

I-007

## 音響情報の時空間配置照合によるビデオ間の部分一致検索 Identical Segment Search between Video data by Time-space Matching of Acoustic Information

江六前 政宏<sup>†</sup>    伊藤 慶明<sup>‡</sup>    小嶋 和徳<sup>‡</sup>    石亀 昌明<sup>‡</sup>  
Masahiro Erokumae    Yoshiaki Ito    Kazunori Kojima    Masaaki Ishigame

### 1. はじめに

今日では、インターネットを介して大容量のデータを扱うことが可能となり、動画を共有する動画サイトも増加している。一方、著作権を侵害する動画の違法アップロードが問題となっており、違反動画を高速に検出できる技術が求められている。

タグ付けが行われていない時系列データの照合方式が提案されているが[1]、MFCC等、特徴ベクトルの抽出やその量子化だけでも多くの計算資源を必要とする。

本研究では、ビデオの音響部のパワーの時系列のみから、そのピーク点の時空間配置情報を特徴量として照合することで、高速な特徴量抽出と照合を実現する。放送された映画データを一度録画し、その実データを用いた照合実験を行い提案手法の有効性を評価する。

### 2. 提案手法

提案手法の全体図を図.1に示す。DVD等のビデオ情報を参照信号、ネット等にあるビデオ情報を入力信号として、各信号から音響を取り出しパワーを算出する。次にパワーの時系列からパワーの大きくなっている箇所(ピーク点)を抽出し、このパワーピークの時空間配置を特徴量とする。ピーク点間の時空間距離を、入力信号と参照信号の局所距離としてDPマッチングを適用し、入力信号内に参照信号が含まれている区間を検出する。以下、それぞれの処理を詳しく説明する。

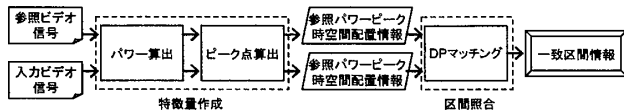


図1 提案手法の処理方式

#### 2.1 パワー算出処理

ビデオから取り出した音響信号  $y(t)$  に対しフレーム時間長  $L_f$ 、フレームシフト  $L_s$  で、以下の式でパワーの時系列データを算出する。

$$P(t) = \sqrt{\sum_{i=0}^{L_p} y(t+i)^2} \quad (t = nL_s) \quad (2.1)$$

#### 2.2 ピーク点抽出処理

算出したパワーの時系列から、値が極大値を示す箇所のみをピーク点として抽出する。

2.1 で求めたパワーの時系列に対し、新たな窓長  $L_p$  (パワー算出で用いた窓とは異なる) 内でパワーの値が局所最大になった場合にピーク点とする。これにより、およそ窓長  $L_p$  の間隔でピーク点が抽出される。

このピーク点はパワーの値とデータ内での位置情報から成り、ビデオの音響特徴を表す時空間配置情報であり、これを用いて照合処理を行う。

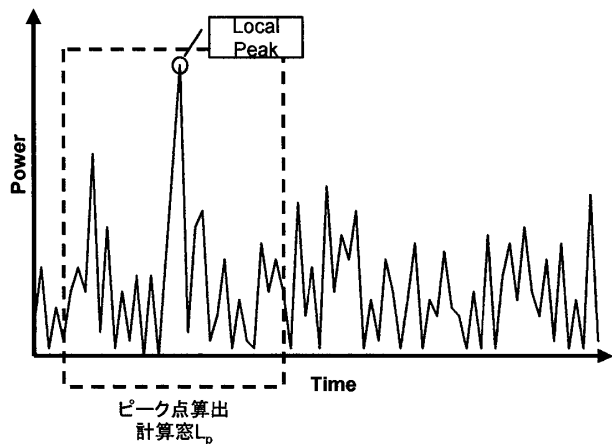


図2 パワー時系列からのピーク点抽出

#### 2.3 時空間配置照合処理

参照信号長  $L_R$  と同一の時間長の入力信号を切り出し、DPマッチングを行う。参照側の  $i$  番目のパワーピーク  $P_i$  の時刻  $t_i$ 、入力側の  $j$  番目のパワーピーク  $P_j$  の時刻  $t_j$  とすると、DPにおける局所距離は、今回は2つのパワーピーク時空間配置の市街地距離として以下の式で求める。

$$d_{ij} = |P_i - P_j| + |t_i - t_j| \quad (2.2)$$

### 3. 評価実験

はじめに述べたように違反動画の検出を前提として、手元にあるビデオの5分程度の区間を多数のビデオの中から高速に検出することを目的とし評価を行う。

#### 3.1 実験データ

実験データには、地上波テレビ放送の映画番組を一旦ハードディスクレコーダに録画保存し、音響データをサンプリング周波数 16kHz、量子化ビット数 16bit で取り出した。計 24 番組の放送された映画を用い、各番組から 1 時間ず

<sup>†</sup> 岩手県立大学大学院 Graduate School, Iwate Prefectural University

<sup>‡</sup> 岩手県立大学 Iwate Prefectural University

つ切り出し繋げた 24 時間の音響データを入力信号とし、参照信号は、各番組から 5 分ずつ、 $\times 10$  個切り出し、合計 240 個を用意した。この際、パワー計算時に参照信号と入力信号の波形が同一にならないよう入力側は 5ms ずらして切り出した。

### 3.2 実験条件と評価方法

提案する検索方式では、パワー算出処理での窓長  $L_f$  とシフト長  $L_s$ 、ピーク点抽出処理での窓長  $L_p$  を設定する必要がある。まずこれらの窓長と性能の関係について調査を行う。

性能は正解区間の再現率(Recall)と適合率(Precision)により評価する。検出した入力信号と、参照信号の時間区間の差が 3 秒以内であれば正解区間とし、正解数は 240 (参照信号の個数) となる。

処理速度は、5 分の参照信号に対し、1 時間の入力信号を処理する時間で評価し、[1][2]の方式との比較を行う。

### 3.3 実験結果

#### 3.3.1 検出精度

本方式には 3 つのパラメータ  $L_f$ 、 $L_s$ 、 $L_p$  がある。フレームシフト  $L_s$  は、シフト長を増加させるに伴って分析が荒くなり検出性能が低下する。一方、シフト長を短くしすぎると性能は向上せずに処理速度が低下する。予備実験の結果より、詳細で冗長な面もあるが窓長を 10ms と設定した。

パワーピーク抽出窓  $L_p$  は 1 秒の場合、5 分間のデータで約 75 点のピーク点が得られる。今回は、 $L_p$  を 2 秒と固定し 150 点のピーク点を時空間情報とした。 $L_p$  の適切な値については今後の課題とし、本稿ではパワー算出の窓長  $L_f$  についての実験結果を考察する。

図 3 に、 $L_f$  を 100ms から 4000ms に変化させた場合の検出精度の変化を示す。 $L_s$  は 10ms、 $L_p$  は 2 秒に固定し評価を行った。パワー算出の窓長  $L_f$  を 2000ms に設定したとき検出性能が最も良くなった。

窓長  $L_f$  を 2000ms より短く設定するとパワー値の増減が大きくなり、ピーク点の抽出箇所が安定しなくなる。長すぎる場合はパワー値の増減が緩やかになるため波形に特徴が現れにくく、精度の低下が見られた。今回の実験ではパワー算出の窓長  $L_f$  とパワーピーク抽出の窓長  $L_p$  を同程度の 2 秒に設定すると、最も高い性能が得られた。このようにパラメータの荒い設定を行った場合でも高い検出性能が得られ、本手法の有効性を確認できた。

#### 3.3.2 処理速度

表 1 に提案手法の特徴量作成と区間照合に要する計算時間、符号化 LPC ケプストラム特徴量[1][2]の計算時間を示す。前者の提案手法の特徴量作成条件は 3.3.1 で最も高い性能を示した、パワー算出窓長 2000ms、パワー算出窓フレームシフト 10ms、ピーク点抽出窓長 2 秒とする。後者の符号化 LPC ケプストラム特徴量の作成条件は、ハミング窓長 16ms、フレームシフト 16ms、LPC ケプストラム分析 16 とし、符号帳サイズは 32[1][2]、これらの計算には SPTK[3]を使用した。

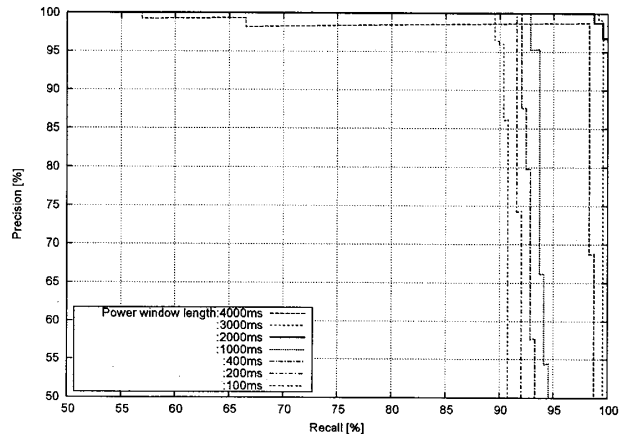


図 3 パワー計算窓長  $L_f$  による性能の推移

表 1 計算時間の比較

提案手法	符号化LPCケプストラム	
パワー算出	6.491秒	ケプストラム算出 11.840秒
ピーク点抽出	0.615秒	符号化 0.973秒
照合	0.289秒	照合
総計	7.844秒	総計 12.813秒

符号化 LPC ケプストラム特徴量の作成に 12 秒以上の計算時間を要するのに対して、提案手法では特徴量作成から区間照合までの処理を 8 秒で完了する。また LPC ケプストラムの符号化と同等の処理時間で、提案手法はピーク点抽出と照合を完了する。以上より本手法の高速性が確認できた。

本手法は MPEG のような圧縮されたデータ中のパワーを表す係数等をそのまま使うことも可能であり、その実装方式を検討したいと考える。

### 4. まとめ

本論文では、動画共有サイトにおける違反動画を高速に検出できるよう、音響情報を用いたビデオ間の一致区間を高速に検出する方式を提案した。パワーの時系列の極大値のみによる時空間配置を特徴量とすることにより、高速かつ高性能な照合が実現できることを、実データを用いた評価実験で確認できた。

今後は放送環境の異なるデータや圧縮データで評価を行っていききたい。また MPEG 等の圧縮データを直接特徴量とする方式についても検討していききたいと考える。

#### 謝辞

本研究の一部は文部科学省科学研究費補助金基盤(C) No.20500096 を受けて実施された。

#### 参考文献

- [1]菅井 康祐, 杉山 雅英, “任意個数の時系列に含まれる類似部分探索法”, 電子情報通信学会技術報告, NLC-2007-78, SP2007-12 (2007).
- [2]杉山 雅英, “複数時系列中の類似セグメント高速探索法”, 情報処理学会論文誌, Vol.49, No.1 (2008).
- [3]徳田・李研究室, Speech Signal Toolkit(SPTK).