

# キャストシャドウの存在する顔画像の照明正規化

松川 徹<sup>†</sup> 岡部 孝弘<sup>†</sup> 佐藤 洋一<sup>†</sup>

<sup>†</sup> 東京大学生産技術研究所 〒 153-8505 東京都目黒区駒場 4-6-1

E-mail: ††{te2,takahiro,ysato}@iis.u-tokyo.ac.jp

あらまし キャストシャドウの影響による照明の急激な変動 (例えば, 帽子の影) は, 屋外画像の顔等にしばしば観測され, 顔認識における解決が必要な重大な問題となっている. 本研究では, キャストシャドウ存在下においても動作が可能な, スモールスケールとラージスケール特徴に基づく照明正規化手法を提案する. 提案手法では, 急激な照明変動による照明正規化の困難性を 2 つのアイデアにより解決している. まず, 顔の反射特性の統計的性質を利用し, 顔画像の高周波成分が, キャストシャドウによって生じたものか, 目や口等の顔の特徴によって生じたものかを区別する. このようにして, キャストシャドウ存在下においても照明変動の影響が受けにくいスモールスケール特徴を抽出する. 次に, キャストシャドウを含む誤差を陽に考慮するように Quotient Image に基づく照明正規化手法を拡張し, 照明変動に頑健なラージスケール特徴を抽出する. 実験により, 提案照明正規化手法が, キャストシャドウの存在する困難な照明条件において, 良好な照明正規化結果を得ることが出来ることを確認した.

キーワード 顔認識, 照明正規化, キャストシャドウ, スモール & ラージスケール特徴, Total Variation, Quotient Image

## 1. はじめに

照明変動問題は, 顔認識において解決が必要な極めて重要な問題の一つである. 主成分分析 (PCA) や線形判別分析 (LDA) 等の従来の顔認識手法は照明の影響を受けやすい. また, LBP 特徴等の照明の影響に頑健な特徴量であっても厳しい照明条件の下では, 必ずしも有効に機能しないことが報告されている [1].

従来の様々な照明条件下における顔認識技術は, 2 つのアプローチに分類できる. 一つは, 照明錐 [2] や球面調和部分空間 [3][4] 等の照明部分空間を利用する方法である. もう一方は, SQI [5], DCT [6], LTV [7], TT [1], NDF [8] 等の照明の正規化手法である. 前者が様々な照明条件下における画像の集合から顔の見えをモデル化するのに対して, 後者では, 照明に不変な (影響を受けにくい) 特徴を 1 枚の画像から抽出する. 1 枚の画像から照明変動を抑制出来ることは, 様々な応用にとって望ましい [9]. そのため, 本研究では後者の照明の正規化手法に着目する.

人間の顔を含む多くの物体は, 反射率 (テクスチャ) と陰影の要素の積としてよく近似できる. 前者は物体表面の内在的性質を示しており, 画像が取得された照明条件とは独立している. 反対に, 後者は物体の形状に加えて照明条件に依存する.

拡散反射成分と陰 (アタッチドシャドウ) の成分は画像の低周波成分である [3], [4]. これを受けて既存の照明正規化手法では, 1 枚の画像からスモールスケール特徴 (主に画像の高周波成分, 上記の反射率に相当) を抽出して認識に用いる一方で, ラージスケール特徴 (主に画像の低周波成分, 上記の陰影に相当) を破棄している. 例えば, MSR [10] と SQI [5] では入力画像を平滑化した同

一の画像で割ることによって正規化を行い, DCT [6] や Wavelet [11] では周波数解析に基づき入力画像の低周波成分を破棄することで正規化を行う.

近年, Xie 等 [12] はスモールスケールとラージスケール特徴に基づく照明正規化手法が, スモールスケール特徴だけをを用いる手法よりも顔認識に有効であることを報告している. この結果は, ラージスケール特徴もまた顔形状等の人物の個人性を含むことを示している.

しかし, 影 (キャストシャドウ) は画像に高周波成分を引き起こす [13], [14]. 例えば, 極端な角度の光源から照らされた鼻等の顔のパーツにより生じる影や帽子等の他の物体によって生じる高周波成分は影の境界に高周波成分を発生させる. 一枚の入力画像のみから画像の高周波成分を顔の反射率によって生じるものか他物体によって生じたものかを区別することは困難であり, そのため, 従来の手法により分離されたスモールスケール特徴とラージスケール特徴はキャストシャドウの影響を受けてしまう. 従って, 極度な照明条件下や他物体からの影を含む顔照明を正規化することは未だに極めて困難な問題として残る.

この問題を解決するため, 本研究ではキャストシャドウが存在する場合にも有効に働く, スモールスケールとラージスケール特徴に基づいた, 照明の影響を受けにくい顔画像の正規化手法を提案する. 具体的に, 顔画像の統計的性質を利用した Total Variation モデル [7] により, スモールスケールとラージスケール特徴の分離を行う手法と照明基底の他に誤差基底を用いることで, キャストシャドウを含む誤差への対応を可能とした Quotient Image [15] に基づくラージスケール特徴の正規化手法を提案する.

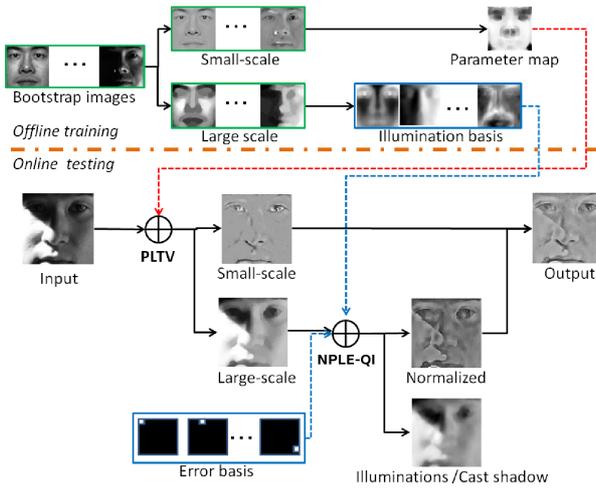


図1 提案手法のブロック線図.

また、公開データベースと独自に収集したデータベースを利用して、評価実験を行う。前者は極度な光源で撮影された画像を含み、後者は他の物体によって投げかけられたキャストシャドウを含む。提案手法が他の最新手法と比較しても同等以上の性能を得ることが出来ることを確認する。

## 2. 提案手法の概要

提案手法では、対象を特定のクラス(正面顔<sup>注1)</sup>)に限定し、照明正規化を行う。図1に示すように提案手法は学習段階とオンラインのテスト段階からなる。

学習段階では、多数の光源方向から照らされた多数の対象物体の顔を含むブートストラップ顔画像から、顔の事前知識を獲得する。まず、各ブートストラップ画像をスモールスケールとラージスケール特徴に分離しておき、これらを用いて、オンラインのテスト段階で利用するための統計的性質を学習する。具体的に、ブートストラップ画像のスモールスケール特徴から、各画素において顔のスモールスケール特徴の出現しやすさを表すパラメータマップを獲得する。また、未知の照明条件下における未知の人物のラージスケール特徴を近似する照明基底を獲得する。

テスト段階では、分離と正規化の各ステップにおいて獲得した統計的性質を用いる。分離のステップでは、Total Variatoin モデル[7]を用いて入力画像をスモールスケールとラージスケール特徴に分離する。スモールスケール特徴の統計的性質を表すパラメータマップが高周波成分を顔によって生じたものか、キャストシャドウによって生じたものかを区別するために用いられる。このように、キャストシャドウの影響の受けにくスモールスケール特徴を獲得する。この分離手法を *PriorLTV (PLTV)* と呼ぶ。

しかしながら、その一方で、個人性を含んでいるラージスケール特徴[12],[16],[17]にキャストシャドウの影

響が押し付けられてしまう。そこで、キャストシャドウを考慮に入れてラージスケール特徴の照明を Quotient Image [15] に基づく手法で正規化する。正規化のステップで、ラージスケール特徴は平均顔の照明変動基底と誤差基底の線形結合で近似的に表現したのち、ラージスケール特徴を近似表現で割って正規化することで、個人性を抽出する。ここで、平均顔のラージスケール特徴は照明基底とエラー基底の線形結合によって表現される。このようにキャストシャドウの影響を受けにくい正規化されたラージスケール特徴を獲得する。この正規化手法を *Non-Point Lighting and Error Quotient Image (NPLE-QI)* と呼ぶ。

最終的に、これらのスモールスケールと正規化されたラージスケール特徴を統合し、顔認識に用いる。

以降の章で、これらの手法の詳細を記述する。3章では、顔の統計量を利用したスモールスケールとラージスケール特徴の分離手法について述べる。4章では、キャストシャドウを陽に考慮にいれたラージスケール特徴の正規化手法について述べる。

## 3. 顔の事前知識を利用した画像分離

まず、入力画像を Logarithmic Total Variation (LTV) [7] に基づく手法を用いて、スモールスケールとラージスケール特徴に分離する。高周波成分を顔の反射率によって生じたものとキャストシャドウによって生じたものとに区別するため、対象の事前知識、即ち、ブートストラップ画像から獲得された顔のスモールスケール特徴の統計的性質を用いる。

### 3.1 PriorLTV

画素  $i(x, y)$  は反射率  $\rho(x, y)$  と陰影  $s(x, y)$  の積として表現できる:

$$i(x, y) = \rho(x, y)s(x, y), \quad (1)$$

ここで  $(x, y)$  は画像平面の座標である。

式(1)の対数を取って、次のように積の形式は加法的な形式に変換することができる:

$$f(x, y) = v(x, y) + u(x, y), \quad (2)$$

ここで  $f(x, y) = \log i(x, y)$ ,  $v(x, y) = \log \rho(x, y)$ , また  $u(x, y) = \log s(x, y)$  である。 $v$  をスモールスケール特徴、また  $u$  をラージスケール特徴と呼ぶ<sup>注2)</sup>。

TV-L1 モデル[19]を適用することによって、LTV は  $u$  と  $v$  を以下のように推定する:

$$u = \arg \min_{\hat{u}} \int \int |\nabla \hat{u}| + \lambda |f - \hat{u}|_1 dx dy, \\ v = f - u. \quad (3)$$

(注2): 式(1)の分離はユニークでない。例えば、Chen 等は  $\rho(x, y)/\rho_l(x, y)$  をスモールスケール特徴、 $s(x, y)\rho_l(x, y)$  をラージスケール特徴と呼んでいる。ここで、 $\rho_l(x, y)$  は顔の反射率の低周波成分である。

(注1): 側面顔等にも同様な処理で対応可能である。

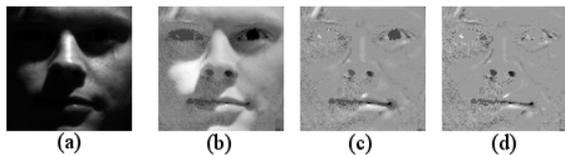


図2 LTVの結果例。(a)入力画像,(b)-(d)スモールスケール特徴 $v$ ,各々は $\lambda = 0.1, 0.3, 0.5$ の場合。パラメータ $\lambda$ を適切に設定しなければ,結果は顔の部分を除外したり,キャストシャドウを保ったりする。

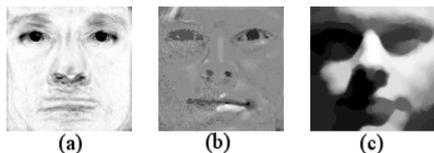


図3 PLTVの結果例。(a)パラメータマップ。明るい画素が $\lambda(x, y)$ が大きいことを示す。(b)(c)図2の画像の分離結果。(b)はスモールスケール特徴,(c)はラージスケール特徴。

ここで,  $\int \int |\nabla u| dx dy$  は  $u$  の Total Variation, すなわち  $|\nabla u|$  は滑らかさの項<sup>(注3)</sup>である。反対に  $|f - u|$  は  $u$  を  $f$  に近似する項, また,  $|v|$  を最小化する項である。スカラーパラメータ  $\lambda$  がこれらの項の間のトレードオフを調整する。

LTVによって抽出されたスモールスケール特徴は顔の反射率のエッジを保存でき, LTVは様々な照明条件下における顔認識の最高水準の手法の一つである。しかしながら, LTVは顔特徴によって生じる高周波成分とキャストシャドウによって生じる高周波成分を分離することは困難である。これはキャストシャドウによって生じる高周波成分をスモールスケール特徴から取り除くコストと顔特徴によって生じる高周波成分を取り除くコストが同一であるためである。図2に示すように, キャストシャドウの影響が $\lambda$ が小さい時にスモールスケール特徴に残っている。また,  $\lambda$ が大きい場合にはスモールスケール特徴にキャストシャドウの影響があまり観測されなくなるが, 顔の判別的な特徴も部分的に失われてしまう。

本研究では, パラメータ $\lambda$ を適応的に画素毎に決定する。即ち, 式(3)における大域的なパラメータを画素毎の局所的なパラメータ $\lambda(x, y)$ に変更する。この $\lambda(x, y)$ をパラメータマップと呼ぶ。 $\lambda(x, y)$ が小さい値を持つ時, 全体のコスト関数における滑らかさ項の寄与が増加する。これはラージスケール特徴 $u(x, y)$ が滑らかになり易く, また, スモールスケール特徴 $v(x, y)$ が $(x, y)$ において高周波成分を含みやすくなる傾向があることを意味する。

従って, パラメータマップは顔の反射率が高周波成分を含む場所では小さな値を取り, 顔の反射率が平坦な部分では大きな値を取るべきである。次の節で, このようなパラメータマップをブートストラップサンプルから学

習する手法を説明する。

顔の事前知識をLTVに活用することから, この手法を *PriorLTV (PLTV)* と呼ぶ。図3にパラメータマップと分離結果を示す。図2と比較して, スモールスケール特徴からキャストシャドウの効果を除外することに成功していることが確認できる。

本研究では, 最適化問題に Parametric Maximum Flow (PMF) アルゴリズム [18] を用いた。PMFは従来のSOCPやPDFに基づくソルバ [19] よりも高速に最適化問題を解くことが出来る。スモールスケール特徴 $v$ にマイナーなアーチファクトが観測されることがしばしば見られた。そこで, 従来手法 [12] と同様な方法で平滑化することによって, この影響を抑制した。

### 3.2 パラメータマップ

顔の反射率は均一ではない; 目や鼻, 口等の領域では高周波成分が支配的であり, 額や頬等の領域では低周波成分が支配的になる。顔のブートストラップ画像から獲得される知識を利用することで, パラメータマップ $\lambda(x, y)$ を高周波成分が支配的な領域では小さく, 逆の場合には大きくするよう設定する。

まず, ブートストラップ画像から主に正面光源から照らされたサンプルを選ぶ。これらを用いて,  $N$ 個のスモールスケール特徴 $\{v_n\}_{n=1}^N$ を大域的パラメータ $\lambda_0$ を用いたLTVを適用することによって獲得する。 $v_n$ においては多くの判別的な顔特徴を残すことが望ましいため,  $\lambda_0$ は照明変動下で用いるLTVよりも小さい値に設定する。元の画像にキャストシャドウを含んでいないため, 計算されたスモールスケール特徴 $v_n$ を正確と考える。

$v_n$ の絶対値は各画素において顔特徴の出現しやすさを示しているため, これらの平均値 $|\bar{v}| = \sum_{n=1}^N |v_n| / N$ を顔の事前知識として用いる。これを用いて, パラメータマップを $|\bar{v}|$ の線形関数として以下のように定義する。

$$\lambda(x, y) = \begin{cases} \lambda_{\max} & (|\bar{v}(x, y)| < T_L) \\ b - a|\bar{v}(x, y)| & (T_L \leq |\bar{v}(x, y)| \leq T_H) \\ \lambda_{\min} & (|\bar{v}(x, y)| > T_H), \end{cases} \quad (4)$$

ここで,  $a = (\lambda_{\max} - \lambda_{\min}) / (T_H - T_L)$ , また  $b = \lambda_{\max} - aT_L$  である。 $T_L$ と $T_H$ は $|v_n|$ のパーセンタイルによって決定される上限と下限である。パラメータマップの例は図3に示している。

### 4. キャストシャドウ存在下での照明正規化

前章で述べた手法により, キャストシャドウの影響を受けにくいスモールスケール特徴を得る。しかし, キャストシャドウの影響はラージスケール特徴に残されている。そこで, この影響を考慮に入れてラージスケール特徴の正規化を行う。

具体的には, 入力画像と同一照明下における平均顔でラージスケール特徴を割ることで, 照明の影響を抑え, 人

(注3): Total Variationの離散(画像)領域では様々な定義がある。本研究では, 16-近傍の anisotropic discretization [18] を用いた。

物固有の特徴を抽出する。NPL-QI [20] と同様な方法により平均顔のラージスケール特徴を表現する照明基底を用いるが、低次元の照明基底はキャストシャドウを上手く表現できない。これはキャストシャドウは高周波成分であること、より重要なことは、顔に影を投げかける物体は任意であるためである。本研究では、この問題に対処するため、任意形状の影に対応できる基底表現を用いてこの誤差を取り除く。また、鏡面反射成分や拡散反射成分およびアタッチドシャドウにおける照明部分空間の誤差も取り除くことが、照明変動に頑健な認識を行うために重要である。提案手法では、このような誤差もキャストシャドウの誤差と同時に取り除くことが可能である。

前章において、スモールスケールとラージスケール特徴を  $v(x, y)$ ,  $u(x, y)$  のように画像座標系に対する関数で表現していた。本章では、記述の簡潔性のため、画素値をラスタスキャンし、各特徴を  $v, u$  と列ベクトルで表現する。

#### 4.1 誤差に対応した Quotient Image

まず、ブートストラップ画像中の様々な照明条件下における複数人の人物の顔画像に対して LTV を用いて分離することによって、ラージスケール特徴の集合  $\{u_m\}_{m=1}^M$  を得る。次に、これらを用いた行列  $[u_1, u_2, \dots, u_M]$  に対して主成分分析を適用することによって得られる上位  $K$  個の固有値に対応する固有ベクトルの行列  $A = [a_1, a_2, \dots, a_K]$  を得る。もしラージスケール特徴がキャストシャドウや鏡面反射の影響を含んでいなければ、 $u = Al$  という近似が行える。ここで、 $l = (l_1, l_2, \dots, l_K)^T$  は線形結合の係数である。しかし、キャストシャドウや鏡面反射の存在しているデータの場合、この近似では誤差が含まれる。そこで、まず、低次元の照明部分空間で表現できない重大な誤差を含むかを判定するために、最小2乗当てはめ、

$$\min_l \|u - Al\|_2^2, \quad (5)$$

を行い、当てはめ誤差  $err = \|u - Al\|_2$  を計算する。誤差  $err$  が  $\epsilon \|u\|_2$  より、大きい場合には Sim らが、画像を低次元の陰影と誤差で表現したのと同様に [21], 誤差項  $e$  を導入してラージスケール特徴を以下のように近似する<sup>(注4)</sup>。

$$u \sim Al + e. \quad (6)$$

Quotient Image [15], [20] と同様な方法によって、ラージスケール特徴とそのブートストラップ画像から計算されて近似されたもの（入力画像と同一照明条件を持つ平均顔）との比により、顔の特徴の判別的な情報を抽出する。ラージスケール特徴が対数空間での議論となること

(注4): なお Sim 等は顔画像における照明基底の誤差に着目しているが、本研究と異なり、学習画像に存在する画像におけるピクセル毎の誤差の統計量を利用しており、任意の誤差に対応していない。

を思い出すと、比は差分になる。従って、ラージスケール特徴の Quotient Image, すなわち正規化されたラージスケール特徴  $q$  を以下のように定義する。

$$q = u - (Al + e). \quad (7)$$

同様に、重大な誤差が存在しないと判断されたデータに対しては、 $q = u - Al$  として定義する。

本研究では次節で説明する方法により、 $Al + e$  の近似を行い、それを用いて正規化を行う。 $q$  を得た後、スモールスケール特徴と正規化されたラージスケール特徴の和  $v + q$  を顔認識に用いる。

このラージスケール特徴の正規化手法を *Non Point Light and Error Quotient Image (NPLE-QI)* と呼ぶ。ここで、NPL という名称は、PCA により獲得された照明基底が非点光源 (*Non Point Light*) の拡散反射を表現する球面調和基底 [3], [4], [20] に線形な関係があることに由来する<sup>(注5)</sup>。

#### 4.2 誤差と陰影基底の当てはめ

次に、NPLE-QI において係数  $l$ , および誤差項  $e$  を推定する手法について述べる。本研究では、誤差項を基底関数の線形結合として表現する:

$$e = Bb. \quad (8)$$

ここで、 $B$  は基底関数の集合、 $b$  はそれらの係数である。

本研究では画素単位での誤差を仮定し、 $B = \xi I_e$  とする。ここで、行列  $I_e$  は  $P \times P$  ( $P$  は画素数) の単位行列である<sup>(注6)</sup>。また、スカラーパラメタ  $\xi$  は最適化において誤差項の寄与を調整する項である。

照明項  $Al$  と誤差項  $e$  を統合、当てはめの誤差を最大  $\epsilon \|u\|_2$  まで許容することにより、係数  $w = (l^T, b^T)^T$  を以下のように得る。

$$\begin{aligned} \min_w & \|w\|_1 \\ \text{subject to} & \|u - Cw\|_2 \leq \epsilon \|u\|_2, \end{aligned} \quad (9)$$

ここで、 $C = [A, B]$  である。この L1 ノルム最小化問題を解くアルゴリズムは多数存在する。本研究では l1-magic の l1qc logbarrier 関数 [22] を用いる。

このモデルは、顔画像という類似性の高い基底と任意の誤差基底からなる cross-and-bouquet モデルとして理論

(注5): 本研究では、PCA を対数空間において行っているが、これは、 $u$  の指数を取ってから基底表現するよりも安定した結果が得られるためである。我々は、対数空間においても非点光源の照明変動が、少数の基底で表現できることを確認している。

(注6): 画素値が大きい場合、計算量が大きい。高速化のため、画像サイズを縮小し、低解像度の空間で陰影・誤差基底表現を求めた後、元の解像度へバイキュービック補間でアップサンプリングする。ラージスケール特徴は、画像の低周波成分であるため、このような処理に問題はない。実験では、 $100 \times 100$  の画像を  $20 \times 20$  に縮小して基底当てはめを行った。

### Algorithm 1 NPLe-QI for large scale feature

**Input:** Large scale feature  $u$ .

- 1:  $\hat{u} = \text{downsample}(u)$ .
- 2:  $\hat{l} = \text{argmin} |\hat{u} - A\hat{l}|_2$ .
- 3: **if**  $|\hat{u} - A\hat{l}|_2 > \epsilon |\hat{u}|_2$  **then**
- 4:    $\hat{w} = \text{arg min} |w|_1, \quad \text{s. t. } |\hat{u} - Cw|_2 \leq \epsilon |\hat{u}|_2$ .
- 5:    $q = u - \text{upsample}(C\hat{w})$ .
- 6: **else**
- 7:    $q = u - \text{upsample}(A\hat{l})$ .
- 8: **end if**

**Output:** Normalized large scale feature  $q$ .

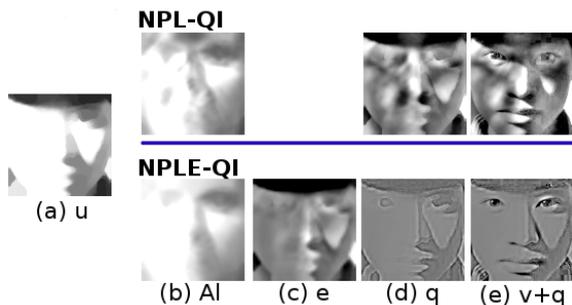


図4 Quotient Image によるラージスケール特徴の正規化.

的な解釈が行われている L1 最小化における密な誤差復元 [23][24] に基づいたものである。本研究では、陰影基底は顔の照明変動を良く近似する直交基底を用いているが、L1 最小化では疎な陰影の解が得られる。しかし本研究目的では、密な誤差を復元することに注目しており、陰影の疎性は必要ではない。上記の最小化を L2 ノルム最小化に置き換えることにより、同様な効果が得られるが、この場合、パラメタ依存度が大きくなるという問題があるため、L1 ノルムを利用している。

NPLe-QI のアルゴリズムを Algorithm.1. にまとめる。また、処理結果の例を図 4 に示す。(a) は入力ラージスケール特徴、(b) は照明項、(c) は誤差項である。(b) のように、提案手法の誤差項を用いない場合に相当する NPL-QI では、帽子のキャストシャドウが良く表現できておらず、また、実際には暗くない左頬に暗い画素が伝播している。これに対して、NPLe-QI では、誤差項がキャストシャドウを上手く表現できており、暗い画素が伝播する現象はあまり発生していない。その結果、(d) の Quotient Image (正規化ラージスケール特徴) 及び、認識に用いる (e) のスモールスケール特徴との和は、NPL-QI よりもキャストシャドウに敏感でないことが確認できる。

## 5. 実験

### 5.1 設定

**データセット:** 実験には次の 4 つのデータセットを用いる。: Extended Yale B [2], CMU PIE [25], CAS-PEAL [26], 独自に収集したキャストシャドウを含むデータセット (CAST と呼ぶ)。全てのデータセットで、正面向きの顔のみを用いる。Extended Yale B は、38 人の被験者が 64 種

類の光源方向から照らされた画像を含んでいる。これらの画像は顔の正面の方向と光源とのなす角度に基づき、5 つのサブセットに分割されている。CMU PIE は、21 個の異なる光源から照らされた 68 人の被験者の画像からなる。CAS-PEAL は、様々な照明方向と装飾品をつけた 1040 人の被験者からなる。本研究では、この照明方向のセットと装飾品セットを用いるが、装飾品のセットからはキャストシャドウの効果にのみ興味があるため、帽子を被った画像のみを選び出したデータを用いる。しかし、このデータベースでは帽子を被ったデータは正面方向からのみ照明が照らされているため、キャストシャドウの影響は大きくない。そこで、本研究では、帽子の影と他の物体から投げかけられるキャストシャドウを含む顔画像を収集した。CAST は 14 人の被験者が、帽子等の 4 種類の障害物と様々な光源から照らされた 50 枚の照明状態の画像からなる。

**認識方法:** 全ての画像を目の座標値に従い切り出し、 $100 \times 100$  画素を持つ 256 階調のグレースケール画像としてリサイズする。多くの関連研究 [5], [7], [8], [16] で用いられているように、正規化相関に基づく最近傍法による識別を行う。各人物に対して、理想的な照明条件 (正面光源) の下で撮影された 1 枚の画像のみを参照画像として登録する。

**ブートストラップ画像:** 様々な照明条件の下で照らされた 10 人の被験者の画像をブートストラップ画像として用いる。CMU PIE を Extended Yale B のブートストラップ画像として用い、Extended Yale B を他のデータベースに対して用いる。

**パラメタ設定:** PLTV のパラメタマップの学習に用いる LTV のパラメタを  $\lambda_0 = 10^{-2}$  とする。また、本研究では、 $T_L$  は  $|\bar{v}_n|$  の 1 番目、 $T_H$  は 99 番目のパーセンタイルとする。 $\lambda_{min}$ ,  $\lambda_{max}$  を上限と下限を示す主なパラメタ値であり、これら変化させて考察を行う。デフォルトでは、 $\lambda_{min} = 0.2$ ,  $\lambda_{max} = 0.4$  とする。NPLe-QI においては、誤差の寄与パラメタ  $\xi$  を 0.1 と設定する。基底による近似の精度を表す  $\epsilon$  は、パラメタを変化させた場合について考察を行うが、デフォルトでは、 $\epsilon = 0.04$  とする。

### 5.2 PriorLTV の効果

この実験では、本研究で提案した PLTV と LTV を比較するため、スモールスケール特徴のみを認識に用いる。様々なパラメタにおける結果: 異なるパラメタにおける LTV と PLTV の性能を評価した。LTV のパラメタは大域的なパラメタ  $\lambda$  である。PLTV のパラメタは  $\lambda_{min}$  と  $\lambda_{max}$  である。実験では、 $\lambda_{min} = \lambda - 0.1$ ,  $\lambda_{max} = \lambda + 0.1$  のように設定した。結果を図 5 に示す。左のグラフが Extended Yale B の平均認識率である。提案した PLTV が、多くの場合で LTV よりも良い結果を示していることが確認できる。右のグラフが Extended Yale B の各サブセットにおける PLTV と LTV との差である。

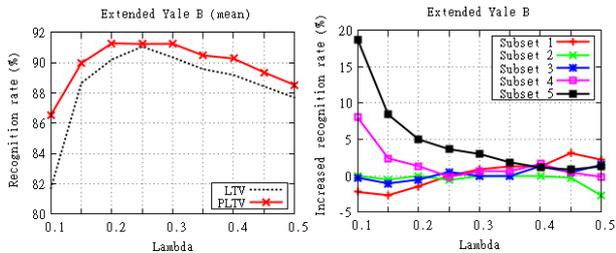


図5 LTV と PLTV の比較. 左: Extended Yale B の平均認識率. 右: 各サブセットの認識率の差 ( PLTV - LTV ).

PLTV と LTV の差は、サブセット 1-3 のような光源方向がほぼ正面の場合には差が見られないが、PLTV は、サブセット 4, 5, すなわち、顔の横方向から照明が照らされ鼻や額等によるキャストシャドウの影響が現れる場合、により有効に機能していることが確認できる。また、パラメータ  $\lambda$  を増加させた時には PLTV と LTV の差が小さいことも確認できる。これは、図 2 の観察と関連しており、 $\lambda$  ( $\lambda_{\min}$ ) を大きくするに従い、顔の判別的な特徴が除外されるためである。 $\lambda$  を大きく設定したとき、PLTV を用いたとしても判別的な特徴を抽出することは困難である。これらの結果より、PLTV や LTV の結果は  $\lambda = 0.3$  の周辺が最良であると考えられる。最適なパラメータを正解情報のないテストデータのみから選定することは課題として残るが、切り出し条件の同じ異なるデータベース間で最適な値は近い値を持つ傾向を確認しており、この選定は難しい。

理想的なパラメータにおける比較: 最適なパラメータ  $\lambda_{\min} = 0.2$  と  $\lambda_{\max} = 0.4$  における認識率を表 1 に示す。PLTV が多くの場合で、LTV よりも優れた結果を生み出していることが解る。なお、CMU PIE は極端な照明条件を含んでおらず、キャストシャドウの影響が少ないため、LTV で高い精度が既に達成されている。CAS-PEAL Hat においては、PLTV によって、LTV の精度より高い認識率が得られなかった。これは、学習に Extended Yale B を用いたため、人物の違いによるずれが生じていたためである。同一データベースの同一人物でパラメータを取得した場合を表の\*付で表記しているが、この場合、CAS-PEAL Hat では、精度が向上していることが確認できる。スモールスケール特徴と正規化を行ったラージスケール特徴とを組み合わせさせた場合 (S&L と表記する) も表 1 に示す。この場合においても、PLTV が LTV よりもほぼすべての場合で優れていることが確認できる。図 7 に結果画像を示す。PLTV によって得られた特徴が、LTV よりも顔の特徴を正確に捉えていることも確認できる。なお、実行段階における PLTV の計算コストは LTV におけるそれと同一であるため、多くの場合では、PLTV を利用した方がよい。

### 5.3 NPLe-QI の効果

この実験では、提案した NPLe-QI と NPL-QI を比較す

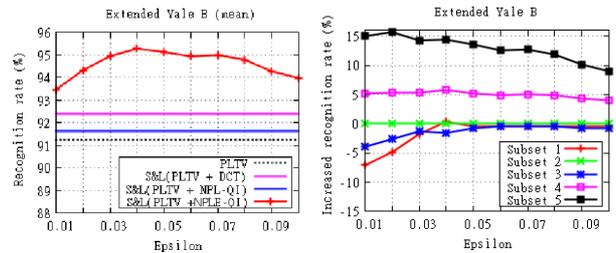


図6 ラージスケール特徴の効果. 左: Extended Yale B の平均認識率. 右: 各サブセットの認識率の差 ( S&L (PLTV + NPE-QI) - S&L(PLTV + NPL-QI) ).

る。参考に、表 1 にラージスケール特徴ではなく、元画像に適用した場合の結果を示す。NPLe-QI は、NPL-QI と比較して多くの場合、特にキャストシャドウが画像中で支配的になる状況において、よい性能を示すことが確認できる。本研究において NPLe-QI は、ラージスケール特徴において判別的な特徴を抽出することを目的としており、これを調査するため、スモールスケール特徴と合わせた場合における評価を詳細に述べる。実験では、PLTV のパラメータを  $\lambda_{\min} = 0.2$  と  $\lambda_{\max} = 0.4$  に設定した。異なるパラメータにおける結果: NPLe-QI において、パラメータ  $\epsilon$  を変化させた時の認識率を調査した。結果を図 6 に示す。Extended YaleB における、全てのサブセットの平均認識率を左の図に示している。S&L(PLTV+NPLe-QI) の性能は、他の組み合わせ (PLTV+DCT, PLTV+NPL-QI) と元の PLTV よりも優れていることが確認できる。右の図は、Extended Yale B の各サブセットにおける S&L(PLTV+NPL-QI) から S&L(PLTV+NPLe-QI) への差を示している。困難な照明条件で撮影されているサブセット 4, 5 において、提案手法である S&L(PLTV+NPLe-QI) が S&L(PLTV+NPL-QI) に対して非常に高い性能を示していることが確認できる。これは、NPL-QI は元の  $u$  を誤差項  $e$  を用いずに近似するため、キャストシャドウが正規化ラージスケール画像  $q$  に影響を与えたためである。一方で、SLe-QI は元のラージスケール特徴  $u$  を誤差項  $e$  を用いて近似するため、キャストシャドウの効果を正規化ラージスケール特徴  $q$  から、取り除くことが出来たためである。キャストシャドウの影響の低いサブセット 1, 2 においては、誤差項の寄与を大きくした場合 ( $\epsilon$  が小さい場合、精度が低下している。これは、誤差項には、顔に関する制約がないために、個人性を含む特徴も誤差として除去されるという負の側面も存在するためである。従って、キャストシャドウの影響の少ない場合には、誤差項の寄与を小さくした方がよい。以降の実験では、パラメータを  $\epsilon = 0.04$  と設定する。

図 7 に S&L(PLTV+DCT), S&L(PLTV+NPL-QI), と S&L(PLTV+NPLe-QI) の結果の画像例を示す。S&L(PLTV+NPLe-QI) が、視覚的にも他の手法よりも優れていることが確認できる。

他手法との比較: 提案手法と近い手法に加えて、次の

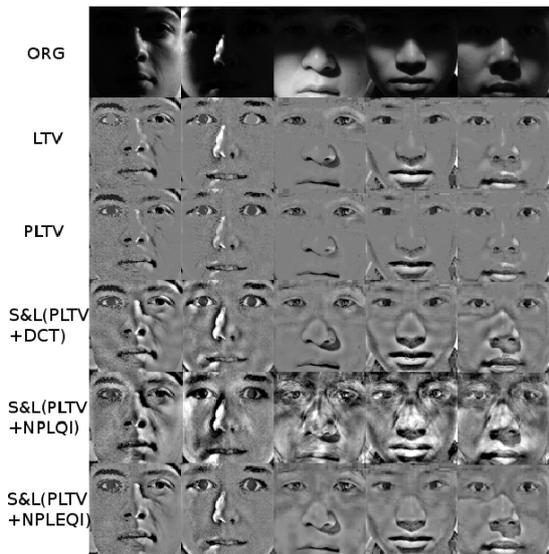


図7 顔正規化画像の例. 左の2つは Extended Yale B, 右の3つは CAST.

他の照明正規化手法との認識率を調査した; ヒストグラム平坦化 (HE), TT [1]<sup>注7)</sup>, DCT [6], NDF [8]. これらの全手法を独自に実装し, 同一の認識条件の下で最良の認識結果が得られるようにパラメタを注意深く設定した. 結果を表1に示す. ここで, 他の手法の結果は元の論文と多少異なっているが, これは, 顔の切り出しの方法や解像度が異なっているためである. 提案手法である S&L(PLTV+NPLe-QI) が多くの場合, 特にキャストシャドウの影響が深刻な場合に, 最もよい性能を示していることが確認できる.

#### 5.4 向き・表情変動のある場合

提案手法では, 正面向き・無表情の顔の統計量を用いて顔照明正規化を行うが, 現実の顔認識の問題では, 向き・表情変動を含む状況が考えられる. そこで, 提案手法を向き・表情変動を含む顔の様々な光源下での画像データベースで評価した. 評価には, Multi-PIE [27] の Session-1 の 249 人の 20 光源変動を用いた. 角度の異なる顔に関して, カメラ番号 05\_1, 05\_0, 04\_1 で撮影された顔, また, 表情の異なる顔に関して笑顔の画像を用いた. 認証には, 各実験条件におけるフラッシュなしの顔を登録した. 図8に Multi-PIE の画像例を, 結果結果を表2に示す.

提案手法は, 微小に向き・表情の異なる顔に対しても精度低下が少ないことが確認できる. PLTV においては,  $\lambda_{min}, \lambda_{max}$  が適切なパラメタの範囲内であれば, 顔の特徴を削る等の大きな悪影響を起さないのである. NPLe-QI においては, ラージスケール特徴は画像の低周波成分であり, さらにずれが生じていても誤差項により取り除く効果があることから, 大きな精度低下が起らなかったためである. このように提案手法は, 向き・表

(注7): TT [1] の全認識アルゴリズムの内, 照明正規化の部分のみを用いた.

表2 Multi-PIE における各手法の比較 (認識率 (%)).\*は対象データベースと同一データベースの同一条件における 10 人の画像をブートストラップサンプルとして用いた場合.

Method	C1	C2	C3	C4
HE	53.31	71.42	63.73	70.78
LTV	99.66	99.26	99.87	99.87
PLTV	99.34	99.07	99.83	99.64
PLTV*	99.43	99.13	99.85	99.64
S&L(LTV + NPL-QI)	99.85	99.75	99.98	99.51
S&L(LTV + NPLe-QI)	99.81	99.77	<b>100</b>	99.62
S&L(LTV + NPL-QI*)	99.92	<b>99.89</b>	99.98	<b>100</b>
S&L(LTV + NPLe-QI*)	<b>99.94</b>	<b>99.89</b>	<b>100</b>	<b>100</b>

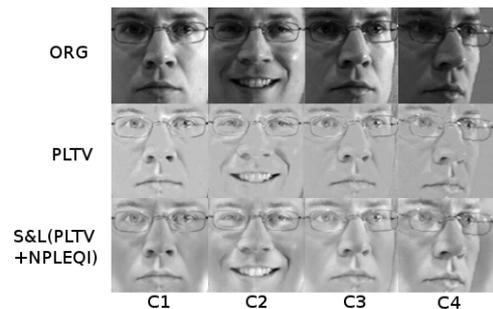


図8 Multi-PIE の例. C1: 正面 (カメラ 5\_1), C2: 正面+表情 (カメラ 5\_1), C3: 側面 1 (カメラ 5\_0), C4: 側面 2 (カメラ 4\_1).

情変動のある場合にも機能するが, 同一の状況から学習した場合の方が性能が高いことも確認できる.

## 6. まとめ

本研究では, 極度な照明条件と他物体からのキャストシャドウが存在する環境下においても動作が可能な, 顔の照明正規化手法を提案した. 顔の反射率の統計的性質とキャストシャドウの誤差表現に基づき, 提案手法は照明の影響を受けにくい顔画像のsmallスケールとlargeスケール特徴を抽出することを可能とした. 実験により, 提案手法が他の最新手法と比較してキャストシャドウの存在する条件で優れた性能, その他の場合では同等の性能を達成できることを確認した.

本研究の拡張には, 多数の方向性が考えられる. 例えば, 図6の中央のグラフの実験結果からパラメタ  $\epsilon$  は入力画像が撮影された照明条件に基づき設定すべきであることが確認できる. 従って, 入力画像の照明状態に応じてパラメタを適応的に設定していくこと [28] と組み合わせることが, 今後の課題としてあげられる. また, 実験で確認したように提案手法は, 顔の向きや表情変化等の微小なずれの影響化においても機能するが, さらに精度向上のためには, 位置合わせを行うことが有効と思われる. そこで, 標準的顔モデルに基づく統計量を入力画像へ合わせて変形させること [29] と組み合わせることとずれに対する頑健性向上を行うことも考えられる.

謝辞 本研究の一部は, 文科省の戦略推進費「安心・安全な社会のための犯罪・テロ対策技術などを実用化するプログラム」による.

表 1 異なるデータベースにおける各手法の比較 (認識率 (%)). \*は対象データベースの本人画像をブートストラップサンプルとして用いた場合.

Method	Extended Yale B						CMU PIE	CAS PEAL		CAST
	Subset1	Subset2	Subset3	Subset4	Subset5	Mean		Light	Hat	
HE	98.22	97.37	53.95	15.70	14.12	55.87	57.92	4.37	25.00	18.43
TT [1]	79.56	99.74	86.32	83.56	84.17	86.67	99.46	18.28	24.73	76.00
DCT [6]	88.89	<b>100</b>	88.16	88.53	87.02	89.53	<b>99.85</b>	20.29	33.66	81.00
NDF [8]	84.44	<b>100</b>	87.89	89.33	<b>91.96</b>	90.57	<b>99.85</b>	21.89	32.50	79.71
LTV [7]	90.22	<b>100</b>	89.74	84.59	87.16	90.34	98.85	21.58	30.54	82.29
PLTV	90.67	98.68	91.05	85.63	91.29	91.46	97.54	22.47	29.64	81.57
PLTV *	90.67	98.68	90.53	86.81	90.15	91.36	98.92	24.37	33.12	80.80
NPL-QI [20]	98.66	<b>100</b>	94.21	72.00	52.35	83.44	96.31	21.89	21.88	53.00
NPLe-QI	92.89	99.74	94.74	<b>90.96</b>	91.44	93.95	99.54	23.32	28.04	68.00
S&L(LTV + DCT) [12]	91.11	<b>100</b>	92.11	89.78	88.30	92.26	98.85	20.60	<b>34.38</b>	82.86
S&L(LTV + NPL-QI) [12]	99.11	<b>100</b>	96.57	85.33	77.18	91.64	98.62	26.13	30.00	79.71
S&L(LTV + NPLe-QI)	99.11	<b>100</b>	95.79	<b>90.96</b>	88.73	94.92	98.92	25.50	31.34	83.71
S&L(PLTV + DCT)	91.56	<b>100</b>	91.32	89.93	89.16	92.39	98.85	20.37	33.84	82.57
S&L(PLTV + NPL-QI)	<b>99.56</b>	<b>100</b>	96.58	85.78	78.17	92.02	98.62	<b>26.04</b>	25.09	79.29
S&L(PLTV + NPLe-QI)	<b>99.56</b>	<b>100</b>	<b>96.84</b>	<b>90.96</b>	91.58	<b>95.78</b>	98.92	25.72	30.18	<b>84.29</b>

## 文 献

- [1] X. Tan and B. Triggs, Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions, *IEEE Trans. on Image Processing*, vol.19, no.6, pp.1635–1650, 2010.
- [2] A.S. Georgiades, P.N. Belhumeur and D.J. Kriegman, From few to many: Illumination cone models for face recognition under variable lighting and pose, *IEEE Trans. on PAMI*, vol.23, no.6, pp.643–660, 2001.
- [3] R. Ramamoorthi, Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of Lambertian Object, *IEEE Trans. on PAMI*, 2002.
- [4] R. Basri and D.W. Jacobs, Lambertian Reflectance and Linear Subspaces, *IEEE Trans. on PAMI*, vol.25, no.2, pp.218–233, 2003.
- [5] H. Wang, S.Z. Li and Y. Wang, Face recognition under varying lighting conditions using self quotient image, In Proc. *FG*, 2004.
- [6] W. Chen, M.J. Er and S. Wu, Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain, *IEEE Trans. on SMC:B*, vol. 36, no.2, pp.458–466, 2006.
- [7] T. Chen, W. Yin, X.S. Zhou, D. Comaniciu and T.S. Huang, Total variation models for variable lighting face recognition, *IEEE Trans. on PAMI*, vol.28, pp.1519–1524, 2006.
- [8] L.-H. Chen, Y.-H. Yang, C.-S. Chen and M.-Y. Cheng, Illumination Invariant Feature Extraction Based on Natural Images Statistics - Taking Face Images as An Example, In Proc. *CVPR*, 2011.
- [9] X. Tan, S. Chen, Z.-H. Zhou and F. Zhang, Face recognition from a single image per person: A Survey, *Pattern Recognition*, vol.39, pp.1725–1745, 2006.
- [10] D.J. Jobson, Z. Rahman and G.A. Woodell, A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes, *IEEE Trans. on Image Processing*, vol.6, no.7, pp.965–976, 1997.
- [11] T. Zhang, B. Fang, Y. Yuan, Y.Y. Tang, Z. Shang, D. Li and F. Lang, Multiscale facial structure representation for face recognition under varying illumination, *Pattern Recognition*, vol.42, pp. 251–258, 2009.
- [12] X. Xie, W.-S. Zheng and P.C. Yuen, Normalization of face Illumination Based on Large- and Small- Scale Features, *IEEE Trans. on IP*, vol.115, pp.1384–1394, 2011.
- [13] R. Ramamoorthi and P.Hanrahan, A Signal-Processing Framework for Inverse Rendering, In Proc. *SIGGRAPH*, 2001.
- [14] T. Okabe, I. Sato and Y. Sato, Spherical Harmonics vs. Haar Wavelets: Basis for Recovering Illumination from Cast Shadows, In Proc. *CVPR*, 2004.
- [15] A. Shashua and T. Riklin-Raviv, The Quotient Image: Class-Based Re-Rendering and Recognition with Varying Illuminations, *IEEE Trans. on PAMI*, vol.23, no.12, pp.129–139, 2001.
- [16] X. Xie, W.-S. Zheng, J. Lai and P.C. Yuen, Face Illumination Normalization on Large and Small Scale Features, In Proc. *CVPR*, 2008.
- [17] G. An, J. Wu and Q. Ruan, Independent Gabor Analysis of Multi-scale Total Variation-Based Quotient Image, *IEEE Signal Processing Letters*, vol.15, pp.186–189, 2008.
- [18] D. Goldfarb and W. Yin, Parametric Maximum Flow Algorithms for Fast Total Variation Minimization, *SIAM journal on Scientific Computing*, vol.32, no.5, pp.3712–1743, 2009.
- [19] A. Chambolle, V. Caselles, D. Cremers, M. Novaga and T. Pock, An Introduction to Total Variation for Image Analysis, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, vol.9, pp.263–310, 2010.
- [20] H. Wang, S.Z. Li and Y. Wang, Generalized Quotient Images, In Proc. *CVPR*, 2004.
- [21] T. Sim, T. Kanade, Combining Models and Exemplars for Face Recognition: An Illumination Example, In Proc. *CVPR Workshop on Models versus Exemplars in Computer Vision*, 2001.
- [22] E. Candes, J. Romberg,  $l^1$  magic : Recovery of Sparse Signals via Convex Programming, 2005.
- [23] J. Wright, A.Y. Yang, A. Ganesh, S.S. Satri and Y. Ma, Robust Face Recognition via Sparse Representation, *IEEE Trans. on PAMI*, vol.115, pp.1384–1394, 2009.
- [24] J. Wright and Y. Ma, Dense Error Correction via L1 Minimization, *IEEE Trans. on Information Theory*, vo.56, no.7, pp.3540–3560, 2010.
- [25] T. Sim, S. Baker and M. Bsat, The cmu pose illumination, and expression (pie) database, *IEEE Trans. on PAMI*, vol.115, pp.1384–1394, 2003.
- [26] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang and D. Zhao, The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations, *IEEE Trans. on SMC*, vol.115, pp.1384–1394, 2008.
- [27] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, Multi-PIE, *Image and Vision Computing*, vol.28, no.5, pp.807–813, 2010.
- [28] H. Han, S. Shan, L. Qing, X. Chen and W. Gao, Lighting Aware Preprocessing for Face Recognition across Varying Illumination, In Proc. *ECCV*, 2010.
- [29] R. Kumar, M. Jones and T.K. Marks, Morphable Reflectance Fields for enhancing face recognition, In Proc. *CVPR*, 2010.