

照明変化を伴う物体認識：顔画像認識への応用

佐藤洋一

情報理工学系研究科電子情報学専攻

概要

顔画像にもとづく個人認識はコンピュータビジョンにおける物体認識技術の一つの応用として捉えることができる。この物体認識には、物体姿勢、視点位置の変化と照明変化の問題があるが、前者が主に幾何学的特性に依存するのに対して、後者は主に光学的特性に依存するため、これらを分けて考えることができる。本研究では、照明変化の問題に着目し、物体姿勢および視点位置一定の条件下で照明のみが変化する場合の物体認識に対し2つの手法を提案した。また、提案手法を顔画像認識に適用した場合の有効性を確かめるために、異なる照明下において撮影された顔画像を含むデータベースを用い、提案手法の性能を代表的な従来法との比較の上で評価した。

1 パターン認識に基づくアプローチ

D 個の画素からなる画像は、全画素値を並べたベクトル、つまり、 D 次元画像空間内の点で表現することができる。画像空間そのものを特徴空間としてパターン認識アルゴリズムを適用する手法は、パターン認識に基づくアプローチと呼ばれる。

1.1 照明錐

物体姿勢および視点位置一定の条件下で照明のみが変化する場合の画像の変化は、重ね合わせの原理¹が成り立つことを仮定すると、画像空間において原点を通る凸錐 (convex cone) に拘束され

¹二つの光源下で撮影された画像が、個々の光源下で撮影された画像の和で表現される。

ることが知られており、この凸錐は照明錐 (illumination cone) と呼ばれている [3]。

反射特性が Lambert モデルで記述される物体の照明錐は、実画像を用いた実験から、低次元部分空間で近似できることが知られていた [6]。最近になって、球面調和関数を用いた周波数空間での理論的解析からも、4次元から9次元程度の低次元部分空間で照明錐を近似できることが示されている [1, 7]。

1.2 提案手法

従来のパターン認識に基づく手法 [2] では、照明変化に伴う画像の変化が照明錐に拘束されることは十分に考慮されていなかった。そこで本稿では、パターン認識に基づくアプローチの枠組みで照明錐を考慮した認識手法を提案する。

各物体に対応する照明錐が、例えば数千次元の画像空間内の低次元部分空間に独立に分布していると仮定すると、個々の照明錐は交わりを持たないことが期待される。

そこで各物体に対応する照明錐が交わりを持たないと仮定すると、照明錐が原点を頂点とする凸錐であることから、任意の二つの照明錐 IC_1, IC_2 は、画像空間の原点を通る超平面 (図1では直線 OP) により分離される。一方、個々の照明錐が交わりを持たないとしても、ある照明錐と他の全ての照明錐が線形分離可能である保証はない。したがって、照明変化を伴う物体認識問題は、 D 次元画像空間において原点を通る線形識別面を用いた2クラス判別問題の組み合わせに帰着する。

さらに、照明の明るさの変化を吸収するために画像を L_1 ノルムで正規化すると、任意の二つの照

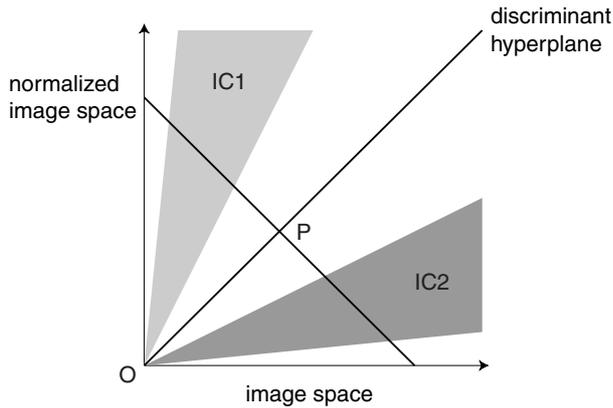


図 1: 提案手法の概念図

明錐は $(D - 1)$ 次元の正規化画像空間における線形識別面 (図 1 では点 P) により分離される。したがって、照明変化を伴う物体認識問題は、 $(D - 1)$ 次元正規化画像空間において線形識別面を用いた 2 クラス判別問題の組み合わせに帰着する。

1.3 評価実験

顔画像データベース Yale Face Database B を用いた評価実験を行った。実験には、10 名の顔を 64 方向の光源下で撮影した 640 枚の画像を用いた。各画像は、光源方向とカメラの光軸のなす角度 θ に基づいて、5 つの部分集合のいずれかに分類されている。図 2 に、各部分集合に属する画像の例を示す。



図 2: 照明変化に伴う画像の変化

以下に示す 4 つの手法を用いた認識実験を行った。

- 最近傍決定則 (NN)
- 固有顔 (EF)
- フィッシャーの線形判別法 (FLD)

表 1: 誤識別率 (%) : パターン認識に基づく手法

Method	Dimension	2	3	4	5
NN-1	4096	5.1	50.8	81.2	85.2
NN-2	4095	0	7.6	56.5	77.2
EF-1	4096 → 5	25.4	72.9	84.8	86.2
EF-2	4095 → 5	7.6	47.5	73.2	78.8
FLD-1	4096 → 5	4.2	37.3	71.7	85.7
FLD-2	4095 → 5	0	13.6	60.9	88.9
SVM-1	4096	2.5	22.9	75.4	87.3
SVM-2	4095	0	0	36.2	88.4

• サポートベクターマシン (SVM)

各手法について、4096 (=64 × 64) 個の全画素値を特徴とした場合と正規化画像の全画素値を特徴とした場合の 2 通り、合計 8 通りの実験を行った。

表 1 に、各人物の学習画像として部分集合 1 に属する 7 枚を用い、部分集合 2 から 5 までの画像をテスト画像とした、光源方向の外挿に関する実験結果を示した。ここで各手法に付けられた番号は、1 が画像の全画素値を特徴とし、2 が正規化画像の全画素値を特徴として用いたことを示している。したがって、FLD-2 と SVM-2 が、照明錐を考慮した認識手法である。

FLD-1 と FLD-2 および SVM-1 と SVM-2 を比較すると、正規化画像を特徴とすることで、認識性能が大幅に改善されていることを確認することができる。これは、FLD-2 と SVM-2 が照明錐を考慮しているのに対して、FLD-1 と SVM-1 では、照明錐を反映した画像空間の原点を通る線形識別面が得られるとは限らないためであると考えられる。また、SVM-2 は、FLD-2 や他の手法よりも優れていることから、照明方向の外挿に関して、照明錐を考慮した認識手法が有効であり、照明錐を分離するような線形識別面を求める手法として、SVM がより効果的であると言える。

なお、各人物の学習画像として部分集合 1 と部分集合 5 に属する 26 枚を用い、部分集合 2 から 4 までの画像をテスト画像とした光源方向の内挿に関する実験においても、照明錐を考慮した SVM が有効であることを確認した。

2 画像合成に基づくアプローチ

画像を構成する成分は、拡散反射成分、ハイライト、attached shadow (陰)、cast shadow (影) の 4 つに分類することができる。反射モデルに基づいて任意照明下の画像を合成して認識を行う手法は、画像合成に

基づくアプローチと呼ばれる。

2.1 光学的アラインメント

拡散反射成分を記述する Lambert モデルを仮定すると、点光源下の画像 x の i 番目の画素値 x_i は、

$$x_i = \rho_i n_i^T s \equiv b_i^T s, \quad (i = 1, 2, \dots, D) \quad (1)$$

で表される。ここで、 ρ_i, n_i は i 番目の画素に対応する物体表面の拡散反射率と単位法線ベクトル、 s は光源の方向を表す単位ベクトルにその強度を掛けたものである。したがって、拡散反射成分のみで構成されている画像は、 b_i^T を D 画素分縦に並べた $D \times 3$ の行列 B を用いて、 $x = Bs$ で表される。これは、任意方向の点光源下の拡散反射成分が 3 枚の画像の線形結合

$$x = \alpha_1 e^{(1)} + \alpha_2 e^{(2)} + \alpha_3 e^{(3)} \quad (2)$$

で表現されることを意味している。また、式 (2) において負の画素値を 0 とすることで、拡散反射成分だけでなく attached shadow も表現することができる。このようにして 3 枚の画像から任意方向の点光源下の拡散反射成分と attached shadow を合成する手法は、光学的アラインメント (photometric alignment) と呼ばれている [8]。

2.2 提案手法

光学的アラインメントを物体認識に適用するには、実画像に含まれるハイライトや shadow などのノイズを考慮する必要がある。光学的アラインメントに基づく物体認識には基底画像の推定 (学習) とテスト画像の再現 (認識) という二つの要素があるが、ノイズを含む学習画像から基底画像 $e^{(i)}$ ($i = 1, 2, 3$) を推定する手法に関しては、照度差ステレオの分野において既に提案されている [5, 9]。そこで本稿では認識過程、特に、結合係数 α_i ($i = 1, 2, 3$) の推定について RANSAC [4] を利用した手法を提案する。

まず、ランダムに 3 画素を選択して式 (2) から結合係数 $\hat{\alpha}_i$ ($i = 1, 2, 3$) を計算し、得られた結合係数と基底画像から

$$\hat{x}_i = \max \left(\sum_{j=1}^3 \hat{\alpha}_j e_i^{(j)}, 0 \right), \quad (i = 1, 2, \dots, D) \quad (3)$$

のようにして拡散反射成分と attached shadow からの画像を合成する。ここで $e_i^{(j)}$ は j 番目の基底画像の i 番目の画素値である。次に、各画素にラベル

$$\xi_i = \begin{cases} 1 & (|x_i - \hat{x}_i| < t : \text{inlier}) \\ 0 & (|x_i - \hat{x}_i| \geq t : \text{outlier}) \end{cases} \quad (4)$$

を付け、 $support = \sum_{i=1}^D \xi_i$ を求める。ここでテスト画像と基底画像が同一のクラスで、テスト画像が拡散

反射成分と attached shadow のみで構成され、かつ、正しい結合係数が得られている場合には、 $support$ は画素数 D に等しい。以上の処理を繰り返して $support$ を最大にする結合係数を求める。最後に、 $\xi_i = 1$ かつ $\hat{x}_i > t$ の画素に対して ξ_i が収束するまで最小自乗法を繰り返し適用し、結合係数および $support$ を得る。

認識においては、クラスごとに基底画像からテスト画像を再現して $support$ を求め、 $support$ が最大となるクラスに分類する。

2.3 評価実験

前節と同様に Yale Face Database B を用いた評価実験を行った。各人物について部分集合 1 に属する 7 枚の画像から、重み付き最小自乗法に基づく手法 [5] により基底画像を推定した (図 3)。

図 4 に処理過程の画像を示した。テスト画像 (a) に対して、同一人物の基底画像を用いて作成した合成画像 (b) と、他の人物の基底画像を用いて作成した合成画像 (c) を示した。各合成画像について、outlier とみなされた画素を白く表示したものが (d), (e) である。同一人物の基底画像から合成した画像が、鼻や目頭の cast shadow を除いてほぼ再現できているのに対して、他の人物の基底画像から合成した画像は、目、鼻、口の周辺を中心として比較的広い範囲で再現に失敗していることがわかる。

表 2 に実験結果をまとめた。ここで SS2, SS3 などはそれぞれ画像データベースの Subset 2, Subset 3 を意味する。提案手法 (PA: RANSAC) の他に、3 枚の基底画像の張る部分空間との距離に基づく照明部分空間法 (IS)、最小自乗法で得られた結合係数を用いて合成された画像とテスト画像との距離に基づく手法 (PA: LS) の結果も示した。表 2 から、拡散反射成分しか表現することのできない照明部分空間法と比較して、attached shadow も表現可能な光学的アラインメントが優れていることがわかる。また、同じ光学的アラインメントに基づく認識手法でも、RANSAC を用いた提案手法が最小二乗法を用いた手法よりも優れているのを確認することができる。なお、提案手法の部分集合 5 に対する誤識別の要因としては、本手法では生成できないハイライトや cast shadow の影響が考えられる。

表 2: 誤識別率 (%) : 画像合成に基づく手法

Method	SS2	SS3	SS4	SS5
IS	0	0	5.1	36.0
PA: LS	0	0	2.9	29.1
PA: RANSAC	0	0	0.7	20.6



図 3: 基底画像

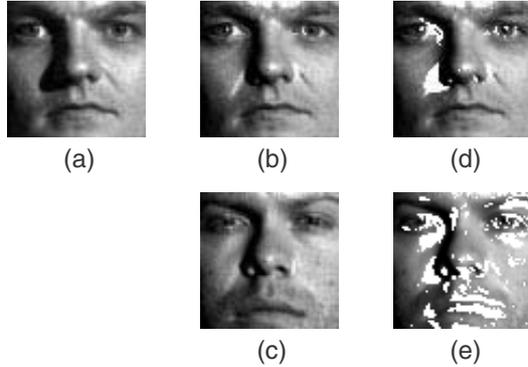


図 4: テスト画像と合成画像

3 まとめ

本研究では、照明変化を伴う顔認識の問題について、二つの異なるアプローチにもとづく手法を提案した。見え方に基づく手法の枠組みでは、照明錐の性質を考慮したうえでサポートベクターマシンを適用し、多様な照明条件下の学習画像が与えられた場合に有効となる簡便な手法を提案した。本研究では顔画像を用いた評価実験を行ったが、提案手法は、Lambert モデルを仮定していないために、生成的手法では扱えないような物体にも適用できる。様々な反射特性を持つ物体への提案手法の適用とその有効性の検証は今後の課題としたい。

一方、光学的アラインメントに基づく生成的手法では、RANSAC を利用し、テスト画像に含まれる鏡面反射成分や cast shadow などの外れ値に対してロバストな顔認識を実現した。今後の課題として、まず、単一点光源を仮定した提案手法の複雑光源への拡張が挙げられる。また、照明変化だけでなく姿勢変化も伴う物体認識への拡張、または、姿勢推定との融合についても検討したい。本研究では鏡面反射成分と cast shadow を外れ値とみなしたが、逆に、これらの成分を効率よくモデル化して積極的に利用することも考えたい。

参考文献

- [1] R. Basri and D. Jacobs, “Lambertian reflectance and linear subspaces”, In *Proc. IEEE ICCV 2001*, pp.383–390, 2001.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection”, *IEEE Trans. PAMI*, 19(7), pp.711–720, 1997.
- [3] P. N. Belhumeur and D. J. Kriegman, “What is the set of images of an object under all possible lighting conditions?”, *Int. J. Computer Vision*, 28(3), pp.245–260, 1998.
- [4] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”, *Communications of the ACM*, 24(6), pp.381–395, 1981.
- [5] A. S. Georgiades, D. J. Kriegman, and P. N. Belhumeur, “Illumination cones for recognition under variable lighting: faces”, In *Proc. IEEE CVPR '98*, pp.52–58, 1998.
- [6] P. W. Hallinan, “A low-dimensional representation of human faces for arbitrary lighting conditions”, In *Proc. IEEE CVPR '94*, pp.995–999, 1994.
- [7] R. Ramamoorthi, and P. Hanrahan, “On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object”, *J. Opt. Soc. Am. A*, 18(10), pp.2448–2459, 2001.
- [8] A. Shashua, “Geometry and photometry in 3D visual recognition”, PhD thesis, MIT, 1992.
- [9] 石井育規, 福井孝太郎, 向川康博, 尺長健, “光学現象の分類基準を利用した画像の線形化”, 画像の認識・理解シンポジウム (MIRU2002), II, pp.167–176, 2002.