

移動物体と背景の描画による手描きスケッチ映像検索システム

瀬 倉 章 宏[†] 戸 田 真 志[†]

本稿では、クエリとして手描きスケッチを利用した映像検索システムを提案する．近年の映像共有サイトの普及に伴い、容易に様々な映像を視聴することが可能となった．しかしその反面、所望の映像を視聴するためにはそれら大量の映像群から映像を検索する必要が生じた．そのような状況の中で、著者らは手描きスケッチを利用した映像検索システムの提案を行ってきた．従来のシステムでは映像中の移動物体の形状とその動きに着目してきた．本稿においては、これまでの2つの要素の描画に加えて背景の描画を取り入れ、移動物体の形状と動きと背景という3つの要素を手描きのスケッチで描画することでクエリを作成し映像検索を行うシステムを提案する．また、提案手法をYouTubeから収集した映像に適用し、その結果も示す．

The Video Retrieval System Using Hand-Drawing Sketch Depicted by Moving Object and Background.

AKIHIRO SEKURA[†] and MASASHI TODA[†]

In this paper, we proposed the video retrieval system using hand-drawing sketch as a query. It became possible to watch various videos, because of the development of video-sharing websites. However, it is necessary to search for the desired video among the accumulation videos. In the past, we have proposed the video retrieval system using hand-drawing sketch. We improve this system by introducing background drawing.

1. はじめに

現在では、デジタルビデオカメラやカメラ機能付き携帯電話の普及により、誰もが容易に映像を撮影することが可能となった．またインターネットにおいては、YouTube¹⁾のような映像共有サイトが人気を博している．それらのサイトには個人が撮影した多くの映像が投稿され、利用者はそれらを視聴し楽しむことができる．しかし多種多様な映像が投稿・蓄積されているが故に、目的の映像を視聴する際には蓄積された大量の映像群からの検索を行う必要が生じる．

例えばYouTubeで検索を行う場合を考えてみる．YouTube上で検索を行う際には、各映像に設定されているタグ情報を利用したキーワード検索が使用される．試しに「猫、かわいい、ジャンプ」「柴犬、公園、ボール」というキーワードで検索を行ってみると、それぞれ1540件、347件の映像がヒットする(2010年10月25日現在)．これらの中から目的の映像を探し出すことは容易ではない．各映像に設定されるキーワ

ードには普遍的なキーワードが使用されていることが多く、また同時にキーワードのみで表現できることには限界があることもあり、検索キーワード数が増えるほど適切なものを設定することは難しくなると考えられる．また、一つ一つ映像を確認するという方法では手間がかかり、全てを確認することは難しい．

この問題を解決する為には、キーワードでは表現しきれないような要素も特徴量として利用することで、映像のさらなる絞り込みを行うことが有効であると考えられる．これまでも、映像特徴を利用して映像検索を行うという試み^{2)~4)}が行われてきた．そして著者らも、手描きスケッチをクエリとした映像検索システム⁵⁾について研究を行ってきた．これまでの提案システムでは映像中の移動物体の形状とその動きに着目しており、映像を検索する際にはそれら2つの要素を手描きのスケッチで表現することでクエリ作成を行っていた．本論文においては、これまでの2つの要素の描画に加えて背景の描画を取り入れ、移動物体の形状と動きと背景という3つの要素を手描きのスケッチで描画することでクエリを作成し映像検索を行うシステムを提案する．映像をスケッチで表現する際には、背景という要素はこれまで描画をサポートしてきた2つ

[†] 公立はこだて未来大学
Future University Hakodate

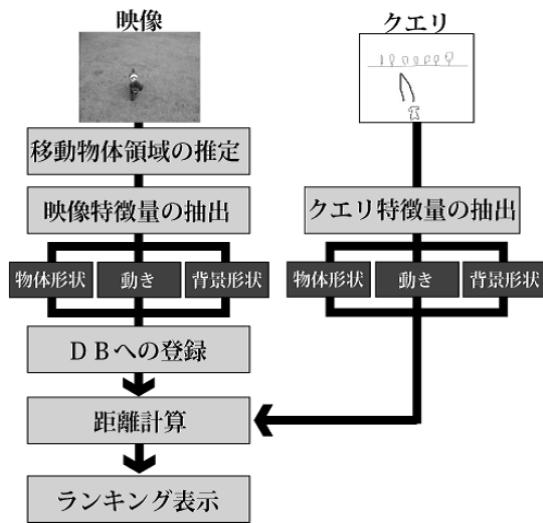


図 1 提案システムの概要図
Fig. 1 Overview of proposed system.

(映像中の撮影対象と動き)と同様に良く描画される要素であり、それを取り入れることでよりユーザの意図を反映したクエリの作成が可能となると考える。

2. 提案手法

提案手法全体の流れを示したものが図 1 である。提案手法では、検索対象とする映像から予め特徴量を抽出して特徴量データベース (DB) を用意しておく。特徴量としては映像中の移動物体の形状情報と動き、それから背景の形状情報を使用する。特徴量を求める際には、移動物体領域を認識し、移動物体領域からは物体の形状情報と動き、背景領域 (移動物体領域以外の領域) からは背景の形状情報をそれぞれ算出する。検索を行う際には、ユーザは検索したい映像のクエリを線画と矢印を使用した手描きのスケッチを描画することで作成する。このとき、物体形状と背景形状は線画を用いて表現し、物体の動きを矢印を用いて表現する。最終的にシステムはクエリと映像の特徴量間の距離を算出し、距離が近い順にランキング形式で結果を提示する。以降本章では、移動物体領域の抽出と特徴量の抽出、入力クエリ、検索時の距離計算についての詳細を述べる。

2.1 移動物体領域抽出

移動物体領域の抽出に関しては、今回のシステムでは手で領域の指定を行う。著者らのこれまでの研究⁵⁾においては、動きベクトルを用いることで移動物体領域抽出を自動で行っており、その手法をそのまま利用することも可能である。しかし今回は、より精度の高い領域抽出・特徴量抽出のこと考え、手動での領域指定を行う。具体的な方法としては、まず手動で領

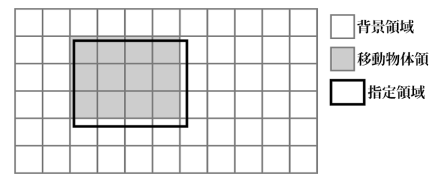


図 2 移動物体領域の指定
Fig. 2 Area definition of moving object.

域指定を行い、その後図 2 に示したようにフレーム画像全体をマクロブロックに分割して小領域単位で判定を行う。移動物体領域とする条件としては、指定領域内に含まれるもしくは、縦横の辺にそれぞれ 1/2 以上重なりを持つ、の 2 つである。

2.2 特徴量抽出

移動物体領域からは物体の形状特徴量と動き特徴量を、背景領域からは背景の形状特徴量をそれぞれ求める。

2.2.1 物体の形状特徴量

形状特徴量としては大橋らのエッジ形状特徴量⁶⁾を利用する。この特徴量は、画像内の線画素の相対的な位置関係を 256 次元のヒストグラムとして表現したものであり、画像の大きさや描画位置の変化に影響されないという特徴を持つ。移動物体と推定された領域に対して、全てのフレームでエッジ形状特徴量を求め、映像のエッジ形状特徴量 Fm とする。ここで i はフレーム数を表している。

$$Fm = \{Fm_1, Fm_2, \dots, Fm_i\} \quad (1)$$

2.2.2 動き特徴量

移動物体として推定された領域は全て同じようなフローベクトルを持つと仮定し、各フレーム画像において移動物体領域内の OF を平均したベクトルを算出する。ただし、映像中で何らかの物体が移動しているということを仮定し、映像中の対象が移動はしないが何か動作をする場合は想定しない。また、ベクトルを平均する際には、カメラワークの影響を取り除くために、領域内の OF は背景の動きとのベクトル差を求めたものを使用する。このようにして得られた各フレームにおけるベクトルをまとめてベクトルの連続データ Vm とし、これを映像中移動物体の動き特徴量とする。

$$Vm = \{Vm_1, Vm_2, \dots, Vm_i\} \quad (2)$$

動き特徴量抽出の模式図を図 3 に示す。

2.2.3 背景の形状特徴量

特徴量としては、物体の形状特徴量と同様に大橋らのエッジ形状特徴量を利用する。背景領域 (移動物体領域以外の領域) に対してエッジ形状特徴量を求める。ただし、背景については短時間で大幅に変化するということはほとんど無い為、すべてのフレームで特徴量

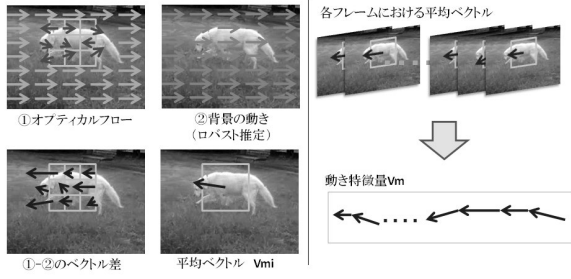


図 3 動き特徴量抽出

Fig. 3 Estimating motion feature.

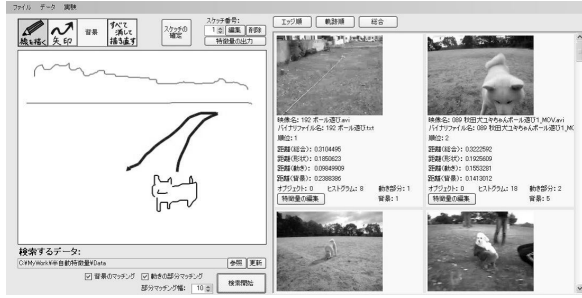


図 4 インタフェースとクエリ例

Fig. 4 The interface and an example of query.

を求めるのではなく、一定フレームごとに求めるものとする。求めた背景の形状特徴量をまとめて、映像の背景形状特徴量 Bm とする。ここで j は形状特徴量の数を表わす。

$$Bm = \{Bm_1, Bm_2, \dots, Bm_j\} \quad (3)$$

2.3 入力クエリ

クエリは線画と矢印を直接描画することで行う。クエリの入力例を図 4 に示す。今回のシステムでは物体と動きと背景のどれを描画するかについては、その都度ユーザが指示するようにしている。そのため線画と矢印、物体と背景等の分離については考える必要はないものとする。

2.4 検索時の距離計算

クエリと映像特徴量間の距離は、物体形状と動きと背景形状のそれぞれの距離を合計することによって算出する。最終的な距離 D は以下の式で表わされる。

$$D = D_V + D_F + D_B \quad (4)$$

D_F は物体形状特徴量間の距離を、 D_V は動き特徴量間の距離を、 D_B は背景形状特徴量間の距離をそれぞれ表わしている。以降、それぞれの距離計算方法について述べる。

2.4.1 物体形状特徴量

物体形状間の距離 D_F を算出するには、ユーザが描いたクエリから抽出した物体の形状特徴量 Fq と映像中の移動物体の形状特徴量 Fm とのユークリッド距離を求める。その際、特徴量 Fm は移動物体が存在

すると推定されたフレーム分だけ 256 次元の特徴量を持つが、その中で最小のものを採用する。

$$D_F = \min_{i=1,2,\dots,I} \left\{ \sqrt{\sum_{k=0}^{255} (Fq_k - Fm_{ik})^2} \right\} \quad (5)$$

ここで、 i は移動物体が存在するフレーム数、 k はヒストグラムの度数を表わす。

2.4.2 動き特徴量

動きの特徴量間の距離 D_V は、クエリから抽出されたベクトルの連続データ Vq と移動物体から抽出されたベクトルの連続データ Vm との距離を算出することで求める。 Vq の抽出方法であるが、まずクエリとして描画された矢印を一定ストロークごとに分割する。分割された線分ごとにベクトル化し、ベクトルの連続データ Vq とする。 D_V の計算には DP マッチングを用いる。DP マッチングを行う際には、対応させた 2 つのベクトルの角度を比較し距離 $d(l, m)$ の計算を行う。ここで N, M はそれぞれ Vm, Vq のベクトル数であり、 w は warping path、 K は warping path の総数である。

$$D_V = \frac{1}{L+N} \sum_{k=0}^K g(w_k) \quad (6)$$

$$g(w_k) = d(l, n) = 1 - \frac{Vm_l \cdot Vq_n}{|Vm_l| |Vq_n|} \quad (7)$$

加えて、前述の DP マッチングの他に部分的な DP マッチングも行う。部分的な DP マッチングは、 Vm を一定区間ごとに部分分割し、複数の連続データ $Vm_{part0}, Vm_{part1}, \dots, Vm_{partI}$ を作成することによって実現する。 Vm と同様に Vm_{part} についても DP マッチングにより Vq との距離 D_V を求め、それらの中で最小のものを移動物体との距離 D_V として採用する。

2.4.3 背景形状特徴量

背景形状間の距離 D_B を算出するには、 D_F と同様に、ユーザが描いたクエリから抽出した背景の形状特徴量 Bq と映像中の移動物体の形状特徴量 Bm とのユークリッド距離を求める。映像の背景形状特徴量として複数存在するヒストグラムのうち、距離が最小のものを採用する。

$$D_B = \min_{j=1,2,\dots,J} \left\{ \sqrt{\sum_{k=0}^{255} (Fq_k - Fm_{jk})^2} \right\} \quad (8)$$

3. 検索と考察

提案手法の効果を確認する為に YouTube に投稿されている映像を取得し、それらを用いた検索実験を行っ

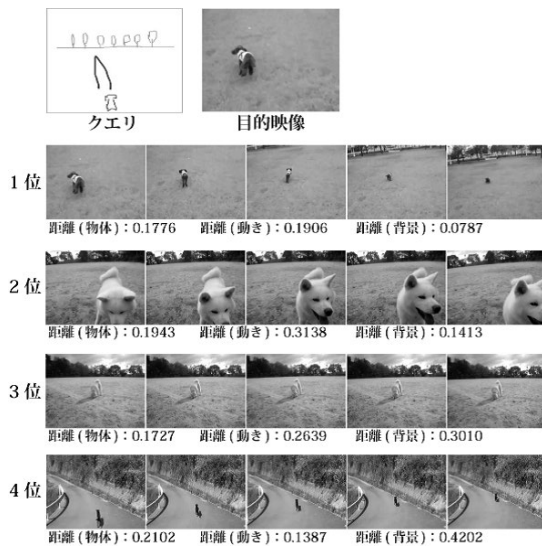


図 5 結果の例 1

Fig. 5 Example.1 of result.

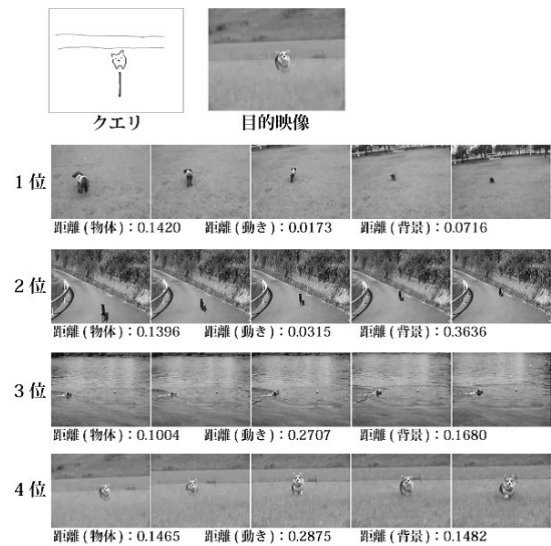


図 6 結果の例 2

Fig. 6 Example.2 of result.

た．映像を検索するキーワードとしては、「犬」「ボール」「遊ぶ」という3つを設定し、該当映像1160件の内から上位407件を取得した．今回の実験においては取得した映像407件のうち、再生時間が30秒以下の映像35個を用いた．

3.1 検索結果

検索結果の例を図5と図6に示す．ここに示した例では目的の映像が上位に絞り込めている．また、背景形状のマッチングをしなかった場合、図5の目的映像は第3位に、図6の目的映像は第10位にそれぞれ検索順位が下がってしまう．映像にもよるが、背景の描画を取り入れることにより、よりユーザの意図を反映した形で検索が行えるようになったと考えられる．

3.2 検索の考察

図5と図6で示した例は背景の描画を取り入れることにより上手く検索することができた例であるが、当然上手く検索できないような場合も存在する．例えば、背景に大量の物体が存在するような場合、ユーザの描画と背景が異なってしまう、特徴量が一致しないという問題が生じる．この問題を解決するためには、映像の背景からそのまま特徴量を求めるのではなく、背景内の物体ごとに領域分割を行い領域ごとに形状特徴量を抽出し、距離計算を行う際にはクエリに合わせてそれらの組み合わせの中から最適な背景特徴量を作るといった工夫が必要となると考えられる．

4. おわりに

本論文では、移動物体の形状と動きと背景を手描きのスケッチで描画することでクエリを作成し映像検索

を行うシステムを提案した．また YouTube の映像を用いた実験により、背景の描画を取り入れることで、従来手法に比べてさらに上位に検索することができる場合があることを示した．

今後の展望としては、背景領域からそのまま特徴量を求めるだけでは上手くいかない場合があることが確認されているので、それらに対応すべく特徴量抽出手法の改善に取り組んでいくことを考えている．

参考文献

- 1) YouTube, <http://www.youtube.com/>
- 2) 三上弾, 東正造, 森本正志: カメラ注目点指示によるクエリ動画作成および LHI に基づく映像シーン検索, 信学技報.PRMU, Vol.106, No.606, pp.67-72, (2007).
- 3) 藤崎宣彦, 宮田高道, 小林亜樹ほか: 適合性フィードバックを用いたクエリ作成支援型動画検索システム, 信学技報.ITS, Vol.105, No.608, pp.95-100, (2006) .
- 4) 青木秀一, 青木輝勝, 安田浩: 動画からのシーン検索のための略画処理手法の提案, 情報処理学会研究報告.CVIM, Vol.2002, No.2, pp.133-140, (2002) .
- 5) 瀬倉章宏, 戸田真志: 動きと形状特徴量を用いた手描きスケッチによる映像検索システムとその利用, 信学技報, Vol.110, No.238, pp.109-114, (2010).
- 6) 大橋剛介, 長島康剛, 望月圭太, 下平美文: エッジ形状に基づいた手書きスケッチ画像検索, 映像メディア学会誌, Vol.56, No.4, pp.653-658, (2002) .