

辞書学習を用いた圧縮センシングによる 多波長・多方向光源下画像の効率的な獲得

小林 直人¹ 岡部 孝弘²

概要: 物体の見えは、物体を照らす光源の波長と方向の両方に依存する。近年、光源装置の進化により多波長・多方向光源下での物体の撮影が可能になったが、光源数の増加に伴い撮影に要する時間も増大してしまうという問題がある。そこで本稿では、スパース性に基づく圧縮センシングの枠組みで、多波長・多方向光源下の画像を効率的に獲得する手法を提案する。提案手法では、辞書学習により解空間を制限することで、圧縮率・復元精度の向上を目指す。特に、訓練データとしてシミュレーションデータを用いることで実画像訓練データを取得する手間を省くとともに、画素値がスペクトル項と幾何学項の積の和で表現されることに着目して、効率的に辞書を学習する。

1. はじめに

物体の見えは、物体を照らす光源の波長と方向の両方に依存する。近年、光源装置の進化により多波長・多方向光源下での物体の撮影が可能になり、そのイメージベーストモデリング [5] や材質 (マテリアル) 認識 [4] への応用が注目を集めている。これらの応用では、光源装置の光源数を増やせば増やすほど精度が向上すると考えられるが、その一方で、撮影に要する時間も増大してしまうという問題がある。

そこで本稿では、圧縮センシング [3] の枠組みで、多波長・多方向光源下における画像を効率的に獲得する手法を提案する。提案手法では、スパース性に基づく圧縮センシングにおいて、訓練データを用いた辞書学習 [1] により解空間を制限することで、圧縮率・復元精度の向上を目指す。

辞書を学習するにあたって、訓練データを増やせば増やすほど、様々な物体に柔軟に対応できる辞書の作成が可能であると考えられる。しかしながら、訓練データを増やすには、様々な物体の多波長・多方向光源下の画像を撮影する必要がある。つまり、効率的に画像を獲得するためには、予め非効率的に膨大な枚数の画像を獲得する必要があり、主客転倒である。

そこで本研究では、実画像訓練データを取得する手間を省くために、シミュレーションデータを利用する。例えば、単純な手法として、反射特性 (物体表面のなめらかさなど) や分光反射率 (物体表面に固有の色) の異なる様々

な物体の合成画像を訓練データとして用いることが考えられる。しかしながら、このような単純な手法では、様々な反射特性と分光反射率の組み合わせが考えられるために、訓練データの量が膨大になってしまう。この問題を解決するために、提案手法では、画素値がスペクトル項と幾何学項の積の和で表現されることに着目して、効率的に辞書を学習する。具体的には、スペクトル項を表現するための辞書と幾何学項を表現するための辞書を各々作成して、それらのクロネッカー積により多波長・多方向光源下の画素値を表現する辞書を得る。

二つの辞書のクロネッカー積に基づく提案手法には、少ない訓練データで多様な物体に柔軟に対応できるという利点がある。実画像を用いた実験を行ったところ、提案手法の訓練データのサイズは単純な手法の約 $1/33$ であるにも関わらず、提案手法を用いた復元画像の PSNR は単純な手法を用いたものよりも大きいという結果が得られた。

2. 関連研究

2.1 辞書学習による圧縮センシング

ある N 次元のベクトル \mathbf{x} を原信号とし、 M 行 N 列 ($M < N$) の観測行列 \mathbf{A} を用いて、観測信号 \mathbf{y} が $\mathbf{y} = \mathbf{A}\mathbf{x}$ と表されているとする。辞書学習による圧縮センシングでは、辞書 (N 次元基底ベクトルの集合) \mathbf{D} と未知の結合係数 \mathbf{c} を用いて、原信号を $\mathbf{x} = \mathbf{D}\mathbf{c}$ のように表現する。ここで辞書 \mathbf{D} は、結合係数がスパースになるように、訓練データを用いて予め学習されているものとする。観測行列 \mathbf{A} と辞書 \mathbf{D} の積を新たな観測行列 $\mathbf{A}' = \mathbf{A}\mathbf{D}$ と考えることができるので、 $\mathbf{y} = \mathbf{A}'\mathbf{c}$ に対して結合係数 \mathbf{c} がスパース

¹ 九州工業大学大学院情報工学府先端情報工学専攻
² 九州工業大学大学院情報工学研究院知能情報工学研究系

になるように L_0 再構成を行うことで、結合係数 \mathbf{c} を推定することができる。推定した結合係数 \mathbf{c} に辞書 \mathbf{D} を掛けることで、原信号 \mathbf{x} を復元することができる。

2.2 画像の効率的な獲得への応用

Ajdin ら [2] は、圧縮センシングの枠組みで、多波長・多方向光源下の画像を効率的に獲得する手法を提案している。具体的には、多波長・多方向光源下において物体表面上のある点で観察される画素値を並べた行列が、低ランク行列とスパース行列の和で表現できると仮定して、低ランク性とスパース性の両方を考慮した Compressive Sparse and Low-Rank (CSRL) と呼ばれる再構成法を提案している。

これに対して提案手法は、同じ圧縮センシングの枠組みではあるが、辞書学習に基づいている点が異なる。提案手法では、多波長・多方向光源下において物体表面上のある点で観察される画素値を並べた行列が低ランク行列になることを暗に仮定している。しかしながら、訓練データを用いて辞書を学習することで、任意の低ランク行列を解として許容するのではなく、実物体の反射特性や分光反射率を考慮して解空間を制限し、圧縮率・復元精度の向上を目指す。

3. 提案手法

3.1 拡散反射と鏡面反射

Shafer[7] は、物体表面で観察される反射光のスペクトルが拡散反射によるスペクトルと鏡面反射によるスペクトルの和で表現されるという二色性反射モデルを提案した。例えば、拡散反射のモデルとして Lambert モデルを、鏡面反射のモデルとして Phong モデルを仮定すると、物体表面上のある点の画素値 r は

$$r = \int l(\lambda)\rho(\lambda)c(\lambda)d\lambda \cos\theta + \int l(\lambda)c(\lambda)d\lambda \cos^n\alpha \quad (1)$$

のように表現される。第1項は拡散反射成分で、 $l(\lambda)$, $\rho(\lambda)$, $c(\lambda)$, および、 θ は、光源の分光分布、物体表面の分光反射率、カメラの分光感度、および、法線と光源方向のなす角度を表す。第2項は鏡面反射成分で、 α , および、 n は、正反射方向と視線方向のなす角度、および、鏡面反射の鋭さを表す指数である。

式(1)の第1項は、光源の分光分布に依存するスペクトル項 $\int l(\lambda)\rho(\lambda)c(\lambda)d\lambda$ と光源の方向に依存する幾何学項 $\cos\theta$ の積で表現されている。同様に、第2項もスペクトル項 $\int l(\lambda)c(\lambda)d\lambda$ と幾何学項 $\cos^n\alpha$ の積で表現されている。したがって、画素値はスペクトル項と幾何学項の積の和で表現される。この性質は、Lambert モデルや Phong モデルに限らず、多くの反射モデルに近似的にあてはまる。

3.2 辞書の学習

図1に示すように、 I 個の光源クラスタ、各クラスタあ

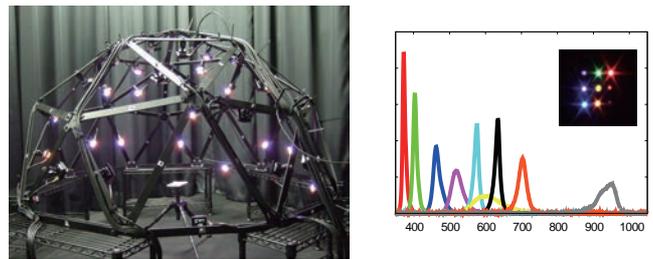


図1 光源装置

たり J 個の異なる分光分布を持つ LED から構成される光源装置を用いて多波長・多方向光源下の画像を獲得する。物体表面上のある点において、 I 個の光源方向、各光源方向あたり J 個の光源色の下で観察される画素値を並べた $N(=I \times J)$ 次元のベクトルを原信号とする。つまり、画素ごとに独立に再構成を行う。

単純な辞書学習法では、訓練データとして、様々な物体の多波長・多方向光源下の画像(物体ごとに N 枚)を用いる。しかしながら、単純な手法では、様々な反射特性と分光反射率の組み合わせが考えられるために、訓練データの量が膨大になってしまう。具体的には、 O_1 種類の物体を用いるとき、 $O_1 = R_1$ (拡散反射や異なるパラメータを持つ鏡面反射などの反射特性の種類) $\times C_1$ (分光反射率の種類) なので、訓練画像の画素数を P_1 とし、画素値を K バイトで表現すると、訓練データの総量は $R_1 \times C_1 \times P_1 \times I \times J \times 3$ (カラー画像の RGB) $\times K$ バイトとなる。

提案手法では、3.1で述べたように画素値がスペクトル項と幾何学項の積の和で表現されることに着目して、効率的に辞書を学習する。具体的には、 I 次元基底ベクトルの集合である幾何学項の辞書と J 次元基底ベクトルの集合であるスペクトル項の辞書を学習しておき、それらのクロネッカー積により、 N 次元基底ベクトルの集合である辞書を得る。クロネッカー積とは任意のサイズの行列の間に定義される二項演算で、その結果は区分行列として与えられる。幾何学項の辞書を I 行 I' 列の行列 $\mathbf{G} = (g_{ii'})$, スペクトル項の辞書を J 行 J' 列の行列 $\mathbf{S} = (s_{jj'})$ とすると、クロネッカー積 $\mathbf{G} \otimes \mathbf{S}$ は

$$\mathbf{G} \otimes \mathbf{S} = \begin{pmatrix} g_{11}\mathbf{S} & \cdots & g_{1I'}\mathbf{S} \\ \vdots & \ddots & \vdots \\ g_{I1}\mathbf{S} & \cdots & g_{II'}\mathbf{S} \end{pmatrix} \quad (2)$$

で与えられる。

幾何学項の学習には球の画像を用いる。球面上の法線は任意の法線を含むため、球の画像を用いて学習した辞書は、あらゆる向きの表面に対応できると考えられる。幾何学項の辞書は色情報とは無関係なので、分光反射率の種類 $C_2 = 1$, かつ、合成画像のバンド数 = 1 となる。したがって、反射特性(拡散反射や異なるパラメータを持つ鏡面反射)の数を R_2 , 訓練画像の画素数を P_2 とすると、幾何学項の

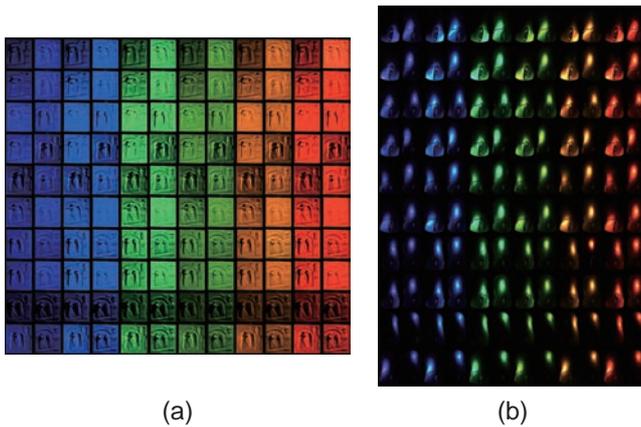


図 2 多波長・多方向光源下の原画像: (a) レリーフ, (b) マウス

訓練データの総量は $R_2 \times P_2 \times I \times K$ バイトとなる。

スペクトル項の学習には、様々な光源の分光分布、物体表面の分光反射率、および、カメラの分光感度に対する色のシミュレーションデータを用いる。分光反射率の種類を D 個とすると、スペクトル項の訓練データの総量は $D \times J \times 3 \times K$ バイトとなる。ところが実際には、画像の画素数 P_2 は分光反射率の種類 D よりも十分に大きいために、スペクトル項の辞書を学習するための訓練データのサイズは、幾何学項の辞書を学習するための訓練データのサイズよりも十分に小さい。したがって、提案手法が要する訓練データの総量は、およそ $R_2 \times P_2 \times I \times K$ となる。

以上をまとめると、単純な手法の要する訓練データのサイズが $R_1 \times C_1 \times P_1 \times I \times J \times 3 \times K$ バイトであるのに対して、提案手法の訓練データの総量は $R_2 \times P_2 \times I \times K$ である。したがって、反射特性の種類が等しく ($R_1 = R_2$)、かつ、画素数も等しい ($P_1 = P_2$) とし、提案手法の訓練データのサイズは、単純な手法の $1/(3C_1J)$ になる。逆に、両方で訓練データのサイズをそろえると、提案手法は、単純な手法よりもより多くの反射特性や画素数を扱うことができる。

4. 実験

4.1 実験条件

実験では、先行研究 [2], [4], [5] と同様の光源装置により、 $20(= I)$ 方向・ $6(= J)$ 色の光源下で撮影された合計 $120(= N = I \times J)$ 枚の画像を用いた。図 2 に石膏製のレリーフとプラスチック製のマウスの各々 120 枚の画像を示す。最初の 2 列が紫色の光源で照らしたときの 20 枚の画像、以下、青色、緑色、黄緑色、オレンジ色、赤色の光源で照らした各々 20 枚の画像であり、物体表面で観察される色が光源色に依存して変化する様子がわかる。また、同じ色の光源に照らされた 20 枚の画像に着目すると、物体表面で観察される濃淡が光源方向に依存して変化する様子もわかる。実験では、これらの 120 枚の画像を、 $M(= 6, 12, 18, \dots, 120)$ 枚の観察画像から復元した。観測行列はベルヌーイラン

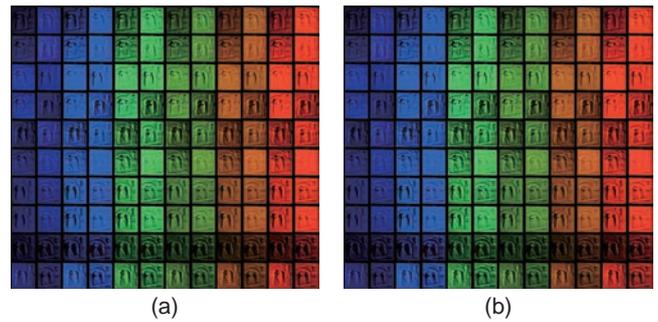


図 3 圧縮率 50%のときの復元結果 (レリーフ): (a) 単純な手法, (b) 提案手法

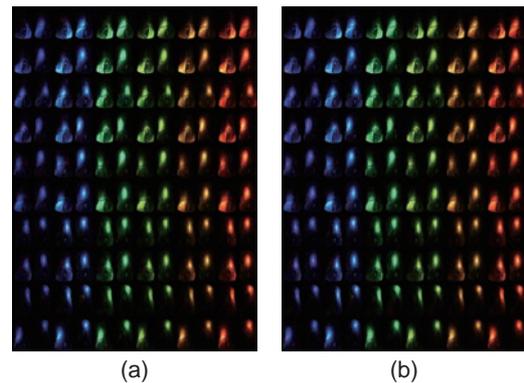


図 4 圧縮率 50%のときの復元結果 (マウス): (a) 単純な手法, (b) 提案手法

ダム行列 [8] とし、復元には SPAMS の ElasticNet[9] を用いた。

提案手法では、 $11(= R_2)$ 種類の反射特性 (拡散反射と 10 種類の鏡面反射) の球の合成画像から幾何学項の辞書を、 $1250(= D)$ 種類の分光反射率 [6] から計算した色データからスペクトル項の辞書を作成した。比較のために、単純な手法では、 $20(= R_1 \times C_1)$ 種類の球 (10 色の拡散反射面と鏡面反射面) の画像から辞書を作成した。画像の画素数は $180 \times 240(= P_1 = P_2)$ として、画素値は double 型 ($K=8$) とした。辞書学習には K-SVD 法 [1] を用いた。

4.2 実験結果

図 3 から図 6 に、圧縮率が 50%のときの単純な手法と提案手法による両物体の画像の復元結果を示す。図 5 と図 6 の拡大画像、および、原画像との差分画像^{*1} から、単純な手法も提案手法も、拡散反射が支配的な画素の再構成精度が良い一方で、鏡面反射成分やレリーフの凹領域などで観察される影 (キャストシャドウ) の再構成精度が良くないことがわかる。また、微妙な色合いや鏡面反射成分の復元精度において、提案手法が単純な手法よりも優れていることがわかる。図 7 に、圧縮率を変えたときの復元画像の PSNR を示す。定量的にも、提案手法が単純な手法よりも

*1 RGB ごとに画素値の差分の絶対値を求め、差分が見やすくなるように定数倍した。

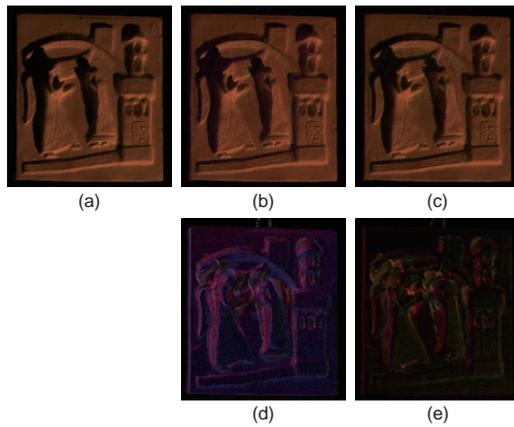


図 5 復元結果と原画像の差分（レリーフ）：(a) 原画像，(b) 単純な手法，(c) 提案手法，(d) 単純な手法の差分，(e) 提案手法の差分

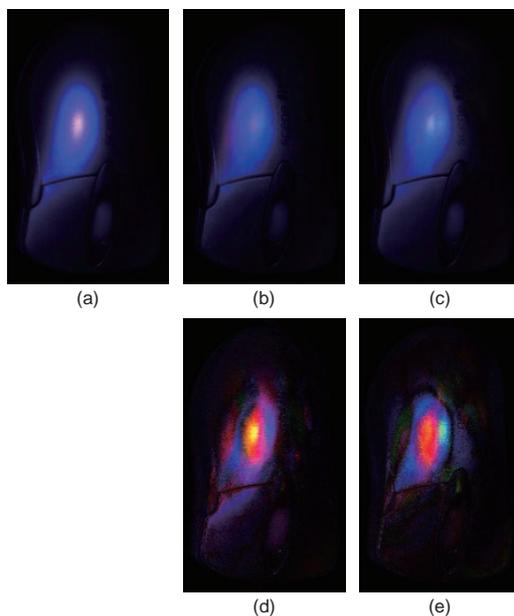


図 6 復元結果と原画像の差分（マウス）：(a) 原画像，(b) 単純な手法，(c) 提案手法，(d) 単純な手法の差分，(e) 提案手法の差分

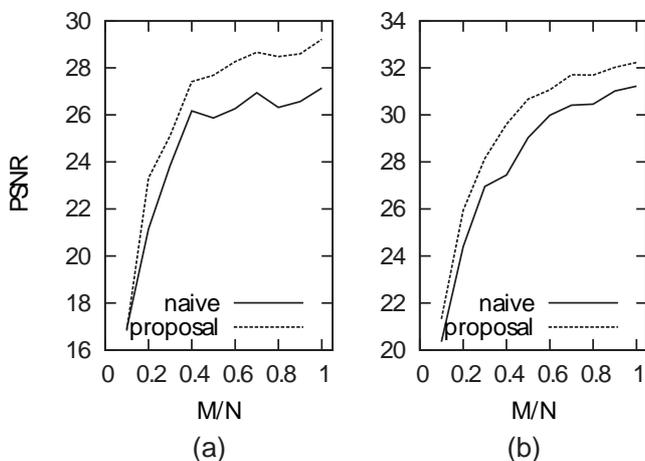


図 7 圧縮率と PSNR の関係：(a) レリーフ，(b) マウス

有効であることがわかる．なお，訓練データのサイズは，提案手法が 72.5MB，単純な手法が 2.32GB であった．以上のことから，提案手法の訓練データのサイズは単純な手法の約 1/33 であるにもかかわらず，定性的にも定量的にも提案手法が単純な手法よりも復元精度が高いことがわかる．

5. むすび

多波長・多方向光源下の画像を辞書学習に基づく圧縮センシングの枠組みで獲得する手法を提案した．特に，画素値がスペクトル項と幾何学項の積の和で表現されることに着目して，効率的に辞書を学習した．実画像実験を行い，提案手法は訓練データの量が大幅に少ないにもかかわらず，単純な手法よりも定性的・定量的に復元結果が優れていることを示した．鏡面反射成分や影の復元精度の向上は今後の課題である．

謝辞 本研究の一部は，JSPS 科研費（課題番号：25280057）の助成を受けた．

参考文献

- [1] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation”, *IEEE Trans. Signal Processing*, Vol.54, Issue 11, pp.4311–4322, 2006.
- [2] B. Ajdin and M. Finckh and C. Fuchs and J. Hanika and H. Lensch, “Compressive higher-order sparse and low-rank acquisition with a hyperspectral light stage”, Technical report, Eberhard Karls Universität Tübingen, WSI-2012-01, 2012.
- [3] D. Donoho, “Compressed sensing”, *IEEE Trans. Information Theory*, Vol.52, Issue 4, pp.1289–1306, 2006.
- [4] J. Gu and C. Liu, “Discriminative illumination: per-pixel classification of raw materials based on optimal projections of spectral BRDF”, In *Proc. IEEE CVPR2012*, pp.797–804, 2012.
- [5] M. Kitahara, T. Okabe, C. Fuchs, and H. Lensch, “Simultaneous estimation of spectral reflectance and normal from a small number of images”, In *Proc. VISAPP2015*, pp.303–313, 2015.
- [6] J. Parkkinen and J. Hallikainen and T. Jaaskelainen, “Characteristic spectra of Munsell colors”, *JOSA A*, Vol.9, No.2, pp.318–322, 1989.
- [7] S. Shafer, “Using color to separate reflection components”, *Color Research & Application*, Vol.10, Issue 4, pp.210–218, 1985.
- [8] G. Zhang, S. Jiao, X. Xu, and L. Wang, “Compressed sensing and reconstruction with Bernoulli matrices”, In *Proc. IEEE ICIA2010*, pp.455–460, 2010.
- [9] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net”, *Journal of the Royal Statistical Society, Series B*, pp.301–320, 2005.