

代表特徴量を用いたビデオ識別処理の高速化

Acceleration of Video Identification Process using Group-of-Frame Feature

粕谷 英司†
Eiji Kasutani

山田 昭雄†
Akio Yamada

1. まえがき

放送の多チャンネル化やインターネットの高帯域常時接続環境の発展に伴い、大量の映像コンテンツが配信されるようになってきている。このため、これらの映像コンテンツを識別するしくみが、コンテンツの利用実態調査、著作権保護、放映動向分析、配信確認、利用監視などさまざまなアプリケーションで要求されている。コンテンツに対して事前に識別子を埋め込むことなくコンテンツを判別する方法として、筆者らはこれまでデータベースで管理されている大量のコンテンツの信号と逐次受信した映像信号の双方の特徴量を比較する方法について検討してきた [1]。しかしながら、フレーム単位で特徴量の照合をおこなっているため、照合回数が多いという問題点がある。そこで本稿では、識別精度を落とさずビデオ識別処理を高速化する方法として、コンテンツを代表する特徴量の生成方法および代表特徴量を用いたフィルタリングによる高速化方法について報告する。

2. 従来方式

ビデオ識別の高速化には、特徴記述の粒度と処理速度とのトレードオフを検討する必要がある。ビデオの記述方法としては、ビデオシーケンスよりキーフレームを選定して、キーフレームより抽出した特徴量を、シーケンスを代表する特徴量として用いる方法がよく知られている [2]。しかしながら、選定されるキーフレームにより特徴量が変わるため、ビデオ識別には向かない。この問題点に対応すべく、ビデオシーケンスに含まれる全てのフレームの集合よりシーケンスを代表する特徴量を生成する方法が Ferman により提案されている [3]。本方式では比較する特徴量を少なく抑えることで高速な識別を実現できるが、特徴量の時系列性が排除されてしまうことに起因して、ビデオシーケンスを識別するのに十分な精度が得られないという問題点がある。一方、ビデオシーケンスに含まれる各フレームの特徴量を利用した方法 [1] では、フレーム単位で特徴量比較することにより高い識別性能が得られるかわりに、フレーム単位で特徴量比較するため照合回数が多く処理コストがかかるという問題点がある。従って、高いビデオ識別精度を保ちつつ処理コストを削減するビデオ識別方法の開発が求められている。

3. システム概要

ビデオシーケンス全体を記述する代表特徴量を用いてフレーム単位で照合を行うビデオシーケンスを限定することにより、ビデオ識別精度を落とすことなく処理速度の高速化を図る。

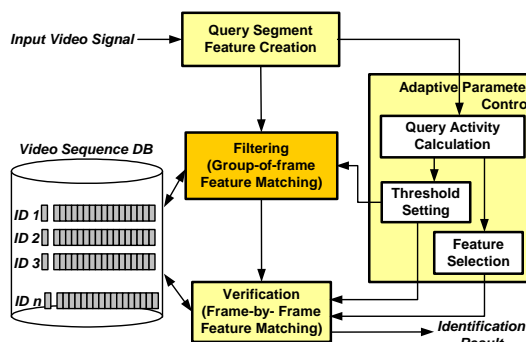


図1. ビデオ識別処理

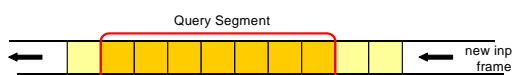


図2. クエリ生成

図1はビデオ識別システムの概要を示す。入力映像より最新N秒分の特徴量を逐次生成（クエリと呼ぶ。図2参照）しながら、映像素材データベースに登録された各コンテンツの特徴量（検索対象と呼ぶ）を順次比較することにより、現在入力されているコンテンツが何であるかを識別する。比較の際には、まずクエリと検索対象の代表特徴量を比較することによりフレーム単位の照合が必要なシーケンスを選定する。そして、選定されたシーケンスの一部がクエリと一致するか否かを、各特徴量をフレーム単位で比較することにより確定する。なお、照合の際のパラメータはクエリの性質に合わせて適応的に定められる [1]。入力信号より新規フレームをキャプチャするたびに、新規フレームを含む最新N秒のクエリを生成し、クエリに対する識別処理を行う。このため、リアルタイム処理の実現には、新たな問合せ信号が生成される前に識別処理が完了することが要求される。

4. 代表特徴量によるビデオ識別処理の高速化

4.1 特徴量記述

シーケンスに含まれる全てのフレームより抽出した特徴量に加えて、シーケンス全体の特徴を表現する代表特徴量をクエリおよび検索対象の特徴量として生成する。代表特徴量は、シーケンスに含まれる全てのフレームの特徴量のアグリゲーションにより生成する。アグリゲーション方法としては、平均化、メジアン算出を導入する。全てのフレームを利用することにより、キーフレームを選出する場合に比べて安定した特徴量を生成することができる。また、ハードウェア負荷を考えると、代表特徴量としてフレーム単位の照合に用いる特徴量と同じものを用いることが望ましい。そこで、フレーム特徴量、代表特徴量の双方について文献 [1] 同様 縮小画像の低次 DCT 係数列である Color Layout を用いる。Color Layout は以下のように定義される。

†NEC マルチメディア研究所

$$D = \{ n_Y, Y_i (i=1,2,\dots,n_Y), n_C, Cb_j, Cr_j (j=1,2,\dots,n_C) \} \dots(1)$$

n_Y および n_C はそれぞれ輝度成分および色差成分係数の数, Y_i, Cb_j, Cr_j はそれぞれ Y,Cb,Cr 成分の係数である.

アグリゲーション方法として平均化を用いる場合について説明する. まず代表特徴量 D_r を以下に定める.

$$D_r = \{ n_Y, RY_i (i=1,2,\dots,n_Y), n_C, RCb_j, RCr_j (j=1,2,\dots,n_C) \} \dots(2)$$

シーケンスに含まれる全フレーム数を M , m フレーム目の特徴量における Y, Cb, Cr 成分の係数を $Y_{i,m}, Cb_{j,m}, Cr_{j,m}$ とすると, 代表特徴量は式(3)により算出できる.

$$RY_i = \frac{1}{M} \sum_{m=1}^M Y_{i,m}, \quad RCb_j = \frac{1}{M} \sum_{m=1}^M Cb_{j,m}, \quad RCr_{j,m} = \frac{1}{M} \sum_{m=1}^M Cr_{j,m} \dots(3)$$

一方, メジアンを用いる場合は式(4)により算出できる.

$$\begin{aligned} RY_i &= \text{median}\{Y_{i,1}, Y_{i,2}, \dots, Y_{i,M-1}, Y_{i,M}\} \\ RCb_j &= \text{median}\{Cb_{j,1}, Cb_{j,2}, \dots, Cb_{j,M-1}, Cb_{j,M}\} \\ RCr_j &= \text{median}\{Cr_{j,1}, Cr_{j,2}, \dots, Cr_{j,M-1}, Cr_{j,M}\} \end{aligned} \dots(4)$$

4.2 代表特徴量によるフィルタリング

クエリーと検索対象の代表特徴量を比較することによりフレーム単位の照合が必要なシーケンスを選定する. 代表特徴量間の距離 d を算出し, d が式(5)を満たす場合, 該当するシーケンスを類似シーケンスとして選定する.

$$d < \theta + e \dots(5)$$

ここで, θ は類似シーケンスとみなすか否かを判定するしきい値で, e はクエリーと検索対象が構成するフレーム数の違いを吸収するための補正項である.

5. 評価実験

実際の放送波から取得した映像を用いて, 代表特徴量による絞り込みおよび識別速度について評価実験を行った.

実験 1. 代表特徴量による候補絞り込み

代表特徴量によるフィルタリングの実験を行った. テスト映像として, 解像度 320x240, ビットレート 480Kbps, フレームレート 29.97fps の条件で MPEG-4 符号化した映像, およびストリーミング配信を想定して同映像を低解像度 (160x120), 低ビットレート (128Kbps, 10fps) にトランスコードした MPEG-4 符号化映像を利用する. 双方の映像を 10 秒ごとのセグメントに分解し, 各セグメントから代表特徴量を生成しておく. 前者の映像を検索対象とし, 後者の映像に含まれる各セグメントをクエリーとして候補絞り込み実験を行う. 図 3 は, 平均化, およびメジアンにより生成された代表特徴量を用いて, しきい値 d の値を変えながら, 1)クエリーと同一箇所が検索対象から正しく選定された割合 2)絞り込まれたセグメント数の割合, について調査を行った結果を示す. 同一箇所が全て選定されるようにしきい値を設定した場合, 平均化によるアグリゲーションの場合には 20% 程度のセグメントに絞り込むことができるのに対して, メジアンを用いた場合には 90% 以上のセグメントが選定されてしまいフィルタリング効果は低い. このため, 良好な結果が得られる平均化をアグリゲーション方法として用いる.

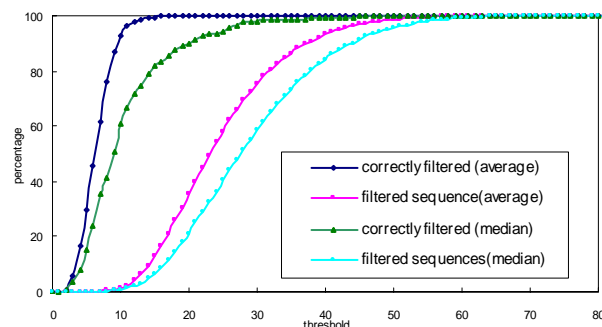


図 3. 識別精度比較

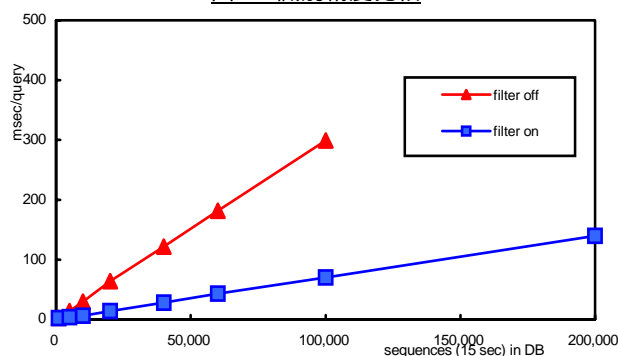


図 4. 識別速度比較

実験 2. フィルタリングによる識別速度

実験 1 で定められたしきい値を用いて識別速度の評価実験を行った. Intel Pentium IV 2GHz の CPU を用いて, 登録シーケンスの数と 1 つのクエリーの識別処理に要する平均時間との関係を求めた結果を図 4 示す. 図 4 より, 代表特徴量によるフィルタリングの導入により, フレーム単位で照合を行う場合に比べて 5 倍程度高速化できることがわかる.

6. まとめ

代表特徴量を用いたビデオ識別処理の高速化手法を提案した. 本手法は, 識別対象となる映像全体に含まれるフレームの特徴量を平均化することにより映像を代表する特徴量を新たに生成する点, 識別処理において代表特徴量によりフレーム単位で照合する映像を限定する点が特徴である. 実験の結果, 代表特徴量の導入により識別精度を保ったまま識別処理時間を 5 倍程度高速化することができた.

参考文献

- [1] 粕谷, 山田, "MPEG-7 Color Layout Descriptor によるシーン適応型実時間ビデオ識別法", 信学総大 SD-3 Mar. 2002.
- [2] P. Aigrain, H. Zhang, and D. Petkovic. "Content-based representation and retrieval of visual media: A state-of-the-art review", Multimedia tools and applications, volume 3, pages 179--202. Kluwer Academic Publishers, 1996.
- [3] A. Ferman, S. Krishnamachari, A. Tekalp, M. Abdel Mottaleb, R. Mehrotra, "Group-Of-Frames/Pictures Color Histogram Descriptors for Multimedia Applications", Proc. of ICIP2000, pp65-68, Sep. 2000.