

任意照明下顔認識のための顔表面の位置相関を考慮したMAP推定

島野 美保子[†] 長尾 健司[†] 岡部 孝弘^{††}
佐藤 いまり^{†††} 佐藤 洋一^{††}

本研究では、各人物について1枚の登録画像のみが与えられている場合の任意照明下のテスト画像に対する顔認識手法を提案する。提案手法は、複数人物を様々な照明条件下で撮影した大量の画像集合から得られる統計量とランバートモデルとを組み合わせた統計的 shape from shading に基づいている。従来手法と比較した際の利点は、MAP 推定時に顔表面位置間における法線方向やアルベドの相関を考慮することにより、部分的に影になった領域においても、新しい照明条件下の顔画像を正しく合成することを可能にした点である。さらに、画素のグルーピングや信頼性指標を導入し、精度を維持しつつ計算コストの削減も実現した。本論文では、提案手法のアルゴリズムを説明し、実顔画像を用いた実験結果により提案手法の有効性を示す。

Face Recognition under Varying Illumination Based on MAP Estimation Incorporating Correlation between Surface Points

MIHOKO SHIMANO,[†] KENJI NAGAO,[†] TAKAHIRO OKABE,^{††}
IMARI SATO^{†††} and YOICHI SATO^{††}

In this paper, we propose a new method for face recognition under varying illumination conditions, using a single inputted image. Our method is based on a statistical shape-from-shading method, which combines the strengths of the Lambertian model and statistical information obtained from a large number of images of different people under varying illumination. The main advantage of our method over previous methods is that our method explicitly incorporates a correlation between the surface points on a face in the MAP estimation of the surface normals and albedos, so that a new image of the same face under novel illumination can be correctly synthesized even when the face is partially shadowed. In addition, our method introduces pixel grouping and a reliability measure in the MAP estimation to reduce computational costs while maintaining accuracy. We demonstrate the effectiveness of our proposed method via experiments with real images.

1. はじめに

顔認識は、コンピュータビジョンの分野においてさかんに研究が行われており、画像から人の顔を認識する手法が数多く提案されている^{(6),(20)}。このような顔認識技術の発展はセキュリティ分野やインタフェース分野の様々な応用の実現を可能とする。たとえば、証明書用写真やパスポート写真のように各人物について画像1枚のみが登録画像として利用可能な場合の個人同定もその1つである。本研究では、特にこのような場

合に着目し、1人につき1枚のみ登録画像がある場合の任意照明下での顔認識問題を扱う。一般に、顔の見え方は姿勢や照明条件、それに表情変化等の要因に影響されるが、ここでは特に照明条件に着目し、照明が変化した場合の人物の正面顔の見え方の変動のモデル化に取り組むこととする。

各人物ごとに異なる照明条件下で撮影された登録画像が十分な枚数と与えられている場合には、各人物の顔の見え方が照明の変化にともなってどのように変動するかをあらかじめ学習することが可能となり、顔認識の問題は比較的容易になると考えられる。たとえば、照明変化にともなう顔画像の変動は、ある程度低次元の部分空間によって表現されることが知られており^{(2),(4),(7)}、異なる照明下で撮影された十分な量の登録画像が得られるならば、既存の顔認識手法により高い認識精度が実現されることが報告されている^{(3),(7),(10)~(14),(16)}。

[†] 松下電器産業株式会社
Matsushita Electric Industrial Co., Ltd.

^{††} 東京大学生産技術研究所
Institute of Industrial Science, The University of Tokyo

^{†††} 国立情報学研究所
National Institute of Informatics

しかしながら、前述の応用例のように、現実的には各人につき大量の登録画像を用意できない場合も多い。このような場合には、照明変動にともなう顔の見えの変動をモデル化できないため、任意照明下における顔認識は非常に困難な問題となる。これは、1枚の登録画像以外の情報がまったくないときには、形状、反射係数、および、照明条件の同時推定が制約不足 (under-constrained) となるために、顔のモデルを獲得することが原理的に不可能であることから分かる。

これに対し、統計的な制約を利用する手法が提案されている。具体的には、登録画像とは別に、異なる照明下で撮影した複数人物の顔画像集合から得られる統計モデルや、複数人物の頭部形状をレーザスキャナ等で直接計測したデータから得られる統計モデルを利用することにより、新たな人物を撮影した1枚の顔画像から形状や反射係数を求める手法がいくつか提案されている^{1),5),15),17),21)}。不均一な反射率を持つ物体であっても1枚の入力画像から形状と反射率を推定する shape from shading も提案されているが、これらの shape from shading では可積分条件⁸⁾や顔の対称性¹⁹⁾のような明示的な仮定が用いられるのに対し、上述の手法^{1),5)}ではランバートモデル等の反射モデルに加え、あらかじめ準備した統計モデルを用いて1枚の画像から形状と反射特性を推定する点が大きく異なっている。

これらの手法では、異なる照明条件下における顔の見えの変動を予測したが、Simらの手法¹⁷⁾を除いて共通の問題点をかかえていた。すなわち、ランバートモデルや Phong モデルのような単純な反射モデルから外れる反射成分を、これらの手法で正しく再合成することはできなかった。たとえば、Atickらや Zhouらの手法は、顔はランバートモデルで表現されるという仮定に基づいていた^{1),21)}。山元らの手法¹⁵⁾は、照明変化にともなう見えの変化を固有空間内の線形変換で表したため、影やハイライト等の非線形な成分の表現は困難であると考えられる。Blanzらの手法⁵⁾では、ランバートモデルに加え Phong モデルも考慮して顔表面上の陰影を表現しているが、顔表面における反射は単純なランバートモデルや Phong モデルから大きく逸脱することが知られている⁹⁾。

本論文で提案する手法は、1枚の登録画像を入力とし、顔表面の法線方向や拡散反射係数を異なる人物の画像集合から得られた統計量を基に最大事後確率 (MAP) 推定により求めるという考え方に基づいており、Simらの手法¹⁷⁾と関連が深い。両手法とも、ランバートモデルで表現される拡散反射成分以外の反射

成分も誤差成分の MAP 推定により求めることができるため、ハイライトや相互反射等の複雑な反射成分を含めて新しい照明条件下で画像を生成することができるという、その他の統計的な shape from shading 手法にはない特長となっている。

我々の手法と Sim らの手法の最も重要な相違点は、MAP 推定時に法線方向や拡散反射係数の顔表面位置間における相関を明示的に考慮していることである。たとえば、我々の手法では、部分的に影になった領域においても新しい照明条件下の顔画像を正しく合成できるのに対し、Sim らの手法では各画素を独立に取り扱っているために正しく推定することができない。また、合成精度を維持しつつ計算コストを削減するための方策として、画素のグルーピング、法線推定の信頼性指標の導入という工夫を新たに提案する。さらに、画像間の照明の明るさ変動も考慮する。

2. 提案手法

提案手法は、i) 学習処理と ii) モデリング処理と iii) レンダリング処理の、3つの処理より構成される。

学習処理では、顔の統計モデル、すなわち、登録者と異なる人物の照明変動下の顔画像集合を用いて、アルベドを含む法線方向の統計モデルと、拡散成分以外の反射成分に対応する誤差項の統計モデルを生成する。以降、この登録者と異なる人物の顔画像集合をブートストラップ画像集合と呼ぶ。

モデリング処理では、学習処理で生成した統計モデルを用いて、免許証写真やパスポート写真のような1枚のみの登録画像から見えの変動をモデル化する。ここでまず、1枚の登録画像が撮影されたときの照明条件を推定する。次に、推定した照明条件と学習した統計量を基に MAP 推定によってその顔の形状と反射特性を復元する。

最後に、レンダリング処理では、新しい照明条件下の誤差項も MAP 推定によって算出し、同じく新しい照明条件下の拡散成分に加えて、新しい見えを生成する。このように合成した画像集合を、あたかも照明条件の異なる複数の登録画像が得られたように、クラス分類手法の学習サンプルとすることによって認識を可能とする。以下にその詳細を記す。

2.1 反射モデルの式

提案手法は、学習処理におけるブートストラップ画像集合と、モデリングを行うための1枚の登録画像の撮影光源として、ともに単一无限遠点光源を仮定する。このとき、 p 番目画素の輝度 i_p は、拡散成分とその他の成分の和によって

$i_p = n_p^T s + e_p(s)$, (1)
と表される¹⁷⁾. ここで, $n_p = (n_{px}, n_{py}, n_{pz})^T$ は,
 p 番目の画素のアルベドと単位法線ベクトルの積で,
 $s = (s_x, s_y, s_z)^T$ は, 照明強度と照明方向単位ベクトル
の積である. 誤差項 $e_p(s)$ は, ハイライトや相互反射
や影のような拡散反射成分以外の反射成分を表し,
画素 p と光源 s に依存すると考える.

式 (1) では各画素の輝度を個別に取り扱う表現で
あったのに対し, 明示的に顔表面位置間の相関を考慮
するために, 各画像に含まれる P 画素の輝度をまと
めて次式のように表現する.

$$i = S^T b + e(s), \quad (2)$$

$$\begin{pmatrix} i_1 \\ i_2 \\ \vdots \\ i_P \end{pmatrix} = \begin{pmatrix} s^T & 0 & \dots & 0 \\ 0 & s^T & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & s^T \end{pmatrix} \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_P \end{pmatrix} + \begin{pmatrix} e_1(s) \\ e_2(s) \\ \vdots \\ e_P(s) \end{pmatrix}.$$

ここで, 画像ベクトル i , s^T を対角上に配置した照
明行列 S^T , 法線アルベドベクトル b , 誤差ベクトル
 e は, 各々, P 次元ベクトル, $P \times 3P$ 行列, $3P$ 次
元ベクトル, および P 次元ベクトルである.

2.2 統計量の算出

次に, ブートストラップ画像集合を用いて, 顔表面
上の異なる位置において観察される明るさから, 顔表
面位置間の形状や反射特性の相関を考慮した統計量を
算出する. 具体的には, 登録画像とは別にあらかじめ
準備したブートストラップ画像集合の各人物に関して
法線やアルベド, 誤差を算出し, それらの算出した複
数人分の法線やアルベド, 誤差からその統計量を計算
する.

ここで用いるブートストラップ画像集合は, L 人
の人物について J 個の既知の照明条件下 s_j ($j =$
 $1, 2, \dots, J$) で撮影された画像集合により構成される.
 l 番目の人物について, J 個の光源下で撮影された画
像集合 $I^{(l)}$ は下記のように

$$I^{(l)} = B^{(l)T} S' + E^{(l)}, \quad (3)$$

と表される. ここで各々, 人物 l の J 個
の照明条件について得られた画像集合 $I^{(l)} =$
 $(i_1^{(l)}, i_2^{(l)}, \dots, i_J^{(l)})$, 人物 l の顔の法線アルベド行列
 $B^{(l)} = (n_1^{(l)}, n_2^{(l)}, \dots, n_P^{(l)})$, J 個の照明方向とその強
度を表した照明集合 $S' = (s_1, s_2, \dots, s_J)$, J 個の照
明条件に対する誤差行列 $E^{(l)} = (e^{(l)}(s_1), e^{(l)}(s_2), \dots,$
 $e^{(l)}(s_J))$ である. また, ブートストラップ画像撮影時
の光源の明るさ ($|s_j|$) は一定であると仮定する (以
降, $|s_j| = 1$ として説明する).

次に, 誤差行列 $E^{(l)}$ をガウスノイズと考え, 最小
二乗法によって $B^{(l)}$ を以下のように

$$B^{(l)} = (S' S'^T)^{-1} S' I^{(l)}, \quad (4)$$

と算出する. そして, 次のように拡散成分を引いた残
りが誤差行列 $E^{(l)}$ である.

$$E^{(l)} = I^{(l)} - B^{(l)T} S'. \quad (5)$$

最後に, 顔の法線と誤差項の統計量を, ブートスト
ラップ画像集合のすべての人物について推定された顔
の法線アルベド行列 $B^{(l)}$ と誤差行列 $E^{(l)}$ から次の
ように求める. 法線アルベドベクトル b のように, L
人について各行列 $B^{(l)}$ を $3P$ 次元ベクトルに展開し,
その平均ベクトル μ_b ($3P$ 次元ベクトル) と共分散行
列 C_b ($3P \times 3P$ 行列) を求める. 誤差項 $e(s)$ の統
計量についても, 平均ベクトル $\mu_e(s_j)$ は, 各照明方
向 s_j について L 個の誤差ベクトル $e^{(l)}(s_j)$ から計
算できる. また, $PJ \times PJ$ の共分散行列 C_e は, C_b
を $B^{(l)}$ から求める方法と同様に L 個の誤差行列 $E^{(l)}$
から算出される. このように, 提案手法においては,
従来手法の各画素を独立に扱っていた点を, 顔表面位
置間の相関を考慮することができるアルゴリズムに拡
張している.

外れ値除去

B の統計量は, 式 (4) で推定された法線から導出さ
れる. しかしながら, ブートストラップ画像集合には
ハイライトや影の成分も含んでいる. したがって, こ
れらの成分を主とする画素による影響を排除できるよ
うに, ハイライトや影を抽出するための閾値よりも輝
度が高い, あるいは低い画素を外れ値と見なす. これ
により, 外れ値を除去したブートストラップ画像集
合から, 顔の法線を推定することができる. この後, 誤
差項の統計量をブートストラップ画像集合から算出
する際には, 再びすべての画素を用いる点に留意する.

2.3 登録画像 1 枚からのモデリング

顔表面位置間の相関を考慮しつつ, 登録画像 1 枚か
らモデリングを行う方法について, 詳細に説明する.
まず登録画像の照明方向を推定し, 次に前章で求めた
統計量を基に法線やアルベドを復元するという 2 つの
処理により構成される.

照明条件の推定

1 枚の登録画像 i が撮影された照明方向と照明強度
を推定する. ここでは, ランバートモデルおよび, 対
象顔として平均の顔形状と反射特性を仮定し, 最小二

$B^{(l)}$ を正しく復元するために, ブートストラップ画像集合に含
まれているハイライトや影といった外れ値を除去した (2.2 節外
れ値除去参照).

乗法により照明条件を推定する．Sim らの手法では登録画像とブートストラップ画像の照明の明るさは一致していると仮定し、影やハイライトの強度については考慮していないが¹⁷⁾、本手法では画像間の照明の明るさ変動をも考慮する．1 枚の登録画像の照明条件を、ブートストラップ画像 $i_j^{(l)}$ の統計量算出で求めた法線アルベド行列 $B^{(l)}$ の平均 B_{avr} を用いて、

$$s = B_{avr}^{T+} i = (B_{avr} B_{avr}^T)^{-1} B_{avr} i, \quad (6)$$

と推定する．推定された照明条件 s の大きさ $\alpha = |s|$ は、ブートストラップ画像集合の基準となる照明強度に対する強度比を表す．この照明強度に比例した拡散反射成分や誤差項を想定することにより、登録画像と照明強度の異なるブートストラップ画像集合の統計モデルを用いても、法線および新しい照明下の誤差項の推定を行うことができる．

顔の法線の推定

顔の法線やアルベドを、2.2 節で算出した法線アルベドや誤差項の統計量と、求めた登録画像の照明条件から、ブートストラップ画像と登録画像の照明の明るさ変動にも対応した MAP 推定によって求める．これらの統計量は、式 (2) のように複数画素の反射モデル式で表現し、2.2 節で記載したとおり顔表面位置間の形状や反射特性の相関を考慮して算出される．したがって、これらの統計量を用いて算出される登録画像の顔の法線アルベドの MAP 推定値は、与えられた登録画像から顔表面位置間の相関を考慮した結果として得られ、 $b_{MAP} = \arg \max_b P(b|i)$ と復元できる．これはベイズの定理に従って式 (7) のように、

$$b_{MAP} = \arg \max_b P(i|b)P(b), \quad (7)$$

と表される． b の確率密度関数はガウス分布であると仮定しているため、 $P(b)$ は、 μ_b と C_b によって記述される．また、 e の確率密度関数もガウス分布と仮定することにより、 $P(i|b)$ は、反射モデルの式 (2) に従って、平均 $S^T b + \mu_e(s)$ と分散 Σ_e のガウス分布によって表される．ここで求める値は、登録画像の照明方向 s に対する統計量である誤差の平均 μ_e と分散 Σ_e である．これらの s に対する統計量は、既知の照明条件 s_j におけるブートストラップ画像集合の統計量から kernel regression によって補間される．具体的には、平均 $\mu_e(s)$ は、

$$\mu_e(s) = \alpha\beta \frac{\sum_{j=1}^J w_j \mu_e(s_j)}{\sum_{j=1}^J w_j}, \quad (8)$$

のように、既知の照明ベクトル s_j を中心とするカーネル関数の線形和で求められる．このとき、本論文ではガウス基底関数 $w_j = \exp(-(D(s, s_j)/\sigma_j)^2/2)$ 、 $D(s, s_j) = |s/\alpha - s_j|$ を用いる． α は、照明条件の推定で説明したように、登録画像とブートストラップ画像の照明強度の違いを考慮するための係数である．また、ここで β を導入し、誤差項を線形和でモデル化することによる照明強度の変化を補正する．つまり、 β はベクトルの大きさを 1 にする照明ベクトルの線形和の大きさ調整係数であり、 $s/\alpha = \beta(\sum_{j=1}^J w_j s_j) / \sum_{j=1}^J w_j$ を満たす． Σ_e の要素もまた、平均 μ_e と同様に、事前に算出された統計量 C_e から kernel regression によって補間される．

$$\sigma_e(s)^2 = \alpha^2 \beta^2 \frac{\sum_{j=1}^J w_j \sigma_e(s_j)^2}{\sum_{j=1}^J w_j}. \quad (9)$$

上記ベクトルや行列を式 (7) に代入すると、 b_{MAP} は以下のように

$$b_{MAP} = (S \Sigma_e^{-1} S^T + C_b^{-1})^{-1} (S \Sigma_e^{-1} (i - \mu_e) + C_b^{-1} \mu_b), \quad (10)$$

と記述される．導出の詳細は付録で記述する．

2.4 新しい照明条件下のレンダリング

新しい照明条件下の画像を合成するために、顔表面位置間と照明方向間の両方の相関を考慮しながら、新しい照明条件下の誤差ベクトルを推定する．ここでも、登録画像と新しく生成する画像の照明の明るさ変動を考慮する．初めに、登録画像の実際の誤差ベクトルを、前節で求めた法線アルベドベクトル b_{MAP} を用いて $e = i - S^T b_{MAP}$ と計算する．次に、この実際の誤差ベクトル e と、新しい照明条件下 S_{new} の誤差成分の確率密度関数として jointly Gaussian distribution を仮定することにより、新しい照明条件下の誤差ベクトルを MAP 推定で

$$e_{MAP} = \mu_{e_{new}} + R^T \Sigma_e^{-1} (e - \mu_e), \quad (11)$$

と求める．ここで、 $\mu_{e_{new}}$ と R は、各々新しい照明条件下の誤差の平均と、登録画像の照明条件と新しい照明条件間における誤差項の共分散である．導出の詳細については付録で記述する．この際、 $\mu_{e_{new}}$ と R の要素の値も、 μ_e と Σ_e から kernel regression によって補間される．前節と同様な方法で算出するが、今回はこれらの値を s 、 s_{new} の各々に対して、 s と既知照明条件 s_j 、 s_{new} と s_j' の類似度を足し合わせたカーネル関数より求める．

これらを用いて、新しい照明条件下の新しい顔画像は、推定された顔の法線アルベドベクトルと誤差ベクトルを用いて次のように

登録画像の照明条件を正しく推定するために、登録画像に含まれているハイライトや影といった外れ値を除去した(2.2 節参照)。

$$i_{\text{new}} = S_{\text{new}}^T b_{\text{MAP}} + e_{\text{MAP}}, \quad (12)$$

と合成される。

3. 統計量算出とモデリングの改良

顔表面位置間の相関の導入に加えて、統計量算出とモデリングの精度向上のために、提案手法では次の2項目の改良を行っている。

3.1 画素のグルーピング

ここまでは、 P を画像の全画素数として説明してきた。しかし、相関演算時間は、画素数 P に関して $O(P^3)$ 必要となる。そこで、計算コストを削減し、本手法をより簡便にするため、画像を部分領域に分割し、各領域内での相関のみを考慮した。なお、提案手法の記述において、 P を部分領域の画素数と見なすことが可能である。

また、画素をグルーピングした部分領域として、顔の法線を推定する際と誤差項を推定する際に同じ部分領域である必要はない。我々の実験では、法線とアルベドの推定においては、上記のような顔表面位置間の相関を導入して算出した。一方、誤差項については、 $PJ \times PJ$ の共分散行列を用いる計算は非常に計算コストがかかるため、画素間は独立に扱い、画素内の照明条件間の相関のみを考慮した。

3.2 信頼性指標の導入

提案手法の目的は、影領域やノイズを含んだ画素に対応する法線を、顔表面位置間の相関を考慮することによって正しく復元することである。しかしながら、この相関を考慮する複数画素の中に法線推定精度の低い画素を含む場合は、これら複数画素の正しい法線復元を阻害すると考えられる。したがって、この法線の推定精度を各画素の信頼性と見え、信頼性の低い画素の影響を小さくした複数画素間の相関を考慮して法線推定を行う。

この信頼性の評価として、統計量を用いて予測した各画素値がどの程度信頼できるかを表した信頼性指標を導入する。具体的には、誤差の分散 Σ_e の値を信頼性指標として用い、ある画素の誤差項の分散が大きい場合には、その画素は C_b で表される顔表面位置間の相関への寄与を小さくする。この際、誤差項の分散の大きさに注目し、誤差項の絶対値の大きさに影響されない点に留意する。

我々の実験では、予備実験で決定した閾値よりも大きい分散を持つ画素の相関への寄与を小さくし、対象画素の法線とアルベドを推定した。この信頼性指標に基づき、顔表面位置間の相関を考慮する画素数を削減することによって、本手法の計算コストも削減するこ

とができる。

4. 実験

提案手法の性能を評価するため、最初の3つの実験では相関を考慮することの有効性の検証を、4番目の実験ではランパートモデルの拡散成分以外の反射成分を考慮することの有効性の検証を目的とする4種類の顔認識実験を行った。すべての実験において、Yale face database B³⁾を用いて算出した共通の統計モデルを用いた。Yale face database Bは、顔の姿勢ごとに様々な照明条件下の画像(各人64枚)が10人分、計640枚用意されている。そのうち、正面向きの各人40枚(極端な照明条件下で過大な影領域を含む24枚を除く)の10人分の400枚の画像を用いた。各画像は、手作業で眼位置を揃えて 40×30 の顔領域を切り出した。

図1に示すように、我々の顔認識実験では、画素のグルーピング領域として、顔部位ごとに分割した少画素数よりなる部分領域と、矩形に分割した多画素数よりなる部分領域の、手作業で定めた2種類で試した。図1(a)において、左頬と右頬の領域は隣接していないが、1つの領域であることに注意する。

図2に、提案手法により合成された画像例を示す。右から照らされた顔を登録画像とし(図2(a))、正面から照らされた画像を、図1(a)に示す部分領域を用いた提案手法によって合成した(図2(b))。参考に、正面から照らされた同一人物の顔の実画像を図2(d)に示す。鼻で光源が遮られることにより生じたキャストシャドウや頬のアタッチドシャドウのような領域においても、拡散成分とハイライトともに正しく合成されていることが分かる。一方、Simらの手法により相関を考慮せずに合成された画像は、図2(c)に見られるように、影領域の推定に限界があることがはっきり

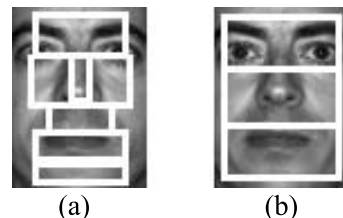


図1 本実験の画素のグルーピングに用いた部分領域：(a) 6部分領域と(b) 3部分領域。(a)において左頬と右頬の領域は隣接していないが、1つの領域であることに注意する

Fig. 1 Segmentations used for grouping pixels in our experiments: (a) 6 regions and (b) 3 regions. Note that left and right cheeks compose one region although they are not adjacent in (a).

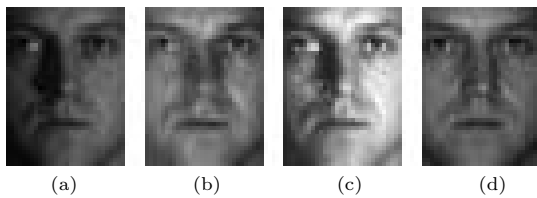


図 2 提案手法による合成画像例: (a) 右から照明が当たっている顔の登録画像, (b) 提案手法により顔表面位置間の相関を考慮した場合の正面方向の光源下での合成画像, (c) 相関演算をしない場合の正面方向の光源下での合成画像, (d) 正面方向の光源下での実顔画像

Fig.2 Example of synthesized images with our method: (a) input image of a face illuminated from right, (b) synthesized image taken under frontal illumination with our method incorporating correlation between surface pixels, (c) synthesized image taken under frontal illumination without correlation, (d) real image taken under frontal illumination.

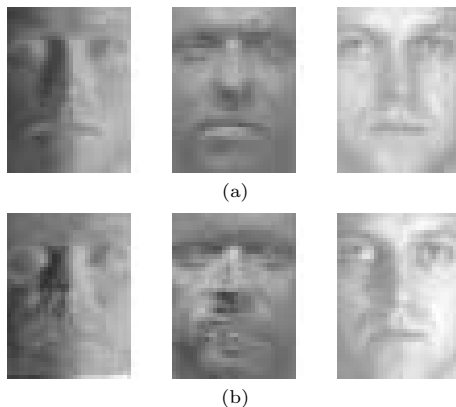


図 3 推定した法線アルベドベクトル例: (a) 提案手法, (b) Sim らの手法; ともに左から各々 x , y , z 方向成分, 登録画像は図 2 (a)

Fig.3 Example of normals including albedo: (a) Our method, (b) Sim and Kanade's method without correlation; Left column is x , center is y , and right is z value. Input image is Fig.2 (a).

見受けられる。

図 3 に, 提案手法および Sim らの手法により推定された法線アルベドを載せる。図 3 の (a), (b) とともに, 左から x , y , z 方向成分を表し, 各々ゼロを 128 にオフセットした値である。提案手法の図 3 (a) に比較して, Sim らの手法により相関を考慮せずに推定された法線アルベドの図 3 (b) は, なめらかさに欠け, かつ影領域の推定にも限界があるといえる。

異なる照明条件下で撮影された 2 種類の画像データベースを用いて, 顔認識実験を行った。一方は, 我々が収集したもので, 各人ごとに 11 枚の異なる照明条件下で撮影された 12 人分の正面向き顔画像集合であ

表 1 Sim らの手法と提案手法の顔認識性能比較 (我々の収集した 12 人の画像データベース利用)。全画素の相関を利用

Table 1 Performance comparison of Sim and Kanade's method and our proposed method incorporating correlation between all pixels by using our face image database of 12 individuals.

Methods	Recognition rate [%]
Sim and Kanade's method (without correlations)	88
Our method (with correlation)	94

る。他方は, CMU-PIE database¹⁸⁾ で, 各人ごとに 21 枚の異なる照明条件下で撮影された 68 人分の正面向き顔画像集合である。

各実験は以下のように行われた。まず, 各人ごとに 1 枚のみが登録画像として用いられ, 異なる照明条件下の 40 枚の画像が, 登録画像と Yale face database B から算出された統計モデルを用いて合成される。これら 41 枚の画像を, 各人ごとの部分空間を生成するのに用いる。最初の実験 (表 1) では, 5 枚のみ画像を合成した。データベースの残りの画像はテスト画像として用い, テスト画像とユークリッド距離が最も近い部分空間を探索することによってクラス分類をするという識別実験を行った。

最初の 3 つの実験では, 相関を考慮することの有効性の検証として, 本手法と関連の深い Sim らの手法との性能比較を行った。

1 番目の実験では, 我々の収集した画像データベースを用いて, 正面から照明を照らして撮影された画像を各人ごとに 1 枚の登録画像とした場合の性能比較を行った。ここで, 全画素の相関を考慮した提案手法によって実験を行った。表 1 は, これら 2 つの手法によって得られた顔認識実験の識別率を示す。これより, 提案手法の顔表面の位置相関を考慮した MAP 推定によって, 認識精度が非常に向上したことが分かる。

2 番目の実験では, CMU-PIE database を利用し, 各人ごとに 1 枚の画像を登録画像として用い, 図 1 (a) に示す部分領域を用いた提案手法によって合成した。登録画像として用いる画像の照明条件を変えながら識別実験を行った結果を表 2 に示す。図 4 に, 用いた登録画像の例を載せる。最初の実験と同様に, 顔表面の位置相関を考慮した MAP 推定によって, 左から照明が当たっている 'f22' を登録画像とした場合, 識別率は 72% から 90% までと大きく向上した。一方, ほぼ正面から照明が当たっている 'f08, f11, f20' を登録画像とした場合は, 向上の効果はそれほど見られなかった。これは, 横から照明が当たっている登録画像

表 2 Sim らの手法と提案手法の顔認識性能比較 (CMU-PIE database 利用). 画素のグルーピングとして図 1 (a) に示す部分領域を利用

Table 2 Performance comparison of Sim and Kanade's method and our proposed method using grouping regions of Fig. 1 (a) by using CMU-PIE database.

Methods	Recognition rate [%]												
	f08	f09	f11	f12	f13	f14	f15	f16	f17	f20	f21	f22	Average
Sim and Kanade's method (without correlation)	93	85	95	88	76	77	69	62	54	95	87	72	80
Our method (with correlation)	95	94	95	95	91	90	87	73	56	95	93	90	88



図 4 登録画像に用いた照明条件の異なる 12 枚の画像例：上段左より f08, f09, f11, f12, f13, f14, 下段左より f15, f16, f17, f20, f21, f22

Fig. 4 Example of training images under different lightings. First row: f08, f09, f11, f12, f13, f14. Second row: f15, f16, f17, f20, f21, f22.

の場合の方が、その影やハイライト等の部分が多く含まれているため、図 2 の合成結果で考察したように相関の考慮によって影やハイライト等の領域も正しく合成できるという本手法の効果がおいに現れたためと考えられる。左から照明が当たっている 'f17' を登録画像とした場合に、その他の登録画像と異なり識別率がそれほど大きく向上しなかった。この理由は、図 4 に見られるように、'f17' は顔領域のほぼ右半分全体を影領域が占めており、相関を考慮しても正しく合成することが難しかったためと考えられる。

3 番目の実験は、我々の収集した画像データベースを用いて、今回は登録画像として斜め上横から照明が当たっている顔画像を用い、画素のグルーピングとして図 1 (b) に示す部分領域を用いたこと以外は最初の実験と同様に行った。この実験では、3 章で説明したように、画素のグルーピングを行う際の信頼性指標を導入した場合と導入しない場合の比較実験も行った。この実験の結果を表 3 に示す。顔表面の位置相関を考慮した MAP 推定によって、識別率は 74% から約 10% も向上し 83% まで達したことが分かる。さらに、信頼性指標による効果も確認できる。

4 番目の実験では、ランバートモデルから外れる反射成分を考慮し、生成することの重要性を検証するため、我々の手法と Zhou らの最近提案した手法²¹⁾ とを比較した。1 枚の登録画像から任意照明条件下の顔

表 3 提案手法の画素のグルーピング (3 領域) と信頼性指標導入による認識性能の向上

Table 3 Performance improvement by grouping pixels (3 areas) and the use of reliability measure in our method.

Methods	Recognition rate [%]
Sim and Kanade's method (without correlation)	74
Our method (Grouping pixels without reliability)	81
Our method (Grouping pixels with reliability)	83

認識という共通の問題設定ではあるが、我々の手法と異なり、Zhou らの手法ではランバートモデルのみによって顔を表現する。両手法の性能比較を行うため、CMU-PIE database を用い、画素のグルーピングとして図 1 (a) に示す部分領域を用いて、彼らの報告²¹⁾ と同じ条件で実験を行った。したがって、上記 3 つの実験で行ったようにすべての照明条件の画像をテスト画像として用いるのではなく、正面および左側から照明が当たっている登録画像とは逆方向の右側から照明が当たっている画像は、テスト画像から排除されていることに留意する必要がある。Zhou らの手法の結果は論文 21) に記載の値を引用した。表 4 から分かるように、我々の手法は Zhou らの手法をすべての登録画像とテスト画像の組合せにおいて大きく上回る性能が得られた。たとえば、'f13' を登録画像、'f16' をテスト画像とした場合、我々の手法の識別率 (100%) は、Zhou らの手法の結果 (59%) よりも高い。これらの登録画像とテスト画像はともに、左側から照明が当たっている画像である。また、正面から照明が当たっている 'f08' を登録画像、左から照明が当たっている 'f15' をテスト画像とした場合、我々の手法 (99%) は、Zhou らの手法 (33%) に比較して高い識別率が得られた。また、全実験の平均を比較しても、我々の手法 (97%) は、Zhou らの手法 (67%) よりも高い識別率が得られた。

本手法が Zhou らの手法を上回る結果が得られた理

表 4 Zhou らの手法と提案手法の顔認識性能比較 (CMU-PIE database 利用). 画素のグルーピングとして図 1 (a) に示す部分領域を利用 (識別率は各セル内で Zhou らの手法/提案手法の順に表示)

Table 4 Performance comparison of Zhou's method (the left number in each cell) and our method (the right number in each cell) using grouping regions of Fig. 1 (a) by using CMU-PIE database.

Recognition rate (Zhou et al.'s method / our method)[%]													
training	f08	f09	f11	f12	f13	f14	f15	f16	f17	f20	f21	f22	Average
test													
f08	-	96/100	96/99	87/100	66/100	60/99	46/96	29/81	22/27	85/97	78/100	53/97	65/90
f09	94/99	-	96/99	96/100	90/100	87/99	56/100	40/96	24/72	84/97	96/100	68/100	75/96
f11	94/99	91/100	-	97/100	72/100	72/99	38/100	28/90	16/51	100/99	94/100	51/100	69/94
f12	88/99	94/100	97/99	-	88/100	93/99	57/100	41/99	28/91	94/99	100/100	76/100	78/99
f13	56/99	87/100	59/99	85/100	-	100/99	90/100	71/100	50/97	54/99	87/100	100/100	76/99
f14	51/99	85/100	63/99	93/100	100/100	-	90/100	66/100	49/100	59/99	91/100	99/100	77/99
f15	33/99	40/99	37/99	49/100	85/100	88/99	-	93/100	78/100	32/99	49/100	97/100	62/99
f16	19/91	26/99	26/90	32/99	59/100	44/99	84/100	-	93/100	26/91	31/100	63/100	46/97
f17	14/75	28/93	19/78	26/97	50/100	41/99	68/100	94/100	-	19/78	26/99	44/100	39/92
f20	90/99	85/100	99/99	97/100	65/100	69/99	38/99	26/90	21/60	-	93/100	53/100	67/95
f21	79/99	94/100	93/99	100/100	88/100	94/99	62/100	49/100	28/96	91/99	-	76/100	78/99
f22	43/99	65/100	46/99	75/100	99/100	99/99	97/100	76/100	59/100	43/99	74/100	-	70/99
Average	60/96	72/99	66/96	76/99	78/100	77/99	66/99	56/96	42/81	63/96	74/99	71/99	67/97

由は、以下の 2 点に起因すると考えられる。1 つは、Zhou らの手法ではランバートモデルを仮定したのに対し、本手法では鏡面反射成分や影のような拡散成分以外の反射成分を統計的にモデル化した点である。2 つ目は、本手法では、1 枚の登録画像が部分的に影である場合においても新しい照明条件下の顔画像を正しく合成できるように、顔表面位置間の相関を考慮した点である。

5. む す び

各人物ごとに 1 枚のみ登録画像がある場合の、任意照明条件下における統計的な shape from shading の枠組みに基づいた新しい顔認識手法を提案した。本手法は、初めに、複数人物を様々な照明条件下で撮影した画像集合を用いて統計モデルを学習する。次に、新しい人物顔の形状と反射特性を、学習した統計モデルと 1 枚の登録画像から MAP 推定によって復元する。最後に、新しい照明条件下の顔画像を、求めた形状と反射特性、MAP 推定で得られた誤差項によって生成する。

従来手法と比較して、提案手法の最大の利点は、顔表面の法線とアルベドの MAP 推定を行う際に顔表面位置間の相関を考慮した点である。その結果、登録画像の顔が部分的に影になっている場合においても、顔の法線とアルベドを正しく推定することができた。さらに、統計量算出とモデリングの精度向上を図るため、以下の 2 つの工夫を行っている。画像内の画素のグルーピング、および法線とアルベドの MAP 推定の

際に有効な画素を選択する信頼性指標の導入である。実画像を用いた顔認識実験を行い、従来手法と比較して提案手法が有効であることを示した。

本実験では相関を考慮する部分領域を手作業で定めたが、自動領域分割は今後の課題とする。また、提案手法における誤差項を調査し、たとえば鏡面反射、相互反射、影、その他の成分に分解していくことが効果的であると考えられる。さらに顔の姿勢変化のように、照明条件以外で顔認識性能を低下させるその他の要因に対応する見えの変化をモデル化することも有効と考える。なお、本研究は長尾健司氏の生前に共同で行っていたものである。

参 考 文 献

- 1) Atick, J.J., Griffin, P.A. and Redlich, A.N.: Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images, *Neural Computation*, Vol.8, No.6, pp.1321-1340 (1996).
- 2) Basri, R., Roth, D. and Jacobs, D.: Clustering appearances of 3D objects, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.414-420 (1998).
- 3) Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.711-720 (1997).

- 4) Belhumeur, P.N. and Kriegman, D.J.: What is the set of images of an object under all possible lighting conditions?, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.270–277 (1996).
- 5) Blanz, V. and Vetter, T.: Face Recognition Based on Fitting a 3D Morphable Model, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.25, No.9, pp.1063–1074 (2003).
- 6) Chellappa, R., Wilson, C.L. and Sirohey, S.: Human and Machine Recognition of Faces, A Survey, *Proc. IEEE*, Vol.83, pp.705–740 (1995).
- 7) Georghiades, A.S., Kriegman, D.J. and Belhumeur, P.N.: Illumination cones for recognition under variable lighting: faces, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.52–58 (1998).
- 8) Horn, B.K.P.: Height and gradient from shading, *IJCV*, Vol.5, pp.37–75 (1990).
- 9) Marschner, S.R., Westin, S.H., Lafortune, E.P.F., Torrance, K.E. and Greenberg, D.P.: Image-based BRDF Measurement Including Human Skin, *Proc. 10th Eurographics Workshop on Rendering*, pp.139–152 (1999).
- 10) Murase, H. and Nayar, S.K.: Visual learning and Recognition of 3-D Objects from Appearance, *Int. J. Computer Vision*, Vol.14, pp.5–24 (1995).
- 11) Nagao, K.: Face recognition by distribution specific feature extraction, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.I-1278–1285 (2000).
- 12) Oja, E.: *Subspace Methods for Pattern Recognition*, Research Studies Press Ltd. (1980).
- 13) Okabe, T. and Sato, Y.: Object Recognition Based on Photometric Alignment Using RANSAC, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.I-221–228 (2003).
- 14) Okabe, T. and Sato, Y.: Support Vector Machines for Object Recognition under Varying Illumination Conditions, *Proc. Asian Conf. Computer Vision*, pp.724–729 (2004).
- 15) 山元直政, 重成一真, 尺長 健, 仮想部分空間法による照明に影響されない顔認識, 電子情報通信学会論文誌, Vol.J84-D-II, No.8, pp.1753–1761 (2001).
- 16) Shimano, M. and Nagao, K.: Simultaneous Optimization of Class Configuration and Feature Space for Object Recognition, *Proc. Int. Conf. Pattern Recognition*, No.2, pp.7–10 (2004).
- 17) Sim, T. and Kanade, T.: Combining Models and Exemplars for Face Recognition: An Illuminating Example, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition Workshop on Models versus Exemplars in Computer Vision* (2001).
- 18) Sim, T., Baker, S. and Bsat, M.: The CMU Pose, Illumination, and Expression (PIE) Database, *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp.53–58 (2002).
- 19) Zhao, W. and Chellappa, R.: Illumination-insensitive face recognition using symmetric shape-from-shading, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Vol.1, pp.286–293 (2000).
- 20) Zhao, W., Chellappa, R., Phillips, J. and Rosenfeld, A.: Face Recognition in Still and Video Images: A Literature Survey, *ACM Computing Surveys*, Vol.35, pp.399–458 (Dec.2003).
- 21) Zhou, S.K., Chellappa, R. and Jacobs, D.W.: Characterization of Human Faces under Illumination Variations Using Rank, Integrability, and Symmetry Constraints, *Proc. European Conf. Computer Vision (ECCV2004)*, Vol.1, pp.588–601 (2004).

付 録

A.1 b_{MAP} の導出

ガウス分布を, 平均ベクトル μ と共分散行列 C を用いて $N(\mu, C)$ と記述することにする. b を与えられたときの i の条件付分布 $P(i|b)$ と b の分布 $P(b)$ を式 (2) で表される反射モデルより, 以下のように定義すると,

$$P(i|b) = N(S^T b + \mu_e, \Sigma_e),$$

$$P(b) = N(\mu_b, C_b),$$

i を与えられたときの b の条件付分布 $P(b|i)$ はガウス分布となり,

$$P(b|i) = N(b_{MAP}, R_b),$$

その平均と共分散を各々 b_{MAP} と R_b とする. ベイズの定理に従って, 下記を得ることができる.

$$\begin{aligned} P(b|i) &\propto P(i|b)P(b) \\ &\propto N(S^T b + \mu_e, \Sigma_e) \times N(\mu_b, C_b) \\ &\propto \exp(-G/2), \end{aligned}$$

ここで, G は以下に示されるようなスカラー量である.

$$\begin{aligned} G &= (i - S^T b - \mu_e)^T \Sigma_e^{-1} (i - S^T b - \mu_e) \\ &\quad + (b - \mu_b)^T C_b^{-1} (b - \mu_b) \\ &= b^T (S \Sigma_e^{-1} S^T + C_b^{-1}) b \\ &\quad - 2((i - \mu_e)^T \Sigma_e^{-1} S^T + \mu_b^T C_b^{-1}) b + const. \end{aligned}$$

$P(b|i)$ を最大化する b を導出するため, b に関して G の微分をとり, これを 0 とする.

$$\frac{\partial G}{\partial \mathbf{b}} = 2(S\Sigma_e^{-1}S^T + C_b^{-1})\mathbf{b} - 2((\mathbf{i} - \boldsymbol{\mu}_e)^T \Sigma_e^{-1}S^T + \boldsymbol{\mu}_b^T C_b^{-1})^T = 0.$$

これによって、 \mathbf{b} の MAP 推定は以下のように

$\mathbf{b}_{\text{MAP}} = R_b(S\Sigma_e^{-1}(\mathbf{i} - \boldsymbol{\mu}_e) + C_b^{-1}\boldsymbol{\mu}_b)$,
得られる。ここで、 $G = \mathbf{b}^T R_b^{-1}\mathbf{b} - 2\mathbf{b}_{\text{MAP}}^T R_b^{-1}\mathbf{b} + \text{const}$
と表され、 $R_b^{-1} = S\Sigma_e^{-1}S^T + C_b^{-1}$ である。

A.2 e_{MAP} の導出

確率ベクトル \mathbf{x} と \mathbf{y} が jointly Gaussian distribution に従うと仮定するならば、

$$P\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}\right) = N\left(\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{pmatrix}\right),$$

と記述される。 \mathbf{y} が与えられる場合の \mathbf{x} の条件付分布は次のように

$$P(\mathbf{x}|\mathbf{y}) = N(\mathbf{m}, C_m), \quad (13)$$

記述される。ここで、 $\mathbf{m} = \boldsymbol{\mu}_x + R_{xy}R_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$ であり、 $C_m = R - R_{xy}R_{yy}^{-1}R_{yx}$ である。

異なる照明条件間の誤差項を jointly Gaussian distribution としてモデル化しているため、

$$P\left(\begin{pmatrix} e(s_{\text{new}}) \\ e(s) \end{pmatrix}\right) = N\left(\begin{pmatrix} \boldsymbol{\mu}_{e_{\text{new}}} \\ \boldsymbol{\mu}_e \end{pmatrix}, \begin{pmatrix} \Sigma_{e_{\text{new}}} & R^T \\ R & \Sigma_e \end{pmatrix}\right),$$

と表される。誤差項 $e(s_{\text{new}})$ は、 $e(s)$ が与えられ、ガウス分布として

$$P(e(s_{\text{new}})|e(s)) = N(e_{\text{MAP}}, \Sigma), \quad (14)$$

と表される。ここで、 $e_{\text{MAP}} = \boldsymbol{\mu}_{e_{\text{new}}} + R\Sigma_e^{-1}(e(s) - \boldsymbol{\mu}_e)$ であり、 $\Sigma = \Sigma_{e_{\text{new}}} - R^T\Sigma_e^{-1}R$ である。

(平成 17 年 9 月 20 日受付)

(平成 18 年 3 月 20 日採録)

(担当編集委員 尺長 健)



島野美保子 (正会員)

1995 年東京大学工学部物理工学科卒業。1997 年同大学大学院工学系研究科物理工学専攻修士課程修了。同年より松下電器産業(株)先端技術研究所等にて、画像認識やコンピュータビジョンに関する研究に従事。モデルベースと事例ベースの融合に関心あり。電子情報通信学会会員。



長尾 健司

1985 年東京大学工学部精密工学科卒業、工学博士(東大電子情報専攻)。1985 年より、松下電器産業(株)の研究所にて、自然言語処理、コンピュータビジョンの研究に従事。1992 年 10 月~1995 年 2 月まで米国マサチューセッツ工科大学 AILAB 客員研究員。動画画像を用いた構造復元、物体認識、顔画像認識に従事。2005 年 1 月逝去。



岡部 孝弘 (正会員)

1997 年東京大学理学部物理学卒業。1999 年同大学大学院理学系研究科物理学専攻修士課程修了。2000 年同博士課程中退。2001 年より東京大学生産技術研究所技官(技術職員)。コンピュータビジョン、コンピュータグラフィックスに関する研究に従事。2005 年度電子情報通信学会論文賞、MIRU2004、MIRU2005 優秀論文賞、平成 16 年度 PRMU 研究奨励賞を受賞。電子情報通信学会、IEEE 各会員。



佐藤いまり

1994 年慶應義塾大学総合政策学部卒業、2002 年東京大学大学院学際情報学府修士課程修了。2005 年同大学院博士課程修了、学際情報学博士。現在、国立情報学研究所助手、科学技術振興機構さきがけ研究員。光源環境の推定、明るさ解析に関する研究に従事。1992~1993 年 Carnegie Mellon University (CMU), Center for Machine Translation, Research Assistant, 1994~1996 年 CMU, The Robotics Institute, Visiting Scholar。1997~2000 年東京大学生産技術研究所、研究補助。2002~2004 年日本学術振興会特別研究員。



佐藤 洋一（正会員）

1997年カーネギーメロン大学計算機科学部ロボティクス学科博士課程修了。同年より東京大学生産技術研究所研究機関研究員，講師，助教授を経て，現在同大学大学院情報学環助教授。Ph.D. in Robotics。コンピュータビジョン，ヒューマン・コンピュータ・インタラクション，コンピュータグラフィックスに関する研究に従事。MIRU2000 最優秀論文賞，MIRU 優秀論文賞（MIRU2004，MIRU2005），1999年度山下記念研究賞，1999年度日本バーチャルリアリティ学会論文誌論文賞，2005年度電子情報通信学会論文賞等を受賞。電子情報通信学会，日本バーチャルリアリティ学会，ACM，IEEE 各会員。
