研究速報

特集 5

UDC 681.3.016：621.397.3

シーン記述言語を用いたマルチメディア検索システム

A Multimedia Retrieval System Using Video Scene Description Language

佐藤　隆*・山根　淳*・龔　怡虹**・坂内正夫*

Takashi SATOH, Jun YAMANE, Gong Yee-Hong and Masao SAKAUCHI

1. はじめに

近年，映像メディアへのニーズが拡大し，ビデオ情報の蓄積が増大するにつれ，画像・映像情報を中心とするマルチメディアデータベースシステムへの期待が高まってきている．特に，画像・映像の認識技術を用いて，自動的な方法によって情報の認識・理解を行い，データベース化の際の検索情報とすることにより，高度なシステムが期待できる．これについては，いくつかの試みが行われている．たとえば，カット変化の自度検出による索引付け手法や[1),2)]，カメラワークの規定による動画像の索引付け手法[3)]，動画像内容の自動記述生成[4)]など，興味深い方式が提案されている．しかしながら，これらの手法は，映像の統計的な性質を解析するに留まって構造に踏み込んでいなかったり，ユーザの検索意図を反映していないなどの問題があった．

ここでわれわれが提案するシステムは，色セグメントの大きさや動きなどの情報を用いて映像を解析・分類し，**映像シーン記述言語（VSDL）**によってユーザの検索意図を記述するものである．これにより，これらの問題を解決しようとするものである．

2. システムの概略

映像は「シーン」の集合体とみなすことができる．ここで，「シーン」とは内容に大きな変化のない複数フレームのまとまりである．シーンには背景や物体があるが，それぞれ特徴的な色や配置，動き情報を検出できれば，シーン構造が把握できることになる．このような観点から，映像シーン記述言語（VSDL）を用いた動画像検索システムを提案する．

システムの使用形態を図1に示す．ユーザはVSDLを用いて検出・分類したいシーンを記述する．記述は複数あってもかまわない．システムはVSDLを解釈し，

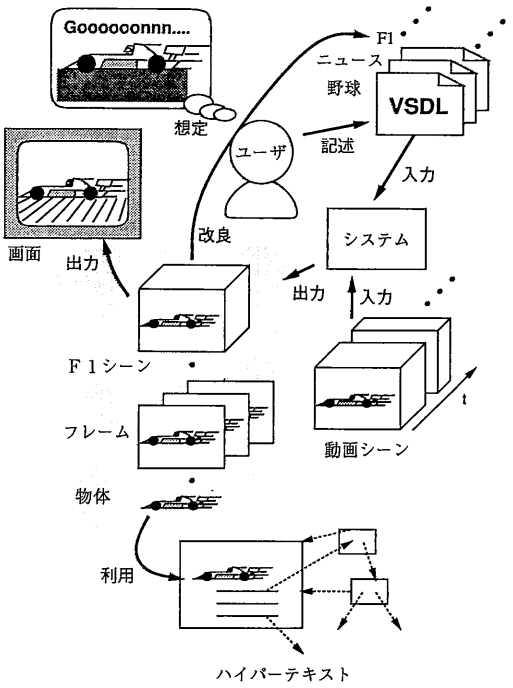

図1　システムの使用形態

入力映像との間でマッチングをとる．マッチングが成功すれば，ユーザの意図したシーンが入力映像から検出されたとする．

システムの内部構造を図2に示す．システムはMeasurement Class（MC），Sub Area Measurement Class（SAMC），Scene Class（SC）の3種類のクラスから構成されている．

SAMCは画面の分割領域について色の分布を調べ，MCに送る働きをしている．

MCはシステムの中核をなし，色のデータベースやクラス間の通信の機能を持っている．SAMCから送られてくる情報をもとにして，セグメントの色，配置や動きといった情報を検出しSCに送る．

*東京大学生産技術研究所　第3部

**Nanyang工科大学（元大学院生）

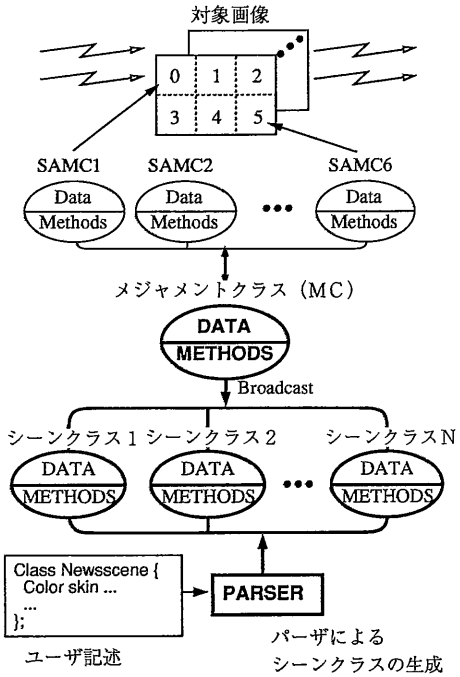

図2　システムの構造

SCは，VSDLによって記述されたファイルを読み込んで生成される．SCは検出・分類したいシーンに対応しており，シーンの属性テーブルを持ち，MCから送られてくる情報と自分の持つ属性とのマッチングをとる機能を持っている．

MCとSCは密接に通信しあい効率のよい処理を可能としている．まず，MCはSCが必要としている情報に着目して映像を解析する．たとえば，比較的高価な処理となる動き検出は，すべての色セグメントに行うのではなく，動き検出を必要としているSCに問い合わせて，追いかける色を絞り込むようにしている．

また，SCはMCから送られてくる情報と自分の属性情報とを照らし合わせ，マッチングに成功するメドが立たなくなった時点で，自分をスリープ状態にし，以後の処理を行わない．逆に，重要な属性のマッチングに成功すれば，自分の属性に関する情報を他のSCより優先してMCに調べてもらうようにする．こうして，SCはもっとも少ない処理でゴールに到達することができる．

以上をまとめると，このシステム構成は次のような特徴があることがわかる．

1) ユーザ定義シーンの検出時間はシーン定義の複雑度に比例する．
2) システムの各部は高い独立性を持っているので，将来の仕様拡張や，画像処理技術の進歩にともなう情報抽出のためのメソッドのバージョンアップに高い自由度を残している．
3) システムは並列処理に向いている．
4) ユーザはシーン記述言語により，比較的簡単に検出・分類したいシーンを定義できる．

3. 映像からの情報抽出

前章で述べたように，本システムでは，映像からセグメントの色の種類と配置，動きの情報を検出し，それらを認識や記述の情報要素としている．

色抽出については，筆者らによる既述の方法[6]を改良して利用している．

色セグメントの動きの抽出については，次のとおりである．

動きの検出方法には，Block Matching法や，Optical Flow方法，Spatio Temporal法などが考案されているが，本システムはこれらの手法が提供する程の精度を必要とせず，むしろ，次の条件を満たした手法であることが望ましい．

・画素単位の動き情報より，色セグメント単位の大まかな動き情報を抽出できること．
・高速処理が可能であること．
・画像ノイズに強く，一般映像への適用が可能であること．

このため，次のような簡略な動き検出方法を用いることにした．

1) 比較的近い2つのフレーム N_1，N_2で，同じ色空間のクラスタリング結果を用いて限定色表示をする．
2) ある代表色について，8連結領域を求め，セグメントとする．
3) N_1のセグメント A_1と重なり合い，かつ，面積が近い N_2のセグメント A_2を求め，対応セグメントとする．
4) A_1の重心から A_2の重心へのベクトルをセグメントの動きベクトルとみなす．

ただ，この方法では，セグメントのなかに非常に細長いものや形状が複雑なものが存在すると，フレームによって2つに分離したり融合したりして不安定となる．このため，式1に示す占有率を計算し，この値が20%以下のセグメントを処理から除去している．

占有率＝図形の面積÷図形の外接長方形の面積　(1)

この手法を使ってすべての色のセグメントについて動きベクトルを求めることにより，ズーミングやパニングといったカメラ操作を検出することも可能である．

研究速報

4. 映像シーン記述言語

本手法では，ユーザが所望の映像シーンを表現できる手段が必要である．検出・分類したいシーンを抽象的に記述するために，本システムでは映像シーン記述言語（VSDL）を提案している．

VSDLには，表1に示す19個のキーワードが存在する．

これらは以下の3つの記述階層でそれぞれ使用される．

4.1 色定義階層

Colorは色を定義するためのキーワードである．色の定義にはHVC表色系における色相，明度，彩度の範囲をパラメータとして渡す．パラメータをDon't careにすることもできる．

4.2 セグメント定義階層

Segmentを用いて，色セグメントを定義する．定義中では，まず，colorによって，セグメントの色を定義する．これには，前節で述べた色定義を用いる．

セグメントの面積の指定のしかたには2種類あり，Areaによってセグメントの画素数を，あるいは，Densityによって，画面の分割領域における画素の割合を指定する．後者の方法は，テクスチャー領域などの指定に適する．

セグメントの位置の指定のしかたには3種類ある．ひとつは，絶対位置の指定として，Positionによって，セグメントの存在する画面の分割領域の識別番号を指定する．または，相対位置の指定として，Relationにより，他のセグメントとの位置関係を記述できる．ここでは，Upper，Lower，Left，Rightによって関係を記述する．もう一つの位置指定方法は，Extentによって，セグメントの出現する画面の分割領域の数を指定するもので，出現場所を特定しない．

セグメントの動きについては，ColorMotionによって与える．パラメータとして動きベクトルの向きと大きさを与えるが，Don't careを設定することもできる．

4.3 シーンクラス定義階層

Classによって，これまで定義してきたセグメントを集めて，シーンクラス（SC）を定義する．まず，ConsiftOfにより，シーンを構成するセグメントを定義する．Appear，Disappearにより，シーンの途中から出現/消滅するセグメントを記述することも可能である．

シーンの内容の変化を示す指標として，ColchgRateによって，2フレーム間の色分布の変化率を記述することができる．たとえば，テレビのニュースや天気予報のシーンは，画面の構成が一定であり，色分布の変化率は小さい．変化率は2枚のフレームを限定色処理したあと，式2により計算される．

$$\text{変化率}=\max\left(R_0, R_1, \ldots R_{S-1}\right) \tag{2}$$

$$R_i=\sum_{j=1}^{N}\frac{H_i\left(f_1, j\right)-H_i\left(f_2, j\right)}{H_i\left(f_1, j\right)} \tag{3}$$

ここで，$H_1\left(f_k, j\right)$は，フレームf_kの分割領域i

図3　典型的なニュースシーン

表1　VSDLのキーワード

色定義階層	Color
セグメント定義階層	Segment Color Density Area Position Extent Relation Upper Lower Left Right Color Motion
SC定義階層	Class Consist Of Appear Disappear Colchg Rate Zooming Panning

```
//肌色の定義
(Color skin_c ((0.436 0.968) nil (12.0 25.0)))
//背景色の定義
(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          //肌色セグメントの定義
  (Color skin skin_c)//肌色
  (ColorMotion         //ほとんど動かない
      ((0 2.0) nil))
  (Area >= 1000)  //面積は1000以上
  (Position (or 1 4))//画面分割領域1,4に存在
)
(Segment bg            //背景色セグメントの定義
  (Color bg bg_c)  //背景色
  (ColorMotion         //ほとんど動かない
      ((0 2.0) nil))
  (Extent >= 4)    //4つ以上の分割領域に存在
  (Density >= 50)//各分割領域に占める割合は
                 //50%以上
)
(Class News            //ニュースシーンの定義
//定義するシーンは'bg'と'skin'との
//2つのセグメントから構成される
  (ConsistOf (and bg skin))
  (ColchgRate < 2)//色分布の変化率は2%以下
  (Zooming NO)     //ズーミングなし
  (Panning NO)     //パニングなし
)
```

図4　VSDLによる記述例

表2　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	中	○
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	×

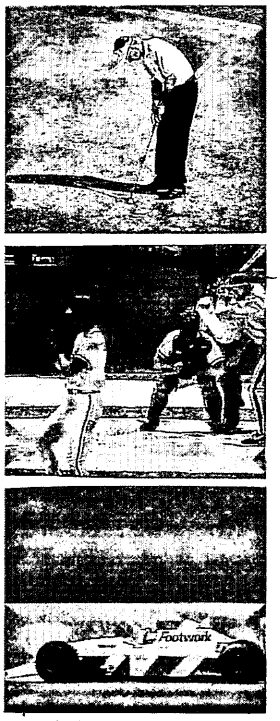

図5　実験に用いた映像の一部

$(0 \leq i \leq S-1)$ における画素数を表している．つまり，N色限定色画像 f_1, f_2 について，画面分割領域ごとに，色ヒストグラム $H_i(f_k, j)$ の変化率を求め，その最大値を色分布の変化率とする．

この他，シーンをカメラワークによって規定するため，Zooming，Panningによって，ズーミングとパンニングの有無を記述することができる．

4.4 シーン記述の例

図3のような典型的なニュースシーンをVSDLによって記述した例を図4に示す．印刷の都合上わからないが，画面では青い背景の中央に灰色の背広を着たアナウンサーがいる．

5. 評 価 実 験

本システムをワークステーション上に実装し実験を行った．評価実験には，表2に示す19種類のシーンを用い，限定色数16，画面分割数6とした．映像の一部を図5に示す．表中の難易度は，VSDLで記述するときの主観的な難しさを示している．分類結果は，対応シーンを検出できたかどうかを示し，○は成功×は失敗を示す．

表からわかる通り，かなり幅広い種類の映像をVSDLによって記述・検出でき，本システムの採用した色，動きなどの情報を用いるシーン記述の有効性を示した．

6. お わ り に

本論文では，まずユーザの意図にそって映像の内容や構造を検出するシステムを提案し，その内部構造を述べた．つづいて，映像から色や動きなどの情報を効率よく抽出する方法について述べた．また，ユーザの意図を表現するための映像シーン記述言語について，そのあらましを述べた．最後に実験によって本システムを評価し，その有効性を明らかにした．今後は，より複雑なシーンにも柔軟に対応するように，VSDLの仕様を再検討して発展させていく予定である．　(1992年9月1日受理)

参 考 文 献

1) 上田：“インタラクティブな動画像編集方式の提案”，信学技報，**IE0-6**，(1990)
2) 長坂，田中：“カラービデオ映像における自動索引付け法と物体探索法”情処学論，**vol. 33**, No. 4, pp. 543−549 (1992)
3) 阿久津，外村，大庭：“動画像インデキシングを目的としたカメラ操作の規定方法”，信学第2回機能図形情報システムシンポジウム，pp. 107−112 (1991)
4) 君山，清末，大庭：“動画像中物体に対する自動記述生成の検討<”，情処第44回全体 5B-8 (1992)
5) 宮原，吉田：“色データ (R, G, B) ↔ (H, V, C) の数学的変換方法”，TV雑誌 Vol. 43, No. 10 pp. 1129−1136 (1989)
6) 龔怡虹，大沢裕，全炳東，坂内正夫：“色割当の安定性を重視した動面像の限定色表示方式”，テレビジョン学会誌，**Vol. 45**, No. 11, pp. 1446−1454 (1991)