

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 伊藤 貴之

お茶の水女子大学理学部 〒112-8610 東京都文京区大塚 2-1-1

E-mail: saki@itolab.is.ocha.ac.jp, itot@is.ocha.ac.jp

あらまし 本報告では、複数の動画の要約情報となるフレームを一覧表示する動画像の要約可視化インタフェースの一手法を提案する。本手法ではズーム操作によって、ユーザが要約の詳細度を調節することができる。これによって、ズームアウト操作によって複数の動画を表示させ、内容の比較や、所望の動画の探索が容易になる。逆にズームイン操作によって、一つの動画に関する詳細内容の把握や、特定シーンの発見が容易になる。動画ファイルの要約表示には、動画中の各シーンの代表画像を抽出し、それぞれに対応する時間軸上の位置に配置する。ここで時間軸を考慮して、要約結果画像を年表のように可視化することにより、ユーザが一目で動画ファイル全体の流れ(ストーリー)を把握でき、同時に動画全体の長さや各シーンの位置を直感的にユーザに伝えられる。

キーワード 動画, 要約, 可視化インタフェース

A Visual User Interface for Summarized Multiple Movies

Saki KASAMATSU and Takayuki ITOH

Faculty of Science, Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

E-mail: saki@itolab.is.ocha.ac.jp, itot@is.ocha.ac.jp

Abstract This paper propose a visual user interface that displays frames of summarization of multiple movies. The technique provides a zooming user interface that enables to control level-of-detail by users. It makes easier to compare the contents of multiple movies or specify the movies users want, by the zoom out operation. It also makes easier to understand the detailed stories of movies or discover the specific scenes users want, by the zoom in operation. The technique extracts representative frames from movies, and places them onto the specific position of time axis on the display. Consequently, it displays the frames like a chronological table. It is convenient to understand the story, length of the movies, and the times of scenes.

Keyword Movie, Summarization, Visual User Interface.

1. はじめに

近年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、個々の内容を確認する、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。また、要約結果の可視化手法として、画像の大きさをシーンの重要度によって割り当てる、開発者が要約の長さを調節できる、などの先行研究が存在する。しかし、これら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ものであり、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。

そこで本報告では、動画要約の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。

我々は本手法の提案において、一度見た動画ファイルの内容確認や、蓄積された多数の動画ファイルの比較に主眼を置いている。多数の動画ファイルの要約結果を一画面に全部表示するためには、一動画ファイルが費やす画面空間を節約する必要がある。そのため本手法では、一動画ファイルあたりの表示画像数を抑えるように可視化インタフェースを開発している。特にドラマや映画などの動画はストーリー性が強く、時刻によって画像が大きく異なることが多いため、印象的なシーンの画像を

手がかりにすれば、少ない画像でも比較的早く内容を把握できると考えている。



図 1: 動画の要約結果を時間軸上に配置

本論文の構成は以下の通りである。2 章で関連研究について説明した後、3 章で本研究の提案する手法の説明をする。4 章でまとめ、今後の課題を述べる。

2. 関連研究

動画の普及が進む昨今では、動画から代表的な画像（以下、キーフレームと呼ぶ）を複数個抽出し表示することで、動画の要約を自動生成する研究が盛んである。その際、動画をショット（連続したフレームの集合）ごとに分割するショット境界検出技術が重要となるが、従来のフレーム間差分を用いたものに加え、カメラワークやフラッシュなどによる誤検出を防ぐためにオプティカルフロー、HSV 色情報やエッジ差分を用いて精度を高めた手法[1]がある。更に、パターン認識による学習手法であるサポートベクタマシンを用いたカット検出手法[1][2][3]も研究されている。

また動画要約では、動画中の重要な部分を除去せずにより簡潔に内容を表現するべく、キーフレームの選択手法が重要となってくる。従来の手法[4]では、各ショット中のフレームをクラスタリングすることでキーフレームを選択するが、Smith らの手法[5]では、映像中の字幕や人物・物体の検出などの映像解析技術を用いて各フレームの重要度を算出し、重要なフレームの自動選択を実現している。さらに、映像中の特徴的な動きを検出することで料理映像中の重要なシーンを自動抽出する手法[6]や、ショット構成を解析しスポーツ映像から特定シーンを抽出することで動画要約を行う手法[7]などは、対象とする動画の種類を限定して、要約手法の精度をより高めている。

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[8]がある。Luo らの手法[9]では、長時間に及ぶニュース映像の各フレームにおける重要度を測り、ユーザに

とってより重要な映像を中心に表示するという考え方で、映像の一覧可視化を実現している。また、要約中の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ツール[10][11]の開発もされている。しかし、これらの可視化手法はほとんどが、1 つの動画ファイルの要約のみを表示させる目的で開発されているため、特定の動画の内容把握や、その中から特定のシーンを探索するには向いているが、逆に複数のファイルにまたがる確認や検索には向いていない。また、要約の長さを調節できるのは開発者のみであり、ユーザ自身が用途に応じて短い要約と詳しい要約の表示を使い分けることは不可能である。

3. 提案手法

本報告では、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。その具体的な手法を以下に論じる。

本研究では、動画要約については先行手法に類似する手法を採用しており、動画要約結果の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ていき、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

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、複数の人物が検出されると重要度を更に高くする．

② クローズアップ

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．クローズアップされた画像は小さく表示しても認識しやすいことを考慮している．

③ 彩度

より彩度の高い画像は、小さく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．

④ 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．

⑤ 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．例えば映画やドラマなどの動画では、後半の時刻にはクライマックスや盛り上がるシーンが多いことから、そのフレームが動画の後半に位置するものであれば、重要度を高くする．逆に動画の初めに位置するフレームは重要度を低く割り当てる．

これらの情報を用いて、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する際に、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、最大重要度の 1/8 未満だったら、サイズ 4、1/8 以上 1/4 未満だったらサイズ 3、1/4 以上 1/2 未満だったらサイズ 2、1/2 以上だったらサイズ 1 と割り振るなど.) 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)．

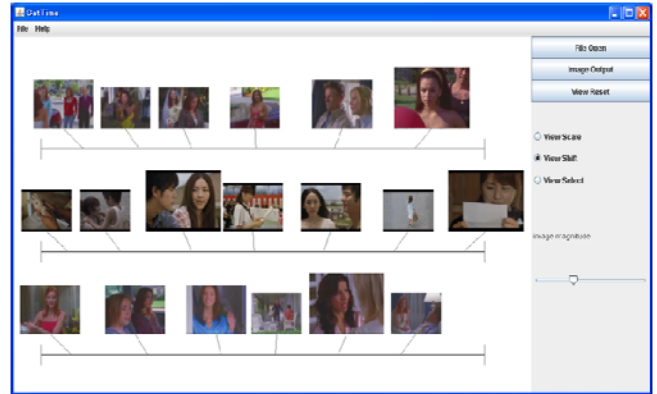


図 2: 実行結果例(1)

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、単位時間あたりに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．

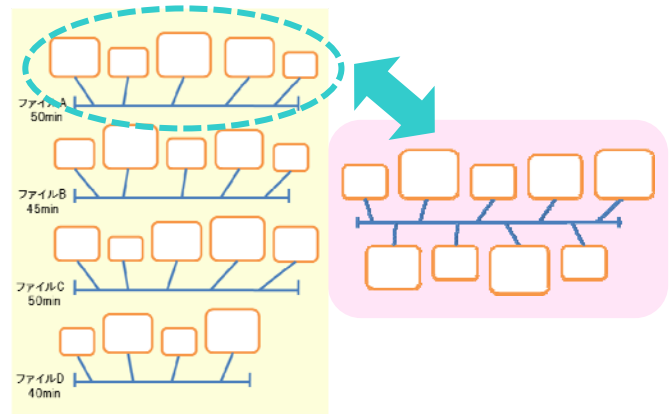


図 3: 時間軸の上下部にキーフレームを表示

ズームアウト操作にもなって本手法は、近くに位置するキーフレーム間で重要度を自動比較し、重要度がより低いキーフレームを排除することで、キーフレームの数を減らす．逆にズームイン操作にもなって本手法は、排除されたフレームを復活させて、フレームの数を増やす．また、縦に拡大していくと、要約結果が 2 段に亘って表示され、時間軸の下部にもキーフレームが表示される (図 3 参照)．キーフレームを時間軸の上下に 2 段で表示することで、キーフレーム数を増やした場合にも、効率的な表示が実現できると考えら

れる.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,ファイルによって表示キーフレーム数が増えるため,表示するキーフレームの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,重要度の高いものから優先する.

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. 実行結果

3章の提案内容を実装した結果を図2及び図4に示す.図2の実行結果において,ユーザが縦に拡大操作すると,本手法は図4のように,フレームを2段にわたって表示する.このように,ユーザのマウス操作による詳細度の調節を可能にした.

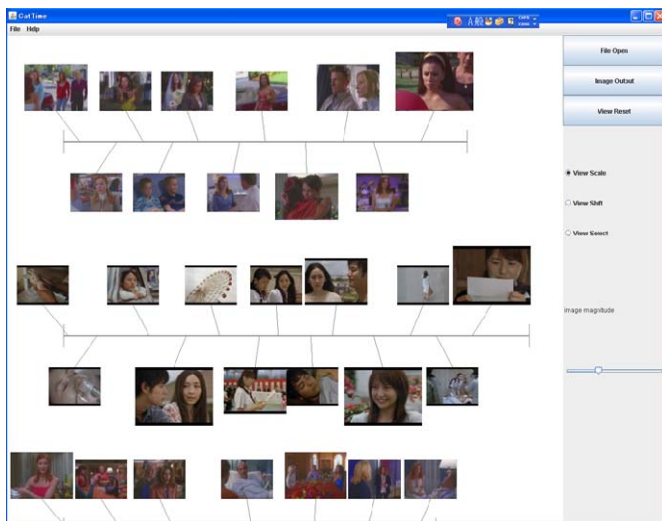


図4: 実行結果例(2)

5. まとめと今後の課題

本報告では,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開発手法を提案した.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での内容確認・ファイル探索に主眼をおいている.

今後の課題としては,重要度算出の手法の正当性の検証や,精度向上のための新たな要素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,現時点ではキーフレームは時間軸に対応する場所に配置されているが,フレーム間のスペースを有

効活用し,フレームを等間隔に配置することも考慮している.

さらには,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] 宮村, 中村, 篠田, 古井, “多段 SVM を用いた頑健な動画ショット境界検出”, 画像の認識・理解シンポジウム (MIRU 2007) IS-2-19, pp.815-820 (2007-7).
- [2] K. Matsumoto, M. Naito, K. Hoashi, and F. Sugaya, “SVM-Based Shot Boundary Detection with a Novel Feature”. ICME 2006: 1837-1840
- [3] Arnon Amir. The IBM Shot Boundary Detection System at TRECVID 2003.
<http://www-nlpir.nist.gov/projects/tvpubs/tvpapers03/IBM%20TREC03%20SBD.slides.pdf>
- [4] Y. Zhuang, Y. Rui, T.S.Huang, and S. Mehrotra, “Adaptive Key Frame Extraction Using Unsupervised Clustering,” in *Proc. ICIP '98*, Vol. I, pp. 866-870, Oct 1998.
- [5] 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.
- [6] 浜田, 三浦, 井手, 佐藤, 坂井, 田中: “動きに基づく料理映像の自動要約手法,” 画像の認識・理解シンポジウム (MIRU2002) 論文集, Vol.2, pp.203-208, Aug. 2002.
- [7] B. Li, and M. I. Sezan, “Event Detection and Summarization in Sports Video,” Content-Based Access of Image and Video Libraries, 2001. (CBAIVL 2001). IEEE Workshop on 14 Dec. 2001 pp. 132 - 138
- [8] S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky, “Video Manga,” Proceedings of the Seventh ACM International Conference on Multimedia, pp.383-392, Orlando, United States, 1999.
- [9] 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.
- [10] 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, March 1995.
- [11] B. Yeo, and M. M. Yeung, “Retrieving and visualizing video,” *Communications of the ACM*, Vol. 40, Issue 12, pp. 43-52, Dec 1997.