C-21

動きの階層性による動きの類似性を考慮した動画検索

Video Retrieval Using Motion Similarity in Term of Motion's Abstract Levels

池田 朋弘†　　野宮 浩揮†　　宝珍 輝尚†
Tomohiro Ikeda　　Hiroki Nomiya　　Teruhisa Hochin

１．はじめに

近年, コンピュータの性能の向上や HDD レコーダ, 動画共有サイトの普及により, 個人が大量の動画データを扱う機会が増加している。そこで, 大量の動画データからうまく目的の動画データが検索できるように, 動画データから多角的に情報を取り出し, 検索に活用する方法が求められている。

このような要求に答えるべく取り組まれている研究の方針として, 大きく２つのアプローチがある。

一つ目のアプローチは, 動画の内容を適切に記述することにより動画の意味的な情報を使用する意味アプローチである。これにより, “人が棒を切るシーン”といった検索が可能となる。意味アプローチに該当する研究として, 映像制作の観点からの内容記述[1]やユーザの検索要求に対して柔軟に動画の内容を定義できる内容記述[2]などが提案されている。これらは, 人が映像を観て捉える映像の意味に良く一致するものであり, 直感的な検索となる。しかし, 記述に用いる概念が抽象的であるため, 動画から人手でこの概念を抽出しなければならないという問題がある。

二つ目のアプローチは, 動画の信号的な特徴に注目し情報を取り出す特徴アプローチである。これにより, “赤色の画面領域が右に５画素移動するシーン”といった検索が可能となる。特徴アプローチに該当する研究として, 動体の色情報と移動距離を特徴とした手法[3]や, 動画を時間と空間（平面）からなる３次元の信号と捉え, ３次元周波数成分を特徴とした手法[4]などが提案されている。これらは, 自動的または半自動的に特徴を抽出できるが, 意味アプローチと比べて非直感的である。

そこで本研究では, 両アプローチの欠点を補い, 半自動的な処理で, かつ, 直感的な検索を可能とすることを目的とする。そのために, 特徴アプローチに準じて動画から半自動的に登場オブジェクトの動きを抽出する方針をとる。さらに入力を動画データとして, その動画データに含まれるオブジェクトとその動きが類似するシーンを提示する検索手法を提案する。特に, オブジェクトの動きの類似性の決定に利用するために, 動きの空間領域と時間領域の階層性に着目する。

以後, 2. では動画の持つ階層性および動きベクトルに注目した関連研究について述べる。3. では提案手法について述べ, 4. で実験とその結果を述べる。最後に 5. でまとめを行う。

†京都工芸繊維大学大学院, Kyoto Institute of Technology

２．関連研究

2.1　動画の階層性に注目した関連研究

動画は静止画の時系列として構成されるため, 動画には空間領域と時間領域が存在する。さらに, 動画は空間領域と時間領域のそれぞれに対して階層性を持つ。マルチメディアの高速な内容検索を目的として規格された MPEG-7[5]で導入されている空間領域の階層性を図１に示し, 時間領域における階層性を図２に示す。図１では, あるフレームの背景とオブジェクト, オブジェクトの頭と体による階層構造を表している。図２では, 動画全体がシーンに分割される階層構造を表している。

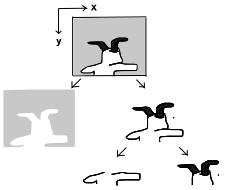

図１　MPEG-7 の空間領域の階層性の例

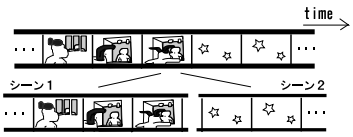

図２　MPEG-7 の時間領域の階層性の例

動画データを含むマルチメディアデータの内容を表す上でこのような階層性を考慮することは, データの内容をより特徴的に捉えることや, データの構造を表すこと, そして高速な処理を行うことなどに対して有効であり, 様々な研究で利用されている。本研究においても, 動きベクトルの階層性を考慮する。

2.2　動きベクトルに注目した関連研究

動画データにおいて, 隣接する２フレームを考える。このとき, 空間領域におけるある一点から得られる特徴量に対して, 図３に示すように点の変化量と２フレーム間の時間差によって動きベクトルが表現できる。

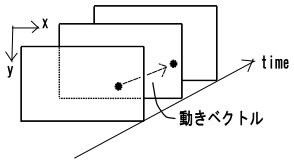

図３　動画の構成と動きベクトル

ここで, 例えば, 特徴を画素の輝度として動きベクトルを表現すると, 動画データに登場するオブジェクトの領域内の各画素に対応する動きベクトルが, 全て同じとなるのが理想的である。しかし, 実際には撮影時の照明といった外的なノイズが混入するため, 同じベクトルを持つ領域を追跡するだけではオブジェクトの動き情報を正確に得ることは困難である。

その例として図４に動画データ中のある２フレームの様子を示す。各フレームの左下にオブジェクト A, 右上にオブジェクト B の領域がある。A の領域に含まれる画素と対応する動きベクトルは全て等しく理想的であるが, 実際には B のように理想とは異なることがある。

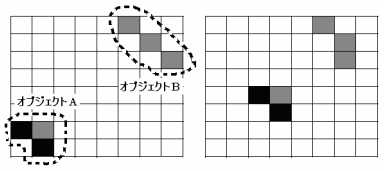

図４　動画データ中のある２フレームの様子
（色の濃さは画素の輝度を表す）

したがって, オブジェクトの動きの抽出や動きによるオブジェクト領域の特定には, オブジェクト領域内の動きベクトルの扱い方を工夫する必要がある。文献[3]では, オブジェクトの動きの取得は, オブジェクト領域内の動きベクトルの平均値を用いている。

本研究でも, オブジェクト領域内の動きベクトルからオブジェクトの動きを決定する際に, 最も度数の多い動きベクトルや最も度数の多い方向に属する動きベクトルの平均値をオブジェクトの動きとして採用する。

3. 提案手法

3.1　空間領域上での動きベクトルの向きに着目した階層性の利用

動画の空間領域の階層性を動きベクトルに導入する。これにより, 動きベクトルの特徴をより的確に捉えられ, オブジェクト領域内の動きベクトルに含まれるノイズの軽減とオブジェクトの動きの類似性の決定に利用することができる。

提案手法では, 動きベクトルの向きに注目する。動きベクトルを, 図５で示される空間領域上の中心位置に始点を置き, 終点の方向に近い灰色で示される領域を選ぶ。これによって, 方向を割り当てる。さらに割り当てに用いる領域を細かくすることで, より詳細な割り当てが可能となる。また, 図５において, 割り当て D17 の D17[1], D17[2], D17[16]の方向が, 割り当て D9 では, D9[1]の方向で表されるように, 動きの向きに着目した階層構造が構築できる。

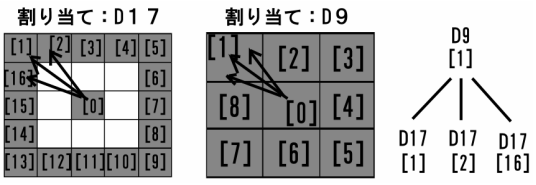

図５　17 方向（左）と 9 方向（中）への割り当てと階層構造の例（右）

3.2　時間領域上での動きベクトルの階層化

時間領域上での動きベクトルの階層化はベクトルの合成によって行う。すなわち, 階層化のための処理は以下の式で表される。

$$V_{i,i+1} + V_{i+1,i+2} = V_{i,i+2} \tag{1}$$

ここで, $V_{i,j}$ はフレーム i から j へのある領域における動きベクトルを表す。これを繰り返して用いることにより, 図６のような階層構造を作る。動きベクトルの時間領域の階層化によって, 検索の高速化や, オブジェクトの動きの類似性の決定に利用することができる。

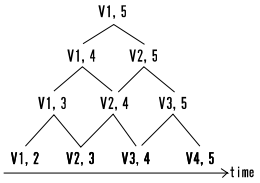

図６　動きベクトルの時間領域の階層構造

3.3　システムの構成

動きの階層性による動きの類似性を考慮した動画検索を実現するためのシステムの概略を図７に示す。

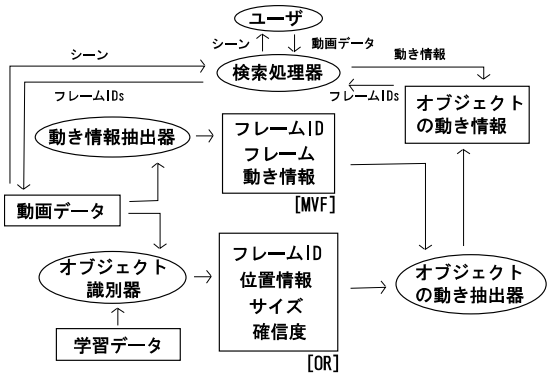

図７　システムの概略

このシステムは, ユーザが検索したいと思うオブジェクトとそのオブジェクトの動きが映っている動画データを入力すると, 類似するオブジェクトの動きが含まれるシーンを検索結果として返す。これは, 入力と検索対象のオブジェクトの動き情報を照合することによって得られる。

提案するシステムでは, ユーザの入力に先立って, 検索対象となる動画データに対して予めオブジェクトの動きを抽出するための処理を行う必要がある。この処理には, まずフレーム全体の動き情報 MVF とフレーム上のオブジェクトの領域情報 OR が必要となる。OR は, オブジェクトのサイズ, 位置, 認識の確信度, フレーム ID からなる。MVF は, 動き情報抽出器によって取得され, OR は, オブジェクト識別器によって取得される。そして, 取得された MVF とフレーム上の OR をオブジェクトの動き抽出器に入力することによってオブジェクトの動きを抽出する。

3.4　動き情報抽出器

動画データより隣り合う２フレームの情報を取得し, フレーム全体の動き情報 MVF を出力する。実装には OpenCV[6]と呼ばれるコンピュータビジョン向けのライブラリを使用する。ここで, 抽出精度と処理コストのトレー

ドオフを図るため, ブロックマッチング法を用いて抽出を行う。

3.5 オブジェクト識別器

動画データよりフレーム情報を取得し, フレーム情報から検索の対象とするオブジェクトを認識し, オブジェクト領域情報 OR を出力する。

オブジェクトの識別には, Viola らによるカスケード型識別器[7]を用いる。これは, 識別対象のイメージ領域上の輝度の差という簡単な情報を利用する識別器を複数用いることによって, 高速で精度の高い識別器を構築する。

3.6 オブジェクトの動き情報の抽出器

入力されるフレーム全体の動き情報 MVF とオブジェクト領域情報 OR に含まれる誤差の影響を軽減することによって, より精度の高いオブジェクトの動き情報を抽出し出力する。そのために, 図８のような木を構成し利用する。この木は, 入力される MVF と OR から推定されるオブジェクトの動きと, その動きに基づいて推定されるオブジェクト領域（EOR）を複数表している。実装では, 入力される OR の数は１フレームにつき１つとしている。

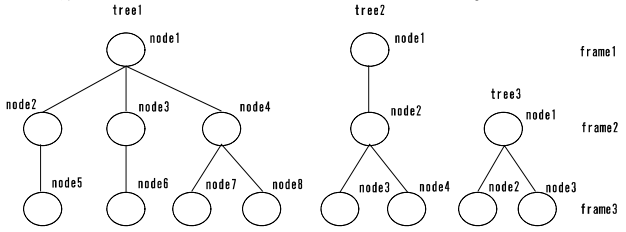

図８　オブジェクトの動き情報の抽出に用いる木構造

各木はある１つのオブジェクトに対応する。図８は, ３つの木があるため, ３つのオブジェクトが推定の対象になっている状態を示している。

さらに, 木の各ノードは, あるフレーム上での EOR を表し, 階層の深さと時系列が対応する。したがって, 同じ階層にあるノードは全て同じフレーム上での EOR を表す。例えば, frame2 に対応するフレーム上で, tree1 のオブジェクトの動きが３つ推定され（node2, 3, 4）, 同様に tree2 では１つ（node2）, そして新たに tree3 の初期領域であるルートが生成されている（node1）。

木のルートは, 入力される OR が既に生成されているノードが示す EOR と重なっていない場合にのみ生成される。子ノードは, 親ノードが表す EOR とその階層に対応する MVF を用いて, EOR の動きベクトルに対して生成条件を設定することによって生成する。生成条件に関しては, 4.1 の実験方法で扱う。その際に, 生成に用いた動きベクトルの割合を存在率として, 子ノードに格納する。さらに, 子ノードとして推定されたオブジェクト領域が, 識別器から得られたオブジェクト領域と重なれば, その確信度も子ノードに格納する。

最終的に, 構成された木のルートから葉ノードまでの１つの経路を確定し, その経路が表す動き情報の時系列をオブジェクトの動き情報とする。経路の決定条件は, 注目ノードの子ノードに対して, 確信度が最も大きいものを選ぶ。全ての子ノードに確信度がない場合は, 存在率が最も大きいものを選ぶ。

3.7 検索処理器

現在, この機構は未実装であるので実装方針を述べる。入力動画データからも検索対象の動画データと同様にオブジェクトの動き情報を抽出する。その後, 階層性を考慮した類似性の導入と効率的な検索を目的として, 各動き情報に含まれる動き時系列を階層化処理し, 図９のような検索用の木を構成する。ここで, $D^*(i,j)$ は図５と対応する D^* の割り当てで, フレーム i から j に対して時間領域の階層化を適用したものである。実際どの層でどちらの空間または時間領域の階層化処理を用いるかは, パラメータで設定できるようにする。

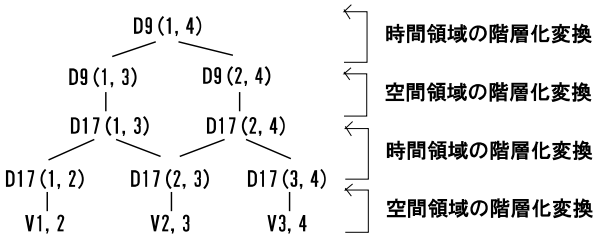

図９　検索用に構成する木

そして, 入力と検索対象の２つの木に対して, 上位層から順に動きを比較し, 類似度を求める。最終的に, すべてのオブジェクトの動きの類似度を求め, その値より最も類似するオブジェクトの動きと対応するシーンを検索結果とする。

4. 実験

4.1 実験方法

図７で示した提案システムのうち検索処理器が未実装であるため, 動画データに対してフレーム全体の動き情報とフレーム上のオブジェクトの領域情報を取得し, オブジェクトの動きを抽出し階層化する。そして, その結果より動きの類似性を決めるために導入した動きの階層性の影響を確認する。

ここでは, 茶道でのお点前の映像（図 10）から茶筅の動きの抽出を試みる。カメラと茶道に用いるオブジェクトの位置関係がほぼ同じである５つの映像から, 茶筅が動いているシーンを手作業で特定し, 実験用データを作成した。実験用データについて表１に示す。

図 10　使用する動画データの一部

表１　実験に使用するデータ

データ ID	DATA1	DATA2	DATA3	DATA4	DATA5
#frames	480	437	407	520	326
frame size	720×480				
frame rate	29.97[fps]				

また, 動き抽出器に対して以下の条件を設定する。
[A]子ノードの生成条件
(A－ｉ)オブジェクト領域内の動きベクトルのうち最も度数の多いもので子ノードを生成
(A－ii)オブジェクト領域の動きベクトルを空間領域の階層レベルに対応する 9 方向（図５参照）に割り当て, 最も数の多い割り当て方向に属する動きベクトルの平均値より子ノードを生成
さらに, 検索処理器の実装を想定して, 抽出されたオブジェクトの動きに対して以下の処理を行う。
[B]オブジェクトの動きを階層性に着目して変換
(B－ｉ)図５で示される 9 方向に変換
(B－ii)図５で示される 17 方向に変換

4.2　実験結果

子ノードの生成条件[A]を二通り適用させて得たオブジェクトと, その動きと対応する領域情報を動画データと視覚的に照らし合わせて, 抽出したオブジェクト領域に茶筅が映っているか確認した。その結果, 茶筅の領域の特定が不十分だったため, 茶筅の動きの抽出は DATA1 において部分的にのみしか確認できなかった。
また, 視覚的に条件[A]によるオブジェクト領域の動きを調べたところ, A－ｉでは, 視覚的に確認できる動きに反することはなかった。A－iiは, A－ｉよりも視覚的に細かな動きにも対応する場合が多かったが, 視覚的な動きと反する動きも目立った。これは, 動きベクトルのノイズの影響によるものであろうと考えられる。
確認できた結果を表２に示す。さらに, その中の motion ｉの動きを図 11 に図示する。動きの始点をフレームの中心として, 動きベクトルのスケールを２倍に拡大している。

表２　視覚的に確認できた茶筅の動き

識別 ID	生成条件	抽出区間
motion ｉ	A- ｉ	第 20～第 47 フレーム
motion ii -1	A- ii	第 18～第 23 フレーム
motion ii -2		第 460～第 470 フレーム

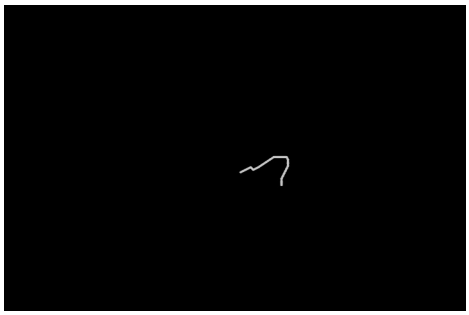

図 11　motion ｉの動き

さらに, motion ｉに対して処理[B]を適応した結果を図 12 に示す。

図 12　B－ii（左）と B－ｉ（右）による motion ｉの動き

図 11 および図 12 より, 空間領域における階層性に着目した変換ではなく, motion ｉの動き自身の性質による影響もあるが, 階層レベルを上げることによって, より的確に動きの特徴を捉えることができたと言える。

5.　まとめ

本稿では, 動きの階層性を考慮し, 動画データからオブジェクトとその動きを抽出し照合することによって, 入力動画データにて表されるオブジェクトとその動きの類似検索を行う手法を提案した。そして, オブジェクトの動き抽出に関する手法と動きへの階層性の導入手法の有効性の確認を実験によって試みた。
しかし, 今回はオブジェクトの動き抽出の精度が悪く, 有効性の確認までにいたらなかった。今後は, オブジェクト識別器の識別精度を改善し, 子ノードの生成条件をより工夫するなどオブジェクトの動き抽出の精度を向上させつつ, さらに検索処理器を実装してシステムを完成させて, 提案手法の有効性を確認する予定である。

謝辞

本研究は, 一部, 文部科学省科学研究費補助金（課題番号：20300037）による。

参考文献

[1] 柴田正啓：“映像の内容記述モデルとその映像構造化への応用”, 信学論, Vol.J78-D-Ⅱ, No5, pp.754-764（1995）
[2] 是津耕司, 上原邦昭, 田中克己：“時刻印付きオーサリンググラフによるビデオ映像のシーン検索”, 情処論, Vol.39, No.4, pp.923-932（1998）
[3] 加藤光幾, 石川博：“動画像を対象とする内容検索方式”, 情処研報 DBS, No.111-12, pp.87-94（1997）
[4] 山名信弘, 井辺昭人, 三浦文裕, 前島謙宣, 森島繁生：“動画の３次元周波数成分を用いた顔認証システム”, 信学技報, PRMU2006-22, MI2006-22, pp.13-18（2006）
[5] 國枝孝之, 脇田由喜, 高橋望：“MPEG‐7 と映像検索―マルチメディア情報検索の新手法を詳述 ”, CQ 出版, 第１版（2004）
[6] 怡土順一, 上田悦子, 小枝正直, 竹村憲太郎：”opencv.jpートップページー”, http://opencv.jp/
[7] P. Viola and M. Jones: “Rapid Object Detection using a Boosted Cascade of Simple Features,” In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.511-518（2001）