

# テレビ番組から視聴したい定番コーナーをすばやく探索するシリーズコーナー検出機能

Regular Feature Detection Function for Fast Identification of Regularly Broadcast Segments in TV Programs

大盛 善啓

■ OHMORI Yoshihiro

映像の内容を解析して自動的に見出しを付けるインデクシング技術が近年注目され、繰返し・盛り出し検出などの機能がこれまでに製品化されている。しかし、これらの機能は、ニュース番組で毎回放送されるスポーツコーナーのように、複数の録画映像の解析を必要とする定番コーナーを検出できなかった。

東芝は、複数の映像に共通に含まれる動画像と音声を高速に探索することによって、冒頭にいつも決まった動画像や音声を含む定番コーナーに見出しを付けるシリーズコーナー検出機能を開発した。この機能で利用している探索アルゴリズムは約16 Kバイトの少量のメモリで高速に動作するため、AVノートPC（パソコン）や組込み機器に適用できる。

Automatic indexing techniques for videos have become a focus of attention in recent years due to their usefulness in searching for desired scenes in large volumes of accumulated TV programs, and indexing functions such as repetitious scene detection and climax detection have already been released. However, these functions are unable to adequately detect regular features such as a sports segment in a news program because the analysis of multiple recorded programs is required.

Toshiba has developed a regular feature detection function for TV programs that can search for video and audio fragments commonly contained in multiple programs. The search algorithm of this function is capable of fast operation using a small amount of memory of about 16 Kbytes, and is suitable for audiovisual PCs and embedded systems. This function provides a new style of viewing such as jumping among regular features in different days' broadcasts.

## 1 まえがき

近年の映像圧縮技術の進歩や記憶装置の大容量化によって、ハードディスクレコーダや録画機能付きテレビ（TV）、AVノートPCなどが普及し、一般家庭で大量の映像を録画できるようになった。

映像は一覧して内容を把握することが難しいため、大量の録画映像の中から見たい映像やシーンを効率良く探す機能への要求が強くなってきた。そこで、映像を短縮視聴するために内容を解析してインデックスを付加する技術が注目されている。東芝は、ハードディスクレコーダのヴァルディア（VARDIA）<sup>TM</sup>にTV番組内の繰返しシーンにチャプターを付けるマジックチャプター<sup>TM</sup>や、AVノートPC Qosmio<sup>TM</sup>にTV番組の登場人物の顔を一覧する顔deナビ<sup>TM</sup>を搭載してきた。これらの機能は単数の映像を解析するだけで実現できるが、ニュース番組で毎回放送されるスポーツコーナーのように、複数の録画映像を解析する必要がある定番コーナーを検出できなかった。

当社は、複数の動画像と音声から類似区間を探索することにより、TV番組に継続して含まれる定番のコーナー（以下、シリーズコーナーと呼ぶ）を検出するシリーズコーナー検出機能を開発した。

ここでは、動画像と音声から類似区間を高速に探索する技

術の概要と、シリーズコーナー検出に適した高速探索のPRFS（Predictive Region Fitting Search）法などについて述べる<sup>(1)</sup>。

## 2 シリーズコーナー検出機能

シリーズコーナー検出機能は、異なる放送日の映像に毎回含まれる同じコーナーを検出して、サムネイル画像を一覧表示す

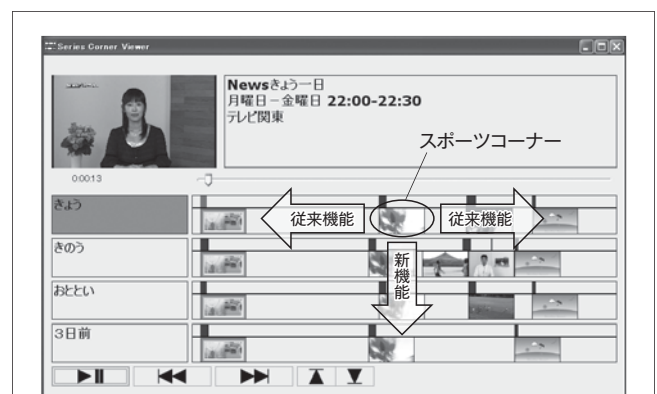


図1. シリーズコーナー検出機能の画面例 — きょうのニュース番組のスポーツコーナーを視聴中に、この画面を呼び出して下方に移動することで、きのうの野球の結果を視聴できる。

Example of regular feature detection function display

る(図1)。例えば、きょうのニュース番組のスポーツコーナーを視聴中にこの画面を呼び出し、きのうの野球の結果を視聴できる。従来の機能では、このような複数の映像間を移動する視聴は実現できなかった。

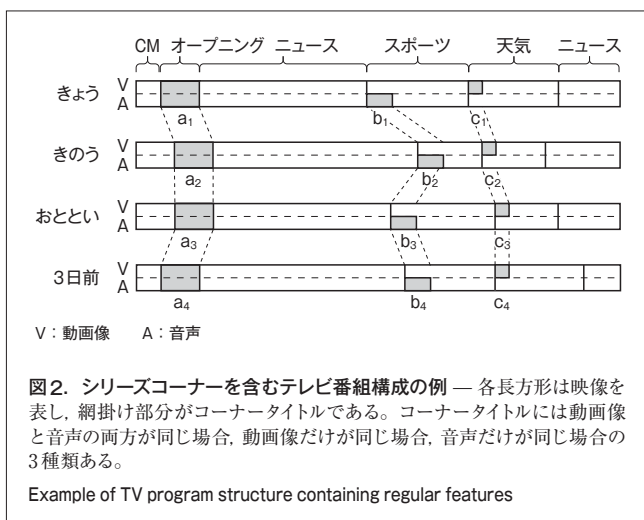
### 3 シリーズコーナーの検出方法

シリーズコーナーの中には、放送日が異なっても同一の動画像や音声を繰り返し含むものがある。例えば、ニュース番組に含まれるスポーツや株価のコーナーなどは、冒頭に毎回同じCG(コンピュータグラフィックス)動画像や音声を含むことが多く、これらをコーナータイトルと呼ぶ。国内だけでなく海外の放送番組でも、コーナータイトルを含むシリーズコーナーが多用されており、これらを検出すればシリーズコーナーの冒頭にインデックスを付加できる。

シリーズコーナーを含むニュース番組の構成の一例を図2に示す。この番組はCM(コマーシャル)、オープニング、ニュース、スポーツ、天気、ニュースのコーナーから成り、番組の構成は同じだが各コーナーの開始時刻が放送日によって異なる。したがって、映像を視聴して希望するコーナーを探すには時間を要する。

オープニング、スポーツ、天気のコーナーはコーナータイトルa, b, cを含むシリーズコーナーである。aは動画像と音声の両方、bは音声だけ、cは動画像だけが同じである。きょうときの映像間で動画像の類似区間を検出すると(a<sub>1</sub>, a<sub>2</sub>), (c<sub>1</sub>, c<sub>2</sub>)を検出でき、音声の類似区間を検出すると(a<sub>1</sub>, a<sub>2</sub>), (b<sub>1</sub>, b<sub>2</sub>)を検出できる。音声と動画像の検出区間を統合するとすべてのコーナータイトルを検出できる。きのうとおととい、おとといと3日前の間で同様に処理すれば4日分のコーナータイトルを検出できる。

映像間に共通に含まれる動画像と音声の類似区間を探す処理は、非常に計算量が多いため高速化が実用化の鍵となる。



## 4 部分類似区間の高速探索

### 4.1 基本手法と従来の高速化手法

きょうときのうの二つの動画像から部分類似区間を検出する方法について述べる。

まず、二つの動画像を構成するフレーム画像の平均輝度値を符号化して特徴ベクトル系列を算出する。次に、二つの特徴ベクトル系列のそれぞれに固定長の探索窓を設定して、探索窓の中にある特徴ベクトル間の類似度を算出する。最後に、探索窓を任意の位置に移動して類似度の大きい位置を統合すれば、部分類似区間を検出できる。しかし、このしらみつぶし探索法は処理量が非常に大きい。

探索を高速化するために、部分類似区間の特徴を抽出して番組のモデルを作成する手法があるが、照合するモデルを更新するために前後の録画映像との間で探索しなければならないので、前後の録画映像の間だけで探索するシリーズコーナー検出と処理量が変わらない。

また、すべての特徴パターンについてそのパターンが存在する映像上の位置を記憶した早見表をあらかじめ作成する方法がよく用いられる<sup>(2)</sup>。照合する代わりに早見表を引くだけで済むため高速化できる。あまり変化しない大容量のデータに対して何度も検索する場合に効率が良いが、大量のメモリを消費する。

そのほかに、類似度が明らかにしきい値を超えない範囲で照合を省略して高速化する手法があり<sup>(3)</sup>、事前処理が不要で消費メモリ量が少ないのでシリーズコーナー検出に適している。当社は、この手法を更に高速化したPRFS法を開発してシリーズコーナー検出に応用した<sup>(1)</sup>。

### 4.2 特徴ベクトルと類似度

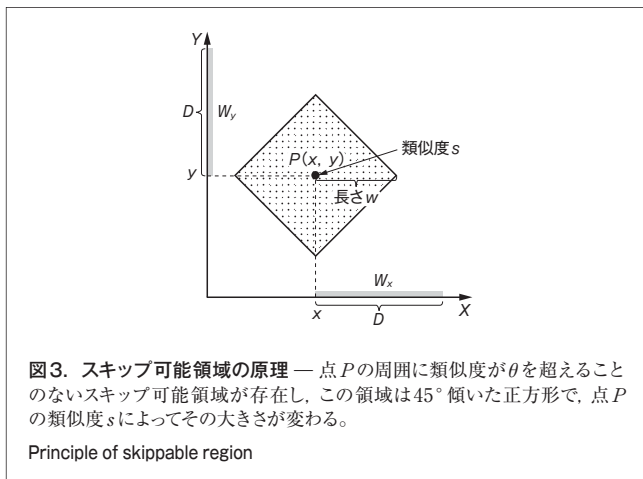
PRFS法では、動画像と音声から抽出した特徴ベクトルを用いて類似度を算出する。

動画像の特徴ベクトル抽出では、1枚のフレーム画像を9分割してそれぞれの領域の平均輝度値を4段階に量子化し、フレーム画像全体の平均輝度も求めて16段階に量子化した値を連結して特徴ベクトルとする。音声の特徴ベクトル抽出では、7チャンネルのバンドパスフィルタでパワーを算出し、各チャンネルのパワーを4段階に量子化した値を連結して特徴ベクトルとする。動画像と音声の標本化周波数は10 Hzで、特徴抽出にかかる処理量は非常に小さい。

類似度は正規化ヒストグラム重なり率で求める。正規化ヒストグラム重なり率とは、二つのヒストグラムが重なった部分の度数とヒストグラムの総度数との比で、0から1の間の値をとる。二つのヒストグラムが一致する場合に1になる。

### 4.3 高速化の原理

きょうときのうの映像から抽出した二つの特徴ベクトル系列XとYから、類似度がしきい値 $\theta$ を超える位置をPRFS法で探す方法について述べる。図3に示すように、X上とY上の



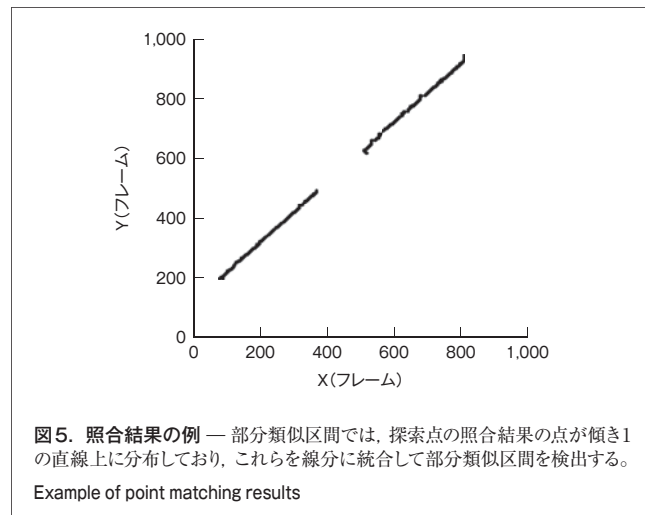
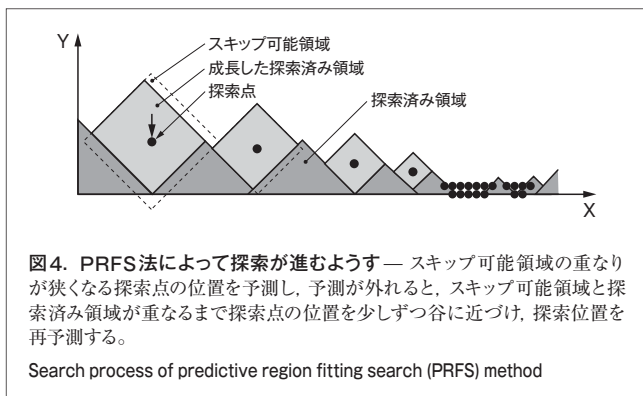
任意の位置に設定した探索窓を移動して探索するので、探索空間は2次元となる。点P(x, y)は、X上で位置xを先頭とする固定長Dの探索窓Wxと、Y上で位置yを先頭とする長さDの探索窓Wyを設定することを表す。

点Pの類似度をsとすると、点Pの周囲に類似度が $\theta$ を超えることのないスキップ可能領域は、類似度sによって正方形の大きさが変わる45°傾いた正方形となる<sup>(3)</sup>。対角線の長さを2wとするとwは(1)式で表される。ここで、 $\text{floor}(t)$ はtを超えない整数を返す関数とする。

$$w = \text{floor}(D(\theta - s)) + 1 \quad (1)$$

スキップ可能領域の内部で照合を省略することで探索を高速化できる。図4に示すように、PRFS法では探索空間全体をスキップ可能領域で敷き詰めて探索を進める。図中の濃い網掛け部分は探索済み領域、点は類似度を求めた位置(探索点)、点線はスキップ可能領域を表す。のこぎり状の探索済み領域の谷の部分にスキップ可能領域を当てはめて探索済み領域を成長させる。探索済み領域と重なったり、のこぎり状の形状からはみ出したりした部分は利用されない。

のこぎり状の探索済み領域とスキップ可能領域の重なりが狭くなるようにスキップ可能領域を敷き詰めれば効率良く探索できるが、類似度によってスキップ可能領域の大きさが変わる



ため、最適な探索点の位置はわからない。そこで、PRFS法では探索点の位置を予測し、予測が外れた場合には、スキップ可能領域と探索済み領域が重なるまで探索点の位置を少しずつ谷に近づけて、探索位置を再予測する処理を繰り返す。

#### 4.4 探索結果の統合

探索点の照合結果の例を図5に示す。類似度が0.7以上の点を表しており、部分類似区間では照合結果の点が傾き1の直線上に分布する。これらをつなげて線分を生成し、部分類似区間を検出する。図5は一つの類似区間の検出結果を表すが、二つに分かれている。そこで、間隔があらかじめ与えたい値以下の場合に、類似区間を連結して誤検出を抑制する。

## 5 PRFS法と従来手法の比較

### 5.1 処理時間

二つの50分間の映像から抽出した特徴ベクトル系列から類似区間をPRFS法で検出する処理時間を調べた。PRFS法と従来手法の探索処理時間を図6に示す。しらみつぶし探索法(ES)、時系列アクティブ探索反復法(rTAS)、参照区間自由時系列アクティブ探索法<sup>(3)</sup>(RIFAS)などの従来手法と、PRFS法を比較するため、インテル® Pentium®(注1)4プロセッサ、3 GHz、2 GバイトメモリのPC上で実験した。

ES法の処理時間は一定で、そのほかの手法は $\theta$ が大きいかほど処理時間が減少し、0.2より大きい場合PRFS法が最速だった。 $\theta$ が実用的な0.7の場合、PRFS法では従来最速だったRIFAS法よりも処理時間が4.7倍短くなった。映像のばらつきを考慮すると、1時間の映像において動画像と音声の両方の探索と検出結果の統合に、ES法では約6時間かかるが、PRFS法は4分以内で処理できる。組込み機器での処理時間はPCの場合の数倍にとどまると推測されるため、PRFS手法は組

(注1) インテル、Pentiumは、米国及びその他の国における米国Intel Corporation又は子会社の登録商標又は商標。



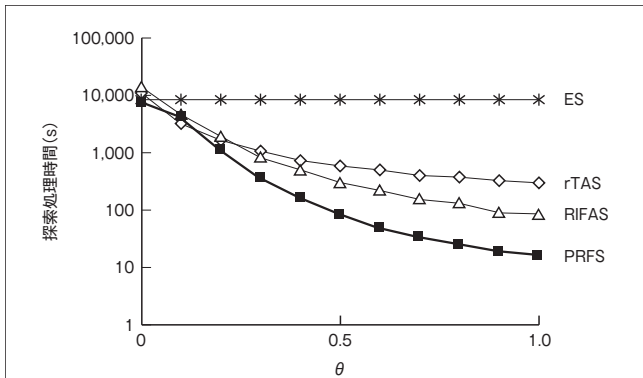


図6. 探索処理時間の比較 —  $\theta$ が実用的な値の0.7の場合、PRFS法では従来最速だったRIFAS法よりも処理時間が4.7倍短くなった。

Comparison of search times

込み機器にも実装可能と考えられる。

## 5.2 メモリ消費量

PRFS法が消費するメモリ量についても調べた。PRFS法とRIFAS法は、探索済み領域の形状を記憶するためにメモリを消費し、探索済み領域で挟まれた谷の底の点(活性点)をすべて記憶する。したがって、RIFAS法とPRFS法では、特徴ベクトル系列の長さに比例する数の活性点を記憶する必要がある。探索プログラム本体を除けば、消費メモリ量のほとんどは活性点の記憶によるものである。

50分の映像について、探索の過程で記憶した活性点の個数を図7に示す。PRFS法で使用した活性点の数がRIFAS法よりも平均0.38倍少なく、PRFS法で活性点の記憶に必要なメモリ量がRIFAS法の半分以下であることを意味している。

1時間の映像を処理するためにRIFAS法では4,000個の活性点を記憶する必要があるが、PRFS法では2,000個だけで十分である。一つの活性点の記憶に8バイト必要なので、1時間の映像では、PRFS法はわずか16 Kバイトのメモリしか消費しない。RIFAS法は32 Kバイトで、早見表を用いる手法ではPRFS法の約400倍の約7 Mバイトものメモリを消費する。

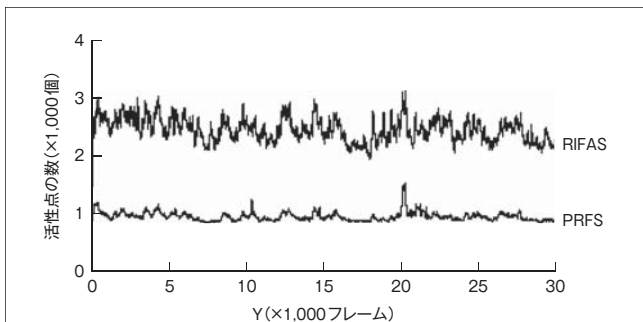


図7. 活性点の数の比較 — PRFS法で使用した活性点の数がRIFAS法よりも平均0.38倍少なく、PRFS法に必要なメモリ量はRIFAS法の半分以下である。

Comparison of number of active points

## 5.3 検出精度

コーナータイトルを含むシリーズコーナーをPRFS法を用いて検出した場合の適合率と再現率を表1に示す。映像の長さは1時間で、処理した映像数はニュース番組が4で音楽番組が2である。 $\theta$ は番組ごとに求め、検出区間が正解区間と少しでも重複すれば正解とした。

再現率は1.00でコーナータイトルをすべて検出できた。適合率は0.42と0.59で正解数の約2倍の数の過剰検出があったが、この程度の数であれば視聴時にスキップすればよく、視聴への影響は少ないと考えられる。

表1. PRFS法の検出精度

Detection accuracy of PRFS method

ジャンル	適合率	再現率
ニュース	0.42	1.00
音楽	0.59	1.00

## 6 あとがき

当社は、複数の動画像と音声から類似区間を高速に探索するPRFS法を開発し、TV番組に継続して含まれる定番のコーナーを検出するシリーズコーナー検出機能を試作した。この機能をPC上に実装すると、1時間の映像を約4分で処理できる。また、PRFS法では、活性点を記憶するのに約16 Kバイトのメモリしか消費しないので、組込み機器にも適用できる。この機能により、複数の映像間を移動する新たな視聴スタイルを提供できるようになる。

映像機器の進歩とともに、映像視聴の支援技術は更に重要度を増していくと考えられ、当社は、今後も映像インデクシング技術を更に発展させていく。

## 文献

- (1) 大盛善啓. 探索位置の予測による時系列データの高速類似区間検出. 電子情報通信学会論文誌D. J92-D. 6, 2009, p.810-821.
- (2) Yang X-F., et al. Efficient Short Video Repeat Identification With Application to News Video Structure Analysis. IEEE Trans. Multimedia. 9, 3, 2007, p.600-609.
- (3) 西村拓一, ほか. アクティブ探索法による時系列データ中の一致区間検出 — 参照区間自由時系列アクティブ探索法 —. 電子情報通信学会論文誌D-II. J84-D-II. 8, 2001, p.1826-1837.



大盛 善啓 OHMORI Yoshihiro

研究開発センター マルチメディアラボラトリー 研究主務。  
ウェアラブル端末、信号探索、及び映像インデクシングの研究に従事。電子情報通信学会、情報処理学会会員。  
Multimedia Lab.