

映像シーン記述言語を用いた動画像検索システムの提案

4 J-6

山根 淳, 佐藤 隆, 龔 怡虹, 坂内正夫
東京大学生産技術研究所

1 はじめに

今後 CATV や 高速高帯域 ISDN の発展・普及によって多種多用の映像メディアが多数のチャンネルによって一般に流通することが予想される。また現在、映像メディアを中心とするマルチメディア情報をコンピュータシステムの中心に位置付けようとする方向性をもったハイパーメディアシステムへのニーズが高まっている。

したがって将来予想される膨大な量の映像メディアから必要な時に必要な情報を獲得する手法への要求が生じてくるが、特に映像メディアの場合データ量が膨大で抽象化レベルが低いため、従来の人手に頼った方法ではおのずと限界がある。従って、自動認識・理解を積極的に用いた情報抽出手法が重要な要素となる。

このような観点から、我々はこれまでに映像メディアから自動獲得したい情報を一般的に記述できるハイパーメディアの実現に向けて、映像シーン記述言語 (Video Scene Description Language, 以下 VSDL と略記) を提案してきた [1] が、本発表においては予備実験を行なうことによってその分類・記述能力の評価を行ない、また VSDL を用いたユーザ所望のシーンを検出するシステムの提案を行なう。

2 映像シーン記述言語 (VSDL)

2.1 VSDL の用いる特徴パラメータの抽出

VSDL では、画像の特徴パラメータとして、色とその動きの情報に着目している。これらの情報は次の処理によって得る [1]。

(1) 原画像から主要な色を抽出し、N 色限定色画像を生成する

フルカラー画像を RGB 表色系から $L^*a^*b^*$ 表色系に変換して、出現頻度の高い色を代表色とする。代表色を中心として球状に色空間をクラスタリングし、フルカラー画像を N 色に限定する。

このとき、画像をサブサンプリングしても色空間のクラスタリングに影響が少ないことが実験で確かめられたので、画像の画素数を 1/25 程度まで低減して、文献 [2] の方法を高速化している。

(2) 2つのフレーム間で、色領域の動きを検出する

2つのフレームで、同じ色割当テーブルを用いて限定色表示することで、フレーム間で近い色分布を有する

```
(Color skin_c ((0.436 0.968) nil (12.0 25.0)))
(Color bg1 (nil nil (0.0 8.0)))
(Color bg2 ((3.362 5.0) nil (8.0 30.0)))
(Color bg_c (or bg1 bg2))
```

```
(Segment skin
(Color skin skin_c)
(ColorMotion 2)
(Position (and (or 0 3) (or 2 5)))
(Area >= 200)
)
```

```
(Segment bg
(Color bg bg_c)
(ColorMotion 2)
(Extent >= 4)
(Density >= 40)
)
```

```
(Class News
(Exist (and bg skin))
(ColorChange 2)
(SceneChange 0)
(Zooming 0)
(Panning 0)
)
```

図 1: VSDL によるニュースシーンの記述の例

領域同士の対応が容易になる。方法の詳細は文献 [1] に譲る。

2.2 VSDL の記述階層

VSDL は、

- (a) 情報獲得意図を表現するための言語
- (b) 情報の構造を認識するための言語
- (c) データ蓄積の内部表現のための言語
- (d) 情報のプレゼンテーションのための言語

の機能を統合した言語として構想・設計している。以下の実験では (a)(b) について VSDL の部分的実現を行なった。なお、現在 (c)(d) の機能や、さらに柔軟な意図の表現、また、形状など複雑な認識が可能ないように、VSDL を拡張中である。

さて、VSDL には現時点で 18 個のキーワードが存在する。これらは、次のように 3つの記述階層で用いられる (ニュースシーンの例を図 1 に示す)。

(1) 色定義階層

Color を用い、色を定義する。

(2) 領域定義階層

Segment を用い色領域を定義する。定義の中では、Density, Area で面積を、Position, Extent,

A Proposal of Moving Image Retrieval System Using Video Scene Description Language

Jun YAMANE, Takashi SATOU, Yihong GONG, Masao SAKAUCHI

Institute of Industrial Science, Univ. of Tokyo

表 1: 評価実験に用いたシーンのリスト

番号	映像名	映像内容	記述の 難易度	分類 結果
1	ニュース1	中央にアナウンサ	易	○
2	ニュース2	二人の対談	易	○
3	天気予報	天気予報パネル	易	○
4	大相撲1	力士のアップ	易	○
5	大相撲2	取り組みシーン	中	○
6	ゴルフ	ボールを打つシーン	中	○
7	シンクロ1	脚の演技	易	○
8	シンクロ2	頭を水面に出している	中	○
9	プロ野球1	打者のアップ	易	○
10	プロ野球2	投手と打者	難	○
11	カーレース1	車のアップ	中	○
12	カーレース2	会場全景	極難	×
13	プロレス	格闘シーン	難	○
14	テニス	打ち合いシーン	中	○
15	体操1	床競技	中	○
16	体操2	平行棒競技	中	○
17	体操3	鞍馬競技	中	○
18	新体操	ボール競技	中	○
19	バレーボール 試合	ネットをはさむ 攻防シーン	極難	×

Upper, Lower, Left, Right で位置を、ColorMotion で動きを指定できる。

(3) シーンクラス定義

Class を用いシーンクラスを定義する。定義の中では、ConsistOf, Appear, Disappear によって色領域の出現を、ColchgRate によってフレーム間の色分布の変化を、Zooming, Panning によってカメラワークを記述できる。

2.3 VSDL の分類能力評価実験

19 種類のシーンを用いて評価実験を行なった(表1)。17 種類のシーンを検出することができ、本アプローチが有効であることがわかる。複雑な形状をした領域が多数現れるようなシーンや、背景と物体との分離が難しいシーンでは、VSDL での記述が難しく、検出に失敗している。

3 VSDL を用いたシーン検出システム (図 2)

上記の実験により、VSDL のシーン分類能力が示された。以下では、一連のビデオ映像をリアルタイムに解析してユーザ所望のシーンになったらユーザに知らせる動画シーン検出システムについて述べる。

まず時刻 $t = t$ における画像(一枚目の画像)について N 色による限定色画像を作成する。その限定色画像から色情報を抽出し、その結果と、VSDL の記述によって与えられた属性テーブルに記述されている静的な属性(色分布など)とのマッチングをとる。このマッチングに失敗したらシーンチェンジが検出されるまでスリープ状態に入る。マッチングに成功したら $t = t+1$ の画像(二枚目の画像)についてもマッチングを行なう。この場合は静的な属性のマッチングののち、物体の動きなどの動的な属性についてもマッチングをとることにな

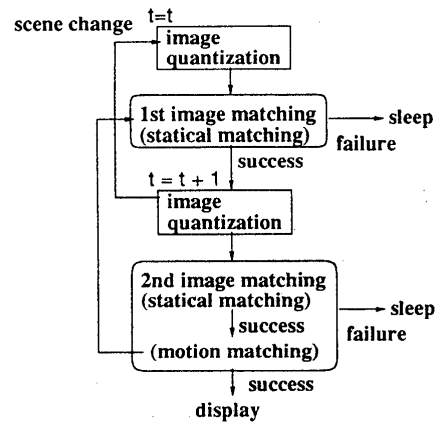


図 2: シーン検出システムのフロー

る。すべてのマッチングに成功した場合システムはユーザに検出されたシーンを提示する。

$t = t+1$ の画像については、静的な属性のマッチングに失敗した場合は $t = t$ の画像と同様、シーンチェンジが検出されるまでスリープ状態に入る。動的なマッチングのみに失敗した時は $t = t+1$ の画像を一枚目の画像とし、 $t = t+2$ の画像に対して $t = t+1$ の画像に行なった処理を繰り返す。

4 おわりに

本稿においては、映像メディアから必要な情報を自動的に獲得するハイパーメディアの実現のために、色の種類と動きに着目するシーン記述言語 VSDL を実現し、実験結果によりその分類能力の有効性が示された。また、この VSDL を用いて、所望のシーンになったらユーザに知らせることのできるシーン検出システムについての提案を行なった。

現在、さらに柔軟な表現が可能になるようにユーザインタフェースの観点から VSDL の機能拡張を行なっている [3]。また、シーン検出システムを実装し、評価することによってアプリケーションの立場から見た VSDL の改良も行なっていく予定である。

参考文献

- [1] 龔, 坂内: 動画像シーン分類に関する一提案情処 44 回前期全大, 1C-7(1992).
- [2] 龔, 全, 大沢, 坂内: 色割当の安定性を重視した動画像の限定色表示方式, デレビ誌, Vol.45, No.11, 1991-11.
- [3] 佐藤, 山根, 龔, 坂内: ハイパーメディアシステムのための映像シーン記述言語, 情処 44 回後期全大, 4J-5(1992).