

動画像を対象とする内容検索方式

加藤光幾 石川博

(株)富士通研究所

要約

動画像を対象とする内容検索方式を提案する。動画像から動体領域を自動的に検出し、動体領域内の特徴量を抽出し検索に用いる。動画像データとして MPEG video を用い、マクロブロックごとの動ベクトル情報および DC 成分を用いて動体領域を検出し、移動方向を求める。また、動体領域内をいくつかの色領域に分割し、それらの代表色および相対方向、動体の移動方向を特徴量とする。検索時には、ユーザが検索したい物体の色配置と移動方向を入力し、動画像から抽出した特徴量との類似度を計算する。特徴量には背景領域の情報が含まれていないので、背景の影響を受けない検索が可能となる。本方式に基づいたプロトタイプを作成した。

Content-based Retrieval System for Video Data

Koki Kato and Hiroshi Ishikawa

Fujitsu Laboratories Ltd.

Abstract

A new content-based retrieval system for the video data is proposed. The moving object regions and the feature information of each moving object are automatically extracted from the video data (MPEG video) for the retrieval. The moving object regions and the moving direction are calculated with the motion vectors and the DC coefficient of each macroblock. The similarity scores between the desired object and the extracted objects are calculated with the color values, their locations of each object region, and the moving directions so that the scores do not contain the interference with the background. A prototype system based on this method has been developed.

1. はじめに

コンピュータの高速化により、デジタル化された静止画や動画像が多く使われるようになってきた。またさまざまな研究機関で開発が行われている電子図書館でも、従来の図書館に蓄積されてきたテキスト、絵画、図、ビデオテープなどが電子化されようとしている。これらの膨大なデータを有効に活用するには、ユーザが必要とする情報に容易にアクセスし利用することを可能とする技術が必要である。

われわれはマルチメディアデータを効率的に獲得・蓄積、検索、表示するための環境として、マルチメディアデータベース“InfoServer”を開発してきた[1][2]。InfoServer はオブジェクト指向データベースの上に構築しており、メディアデータの一部を示す位置情報をもつオブジェクト(ビューオブジェクト)をアクセス単位としている。ビューオブジェクトを用いることにより、余分なデータの複製を防ぎながらビューオブジェクトを単位とするデータの再利用を容易に行える。InfoServer はビューオブジェクトを時間・空間的に組み合わせる方法を記述するスクリプト、スクリプトを限られた資源で実行するため QOS 制御を行うスケジューラ、メディアデータの論理情報と物理情報を管理するデータモデルを含む。

InfoServer ではビューオブジェクトの検索にキーワードを用いてきた。しかしキーワード付けに手間がかかるなどの問題があった。

マルチメディアデータを扱うときの重要な点の一つは検索手段である。テキストデータに対しては、全文検索技術が開発され実用化されている。しかし、静止画、動画、音声などのデータはこれまで多く使われてきたテキストデータに比較すると抽象度が低く、このままでは検索することができない。そのため検索用の付加情報として人間が手作業でキーワードをつけることが多かった。しかし大量のデータにキーワードをつける作業は大変であり、作業者によって質にばらつきが生じる。またデータが持つさまざまな情報を

すべてキーワードにするのは不可能である。そのため生データから自動的に特徴量を抽出し検索に用いる内容検索技術が望まれる。マルチメディアデータは注目する特徴量によって様々な検索が可能となる。そのため将来的には複数の検索方法を用いて検索するようになると思われる。

本稿はそのような多様な検索方式の一つとして、動画像内に存在する動体を単位とし、動体内の色情報および移動方向を用いる内容検索方法を提案する。

以下では従来行われてきた内容検索の研究、本稿での提案方式、提案方式に基づくプロトタイプとそれを用いた実験結果、考察の順に述べる。

2. 従来研究

内容検索は従来静止画を対象として多く研究されてきた。QBIC (IBM Almaden Research Center)[3]は静止画内の色、テキストチャ、(背景が単純なときは)形を特徴量として用いた検索を行える。しかしこれらの特徴量は画面全体に関するものであり、空間の配置は考慮していない。一方、動画像に対してはショット検出と代表フレーム選択などの機能を持つ。VisualSEEK (Columbia Univ.)[4]は静止画を対象とし、色、テキストチャ、大きさに加え、色領域の位置、色領域間の位置関係も特徴量とした検索が行える。しかしこのシステムでも画面全体から求めた特徴量しか用いておらず、画像内の物体それぞれの特徴を現すことができない。そのためユーザが画像内の物体を検索したいときは、検索に用いる特徴量が物体以外の背景領域に影響され、適切な検索を行うのが困難だと考えられる。

動画像を対象とする研究としては、JACOB (Univ. of Palermo)[5]がある。JACOB では動画像をショットに分解し、ショットの代表フレーム(r-frame)を選択する。特徴量としてr-frame から抽出した色、テキストチャ、動き情報を用いる。動き情報は r-frame の前後のフレームから計算したオプティカルフローを用いる。動きを用いた領域分割も可能であるが半自動であり、ユーザが興味ある

オブジェクトをクリックするとシステムがオブジェクトの領域を計算し、領域にラベルをつける。JACOBもQBICなどと同様にフレーム全体の特徴量を用いているため、物体それぞれの特徴量を用いた検索は困難であると考えられる。

3. 動画像を対象とする内容検索

3.1. 本方式の目的

ユーザが物体を検索したいときユーザが想起するのは物体に関する特徴であり、背景がどうなっているかを気にしない場合も多い。しかし従来の画像を対象とする内容検索方式の多くは画像全体の特徴を用いており、特徴量には物体とそれ以外の背景領域がそれぞれ持つ特徴量が混合している。そのため、物体が持つ特徴量を用いて検索しようとしても、背景領域に影響されてしまい期待したような検索ができない。

我々は動画像を検索対象とし、画像内の物体ごとに求めた特徴量を用いる内容検索方式を提案する。しかしながら、任意の背景を持つ動画像から物体領域を検出するためには、あらかじめ物体の形状に関する知識が必要となるため難しい。我々は検索対象とする物体を、画像内の動いている物体(動体)に限定することにより、領域検出を単純化する。動体領域は隣接フレームにおける差異を用いることにより検出可能であり、動体の形状に関する知識を必要としない。特徴量は動体領域内部から計算されるため、それぞれの動体の特徴のみからなり、背景領域からの影響を受けておらず、動体を精度よく検索できることが期待できる。

本検索方式を用いて、ユーザが想起する物体の色に関する特徴を用いた検索や、それに加えて物体の移動方向を用いた検索が行える。例えば、

- 特定の物体が現れるショットの検索(服が変わらないと仮定したときに、特定の人物が出てくる場面を選択する)
- 動いている物体が存在する場面、存在しない場面の切り出し(画面上で激しい動きの場面、動きのない場面を選択する)

などの検索が可能となると考えられる。

本稿で述べる検索方式では、動画像データとしてMPEG videoフォーマットデータを用いる。MPEG videoデータは動画像フォーマットとして広く用いられている上、内容検索を行う上で有用な情報を含んでいる。MPEG videoデータには直前に復号した画像を元に動き補償を用いて復号するためのデータとしてマクロブロック(16x16画素)ごとの動ベクトル情報と色情報が含まれている。あるマクロブロックに含まれる動ベクトル情報は、そのマクロブロックの部分を前の画像の別の部分から動いてきたとみなして予測した時の移動量を現している。

3.2. 特徴量抽出処理の流れ

MPEG videoデータから検索に用いるための特徴量は次の手順で求める。

- (1) MPEG videoデータから、動体領域検出に用いる動ベクトル情報とマクロブロックの色情報(DC成分)を取り出す。なお、本稿での動ベクトルは移動方向と一致するように符号を変えている。
 - (2) 動画像からカット点を検出し、ショットごとに分割する。カット点抽出アルゴリズムは既にいくつかの研究が発表されており([6]など)、それらを用いることができる。以下の処理はショット単位で行う。
 - (3) 動ベクトル情報を用いてチルトやズームなどのカメラワークを検出する。動ベクトル情報をオプティカルフローの近似として扱うことにより、フローの傾向を調べカメラワークを求めることができる。この情報に基づき、背景領域が静止しているように動ベクトルの補正を行う。ただし後述のプロトタイプにおいてカメラワークの検出・補正は未実装である。
 - (4) 動体領域を検出する。
 - (5) 色分布を抽出する。
- (4)と(5)に関して詳細を以降に述べる。

3.3. 動体領域検出

静止画から物体の領域を検出するためには、物

体の形状についてあらかじめ知識を持っていることが必要である。動画像中の動いている物体(動体)を検出するのはそれより容易であり、動体領域はフレーム内の部分ごとの動きを示すオプティカルフローを計算することにより検出することができるが、その計算には時間がかかる。

我々は MPEG video データに含まれる情報を用いることにより、動体領域検出をより少ない時間で行う。検出方法は2種類用いている。一つは動ベクトル情報を用いる方式で、もう一つはマクロブロックの色の DC 成分を用いる方式である。本方式ではこれらの情報を用いてマクロブロック単位で動体領域を決定する。

動ベクトル情報を用いる方式では、動ベクトル情報をオプティカルフローの近似として用い、カメラワークの補正後の動ベクトル情報に対し、動ベクトル情報のベクトル長が閾値以上のマクロブロックを選択して、領域の固まりを求め、それぞれを動体領域とする。また、特徴量の一つである動体領域の移動方向は、動体領域内の動ベクトルの平均を計算することで求めている。

I-フレームには動ベクトル情報が含まれていないため、動体領域検出が行えない。そのため直前のマクロブロックの動ベクトル情報を用いて動体領域を推定する。ただし後述のように動ベクトルは必ずしも正確ではないので、重みをつけることにより推定が外れすぎないようにしている。

一方、動体が静止すると動ベクトルが0になるため動体領域がなくなり、物体が消滅したように見える。しかし実際には物体が画面上に存在する場合があるため、ある程度の時間の間はその領域を保持する。ただし本当に物体が消滅した場合もあるので、一定時間経過する間に再び動体領域が現れないときはその領域を消去する。

動ベクトル情報を用いて動体領域検出を行う場合に考慮する必要があるのは、動ベクトルが動画像を圧縮するために使用されていることから生じる傾向があることである。画像内の均一な色を持つ部分(単一色の壁や CG 画像)では、直前に復号した画像との差が少ない。この場合、あるマ

クロブロックに対して一番類似したブロックは自身の場所ではなく近傍にある可能性があり、動ベクトルがその方向を向くことがある。これは MPEG video データの動ベクトル情報は正しいオプティカルフローにノイズがのっていると考えられる。このノイズの影響を防ぎ、より精度よく動体領域検出を行うために、背景との色の変化を用いた動体領域検出法も用いている。

本検出方式はI-フレームのマクロブロックに含まれる色情報(DC 成分)を用いる。マクロブロックごとにそれが背景と判断されたときの色の平均値を求めておく。閾値以上に色が変わった場合はそのマクロブロックに物体が差し掛かると判断する。この方法はカメラが固定している状態にのみ有効である。

上記の動ベクトル情報を用いた方式や色情報を用いた方式で検出した動体領域ごとにラベルをつける。ラベル番号は隣接フレームでの領域の重なりを用いて計算しており、同一動体に対し固有のラベル番号が振られる。この情報により、ある動体が画面内に存在する期間やフレーム内の動きを調べることができる。

3.4. 色分布抽出

動体領域は幾つかの単一色領域から構成されていると仮定することができる。この性質を用い、動体領域をいくつかの色領域に分割し、各色領域を代表する色(代表色)、色領域の重心座標、色領域の縦横比を計算し、これらを検索に用いる特徴量としてデータベースに登録する。このとき、面積のより大きい色領域の方が特徴量として重要度が高いと仮定している。今のところ、テクスチャは扱っていない。

色分布抽出は次の処理手順で行う。

まず、あるフレームのデコード済みのフレームデータ(色は YCbCr フォーマットである)を読み込み、色情報を YCbCr から $L^*a^*b^*$ 表色系に変換し、以降は $L^*a^*b^*$ 空間で計算する。2点間の距離を計算するときに $L^*a^*b^*$ の方が線形性が保たれているからである。

次に、動体領域ごとに代表色を計算する。プロ

トタイプでは代表色を8色にしており、動体領域内の各色が占める面積を用いて計算する。

動体領域内を、求めた代表色で分割する。これは動体領域内の画素ごとに、もっとも近い代表色を探しその色で置き換えることを行っている。色による分割ができたなら、色領域それぞれを整形する。この際、面積が小さく重要性の低い色領域を除去する。この結果を用い、色領域ごとに代表色、面積、中心座標、縦横比を求めデータベースに登録する。

3.5. 検索アルゴリズム

動体の検索は検索用 GUI を用いて行う。ユーザは検索したい動体を持つ色配置と移動方向を指定して検索を指示すると、それに類似する動体を含む画像が類似度に応じて表示される。GUI は色配置用ウィンドウ、カラーパレット、動き方向を指定するためのウィンドウなどからなる。検索するとき、ユーザが検索したい物体を持つ色をカラーパレットから選択し、色配置用ウィンドウ上に長方形として配置する。その際、長方形の大きさや形、位置を設定できる。また、動体の移動方向を指定することもできる。

類似度は色とそれらの相対位置、領域の縦横比、移動方向を用いて行う。類似度計算は次のように行っている。

まず移動方向が指定されている場合は、移動方向が同じ動体をデータベースから選択し類似度を計算する候補とする。GUI で移動方向の指定がされていないときは、データベース内の全ての動体を対象に以下のように色を用いて類似度を計算する。

ユーザが入力した色配置の色ブロックを面積順にソートし、面積の上位 N 個だけを選択する。色ブロックのうち、面積が最大のものをピボット(相対方向を計算するときの原点)として設定し、残りの色ブロックへの相対方向を求める。

移動方向を用いて選択された動体ごとに、面積が上位 N 個の色ブロックを選択する。ある色ブロックをピボットとしたときの、残りの色ブロックへの相対方向を計算し、ユーザが指定した色配

置の相対方向と一致するかを調べる。一致したときはさらに色の距離を $L*a*b^*$ 表色系上で計算するとともに縦横比の一致度を求める(ただしプロトタイプでは縦横比を用いていない)。この時の N 個の色差と縦横比の一致度の合計を、ある色ブロックをピボットとして選択したときの動体領域のスコアとし、ピボットとする色ブロックを変えたとき最小のスコアを、その動体の類似度とする。

検索結果として、類似度の高い(スコアの小さい)ものをユーザに提示する。

4. プロトタイプシステム

以上で説明した方式を用いた内容検索システムのプロトタイプを作成し、簡単な評価を行った。

カメラを静止した状態で撮影した画像を対象とし、カメラワークの補正は行っていない。実験用の画像として、videocam で撮影したショッピングモールの画像と、室内で人間が歩いている画像を用いた。一旦録画したものを Sun WS の SunVideo を用いて 320x240 のサイズでキャプチャし、ソフトウェアエンコーダ(MPEG Software Simulation Group が開発した MPEG-2 Encoder Version 1.2)を用いて MPEG-1 video フォーマットに変換した。また色分布抽出で用いるデコーダは上記グループが開発した MPEG-2 Decoder を用いた。今のところ、特徴量はファイルに記録しており、データベースとの関係機能は開発中である。

検索ウィンドウは X-Window 上で作成した。色分布の入力は、カラーパレットから色を選択し、色分布入力部をクリックすると長方形が生成されるので、マウスを用いて長方形を適当な大きさ(面積、縦横比)に変え、画面上に配置する。検索ボタンを押すと色入力部の長方形それぞれの色、位置座標、面積、縦横比を検索プログラムに渡し、検索プログラムで類似度が高いと評価されたフレームの順に、該当フレームの前後を再生プログラムで動画として表示する。図 1 に検索ウィンドウのハードコピーを示す。画面は実際にはカラーで、左上の色分布入力部には左からオレンジ色、

白、黒の長方形が並んでいる。また動き方向入力部で左から右に移動している物体を検索するように指定している。

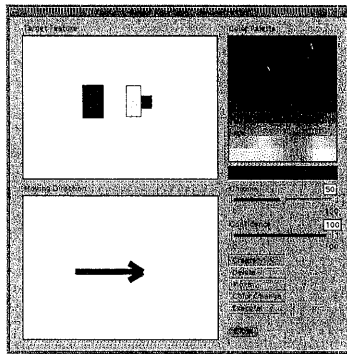


図1 GUI

4.1. 動体領域検出の結果

図 2は人間が画面上で比較的大きく映っている状態で歩行している動画像を用いて動体領域を検出した例である。ハイライト部分が動体として検出された部分である。また下の図は、それぞれ上の画像に対応した動ベクトルを示している。

4.2. 色分布抽出の結果

あるフレームにおける色分布抽出の結果を図 3に示す。左上の図のハイライト部分が色分布を抽出する動体領域であり、二人がベビーカーを押し

ながら右方向に歩いている。その他の図は 8 個の代表色でクラスタリングした結果を代表色ごとに表示したもので、黒色(バッグと頭部)、ベージュ色(壁面)、薄黄色(壁面)、白色(服)、オレンジ色(服)などに対応する。

これらの図は領域を整形する前のデータで、領域が細かく分割されている。特徴量を求めるときはこれらの領域を整形し、大きな領域だけを用いている。

4.3. 検索結果

検索例として、オレンジ色のシャツを着た人と、白い服に黒のバッグを持った人がいっしょに歩いている場面を検索する場合を示す(この場面は以前に見たことがあり検索したい色を知っていると仮定している)。ユーザは検索ウィンドウで図 1に示した色分布を入力し検索する。この検索に対し、一番類似度の高かったものが図 3の左上のフレームである。背景領域にはさまざまな色が存在するが、それらに影響されずに、検索したい物体の色のみで検索できている。

5. 考察

実験データに対する検索結果は比較的良好に思われる。しかしユーザが期待した検索結果と異なる動体領域の類似度が高くなることもある。現在用いているプログラムでは色分布抽出で

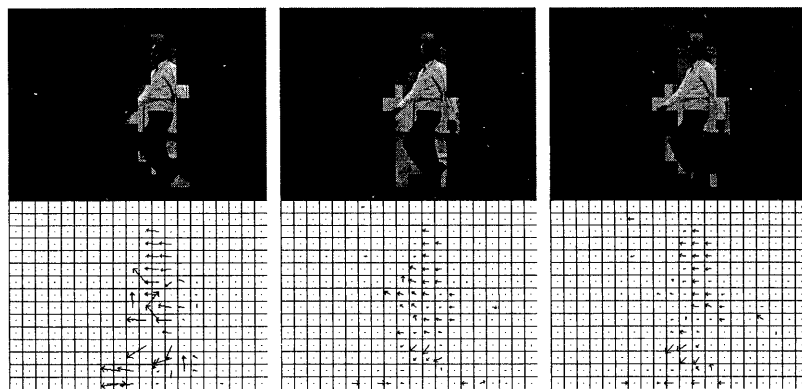


図2 動体領域と動ベクトル

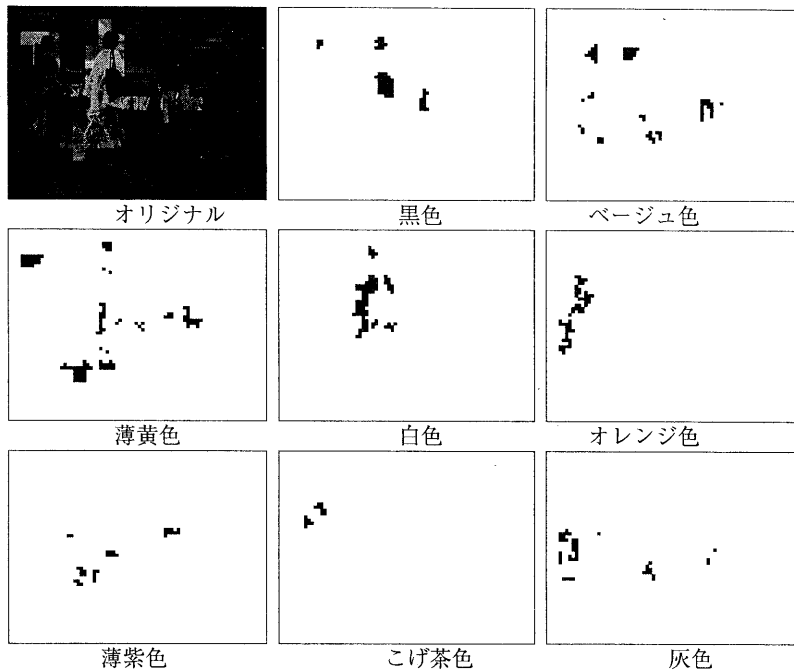


図3 動体領域と色分布

領域が細かく分割される傾向があり、人間が認識する領域と形が異なってしまう。これによる問題点は 2 つある。一つは領域の面積が異なるため、上記アルゴリズムで選択される色領域が合わなくなる。もう一つは形状が異なるため、縦横比が使えなくなる。これらにより検索時にユーザが入力するパターンとマッチングし難い。これらの問題点を解決するため、領域分割のプログラムを改良中である。

GUI では色選択をカラーパレットを用いて行っているが、色を指定するのが難しいように思われる。これはカラーパレットに表示されている色から人間が記憶しているあいまいな色を選択するのが難しいためではないかと考えている。カラーパレットの色の量を調整することにより対処することを検討している。

動体領域検出の精度に関する問題としては、動ベクトル情報に含まれる「ノイズ」による影響で、実際には動体でない領域も動体として認識され

ることがある。特に色変化が少ない領域では広い面積でノイズとなることがあり、その中の色分布を求めたときに領域面積が閾値を超えることがある。検索プログラムでは色領域の面積比は用いているが、面積そのものは利用していないため、それらの類似度が高くなってしまふことがある。この問題に関しては、動体領域検出の精度を上げることで現象の発生を減らそうとしている。

I-フレームの DC 成分を用いた動体領域検出では、領域の識別にマクロブロックの平均の色の変化を用いているため、物体がマクロブロックの一部だけを占めているときは色変化が閾値以下になる場合がある。そのため検出される領域が実際より小さくなりがちである。

現在用いている MPEG データは解像度が低いため、色分布抽出を精度よく行うために処理フレームごとに MPEG データをデコードしており、計算時間がかかる一要因となっている。将来、使用データを MPEG-2 video の MP@ML に変更する

と解像度があがるため、画面上で大きな動体を扱うときはマクロブロック単位で色分布を求めても精度が確保できると思われる。その場合はデコード処理が省けるので、特徴量抽出処理の高速化が期待できる。

6. まとめ

動画像を対象として、色情報と移動方向を特徴量として用いる内容検索方式を提案し、プロトタイプとその実行結果を示した。特徴量を動体領域から求めているため、背景領域の特徴に影響されずに動体を検索することができる。今後の課題としては、各モジュールの精度向上、カメラワーク補正機能の追加、検索機能の充実などがあげられる。また特徴量をデータベースに格納し、色空間からの検索を R-tree などの多次元インデクスを用いて高速化を行いたいと考えている。

参考文献

- [1] H. Ishikawa, et al., "A Next-Generation Industry Multimedia Database System", Proc. of the Twelfth International Conference on Data Engineering, Feb. 1996.
- [2] 加藤など, "マルチメディアデータベース InfoServer - スクリプトと動画再生系", 第7回データ工学ワークショップ(DEWS'96), Mar. 1996.
- [3] M. Flickner (IBM Almaden Research Center), et al., "Query by Image and Video Content: The QBIC System", IEEE Computer, Sep. 1995.
- [4] <http://www.ctr.columbia.edu/~jrsmith/html/pubs/acmmm96/acmfin.html>
- [5] <http://wwwcsai.diepa.unipa.it/research/projects/jacob>
- [6] 谷口など, "映像ショット切換え検出法とその映像アクセスインタフェースへの応用", 信学論(D-II), Vol. J79-D-II, No.4, Apr. 1996.