

特徴空間のマージン付き適応分割による 高速かつ検出漏れのない準同一映像区間検出

高橋 友和[†] 鈴木 秋吾^{††} 井手 一郎^{††,†††} 村瀬 洋^{††}

[†] 岐阜聖徳学園大学経済情報学部 〒500-8288 岐阜県岐阜市中鶯 1-38

^{††} 名古屋大学大学院情報科学研究科

^{†††} 国立情報学研究所

E-mail: †ttakahashi@gifu.shotoku.ac.jp

あらまし 長時間映像中に現れる準同一映像区間 (Near-duplicate video segment) を高速かつ漏れなく検出する手法を提案する。記録される映像量の爆発的な増加とともに、利用者が要求する映像区間を高速に探索し、閲覧性を向上させる技術が求められている。その中に準同一映像区間を検出する技術と、その結果を用いて映像間の関連性を抽出する技術に関する研究がある。準同一映像区間とは、映像の中に複数回出現する、同一もしくは類似した映像区間を指す。映像中の準同一映像区間をすべて検出するには、映像区間の照合を総当りで行うため、映像長の 2 乗に比例した回数の照合が必要になる。それには膨大な計算を要するため、従来は時間的なコストの面から実用化が困難であった。提案手法は、特徴空間を特徴ベクトルの分布に対して適応的に分割して照合回数の削減を図ることにより、高速な検出を実現する。さらに、マージン空間と呼ばれる空間を分割境界付近に設けることによって、分割の際に生じる分割境界付近の検出漏れを防ぎ、漏れのない検出を実現する。24 時間分の一般放送映像を入力映像として用いた実験の結果、提案手法を用いた場合には、総当りの場合と比較して約 500 倍高速に、かつ漏れなく準同一映像区間が検出できることがわかった。

キーワード 準同一映像区間検出, 空間分割, 高速化, 精度保証

Fast and Accurate Near-Duplicate Video Segments Detection by Adaptive Feature Space Division with Margin

Tomokazu TAKAHASHI[†], Shugo SUZUKI^{††}, Ichiro IDE^{††,†††}, and Hiroshi MURASE^{††}

[†] Faculty of Economics and Information, Gifu Shotoku Gakuen University

Nakauzura 1-38, Gifu-shi, Gifu, 500-8288 Japan

^{††} Graduate School of Information Science, Nagoya University, Japan

^{†††} National Institute of Informatics, Japan

E-mail: †ttakahashi@gifu.shotoku.ac.jp

Abstract We propose a method for fast near-duplicate video segments detection with guaranteed accuracy. With an explosive increase of recorded videos, technology that searches for video fast based on users' requests is needed. Among these technologies, there is the technology for detecting near-duplicate video segments, and works applying it to video retrieval. Near-duplicate video segments are identical or extremely similar video segments that appear two or more times in a video stream. Because a huge amount of calculation is needed for detecting all near-duplicate video segments, practical use was difficult. The proposed method reduces the number of comparison by adaptively dividing the feature space according to the distribution of feature points. Moreover, to guarantee no false positives, the proposed method extends the divided spaces at the boundary by adding a margin space. Results of an experiment using 24 hours of an actual broadcast video demonstrates that applying the proposed method reduces the computational time by 1/500 compared to that without applying the proposed method.

Key words Near-duplicate video segment detection, space-division, speed up, accuracy guaranty

1. はじめに

近年、膨大な量の映像を蓄積することが可能となり、映像の探索の効率化および閲覧性の向上が必要となっている。そのための技術として、特定の映像区間をキーとして、それと同一の映像区間を大量に蓄積された映像から高速に探索する技術が多く研究されている。これらの中には、最近傍探索に基づく手法がある。Aryaら[1]は、特徴空間分割を用いて探索領域を絞り込むことで、キーの近傍点の候補となる点との距離計算回数を削減するANN(Approximate Nearest Neighbor)法を提案している。また荒井ら[2]は主成分木と呼ばれる、主成分分析を用いた特徴空間分割を用いて更に高速な検出を実現している。山岸ら[3]は、映像中の画像を特徴空間内で事前にクラスタリングしておくことにより、キーとして与えた映像中の画像と類似した画像を高速に探索する手法を提案している。

一方、別の問題として、特定の映像区間をキーとして与えるのではなく、与えられた映像中に繰り返し現れる同一あるいは類似した映像区間を検出する技術に関する研究が行われている。また、それを映像の意味構造解析に役立てる研究が行われている。このような、映像中に繰り返し現れる同一あるいは類似した映像区間のことを、本稿では準同一映像区間(Near-duplicate video segment)と呼ぶ。例えば、社本ら[4]は、長期にわたって蓄積されたニュース映像中から、準同一映像区間の存在を手がかりとして関連する一連のニューストピックを対応付け、映像の意味構造を抽出する手法を提案している。また、小川ら[5]は準同一映像区間の共起を手がかりに、異なる言語のニュース映像の中から、同一ニューストピックについて言及している映像を自動で検出する言語横断検索手法を提案している。

準同一映像区間を漏れなく検出するには、入力映像中の全映像区間数を N としたとき、すべての区間の組み合わせである $\binom{N}{2}$ 回の膨大な数の総当たりによる照合を要する。これに対し、野田ら[6]は、特徴次元圧縮により、映像区間の照合に要する計算を短縮することで高速化を実現している。文献[4],[5]では、この手法を用いて準同一映像区間の検出を行っている。また、この手法に基づいて井手ら[7]は、映像区間の特徴として、番組ジャンルに適したものをを用いることにより、更なる検出の高速化を図っている。しかしながら、これらの手法では、照合回数は $\binom{N}{2}$ のままである。これに対して、Naturelら[8]は、ハッシュテーブルを用いて、同一のハッシュ値を持つ映像区間のみを照合することで、照合回数の削減を図っている。しかしながら、この手法は近似解法であり、検出結果には漏れが存在する可能性がある。これに対して、西村ら[9]は、村瀬らが提案したアクティブ探索法[10]を準同一映像区間検出に適用することにより、検出漏れがないことを保証しながら照合回数の削減を図

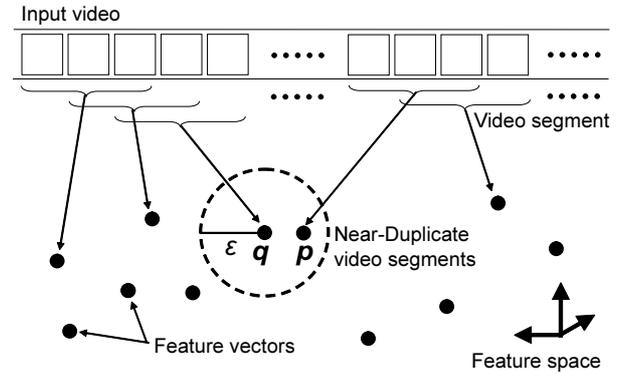


図1 準同一映像区間対

ている。この手法に基づいて大盛ら[11]は、柏野ら[12]が提案したオーバースキッピングの考え方を応用することにより、検出漏れがないことを保証したまま、更なる照合回数の削減を図っている。しかしながら、これらの手法は映像区間の類似尺度にヒストグラムインターセクションを用いることを前提としている。このため、利用可能な映像区間の特徴は、時間的かつ空間的にグローバルなヒストグラム特徴に制限される。これは、例えば時間順序の異なるシーンの区別はできないことを意味する。

これに対して本研究では、準同一映像区間検出に用いる映像区間の類似尺度として、特徴ベクトル間の距離を用いる。これにより、映像区間の汎用な特徴記述が可能となる。そして、特徴空間を特徴ベクトルの分布に対して適応的に分割して照合回数の削減を図ることにより、高速な検出を実現する。さらに、マージン空間と呼ばれる空間を分割境界付近に設けることによって、分割の際に生じる分割境界付近の検出漏れを防ぎ、漏れのない検出を実現する。

以下、2. では、本稿で対象とする準同一映像区間検出の問題を定義するとともに、実際に検出された準同一映像区間の例を示す。3. では、提案手法である特徴空間のマージン付き適応分割について詳細に述べる。4. では、提案手法の評価を目的とし、実際の一般放送映像を対象として行った準同一映像区間検出実験について述べ、その結果を考察する。5. で本稿をまとめる。

2. 準同一映像区間検出

2.1 問題定義

本稿で扱う問題は、入力映像中から以下に定義する準同一映像区間対を漏れなく、かつ誤りなく検出するものである。

図1に示すように、入力映像中の任意の位置、同じ長さの2つの映像区間を表す特徴ベクトルをそれぞれ p, q とし、以下の条件を満たす映像区間対 p, q を「準同一映像区間」対と定義する。

$$\|p - q\| \leq \epsilon \tag{1}$$



(a) Commercial messages



(b) Opening shots of a same TV program



(c) News shots edited from a same source video

図 2 準同一映像区間対の例

ただし, ϵ は準同一映像区間検出のしきい値である.

この問題に対し, 総当りによる手法では, N 個の特徴ベクトルの中から全てのベクトル間の距離を計算するために $\binom{N}{2}$ 回の膨大な数の照合を要する.

2.2 準同一映像区間対の例

放送映像において, 準同一映像区間は映像の意味構造解析を分析する際に重要な意味を持つため, これらの映像区間を検出することに対して多様なニーズがある. 図 2 に準同一映像区間対の例を示す. 検出される準同一映像区間対は, 映像区間の特徴の定め方としきい値の決め方に依存して異なる. ここでは, 文献 [4], [5] で採用されている各フレーム画像の各画素の輝度値に基づいた特徴を用いた場合について示す.

図 2 の (a) はテレビコマーシャル, (b) は番組のオープニングである. これらを検出することで, 新規コマーシャルデータベースの自動構築や繰り返し放送されるテレビ番組の検出などが可能になる. (c) は同一の素材映像にテロップや他の映像が合成されたものである. ニュース番組で何度も採り上げられ, 繰り返し放送される素材映像は, 重要な出来事と関連する可能性が高い.

3. 特徴空間のマージン付き適応分割

準同一映像区間を高速かつ漏れなく検出することを目的として, 特徴空間のマージン付き適応分割手法を提案する.

最初に, 特徴空間の分割による照合回数削減の基

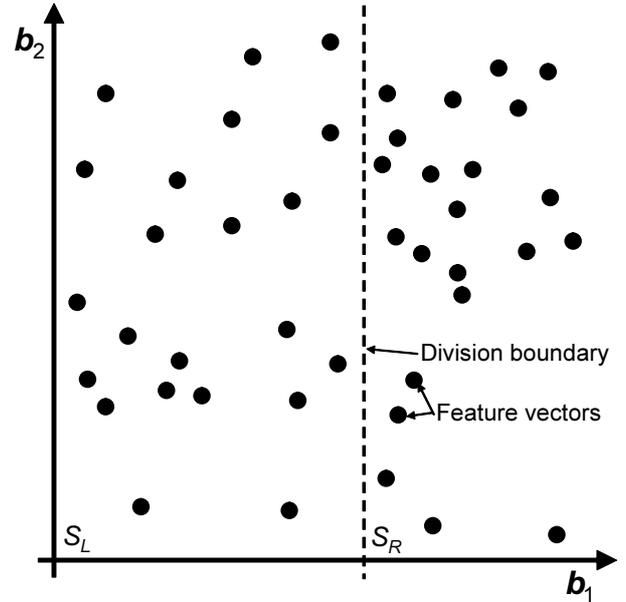


図 3 特徴空間の適応分割の概念図 (1 回目の分割結果)

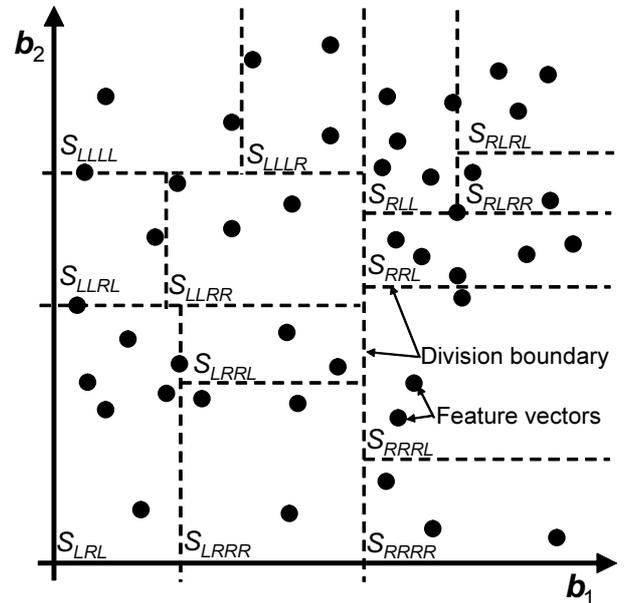


図 4 特徴空間の適応分割の概念図 (最終的な分割結果)

本的な考え方を述べる. N_O 個の特徴ベクトルが分布する特徴空間 S_O を分割し, 複数の分割された小空間 S_t ($t = 1, \dots, T$) を作る. 空間 S_t に含まれる特徴ベクトルの数を N_t としたとき, 分割後の照合回数は $\sum_{t=1}^T \binom{N_t}{2}$ となる. ここで, $N_O = \sum_{t=1}^T N_t$ より $\binom{N_O}{2} \geq \sum_{t=1}^T \binom{N_t}{2}$ が成立するため, これによって照合回数が削減できる.

以降では, 提案手法の詳細について述べる.

3.1 特徴空間の階層的分割による照合回数の削減

基底ベクトル $\{b_1, \dots, b_D\}$ によって形成される特徴空間 S_O に存在する特徴ベクトルの集合を $Q_O = \{q_1, \dots, q_{N_O}\}$ とし, 特徴ベクトルの 1 つを $q_n = (q_{n,1}, \dots, q_{n,D})$ で表す. 以下に階層的分割のアルゴリズムを示す.

Step 1: $S \leftarrow S_O, Q \leftarrow Q_O, N \leftarrow N_O$ として、次に進む。

Step 2: $b_{\hat{d}} (\hat{d} \in \{1, \dots, D\})$ を分割軸として選択する。

Step 3: $\{q_{1,\hat{d}}, \dots, q_{N,\hat{d}}\}$ の中央値 $m_{\hat{d}}$ を分割境界として計算する。中央値を選択することで、分割後の両空間の特徴ベクトルの数が等しくなり、照合回数の削減の効率が良くなる。

Step 4: $b_{\hat{d}}$ と $m_{\hat{d}}$ を用いて、 S を小空間 S_L と S_R に分割する。それぞれの空間に存在する特徴ベクトルの集合は以下ようになる。

$$\begin{aligned} Q_L &= \{q_n | q_{n,\hat{d}} < m_{\hat{d}}\} \\ Q_R &= \{q_n | q_{n,\hat{d}} \geq m_{\hat{d}}\} \end{aligned} \quad (2)$$

Q_L 内の特徴ベクトルの個数を N_L , Q_R 内の特徴ベクトルの個数を N_R とする。

Step 5: $S \leftarrow S_L, Q \leftarrow Q_L, N \leftarrow N_L$ として、Step 2-4 の処理を行い、さらに $S \leftarrow S_R, Q \leftarrow Q_R, N \leftarrow N_R$ として、Step 2-4 の処理を行う。

ある終了条件を満たすまで上記を再帰的に繰り返し行うことにより、階層的な空間分割を行う。ただし、Step 2 における \hat{d} の選び方、ならびに再帰処理の終了条件は 3.3 で述べる。図 3 に 2 次元の特徴空間を考えた場合の 1 回目の分割結果の概念図を、図 4 に最終的な分割結果の概念図を示す。特徴ベクトルの分布に応じて分割が階層的に行われる様子が伺える。

3.2 マージン空間による精度保証

以上の分割処理では、分割境界の付近に存在する準同一映像区間の特徴ベクトル対が、それぞれ異なる空間に分断された場合には、検出漏れが生じる。この問題を解決するため、図 5 のように分割境界 (Division boundary) を挟んで幅 ϵ の空間を設定し、 $|q_{n,\hat{d}} - m_{\hat{d}}| \leq \epsilon/2$ となる場合には、両方の空間に q_n を所属させるように、2 つの空間を部分的に重複させて分割する。この重複した空間をマージン空間 (Margin space) と呼ぶ。 ϵ は準同一映像区間検出の際の距離のしきい値である。これにより、任意の準同一映像区間対は、必ずどちらかの空間内に属することが保証される。このとき、異なる空間に共通して含まれるマージン空間内の特徴ベクトルを共通ベクトル (Shared vector) と呼ぶ。

3.3 適応分割による効率的な高速化

共通ベクトルは分割後の 2 つの空間に重複して存在する。そのため分割する度に特徴ベクトルの総数が増加していき、分割後の照合回数の和が分割前よりも増加する可能性がある。

そこで、特徴ベクトルの分布に応じた空間分割の打ち切りを行う。まず次式を定義する。

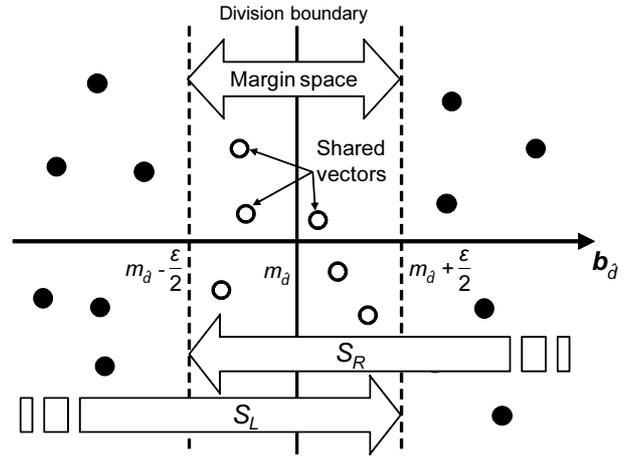


図 5 マージン空間

$$f_S(d) = \binom{N}{2} - \left(\binom{\frac{N}{2} + M_{L,d}}{2} + \binom{\frac{N}{2} + M_{R,d}}{2} \right) \quad (3)$$

ここで、 N は分割前の空間 S の特徴ベクトルの数である。また、 $M_{L,d}$ と $M_{R,d}$ はそれぞれ、空間 S を分割軸 b_d で分割したときの分割後の各空間 S_L, S_R での共通ベクトルの数である。 $f_S(d)$ は分割前と分割後の照合回数の差を示している。よって、3.1 のアルゴリズムの Step 2 における \hat{d} は、以下のように決定できる。

$$\hat{d} = \arg \max_d f_S(d) \quad (4)$$

なぜなら $b_{\hat{d}}$ が最も照合回数を削減できる分割軸となるためである。ただし、以下の条件を満たす d が見つからない場合は分割を終了する。

$$f_S(d) > 0 \quad (5)$$

上記の処理によって、映像区間ベクトルの分布に適応した特徴空間の分割を行うことができる。

4. 評価実験

一般放送映像を入力映像として、特徴空間のマージン付き適応分割を行った。得られた各空間に含まれる特徴ベクトルの数から照合回数を求めることにより、高速化の効果を評価した。準同一映像区間検出の単位となる映像区間長は 5.0 秒 (150 フレーム) とした。映像区間の特徴や検出しきい値に関しては、文献 [4], [5] で採用されているものを参考にした。具体的には、映像区間を各フレーム画像の各画素の輝度値に基づいて 20 次元の特徴ベクトルとして表現し、しきい値は図 2 に示した準同一映像区間が検出できる値に設定した。

なお、提案手法を用いた場合の検出精度は、入力映像中の全映像区間を総当りで照合し、式 (1) を満たす特徴ベクトル対を漏れなく検出する場合と等価となる。また、検出された準同一映像区間が人間の感覚と合っているかという観点からの精度評価は、準同一映像区間を利

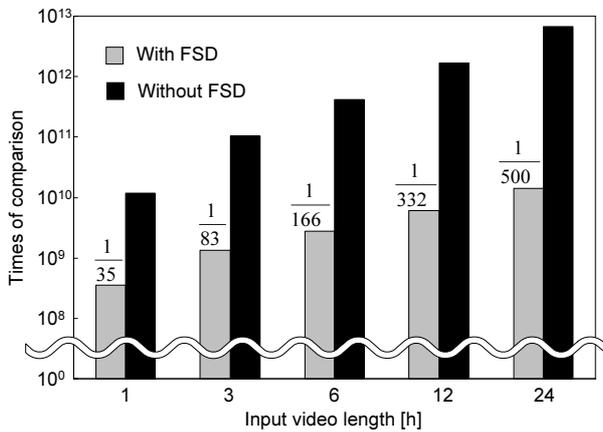


図 6 一般放送映像に対する提案手法の適用による照合回数の変化

用するアプリケーションに依存するため、本稿では扱わず、個別のアプリケーションに対する検出精度の評価は文献 [4], [5] に譲る。

4.1 入力映像時間長の違いに対する評価実験

提案手法の一般的な性能を調査するため、NHK 総合テレビ放送 24 時間分 (2006 年 11 月 8 日午前 5 時 ~ 翌日午前 5 時) を入力映像とした実験を行った。図 6 に、特徴空間のマージン付き適応分割を行った場合 (With FSD) と、行わなかった場合 (Without FSD) の照合回数を示す。ただし、各値の桁に差があるため縦軸は対数目盛りをとっている。

いずれの入力映像長においても、提案手法による大幅な照合回数の削減 (1/35 ~ 1/500) が確認できる。また、入力映像長の増大に従って、照合回数の削減の効果が高くなっていることがわかる。

4.2 異なる番組ジャンルの映像に対する評価実験

放送映像は、番組のジャンルによって、特徴空間中の特徴ベクトルの分布が大きく異なると推定される。例えばスポーツ中継では、単色のフィールドが画面に大きく写される場面や、固定カメラからの固定されたアングルで撮影された場面など、類似映像が多く含まれている。そのため、特徴空間への分布は特定の箇所に集中することが予想される。それに対して、映像が多彩に切り替わるバラエティ番組などでは、分布が広くばらつくと思われる。また、ニュース番組は他の 2 ジャンルの性質の間にあると考えられる。

各ジャンルに対する提案手法の効果を調べるため、バラエティ、ニュース、スポーツ中継 (サッカー) の 3 ジャンルの 6 時間分の映像に対して、提案手法を適用した。

図 7 に各ジャンルでの映像に提案手法を適用したときの照合回数の変化を示す。ここで、入力映像長に応じて映像区間の特徴ベクトルの数が決まるので、提案手法を適用しない場合 (Without FSD) の照合回数はジャンル

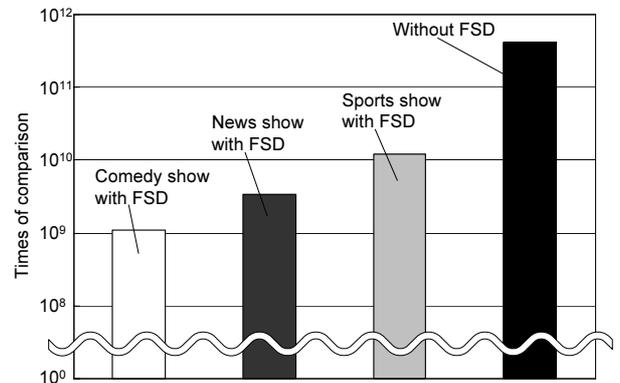


図 7 提案手法の適用によるジャンル別の照合回数の変化

の違いに依らず同数となる。図 7 から、いずれのジャンルに対しても大幅な照合回数の削減が確認できる。また、ジャンル別に照合回数を比較したとき、サッカー (Sport show with FSD)、ニュース (News show with FSD)、バラエティ (Comedy show with FSD) の順に値が大きい。このことから、特徴ベクトルが広域に分布している入力映像であるほど、提案手法が有効に働いていることがわかる。特にサッカー番組は緑色フィールドが画面に大きく写されるシーンが多いため、特徴空間の特定の箇所に特徴ベクトルの分布が集中する傾向が強い。そのため、マージン空間に含まれる共通ベクトルの数が増加して、分割が早い段階で打ち切られたため、他のジャンルと比較して効率的な照合回数の削減が行われなかったと考えられる。一方、バラエティ番組はカメラワークや CG エフェクトなどによって画面の移り変わりが変化に富む。その結果、特徴ベクトルが特徴空間に広く疎らに分布したため、効果的な照合回数の削減が可能となったと考えられる。

4.3 分割処理のオーバーヘッドに関する考察

特定の映像区間をキーとして、それと同一の映像区間を大量に蓄積された映像から高速に探索する場合には、探索を高速化するための特徴空間の分割処理は、事前処理として 1 度だけ行えばよい。処理に十分な時間をかけることができる。

しかし、本研究で対象とする準同一映像区間検出の場合には、特徴空間の分割処理は、入力映像が与えられてから準同一映像区間を検出する処理と同時にされる。つまり、提案手法によって準同一映像区間検出を検出する処理が高速化できたとしても、特徴空間の分割処理の計算コストが高ければ、全体としての処理の高速化は達成できない。提案手法が用いる分割処理の計算量は、中央値を算出する処理が入るため、特徴ベクトル数 N に対して $O(N \log N)$ となる。ここでは、この分割処理の全体の処理に対するオーバーヘッドを調査する。

図 8 に、4.1 で用いたものと同じ入力映像に対する全体の処理に要した計算時間と、そのうちの提案手法によ

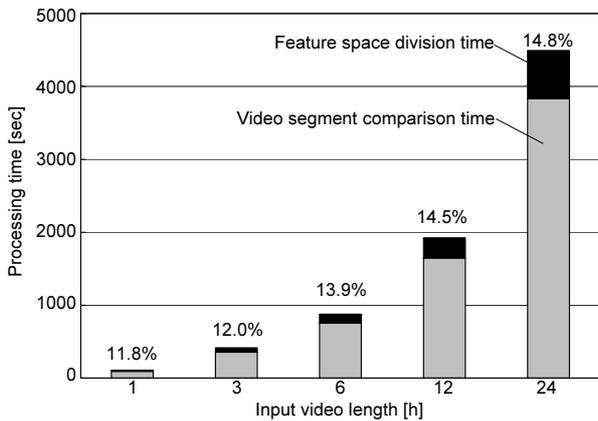


図 8 分割処理のオーバーヘッド

る分割処理に要した時間の割合を示す．例えば，入力映像長が 24 時間のときの分割処理に要する時間の割合は 15%程度であり，分割処理を行わない場合には図 6 で示したように約 500 倍の計算時間を要することを考えると，全体としての処理の高速化が十分達成できていることがわかる．

5. む す び

長時間映像中に現れる準同一映像区間を高速かつ漏れなく検出する手法を提案した．24 時間分の一般放送映像を入力映像として用いた実験の結果，提案手法を用いた場合には，総当りの場合と比較して約 500 倍高速に，かつ漏れなく準同一映像区間が検出できることがわかった．異なる番組ジャンルを対象とした実験では，バラエティ番組など，特徴ベクトルが広域に分布している入力映像であるほど，提案手法が有効に働くことを確認した．また，特徴空間の分割処理の全体の処理に対するオーバーヘッドを調査した結果，分割処理に要する計算時間は分割の結果として削減される計算時間に対して十分に小さいことから，提案手法を適用することによって全体としての処理の高速化が達成できていることがわかった．

今後の課題として，より効果的な高速化を行うための，特徴空間の分割方法の検討が考えられる．特に，共通ベクトル数の削減をねらって特徴ベクトルが密集した場所にマージン空間を配置しない方法を検討したい．

謝辞 日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に深く感謝する．本研究の一部は文部科学省科学研究費補助金による．また，本研究では画像処理に MIST ライブラリ (<http://mist.murase.m.is.nagoya-u.ac.jp/>) を使用した．

文 献

- [1] S . Arya , D . M . Mount , N . S . Netanyahu , R . Silverman, and A . Y . Wu, "An optimal algorithm for approximate nearest neighbor searching," *Journal of the ACM*, vol.45, no.6, pp.891-923, Nov. 1998.
- [2] 荒井英剛, 武本浩二, 加藤丈和, 和田俊和, "階層的固

- 有空間による高次元最近傍探索の高速化," 画像の認識・理解シンポジウム 2006 論文集, pp.291-297, July 2007.
- [3] 山岸史典, 片山紀生, 佐藤真一, 坂内正夫, "疑似高次元クラスタリングによる同一映像断片探索の高速化," 画像の認識・理解シンポジウム 2004 論文集, vol.2, pp.452-457, July 2004.
- [4] 社本裕司, 高橋友和, 井手一郎, 村瀬洋, "同一映像区間を手がかりとしたニュース映像アーカイブのトピック構造解析," 第 3 回デジタルコンテンツシンポジウム講演予稿集, 3-4, CD-ROM, June 2007.
- [5] A. Ogawa, T. Takahashi, I. Ide, and H. Murase, "Cross-lingual retrieval of identical news events by near-duplicate video segment detection," *Proc. the 14th International MultiMedia Modeling Conference*, pp.287-296, Jan. 2008.
- [6] 野田和広, 目加田慶人, 井手一郎, 村瀬洋, "圧縮特徴空間内での繰り返し照合を用いた長時間映像からの同一区間映像の高速検出," 2004 年度映像メディア処理シンポジウム資料, pp.105-106, Nov. 2004.
- [7] I. Ide, K. Noda, T. Takahashi, and H. Murase, "Genre-adaptive near-duplicate video segment detection," *Proc. 2007 IEEE International Conference on Multimedia and Expo*, pp.484-487, July 2007.
- [8] X . Naturel and P . Gros, "A fast shot matching strategy for detecting duplicate sequences in a television stream," *Proc. the Second International Workshop on Computer Vision Meets Databases*, pp.21-27, Oct. 2005.
- [9] 西村拓一, 水野道尚, 小木しのぶ, 関本信博, 岡隆一, "アクティブ探索法による時系列データ中の一致区間検出-参照区間自由系列アクティブ探索法-, " 電子情報通信学会論文誌, vol.J84-D-II, no.8, pp.1826-1837, Aug. 2001.
- [10] 村瀬洋, V. V. Vinod, "局所色情報を用いた高速物体探索-アクティブ探索法-, " 電子情報通信学会論文誌, vol.J81-D-II, no.9, pp.2035-2042, Sep. 1998.
- [11] 大盛善啓, "探索位置の予測に基づく時系列データの高一致区間検出-参照区間自由時系列アクティブ探索法 (RIFAS 法) の改良-, " 情報処理学会研究報告 CVIM, no.2007-CVIM-157, pp.93-100, Jan. 2007.
- [12] 柏野邦夫, 村瀬洋, "オーバースキッピングによる時系列アクティブ探索法の高速化," 日本音響学会研究発表会講演論文集, pp.445-446, Sep. 1999.