

# 動画像データの要約可視化インタフェースの一手法

笠松沙紀（指導教員：伊藤貴之）

## 1. 概要

マルチメディアコンテンツのデジタル化に伴い、録画機器の発達やインターネットでの動画配信が進み、個人が扱う動画の量が増えている。コンピュータ上に大量に蓄積した動画の中から、所望のファイルを探す、個々の内容を確かめる、複数のファイルの内容を見比べるといった作業には非常に労力を費やすことがある。

動画像データは内容を一目で把握しにくいいため、キャプチャ画像で内容を表現する動画要約の手法が盛んである。しかし、要約結果の可視化手法は単にブラウザに画像を並べたものが多く、各画像の表すシーンの時間や、動画全体の長さが伝わりにくい。また、これらの可視化手法の多くは、一つの動画ファイルの要約結果をブラウザに表示することに主眼をおいており、複数のファイルを一覧して内容を見比べることを目的としていない。

そこで本研究では、動画像データの要約結果画像を時間軸上に配置し（図1参照）、1ファイルだけでなく複数のファイルの要約結果を羅列することで、各動画ファイルのストーリーの流れや各シーンの位置を比較するための可視化手法を提案する。本手法では、ユーザ自身がズーム操作によって要約の詳細度（=1ファイルにおける表示画像数）を調節できる。これにより、ズームイン操作によってより詳しい動画の内容を確認する、あるいは逆にズームアウト操作によってより多数の動画を一度に一覧できる、というような適応性のあるユーザインタフェースを実現する。

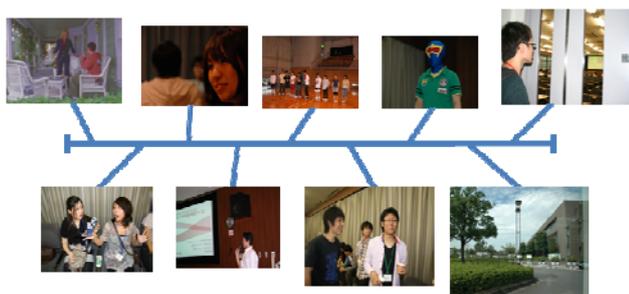


図1: 動画の要約結果を時間軸上に配置

## 2. 関連研究

動画の普及に伴い、動画から代表画像（以下キーフレームと呼ぶ）を抽出する要約手法の研究が盛んである。動画要約では、動画中の重要な部分を除去せず、かつ簡潔に内容を表現するために、キーフレームの選択手法が重要である。従来手法として、各ショット中のフレームのクラスタリングによるキーフレームを選択する手法[1]、映像中の字幕認識や人物・物体検出の結果から重要と判断されるフレームを自動選択する手法[2]、などが知られている。

動画要約の可視化では、要約結果の画像を漫画のコマ割りのように配置し、1ページで動画の内容を表現する手法[3]

や、長時間に及ぶニュース映像を分割して重要度を算出し、重要なキーフレームを中心に表示する一覧可視化手法[4]がある。また、キーフレームを選択するとそのシーンから動画が再生されるブラウザ[5]の開発もされている。

これらの可視化手法は主として、1つの動画ファイルの要約結果を表示させる目的で開発されている。そのため特定の動画の内容把握や、特定のシーンを探索する目的は向いていない。また、ユーザ自身による対話的操作によって、用途に応じて要約の詳細度を調節することは考えられていない。

## 3. 処理手順

本研究では、動画データから抽出したシーンごとのキーフレームを時間軸にそって表示する、という考え方による可視化インタフェースの一手法を提案する。本研究では、動画要約については先行手法を踏襲するとともに、可視化インタフェースの開発に重点をおいている。また本研究において、より有効的な可視化結果を得るために適した動画として、主にドラマや映画を対象にする。

### 3.1. 動画要約

まずキーフレーム抽出による動画要約を行う。

フレーム間ピクセル差分を用いて、動画中の連続した2枚のフレーム間の輝度を比較し、差分が閾値より大きい点で動画をショットごとに分割する。次にショットごとの代表フレームを抽出する。現状では各ショットの時間的中心にあるフレームを代表フレームとするが、将来的には改良する予定である。全ショットの代表フレームを抽出した後、各フレームの YCbCr 値の平均を求め、3次元空間上に配置する。ここで、YCbCr 値の平均値が近いフレーム同士は類似度が高いという前提に基づき、フレーム画像群に対してクラスタリングを行う。類似度によって分類されたフレームの集合を、シーンとする。この各シーン中の重心座標に最も近い YCbCr 値を持つフレームを、シーンの代表フレームとし、これがキーフレームとなる。

### 3.2. キーフレーム重要度の算出

キーフレームを決定した後、各フレームの画像情報をもとに重要度を算出する。従来の動画要約手法では、キーフレームの重要度算出には、そのフレームの属するシーンの意味的重要性を主に考慮している。しかし本研究では、シーンの重要度だけでなく、キーフレームをブラウザに表示した際の画像の見やすさにも重点を置く。以下に重要度のもととなる画像情報について述べる。

#### ① 顔検出数・クローズアップ

フレーム中に人物が映っているシーンは重要度が高いとし、検出された人物の顔がより大きく写っていれば重要な場面と見なす。クローズアップ画像は縮小表示しても認識しやすいことを考慮している。

#### ② 彩度

より彩度の高い画像は、縮小表示しても見やすいため、重要度を高く設定する。彩度は前述の YCbCr 値を用いて算出する。

### ③ シーンの長さ

フレームの属するクラスタの大きさはシーンの長さを意味し、より長いシーンは重要と見なす。

### ④ フレームの位置

動画全体に対するキーフレームの位置によって重要度を変化させる。一般的に動画中でクライマックスや盛り上がるシーンが多いとされる後半のフレームは重要度を高くする。

以上の情報を用いて、各キーフレームの重要度を算出する。

## 3.3. キーフレームのサイズ割り当て

続いて本手法は、算出した重要度から、各キーフレームにサイズを割り当てる。キーフレームを可視化の際に、各フレームの重要度に応じたサイズで表示し、より重要度の高いものを大きく表示することで、重要なシーンを強調した可視化を実現する。

前章で各キーフレームの重要度を算出したが、これらの全キーフレームの中で、最も重要度の高いキーフレームを目安とし、この最大重要度と各キーフレームの重要度の比率でサイズを割り当てていく。予め比率の閾値を決め、数段階のサイズでフレームを表示する。(例えば、あるフレームの重要度が、最大値の 1/8 未満だったらサイズ 1, 1/8 以上 1/4 未満だったらサイズ 2, 1/4 以上 1/2 未満だったらサイズ 3, 1/2 以上だったらサイズ 4 と割り振る。) これにより、異なる動画ファイル間でキーフレームの重要度が大きく変化しても、統一したサイズの割り当てが可能である。

最後に重要度に応じてサイズが割り当てられたキーフレームを、時間軸上に表示する。(図 2 参照)

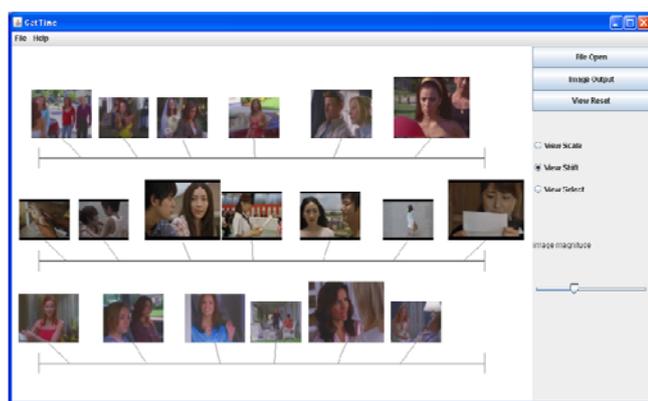


図 2: キーフレームを時間軸上に表示

## 3.4. 要約詳細度の調節

本手法における可視化インタフェースの特徴は、ユーザ自身によるズーム操作によって、要約詳細度を調節できる点である。ここで要約詳細度とは、1つの動画ファイルにつき表示されるキーフレームの数をいう。本手法では画面水平方向のマウス移動によるズーム操作にもなって、表示するキーフレーム数を調節する。このときズームイン

によってキーフレームが増え、より詳細な内容を見ることができる。逆にズームアウトによりキーフレーム数が減り、より簡潔な要約結果を提示できる。

ズームアウト操作にもなって本手法は、近隣するキーフレーム間で重要度を比較し、重要度がより低いキーフレームの表示を省略することで、表示キーフレーム数を減らす。逆にズームイン操作にもなって本手法は、省略されたフレームを復活させて、表示キーフレーム数を増やす。このように本手法は、重要度の高さに基づいて、表示キーフレーム数を制御する。

また、動画ファイルによって表示キーフレーム数が大きく異なる場合があるため、一動画ファイルあたりの表示キーフレーム数を動画の長さから予め決定する。抽出されたキーフレーム数がこの数より多い場合は、重要度の高いものを優先的に表示する。

なお本手法では、画面垂直方向のマウス操作によって、表示する動画ファイル数も同時に調節できる。

## 4. まとめと今後の課題

本研究では、ユーザによる動画ファイルの確認・探索を容易にするために、ユーザ自身による要約詳細度の操作を可能にした、動画要約の可視化インタフェースの一手法を提案した。本手法は、一つの動画データの内容把握を目的とした従来の可視化手法と異なり、蓄積した複数の動画ファイル間の内容比較・ファイル探索に主眼をおいている。

今後の課題としては、重要度算出の要素となる画像情報の正当性の検証や、新たな重要度算出手法の考察が考えられる。特に本研究は複数の動画ファイルを対象としているため、動画要約や重要度算出においても、複数の動画ファイルを互いに比較しながら処理することが望ましい。この点についても今後検討を加えたい。また、動画ファイルの再生機能を加えるなどして、ユーザインタフェースの操作性の向上も今後の課題として挙げられる。

## 参考文献

- [1] Y. Zhuang, Y. Rui, T.S.Huang, and S. Mehrotra, "Adaptive Key Frame Extraction Using Unsupervised Clustering", International Conference on Image Processing, Vol. 1, pp. 866-870, 1998.
- [2] M. Smith, and T. Kanade, "Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97), pp. 775-781, 1997.
- [3] S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky, "Video Manga: generating semantically meaningful video summaries", Proceedings of the Seventh ACM International Conference on Multimedia, pp. 383-392, 1999.
- [4] H. Luo, J. Fan, J. Yang, W. Ribarsky, S. Satoh, "Exploring Large-Scale Video News via Interactive Visualization", IEEE Symposium On Visual Analytics Science And Technology (VAST2006), pp. 75-82, 2006.
- [5] H. Zhang, C. Y. Low and S. W. Smoliar, "Video Parsing and Browsing Using Compressed Data", Multimedia Tools and Applications, vol.1, issue 1, pp. 89-111, 1995.