

繰り返し映像区間検出を高速化するための特徴空間の適応的分割

鈴木 秋吾[†] 高橋 友和^{†,††} 井手一郎^{†,†††} 村瀬 洋[†]

[†] 名古屋大学大学院 情報科学研究科 〒464-8601 愛知県名古屋市中種区不老町

^{††} 日本学術振興会

^{†††} 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-1

E-mail: [†] {aogawa, ttakahashi, ide, murase}@murase.m.is.nagoya-u.ac.jp, ^{†††} ide@nii.ac.jp

あらまし 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。その中に繰り返し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。繰り返し映像区間とは、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 (Near duplicate video segment) を指す。しかし、映像中の繰り返し映像区間をすべて検出するには、膨大な計算量が必要となるため、時間的なコストの面から実用化が困難であった。そこで本研究では、映像の特徴空間を特徴点の分布に応じて適応的に分割し、計算量を削減することで検出を高速化する手法を提案する。また、一般放送映像を用いた実験によって、その有効性を検証する。

キーワード 空間分割, 繰り返し映像, 高速化, 精度保証

Adaptive Division of Feature Space for Fast Detection of Near-Duplicate Video Sequences

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

[†] Graduate School of Information Science, Nagoya University Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

^{††} Japan Society for the Promotion of Science

^{†††} National Institute of Informatics 2-1-1 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: [†] {aogawa, ttakahashi, ide, murase}@murase.m.is.nagoya-u.ac.jp, ^{†††} ide@nii.ac.jp

Abstract With an explosive increase in recorded videos, technology that searches for video for fast and user requests is needed. Among these technologies, there is the technology for detecting near-duplicate video segments, and works application on application to video extraction using that results. Near duplicate video segments are identical or extremely similar video segments that appear two or more times in a video. Because a huge number of calculation is needed for detecting all near duplicate video segments, practice use is difficult. To detect near-duplicate video segments faster, this proposes a method that divides the feature space composed of video segments reduces computation time adaptively. By applying it to actual broadcast video, the effectiveness of the proposed method is verified.

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

1. はじめに

近年、膨大な量の映像を備蓄することが可能となり、映像の探索の効率化及び、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備蓄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 [1]~[5]。これらの中には、最近傍探索に基

づく手法がある。このアプローチでは、映像を特徴空間に投影したとき、入力キーに対する最近傍点を探索する。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 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。

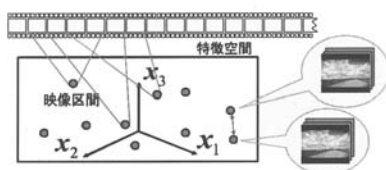


図 1 映像区間の特徴空間への投影

一方、これとは別の問題として、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映像区間をもれなく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 [6] [7] [8] [9]。例えば、社本ら [8] は、長期にわたって備蓄されたニュース映像中から、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 [9] は繰り返し映像区間の共起を手がかりに、異なる言語の放送映像の中から、同一ニューストピックについて放送された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。

しかし、図 2 に示すように、繰り返し映像区間を漏れなく検出するには、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 $N C_2$ 回の膨大な数の照合を要する。これに対し、野田ら [7] は、特徴次元圧縮により、映像区間対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 (図 5)。しかしながら、この手法では、照合回数は $N C_2$ のままであるため、照合回数を減らすことで更なる高速化が考え求められる。

そこで本研究では、特徴空間を分布に適応的に分割することで、照合回数の削減を図る。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、分割前と同じ検出精度を保証する。

2. 繰り返し映像区間検出

2.1 問題定義

本研究では、以下の条件を満たす映像区間対を「繰り返し映像区間」と定義し、これらを映像中から全て検出することを目的とする。

入力映像中の全映像区間を特徴ベクトルとして表現し、 $\mathbf{p}$, $\mathbf{q}$ を任意の映像区間をあらわすベクトルとする (図 1)。
このとき、 $\|\mathbf{p} - \mathbf{q}\| \leq \epsilon$ を満たす $\mathbf{p}$, $\mathbf{q}$ を繰り返し映像区間と定義する。

この閾値 ϵ 以下のベクトル対の探索を、本研究では ϵ -対探索 (ϵ -pair search) と呼ぶ。これは特定のキーの近傍点を探す最近傍探索問題等とは大きく異なる (図 2)。この問題に対し、単純な手法では、 d 次元のベクトル間の距離を計算するのに、 d に比例する計算コストを要し、また、 N 個のベクトルの中から全てのベクトル間の距離を計算するには $N C_2$ 回の照合を要する。つまり高次元になるほど、またベクトル数が多くなるほど探索に時間がかかる。

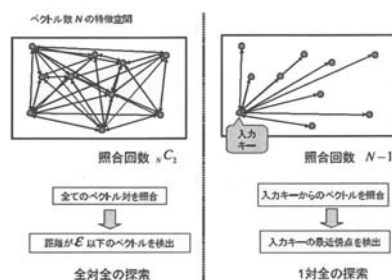


図 2 全対全探索と一対全探索

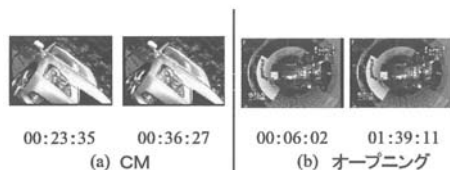


図 3 同一映像区間



図 4 ほぼ同一映像区間

2.2 繰り返し映像区間の検出例

放送映像において、繰り返し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もつ。これら映像区間を検出することに対して、多様なニーズがある。繰り返し映像区間には大きく分けて、図 3 の同一映像区間と、図 4 のほぼ同一 (Near-duplicate) な区間に分けられる。前者は CM など全く同じ映像区間を指す。後者は、細部が異なるが全体としては極めて似ており、特徴空間内では非常に近い箇所に分布する映像区間対を指す。

まずは、同一映像に関して説明する。図 3 の (a),(b) にはテレビコマーシャルと番組のオープニングを記している。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。

次に、ほぼ同一映像について説明する。図 4 の (a) はテロップや他の映像との合成が組み込まれたものである。何度も繰り返

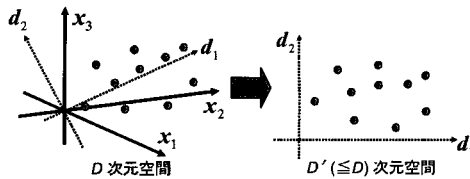


図5 特徴空間の低次元化

返し放送される同一素材ニュース映像は、重要な出来事と関連する可能性が高い、これを利用して社本ら??は関連するニューストピックの意味構造を抽出する手法を提案している。

図4の(b),(c)は同じ番組内で同じカメラから撮影された、極めて似た映像区間である。これらを調べれば、番組内のカメラワークや、番組の大きな節目を見つけることができる。

2.3 従来研究

本研究の関連研究として以下の2つの高速化に関する研究を挙げる。前者は近似解法を用いた手法であり、精度と高速化のトレードオフを利用した技術である。後者は、特徴空間の次元を削減することによって、ベクトル間の距離計算自体の高速化を実現する手法である。

2.3.1 近似解法

その他の高速手法として手法として、ハッシュを用いた階層的クラスタ解析手法である Locality Sensitive Hashing による手法が提案されている。この手法では入力キーの近傍点の候補を、空間ではなく、同じハッシュキーを持つベクトルのみに範囲を限定する。これにより確実な高速化ができる一方で、分布のあらゆる変化に応じた完全なハッシュ関数が存在しない以上、近似解法となる。

2.3.2 距離計算の高速化

特徴空間を次元圧縮することで、ベクトル間の距離計算を高速化する手法である[7]。具体的には、十分長い訓練映像に対して主成分分析を行い、低次元空間で基底に用いる固有ベクトルを求める。そのうち D' 個を選び、 D 次元空間中の映像区間ベクトルを D' 個の固有ベクトル $(\mathbf{d}_1, \dots, \mathbf{d}_{D'})$ を基底とする低次元空間へ、射影する。得られた D' 次元の空間で距離計算することによって、 D と D' の比に応じた計算時間の短縮が可能になる(図5)。

このとき D 次元空間から D' 次元空間へ射影をするとき特徴量の欠落が起きる。そのため、 D' 次元空間で得られた結果は漏れはないが、 ϵ 以上を含んでいる可能性がある。そこで、 D' 次元空間の検出結果を D 次元空間に戻し照合をする。これを詳細照合と呼ぶ。詳細照合を行うことで距離 ϵ 以下の対を全て検出することができる。

3. 特徴空間の適応的分割

繰り返し映像区間検出を高速化するため、特徴空間を分割することによって、照合回数を削減する手法を提案する。映像ベクトルが分布している特徴空間 S を分割し、新たに複数の分割された小空間 S_t ($t = 1, \dots, T$) を作る。その結果、分割された

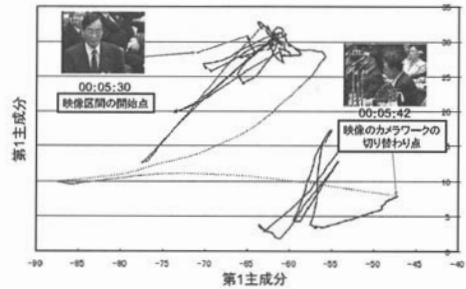


図6 低次元空間への再投影

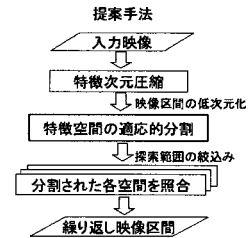


図7 提案手法の処理の流れ

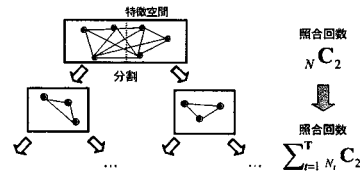


図8 特徴空間の分割による距離計算(照合)回数の削減

各空間 S_t に含まれるベクトル数を N_t としたとき、 T 個の空間に分割した後の照合回数は分割前の $N C_2$ 回から $\sum_{t=1}^T N_t C_t$ 回となる。ここで、 $N = \sum_{t=1}^T N_t$ より $N C_2 \geq \sum_{t=1}^T N_t C_t$ が必ず成立するため、照合回数が削減できる。

以降では、提案手法である特徴空間の適応的分割の詳細について述べる。図7に処理の流れを示す。

3.1 階層的分割による照合回数の削減

低次元特徴ベクトルの集合を $S = \{\mathbf{q}_1, \dots, \mathbf{q}_N\}$ とおく。まず、ベクトル $\mathbf{q}_i$ の要素のうち、特徴空間の1つの基底 $\mathbf{d}_j$ に射影しその値を q_{i,d_j} とする。 $\{q_{1,d_j}, \dots, q_{N,d_j}\}$ の中央値 m_{d_j} を算出し、これを分割境界とする(図9)。中央値を選択することで、分割後の両空間のベクトル数が均一となり、照合回数の削減量が多くなる。 S に含まれる各 $\mathbf{q}_i$ を $q_{i,d_j} < m_{d_j}$ となる空間 S_L とそれ以外の集合 S_R に分割する。以下にそのアルゴリズムを示す。

- (1) N 個のベクトルを含む特徴空間 $S = \{\mathbf{q}_1, \dots, \mathbf{q}_N\}$ につき、 S 内に存在する軸の中から分割軸 $\mathbf{d}_i$ を選択する。
- (2) 全ベクトルを $\mathbf{d}_i$ に射影したとき、その中央値を m_{d_i} とする。この m_{d_i} を用いて、以下に示す条件で m_{d_i} の両側に新

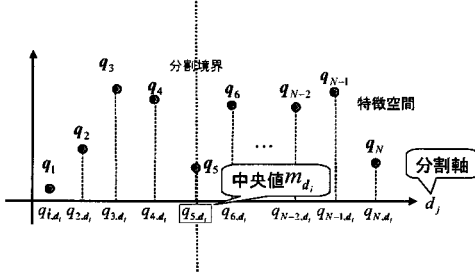


図 9 分割境界の選択

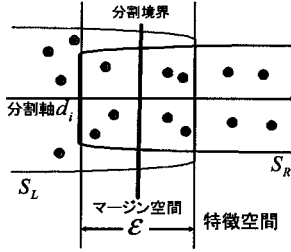


図 10 マージン空間

たな小空間 S_L, S_R を作る.

$$S_L = \{q_i | q_{i,d_j} < m_{d_j} (q_i \in S)\}$$

$$S_R = \{q_i | q_{i,d_j} \geq m_{d_j} (q_i \in S)\}$$

(3) 空間 S_L, S_R に対して, (1), (2) の処理を再帰的に行う. なお, 再帰処理の停止条件を 4.3 節に記す.

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繰り返し映像区間のベクトル対が, 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, 分割境界を挟んで幅 ϵ の空間を設定し, $|q_{i,d_j} - m_{d_j}| \leq \theta$ となる場合には, 両方の空間に q_i 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 (図 10). この重複した領域をマージン空間と呼ぶ. ϵ は繰り返し区間検出の際の距離の閾値とする. これにより任意の q_i に対する繰り返し映像区間は, 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区間内の特徴ベクトルを共通ベクトルと呼ぶ.

3.3 適応的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. そのため分割するたびに全空間の総ベクトル数が増加していき, 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. 図 11 に, ニュース映像 1 時間を対象として階層分割を行ったときの, 各分割数での共通ベクトルの増加を示す. ただし, 元の特徴空間に分布していたベクトル数は 10,000 個である. この流れから, 分割が進むにつれて, 共通ベクトルが指数的に増加するのが分かる.

そこで, 特徴空間のベクトルの分布に応じた分割の打ち切り

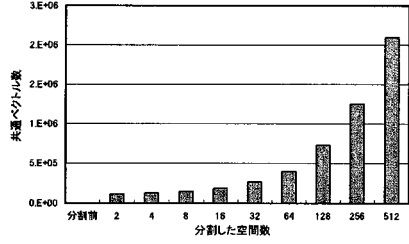


図 11 共通ベクトル数の増加

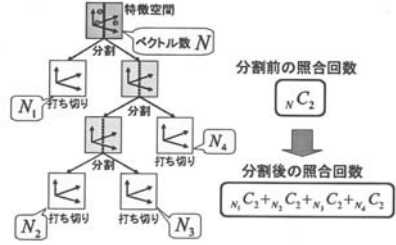


図 12 特徴空間の適応的分割

を行う. まず次式を定義する.

$$f_{S_i}(d_j) = N_i C_2 - \left(\frac{N_i}{2} + M_{L,d_j} \right) C_2 + \left(\frac{N_i}{2} + M_{R,d_j} \right) C_2 \quad (1)$$

ここで, N_i は分割前の空間 S_i のベクトル数である. また, M_{L,d_j} と M_{R,d_j} は空間 S_i を分割軸 d_j で分割したときの分割後の各空間での共通ベクトルの数である. $f_{S_i}(d_j)$ は分割前と分割後の照合回数の差を示している. 式 (1) を用いて, $f_{S_i}(d_j)$ の値が最大となる軸を選択して分割を行う. なぜならその軸が最も照合回数を削減できる軸であるためである. また, $f_{S_i}(d_j) > 0$ を満たす軸が見つからなければ分割を終了する. 上位主成分から順に分割を行うのではなく, 最も条件の良い軸を毎回選択するので, 同じ軸を何度も用いることができる.

上記の処理によって, 映像区間ベクトルの分布に応じた特徴空間分割を行うことができる. その結果, 図 12 に示すような非対称でいびつな分割木ができる. 本論文では, 階層の浅い空間から深い空間にかけて, 1 階層, 2 階層, ... と数えている.

4. 評価実験

提案手法の効果を調査するため, 実際の放送映像に対して, 特徴空間を分割した. 得られた分割空間に対して, 照合回数を分割前と比較した.

4.1 実験データ

入力映像に対して特徴空間の適応的分割を行い, 照合回数を調べる. まず事前処理として, 入力映像中の全映像区間を元の D 次元から特徴次元圧縮し, D' 次元の低次元特徴空間へ射影する. 次に, 得られた特徴空間に対して適応的分割を施し, 各ベクトルを分割された空間へ配分する. そして得られた分割空間に

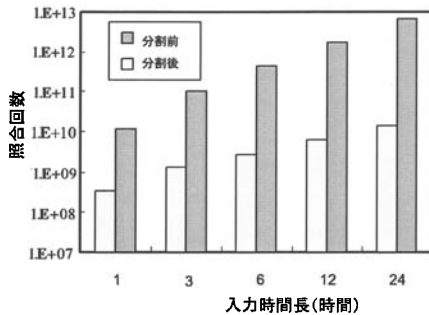


図 13 一般放送映像の適応的分割による照合回数変化

含まれるベクトル数から照合回数を求め、高速化の効果を推定する。実験に用いたデータを以下に示す。

実験データ

- 映像のファイル形式: MPEG1
- 映像区間長: 5.0 sec (150 frame)
- 圧縮前の空間における情報量:
 $D = 360 \times 240 \text{ pixel} \times 30 \text{ frame} \times 5.0 \text{ sec} \times 3 \text{ (RGB)} \text{ 次元}$
- 圧縮後の空間における情報量:
 $D' = 20 \text{ 次元}$

4.2 一般放送映像に対する評価実験

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 (2006 年 11 月 8 日 午前 5 時～翌日 午前 5 時) を対象とした空間分割実験を行った。図 13 に、特徴空間の適応的分割を行う前後の照合回数を示す。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。

両手法を比較すると、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) を確認した。照合回数の変化は、入力映像長の増加とともに指数的に増加している。これは、分割前の照合回数 $N \cdot C_2$ と分割後の照合回数 $\sum_{t=1}^T N_t \cdot C_2$ の両方が、空間に含まれるベクトル数の 2 乗オーダーに従って増加するからである。つまり、入力映像長の増加に従って、照合回数の削減の効果が高くなる。

4.3 異なる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、特徴空間中の特徴点の分布が大きく異なると推定される。図 14 と図 15 にサッカー、バラエティそれぞれを特徴空間に投影したものを示す。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。

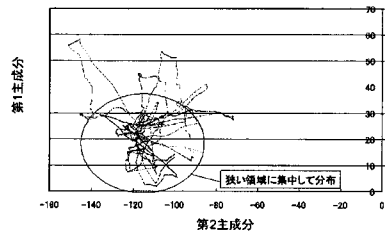


図 14 サッカー中継を特徴空間へ投影した分布

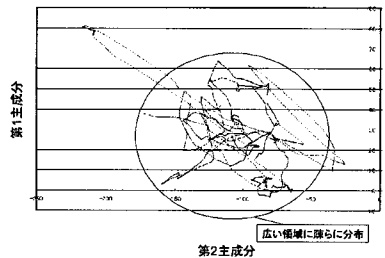


図 15 バラエティ番組を特徴空間へ投影した分布

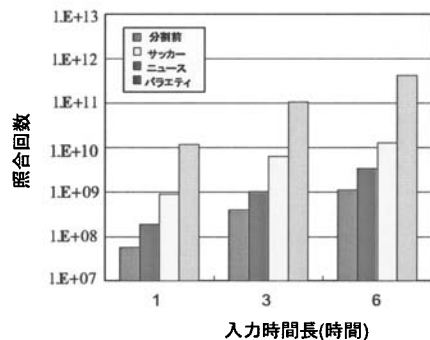


図 16 ジャンル別の適応的分割による照合回数変化

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、サッカー (スポーツ中継) の 3 ジャンルの映像に対して、特徴空間の分割を行った。

図 16 に各ジャンルでの映像に適応的分割を行っ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ベクトルの数が決まるので、特徴空間の分割前の照合回数はジャンルの違いによらない。図 16 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、いずれの入力時間に対してもサッカー、ニュース、バラエティの順に値が大きい。これは照合回数が多い順に、特徴空間の特定の箇所に分布が集中する傾向が強いと考えられるため、早い段階で分割が打ち切られると考えられる。なぜなら、分布が特定の箇所に集中してしまうと、マージン区間に含まれるベクトルの数が増えてしまい、照合回数の削減が効率的に行えないからである。特にサッカー番組は緑色フィー

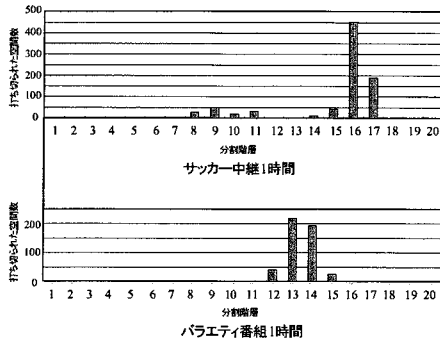


図 17 各階層で打ち切られた空間数

ルドが画面に大きく写されるシーンが多いため、類似した映像区間が増加しやすい。

逆に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、特徴空間に広く疎らに分布するため、効果的な照合回数の削減が可能になる。

一般に放送映像の特徴ベクトルは、特定の箇所分布に偏ることが多く、提案手法の効果はこの偏りに大きく依存する。

図 17 に、上記の実験を行ったときの、分割で空間が打ち切られた階層の深さの頻度を示す。この図から、サッカー映像は分割の打ち切られる階層が浅い箇所と深い箇所に分かれていることが読み取れる。これは、分布が偏っている場合、密集している箇所は浅く分割され、逆に密集している箇所以外は深く分割されるためと考えられる。このことから、提案手法を用いたとき、分布が偏っていると、バランスの悪い分割になってしまい、効率的な照合回数の削減ができないと予想される。

5. む す び

本研究では、繰り返し映像区間検出の高速化のため、特徴空間の適応的分割を用いて、照合回数の削減を行う手法を提案した。

評価実験では、一般放送映像（24 時間）と 3 番組ジャンルを対象とした、提案手法を用いる前後で照合回数の比較を行い、提案手法の有効性を示した。また 3 番組ジャンルを対象とした実験では、広域に分布している特徴空間であるほど、提案手法が有効であることを確認した。

今後の課題として、より効果的な分割を行うための、分割境界を決定する新たな方針の検討が考えられる。特に、今まではベクトルの要素の中央値に依存し、マージン空間に含まれるベクトル数を考慮していなかったため、マージン空間が分布の密集を回避するような指標を見つける必要がある。

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感謝する。また、映像を提供してくれた国立情報学研究所 (NII) に謝辞を述べる。本研究の一部は、文部科学省 21 世紀 COE プログラムおよび科学研究費補助金による。また、本稿の実験では、映像情報処理に MIST ライブラリ

(<http://mist.suenaga.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 (MIRU)2006 論文集, pp.291-297, July 2007.
- [3] 荒井英剛, 中村 恭之, 和田俊和, "主成分木を用いた画像学習法の提案と性能比較", 画像の認識・理解シンポジウム (MIRU)2007 論文集, pp.516-521, July 2007.
- [4] D. Dementhon and D. Doermann, "Video retrieval using spatio-temporal descriptors", *Proceedings of ACM Multimedia 2003*, pp.508-517, Nov. 2002.
- [5] D.Zhang and S.Chang, "Detecting image near-duplicate by stochastic attribute relational graph matching with learning," *Proceedings of ACM Multimedia 2004*, pp.877-884, Oct. 2004
- [6] X. Naturel and P. Gros, "A fast shot matching strategy for detecting duplicate sequences in a television stream", *Proceedings of the Second International Workshop on Computer Vision Meets Databases*, pp.21-27, Oct. 2005
- [7] 野田和広, 高橋友和, 井手一郎, 目加田慶人, 村瀬洋, "適応的特徴選択を用いた長時間放送映像からの高速な繰り返し区間検出," 電子情報通信学会技術研究報告, PRMU2005-289, Mar. 2006
- [8] 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. 200
- [9] 小川晃, 野田和広, 高橋友和, 井手一郎, 村瀬洋, "画像情報を用いた同一ニュースイベントの言語横断検索," 第 3 回デジタルコンテンツシンポジウム講演予稿集, 2-1, CD-ROM, Jun. 2007