

動画像中物体に対する自動記述生成の検討

5B-8

君山 博之 清末 悌之 大庭 有二

NTT ヒューマンインタフェース研究所

1. はじめに

家庭用のビデオデッキやビデオカメラの普及に伴い大量の動画像が世の中に溢れ出してきた。ところが、これらの動画像をアプリケーションから利用する場合、例えば、動画像ファイル中のある場面を検索しようとする場合、そのままの形では先頭から順番に見て探さなければならず、非常に不便である。そこで、そのための情報をインデクスとして付与しておく必要がある。現在、このような動画像を扱うためのインデクスは、人間が動画像ファイルの内容を見て付与している。しかし、ファイルの数が多い場合や細かい単位(例えば、フレーム単位)で扱いたい場合、付与するインデクスの量が膨大となるため、自動化が望まれている。そこで、我々は、自動化を目的として、動画像インデクスとして有効な記述というプリミティブな単語列を動画像から生成する方法の検討を行ってきた[1][2]。本発表では、どのような種類の情報を記述したら良いかという記述カテゴリの検討、被写体情報に関する記述の計算機による自動生成方法、ならびに実際の動画像を用いた記述の生成実験について報告する。

2. 記述カテゴリの検討

今回の検討では、動画像の内容を表現するために被写体に関する情報を記述することにした。まず、被写体に関する情報を4つの大きなカテゴリに分類した。1番目は動画像中の被写体がそれぞれどのような静的特徴を持っているかという、「被写体個々の静的情報」である。このカテゴリでは、被写体の画面上の位置、形、大きさ、色を記述カテゴリとした。2番目は、個々の被写体がどのように動いているかという、「被写体個々の動きの情報」である。このカテゴリでは、位置、形、大きさの時間変化を記述カテゴリとした。ここで色の変化を記述しないのは、色の変化は少ないと考えたからである。3番目は、これらの被写体が動画像の中でどのような位置関係にあるかという「被写体のレイアウト情報」である。このカテゴリでは、被写体の個数と被写体の散らばり方を記述カテゴリとした。4番目は、被写体のレイアウトが時間的にどのように変化していくかという「被写体のレイアウトの時間変化」である。このカテゴリでは、被写体の数の時間変化と被写体の分布の時間変化を記述カテゴリとした。

このように決定した記述カテゴリに対して、複数の記述する単語、例えば、「大きさ」という記述カテゴリに対して「大」、「中」、「小」というように、割り当てた。以上をまとめたものを表1に示す。

表1 記述要素

記述カテゴリ		記述する単語	
個々の静的情報	位置	上、下、右、左、中	
	形	縦長、横長、正方	
	大きさ	大、中、小	
	色	赤、橙、黄、黄緑、緑、青緑、青、青紫、紫、赤紫	
個々の動き情報	位置の変化	上、下、右、左、不変、消滅、出現	速い、遅い、中位、静止
	形の変化	高く、低く、広く、狭く、不変	速い、遅い、中位、静止
	大きさの変化	大きく、小さく、不変	速い、遅い、中位、静止
レイアウト情報	被写体の数	1つ、2つ、...	
	被写体の分布	分散、密集、ばらばら(3体以上) 遠い、中位、近い、接触(2体)	
レイアウトの時間変化	数の変化	増える、減る、不変	速い、遅い、中位、静止
	分布の変化	集中、拡散、ばらばら、そのまま(3体以上) 近づくと、遠ざかる、そのまま、くっつく(2体)	速い、遅い、中位、静止

3. 記述の自動生成方法の検討

動画像からの記述の生成は、被写体領域をフレーム単位で抽出し、抽出した領域の面積などの特徴量を導出した後、特徴量から記述を導出するという手順で行うこととした。処理の流れを図1に示す。

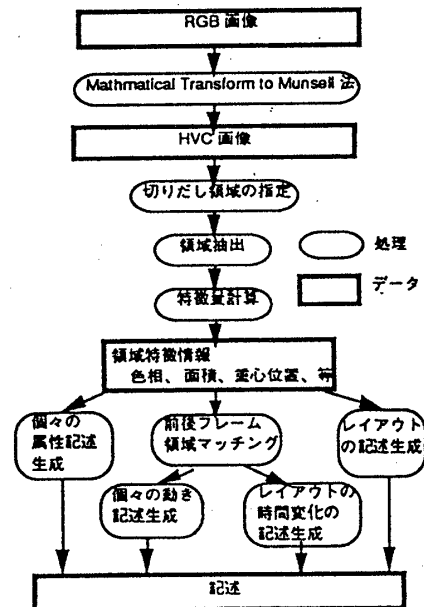


図1 記述の生成フロー

領域の抽出は、RGB画像をMTM法[3]によって、色相、明度、彩度の画像に変換し、予め決定しておいた色相、明度、彩度の値を用いることによって行い、この後、各抽出領域の特徴量を導出する。導出を行う特徴量は、領域の重心座標、高さ、幅、面積、色相の平均値である。これらの特徴量から、例えば、大きさの変化の記述の場合、領域の面積の時間変化の大きさ求め、「速い」「中速」「遅い」「静止」を決定し、時間変化の符号から「大きく」「小さく」「不変」を決定する。表2に、各記述カテゴリに対する記述を生成するために使用する特徴量を示す。

表2 記述の導出方法

記述カテゴリ	各記述カテゴリに対する記述生成に使用する特徴量
位置	領域の重心座標
形	領域の幅と高さの比
大きさ	領域の面積
色	領域の色相の平均値
位置の変化	重心座標の時間変化
形の変化	幅と高さの比の時間変化
大きさの変化	面積の時間変化
被写体の数	抽出領域の数
レイアウト	全抽出領域の重心と各抽出領域の重心との距離の分布
数の変化	抽出領域の数の時間変化
レイアウトの変化	全領域の重心と各領域の重心との距離の分布の時間変化

4. 実際の動画像からの記述の自動生成

実際に放映された動画像を用いて、記述の生成の実験を行った。処理に用いた動画像は、大相撲の画像である。まず、各フレーム単位で被写体領域の抽出処理を行う。今回、被写体領域としたのは力士の領域である。力士の膚とまわしの色相、彩度、明度を予め指定しておくことによって抽出した。その結果の一例を図2に示す。

このように抽出した領域を、時間軸方向につなげていった後、記述の導出を行った。表3に自動的に導出された記述の一例を示す。この記述を行った場面の力士の横方向の動きを図3に示す。この場面は、2人の力士が向い合い、構えて組み合う場面である。この図3と、表3とを比較することによって、導出された記述と動画像の間に矛盾がないことが確認出来る。

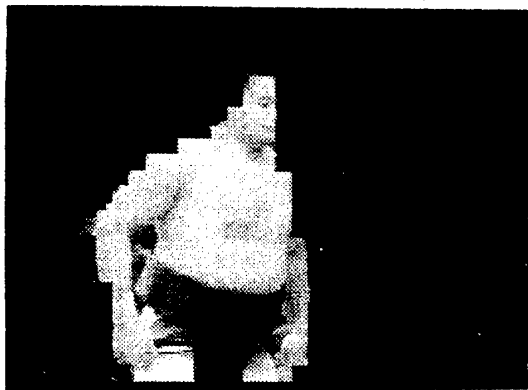


図2 領域抽出結果

5. まとめ

動画像インデクスとして利用するための動画像内容に関する自動生成可能な記述の検討を行い、実際の動画像を処理することによって、動画像の内容を表す記述が導出されることを確認した。今後、検索等のアプリケーションに適用して、記述の有効性を確認していく予定である。

参考文献

- [1] 君山、清末、田中、外村、大庭 “動画像インデクスの構造的記述方法の検討” 信学秋季全国大会 D-193 1991
- [2] 君山、清末、大庭 “動画像の自動記述方法の検討” 信学技報, IE-91 (1992年1月 発表予定)
- [3] 宮原、吉田 “色データ(RGB)-(HVC)数学的変換方法” テレビ学会誌 vol.43, No.10, pp. 1129-1136, 1990

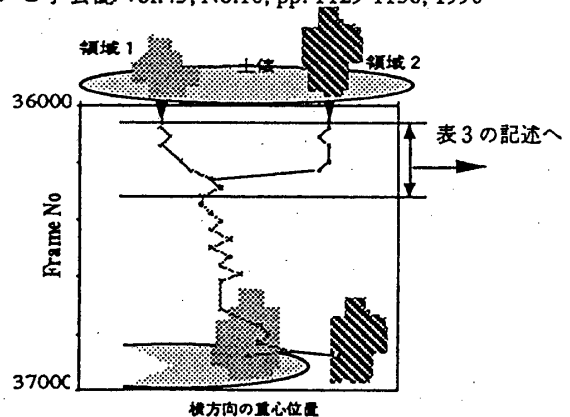


図3 力士の横方向の動き

表3 自動生成された記述

a. 個々の領域に対する記述

領域1 Frame No.	属性 大きさ 色 形	動き(移動)		動き(変形)		動き 大きさ
		速度 方向	速度 横方向 縦方向	速度 横方向 縦方向	速度 横方向 縦方向	
36081	中黄色 正方	静止	—	速い	狭く 低く	大きく
36111	小黄色 縦長	速い	左上	速い	狭く 高く	大きく
36141	小黄色 縦長	静止	—	中位	広く 不変	不変
36201	中黄色 縦長	速い	右下	中位	広く 低く	大きく
36231	中黄色 縦長	速い	右上	速い	広く 高く	大きく
36261	中黄色 正方	速い	右下	速い	不変 高く	大きく

領域2 Frame No.	属性 大きさ 色 形	動き(移動)		動き(変形)		動き 大きさ
		速度 方向	速度 横方向 縦方向	速度 横方向 縦方向	速度 横方向 縦方向	
36081	小黄色 縦長	静止	—	速い	広く 不変	小さく
36111	小黄色 縦長	静止	—	速い	広く 不変	大きく
36141	小黄色 縦長	静止	—	静止	不変 不変	大きく
36201	小黄色 縦長	静止	—	中位	広く 低く	小さく
36231	中黄色 正方	速い	左上	速い	広く 低く	大きく
36261	中黄色 正方	速い	右下	速い	不変 高く	大きく

b. レイアウトの記述

Frame No.	被写体の数	被写体の散らばり	被写体の数の変化	被写体の散らばりの変化
36081	2つ	速い	不変	近づく 速い
36111	2つ	中位	不変	遠ざかる 速い
36141	2つ	速い	不変	近づく 速い
36201	2つ	中位	不変	近づく 速い
36231	2つ	中位	不変	近づく 速い
36261	2つ	近い	不変	くっつく 静止