

マルチメディアデータベースにおける映像情報の入力

2W-8

草場匡宏 高橋淳一 洪政国

日本アイ・ビー・エム(株)東京基礎研究所

1 はじめに

従来コンピュータで扱われてきた文字・数値や静止画像にくらべ、ビデオを中心とする映像データは莫大な情報量をもつためにそれを管理する手段がマルチメディア・データベースを構築する上での大きな課題となってきた。またビデオ機器の急速な普及により、メディア産業のみならず一般の家庭内においてもビデオテープの数は増え続けもはやタイトル等の文字情報で管理できる限界を越えている。今日のMPEG/JPEGにみられるようなメディアのデジタル化を考えると、コンピュータ上であらゆるメディアを統合し利用を図るマルチメディア・データベースの重要性はますます大きくなっており、これらを統一的に管理するマルチメディア・データベースの構築が急務となっている[1]。

我々は映像を物理的・意味的なシーンに分割し、従来人手に頼って行なわれてきたカット割りとは映像情報の入力と管理を支援するユーザインタフェースの研究を行なっている。本研究は現在国立民族学博物館(民博)との共同研究として進められており、現在までにプロトタイプシステム(M-CIDB: Motion Color Image Database)の構築を終え、ユーザーの試用によって問題点を明らかにしながら改良を行なっている。本稿では映像情報を入力する方法について考察すると共に民博での評価実験の結果を報告する。

2 映像データベース構築の課題

映像は静止画像(フレーム)と音声の集合からなりたっているが、ある特定時間でのデータだけでは意味をもたず、連続した時間を単位とするシーンの集合からなりたっていると考えられる[2]。映像を扱うデータベースやオーサリング・システムに関してこれまでいくつか研究が行われているが、映像データベースにおける課題として次のような要件があげられている[3]。

- (1) シーンとその属性を記述するためのデータ構造
- (2) シーンやフレームの容易な検索
- (3) 直感的な検索を行なうユーザインタフェース
- (4) 動画像と音声データの同期

“Video archiving method in a multimedia database”
 Masahiro Kusaba Junichi Takahashi Hong Jung-kook
 IBM Tokyo Research Laboratory, IBM Japan Ltd., 1623-14,
 Shimotsuruma, Yamato, Kanagawa, 242, Japan

(5) 映像情報(インデックス)の作成

(6) システムの互換性

しかしこれらの課題に対して十分な解答を与える研究は行なわれていない。我々はこれらの要件の中でとくに映像情報の入力に関する課題すなわち(1)および(2)に注目して研究を行なった。

3 映像情報の構造化

映像データベースを構築するためのデータ構造を次のように定義する。

3.1 フレーム情報

映像は2次元の静止画像(フレーム)の集合体としてとらえられる。各フレームには先頭から連番あるいは時間を示すフレーム番号が付いている。各々のシーンは映像の一部であり連続したフレーム番号によって位置を特定することができる。フレーム情報としてシーンを特定する先頭・終了フレーム番号の対およびシーンの代表フレームを定義する。

3.2 シーン階層情報

シーンは図1に示すように、論理的に映像全体がより小さい単位のシーンに分割され、さらにそれぞれのシーンはより小さいシーンから成立つ。シーン階層は最終的に物理的に分割できないカットにいたる階層構造として定義する。

3.3 時系列キーワード情報

シーンの意味的な内容を示す属性は、フレーム情報とは別の時系列を基準とし、フレーム情報とは独立した時系列を各キーワードに与える。この時系列キーワード情報によって特定の内容を持ったシーンを選び出すことができる。

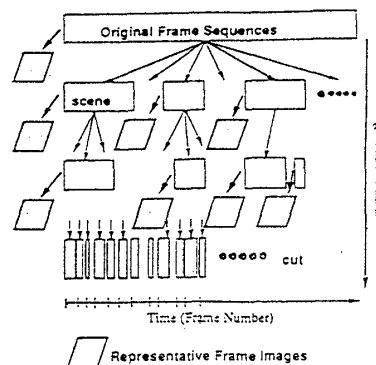


図1 シーンの論理構造

4 映像情報の入力インタフェース

映像情報の構造化によってユーザーが任意のシーンを検索する労力は減少するが、入力するためのコストは大きくなり結果的にデータベースの構築が遅れる。コンピュータが画像の内容を抽出して自動的にシーン情報を入力することは困難なため [4]、映像情報を入力する人間の操作を支援することが不可欠となる。M-CIDB では対話的にシーンのインデクシングを支援するインタフェースとしてシーン・ディテクタおよびシーン・エディタの2つのインタフェースを提供している。

シーン・ディテクタは、フレーム画像のいくつかの特徴量の時間変化を追跡することによって物理的なシーンの変化を検出し、シーンの先頭フレームのタイムコード（フレーム番号）を記録する [5]。検出されたフレームの直前のフレームは1つ前のシーンの最終フレームに相当する。検出したフレーム画像は代表画像としてリアルタイムにスクリーン上に表示する。

シーン・ディテクタによるシーン変化点の検出結果には検出ミスや誤検出が含まれる可能性がある。またシーン・ディテクタによって得られた代表画像がユーザーの要求と一致しない場合も起こり得る。そこでシーン・エディタはシーン情報をデータベースに登録するために、シーン情報を対話的に入力するインタフェースおよびユーザーによるフレーム情報の確認と修正のためのインタフェースを提供する。シーンは代表画像をシンボルとするオブジェクトとして表示されユーザーは全ての操作をオブジェクトに対する直接操作として行なう。

5 プロトタイプピング

プロトタイプシステムは、ビデオ機器の制御と画像・文字情報の入力を行なうPCとデータの蓄積および検索処理を行なうホスト計算機によって構成されている。PC側はPCおよびディスプレイとビデオ機器およびTVモニターによるハイブリッド方式を採用した。ビデオ機器としてはタイムコードによってフレームを管理できるレーザーディスクあるいは8ミリビデオデッキを使用し、RS232Cインタフェースを通じてPCからコントロールをおこなっている。映像はビデオ入力アダプタにより毎秒30フレームで取込みPC上のファイルに保管される。またシーン・エディタによって入力された画像および文字情報はホスト計算機上の関係データベースとして保存する。

シーン情報の入力について手作業によるインデクシングと本プロトタイプを使用した場合を比較した結果、3倍から6倍の効率の向上が得られた [6]。

また、民博内に本プロトタイプを設置して研究員、学芸員などの専門家および来館者への案内人など31人の被験者を対象に評価試験を行なった。その結果自分の職場で使いたいといった意見も多く肯定的な評価が多かったが、操作性とデータ量の充実に改善が期待された。

6 まとめ

本稿は、民博との共同研究の結果に基づいて映像情報の入力について論じるとともにプロトタイプ・システムの紹介と評価試験の結果について報告を行った。

今後は博物館やメディア産業のような大規模なシステムに留まらず一般家庭でのビデオテープの管理のような小規模なシステムにも研究の対象を広げていきたい。さらにこれらのシステムによって蓄えられたマルチメディアをもとにオーサリング等の新たな創作活動を行なうための基礎となるアーカイブシステムを目指して今後もインタフェースを中心とした研究を進める予定である。

7 謝辞

本報告は、国立民族学博物館と日本アイ・ビー・エム（株）との共同研究「博物館におけるマルチメディアの有効利用についての基礎研究」の一環として実施している成果の一部であり、ここで使用しているデータは民博の所有物である。

参考文献

- [1] Hong, J-K. et al., "A Motion Archiving Technology and its Application in an Ethnological Museum," DEXA92, pp207-214, 1992.
- [2] 大喜田 他, "意味データベースに基づく映像情報のモデリング," 信学会, データベースシンポジウム, pp.7-16, Feb. 1990.
- [3] 外村 他, "動画データベースハンドリングに関する検討," 信学技研, IE89-33, pp.49-56, 1989.
- [4] 上田, "インタラクティブな動画編集方式の提案," 信学技研, IE90-56, pp.39-46, 1990.
- [5] Ioka, M., "Method of detecting scene change in moving picture," IBM Technical Disclosure Bulletin, Vol. 34, No. 10A, Mar. 1992.
- [6] 草場 他, "映像データベースにおける情報の入力と管理," 情処研報, 92-CH-15-2, 1992.