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

山下 紗季[1,a)]　伊藤 貴之[1,b)]

概要：本報告では，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する．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．なお本手法では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ことを目的とする．

A User Interface for Digest Movie Creation Focusing on Specific Persons

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

Abstract: This paper presents a 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 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．

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す．ショットを提示する手法については，これまではサムネイルを一覧表示するユーザインタフェースを提案し，報告してきた．本報告では，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の選択ができる新たなユーザインタフェースを提案する．

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 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，平井らは [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

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

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．

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 [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 Azure Media Services[4] 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．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 $Sim(A, B_i)$ としたときに，以下の式 (1) で表される実数をショット A の得点を P_A とする．

$$P_A = max(\ P_{B_i} Sim(A, B_i)\ ) \tag{1}$$

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 [5] 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，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．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採用する．得点が s と t の間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．

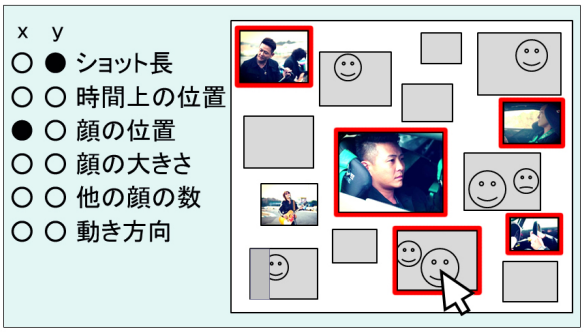

図 1 一覧表示型のインタフェース

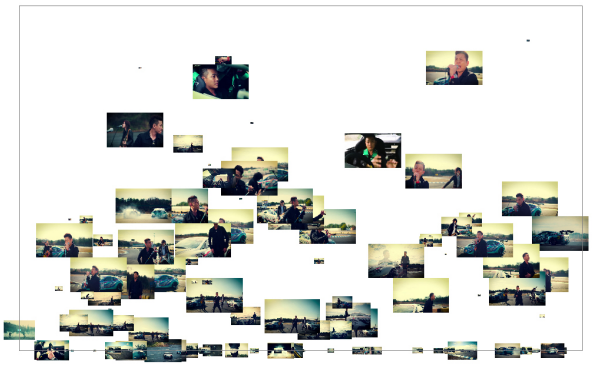

図 2 一覧表示型のインタフェースの実行例

4. ショット選択のためのユーザインタフェース

本章では，ユーザがショットを選択するためのユーザインタフェースの設計について述べる．

4.1 一覧表示型のユーザインタフェース

図 1 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．このユーザインタフェースでは，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，右側に各ショットから生成されたサムネイル群を一覧表示する．ただし全てのショットに対応するサムネイルを表示するのではなく，3.2 節にあるように指定人物が含まれる可能性はあるが確定的ではないショットのみを表示する．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，サムネイルが画面配置される．図 2 では一例として，x 軸にショットの時間上の位置を，y 軸に顔領域の大きさを選択している．この操作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，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，というように一定の意思にもとづいたダイジェスト動画を生成できる．また，表示されているサムネイルをクリック操作で非表示にする機能も実装している．これにより選択操作の最中にユーザが不要と判断したショット表示画面から除外することができるため，より効率的にショットを選択できる．

現時点で我々は 3 種類のショット連結方法を想定している．1 つめは入力動画の時系列順，2 つめはユーザがショットを選択した順，3 つめは隣接ショット間の差分が小さくなる順序である．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をコストとした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．

4.2 ストーリーボード型のユーザインタフェース

4.1 節の手法はユーザによるショット選択の一助となるが，膨大な数のショットに分割される長時間の動画において使いやすくない，ショット選択の段階ではユーザに連結結果が示されない，という問題がある．そこで本節では，ショット連結順を確認しながらショット選択ができるストーリーボード型のユーザインタフェースを提案する．本報告におけるストーリーボードとは，再生される順番にサムネイルを並べた動画編集画面を指す．

まず，3.2 節にあるように，求められた得点をもとにショットを 3 つに分類する．そして得点が十分高いショットのみを対象として仮の順番を決定し，これらを連結表示する．続いて得点が中程度のショットを対象として，得点の算出過程で得られる他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．そして図 3 のように仮決定されたショット群と，これから挿入されうるクラスタを，グラフとして表示する．このときノードはショットまたはクラスタのサムネイル，エッジはショット間の連結もしくは挿入先候補位置への接続となる．ユーザは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する．

これにより，長時間の動画を用いても膨大な数のサムネイルが表示されることはなくなる．また選択中の画面から

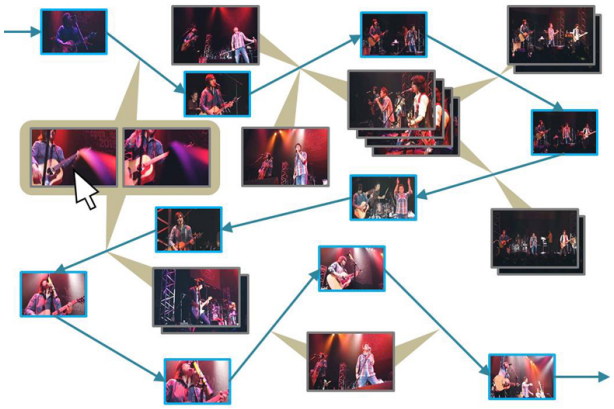

図 3 ストーリーボード型のインタフェース

ユーザがダイジェスト動画の生成結果を想像しやすくなる．

5. 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選出手法を提案した．特に，クラスタリングとグラフの表示を用いたストーリーボード型のユーザインタフェースを新たに提案した．

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．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があげられる．また，サムネイルのみの表示ではショットの区別がつきにくいという問題があるため，ショットのプレビュー画面の追加にも取り組みたい．将来的には，ショットの切れ目に音声処理を施すことやユーザが指定した長さでダイジェスト動画を生成する機能の実装にも取り組みたい．

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] 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.

[4] Microsoft : Media Analytics — Azure Media Services, 入手先 ⟨https://azure.microsoft.com/ja-jp/services/media-services/media-analytics/⟩ (2017.07.27).

[5] Pablo F. Alcantarilla, Jess Nuevo and Adrien Bartoli, Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces, British Machine Vision Conference (BMVC), 2013.