

SIFT 特徴量の検出位置を考慮した同一人物検索手法に関する研究 A Method for Facial Image Search of the Same Person Considering the Detection Position of SIFT Feature

竹添 亮平[†] 大野 将樹[‡] 獅々堀 正幹[‡]
Ryohei TAKEZOE[†] Masaki OONO[‡] Masami SHISHIBORI[‡]

1. はじめに

近年, HD やフラッシュメモリなどの大容量外部記録メディアの普及に伴って大量の映像が容易に保存可能となっている. 保存した映像が Web 上にアップロードされているため Web 上にも大量の映像が存在する. しかし, それらすべてを人の手によって分類や検索することは困難である. そこで大量の映像データの中から目的のシーンを探す技術, 映像検索技術が必要となる. 映像を検索する場合, 特定の物体をクエリとして映像シーンを探すよりも, 特定の人物をクエリとして検索することが多い. 本研究では, 大量の映像データからクエリと同一の人物が映っているシーンを検索する.

検索には, クエリ画像と同一の物体を認識し検索する特定物体検索がある. 一般的な手法として, SIFT 特徴量[1]を用いて類似度を算出する手法が挙げられる. しかし, この手法では物体検索に対してはある程度の精度が得られるが, 人物検索に対しては良い精度とはいえない. そこで本研究では, 顔の特徴点に限定し, さらに顔の特徴に重みづけを行うことで検索精度を向上させる.

2. 関連研究

同一人物検索は, 大きく2つの手法に分類することができる. 1つはあらかじめ画像に付加された言語メタデータをもとにキーワードから画像を検索する言語メタデータ型画像検索手法である. もう1つは画像から計算で得られた特徴量に基づく内容型画像検索手法である. 本研究では, 内容型画像検索手法の検討を行った.

内容型画像検索手法として, 局所特徴量を用いた高速顔認識[2]などがある. 正面画像においては非常に高い精度であるが, データベースは背景のない正面画像のみで作成されており, 映像からデータベースを作成した場合, 背景との誤対応や顔に角度がつくため, 特定人物を検索するのには向いていない.

3. 提案手法

3.1 特定物体の検索システムの概要

特定の人物を類似度ごとに出力する検索システムを実装した. 実装した検索システムの流れを図1に示す. 動画から抽出したキーフレーム画像を用いて顔画像を作成後, 特徴量を抽出して, データベースに

登録する. クエリ画像からも特徴量を抽出し, データベースに登録された特徴量との比較を行う.

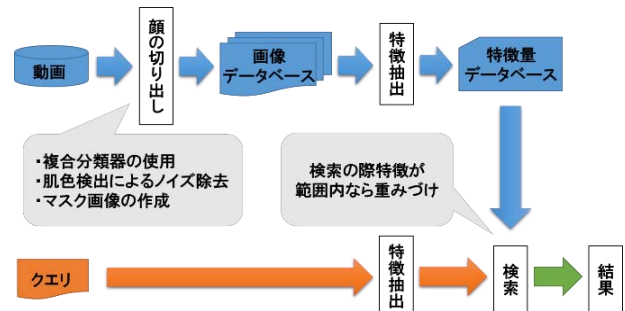


図1: 検索システムの流れ

3.2 画像データベース作成

キーフレーム画像から Haar-like 特徴量の分類器を用いて顔検出を行い, 検出位置で画像を作成し画像データベースとする. 一般的に分類器は単一で用いられているが, 本研究では複数の分類器を組み合わせた複合型分類器を用いた. 背景などの誤対応を防ぐため, 肌色画素が閾値以下の画像を画像データベースから削除した. また, 肌色以外の誤検出を防ぐため, 画像データベースに対して肌色検出を行い, マスク画像を作成する.

3.3 特徴量データベース作成

画像データベースから SIFT 特徴量を抽出し特徴量データベースとする. その際, HSV を用いて肌色検出を行い作成したマスク画像内の特徴量を抽出する. 通常の特徴量抽出とマスク内の特徴量抽出の例を図2に示す. これにより, マスク画像の白の部分のみの特徴量を特徴量データベースに登録する.

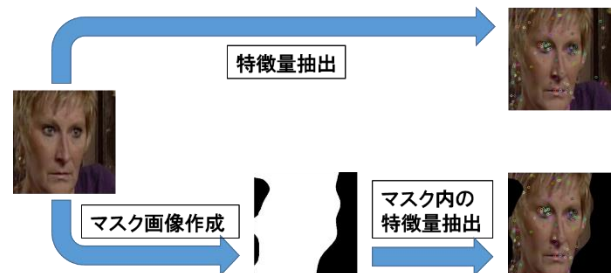


図2: 特徴量抽出の例

3.4 検索手法

画像データベースのマスク内と, クエリ画像の SIFT 特徴量のユークリッド距離の合計を計算することで求める. その際, すべての顔画像を300×300のサ

[†] 徳島大学大学院 先端技術科学教育部 Graduate School of Advanced Technology and Science, Tokushima University

[‡] 徳島大学大学院 ソシオテクノサイエンス研究部 Institute of Technology and Science, Tokushima University

イズにして、図3の赤枠を重みづけ範囲とする。範囲内であれば重みづけを行う。



図3：重みづけ範囲

4. 実験

4.1 データベース

データベースには、映像コンテンツの解析や検索の高度化を目的とした競争型プロジェクトであるTRECVID2014[4]のインスタンスサーチタスクで用いられる映像データを一定のフレーム間隔で分割した画像を用いる。キーフレーム画像は一枚のクエリ画像に対しておよそ2,000枚となり、画質や角度、明度が異なっている人物が写っている。

4.2 実験内容

手法の比較として以下の3つの実験を行った。

- ・実験1：画像から顔を切り出し、SIFT特徴量をそのままデータベースに登録し、類似度を求める。
- ・実験2：実験1にさらに顔と判別した特徴点に重みづけを行う。
- ・実験3：特徴を検出する際、マスク内の特徴のみをデータベースに登録し、実験2と同様に重みづけを行う。

また、検索するクエリ画像を図4に示す。検索結果の上位300件中にどれだけ検索対象が含まれているかを評価対象とし、適合率と再現率のグラフで比較を行う。データベース2000枚に含まれる正解画像はクエリ画像によって枚数が違う。



図4：クエリ画像

4.3 実験結果

クエリ画像の全体平均を図5に示す。適合率、再現率グラフは右上にあるほど良い精度であるといえるので、実験1から実験3に進むにつれ、適合率、再現率ともに精度が向上していることが分かる。最終的な結果はどのクエリ画像においても、実験3が最も良い精度だった。

5. 考察

まず、実験1と実験2の比較を行う。重みづけをしたことにより多くのクエリ画像に対して精度の向上がみられた。特に、メガネやひげといった特徴が

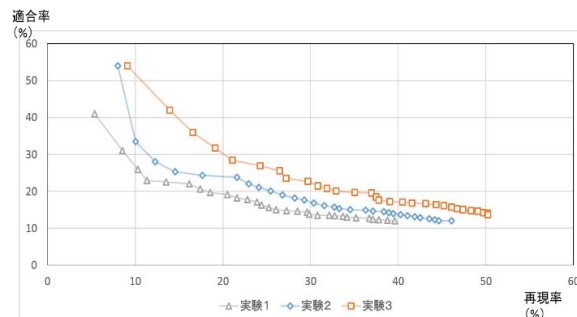


図5：実験結果

ある人物の結果が良かった。しかし、クエリ画像の目と別の人物と目が対応するといった誤対応もあった。また、重みづけ範囲によって少なからず背景との対応点にも重みづけされてしまい、背景が似ている別の人物の画像が上位になるという問題点があった。

次に実験2と実験3の比較を行う。実験3は実験2の背景との誤対応を防ぐことを目的とした手法である。結果は、ほとんどのクエリ画像において背景と誤対応していた画像が下位になり、結果的に正解画像が上位になった。しかし、一部クエリ画像においては、背景が似ているため上位になっていた正解画像が下位になった。

また、どの実験においても正確に目と目、鼻と鼻が対応しているとは言えない。さらに、Haar-like特徴量を用いて顔の分類を行ったが、すべての顔を検出できていないわけではなかった。

6. まとめ

本稿では、SIFT特徴量を用いた映像データに対する人物検索手法を提案した。本手法では、Haar-like特徴を用いた複合分類器により画像から顔を切り出してデータベースを作成した。そして、背景の特徴量の誤検出、誤対応を減らすために肌色検出を用いてマスク画像を作成し、マスク内の特徴点のみを用いて検索を行った。さらに、顔との対応点に重みづけを行ったところ、精度の向上がみられた。

今後は、複合分類器を用いて検出できなかった顔の検出と、同じ部位どうしの対応をとる手法を行い、さらなる精度向上を目指したい。

参考文献

- [1] D. G. Lowe, "Object recognition from local scaleinvariant features", Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 1150-1157, (1999).
- [2] 内海 ゆず子, 坂野 悠司, 前川 敬介, 岩村 雅一, 黄瀬 浩一, "局所特徴量と近似最近傍検索を用いた大規模データベースに対する高速顔認識", 情報処理学会, Vol.2013, No.4, (2013)
- [3] P. Viola, M. Jones : "Rapid object detection using a boosted cascade of simple features," Proc. IEEE Conf. CVPR, pp.511-518 (2001)
- [4] TRECVID, <http://trecvid.nist.gov>