

## リアルタイムTVインデクシング

青木 恒 堀 修

(株)東芝 研究開発センター 情報・通信システム研究所

〒210-8582 川崎市幸区小向東芝町1 / Tel: 044-549-2288 / Fax: 044-520-1308

hisashi.aoki@toshiba.co.jp / osamu.hori@toshiba.co.jp

## 1. はじめに

テレビ放送は過去40年にわたり、そして現在も家庭にとって最先端の情報窓口である。放送局数も徐々に増加し、近年では衛星デジタル放送の実用化に伴って数百チャンネルの放送番組を見ることができ、一方、テレビ放送から得られる映像や音声を認識する技術も進歩しており、カット検出を応用して内容一覧を作成したり<sup>[1]</sup>、音声認識を応用してニュース番組のトピック分離を行う試みがなされている<sup>[2]</sup>。既に処理済みの番組であれば、自然言語入力に興味あるニュース場面を検索できるシステムも試作されている<sup>[3]</sup>。

しかし、これまでの解析技術は高性能の計算機によって実現するものであったり、処理のために数十分～数時間を要するものであることが多く、家庭にこれらのメディア解析技術が導入されるためには「廉価な機械でも」「待たずに」反応できるプラットフォームの開発が不可欠である。

筆者らは、家庭用コンピュータの性能で、かつ映像入力と同時に映像解析を行う（すなわち、受信完了時に処理が終了している）アプリケーションを開発した。このシステムでは、これまで普及してきたVTRの機能、操作性を踏襲した上で、筆者らがすでに開発した類似ショット検出アルゴリズム<sup>[4]</sup>を用いて、少ないアイコン数の縮約画面で録画済み番組内容を確認したり、あらかじめ登録された画面の登場を待って録画を開始させることができるなど、映像解析技術を応用した新しい付加機能が搭載されている。

今回の展示では、上述のリアルタイムTVインデクシングシステムの機能と処理手法を説明する。

## 2. プラットフォームの設計

図1にプログラムの処理フローを示す。入力部ではキャプチャ映像、AVI、MPEGファイルからの入力に対する処理がオブジェクト化されており、

将来新しい映像形式が登場した場合にも、それを読み込むオブジェクトを追加するだけで済むように設計されている。動作中、この入力部は後段の解析部での解析処理が終わり次第、新しいフレームを1枚読み込み、解析部に送る。入力源がキャプチャ画面である場合には、後段の処理に遅れが生じないように、この入力部はフレーム入力をスキップする。したがって、解析に時間を要する低性能のコンピュータを用いた場合には、自動的に入力するフレームレートが低く調整される。

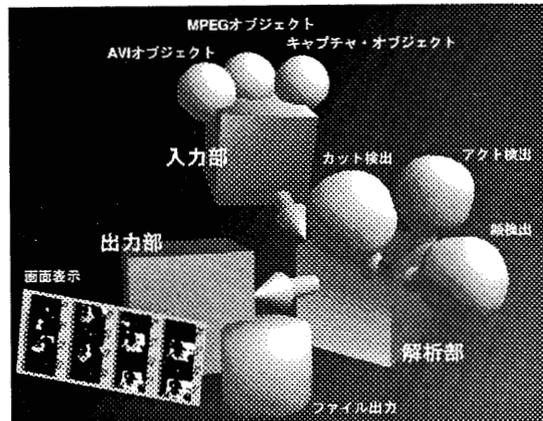


図1 プログラム内部の処理フロー

解析部では各種の解析エンジンをプラグインとして取り込むことが可能である。このため、アプリケーション本体をバージョンアップすることなく、新たな解析機能を追加することができる。プラグイン相互の協調動作も可能で、「カットがあったときに類似ショットを探索する」「カットがあったら顔の存在をチェックする」などの組み合わせがプラグインの製作だけで実現できる。プラグインは機能に応じたIDを持つルールになっており、アプリケーション本体は各プラグインの機能にしたがってコールの順番を並べ替えている。たとえばアクト検出プラグインは、カットが生じると同時に類似ショット検出結果をレポートするので、カット検出プラグインより後に呼び出されるようにアプリケーション本体が並べ替える。

出力部では処理結果を逐次表示するとともに、独自形式のファイルに書き出す。このファイルには構造化情報が含まれており、「録画」済みの映像ファイルは、ショット、あるいはアクト（「ショット」の上位構造、類似ショット統合によって

“Real Time TV Indexing,”

Hisashi AOKI and Osamu HORI,  
R&D Center, Toshiba Corporation,  
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki  
210-8582, Japan.

検出された数ショットからなる区間)の2階層で見ることができる。

### 3. 搭載プラグイン

筆者らは本システムのために現在のところ以下の4種類のプラグインを作成した。これらは必要に応じて機能をオン・オフすることができるので、性能の低いコンピュータで実行する場合などには機能を選択すれば速度を向上させることができる。

#### A. カット検出プラグイン

場面転換を検知し、アイコンを生成する最も基本的な構造化機能を実現する。入力源がMPEGである場合には、このプラグインは機能せず、入力部のMPEGオブジェクト内で高速カット検出を行う。

#### B. アクト検出プラグイン

カット検出によって検知された小場面(ショット)中で類似であるものを推定する。連続する2ショットが類似であると判定された場合にはカット検出を取り消す。たとえばフラッシュや大物体の横切りによってカットが誤検出されることがあるが、この機能によって自動的に修正される。2ショット以上前と類似であると判定された場合には、類似であったそれらのショットに挟まるショットは同じ「アクト」(=連続するショットのまとまり)に属すると推定し、ラベル付けする。たとえば対話のシーンなどにおいては2者のショットが交互に現れるので、対話区間を一つのアクトとして検出することができる。ショットごとのアイコンを表示するよりも、アクトのアイコンを表示した方がアイコン数は56~86%程度削減することができる。

#### C. 待ち受けプラグイン

フレームをこのプラグインに登録しておく、合致するフレームが出現した際に「ユーザに通知」「録画を開始」などのアクションを起こさせることができる。毎週決まった画面から始まる番組であれば、この機能を用いて録画(ファイルへの記録)を開始することができる。プログラムは常に過去数十秒の画像をバッファリングすることもできるので、登録画面が登場する何秒前、何ショット前といった録画開始を指示することもできる。

#### D. 顔検出プラグイン

各ショットの先頭において、肌色領域のまとまり方やサイズを元に、顔の存在を類推する。

### 4. 操作例

図2に、本アプリケーションの操作画面例を示す。「ショット検出」「待ち受け」を併用することにより、ショット先頭ではないフレームを待ち

受けキーに登録しても、ショット先頭から録画を開始することができる。これは従来のVTRにはなかった新しい「留守録」のスタイルとして受け入れられるものと期待している。また「待ち受け」機能では登録ショットの登場順を指定することができるので、ショット構成が既知であるCMなどに適用すれば「CMとばし」を観察するアプリケーションとしても用いることができる。

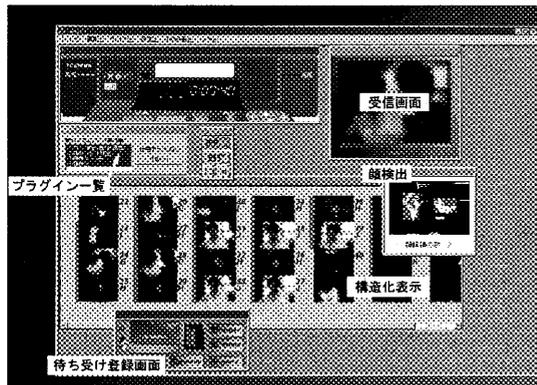


図2 リアルタイムTVインデクシングシステム

### 5. おわりに

筆者らはパーソナルコンピュータを利用した次世代VTRの姿として、マルチメディア解析技術を容易に導入、拡張できるリアルタイムTVインデクシングシステムを開発した。今後は、字幕認識や文字認識を利用したテキストインデクスの作成機能や、音声認識を利用した話者インデクス、スピーチ内容インデクスなどもプラグインとして搭載していく予定である。

### 参考文献

- [1] 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.J-75-D-II, no.2, pp.216-225(1992).
- [2] 西田, 有木「話者照合に基づく映像中の話者の自動索引付け」第3回知能情報メディアシンポジウム論文集, pp.55-60(1997).
- [3] 金出, 佐藤「Informedia:CMUデジタルビデオライブラリプロジェクト」情報処理 vol.37, no.9, pp.841-847(1996).
- [4] 青木, 堀「繰り返しショットの統合による階層化アイコンを用いたビデオ・インタフェース」情処学論, vol.39, no.3 (1998) (掲載予定)
- [5] 金子, 堀「動きベクトル符号量を用いたMPEG映像からの高速カット検出」, 信学会研究報告 PRMU96-100(1996).