ヒューマンインタフェース 72－7
（1997. 5. 16）

映像解析技術を利用した映像メディアのためのGUI

堀 修　青木 恒　窪田 進　金子 敏充
(株) 東芝 研究開発センター 情報・通信システム研究所
〒210 川崎市幸区小向東芝町1
{hori,aoki,skubota,kaneko}@eel.rdc.toshiba.co.jp

映像解析技術を利用することにより映像編集の経験のない人でも容易に映像メディアを操作できるGUIを提供する．本システムの特徴は，カット検出（シーンが突然変化するところ），顔検出，カメラの動き推定などの映像解析を行い，その情報に基づいた静止画の見出しを3次元の直方体で表した映像メタファの上につけて検索しやすくした点にある．また，操作が直観的に行えるように映像の部分的な切り出し位置をインターラクティブに見ながら確認できるようにした．これにより，ユーザがフレーム番号やタイムスタンプを全く気にすることなく，あたかも紙の文書を切り貼りする感覚で映像を操作することができるようになった．その他にもパンした映像からパノラマ画像を生成するなどハイパーメディアに有用な素材を作成する機能がある．

GUI for Digital Video Media using Video Analysis Technology

O. Hori, H.Aoki, S. Kubota, and T. Kaneko
R & D Center, TOSHIBA Corporation
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki, Kanagawa 210, Japan

We propose a video editing GUI employing video analysis technology for laymen. This system represents a video frame sequence as a long rectangular parallelepiped and it also indexes the video by extracting scene change points, extracting human faces, and estimating camera motion parameters such as panning and zooming for easy retrieval. It allows laymen to intuitively handle a video sequence without knowing frame numbers or time stamps by giving interactive frame viewing environment to check start and end points of video clipping. In addition, it has a function to generate a panorama image from a panning video sequence.

1 はじめに

計算機の急激な進歩により画像だけでなく映像を操作することが容易になりつつある．それにともなって，文字，図，ＣＧ，画像，映像等のメディアどうしをリンクし，インターラクティブに操作するハイパーメディアが実現できるようになり，いままで，計算機能力の低さから十分に普及していなかった映像メディアがハイパーメディアの重要なコンテンツの一部として活躍できる環境が整ってきた．パソコンにおいても画像処理を高速に行えるＣＰＵが提供しはじめられ，特殊なハードウェアを用いずに動画を再生できるようになってきた．また，ハードディスクが大容量化すると同時に価格も下がり映像を保存し作業できるようになってきている．さらに，リムーバブルな大容量媒体としてＣＤ－Ｒ（約700MB）が登場し，近い将来にはＤＶＤ－ＲＡＭ（約2GB）などの登場でデジタル映像を容易に格納できるようになりつつある．ネットワーク回線においても１００ＭＢＰＳの高速回線が一般に使えるようになってきた．このような環境下においては，従来のテキストメディアだけでなく映像を含んだハイパーメディア等の，より表現力の高いメディアへの移行が進むと予想される．テキスト，画像，ＣＧなどメディアを編集・加工する技術は一般の人でも比較的簡単に操作する環境が提供されているものの，デジタル映像メディアの編集・加工操作はノンリニア編集と呼ばれ，プロ向けの編集システムが市販されているだけである．しかも，その操作は，一般の人が操作するには敷居が高い．また，市販されているノンリニア編集は映像を一本のストーリーを作成するための環境を与えているが，ハイパーメディアの場合は映像も全体のメディアの一部であり，必ずしも一本のストーリーに編集する必要はない．映像をひとつの素材として貼り込み，他のメディアとリンクを張ることにより，比較的簡単にインターラクティブに鑑賞する作品を作ることができ一般の人でも映像コンテンツを作りやすい．ハイパーメディア用の映像素材を作成する場合は，映像を繋合わせる処理よりも，取材した映像から必要な部分を切り出す作業が主たる操作となる．よって，いかに自分が必要とする場所をすばやく捜し出し，必要な領域を切り出せるかが重要となる．筆者らは，これまでに編集済みの映像を解析し，その情報を利用した映像ブラウジングシステム [1, 2, 3] を開発してきた．今回，映像編集操作を支援するために，映像解析技術を用いて編集前の映像に自動的に見出しをつけることによって，ユーザが容易に所望の場所を探せるシステムを開発した．第2章では，本システムの設計方針について述べる．第3章では，本システムの特徴となる映像解析を用いた見出し付け機能について述べる．第4章では，設計方針に基づいて，見出しを利用した映像メディア編集のためのＧＵＩ機能について述べる．

2 ＧＵＩの基本設計方針

これまでにも，映像解析技術を応用した対話型映像編集方式が提案されている [4][5]．これらは，映像の解析結果に従って最初から映像を分割しているため，本来の映像が連続メディアであることが希薄になっている．本システムでは，解析結果を表現する際においても映像の連続性を損なわないようにし，映像上の事象の起こる順序や時間的位置関係が直感的にわかるようにＧＵＩを設計した．そして，ユーザが所望の映像の場所を探しだし，その一部を容易に切り出せることを重要な操作とした．基本設計方針して，1) 所望の映像が記録されている場所がだいたいどのあたりにあるかを予測できるように映像の連続性を損なわないように見出しを付ける．2) 所望の映像が記録されている場所の見当をつけた後は，その場所の内容を簡単に外観できるようにインターラクティブに映像を再生できるようにする．しかし，詳細に内容を見たい場合はその部分を切り出し，音声を含めて完全な再生を行えるようにする．3) ユーザにフレーム番号やタイムスタンプを意識させないで編集ができるようにする．4) 操作が直観的に行えるようにＧＵＩにおいてドラッグアンドプレーの操作性を取り入れ，編集に際して操作をモードレスにする．以上の基本設計方針に基づいて，図1に示すような環境で試作システムを作成した．

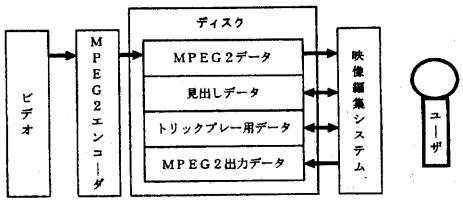

図 1: 試作システムの環境

ＭＰＥＧが将来主流の映像形態と考え映像データとして採用した．しかし，今回は，試作システムということで一般ユーザに馴染みのあるパソコンではなく開発効率を優先しＵＮＩＸベースのＥＷＳの上に構築した．将来的には，性能が向上すればパソコン上に構築できると考えている．ＭＥＰＥＧ２映

像データを映像解析を行って見出しデータとＧＵＩ機能におけるトリックプレーのためのデータをあらかじめ生成し，それらとＭＰＥＧ２データを用いて編集するようにした．編集された映像はＭＰＥＧ２映像データとして出力され格納される．

3　映像解析による自動見出し付け

この章では，映像に見出しを自動的につける方法について述べる．自分で撮影した映像であってもどの部分にどのような映像を撮影したかを知るために，従来のＶＴＲの早送り機能を使って探すことはたいへんであった．ランダムアクセス可能なデジタル映像においても，再生機能しかなければその機能を頼りに適当に探すしかない．もし，映像の適切な場所に静止画の見出しがついていれば所望の映像を探す出すことは比較的容易である．そこで，検索に有用な見出しを計算機が自動的に作成すればユーザの負担は少なくなる．今までに，最も多い見出しの付け方はシーンの変わるところを自動的に検出する方法（カット検出）[6]-[12]で，生素材の映像の場合はビデオカメラのスイッチをＯＮ／ＯＦＦした場所に該当する．その場所にひとつのフレームの静止画を見出しとしてつけることにより自分が撮影した場所ごとに映像を分割できることになる．筆者らはさらに分割されたシーンごとに映像に対して撮影者の意図を代表する見出しを作成することを試みた．その一つがシーンに人がいるかどうかを自動的に判別することであり，もう一つは，パン，ズーム，静止などのカメラの動きを推定しそのカメラの動き（ユーザのカメラの操作）ごとに映像を分割し見出しを付けることである．以下に，それぞれの見出しの検出方法について述べる．

3.1　カット検出

本システムで用いたカット検出手法は，画像間の類似度の定義としては，ＭＰＥＧにより圧縮符号化された映像データ中に含まれる動きベクトルデータの符号量を利用している．動きベクトルの符号量とカットとの関係を考察すると，1) カットがない場合には画像間の予測符号化が使われ，しかも隣り合うブロックは類似した動きベクトルを有するため，動きベクトルの符号量は小さくなる，2) カットにより全く異なる画像に変わると，カットを越えた画像間の予測が使われず，動きベクトルが符号化されなくなる，3) カット後も同じ様な色調であるためにカットを越えて画像間予測が行われる場合には，動きベクトルは不揃いになるため，動きベクトルの符号量は多くなる，という関係がある．従って動きベクトルの符号量と画像間の類似度とは反比例の関係にある．よって，ＭＰＥＧ２映像を復号することなく動きベクトルの符号量のみをみることで高速にカット検出を行うことができる．アルゴリズムの詳細については文献 [13] を参照されたい．

3.2　人物検出

映像の中に人物がいるかどうかは有用な情報である．特に，一般の人が趣味で映像を撮る場合は身近な人物を撮影しその部分を検索したいというニーズが高い．カット検出された映像に対して人物がいるかどうかを人物の正面顔があるかどうかを検出することによって判断する．ここで，問題になるのは従来からよく行われている顔の検出の研究が固定のカメラによる背景固定である場合が多いのに対して，今回扱う必要のある画像が任意の背景を持つことである．

映像データはデジタル映像としてフレームに対して以下の処理が行なわれる．実際に処理されるフレームは予め選択されたものを用いる．例えば，すべてのフレームを処理すると時間がかかるのでシーンチェンジの起こった最初のフレームだけを処理対象とする．フレーム画像に対して顔の存在する候補を抽出する．顔のサンプル画像から学習した顔の辞書である固有顔 [14] を求め，画像のスケールを変えテンプレートマッチングを行う．類似度が大きいところを顔の存在する候補とする．しかし，その候補に対して真顔（本当の顔）の他に多くの疑似顔（顔に似ている背景の一部）が発生する問題がある．その問題を解決するために，疑似顔をサンプルを集め真の顔のサンプルの間で２クラスを分類するための判別分析 [15] を行い候補から真の顔のみを検出する．

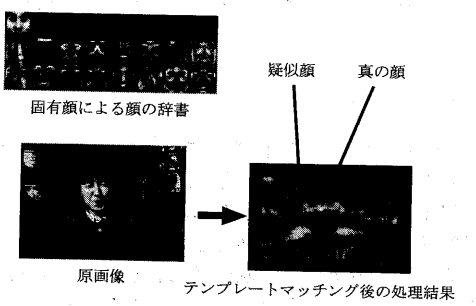

図 2: 顔検出の処理例

3.3 カメラの動き推定による映像の分割

カメラの動きの推定は，映像における隣合うフレーム間の画像から求められるオプティカルフローを用いて行う．求められたオプティカルフローに計算誤りがあったり，動いている物体があるため，それらを考慮してカメラの動きを推定する必要がある．カメラの動きに基づくオプティカルフローは背景から生じる．筆者らは，ハフ変換を用いて高ロバストなカメラの動き推定を行う手法を開発した [16]．求められたオプティカルフローから FOE(Focus of Expansion) の候補をハフ空間に投票し求め大局的なカメラの動きを推定する．これをもとに，背景と移動物体から来るオプティカルフローを分離することができ，背景からくるオプティカルフローからカメラの動きとその速度を計算することが可能となる．求められたカメラの動きからパン，ズームなどの開始点から終点をもとめ映像を分割するための情報とする．

4 映像メディア操作のためのＧＵＩ

従来は，映像を編集する場合映像信号として保存されたＶＴＲテープを早送り，再生，巻き戻し機能を組合せ必要な場所を捜し出し，テープカウンターやタイムスタンプを記録しながら所望の場所を探すという作業をしていた．この場合，テープのどの部分にどんな映像を撮影したかを撮影時に記録しておく必要がある．しかし，一般の人が趣味用に撮影を行う場合そのような手間をかけることは考えにくい．

そこで，計算機上でデジタル映像を編集するためのノンリニア編集システムが研究され，最近多くの実用システムが市販され始めた．これらのシステムでは映像をシーンチェンジ（カット）の部分で分割した単位で表現されている．しかし，これでは，映像が本来連続したメディアであるという表現が希薄になる．

4.1 映像メタファ

本システムで，映像のメタファを連続したフレームの塊を横に長く横たわる直方体として表し，映像が撮影された順に左手前から右奥へフレームを重ねて伸ばした３次元の物体とした．その直方体にシーンチェンジがあった位置に切れ目を入れ，分割した単位に代表フレームを表示する．この表示によって，どのような事象がどの順でどの程度の長さで撮影されたかが一見してわかる．また,『人が映っている』とか『カメラがパンしている』などの情報を直方体の側面に表示することによって撮影された映像に付随する情報を与え，ユーザに映像の内容を想起させる．例えば，Ａさんが映っていた場所でズームした場面が欲しい場所には，この映像メタファから容易に見つけ出すことができる．３次元映像メタファは拡大縮小／回転ができ必要な部分を詳細に見ることが出来る．ただし，見出しは常に正面方向を向いて表示されるようになっている．分割された単位の見出しをドラッグすることで，別の直方体の映像メタファとして分離することができカメラの動きごとに分割することもできる.(図3参照)

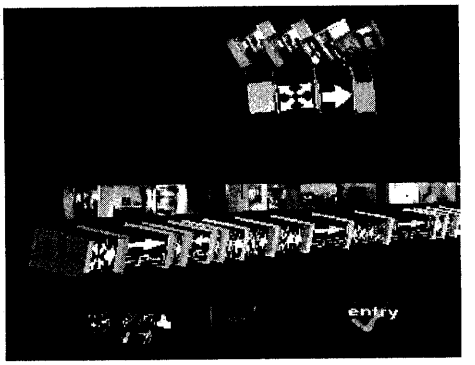

図 3: 映像メタファ．映像の一部を切り出した例.

4.2 インターラクティブムービングアイコン

所望の場所をだいたい探し当てた後は，実際に映像の中身を動画として再生／確認したい場合が生じる．その場合は，直方体の上部をマウスでドラッグすることにより指し示すフレームの画像を小さなアイコンとして表示することができる．このアイコンをインターラクティブムービングアイコンと呼ぶことにする．マウスをドラッグすることでアイコンが動画となり任意の速さで疑似的な再生／早送り再生／逆回し再生ができ，映像の内容をインターラクティブに確認できる．従来のＶＴＲのシャトルと異なり，再生位置へのランダムなアクセスと局所的な可変再生が同時に一つの操作できる点が特徴である．(図4参照)

4.3 インターラクティブクリッピング

映像の所望の部分を確認できた後は，その部分を切り出す作業を行う必要がある．従来は，フレーム番号かタイムスタンプを頼りに切り出しの位置を決めていたが，このシステムでは，前述のインターラ

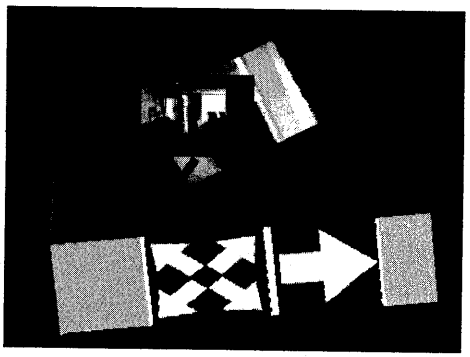

図 4: インターラクティブムービングアイコンの例.

クティブムービングアイコンを利用することによって，それらのコードを意識することなく切り出しができる．図5に示すように，インターラクティブムービングアイコンを利用し，左ボタンを押下することにより切り出しの先頭位置を決定する．次に，そのままマウスをドラッグしながら，先頭位置を示すアイコンの下に表示されるインターラクティブムービングアイコンを見て，終端位置をマウスの左ボタンをリリースすることによって決定する．切り出した領域は，別の直方体の映像メタファとして自動生成される．

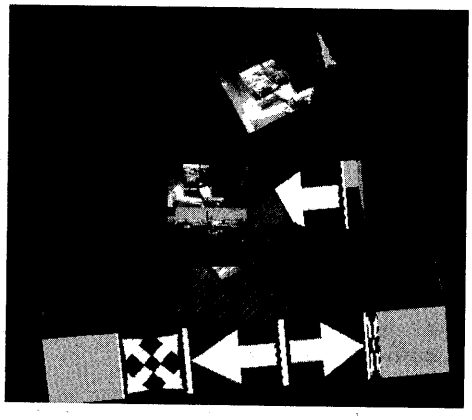

図 5: インターラクティブムービングアイコンを利用して所望の映像領域を切り出した例.

4.4 3次元空間におけるオブジェクト指向インターフェース

すべての操作がオブジェクト指向のインターフェースとなっている．映像そのものが3次元の直方体として表現され，クリッピングされた映像の一部も同様の直方体の塊として表現される．その直方体の映像メタファをアイコンの上にドラッグ・アンド・ドロップすることで再生や加工を行える．例えば，切り出した部分をMPEG2再生をしたい時は，映像メタファをテレビの形をしたアイコンの上にドロップする．また，切り出した部分を保存したい場合は“ENTRY”と書いたアイコンの上にドロップすることによって登録を行う．このように，全ての操作がマウスだけでほとんど出来るように設計されている．

4.5 動画の変化に即した可変速早回し機能

以上述べてきたように，見出し，インターラクティブムービングアイコン，MPEG2再生を用いることによって迅速な映像の内容把握が可能である．しかし，見出しではわからないようなすべての事象をもれなく見るためには，全体を通して再生することが必要な場合がある．そのため，全体を通して高速再生する可変早回し機能がある．従来の早回し機能は可変であっても人が制御する必要があったが，今回提供する可変速早回し機能は動画像の変化に即して自動的に早回し再生速度が制御される．前章で述べたカメラの動き推定で得られた情報から動画像の変化の度合が計算される．早回しする速度は動画像があまり変化していないところは早めに再生し，変化の大きいところは遅めに再生することにより再生表示画面の動画像の変化率をほぼ一定にすることで，高速でかつ見やすい早回し再生表示が可能となる．ただし，不自然な表示を避けるには表示速度がなめらかに変化するように制御したり，静止画が続くところが瞬時に終わらないように非線形な制御を行っている．

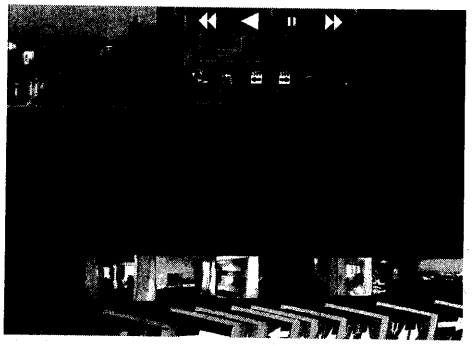

図 6: 動画像の変化に対応して可変速で再生表示するビューワー.

4.6 映像からのパノラマ画像生成

レンズの取り換えができない家庭用のビデオカメラを用いて広い領域を撮影する場合，しばしばパンをして撮影を行う．ユーザの気持ちとしてはパノラマ画像を撮影したいという場合がある．ハイパーメディアの素材としてパノラマ画像をビデオで撮影できれば素材の収集として魅力的である．将来は映像からパノラマ画像のような2次元画像及び3次元物体を再構成しハイパーメディアの素材にするということも考えられる．今回のシステムではパンした映像からパノラマ画像を作成を試みた．その場合，映像を精密に張り合わせるために高精度なオプティカルフローを算出する必要がある．筆者らは，高精度なオプティカルフローを計算する方法を開発し [17]，それを用いて映像のフレーム間を画素レベルで位置合わせして足し合わせることによりパノラマ画像を生成した．図7に示すように，時間軸方向の映像情報を空間的な広がりを持つパノラマ静止画像に変換できる．

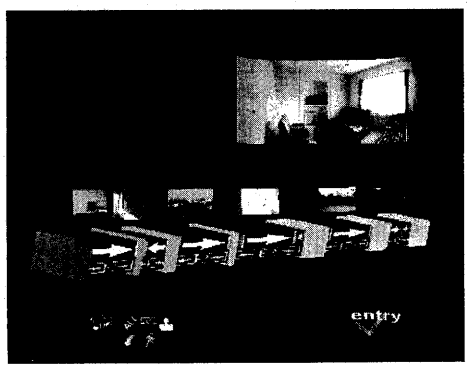

図 7: パノラマ画像の生成の例．

5 おわりに

本稿では，映像メディアの編集の経験のない人でも直感的に操作できる映像解析結果を利用したＧＵＩを提案した．従来のノンリニア編集システムがフレーム単位の編集を容易にしているのに対して，提案したＧＵＩではフレームを重ねて連続性を失わない3次元の直方体表現を映像メタファとして用い，見出しの静止画像をつけたり，側面に映像情報（カメラの動きや人物の有無）を表示することにより，映像の内容をすばやく把握できるようにすることによって映像を編集しやすくした．その結果，ユーザからフレーム番号やタイムスタンプを隠蔽し，映像を文書の記事を切り取るような感覚で編集できるようになった．今後は，実際にユーザ使ってもらいシステムの評価を行うことでＧＵＩの改善をはかっていく．

参考文献

[1] 青木, 下辻, 堀,"映像ブラウジングのための類似ショット統合," 情処研報,96-HI-67,1996.

[2] H.Aoki, S.Shimotsuji, O.Hori,"A Shot Classification Method of Selecting Effective Key-Frames for Video Browsing," Proceedings of ACM Multimedia 96, pp.1-10,1996.

[3] 青木, 堀,"映像構造を利用した代表フレーム表示," 情報処理学会 インタラクション'97,pp.9-16,1997.

[4] 上田, 宮武, 吉澤," 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, 1994.

[5] 外村, 大辻, 阿久津, 大庭,"蓄積映像ハンドリング技術," NTT R&D,pp.61-70,1993.

[6] 大辻, 外村, 大庭, "輝度情報を使った動画像ブラウジング, " 信学技報 IE90-103, 1991.

[7] 菊川, 川淵, "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, 1992.

[8] 大辻, 外村, 大庭, "動画カット検出," 信学技報 IE91-116, 1992.

[9] 外村, 安部, "動画像データベースハンドリングに関する検討," 信学技報 IE89-33, 1989.

[10] 長坂, 田中, "カラービデオ映像における自動索引付け法と物体探索法," 情報処理学会論文誌, 4, pp.543-550, 1992.

[11] 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.

[12] 大辻, 外村, 大庭, "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, 1994.

[13] 金子, 堀,"動きベクトル符号量を用いたＭＰＥＧ動画像からの高速カット検出," 信学技法 PRMU96-100, pp.55-62, 1996.

[14] M. Turk and A. Pentland, "Eigen faces for Recognition," Journal of Cognitive Neuroscience, vol.3, no.8, pp.71-86, 1991.

[15] K. Fukunaga, "Introduction to Statistical Pattern Recognition, Second Edition," Academic Press, pp.445–460,1990.

[16] 窪田, 下辻 ,"ハフ変換を用いたカメラパラメータの推定及び動画像からの移動物体の分離," MIRU96-II, pp.121-126,1996.

[17] 窪田, 堀,"定量的信頼度を伴うオプティカルフローの計算法," 信学技法 PRMU97-102,1997.