

H-008

ショット毎の登場人物の把握に適した代表画像抽出・閲覧システムの構築

Development of Representative Frame Extraction / Video Browsing System Suitable for Grasping People Appearing in Each Shot

石川真澄† 大網亮磨†
Masumi Ishikawa Ryoma Oami

1. はじめに

地上デジタル放送開始に伴うテープレース化の進展や蓄積機器の大容量化を背景に、各放送局に膨大な素材映像が蓄積されるようになった。これらの素材映像から編集用区間の探索や不要になった映像の廃棄判断を効率的に行うために、映像内容を即座に把握したいという要望がある。映像を全て見ることなく素早く映像内容を把握可能にするために、代表画像を利用した映像閲覧方式が提案されている。

各ショットの先頭フレームを代表画像として提示する閲覧システム[1]では、各ショット内容を概観できるもののショット途中に出現する重要対象を把握できない。一方[2][3]では、映像の登場人物は内容把握において重要な役割を担うとして、各人物の顔画像を提示する。しかし、顔画像を切り出して提示するため登場場面を概観できない。そこで我々は、各ショットの登場人物と登場場面を即座に把握可能な代表画像選択方式を検討してきた[4]。

本稿では、[4]による登場人物および登場場面の把握効果を確認するために構築した代表画像抽出・閲覧システムについて述べる。本システムを用いて映像閲覧を行い、2種の代表画像による内容把握への有効性を報告する。

2. システム構成

本システムは、抽出システムと閲覧システムから成る。抽出システムは、ショットの先頭フレームと[4]に基づき決定したフレームを代表画像として選択する。閲覧システムは、代表画像による内容把握支援機能を提供する閲覧画面を生成する。本システムは、代表画像抽出端末、Webサーバ、クライアント端末の3つのPCにより実現される(いずれもCPU3GHz、メモリ1GB、Windows OSのPC)。システム構成を図1に示し、各PCの動作概要を述べる。

代表画像抽出端末

抽出システムが動作するPC。各映像に対し代表画像抽出処理を実行し、抽出した代表画像をWebサーバに格納する。抽出処理の各ステップで得た映像情報(映像名・映像長・生成日時・映像および代表画像のURL・処理の進捗情報)をデータベースに登録管理し、Webサーバに提供する。

Webサーバ

閲覧システムが動作するPC。代表画像抽出端末における処理の進捗状況を確認し、進捗に応じて表示可能となる3種の提示モードを制御する。クライアント端末を介して入力されるユーザの要求に応じて、代表画像抽出端末から取得した映像情報をもとに閲覧画面を生成し、クライアント端末に送信する。

クライアント端末

Webサーバの生成した閲覧画面をユーザに表示する。

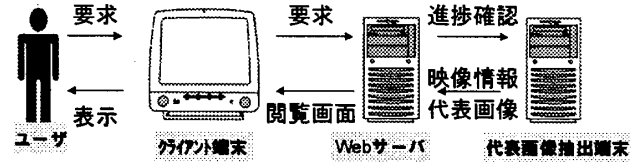


図1. システム構成

3. 抽出システム

抽出システムは、2種の代表画像を抽出し、処理の過程で得た映像情報をデータベースに登録する。代表画像抽出処理は、以下の3ステップで構成される。

ステップ1. 入力映像の検知

代表画像抽出端末はWebサーバ上での映像生成を定期的にチェックし、新規映像を検知すると作業フォルダにコピーして抽出処理を開始する。同時に、映像名、映像長、生成日時、映像のURL、ステップ1の完了フラグをデータベースに登録する。

ステップ2. ショットに基づく代表画像の生成

視覚的特徴量の軌跡の変化に基づくショット分割[5]を実行し、各ショットの先頭フレームを代表画像として選択する。代表画像をWebサーバに格納し、代表画像のURLとステップ2の完了フラグをデータベースに登録する。

ステップ3. 人物に基づく代表画像の生成

各フレームに対し顔検出と顔特徴量抽出を行い、顔特徴量の類似性と目位置距離を利用して、ショット単位で人物ごとに顔をグループ化する。各顔グループの代表的な顔を含むフレームから人物の重複しないフレームを、代表画像として選択する[4]。代表画像をWebサーバに格納し、代表画像のURLとステップ3の完了フラグをデータベースに登録する。

なお、ステップ3の途中で新規映像の入力が検知されると、ステップ3を中断して全新規映像に対しステップ2までを実行し、完了後にステップ3の中断フレームから処理を再開する。ステップ3と比較して処理の軽いステップ2を優先的に実行し、出来るだけ多くの映像からショットに基づく代表画像を抽出することで、早い時点で多数の映像を代表画像により閲覧できるようにする効果がある。

4. 閲覧システム

閲覧システムは、内容把握支援機能として、抽出処理の進捗に応じた代表画像の表示、各代表画像の出現位置の表示、映像の部分再生を実現する閲覧画面を生成する。

4.1 閲覧画面

閲覧画面を図2に示し、各機能の概要と効果を説明する。

(1) 代表画像の表示

各ショットの内容を把握可能な代表画像を表示する。1ショットから代表画像が複数枚選択される場合でもショット境界が明確であるように、ショット境界を赤縦線で表示

† 日本電気株式会社, NEC Corporation



図2. 閲覧画面

する。また、ユーザの操作を明示するため、ユーザがマウスオンした代表画像をハイライト表示する。代表画像を以下3種の提示モードで表示する。

A. ショットモード

ステップ2.で得た代表画像を表示する。これにより、各ショットの場面の概要を把握できる。また、代表画像を順に見ることで、ショット間のつながりを掴める。

B. 人物モード

ステップ3.で得た代表画像を表示する。これにより、ショット内の登場人物とその登場場面を把握できる。また、代表画像の有無から登場人物の有無を確認できる。

C. ショット+人物モード

人物が登場するショットでは[4]に基づき決定したフレーム、登場しないショットではショット頭のフレームを表示する。これにより、映像全体の流れを把握するとともに、人物の登場するショットを確認できる。

(2) 代表画像の出現位置の表示

各代表画像の出現位置をタイムバー上に表示し、ユーザがマウスオンした代表画像の出現位置と代表画像を含むショットを、タイムバー上にハイライト表示する。これにより、各代表画像に相当する場面が、映像のどこに出現するか把握できる。

(3) 部分再生

ユーザがクリックした代表画像の出現時刻の前後数秒を再生する。また、ユーザがタイムバー上でクリックしたショット区間を再生する。これにより、ユーザが選択した箇所について、部分的に具体的な内容を確認できる。

4.2 提示モードの制御

閲覧システムは、抽出処理の進捗に応じて選択可能な提示モードおよび映像再生の可否を制御する。これにより、ユーザが閲覧要求を出した際にその映像に対する代表画像抽出処理が途中であっても、一部情報の提示が可能になる。抽出処理の各ステップ完了時で表示可能な情報を表1にまとめる。

表1. 各処理の完了時で表示可能な情報

完了処理	表示可能な情報
なし	なし
ステップ1.	映像の再生
ステップ2.	ショットモードのみ、映像の再生
ステップ3.	提示モード3種全て、映像の再生

5. システムによる閲覧効果の確認

筆者らは、本システムを利用して1分映像40本を閲覧した。閲覧画面の各機能について確認された効果を示す。

(1) 代表画像の表示による効果

A. ショットモード

登場人物の有無に関係なく、映像全体の情景を大まかに把握できる点で有効であった。

B. 人物モード

インタビューなどのシーンで、内容把握において有用な人物がショット途中に現れる場合(例えば、ズームやパンによってショットの途中に人物が登場する場合、人物が顔向きを変えて正面向きの顔がショットの途中に存在する場合)に有効であった。また、登場人物の有無、代表画像に映った人物が目的の人物か否か、登場場面が編集に使えるか否かを判断できる点で有効であった。

C. ショット+人物モード

人物の各登場場面について、前後のショットとのつながりを含めて把握できる点で有効であった。

(2) 代表画像の出現位置の表示による効果

代表画像が出現する時間間隔から、各代表画像が代表する映像長がわかる。これにより、長いショットから選択された代表画像が少数枚の場合にショットを部分視聴する、などの判断ができる点で有効だった。

(3) 部分再生

インタビューシーンのように、内容把握において有用な人物が登場する場面を代表画像から特定して部分再生することにより、発話内容や立ち振る舞いを確認できる点で有効だった。

6. まとめ

本稿では、各ショットの登場人物を把握可能な代表画像による表画像抽出・閲覧システムの構築について報告した。本システムは、抽出システムと閲覧システムとで構成される。抽出システムは2種の代表画像を抽出し、閲覧システムは代表画像に基づく内容把握支援機能を備えた閲覧画面を生成する。本システムによる映像閲覧の結果、登場人物有りのショットでは正面向きの顔が途中から現れる場合であっても登場人物と場面とを把握でき、登場人物無しショットでは情景を概観できる点で、2種の代表画像が内容把握に有効であることを確認した。

参考文献

- [1] Y.Tonomura, A.Akutsu, Y. Taniguchi, "Structured video computing," IEEE Multimedia, vol.1, no.3, pp.34-43, 1994.
- [2] Zhu Liu and Yao Wang, "Major Cast Detection in Video Using Both Audio and Visual Information," ICASSP'01, vol.3 7-11, pp.1413-1416, Apr.2001.
- [3] Masumi ISHIKAWA, Ryoma OAMI, Kyoji HIRATA, "TV-program Selection with Performer Representative Faces for Personal Video Recorders," IEEE ICCE2008 1.3-2 2008.1.
- [4] 石川真澄, 大網亮磨, 平田恭二, "目位置遷移特性に基づくグループ化と重複人物検知による代表画像抽出方式," D-12-2 IEICE2008.
- [5] 岩元浩太, 山田昭雄, "多次元特徴空間解析に基づく映像のカット検出手法", I-027 FIT2005.