

# Bucket Distance HashingとMetric Learningを組み合わせた表情変化に頑健かつ高速な顔認識

水野 智也<sup>1,a)</sup> 内海 ゆづ子<sup>1,b)</sup> 岩村 雅一<sup>1,c)</sup> 黄瀬 浩一<sup>1,d)</sup>

概要：大規模なデータベース (DB) に対して高速な顔認識手法の一つに内海らの手法 [1] がある。内海らの手法はクエリの特徴量とのユークリッド距離に近い DB の特徴量を高速に探索する近似最近傍探索手法を用いることで高速な顔認識を実現した。しかし、内海らの手法はクエリの特徴量と同一の特徴量を探索する手法であるため、表情変化に弱いという問題がある。表情変化に頑健な顔認識手法に Cao らの手法 [2] がある。Cao らの手法は Metric Learning を用いることで、表情変化に頑健な顔認識を実現した。しかし、Cao らの手法には処理速度が遅いという問題がある。本稿では、表情変化に頑健な Cao らの手法に内海らの手法で使用されている高速な近似最近傍探索手法を導入する。内海らの手法と Cao らの手法では評価関数が異なるため、近似最近傍探索を直接 Cao らの手法に適用できない。そこで、Cao らの類似度評価関数をユークリッド距離で表現することで、高速化を実現する。100 枚の顔画像 DB での顔認識実験を行った結果、内海らの手法と比べて認識速度を保ったまま、認識率を 10% 上昇させることに成功した。

## 1. はじめに

近年の防犯意識の高まりに伴い、セキュリティシステムへの導入を目的とした様々な個人認識手法が提案されている。それらの個人認識手法の幾つかは、既にサービス化され私達の生活を支えている。セキュリティシステムに用いられている個人認識手法として、虹彩認識、静脈認識、顔認識などがある。このうち、虹彩認識、静脈認識は、識別装置に近付かなければ認識できないため、対象者は自分が認識されていることを自覚している。それらの認識手法と比べて、顔認識は数 m 程度離れた距離からでも認識できるため、認識対象者の意志に関わらず認識できるという利点がある。従って、顔認識はセキュリティシステムだけでなく、防犯カメラを使用して指名手配犯を発見するといった犯罪捜査にも利用可能である。

犯罪捜査に顔認識を用いる場合、犯行現場で撮影された画像を元に、DB に登録されている犯罪者の候補を自動で絞り込むという使い方が考えられる。その際、犯罪歴を持つ人物が多数存在するため、DB が大規模になる。また、犯罪者の顔画像は入手が困難であるため、DB として一人

あたり複数枚の顔画像が必ず使用できるわけではない。さらに、人はいつも同じ表情ではないため、DB とクエリの顔画像の表情が異なる。これらの場合でも高精度な認識を行うためには、表情変化に強く高速な顔認識手法であることが望ましい。大規模な DB に対して高速な手法として内海らの手法 [1]、表情変化に頑健な手法として Cao らの手法 [2] がある。どちらの手法も DB の顔画像として、一人あたり一枚使用することを前提とした手法である。内海らの手法では、クエリと DB の特徴量をユークリッド距離を用いて照合することで認識を行う。この際、Bucket Distance Hashing (BDH) [3] と呼ばれるハッシュを用いた近似最近傍探索手法を導入し、照合の高速化を実現した。しかし、内海らの手法はクエリの特徴量と全く同じ特徴量を探索する手法であるため、表情変化に弱いという問題がある。一方、Cao らの手法では、照合の際に Metric Learning で学習した類似度評価関数を使用することで表情変化への頑健性を実現する。Cao らの手法ではこの Metric Learning により、異なる人物から抽出される特徴量の差を学習する。これにより、異なる人物から抽出された特徴量の距離は遠く、表情が異なる同一人物から抽出された特徴量の距離は近くなるような類似度評価関数が生成される。Cao らの手法では、認識の際クエリの顔画像と DB の全ての顔画像を照合する。従って、DB が大規模になった場合、照合する画像枚数が増えるため、処理速度が低下する。

そこで、本稿では Cao らの手法に内海らの手法で使用さ

<sup>1</sup> 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) mizuno@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

れている BDH を導入することにより、表情変化に頑健かつ高速な顔認識手法を提案する。具体的には、Cao らの手法において、クエリの特徴量との類似度が大きい特徴量を探索する処理に BDH を導入し高速化する。しかし、BDH は近傍点の計算にユークリッド距離を用いているため、バイリニアシミラリティとマハラノビス距離を用いている Cao らの類似度評価関数に直接適用することができない。そこで本稿では、Cao らの類似度評価関数をユークリッド距離で表現する。これにより、Cao らの手法を BDH により高速化できる。

## 2. 関連研究

多くの研究者が表情変化、照明変化、隠れなどの様々な顔画像の撮影条件の変化に対応するための顔認識手法を提案している。Weng ら [4] は Metric Learning を使用して顔の隠れに強い顔認識手法を提案した。Weng らの手法では、クエリの特徴点座標全てを、DB の特徴点の配置と最も重なるように射影変換する。変換後、特徴点の配置がクエリと最も近くなる画像が認識結果となる。この手法では、特徴点をフィルタリングし信頼性の低い特徴点を除外する。これにより、隠れにより特徴点の一部が隠れていても、残りの信頼性の高い特徴点のみを使用したマッチングにより、認識を成功させることができる。この手法は隠れに対して頑健であるが、本稿で想定しているようなクエリと DB の顔画像の表情が異なる場合には適さない。

Yang ら [5] と John ら [6] は Sparse Coding を使用し表情変化に頑健な顔認識手法を提案した。sparse coding では、顔画像から表情変化などのノイズを除いた人物の類似性のみを抽出し、複数枚の顔画像を集めて構成される画像辞書とパラメータの積で表現することができる。従って、認識の際 DB とクエリの人物の類似性のみを比較することができる。また、松井ら [7] は様々な向きの顔画像を登録し、顔の変形に対応可能な可変テンプレートマッチングを採用することにより表情変化に頑健な顔認識を実現した。Kakadiaris ら [8] は顔の目や鼻などのパーツの位置や顔向きを二次元モデルよりも正確に捉えることができる三次元モデルを使用することにより、顔のパーツの位置や顔向きを DB の画像と正確に揃えた状態で比較する表情変化に頑健な顔認識を実現した。福井ら [9] は一人あたり複数枚の顔画像を使用して部分空間を作成し、その部分空間に対して制約部分空間法を適用した表情変化にロバストな顔認識手法を提案した。しかし、これらの手法は全て計算量が大きい。従って、本稿で想定しているような DB が大規模な場合には、処理時間が膨大になってしまうため適さない。

## 3. 従来手法

本章では、提案手法のベースとなる内海らの手法 [1] と Cao らの手法 [2] の概要について説明する。

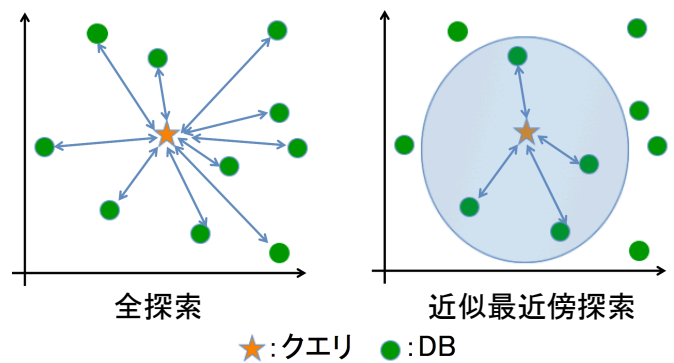


図 1 近似最近傍探索による探索の高速化

### 3.1 内海らの手法

本節では、内海らの手法の特徴抽出と認識処理について説明する。

#### 3.1.1 特徴抽出

内海らの手法では、まず、DB として用いる全ての顔画像から PCA-SIFT[10] 特徴量を抽出する。PCA-SIFT 特徴量は回転、スケール変化などに頑健な特徴量である。そして、全画像から抽出した全ての特徴量を DB に登録する。

#### 3.1.2 認識

認識の際には、クエリからも DB の画像と同様に PCA-SIFT 特徴量を抽出する。そして、クエリの特徴量とのユークリッド距離が近い特徴量を探索し、距離の逆数を重みとして対応する人物に投票する。この投票処理をクエリの特徴量全てに対して行い、得票の最も多い上位  $n$  人を認識結果とする。クエリの特徴量とのユークリッド距離が近い特徴量を探索する際、DB の全ての特徴量と距離計算すると、画像枚数が多くなるにつれて、距離計算しなければならぬ特徴量の数も増え、処理時間が膨大になってしまう。従って、内海らの手法では、精度を犠牲にして距離計算の回数を減らす近似最近傍探索を行う。近似最近傍探索では、上記のような最近傍探索問題において、図 1 のように、クエリの特徴量とのユークリッド距離が近い  $k$  近傍の候補を選択し、選択した特徴量とだけ距離計算する。これにより距離計算の回数を大幅に削減できる。内海らの手法ではハッシュ関数を用いた近似最近傍探索手法である Bucket Distance Hashing(BDH)[3] を用いて高速化を実現した。

### 3.2 Cao らの手法

本節では、Cao らの手法の Metric Learning による類似度評価関数の学習と、認識処理について説明する。

#### 3.2.1 Metric Learning による類似度評価関数の学習

Cao らの手法では、まず、DB として用いる全ての顔画像から  $d$  次元の大域的特徴量を抽出する。そして、DB の全ての特徴量を使用して Metric Learning を行い異なる人物から抽出される特徴量の差を学習する。一般的に Metric Learning では一人あたり複数枚の画像を学習に使用する

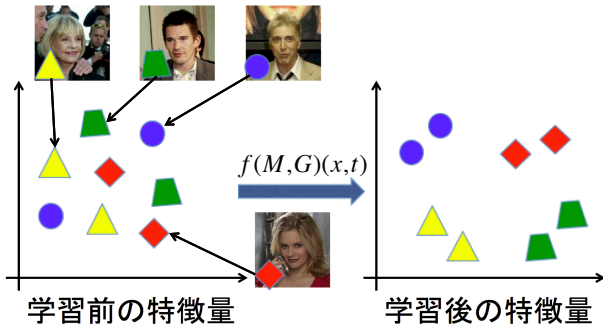


図2 Metric Learningによる距離学習

が, Cao らの手法では一人あたり一枚の画像のみを学習に使用する. 図2にCao らの手法での Metric Learning により, DB の特徴量の距離を学習した例を示す. 図のように, 学習後では異なる人物から抽出される特徴量の距離が遠くなっているため, 表情の変化によりクエリから抽出される特徴量が少し変化しても, 探索の際正しい特徴量に対応づける. Cao らの手法では学習により, このような距離関係を得られる類似度評価関数が生成される. Cao らの手法で類似度評価関数は以下のものが使用される.

$$f(M, G)(x, t) = s_G(x, t) - d_M(x, t) \quad (1)$$

ここで  $s_G(x, t) = x^T G t$ ,  $d_M(x, t) = (x - t)^T M (x - t)$  である.  $x, t$  はそれぞれクエリとDBの  $d$  次元の大域的特徴量を表すベクトル,  $s_G(x, t)$  はバイリニアシミュラリティ,  $d_M(x, t)$  はマハラノビス距離である. また,  $G, M$  はそれぞれ特徴量  $x$  と  $t$  の相関,  $x$  と  $t$  の差の相関を表す対称行列である. Cao らの手法では  $G, M$  を Metric Learning で学習する.

### 3.2.2 認識処理

検索の際には, クエリからもDBの画像と同様に  $d$  次元の大域的特徴量を作成し, クエリの特徴量と類似度が大きなDBの特徴量を全探索により探索する. この際, 学習した類似度評価関数を使用する. そして, 類似度の大きな特徴量が抽出された上位  $n$  人を認識結果とする.

## 4. 提案手法

本章では, Cao らの手法に内海らの手法で使用されているBDHを導入した提案手法について述べる.

### 4.1 Cao らの手法へのBDHの導入方法

本稿では表情変化に頑健かつ高速な顔認識手法を実現する. 具体的には, Cao らの手法において, クエリの特徴量との類似度が大きい特徴量を探索する処理にBDHを導入し高速化する. しかし, BDHは近傍点の計算にユークリッド距離を用いているため, 式(1)のようにバイリニアシミュラリティとマハラノビス距離を用いているCao らの類似度評価関数に直接適用することができない. そこで, 本手法ではまず, 2つの距離尺度で表されているCao らの類似度

評価関数を1つの距離尺度にまとめる. そして, まとめたものをユークリッド距離で表現する. 2つの距離尺度を1つにまとめるために, まず, 式(1)を変形すると,

$$f(M, G)(x, t) = x^T y - t^T M t \quad (2)$$

となる.  $y = (G + 2M)t$  と定義した. ここで, 類似度評価関数  $f(M, G)(x, t)$  が, DB の特徴量  $t$  を行列で射影した  $y$  とクエリ特徴量  $x$  の内積で表されていることに着目する. そして, 今着目している2つの特徴量のユークリッド距離  $\|x - y\|^2$  と類似度評価関数  $f(M, G)(x, t)$  の関係を表すと,

$$-2f(M, G)(x, t) = \|x - y\|^2 - \|x\|^2 + L(t) \quad (3)$$

となる. ここで  $L(t) = t^T \{2M - (G + 2M)^T (G + 2M)\} t$  であり, DB 特徴量  $t$  のみに依存する項である. この段階では,  $-\|x\|^2 + L(t)$  があるため, まだ類似度評価関数  $f(M, G)(x, t)$  を完全にユークリッド距離で表現できていない. そこで, まず  $L(t)$  をユークリッド距離  $\|x - y\|^2$  に加えるために, クエリとDBの特徴量の次元を1次元増加させる. 具体的には

$$x' = (x^T, 0)^T \quad (4)$$

$$y' = (y^T, \sqrt{L(t)})^T \quad (5)$$

のように  $d+1$  次元目の値として,  $x$  に0を,  $y$  に  $\sqrt{L(t)}$  を追加した  $x'$  と  $y'$  を定義する. 1次元追加する前の特徴量のユークリッド距離  $\|x - y\|^2$  と1次元追加した後のユークリッド距離  $\|x' - y'\|^2$  の関係を式で表すと

$$\|x' - y'\|^2 = \|x - y\|^2 + L(t) \quad (6)$$

となる.  $\sqrt{L(t)}$  はDB特徴量のみ依存するので, クエリの特徴量を与えられる前に計算しておくことができる. このように定義された  $d+1$  次元の特徴量  $x'$  と  $y'$  のユークリッド距離を使用して, 式(3)は次のように表される.

$$-2f(M, G)(x, t) = \|x' - y'\|^2 - \|x\|^2 \quad (7)$$

DBの特徴量を探索する際,  $\|x\|^2$  は一定の値なので, 類似度を計算する際  $\|x\|^2$  を無視して考えることができる. 式(7)より,  $d+1$  次元の特徴量のユークリッド距離  $\|x' - y'\|^2$  とCao らの類似度  $f(M, G)(x, t)$  は逆相関の関係を持つことになり,  $f(M, G)(x, t)$  にBDHを適用することができる.  $\sqrt{L(t)}$  を計算する際,

$$L(t) = t^T \{2M - (G + 2M)^T (G + 2M)\} t \geq 0 \quad (8)$$

となる必要がある. そのためには,  $2M - (G + 2M)^T (G + 2M)$  が半正定値行列にならなければならない. さらにそのためには, 少なくとも  $\|M\| \leq 0.5$  となる必要がある.

## 4.2 提案手法の流れ

まず、DBとして用いる全ての顔画像から  $d$  次元の大域的特徴量を作成する。そして、DBの特徴量全てを使い、Metric Learningで相関行列  $G$ 、 $M$  を学習する。求めた行列を使用して大域的特徴量の  $d+1$  次元目の値を計算する。前述の通り、この  $d+1$  次元目の値はクエリが与えられる前に計算できる。検索の際は、クエリについてもDBと同様に大域的特徴量を作成し、BDHを利用して近似最近傍探索を行うことで、クエリの特徴量とユークリッド距離が近い特徴量を探索する。そして、類似度の大きな特徴量が抽出された上位  $n$  人を認識結果とする。

## 5. 提案手法の評価実験

提案手法の認識率と処理時間の評価をするために、提案手法、内海らの手法とCaoらの手法の認識率、処理時間の比較実験をした。

### 5.1 実験条件

実験にはFace in the wild dataset[11]の顔画像を使用した。このデータセットには5749人分の合計13233枚の画像があり、この内1680人分は1人あたり2枚以上ある。また、このデータセットはインターネットから著名人の画像を集めることにより作成されたため、表情変化、照明変化や顔の一部が物体と重なり隠れている画像が多数ある。このデータセットの画像から、顔の切り出しを行い、目や鼻などの位置を揃える正規化と、顔が正面を向くように向きの正規化を行った。実験では正規化に失敗した画像は除外した。画像はすべてグレースケールで、解像度は  $512 \times 512$ [pixel]である。DBとしてこのデータセットの画像を1人につき1枚、合計100枚を使用した。また、クエリとして、DBと同じ人物の異なる表情の顔画像を合計100枚を使用した。クエリとDBの画像例を図3に示す。

クエリとDBから抽出する局所特徴量としてPCA-SIFT[10]特徴量を使用した。特徴量は図4の9箇所から、2, 6, 10の3通りのscaleで抽出した。これらの9箇所から抽出された27個の特徴量は、他の位置から抽出された特徴量と比べて表情変化や照明変化に対して頑健な特徴量となる[12]。また予備実験から10より小さいscaleで抽出した特徴量が認識に寄与することが分かっている。内海らの手法の局所特徴量は27個の局所特徴量をそのまま使用した。Caoらの手法と提案手法の大域的特徴量は、27個の局所特徴量を結合したものを主成分分析により100次元に圧縮することで作成した。100次元に圧縮することで精度を保ったままに、処理時間を最も短くすることができる。認識の際は、認識結果の上位10人の内に正解の画像が含まれている場合、認識成功と判定した。実験に使用した計算機は、CPUがIntel(R) Xeon(R) E5-4627 v2(3.30GHz)、メモリは512GBである。処理時間は特徴量



図3 Face in the wild datasetの画像例



図4 特徴量抽出位置

の検索にかかった時間のみを測定し、画像の正規化や特徴抽出、Metric Learningによる行列学習の時間は含まない。

### 5.2 結果・考察

提案手法、Caoらの手法、内海らの手法の認識率と処理時間を表1に示す。実験の結果、提案手法と内海らの手法を比べると認識率が10%上昇した。これは、Metric Learningにより表情変化に頑健な類似度評価関数を学習できたためと考えられる。また提案手法は内海らの手法と比べて処理時間が低下した。これは内海らの手法は局所特徴量を使用するのに対して、提案手法では大域的特徴量を使用する。従って、内海らの手法では探索の際、1枚のクエリにつき27回探索するのに対して、提案手法では1回探索するだけで良い。また、内海らの手法の方が提案手法と比べてDBの特徴量数が多いため、1回の探索にかかる時間が長いことも要因と考えられる。

次に、提案手法とCaoらの手法を比べると認識率を保ったまま処理時間が約9分の1になった。これは、Caoらの手法の類似度評価関数には行列が含まれているので、クエリとDBの特徴量の類似度を計算する際、計算コストが高い行列計算をしなければならない。これに対して、提案手法は学習時に  $d+1$  次元目の値を求める処理として行列計算を行う。そのため、認識時には、ユークリッド距離の計算を行うだけでよい。このため計算時間が大幅に削減されたと考えられる。実際にCaoらの手法の一人あたりの平均処理時間2.7[msec]のうち、行列計算の時間は一人あたり平均2.2[msec]かかっていた。

表 1 DB100 枚における提案手法と従来手法の比較

条件	認識率 (%)	一人あたりの処理時間 (msec)
内海らの手法	46	0.80
Cao らの手法	56	2.7
提案手法	56	0.30

## 6. まとめ

本稿では, Cao らの手法に内海らの手法で使用されている BDH を導入することにより, 表情変化に頑健で高速な顔認識手法を実現した. その結果, 内海らの手法と比べて認識率が 10% 上昇し, Cao らの手法と比べて処理時間が約 9 分の 1 になった. 今後の課題として, 大規模な DB で実験することが挙げられる.

謝辞 本研究は JSPS 科研費 25240028 の助成を受けた.

## 参考文献

- [1] 内海ゆづ子, 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探索を用いた大規模データベースに対する高速顔認識, 情報処理学会研究報告コンピュータビジョンとイメージメディア, Vol. 2013-CVIM-186, No. 4, pp. 1-7 (2013).
- [2] Qiong, C., Ying, Y. and Li, P.: Similarity Metric Learning for Face Recognition, *Proceedings of International Conference on Computer Vision (ICCV 2013)*, pp. 2408-2415 (2013).
- [3] Iwamura, M., Sato, T. and Kise, K.: What is the most efficient way to select nearest neighbor candidates for fast approximate nearest neighbor search?, *Proceedings of the 14th International Conference on Computer Vision (ICCV 2013)*, pp. 3535-3542 (2013).
- [4] Weng, R., Lu, J., Hu, J., Yang, G. and Tan, Y.-P.: Robust Feature Set Matching for Partial Face Recognition, *Proceedings of International Conference on Computer Vision (ICCV 2013)*, pp. 601-608 (2013).
- [5] Meng, Y., Van, L. and Zhang, L.: Sparse Variation Dictionary Learning for Face Recognition with A Single Training Sample Per Person, *Proceedings of International Conference on Computer Vision (ICCV 2013)*, pp. 689-696 (2013).
- [6] Wright, J., Ganesh, A., Sastry, S. and Ma, Y.: Robust Face Recognition via Sparse Representation, *Pattern Analysis and Machine Intelligence (IEEE 2009)*, Vol. 31, No. 2, pp. 210-227 (2009).
- [7] 松井 淳, Simo, C.: 顔テンプレートの摂動による表情変化への顔認識ロバスト性向上, 電子情報通信学会技術報告, Vol. 103, No. 455, pp. 73-78 (2003).
- [8] Kakadiaris, L. A., Passalis, G., Toderici, G., Murtuza, M. N., Lu, Y., Karampatziakis, N. and Theoharis, T.: Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach, *Pattern Analysis and Machine Intelligence (IEEE 2007)*, Vol. 29, No. 4, pp. 640-649 (2007).
- [9] 福井和広, 山口 修, 鈴木 薫, 前田賢一: 制約相互部分空間法を用いた環境変動にロバストな顔画像認識, 電子情報通信学会論文誌, Vol. 82, No. 4, pp. 613-620 (1999).
- [10] Ke, Y. and Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 506-513 (2004).
- [11] Gary B. Huang, Ramesh, M., Berg, T., Learned-Miller, E.: Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical report, University of Massachusetts, Vol. 1, No. 2 (2007).
- [12] Everingham, M., Sivic, J. and Zisserman, A.: 'Hello! My name is... Buffy'-automatic naming of characters in TV video, *Proceedings of the 17th British Machine Vision Conference (BMVC 2006)* (2006).