

# 顔認識技術における位置合わせの重要性について

田代 茜 (応用技術開発部)  
 笠井 隆志 (応用技術開発部)  
 宇佐見 直人(応用技術開発部)

**概要**：応用技術開発部は従来型の固有顔といわれる特徴を用いた顔認識技術を保有しているが、完全自動化等、性能がシステムに決定的な影響を与える場面で使える程の認識精度は出せていない。一方、Deep Learning の登場により、人間並みの高い認識精度の顔認識技術の Facebook の DeepFace[1]や Pyramid CNN[2]が登場した。これらの顔認識技術は顔の位置ずれへの頑健性を強化していることで、高い認識精度を出している。このことより従来型の手法に顔の位置合わせを取り入れることで、顔認識精度を自動化レベルに到達させることが可能であると仮説を立てた。そこで、本報告では従来型の手法の仕組みを分解して評価することで、顔認識精度と位置合わせの因果関係を明らかにしたい。

**キーワード**：画像認識、固有物体認識、顔認識、主成分分析、線形判別分析

## 1. 序論

応用技術開発部は従来型の固有顔といわれる特徴を用いた顔認識技術を保有しており、それをいくつかのソリューションに適用してきた。しかし、その適用は業務の効率化等人間の補助として顔認識技術を適用しており、セキュリティ用途や完全自動化等、性能がシステムに決定的な影響を与える場面で使える程の認識精度は出せていない。一方、Deep Learning の登場により、画像の機械学習技術が発展し、人間並みの高い認識精度を持った顔認識技術として Facebook の DeepFace[1]や Pyramid CNN[2]が登場した。

通常顔認識に使用されている Deep Learning は畳み込みニューラルネットワーク(Convolutional Neural Network:CNN)であり、畳み込みによって位置のばらつきに対してロバスト性を確保している。さらに DeepFace は3次元的に顔の方向合わせも行うことで人間並みの認識精度を出している[3]。これらの特徴に共通する事項として、顔の位置ずれへの頑健性を強化していることがあげられる。

従来型の手法に顔の位置ずれへの頑健性すなわち顔の位置合わせを取り入れることで、顔認識精度を自動化レベルに到達させることが可能であると仮説を立てた。

そこで、本報告では従来型の手法の仕組みを分解することで、顔認識精度と位置合わせの因果関係を明らかにしたい。

## 2. 固有顔を用いた顔認識技術

### 2.1 顔認識技術の概要

顔認識技術は予め用意された顔写真である学習画像の顔部分と認識したい人物が映っている検査画像の顔部分の類似度を算出し、最も類似度が高かった顔写真の人物と認識する技術である(図 1)。

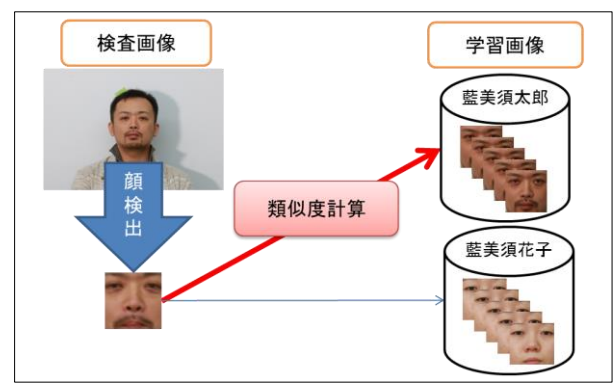


図 1 顔認識技術のイメージ

以下に固有顔といわれる特徴を用いた顔認識技術の手順を示す。

- 1 Haar-like 検出器[4]等を用いて、画像から顔部分を抽出する(顔検出)。
- 2 学習画像群の画素毎の平均をとり、平均顔を作成する。
- 3 学習画像毎に画素値と平均顔の差をとり、差分画像を作成する。
- 4 全ての学習画像の差分画像に主成分分析を行い、学習画像を複数の固有顔といわれる特徴を求める。この時、各学習画像はすべての固有顔と平均顔の重み付き加算で表現することが出来る(図 2)。この時、重みベクトルを各学習画像の特徴ベクトルとする。
- 5 検査画像の各固有顔との一致度(係数)から重みを求めて特徴ベクトルに変換する。
- 6 学習画像群の特徴ベクトルと検査画像の特徴ベクトルのユークリッド距離を算出する。変数を  $P$  個持った 2 ベクトル間のユークリッド距離は式(1)より求める。

$$D(x, y) = \sqrt{\sum_{i=1}^P (x_i - y_i)^2} \quad \dots(1)$$

7 ユークリッド距離が近いほど類似度が高いこととして、検査画像と類似度が最も高かった学習画像の人物と認識する。

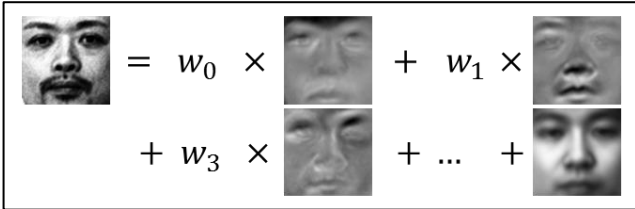


図 2 顔画像と固有顔及び平均顔の関係

## 2.2 顔認識技術の詳細

### 2.2.1 平均顔

学習画像群の画素毎に平均をとった画像である。平均顔の一例を図 3 に示す。これを用いて、各画像との差分画像を求めると、ピクセル毎の平均からの乖離を表したものになり、顔画像の特徴を表す最も基本的なデータとなる。



図 3 実際の平均顔

### 2.2.2 主成分分析

実際の顔画像には顔のパーツのずれや顔の傾き等のノイズが含まれている。そのため、顔画像の単純な画素値を比較して顔認識を行うと、顔の特徴とは関係ない成分も加味してしまう。そこで、顔の特徴だけで比較を行えるよう、顔の特徴をよく表している成分だけを抽出するために、主成分分析を行う。

図 4 のような 2 変数のデータを例にとる。図 4 は、横軸が大きくなるほど縦軸も大きくなるという特徴を表現しているが、図 5 のように最もデータの特徴を表現できる新しい軸を用意すれば、もっと容易にデータの特徴を表現することが出来る。このようにある尺度で特徴が最も表現できる新しい軸を見つけて評価することを特異値分解 (Singular Value Decomposition: SVD) という。

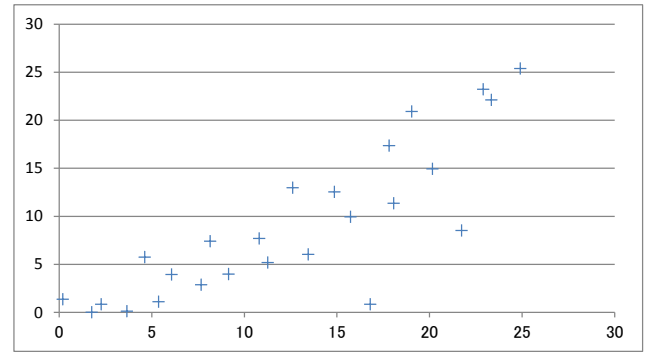


図 4 サンプルデータ

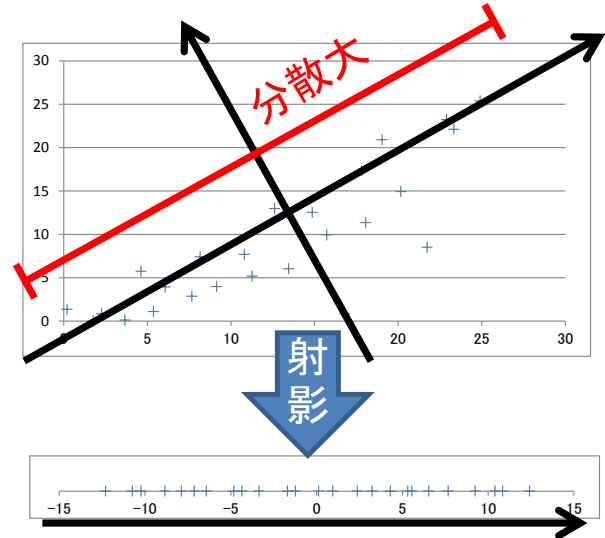


図 5 サンプルデータの軸の変換

図 5 より新しい軸はデータの分散が最も大きくなっている方向に求められており、データの特徴をよく表現する方向は分散であることがわかる。データの分散を尺度に、特異値分解によって固有値、固有ベクトルを得ることを主成分分析 (Principal Component Analysis: PCA) という。主成分とは新しく求められた指標となる軸のことであり、データの値を最も特徴を表している軸の方向を第一主成分、データの値を次に特徴を表している軸の方向を第二主成分という。

図 5 より新しい横軸は、データの特徴を概ね表現できているので、横軸すなわちデータの分散が大きい軸のみでデータを表現しても問題がないと考えられる。データの分散が大きい軸のみで表現すると、図 4 の場合、データの次元を 2 次元から 1 次元に減らすこととなる。1 次元にすることで、一つの値で特徴を表現することが可能となった。このように主成分分析を行うと、結果的にデータの次元が減少することがある。この時、分散の小さい軸の情報が失われる。これを情報の損失という。顔認識技術では、少ない情報量 (低次元) で、質の高い情報 (次元圧縮による情報の損失を最小限にする) を得るために主成分分析を行う。

顔認識では、学習データの差分画像に主成分分析を行うことになる。画像は画素数次元のベクトル情報と考えられるが、多くの場合各画素の情報は互いに相関がある（例えば、暗い画素の隣は暗い可能性が高いなど）。主成分分析を行うことで、学習画像の中で相関のある画素同士をまとめて表現できる新しい軸（固有ベクトル）を求めることが出来る。固有ベクトルを可視化し、上位 50 位までを一覧表示したものが図 6 である。画素同士の相関の強さが濃淡で表現されており、人の顔のようにになっていることがわかる。これを固有顔といい、顔画像は固有顔と平均顔の重み付き加算で表現することができる。

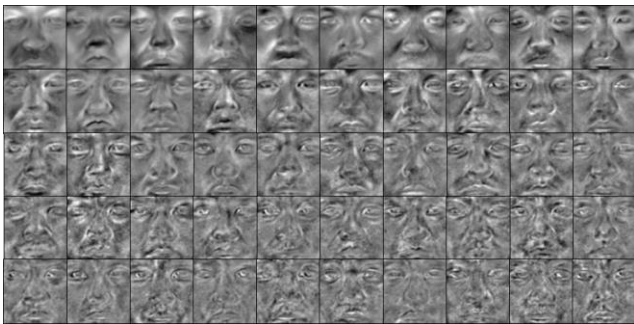


図 6 固有顔一覧(PCA)

### 2.2.3 線形判別分析

主成分分析を顔認識に適応することで各顔画像の特徴を表すのに有効な成分を抽出することが出来た。次に、顔画像間ではなく、人物間で異なる特徴を抽出するために、線形判別分析を行う。

図 7 のような 2 変数、2 クラスのデータを例にとる。図 7 のデータをそれぞれのクラスの特徴が最も出るような表現をしたい。この時、図 8 のような軸をとると、2 クラスを線形に分離できる。新しい軸は、同じクラスに属するデータは一カ所に集まり、クラスが異なるデータは離れた位置になるような方向になっている。このような状態を「クラス間分散が最大かつクラス内分散が最小」という。

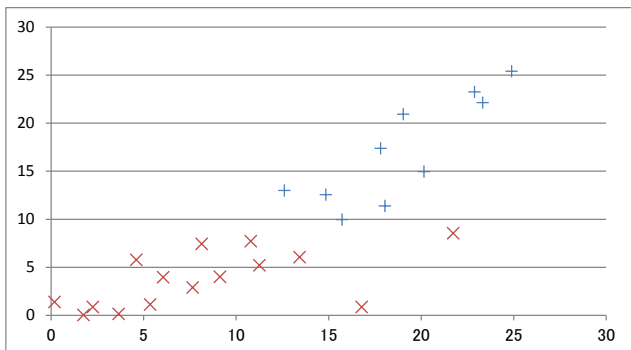


図 7 サンプルデータ 2

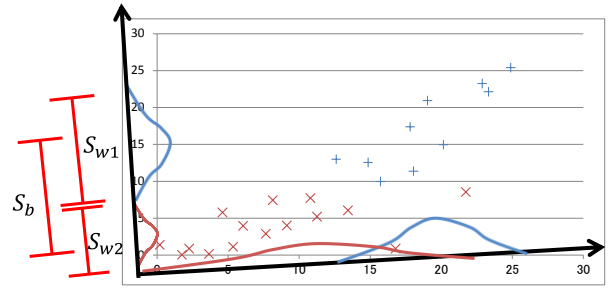


図 8 サンプルデータ 2 の軸の変換

クラス間の分散  $S_b$  は、式(2)から求められる。

$$S_b = \sum_{n=1}^N m_n \|\bar{f}_n - \bar{f}\|^2 \quad \dots(2)$$

- $\bar{f}$  : 全データの平均
- $\bar{f}_n$  : クラス内データの平均
- $m_n$  : クラス内のデータ数
- $N$  : クラス数
- $\|\bar{f}_n - \bar{f}\|^2$  :  $\bar{f}$  と  $\bar{f}_n$  のユークリッド距離

クラス内の分散  $S_w$  は式(3)から求められる。

$$S_w = \sum_{n=1}^N \left( \sum_{m=1}^{m_n} \|f_{nm} - \bar{f}_n\|^2 \right) \quad \dots(3)$$

- $f_{mn}$  : クラス内データ
- $\bar{f}_n$  : クラス内データの平均
- $m_n$  : クラス内のデータ数
- $N$  : クラス数
- $\|f_{mn} - \bar{f}_n\|^2$  :  $f_{mn}$  と  $\bar{f}_n$  のユークリッド距離

したがって、「クラス間分散が最大かつクラス内分散が最小」にするには、 $\frac{S_b}{S_w}$  が最大になるような新しい軸を定めればよい。

このように、クラス情報を教師情報として与え、クラス内分散とクラス間分散を尺度とし、特異値分解によって固有値、固有ベクトルを得ることを線形判別分析(Linear Discriminant Analysis: LDA)という。

実際の学習画像で、主成分分析のみを行った場合とその後線形判別分析を行った場合を比較して、効果の確認を行い、クラス内分散及びクラス間分散の比較を表 1 に示した。線形判別分析を行った方が、クラス内分散におけるクラス間分散の割合が大きくなった。

表 1 クラス内分散とクラス間分散の比較

	PCA	PCA+LDA
クラス内	279121196.48	7507841.92
クラス間	1344191410.15	682277647.01
クラス間 クラス内	4.82	90.88

また、ある人物に注目した第一主成分と第二主成分の散布図を作成し、図 9、図 10 に示した。図 9 では同じ人物の分布が広がって、他の人物との境界が分離できないが、図 10 では分離できていることが見て取れる。以上より、線形判別分析を行ったことで、より個人判別がしやすくなったと考えられる。

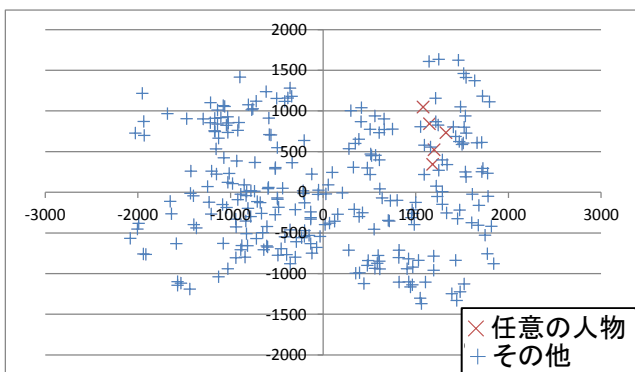


図 9 PCA のみを行った場合の第一主成分、第二主成分の分散図

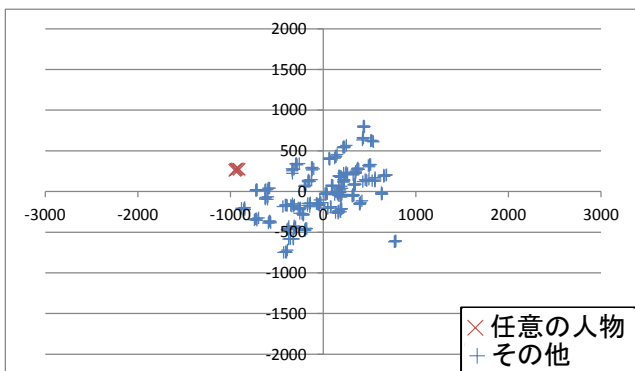


図 10 PCA+LDA を行った場合の第一主成分、第二主成分の分散図

1 人を 1 クラスとし、53 人の学習画像で主成分分析を行った後に線形判別分析を行った時の上位 50 位までの固有顔を図 11 に一覧表示した。

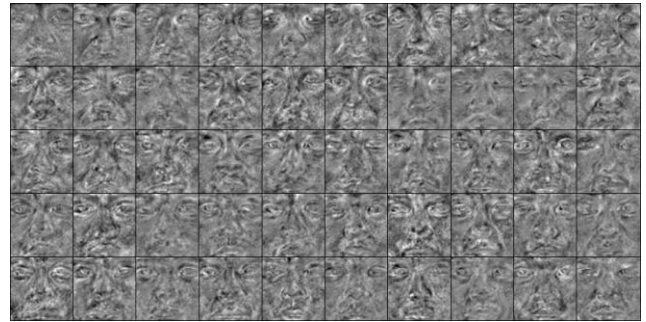


図 11 固有顔一覧(LDA)

### 3. 位置合わせの強化

#### 3.1.1 現在の技術の問題点

顔認識に使用する学習画像や検査画像は、顔検出技術によって画像中から顔部分のみを抽出して作成している。顔検出技術は、指定した領域が顔であるか否かを返す 2 クラス判別器を用いる方法が一般的で、本手法でも Haar-like 特徴量を用いた検出器を使って画像内を均等にサンプリングして検索する方法をとっているため、サンプリング間隔未満の微小な位置ずれを内包する(図 12)。また、平均顔との差分画像を用いて照合を行うため、顔の角度や縮尺の違い等、顔の特徴以外が特徴として抽出されてしまう問題もある(図 13)。

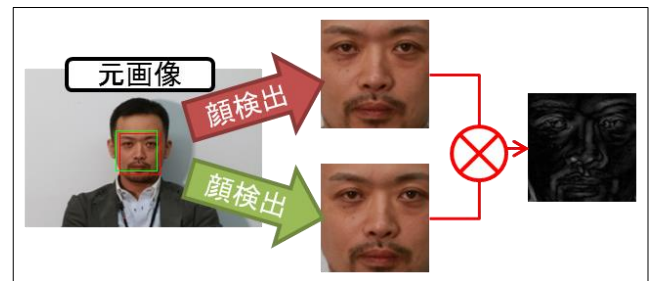


図 12 Haar-like 検出器を用いた顔検出

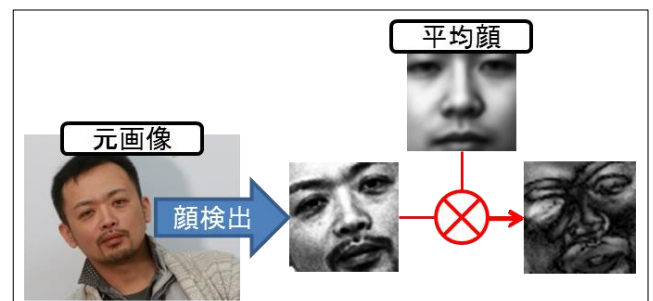


図 13 平均顔と顔の角度が異なる場合の特徴抽出

以上のような顔の位置ずれを解消して顔の特徴だけを用いて照合を行うために、学習画像と検査画像の位置合わせを強化する。

### 3.1.2 学習画像の位置合わせ

- 1 学習画像は平均顔作成時に顔の位置合わせを行う．具体的な手順を下記に示す．学習画像から平均顔を作成する
- 2 画像毎に、平均顔との誤差が少なくなるように変形(平行移動・回転拡大縮小)を行う
- 3 2の手順で得られた変形後の画像を用いて平均顔を作成し直す
- 4 平均顔の変化量が許容誤差以内になるまで 2~3 を繰り返す

平均顔位置合わせした結果を図 14 に示す．位置合わせを行ったことによって、平均顔が鮮明化されたことが確認できる．また、クラス内分散、クラス間分散を見てみても、クラス内分散におけるクラス間分散の割合が位置合わせを行った場合の方が大きくなっており、より位置ずれによるノイズが軽減されて、各クラス(個人)が持つ特徴が抽出されていると考えられる．

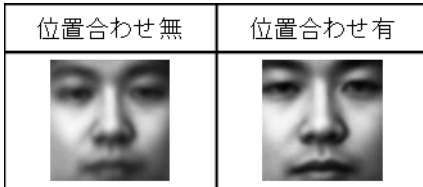


図 14 学習画像の位置合わせを行った時の平均顔の比較

表 2 学習画像の位置合わせによる  
クラス内分散、クラス間分散への影響

	鮮明化無	鮮明化有
クラス内	7507841.92	4269772.01
クラス間	682277647.01	526733967.69
クラス間 クラス内	90.88	123.36

### 3.1.3 検査画像の位置合わせ

検査画像についても、平均顔との誤差が少なくなるように、平行移動・回転・拡大縮小を行うことで位置合わせを行う．一例として図 15 顔検出前の検査画像の検査画像の位置合わせを行わない場合と位置合わせを行った場合の比較を行った．比較結果を図 16 に示した．位置合わせを行わない場合は顔検出結果で顔が傾いているが、位置合わせを行った場合は顔がまっすぐであることがわかる．また、平均顔との差分画像を見ても位置合わせを行わない場合よりも位置合わせを行った場合の方が差分が小さくなり、検査画像の顔の位置合わせが出来ていることを確認することができた．



図 15 顔検出前の検査画像


	位置合わせ無	位置合わせ有
顔検出結果		
平均顔との差分		

図 16 位置合わせ無と位置合わせ有の顔検出結果の比較

## 4. 精度評価

### 4.1.1 実験方法

位置合わせの強化の評価のため、位置合わせ強化を行った場合と行わなかった場合の精度の比較実験を行った．解像度 1368\*912、人数 53 人、合計枚数 739 枚の顔写真を照明状態、撮影場所の条件を同一にし撮影を行い、実験に用いた．その中から無作為に 1 人 5 枚の学習画像を選び、残り 374 枚を検査画像とする顔認識データを 3 セット作成した．顔認識データ毎に位置合わせ強化を行った場合と行わなかった場合の顔認識結果を出力した．認識結果から精度を求め、比較を行なった．なお、精度 A は式(4)から求めた．

$$A = \frac{\text{認識が正しかった画像数}}{\text{検査画像数}} \quad \dots(4)$$

### 4.1.2 実験結果

実験結果を表 3 に示す．表 3 からわかるように位置合わせの強化を行うことで位置合わせ前後で誤差が平均 0.41 倍になり、顔認識精度が上昇したことが分かった．

表 3 各データにおける認識精度

	学習画像数	検査画像数	位置合わせ無 (%)	位置合わせ有 (%)	誤差増加率
合計	-	1122	92.16	96.79	0.41
顔認識データ 1	265	374	91.98	97.98	0.37
顔認識データ 2	265	374	91.71	96.26	0.45
顔認識データ 3	265	374	92.78	97.06	0.41

## 5. 結論

平均顔鮮明化、検査画像の位置合わせにより顔認識の位置合わせの強化を行うことで従来手法の精度を向上させることが出来た。またこの結果より、顔認識精度と位置合わせの因果関係があることが明らかにすることが出来た。

顔認識技術でノイズになる要素は二次元の位置ずれだけでなく、以下にあげられる要素について対策をとる必要がある。

- 1 三次元的の位置ずれ(顔の方向)
- 2 学習画像と検査画像の照明状態の違い
- 3 カメラの画角による対象物の写り方の違い

ディープラーニングの登場によってブレイクスルーが起きており、画像認識はIoTの基本インフラになっていく。世の中の最先端技術に追いつくためにも、顔認識技術を技術の尺度として、これからも自動化性能を追求していきたい。

## 参考文献

- [1] Y. Taigman, M. Yang, M. A. Ranzato and L. Wolf. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. CVPR2014.
- [2] H. Fan, Z. Cao, Y. Jiang, Q. Yin and C. Doudou. Learning Deep Face Representation. arXiv:1403.2802.
- [3] beam2d(2014)「顔認証技術: DeepFace と Pyramid CNN」, <<https://research.preferred.jp/2014/03/face-verification-deepface-and-pyramid-cnn/>>(2016/04/05 アクセス)
- [4] Cru(2010)「サルでもわかる顔検出の原理」, <<http://d.hatena.ne.jp/Cru/20100613/1276436975>>(2016/04/05 アクセス)