

特定人物に注目したダイジェスト動画生成支援のためのインタフェース

山下 紗季[†] 伊藤 貴之[†][†]お茶の水女子大学 理学部 情報科学科 〒112-0012 東京都文京区大塚 2-1-1

1. 概要

ダイジェスト動画の生成は長時間の動画コレクションの中から必要なシーンだけを短時間で鑑賞する有効な手段である。本報告では、映像内に登場する人物に注目したダイジェスト動画生成を支援する一手法を提案する。

本研究におけるダイジェスト動画の定義は、与えられた動画群からユーザが指定した人物が映るシーンを検出して連結させたものである。本研究では動画の内容要約は目的としない。このようなダイジェスト動画が生成されることで、グループ歌手の映像やドラマ映像からユーザが鑑賞したい人物にのみ注目した短い動画を生成することができる。特にユーザがグループ内の特定の個人や特定の俳優のファンである場合に、このようなダイジェスト動画は有用である。

本手法では、指定された人物を含む可能性はあるが確定的ではないショットを一覧表示し、ユーザにショットを選択させる。動画像処理による自動判別とユーザによる選択を組み合わせることで、少ない操作で満足度の高いダイジェスト動画を生成できる。

2. 関連研究

ビデオから特定人物を検出する手法として、まず[1]があげられる。この手法は報道番組の映像に特化しており、顔識別で得られる情報のほかにテキスト情報、タイミング情報なども利用している。そのため、音楽映像やドラマ映像には別の手法を併用する必要がある。[2]は顔に特化した認証手法である。ミュージックビデオに対して実験が行われており、個人アーティストであれば認証率は95%であった。しかし、顔がカメラを向いていないショットや、手元など顔以外の部分にクローズアップしているショット、前髪やマイクで顔が遮蔽されているショットなどは顔領域が検出されず、顔認証ができない。そのため、ユーザが指定した人物が映るショットすべてを検出することは難しい。

3. 提案手法

本章では提案手法の処理手順(図1)を述べる。

3.1 ショット分割

まず、入力された動画をショットに分割する。ショットとは、場面が大きく変化するカット点に挟まれた連続したフレームを指す。このショットが、生成されるダイジェスト動画の一単位となる。分割処理には[3]のプログラムを用いた。このプログラムからは各ショットの始点と終点をフレーム番号で取得できる。

3.2 顔検出、顔識別

続いて各ショット中の顔領域から指定人物を含む可能性を推定し、ショットに得点を与える。

はじめに Azure Media Services を用いてショット中の顔領域を検出する。顔検出できたショットについては、続いてユーザに指定人物の顔画像を入力させ、検出された顔領域との類似度を0から1の実数で算出する。顔検出ができたショットはその類似度を得点とする。

顔領域が検出できなかったショットについては、顔検出できたショットの得点をもとに得点を算出する。得点を求めたいショット A の得点を P_A として、ショット A と顔検出できたショット群 B_i との類似度をそれぞれ求める。類似度を $\text{Sim}(A, B_i)$ としたときに、以下の式で表される実数をショット A の得点 P_A とする。

$$P_A = \max(P_{B_i} \times \text{Sim}(A, B_i))$$

これにより顔領域の条件の差を吸収したショット選出を可能にする。類似度の判定には AKAZE 特徴量を用いる。

以上の手順で算出された得点に比例した大きさで、各ショットのサムネイルを一覧表示する。

3.3 特徴量取得

3.2 節で取得した得点のほかに、各ショットから特徴量を取得する。これは一覧表示する際の画面上の座標値を算出するために用いる。取得する特徴量は、ショットの長さ、分割前の動画における時間上の位置、顔の大きさ、顔の位置、指定人物以外の顔の数、画面の動き方向とする。

A User Interface for creating digest videos focusing on specific persons

[†]Saki Yamashita and Takayuki Itoh, Ochanomizu University

ショットの長さや時間上の位置は 3.1 節で取得したカット点情報から算出できる。顔に関する特徴量は 3.2 節で取得した顔検出の結果から算出できる。画面の動き方向については OpenCV を用いて算出する。

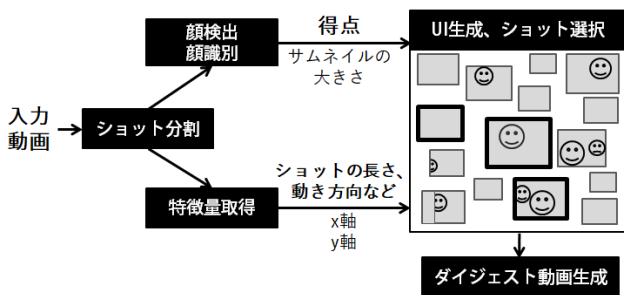


図 1 処理の流れ

3.4 ショット選択のためのインタフェース

図 2 のようにショットを一覧表示にしたユーザインタフェースを生成する。3.3 節であげた各特徴量を画面左側にラジオボタンとして配置し、右側にサムネイルを一覧表示する。表示するサムネイルはショット全てではなく、指定人物が含まれる可能性はあるが確定的ではないものとする。そのため $[0,1]$ の間に閾値を 2 つ定め、得点はその閾値の範囲内となるショットのみを表示する。2 つの閾値を $s, t (s < t)$ としたとき、得点が閾値 s より小さいショットは指定人物が存在しないであろうとして除外する。得点が閾値 t より大きいショットは確実に指定人物を含んでいるとして、あらかじめダイジェスト動画に使用するショットに選出する。そして、ユーザは一覧表示されたサムネイルの中からダイジェスト動画に使用するショットをクリック操作で選択する。このとき、左側のラジオボタンから特徴量を 2 つ選択することで、サムネイルの配置を決定する。これにより、ユーザは選択されるショットの多彩さを調節することができる。たとえば、顔の大きさを指標にして近景と遠景の両方を含むダイジェスト動画を生成することや、ショットの長さを指標にして短いショットだけをまとめたダイジェスト動画を生成することができる。

3.5 ショット連結

選択されたショットを繋ぎ合わせてひとつの動画にする。現時点でショットの連結順を 3 種類実装している。1 つめは入力動画の時系列順に連結する方法、2 つめはユーザがショットを選択し

た順に連結する方法、3 つめは隣接ショット間の差分が小さくなる順序で連結する方法である。3 つめの方法では、ショット A の最終フレームとショット B の先頭フレームの差分から 2 ショット間の距離を算出し、巡回セールスマン問題を解くことでショットの表示順を特定する。



図 2 一覧表示のインタフェース

4. まとめと今後の課題

本報告では、特定の人物に注目したダイジェスト動画生成を支援する手法として、一覧表示のユーザインタフェースを用いることで、自動判別とユーザによる選択を組み合わせたショット選出手法を提案した。

本研究はまだ実装が完成していないので、今後の課題としてまず、ユーザインタフェース部分の完成、および各ショットの得点や特徴量の取得処理を全自動化することがあげられる。そのほかの課題としては、ユーザインタフェースとダイジェスト動画に対する評価手法の検討や、ショットの切れ目に対する音声処理、連結順序をユーザが調整するためのユーザインタフェースの追加があげられる。また、ユーザが指定した長さに合わせてダイジェスト動画を生成する機能や、クロスフェードの機能も検討する。

参考文献

[1] M. Chen, A. Hauptmann, “Searching for a Specific Person in Broadcast News Video”, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings (ICASSP’04), Vol.3, pp.iii-1036, 2004.
 [2] 平井辰典, 中野倫靖, 後藤真孝, 森島繁生, “シーンの連続性と顔類似度に基づく動画コンテンツ中の同一人物登場シーンの同定”, 映像情報メディア学会誌, Vol.66, No.7, pp.J251-J259, 2012.
 [3] P. Sidiropoulos, V. Mezaris, I. Kompatsiaris, J. Kittler, “Differential Edit Distance: A metric for scene segmentation evaluation”, IEEE Transactions on Circuits and Systems for Video Technology, Vol.12, No.6, pp.904-914, 2012