

仮想現実感・マルチメディアデータベース統合 システムの構築に関する研究 マルチメディアデータの抄録法

早手 隆裕 妻鹿 仁彦 吉田 香 打浪 清一

九州工業大学情報工学部
〒820-8502 福岡県飯塚市大字川津 680-4
E-mail:{hayate,mega,kaori,uchinami}@taurus10.cse.kyutech.ac.jp

近年、マルチメディアデータを手軽に扱えるようになり、より高性能なマルチメディアデータベースの開発要求が高まっている。そこで、我々は、概念スキーマを仮想現実感データベースで実現し、真のメディア統合を実現した一元的な仮想現実感データベースシステムと、マルチメディアデータベースと仮想現実感システムを並置し、両方を相互に行き来できる二元的なデータベースシステムを提案し、実装してきた。これらのシステムを構築するにあたって、マルチメディアという膨大な量のデータから必要なデータをすばやく見つけ出すには効率的な抄録が必要である。そこで本稿では、MPEG-7を用いた動画像の内容抄録法を適用することにより、動画像抄録の自動化、効率化を図り、有効なマルチメディアコンテンツの管理法を議論する。

INTEGRATION OF MULTIMEDIA DATABASE WITH A VIRTUAL REALITY SYSTEM Indexing Method of Multimedia Data

Takahiro HAYATE Kimihiko MEGA Kaori YOSHIDA Seiichi UCHINAMI

Faculty of Computer Science and Systems Engineering
Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502 Japan
E-mail:{hayate,mega,kaori,uchinami}@taurus10.cse.kyutech.ac.jp

Recently, many people use multimedia database easily. However, conventional multimedia database cannot compensate media integrity. We proposed and developed two kinds of multimedia database which can integrate each media. The one is a multimedia database which express conceptual schema in a virtual reality database. The other is unified database of multimedia database and virtual reality database. These database system require efficient indexing method in order to find out target data quickly from huge number of data. In this paper, we propose video indexing method using MPEG-7 and show experimental results of video retrieval.

九州工業大学情報工学部

Faculty of Computer Science and Systems Engineering

Kyushu Institute of Technology

1. はじめに

近年、コンピュータの普及、高性能化により、静止画、音声、動画像などの様々なマルチメディアデータを、一括処理できるシステムの要求が高まっている。しかし、従来のマルチメディアデータベースシステムでは、単一メディアによる更新がすべてのメディアに反映されないといった点で、真の意味でのメディア統合ができていないと言える。

そこで、マルチメディアデータを元に仮想現実感システムを作成し、各メディアでの更新を仮想現実内に反映させることによって真のメディア統合を実現するデータベースシステムの実現法として、本研究室では二通りの手法を提案し、実装してきた。

2. マルチメディアデータベースシステム

2.1 仮想現実感マルチメディアデータベースシステム

第一の手法は、概念スキーマを仮想現実感データベースで実現し、マルチメディアデータを用いて仮想空間内に現実世界を忠実に再現することでメディア統合を実現した仮想現実感マルチメディアデータベースシステムである。

このシステムの出力データは、仮想現実の作成及び更新に用いられたデータではなく、現実世界を忠実に再現した仮想現実感システムから生成されたデータとなる。

このシステムは、仮想現実感システムを用いた一元論的なシステムと言える。

仮想現実感データベースシステムの概略を図1に示す。

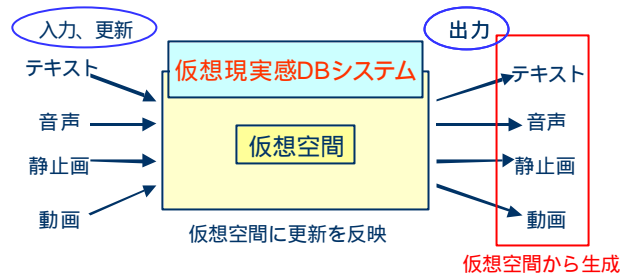


図1 仮想現実感データベースシステム

2.2 並置型仮想現実感・マルチメディアデータベース統合システム

第二の手法は、仮想現実世界を作成する元となるマルチメディアデータを格納したデータベースシステムと仮想現実感データベースシステムのそれぞれの長所を生かし、統合することで、両システムを相互に行き来できる並置型のシステムである。

2.1で述べたシステムの出力が仮想現実感データのみであるのに対し、このシステムではマルチメディアデータベース内に格納されたマルチメディアデータそのものを参照することが可能である。仮想現実データと元のデータを比較すると、自由度の面では仮想現実データが勝るが、現在の技術では元のマルチメディアデータほどのリアリティを持つことはできない。

このシステムは、仮想現実感システムとマルチメディアデータベースの二元論的なシステムと言える。

2.3 マルチメディアデータの抄録

ここで、マルチメディアデータの中でも動画はデータ量が膨大であるため、データベースに質問があってからデータを全てスキャンするのは実用的ではない。そこで、前に述べた二通りのシステムのどちらにおいても、マルチメ

ディデータの抄録が必要である。

2.1 のシステムにおいては、動画像による入力が行われれば、そのデータから仮想現実感システム中に登場する人物や対象物を認識し、それを仮想物体、仮想イベントに変換したうえで仮想現実感データベースシステムに格納する必要がある。

また、2.2 のシステムにおいて、仮想空間に登場する人物や登場物が動く場合、すなわち、イベントをデータベース化する場合、まず、登場人物や登場物がどのような動きをしているのかを把握する必要がある。このためにも動画像等の抄録が必要である。

そこで、本稿では、これらのシステムを構築するにあたって、必要となる動画像の内容抄録法について考察し、膨大な情報量を持つ動画像データベースに関してシステムに有用な抄録法を提案、検討し自動化を図る。

3. 仮想現実感・マルチメディアデータベース統合システム

3.1 システム構成

ここで、動画像抄録と仮想現実感システムの連携動作が重要となる仮想現実感・マルチメディアデータベース統合システムについて、詳しく述べる。マルチメディアデータベースシステムは、現実のデータを格納できる反面、動画像を例にとると、撮影者やデータベース作成者の目を通してしか見ることができない。つまり、撮影された動画像を別の角度から見たいといった要求に応えることはできない。それに対し、仮想現実感システムは、リアリティに欠ける反面、作成された空間内を歩き回ったり、仮想装置を操縦したりといった操作が、利用者の意図

により可能である。つまり、ユーザーが望む自由な角度からデータを体験、評価することができる。この両者を統合することにより、実際に撮影された動画像や、写真はマルチメディアデータベース内から参照でき、またその空間の構造や様子は仮想空間内で体験評価できるといったよりよいシステムが構築できる。図2にシステム構成を示す。

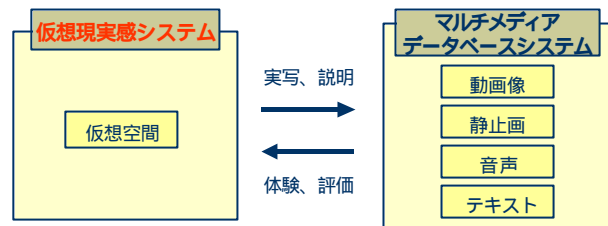


図2 仮想現実感とマルチメディアデータベースの統合

3.2 仮想現実感システム

この統合システムにおいて、仮想現実感システムの役割は、仮想オブジェクトや仮想人物を仮想空間内に表現し、必要に応じてイベントを再生することである。これは VRML データベースシステムを用いて作成する。ユーザーは動画像中のイベントと同じイベントを仮想空間内で体験し、実際の映像が見たい場面ではマルチメディアデータベースにアクセスし、動画像を閲覧することになる。

ここに登場する人物、物の動きは、以下に述べる抄録で求めたデータを用いて作成する。

4. マルチメディアデータの抄録

4.1 MPEG-7 について

MPEG-7 とは、正式名称は Multimedia

Contents Description Interface であり、MPEG-1、2、4 などの映像圧縮技術とは違い、主な標準化対象として、内容記述ツールを定義するための言語 (Description Definition Language)、記述スキーム (Description Scheme)、記述子 (Descriptor) を定めている[3]。また、テキストとして表現される内容記述データをバイナリデータとして圧縮、およびバイナリデータからテキストデータに伸長する仕組みも標準化対象としている。つまり、デジタル映像の持つ特性として挙げられる検索、ブラウズ、ナビゲーション、編集機能を容易に利用できるようにするための技術で、マルチメディアコンテンツの内容を記述するための規格である。

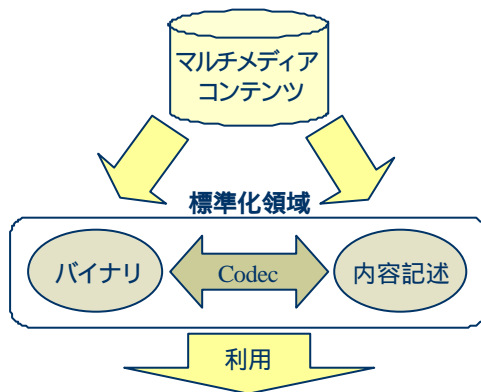


図3 MPEG-7の標準化領域

MPEG-7の内容記述(メタデータ)には、高レベルな記述と低レベルな記述の2種類がある。

図4に示すように、高レベルな記述とは映像の内容を人が見て記述するものである。これは、書籍のモデルを用いて設計されている。書籍全体を読まなくても必要とされる部分だけを拾い読みすることができるように用いられる目次や、キーワードで該当部分を見つけ出せる索引に該当する部分が、映像コンテンツにおけるチャプター・ハイライトシーンとして、また、書籍の要約に該当するものにダイジェストが

対応する。

一方、低レベルな記述とは映像中の画像信号や音声信号の特徴量であり、計算機によって自動的に抄録されるものである。この低レベルな記述として、画像信号については色、テクスチャ、形、動き、領域などの特徴量が規格化されている。よって、低レベルな記述を用いた検索においては、まずは与えられた検索対象の映像の特徴量を計算し、次にその特徴量と検索対象の映像の特徴量との類似度を計算することで、類似度の高い映像を見つけ出すといった流れになる。

また、高レベルな記述では、あらかじめ人手によって映像シーンをテキスト注釈として記述することにより、低レベルな特徴量では検索が困難なものに対して意味レベルでの映像検索を可能としている。

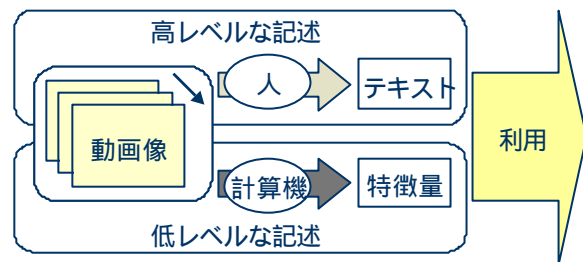


図4 2レベルの内容記述

4.2 MPEG-7を用いた抄録

動画の抄録法として、前述したMPEG-7を利用し高・低2レベルの抄録を用いることにした。すなわち、人の力による抄録と、ソフトウェアによる自動的、機械的な抄録の2種類の方法を用いる。

高レベルな抄録には、動画検索に必要な不可欠と考えられる人物、場所、再生時間などについて、あらかじめ人が見て抄録したものをXMLを用いて記述した。XMLによる記述例を図5に示す。

```

- <Video Number="A-8">
- <Cut Number="1" Interval="25s">
- <Content Number="1">
- <Subject>
  マンク
  <Person>2人</Person>
</Subject>
<Object>供物</Object>
<Action>置く</Action>
<Place>D00</Place>
<Time>6:18</Time>
<Camera>F190,215</Camera>
</Content>

```

図5 タグ付けによる抄録

低レベルな抄録については、今回、被写体検索をする場合に重要になると考えられる Visual 記述子の映像特徴に分類される“色”と“動き”について重点を置き動画像データから自動的に抽出し、抄録した。

色についての情報は、シーン毎にその内容を表す静止画を1枚選出し、それをそのシーンの代表静止画とする。これを8色に減色し、各色が合計何ピクセルあるかを抽出、その色の塊と重心についての特徴量を求めた。

また、動きについての情報は、動画像データからフレーム間差分を取るにより移動体を検出、その移動体の数、移動方向、移動量を算出し、カメラの動作、物体の動き、物体の軌跡などの特徴量を求めた。

この2つの特徴量を用いることにより、動画像中から目的のシーンやカットを検索することが可能になると考えた。

5. データベース設計

今回、研究に用いる動画像データは、打浪教授が南太平洋学術調査交流計画で撮影したヒンズー教寺院の創立祭であるオダランについてのビデオである。これはすでに人の手によってシーン毎の登場人物、動作、時刻、場所、カメラの角度などの抄録が行われている。この動

画像データをビデオからデジタルデータとして取り込み、シーンで分割し、avi形式でデータを記録しデータベース化した。その内、重要なシーンを抽出し実験を行った。データは以下のようになった。

- ・ シーン数 48
- ・ カット数 89722
- ・ 画像サイズ 320×240

今回はこの48シーンについて高レベル、低レベルの抄録を行った。

5.1 高レベルな記述による抄録データベース

MPEG-7において、高レベルな記述は図5に示すようにタグ付けによって抄録されている。これは、記述者以外の第三者が見てもわかりやすく、データを利用しやすい反面、そのままデータベース化したのでは、冗長が多く得策ではない。そこでデータベースとしては、メタ情報部分は構造に含める形で削除し、関係型データベースで構築した。抄録項目を以下に示す。

- ・ ID ビデオ No、カット No、何カット目であるかを表す。
- ・ 主語(subject) 動作主を表す。
- ・ 動詞(verb) 動作主の動作を表す。
- ・ 目的語(object) 動作の目的語を表す。
- ・ 補語(complement) 人物や物の状態を修飾するためのもの
- ・ 座標(x,y) 平面図上の位置座標を表す。
- ・ 場所(place) 地図上の場所名を表す。
- ・ 時刻(time) そのイベントが開始された時刻を表す。
- ・ カメラの関係(camera) カメラのズーム、パンなどを表す。
- ・ シーン番号(sceneNo) そのイベントが含まれるシーンを表す。

高レベルな記述による抄録データベースの

抄録例として、これらの項目をデータベース化したものを表 1 に示す。

表 1 高レベルな抄録データベース

id	subject	X	Y	place	verb
A080101	ブマク	3	16	祭壇前	置く
A080102	ブマク	3	16	NULL	受け取る
A080103	人	3	16	D00あたり	お参りする
A080201	ブマクグデ	3	16	NULL	会釈する
:	:	:	:	:	:

object	complement	time	sceneno	camera
供物を	NULL	6:18	1	固定
供物を	参詣人から	NULL	1	固定
NULL	NULL	NULL	1	固定
NULL	NULL	NULL	2	NULL
:	:	:	:	:

5.2 低レベルな記述による抄録データベース

色についての特徴量としては、シーン毎に選出した代表静止画を 8 色に分類し、各色が合計何ピクセルあるかの統計を取ったものと、各色について塊の大きさとその重心を計算したものを作成した。今回用いたデータは、もともとアナログデータであったためノイズが多くフルカラーで抄録を行っても特徴量とは言い難いものであったので、減色を行い、色の偏りをはっきりさせた。色についての低レベルな抄録データベースの例を以下に示す。

表 2 色の塊とその重心のデータベース

frameno	katamari	jusinx	jusiny
1	2266	305	37
1	783	128	102
1	203	269	166
:	:	:	:

低レベルな動きの抄録については、一般に画像系列の観測時間周波数が大きい場合には、フレーム間差分によって得られる差分領域は移動物体の周辺に出現し、その動きは移動物体の動きに酷似することから、動画データの 1 秒間を 30 フレームに分割し、そのフレーム間

差分画像から検出された差分領域を追跡することで移動物体を検出、抄録することにした。

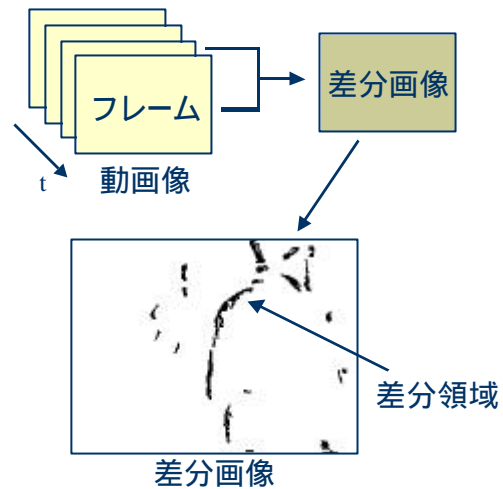


図 6 フレーム間差分

移動物体の追跡については、空間的、時間的に隣り合う差分領域が共に高い差分値であれば、同じ移動物体とみなし、移動方向、移動量を算出した。低レベルな抄録として、この際に検出した差分領域の数、大きさ、位置、移動方向、移動量について特徴量として抄録を行った。動きについての低レベルな抄録データベースの例を以下に示す[4]。

表 3 差分領域の重心のデータベース

frameno	id	jusinx	juxiny
10	101	12	2
10	102	292	134
11	102	290	133
12	102	291	130
:	:	:	:

表 4 差分領域の移動量データベース

sceneno	id	direction	move
11	102	4	2.24
12	102	3	4.12
13	102	3	2.24
13	121	3	3
:	:	:	:

この特徴量を用いて“パン”、“ズーム”、“近づく”、“遠ざかる”、“左右に移動する”といったカメラの動きや移動物体を抄録した。

6. 実験

6.1 検索実験

MPEG-7 を用いた抄録を評価するため、実際に動画像中に登場する被写体を指定して、各レベルでの検索実験を行った。今回は、祭りの中に現れる僧侶であるプマंकを被写体とし、移動するプマंकが存在するシーンを検索した。

6.2 高レベルの検索

以下のような SQL 文によって、プマंकがいるシーン番号を得ることができる。

```
select sceneno
from highlv_table
where subject = 'プマंक'
```

検索を行ったところ、48 シーン 中 12 シーンが検索された。これはあらかじめ人手によって抄録されたデータベースによるものの検索結果である。

6.3 低レベルの検索

低レベルの場合、検索対象であるプマंकについて特徴を考えて検索を行う必要がある。まず、色の特徴としては、プマंकはオレンジ

(赤黄)色の腰巻をして、白いシャツを着ている。すなわち、2 色の重心の x 座標に関してほぼ等しく、白色の重心の y 座標がオレンジ(赤黄)色の重心の y 座標よりも上に位置するという条件で検索をかければよい。オレンジ色については赤色、黄色のどちらの色にもとれることから、赤色、黄色の両方について検索し、その論理積をとることでオレンジ色を求めた。動きについては、カメラの動きを除外した上で、移動体が存在するシーンを検索すればよい。その上で、色の特徴量から検索されたシーンと動きの特徴量から検索されたシーンの論理和をとることにより、移動しているプマंकが存在するシーンを特定した。

6.4 実験結果

検索した結果、色の特徴量から検索されたのは 48 シーン 中 18 シーン。また、動きの特徴量から検索されたのは 16 シーンであった。これらの結果から論理和をとると低レベルの検索では 48 シーン 中 22 シーンが検索された。

この結果の評価として、実際に移動するプマंकが存在するシーンと、高レベルと低レベルの検索結果とを比較した表を表 5 に示す。

表 5 2 レベルの検索結果の比較

scene No	事実(正解)	高or低	高レベル	低レベル
1				
2	x			
3				
4				
5				
6				
7				
8				
9	x			
10	x			
11	x			
12				
13				
14	x			
15				
16				
17	x			
18				
19				
20				
21				
22	x			
23				
24	x			
25				
26				
27	x			
28	x			
29	x			
30	x			
31				
32				
33	x			
34				
35	x			
36	x			
37				
38				
39	x			
40	x			
41				
42				
43				
44				
45				
46				
47				
48				

この表の事実とは、そのシーンの中に移動するオブジェクトが登場していることを表す。つまり、検索されるべきシーンに印がついている。

また、この表をもとにして移動体検出の呼出率と適合率について評価した。

色の特徴量と動きの特徴量の低レベル抄録を組み合わせて検索した結果を表 6 に示す。

表 6 移動体検出の呼出率と適合率

	高 or 低	高レベル	低レベル
呼出率	58.06%	29.03%	41.94%
適合率	81.82%	100.00%	81.25%

7. まとめ

検索結果について見てみると、高レベルの検索の呼出率が低いのが目立つが、これは人の手によって抄録されたということが大きく関わってくる。人によって抄録が行われる場合、印象に残ったものから抄録を行い、特に重要ではないもの、また、気が付かなかったり、背景などのように変化が少ないものについての抄録を行わないためである。また、適合率については、人によって直接抄録が行われているため、抄録されるものに対する認識が間違っていない限り適合率の低下は有り得ない。抄録に対する労力がかかる反面、抄録内容は正確であることが適合率の高い要因である。

一方、機械的に行った低レベルな抄録の結果から、自動的に抄録したものでもある程度の適合率がでていたことがわかった。適合率が高レベルより劣る原因としては、機械的に特徴量を求めたため、人間では気づかないような抄録ができる反面、抄録されるものに対する認識を考慮しない点が要因と考えられる。

また、高レベル、低レベルそれぞれ単独の場合より、両方の論理和を使った場合の結果の方が呼出率、適合率のバランスがとれていることがわかった。

MPEG-7 で提案されている記述項目のうち、被写体とその動きに関する項目を用いて、動画データベースを抄録、検索できるシステムを提案、試作した。

人が逐次抄録をした高レベルの抄録に対し、機械的に行った低レベルの抄録では、人が見逃したり、不必要と判断し無視した移動体までも検出していることがわかった。しかし、その反面、カメラの動きの関係で、重要な被写体の動きを抄録できないという欠点も発見された。

高レベルと低レベルの 2 レベルの抄録がお互いに補完し合うことで、検索機能が豊富になり、また、低レベル抄録を用いることにより、抄録の半自動化が実現できたと言える。

参考文献

- [1] Seiichi Uchinami, Takayuki Tokumaru, et al: "Integration of Multimedia Database with Virtual System", International conference on Virtual System and multimedia 1995.
- [2] Seiichi Uchinami: "Virtual Reality database Systems", Advanced Database Systems for Integration of Media and User environments 1998, World Scientific.
- [3] 堀 修:"マルチメディアコンテンツ内容記述国際標準 MPEG-7 の概要とその応用-内容記述(メタデータ)はどのように作成されるか?-", 信学技報, TECHNICAL REPORT OF IEICE, PRMU 2002-25 (2002-6),p.35-42
- [4] 嶺 直道,八木 康史,谷内田 正彦:"時系列差分画像を用いた複数移動物体の抽出および追跡", 情報処理学会研究報告,93-CV-81,p.51-56,1993.
- [5] 高木 幹雄,下田 陽久:"画像解析ハンドブック",東京大学出版会(1991)