

透明物体検出のための円偏光背景パターンと識別器の同時学習 Utilizing Polarized Patterns for Transparent Object Detection

平井 愛唯奈[†] 吉田 道隆[†] 岡部 孝弘[†]
Ayuna Hirai Mititaka Yoshida Takahiro Okabe

1 はじめに

ガラスやプラスチックなどの無色で透明な物体は身の周りに広く存在し、その検出はロボットピッキングや品質検査などの産業応用においても重要である。無色の透明物体は、背景の光を透過するのに対して、反射光をほとんど生じない。そのため、画像ベースの無色透明物体の検出は容易ではない。本稿では、透過光に基づく無色透明物体（以下、透明物体）の検出という挑戦的な課題に取り組む。

透過光に基づく透明物体の検出方法として、屈折による位置ずれを手掛かりとした手法がある [1]。一般に、透明物体を介して観察される背景は、屈折の影響を受けて歪んで見えることから、背景の歪み、つまり、透過光の位置ずれから透明物体の有無を識別できる。ところが、一様な背景に対してはもちろん、平面状やシート状の透明物体に対しても、屈折の影響が小さいことから、検出精度が低下してしまうという問題がある。

一方、透過光に基づく透明物体の検査方法として、複屈折による位相差を手掛かりとした手法がある [2]。光は横波であり、振動方向の偏った光は偏光と呼ばれる。一般に、ガラスやプラスチックに応力が働いていると、複屈折により位相差が生じ、透過光の偏光状態が変化する。したがって、偏光を用いて応力や内部歪みを検査することができる。また、複屈折が生じている透明物体については、透過光の偏光状態の変化が物体検出の手掛かりになり得る。

そこで本研究では、屈折による位置ずれと複屈折による位相差という相補的な手掛かりを活用したワンショットの透明物体検出法を提案する。提案手法では、円偏光パターンを背景としたときの透過光を入力として、機械学習ベースの物体検出を行う。このとき、円偏光背景パターンと識別器が互いに依存することから、これらを end-to-end に同時に学習することで、検出精度の向上を

目指す。また、ディスプレイ、円偏光板、偏光カメラからなるセットアップを用いた実画像実験を行い、円偏光を活用すること、および、背景パターンを学習することの有効性を示す。

本研究の主要な貢献は下記の3つである。(1) 屈折による位置ずれと複屈折による位相差という相補的な手掛かりを組み合わせた検出法を提案した。(2) 互いに依存する円偏光背景パターンと識別器を、end-to-end に同時に学習する枠組みを提案した。(3) 実画像実験を行い、提案手法の有効性を示した。

2 関連研究

2.1 透明物体検出

透明物体の検出にはさまざまな手法が提案されている。本節では、代表的な先行研究を紹介し、提案手法との違いについて述べる。

Maeno ら [1] は、透明物体による光の屈折に起因する背景の歪みに着目し、ライトフィールドカメラで取得した25視点の画像から、屈折による変位を特徴量とした物体分類を行っている。この手法は自然な背景に対しても適応可能であるが、屈折が微小な平面状・シート状の物体や一様な背景には適さない。また、Kalra ら [3] は非偏光光源下における反射光の偏光状態に基づく透明物体検出を提案している。この手法は、遮蔽輪郭線付近からの反射光が偏光していることを活用しているため、法線がカメラ方向を向いている領域に対しては、偏光度が小さいために検出精度が低下するという課題が残る。

光弾性法 [2] では、物体に力を加えることで生じる内部歪みによる複屈折現象を利用し、内部応力の状態を可視化・定量化する。この手法では、円偏光などを用いて透過光の位相差を測定し、主応力差を推定する。複屈折による位相差は、透明物体検出の手掛かりとなり得るが、内部歪みの無い透明物体の検出には適さない。

従来手法は、物体形状や背景、内部応力の有無に依存するため、検出可能な対象が限られていた。本研究では、屈折による位置ずれと複屈折による位相差という2

[†]岡山大学 (Okayama University)



図1: 計測システム画像と模式図. ディスプレイに表示した背景パターンを円偏光板で円偏光化し、透過した光を偏光カメラで撮像する.

つの相補的な物理的手掛かりを活用する. このアプローチにより, 従来手法では困難であった平面状や内部歪みの無い物体を含む, さまざまな透明物体の高精度な検出が期待できる.

2.2 同時最適化

近年, 光学イメージングモデルをパラメータ化した撮像層をアプリケーション層に接続して, 誤差逆伝播法を用いて end-to-end にモデルを学習し, センサと画像処理の同時最適化を行う手法が提案されている. Chakrabarti[4] はカラー画像の獲得のためにセンサのカラーフィルタと画像再構成手法の同時最適化を行っている. また, Wu ら [5] は深層学習を利用した深度推定のために, レンズに装着する位相マスクと画像再構成の同時最適化を行っている.

このような同時最適化の枠組みは, センサの設計に限らず照明環境の設計にも応用されつつある [6][7]. 本研究では, 透過光に関する重ね合わせの原理に基づいて, 背景パターンと識別器の同時最適化を提案する.

3 提案手法

本節では, 前節で述べた2つの手掛かりを捉えるための撮像システム, および, 円偏光パターンと識別器を同時に学習するネットワークを説明する.

3.1 撮像システム

本研究では, 位置ずれと位相差をワンショットで同時に捉えるため, 図1に示す撮像システムを構築する. 本システムでは, 背景にパターンを表示するディスプレイを設置し, その出力を円偏光板を通すことで円偏光化する. 円偏光化されたパターン光がシーンを通過した後, 偏光カメラにより偏光成分が撮像される. この構成により, 屈折と複屈折による情報を同時に取得できる.

光の屈折は, 物体の形状や屈折率の違いによって透過光の進行方向が変化する現象である. これにより, 透明物体越しに観察される背景パターンの像には, 歪みやズレが生じる. この変化は偏光状態に依存せず, 光の総強度

の空間分布として観測される.

光の偏光状態は4次元のストークスベクトルで表される [8]. ストークスベクトルの各成分 s_0, s_1, s_2, s_3 はそれぞれ, 光の総強度, 水平および垂直偏光成分の強度差, 45° 成分と 135° 成分の強度差, 右回り円偏光と左回り円偏光の強度差を表す.

光が異なる屈折率の媒質を透過する際の偏光状態の変化は, フレネルの式で表される. フレネルの式に基づいて, 透過に伴う偏光状態の変化を表すミューラー行列 M_T は次のように表される [8]

$$M_T = \frac{1}{2} \begin{bmatrix} T_{\perp} + T_{\parallel} & T_{\perp} - T_{\parallel} & 0 & 0 \\ T_{\perp} - T_{\parallel} & T_{\perp} + T_{\parallel} & 0 & 0 \\ 0 & 0 & 2\sqrt{T_{\perp}T_{\parallel}} \cos \delta & 2\sqrt{T_{\perp}T_{\parallel}} \sin \delta \\ 0 & 0 & -2\sqrt{T_{\perp}T_{\parallel}} \sin \delta & 2\sqrt{T_{\perp}T_{\parallel}} \cos \delta \end{bmatrix} \quad (1)$$

ここで, θ_i は入射角, θ_t は屈折角を表し,

$$T_{\perp} = \frac{2n_1 \cos \theta_i}{n_1 \cos \theta_i + n_2 \cos \theta_t}, \quad (2)$$

$$T_{\parallel} = \frac{2n_1 \cos \theta_i}{n_2 \cos \theta_i + n_1 \cos \theta_t} \quad (3)$$

である. また, δ は位相差, T_{\perp} および T_{\parallel} は振幅透過係数, n_1, n_2 は屈折率を表す.

右回り円偏光をストークスベクトル s_C で表すと

$$s_C = (1, 0, 0, 1)^T \quad (4)$$

のように表される. ストークスパラメータの変換の式 $S' = MS$ を用いると, 透明物体を透過後の右回り円偏光 s'_C は次のように変換される.

$$s'_C = \frac{1}{2} (T_{\perp} + T_{\parallel}, T_{\perp} - T_{\parallel}, 2\sqrt{T_{\perp}T_{\parallel}} \sin \delta, 2\sqrt{T_{\perp}T_{\parallel}} \cos \delta)^T \quad (5)$$

この変換式により, $T_{\perp} \neq T_{\parallel}$ のときは s_1 に, $\delta \neq 0$ のときは s_2 に変化を生じさせることが示される. これは複屈折による偏光状態の変化を捉えるために, 円偏光が有効であることを意味する.

また, 仮に平面状の透明物体を光が垂直に透過する場合, 透過後の右回り円偏光 s''_C は以下のように表される.

$$s''_C = \left(\frac{2n_1}{n_1+n_2}, 0, \frac{2n_1}{n_1+n_2} \sin \delta, \frac{2n_1}{n_1+n_2} \cos \delta \right)^T \quad (6)$$

この結果から, たとえ屈折の影響が小さい場合でも, $\delta \neq 0$ であれば s_2 に変化が生じ, 偏光状態の変化を通じて検出が可能となることがわかる.

これにより, 本システムではストークスベクトルの s_0, s_1, s_2 を活用することで, 屈折による位置ずれと複屈折による位相差という両方の手掛かりを偏光カメラでワンショットで観測することが可能となる.

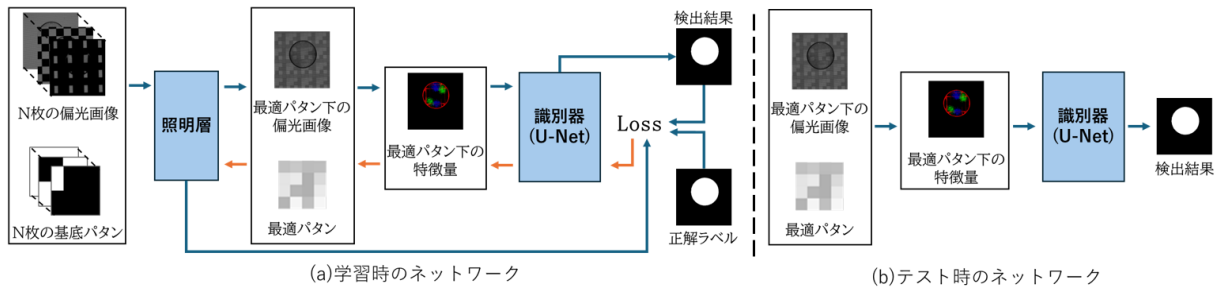


図 2: 照明層と識別器からなる提案ネットワーク。青の矢印が順伝播を、オレンジの矢印が逆伝播を示す。照明層への入力基底パタンのセットとそれらの基底パターンを表示して撮影した画像であり、出力は最適な背景パターンと最適な撮影条件下での画像である。識別器への入力照明層からの出力であり、出力は予測された検出結果である。学習時 (a) は、照明層と識別器は損失関数に基づいて end-to-end で同時に最適化される。テスト時 (b) は最適照明下で画像を撮影し、学習済みの識別器を用いて検出する。

3.2 同時学習

本稿では、円偏光パターンと識別器が相互に依存することから、これらを同時に学習するネットワークを提案する。図 2 に示すように、提案ネットワークは、円偏光パターンを設計する照明層（撮像層）と透明物体の有無を識別する識別器（アプリケーション層）により構成される。

提案手法では、円偏光パターンを N 種類の基底の線形結合で表現する。照明層は、ディスプレイに N 種類の基底パターンを表示して撮影した N 枚の撮影画像を入力とし、最適な円偏光背景パターン、および、それを表示して撮影した 1 枚の撮影画像を出力する。識別器は、この 1 枚の撮影画像を入力とし、透明物体の存在確率を画素ごとに出す。

学習時には、各シーンあたり N 枚の画像と正解ラベルを用いて、照明層と識別器を end-to-end に同時に学習する (図 2(a))。一方、テスト時は、最適化された円偏光パターンを表示して撮影された 1 枚の画像を識別器に入力して、透明物体を検出する (図 2(b))。

3.2.1 照明層

この層では、重ね合わせの原理に基づいて、最適な照明パターンの学習を行う。重ね合わせの原理から、複数のパターンを重ね合わせて表示して撮影したときの透過光画像は、個々のパターンを表示して撮影したときの透過光画像の重ね合わせで表現できる。この性質により、照明パターンと透過光画像を、同一の結合係数を持つ線形結合で表現できるため、照明条件の最適化を連続的かつ勾配ベースに行うことが可能となる。

照明パターンの設計においては、まず小領域に定義されたパターンを空間的に繰り返し配置することで、ディスプレイ全体を覆う周期的な構造を実現する。本研究では各小領域パターンを N 種類の Haar ウェーブレット基底パターン [9] (図 3 は $N = 16$ の例) の重ね合わせとして表現する。各基底パターンに対して得られた透過光の偏光画

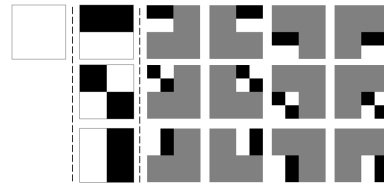


図 3: $N = 16$ の Haar 基底パターン。各パターンは空間的に局所的な輝度変化を持ち、白が +1, 黒が -1, 灰色は 0 を示す。この線形結合で 4×4 の全パターンが表現できる。

像を、重み w_i により線形結合することで、最適な照明条件下で撮影された画像を合成する。

3.2.2 識別器

識別器には、スキップ接続を持つ U-Net [10] をベースとしたセグメンテーションモデルを用いる。識別器には、透明物体を置いたときの偏光画像と置いていないときの偏光画像のストークスベクトルの成分 (s_0, s_1, s_2) の差分を入力する*1。学習には、透明物体を白、背景を黒としたマスク画像を教師データとして使用し、損失関数には Binary Cross Entropy と Dice Loss, Focal Loss の加重和に加え、照明の線形結合結果が負の明るさを持つ場合にペナルティを加える。

4 実験

4.1 実験環境

図 1 に示すディスプレイ、円偏光板、透明物体、および偏光カメラからなる計測システムで画像を取得し、物体ごとに切り分けた。データセットとして、実画像と正解画像の組を 81 組用意した。

正解画像の作成には、各シーンあたり 50 枚の画像を用いた。具体的には、透明物体あり/なしの各々のシーンについて、ディスプレイに表示するチェッカーパターンをシフトさせながら撮影した 25 枚の画像を撮影した。これらから得られる 25 組の差分画像を合成し、大津の

*1 透明物体無しの偏光画像は、すべてのシーンに共通のため、一度だけ撮影すればよい。そのため、提案手法はワンショットの検出法とみなせる。

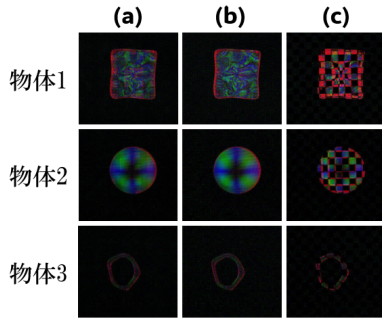


図4: $|s_0|$, $|s_1|$, $|s_2|$ の可視化. それぞれ R,G,B で示す. (a) 提案手法, (b) 背景白, (c) 背景チェッカー

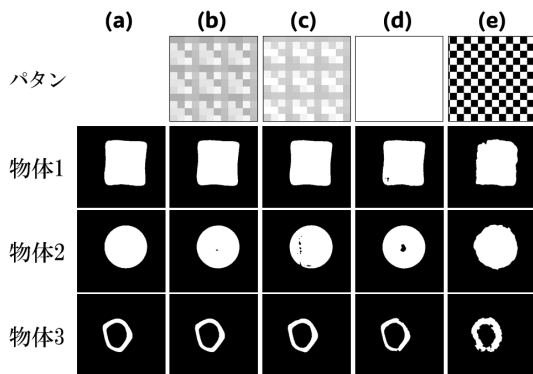


図5: 実験結果. テスト時に利用したパターンと検出結果を示す. (a) 正解ラベル, (b) 提案手法, (c) s_1, s_2 不使用, (d) 背景白, (e) 背景チェッカー

二値化 [11] と膨張・収縮処理を施し, 透明物体を白, 背景を黒とした正解画像を作成した.

4.2 評価と比較手法

定量的な評価指標として, 全画素の正答率 (Acc), 各クラスの IoU の平均 (mIoU), およびクラスごとの正答率の平均 (mAcc) を用いた. また評価は, データセットを 5 分割して交差検証を行い, その平均値を用いた. 比較手法は, s_1, s_2 を用いずに s_0 のみを入力として同時最適化する手法, 背景パターンを全面白にして識別器のみを学習する手法, 背景パターンをチェッカーパターンとして識別器のみを学習する手法の 3 つである.

4.3 実験結果

図4に, それぞれの手法について, ストークスペクトルの 3 成分の差分を, R, G, B として可視化した図を示す. s_2 が s_0 と異なる位置に表れていることや, 空間的パターンの利用により s_0 が新たに発生していることがわかる.

図5に, 学習により得られた最適な背景パターンと検出結果を示す. 図5(a) 正解ラベル (白が物体領域) を表し, (b) は提案手法, (c) は s_1, s_2 を不使用の場合, (d) は背景パターンが白の場合, (e) はチェッカーパターンの場合の結果である. 図5および, 表1の定量的結果よ

手法	all Acc	mean IoU	mean Acc
提案手法	99.82%	99.35%	99.69%
s_1, s_2 を不使用	99.65%	98.71%	99.65%
背景白	99.57%	98.44%	99.21%
背景チェッカー	98.53%	94.70%	97.55%

表1: 提案手法と比較手法の定量的比較. 全画素の正答率 (all Acc), 各クラスの IoU の平均 (mean IoU), 各クラスの正答率の平均 (mean Acc)

り, 提案手法が比較手法よりも優れていることが確認された.

5 むすび

本稿では, 円偏光とパターン照明を用いた透明物体検出手法を提案した. 照明パターンと識別器を end-to-end で同時に学習する枠組みを導入することで, 最適な照明条件を求めた点に特徴がある. また, プロトタイプシステムを用いた実験により, 偏光情報および背景パターンの最適化が検出精度の向上に寄与することを確認した. 今後は, 平面かつ内部歪みが少ない物体に対しても適応可能なモデル構築を目指す.

参考文献

- [1] K. Maeno, H. Nagahara, A. Shimada, and R. Taniguchi, "Light Field Distortion Feature for Transparent Object Recognition", In Proc. IEEE/CVF CVPR2013, pp.2786–2793, 2013.
- [2] K. Ramesh, "Digital Photoelasticity", Springer-Verlag, 2000.
- [3] A. Kalra, V. Taamazyan, S. Rao, K. Venkataraman, R. Raskar, and A. Kadambi, "Deep Polarization Cues for Transparent Object Segmentation", In Proc. IEEE/CVF CVPR2020, pp.8602–8611, 2020.
- [4] A. Chakrabarti, "Learning Sensor Multiplexing Design through Back-propagation", In Proc. NIPS2016, pp.3081–3089, 2016.
- [5] Y. Wu, V. Boominathan, H. Chan, A. Sanjarnarayanan, A. Veeraraghavan, "Phasecam3dlearning phase masks for passive single view depth estimation", In Proc. IEEE ICCP2019, pp.1–12, 2019.
- [6] T. Ueda, R. Kawahara, T. Okabe, "Learning Projection Patterns for Direct-Global Separation", In Proc. NIPS2016, Vol. 3, pp.599–606, 2024.
- [7] T. Hirao, R. Kawahara, T. Okabe, "Using Extended Light Sources for Relighting from a Small Number of Images", In Proc. NIPS2024, Vol. 3, pp.607–615, 2024.
- [8] E. Collett, "Field Guide to Polarization", SPIE Press, 2005.
- [9] A. Haar, "Zur Theorie der orthogonalen Funktionensysteme", Mathematische Annalen, Vol. 69, pp.331–371, 1910.
- [10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation", In Proc. MICCAI2015, pp.234–241, 2015.
- [11] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE Trans. Syst. Man Cybern, Vol.SMC-9, No.4, pp.62–66, 1979.