

顔の個人性を反映した複数の顔モデルによる 特徴点検出精度の向上

森 俊介^{1,a)} 内海 ゆづ子^{1,b)} 岩村 雅一^{1,c)} 黄瀬 浩一^{1,d)}

概要: 顔の特徴点を検出する方法として、顔モデルを用いる方法がある。これは与えられた顔画像に対して、顔モデルをフィッティングさせて顔の形状を推定することで、顔の特徴点を検出する手法である。これまで提案されてきた顔モデルベースの特徴点検出手法では、複数の人物の顔画像を用いて1つの顔モデルを構築していた。しかし、人によって顔の特徴は異なるため、1つのモデルでは顔の個人性に対応することが難しく、検出した特徴点の誤差が大きくなることもある。そこで本研究では、顔の特徴点検出精度の向上を目的とし、複数の顔モデルを構築して特徴点を検出する手法を提案する。提案手法では複数の顔モデルがあり、それぞれの顔モデルで用いられる学習画像は異なる人物の顔画像である。そのため各顔モデルは顔の個人性を反映しており、複数の顔モデルをフィッティングに用いることで顔の個人性に対応可能となる。また最も適合した顔モデルから、与えられた顔画像がどのような顔と類似するか大分類することが可能となる。実験の結果、1人につき1つの顔モデルを構築することで、特徴点検出精度は1.3[pixel]向上し、分類の精度は96.0%となった。

1. はじめに

近年、顔認証や表情認識など顔画像から得られる特徴を利用して、画像に写っている人物や人物の表情を推定する技術が発達しており、広く社会に浸透している。例えば、マンションや会社の入り口に設置した監視カメラから入室した人や退室した人を記録する入退室管理システム^{*1}やテーマパークなどでの本人確認^{*2}、写真を撮る時にカメラに映っている人が笑顔になると自動でシャッターを切る技術^{*3}などがある。このように顔画像を利用した技術が発達することで、これまで人間が行っていた人物の特定や表情の推定の作業を自動的に行うことが可能となり、より便利な社会に繋がるといえる。

顔認証や表情認識には、顔画像中の目や鼻、口、輪郭などの特徴点から抽出した特徴量を用いる。抽出した特徴量は、データベースに登録してある特徴量に検索をかけ、顔

認証や表情認識を行う。データベースに登録してある特徴量は、登録する顔画像の目や鼻、口、輪郭などの特徴点から抽出した特徴量である。もし抽出した特徴点が正解の特徴点からずれば、抽出される特徴量が変化して、顔認証や表情認識に失敗することがある。高精度な顔認証や表情認識を行うためには、まず特徴点を正確に検出しなければならない。

一般的な特徴点の検出方法として、顔の形状や外観の情報を含んだ顔モデルを用いる方法がある [1], [2]。これは与えられた顔画像に対して、あらかじめ構築しておいた顔モデルをフィッティングさせて顔の形状を推定する方法である。顔の形状を推定することで、目や鼻、口、輪郭などの特徴点の座標が求められる。

これまで提案されてきた顔モデルを用いた特徴点検出手法では、顔モデルを構築する時は複数の人物の顔画像を用いて1つのモデルを構築していた。しかし人の顔は目や鼻、口などのパーツの配置や肌の色、輪郭の形など人によって特徴が異なるため、1つのモデルでは顔の個人性に対応することが難しい。そのため検出した特徴点が正解の特徴点からずれて顔認識に失敗する問題がある。

そこで本研究では、顔画像の特徴点検出精度の向上を目的とし、複数の顔モデルを構築して特徴点を検出する手法を提案する。提案手法では、類似した外観を持つ顔ごとに顔モデルを構築する。そして、複数の顔モデルを用いて

¹ 大阪府立大学大学院工学研究科 〒599-8531 堺市中区学園町1-1 Graduate School of Engineering, Osaka Prefecture University 1-1, Gakuencho, Naka, Sakai, Osaka 599-8531, Japan

a) mori_s@m.cs.osakafu-u.ac.jp

b) yuzuko@cs.osakafu-u.ac.jp

c) masa@cs.osakafu-u.ac.jp

d) kise@cs.osakafu-u.ac.jp

*1 http://www.nec-nis.co.jp/ja/products/face_authentication/access.html

*2 <http://jpn.nec.com/face/>

*3 <http://www.sony.jp/CorporateCruise/Press/200709/07-0904/>

フィッティングを行い、一番最適となったモデルとそのモデルにより検出した特徴点を結果として出力する。これにより、人によって異なる顔のパーツの配置や肌の色の違いにも対応可能になると考えられる。また一番最適なモデルを出力するのでクエリがどの顔のクラスに属するか、もしくは近いかがわかり大分類が可能のため、顔認識の補助に繋がる。

2. 関連研究

顔の特徴点検出において、昔からよく利用される手法に Elastic Bunch Graph Matching (EBGM) [3] や Active Shape Model (ASM) [4] を用いた方法がある。EBGM は顔の特徴点間の関係をグラフ構造によって表現し、各ノードは Gabor ウェーブレット特徴量 [5] を用いる。特徴点検出対象の顔画像に対して、各ノードの配置を変化させて最適な位置に適合することで特徴点を検出する。ASM は顔の形状の変化を学習データからモデル化する。顔形状は 2 次元の画像平面上での特徴点の集合で表される。モデルのパラメータを変化させてモデルの形状を変化させることで、特徴点検出対象の顔画像に対して、最適な形状を求めて、特徴点を検出する。ASM を拡張したモデルに Active Appearance Model (AAM) [6] がある。AAM では顔の形状だけでなく外観を学習データからモデル化する。顔の外観は顔領域の輝度値で表される。ASM と同様に、モデルのパラメータを変化させることでモデルの形状や外観が変化させ、特徴点検出対象の顔画像に対して最適な形状と外観を求めることで特徴点を検出する。近年は顔の特徴点の周辺の特徴だけを用いたモデルが注目されている。代表的なモデルとして Constrained Local Model (CLM) [7] がある。CLM では外観は特徴点の周りのアピアランスを用いて表す。形状は特徴点の位置関係を木構造によって表す。このためフィッティングのときには、動的計画法により最適な形状を求めることが可能である。ここで挙げた代表的なものだけでなく、多くのモデルベースの特徴点検出手法は、様々な人物の顔画像を学習データとして、1 つのモデルを構築する。よって、モデルは様々な人物に適合が可能である反面、個人性を失っているため、顔の詳細なフィッティングには向いていない。そこで、提案手法では複数のモデルを用いてモデルに個人性を保つことで、精度の向上を図る。

複数のモデルを構築して特徴点を検出する手法に Zhu らの手法 [8] がある。Zhu らの手法では、顔の向きの変化に対応するために、顔の向きごとにモデルを構築する。そして与えられた顔画像に対して全てのモデルをフィッティングさせ、最も一致したモデルと検出した特徴点の座標を結果として出力する。提案手法では顔の向きではなく、顔の個人性に注目して複数のモデルを構築し、全てのモデルをフィッティングして最も一致したモデルと検出した特徴点



図 1 モデルの例

の座標を結果として出力する。

3. 複数のモデルによる顔の特徴点検出手法

提案手法では、学習用の顔画像をクラスタリングし、各クラスで顔モデルを構築する。そして構築したモデル全てをフィッティングさせて、最も一致したモデルの特徴点の座標を検出結果として出力する。本章では、はじめに提案手法に用いた顔モデルについて述べ、続いてモデルの学習方法について詳しく説明する。

3.1 DPM を用いた顔モデル

提案手法では、DPM を用いて顔の形状と外観を表す。DPM は Felzenszwalb らによって提案されたモデルであり、静止画から人や車などの一般的なカテゴリの対象を見つけて位置を特定する物体検出に用いられる [9]。このモデルは、与えられた特徴点の相対位置を利用した形状と、特徴点の周りのアピアランスを表すパッチを組み合わせている。各パッチの位置が移動可能であるため、物体のパーツの並びが変化した場合や、認識対象物の大きさが変化しても認識可能である。DPM を顔の特徴点検出に利用することで、表情や顔の大きさの変化に柔軟に対応可能となる。

3.1.1 顔モデルの定式化

ここでは、提案手法の顔モデルの定式化を行う。DPM に基づく顔モデルは形状とアピアランスの両方を表す。形状は与えられた特徴点の相対位置、アピアランスは特徴点の周りから抽出された局所特徴量 (Histogram of Orientated Gradients (HOG)) [10] により表される。構築したモデルの例を図 1 に示す。白い四角が各特徴点から得られる局所特徴量を抽出する領域 (パッチ) を表し、四角の内側の白い矢印は各ピクセルにおける勾配方向を表す。また赤い線は特徴点をつないだ顔の形状を表す。

モデルを M 個構築した時のモデル番号を m 、モデル m におけるパッチの数を V_m 、画像を I 、 i 番目の特徴点の座標を $l_i = (x_i, y_i)$ とする。そして学習画像の特徴点の集合

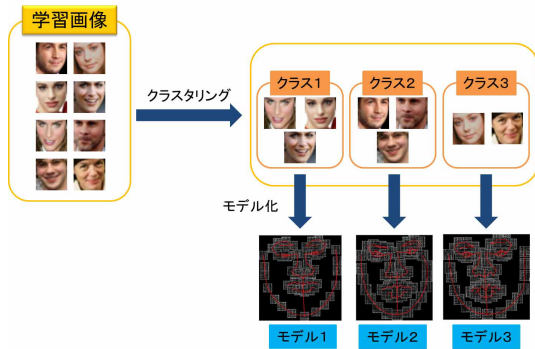


図 2 学習の流れ

を $L = \{l_i : i \subseteq V\}$ とすると, DPM を用いた顔モデルは式 (1) で表される.

$$S(I, L, m) = \text{App}_m(I, L) + \text{Shape}_m(L) + \alpha^m \quad (1)$$

$$\text{App}_m(I, L) = \sum_{i \subseteq V_m} \mathbf{w}_i^m \cdot \phi(I, l_i) \quad (2)$$

$$\text{Shape}_m(L) = \sum_{ij \subseteq E_m} a_{ij}^m dx^2 + b_{ij}^m dx + c_{ij}^m dy^2 + d_{ij}^m dy \quad (3)$$

式 (2) はモデルのアピランスを表す. $\phi(I, l_i)$ は画像 I の座標 l_i から抽出した特徴量であり, \mathbf{w}_i^m はフィルタである. 式 (3) はモデルの形状を表す. $dx = x_i - x_j$ と $dy = y_i - y_j$ は i 番目の特徴点と j 番目の特徴点の x 座標, y 座標それぞれの変位である. 各項によって 2 つの特徴点の位置関係を表す. また, パラメータ $(a_{ij}, b_{ij}, c_{ij}, d_{ij})$ によって各パッチの位置の変化の幅を指定する. α^m はモデルを制御するバイアスである.

3.2 複数のモデルの構築

提案手法では, 顔の個人性に対応するため, 複数の顔モデルを構築して特徴点を検出する. 本節では, 複数のモデルの構築方法について述べる. 図 2 にモデルを構築するための学習の流れを示す. 最初に, 複数のモデルを構築するために, 学習顔画像をクラスタリングする. 続いて, それぞれのクラスごとに顔モデルを構築する. 以下, 学習画像のクラスタリング, モデルの学習方法について詳細に述べる.

3.2.1 学習画像のクラスタリング

複数のモデルを構築するため, 学習画像をクラスタリングする. 本研究では, 類似した顔の外観によって学習画像を分割するために K-means クラスタリングを利用する. また各モデルの構築に用いる学習画像を均等にするためランダムにクラスタリングを行う.

K-means クラスタリングで学習画像を分割する方法について説明する. 学習画像には同一の人物の顔画像が数枚あるため, まず人物ごとに個別の顔モデルを構築する. 顔の外観によって学習画像を分割するため, クラスタリングに

は各モデルのフィルタを利用する. 各モデルのフィルタはベクトルで表されており高次元であるため, 主成分分析により次元圧縮を行う. これを全てのモデルのフィルタについて行い, 圧縮されたベクトルを用いて K-means クラスタリングにより学習画像のクラスタリングを行う. 本研究では, K-means クラスタリングのうち, KKZ 法 [11] を用いてクラスタリングをする.

K-means クラスタリングでは, モデルごとに学習に用いた人数が異なる. このことにより, 特徴点検出精度などに影響が出る可能性が考えられる. そこで, ランダムにクラスタリングを行いモデルを構築して特徴点検出の精度評価をすることで, 学習に用いた人数と精度の関係も明らかにする.

3.2.2 モデルの学習方法

モデルの学習は顔画像をクラスタリングして, クラスごとに学習を行う. Chow-Liu algorithm [12] を用いて与えられたモデルの特徴点の配置が最適となる最尤木構造を見つける. n を学習画像の ID, クラスタリングされた顔画像であるポジティブ画像のサンプルを $\{I_n, L_n, m_n\}$, 顔以外の画像であるネガティブ画像のサンプルを $\{I_n\}$, $z_n = \{L_n, m_n\}$ とする. 式 (1) のパラメータ \mathbf{w} と $(a_{ij}, b_{ij}, c_{ij}, d_{ij})$, α を 1 つのベクトル β にまとめると, β を推定するための目的関数は式 (4) のように表すことができる [13].

$$S(I, z) = \beta \cdot \Phi(I, z) \quad (4)$$

そして式 (5) を制約として, 最適化し, モデルのパラメータを推定する [8].

$$\arg \min_{\beta, \xi_n \geq 0} \frac{1}{2} \beta \cdot \beta + C \sum_n \xi_n \quad (5)$$

$$\text{s.t. } \forall n \in \text{pos} \quad \beta \cdot \Phi(I, z_n) \geq 1 - \xi_n$$

$$\forall n \in \text{neg}, \forall z \quad \beta \cdot \Phi(I, z_n) \leq -1 + \xi_n$$

$$\forall k \in K, \quad \beta_k \leq 0$$

3.3 特徴点の検出と顔画像の分類

3.2 節の学習したモデルをフィッティングすることで, 特徴点の検出と顔画像の分類を行う. 図 3 に特徴点検出の流れを示す. 与えられたクエリに対して, 構築したモデル全てをフィッティングさせる. そして最適なモデルとそのモデルを用いたときに検出した特徴点の座標を結果として出力する. クエリ I が与えられた時に, 構築したモデルを用いて特徴点を検出するために, 式 (1) の $S(I, L, m)$ が最大となる L と m を求める.

$$S^*(I) = \max_m [\max_L S(I, L, m)] \quad (6)$$

式 (6) によって求まる m がクエリに対して最適なモデルであり, L が検出した特徴点の座標となる. 最大値を計算するときは Dynamic Programming (DP) [14] を用いる.

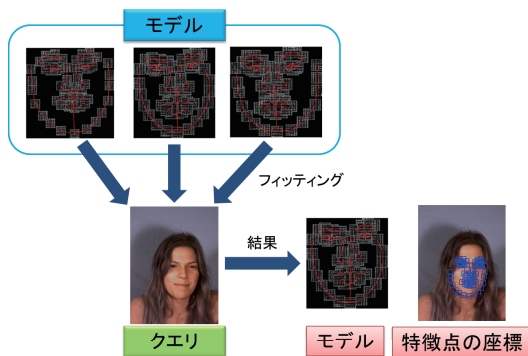


図 3 特徴点検出の流れ

式 (6) により最適なモデルが求まるため、構築する顔モデルが 1 人あたり 1 つであれば、特徴点を検出すると同時に顔認識が可能となる。また 1 つの顔モデルが複数人で構築されている場合、結果として出力したモデルからクエリの人物がどの顔のクラスに属する、もしくは近いかが分かる。

4. 実験

提案手法の特徴点の検出精度を評価するために実験を行った。また、提案モデルを用いた場合に個人認証が可能であるか検証した。

4.1 実験条件

実験には MUCT Face Dataset [15] を用いた。その一例を図 4 に示す。このデータセットには 276 人の異なる人物の顔画像と 1 枚につき 76 点の顔の特徴点が登録されている。また 1 人につき 3 種類の明るさの画像がある。画像の解像度は 480×640 [pixel] である。このうち図 4 の左側と右側の同じ照明条件の画像 2 枚を学習画像とし、図 4 の中央の画像と同じ照明条件の画像をクエリとして用いた。提案手法のモデル構築には 50 人分の画像を用いた。

3.2 節で述べたように、学習画像をクラスタリングし、複数のモデルを構築した。本実験では、学習画像を 2, 3, 4, 5, 10, 15, 20, 25, 50 個に分割し、それぞれ顔モデルを学習して評価をした。モデルの数が 2, 3, 4, 5, 10, 15, 20, 25 個の場合は K-means クラスタリングとランダムにクラスタリングをして複数のモデルを構築した。またモデルを 50 個構築した場合は 1 人の学習画像から 1 つのモデルを構築した。3.2.1 節で述べたように、提案手法で用いた K-means クラスタリングには同じデータに対しては必ず同じクラスタリング結果を返す。よって K-means クラスタリングを用いた場合は、1 回クラスタリングを行い、モデルの学習と特徴点検出精度の評価をした。一方、ランダムにクラスタリングをする場合は、クラスタリング結果は毎回異なる。そこで、クラスタリングを 10 回行い、その都度モデルの学習をして特徴点検出を行い、精度の平均を計



図 4 MUCT Face Dataset の顔画像の一例

算した。これに加えて、提案手法のモデルとの精度の比較をするため、50 人の学習画像全てを用いて 1 つのモデルを構築した。以下、このモデルを従来モデルと呼ぶことにする。学習の際のネガティブ画像は、全てのモデルで同じ非顔の画像集合を用いた。

特徴点検出精度の評価方法は、検出した全ての特徴点と正解の特徴点との距離を計算し、それらの和を平均して誤差を求めた。つまり 1 つの特徴点あたりの誤差の平均値を評価として用いた。また顔の分類の評価方法は、クエリに最も適合したモデルを構築するのに用いた人物とクエリの人物が一致すれば正解とする。使用した計算機の CPU は Opteron6238 2.6GHz, メモリは 512GB, OS は Debian GNU/Linux 6.0 である。

4.2 クエリとモデル構築の人物が同一の場合

クエリの人物とモデルの構築に用いる人物が同一の場合の特徴点検出精度と顔の分類の正解率を評価した。クエリの人物とモデルの構築に用いる人物が同一であるため 50 枚の画像をクエリとして用いた。従来モデルと提案モデルの特徴点の検出精度についての評価を行った結果を図 5 に示す。緑のグラフが K-means クラスタリングした場合の結果で、赤のグラフがランダムにクラスタリングした場合の結果である。検出した特徴点の誤差が 0[pixel] に近いほど検出精度が高い。モデルを 50 個構築した場合の誤差の平均値は 3.47[pixel] となり、従来モデルを用いた場合の誤差の平均値は 4.80[pixel] となった。これにより 1 人につき 1 つにモデルを用いると特徴点の検出精度は向上することが

明らかになった。1人につき1つモデルを構築することで、従来モデルよりも人の顔の個人性に対応できたと考えられる。また K-means クラスタリングした場合は従来モデルによる検出精度と比べてほぼ変化がなかった。K-means クラスタリングした場合、モデル構築に用いる人数はモデルによって異なる。クエリに適合する回数が多かったモデルは、構築に用いる人数が多いモデルであった。モデル構築に用いる人物が多いため、モデルは顔の個人性を反映しておらず特徴点検出精度が向上しなかったと考えられる。またランダムにクラス分けした場合は、モデルの数が多くなるほど特徴点の検出精度は向上した。モデルの数が多いと1つのモデルを構築するのに用いる人の数が少なくなる。このため構築した顔モデルのアピアランスが人のクエリの顔画像のアピアランスに対応しやすくなり、特徴点の検出精度が向上したと考えられる。

提案手法による顔のクラスタリングの精度を評価した。提案モデルを用いた場合の正解率を図6に示す。50個のモデルを用いた場合の正解率は96%であった。この場合、モデルは人物ごとに構築しているため、正解すると個人認証をしていることと同じになる。この正解率から、提案手法は特徴点の検出と認識を同時に行うことが可能であるといえる。ランダムにクラス分けをして複数のモデルを構築した場合は、モデルの数が多くなれば正解率が向上した。ランダムにクラスタリングした場合、各モデルの人数が均等になるようにした。このためモデルの数が多くなるほど1つのモデルに用いる人の数は少なくなる。結果としてモデルが個人性を反映しやすくなり、正解率が向上したと考えられる。

顔認識に失敗したクエリを図7(a)に、図7(a)の人物が学習に用いられたモデルを図7(b)にそれぞれ示す。また図7(a)がクエリとして与えられた時に出力されたモデルを図7(c)に示す。図7(b)のモデルの学習画像2枚は正面を向いていたため、構築したモデルも正面を向いている。しかし図7(a)のクエリは顔が傾いており、クエリの人物が学習に用いられた図7(b)に示すモデルではなく、図7(c)に示す顔が傾いているモデルのほうが適合した。学習画像が2枚と少なかったため、構築したモデルが顔の傾きの変化に対応できなかったと考えられる。K-means クラスタリングによってクラス分けをした場合は、モデルの数が多くなれば正解率が低下した。

K-means クラスタリングを用いたときに正解率が低下した理由を明らかにするために、モデルの学習に用いた人数と適合モデルとして選択された回数との関係を調べた。表1、表2にモデルの数が4個と5個のときのモデルの学習に用いた人数と適合モデルとして選択された回数をそれぞれ示す。K-means クラスタリングでは各クラス的人数が均等にならないため、複数のモデルを構築した場合、1つのモデルを構築するのに用いる人物の数が異なる。表1、表2か

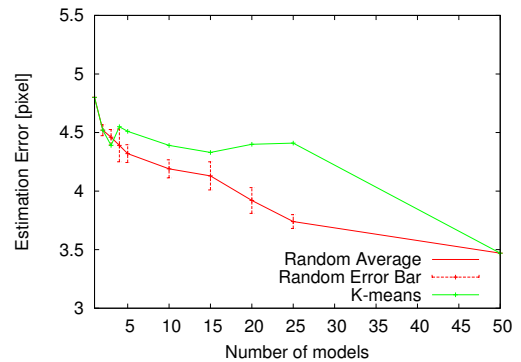


図5 特徴点の検出誤差

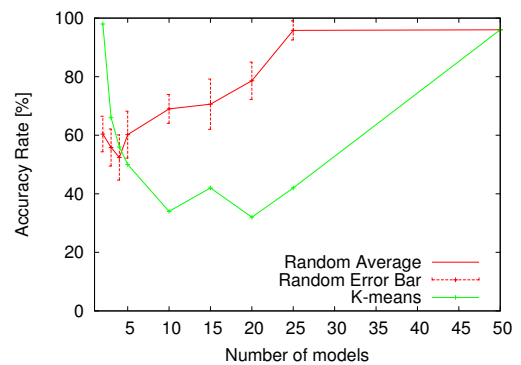


図6 提案モデルを用いた場合の正解率

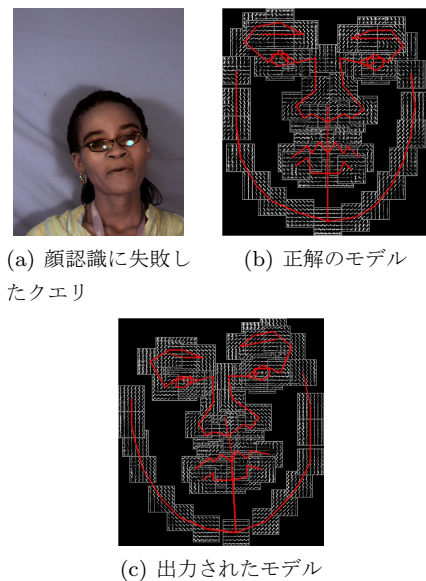


図7 顔認識に失敗したクエリと正解のモデルと出力されたモデル

ら、モデルの数が4個のときも5個のときも、モデルの学習に用いた人数が多いモデルが結果として出力されることが多い。学習に多くの人物を用いたほうが顔の表情の変化や傾きに柔軟に対応できるため、学習に多くの人物を用いてモデルを構築したほうが結果として出力されやすくなると考えられる。

4.3 クエリとモデル構築の人物が異なる場合

モデルの構築に用いた人物と異なる人物の顔画像がクエ

表 1 K-means によるクラスタリングでモデルを 4 個構築した場合の各モデルが学習に用いた人数と結果として出力された回数

モデル ID	1	2	3	4
学習の人数 [人]	1	18	7	24
結果として出力された回数 [回]	0	16	0	34

表 2 K-means によるクラスタリングでモデルを 5 個構築した場合の各モデルが学習に用いた人数と結果として出力された回数

モデル ID	1	2	3	4	5
学習の人数 [人]	1	11	7	24	7
結果として出力された回数 [回]	0	0	0	49	1

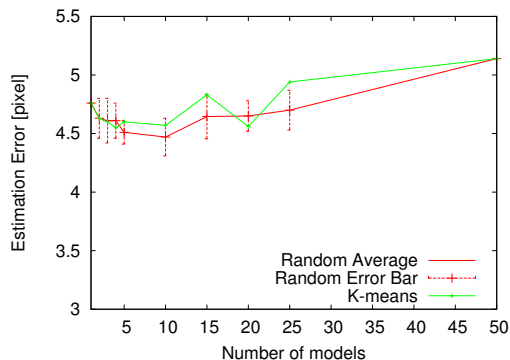


図 8 クエリと学習画像の人物が異なる場合の特徴点の検出誤差

リとして与えられたときに、提案手法を用いることで従来モデルよりも特徴点の検出精度が向上するか検証した。MUCT Face Dataset から学習画像に用いていない人物の顔画像 20 枚をクエリとして、特徴点検出精度を評価した。

特徴点検出精度の評価の結果を図 8 に示す。緑のグラフが K-means クラスタリングした場合の結果で、赤のグラフがランダムにクラスタリングした場合の結果である。従来モデルを用いた場合の誤差の平均値は 4.80[pixel] である。最も誤差の平均値が小さかったのは、ランダムによるクラスタリングで 10 個のモデルを構築した場合で 4.47[pixel] となった。またモデルの数が 50 個の場合の誤差の平均値は 5.14[pixel] となり、従来モデルよりも検出精度が低くなった。1 人につき 1 つのモデルを構築した場合、モデルは学習に用いた人物そのものとなる。クエリが学習に用いた人物とは異なる人物であった場合、モデルとクエリのアピアランスの違いが大きくなりモデルの最適化が困難になる。そのため従来手法よりも特徴点の検出精度が低下したと考えられる。

5. まとめ

本研究では、顔画像の特徴点検出の向上のため、複数の顔モデルを用いる手法を提案した。提案手法では、学習画像をクラスタリングしてクラスごとにモデルを構築し、クエリに対して全てのモデルをフィッティングして、最も適合したモデルと特徴点の座標を検出結果として出力した。

クエリに対して最適なモデルを選択することで、クエリの人物を大分類することが可能となった。人物ごとにモデルを構築した場合、従来モデルのように 1 つのモデルを構築した場合よりも特徴点の検出精度は 1.3[pixel] 向上した。また、クエリと適合した顔モデルの正解率は 96% となった。クエリとして用いる人物をを学習に用いていない人物にした場合は、特徴点の検出精度は低下した。

今後の課題として、学習画像のクラスタリングの方法の変更や、クエリと学習画像の人物が異なる場合の特徴点の検出精度向上、またより大規模なデータベースを用いた実験が挙げられる。

参考文献

- [1] Yu, X., Huang, J., Zhang, S., Yan, W. and Metaxas, D. N.: Pose-Free Facial Landmark Fitting via Optimized Part Mixtures and Cascaded Deformable Shape Model, *Proceedings of The IEEE International Conference on Computer Vision (ICCV)* (2013).
- [2] Zhao, X., Shan, S., Chai, X. and Chen, X.: Cascaded Shape Space Pruning for Robust Facial Landmark Detection, *Proceedings of The IEEE International Conference on Computer Vision (ICCV)* (2013).
- [3] Wiskott, L., Fellous, J.-M., Krüger, N. and von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 775–779 (1997).
- [4] Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J.: Active Shape Models & Mdash; Their Training and Application, *Comput. Vis. Image Underst.*, Vol. 61, No. 1, pp. 38–59 (1995).
- [5] Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P. and Konen, W.: Distortion Invariant Object Recognition in the Dynamic Link Architecture, *IEEE Trans. Comput.*, Vol. 42, No. 3, pp. 300–311 (1993).
- [6] Cootes, T. F., Edwards, G. J. and Taylor, C. J.: Active Appearance Models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 484–498 (1998).
- [7] Cristinacce, D. and Cootes, T.: Feature detection and tracking with constrained local models, *Proceedings of British Machine Vision Conference*, Vol. 3, pp. 929–938 (2006).
- [8] Zhu, X. and Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild, *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2879–2886 (2012).
- [9] Felzenszwalb, P., McAllester, D. and Ramanan, D.: A Discriminatively Trained, Multiscale, Deformable Part Model, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2008).
- [10] Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, *Proceedings of International Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 886–893 (2005).
- [11] Katsavounidis, I., Jay Kuo, C.-C. and Zhang, Z.: A new initialization technique for generalized Lloyd iteration, *Signal Processing Letters, IEEE*, Vol. 1, No. 10, pp. 144–146 (1994).
- [12] Chow, C. K. and Liu, C. N.: Approximating discrete probability distributions with dependence trees, *IEEE Transactions on Information Theory*, Vol. 14, pp. 462–467 (1968).
- [13] Yang, Y. and Ramanan, D.: Articulated Pose Estimation with Flexible Mixtures-of-parts, *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, pp. 1385–1392 (2011).
- [14] Felzenszwalb, P. F. and Huttenlocher, D. P.: Pictorial Structures for Object Recognition, *International Journal of Computer Vision*, Vol. 61, No. 1, pp. 55–79 (2005).
- [15] Milborrow, S., Morkel, J. and Nicolls, F.: The MUCT Landmarked Face Database, *Pattern Recognition Association of South Africa* (2010).