

グローバルな枝刈りを導入した音や映像の高速探索

木村 昭悟[†] 柏野 邦夫[†] 黒住 隆行[†] 村瀬 洋[†]

A Quick Search Method for Multimedia Signals Using Global Pruning

Akisato KIMURA[†], Kunio KASHINO[†], Takayuki KUROZUMI[†], and Hiroshi MURASE[†]

あらまし 探したい音や映像が具体的に与えられ、それとほぼ同じ信号の区間が蓄積された長時間の音や映像の中のどこにあるかを高速に探索する問題に対する新しい手法を提案する。これまでに、高速探索手法の一つとして、時系列アクティブ探索法が提案されている。これは、信号から抽出したヒストグラムに基づく信号探索手法であり、ローカルな枝刈り、すなわち照合窓周辺の類似度に基づいて探索の必要がない区間の照合を省くことにより高速な探索を実現していた。それに対し、提案手法は、ローカルな枝刈りに加え、信号全体の類似性に基づくヒストグラムの分類により信号時系列全体を見て探索の必要がない区間を取り除く、グローバルな枝刈りを導入することにより、更に高速な探索を実現する。本論文では、グローバルな枝刈りの度合と保証される精度との関係を詳しく議論する。例えば提案手法では、128 次元のヒストグラムを 1024 通りに分類したとき、時系列アクティブ探索法に比べ、事前処理の計算時間が信号時系列の再生時間の 1%程度増加するものの、同程度の精度を保ったまま、探索速度を約 9 倍にできることを実験により示す。

キーワード 一致探索、時系列アクティブ探索法、グローバルな枝刈り、ヒストグラム重なり率、 L_2 距離

1. ま え が き

近年、大量の音や映像のデータベースの中から目的の音や映像を高速かつ正確に探し出す技術が強く求められている。

音や映像の検索・探索に関する手法は数多く提案されている。これらの手法の多くは、求める音や映像の内容に関する何らかの条件を指定して、それに適合する具体的な音や映像をデータベースや長時間の素材等から取得するものである [1] ~ [5]。本論文では、この探索方式を内容検索と呼ぶ。

一方、本論文では、探したい音や映像 (参照信号, reference signal) が具体的に与えられ、それとほぼ同じ信号の区間が蓄積された膨大な音や映像 (蓄積信号, stored signal) の中のどこに存在するかを高速かつ正確に検出する手法を提案する。本論文では、この探索方式を一致探索と呼ぶ。一致探索手法は、テレビ放送やラジオ放送を蓄積したデータに対する特定のコマースル・楽曲等の検出や統計情報の作成、インターネット

トにおける音楽・映像等著作物の不正利用防止など、幅広い応用が考えられる。

このような手法の一つとして、ヒストグラムに基づいた信号探索手法である時系列アクティブ探索法 (Time-Series Active Search) が提案されている [6]。時系列アクティブ探索法は、照合時点周辺の類似性に基づいて探索の必要がない区間の照合を省くことにより高速な探索を実現している。本論文では、この高速化手法をローカルな枝刈り (local pruning) と呼ぶ。時系列アクティブ探索法では、あらかじめ特徴抽出を行っておいた場合、60 時間分の蓄積された音や映像から、15 秒分の音や映像の断片と同一の区間を、およそ 1 秒以内で探索することが可能である。しかし、更に膨大な蓄積信号を対象とする場合、より高速に探索を行うことが求められる。

ところが、時系列アクティブ探索法は、照合時点周辺の類似性のみを用いているため、例えば蓄積信号中で参照信号と全く類似していない区間が長く連続していたとしても、その区間の照合を完全には省くことができない。そのため、時系列アクティブ探索法は、参照信号と蓄積信号全体との類似性にかかわらず、蓄積信号の長さに従って探索時間が増大するという問題点がある。

[†] 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所、厚木市
NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi-shi, 243-0198 Japan

これらの問題を解決するためには、蓄積信号全体の類似性を考慮した探索範囲の絞込みが不可欠である。

これまでも、探索範囲を絞り込む手段は、主に静止画像検索などを対象にデータベース技術の分野で数多く発表されている。例えば、事前に超直方体 [7]、超球 [8]、超直方体と超球の共通部分 [9] でそれぞれデータを階層的に分類して探索木を構成することで類似画像検索を高速に行う手法が代表的である。これらの手法は、基本的には事前処理の計算量は問わずに探索時の計算量を最小化することを目的としたものであり、静止画像検索などのように、信号から抽出した特徴量の次元が比較的低い場合には効果が高い。しかし、これらの手法を音や映像の探索に応用した場合、特徴量の次元若しくは信号規模の増大につれて事前処理に要する計算量が爆発的に増加してしまう、探索木の保持に必要な記憶容量が膨大になるなどの問題点がある。

これに対し本論文では、時系列アクティブ探索法におけるローカルな枝刈りに加え、膨大な計算量や記憶容量を必要としない事前処理によって蓄積信号全体の類似性を考慮した探索範囲の絞込みを効率的に行う手法、すなわちグローバルな枝刈り (global pruning) を導入した探索手法を提案する。以下、2. で時系列アクティブ探索法の概要を述べた後、3. で提案手法について具体的に説明し、4. で提案手法の有効性を実験によって示す。5. をむすびとする。

2. 時系列アクティブ探索法

2.1 アルゴリズムの概要

従来法である時系列アクティブ探索法の処理の概要を図 1 に示す。アルゴリズムの詳細は [6], [10] に記載されているので、ここでは要点のみを記す。

まず、参照信号と蓄積信号からそれぞれ特徴ベクトルを計算する。次に、参照信号と蓄積信号の双方に対して同じ長さの時間窓をかけ、窓内の特徴ベクトルを分類してヒストグラムを作る。そして、ヒストグラム同士の類似度の値 (類似値) があらかじめ設定した値 (探索しきい値, search threshold) を超えるかどうかで、参照信号の有無を判定する。このとき、類似値と探索しきい値から、探索漏れを起こさないことを保証したまま探索を時間方向にスキップできる時間幅 (スキップ可能幅) を求めることができ、その分だけ蓄積信号にかかる窓をずらして探索を進める。

特徴ベクトルは音響信号・映像信号ともに [10] と同様の方法で計算する。ヒストグラム同士の類似度はヒ

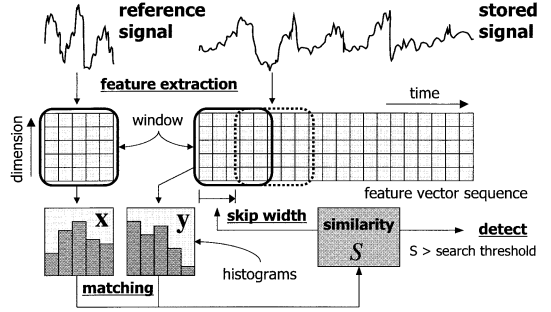


図 1 時系列アクティブ探索法の概要
Fig. 1 Overview of time-series active search.

ストグラム重なり率 (histogram intersection) [11] を用いる。ヒストグラム重なり率は次の式で定義される。

$$S_1 = S_1(H_R, H_S) \stackrel{\text{def}}{=} \frac{1}{D} \sum_{i=1}^L \min(h_{Ri}, h_{Si}). \quad (1)$$

ただし、 H_R 及び H_S は、それぞれ参照信号と蓄積信号に対するヒストグラム、 h_{Ri} 及び h_{Si} はそれぞれのヒストグラムの i 番目の次元に含まれる特徴ベクトルの度数、 L はヒストグラムの次元数、 D はヒストグラムの総度数である。

スキップ可能幅 w は、類似値の上限に関する考察 [6] から、次式で求められる。

$$w = \begin{cases} \text{floor}(D(\theta_1 - S_1)) + 1 & (\text{if } S_1 < \theta_1) \\ 1. & (\text{otherwise}) \end{cases} \quad (2)$$

ただし、 $\text{floor}(x)$ は切り下げを、 θ_1 は探索しきい値を表す。

類似度としてヒストグラム重なり率を用いる理由は、(1) 計算が容易である (2) スキップ可能幅が簡単な計算によって求められる (3) 雑音環境下で高い探索精度を示すである。

2.2 ヒストグラム重なり率の有効性

ヒストグラム重なり率の類似尺度としての妥当性を調べるため、探索精度に関してヒストグラム重なり率と、識別尺度として一般的に用いられるヒストグラムの L_2 距離 (ユークリッド距離) とを比較する実験を、音響信号を用いて行った。 L_2 距離は、以下で定義される。

$$d_2(H_R, H_S) \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^L |h_{Ri} - h_{Si}|^2}. \quad (3)$$

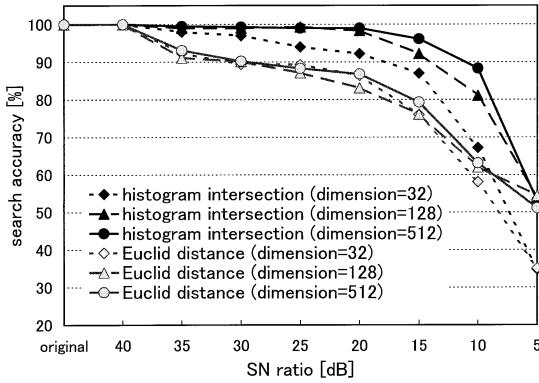


図 2 距離尺度による耐雑音性の違い

Fig. 2 Noise tolerances of two distance measures.

まず、繰返しを含まない約 20 分の音響信号を 2 回に分けて計算機に取り込んだ。このうちの一方から、一定の時間区間をランダムな場所から切り出して参照信号とし、他方を蓄積信号とした。そして、取り込んだ各音響信号から、前述の手法により特徴抽出を行った。蓄積信号に対して白色ガウス雑音を加えた場合についても実験を行った。

本実験では、ヒストグラムの次元数、及び雑音を重畳した際の SN 比をパラメータとした。同一の実験条件において、時系列アクティブ探索法を 100 回繰り返して探索を行い、精度を測定した。適合率と再現率とが等しくなる探索しきい値を設定し、精度は、その探索しきい値での適合率 (= 再現率) で評価した。ここで、適合率とは、探索結果として出力されたもののうち正しいものの割合であり、再現率とは、探索されるべきもののうち探索結果として出力されたものの割合である。

実験結果を図 2 に示す。これによれば、35 dB 以下では、いずれの次元数においても、ヒストグラム重なり率は L_2 距離と比べて 10~15% 程度高い精度を得ている。ヒストグラム重なり率においては、次元数が 128 あれば、SN 比 20 dB までは 99% 以上の探索精度が得られている。

以上の実験から、探索精度に関してヒストグラム重なり率が L_2 距離よりも優位であることが明らかになった。

3. 提案手法

3.1 アルゴリズムの概要

提案法の概要を図 3 に示す。図 3 において、二つ

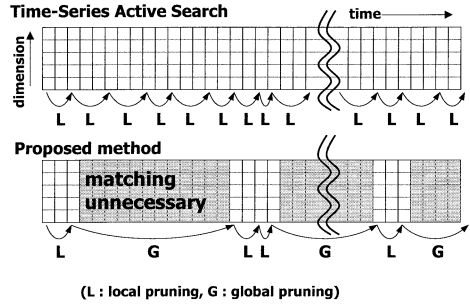


図 3 提案法の概要

Fig. 3 Overview of the proposed method.

の方形はともに蓄積信号を表しており、横軸 1 目盛り信号 1 フレームを表現している。

時系列アクティブ探索法では、式 (2) より、スキップ可能幅の最大値はヒストグラムの総度数すなわち時間窓の幅で上から抑えられている。それに対し、提案手法では、事前処理において探索に先立って蓄積信号全体を調べ、参照信号との類似値が小さい蓄積信号中の区間をあらかじめ求めておく。これにより、大幅に探索範囲を絞り込むことが可能となる。

処理手順を図 4 に示す。

処理は大きく事前処理と探索時の処理に分けられる。事前処理は、具体的な参照信号が与えられる前に行うことができる。

事前処理は四つの手順からなる。

- (1) 蓄積信号から特徴ベクトルを計算する。
- (2) (1) に時間窓をかけ、窓を 1 フレームずつずらしながらヒストグラムを作成する。
- (3) ヒストグラム空間を分割する。
- (4) ヒストグラムを各分割空間に分類する。

なお、時間窓の長さは、想定される参照信号の長さとする (例えば 15 秒)。また、ヒストグラムは、特徴ベクトルをベクトル量子化 (VQ) し、VQ 符号語ごとの特徴ベクトル数を数えることによって作成する。

以下では、蓄積信号から作成された各ヒストグラムを蓄積ヒストグラム、蓄積ヒストグラムの時系列を蓄積ヒストグラム系列と呼ぶ。時系列アクティブ探索法においては、蓄積ヒストグラムを事前に作成しないため、時間窓の大きさは与えられる参照信号の長さごとに変化することを許容していたが、本手法では蓄積ヒストグラムを事前に作成する必要があるため、時間窓の大きさは固定される。

以上の事前処理によって、各ヒストグラムの分類結

果が出力される。

探索時の処理は四つの手順からなる。

- (1) 参照信号から特徴ベクトルを計算する。
- (2) (1) からヒストグラムを作成する。
- (3) ヒストグラムの分類を用いてグローバルな枝刈りを行う。
- (4) 選択された分類に属するヒストグラムに対してのみ時系列アクティブ探索法に基づくヒストグラム照合を行う。

参照信号から作成されたヒストグラムを、参照ヒストグラムと呼ぶ。

時系列アクティブ探索法において、蓄積ヒストグラムは、ヒストグラム照合を行いながら、照合が必要な箇所のヒストグラムのみをそのつど作成していた。これに対し提案法では、事前処理においてすべての箇所における蓄積ヒストグラムを作成するので、探索時にも、この作成済の蓄積ヒストグラムを用いて照合を行うことも可能である。しかし、この方法では蓄積ヒストグラムを保持しておくのに大きな記憶容量が必要となり、長時間の蓄積信号に対する探索が困難となる。

そのため本論文では、事前処理における蓄積ヒストグラムの作成は、あくまでヒストグラムの各分類に属する時間区間を得ることだけを目的とするものとし、探索時には、時系列アクティブ探索法と同様に、照合を行いながら、必要な箇所における蓄積ヒストグラムを改めて作成する方法を取った。

提案法は、時系列アクティブ探索法と比較すると、事前処理 (2) において、探索に先立って蓄積信号に対するヒストグラムを作成しておくとともに、事前処理 (3) (4) によってあらかじめそれらのヒストグラムを分類しておき、探索時の処理 (3) で探索範囲を絞り込むことによって高速化を実現している。

以下では、グローバルな枝刈り手法における主要な技術について、詳しく説明していく。

3.2 ヒストグラム空間の分割

ヒストグラム空間の分割では、VQ における符号帳学習アルゴリズム、例えば LBG アルゴリズム [12] を用いる。VQ における距離尺度としては、 L_2 距離を用いる。

3.3 ヒストグラムの分類

ヒストグラムの分類の概要を図 5 に示す。

あらかじめ作成された蓄積信号側の各ヒストグラムは、VQ アルゴリズムを用いて分類される。すなわち、ヒストグラムからの L_2 距離が最小となる代表ヒストグラム (representative histogram) をもつ分類に割り振られる。代表ヒストグラムとは、VQ の符号語に対応するヒストグラムを指し、VQ の符号帳の学習に LBG アルゴリズムを用いた場合、同じ符号語が割り当てられた学習サンプルの重心となる。以上の過程で作成されたヒストグラムの分類を、クラスと呼ぶ。

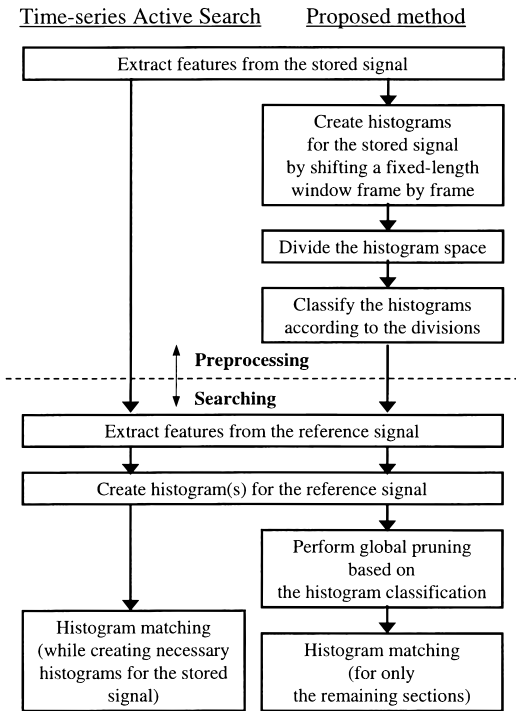


図 4 処理手順
Fig. 4 Processing procedure.

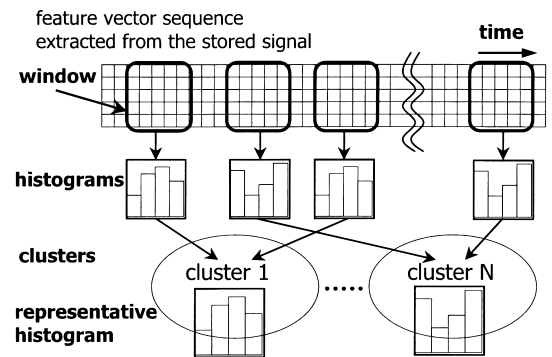


図 5 ヒストグラムの分類の概要
Fig. 5 Histogram classification scheme.

本節と前節で説明した処理は参照信号を必要としないので、すべて探索の前に行うことができる。

3.4 グローバルな枝刈り

3.4.1 概要

本節で述べる処理は、前節以前で述べた処理によって得られたクラスタを用いて探索されるべき蓄積信号の区間を見つけ出すもので、参照信号が与えられてから行う。

参照ヒストグラムは、前節 3.3 で述べたヒストグラムの分類によって、 L_2 距離が最小となる代表ヒストグラムをもつクラスタ(参照クラスタ)に分類される。次に、ある選択条件を満たすクラスタを選択し、選択されたクラスタに分類されているヒストグラムを探索の対象とする。このようにして探索されるべき蓄積ヒストグラムを決定し、それらのヒストグラムに対してのみ時系列アクティブ探索法に基づくヒストグラム照合を行う。

ここで重要な点は、クラスタの選択条件を具体的にどのように定めるかである。以下では、その点について述べる。

3.4.2 クラスタの選択条件

クラスタの選択条件として、参照ヒストグラムとの L_2 距離があらかじめ定められた値である選択しきい値 (selection threshold) を下回るようなヒストグラムを含む可能性があるかどうかを調べる (図 6)。この条件を満たすクラスタのみを選択し、満たさなければ選択しないものとする。以下では、この選択条件の判別式を導出する。

まず、参照クラスタに隣接するクラスタについて考える。

図 7 は、3 点 R, C_1, C_2 が乗るような平面上で L 次元ヒストグラム空間を切り出した様子を表している。ここで、 R は参照ヒストグラム、 C_1 は参照クラスタの代表ヒストグラム、 C_2 は参照クラスタに隣接するクラスタの代表ヒストグラムを表し、 d_{R1}, d_{R2}, d_{12} はそれぞれ、 R と C_1 、 R と C_2 、 C_1 と C_2 の L_2 距離を示す。

ここでの目的は、 C_2 を代表ヒストグラムとするクラスタに属するヒストグラムの中に、参照ヒストグラムとの L_2 距離が選択しきい値 θ_2 を超えるヒストグラムが存在する可能性があるかどうかを判別する式を導出することである。

参照ヒストグラム R からの距離が選択しきい値 θ_2 以内であるヒストグラム、すなわち、 R を中心とする

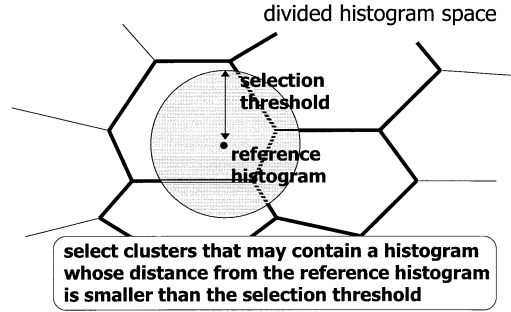


図 6 グローバルな枝刈りによって選択されるクラスタ
Fig. 6 Clusters selected by the global pruning.

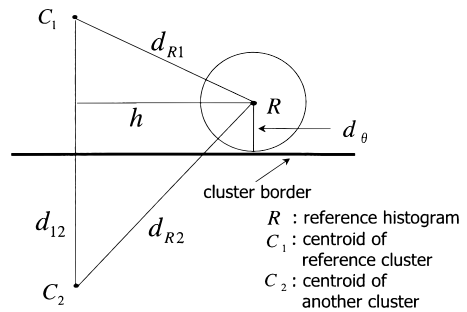


図 7 クラスタの選択条件
Fig. 7 Condition of cluster selection.

半径 θ_2 の超球の内部にあるヒストグラムが探索の対象となるので、選択しきい値 θ_2 が参照ヒストグラムとクラスタとの最短距離 d_θ より大きくなる (d_θ が選択しきい値 θ_2 よりも小さくなる) ときに、 C_2 を代表ヒストグラムとするクラスタを選択すればよい。

図 7 より、次の式が成り立つ。

$$\begin{aligned} h^2 &= d_{R1}^2 - \left(\frac{1}{2}d_{12} - d_\theta\right)^2 \\ &= d_{R2}^2 - \left(\frac{1}{2}d_{12} + d_\theta\right)^2 \end{aligned} \quad (4)$$

式 (4) の後半を解くことによって、 d_θ が得られる。

$$d_\theta = \frac{d_{R2}^2 - d_{R1}^2}{2d_{12}} \quad (5)$$

参照クラスタに隣接しないクラスタの場合、参照ヒストグラムとクラスタとの最短距離は、式 (5) で計算される値 d_θ よりも必ず大きくなる。すなわち、 d_θ が選択しきい値 θ_2 以下となるクラスタを選択すれば、選択すべきクラスタを選択し損なうことはない。

よって、以上の議論より、

$$\theta_2 \geq \frac{d_{R2}^2 - d_{R1}^2}{2d_{12}} \quad (6)$$

を満たすクラスタを選択すればよい。

3.4.3 類似値上限・下限に基づくしきい値設定

本手法では、二つの独立なしきい値、探索しきい値と選択しきい値を設定する必要がある。探索しきい値と選択しきい値との関係により、理論的に精度が保証される範囲が変化する。そこで以下では、ヒストグラム重なり率と L_2 距離との間に成り立つ関係に基づいて、二つのしきい値と理論的に精度が保証される範囲との関係を導き、所望の保証範囲を与えるように探索しきい値 θ_1 から選択しきい値 θ_2 を決定する指針を与える。

ここで「精度を保証する」とは、蓄積信号の中から参照信号との類似値（距離値）がしきい値を超える（しきい値未満となる）箇所をすべて検出し、しきい値以下となる（しきい値以上となる）箇所を一つも検出しないことを意味する。

ヒストグラム重なり率（式 (1)）は、 L_1 距離

$$d_1 = d_1(H_R, H_S) \stackrel{\text{def}}{=} \sum_{i=1}^L |h_{Ri} - h_{Si}| \quad (7)$$

を用いて、以下のように表すことができる。

$$S_1 = S_1(H_R, H_S) = 1 - \frac{1}{2D} d_1 \quad (8)$$

ただし、式 (1) から式 (8) への変形では、

$$\min(x, y) = \frac{1}{2} \{ (x + y) - |x - y| \}$$

という関係を用いている。また、 L_1 距離と L_2 距離との間には、一般的に次の関係が成り立つ。

$$d_2 = d_2(H_R, H_S) \leq d_1 \leq \sqrt{L} d_2. \quad (9)$$

式 (8)(9) より、

$$\frac{2D}{\sqrt{L}} (1 - S_1) \leq d_2 \leq 2D (1 - S_1) \quad (10)$$

が得られる。ここで、選択しきい値 θ_2 が

$$\theta_2 \geq 2D (1 - \theta_1)$$

を満たすとき、式 (10) より、 $S_1 \geq \theta_1$ を満たすすべてのヒストグラムは $d_2 \leq \theta_2$ を満たす。すなわち、グローバルな枝刈りを行っても $S_1 \geq \theta_1$ を満たすヒストグラムをもしも選べなくも選べる。対して、選択しきい値

θ_2 が

$$\theta_2 \leq \frac{2D}{\sqrt{L}} (1 - \theta_1)$$

を満たすとき、式 (10) より、 $d_2 \leq \theta_2$ を満たすすべてのヒストグラムは $S_1 \geq \theta_1$ を満たす。すなわち、ローカルな枝刈りを行っても $d_2 \leq \theta_2$ を満たすヒストグラムは必ず検出される。

探索しきい値 θ_2 を小さくするにつれて、グローバルな枝刈りによって探索範囲が大幅に絞り込まれ、探索速度の向上が期待される。しかし、探索範囲の削減が過剰になると、探索漏れが発生しやすい状況を作り出してしまふ。更に、精度保証範囲が L_2 距離基準の精度保証範囲に近づくため、 L_2 距離がヒストグラム重なり率と比較して探索精度が低いという事実 (2.2) から、探索精度が低下することが予想される。

そこで本論文では、パラメータ p を導入することによって、探索しきい値から選択しきい値を決定する式を記述し、 p を変化させることによって、探索速度と探索精度の関係を調べる。選択しきい値の決定式を以下に示す。

$$\theta_2 = \frac{2D}{(\sqrt{L})^p} (1 - \theta_1). \quad (11)$$

図 8 は、 p の変化によって精度が保証される範囲がどのように変化するかを示している。図 6 及び図 7 同様、簡単のためヒストグラム空間を 2 次元で表現している。

参照ヒストグラムとの類似値（ヒストグラム重なり率の値）が探索しきい値 θ_1 を超える領域は、 L 次元空間における正 2^L 面体を形成する。2 次元の場合、

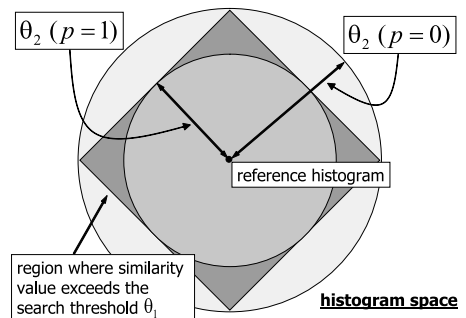


図 8 半径パラメータ p による選択しきい値 θ_2 の変化と探索しきい値 θ_1 との関係

Fig.8 Relationship between the search threshold and the selection threshold when $p = 0$ and $p = 1$.

図 8 にあるように、正方形になる．ここで $p = 0$ のとき、 $\theta_2 = 2D(1 - \theta_1)$ となり、参照ヒストグラムとの (L_2) 距離値が選択しきい値 θ_2 未満となる領域は、参照ヒストグラムとの類似値が探索しきい値を超える領域にちょうど外接する超球を形成する．すなわち、 $p = 0$ における選択しきい値は、 L_1 距離のもとで精度を保証する最小の選択しきい値である．また $p = 1$ のとき、 $\theta_2 = \frac{2D}{\sqrt{L}}(1 - \theta_1)$ となり、参照ヒストグラムとの距離値が選択しきい値未満となる領域は、参照ヒストグラムとの類似値が探索しきい値を超える領域にちょうど内接する超球を形成する．

以上の考察より、 p は参照ヒストグラムとの距離値が選択しきい値未満となる領域が形成する超球の半径を制御しているため、以降、 p を半径パラメータと呼ぶ．

4. 実 験

4.1 しきい値設定

提案法の有効性を示すため、200 時間分の音響信号データから、特定の 15 秒の音響信号を探索するのに要する時間と探索精度を測定した．実験に用いた計算機の CPU は、Pentium III 966 MHz とした．

まず、テレビ放送から音響信号を取り込み、MP3 形式にエンコードして外部記憶装置に記録した．ここで、MP3 形式を用いたのは、原信号に比べ記憶容量を小さくするためである．取込みは蓄積信号用として 200 時間分を 1 回取り込んだほか、参照信号用としてこれとは別のテレビ局の放送から音響信号を同様に取り込み、無作為に 15 秒の異なる CM を 1000 個選択して記録した．いずれの場合も、原信号の取込みは標準化周波数 32 kHz、量子化精度 16 bit 直線、モノラルで行い、MP3 符号化はビットレート 56 bit/s で行った．

取り込んだ各音響信号から [6] と同様の方法で特徴抽出を行った．特徴ベクトルの時間幅は 60 ms、特徴ベクトルの時間刻みは 10 ms とした．各特徴ベクトルをあらかじめ作成されたサイズ L の符号帳を用いてベクトル量子化し、ヒストグラムを作成した．蓄積ヒストグラム系列は、3.2 及び 3.3 で述べた処理により C 個のクラスタに分類した．探索しきい値 θ_1 は $\theta_1 = 0.85$ とした．

半径パラメータ p の変動による探索精度と探索速度との変化を評価するため、 p を 0 から 2 まで変化させて、探索に要する時間と探索精度を測定した．本節において、ヒストグラムのピン数 $L = 128$ 、クラスタ数

$C = 1024$ とした．

なお以下の議論において、時間はいずれも CPU 時間で測定し、1000 個の参照信号について測定を行った平均値を示している．

(1) 特徴抽出時間

200 時間分の蓄積信号と 15 秒の参照信号から MP3 のデコードを行いながら特徴抽出を行うのに要する CPU 時間は、7 時間 34 分 35 秒 (再生時間の約 4%) であった．

(2) ベクトル量子化時間

200 時間分の蓄積信号と 15 秒分の参照信号から抽出された特徴ベクトル系列のベクトル量子化に要する CPU 時間は、12 分 30 秒 (再生時間の約 0.1%) であった．これは、すべての特徴ベクトルをメモリにロードしてからメモリ上で処理する時間を測定したものである．

(3) ヒストグラム分類時間

200 時間分の蓄積信号と 15 秒分の参照信号から作成されたヒストグラム系列の分類に要する CPU 時間は、1 時間 30 分 37 秒 (再生時間の約 0.8%) であった．これは、すべてのベクトル量子化後の特徴ベクトル系列をメモリにロードし、ヒストグラムを順次作成していきながらメモリ上で処理する時間を測定したものである．

(4) 枝刈り時間

グローバルな枝刈りとは、参照ヒストグラムからの距離が最小となるクラスタの重心を探すこと、選択条件の判別式 (6) を計算すること、及び分類されているヒストグラムを選択することである．上記手順の計算量はヒストグラムの次元数 L 及びクラスタ数 C のみに依存する．よって、枝刈り時間は半径パラメータ p の大小によらず一定値を示す．

本節の実験 ($L = 128$, $C = 1024$) において、枝刈り時間は 0.02 秒であった．後に示すように、これは探索実行時間の約 5~50% になる．しかし、あらかじめ分類されたヒストグラムの選択という操作を無視すれば、枝刈り時間は蓄積信号の長さにもよらず一定であるので、蓄積信号が更に長時間になれば無視できる．

(5) 探索実行時間

実験結果を図 9 に示す．

図 9 は、時系列アクティブ探索法と提案法の探索実行時間を比較した結果を示している．ここで、水平軸は半径パラメータ p 、垂直軸は探索実行時間 (search time, 左) と探索時間短縮比 (time reduction rate,

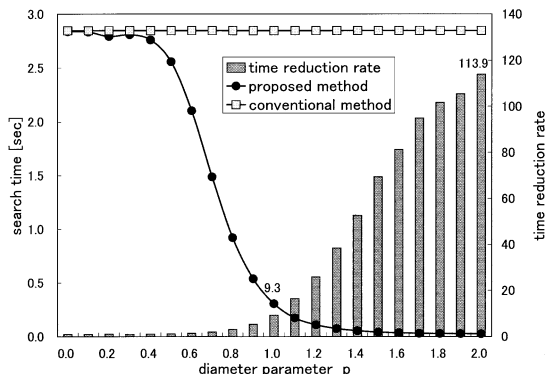


図 9 半径パラメータと 200 時間分の蓄積音響信号の探索に要する探索時間との関係

Fig. 9 Relationship between the diameter parameter and search time for a stored audio signal of 200 hours.

右)である。探索時間短縮比とは、時系列アクティブ探索法に対して提案法がどの程度探索実行時間を削減できたかを示すもので、提案法における探索実行時間に対する時系列アクティブ探索法における探索実行時間の比で定義される。

図 9 に示されるように、提案法は p が 0.4 より大きいとき、半径パラメータが大きくなるにつれて探索実行時間を短縮していき、 $p = 1.0$ のときには約 0.3 秒で、 $p = 2.0$ のときには約 0.025 秒で探索が実行されている。このとき、提案法は時系列アクティブ探索法と比較してそれぞれ約 9 倍 ($p = 1.0$)、約 110 倍 ($p = 2.0$) 高速な探索を実現している。

$p = 0$ から $p = 0.4$ にかけては、探索法の探索実行時間が p の増加に伴って単調に減少しないことがある。この現象は次のように説明される。

半径パラメータ p を大きくしていくと、探索されるべき蓄積信号の量は単調に減少するが、蓄積信号全体に対して探索されるべき蓄積信号の占める割合が比較的大きい場合には、探索されるべき区間が短い間隔で多数存在することがしばしば起こり得る。この状態で各区間ごとに時系列アクティブ探索法に基づく照合及びスキップを行うと、従来の時系列アクティブ探索法ではスキップできていた箇所が探索されるべき区間の切れ目に当たる確率が高くなり、かえってスキップ幅を縮めてしまうことがある。

探索に要する時間以外に、

(6) 特徴ベクトルのベクトル量子化に用いる符号帳の作成に要する時間 (符号帳作成時間)

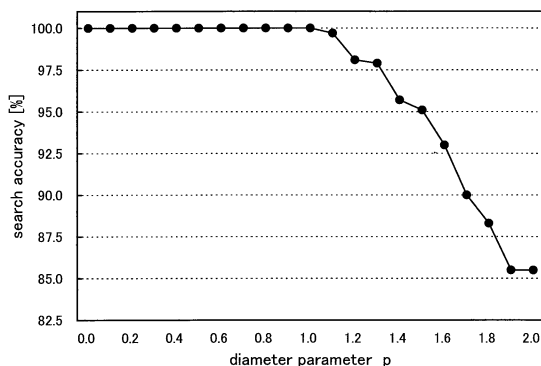


図 10 半径パラメータと探索精度との関係

Fig. 10 Relationship between the diameter parameter and search accuracy.

(7) ヒストグラム空間の分割に要する時間 (ヒストグラム空間分割時間)

が必要となる。しかし、事前に様々な種類の信号から十分な数だけサンプリングして符号帳の作成及びヒストグラム空間の分割を行えば、これらの処理は蓄積信号ごとにやり直す必要がない。よって本論文では、符号帳作成時間及びヒストグラム空間分割時間は探索に要する時間に含めていない。

(8) 探索精度

実験結果を図 10 に示す。

図 10 は、提案法の探索精度を測定した結果を示している。ここで、水平軸は半径パラメータ p 、垂直軸は探索精度 (search accuracy) である。探索精度は時系列アクティブ探索法の探索結果を正解としたときの適合率によって評価した。

図 10 に示されるように、 $p = 1.0$ を境に探索精度が低下し始め、 $p = 2.0$ において 85% まで精度が悪化している。

4.2 探索パラメータ

ヒストグラムの次元数 L 及びクラス数 C と、探索精度及び探索速度との関係を調べるため、 L 及び C を変化させて探索に要する時間と探索精度を測定した。なお、本節において、使用した音響信号及び特徴抽出パラメータは前節の実験と同一である。探索しきい値 θ_1 は、 $\theta_1 = 0.85$ ($L = 128$)、 $\theta_1 = 0.8$ ($L = 256$)、 $\theta_1 = 0.75$ ($L = 512$) とした。半径パラメータは $p = 1$ とした。

(1) 特徴抽出時間 (2) ベクトル量子化時間及び (3) ヒストグラム分類時間は前節と同一なので、省略

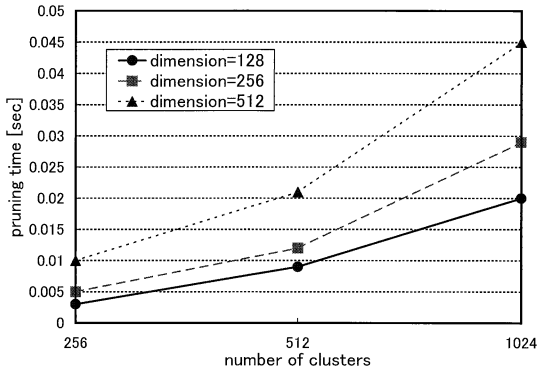


図 11 200 時間分の蓄積音響信号に対するクラスタ数と枝刈り時間の関係
 Fig. 11 Relationship between the number of clusters and pruning time for a stored audio signal of 200 hours.

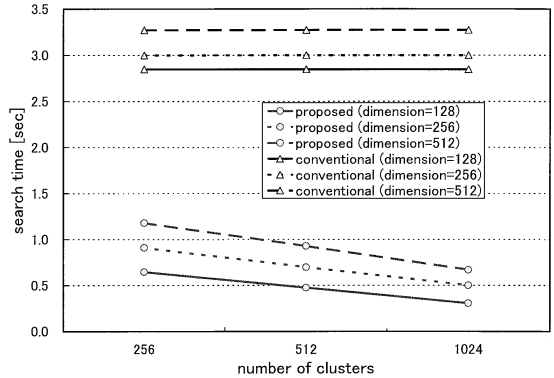


図 12 200 時間分の蓄積音響信号に対するクラスタ数と探索時間との関係
 Fig. 12 Relationship between the number of clusters and search time for a stored audio signal of 200 hours.

する。

(4) 枝刈り時間

実験結果を図 11 に示す。

図 11 は、ヒストグラムの次元数をパラメータとして、提案法の枝刈り時間を測定した結果を示している。ここで、水平軸はクラスタ数、垂直軸は枝刈り時間である。

図 11 に示されるように、各次元数について枝刈り時間はクラスタ数に対しほぼ線形に増加している。また、次元数が増加すると枝刈り時間も増加する傾向がある。例えば、次元数が 128、クラスタ数が 1024 のとき、枝刈り時間は 0.02 秒である。

(5) 探索実行時間

実験結果を図 12 に示す。

図 12 は、ヒストグラムの次元数をパラメータとして、時系列アクティブ探索法と提案法の探索実行時間を比較した結果を示している。ここで、水平軸はクラスタ数、垂直軸は探索実行時間(左)と探索時間短縮比(右)である。

図 12 に示されるように、提案法はクラスタ数が大きくなるにつれて探索実行時間を短縮しており、クラスタ数が 1024 のときにはいずれの次元数においても 1 秒以下で探索が実行されている。また、いずれの手法においても、次元数が大きくなるにつれて探索実行時間が増大する傾向にある。

(6) 探索精度

実験結果を図 13 に示す。

図 13 は、提案法の探索精度を測定した結果を示し

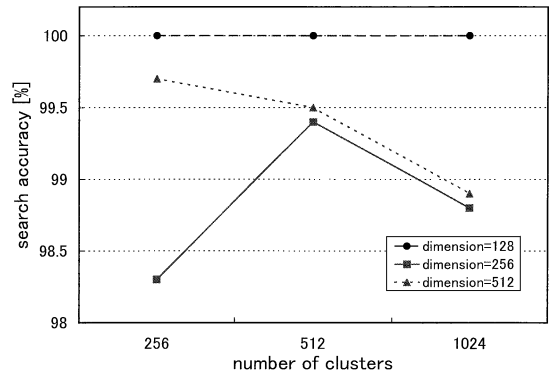


図 13 クラスタ数と探索精度との関係
 Fig. 13 Relationship between the number of clusters and search accuracy.

ている。ここで、水平軸はクラスタ数、垂直軸は探索精度である。探索精度は時系列アクティブ探索法の探索結果を正解としたときの適合率によって評価した。

図 13 に示されるように、いずれの次元数においても 98% 以上の高い探索精度を示しており、特に次元数が 128 のときにはいずれのクラスタ数においても 100% の探索精度、すなわち時系列アクティブ探索法と同一の探索精度を示している。

これらの結果で良好な性能を示した例として、ヒストグラムの次元数 $L = 128$ 、クラスタ数 $C = 1024$ 、半径パラメータ $p = 1$ としたときの測定結果について、表 1 に示す。

4.3 映像信号

提案法が映像信号に対しても適用可能であることを

表 1 200 時間分の蓄積音響信号に対する実験結果
Table 1 Experimental results for a 200-hour stored audio signal. (summary)

探索実行時間		探索時間削減比
提案法	従来法	
306 ms	2847 ms	9.3
照合回数		照合回数削減比
提案法	従来法	
13256 回	112120 回	8.5

表 2 24 時間分の蓄積映像信号に対する実験結果
Table 2 Experimental results for a 24-hour stored video signal. (summary)

探索実行時間		探索時間削減比
提案法	従来手法	
4.2 ms	90.0 ms	21.4
照合回数		照合回数削減比
提案法	従来手法	
1153 回	8597 回	7.5

示すため、24 時間分の映像信号データから、特定の 15 秒の映像信号を探索するのに要する時間と探索精度を調査した。

まず、家庭用ビデオデッキで録画した 24 時間分のテープ (VHS HiFi, 3 倍モード) を再生して、映像をワークステーションに取り込んだ。取込みは、蓄積信号用として 24 時間分を 1 回取り込んだほか、参照信号用として同じテープから無作為に 15 秒の異なる信号を 10 本選択して再生し、入力信号用とは別に取り込んだ。いずれの場合も、取込みはフレームレート 29.97 Hz, Motion JPEG, 画面サイズ 320 × 240 で行った。取り込んだ各映像信号から [10] と同様の方法で特徴抽出を行った。特徴抽出の分割数は $W = 6$ (横方向 3 等分割, 縦方向 2 等分割) とした。

ヒストグラムの次元数は $L = 256$, クラスタ数は $C = 1024$ とした。また、探索しきい値 θ_1 は、時系列アクティブ探索法を用いたときに適合率と再現率がともに 100% になるように、 $\theta_1 = 0.8$ とした。半径パラメータは $p = 1$ とした。

探索実行時間及び照合回数について、測定結果を表 2 に示す。探索実行時間及び照合回数は、10 本の参照信号について測定を行った平均値を示している。なお、本節の実験において、10 本の参照信号すべてに対して探索結果は正しいことを確認した。この結果より、提案法は映像信号に対しても有効であることが示された。

5. むすび

本論文では、時系列アクティブ探索におけるローカルな枝刈りに加え、新たにグローバルな枝刈りを導入した探索手法を提案した。グローバルな枝刈りにおいて、半径パラメータ p を導入することによって、探索しきい値から選択しきい値を決定し、 p を大きくしていくにつれて、探索精度を保ったまま探索時間及び照合回数が削減されることを実験により示した。また、クラスタ数を大きくしていくにつれて、探索精度を保ったまま探索実行時間が削減されることを実験的に示した。特に、ヒストグラムの次元数を 128, クラスタ数を 1024 としたとき、 $p = 1$ で探索時間を約 9 分の 1 に削減することができた。更に、提案法は時系列アクティブ探索法同様、音にも映像にも適用可能であることを示した。映像の場合、ヒストグラムの次元数を 256, クラスタの数を 1024 としたとき、 $p = 1$ で探索時間を約 20 分の 1 に削減することができた。

今後は、探索精度を保ったまま、より膨大な時系列信号をより高速に探索することができる手法について検討を進める予定である。

謝辞 日ごろ御指導を頂く NTT コミュニケーション科学基礎研究所の石井健一郎所長及び管村昇部長に感謝する。また、日ごろ御協力を頂く同研究所メディア認識研究グループの諸氏に感謝する。

文 献

- [1] S. Dharanipragada and S. Roukos, "A fast vocabulary independent algorithm for spotting words in speech," Proc. ICASSP98, vol.1, pp.233-236, May 1998.
- [2] D.A. James and S.J. Young, "A fast lattice-based approach to vocabulary independent word spotting," Proc. ICASSP94, vol.1, pp.377-380, April 1994.
- [3] 小杉尚子, 西原祐一, 紺谷精一, 山室雅司, 串間和彦, "ハミングを用いた音楽検索システム," 信学技報, DE99-18, July 1999.
- [4] H.D. Wactlar, "Informedia — Search and summarization in the video medium," Proc. Imagina 2000 Conference, Jan.-Feb. 2000.
- [5] R. Mohan, "Video sequence matching," Proc. ICASSP98, vol.6, pp.3697-3700, May 1998.
- [6] 柏野邦夫, ガビンスミス, 村瀬 洋, "ヒストグラム特徴を用いた音響信号の高速探索法," 信学論 (D-II), vol.J82-D-II, no.9, pp.1365-1373, Sept. 1999.
- [7] N. Beckman and H.P. Kriegel, "The R⁺-tree: An efficient and robust access method for points and rectangles," Proc. ACM SIGMOD 90, pp.322-331, 1990.
- [8] D.A. White and R. Jain, "Similarity indexing with the SS-tree," Proc. ICDE96, pp.516-523, Feb.-March

1996.

- [9] 片山紀生, 佐藤真一, “SR-Tree: 高次元データに対する最近接探索のためのインデックス構造の提案,” 信学論(D-I), vol.J80-D-I, no.8, pp.703-717, Aug. 1997.
- [10] 柏野邦夫, 黒住隆行, 村瀬 洋, “ヒストグラム特徴を用いた音や映像の高速 AND/OR 探索,” 信学論(D-II), vol.J83-D-II, no.12, pp.2735-2744, Dec. 2000.
- [11] V.V. Vinod and H. Murase, “Focused color intersection with effective searching for object extraction,” Pattern Recognit., vol.30, no.10, 1997.
- [12] A. Garsho and R.M. Gray, Vector quantization and signal compression, Kluwer Academic, MA, USA, 1992.
- [13] 杉山雅英, “セグメントの高速探索法,” 信学技報, SP98-141, Feb. 1999.

(平成 13 年 11 月 30 日受付, 14 年 3 月 25 日再受付)



村瀬 洋 (正員)

昭 53 名大・工・電子卒・昭 55 同大学院修士課程了。同年日本電信電話公社(現日本電信電話(株))入社。以来,文字・図形認識,コンピュータビジョン,マルチメディア認識の研究に従事。平 4 から 1 年間米国コロンビア大客員研究員。現在,日本電信電話(株)NTT コミュニケーション科学基礎研究所メディア情報研究部長。工博。昭 60 本会学術奨励賞,平 4 電気通信普及財団テレコムシステム技術賞,平 6 IEEE-CVPR 国際会議最優秀論文賞,平 7 情報処理学会山下記念研究賞,平 8 IEEE-ICRA 国際会議最優秀ビデオ賞,平 13 電子科学技術振興財団高柳記念奨励賞,平 13 本会ソサイエティ論文賞,平 14 本会業績賞各受賞。情報処理学会,IEEE 各会員。



木村 昭悟 (正員)

平 10 東工大・工・電気電子卒。平 12 同大学院電気電子工学専攻修士課程了。同年日本電信電話(株)に入社。現在,日本電信電話(株)NTT コミュニケーション科学基礎研究所に所属。メディア情報認識・探索の研究に従事。情報理論,学習理論に

興味をもつ。情報理論とその応用学会,日本音響学会,IEEE 各会員。



柏野 邦夫 (正員)

平 2 東大・工・電子卒。平 7 同大学院電気工学専攻博士課程了。工博。同年日本電信電話(株)に入社。現在,日本電信電話(株)NTT コミュニケーション科学基礎研究所主任研究員。音響信号の認識・分離・探索,及び情報統合の研究に従事。メ

ディア情報を対象とする信号処理及び知識処理に興味をもつ。平 5 情報処理学会奨励賞,平 9 人工知能学会全国大会優秀論文賞,平 11 日本音響学会粟屋潔学術奨励賞,平 13 本会ソサイエティ論文賞,平 14 本会業績賞各受賞。情報処理学会,日本音響学会,人工知能学会,日本音楽知覚認知学会,IEEE 各会員。



黒住 隆行 (正員)

平 9 都立大・理・物理卒。平 11 北陸先端科学技術大学院大学情報科学研究科博士前期課程了。同年日本電信電話(株)に入社。現在,日本電信電話(株)NTT コミュニケーション科学基礎研究所に所属。パターン認識,画像処理に興味をもつ。平 14 本

会業績賞受賞。