

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 27 日現在

機関番号：17401

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500111

研究課題名（和文）映像想起とそのスケッチ表現メカニズムを利用した未整理映像の検索方式

研究課題名（英文）Video Retrieval System Using Scene Remembrance and Its Hand-drawing Sketch Representation

研究代表者

戸田 真志（TODA MASASHI）

熊本大学・総合情報基盤センター・教授

研究者番号：40336417

### 研究成果の概要（和文）：

映像情報爆発時代における大量の未整理映像を対象として、映像中のオブジェクトとその動きを時空間的に重畳した手描きスケッチ画像をクエリとする映像検索システムを実現した。手描きスケッチ画像は、一連の映像における心象を形作るオブジェクトやその動きを時空間的に重畳したものであるため、これら各要素を分解し、映像クエリとして再整形する。具体的には、ユーザは、検索したい映像のクエリとして、線画と矢印を使用した一枚のスケッチを描画し、映像中に含まれるオブジェクトおよびその動き情報とのマッチング処理を行うことで、合致度の高い映像記録を検索結果として提示している。YouTube から取得した映像に対し提案手法を適用し、その有効性を検証した。

### 研究成果の概要（英文）：

In this research, we have realized the video retrieval system using hand-drawing sketch representation, which has an object, its motion and the background figure. A user make a sketch by line drawing and arrow figure, and our system does a matching processing between user's sketch representation and characteristics which are acquired by video. The proposal technique was applied to the video acquired from YouTube, and the validity was verified.

### 交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010 年度	1,200,000	360,000	1,560,000
2011 年度	1,000,000	300,000	1,300,000
2012 年度	1,200,000	360,000	1,560,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：画像工学・画像計測・メディア工学・情報検索

科研費の分科・細目：メディア情報学・データベース

キーワード：スケッチ描画・心象画・映像検索・未整理映像・エッジ特徴・動きベクトル

#### 1. 研究開始当初の背景

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、YouTube 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す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（図 1）。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。ところで、我々が映像、特に一度見たことのある映像を想起しつつ他人に説明するとき、その内容を、図 2 に示すような対象（オブジェクト）形状の線画と対象の動き情報（矢印の場合が多い）、および背景を融合させた簡単なスケッチにて「こんな感じ」と説明/表現することがしばしばあることに気がつく。いわゆる心象画である。心象画としての手書きスケッチは、一連の映像の中に登場するオブジェクトやその動きを時空間的に重畳し、長時間の映像もわずか数枚の画像として表現している。ここで重要なことは、心象画としての手書きスケッチは、我々にとって日常的な表現行為であり、作成の負担が極めて低い、ということである。このような、我々が日常行っている映像想起とその表現のメカニズムを充実再現することにより、ユーザのクエリ作成の負担の少ない、全く新しい映像検索/再検索の方法論が提案できると考えた。



図 1 キーワードによる映像検索



図 2 スケッチ描画例

## 2. 研究の目的

上記を受けて、本研究では、心象画としての手書きスケッチ画像を検索クエリとする映像検索手法を検討する。手書きスケッチ画像は、一連の映像における心象を形作るオブジェクトやその動き、および背景を時空間的に重畳したものであるため、これら各要素を分解し、映像クエリとして再整形することが中心課題となる。具体的には、手書きスケッチとして描かれた線画について、スケッチ中のオブジェクトの形状情報とその動き情報、および背景情報とに分離し、それらとのマッチング度合いの高い映像記録を検索結果として提示する。提案システムの概念図を図 3 に示す。



図 3 スケッチによる映像検索

## 3. 研究の方法

提案手法全体の流れを示したものが図 4 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、距離に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。

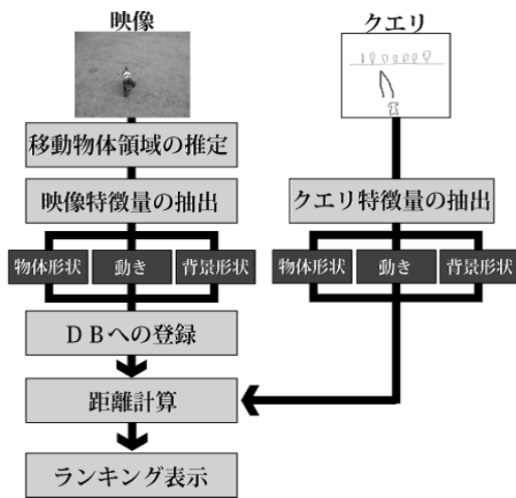


図 4 提案システムの概要図

(1) 移動物体領域抽出

映像中の移動物体は、映像のオプティカルフローを用いることで抽出する。撮影カメラの移動パラメータと移動物体とを区別するために、あるシーン全体のオプティカルフローをロバストクラスタリングすることで、カメラの移動パラメータと移動物体とを区別する。

(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。

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  $F_m$  とする。ここで  $i$  はフレーム数を表している。

$$F_m = \{F_{m_1}, F_{m_2}, \dots, F_{m_i}\}$$

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オプティカルフロー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、領域内のオプティカルフロー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  $V_m$  とし、これを映像中移動物体の動き特徴量とする。

$$V_m = \{V_{m_1}, V_{m_2}, \dots, V_{m_i}\}$$

動き特徴量抽出の様子を

図 5 に示す。

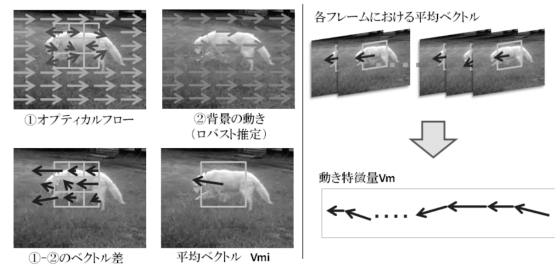


図 5 動き特徴量抽出の様子

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  $B_m$  とする。ここで  $j$  は形状特徴量の数を表す。

$$B_m = \{B_{m_1}, B_{m_2}, \dots, B_{m_j}\}$$

(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 6 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。



図 6 インターフェースとクエリ例



(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。

$$D = D_V + D_F + D_B$$

$D_F$  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している。以降、それぞれの距離計算方法について述べる。

① 物体形状特徴量

物体の形状特徴量  $D_F$  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  $F_q$  と映像中の移動物体の形状特徴量  $F_m$  とのユークリッド距離を求める。その際、特徴量  $F_m$  は移動物体が存在すると推定されたフレームのみ 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。なお、下式にて、 $i$  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す。

$$D_F = \min_{i=1,2,\dots,J} \left\{ \sum_{k=0}^{255} (F_{qk} - F_{m_{ik}})^2 \right\}$$

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  $V_q$  と移動物体から抽出されたベクトルの連続データ  $V_m$  との距離を算出することで求める。 $V_q$  の抽出方法は、まずクエリとして描画された矢印を一定ストローク毎に分割する。分割された線分毎にベクトル化し、ベクトルの連続データ  $V_q$  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  $V_m$ 、 $V_q$  のベクトル数であり、 $\omega$  は warping path、 $K$  は warping path の総数である。

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

$$g(\omega_k) = d(l, m) = 1 - \frac{V_{m_l} \cdot V_{q_m}}{|V_{m_l}| |V_{q_m}|}$$

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  $V_m$  を一定区間毎に部分分割し、複数の連続データを作成することによって実現する。これらの部分データに関しても DP マッチングにより  $V_q$  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。

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  $B_q$  と映像中の移動物体の形状特徴量  $B_m$  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。

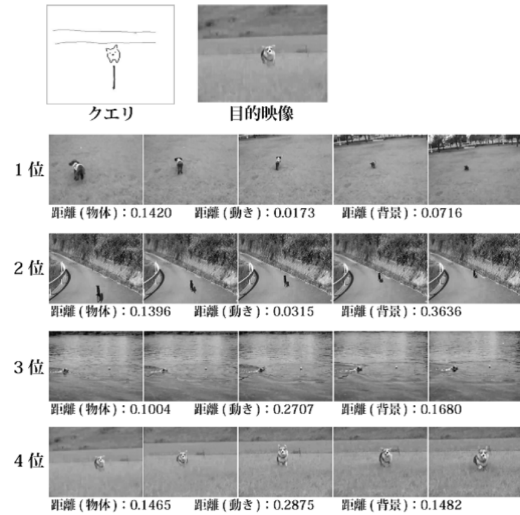


図 7 実験結果例 (1)

4. 研究成果

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。検索結果の例を図 7 と図 8 に示す。目的の映像が上位に絞り込めていることがわかる。

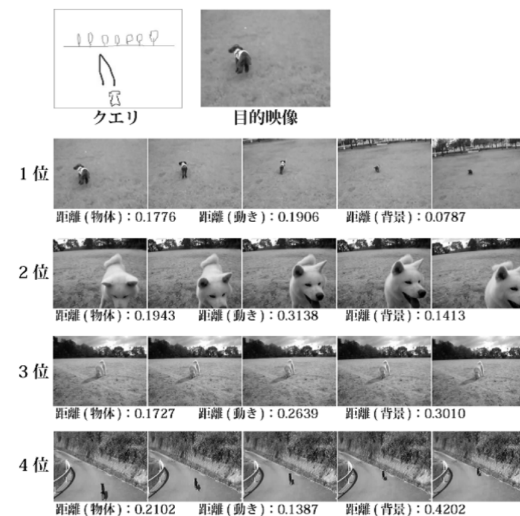


図 8 実験結果例 (2)

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 1 件)

① Hiroki Kobayashi、 and Masashi Toda、  
“Utilization of Textural Features in  
Video Retrieval System by Hand-writing  
Sketch”、 International Journal of Image、  
Graphics and Signal Processing (IJIGSP)、  
Vol.4、 No.8、 pp.1-7、 2012、 査読有。

[学会発表] (計 7 件)

②相田彰大、戸田真志、 “適合性フィードバックを用いた手書きスケッチ映像検索システム”、 情報処理学会エンタテインメントコンピューティング研究会、 2012年3月26日、 神戸大学、 査読無。

③小林広樹、 戸田真志、 “手描きスケッチ映像検索システムにおけるテクスチャ特徴量の利用”、 動的画像処理実利用化ワークショップ2012、 2012年3月8日、 公立はこだて未来大学、 査読有。

④Hiroki Kobayashi、 and Masashi Toda、  
“Video Retrieval System with Texture  
Feature by Hand-writing Sketch”、 18<sup>th</sup>  
Japan-Korea Joint Workshop on Frontier of  
Computer Vision (FCV2012)、 2012年2月3  
日、 川崎市国際交流センター、 査読有。

⑤ Akihiko Aita、 and Masashi Toda、  
“Relevance Feedback For Video Retrieval  
System Using Handwriting Sketch”、 18<sup>th</sup>  
Japan-Korea Joint Workshop on Frontier of  
Computer Vision (FCV2012)、 2012年2月3  
日、 川崎市国際交流センター、 査読有。

⑥瀬倉章宏、戸田真志、 “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”、 インタラクション2011、 2011年3月10日～12日、 日本科学未来館、 査読有。

⑦五十嵐勇一、 戸田真志、 “テクスチャを考慮した手書きスケッチによる画像検索に関する研究”、 平成22年度 IEICE 北海道支部学生会インターネットシンポジウム、 2011年2月16～25日、 査読無。

⑧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”、 電子情報通信学会 MVE 研究会、 2010年10月21日～22日、 とちぎプラザ、 査読無。

## 6. 研究組織

### (1) 研究代表者

戸田 真志 (TODA MASASHI)

熊本大学・総合情報基盤センター・教授  
研究者番号：40336417