

## ■2群 (画像・音・言語) -11編 (マルチメディア)

# 4章 マルチメディアアプリケーション

(執筆著：佐藤真一) [2011年2月 受領]

### ■概要■

本章では、マルチメディアを活用した様々な応用技術について述べる。これまで述べてきたように、マルチメディア、特に音響・画像・映像などの情報は、人間には直感的で理解しやすいが、計算機には扱いにくい情報であると言える。これまでの章で、こうした特性を明らかにするとともに、それを解決するための様々な技術、すなわちマルチメディア情報のデジタル化、意味情報を抽出するためのコンテンツ解析技術、様々な加工技術、メタデータ関連技術、圧縮技術などについて述べた。これらの技術により、マルチメディア情報は、人間に対する直感的で理解しやすいという特性を最大限に利用しながら、計算機による様々な処理が可能となってきた。こうした技術に支えられ、マルチメディアを活用した様々な応用技術・サービスが生まれている。本章では具体的にこれらの応用について紹介する。

まずは、教育支援として、講義のマルチメディアコンテンツ化、遠隔講義などの技術について概説する。ついで、エンターテインメント・ゲームとして、産業的に大きな成功を収めているゲーム技術を中心にエンターテインメントにまつわる応用について概説する。モバイルマルチメディアでは、携帯電話並びに関連するサービスについて、特にマルチメディアとのかかわりを中心に述べる。放送応用では、放送のための制作支援にまつわるコンテンツ解析技術、メタデータ関連技術、更には視聴者のための技術などについて概説する。検索システムでは、実際に運用されているマルチメディア検索技術も含めて関連技術について述べる。

### 【本章の構成】

本章では、教育支援(4-1節)、エンターテインメント・ゲーム(4-2節)、モバイルマルチメディア(4-3節)、放送応用(4-4節)、検索システム(4-5節)について述べる。

## ■2群 - 11編 - 4章

### 4-1 教育支援

(執筆者：角所 考) [2010年2月 受領]

教育は教員と学習者のコミュニケーションであり、マルチメディア技術は人間同士の情報伝達媒体(メディア)を扱う技術であることから、両者は親和性が高く、様々なアプリケーションが考えられるが、これまでのところは、教育機関における主たる教育形態が一斉授業型の講義であることから、遠隔講義や講義アーカイブ化など、講義の受講における時空間的制約を緩和することを目指したアプリケーションが多い。また、eラーニングの普及により、今後はこれを対象としたアプリケーションも重要となると思われる。

教育支援とはつまるところ教員-学習者間のコミュニケーション支援であり、技術的には一般のコミュニケーション支援と重複する部分も多い。その一方でコミュニケーションの目的が明確であり、かつそれが行われる環境が特徴的であることから、特有の要求や解決策も存在する。本節ではこれらについて概観する。

#### 4-1-1 人物位置検出 - 講師・受講者位置の検出

##### (1) 講師位置の検出

遠隔講義や講義アーカイブ化のために講師を追跡撮影するには、その位置を検出する必要がある。講義室は外光の影響を受けにくい設計になっており、照明条件の変化は比較的少ないが、講師の後ろには黒板やスクリーンが存在することが多く、動的な背景変化は避けられない。この解決策としては、板書による背景変化は講師の存在する部分のみで生じる、スライドによる背景変化は内容が既知であるといった講義特有の知識の利用が有効である。

##### (2) 受講者位置の検出

講義映像の臨場感を増す上で受講者映像の挿入は有効である。このための受講者位置検出には、天井カメラの利用が有効である。この画像の背景差分やフレーム間差分により物体傾

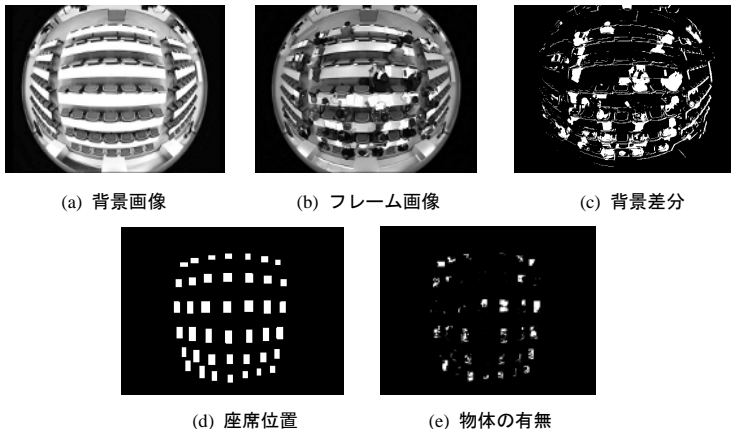


図 4・1 天井カメラと座席位置制約による物体の有無の検出例<sup>1)</sup>

域や動きのある領域が得られる。また、講義室中の座席位置は決まっているため、これを制約に用いることで正確な位置が絞りやすい。図 4・1 は背景差分にこの制約を適用した例である。

### (3) 質問者位置の検出

質問者は受講者中で唯一の発話者であることから、マイクアレイを用いた音源定位が利用できる。講義室では受講者の着席状態によって反響特性が変化するうえ、机による反射も起きやすいため、一般には正確な位置推定が難しいが、(2)項で述べた方法で受講者の位置を求め、これを制約に用いると、座席単位で質問者位置を推定することが可能である。

## 4-1-2 動作・状況認識 — 講義状況の識別

### (1) 講義状況の識別

講義映像は長時間のため、必要な箇所を見つけるためのインデキシングが欠かせない。スライド説明の部分では個々のスライド自体をインデックスとして活用できるが、説明済のスライドがそのまま放置されている状況も多いことから、スライド説明の有無を識別する必要がある。他にも板書説明や受講者への語りかけなど、識別が有用な講義状況は多いが、識別の手がかりとなる観測特徴は曖昧である。例えば指示棒を利用してスライド説明を識別しようとしても、スライド説明中に指示棒がスクリーンに接近しない状態は頻発するため、信頼性は高くない。しかし各状況での行動パターンや状況遷移のパターンを長期間観測してみると、講師ごとに特徴が見られることがあり、これをモデル化すると状況識別に有効である。

### (2) 指示対象の認識

スライド説明時にスライド中の特定箇所が指示棒で明示的に指示される場合には、それをより粒度の細かいインデックスとして利用できる。このとき、指示対象が点の場合には指示棒の位置が、領域の場合には指示棒の軌跡内部が、物体形状の場合には指示棒の軌跡形状が、物体配置の場合には指示棒の移動方向が、それぞれ指示対象を表すことになり、指示棒の位置と指示対象の関係は指示対象の種類に依存する(図 4・2)。そこでこのような指示対象の種類を、指示棒の動きや講師の立ち位置・姿勢などから認識すれば、指示対象の特定に役立つ。

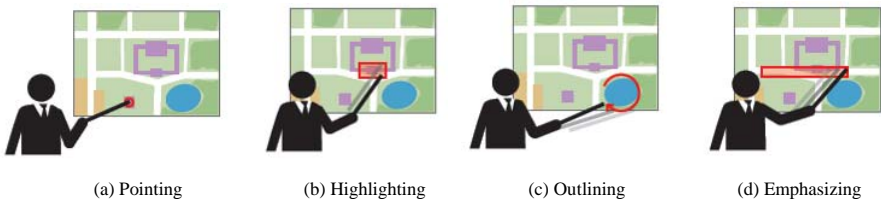


図 4・2 指示動作の種類<sup>2)</sup>

## 4-1-3 イメージモザイクング — 広視野高解像度の講義室画像合成

講義では、講師の姿や黒板、スライドなど、注視すべき複数の対象が存在するが、通常の講義映像は、カメラの画角と解像度の制約から、各時点で重要と思われる注視対象のみを選択的に撮影している。これに対し、イメージモザイクングによって講義室の広視野高解像度

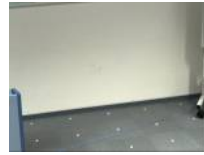
画像を合成できれば、注視対象の選択自由度が向上する。このとき平面射影変換を適用すると、講師が平面でないため、最も重要な注視対象である講師部分に歪みが生じる。また講師の姿勢は時間変化するため、その3次元形状を予めモデル化しておくことも難しい。更に黒板やスクリーンは平面であるが、やはり時間変化する。この問題に対し、4-1-1項で述べた講師の追跡撮影処理が活用できる。講師追跡撮影にレンズ中心でパンチルトズーム (PTZ) 可能なカメラを用いれば、その映像は同一 PTZ での講師映像に変換できるため、スクリーンと黒板を専用固定カメラで常時撮影し、平面射影変換して講師映像に重畳すれば、原理的に歪みのない広視野高解像度画像が合成できる (図 4・3)。



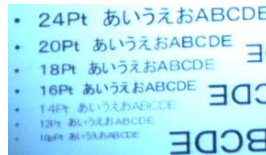
(a) 講師追跡画像



(b) 講義室の壁面画像の例



(c) スクリーン画像



(d) (c)を拡大したもの



(e) (a)に(b)(c)などを合成した広視野高解像度画像

図 4・3 広視野高解像度画像の合成例<sup>3)</sup>

#### 4-1-4 顔画像認識 - 受講者の継続認証

授業単位認定のためには、当該授業時間を通じた継続受講の事実が確認できる必要がある。教員が同席していない e ラーニングや遠隔講義でこれを実現するには、受講者認証が必要となるが、システムへのログインや IC カードの利用などによる能動的方法で授業中継続して認証を繰り返すのは受講の妨げとなるため、受講者の顔画像認識に基づく受動的方法での継続的認証が必要となる。このためには、講義室内の受講者を前方から撮影する必要があるが、

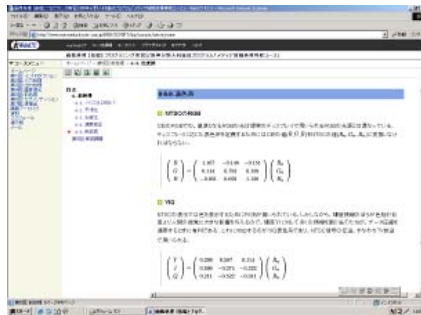
受講者同士の隠れが生じやすい。このような問題に対しては、隠れのためにカメラの各視線方向ごとに時々刻々出現・消滅する顔パターンに対し、受講者の座席位置は授業中変化しない、顔の隠れはカメラの視線方向沿いの座席位置に座っている受講者間で生じる、といった座席位置の制約を利用し、受講者の同定と座席位置の推定を同時に行うアプローチが有効である。

#### 4-1-5 顔画像合成 — 学習者の映像ログの獲得

対面型授業において、講師は受講者の様子を見ながら興味や理解度、集中度を推測し、授業の進め方を動的に変えることができる。一方、eラーニングでは、教師が学習者に付き添っていないため、学習者の様子を見ながらの指導ができず、教材を一方向的に与えるのみになりやすい。この問題を解決するには、eラーニング中の学習者の様子を教師に伝えるための学習ログが必要となる。例えば、学習者端末にカメラを設置し、映像の形で学習ログを残すことが考えられるが、その場合、学習者が教材のどのページのどの箇所に対してそのような様子を示していたのかを伝える情報も併せて必要となる。図4・4は、学習者端末に設置された2台のカメラ画像と教材画面から、画面を透かして反対側から眺めたような映像を学習ログとして合成した例である。



(a) カメラ画像



(b) 学習画面



(c) (a)~(c)からの合成画像

図4・4 学習者映像を用いた学習ログの合成例<sup>4)</sup>

■参考文献

- 1) 西口 他, “講義自動撮影における話者位置推定のための視聴覚情報の統合,” 電学論, vol.124-C, no.3, pp.729-739, 2004.
- 2) L. L. Pozzer-Ardenghi, W-M Roth, “Gestures: Helping Students to Understand Photographs in Lectures,” Connections '03, pp.1-30, 2003.
- 3) 八代 他, “講師追跡撮影カメラと平面对象撮影カメラを併用した講義室の高解像度画像合成,” 信学論(D), vol.J92-D, no.2, pp.236-246, 2009.
- 4) 中村 他, “e-learning 環境における学習者の顔情報把握のための視覚的インタフェース,” ヒューマンインタフェース学会論文誌, vol.8, no.4, pp.527-536, 2006.

## ■2群 - 11編 - 4章

### 4-2 エンターテインメント・ゲーム

(執筆者：大場章男) [2010年2月 受領]

ゲーム機は、元来、文字、映像、動画、音声など複数の情報とそれらを対話的に処理するプログラムをパッケージ化したコンテンツのプレイヤー（メディアプレイヤー）であり、マルチメディアプレイヤーである。

特に大容量メディアであるCDROMを搭載したプレイステーション世代以降では、自然動画や、CDサウンドといった大容量のデータが扱える。更に半導体などの最新の技術導入により、3次元ポリゴンによる3次元リアルタイムコンピュータグラフィクスも可能にした。その後は、インターネットの普及に伴い、ネットワークゲームのように多人数双方向のVRによる対話型環境も実現してきている。このようにゲーム機は常にスーパーセットのマルチメディア環境とビジネスを提供し、リードしてきた。

近年は、テレビのハイビジョン化に対応した高画質化が行われる一方、従来のキypadとGUIによるインタフェースに加え、加速度センサや振動によるフォースフィードバックなど、多様な入力装置に導入され、より直感的な表現と入力インタフェースを実現している。マイク入力や、カメラ入力による音声認識、画像認識（動き、顔認識など）技術も導入され、ゲーム機に対する操作入力もマルチメディア化が進んでいる。

本節では、4-2-1項でゲームにおけるマルチメディアコンテンツの発展、4-2-2項でそれらを支える技術、4-2-3項で最近の動向、最後の4-2-4項で課題として、使う側のリテラシー、リテラシーを育むオーサリング環境の重要性について述べる。

#### 4-2-1 ゲームにおけるマルチメディアコンテンツの発展

家庭用ゲーム機は、家庭のテレビを使い、文字、映像、動画、音声など複数の情報とそれらを対話的に処理し、ユーザに提示してきた。ここでは、家庭用ゲームコンソールに注目し、1983年に登場したファミリーコンピュータからの2005年のプレイステーション3にいたる代表的ゲーム機におけるマルチメディアコンテンツの構成要素の進展を示す。

表4・1に各世代のゲーム機で使われた文字、映像、動画、音声の表現法を示す。「文字、映像、動画」、「音声」にはそれぞれ専用のハードウェア回路が使われた。特に初期のゲーム機では、効率的なスプライト表現をグラフィクスに用いることにより、PCに比べ非力な計算リソースにもかかわらずリアルタイムアニメーションによる対話表現をいち早く実現した。

プレイステーション世代からは、大容量の光メディアと最先端半導体技術が投入され、MotionJpegによる動画像再生やCD品質の音声再生に加え、テクスチャマッピングされたポリゴンを用いた3次元CG表現によるリアルタイムアニメーションも実現した。リアルタイムの3次元CGにより、フライトシミュレータ、ドライビングシミュレータなどのシミュレータ型のゲームも登場し、リアルな挙動、映像生成のためにより多くの計算リソースが投入されるようになった。また、3次元CG表現はインターネットを使ったバーチャルリアリティによるコミュニケーションゲームも可能にした。

表 4・1

	文字	画像	動画	音声
FC 1983~	ドットフォント	ドット絵、スプライト合成	スプライトアニメーション	合成音(波形合成)
SFC 1990~	ドットフォント	ドット絵、スプライト合成	スプライトアニメーション	サンプリング音声合成(8音) 16bit ステレオADPCM
PS 1994~	ドットフォント 階調フォント	ドット絵、スプライト(ポリゴン)合成 写真、CG画像(ビットマップ) 3次元CG(ポリゴン)	CG動画、自然動画(MotionJpeg) リアルタイム3次元CGアニメーション	サンプリング音声合成(24音) 16bit ステレオADPCM
PS2 2000~	ドットフォント 階調フォント	写真、CG画像(24bitカラー) 3次元CG(ポリゴン、曲面)	CG動画、自然動画SD(DVD:MPEG2) リアルタイム3次元CGアニメーション	サンプリング音声合成(48音) 16bit ステレオADPCM CD Sound
PS3 2006~	ドットフォント 階調フォント ベクターフォント	写真、CG画像 3次元CG(ポリゴン、曲面) プログラマブルシェーダー	CG動画、自然動画HD(BD:AVC) リアルタイム3次元CGアニメーション 物理シミュレーション	サンプリング音声合成 圧縮音声フォーマット(MP3,AACなど) 物理音源など、ソフトウェア

\* FC：ファミリーコンピュータ，SFC：スーパーファミコン，PS：プレイステーション，PS2：プレイステーション2，PS3：プレイステーション3

\* スプライトアニメーション：小画像（スプライト）群と背景画像（BG）の合成で画像を構成するスプライトグラフィクスにおいて、スプライトの移動や切り替え、背景のスクロールなどでアニメーションする手法。

#### 4-2-2 支える技術（半導体，コンテンツメディア，通信など）

ゲーム機は、その利用形態、ビジネスモデルから、同様な機能、技術で構成される PC とは異なった発展をしている。半導体や、メディア、記憶装置、通信などの PC とゲーム機の両者を支える技術は、いわゆる情報家電とともに共有され、市場形成と資本投資を促進しで相互にその発展を支えてきた。

表 4・2

	CPU	主記憶 Main Memory	メディア	画像 Graphics	通信その他
FC 1983~	CPU:8bitカスタム (1.79MHz)	SRAM:2KB 他:VRAM:2KB	マスクROM(数KB)	スプライト(52色)	
SFC 1990~	CPU:16bitカスタム (3.58MHz)	DRAM:128KB 他:VRAM:64KB, Sound:64KB	マスクROM(数MB)	スプライト(32768色)	
PS 1994~	CPU:32bitカスタム (34MHz) 66MIPS	DRAM:2MB 他:VRAM:1MB, Sound:512KB	CD(650MB)	フレームバッファ テクスチャャーポリゴン JPEGデコード	
PS2 2000~	CPU:EmotionEngine (295MHz) 6.2GFLOPS [1]	DRAM:32MB 他:VRAM:4MB, Sound:2MB	DVD(4.7GB)	広帯域バッファ、2バッファ 混載DRAM技術 MPEG2デコード	USB1.0(iLink) LAN(100M)
PS3 2006~	CPU:CellBE (3.2GHz) 200GFLOPS [2][3]	DRAM:256MB 他:VRAM:256MB	BD(50GB)	FullHD:HDMI1.3 1920x1080p 60fps Programmable Shader	USB2.0 LAN(1G) HDD

\* FC：ファミリーコンピュータ，SFC：スーパーファミコン，PS：プレイステーション，PS2：プレイステーション2，PS3：プレイステーション3

表 4・2 に各世代のゲーム機に使われた基幹技術の発展を示す。当初、数 K バイトの半導体 ROM であったコンテンツメディアは、光メディア導入後は CD (650 MB)、DVD (4.7 GB)、BD (50 GB) と進み、HD 動画で数時間に及ぶ大容量コンテンツも配信できるようになって



きている。

CPU、GPU、メモリとゲーム機の基幹部品を構成する半導体は、1.5年で性能と集積度が2倍になるというムーアの法則に従い、ゲーム機のほぼ5年周期の世代交代の間に性能（動作周波数<sup>\*1</sup>）と集積度（トランジスタ数）がそれぞれ10倍になっている。動作周波数の向上と並列化技術により世代間では計算力が100倍にも増加している。

この計算力はリアルタイムの3次元CGや動画再生などゲーム機における新しいコンテンツ表現や高品質化に使われ、ゲーム機のコンテンツへのこれらの新しいニーズが記憶メディア、半導体、コンピュータそのものの発展を促してきた。

#### 4-2-3 近年の動向

ボタン操作主体のキーボード操作がゲームの基本入力インターフェースになっている。近年は、ジャイロや加速度センサ、カメラ入力など多様なセンサを使った身体操作系のゲームが注目されブームになっている。また、2011年のアナログ停波に向け、家庭用テレビもHD化が始まっており、ゲーム機もプレイステーション3世代ではHD出力可能になり、ゲームコンテンツもHD化している。

デジタル化、HD化におけるディスプレイの多様化に伴い、近年、急速に注目されている両眼視ステレオ画像表示も家庭用テレビの視野に入ってきた。リアルタイムに映像を生成する3次元リアルタイムコンピュータグラフィクスは、両眼視ステレオ画像の生成が容易であり、両眼視ステレオ対応ゲームが注目されている。

家庭へのインターネットの普及が進行し、ゲーム機を使った多人数双方向ネットワークゲームや、ゲームコンテンツを含むコンテンツ配信などインターネットサービスが始まっている。インターネット回線とカメラやマイクを使ったAVチャットのようなマルチメディア通信環境も導入され、ゲームならではの付加価値サービスとして顔認識や音声認識などの技術の利用が今後、期待される。

#### 4-2-4 課題

近年のゲーム機の急速な技術的発達において、その品質と表現力を使いこなすユーザリテラシーの育成が追いついていない。健全な情報配信、市場を形成するため著作権保護機構やセキュリティ技術と新しい高品質な表現力のリテラシーを育む自由な、試行やオーサリングが両立するシステム環境の整備が求められている。

#### ■参考文献

- 1) K. Kutaragi et al., "A Microprocessor with a 128b CPU, 10 Floating-Point MACs, 4 Floating-Point Dividers, and an MPEG2 Decoder," ISSCC Dig. Tech. Papers, pp.256-257, Feb. 1999.
- 2) Dac Pham et al., "The Design and Implementation of a First-Generation CELL Processor," In Proceedings of the IEEE International Solid-State Circuits Conference, 2005.
- 3) T. Chen, R. Raghavan, J. N. Dale, E. Iwata, "Cell broadband engine architecture and its first implementation: a performance view," IBM Journal of Research and Development, vol.51, no.5, pp.559-572, Sep. 2007.

\*1 プレイステーション3以降では、動作周波数の向上については鈍化している。

## ■2群 - 11編 - 4章

---

### 4-3 モバイルマルチメディア

執筆中

## ■2群 - 11編 - 4章

### 4-4 放送応用

(執筆著：八木伸行) [2010年2月 受領]

アナログ放送からデジタル放送への移行に伴い、EPG (Electronic Program Guides, 電子番組ガイド) やデータ放送などのマルチメディアサービスが始まった。大容量蓄積媒体をもつ受信機で、ダイジェスト試聴など様々なスタイルで視聴することができる TV-Anytime サービスも開発されている。また、HDD/DVD レコーダやインターネットのユーザでは、映像コンテンツを検索し様々なスタイルで視聴したいという要求が増えている。映像コンテンツの多様な利用の要望を満たすためには、シーン単位でコンテンツを説明するセグメントメタデータが必要である。現在、このメタデータの自動付与技術やフレームワークの研究開発が注目されている。また、コンテンツ自体をリッチにするための研究開発も活発である。

#### 4-4-1 映像メディア解析技術を活用した映像効果

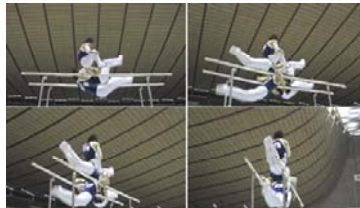
スポーツ番組では、プレイ状況や戦術などを視聴者にわかりやすく説明したり、競技を楽しませるために、様々な映像効果が開発されている。その一つが、バーチャルスタジオ技術である。例えば水泳では、世界記録ラインを撮影映像にCGで重畳して表示している。この技術に画像処理・認識技術を組み合わせることにより、更にダイナミックな映像を作り出すことができる。図4・4・1は、NHKのスポーツ中継放送で使われた例である。



(a) 投球軌跡表示例



(b) オフサイドライン表示例



(c) 多視点映像の表示例

図4・4・1 映像メディア解析技術を活用した映像効果の例

(a)は野球の投球軌跡表示例である。白いボールが小さくぼけており、白いユニフォームや背景の看板の白文字と重なると抽出が急に困難になるため、大きさや形状によるフィルタリ

ングに加え、ボールの運動法則を使い、ボールがありそうな位置を逐次予測してボールを抽出している。選手のユニフォームに重なって仮に抽出に失敗した場合にも予測を続け、ユニフォーム上を通り過ぎた後、うまくリカバーして抽出できるようにしている<sup>2)</sup>。

(b)は、サッカー競技のオフサイドラインの表示例(黄線)である。2台のハイビジョンカメラでフィールド全体を撮影し、背景差分と色により選手を抽出し、ユニフォームの色から両チームの選手、キーパー、審判を判別し、オフサイドラインを決定している。日照変化による色、影の変化への適応処理、面積や動き予測などによるクロスプレイへの適応処理により、ロバストに選手の抽出、追跡を行っている。得られたオフサイドラインは、座標変換をして、オンエアカメラ映像上に重ねている<sup>3)</sup>。

(c)は、体操競技の演技を、選手の運動履歴を残しながら多視点で表示した例である。ハイビジョンカメラ12台を選手に向けて固定して並べ撮影する。撮影された映像を選択した注視点を中心にるように仮想的に射影変換し、それを連続して提示することで注視点を中心にした多視点映像を作っている。更に、多視点の映像を対象にフレーム間差分映像を視体積交差法を用いて選手領域を抽出し、軌跡を重ね合わせている<sup>4)</sup>。

#### 4-4-2 メタデータ制作

多くの映像コンテンツが制作され放送されている。例えば、NHK アーカイブス<sup>5)</sup>には約60万本の放送済みのビデオテープが保管されている。このコンテンツを活用するために重要なメタデータに関しては、タイトルなど番組全体の情報はあるものの、セグメントメタデータは、ほとんど付与されていない。このセグメントメタデータを、番組制作過程で付与するシステムが開発されている。NHKのベアトス<sup>6)</sup>が、例である。それでもセグメントメタデータの付与作業は大変で、生番組をはじめとして、すべての番組には適用できてはいない。

これを補うために、映像メディア解析により自動的に付与する研究が活発である(2群11編2章参照)。しかし、実際にサービスに利用されるメタデータには高い信頼度が要求されるため、自動付与されたメタデータを人手で編集する必要がある。メタデータには、権利情報や書誌の情報など自動では生成できないデータもあり、これらを含めた統合的なメタデータ管理システムが研究開発されている。aceMedia プロジェクトの M-OntoMat-Annotizer<sup>6)</sup>、PrestoSpace プロジェクトの MAD<sup>7)</sup>、NHK の MPF<sup>8)</sup> などがある。MPF は、各所で開発した処理手法を統合し、手動制作もまじえて高精度のメタデータを制作するフレームワークである。このために MPEG-7 ベースのメタデータモデル、メタデータハンドリングのための API、メタデータ抽出処理をモジュール化するための共通インタフェースを規定している。統一されたメタデータ仕様により、映像コンテンツの流通も促進される。MPF では、仕様だけでなく、リファレンスソフトウェアである無料のメタデータエディタと、いくつかの解析処理を行うサンプルモジュールもソース付きで公開されている<sup>9)</sup>。

#### 4-4-3 マルチメディアコンテンツガイド

現在、日本では、かなりの地域で10チャンネル以上の無料放送を見ることができている。多チャンネル化が進み、多くの番組の中から見たい番組を探すことは困難な状況にあり、検索が目立ってある。パソコンでは一般的な検索であるが、テレビの視聴スタイルは基本的に受動的で、リモコンしかないテレビ受信機で能動的な検索を行うことは、従来の受動的視聴

スタイルと大きく異なることになる。この受動的視聴スタイルと検索を融合する提案として、CurioView<sup>10)</sup>などがある。

CurioView では、視聴しているコンテンツが興味領域であると推定し、それに関連するコンテンツを検索し提示する。関連性の評価にコンテンツ解析技術やメタデータを使用し、興味領域から広がりがあったり深みがあるコンテンツを提示する。提示されるコンテンツも番組の場合は、要約映像を提示する。要約生成には、サマリ生成、ダイジェスト生成の研究例が多くあるが、デジタル放送の特性を生かし EPG と番組中の字幕データの関連性を利用する。すなわち、EPG が番組紹介的な情報のエッセンスであると考えて、この文章に最も類似する字幕部分を抽出し、それに対応する映像区間を集めて再生する<sup>11)</sup>。

#### 4-4-4 コンテンツ活用

TV 番組には貴重な映像も多く含まれており、教育目的など様々な展開が期待される。例えば、自然番組の中から動物のシーンを抽出し整理すると、マルチメディア動物百科事典が自動生成できる。字幕（クローズドキャプション）に含まれる被写体を描写する表現を機械学習により抽出することで、動物名とその動作を少ない誤り率で抽出することができる<sup>12)</sup>。図 4・4・2 は処理結果の例で、約 20 本の番組から約 80 種の動物が映っている約 300 シーンを抽出している。



図 4・4・2 マルチメディア動物百科事典

映像クイズの自動生成例では、ニュース番組から機械学習を用いて主題がはっきりしている映像を画像認識により抽出し、3択の答えを字幕から抽出する。1択は、映像に付随したものであるが、残り二つは別の映像に付随するものである。クイズとして成立するためには、以て非なる選択肢の文を抽出するところがポイントとなる<sup>13)</sup>。

#### 4-4-5 コンテンツ生成

映像コンテンツの制作は、プロ以外には難易度の高い仕事である。ひとつの理由は映像編集やCG制作などのツールの操作が難しいことで、もうひとつの理由はストーリー、台本作りの難しさである。後者については、他書<sup>14)</sup>に譲り、ここでは、前者について述べる。

TVML (TV program Making Language) は、テレビ番組生成のために開発されたスクリプト言語で、簡単な記述で3次元のスタジオセットの中で、CGキャラクターに演技させることができる。通常のCG制作ツールのようにカメラの位置、キャラクターの動きを細かく制御する必要がない。また、撮影映像を入れ込むことも可能である。これらにより、プロ以外でも簡単に番組を制作することができる。図4・4・3に、TVMLによる映像生成例を示す。TVMLに類したものと、e-learning教材作成用のCPSL<sup>15)</sup>、擬人化エージェント用のMPML-VR<sup>16)</sup>などがある。TVMLよりも簡単に番組を記述できるTVCreatorも開発されている<sup>17)</sup>。ワープロ風のエディタで簡単なテキスト台本を書くだけで、演出が付加されたテレビ番組が生成される。スタジオセット、キャラクターの演技やカメラワークなどの演出を決めることはプロ以外には難しいが、この演出が自動で付与される。TVML、TVCreatorは、フリーで公開されている<sup>18)</sup>。



図4・4・3 TVMLによる映像生成例

#### ■参考文献

- 1) 石川清彦, 柴田正啓, 八木伸行, “映像放送サービスにおけるメタデータ制作・活用技術の動向,” 信学論, vol.J91-A, no.3, pp.296-305, 2008.
- 2) 高橋正樹, 三須俊彦, 合志清一, 藤田欣裕, “画像内の物体抽出技術を用いた高速投球軌跡作画手法,” 信学論, vol.J88-D-2, no.8, pp.1672-1680, 2005.
- 3) 三須俊彦, 高橋正樹, 合志清一, 蓼沼 眞, 藤田欣裕, 八木伸行, “実時間画像処理に基づくオフサイドライン可視化システム,” 信学論, vol.J88-D-II, no.8, pp.1681-1692, 2005.
- 4) 富山仁博, 岩館祐一, “多視点ハイビジョン映像生成システムの開発と運用,” 第11回画像の認識・理解シンポジウム論文集(MIRU 2008), pp.1063-1068, 2008.
- 5) 大久保達也, 八木伸行, “NHK アーカイブス—制作から保存まで—,” 人工知能学会誌, vol.18, no.3,

pp.224-229, 2003.

- 6) K. Petridis, D. Anastasopoulos, C. Saathoff, N. Timmermann, I. Kompatsiaris and S. Staab, "M-OntoMat-Annotizer: Image Annotation. Linking Ontologies and Multimedia Low-Level Features," 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems, 2006.
- 7) A. Messina, L. Boch, G. Dimino, W. Bailer, P. Schallauer, W. Allasia, M. Groppo, M. Vigilante, and R. Basili, "Creating Rich Metadata in the TV Broadcast Archives Environment: The PrestoSpace Project," 2nd Int. Conference on Automated Production of Cross Media Content For Multi-Channel Distribution, 2006.
- 8) 住吉英樹, 佐野雅規, 八木伸行, "メタデータ制作フレームワーク," 映情学誌, vol.61, no.2, pp.152-157, 2007.
- 9) <http://www.nhk.or.jp/strl/mpf/>
- 10) H. Sumiyoshi, M. Sano, J. Goto, T. Mochizuki, M. Miyazaki, M. Fujii, M. Shibata, N. Yagi: CurioView: TV Recommendations Related to Content Being Viewed, IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, mm2010-21, 2010.
- 11) 河合吉彦, 住吉英樹, 八木伸行, "電子番組表における紹介文を利用した番組紹介映像の自動生成手法," 信学論, vol.J91-D, no.8, pp.2157-2165, 2008.
- 12) 三浦菊佳, 山田一郎, 住吉英樹, 八木伸行, 奥村 学, 徳永健伸, "放送番組を素材としたマルチメディア百科事典の自動構築," 映情学誌, vol.62, no.1, pp.110-116, 2008.
- 13) 佐野雅規, 八木伸行, 片山紀生, 佐藤真一, "蓄積されたニュース番組からの画像付きクイズ生成手法の提案," 信学論, vol.J92-D, no.1, pp.141-152, 2009.
- 14) D.Arijon (岩本憲児, 出口丈人訳), "映画の文法—実作品に見る撮影と編集の技法," 紀伊国屋書店, 1980.
- 15) 新藤義昭, 松田 洋, 鈴木誠史, "3D-CGAnimation のシナリオ記述言語 CPSL と Cyber Teaching Assistant の開発," 情処論, vol.43, no.8, pp.2782-2796, 2002.
- 16) 岡崎直観, Santi Saeyor, 土肥 浩, 石塚 満, "マルチモーダルプレゼンテーション記述言語 MPML の 3 次元 VRML 空間への拡張," 信学論, vol.J85-D-I, no.9, pp.915-926, 2002.
- 17) 浜口斉周, 道家守, 林正樹, 八木伸行, "演出スタイルシートを用いたブログ型テレビ番組制作・公開・視聴システム," 信学論, vol.J89-DII, no.10, pp.2194-2205, 2006.
- 18) <http://www.nhk.or.jp/strl/tvml/>

■2群 - 11編 - 4章

---

## 4-5 検索システム

執筆中