MPEG-4の動向

渡辺 裕

SC29専門委員会幹事, NTT

239-0847 横須賀市光の丘1-1
NTTヒューマンインタフェース研究所 画像通信研究部 922A
phone: 0468-59-2823, fax: 0468-59-2829, e-mail: *hiroshi@nttvdt.hil.ntt.co.jp*

MPEG-4は, 単純に高圧縮を目指した従来のマルチメディア符号化と異なり, オブジェクトベースの符号化を採用することによって, 新しいインタラクティビティの実現を目指している. 符号化の対象は自然画像や自然音響だけでなく, 合成画像や合成音声も含まれる. さらに複数のオブジェクトを自由に操作して合成, 表示することができる. また, モバイルアプリケーションを意識して誤り耐性が強化されている点も従来にない特徴である. MPEG-4東京会合 (98/3/16-20) から2カ月後に, 各パートのFCD (Final Committee Daft) が作成され, これからFCD投票に移行するところである. 本稿では, 来年にバージョン1の標準化作業が終了する予定であるMPEG-4の動向について述べる.

Trends on MPEG-4

Hiroshi Watanabe

Japanese SC29 committee Secretariat, NTT

1-1 Hikarinooka, Yokosuka, 239-0847 JAPAN
Visual Communication Lab., NTT Human Interface Labs.
phone: 0468-59-2823, fax: 0468-59-2829, e-mail: *hiroshi@nttvdt.hil.ntt.co.jp*

Main target of MPEG-4 is realization of new interactivity given by object based coding. This feature is quite different from the conventional multimedia coding which pursues high compression. Inputs to an MPEG-4 encoder are synthetic video and audio signal as well as natural ones. In addition, it is possible to display video and to playback audio by manipulating each objects. Strong error resilience is given to MPEG-4 by taking mobile applications into account. A FCD (Final Committee Draft) of each part was created two months later after the MPEG-4 Tokyo meeting (98/3/16-20), and it now goes into the balloting phase. In this report, trends on MPEG-4 in which the standardization work for version-1 will be finished next year are described.

1 まえがき

MPEG-4では, 自然 / 合成のビデオ, オーディオ情報をオブジェクト単位に符号化し, シーン記述情報に基づいて合成して表示する符号化標準である. MPEG-4を技術面から大別すると, CG (Computer Graphics) を含む映像の符号化, 音声や音響の符号化, 同期や多重化分離を表示合成のための処理の3つに分けられる.

MPEG-4の映像のうち, ビデオについては, 当初圧縮効率の向上を目標としていた. しかし実験の結果, 圧縮性能は大幅には改善できないことが判明した. そこで, オブジェクト単位のビデオ符号化に目標が切り替えられた. オブジェクトは自然動画に含まれるものに限らず, 3次元のメッシュとテクスチャ符号化で表現されるCGも含んでいる. 表示の段階で, VRMLにより視点の移動に対応した映像を得ることができる. また, アニメーションもビデオオブジェクトの一つである. 顔と胴体を表現するためのパラメータ符号化が検討されている. MPEG-4では, 自然動画だけでなくCGも扱うため, 規格はビデオ (Video) ではなくビジュアル (Visual) と呼ばれる. 小さな差であるが, MPEG-1やMPEG-2とは別次元のものであることを示している.

MPEG-4オーディオは, ビットレートと入力の違いによって, 楽音の他に音声, パラメトリック音声 / 楽音, 合成音声 / 楽音の符号化に分けられる. 通常のオーディオ符号化は, 画像と同じようにMPEG-2型の何らかの変換符号化が用いられる. 音声に対しては符号励起線形予測符号化, CELP(Code Excited Linear Prediction Coding) が適している. より低いビットレートではパラメータで音声を表現する手法, HVEC(Harmonic Vector Excitation Coding) とパラメータで楽音を表現する手法, IL(Indivisual Line Coding) が検討されている. また合成音声はTTS(Text-to-Speech) と呼ばれ, テキストを音声に変換する技術である.

MPEG-4システムでは, 任意の個数のビジュアルオブジェクトとオーディオオブジェクトをシーン記述情報に基づいて組合せて表示する. DMIF(Delivery Multimedia Integration Framework) では, ディジタル蓄積メディアへのインタフェースとリアルタイムインタフェースを規定する. システム中の仕組みとして, 符号化データであるビットストリームを違法コピーから保護するための著作権 (IPR, Intellectual Property Right) 保護の仕組みが検討されている. さらに, オブジェクトの内容の検索を行うための情報 (OCI, Object Content Information) が付加できる (この機能を強化したものを標準化しようという試みがMPEG-7である).

本稿では, これらMPEG-4の規格の各パートの内容について解説し, 今後どのような方向に向かって行くのかについて述べる.

2 MPEG-4の作業分担と規格の構成

2.1 サブグループと坦務

MPEG-4の標準化作業は表1に示すように9つのサブグループによって進められている. システムグループは同期, 多重化およびシーン記述による表示手法に関する規格の作成を行う. ビデオグループは自然動画像の符号化の規格作成を行う. SNHCグループは合成画像 / 合成オーディオの符号化の規格作成を行う. オーディオグループは自然音響 / 音声の符号化の規格作成を行う. DMIFグループはストリーム配送のための制御インタフェースの規格作成を行う. リクワイヤメントグループは, 要求条件の整理とプロファイルとレベルの作成を行う. これらのプロファイルとレベルの仕様は, それぞれの規格に含まれる. テストグループは主観品質試験手法を確立し, 評価試験を行う. インプリメンテーショングループはソフトウェアとハードウェアによるMPEG-4エンコーダ / デコーダの演算量の評価を行う. リエゾングループは外部標準化機関との調整を坦務としている.

2.2 規格の構成

規格は, 前記のサブグループがそのまま各パートに対応しているわけではなく, メディア別に統合されている. 表2にPart1からPart6の規格作成に関する今後の作業予定を示す. Version-1は, システム (Systems), ビジュアル (Visual), オーディオ (Audio), 適合性試験 (Conformance Testing), 規範ソフトウェア (Reference Software), マルチメディア制御 (DMIF) の6部構成からなる. 従来の動画像という意味でのビデオ (Video) に加えて, 合成画像 (SNHC) も扱われることから, ビデオ符号化の規格はビジュアルと呼ばれる. まずVersion-

表 1: MPEG-4 のサブグループ構成

サブグループ名	業務内容
Systems	同期，多重化およびシーン記述による表示手法の規格作成
Video	自然動画像の符号化の規格作成
SNHC	合成画像 / 合成オーディオの符号化の規格作成
Audio	自然音響 / 音声の符号化の規格作成
DMIF	ストリーム配送のための制御インタフェースの規格作成
Requirements	要求条件の整理とプロファイルとレベルの作成
Test	主観品質試験手法の確立と品質評価
Implementation	ソフトウェア / ハードウェア実現した際の演算量の評価
Liaison	外部標準化機関との調整

1 がリリースされ，その後残ったツール群を整理して Version-2 が作成される予定である．Version-2 は Version-1 とは異なったものではなく，Version-1 を包含する形になる．したがって，規格は修正票 (Amendment) として発行される．また，DMIF はデコーダ側からエンコーダ側にアプリケーションに依存した情報を送るためのインタフェースを規定する．MPEG1 のように蓄積メディアからのデータの一方的な読みだしにとどまらず，逆にデコーダの条件に応じてエンコーダを制御することが可能である．通信などの双方向アプリケーションへの仕組みは DMIF に含まれる．

表 2: MPEG-4 の Work Plan

Number	Title	WD	CD PDAM PDTR	FCD FPDAM FPDTR	DIS DAM DTR	IS AMD TR
14496-1	Systems		97.10	98.7	98.10	98.12
14496-2	Visual		97.10	98.7	98.10	98.12
14496-3	Audio		97.10	98.7	98.10	98.12
14496-4	Conformance Testing		98.12	99.7	99.12	00.02
14496-5	Reference Software		97.10	98.7	98.10	98.12
14496-6	DMIF		97.10	98.7	98.10	98.12
14496-1/Amd 1	Systems extensions	97.10	98.12	99.7	99.12	00.02
14496-2/Amd 2	Visual extensions	97.10	98.12	99.7	99.12	00.02
14496-3/Amd 3	Audio extensions	97.10	98.12	99.7	99.12	00.02
14496-5/Amd 5	Reference software extensions	97.10	98.12	99.7	99.12	00.02
14496-6/Amd 6	DMIF extensions	97.10	98.12	99.7	99.12	00.02

3 規格の内容

3.1 ビジュアル

ビジュアルには，複数のツール (技術) を集めてある機能を満たすことのできる仕様 (プロファイル) とレベル (規模のパラメータ) が定義される．プロファイルはビットストリームのシンタクスのサブセットである．現在の定義では，コンビネーションプロファイルが，ビットストリーム互換の基準点に相当する．コンビネーションプロファイルは複数のオブジェクトプロファイルの組合せからなる．MPEG-2 ではビデオだけであったから，コンビネーションプロファイルという概念はなく，単一のオブジェクトプロファイルのみであった．MPEG-4 では，アニメーションのプロファイルやスケーラブルテクスチャーのプロファイルなど，組み合わせて使用する

技術が出現する．そのためコンビネーションプロファイルが導入された．これらの名称が混乱を招きやすいため，現在名称は再検討されている．

ビジュアルコンビネーションプロファイルがどのようなオブジェクトプロファイルからなるかを表3に示す．またビジュアルオブジェクトプロファイルがどのようなツールで成立っているかを表6に示す．

表 3: ビジュアルコンビネーションプロファイル

Combination Profile	Object Profiles									
	Simple	Core	Main	Simple Scalable	12bit	Basic Anim. 2D	Anim. 2D Mesh	Simple Face	Simple Scalable Texture	Core Scalable Texture
Simple	√									
Simple B-VOP Scalable	√			√						
Core	√	√								
Main	√	√	√						√	√
12bit	√	√			√					
Simple Scalable Texture									√	
Simple FA								√		
Hybrid	√	√				√	√	√	√	√
Basic Anim. 2D Texute						√		√	√	√

3.2 オーディオ

オーディオのコンビネーションプロファイルを表4に示す．ヴィジュアルと同様にコンビネーションプロファイルはオブジェクトプロファイルの組合せからなる．AAC, TF(Time Frequency Mapping Coder), Twin VQ は音響の符号化技術, CELP は音声, HVXC と HVLN はパラメトリック楽音, TTSI はテキスト音声変換である．AAC も Twin VQ も時間信号の周波数成分への変換符号化である TF コーダの一種である．

表 4: オーディオコンビネーションプロファイル

Combination Profile	Object Profiles									
	AAC Main LC, SSR	TF, TF Main Scalable	TF LC Scalable	TwinVQ Core	CELP	HVXC	HILN	Main Synthetic	Wavelet Synthesis	TTSI
Low Rate Synthesis									√	√
Speech					√	√				√
Scalable	√	(√)	√	√	√	√	√		√	√
Main	√	√	√	√	√	√	√	√		√

3.3 システム

システムは同期多重に関して複数のプロファイルを持たないが, シーン記述に関してはいくつかのプロファイルを持っている. それらを表5に示す. シーン記述は BIFS(BInary Format for Scenes) により行われる. BIFS は VRML の拡張になっている. 2D やアニメーションやオーディオへのノードは MPEG-4 固有のものである.

表 5: システムシーン記述プロファイル

Scene Description Profile	Nodes		
	ROUTEs	BIFS Animation	BIFS Updates
Simple			√
2D	√	√	√
VRML	√		√
Audio	√		√
Compleite	√	√	√

4 課題

著作権保護および内容検索のためのデータはオブジェクト単位に附属することになる. Version-1 の MPEG-4 ビットストリームには著作権保護機能が含まれていない. これは Version-2 に含まれることになるが, その時点で Version-2 デコーダは Version-1 ビットストリームがそのままでは再生できない仕組みとなる. すなわち, 鍵の合わない海賊版のビットストリームを再生しないデコーダとなる可能性があり, ユーザの利便性と対立する. 著作権保護および内容検索のための具体的な仕組みは今後の課題となる.

また, デバイスやユーザインタフェースに依存せずに API をデータと共にダウンロードした後に動作するような環境を提供できるような MPEG-4 API の標準化も Version-2 に含まれている. これには Java をベースとすることが決定されているが, 真にプラットフォームに依存しないかどうか, 注意する必要がある.

5 むすび

本稿では, MPEG-4 の動向について述べた. MPEG-4 は圧縮効率よりもオブジェクト単位の操作, 著作権保護, 内容検索機能, インタラクティブな操作といった機能の高度化という観点で標準化作業が進められていることを示した.

6 略語・用語

AAC	Advanced Audio Coding
BIFS	Binary Format for Scene description
CELP	Code Exited Linear Prediction
DMIF	Delivery Multimedia Integration Framework
FBA	Facial and Body Animation
HVXC	Harmonic Vector eXcitation Coding
IL	Indivisual Line Coding
TTS	Text To Speech
VRML	Virtual Reality Modeling Language
Alpha	画像を重複させるとき透過率を指定する 2 次元データ
Mesh	幾何形状の表面を連続に覆う網構造のデータ
Sprite	背景画像データ

表 6: ビジュアルオブジェクトプロファイル

Visual Tools	Object Profiles									
	Simple	Core	Main	Simple Scalable	12bit	Basic Anim. 2D	Anim. 2D Mesh	Simple Face	Simple Scalable Texture	Core Scalable Texture
Intra Coding Mode (I-VOP)	√	√	√	√	√		√			
Inter Prediction Mode (P-VOP)	√	√	√	√	√		√			
AC/DC Prediction	√	√	√	√	√		√			
Slice Resyncronization	√	√	√	√	√		√			
Data Partitioning	√	√	√	√	√		√			
Reversible VLC	√	√	√	√	√		√			
4MV, Unrestricted MV	√	√	√	√	√		√			
Binary Shape Coding		√	√		√	√	√			
H.263/MPEG-2 Quantization Tables		√	√		√		√			
P-VOP based temporal scalability Rectangular Shape		√	√	√	√		√			
P-VOP based temporal scalability Arbitrary Shape		√	√		√		√			
Bi-directional Prediction Mode (B-VOP)		√	√	√						
OBMC			√							
Temporal Scalability Rectangular Shape				√						
Temporal Scalability Arbitrary Shape										
Spatial Scalability Rectangular Shape				√						
Static Sprites			√							
Interlaced tools			√							
Grayscale Alpha Shape Coding			√							
4 to 12bit pixel depth					√					
2D Dynamic Mesh with Uniform Topology						√	√			
2D Dynamic Mesh with Delauny Topology							√			
Facial Animation Parameters								√		
Scalable Wavelet Texture rectangular						√	√			√
Scalable Wavelet Texture spatial scalable							√		√	√
Scalable Wavelet Texture shape adaptive						√	√			