

# 時系列アクティブ探索法のための特徴ひずみに頑健な確率ディザボータイング

黒住 隆行<sup>†</sup>      柏野 邦夫<sup>†</sup>      村瀬 洋<sup>†</sup>

Probabilistic Dither-Voting: Improving Robustness of Time-Series Active Search with Respect to Feature Distortions

Takayuki KUROZUMI<sup>†</sup>, Kunio KASHINO<sup>†</sup>, and Hiroshi MURASE<sup>†</sup>

あらまし 既知の音や映像（目的信号）が長時間の音や映像（蓄積信号）のどの時点にあるか探索する問題（時系列探索）において、高速かつ高精度に探索する手法を提案する。時系列探索における問題点は、音や映像の特徴にビデオのダビングや各種圧縮方式などによるひずみが発生することである。そのようなひずみが激しく起こると、探索精度が低下する。本論文では、そのようなひずみを吸収するための手法として、確率ディザボータイングを提案する。これは、ひずみの確率分布を学習により求め、その確率分布をヒストグラム上に表現するものである。ビデオのダビングや圧縮が起こった映像を探索する実験では、いずれのひずみにおいても探索精度の改善が見られた。例えば、ダビングを4回行った1時間の蓄積信号から5秒の目的信号を探索する場合には、従来法より探索精度が4.5%改善し、本手法の有効性が確認された。本手法により、様々なひずみに対して頑健なマルチメディア探索が可能になると考えられる。

キーワード 検索, 映像, 時系列アクティブ探索法, ベクトル量子化, ノイズ

## 1. ま え が き

近年のネットワークの高速化やコンピュータの高機能化に伴い、インターネット上での情報の流通は、文字情報だけでなく音や映像のデータも大量に流通するようになってきた。このため、音や映像の内容検索や時系列探索の必要性が増している。

ここで内容検索とは、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膨大なデータベースから取得することをいう。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いる。音に関しては、時間領域や周波数領域などの特徴量に基づくもの[1]と、ワードスポッティングに基づくもの[2]、映像に関しては、カット点検出によりワンシーンを切り出した後、色特徴や周波数特徴、動き情報に基づいたインデクシング[3]、ビデオOCRに基づく方法[4]、また、音と映像の両方を用いるもの

として、音声やテロップの認識結果に基づいた手法[5]などが挙げられる。これらの内容検索に用いられている探索手法は、いったんデータベースの情報をインデクシングした後、文字情報のようなシンボリックな情報を検索キーとして探索を行うものである。一方、時系列探索とは、具体的な音や映像（目的信号）を指定して、それに適合する音や映像を膨大なデータベース（蓄積信号）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後者の時系列探索である。このような時系列探索を実現する手法として、特徴を1フレームずつずらしながら照合する特徴ずらし照合法が考えられるが、この方法は探索精度が良いものの探索時間が膨大であるという問題がある。また、時系列特徴に対しいったんベクトル量子化（VQ）を行って符号時系列にした後、BM法[6]のような文字列探索法を用いて探索する方法や文字列探索とユークリッド距離による探索を組み合わせた手法[7]がある。しかし、これらの手法は高速に探索できるが、特徴にひずみがある場合、探索精度が低いという問題がある。インターネット上のコンテンツを探索するような応用を考えた場合、MPEG圧縮やビデオのダビングなど

<sup>†</sup> 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

によるひずみの影響のために、信号が変形してしまい探索精度が低下するという問題がある。そのため、何らかの方法により特徴のひずみを吸収し、高速かつ高精度な探索を実現する必要がある。

ところで、既に我々は音や映像の高速探索の研究を進めており、「時系列アクティブ探索法」(TAS; Time-series Active Search)と呼ぶアルゴリズムを提案した [8]。時系列アクティブ探索法は、探索したい具体的な信号(目的信号)が与えられたとき、それと同一の特徴をもつ信号区間の出現位置を長時間の信号(蓄積信号)中からもれなく検出する手法である。これによれば、あらかじめ特徴抽出などの前処理を行っておけば、例えば 15 秒間の目的信号を一致探索する場合、通常の PC でも処理 1 秒当り 24 時間分の蓄積信号に対する探索を完了できる。この高速性により幅広い応用を考えることができる [9]。

我々が提案した時系列アクティブ探索法では、探索に使用する特徴ベクトルをベクトル量子化(VQ)し、VQ 符号により表現していた。そこで使用した VQ の方法は、ベクトルを各要素ごとにスカラ量子化(SQ)し、その SQ 符号の組合せで表現される領域ごとに VQ していた [8]。しかし、実際の特徴ベクトルは、ベクトル空間上で偏りをもった分布をしており、ベクトルの各要素ごとに分割することは、必ずしも効率的に特徴を表現しているとはいえない。

また、従来の時系列アクティブ探索法は目的信号との類似度があるしきい値以上の区間を検出する手法であるが、特徴がノイズによって変形する場合、例えば、ノイズによって画像の輝度や色が激しく変化した場合、著しく類似度が低下し検出洩れを起こす可能性があった。しかし、これではインターネット上での映像の使用をチェックするような応用を考えた場合、ビデオデッキによるダビングや MPEG によって圧縮された映像を検出することが非常に困難である。

本論文では、以上の二つの問題点を解決するため、次の二つの改良を行う。前者の問題の解決には、特徴ベクトルの分布に基づいて VQ することにより、効率的に VQ コードを割り当て、探索精度を向上させる。一方、後者の問題の解決には、目的信号、蓄積信号に含まれるダビングや圧縮等によるノイズの確率分布をヒストグラム上に表現し、それを探索に用いることで探索精度を向上させる [10], [11]。特に後者は、様々なノイズに対応した目的信号を OR 探索 [12], [13] することで、ひずみに頑健な時系列探索を可能にする考

えられる。

## 2. 時系列アクティブ探索法

### 2.1 時系列アクティブ探索法の概要

以下に時系列アクティブ探索法の概要を簡単に要約する。処理の流れを図 1 に示す。まず、目的信号(探索のキーとなる短時間の信号)と蓄積信号(長時間の信号)からそれぞれ特徴を抽出する。次に、目的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 VQ して、各 VQ コード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(探索しきい値)を超えるかどうかで、目的信号の有無を判定する。このとき、類似度と探索しきい値とから、探索位置をスキップさせても目的信号を見逃さないことが保証されたスキップ幅  $w$  を求めることができる [8]。そこで、 $w$  だけ蓄積信号に対する窓をずらして探索を進める。例えば、類似度としてヒストグラムの重なり率  $S$  を用いた場合、

$$S = S(H_Q, H_S) = \frac{1}{D} \sum_{l=1}^L \min(h_{Ql}, h_{Sl}) \quad (1)$$

である。ここで、 $H_Q, H_S$  は、それぞれ目的信号と蓄積信号の照合区間に対するヒストグラムであり、 $h_{Ql}, h_{Sl}$  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 VQ 符号帳のサイズ(ヒストグラムのビンの数)、 $D$  はヒストグラムの総度数(照合区間の長さ)である。このとき、その時刻における類似度があらかじめ決めておいた探索しきい値  $\theta$  を下回っていると、次式で表される区間  $w$  だけ探索点をスキップさせても目的信号を見逃さないことが保証されている。

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

ただし、 $[\cdot]$  は切下げを表す。

### 2.2 映像特徴の抽出

時系列アクティブ探索法は、音や映像のような時系列信号であれば適用できる手法 [14] であるが、ここでは映像探索について扱う。本論文では、映像特徴としてカラー縮小画像を用いた。ここで、映像特徴ベクトル  $x(k)$  を、

$$x(k) = (x_{1r}(k), x_{1g}(k), x_{1b}(k), \dots, x_{jc}(k), \dots, x_{wr}(k), x_{wg}(k), x_{wb}(k)) \quad (3)$$

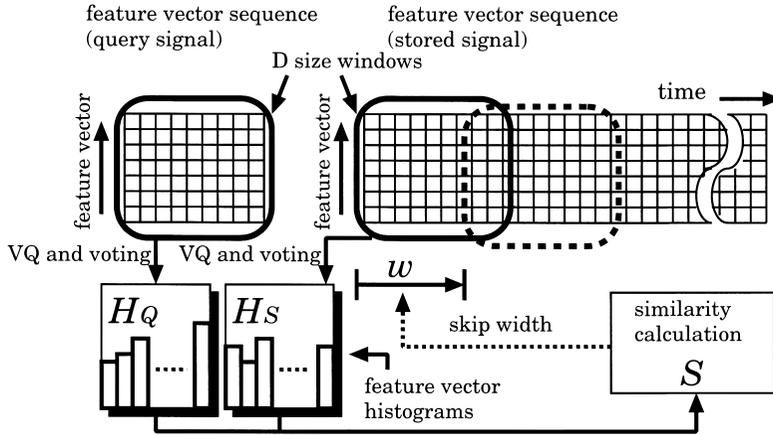


図1 時系列アクティブ探索法の概要  
Fig. 1 Overview of time-series active search.

と定義する．ここで  $k$  はフレームの時刻であり， $x$  の添え字  $j$  は各フレームの画像を  $W$  個のサブ画像に分割した分割番号，添え字  $c$  は RGB を表す．ただし， $W$  は経験的に与えられる値である．なお，一般に， $W$  を大きくすると探索精度が向上するが，特徴抽出等の処理速度が低下する． $x_{jc}$  は各画素の輝度値をサブ画像内で RGB のうちの単一色に関して平均し正規化した値であり，

$$x_{jc}(k) = \frac{\bar{y}_{jc}(k) - \min_i \bar{y}_{ic}(k)}{\max_i \bar{y}_{ic}(k) - \min_i \bar{y}_{ic}(k)} \quad (4)$$

である．ただし，

$$\bar{y}_{ic}(k) = \frac{1}{|I|} \sum_{p \in I} y_{pc}(k) \quad (5)$$

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， $|I|$  は  $I$  の画素数， $y_{pc}(k)$  は画素  $p$  の RGB のうちの単一色  $c$  の画素値を表す．

### 3. 特徴ベクトルの分布とひずみの学習

#### 3.1 VQ

実際の映像から抽出された特徴ベクトルは，ベクトル空間上での分布に偏りがある．特に，RGB の値の相関は高いということは，一般的によく知られている．従来のアクティブ探索法では，特徴ベクトルの各要素ごとに等間隔に SQ し，その SQ 符号の組合せで表現される領域ごとに VQ していた．しかし，その手法よりも，実際の特徴ベクトルの分布に従って出現頻度の

高い領域を細かく分割し，出現頻度の低い領域を粗く分割する方が効率的に符号を与えることができると考えられる．本論文では，特徴ベクトルの分布に基づいて VQ する方法として，代表的なクラスタリングアルゴリズムである Linde-Buzo-Gray (LBG) アルゴリズムを用いて代表ベクトルを求めた．ここで，LBG アルゴリズムとは，学習サンプルに対し階層的にクラスタ数を増やしながらか平均アルゴリズムを用いてクラスタリングを行うアルゴリズムである [15]．また，特徴ベクトルを距離が最も近い代表ベクトルで近似することで量子化を行った．このように実際の特徴ベクトルの分布に基づいて VQ することにより，より精度の高い探索ができると考えられる．

#### 3.2 確率ディザボータイング

確率ディザボータイングとは，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．図 2 のように，入力特徴ベクトルをある確率分布に従ってばらつかせ，確率分布を複数の特徴ベクトルを用いて表現する．そして，それらの特徴ベクトルを VQ する．このように，特徴ベクトルを確率分布に基づいてばらつかせると，ノイズ等のひずみによって投票される可能性があるところに投票され，ひずみに頑健になると考えられる．なお，投票する票の数は，一つの特徴ベクトル当たり複数用いた方が，確率分布をより反映するが，高速性を維持するためには，一つの特徴ベクトル当たり少数の投票を行った方がよい．

同様に複数のビンに投票する手法として杉山は Fuzzy Active 探索法を提案している [16]．しかし，Fuzzy Ac-

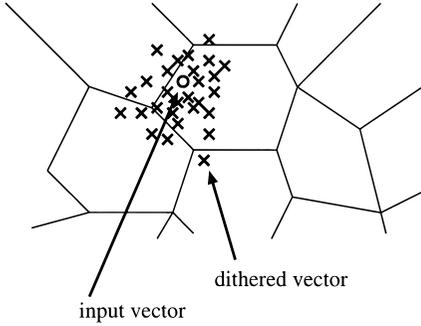


図2 確率分布に基づいてばらつかせた特徴ベクトル  
Fig. 2 An example of dithering feature vectors based on the probabilistic distribution.

tive 探索法は特徴ベクトルと周囲のピンに対応する代表ベクトルとの距離計算に基づき、実数で定義された計数を複数のピンに投票する必要があるために計算時間がかかる。それに対し、提案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確率的に投票を行う手法で、各ピンの計数を整数のまま扱えるため高速なヒストグラム作成が可能である。

ここで、特徴ベクトルの確率分布について考察する。目的信号と蓄積信号ともに、探索が行われるまでに、何らかの要因によりひずみが発生している。例えば、ビデオのダビングでは、録画・再生を繰り返すことによるビデオ信号の輝度、色の信号の劣化によるひずみ、JPEG や MPEG では、量子化誤差によるモスキートノイズやブロックノイズなどがある。これらのひずみは発生のメカニズムや性質が全く異なるが、これらのひずみのすべてを詳細にモデル化し考慮すると、処理が複雑になってしまうという問題がある。そこで、簡単のため、ノイズは加法性雑音、システムは線形システムと仮定し、すべてのひずみを同一のモデルにより表現する。探索に用いる目的信号は、ビデオテープやテレビの放送などからキャプチャ装置を介して得られたものである。図3のように、元信号  $x(k)$  がビデオデッキやテレビの受信器、PC のキャプチャ装置等のシステム  $Z_Q$  を介して、目的信号  $x_Q(k, t_Q)$  が得られる。すなわち、

$$x_Q(k, t_Q) = x(k) + \varepsilon_Q(k + t_Q) \quad (6)$$

である。ただし、 $k$  はフレームの時刻、 $t_Q$  は目的信号のキャプチャを開始した時刻、 $\varepsilon_Q(t)$  は、システム  $Z_Q$  によるノイズを表す確率ベクトルである。一方、ビデオのダビングを複数回行ったり、MPEG による

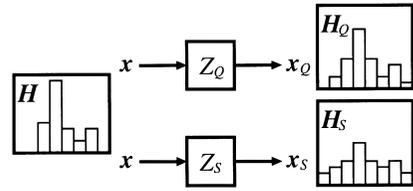


図3 元信号と目的信号、蓄積信号の関係  
Fig. 3 Relationship between a query signal and a stored signal.

圧縮を行う場合に相当するシステム  $Z_S$  を介して、蓄積信号  $x_S(k, t_S)$  を得る。すなわち、

$$x_S(k, t_S) = x(k) + \varepsilon_S(k + t_S) \quad (7)$$

である。ただし、 $t_S$  は蓄積信号のキャプチャを開始した時刻、 $\varepsilon_S(t)$  は、システム  $Z_S$  によるノイズを表す確率ベクトルである。ここで、照合区間の長さを  $D$  とすると、 $x(k)$  から  $x(k + D)$  までのヒストグラム  $H(k)$  の VQ コード  $l$  の計数  $h_l(k)$  は、

$$h_l(k) = \sum_{n=k}^{k+D} \int \delta(\mathbf{y} - \mathbf{x}(n)) \phi_l(\mathbf{y}) d\mathbf{y} \quad (8)$$

と表される。ただし、 $\delta(\mathbf{y})$  はデルタ関数、

$$\phi_l(\mathbf{y}) = \begin{cases} 1 & (\mathbf{y} \text{ が VQ コード } l \text{ の領域内}) \\ 0 & (\text{上記以外}) \end{cases} \quad (9)$$

とする。ここで、目的信号  $x_Q(k, t_Q)$  のヒストグラム  $H_Q(k, t_Q)$  の VQ コード  $l$  の計数  $h_{Ql}(k, t_Q)$  の期待値は、

$$\begin{aligned} E\{h_{Ql}(k, t_Q)\} &= \sum_{n=k}^{k+D} \int E\{\delta(\mathbf{y} - \mathbf{x}_Q(n, t_Q))\} \\ &\quad \times \phi_l(\mathbf{y}) d\mathbf{y} \\ &= \sum_{n=k}^{k+D} \int \int \delta(\mathbf{y} - \mathbf{x}(n) - \varepsilon_Q) \\ &\quad \times P_Q(\varepsilon_Q) \phi_l(\mathbf{y}) d\varepsilon_Q d\mathbf{y} \\ &= \sum_{n=k}^{k+D} \int P_Q(\mathbf{y} - \mathbf{x}(n)) \phi_l(\mathbf{y}) d\mathbf{y} \end{aligned} \quad (10)$$

となる。ただし、 $E$  は  $t_Q$  に関する平均、 $P_Q(\varepsilon_Q)$  は  $\varepsilon_Q$  の確率密度関数とする。同様に、蓄積信号  $x_S(k, t_S)$

のヒストグラム  $H_S(k, t_S)$  の VQ コード  $l$  の計数  $h_{Sl}(k, t_S)$  の期待値は,

$$E\{h_{Sl}(k, t_S)\} = \sum_{n=k}^{k+D} \int P_S(\mathbf{y} - \mathbf{x}(n)) \phi_l(\mathbf{y}) d\mathbf{y} \quad (11)$$

となる。ただし、 $P_S(\epsilon_S)$  は  $\epsilon_S$  の確率密度関数とする。ここで、一般的に  $P_Q(\mathbf{y}) \neq P_S(\mathbf{y})$  なので、 $H_Q(k, t_Q)$  と  $H_S(k, t_S)$  の類似度  $S(H_Q(k, t_Q), H_S(k, t_S))$  の期待値は、たとえ目的信号と蓄積信号の元信号が等しかったとしても 1 とはならない。そこで、図 4 のように目的信号  $x_Q(k, t_Q)$  にシステム  $Z_S$  を介して変形目的信号  $x_{QS}(k, t_Q)$ 、蓄積信号  $x_S(k, t_S)$  にシステム  $Z_Q$  を介して変形蓄積信号  $x_{SQ}(k, t_S)$  を生成する。ここで、変形目的信号  $x_{QS}(k, t_Q)$  のヒストグラム  $H_{QS}(k, t_Q)$  の VQ コード  $l$  の計数  $h_{QSl}(k, t_Q)$  の期待値は、

$$E\{h_{QSl}(k, t_Q)\} = \sum_{n=k}^{k+D} \int P_{QS}(\mathbf{y} - \mathbf{x}(n)) \phi_l(\mathbf{y}) d\mathbf{y} \quad (12)$$

となる。ただし、 $P_{QS}(\epsilon_{QS})$  は  $\epsilon_{QS}(= \epsilon_Q + \epsilon_S)$  の確率密度関数とする。同様に、変形蓄積信号  $x_{SQ}(k, t_S)$  のヒストグラム  $H_{SQ}(k, t_S)$  の VQ 符号  $l$  の計数  $h_{SQt}(k, t_S)$  の期待値は、

$$E\{h_{SQt}(k, t_S)\} = \sum_{n=k}^{k+D} \int P_{SQ}(\mathbf{y} - \mathbf{x}(n)) \phi_l(\mathbf{y}) d\mathbf{y} \quad (13)$$

となる。ただし、 $P_{SQ}(\epsilon_{SQ})$  は  $\epsilon_{SQ}(= \epsilon_S + \epsilon_Q)$  の確率密度関数とする。ここで、システム  $Z_S$  と  $Z_Q$  は可換であるので、 $P_{QS}(\mathbf{y}) = P_{SQ}(\mathbf{y})$  となる。よ

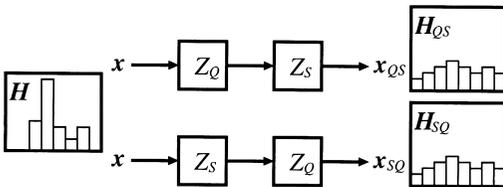


図 4 元信号と変形目的信号と変形蓄積信号の関係  
Fig. 4 Relationship between the modified query signal and the modified stored signal.

って、目的信号と蓄積信号の元信号が等しければ、 $S(H_{QS}(k, t_Q), H_{SQ}(k, t_S))$  の期待値は 1 となる。

以上の考察に基づき、確率ディザボータイングを用いて変形目的信号と変形蓄積信号のそれぞれを得る。つまり、目的信号の特徴ベクトルを  $\epsilon_S$  の確率分布に従ってばらつかせて、変形目的信号を表現する。一方、蓄積信号の特徴ベクトルを  $\epsilon_Q$  の確率分布に従ってばらつかせて、変形蓄積信号を表現する。このようにして得られた変形目的信号、変形蓄積信号それぞれの特徴ベクトルを VQ し、それを探索に用いる。

### 3.3 確率分布の学習

ある信号が、システムを介して変形する場合、そのひずみの確率分布は一般的に求めることは困難である。そこで、本論文では簡単のため、システムを介した信号は、正規分布に従って変形すると仮定する。すなわち、 $\epsilon_Q, \epsilon_S$  は正規確率過程とする。

$x(k)$  をビデオやテレビ放送からの信号とすると、通常  $x(k)$  を直接観測することは不可能である。よって、 $x_Q(k, t_Q)$  と  $x(k)$  から  $\epsilon_Q$  の特性を観測するのではなく、別々にとった、 $x_Q(k, t_Q)$  から  $\epsilon_Q$  の特性を観測する。

まず、 $\epsilon_Q$  の特性の学習用の信号として、同一の元信号からシステム  $Z_Q$  を介して得た信号を二つ用意する。ただし、二つの信号それぞれを  $x_Q(k, t_{Q1}), x_Q(k, t_{Q2})$  とする。ここで、それぞれのフレームごとの  $x_Q(k, t_{Q1})$  と  $x_Q(k, t_{Q2})$  との差分値から、平均 2 乗誤差  $MSE(x_Q(k, t_{Q1}), x_Q(k, t_{Q2}))$  を求める。ただし、

$$\begin{aligned} MSE(\mathbf{x}(k), \mathbf{y}(k)) &= E\{(\mathbf{x}(k) - \mathbf{y}(k)) \cdot (\mathbf{x}(k) - \mathbf{y}(k))\}, \quad (14) \end{aligned}$$

$E$  は  $k$  についての平均である。ここで、 $\epsilon_Q$  の分散を  $\sigma_Q^2$  とすると、

$$\begin{aligned} MSE(x_Q(k, t_{Q1}), x_Q(k, t_{Q2})) &= E\{\epsilon_Q(k, t_{Q1})^2\} \\ &\quad - 2E\{\epsilon_Q(k, t_{Q1}) \cdot \epsilon_Q(k, t_{Q2})\} \\ &\quad + E\{\epsilon_Q(k, t_{Q2})^2\} \\ &= E\{\epsilon_Q(k, t_{Q1})^2\} + E\{\epsilon_Q(k, t_{Q2})^2\} \\ &= 2\sigma_Q^2 \quad (15) \end{aligned}$$

となる。同様に、 $\epsilon_Q, \epsilon_S$  の特性の学習用の信号として、元信号からシステム  $Z_Q, Z_S$  それぞれを介して

得た信号  $x_Q(k, t_Q)$ ,  $x_S(k, t_S)$  を用意する．ここで、平均 2 乗誤差は

$$\begin{aligned} \text{MSE}(x_Q(k, t_Q), x_S(k, t_S)) \\ &= E\{\varepsilon_Q(k, t_Q)^2\} + E\{\varepsilon_S(k, t_S)^2\} \\ &= \sigma_Q^2 + \sigma_S^2 \end{aligned} \quad (16)$$

となる．ただし、 $\varepsilon_Q$ ,  $\varepsilon_S$  の分散をそれぞれ  $\sigma_Q^2$ ,  $\sigma_S^2$  とする．

式 (15), 式 (16) のそれぞれは前もって学習により求める必要がある．平均 2 乗誤差の計算は、二つの信号を用意し、それぞれのフレームを対応させてから平均 2 乗誤差を計算する．この平均 2 乗誤差から計算して得た  $\sigma_Q^2$ ,  $\sigma_S^2$  を分散とする正規乱数を用いて目的信号と蓄積信号のそれぞれをばらつかせて確率ディザポータリングを行う．

## 4. 実 験

### 4.1 VQ

特徴ベクトルの分布に基づく VQ の有効性を示すため、テレビ放送の録画を用いて探索実験を行った．2000 年 5 月 1 日に、テレビ放送を録画し、異なる CM をつなげて 1 時間に編集した．これは一つの目的信号に対し探索対象の映像が 2 度以上出現しない方が実験で正解を集計する際に都合がよいためである<sup>(注1)</sup>．この映像を再生してワークステーションに 2 回に分けて取り込んだ<sup>(注2)</sup>．このうち的一方から、一定の時間区間をランダムな場所から切り出して目的信号とし、他方を蓄積信号として探索を行った．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．本実験では、探索しきい値を次式によって定めた．

$$\theta = m + t\nu. \quad (17)$$

ここで、 $m$  と  $\nu$  はそれぞれ与えられた目的信号に対して蓄積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の平均と標準偏差であり、 $t$  は経験的に与えられる係数である．ただし、式 (17) において、 $\theta$  が 1 を超えるときは、 $\theta = 1$ 、0 を下回るときは  $\theta = 0$  とした．本実験では、式 (17) における  $t$  の値を 200 回繰り返して中一定とし、その値を適合率 (precision rate) と再現率 (recall rate) がほぼ等しくなるように調節し、その適合率と再現率の値を精度とした．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．

$x(k)$  の各要素を 0.5 を境界として等分割しそれぞれの領域で VQ を行う方法と、LBG アルゴリズムにより代表ベクトルを求め、特徴ベクトルを距離が最も近い代表ベクトルで近似する方法のそれぞれを用いて探索精度を測定した．ただし、式 (3) におけるサブ画像数  $W$  は縦方向に 2 分割、横方向に 2 分割としたので、 $W = 4$  とした．また、分布に基づく VQ の代表ベクトルは、蓄積信号とは別の 20 分の CM をつなげた映像を用いて LBG アルゴリズムにより学習した．なお、代表ベクトル数は 4096 とした．ここで、 $W = 4$  とした理由は、VQ 符号数を等分割に基づく VQ と分布の偏りに基づく VQ で等しくし、VQ 符号数の差が評価に影響しないようにするためである．すなわち、本実験では等分割に基づく VQ は各要素ごとに 2 分割したので、 $2^{4 \times 3}$  の VQ 符号数となり、 $W = 4$  とした場合にそれぞれの VQ 符号数が等しくなる．図 5 に等分割に基づく VQ と分布の偏りに基づく VQ を行ったときそれぞれの目的信号の長さや探索精度の関係を示す．これによれば、1.0~20.0s のいずれにおいても、数%の探索精度の改善が見られる．特に、等分割に基づく VQ では目的信号の長さが 2.0s のとき探索精度が 88.6% であるのに対し、特徴ベクトルの分布に基づ

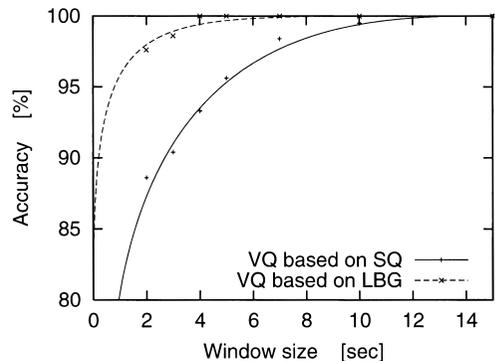


図 5 照会区間の長さや探索精度の関係  
Fig. 5 Window size vs. search accuracy.

(注 1) : ニュース番組や討論番組のように似たような画面構成が多く含まれる映像を用いて探索した場合、本論文の実験結果よりも探索精度が低下する場合がありますので、注意が必要である．

(注 2) : 取込みはフレームレート 29.97 Hz、フレームごとに JPEG 圧縮、画像サイズ 320 × 240 で行った．なお、ビットレートは 3.572 Mbit/s である．

く VQ では探索精度が 97.6% となり, 9.0% の探索精度の改善が見られる .

以上の実験から, 特徴ベクトルの分布に基づいて VQ することの有効性が明らかとなった .

#### 4.2 確率ディザボータイング

次に, 確率ディザボータイングにより, ダビングに対する探索の頑健性を示すため, 実験を行った . 4.1 で用いた 1 時間の映像と同じものを (A) 探索精度の実験用, それとは別の 20 分の映像を (B) ノイズの分散の学習用として用意した . (A)(B) それぞれの映像について,

(1) そのままワークステーションに取り込んだもの,

(2) (1) とは別にそのままワークステーションに取り込んだもの,

(3) ビデオデッキで 2 回ダビングしてからワークステーションに取り込んだもの,

(4) ビデオデッキで 4 回ダビングしてからワークステーションに取り込んだもの,

(5) (1) を JPEG で圧縮したもの,

(6) (1) を MPEG で圧縮したもの

の合計 12 本用意した<sup>(注3)</sup> .

まず, 4 回ダビングした映像 (A-4) を蓄積信号として実験を行った . 目的信号のノイズの分散  $\sigma_Q^2$  については (B-1) と (B-2) の平均 2 乗誤差から式 (15) により計算し, 蓄積信号のノイズの分散  $\sigma_S^2$  については (B-1) と (B-4) の平均 2 乗誤差から式 (16) により計算した . なお, 式 (16) における  $\sigma_Q^2$  は, (B-1) と (B-2) の平均 2 乗誤差から計算された目的信号のノイズの分散を用いた . ただし, 式 (3) におけるサブ画像数  $W$  は縦方向に 3 分割, 横方向に 4 分割としたので,  $W = 12$  とした . また, 分布に基づく VQ の代表ベクトルは, 4.1 で VQ の学習に用いた映像と同様のものを用いて LBG アルゴリズムにより学習した . なお, 代表ベクトル数は 4096 とし, 確率ディザボータイングは一つの特徴ベクトル当り 10 票投票した . また, ディザリング処理は正規乱数テーブルを事前に作成して行った . 処理時間は, SGI 社  $O_2$  R10000 (250 MHz) を用いた場合, 1 時間の特徴ベクトル列をディザリングするのに 13.6 秒かかった . 図 6 に目的信号を (A-1), 蓄積信号を (A-4) としたときの, 入力特徴ベクトルをそのまま VQ したものの (従来法) と確率ディザボータイングを行ったもののそれぞれの照合区間の長さ探索精度の関係を示す . これによれば, 1.0 ~ 20.0 s のい

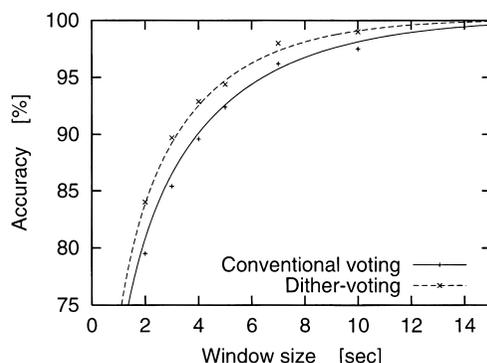


図 6 照合区間の長さ探索精度の関係  
Fig. 6 Window size vs. search accuracy.

表 1 ダビング, 圧縮を行った映像での探索精度  
Table 1 Search accuracy for dubbed or compressed videos.

	従来法 [%]	提案法 [%]
別々にキャプチャ	100.0	100.0
2 回ダビング	92.3	95.5
4 回ダビング	92.4	94.4
JPEG 圧縮	99.4	99.5
MPEG 圧縮	98.8	99.4

れにおいても, 数% の探索精度の改善が見られる . 特に, 従来法では目的信号の長さが 2.0 s のとき探索精度が 79.5% であるのに対し, 確率ディザボータイングでは探索精度が 84.0% となり, 4.5% の探索精度の改善が見られる .

次に, 目的信号を (A-1), 蓄積信号を (A-2,3,4,5,6) とし, 照合区間の長さを 5.0 s に固定して実験を行った . 表 1 に, その結果を示す . ただし, (A-5,6) を蓄積信号とする実験については, 目的信号のノイズの分散を  $\sigma_Q^2 = 0$  としている . いずれのノイズにおいても, 従来法に比べて確率ディザボータイングを用いた方が探索精度が数% 向上している . また, よく似た二つのシーンが含まれる場合に, 従来法で判別できていた映像が, 確率ディザボータイングにより判別ができなくなる可能性も考えられる . そこで, 表 1 の 4 回ダビングの実験において, 従来法の探索の類似度が探索しきい値  $\theta$  から  $\theta - 0.1$  の範囲の映像 163 件について, 確率ディザボータイングを行った場合に類似度のどのように変化するかを調べた . その結果, 確率ディザボータイングを行わなかった場合, 探索しきい値と

(注 3): (5) は  $64 \times 48$  画素に縮小した後, ビットレート 330 kbit/s の JPEG 圧縮を行った (6) は  $64 \times 48$  画素に縮小した後, ビットレート 100 kbit/s の MPEG 圧縮を行った .

の差分値  $\theta - S$  の平均値は 0.061 であったのに対し、確率ディザボータリングを行った場合には、0.101 に上昇した。よって、確率ディザボータリングにより判別性能が平均的に向上することが確認された。以上の実験から、確率ディザボータリングにより特徴ベクトルをシステムのノイズの特性に基づいてばらつかせることにより頑健性が向上することが明らかとなった。

## 5. む す び

本論文では、時系列アクティブ探索法において、特徴ベクトルの分布に基づいて VQ することにより、効率的に VQ コードを割り当てられ、探索精度が向上することを示した。また、蓄積信号に含まれるビデオのダビングや MPEG などの圧縮等によるノイズの確率分布を学習により求め、その確率分布を確率ディザボータリングを用いてヒストグラム上に表現することで、様々なノイズに対応した探索が可能になることを示した。なお、本論文で示した加法性の正規分布を用いたひずみモデルはモデルの一例にすぎず、必ずしも最適なモデルとはいえない。しかし、実験結果が示すように、本論文で扱った範囲のひずみでは、このようなモデル化を行うことに合理性があると考えられる。

本論文で検討したような、ノイズに頑健な探索は、ダビングや圧縮等で様々に変化したマルチメディアデータの高速探索の基礎になると考えられる。また、本手法は、時系列探索に限らず、静止画の探索 [17], [18] にも適用できると考えられる。

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、管村昇部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] 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.
- [2] S.R. Subramanya, R. Simha, B. Narahari, and A. Youssef, "Transform-based indexing of audio data for multimedia databases," *Proc. IEEE Conf. Multimedia Computing and Systems*, no.3, pp.211-218, 1997.
- [3] Y. Gong, G. Proietti, and C. Faloutsos, "Image indexing and retrieval based on human perceptual color clustering," *Proc. CVPR*, pp.578-583, 1998.
- [4] H.D. Wactlar, M.G. Christel, and A.G. Hauptmann, "Lessons learned from building a terabyte digital video library," *Computer*, pp.66-72, Feb. 1999.
- [5] 鷹尾誠一, 舟本純一, 有木康雄, 緒方 淳, "ニュースデー

タベースに対するクロスメディア検索," *MIRU2000*, July 2000.

- [6] R.S. Boyer and J.K. Moore, "A fast string searching algorithm," *Commun. ACM*, vol.20, no.10, pp.762-772, 1977.
- [7] E.J. Keogh and M.J. Pazzani, "An indexing scheme for fast similarity search in large time series database," *Proc. ICSSDM*, pp.56-67, 1999.
- [8] 柏野邦夫, ガビンズミス, 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] T. Nishimura, N. Sekimoto, J. Xin Zhang, M. Ihara, T. Akasaka, H. Takahashi, and R. Oka, "Methodology for retrieving time sequence pattern," *IWHIT/SW'99*, Oct. 1999.
- [10] K. Kashino, T. Kurozumi, and H. Murase, "Feature fluctuation absorption for a quick audio retrieval from long recordings," *Proc. ICPR-00*, vol.3, pp.102-105, Sept. 2000.
- [11] 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リング," *MIRU2000*, July 2000.
- [12] 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, June 1999.
- [13] 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.
- [14] 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.
- [15] 守谷健弘, 音声符号化, 電子情報通信学会, 1998.
- [16] 杉山雅英, "セグメントの高速探索法," *信学技報*, SP98-141, Feb. 1999.
- [17] 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.
- [18] 村瀬 洋, 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-D-II, no.9, pp.2035-2042, Sept. 1998.

(平成 12 年 10 月 2 日受付, 13 年 1 月 30 日再受付)



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、同社 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。



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．同年日本電信電  
話(株)に入社．現在，同社 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．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．



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)に入社．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認識  
研究グループリーダー．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 高柳記念電子科  
学技術振興財団高柳記念奨励賞各受賞．情報処理学会，IEEE  
各会員．