタイム通信への対応が不十分でした。

これらの課題が新たな潮流を生み出すきっかけとなります。

7.2 ロイヤリティフリーに挑戦する AV1

H.265 の複雑なライセンス状況は、映像業界に大きな転換をもたらしました。

2015 年 9 月、Amazon、Cisco、Google、Intel、Microsoft、Mozilla、Netflix の 7 社が中心となり、Alliance for Open Media (AOMedia) が設立されました。現在では、Apple、Facebook (現 Meta)、Samsung、ARM、NVIDIA なども加わり、40 社以上が参加しています。AOMedia の理念は、オープンでロイヤリティフリーな映像技術の開発です。

7.2.1 AV1 の技術的革新

2018 年 3 月に公開された AV1 は、Google の VP9 と Xiph の Daala の技術を統合したロイヤリティフリーコーデックです。

AV1 最大の特徴は、フィルムグレイン合成機能です。映画のフィルムノイズを分離して符号化し、デコード時に再合成することで、芸術的表現を低ビットレートで実現しています。画質向上では、Constrained Directional Enhancement Filter (CDEF) と Loop Restoration Filter (LRF) の 2 段階ループフィルターを導入しました。

予測技術では Warped Motion 機能により、アフィン変換や透視変換を用いた複雑