

## 注目領域中の画像類似度に基づく 動画中のキャラクター登場シーンの推薦手法

増田 太郎 平井 辰典 大矢 隼士 森島 繁生

早稲田大学 / JST CREST

### 1. はじめに

近年、インターネット上での動画コンテンツの数は増加の一途をたどっている。そんな中、現在の動画検索は主に動画名やタグなどのテキスト情報に基づき行われている。しかし、テキストによる検索では情報量の多い動画コンテンツを適切に表現することは難しい。例えば、同一キャラクターにおける微妙な作風の違いなど、表現に幅があるような動画の検索を行う際には、映像内容そのものに基づく検索が必要である。また、映像内容に基づく検索であれば、動画中の時系列において何が映っているかを考慮できる。そこで本研究では、画像類似度に基づく類似キャラクターの検索手法について検討することに加え、ユーザが注目したいキャラクターが映る他の動画を検索、推薦する手法を提案する。

### 2. 関連研究

キャラクター登場シーンに基づく動画検索に関する研究として、佃らはニコニコ動画<sup>1</sup>の動画に時刻に同期して付与されたコメント情報を利用し、キャラクターの登場シーンと活躍度を推定した[1]。しかし、短いテキスト情報であるコメントからは動画の詳細なニュアンスを推定するのが困難なため、多くの場合同一キャラクターはその作風の幅に限らずひとまとめにされてしまう。

高山らは、画像中のエッジ特徴や肌色領域を抽出することで、手描きキャラクターの認識・特定手法を提案した[2]。また、映像中からキャラクターを検索する研究には橋爪ら[3]の局所色ヒストグラムを用いた手法などがある。これらは単一画像（フレーム）ごとにキャラクターを検出する手法であり、またアノテーション情報を使わないため、その精度には限界がある。

本手法では、全てを画像ベースで行うのではなく、動画に付与されるタグ情報を補助的に用いることで、キャラクターが登場する動画を絞り込む。その上で動画中の登場キャラクターに関する複数フレームを利用することで、映像の内容に基づくキャラクターの登場シーンの推定を行う。

### Recommending Scenes of 2D Characters Based on Image Similarity in Region of Interest

Taro Masuda, Tatsunori Hirai, Hayato Ohya and Shigeo Morishima

Waseda University / JST CREST

<sup>1</sup> <http://www.nicovideo.jp/>

### 3. 提案手法

本手法では、データベース動画群に対してキャラクターの顔領域の検出を行い、顔領域中の特徴を抽出する。それに対してユーザが視聴したいキャラクター名と、そのキャラクターの中でも好みの作風で描かれた顔を入力することで、入力キャラクターが登場するデータベース動画群を抽出し、それらの顔情報と入力顔領域における顔情報との間の類似度を算出する。それにより、選択したキャラクターと視覚的に類似したキャラクターを含む動画シーンの検索及び推薦を行う。

#### 3.1 ショット検出

動画において、シーンやカメラの切り替わりが無く、フレームの連続性が保たれる最小区間をショットという。同一ショットに登場するキャラクターは同一人物である場合が多い。そこで事前計算として、データベース中の各動画をショット単位に分割する。ショットの切り替わり箇所検出には、動画の各隣接フレームとの輝度ヒストグラムの距離を特徴量として用いた。また、入力したキャラクターの顔を含む入力動画についても、同様にショットを検出し、入力した箇所を含むショット内の全フレームを、後述のキャラクターの類似度算出のために用いる。

#### 3.2 顔領域検出

動画中のキャラクターの顔領域検出のために、高山ら[2]の手法を参考に肌色領域抽出を行う。得られた領域には顔以外の領域も含まれるので、その中から顔領域を選別するために周辺画素情報を用いる。実際の顔領域の上部には、そのキャラクター特有の髪色が現れると想定される。まず、顔領域周辺の上部の画素について色相のヒストグラムを求める。このヒストグラムのピークが、入力した顔領域上部の色相のヒストグラムのピークと一致すれば、その領域を顔領域とする。ただし、画素数や縦横比に閾値を設け、顔領域として相応しくない領域を排除した。

#### 3.3 フレームの連続性を用いた顔の再検出

ショット内の検出顔領域のうち最も肌色領域の比率が高いものは、正しく検出できた顔領域である可能性が高いため、これをテンプレートとする。顔の誤検出を改善するために、ショット内の前後フレームに対し、テンプレートマッ

チングにより顔をトラッキングする。類似度判定には SAD を用いた。マッチした領域を新たな正解顔領域とみなし、フレームを前後させるごとに各フレームでの顔領域を更新する。15 フレーム以上にわたるトラッキングの場合には、テンプレートを別のフレームの顔領域で置換する。

### 3.4 特徴量計算

ここまでで得られた顔領域から、色および形状に関する以下の 2 種類の特徴量を用いて顔領域間の類似度を算出する。1 つ目は、色合いに関する特徴量であるカラーヒストグラムである。まず、顔領域を均一に 16 分割し、各領域に対して 64 色に減色処理を行い、1024 次元の特徴量とする。2 つ目として、テクスチャに関する特徴量である LBP (Local Binary Pattern) を用いる。LBP は、局所領域の輝度値分布を用いた、濃度変化に頑健な特徴量であるため、動画間で顔の明るさの異なる場合などでもロバストにキャラクターの特徴記述ができる。LBP を用いて 256 種類の輝度パターンのヒストグラムをテクスチャ特徴量とする。2 種類の特徴量は、総画素数で正規化処理を行った。以上の手順でショットの全顔領域に対して特徴量を算出し、フレーム数で正規化を行ったものをキャラクターの特徴量とする。

### 3.5 キャラクターの検索

まず、ユーザが視聴したいキャラクター名を入力すると、入力キャラクター名に関連するタグの付与された動画が集められる。同時に、ユーザはそのキャラクターが映っている既知の動画の中から好みの画風で描かれたキャラクターの顔領域を選択する。選択した顔領域をクエリとしてデータベース動画の中からキャラクターを検索する。

#### 3.5.1 顔領域間の類似度算出

2 つのショットから得られた特徴量間の類似度を、ヒストグラム交差法により計算する。カラーヒストグラムと LBP それぞれについて、類似度  $S_C$ 、 $S_L$  を計算し、重みを  $\alpha$  とすると、最終的なショット間のキャラクターの類似度  $S$  は以下の式(1)で表される。

$$S = \alpha S_C + (1 - \alpha) S_L \quad (0 < \alpha < 1) \quad (1)$$

ここで、重み  $\alpha$  はカラーヒストグラムと LBP の分散を考慮して決定される。

入力した顔との類似度が高いショットから順に推薦結果を提示することで、所望のキャラクターに見た目が近いキャラクターが推薦される。

## 4. 動画ハイパーリンクによる映像内容に基づく動画の検索システム

本手法では、任意の顔領域を入力として求めたが、これは、動画を視聴しながら気に入った

キャラクターをクリックする操作を想定している。それによって、動画に映るキャラクターやオブジェクトを仮想のハイパーリンクと見立てることができる。このシステムのプロットを図 1 に示す。

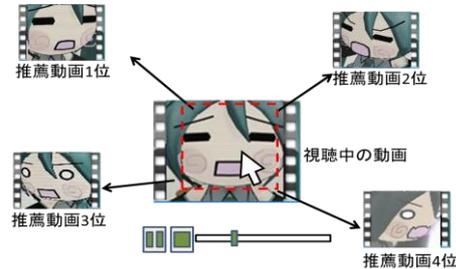


図 1 動画検索システムの概念図

## 5. 実験と結果

本手法を用いて、実際の動画からキャラクターの顔領域抽出およびキャラクター登場シーンの推薦を行った。本実験では、「初音ミク」というキャラクター名に関連するタグが付与された動画を使用した。データベースとして用意した 2 つの動画における顔領域検出率を表 1 に示す。また、ある入力領域に対して、類似度が高い推薦結果として得られたシーンの例を図 2 に示す。

表 1 「初音ミク」の顔検出精度

顔登場 フレーム数	顔検出成功 フレーム数	誤検出 フレーム数	顔検出率 (%)
4104	652	1453	15.9



図 2 入力顔領域 (左) と推薦結果上位 3 例 (右)

## 6. まとめ

本稿では、キャラクター名およびユーザの指定した顔領域を入力として、動画コンテンツ中の作風まで考慮した特定キャラクター登場シーンを推薦する手法を提案し、そのシステム応用の可能性について検討した。

今後は、キャラクターの顔領域だけでなく、より一般的なオブジェクトについて同様の推薦が行えるように本システムを汎用化し、動画像中の様々な物体に対し類似動画シーンへのリンクを張ることを目指したい。

### 文献

- [1] 佃洗撰ら, " 視聴者のコメントに基づく動画検索および推薦システムの提案", WISS 2011 論文集, 2011.
- [2] 高山耕平ら, " 特徴抽出によるアニメキャラクターの顔認識とキャラクター検索", VC/GCAD 合同シンポジウム 2012, 2012.
- [3] 橋爪千枝ら, " 局所的な色照合を利用した映像からのキャラクター検索", IEICE 総合大会講演論文集, 1996.