

多種類の特徴歪みを考慮した 音響信号の高速探索法

柏野 邦夫 黒住 隆行 村瀬 洋

NTTコミュニケーション科学基礎研究所
〒243-0198 神奈川県厚木市森の里若宮 3-1

{kunio, kurozumi, murase}@eye.brl.ntt.co.jp

あらまし 本稿では、既に提案した音響信号の高速 OR 探索法の一応用として、特徴歪みのある場合の音響信号の高速探索法を提案する。提案法は、予め特徴の歪み方の確率分布を学習しておき、その確率分布をヒストグラム上に表現することを特徴とし、複数の確率分布を考慮するために OR 探索を行うものである。特徴歪みの一例としてマイクロホン収録音を用いて実験を行ったところ、確率分布を考慮しない従来の方法に比べて探索時間をあまり増やすことなく、従来法による探索誤りを半減できる可能性が示された。

キーワード 時系列探索, 時系列アクティブ探索, 特徴歪み, OR 探索

A Quick Search Method Coping With Multiple Kinds of Feature Distortions

Kunio Kashino, Takayuki Kurozumi, Hiroshi Murase

NTT Communication Science Laboratories
3-1 Morinosato-Wakamiya, Atsugi-shi,
243-0198, Kanagawa, Japan.

Abstract Here we propose a quick search method for distorted audio signals. The method is an application of the quick OR-search method that we previously developed. In the present method, feature distortions are first learned as probabilistic distributions, and then the multiple distributions are efficiently considered using the OR-search method. The experiments using microphone-recorded signals show that the proposed method achieves approximately 3 to 5 % better accuracy than the old method with little increase of search time.

key words time-series search, time-series active search, feature distortion, OR search

1 まえがき

我々は、現在広く用いられている文字列探索の高速アルゴリズムと同様に、音や映像に関する高速探索アルゴリズムの実現が重要であるとの認識に立って、音や映像の高速探索に関する研究を進めている。既に、98年に「時系列アクティブ探索法」と呼ぶ手法を提案した[1, 2, 3]。時系列アクティブ探索法は、音や映像の特徴ベクトルを分類（ベクトル量子化）し、各分類（量子化符号）ごとの出現回数を数えてヒストグラムを作り、そのヒストグラムどうしの照合によって探索を進めることを特徴とする。ヒストグラムどうしの照合は、特徴どうしを直接照合するのに比べて照合自体の計算量を少なくできる上、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、特徴ベクトルの直接照合に比べて数百倍の探索速度が実現された[2]。

各研究機関等でも、関連する研究が進展してきている。例えば杉山は、ヒストグラム照合におけるスキップについて一般的な考察を与えとともに、照合時のベクトル量子化をファジィ化した「ファジィアクティブ探索法」を提案した[4, 5]。また西村らは、照合する信号区間の長さをフリーとする「RIFAS」（リファレンスインターバルフリーアクティブ探索法）などの時系列探索法を提案した[6]。また我々自身も、より柔軟で適用範囲の広い高速探索の実現を目的として、AND探索・OR探索¹の高速アルゴリズムの提案や[7, 8]、「確率ディザボータイング」と呼ぶ、特徴の歪みに頑健なヒストグラムの作成法などを提案してきた[9]。

ところで、時系列アクティブ探索法では、これまで照合する信号の特徴歪みが小さいことを仮定していた。もし何らかの原因で特徴が歪んでいた場合には、検出すべき信号区間であっても信号同士の類似度が低下するため、検出に失敗することがあり得る。しかし実際の応用を考えると、例えばマイクロホンを経由した音とオリジナルの音

¹AND探索とは、複数の条件を共に満たす部分を探索すること、OR探索とは、複数の条件のいずれかを満たす部分を探索することである。

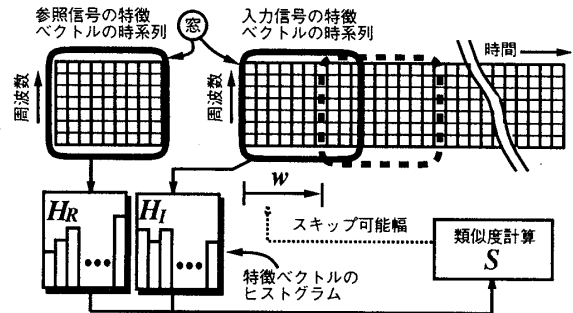


図 1: 時系列アクティブ探索法の概要

との照合など、各種の特徴歪みを考慮して照合しなければならない場面も多い。

そこで本稿では、我々が前回提案していたOR探索の高速アルゴリズムの一応用として、複数種類の特徴歪みを考慮しながら音響信号を探索する実験を行ったので報告する。以下2.で、時系列アクティブ探索法と、OR探索の効率的なアルゴリズムについて要約したのち、3.で、確率ディザボータイングとOR探索を併用した特徴歪みへの対処法を提案する。続いて4.で実験的検討を行ったのち、5.をむすびとする。

2 探索法の要約

2.1 時系列アクティブ探索法

まず時系列アクティブ探索法について要約しておく。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分類してヒストグラムを作る。そして、ヒストグラムどうしの類似度が、予め設定した値（これを探索閾値と呼ぶ）を越えるかどうかで、参照信号の有無を判定する。このとき、類似度の値と探索閾値とから、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。

特徴ベクトルの作成方法や分類の仕方は各種考えられる。本稿では、計算量の少なさを重視して、帯域フィルタバンクを用いて計算した短時間

パワースペクトルを特徴ベクトルとして用いる。

また、特徴ベクトルの各要素の値をいくつかのビン (bin) に分割することによって分類している。この分類は、ベクトル量子化の簡単な場合に相当する。

また、ヒストグラムどうしの類似度についても様々なものが考えられるが、我々は特にヒストグラム重なり率に着目している。ヒストグラム重なり率 S は、次のように定義される。

$$S = \frac{1}{D} \sum_{l=1}^L \min(h_{Il}, h_{Rl}) \quad (1)$$

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_{Il}, h_{Rl} はそれぞれの l 番目のビンに含まれる度数である。また L はヒストグラムのビンの数、 D はヒストグラムの総度数 (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、類似度の上限值に関する考察から、次式で求められる。

$$w = \begin{cases} \lfloor D(\theta - S) \rfloor + 1 & (S < \theta) \\ 1 & (\text{上記以外}) \end{cases} \quad (2)$$

ただし w の単位は特徴ベクトルの個数であり、 $\lfloor \cdot \rfloor$ は切り下げを表し、 θ は探索閾値である。類似度が θ を越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。

2.2 複数参照信号の OR 探索

同一の入力信号に対して複数の参照信号を OR 探索する際には、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ことができる。これは、ある一つの参照信号について照合を行ったとき、事前に計算しておいた参照信号同士の類似度を利用することで、他の参照信号について、照合を行わなくても時間窓のスキップを行うことができる場合があることによる。

すなわち、図 2 に示すように、 N 個の参照信号 R_j ($j = 1, 2, \dots, N$) 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 は全て等しく、 $D_j = D$ であるとする。いま、 $j = m$ である H_{Rm} と H_I について式 (1) を計算して類似度

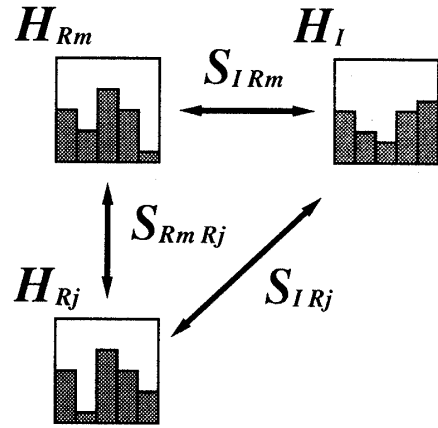


図 2: 類似度同士の関係の説明図

S_{IRm} を得たとする (添字は信号を表す)。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、

$$S_{IRj} \leq 1 - |S_{IRm} - S_{RmRj}| \quad (3)$$

が成り立つことが示される [8]。

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、参照信号どうし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、全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。

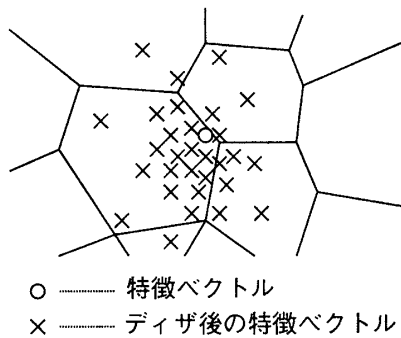


図 3: 確率分布に基づいてばらつかせた特徴ベクトル

3 特徴歪みの考慮

特徴歪みに対処する方法としては、大別して、歪みに頑健な特徴の工夫や特徴の正規化の工夫といった特徴指向のアプローチと、マルチテンプレート化のような照合指向のアプローチが考えられる。マルチテンプレートとは、パターン認識において、一つのカテゴリに対し複数の辞書パターンを用いて照合することを指す。これらのうち本稿では後者のアプローチに着目し、以下に述べる確率ディザボーティングを用いて特徴歪みに対処することを考える。

確率ディザボーティングとは、特徴ベクトルをある確率分布に基づいてヒストグラムのビンに投票する手法である [9]。図 3 のように、特徴ベクトルを、ある確率分布にしたがってばらつかせ、確率分布を反映した形で特徴ベクトルを表現する。その後、各特徴ベクトルをベクトル量子化してヒストグラムに投票する。ばらつかせる確率分布を、特徴歪みを反映したものとするにより、特徴歪みの影響をキャンセルできると考えられる。すなわち、図 4 のように、原信号 $s(k)$ が何らかのシステム Z_R を介してシステムにキャプチャされ、参照信号 $s_R(k, t_R)$ が得られたとする。すなわち

$$s_R(k, t_R) = s(k) + \epsilon_R(k + t_R) \quad (4)$$

である。ただし k はフレームの時刻、 t_R は参照信号のキャプチャを開始した時刻、 $\epsilon_R(t)$ はシステム Z_R による歪みを表す確率ベクトルである。一方、 Z_R とは異なる別のシステム Z_I を介して入力信号 s_I を得る。すなわち

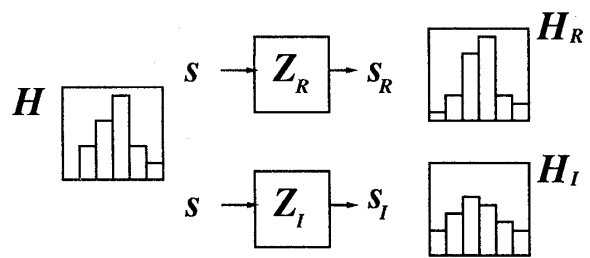


図 4: 原信号と参照信号・入力信号の関係

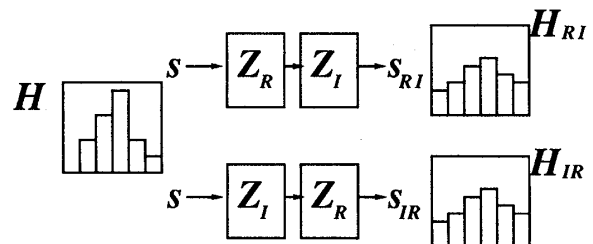


図 5: 原信号と補償参照信号・補償入力信号の関係

$$s_I(k, t_I) = s(k) + \epsilon_I(k + t_I) \quad (5)$$

である。ただし t_I は入力信号のキャプチャを開始した時刻、 $\epsilon_I(t)$ はシステム Z_I による歪みを表す確率ベクトルである。

ここで、ヒストグラム $H_R(k), H_I(k)$ における l 番目のベクトル量子化符号の計数 $h_{RI}(k), h_{II}(k)$ の期待値は、それぞれ

$$E \{h_{RI}(k, t_R)\} = \sum_{n=k}^{k+D} \int P_R(x - s(n)) \phi_l(x) dx \quad (6)$$

$$E \{h_{II}(k, t_I)\} = \sum_{n=k}^{k+D} \int P_I(x - s(n)) \phi_l(x) dx \quad (7)$$

となる。ただし D は時間窓の長さであり、 $P_R(x), P_I(x)$ はそれぞれ ϵ_R, ϵ_I の確率密度関数を表し、 $\phi_l(x)$ は、 x が l 番目のベクトル量子化符号領域内にあるとき 1、そうでないとき 0 と定義された関数とする。これらの式は、信号のまわりの確率分布を該当領域内で積分し、ヒストグラムの計数に加算することを表している。

これらは一般に等しくないため、原信号が同一であってもヒストグラムの類似度は 1 とならない。このため Z_R, Z_I による歪み ϵ_R, ϵ_I を学習により推定して図 5 のように変形（補償）することにより、両者の期待値を

$$E\{h_{RI}(k, t_R)\} = \sum_{n=k}^{k+D} \int P_{RI}(\mathbf{x} - \mathbf{s}(n)) \phi_I(\mathbf{x}) d\mathbf{x} \quad (8)$$

$$E\{h_{IR}(k, t_I)\} = \sum_{n=k}^{k+D} \int P_{IR}(\mathbf{x} - \mathbf{s}(n)) \phi_I(\mathbf{x}) d\mathbf{x} \quad (9)$$

とすることができ、システムが可換であれば $P_{IR} = P_{RI}$ となって、原信号が同一であればヒストグラムの類似度の期待値も 1 とすることができる。

ところが、 ϵ_I, ϵ_R を直接観測することはできないので、実際には、例えば観測された参照信号を基準として、観測された入力信号の特徴歪みを学習することにする。すなわち、学習用の信号を用意し、学習したい特徴歪みを実際に加えたものを観測する。例えばマイクロホン収録音を学習する場合、オリジナルの音を様々な環境で実際にスピーカで再生したものを様々なマイクロホンで収録することになる。このようにして収録した音と、オリジナルの音との時間軸を合わせて十分な時間観測することで、ある特徴ベクトルの歪みの確率分布を学習することができる。

なお本稿では、学習はベクトル量子化の符号領域ごとに行い、また特徴ベクトルの歪みは特徴空間内での並行移動と正規分布でモデル化できるものと仮定した。すなわち並行移動量と正規分布の分散という 2 つのパラメータが学習の対象である。また、あらゆる特徴歪み全体について一度に学習するのではなく、特徴歪みをいくつかクラス分けし、そのクラス毎に学習を行って、探索時にそれらについての OR 探索を行うことにする。

確率分布を表現するために投票する票の数は、特徴ベクトル一つあたり複数票を用いた方が確率分布をより良く反映するが、高速性を維持するためには、1 票の方がよい。照合区間が十分長ければ、累積されたヒストグラムは確率分布を反映していると考えられる。

杉山によるファジィアクティブ探索法は、複数のヒストグラムビンに投票を行うという点で本手法に先行するものである [5]。ファジィアクティブ探索法は特徴ベクトルと代表ベクトルの距離に基づいて投票するのに対し、本手法は、学習によって求めた特徴ベクトルの生起確率分布に基づいて投票する。

4 実験

4.1 OR 探索の性質の基本検討

まず、OR 探索の性質に関する基本検討として、複数の参照信号を別々に探索した場合に比べて、提案アルゴリズムによりどの程度高速化できるかを調べた。

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。

2. 章の方法によって 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、探索閾値 $\theta = 0.8$ とした。

図 6 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およ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 6 では、照合回数・探索実行時間とも、参照信号どうしの事前の照合を含んだ参照信号 5 個分の値である。図 6 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およ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

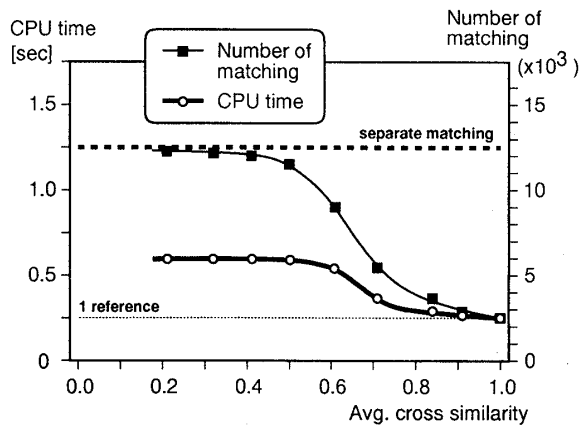


図 6: 5 個の参照信号の OR 探索における平均照合回数と平均探索実行時間

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付いた場合にも、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。確率ディザボーテイングによる OR 探索で特徴歪みに対処しようとする場合には、通常参照信号間の相互類似度が高いと考えられるので、本実験結果から、効率的な探索が期待できると言える。

なお、参照信号どうし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 乗のオーダーで増大する。例えば参照信号が 1000 個の場合、参照信号どうしの類似度の計算に約 50 万回の照合が必要となる。

4.2 マイクロホン収録音による検討

次に、実際の特徴歪みの例として、マイクロホン収録による特徴歪みを取り上げ、提案法の効果を調べた。

本実験では、入力信号を 15 分間の音楽信号とした。これは原信号をスピーカで再生し、静かな室内（通常の実験室であり無響室ではない）でマ

表 1: マイクロホン収録音を用いた実験結果

	精度 [%]		時間 [ms]	
	提案	従来	提案	従来
Mic 0.1 m	98	95	60	49
Mic 0.3 m	98	94	52	44
Mic 1.0 m	97	92	52	52
原信号	100	100	51	49

イクロホンで収録したものである。スピーカを含む再生系は 1 種類とし、マイクロホンも 1 種類としたが、マイクロホンとスピーカとの距離を 0.1 m, 0.3 m, 1.0 m の 3 種類に変えて収録し、計 3 つの信号を用意した。一方、参照信号は原信号とし、15 分間の中からランダムに 10 秒間の区間を 100 個選択した。また、学習は、前記 3 種類の条件のそれぞれについて、入力信号とは別の学習用信号 5 分間を用いて行った。

探索閾値を除く探索のパラメータは前節と同様とした。探索閾値は固定とせず、各条件単位で探索閾値を上下させ各条件について、適合率と再現率が一致した値とした。そしてその一致した値をもって精度の値とした。

表 1 に実験結果を示す。表中で「提案」とは、提案法による結果を示す。提案法では、3 種類の学習結果を反映したものと、原信号のままの合計 4 つの参照信号の OR 探索を行ったものである。ボーテイングは、原信号の特徴ベクトル 1 つにつき 1 票とした。また「従来」とは、原信号のみを用いて探索した場合を表す。

表 1 によれば、確率分布を考慮しない従来の方法に比べて、探索時間の増加は 2 割程度以内であるが、探索誤りをおよそ半減できていることが分かる。

5 むすび

本稿では、既に提案した音響信号の高速 OR 探索法の一応用として、特徴歪みのある場合の音響信号の高速探索法を提案した。提案法は、予め特徴の歪み方の確率分布を学習しておき、その確率分布をヒストグラム上に表現する手法である確

率ディザボータイングを用いて、複数の確率分布を考慮した参照信号について OR 探索を行うものである。実験では、まず OR 探索の基本特性を調べ、参照信号の相互類似度が 0.84 の時には、1つの参照信号についての探索の約 1.1 倍の計算時間で 5つの参照信号についての探索を行えることを示した。次に、実際のマイクロホン収録音を用いて特徴歪みのある音についての探索性能を調べ、確率分布を考慮しない従来の方法に比べて探索時間をあまり増やすことなく、従来よりも探索精度を改善できる可能性があることを示した。

今回の実験は、予備的なものとして、学習とテストとは同一の環境で収録した信号を用いたが、実際にはテスト環境と全く同一の環境で学習できるとは限らない。そのため、今後は、本稿のように特徴歪みを網羅的に学習しておくアプローチがどのような範囲で有効かの検証や、特徴歪みのクラスタリング等についての検討を行う予定である。

謝辞

ご指導頂く 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、萩田紀博部長、およびメディア認識研究グループの諸氏に感謝する。

参考文献

- [1] 柏野 邦夫, ガビンスミス, 村瀬 洋: “ヒストグラム特徴系列に基づく長時間音響信号の高速探索”, 音響講論集 98 秋, pp.561-562 (Sep.1998).
- [2] 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 (Sep.1999).
- [3] 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).
- [4] 杉山 雅英: “ビデオデータにおける指定人物の検出と追跡”, 映像情報メディア学会技術報告, Vol.22, No.66, pp.7-12 (Nov.1998).
- [5] 杉山 雅英: “セグメントの高速探索法”, 信学技報, SP98-141, pp.39-45 (Feb.1999).
- [6] T. Nishimura, N. Sekimoto, J. XinZhang, M. Ihara, T. Akasaka, H. Takahashi, and R. Oka: “Methodology for Retrieving Time Sequence Pattern”, *Proc. IWHIT/SW99*, pp.1-8 (Oct.1999).
- [7] 柏野 邦夫, 村瀬 洋: “複数の音響信号の高速探索”, 音響講論集 99 春, pp.447-448 (Mar.1999).
- [8] 柏野 邦夫, 村瀬 洋: “時系列アクティブ探索法における音響信号の高速 AND/OR 探索”, 信学技報, SP99-23, pp.41-48 (Jun.1999).
- [9] 黒住 隆行, 柏野 邦夫, 村瀬 洋: “時系列アクティブ探索法のための特徴変動に頑健な確率ディザボータイング”, 画像の認識理解シンポジウム MIRU2000, vol.1, pp.313-318 (Aug.2000).