論　文

ヒストグラム特徴を用いた音や映像の高速AND/OR探索

柏野　邦夫[†]　　黒住　隆行[†]　　村瀬　洋[†]

Quick AND/OR Search for Multimedia Signals Based on Histogram Features

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

あらまし　既知の音や映像（参照信号）が長時間の音や映像（入力信号）のどの時点にあるかを探索する問題を，文字列探索と対比して時系列探索と呼ぶ．時系列探索では，音と映像を組み合わせたり，複数の探索条件を論理式で指定して，高速に探索を行えることが望まれる．そこで本論文では，まず，我々が前論文で提案した音響信号探索法である時系列アクティブ探索法を，映像の探索にも適用できることを述べる．次に，参照信号についてのAND探索及びOR探索の効率的なアルゴリズムを提案する．更に，音と映像を組み合わせたマルチモーダルAND探索アルゴリズムを提案する．提案する各アルゴリズムは，それぞれ，個別に探索を行った結果を組み合わせる場合に比べ高速である．例えば参照信号についてのOR探索では，参照信号の相互類似度が0.8以上の場合に，1個の参照信号を探索する場合の約1.2倍以下の探索時間で，5個の参照信号を探索できることが示された．

キーワード　時系列探索，マルチメディア探索，アクティブ探索，高速探索

1. 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]〜[7]．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すことをいう．

これらのうち，本論文で指向するのは高速な探索技術である．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．実際，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．

我々は，既に前論文において音響信号の高速探索法である時系列アクティブ探索法を提案した[8]．しかし，時系列探索の利便性を高めるには，（1）音と映像を自由に組み合わせて探索できること，及び（2）複数の探索条件を論理式（AND/OR）で指定して高速に探索できることが望まれる．

そこで本論文では，このような柔軟性の高い時系列探索のための基本的なアルゴリズムを提案する．まず**2.**で時系列アクティブ探索法の概略を説明するとともに，新たに映像特徴について述べる．次に，**3.**で参照信号についてのOR探索，**4.**で参照信号についてのAND探索について，照合回数の削減法を議論する．更に**5.**で，マルチモーダルAND探索，すなわち入力信号として同期した音と映像が与えられた場合の，モダリティに関するAND探索について議論する．続いて**6.**で，各章で議論した手法の有効性を実験的に検討し，**7.**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

時系列探索の最も基本的な方法は，信号（音や映像）自身や信号から抽出した特徴の相関に基づいて信号検

[†] NTTコミュニケーション科学基礎研究所，厚木市
NTT Communication Science Laboratories, 3–1 Morinosato-Wakamiya, Atsugi-shi, 243–0198 Japan.

出 [9] を行う方法である．しかしこの方法は，長時間の信号に適用した場合膨大な処理時間がかかるという問題があるので，何らかの高速な方法が必要である．あらかじめ音声認識や物体認識などによって音や映像の情報をテキストや記号に変換しておき，探索時には文字列探索を行えば，探索は高速に行えると考えられるが，現在の認識技術では，テキストや記号への変換精度は必ずしも十分ではない．信号をベクトル量子化（VQ）などの方法で記号列に変換し，文字列探索に帰着する方法もあるが，VQ 符号同士を直接照合する方法では，**6.** で述べるように，依然としてかなりの処理時間がかかるという欠点がある．

前論文 [8] において，我々は，音響信号の高速探索アルゴリズムである時系列アクティブ探索法を提案した．これは VQ 符号のヒストグラム同士の照合に基づくことを特徴とする．ヒストグラムは累積特徴であるために，信号の変形による悪影響を受けにくい．また，ヒストグラム同士の照合は，特徴同士を直接照合するのに比べて照合自体の計算量が少ない上 [10]，時間軸方向で照合が不要な区間を求めて探索をスキップすることで，無駄な照合を省くことができる．これらの効果により，実用上十分な探索精度を保ったまま，スペクトル特徴ベクトルの内積に基づく方法に比べて数百倍の探索速度が得られることを報告した．

以下に時系列アクティブ探索法を要約する．処理の流れを図 1 に示す．まず参照信号（探索のキーとなる短時間の信号）と入力信号（長時間の信号）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（分類）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作る．そして，ヒストグラム同士の類似度が，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，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．

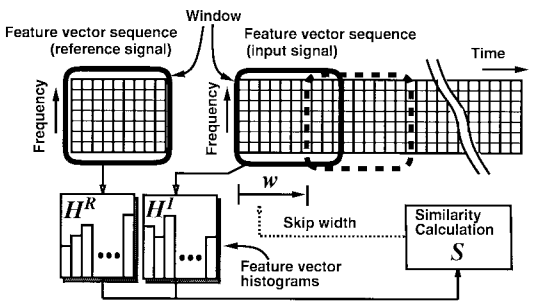


図 1　時系列アクティブ探索法の概要
Fig. 1　Overview of Time-series Active Search.

このアルゴリズム自体は，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ものが考えられる．また，ヒストグラム同士の類似度の定義についても様々なものが考えられる．これらのうちで，我々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．

$$S_{IR} = S(H_I, H_R) = \frac{1}{D}\sum_{l=1}^{L}\min(h_{Il},\, h_{Rl}) \tag{1}$$

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_{Il}, h_{Rl} はそれぞれの l 番目の符号に量子化される特徴ベクトルの数である．また L は VQ 符号帳のサイズ，D は参照信号の長さ（参照信号から導かれた特徴ベクトルの総数）である．

このとき，スキップ可能幅 w は，類似度の上限値に関する考察から，次式で求められる．

$$w = \begin{cases} \text{floor}\,(D(\theta - S_{IR})) + 1 & (S_{IR} < \theta) \\ 1 & (\text{上記以外}) \end{cases} \tag{2}$$

ただし floor(·) 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．

式 (1) において，各種の定義が考えられる中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）時間窓を移動して得られるヒストグラムの類似度の上限値が簡単な計算によって求められること，及び（3）既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 [11]〜[13]，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 [14].

2.2 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，判別性能が良い（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，及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，我々が前論文で用いた特徴ベクトルは，帯域通過フィルタを用いて，分析フレームを移動させながら短時間パワースペクトルを求め，そのスペクトルを周波数チャネルに関して正規化したものである．すなわち音響特徴ベクトル $\boldsymbol{f}(k)$ は，

$$\boldsymbol{f}(k) = (f_1(k),\ f_2(k),\ \cdots,\ f_V(k))\ , \tag{3}$$

と書くことができる．ここで k は分析フレームの位置を表す離散的な時刻 $(k = 1, 2, \cdots)$，V は特徴の次元数であり，$\boldsymbol{f}(k)$ の各要素は，

$$f_j(k) = \alpha(k)\ Y_j(k), \tag{4}$$

である．ここで $Y_j(k)$ は j 番目の帯域通過フィルタの出力波形の，分析フレーム内での 2 乗平均値である．また $\alpha(k)$ は正規化のための係数であり，

$$\alpha(k) = \frac{1}{\max\limits_{j}(Y_j(k))} \tag{5}$$

と定義される．

2.3 映像特徴

時系列アクティブ探索法は映像特徴を用いることにより映像（画像の時系列）の探索にも適用可能である [15]．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 $\boldsymbol{g}(k)$ は，

$$\boldsymbol{g}(k) = (g_1(k), \cdots, g_W(k)) \tag{6}$$

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，

$$g_j(k) = \frac{\bar{x}_j(k) - \min\limits_{i} \bar{x}_i(k)}{\max\limits_{i} \bar{x}_i(k) - \min\limits_{i} \bar{x}_i(k)} \tag{7}$$

である．ただし，

$$\bar{x}_i(k) = \mathop{\mathrm{E}}_{p \in \Omega}[x_p(k)] \tag{8}$$

である．ここで Ω は i 番目のサブ画像内の画素 p の集合であり，E は Ω についての平均を表す．また，画素 p の R，G，B 値を r_p, g_p, b_p とすると，

$$x_p = 0.299 r_p + 0.587 g_p + 0.114 b_p \tag{9}$$

とした．これは NTSC 方式における RGB 値と輝度値との関係式である [16].

3. 複数参照信号の OR 探索

時系列探索の応用分野として，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シャル（CM）や楽曲等の出現回数のカウント，インターネットにおける音響信号の探索エンジン等が考えられる．これらの応用では，同時に数多くの参照信号に対して探索を行いたい場合が多い（OR 探索）．例えば，放送データに対する CM のカウントを行う場合，仮に同一商品の CM に着目したとしても，わずかに異なる複数の類似 CM が同時期に放送されているのが通例である．また楽曲使用回数の場合にも，同時に複数の楽曲の複数の箇所について探索を行いたいことが多い．そこで，同一の入力信号に対して複数の参照信号を OR 探索する際に，単に探索を繰り返した場合よりも照合計算回数を削減する方法を示すことが本章の目的である．

そこで，N 個の参照信号 R_j $(j = 1, 2, \cdots, N)$ からそれぞれヒストグラム H_{Rj} が作成され，入力信号 I の現在の時間窓位置からヒストグラム H_I が作成されていたとする．ただし H_{Rj} の総度数 D_j はすべて等しく，$D_j = D$ であるとする．今，$j = m$ である H_{Rm} と H_I とについて式 (1) を計算して類似度 S_{IRm} を得たとする．我々は，H_I と H_{Rj} $(j \neq m)$ を具体的に照合することなく S_{IRj} の上限値を得ることに興味がある．そこで S_{IRj} の上限値について考察すると，

$$S_{IRj} \leqq 1 - |S_{IRm} - S_{RmRj}| \tag{10}$$

が成り立つことが示される（付録参照）．

このことから，複数の参照信号についての OR 探索は，以下のように行うことができる．

（1） 前処理として，参照信号同士の類似度をすべての組合せについて計算しておく．

（2） 現在位置を入力信号の最初に位置付ける．（ここから探索過程）

（3） スキップ可能位置が現在位置に最も近い参照信号を一つ選択し，現在位置をそのスキップ可能位置とする．

（4） 選択した参照信号と，現在位置の入力信号と

を照合し，類似度を求める．

（5） 得られた類似度をもとに，すべての参照信号に対するスキップ可能幅を更新する

（6） （3）に戻る

これによって，探索過程における照合回数を，参照信号を別々に照合した場合以下とすることができる．なお，D_j が等しくないときには，D_j のうちの最小値を D とすると，長さ D の部分区間について上記の議論が成り立つ．

4. 複数参照信号の AND 探索

複数参照信号の AND 探索とは，複数の参照信号 $R_1, \cdots, R_N$ を，開始時刻にそれぞれ時間遅れ $\tau_1, \cdots, \tau_N$ をつけて時間軸上に配置したとき，入力信号の中で，R_j $(j = 1, \cdots, N)$ のいずれに対しても探索しきい値を超える類似度をもつ区間を求めることである．探索される区間の長さ t_d は，

$$t_d = \max_j(\tau_j + D_j) - \min_j(\tau_j) + 1 \tag{11}$$

となる．ここで，D_j は R_j の継続時間である．t_d の時間区間における全体の類似度 S を以下のように定義すると，問題は，S が探索しきい値を上回る箇所を探索することである．

$$S = \min_j(S_j) \tag{12}$$

ここで，S_j は，j 番目の参照信号と，入力信号の対応する区間との類似度である．

参照信号の AND 探索も，応用上重要である．時系列アクティブ探索法では，累積された特徴によって照合を行っているために，多少の信号の変形に影響されにくい反面，時系列としての時間構造を保存していないので，出現の順序に関する違いを区別しにくいという性質がある．その場合，参照信号を区分し，各区分について AND 探索を行えば，正確な区別が可能となる．

AND 探索の基本的な方法は，各参照信号について順に探索を行うことである．本論文ではこれを順次法と呼ぶ．まず，式 (2) を用いて，全体時間窓（長さ t_d）の現在の位置における j 番目のスキップ可能幅 w_j を求める．すると，全体時間窓のスキップ可能幅 w を

$$w = \max_j(w_j) \tag{13}$$

と求めることができる．これは，式 (12) より，S_j のうちの一つでも θ 以下であれば，S も必ず θ 以下となるので，w_j のうちの j についての最大幅まで時間窓を移動させても，移動中に S が θ を超えることはないからである．

実際には，すべての j について照合を行うよりも，探索しきい値 θ に満たない S_j が発見された時点で，直ちに全体時間窓をスキップする方が照合計算回数は少なくて済むことが多い．これを本論文では順次中断法と呼ぶ．

ところで，特に，もとの参照信号を時間的に分割してできた複数の参照信号について AND 探索を行う場合（つまり AND 探索を行う複数の参照信号が時間的に隣接している場合）には，

$$S_A \geqq \min_j(S_j) \tag{14}$$

が成り立つ（付録参照）．ここで S_A はもとの参照信号における類似度，S_j は分割された参照信号に対する類似度である．これは，AND 探索において類似度が探索しきい値を超える部分は，もとの参照信号についての探索でも決してスキップされないことを意味している．このため，まずもとの参照信号について単独探索を行い，必要な部分のみ各参照信号についての照合を行うことが考えられる．これを本論文では併合法と呼ぶ．

同一の参照信号と入力信号の組合せにおいて，併合法と順次法（あるいは順次中断法）のどちらが平均的な照合計算回数が少ないであろうか．例えば参照信号が N 等分割の場合，併合法におけるスキップ可能幅を w_A として，

$$w_A \geqq \max_j(w_j) \tag{15}$$

となる条件は，

$$S_A \leqq \theta - \frac{1}{N}(\theta - \min_j(S_j)) \tag{16}$$

と計算できる．式 (16) は常に成り立つわけではないが，もし各分割における類似度が均一である（つまり $\min_j(S_j) = S_A$）と仮定すると，$S_A \leqq \theta$ に対して常に式 (16) が成り立つ．このことから，多くの場合に併合法が順次法よりも有利であると考えられる．

5. マルチモーダル AND 探索

前章までに，同一の入力信号に対して，複数の参照信号を OR 探索や AND 探索する場合について議論

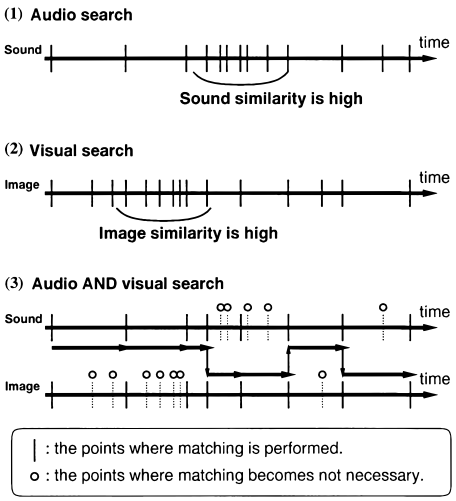


図 2　マルチモーダル AND 探索におけるスキップ可能幅の例
Fig. 2　Skip width example in the search combining audio and video (multimodal AND-search).

したが，このほかに，同期した複数の入力信号に対して，それぞれ対応する参照信号についての OR 探索や AND 探索を行いたい場合もある．ここでは特に，マルチモーダル AND 探索，すなわち同期した音と映像の信号に対し，そのいずれもが参照信号と類似している箇所を探索したい場合について検討する．

マルチモーダル AND 探索では，複数参照信号の AND 探索と同様に，各入力信号についての類似度 S_j のうちの最小値を，すべての入力信号についての類似度 S と定義する．そして，各入力信号について求めたスキップ可能幅 w_j を実時間に換算し，実時間上で最も大きいスキップ幅を採用することによって，照合回数を，それぞれの入力信号について別々に探索する場合以下にすることができる．これを，図 2 に示す．

6. 実　　験

6.1 映像特徴

まず，映像特徴の有効性について検討するため，**2.** に述べた映像特徴について，探索速度と探索精度に関する実験を行った．実験に用いたワークステーションの仕様を表 1 に示す．

6.1.1 探索速度

探索速度を評価するため，テレビ放送 6 時間分の映像から，特定の 15 秒の CM を探索するのに要する時間を測定した．

表 1　実験に用いた計算機の仕様
Table 1　Specification of the workstation used in the experiments.

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

まず，ある民放テレビ局の放送を家庭用ビデオデッキで 6 時間録画した（VHS HiFi，3 倍モード）．次に，この録画テープを再生して，映像を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 15 秒の異なる CM 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，非圧縮 RGB，画面サイズ 80×60 で行った．特徴ベクトルの各次元当りのビン数は 2 とした．また $W = 12$（横方向 4 等分割，縦方向 3 等分割）とした．

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），の三つからなる．なお，以下の議論において時間はいずれも CPU 時間で測定した．CPU 時間は測定ごとに数%程度のばらつきが見られたので，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．

（1） 特徴抽出時間については，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．

（2） ベクトル量子化については，6 時間と 15 秒分の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(注 1)．

（3） 探索実行時間については，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，及び探

(注 1)：本論文の実装では，12 次元特徴ベクトル 1 個を 12 バイトで保持しているので，6 時間分の特徴ベクトルは正味約 7.8 M バイトとなる．

索しきい値に依存する．表 2 に示した CPU 時間は，10 本の参照信号（CM）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．表 2 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．なお本実験では，10 本の CM 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）ことを確認した．

参考のため，図 3 に，本実験における類似度の時間変化パターンの例（ある CM を参照信号とした場合）を示す．○印が探索された時点を，破線が探索しきい値を示している．

なお，VQ 符号を文字列と見て直接比較する方法として，対応する VQ 符号同士が一致しているものの割合を順次数える方法が考えられる．この場合の探索実行時間は本実験の設定において約 2.4 秒であった．これは提案法の約 12 倍の時間である．

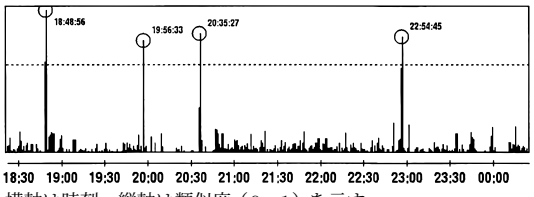
6. 1. 2 探 索 精 度

提案法の探索精度を調べるため，実験 1 とは別のテレビ放送の録画を用いて実験を行った．まず，実験 1 と同様の方法でテレビ放送を録画し，異なる CM の部分をつなげて 1 時間に編集した．これは，2 度以上同一の映像が出て来ない試料を用いた方が，実験の自動化に都合が良いためである．このビデオテープを再生し，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 R，G，B 値に対してそれぞれ白色ガウス雑音を加えた場合についても実験を行った．

表 2　映像特徴による探索速度
Table 2　Search speed based on the image features.

探索実行時間		速度向上	照合回数の比	参照
全探索	提案法			
22.5 s	0.20 s	112 倍	1/207	図 3

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．



横軸は時刻，縦軸は類似度（0〜1）を示す．

図 3　映像探索の例
Fig. 3　Search result by the image feature.

本実験では，参照信号の長さ及び SN 比をパラメータとした．SN 比は，入力信号 1 時間の，R，G，B 値の 2 乗平均値に対して，雑音の分散を設定することによって制御した．同一の実験条件において，200 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，次式の c を制御することによって探索しきい値を変化させた．

$$\theta = m + c\,\sigma \tag{17}$$

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る．式 (17) は予備実験によって経験的に定めたものである．また，式 (17) において θ が 0.9 を超えるときは $\theta = 0.9$，0.1 を下回るときは $\theta = 0.1$ とした．本実験では，式 (17) における c の値を 200 回の繰返し中一定とし，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．

式 (17) は，個々の参照信号ごとに，類似度の分布を考慮して異なる値の θ を設定することを表している．なお本実験における具体的な c の値は 4.8〜23 の範囲にわたっており，雑音のパワーが大きくなるほど，精度を最大化する c の値は小さくなる傾向が見られた．

その他の取込みや探索のパラメータは前節の実験と同様とした．

実験結果を図 4 に示す．これによれば，参照信号が 15 秒間あれば，SN 比 2 dB まで探索も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，99%程度以上の探索精度が得られることがわかる．

以上の実験から，映像に対しても，良好な探索精度と，音響信号に対する探索と同程度の探索速度とを実現できることが明らかになった．ただし，実際の映像系に見られる雑音は白色雑音からかけ離れた性質をもつものも多いので，上記探索精度に関する実験の結果をそのまま実際の映像系における耐雑音特性として解

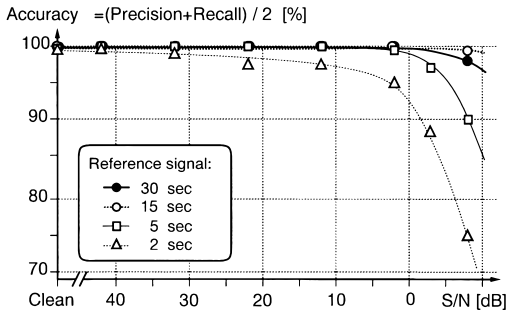


図 4　映像特徴による探索精度
Fig. 4　Search accuracy based on the image features.

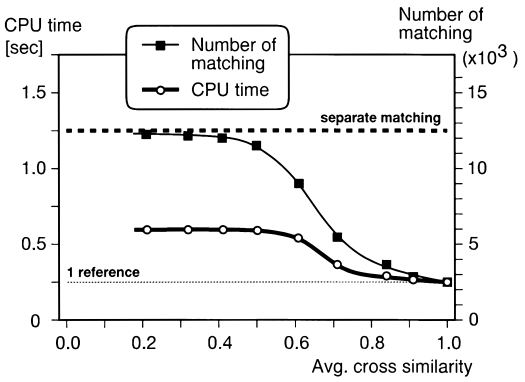


図 5　5 個の参照信号の OR 探索における平均照合回数と平均探索実行時間
Fig. 5　Number of matches and CPU time in the OR-search with 5 reference signals.

釈できないことはいうまでもない．

6.2　複数参照信号の OR 探索

提案法により，複数の参照信号を別々に探索した場合に比べてどの程度高速化できるかを調べるため，実験を行った．本論文で述べる OR 探索や AND 探索のアルゴリズムは，対象が映像であっても音であっても同様に用いることができるが，ここでは音響探索の例を示す．すなわち本実験では，入力信号を 6 時間分のテレビ放送の音響信号とし，参照信号を 5 個の 15 秒間の信号とした．提案法は，精度に関しては参照信号を別々に照合した場合と同一なので，照合回数と探索速度（探索実行時間）を比較する．

3. の議論から，提案法によって OR 探索を効率化できる度合は参照信号相互の類似度（相互類似度）に依存する．そこで，5 個の参照信号は，共通の音響信号（ある CM の一部）と，共通でない音響信号を接続することによって作成し，共通部分の長さを制御することで相互類似度を制御した．また入力信号は，参照信号の作成に用いた音響信号を含まない信号とした．

探索のパラメータは，サンプリング周波数=11.025 kHz，特徴次元数=7，分析フレームの長さ=60 ms，分析フレームの移動幅=10 ms，各特徴次元におけるビン数=3，探索しきい値 $\theta = 0.8$ とした．

図 5 に実験結果を示す．図中の太い破線は 5 個の参照信号を個別に探索した場合の合計照合回数及び合計探索実行時間を示し，細い破線はその 1/5 の値を示す．図 5 では，照合回数・探索実行時間とも，参照信号同士の事前の照合を含んだ参照信号 5 個分の値である．図 5 に示されるように，平均相互類似度が 1 に近いときは 1 個の参照信号の探索に近い照合回数及び探索実行時間で探索可能である．例えば平均相互類似度が 0.84 のとき，1 個の参照信号を探索する場合に比べ，照合回数は約 1.5 倍，探索実行時間は約 1.1 倍であった．また，相互類似度が低くなり，照合回数が個別探索の場合に近づいた場合にも，探索実行時間は個別探索の場合の約 0.47 倍と低くとどまっている．これは，入力信号に対するヒストグラム作成のコストが，個別探索の合計よりも OR 探索の方が少ないことによるものと考えられる．

提案法において，参照信号同士の照合に必要な照合回数は，N 個の参照信号に対して $N(N-1)/2$ 回である．本実験のように 5 個の参照信号の場合には，この照合回数は 10 回であり，無視できる程度である．ただし N の増大に対して照合回数は N の 2 乗のオーダで増大する．例えば参照信号が 1000 個の場合，参照信号同士の類似度の計算に約 50 万回の照合が必要となる．個別探索では 1 個の参照信号の照合に約 2500 回の照合が必要なため，1000 個では約 250 万回であるが，参照信号間の相互類似度が低い（例えば 0.2）とすると，提案法では合計約 300 万回の照合が必要になると見込まれる．

6.3　複数参照信号の AND 探索

参照信号の AND 探索を行った場合の照合回数及び探索速度（探索実行時間）を，実験により確かめた．**6.1** の映像探索実験と同様に，6 時間分のテレビ放送の映像信号を入力信号とし，無作為に選択した 10 本の 15 秒 CM を探索対象とした．それぞれの CM を，重複がないように均等な長さに 3 分割して 3 個の参照信号とし，これらの AND 探索を行うことにした．探

表 3　参照信号 3 分割 AND 探索における平均照合回数と平均探索実行時間
Table 3　The number of matches and CPU time in the AND-search with respect to the reference signals.

探索の種類	照合回数 ($\times 10^3$)	探索実行時間
個　別	22.49	1.40 s
(a) AND（順次法）	21.86	1.43 s
(b) AND（順次中断法）	7.52	0.34 s
(c) AND（併合法）	2.63	0.20 s

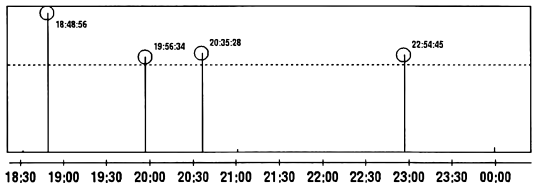


図 6　探索結果（参照信号 3 分割 AND 探索の場合）
Fig. 6　Search result of the AND-search with 3 reference signals.

索のパラメータは **6.1** と同様とした．

表 3 に照合回数及び探索実行時間の測定結果を示す．(a) 順次法，(b) 順次中断法，及び (c) 併合法のすべての場合において，個別探索を 3 個の参照信号について行った場合に比べ照合回数が削減されている．なお，(a)，(b)，(c) の探索結果は同一である（個別に探索を行った場合は，参照信号が異なるため (a)〜(c) の結果とは比べられない）．個別探索に比べ，(a) の照合回数が削減されているのは，式 (13) において最大値をとることによるものである．本実験においては (a)〜(c) の中で (c) の併合法の効果が最も顕著であった．

一方，図 6 は，CM を 3 分割し AND 探索を行った場合の類似度のパターンを示している．これを，15 秒間の CM 全体を 1 個の参照信号として探索した図 3 の場合に比較すると，図 6 では正解位置以外での類似度が極めて小さく，類似度設定におけるマージンが大きくなっていることがわかる．

6.4　マルチモーダル AND 探索

マルチモーダル AND 探索を行った場合の照合回数及び探索速度（探索実行時間）を測定した．前章の実験と同様に，6 時間分のテレビ放送の音響信号及び映像を入力信号とし，10 本の 15 秒 CM の映像及び音響信号を参照信号とした．ただし，通常の CM 探索では，音のみあるいは映像のみの探索で十分であり，マルチモーダル AND 探索を用いる意味が薄い．むしろマルチモーダル AND 探索は，音だけ，映像だけでは

表 4　マルチモーダル AND 探索における平均照合回数と平均探索実行時間
Table 4　The number of matches and CPU time in the search combining audio and video.

探索の種類	照合回数 ($\times 10^3$)	探索実行時間
音　響	11.39	0.45 s
映　像	3.39	0.20 s
AND	3.04	0.18 s

探索結果の候補がたくさん出て来るような場合に，音と映像の両方を使うことで探索結果を絞り込むような状況で使うことが想定される．そこで本実験では，探索しきい値を $\theta = 0.5$ と低めにし，単一モダリティでは多くの探索結果が得られるように設定した．なお本実験では，音または映像の一方の照合で一定以上（予備実験により 1.5 秒に設定）のスキップ可能幅が得られたときは，同じ時間窓位置における他のモダリティでの照合を省略した．

実験結果を表 4 に示す．提案法によって，照合回数・探索実行時間とも，どちらかのモダリティだけの場合よりも減少していることがわかる．個別に探索を行った場合の和に比較すると，AND 探索の照合回数は約 21%，探索実行時間は約 28%となっている．

7. むすび

本論文では，時系列アクティブ探索法を映像探索に適用可能であることを示すとともに，複数の参照信号の AND 探索と OR 探索，及びマルチモーダル AND 探索を効率的に実行する方法を提案した．参照信号の OR 探索では，事前に参照信号同士の照合を行っておくことにより，類似した参照信号について探索を行った場合の照合計算回数が，別々に探索を行った場合以下にできることを示した．また実際の探索速度も高速化され，特に参照信号間の平均相互類似度が 0.8 以上の場合に，一つの参照信号に対する探索の約 1.2 倍以下の探索実行時間で 5 個の参照信号に対する OR 探索を行うことができた．また参照信号の AND 探索では，まず隣接区間を併合して探索を行うことにより，効率的な探索が行えることを実験的に示した．

本論文で検討したような AND 探索と OR 探索は，例えば UNIX 環境における文字列探索ツールである grep ファミリーのように，柔軟で利便性の高い探索をマルチメディアデータに対して実現するための基礎になると考えられる．今後は，高速探索という特性を保ったままで，更に信号の変形やバリエーションを許

容することのできる探索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，及び萩田紀博部長に感謝する．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] 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. Stede, and P. Yanker, “Query by image and video content: The QBIC system,” IEEE Comput., 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. 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. 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] 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—時系列アクティブ探索法,” 信学論（D-II），vol.J82-D-II, no.9, pp.1365–1373, Sept. 1999.

[9] J.C. Hancock and P.A. Wintz, Signal Detection Theory, McGraw-Hill, 1966.

[10] 沢井英文，米山正秀，中川聖一，“大語彙単語音声認識の高速化のための種々の検討,” 音響誌，vol.43, no.11, pp.858–867, 1987.

[11] M.J. Swain and D.H. Ballard, “Color indexing,” Int. J. Computer Vision, vol.7, no.1, pp.11–32, 1991.

[12] V.V. Vinod and H. Murase, “Focused color intersection with efficient searching for object extraction,” Pattern Recognit., vol.30, no.10, pp.1787–1797, 1997.

[13] 村瀬　洋，V.V. Vinod, “局所色情報を用いた高速物体検索—アクティブ探索法,” 信学論（D-II），vol.J81-D-II, no.9, pp.2035–2042, Sept. 1998.

[14] 杉山雅英，“セグメントの高速探索法,” 信学技報，SP98-141, Feb. 1999.

[15] K. Kashino, G. Smith, and H. Murase, “Time-series active search for quick retrieval of audio and video,” Proc. ICASSP-99, vol.6, pp.2993–2996, March 1999.

[16] 高木幹雄，下田陽久（監修），画像解析ハンドブック，東京大学出版会，p.103, 1991.

付　　録

1. 式 (10) の証明

類似度間の関係を図 A·1 に示す．今，m 番目の参照信号と入力信号との照合を実際に行って S_{IRm} が得られたところである．また，S_{RmRj} は事前に計算され定まっている．

（i）　$S_{IRm} \leqq S_{RmRj}$ のとき

m 番目の参照信号と j 番目の参照信号とが非常に類似した場合を想定すると理解しやすい．

H_{Rm} の要素（H_{Rm} に含まれる特徴ベクトル）のうち，H_I との類似度に寄与したものの集合を $\{H_{Rm} \cap H_I\}$ で表し，その要素数を $|H_{Rm} \cap H_I|$ と表すと，式 (1) より

$$|H_{Rm} \cap H_I| = DS_{IRm} \tag{A·1}$$

である．また同様に

$$|H_{Rj} \cap H_I| = DS_{IRj} \tag{A·2}$$

である．

今 $S_{IRm} \leqq S_{RmRj}$ であるから，図 A·2 の包含関係が成り立つ．図 A·2 左側は，実際の照合によって定まった H_{Rm} と H_I との包含関係を表している．今，H_{Rj} の包含関係がどのような場合に $|H_{Rj} \cap H_I|$ が最大となるかを考えると，その条件は，(1) $\{H_{Rm} \cap H_I\}$ の要素がすべて H_{Rj} に含まれ，かつ，(2) H_{Rj} の要素のうちで，既に要素数が定まっている $\{H_{Rm} \cap H_{Rj}\}$ の要素でないもの（図 A·2 の網掛け部分）がすべて H_I との類似度に寄与することである．つまりこれは，H_{Rj} が図 A·2 右の太線で表したような包含関係となるときである．これを式で表すと

$$|H_{Rj} \cap H_I| \leqq |H_{Rm} \cap H_I| + (D - |H_{Rm} \cap H_{Rj}|) \tag{A·3}$$

となる．ゆえに，式 (A·1) と式 (A·2) を用いて

$$S_{IRj} \leqq S_{IRm} + (1 - S_{RmRj}). \tag{A·4}$$

（ii）　$S_{IRm} \geqq S_{RmRj}$ のとき

m 番目の参照信号と j 番目の参照信号とがほとん

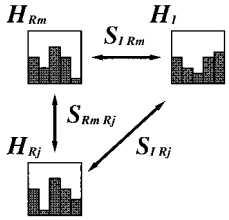


図 A·1　類似度同士の関係の説明図
Fig. A·1　Relations between the similarities.

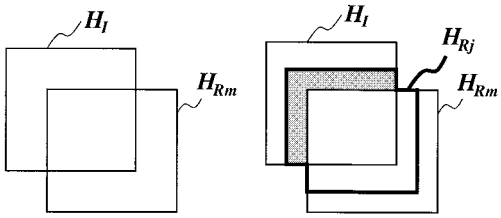


図 A·2　ヒストグラムの要素に関する包含関係（1）
Fig. A·2　Relations between histograms (1).

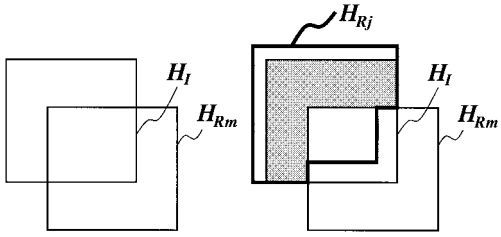


図 A·3　ヒストグラムの要素に関する包含関係（2）
Fig. A·3　Relations between histograms (2).

ど類似していない場合を想定すると理解しやすい．

同様の考察により，図 A·3 の包含関係が成り立つ．したがって，$|H_{Rj} \cap H_I|$ が最大となる条件は，(1) $\{H_{Rj} \cap H_{Rm}\}$ の要素がすべて H_I に含まれ，かつ，(2) H_I の要素のうちで $\{H_{Rm} \cap H_I\}$ の要素でないもの（図 A·3 の網掛け部分）がすべて H_{Rj} との類似度に寄与することである．つまり，

$$|H_{Rj} \cap H_I| \leqq |H_{Rm} \cap H_{Rj}| + (D - |H_{Rm} \cap H_I|) \qquad \text{(A·5)}$$

が成り立つ．ゆえに

$$S_{IRj} \leqq S_{RmRj} + (1 - S_{IRm}). \qquad \text{(A·6)}$$

式 (A·4) と式 (A·6) をまとめて，式 (10) を得る．(終)

2. 式 (14) の証明

分割した個々の区間の長さを $D_1, D_2, \cdots, D_N$ とすると，分割前のヒストグラムと，各分割におけるヒストグラムとにおいて，入力信号との類似度に寄与した要素の数について

$$DS_A = D_1S_1 + D_2S_2 + \cdots + D_NS_N \qquad \text{(A·7)}$$

が成り立つ．各分割は隣接しているので，明らかに

$$S_A \geqq \frac{D_1S_{min} + D_2S_{min} + \cdots + D_NS_{min}}{D_1 + D_2 + \cdots + D_N} \qquad \text{(A·8)}$$

である（ただし $S_{min} = \min_j(S_j)$）．すなわち

$$S_A \geqq S_{min} \qquad \text{(A·9)}$$

であり，式 (14) が成り立つ．(終)

（平成 11 年 11 月 24 日受付，12 年 5 月 8 日再受付）

柏野　邦夫　（正員）

平 2 東大・工・電子卒．平 7 同大大学院電気工学専攻博士課程了．同年 NTT に入社．現在，NTT コミュニケーション科学基礎研究所研究主任．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．工博．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

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 NTT に入社．現在，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．

村瀬　　洋　（正員）

昭 53 名大・工・電子卒．昭 55 同大大学院修士課程了．同年日本電信電話公社（現 NTT）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，NTT コミュニケーション科学基礎研究所メディア認識研究グループリーダ．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各受賞．情報処理学会，IEEE 各会員．