

山岸 史典<sup>†</sup>, 佐藤 真一<sup>‡</sup>, 浜田 喬<sup>†‡</sup>Fuminori Yamagishi<sup>†</sup>, Shin'ichi Satoh<sup>‡</sup> and Takashi Hamada<sup>†‡</sup>

## 1 はじめに

現在 PC 等のハードディスク容量は 3 年で 10 倍の伸びを示しており, 5 年後には 10TB の容量を有するホームサーバが現実のものとなり, 放送映像を残さずすべて蓄積した後, 探索・ブラウズして視聴する形態が可能になると考えられる. このような環境では, 大量の映像の中から重要な映像を検出したり, 映像間の関連性を自動的に見いだすことが強く求められる. そのための技術としてわれわれは映像断片照合が重要な役割を果たすと考えている. 本稿では大規模映像アーカイブにおける効率のよい映像断片照合の実現方法について検討を行う.

## 2 映像断片照合

本稿では, 蓄積された大量の映像データのなかから, “同一” の映像断片 (= セグメント) 対を発見することを映像断片照合と定義する. 実際に数日分の NHK のニュース 7 の映像の中で, 同一映像断片対とすべき例を図 1 に示す. このようにここでい

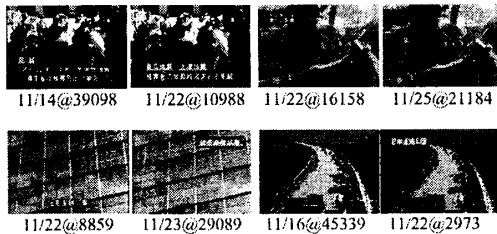


図 1: 映像断片照合の例

う “同一” では, 素材映像そのものが同じものであれば, それに重畳されているテロップなどの違いは許容している. 一般に重要な資料映像等は繰り返し使われる傾向にある. このほかにも事故映像, スポーツ中継の決定的瞬間などがその例である. このように大量の放送映像に対して, 同一映像断片探索を行うと重要度が高い映像を抽出することが可能である. また同時に抽出された同一映像を含む各部分は互に関連した内容を取り扱っていると考えられるので, 関連項目へのリンクを自動生成していることにもなっている.

映像断片照合では, 本質的に “同一” の素材映像の照合を行う必要があるが, DCT の低周波成分等を用いた “類似” 判定手法は利用できない. また, 素材映像のごく一部を使用した映像断片対をもれなく検出するため, フレーム単位の照合が必要である. 問い合わせ映像と照合する部分を探索する手法についての先行研究が存在するが [1][2], 完全照合を目的としたものが多く, かつアーカイブ中の照合対を発見する検討はなされていない. 本質的にきわめて大規模な映像アーカイブ中のごくわずかな同一映像断片対をもれなく発見することが重要となるため, 精密であると同時に効率のよい手法の実現が必須である.

## 2.1 研究の背景

近年のマイクロプロセッサ・ストレージ技術のめざましい向上により, 数年程度で個人が TB 級のハードディスクを備え,

大量の映像蓄積が可能なホームサーバを持つようになるであろう. 本研究はこのような環境で利用されることを想定している. われわれはこの考えに基づき, 地上波東京キー局 7 局の放送映像を 24 時間, MPEG-1 および 2 にてキャプチャし 10TB (約 7000 時間) の容量で蓄積するシステムを構築した [3]. 本システムでは, 毎時 7 チャンネル分 7 時間の映像が新規に獲得される. 従って, この環境において本稿で述べる映像断片照合を定期的に運用するための技術課題は, 蓄積されている数千時間の映像と新規獲得分の 7 時間の映像との間の照合を 1 時間以内に終了するように実現することとなる.

## 2.2 映像断片照合の定義

本研究では映像断片照合の対象となる映像断片を, 連続したフレームにより構成される映像の一部と定義する. 照合可能な映像断片対とは, 対応するフレームがすべて照合可能な映像対であると定義する. 本研究では, 映像断片対は少なくとも 1 秒 (30 フレーム) の長さを持つと仮定する. これにより 2 本の映像間での映像断片照合において, 一方については 30 フレームおきのフレーム照合判定によりもれなく映像断片照合が可能である. 一般にはショットと映像断片とは一致せず, あくまで映像断片はショットの一部分となるため, 基本的にはショット単位ではなくフレーム単位の照合判定が必要である.

フレーム間の照合判定として, 全画面での精密な照合が必要であるため, ここではフレーム間の正規化相互相関 (Normalized Cross Correlation; NCC) を基準として用いることにした. 本研究では, 相関値がきわめて 1 に近い (具体的には 0.85~0.90 以上) 場合を照合したとする. 実際の映像で検証した結果, 想定した照合可能としたい対をほとんど過不足なく検出できることを確認している.

NCC により精密な照合が実現できる反面, すべての画素値から相関値を計算する必要があるため, 計算時間が大きいという問題がある. また, こうした多次元データの照合の高速化のため, 木構造等のデータ構造が有効であることが知られているが [4], データの次元数が数百を越えると大幅に性能が劣化することが知られており, 各フレームが画素数分の次元数を有するとして扱う NCC によるフレーム照合の高速化にはそのままでは利用できない.

## 3 映像断片照合の高速化

処理量の低減と同時に木構造の適用による高速化を図るため, NCC による精密な照合に先立ち, 各フレームを 100 前後の次元数の色ヒストグラムに変換し, これを用いた候補の絞込みにより処理効率の飛躍的向上を図る. 色ヒストグラム間の  $L^1$  距離は, 画像照合に比較的良好な性能を示すことが知られている [1]. 画像を構成する画素間の関係が全く無相関ならば, 色ヒストグラムがほとんど同一でありながら全く異なる画像対は無数に存在し, 画像照合における色ヒストグラムの有効性は低いと考えられる. しかし, 例えば放送映像アーカイブのような特殊なデータセットでは画像の分布の極端な偏りから, 色ヒストグラムがきわめて近い画像対はほとんど同一の画像になると期待できる (局所的偏り). まずこれを実験により検証する.

## 3.1 色ヒストグラムによる絞込み

照合すべき画像対に対して, 絞込みのため, 色ヒストグラム間の  $L^1$  距離がある閾値以下の対のみを抽出するが, この際

<sup>†</sup> 東京大学大学院 情報理工学系研究科  
Graduate School of Information Science and Technology,  
The University of Tokyo  
<sup>‡</sup> 国立情報学研究所  
National Institute of Informatics

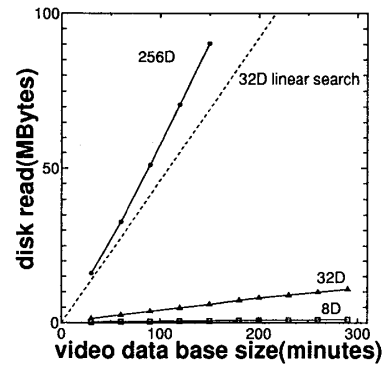
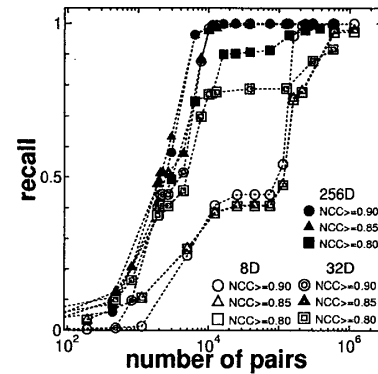
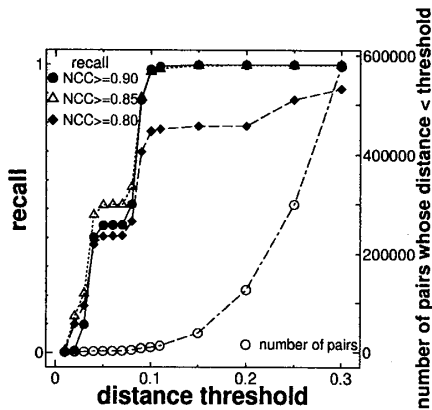


図 2: 色ヒストグラム距離の閾値と呼出率  
と絞り込まれた画像対の数

図 3: 呼出率と絞り込まれる画像対の数  
と絞り込まれた画像対の数

図 4: 蓄積映像量とディスク読み出し量

に 2.2 節で定義した照合可能画像対 (NCC が一定値以上) の取りこぼしがほとんど発生しないことが求められる。同時に効率化の観点から絞り込まれた画像対がなるべく少ないことが望ましい。そこで、色ヒストグラムによる絞り込みで得られる画像対集合に対する実際の照合可能画像対の呼び出し率と、絞り込みで得られる対の数との関係を調査した。2002 年 5 月 11 日と 16 日に放映された、NHK ニュース 7 の映像 (いずれも 30 分、54000 フレーム) 2 本から 10 フレームごとに 1 フレームずつ抜き出し、5400 フレームの集合を 2 組用意し、相互に照合候補 (5400<sup>2</sup> 対) とした。画像のサイズは 352 × 240 である。

色ヒストグラムとして、RGB および縦横それぞれ  $4^3 \times 2^2 = 256$ ,  $2^3 \times 2^2 = 32$ ,  $2^3 \times 1^2 = 8$ , の分割により、3 通りを用意した。32 次元の色ヒストグラムについて、NCC の下限を 0.90, 0.85, 0.80 と変えながら、 $L^1$  距離の閾値をパラメータとして、呼出率と絞り込みで得られた対の数をプロットしたものを図 2 に示す。実際に画像照合で利用する  $NCC \geq 0.90, 0.85$  の場合、 $L^1$  距離の閾値が 0.11 程度で呼出率も十分高く、絞り込みで得られる対の数も 10000 程度で効率的な絞り込みが可能であることが示されている。一方  $NCC \geq 0.80$  においてはいくら取りこぼしを許さなければ、後処理である相関値計算を短時間で終えることができなくなり絞り込みの意味がなくなる。

3 種類の色ヒストグラムの次元に対する、呼出率と絞り込みで得られる対の数の関係を図 3 に示す。32 次元の色ヒストグラムは 256 次元のものとはほぼ変わらない絞り込みの効果を上げることがわかる。それに対して、8 次元まで次元を落とすしまうと絞り込みの効果は格段に小さくなり、呼出率を 1 に近くとることが後の NCC 計算の手間を考慮すると難しいことが明らかである。一方、元の画像の持つ  $352 \times 240 \times 3$  次元という極めて高い自由度に対し、32 次元程度まで次元圧縮を行っても、極めて局所的には高い呼出率を実現可能であることがわかり、映像アーカイブのもつ強い局所的偏りが示されている。

### 3.2 SR-Tree による高速化の検討

色ヒストグラムを用いた絞り込みを効率的に実現するのに木構造を用いる有効性を検証する実験を行った。\*木構造としては高次元空間を効率的に取り扱えることが知られている SR-Tree [4] に対し、 $L^1$  距離を扱う改造を施して利用した。

実験の方法は次の通りである。まず NHK ニュース 7 の映像をフレームごとに色ヒストグラムに変換し、SR-Tree に全フレーム分投入したものを投入する映像量を変えて用意した。次にそれらに対して、投入したものは別のニュース 7 の映像 30 分から周期的に 30 フレームに 1 フレームをクエリ画像として、この画像から一定の色ヒストグラム距離内にある画像すべてを先の SR-Tree から検索した。これにより、2.2 節で定義した映像断片照合のための絞り込みを取りこぼしなく実現できることになる。色ヒストグラムは前項の 256, 32, 8 次元のそれぞれ

について実験を行った。距離の閾値は、前項の実験で同一画像対の候補として約 10000 対程度に絞り込まれる値を用いた。実験は PC (CPU: Pentium iii-1.13GHz, Memory: 512MBytes, OS: Linux 2.4) 上にて C++ 言語 (Compiler: gcc) を用いて行った。

通常、このような検索においてはプロセッサは十分に速くディスク I/O が律速段階であると考えられるので、ディスク I/O 量で処理効率を評価した。図 4 に 1 検索あたりにかかるディスク I/O 量を示す。例えば 32 次元において図 4 上の点線で示したような単純な線形探索に比べ、読み込みデータ量は 10 分の 1 程度におさまっているため、効率化は図られていることが示されている。

今回の手法により、32 次元ヒストグラムを用いた場合、5 時間のアーカイブと 1 時間の映像断片照合に約 8 分間かかることが実測されている。仮に今回の実験の手法をそのまま 7000 時間の映像アーカイブに適用することを考えると、2.1 節で述べた目標の達成のためには約 200 台のコンピュータで分散処理を行う必要があることになり、さらなる効率化が必要であろう。

## 4 まとめ

本稿では大規模映像アーカイブで利用することを前提とした映像断片照合を提案し、正規化相互相関を用いて定義を示した。しかしながら、これは計算量がきわめて多く、実際の大規模アーカイブにはそのままでは適用できない。そこで高速化のための手法として、色ヒストグラム間の距離による絞り込みが有効であることを実験により示した。さらに絞り込みを効率的に行うために SR-Tree を応用し、一定の高速化性能が得られることを示した。今後は大規模映像アーカイブでの実用化のために、より絞り込み性能のよい特徴量の利用、処理の並列化等について検討する予定である。

## 参考文献

- [1] 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, vol. J83-D-II 2000-12, pp.2735-2744, 2000.
- [2] 関本 信博, 西村 拓一, 高橋 裕信, 岡 隆一, “逐次フレーム入力されるクエリによる時系列データの類似区間スポットティング検索”, 電子情報通信学会論文誌, vol. J84-D-II 2001-5, pp.747-757, 2001.
- [3] 片山 紀生, 孟 洋, 佐藤 真一, “映像インデクシング研究のための大規模映像アーカイブシステムの試作”, 情報処理学会研究報告, 2002-DBS-127-3, 2002-FI-67-3, pp.17-23, 2002.
- [4] 片山 紀生, 佐藤 真一, “SR-Tree: 高次元データに対する最近接点検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, vol. J80-D-I 1997-8, pp.703-717, 1997.