

顔認識における特定物体認識と一般物体認識の比較

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

概要：物体認識手法は扱う問題によって特定物体認識と一般物体認識の二種類が使い分けられる。しかし顔認識は両方のアプローチで高精度な認識に成功しており、どちらのアプローチが適しているのかは明らかにされていない。どちらかのアプローチの優位性や、それぞれのアプローチが特に有効なタスクなどを示すことができれば、今後の顔認識技術の発展につながると思われる。本稿では既存手法を利用した比較実験を通して、顔認識における特定物体認識と一般物体認識それぞれの有用性について考察する。また顔認識における各要素の重要度を示すため、一般物体認識による顔認識においてどのような visual words が有効であるのかを調べる。実験の結果、認識率の低かった表情変化の大きいクエリについて他のクエリよりも大きな差が生まれ、一般物体認識の方が高い認識率を示した。visual words に関する検証実験では、顔認識において位置情報が必要不可欠な要素であるということを示した。

1. はじめに

物体認識手法は扱う問題によって特定物体認識と一般物体認識の二種類が使い分けられる。固有名詞で表されるような特定の物体を認識する問題には特定物体認識とよばれる手法が用いられ、対象とする物体があるカテゴリに属するかどうかを判定する問題には一般物体認識とよばれる手法が用いられる。カテゴリと一口に言っても大小様々なものがあり、小さなカテゴリを扱う問題の中にはどちらの手法が認識するのに適しているのか曖昧な問題もある。その中の一つが顔認識である。

顔認識の扱う問題は人物認識であり、特定の人物の顔の認識は前述の定義からすると特定物体認識の扱う問題と考えられる。しかし、人の顔には表情変化があり、表情変化によって人物認識の精度が下がるという問題がある。認識できない場合があるということは、人の顔が表情変化によって別のものに変化していると考えられる。その場合、特定の人物の顔というのは複数の表情を含む一つのカテゴリだと言え、人物判別問題は一般物体認識を用いるべきカテゴリ分けの問題ということになる。実際に特定物体認識と一般物体認識それぞれのアプローチで高精度な顔認識に成功している手法が既に発表されている。しか

し、特定物体認識と一般物体認識のアプローチは全く異なるものであるため、それぞれの顔認識技術の研究は独自に発展してきた。そこで、特定物体認識と一般物体認識のどちらが優れているのか、またそれぞれのアプローチがどのようなタスクに対してより頑健であるのかを明らかにすることで今後の顔認識技術の発展に貢献できる。どちらかの手法が優れていると明らかにできれば、今後は主に優れている方の手法を研究することや、一方が他方よりも優れていた理由を分析することで顔認識技術の向上につながる。また、それぞれのアプローチが得意とするタスクを明らかにすることができれば特定物体認識と一般物体認識それぞれの今後の顔認識技術の研究につながる。

近年、顔認識技術は様々な場面で実用化されており、セキュリティシステムとして使用されるだけでなく防犯カメラなどの映像から特定の人物を検索するのにも利用されるようになった。また、このような顔認識の需要の増加に伴って大規模なデータベースを想定した顔認識技術の需要も高まっている。例えば犯罪捜査において、映像中に指名手配されている人物が映っていないか調べるためには、大量の映像の中から多くの指名手配犯を登録した顔データベースを用いて探索する必要がある。他にも空港の入国管理システムなど、大規模なデータベースにおける顔認識技術の需要は様々ある。本稿で述べる比較実験は顔認識技術の今後の発展のためのものであり、こうした背景を鑑みるに大規模なデータベースを用いた認識も想定している顔認識手法を比較に用いるべきだと考える。本稿では大規模なデータベースにおいて高精度な顔認識に成功している手法として、内海らの手法 [1] と Wu らの手法 [2] という二つ

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

の手法を扱う。内海らの手法は局所特徴量のマッチングを利用した特定物体認識の手法である。一方 Wu らの手法は visual words を利用した一般物体認識の手法である。内海らの手法と Wu らの手法はともにそれぞれの実験条件でその有効性を示している。しかし、両者の比較実験の結果はまだ無い。また、Wu らの手法では一般的に用いられる一般物体認識手法による顔認識の精度が良くなかったため、特殊な visual words を用いることで顔認識の精度を向上させることに成功している。Wu らの手法の visual words と一般的な visual words との相違点は複数あるが、各要素がどの程度の影響を及ぼしたのかについては言及されていない。各要素による影響の大きさを調べることで、顔認識における各要素の重要度を明らかにすることができる。

本稿ではまず、この二つの手法の比較実験をすることにより、顔認識における特定物体認識と一般物体認識の有効性について考察する。また、一般物体認識の visual words を変えて比較実験することで、顔認識に有効な visual words において重要な要素を調べる。以下 2 章で特定物体認識と一般物体認識に関する研究について述べ、3 章では本稿で扱う内海らの手法と Wu らの手法について述べる。そして 4 章では特定物体認識と一般物体認識の比較、異なる方法で作成した visual words の比較それぞれの内容について述べる。5 章で実験条件の説明と実験結果についての考察をし、6 章で本稿についてまとめる。

2. 関連研究

2.1 特定物体認識

特定物体認識では画像から抽出された局所特徴量を認識に用いる。各特徴点から抽出された特徴量同士を直接マッチングすることで比較し、その結果が各画像同士の類似度計算に用いられる。多くの特徴点を用いることで物体の向きの変化や、一部が隠れることに対してより頑健になる。

局所特徴量を用いて認識を行う特定物体認識手法はこれまで様々なアプローチでの研究がなされている。局所特徴量の記憶に必要なメモリ量を量子化によって減らしたり、索引付けを行うことで処理時間を短縮したりといった手法がある [3]。内海らの手法は特に大規模なデータベースにおける顔認識に焦点を当てた手法である。内海らの手法は近似最近傍探索を用いて大規模なデータベースにおける高速な認識を実現しており、かつ投票処理によって高精度の維持にも成功している。将来的に大規模なデータベースに対して有効な顔認識手法の需要はあるはずである。何より、比較対象である一般物体認識は大規模なデータベースにおける認識に適した手法である。そのため、本稿では内海らの手法を特定物体認識の代表として用いて比較実験を行う。

2.2 一般物体認識

文章検索モデルの一つに bag of words というものがあ

る。bag of words では文章中の単語の並び順などは考慮せず、含まれる各単語の数を表すヒストグラムで文章を表す。文章のカテゴリによって含まれやすい単語には傾向があるため、このヒストグラムを用いれば文章のカテゴリ分けができる。一般物体認識はこの bag of words を画像検索に応用した手法である。画像を特徴量の集合だと考えると、例えば bag of words における文章を画像、単語を特徴量といった様に置き換えて考えることができる。一般物体認識では画像から得られた特徴量を用いてヒストグラムを作成し、そのヒストグラム同士を比較することで認識を行う。

一般物体認識では bag of words を画像に応用するためヒストグラム作成の際に visual words という特徴量の辞書を用いる。bag of words における単語とは違い、特徴量は全く同じものが得られる可能性が低い。そこで、一般物体認識では特徴量をクラスタリングし、各クラスタに含まれる特徴量の数をカウントすることでヒストグラムを作成する。そのため一般物体認識ではまずクラスタとして、特徴量の辞書である visual words を作成する。一般的には学習用画像から抽出された特徴量を特徴空間上でクラスタリングした際のセントロイドを visual words とする。そしてクエリやデータベースから抽出した特徴量が、特徴空間上での visual word に近いかを調べ、最も近い visual word に投票を行う。この投票作業によって各画像から作られるヒストグラムを元の画像の代わりにクエリやデータベースとして扱う。

本稿で用いた Wu らの手法は顔認識における精度を向上するため一般的な一般物体認識の手法とは異なる点がいくつかある。一般的な一般物体認識手法ではまず、DoG[4] や MSER[5] といった検出器によって配置した特徴点から抽出した特徴量をクラスタリング [6][7] することで学習し作成した visual words を使用する。クエリやデータベースの画像からも同様に特徴量を抽出し、visual words を用いてヒストグラムに変換する。このヒストグラム同士のユークリッド距離などで元の画像同士の類似度の判定をする。しかし、このような一般的な方法は顔画像に適用した途端に精度が急激に下がることがわかっている。そのため、近年顔認識に有効な一般物体認識手法の研究が行われており、特徴量に位置情報を付加するアプローチ [8] が示されている。しかし、顔画像から抽出される高次元の特徴量を扱うと情報量が大きくなってしまい実用的な方法ではない。その中で一般物体認識を顔認識に有効なものにすることに成功した例の一つが Wu らの手法である。この手法では visual words の作成を一般的なものとは異なる方法で行う。また一般的な一般物体認識同様、Wu らの手法の類似度計算の方法は顔画像を変換したヒストグラムを使用するため、データベース画像をヒストグラム化する時間を考慮に入れなければ高速な認識が可能である。よって大規模データベースでも運用可能な手法であるため、一般物体認

識手法の代表として内海らの手法との比較実験に用いるのに適している。

3. 関連手法

3.1 内海らの手法

内海らの手法では前処理として顔領域の切り出しと顔の角度や縮尺の正規化を行い、目や鼻の位置を合わせた顔画像を用いる。特徴量としては顔認識に用いられる特徴量の一つである PCA-SIFT[9] を用いる。特徴点の決定方法は扱う画像によって二通りの方法が使い分けられている。一つは検出器により特徴点を配置する方法であり、もう一つは dense sampling により特徴点を配置する方法である。検出器による特徴点検出ではエッジなどの特徴量変化が大きい点を検出することで効率よく特徴量を抽出できるという利点がある。しかし、解像度が低い画像からは顔認識に用いるために十分な数の特徴点を検出できないという欠点がある。図 1 の顔画像上の赤い点は特徴点を表しており、5 枚の顔画像はそれぞれの範囲から特徴抽出した特徴点の配置である。特徴点を予め固定しておくことにより画像の解像度に依らず一定数の特徴量を得ることができる。

特徴量のマッチングは全ての特徴点同士の組み合わせで行われるわけではなく、座標によりマッチングする特徴量を限定する。検出器を用いて特徴点検出をした場合は、一定距離内の座標にある特徴点同士のみマッチングを行う。dense sampling を用いて特徴点を配置した場合は、同じ座標の特徴点同士でマッチングを行う。顔画像はあらかじめ切り出しと正規化をしてあるため、座標が近い特徴点同士を比較することで目から抽出された特徴量と鼻から抽出された特徴量といった明らかな誤対応である組み合わせのマッチングを省くことができる。そのような組み合わせを省くことで処理時間の短縮になり、誤対応を防ぐことで認識精度の向上にもつながる。

特徴量同士のマッチングにはユークリッド距離に基づいた近似最近傍探索を用いる。検索画像の特徴量との類似度が高いデータベース画像の特徴量を一定数探索し、その特徴量が抽出されたデータベース画像に投票を行う。特徴量を抽出してから、最近傍の特徴点をもつ画像に投票するまでの流れを図 2 に示す。投票は見つかった特徴点とのユークリッド距離に反比例した重み付きで行うことで絞り込みの精度を向上させる。検索画像の全ての特徴量についてマッチングと投票処理を行い、各データベース画像に対する投票結果がクエリ画像との類似度を表す計算結果となる。近似最近傍探索を用いることで処理を高速化できる一方で、精度は実際の最近傍探索には劣る。しかし正解画像へのマッチング結果とは違い、誤対応であるマッチング結果は複数の画像に投票結果が拡散することが多い。そのため一つ一つの特徴量のマッチングの精度はそれほど高くなくとも、投票処理を介することで正解画像の探索精度を上

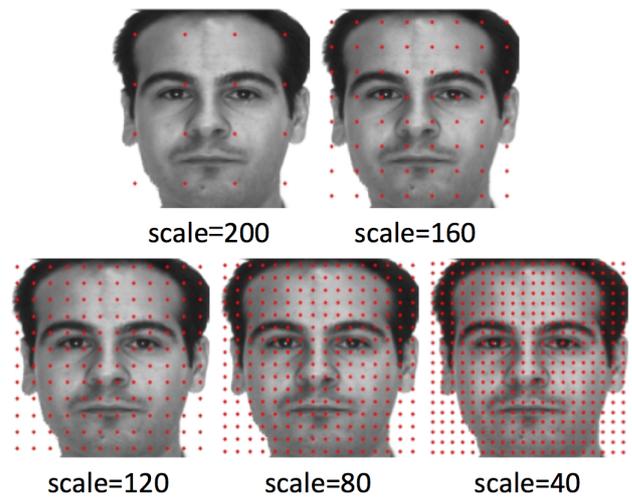


図 1 dense sampling による特徴抽出範囲毎の特徴点の配置

げることができる。また、見つかった近傍点と検索画像の特徴量との距離が大きければ大きいほど、見つかった特徴点の信憑性は低くなる。そのため、距離に応じた重み付きで投票処理を行うことで誤対応の影響を減らし、正解画像の探索精度をさらに上げることができる。

3.2 Wu らの手法

一般的な visual words の作成時は、学習用画像全体から検出器によって検出された特徴点から特徴量が抽出される。この時、特徴量以外のパラメータは扱われず、特徴点の位置情報は失われる。一方で Wu らの手法では目などの特徴が出やすい位置を中心に合計 175 点の固定された各位置に特徴点を配置し、各位置について visual words のセットが作成される。クエリとデータベースにも同じ位置に特徴点が配置され、同じ位置から作成された visual words のセットを用いて投票が行われる。そして全ての位置の投票結果を繋げたヒストグラムが作成される。つまり、位置ごとに一般物体認識を行うことで同じ位置の特徴量同士の認識を行うため、全体で見れば位置情報を利用した一般物体認識を実現している。Wu らの手法の visual words が含む要素と、特徴空間上での分布例のイメージ図を図 4 の (1) に示す。位置情報を利用することで全く違う位置の特徴量との誤対応の可能性を無くすることができる。また各位置の visual words の作成の際に行われるクラスタリングでは、一般的に行われる特徴量によるクラスタリングをせずに学習用画像に写っている人物ごとにクラスタを作成する。特徴量によるクラスタリングを行った場合は、同じような照明条件や同じような表情をした顔画像の特徴量が一つのクラスタを形成する可能性もある。人物ごとにクラスタを作成すればそのような撮影条件などの影響を受けたクラスタをなくすことができ、より人物認識に有用なクラスタを作成できると考えられる。一方で図 4 の (1) を他の 2 種類の visual words と比べると不便な点もある。人物ごとのクラ

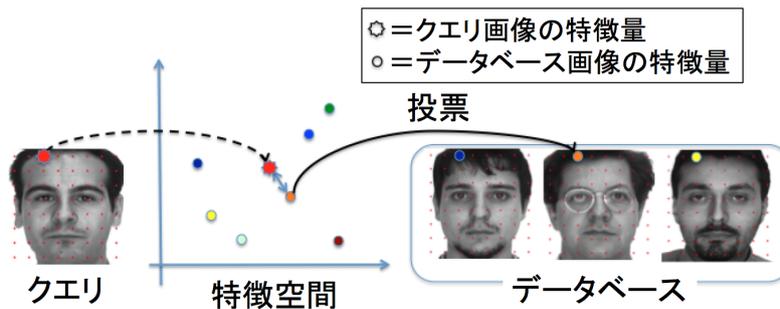


図 2 内海らの手法の大まかな流れ

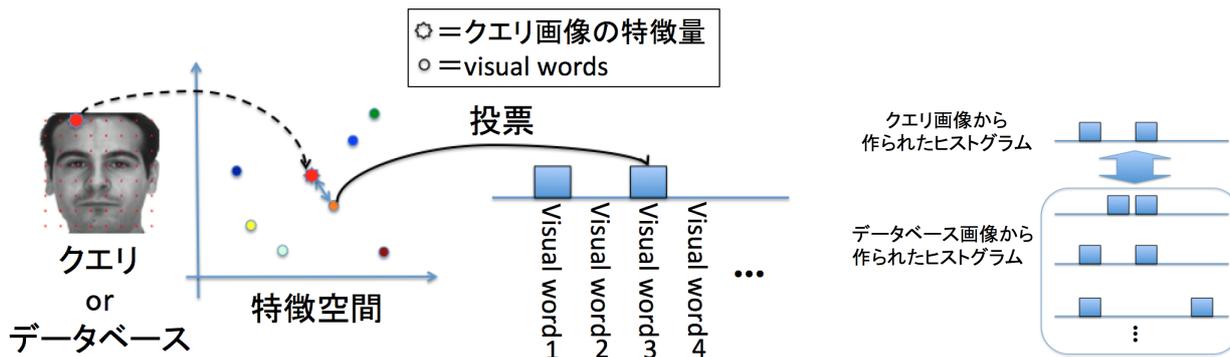


図 3 Wu らの手法の大まかな流れ

スタリングでは特徴量によるクラスタリングとは違い、同一のクラスタに含まれる特徴量が特徴空間上でひとかたまりになっているわけではない。そのため、クラスタごとのセントロイドを利用することができず、ヒストグラム作成の際には特徴量一つ一つとマッチングする必要がある。

Wu らの手法では一般物体認識手法によるクエリとの類似度計算によってデータベースの画像をランク付けした後に、データベース内で特にクエリとの類似度の高かった画像を用いて再ランク付けを行うことで精度を向上させる。まず再ランク付けを行うのに必要な参照画像としてクエリとの類似度が最も高いデータベース画像を選ぶ。そして、その参照画像との類似度が高い画像をデータベース内の上位の画像から選ぶ。新たに選ばれた画像は参照画像の次に再ランク付けされ、参照画像のセットに加わる。この作業を繰り返すことでクエリとの類似度の高い画像の再ランク付けを行う。一枚のクエリ画像のみとの比較で上位にランク付けされた不正解画像が他の正解画像との類似度も高くなる可能性は、正解画像のそれよりも低いと考えられる。よってこの再ランク付け処理によって誤対応を起こした画像を類似度の上位から排除することができ、探索結果の精度向上につながる。

4. 比較内容

4.1 特定物体認識と一般物体認識の比較

特定物体認識と一般物体認識の比較をするため、それぞれ内海らの手法と Wu らの手法を用いて同じ実験条件

での認識率を比較した。Wu らの手法では元々内海らの手法とは異なる特徴量を使用していたが、本実験では共に PCA-SIFT を用いた。これは本実験の目的はあくまでも特定物体認識と一般物体認識という二つのアプローチの比較であり、内海らの手法と Wu らの手法の比較とは異なるためである。同様の理由により、Wu らの手法でデータベースを一度ランク付けした後に行われる再ランク付けによる精度向上も省いた。よって今回一般物体認識の代表として実験に用いた手法は厳密には Wu らの手法とは異なる。しかし、Wu らの手法の visual words を用いているため、以降この手法も Wu らの手法と称することとする。

この比較によって特定物体認識を用いた顔認識手法と一般物体認識を用いた顔認識手法を比べることができる。顔認識において二つのアプローチのどちらが優れているのかを明らかにするとともに、表情変化や照明変化を含むクエリを用いることでそれぞれのアプローチが得意とするタスクの有無も調べられる。

4.2 異なる方法で作成した visual words の顔認識における性能の比較

Wu らの手法では一般的に用いられている visual words とは異なる方法で作成された visual words を用いている。これは一般的な visual words を用いた際の顔認識精度が悪かったためである [2]。このことから一般物体認識においてどのような visual words を用いるかということは認識率に深く関わる要因だということが言える。本稿では特定

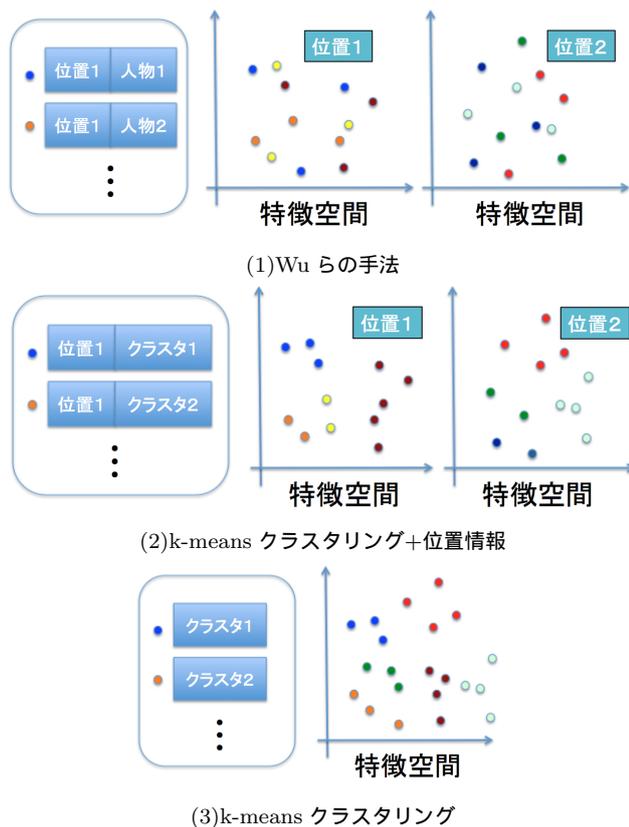


図 4 作成方法毎の visual words

物体認識と一般物体認識の性能比較という目的に加え，具体的に visual words のこういった性質が顔認識精度向上につながるのかという点についても検証をする．

Wu らの手法で用いられた visual words が一般的な visual words と大きく異なる点が二つある．一つは画像上の位置ごとに visual words を作成する点である．これにより特徴点の位置情報を利用する一般物体認識を実現している．もう一つは人物ごとのクラスタを作成している点である．本実験ではこの 2 点の visual words の性質の有効性を検証するため，一般的な方法と同じく特徴量クラスタリングを用いた visual words を位置ごとに作成した．そしてこの検証用 visual words と Wu らの手法の visual words ，一般的な visual words の三つによる顔認識精度を比較した．検証用 visual words と Wu らの手法の visual words を比較することでクラスタリング方法の違いによる認識率への影響を調べられ，検証用 visual words と一般的な visual words を比較することで位置ごとに visual words を作成することによる認識率への影響を調べることができる．

visual words の各要素の顔認識率への影響を明らかにすることで顔認識により有効な visual words の研究につながる．本稿で検証する，人物ごとのクラスタを作成する特徴量クラスタリングの方法が大きな影響を及ぼしていることを明らかにできれば，新たなクラスタリング方法を模索することの有用性を示すことができる．また，特徴点の位置情報が大きな影響を及ぼしていることを明らかにでき

れば，新たな位置情報の利用方法や位置ごとの重みなどといった，位置情報に関する手法の研究の有用性を示すことになる．さらに，顔認識における visual words の要素として重要な値は顔認識全般においても重要な要素であると考えられるため，特定物体認識による顔認識における特徴点の位置情報の重要性の証明にもつながる．

5. 実験

本稿では，特定物体認識と一般物体認識の比較実験と顔認識に有効な visual words の検証実験の二つの実験について述べる．

5.1 実験条件

データベースとして 1 万枚の顔画像データを用いた．1 万枚の顔画像データベースは AR Face Database[10] の無表情，偏りのない照明で撮影した人物の重複のない 132 枚，CAS-PEAL[11]，Surveillance Cameras Face Database[12]，Indian Face Database^{*1} のデータベース画像と，web 上からダウンロードした顔画像から構成される．web 上からダウンロードした顔画像は，重複したものを除いており，全く同じ画像は含まれない．クエリとして，AR Face Database の表情変化と照明変化を 6 パターンを 132 人分を低解像度にしたものを用いた．表情変化は笑っている表情の smile，怒っている表情の anger，叫んでいる表情の scream の 3 セットであり，照明変化は顔の左右それぞれから光を当てた 2 セットと両側から光を当てた 1 セットで構成されている．AR Face Database のクエリ画像は元々 512 × 512 [pixel] だったものをバイキュービック法により 25 × 25 [pixel] に縮小した後，512 × 512 [pixel] に拡大することで低解像度にしたものを用いた．使用した AR Face Database の各パターンの画像例を図 5 に示す．低解像度画像をクエリに用いたため，内海らの手法の特徴点配置は検出器による方法ではなく dense sampling による方法を用いた．また，一般物体認識における学習用画像として，Face in the wild[13] [14] [15] から取得した 158 人の一人当たり 10 枚の合計 1580 枚の画像を用いた．Face in the wild は実環境で撮影された人物画像の顔領域周辺を切り出した顔画像で構成されているデータセットであり，様々な顔の向きや照明条件で撮影された顔画像を含む．顔画像は全て顔の切り出し，顔の角度や縮尺の正規化を行い目や鼻の位置を合わせたものを用いた．k-means クラスタリングによって visual words を作成する際のクラスタ数は，位置ごとに作成する場合は k=50，顔画像全体から作成する場合は k=500 とした．また，100 位までの累積分類率を認識率として評価を行った．実験に使用した計算機の OS は Mac OS 10.10.1，CPU は 2.7GHz Intel Core i7，メモリは

^{*1} <http://vis-www.cs.umass.edu/vidit/IndianFaceDatabase/index.html>

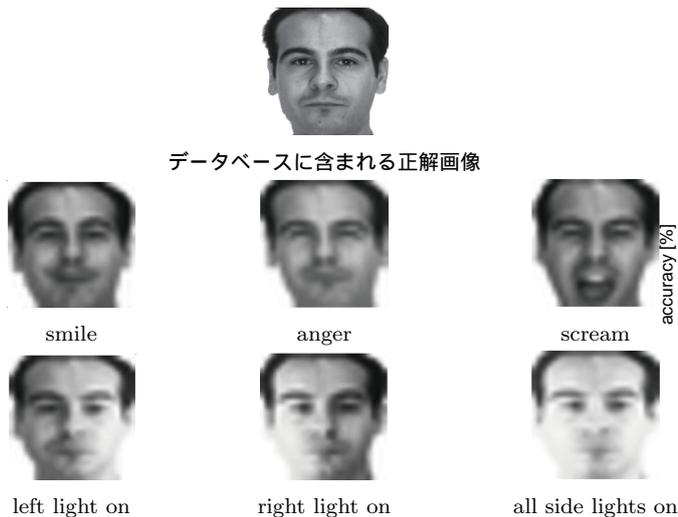


図 5 使用した AR Face Database の画像例

表 1 各手法のクエリのパターンごとの認識率 [%]

	smile	anger	scream	left	right	all	average
(1)	95.5	96.2	18.9	95.5	93.9	59.8	76.6
(2)	91.7	97.0	28.0	95.5	96.2	58.3	77.8
(3)	89.4	95.4	25.8	95.5	87.1	63.6	76.1
(4)	15.2	31.1	5.3	7.6	3.8	2.3	10.9

(1):内海らの手法, (2):Wu らの手法, (3):k-means クラスタリング+位置情報 (4):k-means クラスタリング

16GB 1600MHz DDR3 であった。

5.2 結果と考察

まず特定物体認識と一般物体認識の比較実験の結果について述べる。表 1 の (1) と (2) の 2 行がそれぞれ内海らの手法と Wu らの手法の認識率であり、図 6 はクエリの各種類に対する 2 手法の認識率をグラフ化したものである。表情変化でも照明変化でもどちらか一方の優位性を示す傾向は見られなかった。AR Face Database 内の顔画像は全て正面を向いており、照明変化も意図的に変化させた左右からの光以外の不規則な変化はない。したがって AR Face Database のほとんどのパターンは顔認識を高精度に行うのが難しいデータセットではない。そのため、どちらの手法もほとんどのクエリに対して高精度な認識結果を示したため差が現れにくかったと考えられる。しかし AR Face Database の中でも scream だけは特に極端な表情変化であり認識率が他のパターンよりも低く、scream での認識結果は 9.1% と最も大きな差が出た。このことから、より難易度の高いデータセットを用いれば比較する 2 手法の差別化が図れると考える。したがって Face in the wild などの難易度の高いデータセットを用いた比較をすることが今後の課題として挙げられる。

次に顔認識に有効な visual words の検証実験の結果について述べる。表 1 の (2) と (3) と (4) の 3 行がそれぞれ

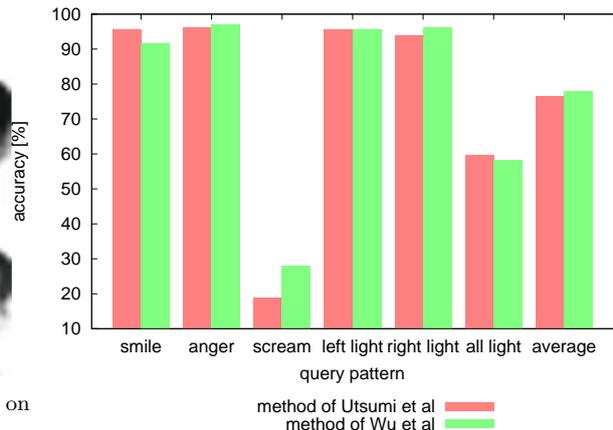


図 6 クエリの各種類に対する内海らの手法と Wu らの手法の認識率

Wu らの手法の visual words と比較用の visual words, 一般的な visual words の認識率であり、図 7 はその 3 種類の visual words を用いた場合のそれぞれの認識率のグラフである。前述の通り一般的な方法で作成された visual words が明らかに低い認識率を示しているのに対して、Wu らの手法の visual words と比較用の visual words の認識率の間にはそれほど大きな差が見られない。このことから、visual words の顔認識における性能を向上させるためには特徴点の位置情報を活用することが特に重要であるとわかる。また、Wu らの手法の visual words と比較用の visual words の認識率の差はそれほど大きいものではなかったものの、ほとんどのパターンで Wu らの手法が上回っていた。このことから人物ごとにクラスタを作成するクラスタリング方法についても認識率を向上させる効果があることがわかる。人物ごとのクラスタを作成することによるデメリットとして、同じクラスタに属する特徴量が特徴空間上で近くに位置するとは限らないのでセントロイドを利用できないという点がある。しかし、このデメリットによって処理が遅くなるのはヒストグラム化の作業までであるため、ヒストグラムを利用する認識処理の段階では影響しない。よって人物ごとのクラスタを作成する方法も有効な手法であると言える。

6. まとめ

本稿では特定物体認識と一般物体認識の顔認識における認識率の比較実験を行った。また、一般物体認識の顔認識における精度を上げるためには visual words のどの要素が重要なのかという検証実験を行った。特定物体認識と一般物体認識の認識率は本実験で用いたクエリに関しては拮抗していた。visual words に関する検証実験では、Wu らの手法の顔認識における精度向上の最も大きな要因は位置情報を含んだ点であるということを示した。

今後の課題としては、Face in the wild のような認識が

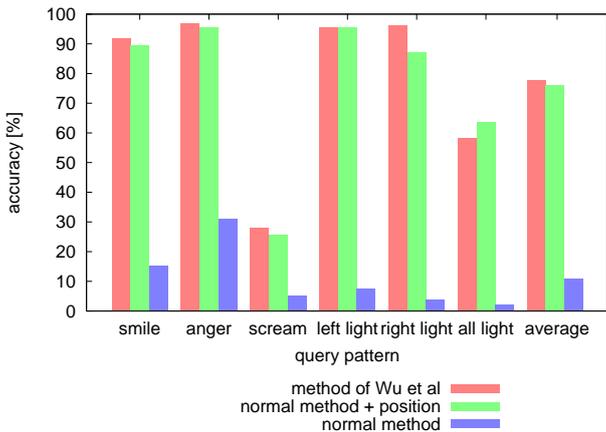


図 7 三種類の手法で作成された visual words によるそれぞれの認識率

より難しいデータセットをクエリとして用いた比較実験で各手法の優位性を示すことができるかを検証することが挙げられる。また、今回行わなかった特徴点の配置やクラスタリングの手法といった visual words の性質に関する検証も今後の課題として挙げられる。

謝辞 本研究は JSPS 科研費 25240028 の助成を受けた。

2015 年 4 月 16 日

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. CVIM, [コンピュータビジョンとイメージメディア], Vol. 2013, No. 4, pp. 1-7 (2013).
- [2] Wu, Z., Ke, Q., Sun, J. and Shum, H.-Y.: Scalable face image retrieval with identity-based quantization and multireference reranking, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 33, No. 10, pp. 1991-2001 (2011).
- [3] 黄瀬浩一: 特定物体認識 (チュートリアル, 一般物体認識・画像特徴量), 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 109, No. 306, pp. 79-87 (2009).
- [4] Lowe, D. G.: Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, Vol. 60, No. 2, pp. 91-110 (2004).
- [5] Matas, J., Chum, O., Urban, M. and Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions, *Image and vision computing*, Vol. 22, No. 10, pp. 761-767 (2004).
- [6] Nister, D. and Stewenius, H.: Scalable recognition with a vocabulary tree, *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Vol. 2, IEEE, pp. 2161-2168 (2006).
- [7] Philbin, J., Chum, O., Isard, M., Sivic, J. and Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching, *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, IEEE, pp. 1-8 (2007).
- [8] Wright, J. and Hua, G.: Implicit elastic matching

with random projections for pose-variant face recognition, *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, IEEE, pp. 1502-1509 (2009).

- [9] Ke, Y. and Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors, *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Vol. 2, IEEE, pp. II-506 (2004).
- [10] Martinez, A. M.: The AR face database, *CVC Technical Report*, Vol. 24 (1998).
- [11] Gao, W., Cao, B., Shan, S., Zhou, D., Zhang, X., Zhao, D. and Ai, S.: The CAS-PEAL large-scale Chinese face database and evaluation protocols, *Technique Report No. JDL-TR_04_FR_001, Joint Research & Development Laboratory, CAS* (2004).
- [12] Grgic, M., Delac, K. and Grgic, S.: SCface-surveillance cameras face database, *Multimedia tools and applications*, Vol. 51, No. 3, pp. 863-879 (2011).
- [13] Wolf, L., Hassner, T. and Taigman, Y.: Effective unconstrained face recognition by combining multiple descriptors and learned background statistics, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 33, No. 10, pp. 1978-1990 (2011).
- [14] Wolf, L., Hassner, T. and Taigman, Y.: Similarity scores based on background samples, *Computer Vision-ACCV 2009*, Springer, pp. 88-97 (2010).
- [15] Taigman, Y., Wolf, L., Hassner, T. et al.: Multiple One-Shots for Utilizing Class Label Information., *BMVC*, pp. 1-12 (2009).