研究ノート

映像要約とメタデータ・ハイディング

馬場口　登*

Video Abstraction and Metadata Hiding

Key Words: Video Abstraction, MPEG-7, Metadata, Information Hiding

1．はじめに

ニュース・ドラマ・スポーツなどのテレビ番組に代表される放送映像(broadcast video)は日々生産され，WEB空間に匹敵する大規模なマルチメディア情報源を形成している．放送映像を蓄積し，映像・画像解析を利用して，映像メディアを知的にハンドリングする手法に対する期待は大きい．すなわち，ピンポイントで所望の映像を検索することやコンパクトな映像を新たに作成することである．具体的にはスポーツ映像に対し，

・好きな選手の活躍したシーンや得点の入ったシーンを探す．

・試合全体の経過が分かる短時間の映像を作る．

ということが実現できると，マスコミュニケーションである放送をパーソナルメディアに変貌させる可能性を有する．

これらに応える映像技術として，映像インデクシング(video indexing)や映像要約(video abstraction/summarization)が考えられている．いずれも映像のセマンティックス(semantics)に関連する処理であるが，物理的な信号レベルと概念的な記号レベルとのギャップ(セマンティック・ギャップと呼ぶ)の存在から一筋縄では行かない問題である．

本稿では，筆者らが進めているMPEG-7メタデータを用いた映像要約，及び情報ハイディング(information hiding)技術を応用した映像コンテンツ管理法を紹介する．

* Noboru BABAGUCHI
1957年2月生
1981年大阪大学大学院工学研究科通信工学専攻前期課程修了
現在，大阪大学(工学部)工学研究科・通信工学専攻，教授，工学博士，画像・映像処理
TEL 06−6879−7744
FAX 06−6879−7684
E-Mail babaguchi@comm.eng.osaka-u.ac.jp

2．映像要約

映像要約の定義は，元となる映像から主要な部分(ハイライト)を抽出して，元の映像より短い記述(映像表現)を生成することである[1]．また，生成された記述から映像コンテンツの全容が把握されることが望ましい．例えば，2時間のスポーツ中継番組から試合全体の経過が分かる2分間の映像を作成することがこれに相当する．映像要約はモバイル環境での映像デリバリや映像ポータルなど様々な応用の可能性を有する処理であり，映像メディア処理においても現在のところ最も活発に研究されているテーマの一つである．

さて，映像要約は要約における表現形式に従い，空間展開型(spatial expansion)と時間圧縮型(time compression)の2タイプに大別される．

空間展開型は，画像キーフレーム(静止画)を2次元平面(ウインドウやスクリーン)上に何らかの基準で配置して提示するもので，ビデオポスタやストーリーボード(絵コンテ)のような表現となる．この表現形式はコンテンツ全体の一覧性，了解性が良いという特徴をもち，映像のコンテンツ全体を把握するには適した形式である．しかしながら，ショット等の部分映像を代表させるのに，どのキーフレームが妥当であるか，また，数千に及ぶキーフレームが出現する場合に，如何に配置するか，などの問題もある．

一方，時間圧縮型は元の映像を時間軸上で圧縮するもので，映像スキミングとも呼ばれる．表現形式は映像クリップで，我々が日常，眼にするもので例えると，スポーツニュースのダイジェストや映画の

予告編がこれに相当する．このタイプの要約では，セマンティックな観点から要約映像の構成要素を決める必要があり，映像編集の技術とも関係が深い．

ここで，映像要約において重要となるハイライトの検出について触れておく．これまでハイライトの自動検出[2]を目指して，画像解析やマルチモーダル解析を利用したものが提案されてきたが，現状でのパターン認識技術の脆弱さから，完全にハイライトを検出することは不可能である．そこで本手法では，映像データに関するセマンティックな記述がメタデータ(データのデータという意味)として得られていることを前提とする．メタデータはマルチメディアの検索などの高度利用に大きく寄与するものである．

映像を含むマルチメディアに対するメタデータの記述体系を定めたものがMPEG-7であり，2001年12月に国際標準化がなされた．MPEG-7はXML schemaなるXMLベースの記述体系であるため，インターネットとの親和性や情報共有の容易性に優れている．MPEG-7では，低レベル(信号特徴)から高レベル(セマンティックス)までを記述することが可能となる．

以下では，我々が開発したスポーツ(野球)映像に関する手法を述べる[3]．この手法は，映像要約を制約充足問題として定式化したもので，上記2タイプは何れも自然に記述できる．提案手法では，ハイライトシーンを定めるために，シーンの重要度を算出するが，この算出にプレイタイプ，プレイヤー，得点，リプレイシーンの有無などのMPEG-7メタデータ記述を参照する．

(a) 時間圧縮型

時間圧縮型の映像要約では，シーンの総数N，それぞれのシーンp_iの重要度$s(p_i)$と長さ$l(p_i)$を持つ映像に対して，要約映像の長さ(要約時間)Lが与えられたときに，どのシーンをどれだけの長さにして，Lに収めるか，すなわち，選ばれたシーンの重要度の和が最大となり，かつLに収まるようにシーンを組み合わせる，という問題に帰着できる．つまり，あるシーンp_iの時間長を変化させる関数を$\phi(l(p_i))$ $(0<\phi(l(p_i))\leqq l(p_i))$とすると，

シーン集合 $P=\{p_1, p_2, \ldots, p_N\}$ から

$$\text{制約条件} \sum_{p_j\in P'} s(p_j)\rightarrow \max, \sum_{p_j\in P'} \phi(l(p_j))\leqq L$$

の下で部分集合$P'=\{p_j \mid j=1, 2, \ldots, k\}(1\leqq k\leqq N)$を選ぶ問題に定式化される．

ここで，選択法として以下の3つの方式を考える．

1．ベーシック法：重要度の高いシーンから選ぶ．
2．グリーディ法：単位時間当たり重要度の高いシーンから選ぶ．
3．プレイカット法：$\phi(l(p_i))=\min[l(p_i), l_{th}(p_i)+\delta L]$として，重要度の高いシーンから選ぶ．ただし，$l_{th}(p_i)$はユーザがシーン内容を理解するのに最低限必要な時間長で，δは要約時間に応じてシーンの長さを変化させるパラメータである．

実際にテレビで放送された実要約映像(ダイジェスト)のプレイシーンを正解集合とみなして，各手法と比較した結果，プレイカット法が最も良好な特性を示し，70～80%のシーン一致度を得た．

(b) 空間展開型

空間展開型の映像要約では，キーフレームの総数N，それぞれのキーフレームf_iの重要度$s(f_i)$と表示面積$a(f_i)$を持つ映像に対して，映像要約を表示するスペースの面積Aが与えられたときに，どのキーフレームをどのような基準でどのように表示スペース内に配置するか，すなわち，選ばれたフレームの重要度の和が最大となり，かつAに収まるようにキーフレームを組み合わせる，という問題に帰着できる．つまり，あるキーフレームの表示面積を変化させる関数を$\phi(a(f_i))$ $(0<\phi(a(f_i))\leqq A)$とすると，

シーン集合 $F=\{f_1, f_2, \ldots, f_n\}$ から

$$\text{制約条件} \sum_{f_j\in F'} s(f_j)\rightarrow \max, \sum_{f_j\in F'} \phi(a(f_j))\leqq A$$

の下で部分集合$F'=\{f_j \mid j=1, 2, \ldots, k\}(1\leqq k\leqq N)$を選ぶ問題に定式化される．

本研究では野球映像を対象映像として扱っており，$\phi(a(f_i))$を一定値として，図1のように縦方向を上から下に向かって時間軸，横方向を左から右に向ってイベント軸としてとり(イベント-時間軸表現)，イニングごとに区切って表示する．また，以下の操作を実装している．

フレームの絞込み：現在表示されているフレームのうち重要度の高いフレームだけを残し，低いフレームを消す．この操作を繰り返すことで表示されるフレーム数は減っていく．

プレイシーンごとの再生：画面上のキーフレームをクリックすることでそのシーンを再生する．
アノテーションの提示：キーフレームの下に選手名，プレイ名を表示する．

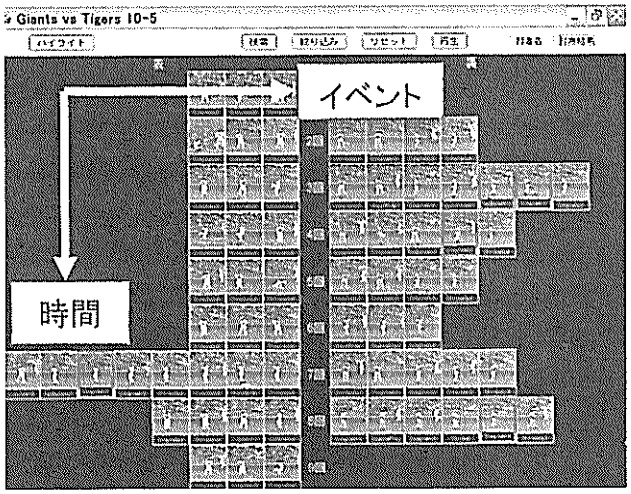

図1　イベント-時間軸表現による空間展開型要約

(c) 個人適応の導入

筆者らは，映像要約において個人適応(personalization)の考え方を導入した．筆者らの主張は，シーンの重要度はユーザの選好や興味によって変化し得るものであるため，それに応じて要約も変化すべきという点である．例えば，Seattle Marinersの試合において，通常のダイジェスト(アメリカでの放送)は得点経過に即したハイライトから構成される映像になるであろうが，Ichiroの活躍に関心をもつ者(日本の放送)には，そのようなダイジェストよりIchiroの全打席のシーンがより重要であろう．そこで提案手法では，好きな選手，好きなチーム名，見たいイベントなどの情報をあらかじめプロファイル(profile)に記述しておくことによって，その内容に応じて，各個人に適応した要約映像を生成する．詳細は[4]を参照されたい．

3．メタデータ・ハイディング

以上に述べたように，メタデータは映像の検索や要約に大きく寄与し，コンテンツ高度利用の一翼を担うことになる．ところが，映像データとメタデータが別個に存在すると，コンテンツ管理に厄介なことが生じるため，筆者らは情報ハイディング技術を援用して，メタデータ(MPEG-7)を映像データ(MPEG-1)に埋め込む手法を開発した[5]．

情報ハイディングとは，画像・映像(動画)・音声などのメディアに密かに情報を埋め込む技術で，コンテンツ流通の際のセキュリティの一手段である．目的の違いから，電子透かし(digital watermarking)とステガノグラフィ(steganography)に大別される．電子透かしは主にコンテンツの著作権などの知的所有権保護を目的とし，著作権者のディジタル著名を埋め込み，コンテンツと分離できないようにする．電子透かしで重要な情報はコンテンツであることに注意されたい．一方，ステガノグラフィとは，元々「covered writing(秘匿された書類)」を意味し，埋め込まれたデータの密かな伝達を目的とするため，ステガノグラフィで重要な情報は埋め込まれたデータである．

我々の手法では，元のMPEG-7メタデータからシーンごとに独立したメタデータを作成し，それらを関連したシーンを構成するフレームに埋め込む．本手法の利点は，(1)メタデータの保存のために，物理的に別個の記憶域が不要である(映像データ内に保存される)，(2)コンテンツ流通のためのネットワーク負荷を軽減できる，(3)映像データとメタデータを統一的に管理できる．(4)メタデータの秘匿を視覚的に認知できないので，メタデータの利用を限られたユーザ(抽出鍵を持っているユーザ)のみに制限できるなどである．これによってマルチメディアデー

図2　オリジナル映像

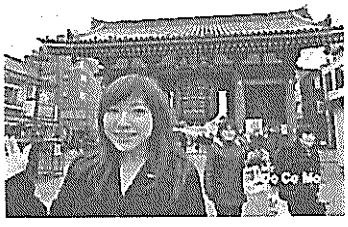

図3　埋め込み映像

タとそのメタデータの流通手段を極めてシンプルにすることができる．

埋め込み手法には，MPEG-1のDCT (Discrete Cosine Transform)領域に埋め込む方法と動きベクトル領域に埋め込む方法を用いる．DCT領域に埋め込む方法は，DCT係数の高周波成分を書き換えることによりデータ埋込を実現する．動きベクトル領域に埋め込む方法は，動きベクトルのベクトル値を表すVLC(Variable Length Code)の最下位ビットを書き換えることでデータ埋込を実現する．DCT領域，および動きベクトルに埋め込む方式各々において，40db，30dbのピークSN比を達成した．最後に，図2に元の映像，図3に埋め込み映像のフレームを示す．ほとんど違いはないことがお分かり頂けよう．

4．むすび

本稿では，映像メディアを気の利いた形に加工し，さらに安全かつ簡便に映像コンテンツを流通させ，高度利用を図る方法を紹介した．このような手法をより発展させることによって，マルチメディア情報が豊かなコミュニケーションをもたらす重要な役割を果たすであろう．

なお，本研究の一部は，情報通信研究機構・創造的情報通信技術研究開発制度，及び日本学術振興会科学研究費(学術創成研究)の補助によることを付記する．

文　献

[1] 馬場口登，上原邦昭，有木康雄："マルチメディア情報の高次処理"，人工知能学会誌，Vol.18，No.3，pp.307-316(2003-05)．

[2] N.Babaguchi, Y.Kawai, and T.Kitahashi："Event Based Indexing of Broadcasted Sports Video by Intermodal Collaboration"，IEEE Trans. Multimedia，Vol.4，No.1，pp.68-75 (2002-03)．

[3] Y.Takahashi, N.Nitta and N.Babaguchi："Automatic Video Summarization of Sports Videos Using Metadata"，Proceedings of Fifth IEEE Pacific-Rim Conference on Multimedia(PCM2004)，Tokyo(2004-12)．

[4] N.Babaguchi, Y.Kawai, T.Ogura and T.Kitahashi："Personalized Abstraction of Broadcasted American Football Video by Highlight Selection"，IEEE Trans. Multimedia，Vol.6，No.4，pp.575-586(2004-08)．

[5] S.Taniguchi, N.Nitta and N.Babaguchi："Embedding MPEG-7 Description in MPEG Video Data by Focusing on DCT-coefficients and Motion Vectors"，Proceedings of Pacific Rim Workshop on Digital Steganography 2004(STEG'04)，Fukuoka(2004-11)．