

音や映像を瞬時に探す時系列アクティブ探索法

Time-series Active Search : Instantaneous Searching of Audio and Video

柏野 邦夫[†]

Kunio KASHINO

村瀬 洋[†]

Hiroshi MURASE

あらまし

放送など長時間の音や映像中から、特定の番組タイトルやCMなど、目的とする音や映像の有無および時刻を高速に探索する方法を開発した。従来の、特徴ベクトルのずらし照合に基づく探索では、長時間の音や映像を探索する場合に計算量が膨大となるという問題があった。これに対し今回開発した「時系列アクティブ探索法」では、特徴ベクトルのヒストグラムに基づいて探索を行うことにより、大幅に計算時間を短縮することができる。ワークステーションを用いた実験では、あらかじめ特徴ベクトルを抽出しておいた場合、6時間の信号から15秒の目的部分を探索するのにかかる時間は、音の場合約2.3秒、映像の場合約1.0秒であり、しかも探索結果は正確であった。また、音や映像に対し白色カウス雑音を重畳した場合の頑健性も確認した。

Abstract

This paper proposes a search method that can quickly detect and locate a desired audio or video segment in a long audio or video stream. The method, called time-series active search, is based on histogram modeling of audio or video features. Histogram modeling enables us to reduce the number of candidate matches between reference and input signals by approximately 100 to 500 times compared to exhaustive search, while guaranteeing the same retrieval accuracy. Experiments show that the proposed method can correctly retrieve a 15-s segment from a 6-h TV broadcast-quality recording within 2.3 s (for audio) or 1.0 s (for video), once the feature vectors have been extracted.

い ま え が き

近年、音や映像のデータが、身の回りに大量に流通するようになってきた。このため、音や映像の検索技術や探索技術の必要性が増している。ここで検索とは、求める音や映像の内容に関するなん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ことをいい、内容検索とも呼ばれる。音や映像の内容検索に関しては数多くの研究が報告されてきている^{1)~4)}。一方、探索とは、具体的な音や映像（参照信号）を指定して、それらがデータベースや長時間の素材等（入力信号）のどこに存在するかを探さることをいう。

これらのうち、本論文で着目するのは探索の技術である。テキストデータのハンドリングにおいて、高速な文字列探索アルゴリズムが重要な役割を果たしているのと同様、マルチメディア

データのハンドリングにおいても、高速な時系列探索アルゴリズムは重要であると考えられる。例えばインターネットにおいて、音楽や映像等の著作物の不正使用を抑止するために、高速な時系列探索法が求められている。また、高速な時系列探索法があれば、長時間のテレビ放送データから、番組タイトルやコマーシャル（CM）など特定の音や映像の放送日時を短時間でピックアップすることなども可能となる。

もちろん、音や映像の探索自体は、従来技術であるスペクトルや波形などのずらし照合で解決できる問題である。また、既知の信号の検出は、信号検出理論として体系化も行われている⁵⁾。しかし従来法では、長時間の入力信号や、多数の参照信号を対象とする場合には、計算量が膨大となるため実用的な処理時間で漏れなく検出あるいは探索することは難しかった。高速化のために、照合の仕方や時間方向のずらし方を粗くすることも考えられるが、その場合は探索漏れの発生など探索精度の

[†] NTTコミュニケーション科学基礎研究所 Communication Science Laboratories, NTT

低下が避けられないという問題があった。

そこで著者らは、「時系列アクティブ探索法」(time-series active search)と呼ぶ音や映像の高速・高精度の探索技術を開発した¹⁰⁾。時系列アクティブ探索法は、特徴ベクトルの量子化符号のヒストグラムに基づくことを特徴とする。これにより、全探索と同じ精度を保証したままで、実際に行う照合計算の回数を全探索に比べ1/100から1/500程度に削減することができ、また照合計算自体も、特徴を直接比較するよりも計算コストが小さい方法を用いることができる。これらの効果により、従来のずらし照合に比べきわめて高速に探索を行うことができる。

以下2章において、時系列アクティブ探索法の概要を説明する。3章において、提案法の探索速度と探索精度に関して評価実験を行う。4章をむすびとする。

2 時系列アクティブ探索法

2.1 手法の概要

図1に、時系列アクティブ探索法の概要を示す。まず参照信号と入力信号からそれぞれ特徴ベクトルを抽出する。次に、参照信号と入力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ベクトル量子化して、各量子化符号の出現回数を計数してヒストグラムをつくる。そして、ヒストグラムどうしの類似度が、あらかじめ設定した値(これを探索しきい値と呼ぶ)を超えるかどうかで、参照信号の有無を判定する。このとき後述のように、類似度の値と設定値とから、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るので、その分だけ入力信号に対する窓をずらして探索を進める。

このアルゴリズムは、特定の特徴ベクトルやその量子化の仕方に依存するものではなく、それらについては各種のモデルを用いることができる。また、ヒストグラムどうしの類似度の定義についても様々なものが考えられる。

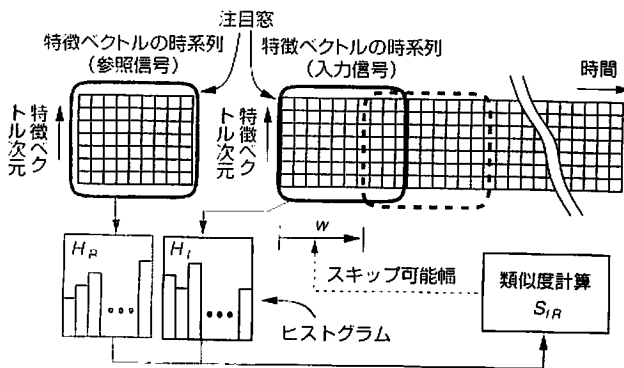


図1 時系列アクティブ探索法の概要

2.2 類似度とスキップ可能幅

ヒストグラムどうしの類似度として、著者らは特にヒストグラム重なり率に着目している。ヒストグラム H_I と H_R の重なり率 S_{IR} は、次のように定義される。

$$S_{IR} = S(H_I, H_R) = \frac{1}{D} \sum_{l=1}^L \min(h_{I_l}, h_{R_l}) \quad (1)$$

ここで H_I と H_R は、それぞれ入力信号と参照信号に対するヒストグラムであり、 h_{I_l} 、 h_{R_l} はそれぞれの l 番目の符号に量子化された特徴ベクトルの数である。また L は量子化符号帳のサイズ、 D は参照信号から導かれた特徴ベクトルの総数である。

本論文でヒストグラム重なり率による類似度の定義を用いる理由は、①類似度計算が簡単であること、②時間窓のスキップ可能幅が後述のように簡単な計算によって求められること、および③すでに画像の物体認識などに適用され、好ましい結果が得られていること⁽¹⁰⁾⁻⁽¹²⁾、の3点である。なお、時系列アクティブ探索法における類似度の定義と類似度の上限値に関して、杉山が系統的に考察している⁽¹³⁾。

さて、時間窓のスキップが可能であるのは次の理由による。ヒストグラムは特徴ベクトルの時系列を分類し累積したものであるから、入力信号の特徴ベクトルに対する時間窓の移動に伴って、式(1)の類似度が急激に変化することはない。例えば、時間窓が一単位移動したとき、最も急に類似度が増加するのは、今まで類似度(本論文の場合、参照信号のヒストグラムとの重なり)に寄与していなかった特徴ベクトルが1つ時間窓外に出て、新たに時間窓内に入ってきた特徴ベクトルが類似度に寄与する場合である。このことから、式(1)における類似度 S の変化率(時間窓の一単位の移動における S の増分)の絶対値は、決して $1/D$ を超えないことが分かる。

すなわち、入力信号に対する時間窓の先頭が n_1 番目の特徴ベクトルであるときの類似度を $S_{IR}(n_1)$ とすると、時間窓が n_2 番目の特徴ベクトルまで移動したときの類似度の上限値 $S_{IR}^*(n_2)$ は、 $n_1 < n_2 < n_1 + D$ のとき

$$S_{IR}^*(n_2) = S_{IR}(n_1) + \frac{n_2 - n_1}{D} \quad (2)$$

で与えられる。ただし、式(1)の定義から S は1を超えないので、式(2)で与えられる $S_{IR}^*(n_2)$ が1を超えるときは、1が上限値となる。

ところで、いま著者らは、類似度が探索しきい値 θ を超える箇所を探し出そうとしている。したがって、式(2)で与えられる上限値が θ 以下となる区間に対しては、照合を行う必要がない。そこで、式(2)において、上限値を θ で置き換え、 $n_2 - n_1$ を w と置いて整理することにより、スキップ可能幅 w を求めることができる。

$$w = \begin{cases} \text{floor}(D(\theta - S_{IR})) + 1 & (S_{IR} < \theta) \\ 1 & (\text{上記以外}) \end{cases} \quad (3)$$

ただし $\text{floor}(\cdot)$ は切下げを表し、 θ は探索しきい値である。類似度が θ を超える箇所については全探索を行う（時間窓を一単位ずつずらす）こととしている。

時系列アクティブ探索法は、類似度を計算することに w を求め、時間窓を特徴ベクトル w 個分だけずらして、また類似度を計算するという操作の繰返して構成される。この操作により、類似度が低い時点では大きくスキップし、類似度が高い時点では緻密に照合するという適応的な探索動作が、探索しきい値を設定するだけで自動的に行われることになる。このとき重要なのは、経験則によるスキップなどとは異なり、上記の議論から明らかのように、類似度が探索しきい値を超える箇所は漏れなく探索できることが保証されている点である。

なお、現在著者らが実装している探索システムの最終的な出力としては、類似度が探索しきい値を超えた箇所をすべて出力するのではなく、類似度が探索しきい値を超えた箇所のうちで、類似度が極大となっている時点のみを出力するようにしている。

2.3 時間順序の考慮

参照信号および入力信号に対する時間窓を N_{div} 個に分割し、分割した時間窓についての類似度の最小値を全体の類似度と定義することにより、信号の時間順序を考慮した探索を行うことができる。この場合のスキップ可能幅は、個々の分割された時間窓についてのスキップ可能幅のうちの最大値とすればよい。

2.4 音響特徴

時系列アクティブ探索法における特徴ベクトルの条件としては、判別性能がよい（目的部分における類似度が高く、目的以外の部分における類似度が低い）こと、頑健である（録音・録画の条件やノイズ等に影響されにくい）こと、および特徴抽出に必要な計算量が過大でないことが要求される。

これらの点を考慮して、音響特徴ベクトルとしては、帯域通過フィルタで求めた短時間パワースペクトルを周波数チャンネルに関して正規化したものを用いることができる。すなわち音響特徴ベクトル $f(k)$ は、

$$f(k) = (f_1(k), f_2(k), \dots, f_V(k)) \quad (4)$$

と書くことができる。ここで k は時刻、 V は特徴の次元数である。 $f(k)$ の各要素は、正規化された短時間パワースペクトルである。すなわち

$$f_j(k) = \alpha(k) Y_j(k) \quad (5)$$

$$Y_j(k) = \sum_{t=k-K+1}^k y_j^2(t) \quad (6)$$

$$k = lM \quad (l = 1, 2, \dots) \quad (7)$$

である。ここで $y_j(t)$ は時刻 t における帯域通過フィルタ j の出力波形、 M は特徴ベクトルの時間刻み、 K は特徴ベクトルの計算に用いる時間区間の長さである。また $\alpha(k)$ は正規化のための係数であり、

$$\alpha(k) = \frac{1}{\max_j (Y_j(k))} \quad (8)$$

と定義される。

帯域通過フィルタ $y_j(t)$ は、例えば2次のIIRフィルタで実装することが可能である。このほか特徴ベクトルとしては、LPC分析に基づくものや、MFCC（メル周波数ケプストラム係数）などを用いることも可能である。

2.5 映像特徴

本論文では、各種ビデオ機器等の特性差に対する頑健性を考慮して、輝度に着目した。すなわち映像特徴ベクトル $g(k)$ は、

$$g(k) = (g_1(k), \dots, g_w(k)) \quad (9)$$

と書くことができる。ここで k はフレームの時刻であり、 g の添字は各フレームの画像を W 個のサブ画像に分割した分割の番号を表す。 g_j は各画素の輝度値をサブ画像内で平均し正規化した値であり、

$$g_j(k) = \frac{\bar{x}_j(k) - \min_i \bar{x}_i(k)}{\max_i \bar{x}_i(k) - \min_i \bar{x}_i(k)} \quad (10)$$

である。ただし、

$$\bar{x}_i = \mathbb{E}_{p \in I} [x_p(k)] \quad (11)$$

である。ここで I は i 番目のサブ画像内の画素 p の集合であり、 \mathbb{E} は I についての平均を表す。また、画素 p の RGB 値を r_p, g_p, b_p とすると、

$$x_p = 0.299r_p + 0.587g_p + 0.114b_p \quad (12)$$

となる。これは NTSC 方式における RGB 値と輝度値との関係式である。

3 実験

前章で述べた探索法を小型ワークステーションに実装し、探索速度と探索精度に関して実験を行った。実験に用いたワーク

表1 実験に用いたワークステーションの仕様

項目	内容
モデル名	SGI社 O ₂
CPU	R10000 (250 MHz)
メモリ	384 MByte
OS	IRIX Release 6.3
コンパイラ	MIPSPRO C Compiler ver 7.00

ワークステーションの仕様を表1に示す。

3.1 実験1: 探索速度

提案法の探索速度を評価するため、テレビ放送相当の音響信号と映像のデータから、特定の15秒の区間 (CMに相当) を探索するのに要する時間を測定した。

まず、6時間にわたるテレビ放送相当の素材が録画されたビデオテープを用意した (VHS HiFi, 3倍モード)。次に、この録画テープを再生し、記録されている音響信号と映像をそれぞれ上記ワークステーションに取り込んだ。取込みは、入力信号用として6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異なる15秒の区間を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、音響信号の取込みは標準化周波数11.0 kHz、量子化精度8 bit直線、モノラルで行った。映像の取込みはフレームレート29.97 Hz、サイズ80×60画素、非圧縮で行った。

探索に要する時間は、①特徴抽出に要する時間 (特徴抽出時間)、②特徴ベクトルのベクトル量子化に要する時間 (ベクトル量子化時間)、③ベクトル量子化の結果を用いて探索を実行するのに要する時間 (探索実行時間、すなわちヒストグラムの作成、類似度の算出、窓の移動の繰返しに要する時間) の3つからなる。これらを順に議論する。なお、時間はいずれもCPU時間で測定した。CPU時間は測定ごとに数%程度のばらつきがみられたので、以下では、各値とも5回同じ測定を行った平均値を示している。

3.1.1 音響特徴抽出時間

特徴抽出時間は M 、 V 、 K の値によって異なるが、本実験では、6時間分の入力信号と15秒の参照信号から特徴を計算するのに要するCPU時間は約217秒であった。すなわち、実時間の約1%の時間で特徴抽出が可能である。したがって、仮に信号の計算機への取込みと同時に処理を行うとすれば、約1%のCPU負荷で特徴抽出が行える。

3.1.2 映像特徴抽出時間

本実験では、6時間分の入力信号と15秒の参照信号から特徴を計算するのに要するCPU時間は約650秒であった。すなわ

ち、実時間の約3%の時間で特徴抽出が可能である。したがって、仮に信号の計算機への取込みと同時に処理を行うとすれば、約3%のCPU負荷で特徴抽出が行える。

3.1.3 音響特徴ベクトル量子化時間

本実験では、次元ごとのスカラー量子化の組合せによりベクトル量子化を行った。6時間の入力信号と15秒の参照信号に対して、音響特徴ベクトルのベクトル量子化に要するCPU時間は、約1.7秒であった。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、オンメモリで処理する時間を計測したものである。

3.1.4 映像特徴ベクトル量子化時間

6時間の入力信号と15秒の参照信号に対して、映像特徴ベクトルのベクトル量子化に要するCPU時間は、約0.86秒であった。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、オンメモリで処理する時間を計測したものである。

3.1.5 音響探索実行時間

探索実行時間の測定結果を表2に示す。探索実行時間は、参照信号、入力信号、 N_{div} 、および探索しきい値に依存する。表2に示したCPU時間は、10個の参照信号について5回ずつ測定した平均値である。また本実験では、探索しきい値は $\theta = 0.7$ とした。また表2では、提案法における照合回数か全探索の場合に比べ平均でどれだけ削減されたか (照合回数の比) も併せて記した。

なお表2のいずれの場合も、10個の参照信号すべてに対して探索結果は正しい (探索漏れも、余分な探索もなく、探索結果の時間誤差は1秒以内である) ことを確認している。

全探索の場合に比べた提案法における照合回数は、 $N_{div} =$ の場合が最も削減の効果が大きくなっているが、これはヒストグラム1つ当りの時間幅が大きいほど平均的なスキップ可能幅が大きくとれることによると考えられる。また、いずれの場合も実際の速度向上の比が照合回数の比よりも小さくなっている

表2 音響探索実行時間

N_{div}	CPU時間		速度向上	照合回数の比	参照
	全探索*	提案法			
1	35.4 s	0.59 s	60倍	1/444	図2
2	35.3 s	0.65 s	54倍	1/339	図3
4	35.2 s	0.72 s	49倍	1/200	図4
8	35.2 s	0.87 s	40倍	1/109	図5

* 全探索とは、スキップ可能幅 w を1に固定した探索のことである。ただし、その場合も $N_{div} > 1$ のときは各分割ごとに順に照合を行い、探索しきい値未満の類似度が発見された時点で直ちに窓を移動させている。

が、これは、ヒストグラム1つ当りの作成コストが、全探索のように特徴ベクトル1つ分ずつ時間窓がずれていく場合のほうが小さいことと、提案法の場合にはスキップ可能量の算出が必

要なことの2点が理由であると考えられる。

一方、本実験で用いたのと同じの特徴ベクトルを用いて、従来法である、特徴ベクトルどうしの相関を用いたずらし照合法に基づく探索（これを本論文では特徴ずらし照合法と呼ぶ）を行ったところ、すべての特徴ベクトルをメモリにロードした時点から計測して、約685秒のCPU時間を要した。提案法では、例えば $N_{div}=1$ のとき、ベクトル量子化時間と探索実行時間の和は約2.3秒であるので、特徴ずらし照合法に比べて約300倍探索が高速化されたことになる。

なお、提案法においては、特徴抽出とベクトル量子化は、探索に先立ってあらかじめ行っておくことができ、1度行っておけば、次からは行う必要がないという特徴がある。したがって、多数の参照信号について探索を行う場合などでは、特徴ずらし照合法に比べて一層効率的な処理が可能である。例えば、100本の異なる15秒のCMを同一の入力信号に対して探索する場合には、特徴抽出後、上記ずらし照合法では約685秒 \times 100=約68500秒（約19時間）のCPU時間を要する。これに対し提案法では、探索実行時間の100倍にベクトル量子化時間（6時間と1500秒分）を加えた時間があれば探索できる。本実験の結果を用いると、その時間は $N_{div}=1$ のとき約61秒と計算でき、これは特徴ずらし照合法の約1100倍の探索速度に相当する。

参考のため、図2から図5に、表2の各場合における類似度の時間変化パターン（ある同一の参照信号を探索した場合）を示す。各図において、○印は探索された時点を、破線は探索しきい値を示す。 $N_{div}>1$ とすることにより時間情報が加味され、しきい値設定に対するマージンが増大していることが分かる。また図6に、上記特徴ずらし照合法による類似度（参照信号と時間窓内の入力信号における特徴ベクトルの相関値）の時間変化パターンを示す。

3.1.6 映像探索実行時間

映像探索実行時間の測定結果を表3に示す。表3に示したCPU時間は、10個の参照信号について5回ずつ測定した平均値である。また本実験では、探索しきい値は $\theta=0.6$ とした。また表3では、提案法における照合回数が全探索の場合に比べ平均でどれだけ削減されたか（照合回数の比）も併せて記した。

参考のため、図7に、表3に対応する類似度の時間変化パターンを示す。

表3 映像探索実行時間

N_{div}	CPU時間		速度向上	照合回数 の比	参照
	全探索	提案法			
1	22.5 s	0.20 s	112倍	1/207	図7

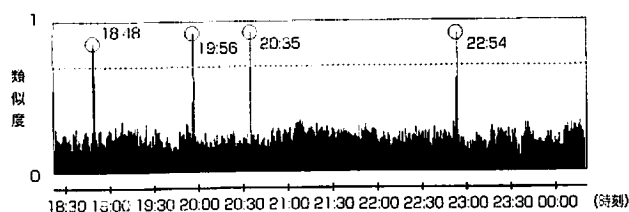


図2 音響探索結果 ($N_{div}=1$)

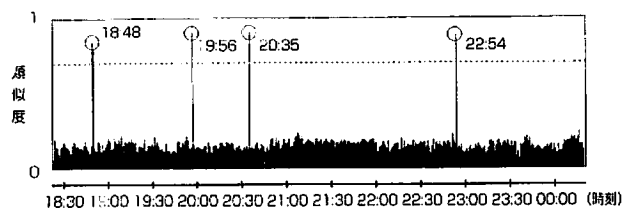


図3 音響探索結果 ($N_{div}=2$)

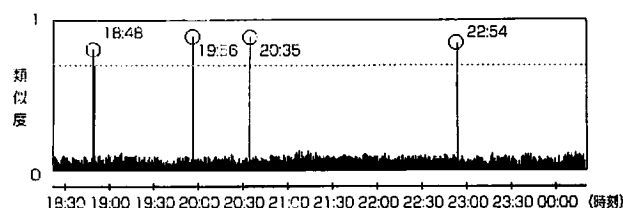


図4 音響探索結果 ($N_{div}=4$)

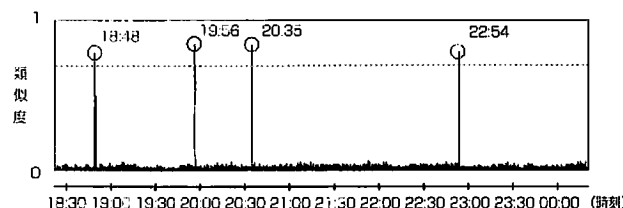


図5 音響探索結果 ($N_{div}=8$)

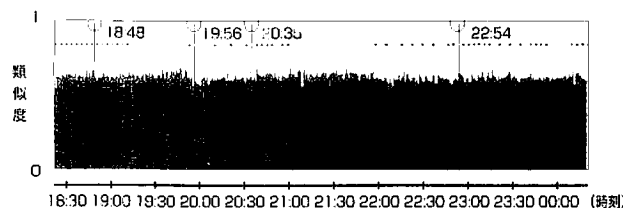


図6 音響探索結果（従来法である特徴ずらし照合法を用いた場合）

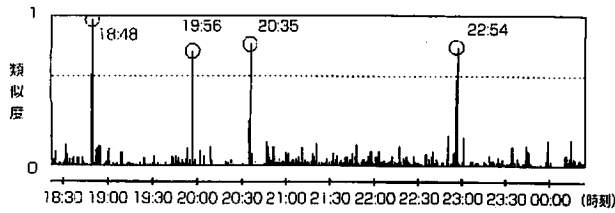


図7 映像探索結果 ($N_{div}=1$)

3.2 実験2：探索精度

提案法の探索精度を調べるため、テレビ放送相当の素材が記録された20分のビデオテープを用いて実験を行った。この20分のビデオテープを再生し、記録されている音響信号と映像を2回に分けてワークステーションに取り込んだ。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、他方を入力信号として探索を行った。入力信号の音や映像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、参照信号の長さ、時間分割数 (N_{div})、およ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、100回繰り返して探索を行い、精度を測定した。精度は、適合率 (precision rate) と再現率 (recall rate) の平均値で評価した。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。適合率と再現率がともに100%であれば、探索漏れも余分な探索もないことを意味する。適合率や再現率は、探索しきい値の設定によって変化するが、本実験では、次式によって探索しきい値を定めた。

$$\theta = m + c\sigma \tag{13}$$

ここで、 m と σ は、それぞれ、与えられた参照信号に対して入力信号をサンプリングし、予備的に類似度の計算を行って収集した類似度値の平均と標準偏差であり、 c は一定の係数である。すなわち、式(13)における c の値を100回の繰り返し中一定とし、その一定値を調節することによって、精度を最大化した値を評価値とした。そのほかの取込みや探索のパラメータは実験1と同様とした。

3.2.1 音響探索精度

音響探索精度の測定結果を図8と図9に示す。

図8は、時間窓の分割数を $N_{div}=1$ とし、参照信号の長さをパラメータとして探索精度を測定したものである。図8に示されるように、参照時間の長さが15秒以上確保できれば、SN比20 dBまで探索漏れも余分な探索も生じていない。また、SN比が30 dB以上であれば、参照信号が2秒であっても、98%程

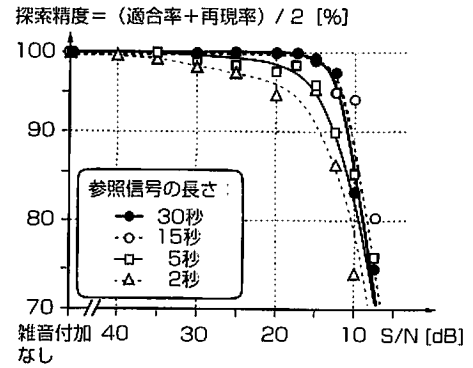


図8 音響探索精度 ($N_{div}=1$ のとき)

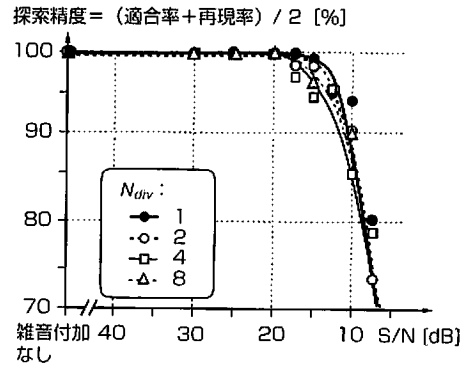


図9 音響探索精度 (参照信号が15秒のとき)

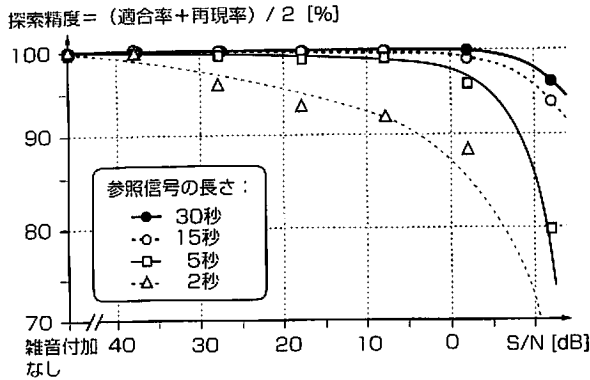
度以上の探索精度が得られることが分かる。

図9は、参照信号の長さを15秒とし、時間窓の分割数 (N_{div}) をパラメータとして探索精度を測定したものである。図2から図5においては、 N_{div} が大きいほど、探索されるべき箇所と探索されるべきでない箇所とにおける類似度値の差 (探索しきい値の-margin) が大きいことが示されていたが、図9によれば、探索精度の値は N_{div} の値にあまり依存していない。すなわち、 N_{div} を大きくしても、雑音が重畳されたときの探索精度を向上させる効果は明らかでない。これは、探索しきい値の-marginの拡大の効果と、雑音の重畳によって入力信号中の正解場所における類似度が低下する効果とが拮抗しているためであると考えられる。

3.2.2 映像探索精度

映像探索精度の測定結果を図10に示す。これによれば、参照信号が15秒間あれば、SN比8 dBまで探索漏れも余分な探索も生じていない。また、SN比が28 dB以上であれば、参照信号が2秒であっても、96%程度以上の探索精度が得られることが分かる。

以上の実験結果から、提案法の精度は、音響信号を用いた場合と映像を用いた場合の双方とも、テレビ放送からの特定の

図10 映像探索精度 ($N_{div}=1$ のとき)

CMの探索などの応用に対しては十分実用的な水準にあると判断することができる。

4 む す び

本論文では、参照信号として具体的な音または映像が与えられたとき、それが長時間の入力信号中のどこに存在するかを高速に探索する手法を提案した。提案法は、各信号の特徴ベクトルをベクトル量子化し、量子化した各符号に対するヒストグラムをつくって、ヒストグラムどうしの照合によって参照信号と入力信号との類似度を判定するというものである。この過程で、ヒストグラムの代数的性質を利用することによって不要な照合計算を省略できる。テレビ放送相当の音と映像を用いて実験を行ったところ、あらかじめ特徴抽出を行っておいた場合、小型ワークステーションを用いて6時間の入力信号から15秒の参照信号を漏れなく探索するのに要する時間は、音を用いた場合約2.3秒、映像を用いた場合約1.0秒であった。また音や映像への白色ガウス雑音の重畳に対しても十分な頑健性を持つことが示された。一方、ベクトル量子化までをあらかじめ行っておいた場合には、上記探索の所要時間は音の場合約0.59秒、映像の場合約0.20秒（いずれも時間窓の分割を行わない場合）である。したがって実用上は、入力信号や参照信号に対して、特徴ベクトルのベクトル量子化までを行った情報を蓄積しておくことにより、効率的な探索が可能である。今後は、さらに高速で適用範囲の広いマルチメディア探索技術の開発を進める予定である。

文 献

(1) J. K. Wu, A. D. Narasimhalu, B. M. Mehtre, C. P. Lam, and Y. J. Gao: "CORE: A Content-based Retrieval Engine for Multimedia

Information Systems," ACM Multimedia Systems, Vol. 3, No. 1, pp. 25-41, 1995.

- (2) M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Yonkani, J. Hafner, D. Lee, D. Petkovic, D. Stedc, and P. Yanker: "Query by Image and Video Content: The QBIC System," IEEE Computer, Vol. 28, No. 9, pp. 23-32, 1995.
- (3) E. Wold, T. Blum, D. Keislar and J. Wheaton: "Content-based classification, search, and retrieval of audio," IEEE Multimedia, Vol. 3, No. 3, pp. 27-36, 1996.
- (4) S. J. Young, M. G. Brown, J. T. Foote, G. J. F. Jones and K. S. Jones: "Acoustic indexing for multimedia retrieval and browsing," Proc. of ICASSP-97, Vol. 1, pp. 199-202, 1997.
- (5) S. Pfeiffer, S. Fischer and W. Effelsberg: "Automatic audio content analysis," Proc. ACM Multimedia, pp. 21-30, 1996.
- (6) J. Saunders: "Real-time discrimination of broadcast speech/music," Proc. of ICASSP-96, Vol. 2, pp. 993-996, 1996.
- (7) S. R. Subramanya, R. Simha, B. Narahari and A. Youssef: "Transform-based indexing of audio data for multimedia Databases," Proc. IEEE Conf. on Multimedia Computing and Systems, pp. 211-218, 1997.
- (8) J. C. Hancock and P. A. Wintz: "Signal Detection Theory," McGraw-Hill, 1966.
- (9) K. Kashino, G. Smith and H. Murase: "Time-series active search for quick retrieval of audio and video," Proc. of ICASSP-99, Vol. 6, pp. 2993-2996, Mar. 1999.
- (10) V. V. Vinod and H. Murase: "Focused color intersection with efficient searching for object extraction," Pattern Recognition, Vol. 30, No. 10, pp. 1787-1797, 1997.
- (11) 村瀬・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—," 信学論D-II, Vol. J81-DII, No. 9, pp. 2035-2042, 1998.
- (12) M. J. Swain and D. H. Ballard: "Color indexing," Int. J. Computer Vision, Vol. 7, No. 1, pp. 11-32, 1991.
- (13) 杉山: "セグメントの高速探索法," 信学技報, SP98-141, pp. 39-45, Feb. 1999.

著者紹介

柏野 邦夫

NTTコミュニケーション科学基礎研究所研究主任

平成7年入社。主に音響認識、マルチメディア探索の研究に従事。

平成2年東京大学工学部電子工学科卒業。4年同大学院工学系研究科修士課程修了。7年同大学院工学系研究科博士課程修了。同年工学博士（同大学）。

電子情報通信学会・情報処理学会・日本音響学会・人工知能学会・IEEE 会員。

村瀬 洋

NTTコミュニケーション科学基礎研究所主幹研究員

昭和55年入社。主にマルチメディア情報のパタン認識の研究に従事。

昭和53年名古屋大学工学部電子工学科卒業。55年同大学院電子工学科修士課程修了。62年工学博士（同大学）。

電子情報通信学会・情報処理学会・IEEE 会員。