

Title	メディアの認識・処理において高速かつ高精度な探索を可能とする特徴の表現方法の研究
Author(s)	黒住, 隆行
Citation	
Issue Date	2007-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/3753
Rights	
Description	Supervisor:小谷 一孔, 情報科学研究科, 博士

博士論文

メディアの認識・処理において高速かつ高精度な探索を 可能とする特徴の表現方法の研究

指導教員 小谷 一孔 准教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

黒住 隆行

2007年9月

要 旨

本論文では、音や映像などの膨大なマルチメディア情報のデータベースの中から目的とするメディア情報を高速かつ高精度に探索するための信号の特徴の表現方法について述べる。

本研究では、目的とするメディア情報を見つけるために、音や映像などの信号を問合せのデータとして用いて検索するというアプローチをとる。このようなアプローチは、実環境中の音や映像そのものをキーとするため、キーワードを用いずに検索を実行することができる。また、与えられたメディアデータがインターネット上でどのように流通しているかを監視するような用途にも利用することで、メディア情報の検索や監視の可能性をさらに拡大することが期待される。

ここで想定している探索のキー（目的信号）は十数秒程度の音や映像の信号の断片であり、探索の対象となるデータベース（蓄積信号）は数万時間規模の膨大な音や映像である。目的信号はテレビ放送からキャプチャされた音や映像のコンテンツの断片や実環境中のスピーカから流れる音やディスプレイに表示されている映像をマイクやカメラでキャプチャして得られる信号の断片であり、蓄積信号は音楽 CD 中の楽曲や映画やテレビ放送などの映像アーカイブなどが想定される。そして、この目的信号を用いて蓄積信号中の同一音源または同一映像信号の存在する箇所を探索し検出することで、目的信号に関連するメタデータや該当する時刻などの情報を取り出すことができる。

このような同一信号を探し出す課題における問題は、目的信号の収録条件によって、信号にさまざまなノイズやひずみなどの変動が含まれてしまうために高い精度で探索することが困難であることと、蓄積信号が長時間になると扱うべき座標や変動パラメータが膨大になるために探索に時間がかかってしまうことである。

本研究では、これらの問題を、縮退特徴、生成特徴、スパースな特徴選択によって解決する。縮退特徴はノイズやひずみなどの変動を吸収して得られる特徴であり、生成特徴はノイズやひずみなどの変動を生成して得られる特徴である。スパースな特徴選択は、頑健な探索を行うために統計量に基づいて信号中の特徴的な箇所を選び出すものである。本論文では、縮退特徴、生成特徴、スパースな特徴選択のそれぞれについて探索実験を行い、その効果について述べる。また、縮退特徴、生成特徴、スパースな特徴選択を組み合わせる縮退生成探索法を提案し、実験によりその効果を確認する。最後に、本研究について総括し、今後の展望と課題について述べる。

Abstract

This thesis describes feature representations of signals used to search media data quickly and accurately from among vastly numerous multimedia data such as audio and video recordings.

The query data used for seeking information are assumed to be audio or video signals. This kind of retrieval approach can be performed without using metadata such as keywords. Moreover, such an approach is applicable to various applications using query signals, such as surveillance of pirated multimedia contents on the internet.

The stored signal in the database is assumed to comprise a huge amount of audio or video data that extend into tens of thousands of hours. The query signal is assumed to be an audio or video segment, such as an approximately ten-second signal captured from audio or video of TV broadcasts, or captured with a speaker and a microphone, or a display and a camera in the real world. The search process using such query signals is obtainable as metadata of the stored signal in the database by detecting and locating a similar signal in the database.

Major problems in this kind of retrieval approach are that the features of the query signal might include noise or distortion caused by environmental conditions and that the calculation time is too long because of the huge parameter space that must be searched.

This paper proposes a solution to those problems using feature degeneration, feature generation, and sparse feature selection. Feature degeneration absorbs noise and distortion; feature generation creates noise and distortion. Moreover, sparse feature selection chooses robust features based on statistical values. This thesis describes feature degeneration, feature generation, and sparse feature selection, along with their effectiveness as shown through experimentation. Subsequently, a combination of feature degeneration, feature generation, and sparse feature selection is proposed along with discussion of present and future work.

目次

1	序論	1
1.1	研究の背景	1
1.2	メディア探索の課題	2
1.3	研究の目的	6
1.4	論文の構成	6
2	メディア探索の問題と解決のアプローチ	12
2.1	メディア探索の問題	12
2.2	探索空間の問題と解決のアプローチ	13
2.2.1	探索空間の問題と従来の探索法	13
2.2.2	解決のアプローチ	14
2.3	変動の問題と解決のアプローチ	17
2.3.1	変動の問題	17
2.3.2	変動の性質と解決のアプローチ	19
2.4	本研究の位置付け	25
3	縮退特徴に基づく探索	29
3.1	はじめに	29
3.2	問題と解決の方針	29
3.3	探索の原理	32
3.3.1	周波数特徴の抽出	32
3.3.2	映像特徴の抽出	33
3.3.3	時間周波数空間上または時間空間上での局所領域ごとの正規化	33
3.3.4	部分空間の構成	34
3.3.5	時系列探索	35

3.4	実験	36
3.4.1	音響信号探索用実験データの収集	36
3.4.2	時間周波数空間上の局所領域ごとの正規化による効果(実験1)	37
3.4.3	部分空間による効果(実験2)	39
3.4.4	騒音下での精度評価(実験3)	40
3.4.5	映像信号探索用実験データの収集	41
3.4.6	時空間上の局所領域ごとの正規化による効果(実験1)	42
3.4.7	部分空間による効果(実験2)	43
3.5	考察	44
3.6	まとめ	44
4	生成特徴に基づく探索	51
4.1	はじめに	51
4.2	問題と解決の方針	51
4.3	時系列アクティブ探索法	53
4.3.1	時系列アクティブ探索法の概要	53
4.3.2	映像特徴の抽出	54
4.4	特徴ベクトルの分布と歪みの学習	55
4.4.1	VQ	55
4.4.2	確率ディザボーティング	55
4.4.3	確率分布の学習	60
4.5	実験	61
4.5.1	VQ	61
4.5.2	確率ディザボーティング	62
4.6	考察	65
4.7	まとめ	66
5	スパースな特徴選択	68
5.1	問題と解決の方針	69
5.2	探索の原理	71

5.2.1	映像特徴の抽出	71
5.2.2	時空間上の局所領域ごとの正規化	72
5.2.3	スパースな特徴選択	72
5.2.4	局所正規化値の量子化	73
5.2.5	量子化値の一致度に基づく時系列探索	73
5.3	実験	73
5.3.1	実験データの収集	74
5.3.2	各処理による効果 (実験 1)	74
5.3.3	長時間の蓄積信号を用いた実験 (実験 2)	76
5.4	合成信号による手法の特性の評価	77
5.5	考察	78
5.6	まとめ	79
6	縮退生成探索	88
6.1	はじめに	88
6.2	問題と解決の方針	88
6.3	探索の原理	90
6.3.1	基本特徴の抽出	91
6.3.2	縮退特徴の抽出	95
6.3.3	変動パラメータの計算	95
6.3.4	蓄積信号の幾何変換	96
6.3.5	類似度の計算	96
6.4	実験	97
6.4.1	提案法による効果	97
6.5	考察	99
6.6	まとめ	99
7	結論	104
7.1	本研究の要約	104
7.2	今後の課題	106
7.3	今後の展望	107

目 次

1.1	時系列探索の概要	2
1.2	携帯電話による楽曲検索の例	9
1.3	携帯電話による動画検索の例	10
1.4	ピクチャインピクチャの例	11
2.1	蓄積特徴	15
2.2	変形して得られた目的特徴	15
2.3	目的特徴と蓄積特徴が一致する箇所	16
2.4	投票空間	17
2.5	蓄積信号から抽出された縮退特徴	18
2.6	目的信号から抽出された縮退特徴	19
2.7	提案法の投票空間上での探索領域	20
2.8	実環境で収録した音響信号の変動の要因	21
2.9	実環境で収録した映像信号の変動の要因	22
2.10	オリジナル信号と欠損した信号	24
2.11	蓄積信号から抽出された縮退特徴	25
2.12	目的信号から抽出された縮退特徴	26
2.13	蓄積信号をスケーリング	27
2.14	蓄積信号をシフト	27
3.1	楽曲の一部分 1	47
3.2	楽曲の一部分 2	47
3.3	音響信号探索の処理手順	48
3.4	映像信号探索の処理手順	49
3.5	音響信号探索の実験装置	49

3.6	映像信号探索の実験装置	50
3.7	探索の例	50
4.1	時系列アクティブ探索法の概要	54
4.2	確率分布に基づいてばらつかせた特徴ベクトル	56
4.3	元信号と目的信号, 蓄積信号の関係	57
4.4	元信号と変形目的信号と変形蓄積信号の関係	59
4.5	照合区間の長さとの探索精度の関係	63
4.6	照合区間の長さとの探索精度の関係	65
5.1	遮蔽がある場合の距離	70
5.2	処理の手順	80
5.3	スパースな特徴選択の例	81
5.4	実験装置	81
5.5	映像探索の例	82
5.6	実験で使った遮蔽物体の配置	83
5.7	適合率と再現率	84
5.8	雑音への耐性	85
5.9	選択数と速度の関係	86
5.10	選択数と精度の関係	87
6.1	蓄積信号から抽出された特徴	89
6.2	変形している目的信号から抽出された特徴	90
6.3	蓄積信号から抽出された縮退特徴	91
6.4	目的信号から抽出された縮退特徴	92
6.5	蓄積信号をスケールリング	93
6.6	蓄積信号をシフト	94
6.7	処理の手順	101
6.8	スパースな特徴選択の例	102
6.9	第二の基本特徴の選択	103

表 目 次

1.1	信号を探索のキーとして探す方法が有効な例	4
1.2	オリジナル信号に対して同一視/区別する基準	6
1.3	オリジナル信号に対して同一視/区別する基準(音の例)	7
1.4	オリジナル信号に対して同一視/区別する基準(映像の例)	8
2.1	探索空間, 変動の一般的性質	13
2.2	変動の性質と要因	23
2.3	変動の性質とその対処法	28
3.1	実環境で収録した音響信号	37
3.2	蓄積信号	38
3.3	実験 1, 2 (予備実験) で使用した音響信号	39
3.4	時間周波数空間上の局所領域ごとの正規化による効果(実験 1)	39
3.5	部分空間による効果(実験 2)	40
3.6	実験 3 で使用した音響信号	41
3.7	探索精度(実験 3)	41
3.8	処理時間	42
3.9	実環境で収録した映像信号	42
3.10	蓄積信号	43
3.11	時空間上の局所領域ごとの正規化による効果(実験 1)	43
3.12	部分空間による効果(実験 2)	44
4.1	ダビング, 圧縮を行った映像での探索精度	66
5.1	実環境で収録した映像信号	74
5.2	蓄積信号	74

5.3	目的信号	75
5.4	各処理の効果(実験 1)	76
5.5	長時間の蓄積信号での探索精度(実験 2)	77
5.6	処理時間(実験 2)	77
6.1	提案法の探索精度	98
6.2	処理時間	99
7.1	縮退生成探索の特性	107
7.2	拡大する研究領域	108

第 1 章

序論

1.1 研究の背景

近年，音や映像などのメディア情報の流通が盛んになってきている．特に，放送の多チャンネル化やビデオオンデマンドの普及，家庭用映像記録媒体の大容量化が進み，我々が視聴できるコンテンツの数が急激に増加している．しかしながら，視聴できるコンテンツの選択肢が増えれば増えるほど，膨大なコンテンツの中から自分の欲しいコンテンツがどこにあるのか探し出すという課題は困難となり，メディアのハンドリングの効率化が求められる．最近では，インターネット上の無料映像投稿サイト [1, 2] や Peer to Peer モデルを利用したファイル共有サービス [3] が登場し，毎日数多くのコンテンツが投稿されている．これらのサイトは，世界中のコンテンツ制作者からさまざまな映像が投稿され，数多くのコンテンツを視聴できる人気サービスである．しかしながら，投稿されるコンテンツの中には著作権を無視したようなものも見られ，社会的な問題となっている．このような問題のコンテンツに対処するために投稿されるコンテンツをひとつひとつ人手でチェックすることは，非常にコストがかかる作業であり，投稿コンテンツの著作権チェックは重要な課題となっている．

このような背景の中，音や映像などのマルチメディア情報の探索技術が極めて重要になってきており，近年では，このようなメディア探索の研究が報告されるようになってきた [4, 5, 6, 7, 8, 9, 10, 11, 12]．ここで，メディア探索で扱う課題は，具体的な信号（目的信号）を指定して，それと類似する信号が膨大なデータ

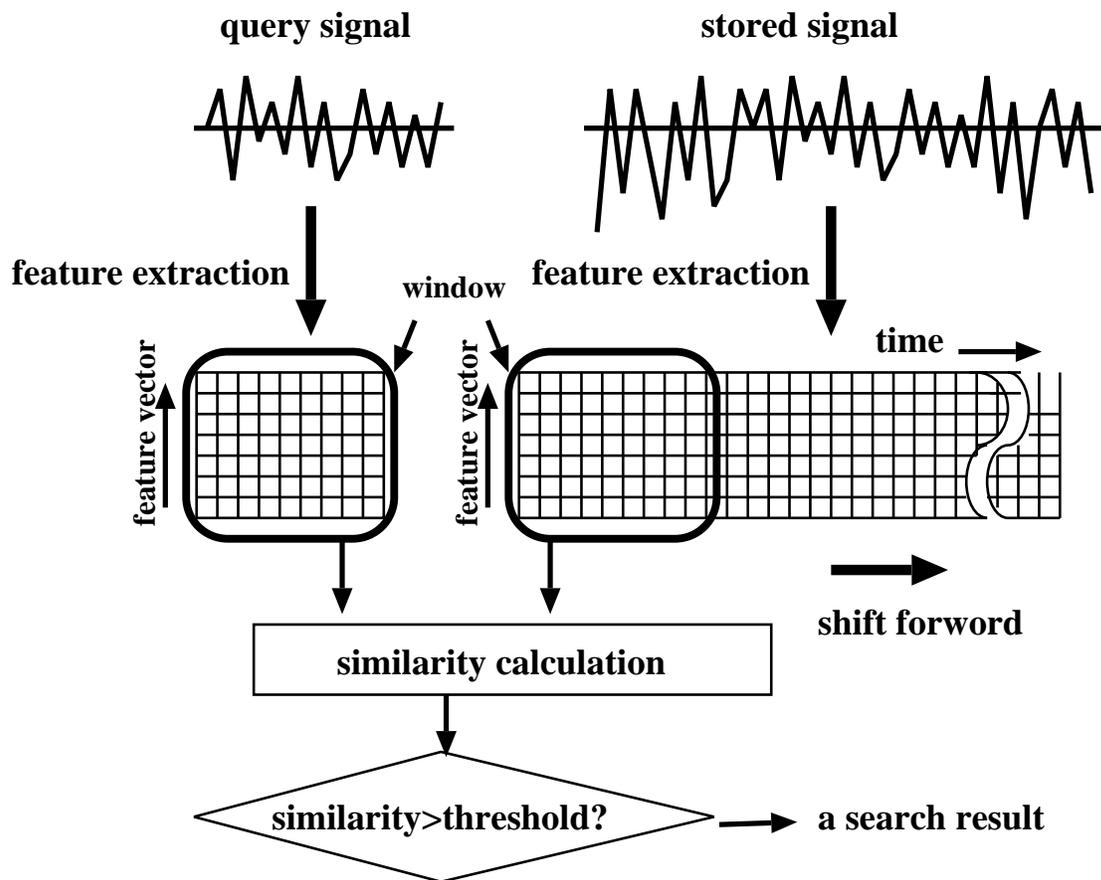


図 1.1: 時系列探索の概要

ベース(蓄積信号)のどこに存在するかを探索するというものである。扱う信号が音や映像などの時系列信号であれば、図 1.1 に示すような探索を行うこととなる。具体的には、目的信号と蓄積信号から特徴抽出を行っておき、蓄積信号中に目的信号と同じ大きさの窓をかけ、目的信号と蓄積信号中の注目箇所との類似度を計算しながら、窓をずらし、目的信号と最も類似した蓄積信号中の箇所を探索結果として出力する時系列探索を行う。

1.2 メディア探索の課題

本研究で対象とするメディア探索は、探索の問合せに信号そのものを用いるものである。音や映像のメディアデータを探し出す従来の方法として、キーワードを

用いる方法 [13] があるが，これは事前に映像に対応するメタデータを準備しておく必要があり，十分管理されたメディアデータにのみ適用できるものである．それに対し，本研究で扱うような探索のキーとして信号を用いる方法では，メタデータが付与されていないメディアデータのように十分な管理下にないデータであっても，そのメディアデータを特定することが可能である．ここで，信号そのものを探索のキーとして探す方法が有効な例を表 1.1 に示す．(1) の音楽のタイトル検索は，手元にタイトルがわからない音楽ファイルがあるような場合に，音楽ファイル中の信号の全て，もしくは，その一部を探索のキーとして利用し，数百万曲に及ぶような楽曲データベースの中のどの部分と一致するかを探索し，同一箇所が見つければ該当する楽曲のタイトルやその楽曲に関するメタデータを取り出すことができるというものである．メディア探索の応用範囲を考えると，必ずしもテレビ放送のような高い品質が確保されたような場面だけにとどまらない．近年では，ノート型コンピュータ，PDA，携帯電話といった携帯端末の普及が著しく，これら携帯端末を用いて音や映像を扱う機会も増えてきている．携帯端末に入力された実環境の音や映像を探索キーとして情報検索ができれば有用性が高い．例えば，(2) のように街角で流れている楽曲を携帯電話に入力することによって，収録された楽曲に関するアーティストやCD，コンサート情報などを検索するといったようなことができると便利である [8, 14, 10, 11] (図 1.2)．また，街頭テレビや家庭用テレビなどに表示されているCM映像を携帯カメラで収録し，収録された映像を用いて検索することによって，映像中のCMに関する情報を検索するといったような応用も想定される [15] (図 1.3)．(3) のテレビ・ラジオ放送の中のCMのカウントは，放送中に何件CMが放送されたかを自動でチェックするものである．(4) のテレビ・ラジオ放送の楽曲プレイリストの作成は，放送された楽曲が何であったかを確かめるために放送された楽曲のリストを自動で作成するものである．(5) のインターネット上の違法コンテンツの監視は，インターネット上で流通しているコンテンツの中に，著作権を無視して流通しているものがないかをチェックするために，音楽や映画などの権利者が既知の著作物のデータベースを利用して自動でチェックするものである．これらの例のように，信号そのものをキーとして探索する手法は，音や映像などの信号が手元にあり，膨大なメディアのデータベースの中から同一のものを探さなければならない場合に有効な方法である．

表 1.1: 信号を探索のキーとして探す方法が有効な例

用途	目的信号(探索のキー)	蓄積信号(データベース)
(1) 音楽のタイトル検索	タイトルのわからない音楽ファイル	楽曲データベース
(2) 携帯電話による音楽検索	街角で流れている音楽	楽曲データベース
(3) テレビ・ラジオ放送の中のCMのカウント	テレビ・ラジオ放送	CMデータベース
(4) テレビ・ラジオ放送の楽曲プレイリストの作成	テレビ・ラジオ放送	楽曲データベース
(5) インターネット上の違法コンテンツの監視	インターネットで流通しているコンテンツ	音楽・映画などの権利者が既知の著作物のデータベース

本研究では、音や映像のメディア探索において探索対象のデータベースとして蓄積される信号は、数万時間を越えるような長時間の音や映像を想定している。具体的には、数ヶ月規模の長時間のテレビ放送や数万本の映画データベース、数百万曲規模の楽曲データベースなどが、そのような蓄積信号の対象となる。一方、探索のキーとなる目的信号は、十数秒程度の音や映像の短い信号の断片である。テレビ放送やFM・AMラジオ放送などでどのようなコンテンツが利用されているかを調べるような応用(プレイリストの自動作成)の場合は、PCに接続されたキャプチャデバイスを介して放送から得られる音や映像の信号の断片が目的信号となる。実環境で収録された信号を用いる場合、目的信号には、収録機器の機器特性や環境雑音による変動、照明や別の物体の写り込みなどが起こり、オリジナルのコンテンツとは大きく異なった信号が入力されることとなる。また、目的信

号として収録の対象となる信号は、必ずしもオリジナルのコンテンツと同一であるとは限らず、編集や加工によって改変されることも有りうる。例えば、放送される映像中にはテロップ表示やピクチャインピクチャ(図 1.4)のような映像編集が頻繁に行われる。映像に限らず、音についても、リミックス版楽曲のように、オリジナルの音楽音源を加工することによって新しい楽曲作品とするようなことも多く行われている。このような場合では、時間を伸縮させるような加工や音のピッチを変更するような加工がしばしば施される。音や映像などのコンテンツの中でどのように他のコンテンツが二次利用されているかを調べるような応用では、このようなさまざまに加工が施されたコンテンツであっても同一のコンテンツとして検出する必要がある。

なお、本論文では、このように実環境で収録された音や映像、加工編集された音や映像を対象とした探索について議論するが、ひとつのマイク、ひとつのカメラで信号を取得することが前提であり、マイクロホンアレーを用いた雑音除去法 [16] のような特殊なデバイスを用いた手法については議論しない。また、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内容検索に関して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 [17] と、ワードスポッティングに基づくもの [18]、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 [19]、ビデオOCRに基づく方法 [20]、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 [21] 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。これらは具体的な音や映像をキーとして検索するものでなく、内容に基づいて検索するものであり、本論文では、このような探索については扱わない。また、編曲された楽曲やハミング探索などの類似音楽探索 [22] や類似画像探索 [23] のような探索についても扱わない。なお、本研究において、オリジナルの信号に対して、同一視するもの、区別するものを、表 1.2 音の例を表 1.3 に、映像の例を表 1.4 にまとめる。これらのように、類似した音や映像であっても、同一のソースでなければ区別するものとする。本論文では、同一視する基準として変

表 1.2: オリジナル信号に対して同一視/区別する基準

変動パラメータによる表現	識別に十分な特徴	同一視/区別
可能	あり	同一視
不可能	あり	区別
可能	なし	区別
不可能	なし	区別

動をパラメータによって表現できるかどうかをひとつの基準とし、特に、音の時間間隔や時間間隔の比(3つの音の相対時刻の比)、映像のカット間隔やカット間隔の比(3つのカット点の相対時刻の比)のように線形変換のような幾何変換があっても保存されるような特徴を持つ信号であり、かつ、識別に十分な特徴が残っていれば同一のものとみなす。

1.3 研究の目的

本研究の目的は、膨大な蓄積信号の中から与えられた目的信号と同一の信号を高速かつ高精度に探索することである。扱う信号は、さまざまな変動を含む信号であるが、同一音源や同一映像であれば、同じものとみなして探索結果として出力することを目指す。

1.4 論文の構成

本論文の構成は次の通りである。第2章では、メディア探索における問題と解決のアプローチについて述べる。第3章では、縮退特徴に基づいた探索手法について述べる。特に、乗法性ひずみと加法性雑音の吸収法について述べ、実験により提案法の有効性を確認した結果を示す。第4章では、生成特徴に基づいた探索手法について述べる。さまざまな変動のモデルを導入しそのモデルに基づいて特徴を生成し探索する手法について述べる。また、映像の探索を例に、さまざまな変動のある信号を探索する実験によりその有効性を確認した結果を示す。第5章

表 1.3: オリジナル信号に対して同一視/区別する基準(音の例)

得られた信号	音の時間間隔	音の時間間隔の比	識別に十分な特徴	同一視/区別
ラジオ放送を介して得られた音楽	変化なし	変化なし	あり	同一視
マイクで収録した街角で流れている音楽	変化なし	変化なし	あり	同一視
時間伸縮加工された音楽	変化あり	変化なし	あり	同一視
ハミングで口ずさんだ音楽	変化あり	変化あり	あり	区別
携帯電話の音量閾値による途切れを含む音	変化なし	変化なし	あり	同一視
大部分がダイナミックレンジを越えてしまい聞き取れないような音	変化なし	変化あり	なし	区別

では、スパースな特徴選択について述べる。統計量に基づいて特徴的な箇所を選択して探索を行う。欠落や遮蔽のある例を音と映像を用いた実験を行った結果を示す。第6章では、縮退生成探索による幾何学的な変動の吸収法について述べる。縮退特徴、生成特徴、スパースな特徴選択、全てを組み合わせた手法であり、これによって幾何学的な変動のある信号を探索し、信号の箇所と変動パラメータの両方を獲得する。第7章では、本研究の要約と今後の展望について述べる。

表 1.4: オリジナル信号に対して同一視/区別する基準(映像の例)

信号	カット点間隔	カット点間隔の比	識別に十分な特徴	同一視/区別
テレビ放送を介して得られた映像	変化なし	変化なし	あり	同一視
ロッキーな映像圧縮により劣化した映像	変化なし	変化なし	あり	同一視
実環境中でカメラで収録された映像	変化なし	変化なし	あり	同一視
スローモーション加工した映像	変化あり	変化なし	あり	同一視
同一人物の出演する別番組	変化あり	変化あり	あり	区別
ピクチャインピクチャによって一部遮蔽されている映像	変化なし	変化なし	あり	同一視
極端に激しいフリッカーを含む映像	変化なし	変化なし	なし	区別



図 1.2: 携帯電話による楽曲検索の例



図 1.3: 携帯電話による動画検索の例



図 1.4: ピクチャインピクチャの例

第 2 章

メディア探索の問題と解決のアプローチ

2.1 メディア探索の問題

1.2 節で述べたようなメディア探索を実現する上での主な問題は、次の 2 つである。

探索空間の問題 音・映像データの膨大な探索空間を高速に探索しなければならないこと

変動の問題 さまざまな収録環境によって発生する信号のノイズやひずみ、コンテンツの編集や加工による変動があっても高精度に探索しなければならないこと

前者に含まれる探索空間には、さまざまな変動を考慮に入れて探索する場合、探索空間は多くの変動のパラメータを含む空間となる。また、時刻や変動パラメータの分解能も制御することができるため、分解能によっては探索空間を大きくすることも小さくすることも可能である。ただし、前者の探索空間は、変動の問題と関わっており、分解能を下げると検出漏れを防げるが誤検出が増え、分解能を上げると誤検出を防げるが検出漏れが増えるという関係にあり、この特性に基づいて問題解決の手法を考えなければならない。また、膨大な探索空間を探索するために、膨大な記憶容量を必要とするような場合もある。そのような場合には、記憶容量の大きさにも考慮が必要である。

表 2.1: 探索空間，変動の一般的性質

	分解能	
	高い	低い
探索空間	大きい	小さい
検出漏れ	増加	減少
誤検出	減少	増加

本章では，これら探索空間，変動についての問題と性質，従来の解決法と本論文のアプローチについて述べる．

2.2 探索空間の問題と解決のアプローチ

2.2.1 探索空間の問題と従来の探索法

さまざまな変動を考慮に入れて探索する場合，探索空間は多くの変動のパラメータを含む空間となり，膨大な大きさの空間を探索しなければならず非常に時間がかかる．また，膨大な探索空間を探索するために，膨大な記憶容量を必要とするような場合もある．そのような場合には，記憶容量の大きさにも注意が必要である．

さまざまな変動パラメータを含む探索空間を探索する従来手法としてハフ変換 [24, 25, 26] に基づいて照合を行う一般化ハフ変換 [27] や幾何的ハッシング [28, 29] が有名である．これはパターンがどういう幾何変換パラメータで変動したかを推定するために，パラメータ空間上の可能性のある座標に投票を行い，得られた得点の高い座標がそのパターンの幾何変換パラメータとなるというものである．ハフ変換に基づいて目的信号と同一の信号を蓄積信号の中から見つけ出す場合，次の手順によって見つけ出すことができる．まず，蓄積信号から抽出した特徴が図 2.1 のように分布しているとする．これに対し蓄積信号を α 倍に伸縮してできた目的信号から抽出した特徴が図 2.2 のように分布しているとする．これらの信号の一致箇所を探すため，同一の特徴を持つ箇所，例えば，同一の周波数を持つ箇所を目的信号と蓄積信号の時間パラメータの空間上に図 2.3 のようにプロットする．

もし同一のパターンで特徴が出現する箇所があれば，図 2.3 のように同一直線上にそのパターンが現れる．その直線を求め目的信号の 0 秒目の直線と交わる点から蓄積信号中の目的信号と同一の信号の出現箇所を求めることができる．ここでこの直線を求めるためにハフ変換を用いる．図 2.4 のような直線の切片と傾きをパラメータにもつ投票空間上で，可能性のあるパラメータの組に投票し，票の大きい座標が決定されるパラメータの組となる．

従来から指摘されているハフ変換の問題は，パラメータの特定に時間がかかることと，投票空間が大きいことである．例えば，オリジナルの信号が時間伸縮の幾何変換によって変動した目的信号をキーとして，蓄積信号中のどこにあるかを探すような場合を考えるとする．目的信号中の特徴の数が N ，蓄積信号中の特徴の数が M ，蓄積信号の時間解像度が V ，時間伸縮解像度が A とし，仮りにハッシュサイズ H のインデックス [30] によって特徴の探索を高速化できるとすると，投票にかかるコストは，

$$O\left(\frac{ANM}{H}\right) \quad (2.1)$$

投票空間のサイズは

$$O(AV) \quad (2.2)$$

である．24 時間分の蓄積信号から 10 秒の目的信号で $\pm 50\%$ の時間伸縮率の範囲を探索する場合，仮に，時間伸縮率の分解能が 0.1% きざみ，時間分解能が 10 ミリ秒きざみで，単位時間あたりの特徴密度が 30 個/秒，ハッシュサイズ 4096 で一様に特徴がインデックスされているとすると，投票にかかるコストは， $1000 \times 300 \times 2.6 \times 10^7 / 4096 = 1.9 \times 10^9$ ，投票空間のサイズ， 3.5×10^9 となり，膨大な時間とメモリ空間が必要となる．

2.2.2 解決のアプローチ

本研究では，このような探索空間が膨大になるという問題に対処するために，縮退特徴および生成特徴を導入し，これらを用いて探索する．

縮退特徴 複数の特徴的な箇所間の相対関係を記述した特徴

生成特徴 目的信号と蓄積信号の相対関係に基づいて目的信号もしくは蓄積信号を幾何変換して生成された特徴

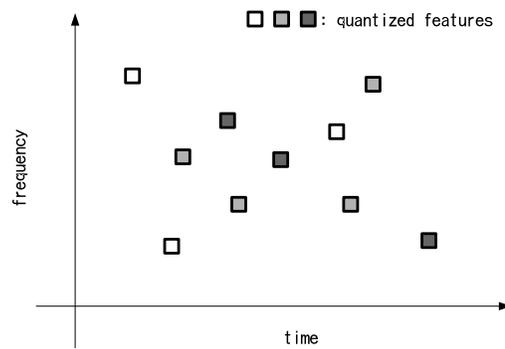


図 2.1: 蓄積特徴

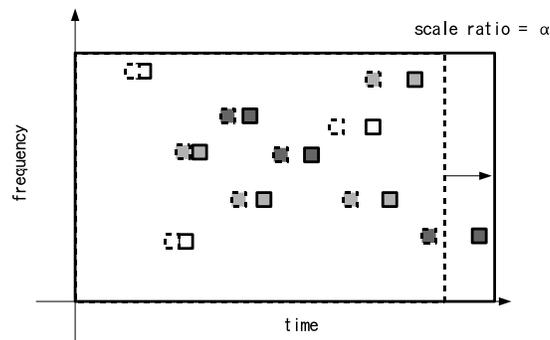


図 2.2: 変形して得られた目的特徴

すなわち、次の手順により類似度を算出する。

1. 目的信号から変動に対して不変な特徴である縮退特徴を抽出，
2. 蓄積信号中の同一縮退特徴を持つ箇所の関係から変動パラメータを計算，
3. 変動パラメータに基づいて蓄積信号から目的信号に整合するように変換して生成特徴を生成，
4. 一致度を計算

例えば、縮退特徴を、特徴から選び出された2つの特徴の組から周波数の比で表すとする(図 2.5, 図 2.6)。そして、同一の周波数比を含む蓄積信号中の箇所を選び出し、目的信号と蓄積信号の関係から得られる変動パラメータを利用して変



図 2.3: 目的特徴と蓄積特徴が一致する箇所

換し該当する箇所のみ照合すればよい．この場合，図 2.7 のように，縮退特徴の一致する箇所のみスコアを計算することとなる．目的信号中の特徴の数が N ，目的信号中の縮退特徴の数が N' ，蓄積信号中の特徴の数が M ，蓄積信号中の縮退特徴の数が M' とし，ハッシュサイズ H' のインデックス [30] によって特徴の探索を高速化できるとすると，投票にかかるコストは，

$$O\left(\frac{NN'M'}{H'}\right) \quad (2.3)$$

である．24 時間分の蓄積信号から 10 秒の目的信号で探索する場合，仮に，単位時間あたりの特徴密度が 30 個/秒，単位時間あたりの縮退特徴密度が 30 個/秒，ハッシュサイズ 4096 で一様に特徴がインデックスされているとすると，投票コストは， $300 \times 300 \times 2.6 \times 10^7 / 4096 = 5.7 \times 10^7$ である．

ここで注目すべきは，式 2.3 には，パラメータの解像度の変数を含まないことである．パラメータ空間が小さい場合には効果はないが，周波数伸縮，時間伸縮，スケール変化，回転など，多くの変動パラメータを扱わなければならないときに，大幅に高速化することができる．また，照合において投票空間を利用してスコアを記憶しておく必要がないため，記憶容量の大幅な削減が可能である．

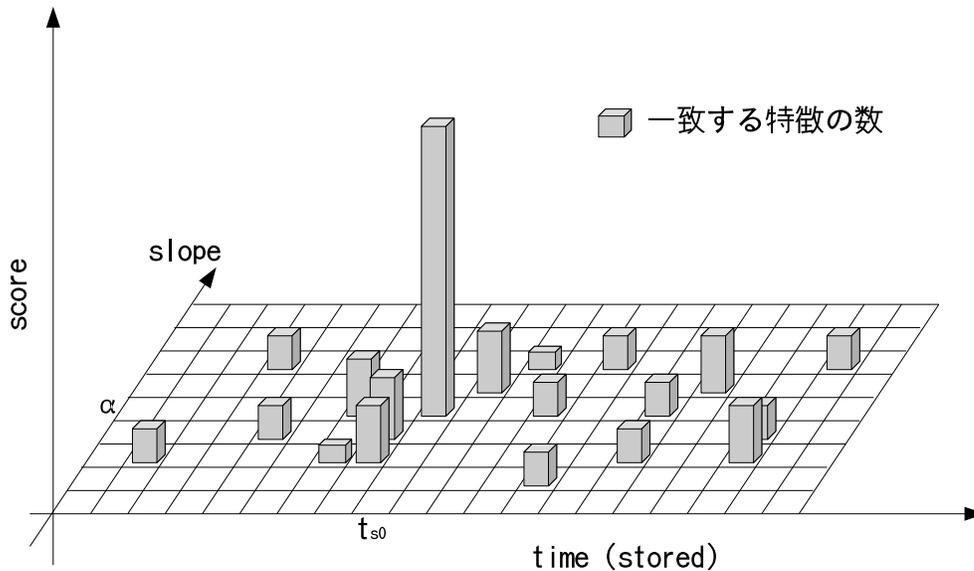


図 2.4: 投票空間

2.3 変動の問題と解決のアプローチ

2.3.1 変動の問題

信号中には、その信号の収録のされ方によってさまざまな変動が含まれる。実環境でマイクで収録したスピーカーの音を目的信号として利用する場合には、その信号に含まれるひずみや雑音として、様々なものが考えられる(図 2.8)。例えば、スピーカーやマイクの機器特性や環境特性の違い(周波数特性の違い)によって発生する乗法性ひずみ、環境雑音によって発生する加法性雑音である。もし、これらのひずみや雑音を全てモデル化し、モデルに基づいてひずみを補正できるならば、探索精度を向上させることができる。しかしながら、現実的には、マイクの特性は既知とできる場合もあるが、スピーカーの特性、環境特性、環境雑音を事前に知ることは難しい。一方、実環境でカメラで収録したディスプレイの映像を目的信号として利用する場合においても、その信号に含まれる変動として、様々なものが考えられる(図 2.9)。例えば、ディスプレイやカメラの機器特性の違いによって発生する乗法性ひずみ、照明変動によって発生する加法性雑音である。こ

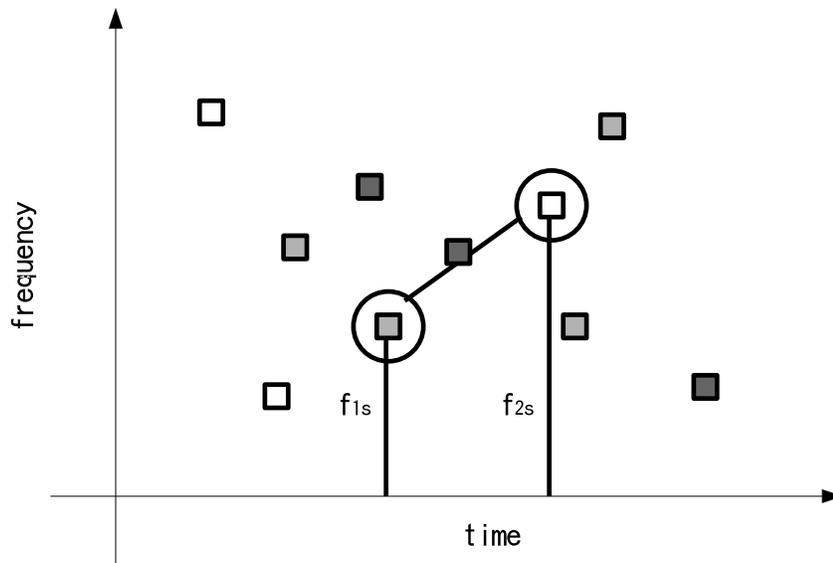


図 2.5: 蓄積信号から抽出された縮退特徴

の場合においても，カメラの特性は既知とできても，ディスプレイの特性や照明変動を事前に知ることは難しい．そこで，信号が入力される収録条件や特性を事前に求めなくとも乗法性ひずみや加法性雑音を吸収し，劣化による変動の少ない特徴を抽出することが必要である．また，携帯電話にある固有の問題として，信号の欠落がある．電波状況が悪いような場合には，音が途切れ途切れになるようなことがある．映像についても同様の信号の欠落がある．例えば，放送される映像中にはテロップ表示やピクチャインピクチャのような映像編集などで，映像が遮蔽され欠落してしまっているような場合がある．このような信号に対しても欠落や遮蔽を回避して探索できる必要がある．また，信号の幾何学的な変動についても無視できない．音の場合は，編集やリミックスなどの人為的な加工において，音響信号の伸縮やピッチの変動などが施されることがある．また，映像の場合は，カメラの配置によって発生する拡大縮小のスケール変化や回転のようなカメラパラメータによるもの，編集による映像の配置の変更などによって幾何学的な変動が起こる．本研究では，以上のような変動(表 2.2)を考慮して探索すること考える．

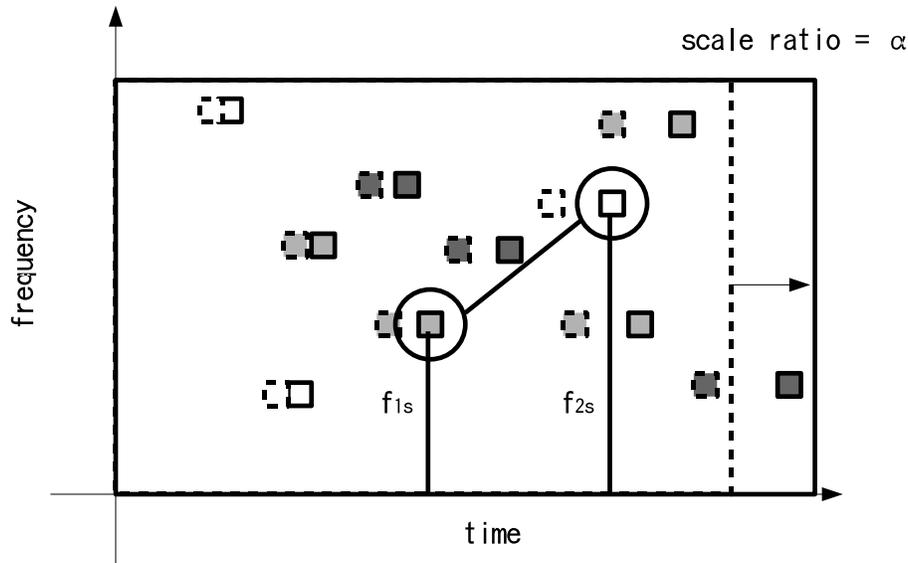


図 2.6: 目的信号から抽出された縮退特徴

2.3.2 変動の性質と解決のアプローチ

本研究で取り扱う変動は、2.3.1 節で述べたように、加法性雑音、乗法性ひずみ、非定常雑音・遮蔽・欠落、幾何変換である。本節では、これら個々の変動の性質とそれぞれに関する対処法について述べる。

加法性雑音と乗法性ひずみによる変動の性質と対処法

ここでは、雑音やひずみの性質が定常な場合のように、統計的な性質が一定で雑音やひずみの性質を予測できるような場合を考える。人のざわめき声や車のエンジン音などによる加法性雑音やマイクの位置の変化による周波数特性の時間変動は、数秒程度の短時間では変動が十分小さいと考えられる。そこで、数秒程度の短時間では、加法性雑音の統計的性質、周波数特性が一定であるとみなして、あるひずみ要因によって変動が起こった信号の時間周波数分布 $Y(t, i)$ が次式のようなモデルで表されると仮定する。

$$Y(t, i) = a(i)X(t, i) + b(i) \quad (2.4)$$

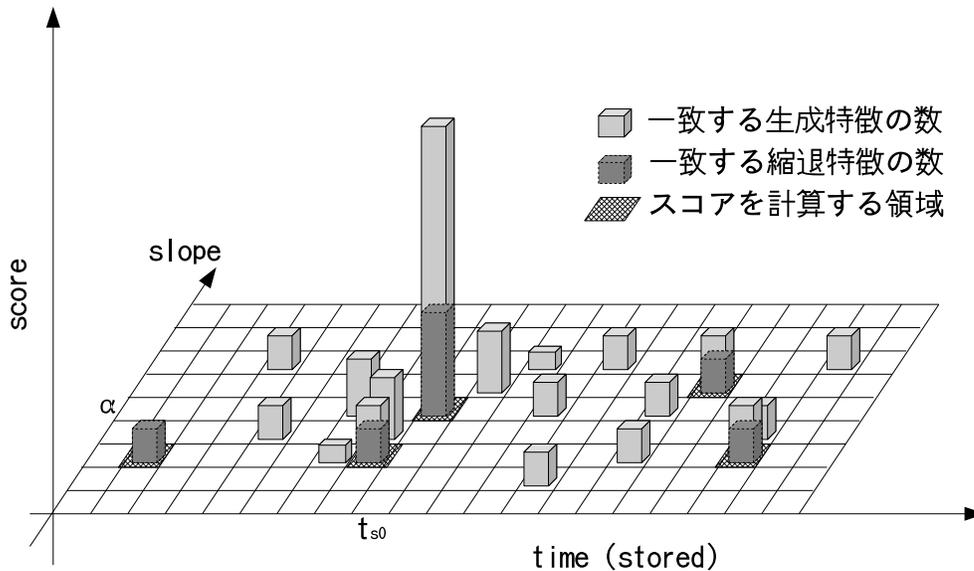


図 2.7: 提案法の投票空間上での探索領域

ただし，時刻 t の原信号の周波数 i のパワーを $X(t, i)$ とし， $a(i)$ は周波数特性の差異によって起こる乗法性ひずみを表す定数， $b(i)$ は加法性雑音を表す定数とする．ここで， $a(i)$ や $b(i)$ をキャンセルするような特徴を選ぶことで，加法性雑音や周波数特性の違いを吸収することができると思われる．

従来，これら加法性雑音と乗法性ひずみへ対処する手法が既に提案されている．加法性雑音へ対処する手法として，スペクトルサブトラクション法 [31] が既に提案されている．この手法は，雑音が定常であると仮定し，得られたスペクトルから，一定区間のスペクトルの平均を差し引くことで，雑音に相当するスペクトルを除去するというものであり，これは，2.4 式の $b(i)$ をキャンセルすることに相当する．しかしながら，この手法は，利用するマイクが固定であったり，周波数特性が既知のマイクを利用するなど，利用環境に制限があるという問題がある．また，乗法性ひずみへ対処する手法として，ケプストラム平均正規化法 [32] がある．この手法は，音声認識のために用いる手法であり，一定区間のケプストラムの平均値をケプストラムから差し引くことで，周波数特性の違いの影響を受けにくい特徴を得ることができる．これは，2.4 式の $a(i)$ をキャンセルすることに相当し，未知の特性のマイクでの音声認識で有効である．これらスペクトルサブトラクション

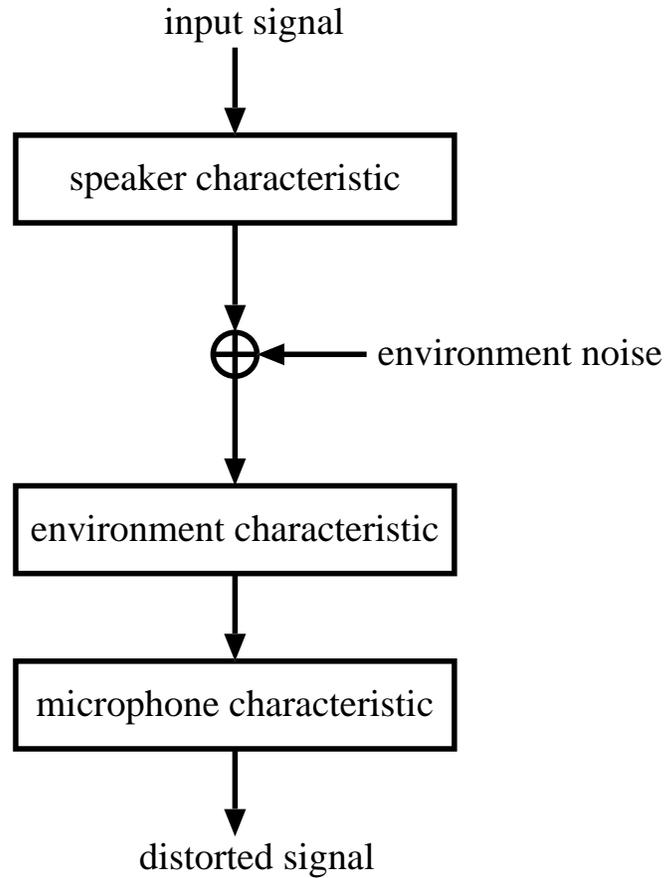


図 2.8: 実環境で収録した音響信号の変動の要因

法とケプストラム平均正規化法は，加法性雑音と乗法性ひずみのいずれか一方をキャンセルするものであり，原理的に同時にキャンセルできるものではなかった．

本研究で導入する局所正規化法は，2.4 式の $a(i)$ と $b(i)$ の両方をキャンセルすることに着目した手法である．これにより，加法性雑音と乗法性ひずみの両方の影響を軽減することが期待できる．

非定常雑音・遮蔽・欠落による変動の性質と対処法

非定常雑音や遮蔽・欠落など，予測が困難な信号の劣化は，2.3.2 節のように予測に基づいて劣化をキャンセルすることができないため，失った信号を復元することは困難である．

非定常雑音や遮蔽・欠落といった信号の劣化は，類似度や距離に大きく影響を

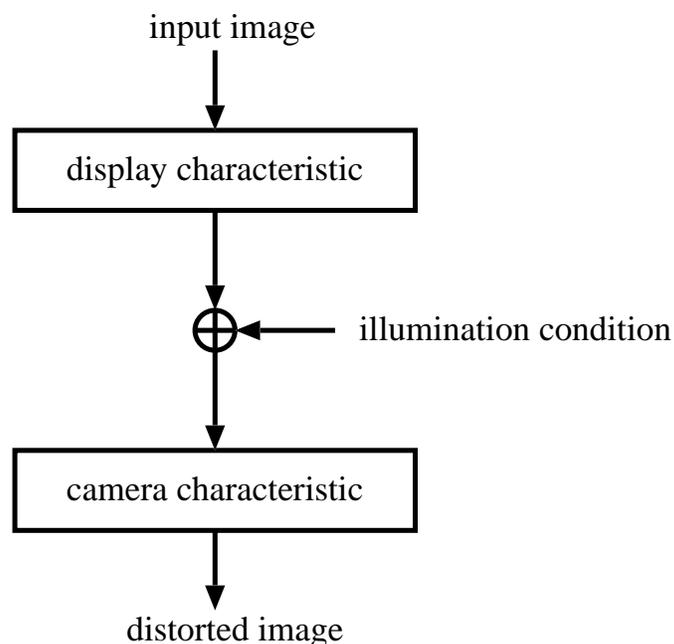


図 2.9: 実環境で収録した映像信号の変動の要因

与え、探索の精度を低下させる要因である。例えば、図 2.10 のようにオリジナルの一次元信号（上）と予測不可能な雑音により欠損した信号が 2 つ（中、下）がある場合を考える。2 つの欠損した信号のそれぞれは、異なる欠損の仕方をしており、L1 距離や L2 距離などのように、差分の割合によって値が変化する距離尺度を用いた場合、欠損した後に埋め込まれた値の内容によって距離が変わってしまう。それに対し、符号の誤り割合を測る尺度であるハミング距離 [33] を用いた場合、誤った区間の長さに相当する量が距離となり、欠損により埋め込まれた信号の影響を受けない。

ハミング距離は通信路の誤り率を測る尺度であるが、従来パターン認識で多く利用されてきた L1 距離や L2 距離のように差分に基づく手法よりも、非定常雑音や遮蔽・欠落によって失った影響を受けにくい距離尺度として有効であると考えられる。

表 2.2: 変動の性質と要因

メディア	変動方向	変動の性質	変動要因
音	パワー	乗法性ひずみ	スピーカ特性の違い
	パワー	乗法性ひずみ	マイク特性の違い
	パワー	乗法性ひずみ	環境吸収特性の違い
	パワー	加法性雑音	環境雑音
	パワー	信号の欠落	劣悪な電波状況
	パワー	非定常雑音	他の音源の重畳
	周波数	幾何変換	ピッチ変動
	時間周波数	幾何変換	時間伸縮
映像	輝度・色	乗法性ひずみ	ディスプレイ特性の違い
	輝度・色	乗法性ひずみ	カメラ特性の違い
	輝度	加法性雑音	照明変動
	輝度	加法性雑音	ディスプレイ表面の反射
	輝度	信号の欠落	テロップ・遮蔽
	時間	幾何変換	スローモーション加工
	空間座標	幾何変換	ピクチャインピクチャ・スキュー
	空間座標	幾何変換	カメラパラメータの違い

幾何変換による変動への対処法

本研究では、信号の伸縮などの変化で発生する幾何変換に対して、頑健に照合するために、前述の縮退特徴および生成特徴を導入する。これらの特徴を利用する基本的なアイデアは、次のとおりである。まず、蓄積信号の特徴の中から2つの特徴を選び出し、それらの相対関係を縮退特徴とし蓄積しておく(図 2.11)。ここで利用する相対関係は、特徴の量子化値や特徴間の周波数の比など、幾何学的な変動があっても不変な量である。この縮退特徴は事前に蓄積信号の全域にわたって抽出しておく。目的信号が入力されると、同様に縮退特徴を計算する(図 2.12)。そして、蓄積信号の中から目的信号と同一の縮退特徴を選び出す。次に、蓄積信

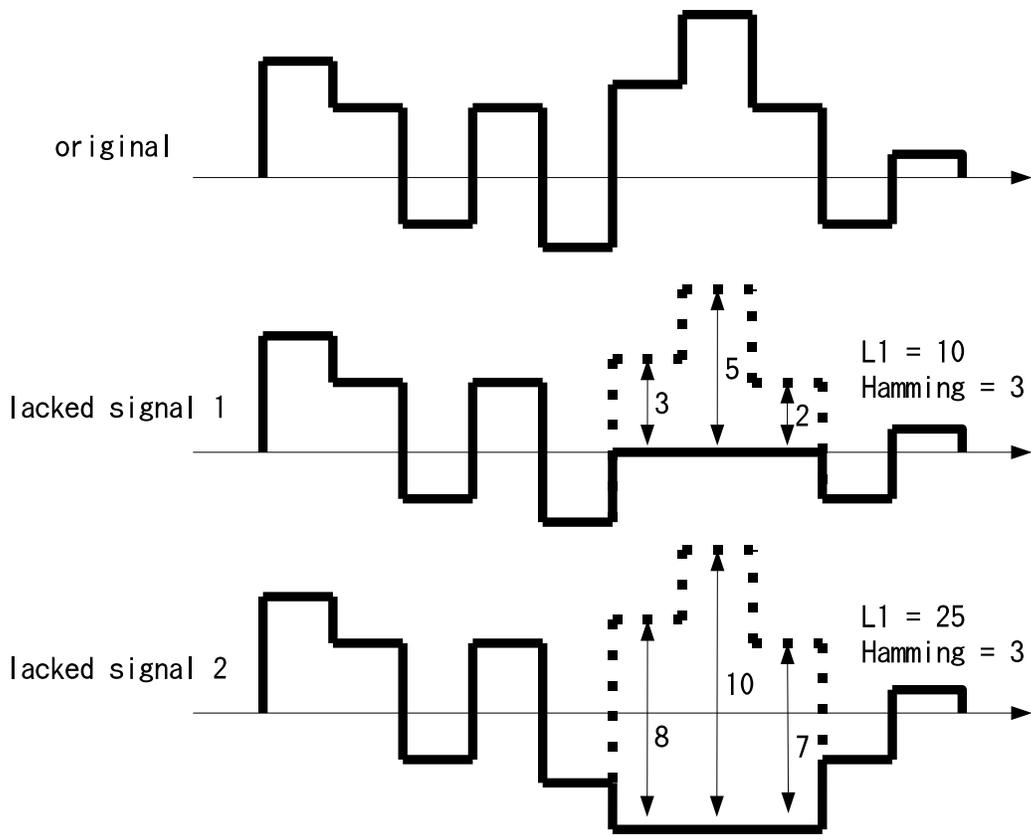


図 2.10: オリジナル信号と欠損した信号

号と目的信号との間の幾何学的変動パラメータを求める．ここで求める変動パラメータは，周波数の比，時間伸縮の比など，目的信号と蓄積信号との間の相対関係を表すパラメータである．次に，蓄積信号を変動パラメータに基づいて目的信号に合わせて幾何変換する(図 2.13，図 2.14)．そして，幾何変換された蓄積信号と目的信号を照合して類似度を得る．

以上のように，本研究では，音や映像から特徴を抽出し探索を行う時，種々の収録条件による変動に対処するために，縮退と生成の両方のアプローチにより問題解決することを検討する．そして，これらのアプローチを融合する方法について提案し，実環境のような多様な変動の起こる劣悪な環境においても高速かつ高精度な探索が可能なメディア探索技術の実現を目指す．

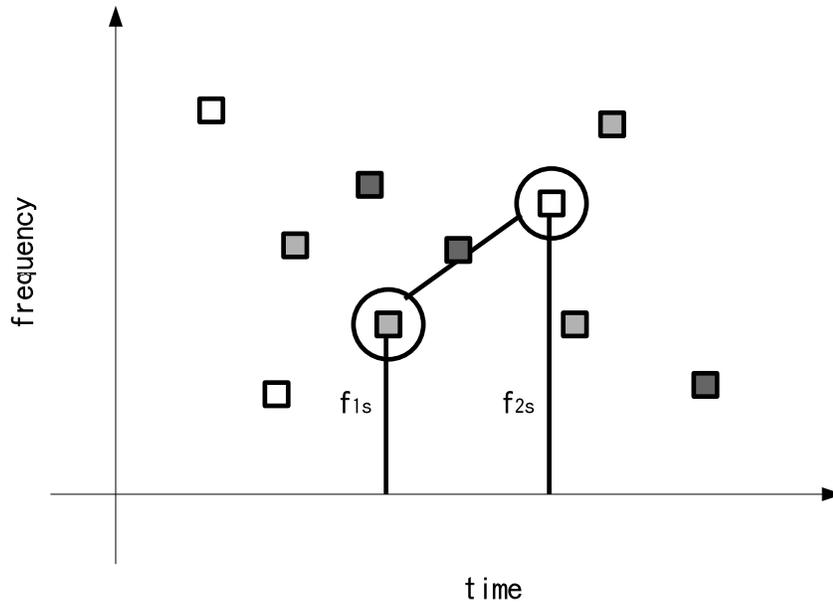


図 2.11: 蓄積信号から抽出された縮退特徴

2.4 本研究の位置付け

ここで、従来手法と提案手法との関係を整理しておく。表 2.3 は、変動の性質とその対処法を従来手法と提案法を区別して示したものである。従来法である、スペクトルサブトラクション、ケプストラム平均正規化、ハミング距離、ハフ変換のそれぞれの手法は、加法性雑音、乗法性ひずみ、非定常雑音・遮蔽・欠落、幾何変換のうちのいずれかに対処する手法であり、全てを同時に満たすものではない。それに対し、提案法の縮退生成探索法は全てを同時に解決するものである。本研究で導入する局所正規化は、スペクトルサブトラクションの加法性雑音の吸収とケプストラム平均正規化の乗法性ひずみの吸収の両方の利点を持つ。特徴選択と符号一致度を用いる探索は、特徴選択された箇所のみハミング距離を計算するものと考えることができる。

提案手法である縮退生成探索法は、縮退特徴、生成特徴、特徴選択の 3 つの考え方に基づく。縮退特徴はノイズやひずみなどの変動を吸収して得られる特徴であり、生成特徴はノイズやひずみなどの変動を生成して得られる特徴である。特徴選択は、頑健な探索を行うために統計量に基づいて信号中の特徴的な箇所を選

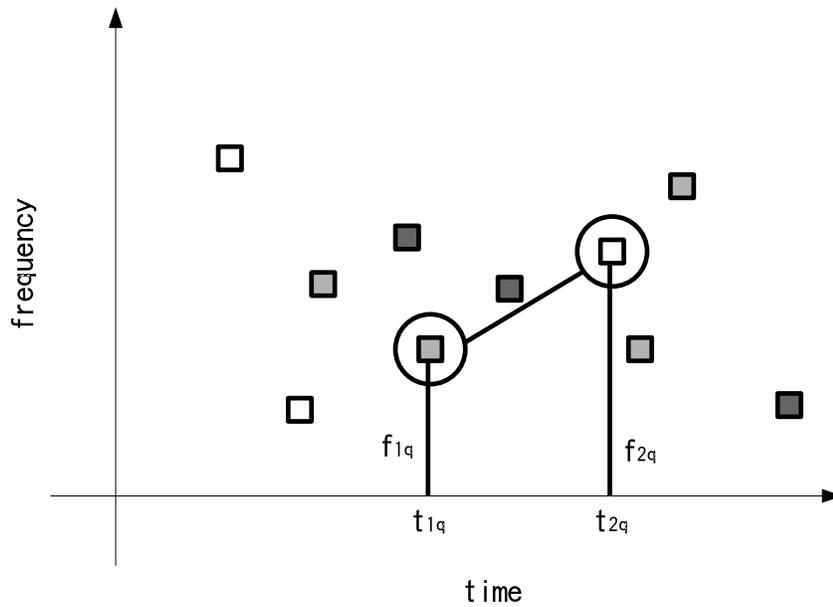


図 2.12: 目的信号から抽出された縮退特徴

び出すものである．縮退生成探索は，選択された特徴に基づいて縮退特徴を導きだし．目的信号と同一の縮退特徴を持つ蓄積信号中の箇所について，幾何変換パラメータを求め，得られた幾何変換パラメータに基づいて特徴を生成し照合を行うものである．本論文では，縮退特徴，生成特徴，特徴選択のそれぞれのアプローチについての議論し，最後に，縮退特徴，生成特徴，特徴選択を組み合わせる縮退生成探索法について議論する．

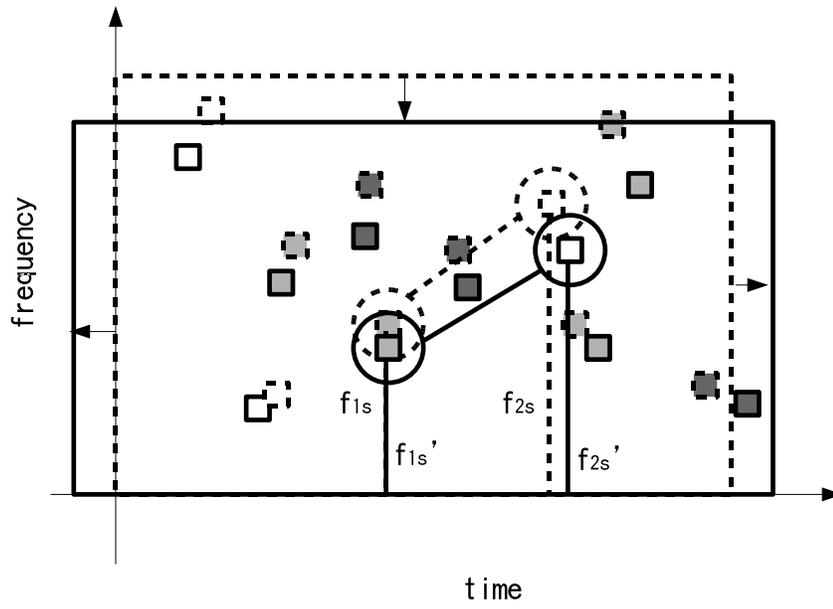


図 2.13: 蓄積信号をスケーリング

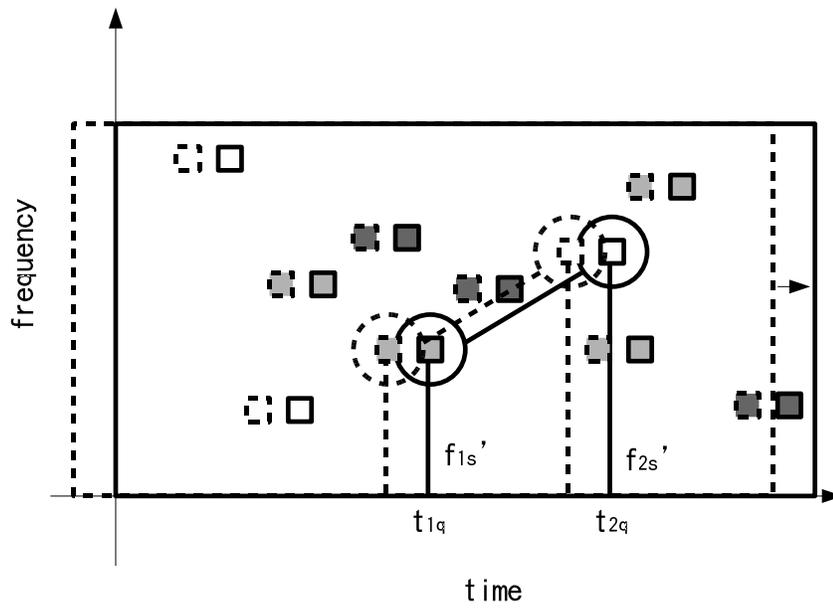


図 2.14: 蓄積信号をシフト

表 2.3: 変動の性質とその対処法

従来/提案	手法	変動の性質			
		加法性雑音	乗法性ひずみ	非定常雑音・遮蔽・欠落	幾何変換
従来法	スペクトルサブトラクション [31]		×	×	×
	ケプストラム平均正規化 [32]	×		×	×
	ハミング距離 [33]	×	×		×
	ハフ変換 [24, 25, 26]	×	×	×	
提案法	局所正規化(縮退特徴)			×	×
	+ 特徴選択と符号一致度 + 縮退生成探索				×

:対応

×:未対応

第 3 章

縮退特徴に基づく探索

3.1 はじめに

本章では，実環境で収録された音やディスプレイに表示されている映像をカメラで収録した映像をキーとしたメディア探索を想定し，縮退特徴に基づく探索のアプローチを用いた手法について述べる [34, 15, 14, 8]．ここでは，乗法性ひずみと加法性雑音による変動を縮退させることで，変動の吸収を試みる．基本的なアイデアは，機器特性や環境特性などによって発生する乗法性ひずみや環境雑音のような加法性ひずみを吸収するために，時間周波数空間または時間空間上の局所領域ごとに正規化を行った後，機器特性や環境雑音などによるひずみに対して頑健な部分空間に射影して照合を行うというものである．以下，提案法の概要と，その妥当性を検討するための実験について順に説明する．

3.2 問題と解決の方針

信号中のひずみを吸収する手法は，これまで，実環境下での音声認識の研究分野においてさまざまなものが提案されている [32, 35, 36, 37]．例えば，加法性雑音を除去するため雑音スペクトルを信号スペクトルから差し引くスペクトルサブトラクション (SS) [31]，乗法性ひずみを吸収するためのケプストラム平均正規化 (CMN) [32] など，簡便で有用性が高い方法が提案されている．しかしながら，これらの研究は，雑音環境下での音声認識に関するものであり，本論文で扱うよ

うな実環境で流れている楽曲の断片をキーとした時系列探索を対象とするものではなかった．そこで本章では，時系列探索に適した汎用的な特徴抽出法という立場から，特徴ひずみに頑健な方法を検討することにする．

さて，実環境で収録された音に含まれる特徴ひずみには，様々なものが考えられる．例えば，スピーカなどの信号発生源の機器の特性によるひずみ，実環境中の雑音，反響，吸収などによるひずみ，入力端末の特性によるひずみなどである．もし，これらのひずみを全てモデル化し，モデルに基づいてひずみを補正できるならば，探索精度を向上させることができるであろう．しかしながら，現実的には，入力端末側の特性は既知とできる場合もあるが，スピーカ特性や環境雑音等を事前に知ることは難しい．そこで，これらのひずみを吸収し，ひずみの変動の少ない特徴を抽出することが必要である．

本章で扱う特徴ひずみを具体的に例示したものが図 3.1 と図 3.2 である．図 3.1 は，ある音楽の同一の部分について複数の条件で収録し，それぞれの信号からフーリエ変換（サンプリング周波数 8000 Hz，フーリエ変換に使用したサンプリング点数 4096）により求めたパワースペクトルを 0 ~ 2000 Hz の平均パワーで割って，さらに 50 Hz ごとにサンプリングして求めた周波数分布を示したものである．収録の条件は，原音（CD 品質），実験室で高音質携帯電話（PHS¹）による収録，実験室で中音質携帯電話（PDC²など）による収録，街頭でのマイク収録，および喫茶店でのマイク収録の 5 種類である．これらの信号間には大きな差異があるが，探索においては，これらを同一のものとみなして認識しなければならない．一方，図 3.2 は，図 3.1 に示したものと別の音楽の一部分の周波数分布である．これら図 3.1 と図 3.2 は別のものとして認識されなければならない．このような信号の探索においては，単純に周波数分布をマッチングしたのでは探索精度が非常に低い．そのため，特徴ひずみに対しては変動が小さくかつ，音響信号の内容に対しては変動を大きくするような特徴抽出法が必要であると考えられる．

本章の基本的なスタンスは，音の雑音や周波数特性の違いに対して，できるだけ不変な特徴を選ぶというものである．人のざわめき声や車のエンジン音など加法性雑音やマイクの位置の変化による周波数特性の時間変動は，数秒程度の短時

¹32kbps ADPCM 方式を用いた簡易型携帯電話

²6.7kbps VSELP 方式を用いた携帯電話

間では変動が十分小さいと考えられる．そこで，数秒程度の短時間では，加法的雑音，周波数特性が一定であるとみなし，あるひずみ要因によって変動が起こった信号の時間周波数分布 $Y(t, i)$ が次式のように表されると仮定する．

$$Y(t, i) = a(i)X(t, i) + b(i) \quad (3.1)$$

ただし，時刻 t の原信号の周波数 i のパワーを $X(t, i)$ とし， $a(i)$ は周波数特性の差異によって起こる乗法的ひずみを表す定数， $b(i)$ は加法的雑音を表す定数とする．ここで， $a(i)$ や $b(i)$ をキャンセルするような特徴を選ぶことで，加法的雑音や周波数特性の違いを吸収することができると考えられる．このために，数秒程度の短時間の周波数特徴を求め，その時間周波数空間上での局所領域の統計量を用いて正規化することを考える．例えば，各周波数帯域ごと独立に数秒程度の周辺領域の平均を差し引き標準偏差で割ることで， $a(i)$ や $b(i)$ に対して不変な特徴が得られる．また，映像の場合も同様に，収録されたディスプレイの映像の明るさやコントラスト比の違いに対して，できるだけ頑健な特徴を選ぶことができる．ディスプレイやカメラの特性の変動や照明の変化は，数秒程度の短時間では変動が十分小さいと考えられる．そこで，数秒程度の短時間では，明るさやコントラスト比は一定であるとみなし，あるひずみ要因によって変動が起こった信号の画素値を $Y(i, t)$ と仮定する．ただし，時刻 t の原信号の画素 i の画素値を $X(i, t)$ とし， $a(i)$ ， $b(i)$ は定数とする．ここで， $a(i)$ や $b(i)$ をキャンセルするような特徴を選ぶことで，明るさやコントラストの特性の違いを吸収することができると考えられる．

また，図 3.1，図 3.2 を見ると，周波数帯域によっては，特に変動の大きい成分，少ない成分などまちまちである．そこで，探索においても特定の帯域の分解能を高めることが探索精度向上につながると考えられる．そこで，周波数特徴を特徴ひずみに対しては変動が小さくかつ音響信号の内容に対しては変動が大きい部分空間に射影することで，より特徴ひずみに頑健な特徴抽出を行うことが考えられる．

3.3 探索の原理

本手法の基本的なアイデアは、特徴ひずみによる変動を吸収するため、2つの変動の吸収処理を行うことにある。すなわち、音の周波数特徴または映像特徴を抽出した後、時間周波数空間上または時間空間上での局所領域ごとに正規化し、特徴ひずみに頑健な部分空間への射影を行う。探索の具体的な処理は、(1) 周波数特徴または映像特徴の抽出、(2) 時間周波数空間上または時間空間上の局所領域ごとの正規化、(3) 部分空間への射影、(4) 時系列探索の順で行う(図 3.3, 図 3.4)。また、部分空間への射影を行うためには、探索に先だてて事前に特徴ひずみに頑健な部分空間を求めておく必要がある。

3.3.1 周波数特徴の抽出

音響特徴としては、単位時間あたりのゼロ交差数、短時間パワースペクトル、LPC ケプストラム、MFCC(Mel frequency cepstral coefficients) など [38] が考えられる。本章では、その中でも代表的な周波数特徴である、フーリエ変換により求めた短時間パワースペクトルを用いた。ここで、時刻 t の音響信号のフーリエ変換後の特徴 $X(t)$ の k 番目の要素を、

$$X(t, k) = \sum_{n=0}^{N-1} x(t+n) e^{-j \frac{2\pi kn}{N}} \quad (3.2)$$

とする。ただし、 $x(t)$ を時刻 t における信号の値、 N は周波数特徴の分析窓の長さ、 k の最大値は標本化定理より $N/2$ である。ここで、パワースペクトル $P(t, k)$ を次式により求める。

$$P(t, k) = |X(t, k)|^2 \quad (3.3)$$

ここで、このパワースペクトルを s きざみで抽出する。すなわち、時間方向 i 番目の周波数特徴 $Q(i, k)$ は、

$$Q(i, k) = P(s_i, k) \quad (3.4)$$

である。

3.3.2 映像特徴の抽出

映像においても特徴を抽出することができる．本研究では，携帯カメラで収録した映像中のディスプレイ表示部分を切り出し，ディスプレイ部分のカラー縮小画像を映像特徴として用いた．ここで，映像特徴ベクトル $x(k)$ を，

$$x(k) = (x_{1r}(k), x_{1g}(k), x_{1b}(k), \dots, x_{jc}(k), \dots, x_{Wr}(k), x_{Wg}(k), x_{Wb}(k)) \quad (3.5)$$

と定義する．ここで k はフレームの時刻であり， x の添字 j は各フレームのディスプレイの表示部分を W 個のサブ画像に分割した分割番号，添字 c は RGB を表す．ただし， W は経験的に与えられる値である． 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た値であり，

$$x_{jc}(k) = \frac{1}{|I|} \sum_{p \in I} v_{pc}(k) \quad (3.6)$$

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， $v_{pc}(k)$ は画素 p の RGB のうちの一色 c の画素値を表す．なお，本稿では，ディスプレイ表示部分の切り出しは手動で行ったが，ある時間区間で画素値の変化の大きい領域を抽出し，アフィン変換により自動的に切り出すことも可能である．

3.3.3 時間周波数空間上または時間空間上での局所領域ごとの正規化

この処理の目的は，数秒程度の短時間で変動の小さい加法性雑音や周波数特性を吸収することである．ここでは，各周波数帯域ごとに，ある時間区間の値から平均と標準偏差を求め，それらを用いて正規化を行う．すなわち，正規化後の周波数特徴 $y(i)$ の k 番目の要素は，

$$y(i, k) = \frac{1}{\sigma(i, k)} (Q(i, k) - m(i, k)) \quad (3.7)$$

である．ただし，

$$m(i, k) = \frac{1}{2M} \sum_{j=-M}^{M-1} Q(i+j, k) \quad (3.8)$$

$$\sigma(i, k)^2 = \frac{1}{2M} \sum_{j=-M}^{M-1} (Q(i+j, k) - m(i, k))^2, \quad (3.9)$$

M は局所時間内の周波数特徴の平均，標準偏差を求めるための時間窓の大きさの半分の値である．

なお，本正規化は，実環境中の音声認識でよく使用される CMN (Cepstrum mean normalization) に類似した手法である．CMN は，周波数特性を吸収するため，短時間のケプストラムの平均を差し引くという処理を行う手法である．この手法は，ケプストラムの平均を計算する時間窓の中で周波数特性が一定であるとする，平均を差し引くことにより周波数特性の差による変動を吸収できるという考え方に基づくものである．それに対し，本正規化は，加法性雑音を吸収するために短時間の周波数特徴の平均 $m(i, k)$ を差し引く，周波数特性を吸収するために短時間の周波数特徴の標準偏差 $\sigma(i, k)$ で割る，という処理を行うものである．またこれは，時間区間内の値をサンプルとみたてたガウシアン正規化と考えることもできる．周波数特徴の平均，標準偏差を計算する時間窓の中で加法性雑音と周波数特性が一定であるとする，この処理により加法性雑音と周波数特性の両方を吸収することができると考えられる．

なお，映像においても，数秒程度の短時間で変動の小さいディスプレイの明るさやコントラスト比の違いを吸収することを目的として，同様の処理を適用することができる．すなわち，各画像領域ごとに，ある時間区間の値から平均と標準偏差を求め，それらを用いて正規化を行う．

3.3.4 部分空間の構成

続いて，正規化後の周波数特徴または映像特徴を部分空間に射影する．この処理の目的は，特徴ひずみに対しては変動が小さくかつ信号の内容に対しては変動が大きい部分空間に射影することで，より特徴ひずみに頑健な特徴抽出を行うことである．

ここで用いる手法は，よく知られた手法である主成分分析 (PCA) と類似した手法であるが，雑音の平均化処理を行うという点で異なる．すなわち，特徴ひずみに対する変動は無視し，信号の内容に対する変動のみを考慮した PCA を行う．

具体的には，CDの音楽などの音響信号中のある周波数特徴を L 個用意し， l 番目の信号部分の正規化後の周波数特徴を y_{l0} とする．また，周波数特徴 y_{l0} それぞれに対して特徴ひずみのある信号を C 種類用意し， c 番目の種類の周波数特徴を y_{lc} とする．すなわち，クラス数 L 個，各クラス C 個の学習サンプルを用意する．

ここで，特徴の平均を用いて共分散行列 R を計算する．

$$R = \frac{1}{L} \sum_{l=1}^L (\bar{y}_l - \bar{y})(\bar{y}_l - \bar{y})^t \quad (3.10)$$

ただし，各クラス毎の平均 \bar{y}_l を

$$\bar{y}_l = \frac{1}{C+1} \sum_{c=0}^C y_{lc}, \quad (3.11)$$

クラス平均の平均 \bar{y} を

$$\bar{y} = \frac{1}{L} \sum_{l=1}^L \bar{y}_l \quad (3.12)$$

とする．ここで， R の固有ベクトルを求める．固有ベクトルは，次式の固有値問題により求める．

$$R\phi_u = \lambda_u \phi_u \quad (3.13)$$

ただし， ϕ_u は互いに直交する固有ベクトルのうちの固有値が u 番目に大きいものである．

ここで得られた，固有ベクトル ϕ_u から成る部分空間に射影した特徴，すなわち，

$$z_u = y\phi_u \quad (3.14)$$

を要素に持つ特徴ベクトル z を用いて探索を行う．

なお，本手法は，従来の PCA に比べて，同じ学習サンプル数でも平均しか使用しないため，比較的少ない計算量で部分空間を求めることができるというメリットもある．本章では，提案の部分空間以外に，従来の PCA，線形判別分析 (LDA) についても比較実験した．

3.3.5 時系列探索

最後に，求められた特徴ベクトルを用いて時系列探索を行う．時系列探索は，蓄積信号中に目的信号と同じ大きさの窓をかけ，目的信号と蓄積信号中の注目箇所

との類似度を計算しながら，窓をずらし，目的信号と最も類似した蓄積信号中の箇所を探索結果として出力する．目的信号の長さが周波数特徴の分析窓の長さよりも大きい場合は，目的信号から周波数特徴のある時間間隔で抽出し，各々の特徴ベクトルの要素から成る特徴ベクトルを探索用の特徴ベクトルとして用いる．また，本章では，類似度としてユークリッド距離を用いることにする．すなわち，

$$D = \sqrt{(Z_q - Z_s)^t(Z_q - Z_s)} \quad (3.15)$$

で定義される距離を用いる．ただし， Z_q は目的信号の特徴ベクトル， Z_s は蓄積信号の照合箇所の特徴ベクトルである．本章では，この距離尺度に基づき探索を行った．

3.4 実験

本手法の有効性について検討するため，さまざまな音質の音響信号，さまざまな品質の映像信号を目的信号として，蓄積信号から同一部分を探索し，探索精度を評価する実験を行った（図 3.5，図 3.7）．ここでは，時間周波数空間上また時空間上での局所領域ごとの正規化による効果を確認する予備実験（実験 1），部分空間による効果を確認する予備実験（実験 2），騒音下での精度評価を行う実験（実験 3）を行った．

3.4.1 音響信号探索用実験データの収集

実験に先だって，楽曲，収録場所，携帯端末を複数選び，実環境下にて楽曲の収録を行った．これらの収録条件を表 3.9 に示す．収録した音響信号は，表 3.9 の楽曲，場所，機器全ての組み合わせ，すなわち合計 1428 個の音響信号（各 42 秒）である．蓄積信号として用いた楽曲は実環境で収録したものの原信号を含む CD 品質の音響信号 116 時間分を用いた（表 3.10）．

表 3.1: 実環境で収録した音響信号

	内容
楽曲	34 曲 (さまざまなジャンル) ・ 学習用データ 17 曲 ・ 評価用データ 17 曲
収録場所	騒音レベル小 4 箇所 ・ 静かな部屋 ・ アイドリング状態の車 ・ コンビニエンスストア ・ カラオケ店 騒音レベル大 3 箇所 ・ 人通りの多い街頭 ・ 騒がしい喫茶店 ・ 交通量の多い交差点付近
携帯端末	・ マイク ・ 高音質携帯電話 (PHS , 通信会社 2 社) ・ 中音質携帯電話 (PDC など , 通信会社 3 社)

3.4.2 時間周波数空間上の局所領域ごとの正規化による効果 (実験 1)

まず，時間周波数空間上の局所領域ごとの正規化による効果を調べるため探索実験を行った．実験は，目的信号の長さを 10 秒とし，目的信号の切り出し部分を 200 回変えて探索精度の測定を行った (音響信号は 8000 Hz で標本化，特徴抽出のパラメータは $s = 512$, $N = 8192$, $M = 16$) . また，目的信号は，表 3.3 の条件で収録した音響信号を合計 12 種類用意し，合計 2400 回の探索を行った．探索時に使用した特徴ベクトルは，160 次元の特徴ベクトル (正規化後の周波数特徴の

表 3.2: 蓄積信号

楽曲	116 時間 (3310 曲)
----	-------------------

0 ~ 4000 Hz を 16 分割し , 各周波数帯域毎に成分を平均化し , 1 秒毎に 10 個の周波数特徴からなる特徴ベクトル) を用いた . 本実験における探索例を図 3.7 に示す . 次に , 実験結果を表 3.11 に示す . ここでは , 探索した時刻が正しい時刻の ± 1 秒以内となった場合に正解とした . なお , 5 位累積分類率とは , 5 位候補までに正しい時刻が入った率とした . また , 比較のため , パワーによる正規化 (PN) , SS , CMN についても同様の実験を行った . 比較に用いたそれぞれの正規化処理は , PN については , 各成分を全体のパワーで割る処理 , すなわち , 特徴ベクトルを

$$y(i, k) = \frac{1}{\sum_{n=0}^K Q(i, n)} Q(i, k), \quad (3.16)$$

とした . ただし , K は FFT の分析における最大周波数とする . SS については , 得られた周波数特徴を 1 秒間について平均し中央の周波数特徴から差し引く処理 , すなわち ,

$$y(i, k) = Q(i, k) - m(i, k), \quad (3.17)$$

とした . CMN については , まず周波数特徴の対数を逆 FFT することによりケプレンシ k' のケプストラム $Q'(i, k')$ を求め , $Q'(i, k')$ を 1 秒間について k' 毎に平均し中央のケプストラムから差し引く . すなわち ,

$$y'(i, k') = Q'(i, k') - m'(i, k'), \quad (3.18)$$

を求める . ただし ,

$$m'(i, k') = \frac{1}{2M} \sum_{i=-M}^{M-1} Q'(i, k'), \quad (3.19)$$

とする . 最後に , $y'(i, k')$ を FFT し , 最後に指数をとる処理をした . 提案の正規化は他の正規化法に比べて探索精度が数十ポイント向上している . これにより , 周波数特徴の時間周波数空間上の局所領域ごとの正規化が有効であることが確認された .

表 3.3: 実験 1, 2 (予備実験) で使用した音響信号

	内容
蓄積信号	12 分 (CD より 17 曲)
目的信号	2400 個の 10 秒の断片 (表 3.9 評価用データより)
目的信号の 収録条件	収録場所： ・騒音レベル小 4 箇所 収録端末： ・マイク, PHS 2 機種

表 3.4: 時間周波数空間上の局所領域ごとの正規化による効果 (実験 1)

	正解率	5 位累積分類率
PN	12.8 %	16.5 %
SS	2.9 %	3.5 %
CMN	20.7 %	25.6 %
提案の正規化	61.5 %	65.9 %

3.4.3 部分空間による効果 (実験 2)

次に, 部分空間に射影することの有効性を調べるため探索実験を行った. ここでは, 表 3.9 の学習用データの中から目的信号と同じ収録場所, 携帯端末で収録した合計 12 個の音響信号からそれぞれ 400 箇所無作為に切り出したものを学習サンプルとして部分空間の学習を行った. 学習により得られた部分空間を用いて, 学習用データとは異なる評価用データ (表 3.9) を用いて探索実験を行った. 実験結果を表 3.12 に示す. なお, ここでは, 比較のため, 3.4.2 節同様に周波数帯域毎に平均したもの, 全学習サンプルに PCA を行ったもの, LDA を行ったものについても同様の実験を行った. 探索時に使用した特徴ベクトルは, 160 次元の特徴ベクトル (使用する固有ベクトルの数を 16 とし, 求めた特徴ベクトルを 10 秒の

表 3.5: 部分空間による効果 (実験 2)

	正解率	5 位累積分類率
提案正規化 + PCA	64.0%	68.9%
提案正規化 + LDA	64.8%	69.4%
提案正規化 + 提案部分空間	71.2%	74.8%

目的信号から 1 秒毎に抽出)を用いた。なお、探索精度は 3.4.2 節と同様の方法で測定した。また、使用した特徴ベクトルの次元数は、3.4.2 節では周波数帯域を 16 分割することにより次元削減されていたのに対し、本節の実験で使用した特徴ベクトルの次元数は、上位 16 の固有値を持つ固有ベクトルのみを使用することで次元削減されている。提案の部分空間は他の手法に比べて探索精度が数ポイント向上している。よって、クラス平均のみを使用した PCA により探索精度が向上することが確認された。

3.4.4 騒音下での精度評価 (実験 3)

次に、さらに騒音の大きい環境や音質の異なる携帯電話を用いて探索精度を評価するため、より多種類のひずみのある目的信号(表 3.6)を用いた実験を行った。表 3.7 に探索精度を示す。なお、括弧内には時刻を特定する探索の精度ではなく、蓄積信号の 3310 曲のうちどの楽曲であるかを特定する探索の精度も示した。また、探索時に使用した特徴ベクトルは、160 次元の特徴ベクトル(使用する固有ベクトルの数を 16 とし、求めた特徴ベクトルを 10 秒の目的信号から 1 秒毎に抽出)を用いた。騒音レベル小では、マイク、高音質携帯電話で収録したもので探索した場合、71.8% の探索精度が得られている。しかしながら、騒音レベル大では、中音質携帯電話で収録したもので探索した場合、探索精度が 15.2% というように著しく低下しており、今後さらに検討が必要である。

最後に参考まで、処理時間の測定を行った。ここでは CPU は Intel 社 Pentium III (1GHz)、ハードディスクは ULTRA160 SCSI 対応を用いた。表 3.8 に結果をまとめる。なお、ここに示す時間は、メモリに読み込んだ後から計測した時間を

表 3.6: 実験 3 で使用した音響信号

	内容
目的信号	それぞれ 2400 個の 10 秒の断片 (表 3.9 評価用データより)
目的信号の 収録条件	収録場所： <ul style="list-style-type: none"> ・騒音レベル小 4 箇所 ・騒音レベル大 3 箇所 収録端末： <ul style="list-style-type: none"> ・マイク, 高音質携帯電話 2 機種 ・中音質携帯電話 3 機種

表 3.7: 探索精度 (実験 3)

	騒音レベル小	騒音レベル大
マイク, 高音質携帯電話	71.8% (72.0%)	27.6% (27.6%)
中音質携帯電話	39.1% (39.5%)	15.2% (15.4%)

示し, 括弧内にはハードディスクからメモリに読み込む時間を含めた時間を示した. ここでは SSDA [39] を用いて探索の高速化を行っている. 前処理は蓄積信号の長さの 15% 程度の時間で完了し, 探索は 100 時間程度の蓄積信号を 2 秒弱で完了することが確認された.

3.4.5 映像信号探索用実験データの収集

実験に先だって, ディスプレイ, カメラを複数選び, 室内環境下にて CM 映像の収録を行った. これらの収録条件を表 3.9 に示す. 収録した映像信号は, 表 3.9 の全ての組み合わせ, すなわち合計 6 条件の映像信号 (各 20 分) である. 蓄積信号として用いた CM 映像は実環境で収録したもののオリジナルの映像である.

表 3.8: 処理時間

	処理時間
前処理時間	18.2 時間 (18.6 時間)
探索時間	1.82 秒 (183.1 秒)

表 3.9: 実環境で収録した映像信号

	内容
映像	<ul style="list-style-type: none"> ・ 評価用映像 (CM 20 分 , 15 秒 CM を繋合わせた映像) ・ 学習用映像 (テレビ放送 10 分)
ディスプレイ	<ul style="list-style-type: none"> ・ LCD ディスプレイ ・ ブラウン管テレビ 各 1 機種
カメラ	<ul style="list-style-type: none"> ・ DV カメラ 異なるメーカーの 3 機種

3.4.6 時空間上の局所領域ごとの正規化による効果 (実験 1)

まず、時空間上の局所領域ごとの正規化による効果を調べるため探索実験を行った。実験は、目的信号の長さを 6 秒とし、目的信号の切り出し部分を 200 回変えて探索精度の測定を行った。なお、映像信号は 29.97 frames/s のフレームレート、 176×144 の画像サイズでキャプチャし、特徴抽出のパラメータは、 $W = 4$ (横方向 2 分割、縦方向 2 分割)、 $M = 45$ とした。また、目的信号は、表 3.9 の条件で収録した評価用映像信号を合計 6 種類用意し、合計 1200 回の探索を行った。探索時に使用した特徴ベクトルは、36 次元の特徴ベクトル (1 秒毎に 3 個抽出した正規化後の映像特徴からなる特徴ベクトル) を用いた。次に、実験結果を表 3.11 に示す。ここでは、探索した時刻が正しい時刻の ± 1 秒以内となった場合に

表 3.10: 蓄積信号

内容	CM 20 分，15 秒 CM を繋合わせた映像，表 3.9 評価用映像と同じ内容)
----	--

表 3.11: 時空間上の局所領域ごとの正規化による効果 (実験 1)

	正解率	5 位累積分類率
提案正規化なし	10.75%	20.25%
提案正規化あり	59.75%	73.83%

正解とした。なお，5 位累積分類率とは，5 位候補までに正しい時刻が入った率とした。また，比較のため，正規化しなかったものについても同様の実験を行った。提案の正規化は他の正規化法に比べて探索精度が数十ポイント向上している。これにより，時空間上の局所領域ごとの正規化が有効であることが確認された。

3.4.7 部分空間による効果 (実験 2)

次に，部分空間に射影することの有効性を調べるため探索実験を行った。ここでは，表 3.9 の学習用データの中から目的信号と同じディスプレイとカメラの条件で収録した映像信号と原信号の合計 7 個から，それぞれ 200 箇所無作為に切り出したものを学習サンプルとして部分空間の学習を行った。なお，特徴抽出のパラメータは， $W = 192$ (横方向 16 分割，縦方向 12 分割)， $M = 45$ とした。学習により得られた部分空間を用いて，評価用データ (表 3.9) を用いて探索実験を行った。実験結果を表 3.12 に示す。なお，ここでは，全学習サンプルに PCA を行ったもの，LDA を行ったものについても同様の実験を行った。探索時に使用した特徴ベクトルは，36 次元の特徴ベクトル (使用する固有ベクトルの数を 12 とし，求めた特徴ベクトルから 1 秒毎に 3 個使用) を用いた。なお，探索精度は 3.4.6 節と同様の方法で測定した。提案の部分空間は他の手法に比べて探索精度が向上し

表 3.12: 部分空間による効果 (実験 2)

	正解率	5 位累積分類率
PCA	86.25%	90.75%
LDA	79.25%	85.67%
提案部分空間	86.33%	91.50%

ていることが確認された。

3.5 考察

本章で利用した局所正規化は、乗法性ひずみと加法性雑音をスピーカーやマイク、ディスプレイやカメラの特性の違いに対して不変な量を求めるという観点で導き出された正規化であり、短時間で乗法性ひずみと加法性雑音が一定であるという仮定が適合する範囲で適応的に変動を吸収することができ良好な結果が得られた。しかしながら、複数の音源が混入する場合や別の物体による遮蔽がある場合など、変動モデルの仮定に合わないような場合には、その変動を吸収できなかったと考えられる。また、部分空間照合は、探索に有利な周波数や画像の色や領域を選択するような効果があると考えられるが、利用したデータは周波数や色・領域に大きな偏りのあるデータでなかったため、大きな改善は見られなかったと考えられる。縮退特徴は、ある基準を設定しそれに対して相対的な値を利用するものである。変動に不変であるため高速に探索できるが、利用する相対値の分解能が十分でなければ、探索精度が低下してしまう。特に幾何変換のような大きな変動を扱う場合には、あまり大きな変動を縮退させすぎると分解能を低下させてしまうため、特徴の抽出の方法には注意が必要であると考えられる。

3.6 まとめ

本章では、実環境で収録された音やディスプレイに表示されている映像をカメラで収録した映像をキーとしたメディア探索を想定し、精度良く探索する縮退に

基づいた探索手法を提案した。本手法は、機器特性や環境特性などによって発生する乗法性ひずみや環境雑音のような加法性ひずみを吸収するために、時間周波数空間または時間空間上の局所領域ごとに正規化を行った後、機器特性や環境雑音などによるひずみに対して頑健な部分空間に射影して照合を行うというものである。実験を行ったところ、これらの効果により探索精度が向上することが確かめられた。例えば、蓄積した CD 品質の音楽音響信号をマイクや PHS で収録した 10 秒間の探索キー信号によって探索する実験では、騒音レベルが比較的小さい場所での収録の場合、71.2% の精度（収録した音響信号の断片が、CD 品質の蓄積信号のどの曲のどの時刻に対応するか特定する）が得られ、本手法を用いない場合の 12.8% に比べてかなりの精度の改善が可能ながかった。また、ディスプレイに表示されている映像を携帯カメラで収録し、その収録映像をキーとして、膨大な映像データからそれと同一の映像を探索する手法についても検討した。映像探索の場合は、ディスプレイやカメラの機器特性や照明変動などに頑健な特徴を抽出するために時空間上の局所領域ごとに正規化を行った後、特徴ひずみの少ない部分空間に射影して照合を行うというものである。実験を行ったところ、これらの効果により探索精度が向上することが確かめられた。例えば、蓄積した CM 映像をカメラで収録した 6 秒間の探索キー信号によって探索する実験では、91.50% の精度（収録した映像信号の断片が、蓄積信号のどの時刻に対応するか特定する）が得られ、本手法を用いない場合の 10.75% に比べてかなりの精度の改善が可能ながかった。

本章で利用した局所正規化は、乗法性ひずみと加法性雑音をスピーカーやマイク、ディスプレイやカメラの特性の違いに対して不変な量を求めるという観点で導き出された正規化であり、短時間で乗法性ひずみと加法性雑音が一定であるという仮定が適合する範囲で適応的に変動を吸収することができ良好な結果が得られた。しかしながら、複数の音源が混入する場合や別の物体による遮蔽がある場合など、変動モデルの仮定に合わないような場合には、その変動を吸収できなかったと考えられる。また、部分空間照合は、探索に有利な周波数や画像の色や領域を選択するような効果があると考えられるが、利用したデータは周波数や色・領域に大きな偏りのあるデータでなかったため、大きな改善は見られなかったと考えられる。縮退特徴は、ある基準を設定しそれに対して相対的な値を利用するも

のである．変動に不変であるため高速に探索できるが，利用する相対値の分解能が十分でなければ，探索精度が低下してしまう．特に幾何変換のような大きな変動を扱う場合には，あまり大きな変動を縮退させすぎると分解能を低下させてしまうため，特徴の抽出の方法には注意が必要であると考えられる．そこで，分解能を維持したまま幾何変換に対して縮退特徴を利用する必要があるが，これについては，第 6 章で述べることとする．

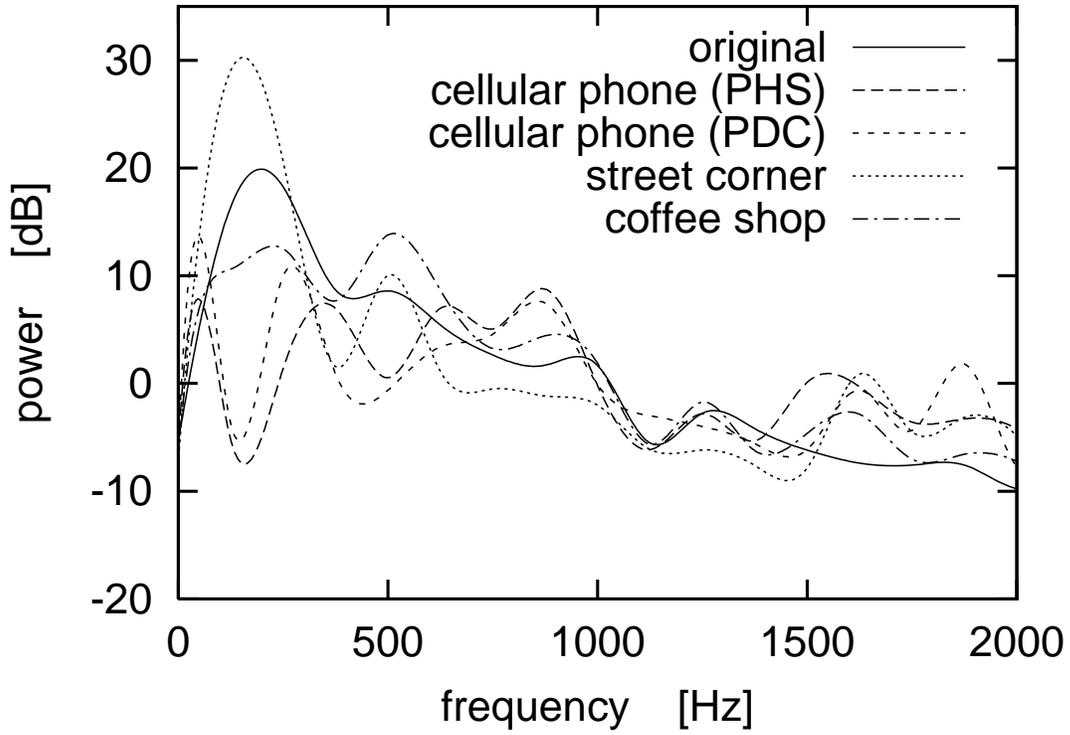


図 3.1: 楽曲の一部分 1

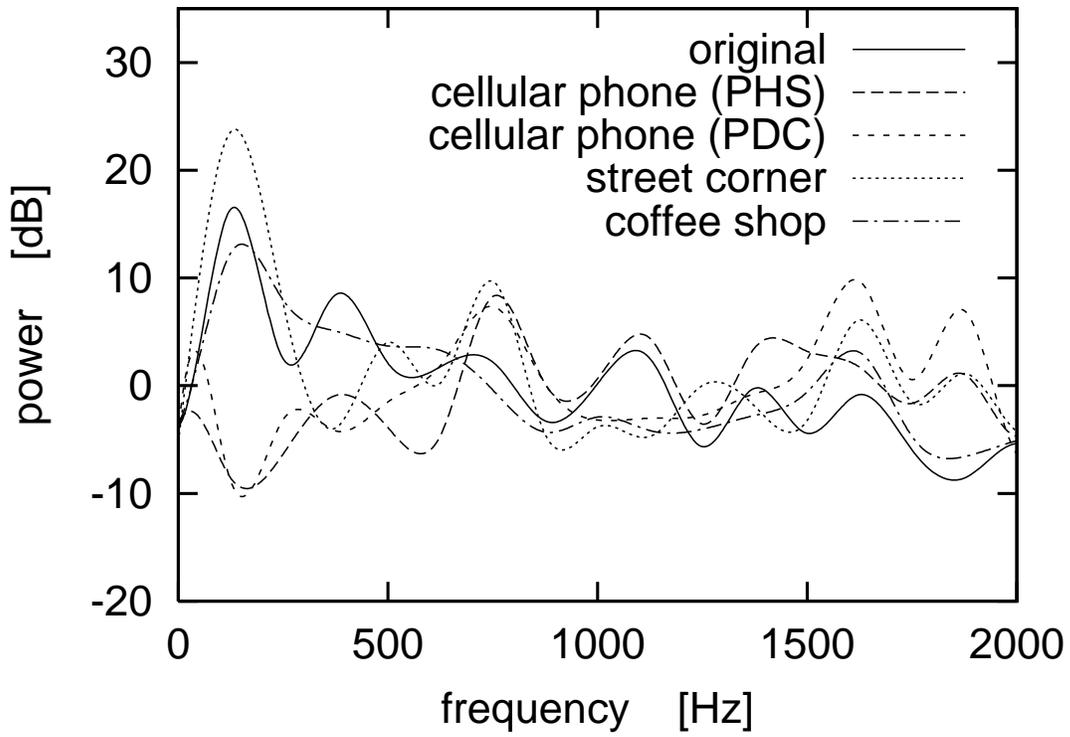


図 3.2: 楽曲の一部分 2

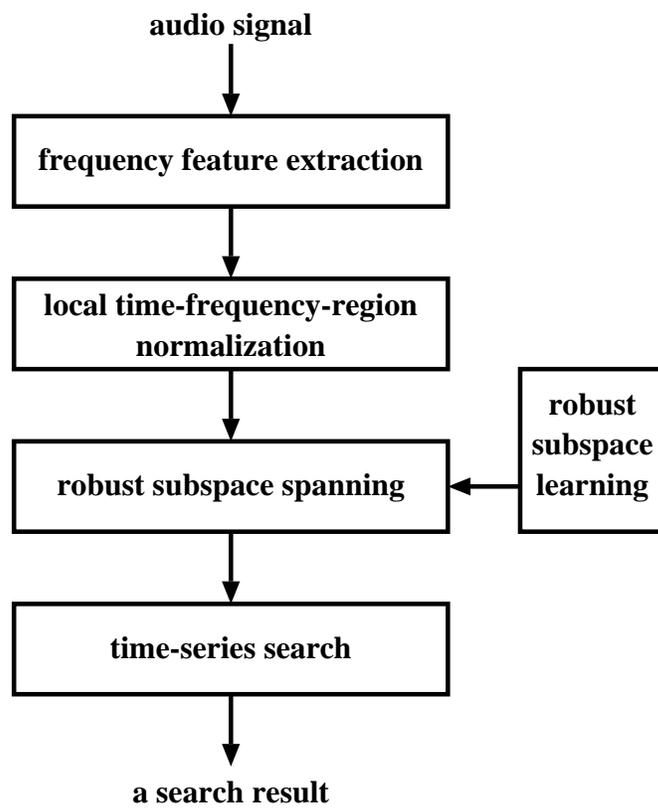


図 3.3: 音響信号探索の処理手順

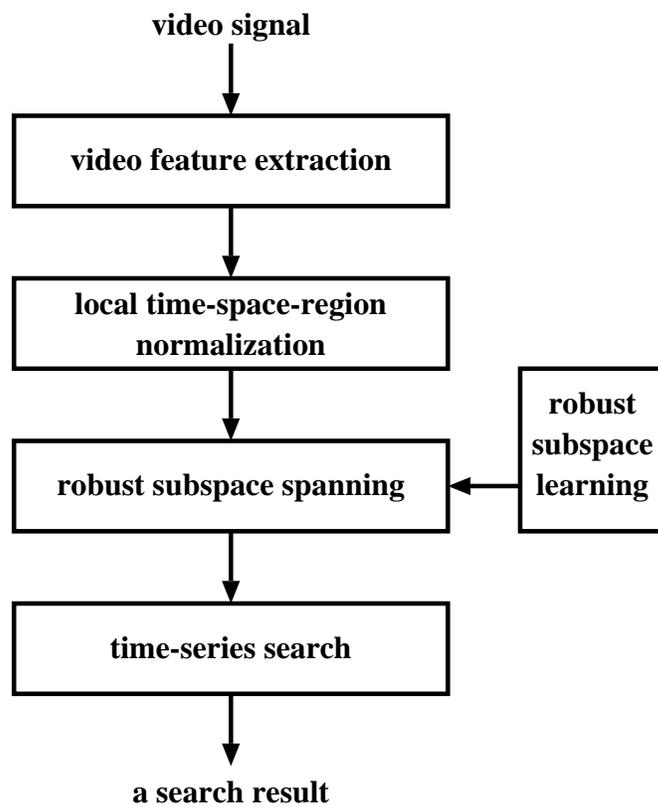


図 3.4: 映像信号探索の処理手順

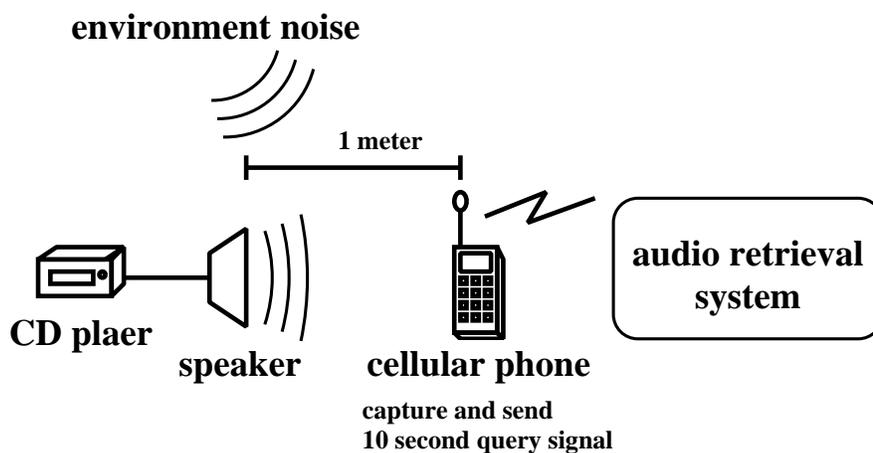


図 3.5: 音響信号探索の実験装置

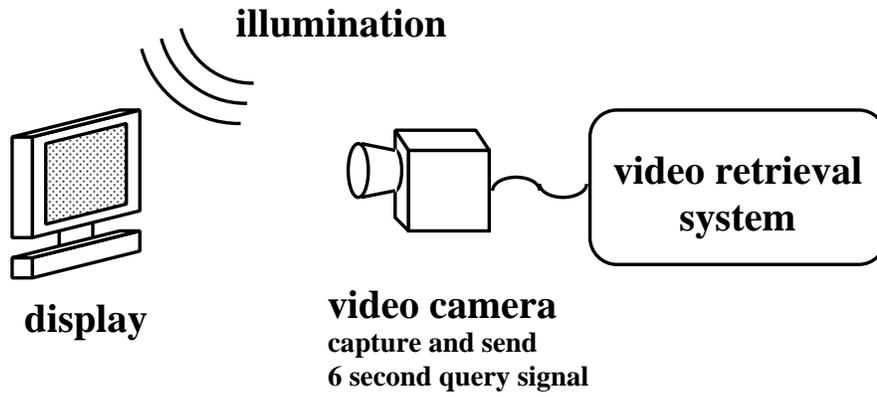


図 3.6: 映像信号探索の実験装置

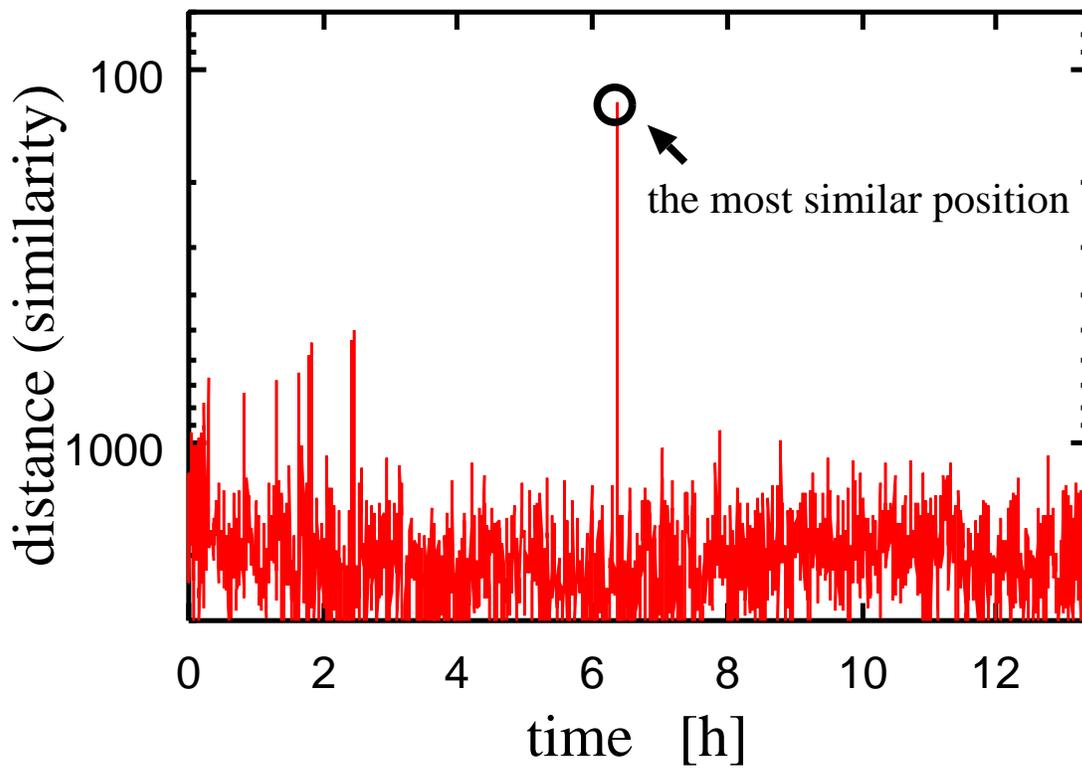


図 3.7: 探索の例

第 4 章

生成特徴に基づく探索

4.1 はじめに

本章では、さまざまな変動を考慮してひずみに対処することを目的として、生成特徴に基づく探索のアプローチを用いた手法について述べる [40, 41]。ここでは、想定されるさまざまなひずみを事前にモデル化しておき、そのモデルに基づいて特徴を生成することで、変動に対処することを試みる。基本的なアイデアは、機器特性や圧縮などによって発生するひずみの統計的な性質を探索に前だって事前に学習しておき、探索時にその性質に基づいて特徴を生成し探索するものである。以下、提案法の概要と、その妥当性を検討するための実験について順に説明する。

4.2 問題と解決の方針

さまざまな変動を考慮して時系列探索を実現するには、考慮する変動が多いとその探索の範囲が膨大になるため、高速に探索することが難しい。例えば、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をさまざまな変動を考慮しながら特徴を生成して探索していくことが考えられるが、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、高速性に主眼をおいた手法として、時系列特徴に対し一旦ベクトル量子化 (VQ) を行って符号時系列にした後、BM 法 [42] 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近似的に探索することで高速化する手法 [43] , クラスタリングと木構造を利用した多次元インデックス手法 [44, 45] などがある . しかしながら , これらの手法は高速に探索できるが , 特徴に歪みがある場合 , 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による歪みの影響のために , 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歪みを考慮し , 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 [6] . 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持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 , 予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 秒あたり 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 [46] .

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ー量子化 (SQ) し , その SQ 符号の組合わせで表現される領域ごとに VQ していた [6] . しかしながら , 実際の特徴ベクトルは , ベクトル空間上で偏りを持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言えない .

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閾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 . しかしながら , 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 .

本章では , 以上の 2 つの問題点を解決するため , 次の 2 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 [47, 40]。特に後者は、さまざまなノイズに対応した目的信号を OR 探索 [48, 49] することで、歪みに頑健な時系列探索を可能にすると考えられる。

4.3 時系列アクティブ探索法

4.3.1 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 4.1 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、ヒストグラムどうしの類似度が、予め設定した値（探索閾値）を越えるかどうかで、目的信号の有無を判定する。このとき、類似度と探索閾値とから、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 [6]。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、

$$S = S(H_Q, H_S) = \frac{1}{D} \sum_{l=1}^L \min(h_{Ql}, h_{Sl}) \quad (4.1)$$

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予め定めておいた探索閾値 θ 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。

$$w = \begin{cases} \lfloor D(\theta - S) \rfloor + 1 & (S < \theta \text{ のとき}) \\ 1 & (\text{上記以外}). \end{cases} \quad (4.2)$$

ただし、 $\lfloor \cdot \rfloor$ は切下げを表す。

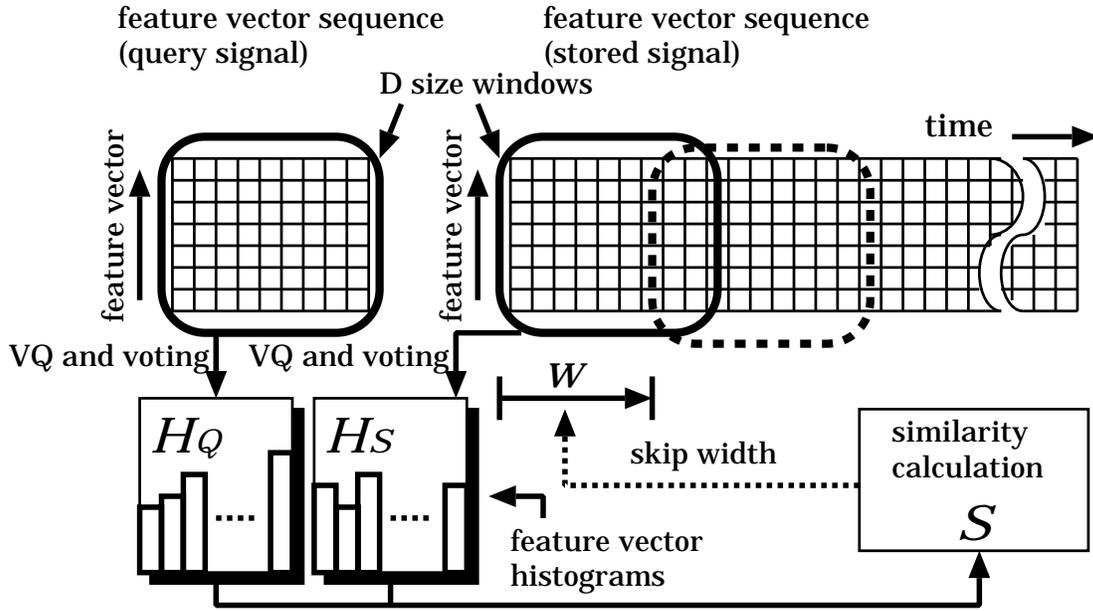


図 4.1: 時系列アクティブ探索法の概要

4.3.2 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 [50] であるが，ここでは映像探索について扱う．本章では，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，

$$x(k) = (x_{1r}(k), x_{1g}(k), x_{1b}(k), \dots, x_{jc}(k), \dots, x_{Wr}(k), x_{Wg}(k), x_{Wb}(k)) \quad (4.3)$$

と定義する．ここで k はフレームの時刻であり， x の添字 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，

$$x_{jc}(k) = \frac{\bar{y}_{jc}(k) - \min_i \bar{y}_{ic}(k)}{\max_i \bar{y}_{ic}(k) - \min_i \bar{y}_{ic}(k)} \quad (4.4)$$

である。ただし，

$$\bar{y}_{ic}(k) = \frac{1}{|I|} \sum_{p \in I} y_{pc}(k) \quad (4.5)$$

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 のうちの一色 c の画素値を表す。

4.4 特徴ベクトルの分布と歪みの学習

4.4.1 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 符号の組合わせで表現される領域ごとに VQ していた。しかしながら，その手法よりも，実際の特徴ベクトルの分布にしたが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。本章では，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 k 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 [51]。また，特徴ベクトルを距離がもっと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。

4.4.2 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 4.6 のように，入力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

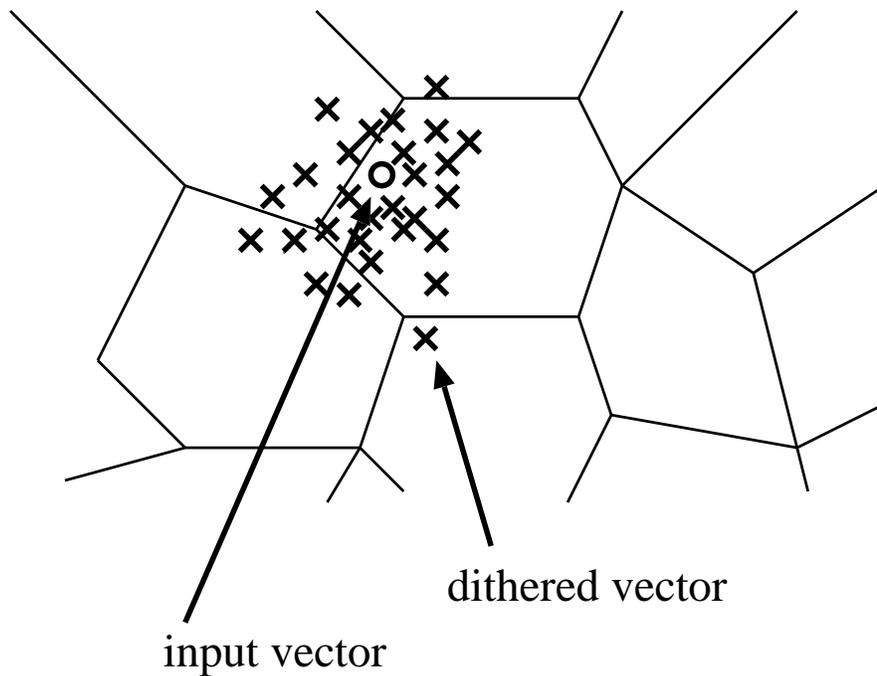


図 4.2: 確率分布に基づいてばらつかせた特徴ベクトル

率分布に基づいてばらつかせると，ノイズ等の歪みによって投票される可能性があるところに投票され，歪みに頑健になると考えられる．なお，投票する票の数は，1つの特徴ベクトルあ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，1つの特徴ベクトルあたり少数の投票を行った方が良い．

同様に複数のピ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 [52]．しかしながら，Fuzzy Active 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歪みが発生している．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歪み，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．これらの歪みは発生メカニズムや性質が全く異

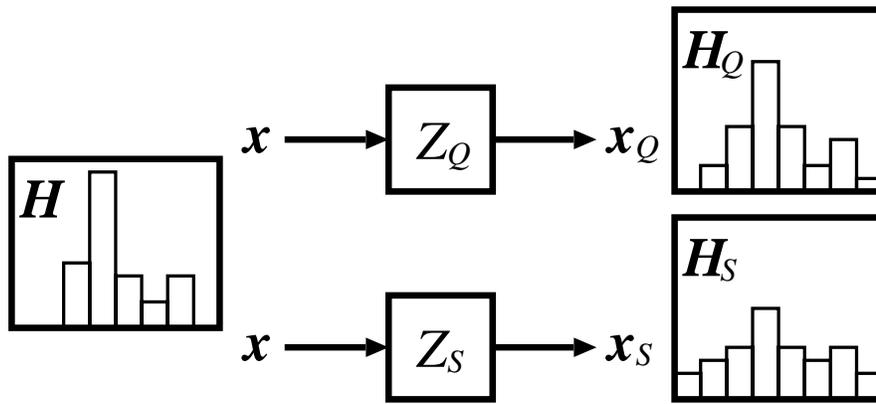


図 4.3: 元信号と目的信号，蓄積信号の関係

なるが，これらの歪みの全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，全ての歪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 4.3 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，

$$x_Q(k, t_Q) = x(k) + \varepsilon_Q(k + t_Q) \quad (4.6)$$

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たり，MPEG による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，

$$x_S(k, t_S) = x(k) + \varepsilon_S(k + t_S) \quad (4.7)$$

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，

$$h_l(k) = \sum_{n=k}^{k+D} \int \delta(\mathbf{y} - \mathbf{x}(n)) \phi_l(\mathbf{y}) d\mathbf{y} \quad (4.8)$$

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

$$\phi_l(\mathbf{y}) = \begin{cases} 1 & (\mathbf{y} \text{ が VQ コード } l \text{ の領域内}) \\ 0 & (\text{上記以外}) \end{cases} \quad (4.9)$$

とする．ここで，目的信号 $\mathbf{x}_Q(k, t_Q)$ 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，

$$\begin{aligned} E\{h_{Ql}(k, t_Q)\} &= \sum_{n=k}^{k+D} \int E\{\delta(\mathbf{y} - \mathbf{x}_Q(n, t_Q))\} \\ &\quad \times \phi_l(\mathbf{y}) d\mathbf{y} \\ &= \sum_{n=k}^{k+D} \int \int \delta(\mathbf{y} - \mathbf{x}(n) - \boldsymbol{\varepsilon}_Q) \\ &\quad \times P_Q(\boldsymbol{\varepsilon}_Q) \phi_l(\mathbf{y}) d\boldsymbol{\varepsilon}_Q d\mathbf{y} \\ &= \sum_{n=k}^{k+D} \int P_Q(\mathbf{y} - \mathbf{x}(n)) \phi_l(\mathbf{y}) d\mathbf{y} \end{aligned} \quad (4.10)$$

となる．ただし， E は t_Q に関する平均， $P_Q(\boldsymbol{\varepsilon}_Q)$ は $\boldsymbol{\varepsilon}_Q$ の確率密度関数とする．同様に，蓄積信号 $\mathbf{x}_S(k, t_S)$ 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，

$$E\{h_{Sl}(k, t_S)\} = \sum_{n=k}^{k+D} \int P_S(\mathbf{y} - \mathbf{x}(n)) \phi_l(\mathbf{y}) d\mathbf{y} \quad (4.11)$$

となる．ただし， $P_S(\boldsymbol{\varepsilon}_S)$ は $\boldsymbol{\varepsilon}_S$ 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 4.4 のように目的信号 $\mathbf{x}_Q(k, t_Q)$ にシステム Z_S を介して変形目的信号 $\mathbf{x}_{QS}(k, t_Q)$ ，蓄積信号 $\mathbf{x}_S(k, t_S)$ にシステム Z_Q を介して変形蓄積信号 $\mathbf{x}_{SQ}(k, t_S)$ を生成する．ここで，変形目的信号 $\mathbf{x}_{QS}(k, t_Q)$ 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，

$$E\{h_{QSl}(k, t_Q)\} = \sum_{n=k}^{k+D} \int P_{QS}(\mathbf{y} - \mathbf{x}(n)) \phi_l(\mathbf{y}) d\mathbf{y} \quad (4.12)$$

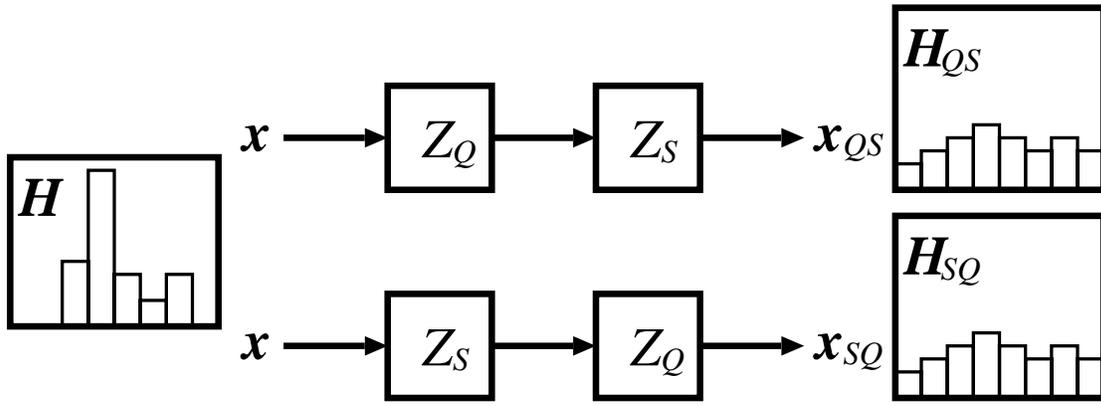


図 4.4: 元信号と変形目的信号と変形蓄積信号の関係

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 $h_{SQl}(k, t_S)$ の期待値は，

$$E\{h_{SQl}(k, t_S)\} = \sum_{n=k}^{k+D} \int P_{SQ}(\mathbf{y} - \mathbf{x}(n)) \phi_l(\mathbf{y}) d\mathbf{y} \quad (4.13)$$

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，確率ディザボート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 ϵ_S の確率分布にしたがってばらつかせて，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 ϵ_Q の確率分布にしたがってばらつかせて，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。

4.4.3 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，その歪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，本章では簡単のため，システムを介した信号は，正規分布にしたがって変形すると仮定する．すなわち， $\varepsilon_Q, \varepsilon_S$ 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 ε_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 ε_Q の特性を観測する．

まず， ε_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2つ用意する．ただし，2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，平均二乗誤差 $\text{MSE}(x_Q(k, t_{Q1}), x_Q(k, t_{Q2}))$ を求める．ただし，

$$\begin{aligned} \text{MSE}(\mathbf{x}(k), \mathbf{y}(k)) \\ = E\{(\mathbf{x}(k) - \mathbf{y}(k)) \cdot (\mathbf{x}(k) - \mathbf{y}(k))\}, \end{aligned} \quad (4.14)$$

E は k についての平均である．ここで， ε_Q の分散を σ_Q^2 とすると，

$$\begin{aligned} \text{MSE}(x_Q(k, t_{Q1}), x_Q(k, t_{Q2})) \\ = E\{\varepsilon_Q(k, t_{Q1})^2\} \\ \quad - 2E\{\varepsilon_Q(k, t_{Q1}) \cdot \varepsilon_Q(k, t_{Q2})\} \\ \quad + E\{\varepsilon_Q(k, t_{Q2})^2\} \\ = E\{\varepsilon_Q(k, t_{Q1})^2\} + E\{\varepsilon_Q(k, t_{Q2})^2\} \\ = 2\sigma_Q^2 \end{aligned} \quad (4.15)$$

となる．同様に， $\varepsilon_Q, \varepsilon_S$ 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，平均二乗誤差は

$$\begin{aligned} \text{MSE}(x_Q(k, t_Q), x_S(k, t_S)) \\ = E\{\varepsilon_Q(k, t_Q)^2\} + E\{\varepsilon_S(k, t_S)^2\} \\ = \sigma_Q^2 + \sigma_S^2 \end{aligned} \quad (4.16)$$

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

式 (4.15)，式 (4.16) のそれぞれは前もって学習により求める必要がある．平均二乗誤差の計算は，2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二乗誤差を計算する．この平均二乗誤差から計算して得た σ_Q^2, σ_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トニングを行う．

4.5 実験

4.5.1 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1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良いためである¹．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²．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，探索閾値を次式によって定めた．

$$\theta = m + t\nu. \quad (4.17)$$

ここで， m と ν 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 (4.17) において， θ が 1 を越えるときは， $\theta = 1$ ，0 を下回るときは $\theta = 0$ とした．本実験では，式 (4.17) における t の値を 200 回繰り返し中一定とし，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

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，本章の実験結果よりも探索精度が低下する場合があるので，注意が必要である．

²取り込みはフレームレート 29.97Hz，フレーム毎に 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 Mbps である．

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もっと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 (4.3) 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 4.5 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との関係を示す．これによれば，1.0 ~ 20.0 秒のいずれにおいても，数ポイント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 2.0 秒のとき探索精度が 88.6% であるのに対し，特徴ベクトルの分布に基づく VQ では探索精度が 97.6% となり，9.0 ポイント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．

4.5.2 確率ディザボーティング

次に，確率ディザボーティングにより，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．4.5.1 節で用いた 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．

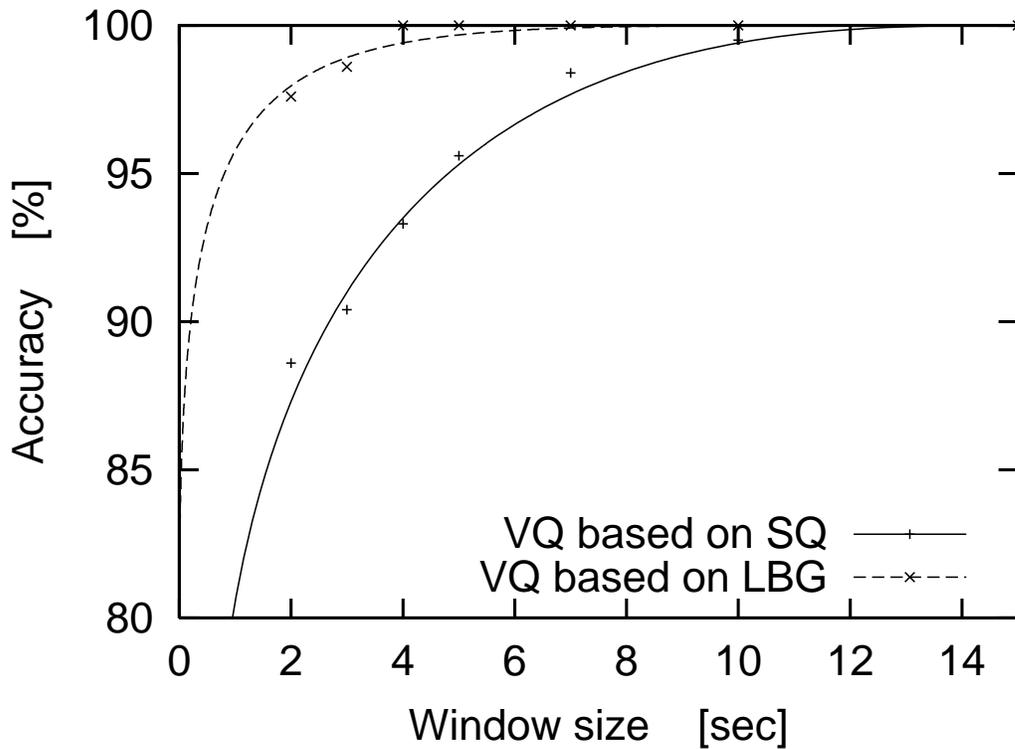


図 4.5: 照合区間の長さ と探索精度の関係

- (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³。

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 σ_Q^2 については (B-1) と (B-2) の平均二乗誤差から式 (4.15) により計算し，蓄積信号のノイズの分散 σ_S^2 については (B-1) と (B-4) の平均二乗誤差から式 (4.16) により計算した．なお，式 (4.16) における σ_Q^2 は，(B-1) と (B-2) の平均二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 (4.3) 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，4.5.1 節で 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，確率ディザボータリングは 1 つの特徴ベクトルあたり 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(250MHz) 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 4.6 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 (従来法) と確率ディザボータリングを行ったもののそれぞれの照合区間の長さや探索精度の関係を示す．これによれば，1.0 ~ 20.0sec のいずれにおいても，数ポイント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 2.0sec のとき探索精度が 79.5% であるのに対し，確率ディザボータリングでは探索精度が 84.0% となり，4.5 ポイント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 5.0sec に固定して実験を行った．表 4.1 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リングを用いた方が探索精度が数ポイント向上している．また，良く似た 2 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，確率ディザボータリ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 4.1 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閾値 θ から $\theta - 0.1$ の範囲の映像 163 件について，確率ディザボータリ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

³(5) は 64×48 画素に縮小した後，ビットレート 330kbps の JPEG 圧縮を行った．(6) は 64×48 画素に縮小した後，ビットレート 100kbps の MPEG 圧縮を行った．

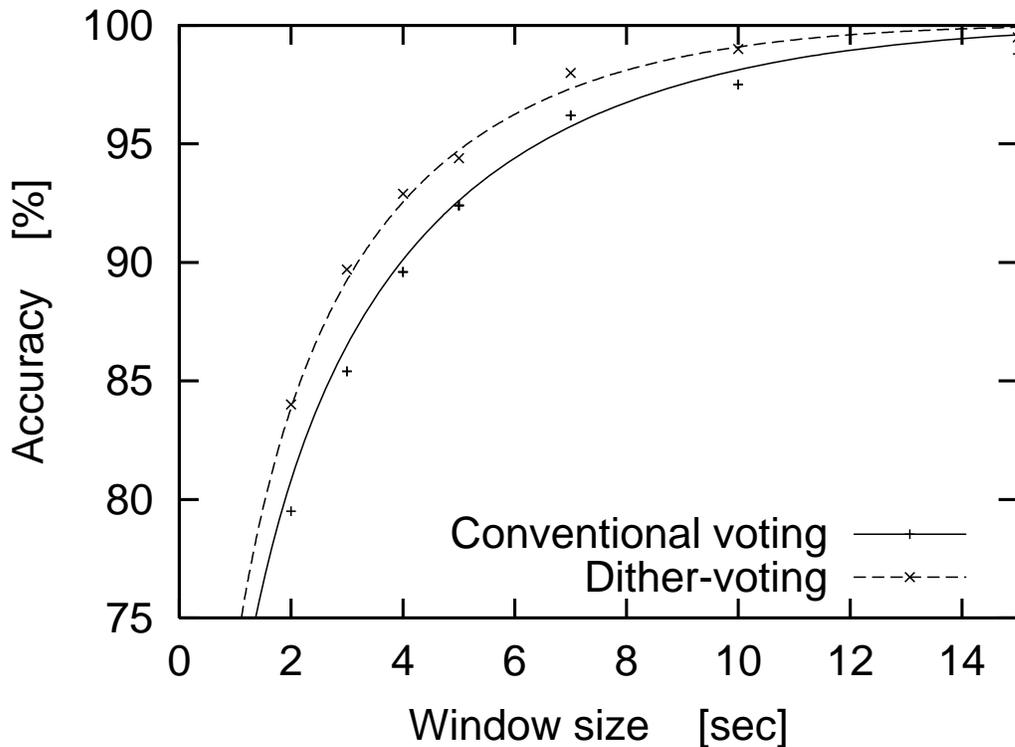


図 4.6: 照合区間の長さ と探索精度の関係

の結果，確率ディザボータイングを行わなかった場合，探索閾値との差分値 $\theta - S$ の平均値は 0.061 であったのに対し，確率ディザボータイングを行った場合では，0.101 に上昇した．よって，確率ディザボータイ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，確率ディザボータイ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．

4.6 考察

本章で導入したひずみモデルは，加法性の正規分布を用いたモデルであった．実験結果が示すように，本章で扱った歪みの範囲では，このようなモデル化を行って特徴を生成することで良好な結果が得られた．しかしながら，ここで扱った変動の範囲は，非常に限定的なものであった．仮りに本手法の延長上で変動の範囲

表 4.1: ダビング，圧縮を行った映像での探索精度

	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

が大きい場合に対処しようとするとき，生成すべき特徴が膨大となり，探索が困難となることが予想される．例えば，幾何変換を生成のアプローチで対処するような場合を考えると，変動の範囲が大きいために全ての変動の範囲を生成していたのでは，膨大な時間がかかると考えられる．

4.7 まとめ

本章では，さまざまな要因によって起こる変動をモデル化し，そのモデルに基づいて特徴を生成して探索する手法について検討した．ここでは，時系列アクティブ探索法を利用した探索を行い，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イングを用いてヒストグラム上に表現することで，さまざまなノイズに対応した探索が可能になることを示した．

本章で導入したひずみモデルは，加法性の正規分布を用いたモデルだった．実験結果が示すように，本章で扱った歪みの範囲では，このようなモデル化を行って特徴を生成することで良好な結果が得られた．しかしながら，ここで扱った変動の範囲は，非常に限定的なものであった．仮りに本手法の延長上で変動の範囲が大きい場合に対処しようとするとき，生成すべき特徴が膨大となり，探索が困難となることが予想される．例えば，幾何変換を生成のアプローチで対処するような

場合を考えると、変動の範囲が大きいため全ての変動の範囲を生成していたのでは、膨大な時間がかかると考えられる。そこで、第3章や後述の第5章のアプローチと組み合わせることで、本アプローチは現実的なものとなると考えられる。

第 5 章

スパースな特徴選択

本章では、頑健な特徴を抽出するという観点から、特徴の選択手法について述べる。これまで検討されてきた音響信号探索に利用する特徴として、周波数スペクトルをベクトル量子化しその量子化値の出現頻度をヒストグラムとして表現するもの [7]、時間周波数空間上で局所的に正規化するもの [14]、周波数差分と時間差分を用いるもの [10]、パワーのピークを用いるもの [11] などが提案されている。[7] の特徴は、パワー周波数分布を効率的に圧縮して表現することができ、高品質な音響信号を探索するためには非常に高速に探索することができ有効であった。しかしながら、携帯電話で受音した音のように、極端に劣化した音を探索する場合には、周波数分布が変形してしまい、十分な探索精度で探索することができないという欠点があった。また、[14] の特徴は、定常的な乗法性ひずみと加法性雑音を吸収でき、実環境での探索を実現するものであるが、非定常な雑音や欠落に対しては十分な精度で探索できないという欠点があった。[10, 11] の特徴は、よりシンプルに特徴を表現するものであるが、音の特徴が周辺の音に影響されやすい可能性があり、特徴の選択方法については、まだ十分な検討が必要と考えられる。また、探索に利用する映像データの特徴表現として、画像フレームをベクトル量子化しその量子化値の出現頻度をヒストグラムとして表現するもの [7]、画像フレームの色ヒストグラムをバイナリで表現するもの [53]、画像フレームの全体のパターンの DCT 低周波成分を量子化して表現するもの [54, 55] などが提案されている。これらの特徴表現は目的信号と蓄積信号の品質が同程度であることが前提となっているが、シンプルな表現であるため、高速な探索に適している。また、

高精度化を主眼とした探索手法としては、既知の特徴ひずみを事前に学習により求めておき、それを用いて原信号にひずみを付加しながら探索する手法 [40] が提案されている。しかしながら、ここで扱っている品質劣化は、データ圧縮によるひずみやダビングによる品質劣化など、比較的少ない劣化の範囲にとどまっている。

このような背景のもと、本章では、与えられた未知の映像断片をキーとして、膨大な映像データからそれと同一の映像を探索する手法を提案する。特に、探索のキーとなる映像断片として、ディスプレイに表示されている映像をカメラで収録した映像やテロップ表示や映像編集などにより映像が遮蔽され欠落してしまっている場合への対応も考慮し、ひずみ、照明変動、遮蔽、欠落に対して頑健に探索する方法について議論する。特に、ここでは、疎らな密度で特徴を選択するスパースな特徴選択について述べる。

5.1 問題と解決の方針

本章の基本的なスタンスは、音響信号や映像信号の収録条件の違いに対して、できるだけ頑健な特徴を選ぶというものである。そこでまず、収録条件の違いによって発生する信号の変動を、定常な変動と非定常な変動に分離し、それぞれの変動の吸収と回避の方法について検討する。

映像の内容によって画素値の変動の大きい色や領域、変動の少ない色や領域があると考えられる。上記のような不変な特徴量を利用した場合、変動の小さい箇所の特徴は、変動の大きい箇所の特徴と比べて、信号の劣化の影響を受けやすい。そこで、特定の色や画像領域を選択し、その選択された箇所のみを照合することが探索精度向上につながると考えられる。

また、上記のような特徴の劣化にはひずみ以外に別のオブジェクトによる遮蔽によって信号が欠落することが考えられる。得られた目的信号には、どの箇所に遮蔽物体があるかを知ることはできないため、遮蔽物体を含む映像を目的信号として探索するためには、遮蔽物体が類似度に与える影響の少ない照合法が必要となる。一般的に画像の照合に用いる代表的な距離として $L1$ 距離や $L2$ 距離が挙げられる。画像がオブジェクトによって遮蔽されているような場合には、欠落部分に入る画像パターンによって距離に与える影響が異なるため、安定した距離値が

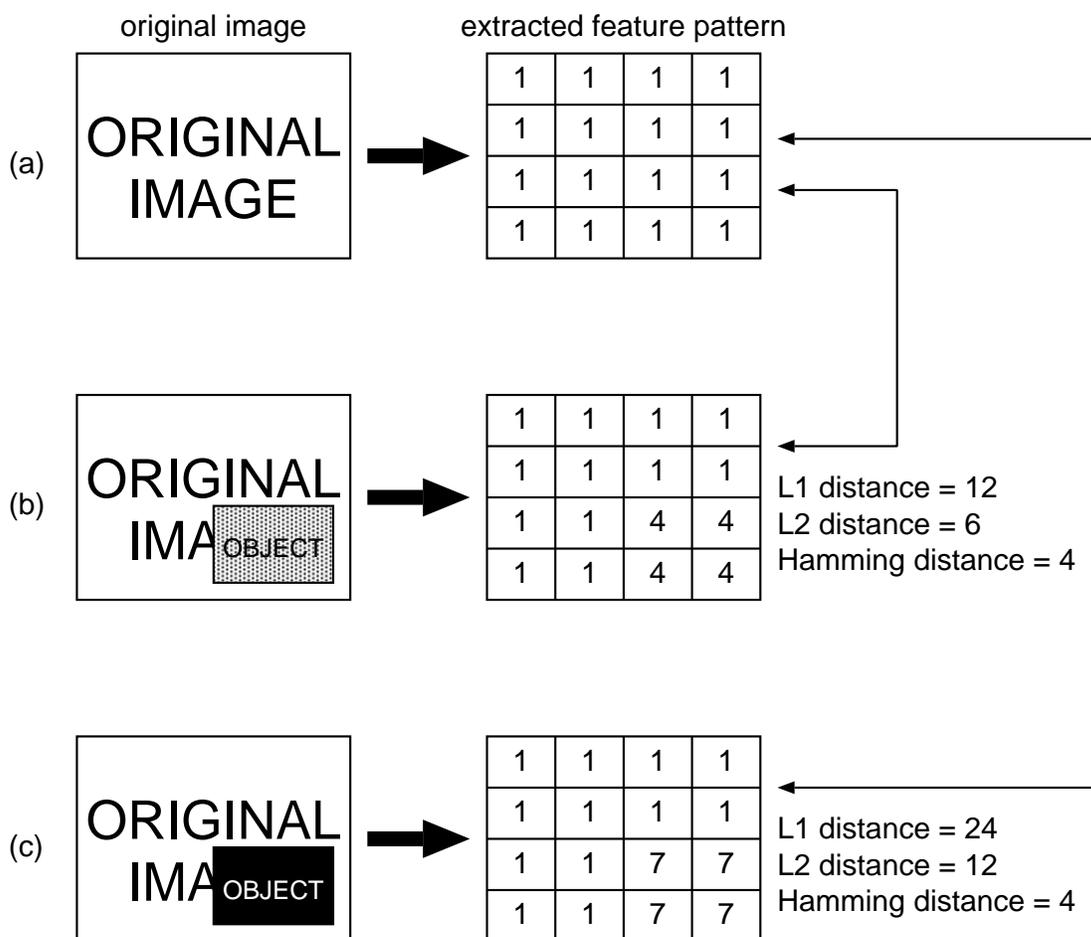


図 5.1: 遮蔽がある場合の距離

得られない．例えば，図 5.1 の (a) のようなオリジナルの画像が (b) や (c) のように面積が同じで画素値の異なるオブジェクトによって遮蔽されている場合，L1 距離や L2 距離を利用した場合，(a) と (b) との間の距離値と，(a) と (c) との距離値が異なる．そこで，遮蔽物体の違いが距離へ影響しないように，特徴ベクトルの各要素をスカラー量子化しその非一致度，すなわち，ハミング距離 [33] を用いることで安定した距離値を獲得できる．ハミング距離を利用すると，遮蔽による距離値への影響が遮蔽部分の画素値に依存せず，距離値が単純に遮蔽物体の面積の値となるため，本章で扱うような目的信号と蓄積信号が同一パターンであることが前提の探索の課題には適した距離尺度である．

5.2 探索の原理

本手法の基本的なアイデアは，変動を吸収するための正規化，統計量に基づいたスパースな特徴選択，遮蔽に頑健な特徴の量子化値の一致度による照合にある．正規化は，ディスプレイ特性，カメラ特性，照明変動などによって発生する目的信号と蓄積信号の間の信号の変動を吸収するための処理である．スパースな特徴選択は，変動による影響の少ない特徴的な箇所のみを照合に利用することにより頑健な探索を行うための処理である．量子化値の一致度による照合は，雑音や欠落などの非定常な変動に対して安定した類似度を得るための処理である．

探索の具体的な処理は，(1) 音響特徴または映像特徴の抽出，(2) 時間空間上での局所領域ごとの正規化，(3) スパースな特徴選択，(4) 局所正規化値の量子化，(5) 時系列探索の順で行う(図 5.2)．

5.2.1 映像特徴の抽出

本章では，携帯カメラで収録した映像中のディスプレイ表示部分を切り出し，ディスプレイ部分のカラー縮小画像を映像特徴として用いた．ここで，映像特徴ベクトル $x(k)$ を，

$$\begin{aligned} x(k) = & (x_{1r}(k), x_{1g}(k), x_{1b}(k), \dots, x_{jc}(k), \\ & \dots, x_{Wr}(k), x_{Wg}(k), x_{Wb}(k)) \end{aligned} \quad (5.1)$$

と定義する．ここで k はフレームの時刻であり， x の添字 j は各フレームのディスプレイの表示部分を W 個のサブ画像に分割した分割番号，添字 c は RGB を表す．ただし， W は経験的に与えられる値である． 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た値であり，

$$x_{jc}(k) = \frac{1}{|I|} \sum_{p \in I} v_{pc}(k) \quad (5.2)$$

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， $v_{pc}(k)$ は画素 p の RGB のうちの単一色 c の画素値を表す．なお，本章では，ディスプレイ表示部分の切り出しは手動で行ったが，自動で切り出す必要がある場合には，ある時間区間で画素値の変化の大きい領域を抽出し，アフィン変換により切り出すなどの工夫が必要である．

5.2.2 時空間上の局所領域ごとの正規化

この処理の目的は、数秒程度の短時間で小さい変動の明るさやコントラスト比の違いを吸収することである [15]。ここでは、各画像領域ごとに、ある時間区間の値から平均と標準偏差を求め、それらを用いて正規化を行う。すなわち、正規化後の映像特徴 $y(i)$ の k 番目の要素は、

$$y(i, k) = \frac{1}{\sigma(i, k)}(x(i, k) - m(i, k)) \quad (5.3)$$

である。ただし、 $x(i, k)$ は、 $x(k)$ の i 番目の要素、

$$m(i, k) = \frac{1}{2M} \sum_{i=-M}^{M-1} x(i, k), \quad (5.4)$$

$$\sigma(i, k)^2 = \frac{1}{2M} \sum_{i=-M}^{M-1} (x(i, k) - m(i, k))^2, \quad (5.5)$$

M は平均、標準偏差を求めるための窓の大きさの半分の値である。またこれは、時間区間内の値をサンプルとみたてたガウシアン正規化と考えることもできる。映像特徴の平均、標準偏差を計算する時間窓の中で加法性雑音と乗法性ひずみが定常であるとすると、この処理により照明による変動とカメラやディスプレイの明るさとコントラスト比の違いを同時に吸収できると考えられる。

5.2.3 スパースな特徴選択

この処理の目的は、ひずみや雑音などによる特徴の変動の影響を受けにくい箇所を選択し、頑健な特徴を得ることにある。ここでは、特徴変動の影響を受けにくい箇所を統計量に基づいて疎らな密度で特徴を選択する。特徴的な箇所のある時間区間の時間平均に対して大きく変動している箇所と仮定し、

$$z(t, f) = |x(t, f) - m(t, f)| \quad (5.6)$$

で得られる平均に対する変動の絶対値の大きい箇所を近傍の時刻から上位 N 個を選択する。なお、選択の様子を図 5.3 に示す。

5.2.4 局所正規化値の量子化

続いて，選択された特徴の量子化を行う．量子化は，局所正規化値に基づいて行う．量子化の方法は非線形量子化やベクトク量子化などが考えられるが，ここでは，単純に線形スカラー量子化を行う．

具体的には，式(5.3)により得られる局所正規化値を次式によって等間隔に L レベルに量子化する．

$$y'(t, f) = \begin{cases} \lfloor \frac{L \times (y(t, f) + l)}{2l} \rfloor & \text{if } |y(t, f)| < l \\ L - 1 & \text{else if } y(t, f) \geq l \\ 0 & \text{otherwise} \end{cases}$$

ただし， l は量子化値を決めるための値， $\lfloor \cdot \rfloor$ は切下げを表す．

5.2.5 量子化値の一致度に基づく時系列探索

最後に，局所正規化値を量子化することにより得られた量子化値を照合しながら時系列探索を行う．時系列探索は，蓄積信号中に目的信号と同じ大きさの窓をかけ，目的信号と蓄積信号中の注目箇所との類似度を計算しながら，窓をずらし，目的信号と最も類似した蓄積信号中の箇所を探索結果として出力する．本手法では，類似度として選択された特徴の量子化値の一致度を用いる．すなわち，目的信号中の符号と蓄積信号中の窓内の符号において，時刻と周波数と量子化値のいずれもが一致するものの個数を類似度とする．これは，選択された箇所のハミング距離を計算することに相当する．

5.3 実験

本手法の有効性について検討するため，さまざまな品質の映像信号を目的信号として，蓄積信号から同一部分を探査し，探査精度を評価する実験を行った(図 5.4，図 5.5)．ここでは，図 5.2 における各処理の有効性を確認する実験(実験 1)，長時間の蓄積信号での探査の精度と速度を確認する実験(実験 2)を行った．

表 5.1: 実環境で収録した映像信号

	内容
映像	CM 60 分 (15 秒 CM を繋合わせた映像)
ディスプレイ	LCD ディスプレイ 2 機種 ブラウン管ディスプレイ 2 機種
カメラ	DV カメラ 4 機種

表 5.2: 蓄積信号

実験 1	CM 60 分
実験 2	CM 60 分, 99 時間分の放送映像, 合計 100 時間分

5.3.1 実験データの収集

実験に先だって, ディスプレイ, カメラを複数選び, 室内環境下にて CM 映像の収録を行った. これらの収録条件を表 5.1 に示す. 実環境で収録した映像信号は, 表 5.1 の全ての組み合わせ, すなわち合計 16 条件の映像信号(各 60 分)である.

また, 遮蔽物体を含む映像信号として, 図 5.6 のように, 蓄積信号と同一の内容の映像の上に画面の 4 分の 1 の面積をグレー領域(256 階調カラーで RGB それぞれが 128 レベルの領域)で覆い隠し, 左上, 右上, 左下, 右下, 中央のそれぞれに配置したもの合計 5 種類を用意した. なお, 編集に用いた映像は, 実環境収録で使用したのと同じ CM 映像である. 用意した目的信号と蓄積信号を表 5.2, 表 5.3 にまとめる.

5.3.2 各処理による効果(実験 1)

実験は, 目的信号の長さを 8 秒とし, 目的信号の切り出し部分を 200 回変えて探索精度の測定を行った. なお, 映像信号は 29.97 frames/s のフレームレート,

表 5.3: 目的信号

オリジナル	蓄積信号の CM 60 分と同一の映像から切り出した 200 個の映像断片
実環境収録	CM 60 分を全てのディスプレイとカメラの組合せ 16 組について収録した映像からディスプレイの領域についてそれぞれ 200 箇所を切り出した合計 3200 個の映像断片
遮蔽物あり	CM 60 分を遮蔽物体ありの映像 5 本からそれぞれ 200 箇所を切り出した合計 1000 個の映像断片

176 × 144 の画像サイズでキャプチャした RGB 各 256 階調の映像を使用し, 15.00 frames/s にフレームレート変換をした後, 映像特徴を $W = 25$ (横方向 5 分割, 縦方向 5 分割) で抽出, その後, 局所正規化を $M = 15$ で行い, 特徴の選択を 1 フレーム内で $N = 14$ で行い, 量子化を $Q = 10, r = 3.2$ で行った. また, 目的信号は, 表 5.3 の条件で収録した映像信号を合計 22 種類用意し, 合計 4400 回の探索を行った.

4400 回の探索において各閾値係数 t に対する適合率と再現率の関係を図 5.7 にまとめる. 探索結果は類似度を閾値処理することにより求めるが, 閾値は次式によって定めた.

$$\theta = m + tv. \quad (5.7)$$

ここで, m と v 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. 本実験では, 式 (5.7) における t の値を全探索について固定とし, 全 4400 回の探索における適合率 (precision rate) の平均値が 99% となるときの再現率 (recall rate) の平均を探索精度として示している. すな

表 5.4: 各処理の効果 (実験 1)

処理	類似度	探索精度 (適合率 99% のときの再現率)		
		オリジナル	実環境収録	遮蔽物体あり
縮小画像	L2 距離	32.00%	8.31%	5.10%
+ 局所正規化	L2 距離	98.50%	50.91%	91.70%
+ 量子化	符号一致度	100.00%	78.81%	99.70%
+ 特徴選択 (提案法)	符号一致度	100.00%	93.72%	99.90%

わち，探索により出力される探索結果のほぼ全てが正解となる．なお，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．なお，ここでは，探索した時刻が正しい時刻の ± 7.5 秒以内となった場合に正解とした．

次に，各処理を付加したときの探索精度を表 5.4 にまとめる．ここでは，処理の効果と比較するために，局所正規化を行わずに映像特徴を時系列に並べたものを特徴ベクトルとし L2 距離 (ユークリッド距離) を類似度尺度としたときの探索精度，局所正規化を行った後に時系列に並べたものを特徴ベクトルとし L2 距離を類似度尺度としたときの探索精度，局所正規化を行った後にスカラー量子化し符号一致度を類似度尺度としたときの探索精度，および，提案法の探索精度を示す．

局所正規化を行わなかった場合と比べて，局所正規化を行った場合，量子化を行った場合，スパースな特徴選択を行った場合，いずれにおいても，実環境収録，遮蔽物体ありの両方で，探索精度の向上が確認された．これにより，局所正規化，スパースな特徴選択，局所正規化値の量子化値の一致度による照合のいずれも有効であることが確認された．

5.3.3 長時間の蓄積信号を用いた実験 (実験 2)

次に長時間の映像データベースでの探索精度と探索速度を評価するため，より長い蓄積信号を用いた実験を行った．ここでは，表 5.3 に示すように実験 1 で使

表 5.5: 長時間の蓄積信号での探索精度 (実験 2)

	オリジナル	実環境収録	遮蔽物体あり
100 時間の蓄積信号	100.00%	93.00%	99.80%

表 5.6: 処理時間 (実験 2)

探索 1 回の平均処理時間	探索速度
8.25 秒	12.12 hours/s

用したものに加えて 99 時間分の映像データを加え合計 100 時間分の映像を蓄積信号として使用した。実験は、実験 1 と同様に合計 4400 回の探索を行った。

実験結果を表 5.5 に示す。なお、探索精度は、実験 1 と同様の方法で測定した。60 分の蓄積信号での探索精度と比べると精度低下が少なく、長時間でも高い精度で探索できることがわかる。

最後に、処理時間の測定を行った。ここで使用した CPU は Intel 社 Xeon (3.2 GHz) である。表 5.6 に結果をまとめる。なお、ここに示す時間は、メモリに読み込んだ後から探索を完了するまでの時間を計測し、全探索について平均したものである。ちなみに、探索時間は、目的信号の長さ、蓄積信号の長さにはほぼ比例する。より長い蓄積信号と目的信号を利用する時はさらに時間がかかるので、注意が必要である。

5.4 合成信号による手法の特性の評価

次に、本手法の頑健性を定量的に評価する。ここでは、2 つの信号をさまざまな混合比で合成して複数の SN 比の音響信号を作成し、本手法の特性を調べた。ここで用いた音響信号の特徴は、3.4.2 節の手法で求めた局所正規化特徴から特徴選択を行ったものである。選択された特徴を量子化し、その量子化値の一致度により類似度を計算した。量子化は 2 値、特徴の選択密度は、毎秒 261 個を局所正規化値の絶対値の大きいものから選択した。図 5.8 に SN 比と探索精度の関係を示す。

なお，ここで用いた目的信号は，20 分の蓄積信号に含まれる楽曲（SN 比では信号として計算）と 20 分の蓄積信号に含まれない楽曲（SN 比ではノイズとして計算）とをミックスしたのから，15 秒の音響信号の断片を 200 箇所切り出して使用した．また，蓄積信号は，13.9 時間分の音響信号を用いた．なお，ここで示した探索精度は，適合率が 99% 時の再現率である．図の示すとおり SN 比がマイナスであっても高い精度で楽曲を特定していることがわかる．

次に， -10dB の目的信号について特徴の選択密度を変化させて探索速度と探索精度を測定した．図 5.9 に探索速度，図 5.10 に探索精度を示す．図のように，選択密度を高くすれば高くするほど探索速度が低下していることがわかる．これは，密度を高くすることによって，一致度を計算するコストが高くなるためである．一方，探索精度は，図のようにピークがあることがわかる．これは，選択密度が高くなると，同一の符号を持つ特徴が増えるために，誤認識が発生しやすくなるためと考えられる．スペクトル全体を投票に用いる方法として，ハフ変換と類似した手法を用いる自己最適化スペクトル相関法 [56] があるが，この手法は，全ての小領域について投票するので，本実験における高密度な場合に相当すると考えられる．実験が示すように，類似度の計算に用いる領域を選択し絞り込むことで，高い精度の検出が可能となるので，適切な選択密度を設定する必要があることがわかる．

5.5 考察

本章で扱った特徴の選択手法は，信号中から頑健な特徴を選択するものであり，探索に有利な部分のみを利用するために高い精度が得られる手法であった．特徴の選択数は，速度や精度に影響を与えるが，これらは単純なトレードオフ関係ではなく，最適な精度を与える特徴の選択数というものが存在した．特に確率的に出現頻度の少ない箇所を中心に選び出されるため，誤って同一箇所が選ばれにくく，探索精度の向上に大きく貢献していると考えられる．また，距離尺度としてハミング距離を用いることで安定した距離値を獲得できた．ハミング距離を利用すると，遮蔽による距離値への影響が遮蔽部分の画素値に依存せず，距離値が単純に遮蔽物体の面積の値となるため，目的信号と蓄積信号が同一パターンであること

が前提の探索の課題には適した距離尺度であると考えられる。しかしながら、本手法は、幾何変換のように特徴の座標そのものが変わるような変換に対して対処できるものではない。実際に幾何変換が行われた信号に対処するためには、選択された特徴を何らかの方法で幾何変換する必要があると考えられる。

5.6 まとめ

本章では、変動に頑健な特徴を抽出するという観点から特徴の選択手法について検討した。本手法は、ディスプレイやカメラの機器特性や照明による変動などの影響の少ない特徴を抽出するために時空間上の局所領域ごとの正規化を行った後、統計量に基づいてスパースな特徴選択を行い、遮蔽や欠落に対応するために特徴量の量子化値の一致度に基づいて類似箇所を検出を行うというものである。実験を行ったところ、これらの効果により探索精度が向上することが確かめられた。例えば、本手法を用いて、実環境でカメラで収録した 8 秒の映像信号の断片が 60 分の蓄積信号のどの時刻に対応するかを調べる課題について実験を行ったところ、単純な縮小画像の照合を用いた場合に 8.31% であった探索精度が 93.72% にまで改善し、本手法の有効性が確認された。

本章で扱った特徴の選択手法は、信号中から頑健な特徴を選択するものであり、探索に有利な部分のみを利用するために高い精度が得られる手法であった。特に確率的に出現頻度の少ない箇所を中心に選り出されるため、誤って同一箇所が選ばれにくく、探索精度の向上に大きく貢献していると考えられる。また、距離尺度としてハミング距離を用いることで安定した距離値を獲得できた。ハミング距離を利用すると、遮蔽による距離値への影響が遮蔽部分の画素値に依存せず、距離値が単純に遮蔽物体の面積の値となるため、目的信号と蓄積信号が同一パターンであることが前提の探索の課題には適した距離尺度であると考えられる。しかしながら、本手法は、幾何変換のように特徴の座標そのものが変わるような変換に対して対処できるものではない。実際に幾何変換が行われた信号に対処するためには、選択された特徴を何らかの方法で幾何変換する必要があると考えられる。そこで、次章で、第 3 章や第 4 章のアプローチと組み合わせることで、より柔軟な探索を実現する。

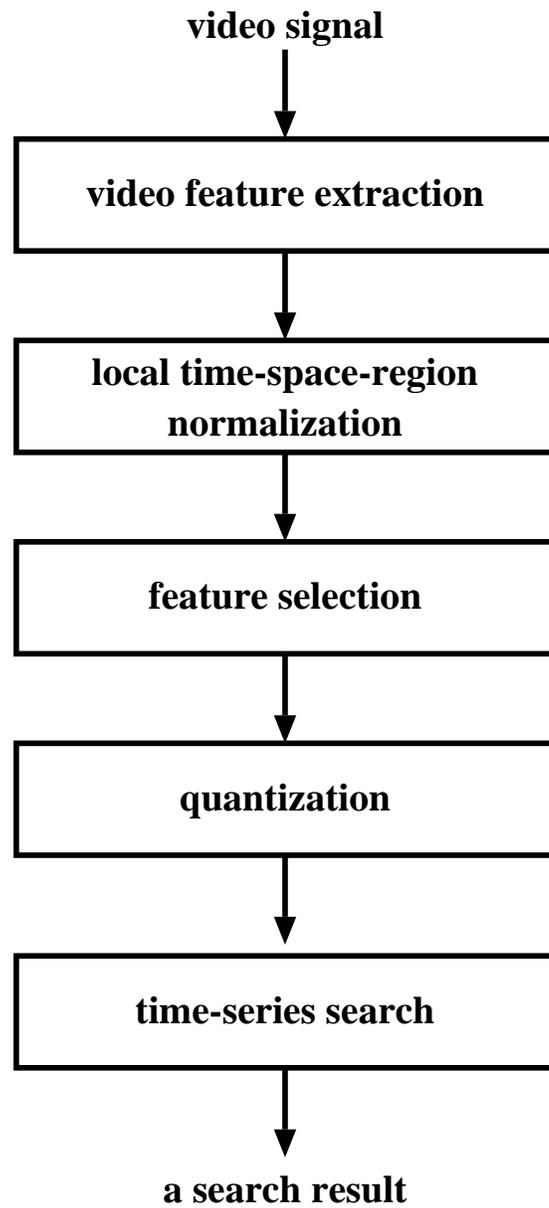


図 5.2: 処理の手順

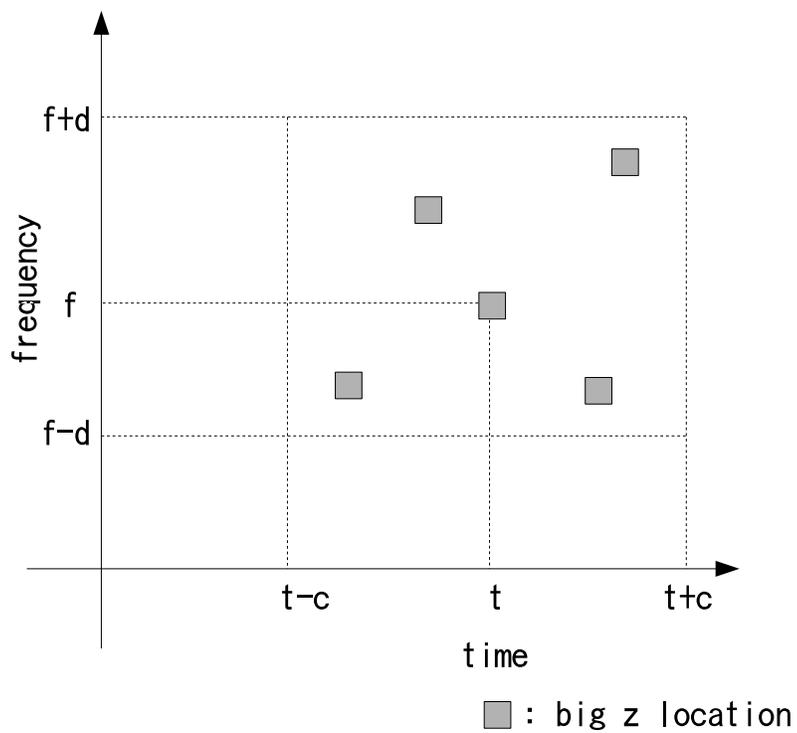


図 5.3: スパースな特徴選択の例

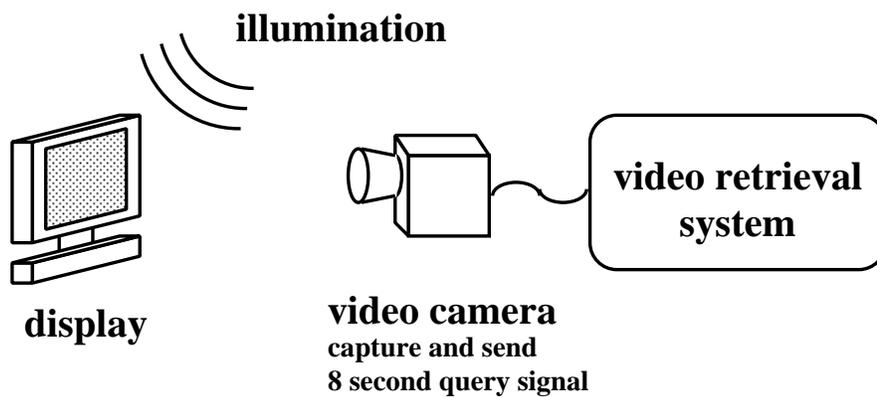


図 5.4: 実験装置

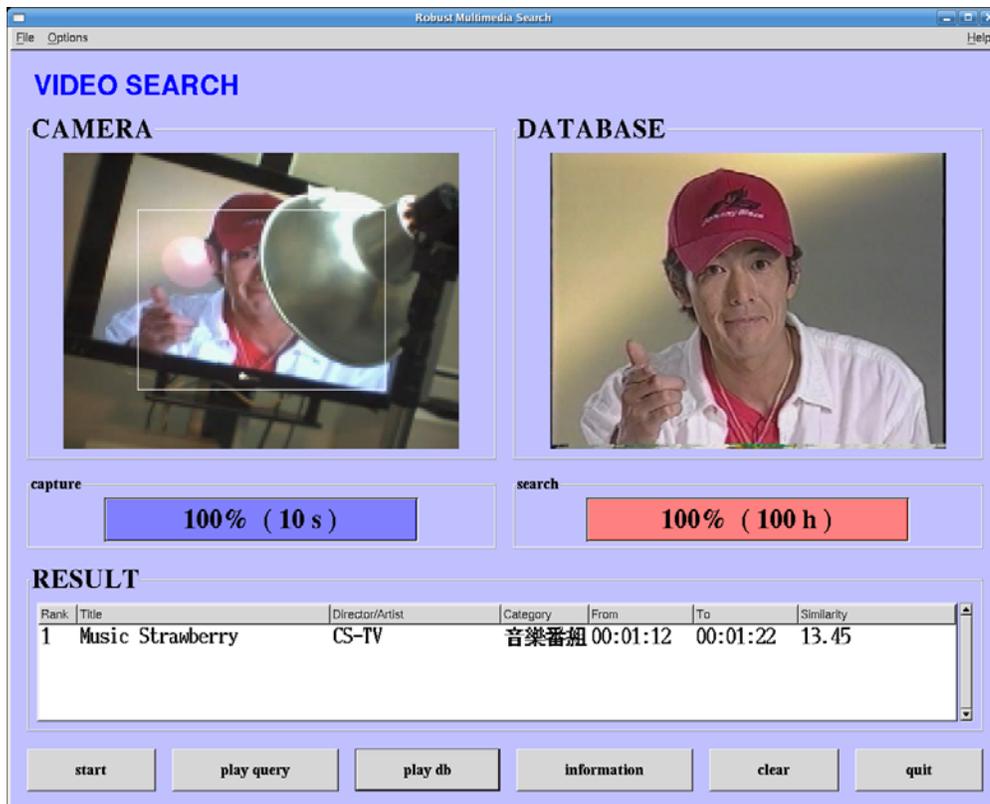


図 5.5: 映像探索の例

an original image



occluded images

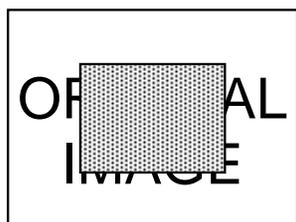
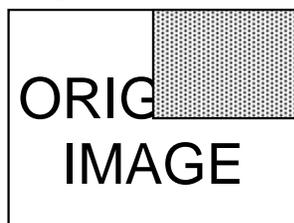
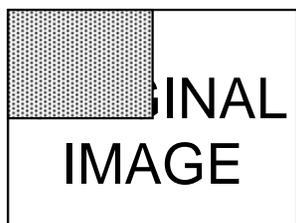


図 5.6: 実験で使用した遮蔽物体の配置

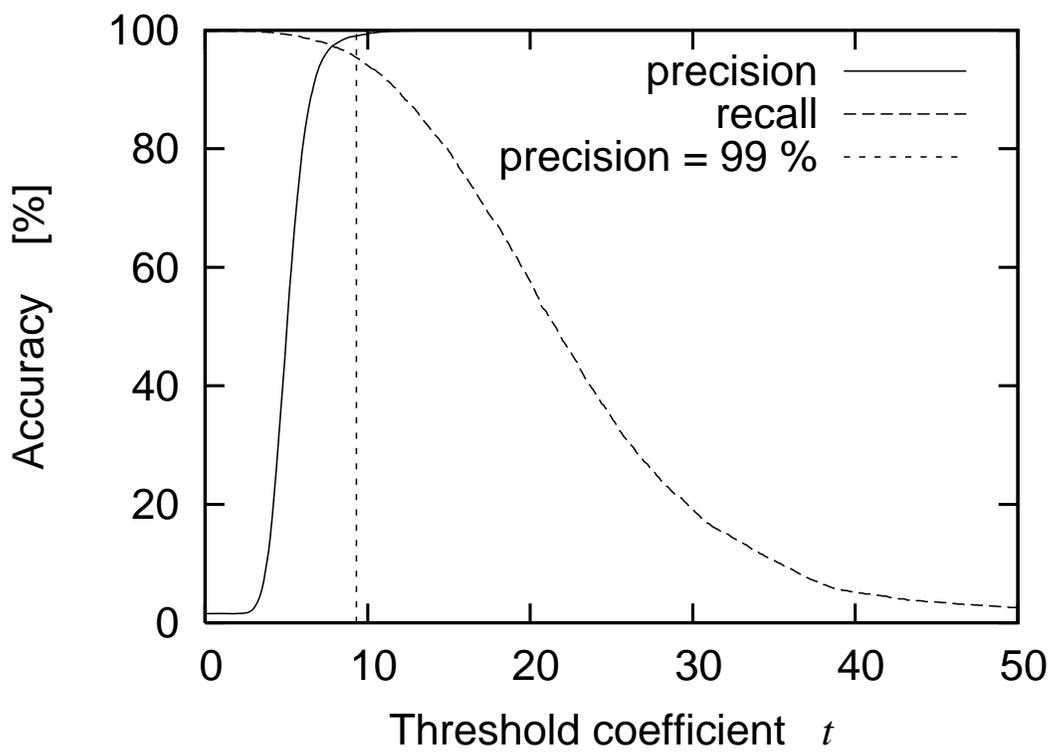


図 5.7: 適合率と再現率

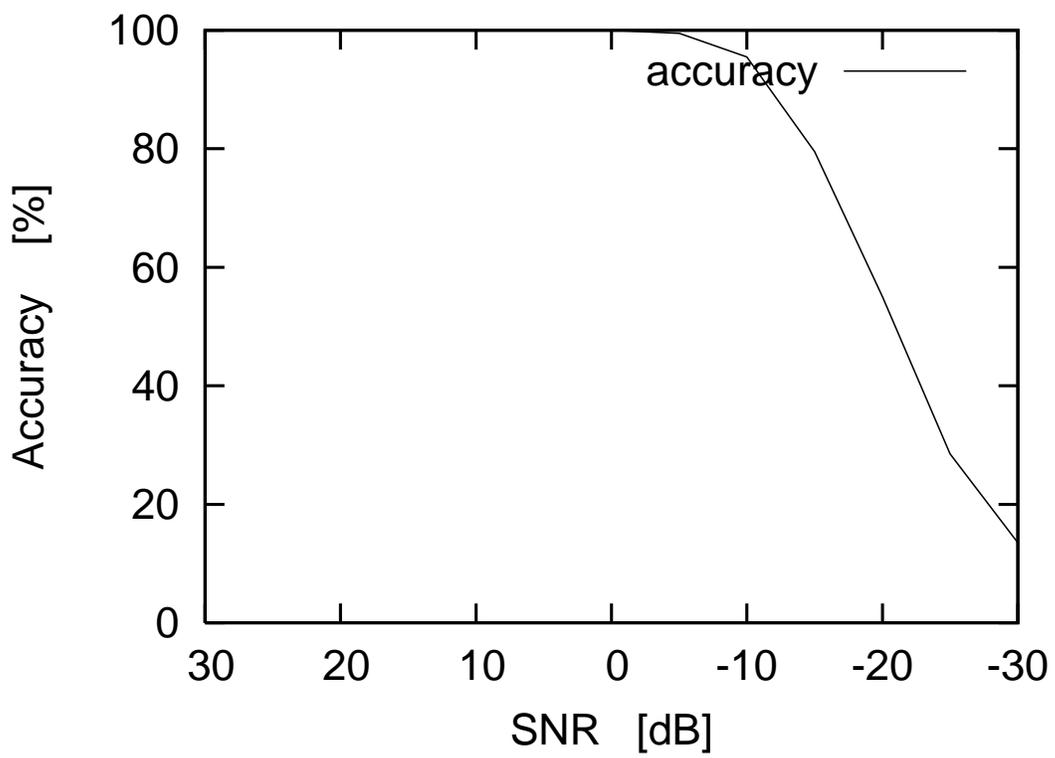


図 5.8: 雑音への耐性

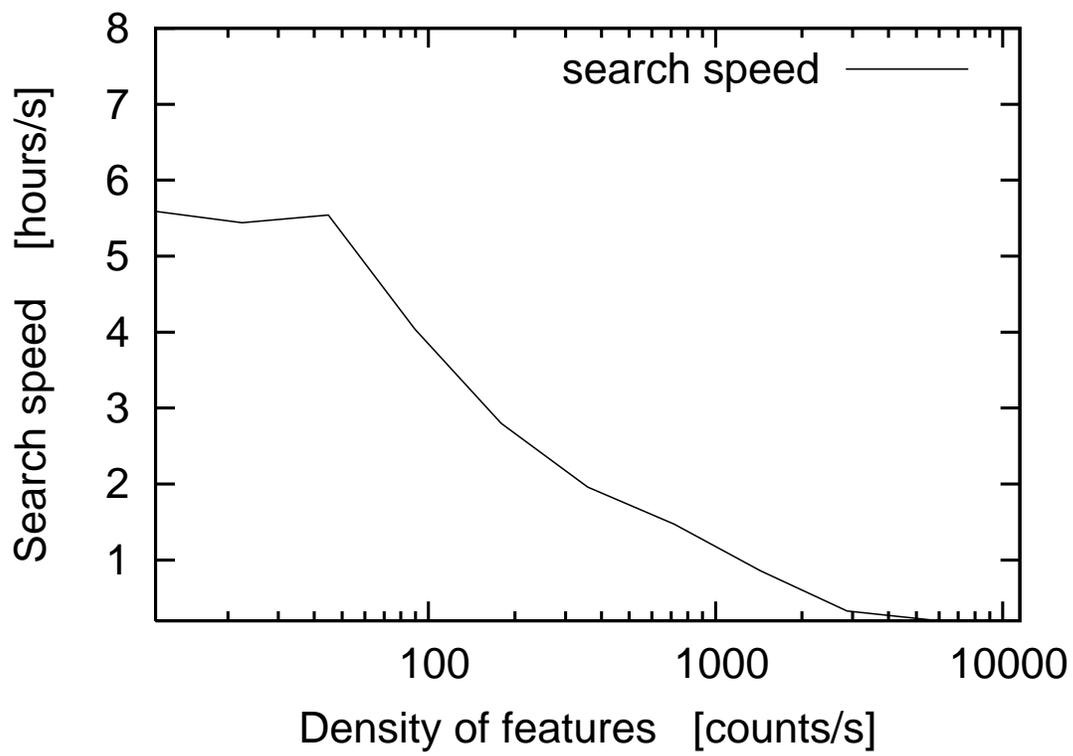


図 5.9: 選択数と速度の関係

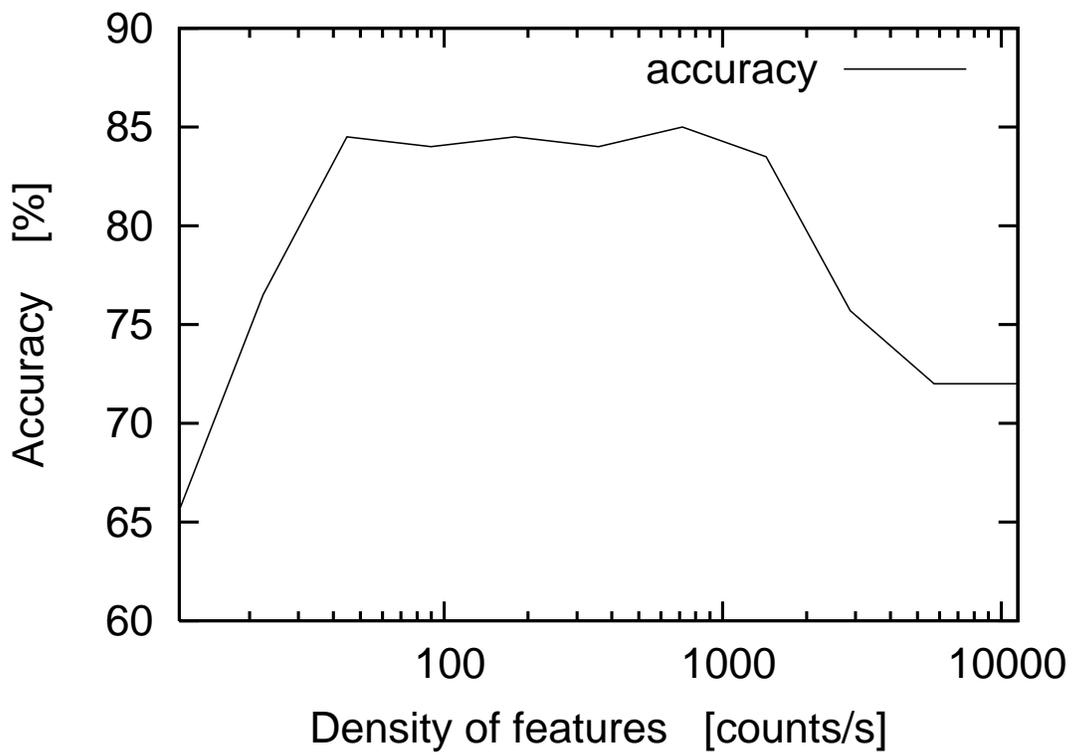


図 5.10: 選択数と精度の関係

第 6 章

縮退生成探索

6.1 はじめに

本章では、幾何変換に対処できる探索法という観点から、縮退生成探索について述べる。縮退生成探索は、これまで述べてきた第 3 章から第 5 章までの全てのアプローチを組み合わせたものである。基本的なアイデアは、幾何変換に不変な縮退特徴を求めておき、目的信号中の縮退特徴と同一の縮退特徴を蓄積信号中から探索し、変動パラメータを求め、得られた変動パラメータに基づいて特徴を生成し照合するものである。以下、提案法の概要と、その妥当性を検討するための実験について順に説明する。なお、本章で議論する提案法は、音にも映像にも適用できる手法と考えられるが、現在、実験での検討が音のみであるため、本章では音を中心に検討結果を述べる。

6.2 問題と解決の方針

本章では、幾何学的な変動のある信号を高速に探索する手法について検討する。ここでは、簡単のため音響信号の場合を取り扱う。例えば、図 6.1 のように、蓄積信号から特徴的な箇所を抽出し、それらが時間周波数空間上の座標と量子化された値で表現されているとする。ここで仮に、図 6.1 の蓄積信号を時間伸縮することによって得られた信号(図 6.2)が目的信号であるとする。これら目的信号と蓄積信号が同一のものとみなして検出されるには、従来、既知の特徴ひずみを事前に

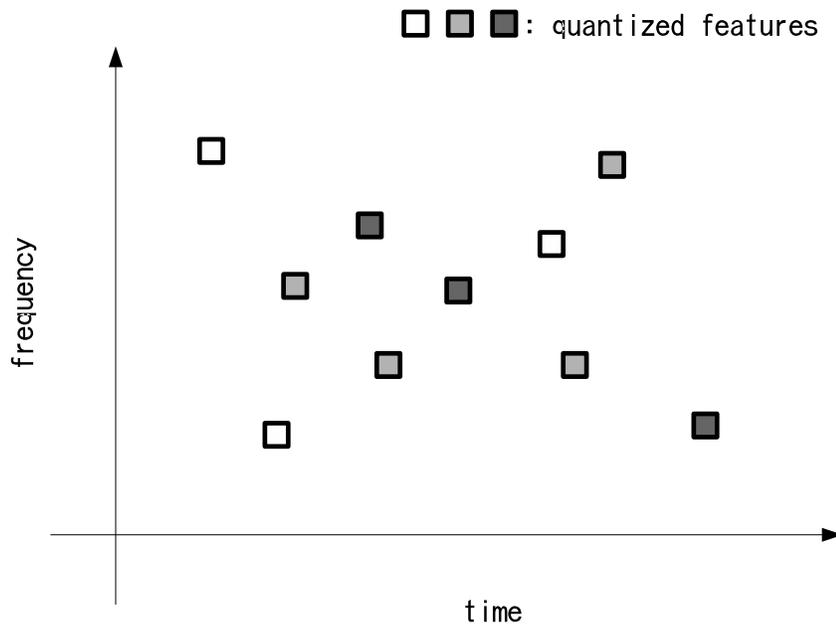


図 6.1: 蓄積信号から抽出された特徴

学習により求めておき，それを用いて原信号にひずみを付加しながら探索する手法 [40]，ハッシュ探索に基づく手法 [10, 30]，ハッシュ探索とハフ変換を組み合わせる手法 [11] などがあった．しかしながら，これらの手法は，幾何学的な変動にも耐えることができる手法であるが，ハフ変換 [25] で投票するためメモリ空間が膨大に必要ななど，たくさんの幾何変換パラメータを考慮すると時間がかかるという問題あった．また，幾何変換に対して不変な特徴を用いて探索することも考えられるが，分解能が低下し誤認識が増えるという問題があり，高精度に探索することは困難である．

そこで，本章では，複数の特徴の間の相対関係を記述した縮退特徴を導入し，効率的に探索する．基本的なアイデアは，次のとおりである．まず，蓄積信号の特徴の中から 2 つの特徴を選び出し，それらの相対関係を縮退特徴とし蓄積しておく（図 6.3）．ここで利用する相対関係は，特徴の量子化値や特徴間の周波数の比など，幾何学的な変動があっても不変な量である．この縮退特徴は事前に蓄積信号の全域にわたって抽出しておく．目的信号が入力されると，同様に縮退特徴を計算する（図 6.4）．そして，蓄積信号の中から目的信号と同一の縮退特徴を選び出す．次に，蓄積信号と目的信号との間の幾何学的変動パラメータを求める．こ

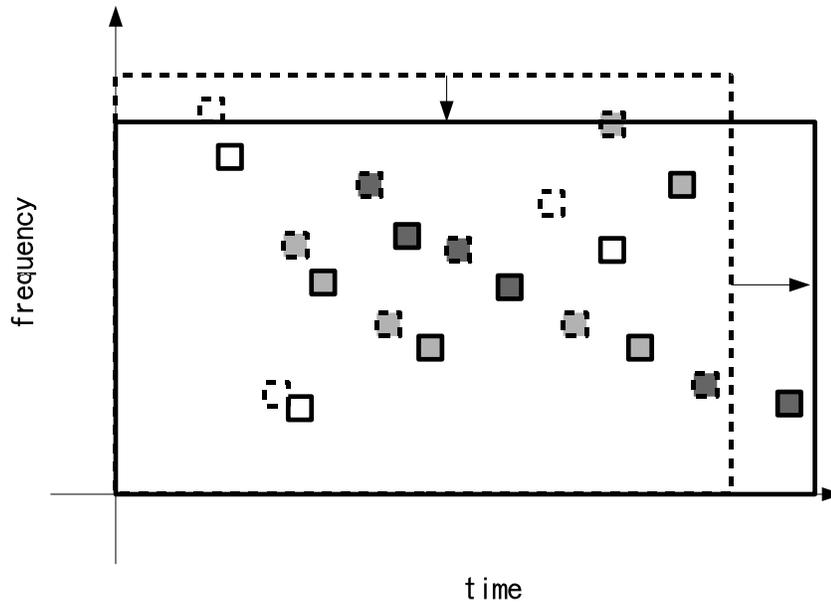


図 6.2: 変形している目的信号から抽出された特徴

ここで求める変動パラメータは，周波数の比，時間伸縮の比など，目的信号と蓄積信号との間の相対関係を表すパラメータである．次に，蓄積信号を変動パラメータに基づいて目的信号に合わせて幾何変換する(図 6.5，図 6.6)．そして，幾何変換された蓄積信号と目的信号を照合して類似度を得る．

6.3 探索の原理

本手法の基本的アイデアは，信号から頑健な特徴を抽出し，複数の特徴の間の相対的な関係を記述した縮退特徴を求め，同一の縮退特徴を持つ目的信号と蓄積信号の箇所を特定し，目的信号と蓄積信号の間の幾何学的変動パラメータを求め，その箇所を中心に幾何学的変動パラメータに基づいて蓄積信号を幾何変換し目的信号と照合するものである．

探索の具体的な処理は，(1) 基本特徴の抽出，(2) 縮退特徴の抽出，(3) 変動パラメータの抽出，(4) 蓄積信号の幾何変換，(5) 類似度の計算の順で行う(図 6.7)．

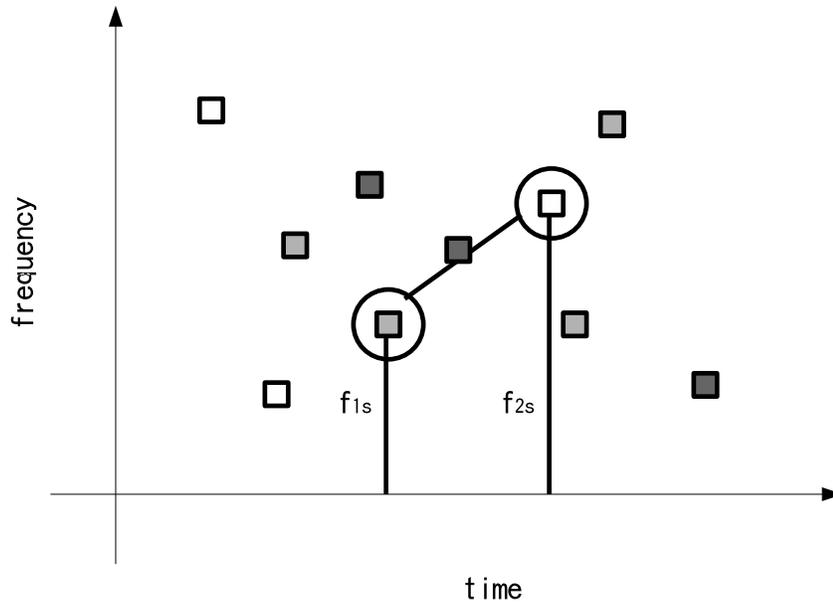


図 6.3: 蓄積信号から抽出された縮退特徴

6.3.1 基本特徴の抽出

まず，探索に使用する基本特徴を抽出する．基本特徴は，音響信号から以下の手順で抽出する．

音響特徴の抽出

音響特徴としては，単位時間あたりのゼロ交差数，短時間パワースペクトル，LPC ケプストラム，MFCC(Mel frequency cepstral coefficients) など [38] が考えられる．本章では，その中でも代表的な周波数特徴である，フーリエ変換により求めた短時間パワースペクトルを用いた．ここで，時刻 t の音響信号のフーリエ変換後の特徴 $X(t)$ の f 番目の要素を，

$$X(t, f) = \sum_{n=0}^{N-1} x(t+n) e^{-j \frac{2\pi f n}{N}} \quad (6.1)$$

とする．ただし， $x(t)$ を時刻 t における信号の値， N は周波数特徴の分析窓の長さ， f の最大値は標本化定理より $N/2$ である．ここで，パワースペクトル $P(t, f)$

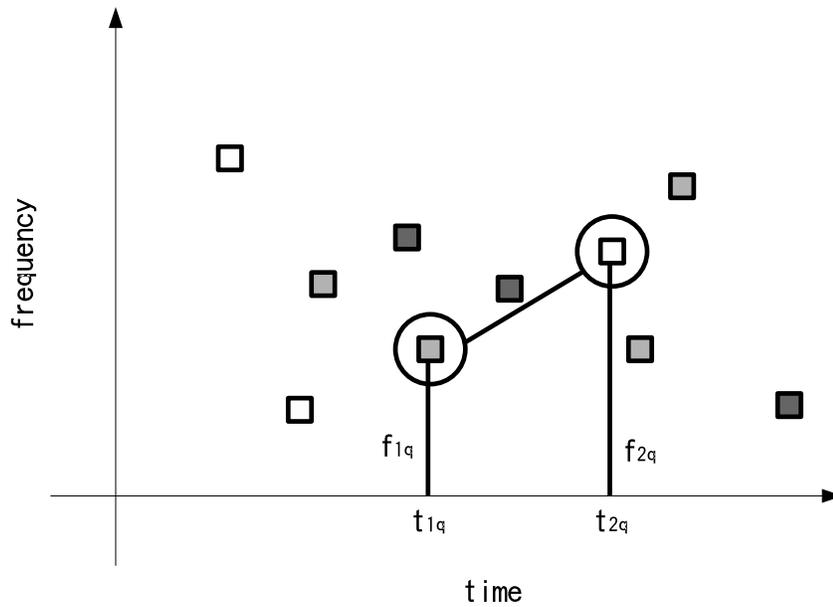


図 6.4: 目的信号から抽出された縮退特徴

を次式により求める．

$$P(t, f) = |X(t, f)|^2 \quad (6.2)$$

ここで，このパワースペクトルを s きざみで抽出する．すなわち，時間方向 t 番目の周波数特徴 $Q(t, f)$ は，

$$Q(t, f) = P(st, f) \quad (6.3)$$

である．

時間周波数空間上の局所領域ごとの正規化

この処理の目的は，数秒程度の短時間で小さい変動を吸収することである [14]．ここでは，各周波数領域ごとに，ある時間区間の値から平均と標準偏差を求め，

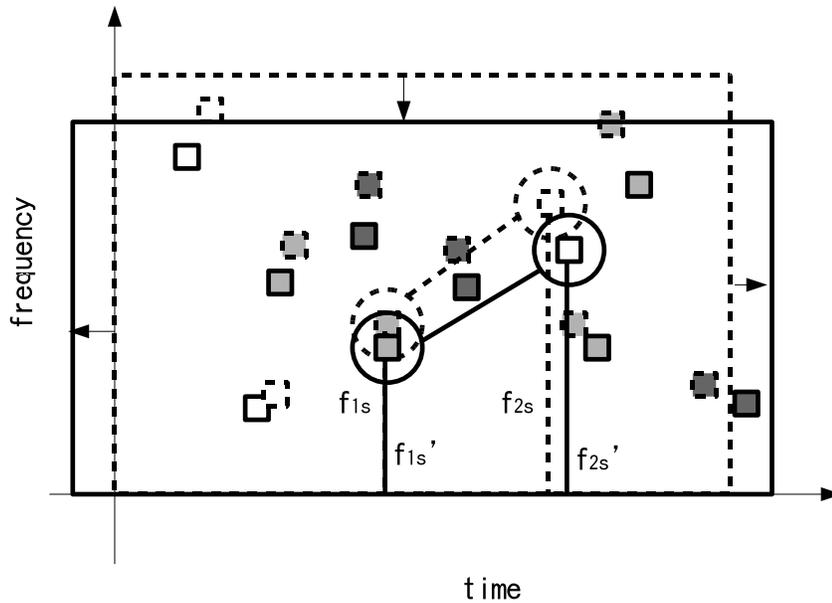


図 6.5: 蓄積信号をスケールリング

それらを用いて正規化を行う．すなわち，正規化後の音響特徴 $y(t)$ の f 番目の要素は，

$$y(t, f) = \frac{1}{\sigma(t, f)}(x(t, f) - m(t, f)), \quad (6.4)$$

である．ただし， $x(t, f)$ は， $x(f)$ の t 番目の要素，

$$m(t, f) = \frac{1}{M} \sum_{i=-\lfloor M/2 \rfloor}^{M-\lfloor M/2 \rfloor-1} x(t+i, f),$$

$$\sigma(t, f)^2 = \frac{1}{M} \sum_{i=-\lfloor M/2 \rfloor}^{M-\lfloor M/2 \rfloor-1} (x(t+i, f) - m(t+i, f))^2,$$

M は平均と標準偏差を求めるための窓の大きさ， $\lfloor \cdot \rfloor$ は切下げを表す．

スパースな特徴選択

この処理の目的は，ひずみや雑音などによる特徴の変動の影響を受けにくい箇所を選択し，頑健な特徴を得ることにある．ここでは，特徴変動の影響を受けにくい箇所を統計量に基づいて選択する．特徴的な箇所をある時間区間の時間平均

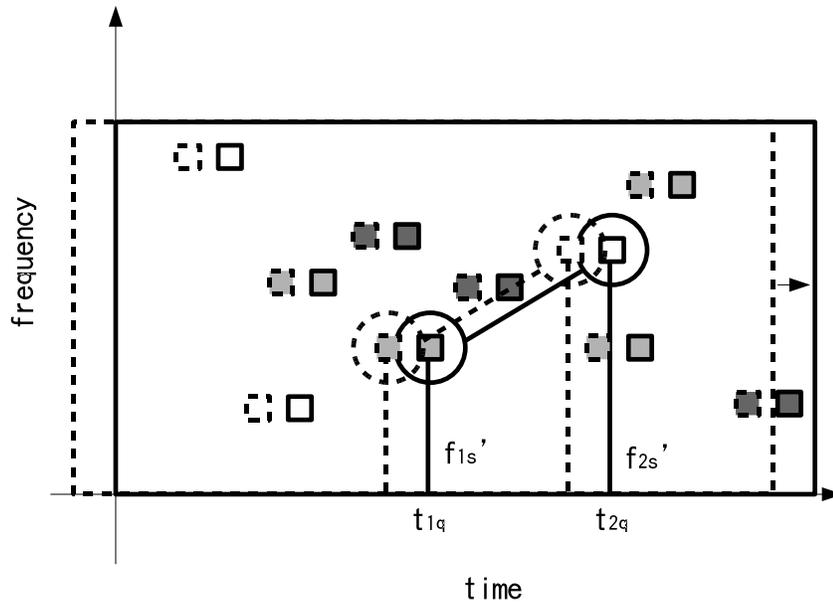


図 6.6: 蓄積信号をシフト

に対して大きく変動している箇所と仮定し，

$$z(t, f) = |x(t, f) - m(t, f)| \quad (6.5)$$

で得られる平均に対する変動の絶対値の大きい箇所を選び出す．選択は，時刻が $t - c$ から $t + c$ ，周波数が $f - d$ から $f + d$ の範囲で上位 N 個を取り出し，注目箇所が含まれていれば，その箇所を特徴として採用する．なお，選択の様子を図 6.8 に示す．

局所正規化値の量子化

続いて，選択された特徴の量子化を行う．量子化は，局所正規化値に基づいて行う．量子化の方法は非線形量子化やベクトク量子化などが考えられるが，ここでは，単純に線形スカラー量子化を行う．

具体的には，式 (6.4) により得られる局所正規化値を次式によって等間隔に L レ

ベルに量子化する．

$$y'(t, f) = \begin{cases} \lfloor \frac{L \times (y(t, f) + l)}{2l} \rfloor & \text{if } |y(t, f)| < l \\ L - 1 & \text{else if } y(t, f) \geq l \\ 0 & \text{otherwise} \end{cases} \quad (6.6)$$

ただし， l は量子化値を決めるための値， $\lfloor \cdot \rfloor$ は切下げを表す．そして，最終的に基本特徴は，選択された特徴の量子化値，時刻，周波数からなる特徴量として表現される．

6.3.2 縮退特徴の抽出

次に，縮退特徴を抽出する．縮退特徴は，上記の基本特徴を複数利用して導く．例えば，第一の基本特徴 $y'(t_1, f_1)$ と第二の基本特徴 $y'(t_2, f_2)$ から計算する場合，

$$D(t_1, f_1, t_2, f_2) = \frac{f_2}{f_1} \quad (6.7)$$

で計算される周波数の比を縮退特徴として利用する．ただし， $t_1 \leq t_2$ とし，この縮退特徴の対応する基本特徴の箇所も対応付けて蓄積しておく．また，縮退特徴に利用する基本特徴は，抽出された全ての基本特徴を第一の基本特徴とし，その第一の基本特徴に対して，時間周波数領域のある領域内に存在する全ての基本特徴を第二の基本特徴として利用する（図 6.9）．

縮退特徴は，式 (6.7) をさらに量子化した値に第一の基本特徴の量子化値と第二の基本特徴の量子化値を組み合わせた値を用いる．具体的には，縮退特徴として，

$$D(t_1, f_1, t_2, f_2) = (\lfloor R \times D(t_1, f_1, t_2, f_2) \rfloor, y'(t_1, f_1), y'(t_2, f_2)) \quad (6.8)$$

という整数値を要素に持つベクトルを利用する．ただし， R は量子化レベルを決めるための値である．

6.3.3 変動パラメータの計算

次に，変動パラメータを計算する．まず，目的信号の縮退特徴と同一の式 (6.8) を持つ蓄積信号の縮退特徴とそれと対応する基本特徴を選び出す．次に，目的信

号の縮退特徴の第一の基本特徴の周波数を f_{1q} , 選び出された目的信号の縮退特徴と同一の蓄積信号の縮退特徴の第一の周波数を f_{1s} , 目的信号の縮退特徴の第一の基本特徴と第二の基本特徴の時間差分を $t_{2q} - t_{1q}$, 蓄積信号の縮退特徴の第一の基本特徴と第二の基本特徴の時間差分を $t_{2s} - t_{1s}$ とすると , 蓄積信号に対する目的信号の周波数の比は ,

$$r_f = \frac{f_{1q}}{f_{1s}} \quad (6.9)$$

蓄積信号に対する目的信号の時間伸縮の比は ,

$$r_t = \frac{t_{2q} - t_{1q}}{t_{2s} - t_{1s}} \quad (6.10)$$

となる . これら式 (6.9) と式 (6.10) が , 蓄積信号から目的信号へ変動したと仮定した場合の変動パラメータである .

6.3.4 蓄積信号の幾何変換

次に , 得られた変動パラメータに基づいて , 蓄積信号を目的信号に合わせて変形する . 蓄積縮退特徴の第一の基本特徴の該当箇所を t_{1s}, f_{1s} とすると , 蓄積信号の縮退特徴の周辺の基本特徴の箇所 t_i, f_i を変換して得られる特徴の箇所 t'_i, f'_i を得る . すなわち ,

$$t'_i = t_{1q} + r_t \times (t_i - t_{1s}), \quad (6.11)$$

$$f'_i = r_f \times f_i \quad (6.12)$$

を計算する . ただし , i は , 蓄積信号中で選択された基本特徴の番号を表す . この変換により蓄積信号の基本特徴を目的信号に合わせて変動パラメータを利用して変換する .

6.3.5 類似度の計算

次に , 類似度を計算する . ここでは , 式 (6.6) で得られる量子化値の一致度を用いる . すなわち , 蓄積信号の基本特徴 $y(t_i, f_i)$ を変換することで得られる量子化

値 $y'(t'_i, f'_i)$ が目的信号の基本特徴の時刻 t'_i , 周波数 f'_i に同一の量子化値で存在するかどうかを判定し , 全ての i において一致する特徴の数をカウントする . 得られたカウント数を類似度とし , その値があらかじめ定められた閾値 θ よりも大きい場合 , 探索結果として出力する .

6.4 実験

本手法の有効性について検討するため , 音響信号を伸縮させたものを目的信号として , 蓄積信号から同一部分を探索し , 探索精度を評価する実験を行った .

6.4.1 提案法による効果

提案法の効果を確認するため , 137 曲の楽曲の音響信号 (8.3 時間) を蓄積信号とし , 蓄積信号に含まれる同一の音楽を加工して得られた信号から無作為に選択した 100 個の目的信号 (15 秒間) に対して探索を行い , 精度を測定した . 蓄積信号としては , 高品質な音響信号 . 目的信号としては , 蓄積信号と同一のオリジナル , ピッチ不変時間変動したもの , 時間不変ピッチ変動したもの , 時間伸縮したものを用いた . なお , 音響信号は 8000 Hz で標本化し , 特徴抽出のパラメータは $s = 128$, $N = 4096$, $M = 15$, $L = 2$ とした .

探索結果は類似度を閾値処理することにより求めるが , 閾値は次式によって定めた .

$$\theta = m + h\nu. \quad (6.13)$$

ここで , m と ν 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 , h は経験的に与えられる係数である . 本実験では , 式 (6.13) における h の値を全探索について固定とし , 全 100 回の探索における適合率 (precision rate) の平均値が 99% となるときの再現率 (recall rate) の平均を探索精度として示している . すなわち , 探索により出力される探索結果のほぼ全てが正解となる . なお , 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

表 6.1: 提案法の探索精度

	縮退生成なし	縮退生成あり
オリジナル	100 %	100 %
ピッチ不変時間変動 +10 %	12 %	71 %
ピッチ不変時間変動 +5 %	50 %	88 %
ピッチ不変時間変動 -5 %	52 %	91 %
ピッチ不変時間変動 -10 %	27 %	73 %
時間不変ピッチ変動 +2 半音	0 %	70 %
時間不変ピッチ変動 +1 半音	0 %	75 %
時間不変ピッチ変動 -1 半音	0 %	73 %
時間不変ピッチ変動 -2 半音	0 %	58 %
時間伸縮 +2 半音	0 %	89 %
時間伸縮 +1 半音	0 %	89 %
時間伸縮 -1 半音	0 %	81 %
時間伸縮 -2 半音	0 %	68 %
変動あり平均	11.8 %	77.2 %

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。なお、ここでは、探索した時刻が正しい時刻の ± 7.5 秒以内となった場合に正解とした。

次に、提案法の探索精度を表 6.1 にまとめる。ここでは、処理の効果を比較するために、縮退生成を行わなかった場合の探索精度も示す。

縮退生成を行わなかった場合と比べて、縮退生成を行った場合、探索精度の向上が確認された。これにより、縮退生成を行って照合を行うことが有効であることが確認された。

次に、本実験における探索の処理時間を測定した。137 曲の楽曲（8.3 時間）を探索するのに、1.32 秒の処理時間がかかった。膨大なデータベースを使用する場合は、現実的でない探索速度となるため注意が必要である。

表 6.2: 処理時間

探索 1 回の平均処理時間	探索速度
1.32 秒	6.28 hours/s

6.5 考察

本手法は、第 3 章から第 5 章までの全てのアプローチを組み合わせたものである。それぞれの弱点を補い効率的に探索することができ、大きな変動があるために探索が困難であった幾何変換を探索の対象とすることを可能とした。しかしながら、ここでは音の場合のみ扱ったが、映像の場合のように変動のパラメータがさらに多くなると、さらに処理が複雑になり実装が困難になることが予想される。また、変動パラメータに基づいて生成特徴を計算するときの誤差にも注意を払わなければならない。特に、距離の近い特徴間で計算された変動パラメータを用いて距離の遠い特徴の生成に利用するときは、誤差を含むため誤差を考慮した探索が必要となる。

また、探索速度は、縮退生成を行わない 5.4 節の探索速度と比べて大幅に低下している。これは、探索空間が時刻に加えて幾何変換パラメータが導入されたことと、縮退特徴の表現のダイナミックレンジが小さいために、探索の枝刈りの効果が小さいためであると考えられる。本実験では、特徴の組から計算される不変量を 4096 に量子化して使用したが、同一の不変量を持つ特徴が多く存在したため、探索コストがかかってしまったと考えられる。

6.6 まとめ

本章では、目的信号と蓄積信号の間の幾何学的変動を考慮して探索する手法を提案した。本手法は、信号から頑健な特徴を抽出し、複数の特徴間の相対的な関係を記述した縮退特徴を求め、同一の縮退特徴を持つ目的信号と蓄積信号の箇所を特定し、目的信号と蓄積信号の間の幾何学的変動パラメータを求め、その変動パラメータに基づいて蓄積信号を幾何変換し目的信号と照合するものである。本

手法を用いて、音響信号を伸縮させた音の断片を 8.3 時間の蓄積信号のどの時刻に対応するかを調べる課題について実験を行ったところ、幾何変換を行わなかった場合に 11.8% であった探索精度が 77.2% にまで改善し、本手法の有効性が確認された。

本手法は、第 3 章から第 5 章までの全てのアプローチを組み合わせたものである。それぞれの弱点を補い効率的に探索することができ、大きな変動があるために探索が困難であった幾何変換を探索の対象とすることを可能とした。しかしながら、ここでは音の場合のみ扱ったが、映像の場合のように変動のパラメータがさらに多くなると、さらに処理が複雑になり実装が困難になることが予想される。また、変動パラメータに基づいて生成特徴を計算するときの誤差にも注意を払わなければならない。特に、距離の近い特徴間で計算された変動パラメータを用いて距離の遠い特徴の生成に利用するときは、誤差を含むため誤差を考慮した探索が必要となる。

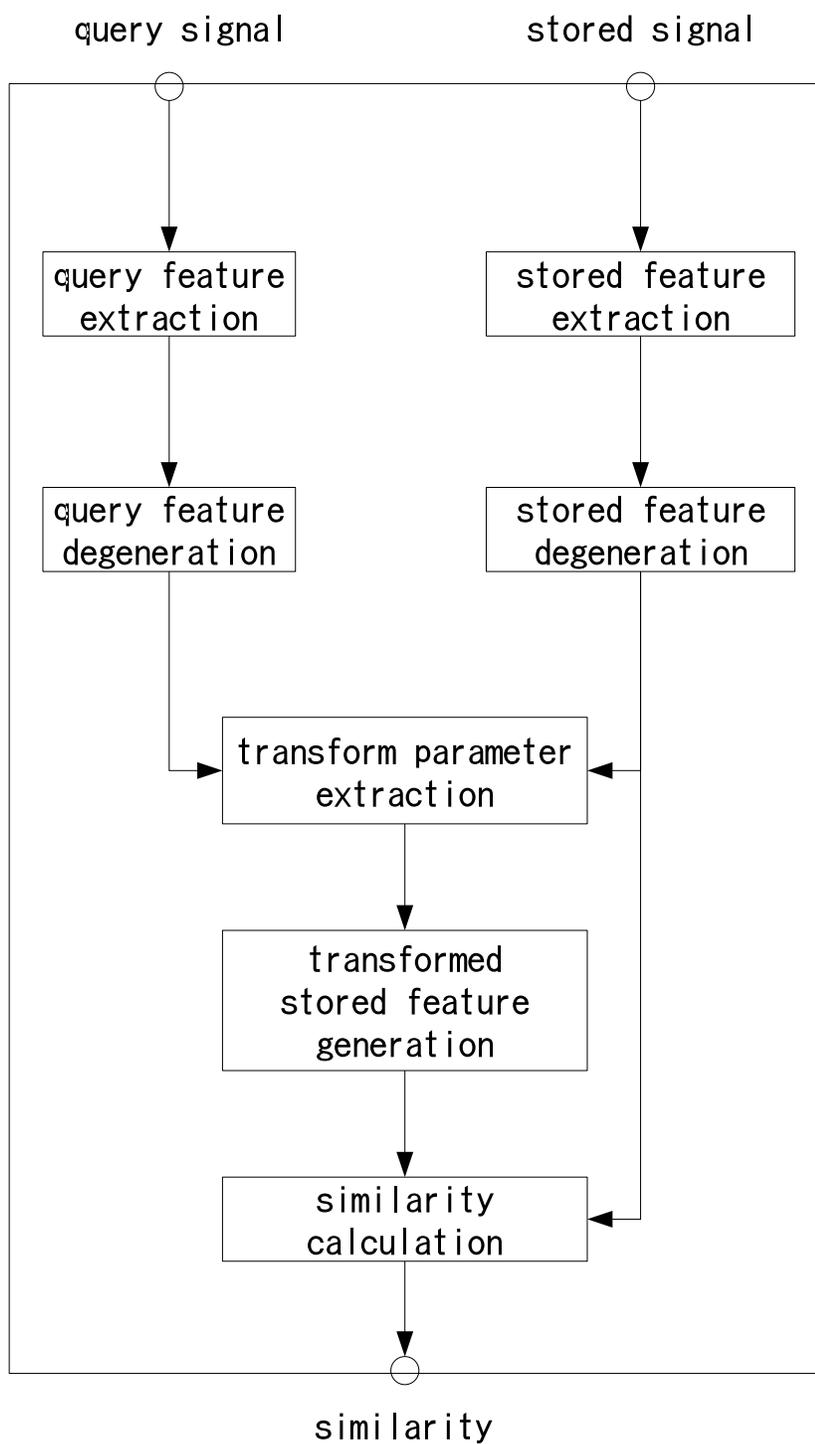


図 6.7: 処理の手順

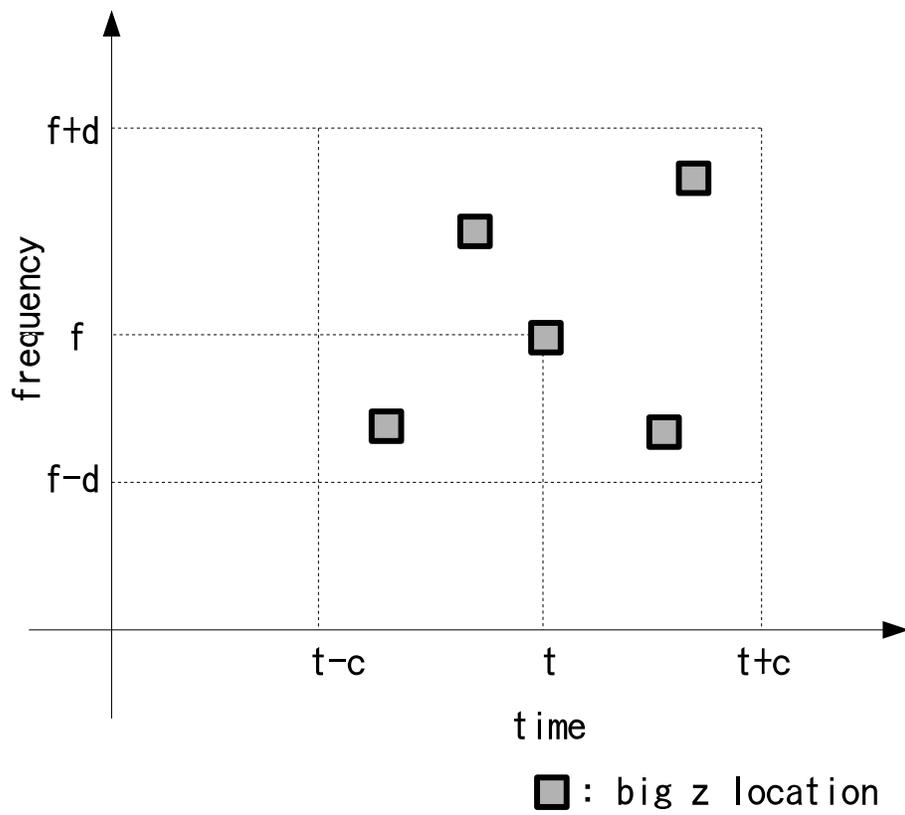


図 6.8: スパースな特徴選択の例

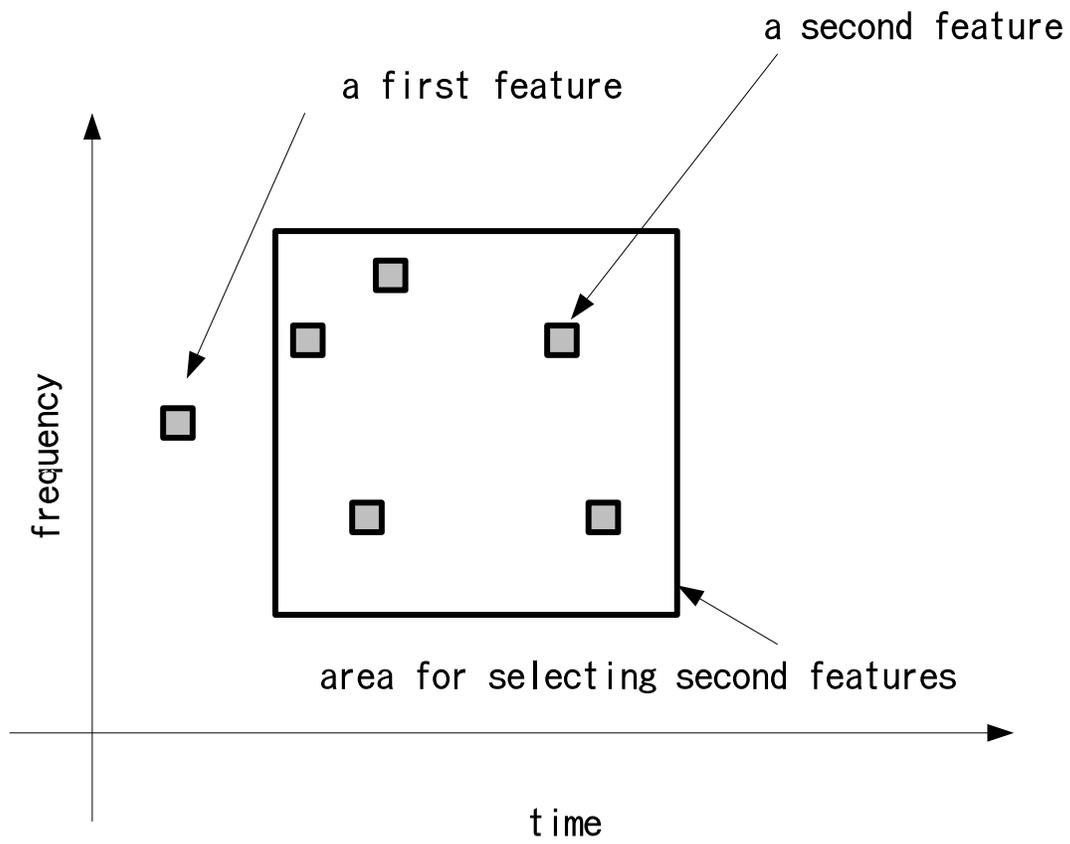


図 6.9: 第二の基本特徴の選択

第 7 章

結論

7.1 本研究の要約

本論文では、音や映像などの膨大なマルチメディア情報の中から、目的とするメディア情報を高速かつ高精度に探索するための特徴の表現方法について述べた。メディア探索を実現するための克服すべき問題は、次の 2 つである。

- 音・映像データの膨大な探索空間を高速に探索しなければならないこと
- さまざまな収録環境によって発生する信号のノイズやひずみ、コンテンツの編集や加工による変動があっても高精度に探索しなければならないこと

また、実環境で収録された音や映像や加工や編集のある音や映像を目的信号とした場合、対処すべき変動には、次のものがあった。

- 機器特性や環境特性などによって発生する乗法性ひずみ
- 雑音や照明の写り込みのような加法性ノイズ
- 携帯電話の電波状況によって発生する信号の欠落や目的の映像の上に他のオブジェクトによって覆われることで発生する遮蔽
- 音や映像の加工による伸縮や拡大縮小・回転やカメラ位置の違いなどで発生する幾何学的な変動

本研究では，これらの変動に対処するために，次の 3 つのアプローチについて検討した．

- 目的信号から変動に対して不変な特徴である縮退特徴を抽出して探索するアプローチ
- 具体的な変動パターンに基づいて変動のある信号を表した特徴である生成特徴を導き探索するアプローチ
- 統計的に特徴的な箇所のみを類似度として使用して照合するアプローチ

縮退特徴を利用するアプローチは，時間周波数空間上または時空間上の局所領域ごとに正規化を行った後，部分空間に射影することで，ひずみや雑音を吸収するものであった．乗法性ひずみと加法性雑音を吸収することができ，実環境で収録した音や映像を使用して探索する課題において有効であることが実験により確認された．縮退特徴は，ある基準を設定しそれに対して相対的な値を利用するものである．変動に不変であるため高速に探索できるが，利用する相対値の分解能が十分でなければ，探索精度が低下してしまう．特に幾何変換のような大きな変動を扱う場合には，あまり大きな変動を縮退させすぎると分解能を低下させてしまうため，特徴の抽出の方法には注意が必要であると考えられる．

生成特徴を利用するアプローチは，ダビングや圧縮によるひずみに対処するために変動のモデルを用意し，それに基づいて特徴を生成し探索するというものであった．ダビングや MPEG による圧縮のある目的信号を用いて探索する実験を行ったところ，従来手法に比べて探索精度が向上し，有効であることが確認された．生成特徴は，得られた特徴を変動モデルに基づいて生成して得られるものである．分解能を維持したまま生成するため高精度であるが，さまざまな変動を考慮すると探索範囲が膨大になるため高速に探索することが困難となる．

スパースな特徴選択を行うアプローチは，統計量に基づいて特徴的な箇所を選択し，その箇所のみを照合するというものであった．実環境で収録した映像や遮蔽のある映像を目的信号として探索する実験を行ったところ，従来手法と比べて探索精度の向上が確認された．これにより遮蔽や欠落のある信号でも探索できることが確認された．スパースな特徴選択は，特徴的な箇所のみを照合に利用する

ものである．頑健かつ高速であるが，幾何学的な変動のように特徴の座標が変わってしまうような場合には，対応策が必要である．

また，縮退特徴，生成特徴，スパースな特徴選択を組み合わせた手法として縮退生成探索を提案した．音の時間伸縮や画像の回転や拡大縮小などの幾何学的な変動に対処するためのものであり，同一の縮退特徴を持つ箇所から変動パラメータを決定し，その変動パラメータに基づいて特徴を生成し，それに基づいて類似度を算出して照合するものであった．時間伸縮やピッチ変動のある信号を用いた探索実験を行ったところ，従来法と比べて探索精度が向上し，有効性が確認された．縮退生成探索は，幾何学的な変動にも対応できる柔軟な探索であると同時に，探索の過程で変動パラメータも特定できるため，変動パラメータの抽出方法としても利用できる．

7.2 今後の課題

縮退生成探索は，加法性雑音，乗法性ひずみ，非定常雑音・遮蔽・欠落，幾何変換に対応できる手法であった．しかしながら，幾何変換に対応することで，探索空間が膨大になったことと，縮退特徴が表現できるダイナミックレンジが小さいために，十分な探索速度を得ることができなかった．高速化のためには特徴から計算される縮退特徴のダイナミックレンジを大きくし，巨大なインデックスを作る必要がある．既に文書画像を高速に探索することを目的に，巨大なインデックスを利用することに着目した，特徴点の局所的配置に基づく手法 [57] など，複数の特徴点の関係をインデックスに利用する手法が提案されている．この手法は本論文のような生成特徴は利用せず，文書画像から計算した不変特徴抽出のみで探索しているが，高速な探索を実現している．本研究における縮退特徴においても，同様に多くの特徴を利用するアプローチをとることで，大きなダイナミックレンジを持つ縮退特徴を導入できる可能性を検討する必要がある．また，本論文では，縮退生成探索については，少数の変動パラメータを考慮する場合についてのみ検討した．静止画や映像探索に応用するとなると，より多くの変動パラメータを扱わなければならない．今後，より柔軟な探索を実現するためには，画像の二次元平面のパラメータや拡大縮小，回転など，より多くの幾何変換パラメータを扱う

表 7.1: 縮退生成探索の特性

加法性雑音	
乗法性ひずみ	
非定常雑音・遮蔽・欠落	
幾何変換	
探索速度	
柔軟性	×

ことのできる縮退生成探索について検討する必要がある。

7.3 今後の展望

さまざまな変動に対処できる高速かつ高精度なメディア探索を実現するという観点から特徴の表現方法に関する基礎的な側面とその応用について今後の展望を述べる。

本論文において、メディアを表現する特徴において、特徴的な特徴のみを利用したスパースな表現が有効であることを示したが、どのような特徴を選択するかについてさらに追求する価値があると考えられる。例えば、粗い量子化、スパースな特徴の表現が高い精度の探索を実現しているため、特徴の選択の方法が重要であることがわかる。特に、出現頻度が低くかつ変動に強い特徴を大きなダイナミックレンジでとらえることができれば、高速性・高精度性がさらに増すと考えられる。

本論文では、乗法性ひずみ、加法性ノイズ、幾何学的変動に対処するために縮退生成探索を提案したが、今後、扱うことのできる変動パラメータとして、より多くの種類を扱いたい。例えば、幾何学的な変動パラメータについては、3次元の座標を表すパラメータや伸縮を表すパラメータ、回転の方向を表すパラメータなどである。これらのパラメータを目的信号から獲得できるようになると、ロボット環境認識に役立つと考えられる。ロボットの周辺にデータベースに登録してあるオブジェクトがあれば、登録されている蓄積信号中のオブジェクトと目的信号中のオブジェクトとの相対関係からロボット自身の座標に関するパラメータを獲得

表 7.2: 拡大する研究領域

統計的頻度に基づく頑健な特徴の選択
単眼カメラによる幾何パラメータの獲得
感性などのパラメータを含む柔軟な表現
ライフログ規模の探索を可能とする超高速メディア探索

することができる。このようなことが実現すれば、複数のカメラを用いなくとも 1 つのカメラで 3 次元情報に関するパラメータを取得することができ、コンピュータビジョン研究における新しい可能性が広がる。また、幾何学的なパラメータ以外も考えられる。例えば、感性に関する情報や人物の年齢や性別、個人性などのさまざまな変動パラメータを扱うことができれば、より柔軟なメディア探索の実現が期待できるであろう。また、データベースの規模も、今後は、より膨大なものになると考えられる。現在のハードディスクの高密度化と低価格化を考えると、大容量記憶装置を携帯し、日常の生活を常に記録するような時代が近い将来に訪れることは想像に難くない。そのような時代になれば、一生分の音や映像から欲しい情報を引き出すようなライフログ規模の超高速・高精度なメディア探索も必要になり、そのようなメディア探索に本研究は大きく貢献するであろう。

本論文で述べたように、縮退特徴、生成特徴、スパースな特徴選択を利用することで、さまざまな変動に対処することができ、さらに、その変動パラメータを獲得できることがわかった。より多くの種類の変動パラメータを扱い、より膨大なデータベースを必要とされるような探索の課題においても、本アプローチは適用できると考えられ、今後、メディア探索技術は、より高速、より高精度、より柔軟な、ユーザの幅広いニーズに応えられるメディア探索技術へと発展していくと考えられる。

謝辞

本研究を行うに当たり，懇切なる御指導，御鞭撻を賜った像情報処理学講座 小谷 一孔 准教授に心より深謝致します．

本研究の遂行にあたり，有益な御助言と御鞭撻を賜った知能情報処理学講座 党 建武 教授，音情報処理学講座 赤木 正人 教授，知能情報処理学講座 徳田 功 准教授に厚く御礼申し上げます．

本研究を行うにあたり，御指導，御鞭撻を賜った北陸先端科学大学 宮原 誠 名誉教授（元 像情報処理学講座 教授）に心より深謝致します．

本研究を進めるにあたり，暖かい御指導をして頂いた名古屋大学大学院情報科学科 村瀬 洋 教授（元 NTT コミュニケーション科学基礎研究所）に深く感謝いたします．

本研究の遂行において多大なる御助言，御鞭撻を賜り，終始御指導を頂いた 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 牧野 昭二 氏，メディア認識研究グループ 大和 淳司 氏，柏野 邦夫 氏に心から感謝いたします．

また，日頃から有益なご意見を頂き御指導いただいた 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 向井 良 氏，大塚 和弘 氏，永野 秀尚 氏，泉谷 知則 氏，木村 昭悟 氏，ぷららネットワークス 川西 隆仁 氏に深く感謝いたします．

最後に，本論文をまとめるに当たって御協力いただいた小谷研究室の諸兄に厚く御礼申し上げます．

参考文献

- [1] YouTube. <http://www.youtube.com/>.
- [2] Cliplife. <http://cliplife.jp/>.
- [3] Napster. <http://www.napster.com/>.
- [4] 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.
- [5] 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—”. 電子情報通信学会論文誌 D-II, Vol. J81-DII, No. 9, pp. 2035–2042, 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—時系列アクティブ探索法—”. 電子情報通信学会論文誌 D-II, Vol. J82-D-II, No. 9, pp. 1365–1373, 1999.
- [7] K. Kashino, T. Kurozumi and H. Murase. “a quick search method for audio and video signals based on histogram pruning”. *IEEE Transactions on Multimedia*, Vol. 5, No. 3, pp. 348–357, September 2003.
- [8] T. Kurozumi, K. Kashino and H. Murase. “A Robust Audio Searching Method for Cellular-Phone-Based Music Information Retrieval”. *Proc. of ICPR2002*, Vol. 3, , August 2002.
- [9] P. Cano, E. Batlle, T. Kalker and J. Haitsma. “A Review of Algorithms for Audio Fingerprinting”. *Proc. of MMSP*, pp. 169–173, 2002.

- [10] J. Haitsma, T. Kalker. “A Highly Robust Audio Fingerprinting System”. *Proc. of ISMIR*, pp. 107–115, 2002.
- [11] A. Wang. “An Industrial Strength Audio Search Algorithm”. *Proc. of ISMIR*, pp. 7–13, 2003.
- [12] Michael Fink, Michele Covell and Shumeet Baluja. “Social- and Interactive-Television Applications Based on Real-Time Ambient-Audio Identification”. *Proc. of Euro-ITV*, 2006.
- [13] MPEG-7. ISO/IEC 15938, Multimedia content description interface.
- [14] 黒住 隆行, 柏野 邦夫, 村瀬 洋. “実環境で受音した楽音をキーとする楽曲探索法”. 電子情報通信学会論文誌 D-II, Vol. J86-DII, No. 12, pp. 1719–1726, 2003.
- [15] 黒住 隆行, 柏野 邦夫, 村瀬 洋. “携帯カメラで収録した映像をキーとする一致映像探索”. 電子情報通信学会技術報告 PRMU2002-130, pp. 7–11, December 2002.
- [16] “小特集—マイクロホンアレー—”. 音響誌, Vol. 51, No. 5, pp. 384–414, 1995.
- [17] 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.
- [18] 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*, No. 3, pp. 211–218, 1997.
- [19] Yihong Gong, Guido Proietti, and Christos Faloutsos. “Image Indexing and Retrieval Based on Human Perceptual Color Clustering”. *Proc. of CVPR*, pp. 578–583, 1998.

- [20] Howard D. Wactlar, Michael G. Christel and Alexander G. Hauptmann. “Lessons Learned from Building a Terabyte Digital Video Library”. *Computer*, pp. 66–72, February 1999.
- [21] 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”. 画像の認識理解シンポジウム MIRU2000, July 2000.
- [22] K. Lemstrom and S. Perttu. “SEMEX - An efficient Music Retrieval Prototype”. *MIR 2000*.
- [23] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele and P. Yanker. “Query by image and video content: the QBIC system”. *IEEE Computer*, Vol. 28, No. 9, pp. 23–32, 1995.
- [24] Paul Hough. “Method and Means for Recognizing Complex Patterns”. *U.S. Patent*, No. 3069654, 1962.
- [25] Duda, R. O. and P. E. Hart. “Use of the Hough Transformation to Detect Lines and Curves in Pictures”. *Comm. ACM*, Vol. 15, pp. 11–15, 2003.
- [26] Frank O’Gorman, MB Clowes. “Generalizing the Hough transform to detect arbitrary shapes”. *Pattern Recognition*, Vol. 13, No. 2, pp. 111–122, 1981.
- [27] Dana H. Ballard. “Finding Picture Edges Through Collinearity of Feature Points”. *IEEE Trans. Computers*, Vol. 25, No. 4, pp. 449–456, 1976.
- [28] Yehezkel Lamdan and Haim J. Wolfson. “Geometric Hashing: A General and Efficient Model-Based Recognition Scheme”. *Proc. of ICCV1988*, pp. 238–249, 1988.
- [29] Haim J. Wolfson. “Geometric Hashing: An Overview”. *IEEE Computational Science and Engineering*, Vol. 4, No. 4, pp. 10–21, October 1997.
- [30] Donald E. Knuth. *The Art of Computer Programming*, Vol. 3. Addison-Wesley, 1973.

- [31] S. F. Boll. “Suppression of acoustic noise in speech using spectral subtraction”. *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-27, No. 2, pp. 113–120, 1979.
- [32] S. Furui. “Cepstral analysis technique for automatic speaker verification”. *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-29, No. 2, pp. 254–272, 1981.
- [33] Richard W. Hamming. “Error-detecting and error-correcting codes”. *Bell System Technical Journal*, Vol. 29, No. 2, pp. 147–160, 1950.
- [34] 黒住 隆行, 柏野 邦夫, 村瀬 洋. “携帯電話で受音した音をキーとする音響信号探索”. *情報科学技術フォーラム 情報技術レターズ*, pp. 159–160, September 2002.
- [35] 藤本 雅清, 有木 康雄. “音声の時間変化モデルに基づく音声信号推定法を用いた非定常雑音下での音声認識”. *電子情報通信学会技術報告*, Vol. SP2000-78, pp. 19–24, 2000.
- [36] 圧境 誠, 中村 哲, 鹿野 清宏. “ケプストラム平均正規化法とHMM合成法に基づくモデル適応化法”. *電子情報通信学会論文誌 D-II*, Vol. J80-D-II, No. 10, pp. 2636–2644, 1997.
- [37] 中川 聖一. “音声認識研究の動向”. *電子情報通信学会論文誌 D-II*, Vol. J83-D-II, No. 2, pp. 433–457, 2000.
- [38] L. R. Rabiner and B. H. Juang. *Fundamentals of speech recognition*. Prentice Hall, 1993.
- [39] D. I. Barnea and H. F. Silverman. “A class of algorithms for fast digital image registration”. *IEEE Trans. Comput.*, Vol. C-21, No. 2, pp. 179–186, 1972.
- [40] 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ティング”. *電子情報通信学会論文誌 D-II*, Vol. J84-D-II, No. 8, pp. 1817–1825, 2001.

- [41] T. Kurozumi, K. Kashino and H. Murase. “A Method for Robust and Quick Video Searching Using Probabilistic Dither-voting”. *Proc. of ICIP2001*, Vol. 2, pp. 653–656, October 2001.
- [42] Boyer R. S. and Moore J. K. “A Fast String Searching Algorithm”. *Comm. of the ACM*, Vol. 20, No. 10, pp. 762–772, 1977.
- [43] Eamonn J. Keogh and Michael J. Pazzani. “An Indexing Scheme for Fast Similarity Search in Large Time Series Database”. *Proc. of ICSSDM*, pp. 56–67, 1999.
- [44] 片山紀生, 佐藤真一. “SR-Tree: 高次元点データに対する最近傍探索のためのインデックス構造の提案”. 電子情報通信学会論文誌 D-I, Vol. J80-D-I, No. 8, pp. 703–717, 1997.
- [45] 山岸史典, 片山紀生, 佐藤 真一, 坂内 正夫. “縮退特徴量を用いた疑似クラスタリングによる高次元近接点探索の高速化”. 情報科学技術フォーラム, Vol. I-040, pp. 89–92, 2004.
- [46] T. Nishimura, N. Sekimoto, J.X. Zhang, M. Ihara, T. Akasaka, H. Takahashi and R. Oka. “Methodology for Retrieving Time Sequence Pattern”. *IWHIT/SW'99*, October 1999.
- [47] K. Kashino, T. Kurozumi and H. Murase. “Feature Fluctuation Absorption for a Quick Audio Retrieval from Long Recordings”. *Proc. of ICPR2000*, Vol. 3, pp. 102–105, September 2000.
- [48] 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”. 電子情報通信学会技術報告, No. SP99-23, June 1999.
- [49] 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”. 電子情報通信学会論文誌 D-II, Vol. J83-D-II, No. 12, pp. 2735–2744, 2000.

- [50] 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, March 1999.
- [51] 守谷 健弘. “音声符号化”. 電子情報通信学会, 1998.
- [52] 杉山 雅英. “セグメントの高速探索法”. 電子情報通信学会技術報告, No. SP98-141, February 1999.
- [53] Yijun Li, Jesse S. Jin and Xiaofang Zhou. “Video Matching Using Binary Signature”. *Proc. of International Symposium on Intelligent Signal Processing and Communication Systems*, pp. 317–320, 2005.
- [54] E. Kasutani and A. Yamada,. “The MPEG-7 color layout descriptor: a compact image featuredescription for high-speed image/video segment retrieval”. *Proc. of ICIP*, Vol. 1, pp. 674–677, 2001.
- [55] MPEG-7, Part 3. ISO/IEC 15938-3, Multimedia content description interface – Part 3.
- [56] 安部 素嗣, 西口 正之. “背景音楽同定のための自己最適化スペクトル相関法”. 電子情報通信学会技術報告 PRMU2001-209, pp. 25–30, January 2002.
- [57] 中居友弘, 黄瀬浩一, 岩村雅一. “特徴点の局所的配置に基づくデジタルカメラを用いた高速文書画像検索”. 電子情報通信学会論文誌 D, Vol. J89-D, No. 9, pp. 2045–2054, 2006.

本研究に関する研究業績

査読付き論文

1. 黒住 隆行, 永野 秀尚, 柏野 邦夫, “実環境で収録された映像断片をキーとする一致映像探索”, 電子情報通信学会論文誌 D, Vol.J90-D, No.8, pp.2223–2231, Aug. 2007.
2. 黒住 隆行, 柏野 邦夫, 村瀬 洋, “実環境で受音した楽音をキーとする楽曲探索法”, 電子情報通信学会論文誌 D-II, Vol.J86-D-II, No.12, pp.1719–1726, Dec. 2003.
3. 黒住 隆行, 柏野 邦夫, 村瀬 洋, “時系列アクティブ探索法の特徴ひずみに頑健な確率ディザボーティング”, 電子情報通信学会論文誌 D-II, Vol.J84-D-II, No.8, pp.1817–1825, Aug. 2001.

その他の論文

1. 川西 隆仁, 久野 和樹, 木村 昭悟, 黒住 隆行, 柏野 邦夫, 高木 茂, “サブテンプレート間距離を用いた適応的のウィンドウスキップによる高速テンプレートマッチング法”, 電子情報通信学会論文誌 D-II, VOL.J88-D-II, No.8, pp.1389–1397, Aug. 2005.
2. Kunio Kashino, Takayuki Kurozumi and Hiroshi Murase, “A Quick Search Method for Audio and Video Signals Based on Histogram Pruning”, IEEE Transactions on Multimedia, Vol. 5, No. 3, pp.348–357, Sep. 2003.
3. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “グローバルな枝刈りを導入した音や映像の高速探索”, 電子情報通信学会論文誌 D-II, Vol.J85-D-II, No.10,

pp.1552–1562, Oct. 2002.

4. 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”, 電子情報通信学会論文誌 D-II, Vol.J83-D-II, No.12, pp.2735–2744, Dec. 2000.

査読付きレター論文

1. 川西 隆仁, 黒住 隆行, 柏野 邦夫, 高木 茂, “参照画像内の部分領域間の残差情報を利用した高速テンプレート照合法”, 情報科学技術フォーラム 情報技術レターズ, Vol. 2, pp. 175-176, Sep. 2003.
2. 黒住 隆行, 柏野 邦夫, 村瀬 洋, “携帯電話で受音した音をキーとする音響信号探索”, 情報科学技術フォーラム 情報技術レターズ, pp.159–160, Sep. 2002.

査読付き国際会議予稿

1. Cheng-Yao Chen, Takayuki Kurozumi and Junji Yamato, “Poster Image Matching by Color Scheme and Layout Information”, Proc. of ICME2006, Jul. 2006.
2. Kunio Kashino, Akisato Kimura and Takayuki Kurozumi, “A Quick Video Search Method Based on Local and Global Feature Clustering”, Proc. of ICPR2004, Aug. 2004.
3. Takahito Kawanishi, Takayuki Kurozumi, Kunio Kashino and Shigeru Takagi, “A Fast Template Matching Algorithm with Adaptive Skipping Using Inner-Subtemplates’ Distances”, Proc. of ICPR2004, Aug. 2004.
4. Takahito Kawanishi, Takayuki Kurozumi, Shigeru Takagi and Kunio Kashino, “Skipping Template Matching Guaranteeing Same Accuracy with Exhaustive Search”, Proc. of ICAPR2003, pp.209-212, Dec. 2003.

5. Akisato Kimura, Kunio Kashino, Takayuki Kurozumi and Hiroshi Murase, “Dynamic-Segmentation-Based Feature Dimension Reduction for Quick Audio/Video Searching”, Proc. of ICME2003, Vol.2, pp.389–392, Jul. 2003.
6. Akisato Kimura, Kunio Kashino, Takayuki Kurozumi and Hiroshi Murase, “Dynamic-Segmentation-Based Feature Dimension Reduction for Quick Audio/Video Searching”, Proc. of ICASSP2003, Vol.3, pp.357–360, Apr. 2003.
7. Takayuki Kurozumi, Kunio Kashino and Hiroshi Murase, “A Robust Audio Searching Method for Cellular-Phone-Based Music Information Retrieval”, Proc. of ICPR2002, Vol.3, Aug. 2002.
8. Akisato Kimura, Kunio Kashino, Takayuki Kurozumi and Hiroshi Murase, “A Quick Search Method for Multimedia Signals Using Feature Compression Based on Piecewise Linear Maps”, Proc. of ICASSP2002, Vol.IV, pp.3656–3659, May. 2002.
9. Takayuki Kurozumi, Kunio Kashino and Hiroshi Murase, “A Method for Robust and Quick Video Searching Using Probabilistic Dither-voting”, Proc. of ICIP2001, Vol. 2, pp.653–656, Oct. 2001.
10. Akisato Kimura, Kunio Kashino, Takayuki Kurozumi and Hiroshi Murase, “Very Quick Audio Searching: Introducing Global Pruning to the Time-Series Active Search”, Proc. of ICASSP2001, Vol.3, pp.1429–1432, May, 2001.
11. Kunio Kashino, Takayuki Kurozumi and Hiroshi Murase, “Feature Fluctuation Absorption for a Quick Audio Retrieval from Long Recordings”, Proc. of ICPR2000, Vol.3, pp.102–105, Sep, 2000.
12. Takayuki Kurozumi, Yoshikazu Shinza, Yukiko Kenmochi and Kazunori Kotani, “Facial Individuality and Expression Analysis by Eigenspace Method Based on Class Features or Multiple Discriminant Analysis”, Proc. of ICIP’99, Vol. 1, pp.648-652, Oct. 1999.

査読付き研究発表予稿

1. 黒住 隆行, 柏野 邦夫, “実環境で収録した映像をキーとする一致映像探索”, 画像の認識理解シンポジウム MIRU2004, I, pp.133–134, Jul. 2004.
2. 柏野 邦夫, 木村 昭悟, 黒住 隆行, “高速な一致映像探索”, 画像の認識理解シンポジウム MIRU2004, I, pp.155–156, Jul. 2004.
3. 川西 隆仁, 黒住 隆行, 柏野 邦夫, 高木 茂, “サブテンプレート間距離を用いた適応的スキップによる高速テンプレートマッチング法—スキッピングテンプレートマッチング法—”, 画像の認識理解シンポジウム MIRU2004, II, pp.19–20, Jul. 2004.
4. 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, Jul. 2000.

研究発表予稿

1. 黒住 隆行, 木村 昭悟, 永野 秀尚, 柏野 邦夫, “幾何変換パラメータを特定する縮退生成探索法”, 電子情報通信学会技術報告 PRMU2006-171 pp.1–6, Dec. 2006.
2. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “音や映像の高速探索のための動的分割に基づく特徴次元削減法”, 電子情報通信学会技術報告 PRMU2002-167 pp.127–131, Dec. 2002.
3. 黒住 隆行, 柏野 邦夫, 村瀬 洋, “携帯カメラで収録した映像をキーとする一致映像探索”, 電子情報通信学会技術報告 PRMU2002-130 pp.7–11, Dec. 2002.
4. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “区分線形写像に基づく映像信号の高速探索”, 電子情報通信学会技術報告 PRMU2001-243, Feb. 2002.

5. 黒住 隆行, 柏野 邦夫, 村瀬 洋, “特徴ひずみに頑健な部分空間を用いた携帯端末による音響信号探索”, 電子情報通信学会技術報告 IE2001-23 PRMU2001-43 MVE2001-22 pp.31-38, Jul. 2001.
6. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “グローバルな枝刈りを導入した長時間音響信号の探索—時系列アクティブ探索の高速化—”, 電子情報通信学会技術報告 PRMU2000-200, pp.53-60, Feb. 2001.
7. 柏野 邦夫, 黒住 隆行, 村瀬 洋, “多種類の特徴歪みを考慮した音響信号の高速探索法”, 電子情報通信学会技術報告 DSP2000-98 SP2000-64, pp.63-69, Sep. 2000.
8. 柏野 邦夫, 黒住 隆行, 村瀬 洋, “二つの音響信号に共通に現れる部分信号区間の高速自動抽出”, 日本音響学会 春季研究発表会, Vol. 1, pp.133-134, 2000.
9. 黒住 隆行, 新座 良和, 剣持 雪子, 小谷 一孔, “固有空間法と重判別分析による顔画像の個人性と表情の解析”, 電子情報通信学会技術報告 CS98-122 IE98-101, pp.57-64, Dec. 1998.

全国大会予稿

1. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “区分線形写像に基づく音や映像の高速探索—特徴系列の分割による次元削減の導入—”, 電子情報通信学会 データ工学ワークショップ DEWS2003 3-A-1, Mar. 2003.
2. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “区分線形写像に基づく音響信号の高速探索”, 日本音響学会 春季研究発表会 3-P-28, Mar. 2002.
3. 黒住 隆行, 柏野 邦夫, 村瀬 洋, “携帯電話で受音した音楽をキーとする高速音響信号探索”, 電子情報通信学会 総合大会講演論文集 情報システム 2, p.224, Mar. 2001.
4. 黒住 隆行, 斎藤 康之, 剣持 雪子, 小谷 一孔, “固有空間法と重判別分析を用いた顔画像の表情解析”, 映像メディア処理シンポジウム IMPS'98, pp.13-14, Oct. 1998.

解説記事等

1. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “グローバルな枝刈りを導入した音や映像の高速探索”, 日本工業出版 画像ラボ, Vol.14, No.11, pp.60–67, Nov. 2003.
2. 柏野 邦夫, 黒住 隆行, 村瀬 洋, “Learning-based Active Search Library”, NTT Review, Vol.15, No.2, pp.38–41, Mar. 2003.
3. 柏野 邦夫, 黒住 隆行, 村瀬 洋, “楽曲や CM の高速探索を実現する学習アクティブ探索ライブラリ”, NTT 技術ジャーナル, Vol.14, No.12, pp.50–52, Dec. 2002.
4. 柏野 邦夫, 黒住 隆行, 村瀬 洋, “音や映像を素早く探すメディア探索技術の未来”, NTT 技術ジャーナル, Vol.14, No.8, pp.59–62, Aug. 2002.
5. 黒住 隆行, 柏野 邦夫, 村瀬 洋, “確率ディザボータイング: 低品質な映像も瞬時に探すためのコア技術”, 日本工業出版 画像ラボ, Vol.13, No.7, pp.41–46, Jul. 2002.
6. 柏野 邦夫, 黒住 隆行, 村瀬 洋, “時系列アクティブ探索法に基づく音や映像の高速 AND/OR 探索”, NTT R&D, Vol.50, No.11, pp.895–901, Nov. 2001.

特許

1. 黒住 隆行, 柏野 邦夫, “信号検出方法、装置、プログラム及びその記録媒体”, 特許 3943060.
2. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “信号圧縮方法、装置、そのプログラムと記録媒体、信号検索方法、装置、プログラムとその記録媒体”, 特許 3884006.
3. 黒住 隆行, 柏野 邦夫, 村瀬 洋, “信号検出方法及び装置、プログラムならびに記録媒体”, 特許 3746690.

4. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “信号検索装置、信号検索方法、信号検索プログラム及び信号検索プログラムを記録した記録媒体”, 特許 3730179.
5. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “信号検出方法、信号検出装置、記録媒体及びプログラム”, 特許 3574075.
6. 柏野 邦夫, 黒住 隆行, 村瀬 洋, “信号検出方法、信号の検索方法及び認識方法並びに記録媒体”, 特許 3537727.
7. 黒住 隆行, 柏野 邦夫, 村瀬 洋, “信号検出方法、装置及びそのプログラム、記録媒体”, 特許 3408800.

公開特許公報

1. 木村 昭悟, 柏野 邦夫, 黒住 隆行, 村瀬 洋, “信号検索方法、装置、プログラムとその記録媒体”, 特開 2007-034325.
2. 黒住 隆行, 柏野 邦夫, “信号探索装置、信号探索方法、信号探索プログラム及び記録媒体”, 特開 2006-293902.
3. 川西 隆仁, 黒住 隆行, 柏野 邦夫, 高木 茂, “テンプレート照合装置および方法”, 特開 2006-031392.
4. 川西 隆仁, 黒住 隆行, 柏野 邦夫, 高木 茂, “テンプレートマッチング方法、装置、プログラム及びその記録媒体”, 特開 2005-063206.
5. 黒住 隆行, 柏野 邦夫, “信号検出方法、装置、プログラム及びその記録媒体”, 特開 2005-062555.
6. 柏野 邦夫, 黒住 隆行, 村瀬 洋, “検索サービスの配信方法及びシステム, 情報検索方法及び装置, 情報検索サーバ, 検索サービス提供方法, そのプログラムおよび該プログラムを記録した記録媒体”, 特開 2005-025770.

7. 木村 昭悟，柏野 邦夫，黒住 隆行，村瀬 洋，“信号圧縮方法、装置、そのプログラムと記録媒体、信号検索方法、装置、プログラムとその記録媒体”，特開 2004-199047.
8. 黒住 隆行，柏野 邦夫，村瀬 洋，“映像信号検出方法および装置、映像信号検出プログラム並びにそのプログラムを記録した記録媒体”，特開 2004-193983.
9. 中川 雄弘，岡 宏美，羽山 政志，浜砂 勝哲，奥野 真衣，佐々 義信，黒住 隆行，青島 史郎，“コンテンツ配信方法および装置とコンテンツ配信プログラムおよび該プログラムを記録したコンピュータ読取り可能な記録媒体”，特開 2003-303295.
10. 木村 昭悟，柏野 邦夫，黒住 隆行，村瀬 洋，“信号検索装置、信号検索方法、信号検索プログラム及び信号検索プログラムを記録した記録媒体”，特開 2003-242510.
11. 黒住 隆行，柏野 邦夫，村瀬 洋，“信号検出方法及び装置、プログラムならびに記録媒体”，特開 2003-022084.
12. 木村 昭悟，柏野 邦夫，黒住 隆行，村瀬 洋，“信号検出方法、信号検出装置、記録媒体及びプログラム”，特開 2002-236496.
13. 柏野 邦夫，黒住 隆行，村瀬 洋，“信号検出方法ならびに装置および同方法がプログラムされ記録された記録媒体”，特開 2002-108385.
14. 黒住 隆行，柏野 邦夫，村瀬 洋，“信号検出方法、装置及びそのプログラム、記録媒体”，特開 2002-044610.
15. 柏野 邦夫，黒住 隆行，村瀬 洋，“検索サービスの配信方法及びシステム，情報検索方法及び装置，情報検索サーバ，検索サービス提供方法，そのプログラムおよび該プログラムを記録した記録媒体”，特開 2002-041569.
16. 柏野 邦夫，黒住 隆行，村瀬 洋，“信号検出方法、信号の検索方法及び認識方法並びに記録媒体”，特開 2001-242880.

表彰

1. 2005年3月22日, 永野 秀尚, 黒住 隆行, 柏野 邦夫, 村瀬 洋, “A. 「多数の小領域スペクトログラムの探索に基づく背景音楽の高速探索法」, B. 「実環境で受音した音をキーとする楽曲探索法」”, 電気通信普及財団 テレコムシステム技術賞 奨励賞.
2. 2004年6月28日, Kunio Kashino, Takayuki Kurozumi and Hiroshi Murase, “A Quick Search Method for Audio and Video Signals Based on Histogram Pruning”, IEEE Transactions on Multimedia Paper Award.
3. 2002年5月28日, 村瀬 洋, 柏野 邦夫, 黒住 隆行, “音・映像信号の高速探索技術に関する先駆的研究”, 電子情報通信学会 業績賞.