特定人物の顔識別にもとづく対話的ダイジェスト動画生成のためのユーザインタフェース

山下 紗季[1,a)]　伊藤 貴之[1,b)]

概要：著者らは映像内に登場する特定人物に注目してダイジェスト映像を生成する一手法を提案している．本手法ではショット選択のためのユーザインタフェースを生成し，その上で顔識別結果にもとづいて自動選択されたショットとユーザによって選択されたショットを連結する．これにより人物が自動検出されなかったショットもダイジェスト動画の一部に選ぶことができ，より満足度の高いダイジェスト動画を生成できる．本報告ではその一環として，人物が自動検出されなかったショットを選出するためのユーザインタフェースを提案する．本手法の主な用途として，俳優やミュージシャンなど人物の魅力的なカットを集めて楽しむという用途があげられる．

A User Interface for Digest Movie Creation Focusing on Specific Persons

Saki Yamashita[1,a)]　Takayuki Itoh[1,b)]

Abstract: We are developing a new method to generate digest videos focusing on specific persons appearing in the video. This method generates a user interface for shot selection. We suppose to manually select shots on the user interface and then combine them with automatically selected shots to generate a digest video. As a result, we can insert the shots in which the target is not automatically detected as part of the digest videos, and generate highly satisfied digest videos. This paper presents a user interface which assists the manual selection of preferable shots. This method aims to collect attractive scenes of specific persons such as actors or musicians.

1. はじめに

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．

著者らは映像内に登場する特定人物に注目してダイジェスト映像を生成する一手法を提案している．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ユーザに提示し，選択させる．動画像処理によって自動選択されたショットとユーザによって選択されたショット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することを目指している．本報告ではその一環として，時間的に隣接するショットの接続関係を表示し，生成されるダイジェスト動画を概観しながらショットを選択できるユーザインタフェースを提案する．

[1] お茶の水女子大学
Ochanomizu Uniersity
[a)] yamashita.saki@is.ocha.ac.jp
[b)] itot@is.ocha.ac.jp

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 Chen らの手法 [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．また，平井ら [2] は顔に特化した認証手法を提案し，ミュージックビデオを対象とした実験で個人アーティストに対して 95%の認証率を実現している．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．

動画編集を支援する手法としては，土田らの手法 [3] があげられる．このシステムでは，複数のカメラから同時に撮影されたダンス映像を自動的に編集し，ユーザインタフェースを生成することで好みのダンス映像に調整できる．自動編集は土田らが調査した動画編集の原則に基づいて行われる．この手法はダンス映像に特化している点，および多視点で同時録画された映像を題材にしている点において本手法とは異なる．

3. ユーザインタフェース生成の前処理

本章では，指定された人物の自動的な検出や，ユーザインタフェース生成のための情報を取得する処理について述べる．

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 Panagiotis ら [4] 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 顔検出と顔識別にもとづく得点付与

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 Microsodt Azure[5] の 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，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範囲 $[0.0, 1.0]$ の実数で算出し，その類似度を得点とする．類似度は Microsodt Azure の Face API を用いて算出する．顔領域が検出されなかったショットについては，顔検出され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 $Sim(A, B_i)$ としたときに，以下の式で表される実数をショット A の得点を P_A とする．

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

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 [6] を用いる．

そして $[0, 1]$ の間に閾値を 2 つ定め，それらを $s, t\ (s < t)$ としたとき，$P_A < s$ となるショットは指定人物が存在しないであろうとしてダイジェスト動画に組み込むショットの候補から除外する．また $P_A > t$ となる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採用する．そして $s \leq P_A \leq t$ であるショットは指定人物を含む可能性はあるが確定的ではないとし，ユーザによる選択でダイジェスト動画に組み込む．

3.3 特徴量算出

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算出する．現状で算出している特徴量は，ショットの長さ，入力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，画面の動き方向である．これらの特徴量は，ユーザによるショット選択時にどのような内容のダイジェスト動画にするかを考慮するための指標として用いられる．ショットの長さと時間上の位置は 3.1 節で取得したカット点情報から算出される．顔に関する特徴量は顔検出の結果から算出される．画面の動き方向はオプティカルフローをもとに算出される．

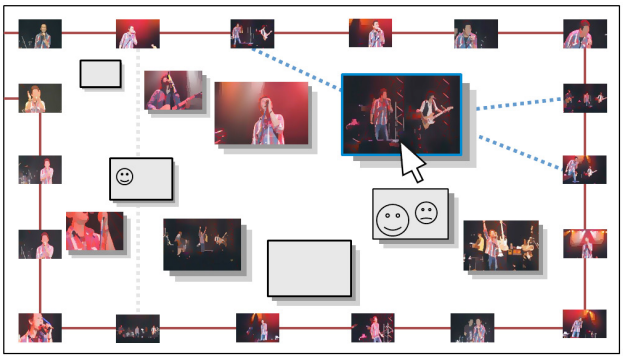

図 1 ユーザインタフェースのイメージ

4. ユーザインタフェースの生成

本章では，ショット連結順を確認しながらショット選択ができるユーザインタフェースを提案する．

まず 3.2 節で説明した得点にもとづきショットを 3 段階に分類する．そして $P_A > t$ であるショットのみを対象として仮の順番を決定し連結する．続いて $s \leq P_A \leq t$ となるショットを対象として，得点の算出過程で得られる他

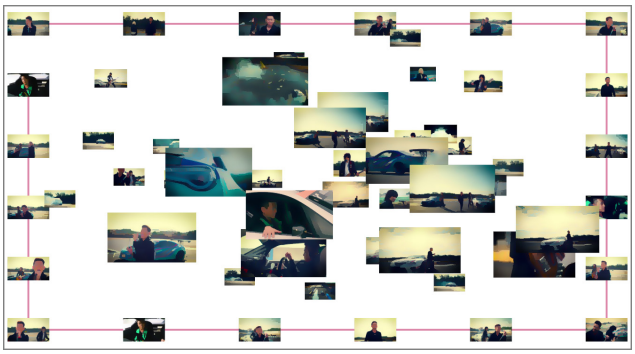

図 2　ユーザインタフェースの実行例

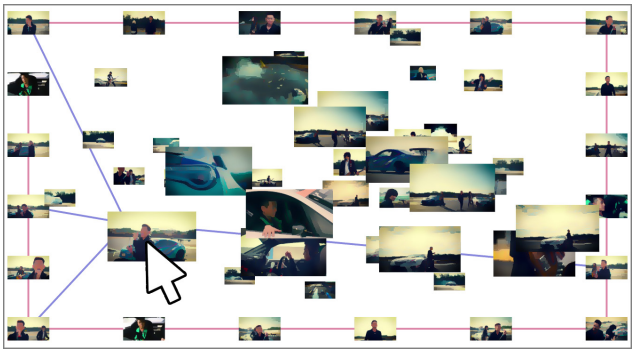

図 3　ユーザインタフェースの実行例，オンマウス時

ショットとの類似度にもとづきクラスタリングを適用する．そして，この処理によって生成された各クラスタが，すでに仮連結されたショットのどの部分に挿入されうるかを判定する．この判定時の評価値が高い場所を挿入の候補位置として推薦する．

続いて図 1 のように，仮決定されたショット群を画面の 4 辺に連結して表示し，その内側にこれから挿入されうるクラスタ群を配置する．ここでノードをショットまたはクラスタのサムネイル，エッジを挿入先候補位置への接続としたグラフを生成し，力学指向ノード配置手法を用いてクラスタの画面上の位置を算出する．そして，クラスタに含まれるショットの得点の最大値に比例する大きさで各クラスタを表示する．あるいは，3.3 節の特徴量のうち一つを選択し，その値が大きいショットを含むクラスタを優先的に表示する．ユーザはショットの連結順を確認しながら，各クラスタに属するショットをダイジェスト動画に採用するかを選択できる．

5. ユーザインタフェースの実行例

本章ではユーザインタフェースの実行例を紹介する．この例では得点が 0.8 より大きいショットをあらかじめ採用し，得点が 0.03 より小さいショットは表示しない．

プログラムを実行すると，まず図 2 のような画面が表示される．挿入候補となる内側のショットと挿入先候補のショットはエッジで接続されているため，ユーザはサムネイルの位置から大まかな挿入先を推察することができる．

マウスカーソルをサムネイルの上に移動させると，図 3 のように挿入先候補となるショットへのエッジが表示される．ユーザはこれにより挿入先候補のショットについて，サムネイルやダイジェスト動画における時間的な位置，前後に仮接続されたショットを確認することができる．これらの情報をもとに，マウスカーソルを置いているショットを採用するか否か，および挿入先の位置を選択する．

6. まとめと今後の課題

本報告では，特定人物に注目したダイジェスト動画生成を支援する手法の研究の一環として，自動判別とユーザによる選択を組み合わせたショットを選出するためのユーザインタフェースを提案した．

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の実装をはじめ，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．ショットの連結順序や挿入先候補の決定については，[3] の文献にあるような動画編集の原則を考慮したい．

また，図 2 において自動車のみのショットが大きく表示されているのが見受けられる．これは 3.2 節で述べた AKAZE 特徴量を用いた類似度の算出において，自動車などエッジを多く含む画像で特徴点が多数検出され，偶発的にマッチング率が高くなるために起きていると考えられる．このようなショットを選択候補から除外するための手法として，一般物体認識を用いて人物を含まないと判定されたショットを減点することや，除外したショットに類似するショット群をユーザインタフェース上で非表示にする機能を追加することがあげられる．

参考文献

[1] Ming-yu Chen and Hauptmann Alexander, Searching for a specific person in broadcast news video, IEEE International Conference on Acoustics, Speech, and Signal Processing (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] 土田修平, 深山覚, 後藤真孝, 多視点ダンス映像のインタラクティブ編集システム, 第 25 回インタラクティブシステムとソフトウェアに関するワークショップ (WISS 2017) pp. 41–46, 2017.
[4] Sidiropoulos Panagiotis, Mezaris Vasileios, Kompatsiaris Ioannis and Kittler Josef, Differential edit distance: A metric for scene segmentation evaluation, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 6, pp. 904–914, 2012.
[5] Microsoft Azure Cloud Computing Platform & Services 2017/12/23 確認 https://azure.microsoft.com/ja-jp/
[6] Pablo F. Alcantarilla, Jess Nuevo and Adrien Bartoli, Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces, British Machine Vision Conference (BMVC), 2013.