

映像ショット切換え検出法とその映像アクセスインタフェースへの応用

谷口 行信[†] 外村 佳伸[†] 浜田 洋[†]

A Method for Detecting Shot Changes and Its Application to Access Interfaces to Video

Yukinobu TANIGUCHI[†], Yoshinobu TONOMURA[†], and Hiroshi HAMADA[†]

あらまし 本論文では、映像をリアルタイムに解析することにより、ショットの切換わりを自動的に検出する新しい方法を提案する。ショットは映像の基本的な単位であり、映像に対するアクセス、編集、検索などのインタフェースを実現する際に有用な情報である。提案法は、瞬時にショットを切り換えるカットだけでなく、フェードやワイプといったゆっくりとしたショット切換えも検出でき、瞬間的なノイズや、被写体の動きに対してもロバストであるという特徴をもつ。提案法は、“隣り合う”フレームの間だけでなく、より間隔をおいた2枚のフレームの間で非類似度を計算し、それらを総合的に評価してショット切換えの判定を行う点が従来の方法とは異なる。更に、検出されたショット切換え情報に基づいて実現される映像アクセスインタフェースとしてPaperVideoとTVRamの二つを提案することによって、本手法の有効性を示す。

キーワード 映像ハンドリング、カット検出、ショット切換え検出、インタフェース

1. ま え が き

コンピュータの高速化、記憶装置の大容量化等に伴い、Video On Demand、映像データベースなどコンピュータによる映像利用の幅が広がってきた。映像情報の量が増加するに従って、その中から本当に必要な情報を効率的に選択し、そこに素早くアクセスできるようにする効率的なインタフェースの重要性が増してきている。現在の映像情報に対するインタフェースは、ビデオデッキのそれ（早送り、再生、巻戻し、等）を模擬したものであり、ある映像の中から必要な部分を見つけ出すには、時間をかけてはじめてから終わりまで見ていくしかなかった。

効率的なアクセスインタフェースを実現するための第一ステップとして、映像を扱いやすく意味のある単位に分割する必要がある。文書には単語、パラグラフ、章、等の単位があるが、映像ではそのような単位が明示的になっていないことが問題である。1台のカメラ

で連続的に撮影された映像区間のことをショットと呼ぶ。ショットは映像の最小単位であり文章の単語にあたるものである。映像を時間順に見ていったとき一つのショットから次のショットに切り換わる変わり目をショット切換えと呼ぶ。ショット切換えの種類を表1に示す[2],[4]。もっとも一般的なのは、瞬時にショットを切り換えるカットである。徐々にショットを切り換える編集効果をまとめて漸次ショット切換えと呼ぶ。漸次ショット切換えはクロマスケーリング切換えと空間的切換えに大別される。クロマスケーリング切換えは、フェード、ディゾルブ等に代表されるものであり、輝度と色差の信号レベルを上げ下げして、一つのショットから次のショットに徐々に切り換えていく編集手法である。空間的切換えは、ワイプに代表されるものであり、一つのショットを一部分からぬぐい去るように消しながら、その後に次のショットを空間的に現していく編集手法である。その他の漸次ショット切換えとしてモーフィングなどがある。

本論文では、各種のショット切換えを画像処理によってリアルタイムに検出する新しい方法を提案する。

本論文の構成は以下のとおりである。2.では、ショッ

[†] NTT ヒューマンインタフェース研究所, 横須賀市
NTT Human Interface Laboratories, Yokosuka-shi, 238-03
Japan

表1 ショット切り換えの種類
Table 1 Types of shot changes.

ショット切り換え	カット
	漸次ショット切り換え
	クロマスケーリング切り換え (フェード、ディゾルブなど)
	空間的切り換え (ワイプなど)
	その他 (モーフィング)

ト切り換え検出法が備えるべき要件を挙げ、ショット切り換え検出の従来法について述べる。3. では、その要件をすべて満たすショット切り換え検出法を提案し、4. で実験結果を示す。5. では映像アクセスインタフェースの例としてPaperVideoとTVRamの二つを挙げ、提案したショット切り換え検出法の有効性を示す。

2. ショット切り換え検出法の要件

ショット切り換え、特にカットの検出方法については従来から多くの研究がなされている [1], [2], [6]~[8], [12]。本章では、まず一般的なショット切り換え検出処理の流れについて述べた上で、ショット切り換え検出法の備えるべき要件を明らかにする。

2.1 ショット切り換え検出処理の流れ

映像を構成する画像列を $f_0, f_1, f_2, \dots$ とする。 $t = 1, 2, 3, \dots$ について隣り合う2枚の画像 f_t, f_{t-1} の絵柄の違いを評価する値である非類似度 $d(t)$ を順次計算し、 $d(t)$ があるしきい値より大きくなった時刻 t をカットの位置として出力することでショット切り換えを検出することができる。例えば非類似度は

$$d_p(t) = \sum_{x,y} 1[|f_t(x,y) - f_{t-1}(x,y)| > T_p] / N_x N_y \quad (1)$$

として計算される。ここで、 N_x, N_y はそれぞれ x, y 方向の画素数を表し、 $1[\cdot]$ は括弧内の条件が成立するとき1を返し、それ以外のとき0を返す関数である。図1に非類似度 d_p の時間変化をグラフに示す。突出したピークがカットに対応している。

この単純な方法には次のような問題点がある。(1) 被写体やカメラの大きな動きが存在する区間で非類似度が大きな値を持つため誤検出が発生する可能性がある (図1(a))、(2) 暗い場面 (コントラストの低い場面) では非類似度の値が全体的に低くなるためショット切り換えを見落とすことがある (しきい値 T_p が固定

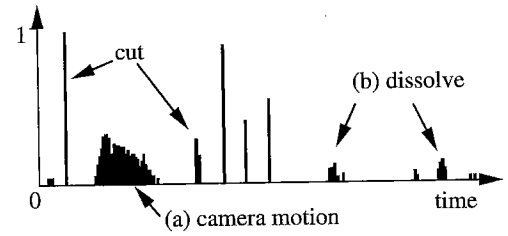


図1 非類似度 d_p の時間変化: (a) カメラの大きな動き, (b) 漸次ショット切り換え

Fig.1 Temporal change in dissimilarity measure d_p : (a) significant object motion, (b) gradual shot transition.

されているため)、(3) 漸次ショット切り換えにおいて非類似度 $d(t)$ はカットのような突出したピークを示さないためそれを検出できない (図1(b))。

2.2 ショット切り換え検出法が満たすべき要件

一般映像 (実験用映像ではなく、テレビ等で一般的に流通している映像) を扱う応用システムにショット切り換え検出処理を組み込むようにするためには、ショット切り換え検出法は以下の要件を満たす必要がある。

- (1) リアルタイム処理が可能であること、
- (2) 被写体やカメラの動きに対してロバストであること、
- (3) フラッシュ光などの瞬間的に発生するノイズに対してロバストであること、
- (4) 暗い場面やノイズの多い場面に適応して検出が可能であること、
- (5) 漸次ショット切り換えを検出できること。

(1) について、映像を標準速度で入力しながらショット切り換えをリアルタイムに検出できることが応用上、重要である。Armanら [1] は、高速化のために、JPEG 圧縮映像から復号化処理を経ないでショット切り換えを検出する方法を提案した。

(2) について、長坂ら [7] は非類似度として色ヒストグラムの χ^2 検定量を用いた。この非類似度は被写体の動きに対して insensitive であるという特長をもつからである。その反面、ショット切り換えの前後で背景色 (大きな面積を占める色) が大きく変化しない場合に検出もれを生じやすいという問題点があった。大辻ら [8] は $d(t)$ に対して突出検出フィルタと呼ばれる時間フィルタを作用することによって動きの影響を抑えた。しかし、このフィルタは漸次ショット切り換えに起因する非類似度の上昇も同時に抑えてしまうため (5) の要件を満たすことができない。

(3) について、静止画カメラのストロボの光によって画面全体の輝度が瞬間的に上昇することがある（ニュース映像に頻発する状況である）。単純なショット切換え検出法はこのノイズをカットと誤認識してしまう。長坂ら [7] はフレームを $4 \times 4 = 16$ 個の小ブロックに分割し、それぞれについて色ヒストグラムの差分を計算し、そのうち、値の大きいものから八つを捨て残りの総和をしきい値処理した。大辻ら [8] は瞬間的ノイズに対応するピークが2の幅をもつことを利用して誤検出を抑えた。

(4) について、1本の映像の中にも暗い場面、明るい場面、ノイズの多い場面など多種多様なものが混在しているため、コントラスト変化やノイズの大きさに適応してしきい値を制御する必要がある。ところが、従来法ではしきい値が固定されていたため、暗い場面のショット切換えを見落とすことが多かった。

(5) について、漸次ショット切換えはカットに比べれば使用頻度は低いものの、ニュース、ドキュメンタリ等では多用される手法であり、検出できることが望ましい。漸次ショット切換えにおいて、非類似度 $d(t)$ はカットの場合のような特徴的なピークを示さないため、カットと同じしきい値処理では検出が困難であった。比較する2枚のフレームの時間間隔を単純に広げればよいように思われるかも知れないが、そうすると被写体の動きをショット切換えと誤認識することが多くなる。

Zhang ら [12] はこの問題点を解決するために twin comparison 法を提案している。非類似度の微小変化を検出し、時間的に累積することで漸次ショット切換えを検出し、その上でカメラ操作とショット切り換えを区別するためにオプティカルフロー解析を後処理として行っている。

Hampapur ら [2] はショット切換えを種類ごとにモデル化するアプローチをとった。しかし、ショット切換えの種類は無限にあるためすべてに対応することはできない。更に、しきい値を含むパラメータの数が増えるため、それらの調整が困難であるという問題があった。

3. 複数スケール比較を用いたショット切換え検出法

本章では、前章で挙げた五つの要件をすべて満たすショット切換え検出法を提案する。

従来法はいずれも“隣り合う”フレームの間の非類

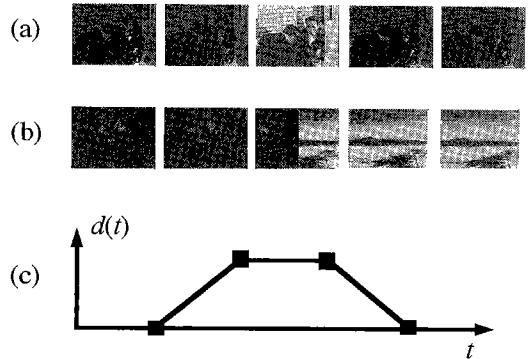


図2 従来法の限界を説明するための図：(a) フラッシュ光を含む画像列、(b) 漸次ショット切換え、(c) (a)、(b) から算出される非類似度 $d(t)$

Fig. 2 A limitation of previous methods: (a) an image sequence with flashing light, (b) an image sequence with a gradual shot change, (c) dissimilarity measure $d(t)$ obtained from (a) and (b).

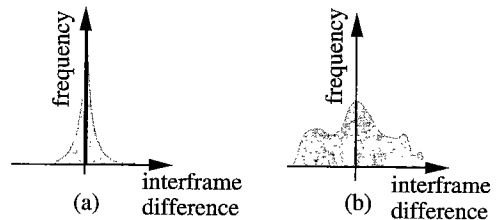


図3 フレーム間差分の分布：(a) 一つのショットに含まれる2枚の画像を比較した場合、(b) ショット切換えをはさむ2枚の画像を比較した場合

Fig. 3 Distribution of inter-frame differences: (a) between two frames inside a shot, (b) between two frames across a shot change.

似度 $d(t)$ に基づいて検出を行っていた。しかし、 $d(t)$ からだけではショット切換えと瞬間的なノイズとを完全に区別することはできない。このことは次のように理解できる。図2に、フラッシュ光を含む画像列 (a) と2フレームで完結する漸次ショット切換えを含む画像列 (b) を示す。これらの画像列に対して非類似度 $d(t)$ を計算すると、いずれの場合も図2(c) に示すように2の幅をもつピークを生じる。従って、“隣り合った”フレームの間の非類似度 $d(t)$ からだけではどちらの状況が生じているかを判別することはできない。瞬間的ノイズを誤認識することを避けようとする、逆に漸次ショット切換えが検出できなくなるため、従来法では (3) と (5) の要件を同時に満たすことができなかった。

このような考察に基づいて、提案法では“隣り合う”フレームの間での非類似度だけでなく、より広い間隔

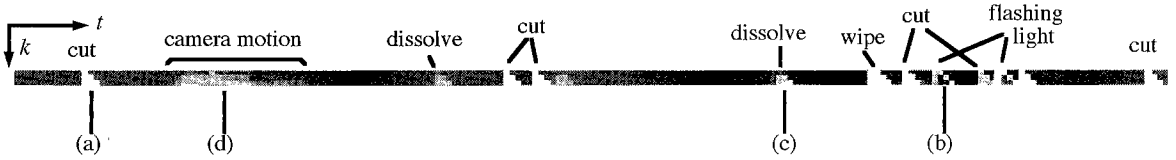


図4 非類似度の時間変化の例：(a) カット, (b) フラッシュ光, (c) 漸次ショット切換え, (d) カメラ操作

Fig. 4 An example of temporal change in dissimilarity measure: (a) cut, (b) flashing light, (c) gradual shot change, (d) camera operation.

をおいた2枚のフレームの間を比較して得られる非類似度を用いる。ここで k の間隔をおいた2枚の画像 f_t, f_{t-k} の間の非類似度を $d^k(t), k = 1, 2, 3, \dots$ と表し, k をスケールと呼ぶ。非類似度の計算方法については3.1で述べる。

問題は、複数スケールについて求めた非類似度の時系列 $d^k(t)$ をどのように統合して最終的な判定を行うか、である。大きなスケールで算出した非類似度は動きの影響を受けやすいし、小さなスケールで算出された非類似度は漸次ショット切換えに対して不感である。最終的な判定の方法については3.2で述べる。

3.1 非類似度の計算方法

フレーム間差分 $D_t^k(x, y) = f_t(x, y) - f_{t-k}(x, y)$ のヒストグラムは図3のような分布を示す。2枚の画像 f_t, f_{t-k} が同じショットに属する場合の典型的な分布を図3(a)に示し、ショット切換えをはさむ場合の分布を図3(b)に示す。(a)の場合は f_t と f_{t-k} が類似の画像であるので、0を中心とした急しゅんな山形の分布を示すのに対して、(b)の場合には大きなフレーム間差分が発生するため裾野の広がった形になる。

非類似度 $d^k(t)$ はこの分布の偏りを評価する次式によって算出する。

$$d^k(t) = \sum_{x,y} 1[|D_t^k(x, y)| > T^k(t)]. \quad (2)$$

但し、フレーム間差分 $D_t^k(x, y)$ の標準偏差を $\sigma^k(t)$ とし、しきい値を $T^k(t) = a \sum_{\tau=1}^{N_t} \sigma^k(t-\tau)$ (a は定数) のように標準偏差の平均に従って制御する。

暗い場面（コントラストの低い場面）では、ショット切換えがあっても顕著なフレーム間差分が発生しないため、従来の固定しきい値ではその微妙な変化を検出できなかった。逆にノイズの多い場面では定常的に大きなフレーム間差分が発生することが誤検出の一因であった。提案法は場面に応じてしきい値 $T^k(t)$ を制御して感度を適切に調整するので、暗い場面での検出

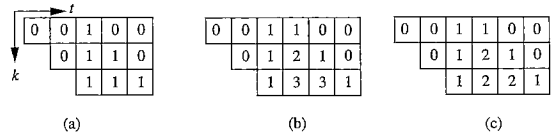


図5 テンプレートの例：(a) カット用, (b) クロマスケーリング切換え用, (c) 空間的切換え用

Fig. 5 Templates for three types of shot changes: (a) cuts, (b) chroma-scaling transitions, (c) spatial transitions.

もれや誤検出を防ぐ効果がある（要件(4)）。

3.2 ショット切換えの判定方法

本節では複数スケールで求められた非類似度の時系列 $d^k(t)$ に基づいて、ショット切換えが存在するか否かの判定を行う手続きについて述べる。

非類似度 $d^k(t), k = 1, 2, 3$ の時系列パターンを例を図4に示す。ここで非類似度の大きさを画素の明るさで表現しており、白色の画素は高い非類似度を表す。カットの発生した箇所では(a)に示すような三角形の特徴的なパターンを示す；非類似度 $d^k(t)$ は2枚の画像 f_t と f_{t-k} の絵柄の違いを評価する量であるから、時刻 t と $t-1$ の間にカットがあれば $d^k(t), d^k(t+1), \dots, d^k(t+k-1)$ が大きな値をもつというわけである。漸次ショット切換えではカットほど、はっきりとしたパターンを生じないがカットと類似した(c)のようなパターンを示す。フラッシュ光による瞬間的なノイズがあった場合には(b)のような、カットとは異なる特徴的なパターンを示す。被写体やカメラの動きによって生じるパターンは(d)のようになる。

提案手法では非類似度の時系列に現れるこのようなパターンに着目することによって、ショット切換えを検出する。ショット切換えの種類（カット、クロマスケーリング切換え、空間的切換え）に対応して図5に示す三種類のテンプレートを導入する。三つのテンプレート $\vec{d}^k(t)$ についてそれぞれ、観測された非類似度の時系列 $d^k(t)$ との相互相関係数

$$C(t) = \frac{\sum_{(\tau,k) \in S} (d^k(t+\tau) - \mu)(\bar{d}^k(\tau) - \bar{\mu})}{\sqrt{\sum_{(\tau,k) \in S} (d^k(t+\tau) - \mu)^2} \sqrt{\sum_{(\tau,k) \in S} (\bar{d}^k(\tau) - \bar{\mu})^2}} \quad (3)$$

を計算し、 $C(t)$ のいずれかがしきい値を超えたとき、時刻 t にショット切換えがあると判定する。ここで、 S はテンプレートの占める領域を表し、 $\mu = \sum_{(\tau,k) \in S} d^k(t+\tau)/|S|$, $\bar{\mu} = \sum_{(\tau,k) \in S} \bar{d}^k(\tau)/|S|$ である。

漸次ショット切換えに伴って現れるパターンは、切換えの種類や絵柄によって変化するので、本来、それぞれについて別々のテンプレートを設定する必要がある。しかし、経験的にクロマスケージング切換えに分類されるフェード、ディゾルブ等は図5(b)と類似のパターンを示し、空間的ショット切換えに分類される各種ワイプは図5(c)と類似のパターンを示す場合が多いことがわかったので、すべての漸次ショット切換えを図5(b), (c)の2種類のテンプレートで代表した。(しかし、高速な漸次ショット切換えがカット用テンプレートによって検出されたり、モーフィングがクロマスケージング用テンプレートによって検出されることもあった。)

瞬間的なノイズやカメラの動きは $d^k(t)$ の上でショット切換えとは明らかに異なるパターンを示すので、一時的なノイズや被写体の動きをショット切換えと区別できる (要件(2), (3))。また、漸次ショット切換え用のテンプレートを用意したことで漸次ショット切換えも検出可能となる (要件(5))。

4. 実験結果と考察

4.1 実験結果

レーザーディスクに記録された映像を標準速度で再生し、ビデオキャプチャボードを介してワークステーションに約8フレーム/秒のレートで縦横64×48画素に縮小した画像を取り込むことで、リアルタイム検出処理を実現した (要件(1))。非類似度を算出する際には輝度値だけでなくRGBの値を用いた。テンプレートとしては図5に示した三つを用いた。表2に、二つのニュース素材と映画に対して行った実験の結果を示す。それぞれの映像は24分の長さをもち、ニュースについては放送を1度VHSテープに録画したものを用いた。ニュース1には空間的ショット切換え (ワイプ) が15個、クロマスケージング切換え (ディゾルブ) が

表2 検出率と誤検出率: 括弧内の数字は (正しく検出されたショット切換えの数/映像に含まれているショット切換えの数) を表す。[] は誤検出数 (本来ショット切換えでないものをショット切換えと誤認識した数) であり、誤検出率は、誤検出数のショット切換えの数に対する割合である。

Table2 Detection and false-detection rate.

	ニュース1	ニュース2	映画
カット	99% (173/175)	94% (154/163)	95% (229/239)
空間的切換え	87% (13/15)	40% (2/5)	—
クロマスケージング切換え	78% (7/9)	50% (4/8)	—
誤検出率	7% [14]	10% [17]	8% [20]

9個含まれていた。そのほとんどが1秒以内で完結する比較的高速なショット切換えであった。ニュース2ではニュース1に比べて漸次ショット切換えの速度が遅かったため、それらを見落とす場合があった。

Hampapurらはカット検出率91%,空間的切換え60%,クロマスケージング切換え76%,誤検出率12%という結果を報告している [2]。Zhangらはカット検出率99%,漸次ショット切換え91%,誤検出率8%という結果を報告している [12]。実験に使用した映像素材がそれぞれ異なり、特に漸次ショット切換えについては速度によって性能が大きく異なるので、これらの結果を単純に比較することはできないが、漸次切換えについてはZhangらの結果に劣るものの、提案法はリアルタイム処理であるにもかかわらず、カット検出率と誤検出率については従来法と同程度の結果を得た。

4.2 考察

以上示した実験結果の他にもさまざまなジャンルの映像について実験を行った。誤検出の原因としては、急激な被写体のカメラの動き、大写しになった被写体の動き、車のライトの写りこみ、大きな面積を占める字幕や小画面の出現、などがあった。逆に検出もれの生じる場合としては、二つのショット切換えが短い間隔 (0.5秒程度以内) で発生する場合、ショット切換えの前後に被写体やカメラの動きがある場合、継続時間の長い (1秒程度以上) 漸次ショット切換え、等があった。特にサッカー中継では、(1) 多くのショットにおいて緑色のフィールドが画面の大部分を占めるためショット切換え時に非類似度が大きな上昇を示さない、(2) カメラが選手を追跡して大きく動くためショット内で非類似度が定常的に高い値を維持する、という二つの理由からショット切換えを見落とす場合が多かった。し

かし、ニュース、ドラマ、教育番組、バラエティ、映画等の映像については、90%程度の検出率が得られた。

漸次ショット切換え用の2種類のテンプレート(図5(b), (c))を導入したことにより高速な漸次ショット切換えの検出が可能となった。この二つのテンプレートは2フレーム(約8フレーム/秒でサンプリングした画像列)を処理している(約0.25秒)で完結する高速な漸次ショット切換えを検出するためのものである。ニュース等に現れる漸次ショット切換えは高速なものがほとんどであるのでこのテンプレートを用いたが、より継続時間の長いショット切換えを検出するためには、より時間幅の広いものが必要である。しかし、時間幅の広いテンプレートを用いるとカメラや被写体の動きをショット切換えと誤認識する可能性が高くなるので、そのトレードオフを考慮する必要がある。漸次ショット切り換えの検出率を上げようとすると誤検出率が高くなる。

ワイプに代表される空間的ショット切換えとカメラや被写体の動きを完全に区別することは画像の内容を計算機が理解しない限り不可能である。例えばカメラマンが屋内から屋外へ移動しながら撮影した映像を考えてみる。映像の中に写っている“ドア”を認識できなければ、それが空間的なショット切換えか単なるカメラや被写体の移動によるものなのかを完全に区別することはできない。しかし、提案法はカメラや被写体の動きが連続的かつ滑らかであることを仮定して現実的な解を与えている。

リアルタイム処理を実現するためには、計算機の処理能力の制限から映像データの空間解像度あるいは時間解像度を落として処理を行う必要がある。時間解像度を下げると、被写体の動きをショット切換えと誤認識したり、短い間隔で発生するショット切換えを見落としがちになるので好ましくない。従って、我々は空間解像度を下げる方法でリアルタイム処理を実現した。単純間引きにより空間解像度を落とすと、細かい絵柄をもつショットでエイリアシングに起因する、ちらつきが原因で誤検出を生じることがある。しかし、空間解像度を落とす際にブロック平均を用いることでこの問題を回避できる。

5. ショット切換え情報を用いた映像アクセスインタフェース

ショット切換え情報の応用として、従来から映像編集[11]や映像サマリー生成[3]等、さまざまなものが

提案されている。本章では、ショット切換え情報を利用した二つの映像アクセスインタフェースを提案する。

ビデオは、時間軸の取扱いの難しさからブラウジングが困難であるという欠点をもつ。例えば、ビデオテープの本数が増えてくると、どのテープに何が入っていたかわからなくなり必要な映像を探し出すのに時間と手間がかかる。

5.1 PaperVideo

PaperVideo[10]は時間軸上に展開する映像の中からショットごとに代表的な画像を抽出して、図6のような一覧表の形で映像の内容と流れを紙の上に静止画表現したものである。映像は一つのメッセージを表現するショットを時間的につなぎ合わせることでより複雑なストーリーを表現するものであるから、このような代表画像の抽出方法は合理的である。代表画像を抽出する最も単純な方法は等時間間隔に画像を抽出するものであるが、同じような画像が連続して抽出されたり、重要な場面が欠落することがあるという欠点がある。

図6の一覧表を参照することによって一目で映像の概略を把握でき、代表画像の下に付記されている時間情報を手掛かりにして、オリジナルビデオに効率的にアクセスすることができる。また、ディスプレイの上ではなく紙の上に表示したことにより、紙が元来もつ良い性質(持ち運びが容易である、書き込みが自由に行える、解像度が高い)を生かしたインタフェースとなっている。

5.2 TV Ram

テレビ放送に代表されるリアルタイム映像は新鮮で多様な情報を伝える情報源である。TV Ram(TV program Real-time Access Monitor)はテレビ放送を一定時間分(例えば、24時間分)プールし、そのプールされた映像に対して効率的なアクセスを実現するシステムである[9]。

リアルタイム映像に対する現在一般的なアクセスインタフェースとして、(1)リアルタイム視聴(テレビ)、(2)録画予約(ビデオデッキ)、の2通りがある。(1)は時間に束縛され、まき戻しすらできないという問題点がある。(2)は(1)の時間的束縛の問題点を解決するものであるが、計画的に録画したものしか見られないという問題点がある。我々がテレビ等のリアルタイム情報源にアクセスする動機は、明確な目的意識に基づくものだけではなく、「何か面白い情報が飛び込んでくるかもしれない」という「情報との偶然の出合いに対する期待感」からであることが多い。後者のよ

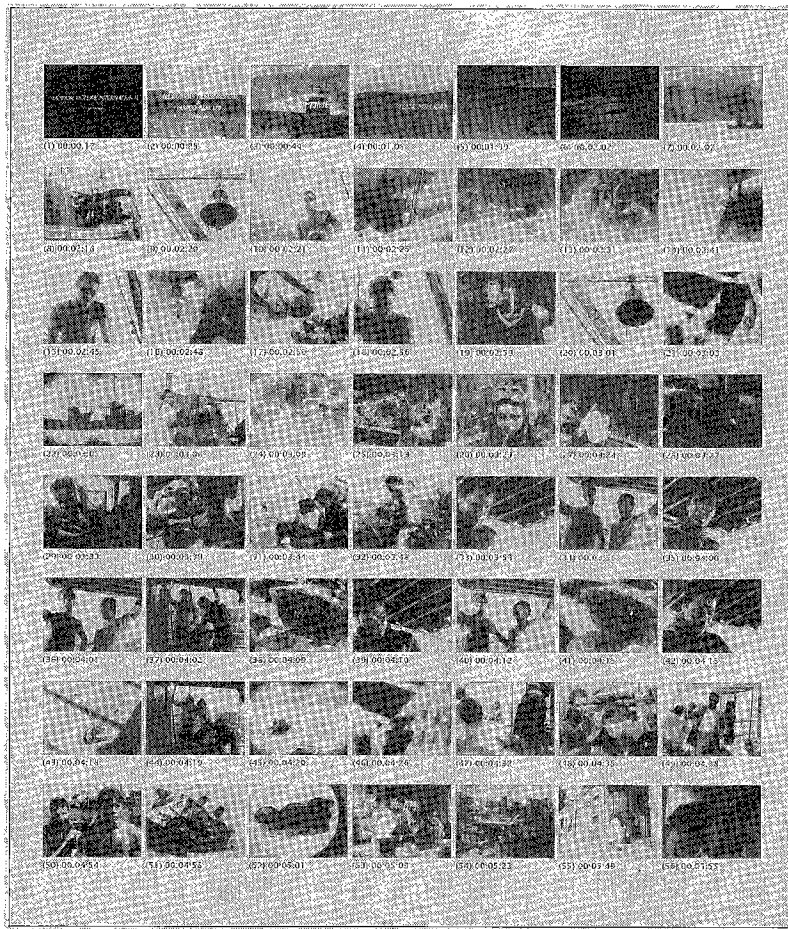


図6 PaperVideo インタフェース

Fig. 6 PaperVideo interface.

うに不明確な目的意識で情報源にアクセスしている状態をグレージング (grazing) と呼ぶ [5]。従ってアクセスインタフェースは、ブラウジングやハンティングといった大なり小なり目的意識のある検索のサポートだけでなく、ユーザをブラウジングの状態に導くためのグレージングのサポートも重要である。TV Ram は長時間の映像情報に対して、グレージング、ブラウジング、鑑賞の三段階を連続的にサポートするアクセスインタフェースを提供する。

処理概要について述べる。テレビ放送を受信しながら、リアルタイムにショット切換えを検出し、ショットの先頭とそこから1秒おきの画像をJPEG圧縮しハードディスクに保存する。音声は連続的に音声ファイルとして保存する。MPEG1で圧縮された24時間分の動画が約16Gバイトの大きさをもつのにに対して、この時

間間引きによって24時間分が約1.6Gバイトという小さな容量で管理可能になる。このような処理は、リアルタイムにショット切換え検出が行えることで初めて可能になるものである。

図7に24時間分のテレビ番組の一覧インタフェースを示す。1時間あたり6枚のインデックス画像をテレビ番組表のように時間順に表示している。テレビ番組は放送局側でスケジューリングされて送出されているので、放送時間を可視化するような一覧インタフェースがグレージングやブラウジングにとって有効である。インデックス画像がコマーシャルの中から抽出されると一覧効果が損なわれるので、それを避けるためにコマーシャル部分ではショット切換えの発生頻度が高いという性質を利用してコマーシャルをインデックス画像として抽出しないようにする。インデックス

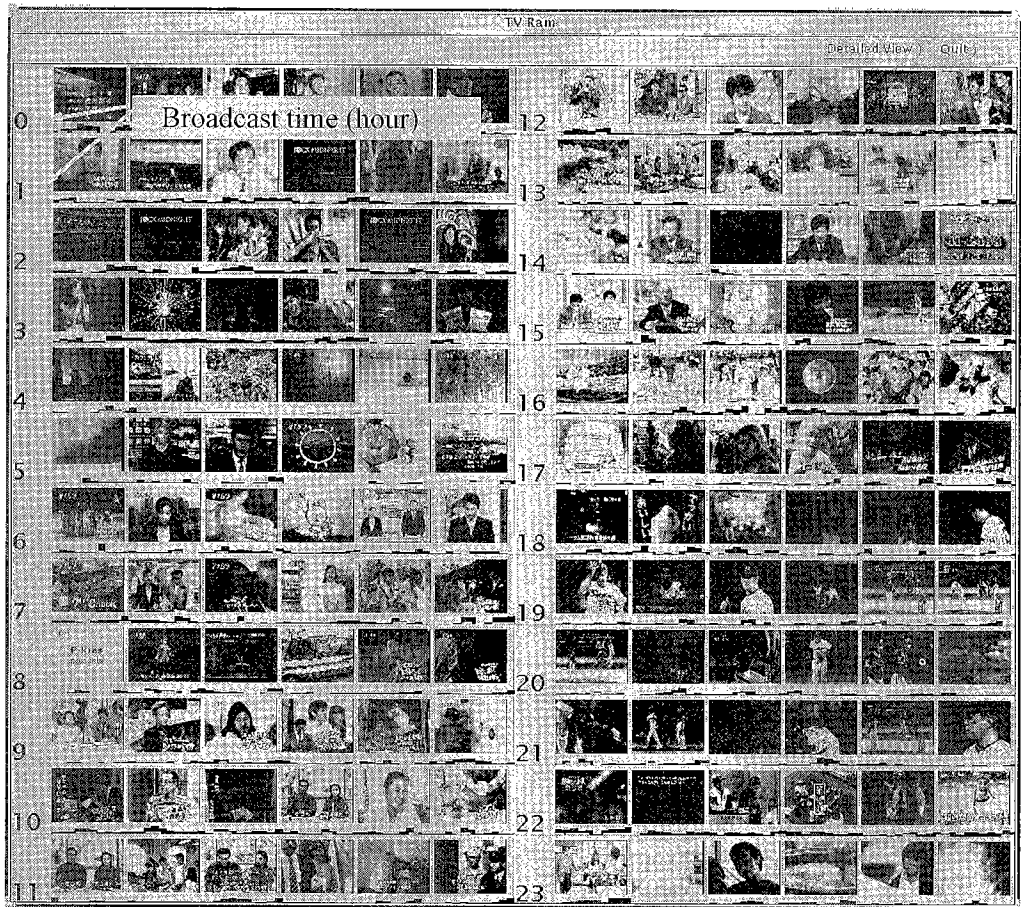


図7 TVRamにおける24時間一覧表示インタフェース
Fig. 7 24 hour overview interface in the TVRam system.

画像をマウスでクリックすると間引き映像が連続的な音声と共に紙芝居的に再生される。この紙芝居的な再生は鑑賞には耐えないが、映像の概略を把握する目的には十分である。この一覧インタフェースによって、面白そうな番組を発見したり、その番組をつまみ見しながら見る見ないの選択が効率的かつ直感的に行えるようになった。24時間の一覧インタフェースの他にも、より細かい単位でブラウジングを可能にするために、ショット切換えに対応する画像をPaperVideoのように一覧表示するインタフェースも備えている。本システムをビデオサーバのインタフェースとして組み込むことで、大量の映像へのアクセスを効率化できる。

6. む す び

今後、マルチメディア情報の量が増加するに従って、

人が情報を素早く取捨選択しアクセスできるようにするインタフェースの重要性が増してくる。その時代に向けて、マルチメディア情報のもつ構造を自動抽出する技術がより重要性を増してくると考えている。

本論文では、複数スケールで算出された非類似度に基づき各種ショット切換えをリアルタイムに検出する方法を提案した。本手法によれば、ニュース映像に対して90%以上の検出率が得られた。更に、ショット切換え情報を用いたアプリケーションの例として、映像の効率的な概略把握を可能にするPaperVideoと、リアルタイム映像へのアクセスインタフェースTV Ramを実現した。

今後の課題は、ショット切換え以外の情報（カメラ操作情報、音声に関する情報等）も利用して新しい映像アクセスインタフェースを実現することである。

謝辞 日ごろから御指導頂くヒューマンインタフェース研究所、映像処理研究部の岸本登美夫前部長、鈴木元、元グループリーダーに感謝いたします。有益な議論を頂いた阿久津明人氏、大辻清太氏をはじめとするグループの皆様に感謝いたします

文 献

- [1] F. Arman, A. Hsu, and M.-Y. Chiu, "Image processing on compressed data for large video databases," Proc. ACM Multimedia'93, ACM Press, pp. 267-272.
- [2] A. Hampapur, R. Jain, and T. E. Weymouth, "Production model based digital video segmentation," Multimedia Tools and Applications, vol.1, no.1, pp. 9-46, 1995.
- [3] 菊川 健, 川淵里美, "映像資料サマリー自動作成システムの開発," 信学論(A), vol.J75-A, no.2, pp. 204-212, Feb. 1992.
- [4] 栗田 稔, 高橋 秀, 竹下強一, "画像のソフトウェア," コロナ社, 1980.
- [5] S. Lambert and S. Ropiequet, eds., "CD-ROM: new papyrus," Microsoft, 1986.
- [6] 宮原 誠, "テレビジョン信号帯域圧縮を対象としたフレーム差信号特性の実測と検討," NHK技術研究, vol.29, no.1, pp.33-49, 1977.
- [7] 長坂晃朗, 田中 譲, "カラービデオ映像における自動索引付け法と物体検索法," 情処学論, vol.33, no.4, pp. 543-550, 1992.
- [8] 大辻清太, 外村佳伸, 大庭有二, "突出検出フィルタを用いた映像カット点検出法," 信学論(D-II), vol.J77-D-II, no.3, pp. 519-528, March 1994.
- [9] Y. Taniguchi, A. Akutsu, Y. Tonomura, and H. Hamada, "An intuitive and efficient access interface to real-time incoming video based on automatic indexing," Proc. ACM Multimedia'95, pp.25-33, ACM Press, 1995.
- [10] 外村佳伸, 谷口行信, 阿久津明人, "PaperVideo: 紙を用いた新しい映像インタフェース," 信学技報, IE94-59, 1994.
- [11] 上田博唯, 宮武孝文, 吉澤 聡, "認識技術を応用した対話型映像編集方式の提案," 信学論(D-II), vol.J75-D-II, no.2, pp. 216-225, Feb. 1992.
- [12] H.-J. Zhang, A. Kankanhalli, and S. W. Smoilar, "Automatic Partitioning of full-motion video," Multimedia Systems, vol.1, no.1, pp. 10-28, 1993.

(平成7年8月15日受付, 11月30日再受付)



谷口 行信 (正員)

平2東大・工・計数卒。平4同大学大学院工学系研究科修士課程了。現在NTTヒューマンインタフェース研究所勤務。映像処理の研究に従事。コンピュータグラフィックス、画像処理などに興味をもつ。



外村 佳伸 (正員)

昭56京都大学大学院修士課程了。同年当時の日本電信電話公社に入社。以来画像を中心としたメディア変換装置の研究・開発、映像ハンドリングの技術研究に従事。昭62~63米国MITメディア研究所、客員研究員。現在NTTヒューマンインタフェース研究所映像処理研究部 主幹研究員。



浜田 洋 (正員)

昭53電通大・電通・通信卒。昭55同大学院通信工学専攻修士課程了。同年日本電信電話公社(現NTT)入社。以来、横須賀電気通信研究所にて音声認識、音声合成、ヒューマンファクタの研究実用化に従事。工博。現在、NTTヒューマンインタフェース研究所主幹研究員。