

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

相田 彰大^{†1} 戸田 真志^{†2}

昨今、動画共有サイトの台頭により、一般ユーザが手軽に動画の投稿・共有を行うようになった。しかし、膨大な量の動画が貯蔵されるようになるにつれて、一般的に動画検索を行う際に用いられている「キーワード検索」だけでは動画を絞り込む事が困難になってきた。ユーザが所望の動作をイメージし、そのシーンに沿ったキーワードをいくつか設定することで検索を行ない、動画を絞り込んでいくという行為は容易ではない。そこで、本研究では、筆者らが既に提案している手書きスケッチを用いた映像検索システムに、適合性フィードバックシステムを適用することでセマンティックギャップを埋め、検索精度の向上を図ることを目的とする。ユーザは、オブジェクトの形状と動作を手書きスケッチで描き、検索クエリを生成する。動画から形状と移動物体の動きを抽出し、それを特徴量に変換し検索クエリとマッチングさせる。そして、この結果を適合性フィードバックシステムを用いてフィードバックすることで検索精度の向上を図る。実験でこの提案システムを適用したところ、従来手法では検索が困難な動画に対してもアプローチすることができ、本手法の有用性を示すことができた。

Video Retrieval System of Handwriting Sketch using Relevance Feedback

AKIHIRO AITA^{†1} and MASASHI TODA^{†2}

It is difficult to represent video scenes using keywords. Therefore, in video retrieval, it is not easy to design keyword queries that enable a viewer to represent and find the desired scenes. We have proposed the video search system which already used the handwriting sketch. In this study, this system propose added relevance feedback that bridge semantic-gap. The user queries the system by drawing a handwriting sketch that includes object shape and motion. The proposed system matches the query and target video by using a histogram of the relative positions of shape edges and the trajectory of moving objects. And, this result is made to learn the relating videos data, and it is feedback. Experimental results showed that simple videos can be fined by this proposed system.

1. はじめに

近年、デジタルカメラやビデオカメラ、カメラ機能付きの携帯電話など映像撮影機器の普及により、動画を記録するという行為が誰でも簡単に行えるようになった。また、「YouTube」¹⁾ や「ニコニコ動画」²⁾ など web 上で動画を公開できる動画共有サイトの台頭で、個人で撮影した動画を手軽に共有する機会が増え、大量の動画が蓄積されるようになってきた。このようなサイトで動画を検索する際には、テキストを用いて行うキーワード検索が一般的であるが、この一般的な方法で大量に蓄積された動画群の中から所望の動画を探し出すためには、キーワード検索を行った後に一つ一つ動画を再生して内容を確認していく作業が必要となる。しかし、すべての動画を一つずつ再生し、内容を確認するという作業は非常に手間がかかり現実的ではない。このような問題を解決するためには、キーワードでは表現できないような他の要素を新たに特徴量として利用し、動画の更なる絞り込みを行うという方法が有効であると考えられる。筆者らは、この新たな特徴量として「手書きスケッチ」を用いて映像検索を行うシステムを既に提案している³⁾。この手法はまず、検索対象とする動画からあらかじめ「形状特徴量」と「動き特徴量」という2つの特徴量を抽出してデータベースに登録しておく。そして、検索を行う際には、ユーザは検索したい所望の動画を線画と矢印を用いた一枚のスケッチとして作成する。動画と同様に、このスケッチから「形状特徴量」と「動き特徴量」の2つの特徴量を抽出し、データベースに既に登録してある動画の各特徴量と照合することで、所望の動画を検索しようとしたものである。

本研究では、この筆者らが既に提案している手書きスケッチを用いた映像検索システムに、適合性フィードバックシステムを適用することで、セマンティックギャップを埋め検索精度の向上を図ることが目的である。手書きスケッチを用いて検索を行う場合、動画内から抽出してきた特徴量よりも、過去に誰かが入力したスケッチの特徴量の方が、類似した特徴量を示す可能性が高い。これは、ユーザがスケッチを描く際、動画を細部まで緻密に再現し描画することがほぼ不可能だからである。また、検索クエリのスケッチは自由描画であるため、ユーザの癖や描き方など個人によって著しく差が生まれてしまう。したがって、データ

^{†1} 公立はこだて未来大学大学院
Graduate School of Future University Hakodate

^{†2} 公立はこだて未来大学
Future University Hakodate

ベース内の動画群に、入力されたスケッチを関連付けすることができれば、動画の各特徴量とユーザの主観のギャップであるセマンティックギャップを埋めることが可能になり、検索精度の向上が期待できる。よって本研究では、この手書きスケッチを用いた映像検索システムに適合性フィードバックシステムを適用し、検索精度を向上させる。

映像検索の分野ではないが、手書きのスケッチをクエリとして利用した研究は数多く存在する。例えば画像検索の分野では、林ら⁴⁾、望月ら⁵⁾、多々良ら⁶⁾の研究などが挙げられる。これらの研究では、それぞれ背景を構成する線、画像中の物体の位置や色やテクスチャ、画像のエッジといった、言語で表現することが困難な要素をスケッチインタフェースを利用して描画することで設定し、画像検索を行っている。また、大橋⁷⁾によって画像検索に適合性フィードバックを適用した研究も行われている。このようなことから、スケッチ表現を用いる方法や適合性フィードバックを適用するという方法は、映像検索においても所望の動画を表現する場合には有効な方法になると考えられる。

2. 提案手法

2.1 システムの概要

映像検索システムの概要を図1に示す。「手描きスケッチによる映像検索システム」では、動画中の「物体形状」と「物体の動線」の2つの情報を用いる。この2つの特徴量を求めるには、動画中の移動物体を認識し、その領域情報に基づいて計算する必要がある。あらかじめ検索対象とする動画から、この「形状特徴量」と「動き特徴量」という2つの特徴量を抽出してデータベースに登録しておく。そして映像検索を行う際には、ユーザは検索したい動画を、線画と矢印を用いた一枚のスケッチとして作成する。この入力したスケッチの特徴量と動画の特徴量間の距離を計算し、距離が近い順にランキング形式で出力することで結果を表示する。このような方法を用いて、スケッチ検索システムを実現した。次節において、特徴量の抽出方法、距離の計算方法を示す。

2.1.1 特徴量の抽出方法

移動物体領域から物体の形状特徴量と動き特徴量の2つを求める。まず、形状特徴量の抽出方法について述べる。形状特徴量については大橋らのエッジ形状特徴量⁸⁾を利用する(図2)。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、動画の

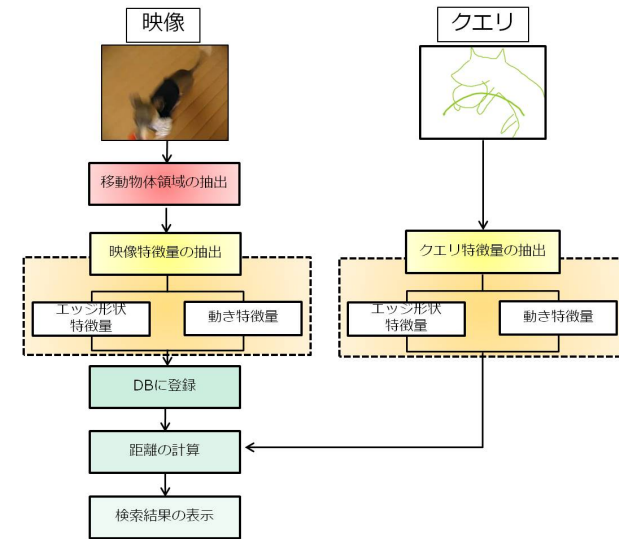


図1 検索システムの概要図

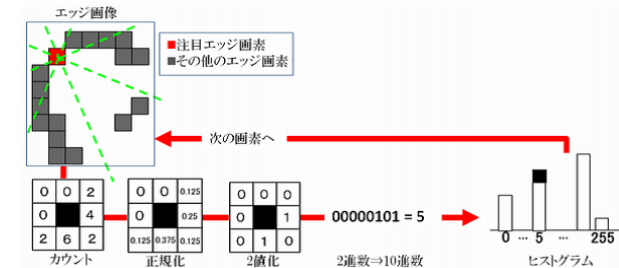


図2 エッジ形状特徴量計算の流れ

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。

$$F_m = \{F_{m1}, F_{m2}, \dots, F_{mi}\} \quad (1)$$

次に、動き特徴量の抽出方法について述べる。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 (OF) を平均したベクトルを算出する。ただし、動画中で何らかの物体が移動しているということを仮定し、動画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

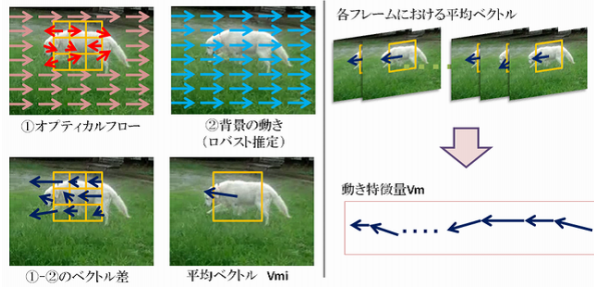


図 3 動き特徴量抽出

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 V_m とし、これを動画中移動物体の動き特徴量とする (図 3)。

$$V_m = \{V_{m1}, V_{m2}, \dots, V_{mi}\} \quad (2)$$

この 2 つの特徴量を動画から抽出してデータベースに登録しておく。

2.1.2 距離の計算方法

クエリとなるスケッチの特徴量と動画特徴量間の距離は、物体形状と動き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 は次式で表わされる。

$$D = CD_V + D_F \quad (3)$$

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それぞれ表わしている。また、 C は重み係数を表わしている。以降、それぞれの距離計算方法について述べる。

まず、物体形状間の距離 D_F を算出する。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 と動画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。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。

$$D_F = \min_{i=1,2,\dots,I} \left\{ \sqrt{\sum_{k=0}^{255} (Fq_k - Fm_{ik})^2} \right\} \quad (4)$$

次に、動きの特徴量間の距離 D_V を算出する。クエリから抽出されたベクトルの連続データと移動物体から抽出されたベクトルの連続データとの距離を算出することで特徴量間の距離を求める。2 つのベクトル連続データの比較を行うに当たっては DTW (Dynamic Time Warping)⁹⁾¹⁰⁾ を利用する。Dynamic Time Warping とは、異なる長さを持つ時系列データを時間的に伸縮させることで、最適なマッチングを行う手法である。

まず、映像の動き特徴量を V_m 、クエリの動き特徴量を V_q とする。距離を求める際には、DTW により V_m と V_q のデータ間の最適な対応付けを求めることになるが、そこで一つ考慮すべき点がある。それは、ユーザが入力した移動物体の動きは映像中の移動物体の動きの一部である可能性があるという点である。 V_m を一定区間ごとに部分分割した $V_{m_{part0}}, V_{m_{part1}}, \dots, V_{m_{partI}}$ を作成することによって動き特徴量の部分マッチングを実現し、ユーザの描画した動きが全体の一部分であった場合に対応する。

$$D_V = \frac{1}{N+M} \sum_{k=0}^K g(w_k) \quad (5)$$

$$g(w_k) = d(i, j) = 1 - \frac{V_{mi} \cdot V_{qj}}{|V_{mi}| |V_{qj}|} \quad (6)$$

ここで N, M はそれぞれ V_m, V_q のベクトル数であり、 w は warping path, K は warping path の総数である。 $d(i, j)$ は DTW で対応づけた各データの距離であり、2 つのベクトルの角度を距離指標として用いる。 V_q と V_m, V_q と各 $V_{m_{part}}$ について D_V を求め、最小のものを最終的な D_V として採用する。

2.2 適合性フィードバックシステムの概要

この検索システムに「適合性フィードバック」を適用することで、システムで抽出した特徴量とユーザが描いたスケッチの特徴量とのセマンティックギャップを埋め検索精度を向上させる。適合性フィードバックとは、ユーザが行った検索結果に「良い」「悪い」などの評価を付与することで、検索結果のブラッシュアップを行なっていく仕組みのことを指す。本手法では、スケッチ検索を行った後に、データベースの中の所望動画に入力したスケッチを関連付けて学習させる。複数の動画に対して既にスケッチの関連付けが行われていると仮定し、この複数枚の関連付けられたスケッチ (以降、スケッチ検索履歴) を用いて適合性フィードバックを行う。

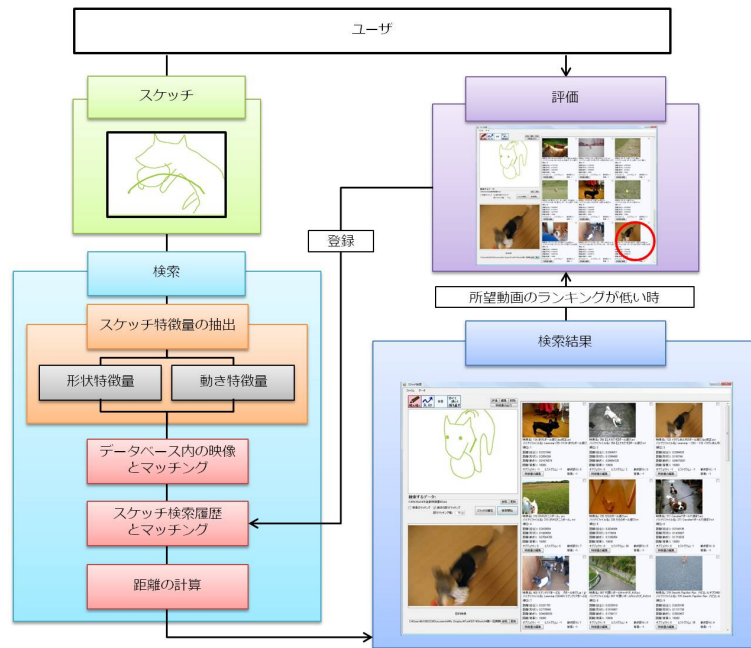


図 4 適合性フィードバックシステムの流れ

2.2.1 検索の流れ

適合性フィードバックを用いた映像検索システムの流れを図 4 に示す。

- (1) ユーザはスケッチを入力する。
- (2) システムはスケッチの特徴量を抽出する。
- (3) データベース内の動画があらかじめ抽出しておいた特徴量とスケッチの特徴量を比較し、距離を計算する。
- (4) 距離に応じてランキングし検索結果を出力する。
- (5) 所望の動画の検索結果が思わしくない時は、検索結果の中から所望の動画にチェックを入れ、所望動画を選択する。
- (6) システムは、ユーザの関連付けに基づき、(2) で抽出したスケッチの特徴量を新たに保存し、所望動画に関連付ける。
- (7) ユーザは再びスケッチを入力する。

- (8) システムはスケッチの特徴量を抽出する。
- (9) スケッチ特徴量とスケッチ検索履歴の各スケッチの特徴量とを比較し最も距離が近いスケッチを取得する。
- (10) 取得したスケッチと関連付けられている動画を取得し、この動画の特徴量をクエリとして検索を行なう。
- (11) 距離に応じてランキングし検索結果を出力する。

以上の処理により、適合性フィードバックを適用した映像検索システムを実現させた。

2.2.2 関連付けたスケッチとのマッチング方法

入力したスケッチと動画に関連付けられたスケッチとのマッチング方法について述べる。まず、入力されたスケッチの特徴量を「形状特徴量」と「動き特徴量」に分割する。次に、入力されたスケッチとスケッチ検索履歴群とで比較を行い、形状特徴量の距離が最も短くなるスケッチ履歴と、動き特徴量の距離が最も短くなるスケッチ履歴とをそれぞれ取得する。取得したそれぞれのスケッチ履歴は、データベース内の、ある動画を検索するために描いたものなので、取得したスケッチはその動画と既に関連付けがなされている。つまり、入力されたスケッチは、その関連付けられた動画のあるシーンと最も類似していると言い換えることができる。よって、入力されたスケッチは、その動画と最も類似しているといえる。この処理を、形状特徴量と動き特徴量双方に行うことで、2本の動画を取得することができる(図 5)。この形状特徴量用の動画特徴量と、動き特徴量用の動画特徴量を用いてデータベース内の動画群とマッチングを行う。

3. 実験と考察

本手法の有効性を検証するため、YouTube に投稿されている動画を取得し、それらに対して従来手法と本手法とで検索を行いその違いを検証した。

3.1 使用動画の取得

YouTube に投稿されている動画を取得し、実験動画として用いる。動画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、該当動画 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動画 407 件のうち、再生時間が 30 秒以下の動画 34 個を用いた。YouTube でのキーワード検索結果をまとめたものが表 1 である。

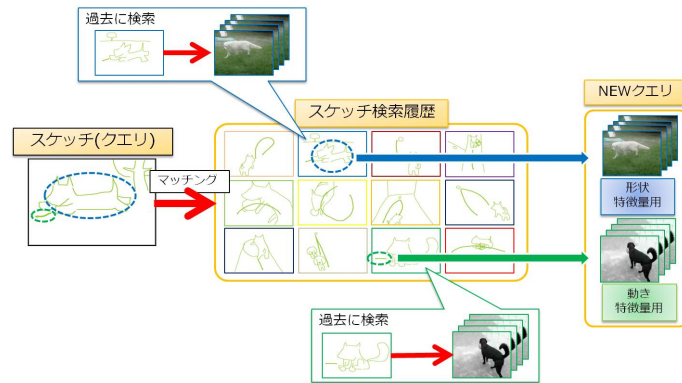


図 5 スケッチ特徴量から映像特徴量への変換

検索日	2010/06/13
設定ワード	「犬」「ボール」「遊ぶ」
表示順	関連度順
該当動画数	1160
取得動画数	407(上位)

3.2 実験の流れ

「ある動画を視聴して記憶し、その映像をスケッチを用いてクエリ化し検索する」という形式で実験を行った。動画は、検索を行なう前に一度だけ見てもらうものとし、再度の動画視聴は不可とした。検索手順を以下に示す。

- (1) PC 上で提示された目的動画を視聴し、記憶する。
- (2) 記憶した後、すぐにスケッチを描く（再度の動画視聴は不可）
- (3) 描いたスケッチで検索を行う。

次に、実験手順について述べる。二種類の動画にスケッチを関連付け、残り一種類の動画を検索する。この動画の検索順位を比較することで本手法の有用性を確認する。

- (1) 二種類の動画に対して事前に検索を行い、入力スケッチを関連付ける。(学習フェーズ)
- (2) 二種類の動画の中間動画について検索を行う。(検索フェーズ)
- (3) この中間動画の検索順位を従来手法と本手法とで比較する。

この手順で実験を行い本手法の有用性を確認した。

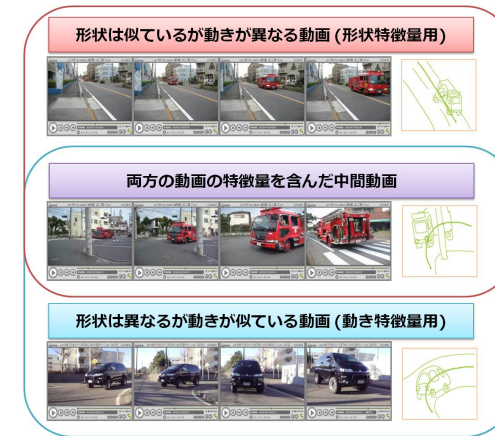


図 6 実験動画について

3.3 実験動画について

実験には三種類の動画を用いる(図 6)。形状特徴量用にした動画、動き特徴量用にした動画、そしてその両方の特徴量を有した中間動画の三種類である。この 3 つの動画のうち、形状特徴量用の動画と動き特徴量用の動画に対して学習フェーズでスケッチを関連付ける。そして、この両方の特徴量を含んだ中間動画を検索対象として検索を行なう。中間動画には関連付けたスケッチは存在しないが、他の二種類の動画には関連付けたスケッチが存在する。この状態で検索を行い、適合性フィードバックを適用する前と後とで変化を確認することで本手法の有用性を確認する。

3.4 実験結果

	従来手法	本手法	変化率
No.286 の動画(図 8)	16/34 位	4/34 位	12 位

3.5 考察

実験に用いた三種類の動画のうち二種類の動画にスケッチを関連付け、その二種類の動画の中間動画を検索した結果、従来手法では 16 位だった動画が 4 位にまで上昇した。適合性

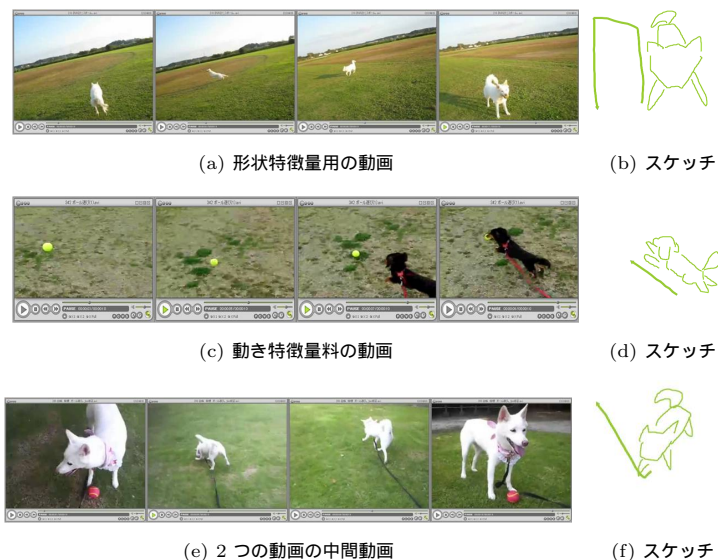


図 7 使用動画と関連付けたスケッチ



図 8 実験結果

フィードバックを適用することで、従来手法では検索が難しい動画にアプローチすることができ、結果として検索精度を向上させることができた。また、形状を抽象的に描いたとしても、既にスケッチ検索履歴が存在する場合、最も距離の近い動画を見つけ出すことができた。

4. おわりに

本研究では、手書きスケッチを用いた映像検索システムに適合性フィードバックシステムを適用した。YouTube から取得した動画に対し提案システムを適用し、実際に動画検索を行なった。その結果、従来手法では検索が難しかった動画に対してもアプローチが可能になり、また、実験結果からも有用性を示すことができた。

今後は、SIFT の様な特徴点抽出手法を用いることで推定精度の向上を図ると共に、背景などの特徴量を増やすことや、各特徴量の重みを変えることで検索精度の向上を図る。

参考文献

- 1) YouTube: <http://www.youtube.com/>.
- 2) Nico Nico Video: <http://www.nicovideo.jp/>.
- 3) 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.MVE, Vol.110, No.238, pp.109-114 (2010).
- 4) 林貴宏, 石川厚, 尾内理紀夫: 「スケッチ線と意味アイコンを用いた風景画像検索」, 信学技報.PRMU, Vol.104, No.573, pp.19-23 (2005).
- 5) 望月貴裕, 藤井真人, 伊藤崇之: 「問い合わせ画像描画による柔軟な画像検索」, 信学技報.IE, Vol.101, No.627, pp.77-82 (2002).
- 6) 多々良友英, 大橋剛介: 「大局的および局所の特徴量を用いたスケッチ画像検索」, 映像情報メディア学会誌, Vol.62, No.12, pp.2059-2062 (2008).
- 7) 大橋剛介, 久森隆史, 望月圭太: 「適合性フィードバックを用いたスケッチ画像検索システム」, 知能と情報: 日本知能情報ファジィ学会誌, Vol.19, No.5, pp.537-545 (2007).
- 8) 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).
- 9) 矢島史, 角谷和俊, 田中克己: 「映像上での動きの直接描画によるサッカー映像検索」, 情報処理学会研究報告, データベース・システム研究会報告, Vol.2002, No.41, pp.33-40 (2002).
- 10) D.J.Berndt, J.Clifford: "Finding Patterns in Time Series:A Dynamic Programming Approach," Advances in Knowledge Discovery and Data Mining, pp.229-248 (1996).