

時空間画像処理による類似動画画像検索

松生 智博[†] 石井 亮司^{††} 前田真一郎^{†††} 岡田 至弘^{†††}

[†] 龍谷大学大学院理工学研究科情報メディア学専攻

^{††} 龍谷大学大学院理工学研究科電子情報学専攻

^{†††} 龍谷大学理工学部情報メディア学科

滋賀県大津市瀬田大江町横谷 1-5

E-mail: {t_matsuike, r_ishii, s_maeda}@afc.ryukoku.ac.jp
okada@rins.ryukoku.ac.jp

あらまし 本報告では、動き特徴に基づく類似動画画像検索手法を提案する。動き特徴として動きの方向と速度を扱う。色特徴によるイベント検出が困難なスポーツ映像を対象に、動画中の動き特徴を時空間画像処理を用いて抽出し、類似動画画像検索に利用できることを示す。時空間画像処理を動き解析に利用することにより、動画中で微小な動きを省き、大きな動きの変化を取り出すことができる。さらに、この大きな動きの変化を考慮した動画画像の内容に基づく検索についても報告する。

Spatio-Temporal Image Processing for Retrieving Similar Video Sequences

Tomohiro MATSUIKE[†], Ryoji ISHII^{††}, Shinichiro MAEDA^{†††}, and Yoshihiro OKADA^{†††}

[†] Division of Media Infomatics, Graduate School of Science and Technology, Ryukoku University

^{††} Division of Electronics and Infomatics, Graduate School of Science and Technology, Ryukoku University

^{†††} Department of Media Infomatics, Faculty of Science and Technology, Ryukoku University

1-5 Yokotani, Seta Oe-cho, Otsu, Shiga, 520-2194 Japan

E-mail: {t_matsuike, r_ishii, s_maeda}@afc.ryukoku.ac.jp
okada@rins.ryukoku.ac.jp

Abstract In this paper, we propose a similar video sequence retrieval method based on spatio-temporal motion features: direction and velocity. First, we show these motion features can be extracted from sports movies by using spatio-temporal image processing. The spatio-temporal image processing enables us to eliminate small movements and extract motion features derived from major change. Then, we show that, by using these features, similar video sequences can be retrieved from sports movies.

1. はじめに

近年、動画画像が大量に蓄積されるようになり、動画画像を対象にした効率的な検索や提示手法が求められている。本報告ではスポーツ映像のような動きの変化が重要となる動画画像を対象とし、利用者が視聴している動画画像と類似する動画画像を検索する方法について報告する。

動画画像の検索手法に色特徴を利用したものがある。これらの手法では、人物や背景といった動画画像に含まれる物体の色特徴を抽出し、その特徴と類似する動画画像を検索する。しかし、こうした検索手法は、背景が画像フレームの大半を占めることが多いサッカーや野球などのスポーツ映像にはあまり有効ではない。これは、カメラがズームアウトすると色特徴の変化が少なくなり、イベ

ントを検出することが困難となるからである。色特徴の変化が少ない動画像からイベントの発生を検出するために、動画像の動き特徴を用いることを考えた。

動きに基づく検索手法として、動画像に付加された文字列による検索がある [1]。しかしこの検索手法には、視聴中に文字列を入力することは利用者にとって大きな負担であること、そして利用者自身が所望する動画像を文字列で正確に表現することがむずかしいという問題がある。そこで本報告では、利用者に視聴中の動画像に対して検索区間の開始点と終了点を指定させ、これに含まれる動き特徴に類似した動画像の検索を行う。

動き特徴によって動画像を検索するとき、利用者が見ているのは、フレーム単位で変化するような微小な動きではなく、動画像の場面変化に関連する大きな動きの変化であると考えられる。動画像の場面変化に関連する大きな動きの例としては、スポーツ映像での攻守の切り替わりやカメラの動きの変化といったものがある。こうした大きな動きの特徴を時空間画像処理によって抽出する。時空間画像処理はある固定長のフレーム列から動き特徴をみることができ、時系列に対して動きの平滑化がされる。その結果、微小な動きによる影響を低減する処理が可能となる。

本報告では、動きの特徴として動きの方向とその速度を扱い、これらを組み合わせて動画像の内容に基づく検索を実現する。例えば、スポーツ映像における攻守の切り替わりでは、画像フレームに対して左方向への動きから右方向への動きの切り替わり、もしくはその逆が起こる。このような動きの方向変化は動画像の内容を表わす特徴であり、動画像の内容に基づく検索の手がかりになると考えられる。また、方向に加えて動きの速度についても動画像の内容を表わす特徴であり、動画像の内容に基づく検索の手がかりになると考えられる。利用者が指定した検索区間に含まれる一連の動きの変化を手がかりとし、1つの動き特徴ごとの検索ではなく、一連の動き特徴を手がかりとして検索を行う。

以下、2章では動画像検索に関する研究について述べ、3章から4章で時空間画像からの動き特徴抽出と類似動画像の検索手法について具体的に説明する。そして5章で本手法の実験とその結果について述べ、6章でまとめとする。

2. 動きに基づく動画像検索

動きは文字列による表現が困難であることから、利用者が描画した動きの記号や軌跡による動画像検索が研究されている [2], [3]。利用者が動きを図として直接描くことが可能なため、文字列による動きの表現よりも利用者は比較的表現しやすい。ただし、利用者が動画像を視聴

中に図を描画する必要があること、描画するためのツールを手にしながらの視聴が必要となることから、今回のように、動画像の視聴中に検索を行うという状況には適さない。

これまでに、動画像中に含まれる動き特徴を用いた検索はされている [4], [5]。動画像のある区間での動き特徴を抽出し、動画像の分類、検索に利用している。しかし、考慮されているのは区間内での動き特徴であり、他の区間との関係については触れられていない。このため、複数の動き変化によって成り立つ動画像の内容表現を扱うことは困難である。

筆者らは文献 [6] で時空間画像処理を用いた動きの解析手法を提案してきた。動画像中に含まれる特徴的な動きとしてカメラワークに着目し、動画像の分類やシーン解析を行った。しかし、動き解析はフレーム単位での処理ではなく、固定長のフレーム列単位に処理していたため、今回のようなフレーム単位で動き特徴の結果を要求されるような場合には適さない。

本報告では動画像中での動き特徴解析で時空間画像処理を行いつつ、動画像のフレーム入力ごとに動き特徴抽出を行う手法について提案する。また、動きの変化を考慮し、内容に基づいた類似動画像検索を行う。

3. 時空間画像からの動き特徴解析

本章では動画像の時空間画像から動き特徴を抽出する手法について説明する。本報告で扱う動き特徴は以下の2つの成分から成るものとする。

- 動き方向
- 動き速度

これらは動画像中で最も影響が大きい動きの移動方向とその速度を示すものであり、時空間画像上に現われる動き特徴からこれらを抽出する。

3.1 時空間画像処理

動画像を画像フレームの x 軸と y 軸、時間方向の t 軸からなる直方体とすると、この t 軸に対して平行な断面図が時空間画像である。動画像中に含まれる動き特徴や動画像を編集した際の特徴が2次元画像上に現れるため、これを利用した動き解析の研究がされている [4], [5], [6], [7], [8]。 y 軸の値を固定することで水平方向の動き、 x 軸の値を固定すると垂直方向の動きにそれぞれ対応する。 y 軸の値を y' で固定し、 y' 上の画素を時間軸方向に並べていくと、図1のような $x-t$ 時空間画像が得られる。この時空間画像には、 y' における水平方向の動きが現れる。同様に、 x 軸の値を x' で固定すると、 x' 上での垂直方向の動きが $y-t$ 時空間画像から得られる。時空間画像処理は F フレームのフレーム列ごとに処理を行う。

$x-t$ 時空間画像処理と $y-t$ 時空間画像処理の違いについ

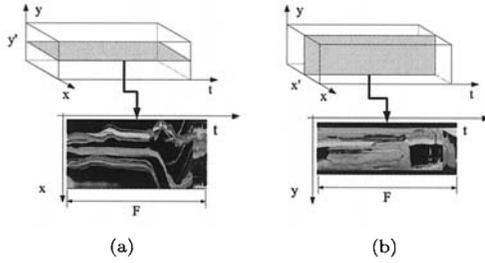


図1 時空間画像の生成.

- (a) y 軸を y' で固定した際に現われる $x-t$ 時空間画像.
- (b) x 軸を x' で固定した際に現われる $y-t$ 時空間画像.



図2 $x-t$ 時空間画像から得られるエッジ画像.

- (a) 物体の動きが含まれていないとき.
- (b) 物体の動きが含まれているとき.

ては固定する軸が異なるのみである。以降、 $y-t$ 時空間画像処理については説明を省略する。

3.2 動き速度の抽出

ここでは動画の時空間画像から動き速度を抽出する手法について説明する。まずはじめに、時空間画像のエッジ抽出を行い、時空間画像から動きの特徴が現われている部分を取り出す。エッジ抽出には Canny オペレータを適用し、図2のようなエッジ画像を得る。

このとき、時空間画像は時間軸を有する2次元画像であるため、時空間画像からエッジ抽出した際の各エッジ点は時間軸に対して正の方向へ移動する。物体が静止しているとき、その物体に対するエッジ点は $\phi = 0$ の方向へ移動し、動きがある場合にはそれ以外の方向へ移動へ移動する。 ϕ の正負符号は動いている方向を表わし、 $|\phi|$ は速度を表わす。また、 ϕ の範囲は $-D_\phi \leq \phi \leq D_\phi$ とする。

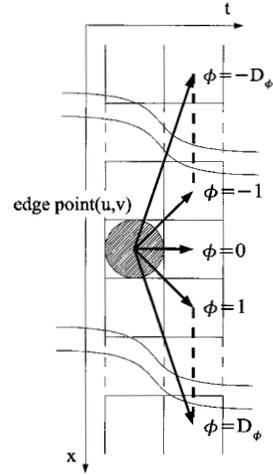


図3 エッジ点 (u, v) の移動方向.

次に、エッジ点の各移動方向の頻度を計算する。時空間画像から得られる、エッジ点の各移動方向の頻度を表わしたものを動きヒストグラムとする。エッジ点 (u, v) が ϕ 方向へ移動しているならば $E(u, v, \phi) = 1$ 、そうでなければ $E(u, v, \phi) = 0$ とすると、時空間画像から得られるエッジ点の移動方向 ϕ の動きヒストグラム M は式1で表わすことができる。

$$M(\phi) = \sum_{v=0}^{F-2} \sum_{u=0}^{W-1} E(u, v, \phi) \quad (1)$$

W は動画のフレーム幅を表わし、 $y-t$ 時空間画像について動きヒストグラムを求める場合には、動画のフレームの高さである H となる。動画の動き速度 σ は動画の中で最も影響を与えている動きとし、動きヒストグラムの最も頻度が大いときの ϕ を σ とする。また、 σ の正負符号は動きの方向であり、これを δ として表わす。

3.3 時空間画像の更新処理

本報告では、利用者が視聴している動画の動き変化に合わせて動画の動き速度をフレーム単位で抽出する。時空間画像処理は F フレームのフレーム列ごとに処理を行うため、新たなフレームを取得するたびに、図4に示すように新たなフレームより F フレーム過去のフレームを取り除いていく。これにより、常に F フレーム分の時空間画像処理を行い、フレームごとに動き速度を抽出する。

ある方向への動きがはじまってからの動きの継続時間を τ とし、この間の平均動き速度 σ_τ を式2により求める。

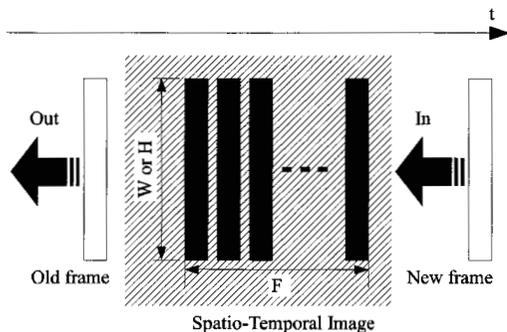


図4 時空間画像の毎フレーム更新。
 新たなフレームを読み込んだときに最も古いフレームを取り除き、 F フレームの時空間画像とする。 W と H はそれぞれ動画のフレーム幅と高さを表す。

$$\bar{\sigma}_\tau = \sum_{i=1}^{\tau} \sigma_i \quad (2)$$

4. 動き特徴パターンの比較

はじめに前処理として、あらかじめ保存されている各動画 $V_d (d = \{0, 1, 2, \dots\})$ について動き特徴解析を行い、動画を動き特徴によって細分化する。この細分化された動画を部分動画 $V_{d_m} (m = \{0, 1, 2, \dots\})$ とする。部分動画に対してそれぞれ動き特徴を付与して蓄積しておく。 V_{d_m} に対する平均動き速度を $\bar{\sigma}_{d_m}$ 、動き方向を δ_{d_m} とする。

利用者により検索が開始されると、利用者が視聴している動画 V_k の解析処理をはじめ、 V_k のフレームごとに動き特徴解析を行い、平均動き速度 $\bar{\sigma}_{k_n} (n = \{0, 1, 2, \dots\})$ と動き方向 δ_{k_n} を得る。検索が開始されてから V_k の動き方向が変化するまではじめの動き方向を δ_{k_0} とし、 V_k の動き方向が変化するごとに $\delta_{k_1}, \delta_{k_2}, \delta_{k_3}, \dots$ とする。この動き方向変化の一連のまとまりをここでは動き特徴のパターン、もしくは単にパターンと呼ぶ。以上の動き特徴を用い、利用者が指定した検索区間の動画と蓄積されている各部分動画との比較を行う。ここで、 V_k に対する類似動画の条件として、 V_k の動き特徴のパターンと一致していることとする。スポーツ映像において動き方向は選手が移動する方向や球技でのボールのゆくえに関わるものであり、動画の内容と深く関係すると考えられるためである。

検索開始後、 V_k の動き方向が変化するまでは動き方向が一致する部分動画の中から速度の差が最小となるものを類似動画として出力する。

利用者が検索中に V_k の動き方向が変化した場合、これまでの動き方向と動きの速度を含めて考慮した検索を行う。

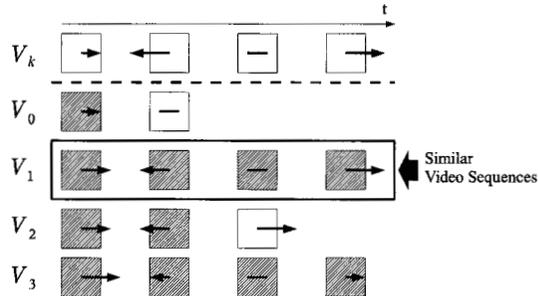


図5 V_k と V_d の動き特徴パターンの比較。
 蓄積されている動画数 d を 4、検索中に動き方向が変化した回数 p を 3 としたときの例。矢印の向きと長さがそれぞれ動きの方向と速度を表す。

これらをまとめると、利用者の検索に対する結果は以下を満たすものとなる。 p は V_k の動き方向が変化した回数 (動き特徴のパターン数) を表す。 $\delta_{k_0} = \delta_{d_m}, \delta_{k_1} = \delta_{d_{m+1}}, \dots, \delta_{k_p} = \delta_{d_{m+p}}$ かつ $\sum_{i=0}^p |\bar{\sigma}_{k_i} - \bar{\sigma}_{d_{m+i}}|$ が最小値をとる V_d の部分動画 V_{d_m} から $V_{d_{m+p}}$ までを類似動画とする。

パターン比較処理の例を図5に示す。 V_0, V_2 は動き特徴のパターンが V_k と一致しなくなった時点で結果対象から外れている。 V_1, V_3 はどちらも V_k と動き特徴のパターンは一致しているが、動きの速度を考慮すると V_1 の方が V_k に近い。このため、 V_1 が V_k に最も類似する動画となっている。

5. 類似動画検索実験

本章では3章、および4章にて提案した手法の実験について述べ、その結果について考察する。

実験対象の動画として、以下の16種類を用意した。

- サッカー映像 9種類
- 野球映像 7種類

このうちサッカー映像5種類、野球映像3種類は利用者が視聴していると想定するクエリ動画、サッカー映像4種類、野球映像4種類は蓄積動画となっている。これらは全て画像フレームサイズ 320×240 pixel、フレームレート 29.97 fps の動画で、クエリ動画は10秒、蓄積動画は10分の動画である。時空間画像処理を適用する際のフレーム数 F を30、エッジ点の移動方向 ϕ の範囲である D_ϕ を画像フレームの横幅と同じ320とした。動きの速度 σ は ϕ と同じ範囲をとるため、 σ の範囲は $[-320, 320]$ となる。

クエリ動画の特徴として、動き特徴のパターン数を表1に示す。Video K1 から Video K5 のサッカー映像は野球映像と比較してパターン数が多い結果が得られた。サッカー映像は野球映像よりも選手やボールがフィールド

表 1 クエリ動画像とそれらに含まれている動き特徴のパターン数.

Video K1 から Video K5 はサッカー映像. Video K6 から Video K8 は野球映像.

動画像のラベル	パターン数
Video K1	6
Video K2	6
Video K3	11
Video K4	9
Video K5	13
Video K6	6
Video K7	3
Video K8	2

ド上で動くことが多い. カメラもそれらに追従して撮影されるため, 動画像内でカメラが頻繁に動く. これらの動きが表 1 のパターン数に影響したと考えられる.

次に, 類似動画像の検索実験について述べる. Video K1 から Video K8 の動き特徴と, これらに対して動き特徴が最も類似している動画像の動き特徴を図 6(a) から図 6(h) に示す. Video K1 と Video K2 はどちらも左右にボールが飛び交う場面である. 図 6(a), 図 6(b) にも左右への動きが現われている. Video K1, Video K2 ともに, 得られた類似動画像はクエリ動画像と同様の場面となっており, 正しい結果が得られたといえる. Video K3 と Video K4 からは選手がアップで映し出される動画像が結果として得られた. しかし, Video K3, Video K4 は選手がアップで映っている動画像ではなく, フィールドを撮影されたものである. 図 6(c), 図 6(d) からわかるように, パターンは一致しているが, 1 つ 1 つの動き方向が切り替わるまでの時間に差がある. Video K3 や Video K4 の動き方向はカメラの動き特徴を表しているのに対し, それぞれの結果の動画像は選手の手や体の動き特徴を表している. このような問題に関しては, 動いている間の時間を考慮することで区別できると考えられる. Video K7 では, 走っている選手を撮影している動画像に対し, 検索結果はテロップがスライドインする部分を含む動画像であった. テロップの動きが選手を追いかけて撮影する動きと一致してしまったため, このような結果となった. Video K7, Video K8 の野球映像は表 1 にも示したとおり, パターン数が少ない動画像である. そのため同一の動き特徴パターンが現われることが多く, パターンによる類似動画像の絞込み効果が小さかった.

6. まとめ

本報告では, スポーツ映像を対象とした動き特徴に基づく類似動画像検索について検討した. 時空間画像処理により抽出した動きの方向や速度といった特徴の組み合

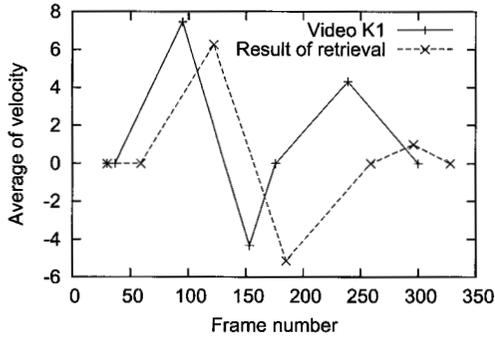
わせを利用し, 動画像の内容に沿った検索を提案した.

実験の結果, サッカー映像での攻守の切り替わりを示すような左右への動きを含む動画像を検索することができた. しかし, 人の目から見て明らかに類似していない動画像が検索結果として出てくることもあり, まだまだ改善の余地はある. 類似していない動画像が検索結果として出てきた主な原因として, 動画像から抽出した動き特徴がカメラの動きから得られたものか, 人物の動きから得られたものかの区別ができていなかった点が挙げられる. これについては, カメラと人物の動きには動きの方向が変化するまでの時間間隔に違いがあるため, 時間の長さを指標に加えることで, これらを区別することができると思われる.

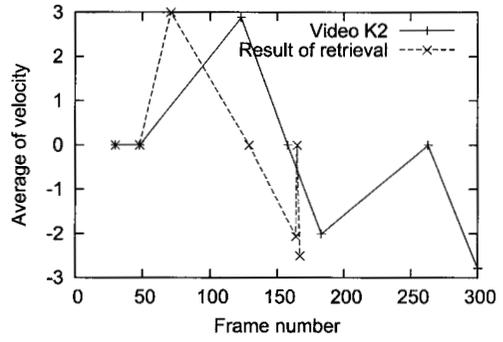
今回の実験はサッカー映像と野球映像を対象として行った. 今後, 様々なスポーツ映像に対して実験を行い, 本手法の有効性を検証していく.

文 献

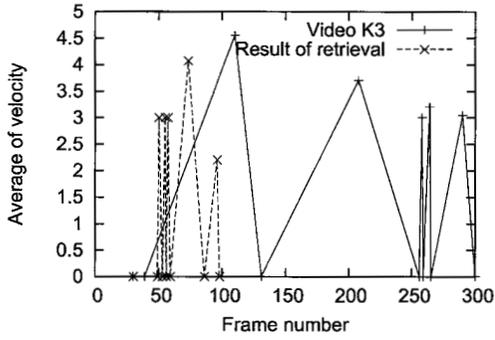
- [1] 宮森恒, 粕谷英司, 富永英義: “動作語を用いた問合せによる映像検索方式”, 電子情報通信学会論文誌. D-II, 情報・システム, II-情報処理, **80**, 6, pp. 1590-1599 (1997).
- [2] 青木秀一, 青木輝勝, 安田浩: “動画像からのシーン検索のための略画処理手法の提案”, 情報処理学会研究報告. CVIM, [コンピュータビジョンとイメージメディア], **2002**, 2, pp. 133-140 (2002).
- [3] 矢島史, 角谷和俊, 田中克己: “映像上での動きの直接描画によるサッカー映像検索”, 情報処理学会研究報告. データベース・システム研究会報告, **2002**, 41, pp. 33-40 (2002).
- [4] C. Ngo, T. Pong and H. Zhang: “On clustering and retrieval of video shots through temporal slices analysis”, *Multimedia, IEEE Transactions on*, **4**, 4, pp. 446-458 (2002).
- [5] C. Ngo, T. Pong and H. Zhang: “On clustering and retrieval of video shots”, *ACM Multimedia*, pp. 51-60 (2001).
- [6] 松生智博, 石井亮司, 岡田至弘: “カメラワークを考慮したダイアログ映像解析”, 電子情報通信学会技術研究報告. IE, 画像工学, **107**, 40, pp. 35-40 (2007).
- [7] 幸裕, 直子, 登: “適合性フィードバックを用いた映像パケットのインタラクティブクラスタリング”, 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理, **88**, 8, pp. 1702-1711 (2005).
- [8] 淳夫, 亮治, 宗: “カメラワークを利用した感性情報の抽出”, 情報処理学会論文誌, **47**, 6, pp. 1696-1707 (2006).



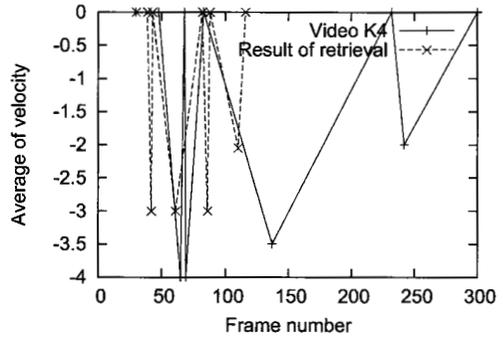
(a)



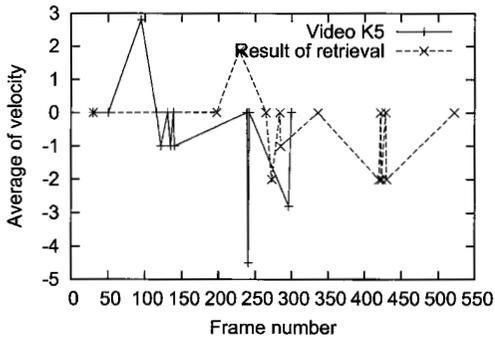
(b)



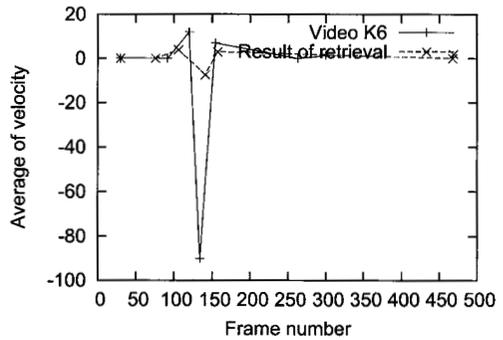
(c)



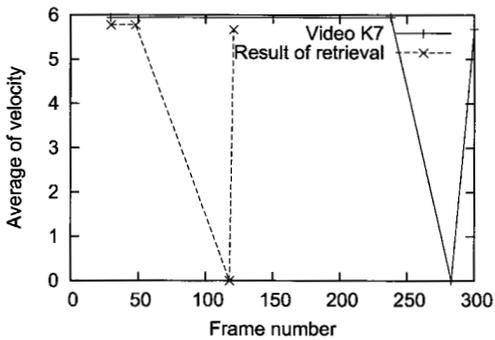
(d)



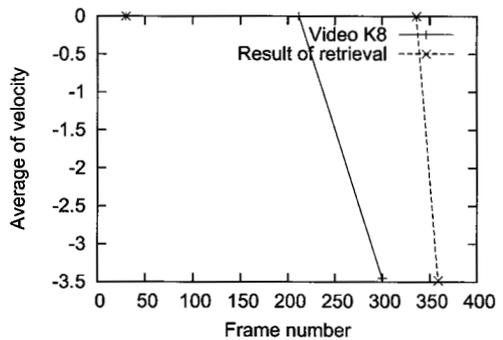
(e)



(f)



(g)



(h)

図 6 類似動画検索の結果.