ビデオ内音響情報の時空間配置を特徴量とした一致区間検索方式の評価

江六前　政宏†　伊藤　慶明†　石亀　昌明†　小嶋　和徳†
岩手県立大学　ソフトウェア情報学研究科†

1. はじめに

近年高速インターネットの整備により，ビデオを共有する動画サイトが一般的になった．一方，著作権を侵害するビデオの違法アップロードが問題となっており，ビデオデータを簡便に検索できる機能が求められている．本研究では，ビデオの音響部からパワーの時系列を求め，その系列中の極大・極小値を示す点を特徴点として抽出し，特徴点の相対値を時空間配置情報として特徴量に使用することで高速にビデオ検索する方式を提案し，実データを用いて提案方式の評価を行う．

2. 提案方式

提案方式の処理の流れを図 1 に示す．本稿ではDVD 等のビデオ情報を参照信号，インターネット等にあるビデオ群情報を入力信号として，各々の信号から音響信号を取り出しパワーを算出する．このパワーの時系列からパワーの極大値，極小値を示す点を特徴点として抽出する．この特徴点は時間情報とパワーの空間情報からなっている．2 つの信号間の音量等の音響的な違いを吸収するため入力信号と参照信号の局所距離は特徴点間の相対的な距離とする．照合は特徴点の湧き出しや脱落を考慮して 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索する．以下，それぞれの処理を詳述する．

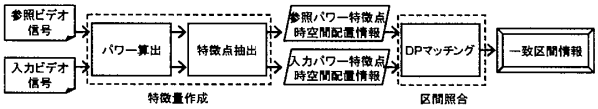

図 1　提案手法の処理方式

2.1 パワー算出処理

ビデオから取り出した標本化周期 T_s の音響信号 $y(t)$に対し，フレーム時間長 L_f 秒，フレームシフト L_s 秒で，以下の式により近似的なパワーを算出しパワーの時系列データとする．実際の計算ではシフト分のパワーを計算しておき，この分の加減算を行うので同一標本値の加算処理は 1 度限りである．

$$P(t)=\frac{T_s}{L_f}\sum_{i=1}^{L_f/Ts}|y(t+T_s\times i)| \quad \begin{pmatrix} t=nL_s \\ n=0,1,2,\cdots \end{pmatrix} \qquad (1)$$

Evaluation of an Identical Segment Search using Acoustical Time-space Feature in Video
†Masahiro Erokumae, Yoshiaki Itoh, Kazunori Kojima, Masaaki Ishigame, Graduate School of Software and Information Science, Iwate Prefectural University

2.2 特徴点抽出処理

算出したパワーの時系列から，極大・極小値を示す箇所のみを特徴点として抽出する．パワーの時系列に対し，新たな特徴点抽出窓長 L_p（L_f とは異なる）内でパワーの値が極大・極小になる場合に特徴点とする．この特徴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と定義する．

2.3 時空間配置照合処理

放送環境・収録環境の違いによる音量の絶対量の差やデータの劣化などにより，同一内容のビデオであってもパワーの値は必ずしも同一でない．この差異を考慮し，2.2 で求めた特徴点のパワーの絶対値ではなく相対値を用いて照合する．参照側の i 番目の特徴点は時刻 t_i の時にパワーP_i^R，入力側の j 番目の特徴点は時刻 t_j の時にパワーP_j^I とする．この 2 つの特徴点間の局所距離は，ユークリッド距離等様々考えられるが，今回は計算負荷が小さい特徴量の市街地距離を用いる．

$$d_{ij}=|(P_i^R-P_0^R)-(P_j^I-P_0^I)|+|(t_i^R-t_0^R)-(t_j^I-t_0^I)| \qquad (2)$$

P_0^R，P_0^I は窓内の始端のパワーを表し，各特徴点のパワーから減算することで窓内の始点からの相対値を取っている．[1]では式(2)のように局所距離をパワー値と時間情報値から均等に算出したが，本稿では放送環境の違いによるパワーの変動を考慮し，パワーと時間情報に対して式(3)のように重みαによって重み付けを行う方式を提案する．

$$d_{ij}=\alpha|(P_i^R-P_0^R)-(P_j^I-P_0^I)|+(1-\alpha)|(t_i^R-t_0^R)-(t_j^I-t_0^I)| \qquad (3)$$

3. 評価実験

評価実験では違反動画の検索を想定して，手元にある DVD の 30 秒から 3 分程度のビデオ区間を多数のビデオの中から検索する設定で行う．

3.1 評価用データ

実験データには，地上波テレビ放送の映画と，それと同一の映画の DVD を用いる．地上波テレビ放送映画は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 で処理しパワー情報とした．DVD 中の映画についても同様である．参照・入力信号は以下のように作成した．

■参照信号：3 分，2 分，1 分，30 秒の 4 つの信号長で 1 本の映画から 10 個ずつ切り出し，7 本の映

画に対しの参照信号を用意した(各信号長について70個ずつ，2分以下は3分の参照信号に含まれる)
■入力信号：3分の参照信号と同一内容の区間を正解区間として設定し，正解区間の切れ目が付加されないようにその前後に正解区間に連続する15秒の区間を付加した3分30秒の区間を抽出し，参照信号数分の10組用意した．同じ番組から参照信号とは異なる2分30秒の区間を切り出し，正解区間の前後に挿入し，1本の映画当り1時間の入力信号とした．残りの17本の映画は1時間ずつ切り出し，正解を含まない．これらを繋げた24時間のデータを入力信号とした．正解数は4つの参照信号長ごとに70個（各参照信号に1つ）となる．

3.2 実験条件と評価方法

実際の使用状況を想定して，DVDを参照用信号，テレビ放送を入力信号とした評価実験を行う．この際に入力信号と同一のテレビ放送信号を参照信号として用いた場合との比較を行う．区間検索性能は，F値により評価する．計算時間の評価は，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，符号化LPCケプストラムを用いる特徴量作成方式[2,3]との比較を行う．手元にあるビデオの音響情報に対して，特徴量作成の計算時間と，照合時間で比較し考察する．

3.3 区間検索性能

参照信号長ごとの区間検索性能について，入力信号にTV，参照信号にTVとDVDを用いて，重みαを0.5と0に設定した場合の結果を図2に示す．重みαを0.5としてパワーと時間情報を均等に用いた時，同一信号のTVを参照信号とすると1分以上の参照長にすれば100%の検索性能が得られた．参照信号をDVDとした場合でも参照長が3分の場合で93.2%のF値と高い検索性能が得られた．αを0として時間情報のみを用いた場合，参照信号長3分の場合F値が94.7%に向上した．これらの結果よりビデオ検索における提案方式の有効性が確認できた．参照長を30秒と短くするとα=0.5としてパワー情報を用いた方が高くなった．参照長が長い場合は時間情報のみで識別するには十分でパワーの

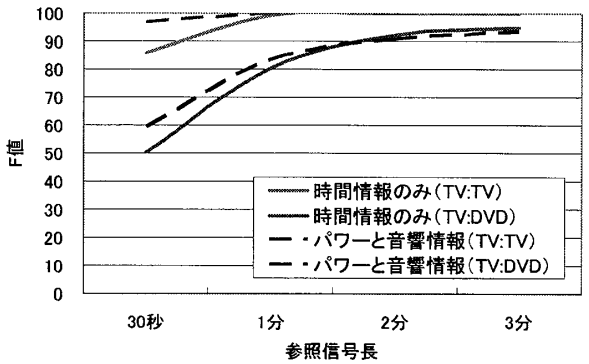

図2 参照信号にTVとDVDを用いた際の検索性能

表1 計算時間の比較

提案手法[s]		符号化LPCケプストラム[s]	
パワー算出	6.44	ケプストラム算出	11.84
特徴点抽出	0.64	符号化	0.97
照合(DP)	0.31	(照合)	-
総計	7.39	総計	12.81

変動を削除した効果があったが，参照長が短いと時間情報のみの照合では区間あたりの情報量が不足したため，パワー情報を用いた方が良くなったと考える．従って参照長が2分以上の場合には，時間情報のみを用いた照合を行い，短い場合にはパワーを用いた照合を行うべきことが確認できた．

3.4 計算時間

表1に提案方式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[2,3]の計算時間を示す．提案方式の特徴量作成条件は予備実験で最も高い性能を示した，パワー算出窓長1秒，パワー算出窓フレームシフト10ms，特徴点抽出窓長1.5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，符号帳サイズは32とし[2,3]，LPCケプストラムと符号化の計算にはSPTK[4]を使用した．表1に示すように符号化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，提案方式では特徴量作成から区間照合までの処理を8秒以内に完了した．

4 まとめ

本稿ではビデオデータ中の音響パワーの特徴点のみの字空間情報を利用することにより高速にビデオ検索する方式を提案した．実験により特徴点の時間情報のみを照合に用いることにより，放送環境の差異が存在する実環境下でも，参照信号長3分の時，従来手法の特徴量作成時間内にF値94%で検索できることを確認できた．今後は，パワーの代わりとなる時系列情報の高速な作成や圧縮データから特徴量を高速に抽出する方式を検討し，処理の高速化も図りたいと考える．

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] 江六前 政宏他，“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”，FIT2009（第8回科学技術フォーラム），pp.255-256，(2009).
[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] 柏野 邦夫他，“ヒストグラム特徴を用いた音響信号の高速探索法，”信学論(D-2)，vol.J82-D-2,no.9,pp.1365-1373 (1999).
[4] 徳田・李研究室, Speech Signal Toolkit(SPTK)