エッジパターン特徴とハイライト特徴を用いた机上透明物体認識

Transparent Object Recognition in Desktop Scenes Using Edge and Highlight Patterns

御堂地 央 (豊橋技科大) 正 三浦 純 (豊橋技科大)

Akira MIDOUCHI, Toyohashi Univ. of Tech. midouchi@aisl.cs.tut.ac.jp
Jun MIURA, Toyohashi Univ. of Tech.

This paper describes a method of recognition for transparent objects on the desk, using a monocular camera. First, we apply a machine learning-based method using edge patterns to the transparent desktop object recognition. Edge patterns which appear in the local area of the image are represented by a probabilistic topic model. Edge pattern features indicating presences of specific edge patterns, are used for training a Support Vector Machine (SVM) binary classifier, which discriminates between transparent objects and backgrounds. Then, we try to detect transparent objects in desk scenes using the classifier. Next, we define a highlight pattern feature describing highlight regions of objects. Highlight pattern features are represented by binarization of highlighted pixel ratio in local areas, and combined with edge pattern features to enhance the previous method. Finally, combined features are tested in cross-validation and object detection and shown to be superior to the previous edge pattern features.

Key Words: Vision, Computer vision, Transparent object recognition

1. はじめに

近年、介護福祉支援や一般生活の補助へのロボットの活用が考えられている。ロボットの家庭内導入を実現するためには、家庭内に数多く存在する透明物体を正しく認識する手法が必要である。

透明物体を認識する従来研究は主に2つのアプローチで行われている。まず、透明物体が持つ光学的特性を利用するアプローチが挙げられる[5][4]。これらは光学モデルを定義し、予め設定された条件下での透明物体が存在する場合の光学的变化(反射強度の変化や表面上でハイライト出現位置)を推定している。しかし、センサと認識対象の位置関係や照明条件を事前に設定する必要があり、多様なシーンでの認識は困難である。次に、エッジベースの認識アプローチがある[1][3]。これらは透明物体特有のエッジ特徴をモデル化して認識を行う。しかし、透明物体はその透明性によりカメラで撮影される際に弱いエッジしか検出されないため、これらのアプローチではエッジ情報を検出しやすくするため、物体の姿勢を限定して、認識を行っている。

そこで、本研究では背景知識を必要としないFritzらの機械学習によるアプローチ[3]をベースに、屋内環境で複数の姿勢を持つ机上の透明物体を認識する手法を開発する。まず、Fritzらの手法を用いて机上に複数配置された透明物体の認識を試みる。次に、認識精度を向上させるため、物体のハイライト領域を記述するハイライト特徴量を定義する。このハイライト特徴量とエッジパターン特徴量とを組み合わせ、机上物体認識を行う。

2. エッジパターン特徴による透明物体認識 [3]

2.1 エッジパターン特徴のモデル化

透明物体全体が映る部分を切り出した画像(図1参照)からエッジパターン特徴をモデル化する。

まず画像に対してエッジ勾配方向ヒストグラム(Histograms of Oriented Gradients:HOG)特徴量を記述する。本研究では画像のサ

イズを130 x 170 pixelに固定した。HOG特徴量記述の手順を以下に示す。

1. 画像 I を20 x 20 pixelのPatchと呼ばれる局所領域に分割する。隣り合うPatchはそのサイズに対する一定の割合 $o_p = 0.75$ で互いに重複し合う。
2. 全てのPatchに対して、
 - (a) Patchを5 x 5 pixelのCellと呼ばれる小領域で均等に16分割する。
 - (b) 全てのCellに対して、
 - bin数9のエッジ勾配方向ヒストグラムを作成する。

よって、Patchは $16 \times 9 = 144$ 次元のHOG特徴量ベクトルを持ち、画像サイズ、PatchのサイズおよびPatchが重複し合う割合から、画像は全体で $23 \times 31 = 713$ 個のPatchを持つ。

更に、確率的トピックモデル supervised LDA(sLDA) with classification [2] を用いてPatch中に出現するエッジパターンのモデル化を行う。具体的には、教師画像から得られたPatchのHOG特徴量ベクトルをサンプルとしてsLDA with classificationの学習を行い、 K 個のエッジパターン特徴 E_k およびエッジパターン特徴の重み η_k を得る ($k = 1, 2, \dots, K$)。

2.2 エッジパターン特徴量の記述

次に、画像中のエッジパターンの有無を示す2値変数のエッジパターン特徴量を記述する。エッジパターン特徴量記述の手順を以下に示す。

1. 画像 I をPatchに分割する。隣り合うPatchは $o_p = 0.75$ で互いに重複し合う。

Fig. 1 Local image examples

Fig. 2 Transparent object recognition using Edge Patterns

2. 全ての Patch に対して、

(a) 学習された sLDA with classification を用いて、Patch におけるエッジパターン特徴尤度 θ_k を推定する。

(b) エッジパターン特徴尤度 θ_k およびエッジパターン特徴の重み η_k から式 (1) の条件を満たす時、その Patch は対応するエッジパターン特徴を持つものとみなす。Patch は K 次元のエッジパターン特徴量ベクトルで表現される。

$$\theta_k = \epsilon / \eta_k \quad (1)$$

ϵ はエッジパターン特徴の有効化水準を示すしきい値である。 $k = 1, 2, \dots, K$ はエッジパターン番号である。 ϵ は [3] と同様に、教師画像の Patch が平均で 2 個のエッジパターン特徴を持つような値を設定する。

3. 画像 I に含まれる全ての Patch のエッジパターン特徴量ベクトルを連結し、画像のエッジパターン特徴量ベクトルとする。画像は $713 \times K$ 次元の特徴量ベクトルを持つ。

透明物体または背景のいずれかのクラスにラベル付けされた教師画像の特徴量ベクトルをサンプルとして、Support Vector Machine (SVM) の学習を行う。本研究では線形 SVM を利用する。

2.3 エッジパターン特徴を用いた透明物体検出

学習された分類器で教師データには含まれていない画像中に存在する透明物体の検出を行う。検出例を図 2 に示す。検出窓の色はその領域に対する信頼度の値が高くなるにつれ、紫から赤へ色相が変化するように設定されている。色が赤に近いほど、分類器はその領域が透明物体領域であると明確に判断している。

3. ハイライト特徴の併用

鏡面反射により滑らかな物体表面に生じるハイライト特徴は、物体の形状と姿勢およびシーンの照明条件に依存する。屋内環境で机上の透明物体が撮影された時、認識対象となる透明物体の形状が特有であるならば、その表面に生じるハイライト特徴の出現位置には他の物体とは異なる特徴的な傾向が生じると考える。そこで、物体のハイライト領域を記述するハイライト特徴量を定義し、従来のエッジパターン特徴と組み合わせ、透明物体の認識を行う。

3.1 ハイライト特徴抽出手法

ハイライト特徴を抽出するために、McHenry ら [6] のヒューリスティックを採用する。ここでは具体的なハイライト抽出手法の計算手順について述べる。適切なハイライト特徴を得るしきい値 T_{best} を決定するために、画像の最大輝度値 T_{max} から順にしきい値 T_n でしきい値処理を行い、ハイライト領域の周囲長 P_n を計算する。ハイライト領域の周囲長として、ハイライト領域の輪郭上のピクセル数を用いる。これまで得られたしきい値と周囲長の観測データ集合 $D = \{(T_n, P_n), \dots, (T_{max}, P_{max})\}$ を最小二乗法

Algorithm 1 Calculate Best Intensity Threshold

Require: $eps \geq 0$

```

1:  $T_n \leftarrow T_{max}$ 
2:  $P_{max} \leftarrow \text{calculate\_perimeter}(T_{max})$ 
3: while  $T_n \geq 0$  and  $err < eps$  do
4:    $T_n \leftarrow T_{n-1}$ 
5:    $P_n \leftarrow \text{calculate\_perimeter}(T_n)$ 
6:    $D \leftarrow \{(T_n, P_n), \dots, (T_{max}, P_{max})\}$ 
7:    $err \leftarrow \text{linear\_least\_squares}(D)$ 
8: end while
9: return  $T_{best} \leftarrow T_{n+1}$ 

```

(a) Input Image (b) Highlights (c) Patch Highlight Patterns (d) Cell Highlight Patterns

Fig. 3 Examples of Patch/Cell Highlight Pattern Extraction

でしきい値 T を変数とする周囲長 P の 1 次関数 $P = aT + b$ へ当てはめ、その最小二乗誤差 err が許容誤差 eps 以上になるまで、 T_n を下げていく。許容誤差を越える直前のしきい値 T_{n+1} を最適なしきい値 T_{best} として採用する。アルゴリズム 1 に疑似コードを示す。

3.2 ハイライト特徴量

次に抽出されたハイライト特徴を用いて、物体のハイライト領域を示す 2 値変数のハイライト特徴量を記述する。また、透明物体の机上位置や照明条件が変化すると、画像上に出現するハイライト領域の位置も変化する。ハイライト領域の位置変化を許容できる最適な特徴量を模索するため、局所領域サイズに Cell, Patch の 2 種類を用意する。Cell ハイライト特徴量記述の手順を以下に示す。

1. 画像 I を局所領域 Patch または Cell に分割する。隣り合う Patch は $o_p = 0.75$ で互いに重複し合う。
2. 全ての Patch または Cell に対して、
 - (a) 局所領域のハイライトピクセル数を求める。
 - (b) 局所領域全体のピクセル数に対するハイライトピクセル数の割合が一定のしきい値を越えるならば、その局所領域をハイライト領域とみなす。
3. 画像 I に含まれる全ての局所領域のハイライト特徴量をまとめ、画像のハイライト特徴量ベクトルとする。

よって、1 つの画像は Patch または Cell ハイライト特徴量を記述する場合、713 または $713 \times 16 = 11408$ 次元のハイライト特徴量ベクトルを持つ。

現実画像のハイライト抽出と Cell/Patch ハイライト特徴量記述の一例を図 3 に示す。

最後に、得られたハイライト特徴量ベクトルを従来のエッジパターン特徴量ベクトルに連結する。連結した特徴量ベクトルを

Fig. 4 Transparent object Detection with Cell Highlight Patterns

Fig. 5 Transparent object Detection with Patch Highlight Patterns

SVM へのサンプルとする。

3.3 ハイライト特徴を併用した透明物体検出

エッジパターン特徴量とハイライト特徴量を組み合わせて学習を行った分類器で、教師データには含まれていない画像中に存在する透明物体の検出を試みる。検出例を図 4, 図 5 に示す。入力したテスト画像は図 2 と同様のものである。

4. 実験

4.1 画像データセット

学習に用いる画像データセットの作成方法について述べる。本研究のデータセットは屋内で机上のセットアップを行い、机上シーンを撮影する。今回は 2 種類の透明物体（ワイングラスおよびタンブラー）を用意し、データセットは透明物体毎に作成する。カメラで撮影されたシーン画像から透明物体を含む画像を切り出し、正のクラスラベルを付与する。同様に、透明物体が含まれない画像も切り出し、負のクラスラベルを付与する。切り出す画像サイズは全て一定である。作成した教師画像の例を図 1 に示す。最終的に各透明物体につき、計 960 枚（正:負=240:720）ずつの教師画像データセットを作成した。

4.2 認識性能評価法

評価手法として、10-fold cross validation を採用する。データセットの分割は各透明物体のデータセットにつき 1 度のみ行い、各手法は全て同じ訓練/テストデータで検証される。線形 SVM による分類で得られるテストデータの出力値をしきい値 $\theta = 0, 0.1, \dots, 1.0$ で 2 値化し、テストデータを 2 クラスへ分類する。

認識性能の評価には、以下の評価値を用いる。

- Recognition Precision $p_{\theta}^r = TP_{\theta}^r / (TP_{\theta}^r + FP_{\theta}^r)$

- Recognition Recall $r_{\theta}^r = TP_{\theta}^r / T_{all}^r$

$TP_{\theta}^r, TN_{\theta}^r, FP_{\theta}^r$ はそれぞれ、テストデータのラベルとしきい値 θ 時の分類結果とを照合した際の True Positive, True Negative, False Positive の数を示す。 T_{all}^r は教師データセット全体で正のクラスラベルを持つデータ数を表す。

4.3 評価結果

従来手法およびハイライト特徴量併用後の性能を比較するため、以下 3 つの手法で SVM を学習する。

Table 1 Settings of highlight pixel ratio threshold

設定名	ハイライトピクセル割合しきい値
CH1	0.5
CH2	0.25
CH3	0.125
PH1	0.03125
PH2	0.015625
PH3	0.0078125

Fig. 6 Wineglass recognition result

Fig. 7 Tumbler recognition result

Edge Pattern エッジパターン特徴量のみ使用

Cell Highlight Pattern(CH) Cell ハイライト特徴量を併用

Patch Highlight Pattern(PH) Patch ハイライト特徴量を併用

特徴量記述に必要なパラメータの内、以下のものは全ての設定において固定とする。

- 確率的トピックモデルのモデルパラメータ
 - ハイパーパラメータ $\alpha = 2$
 - トピック数 $K = 10$
- ハイライト特徴量記述処理のパラメータ
 - 許容誤差 $eps = 1000$

PH, CH については最適なハイライトピクセル割合しきい値を模索するため、表 1 に示す複数のパラメータ設定で学習を行う。

各設定の $\theta = 0, 0.1, \dots, 1.0$ における Precision および Recall を Precision/Recall 曲線としてプロットし、図 6, 図 7 に示す。

認識性能評価に関する考察を述べる。まず図 6, 図 7 の全ての設定で高い精度が出ている。この結果から、今回作成したデータセットがエッジパターン特徴量のみで十分な認識が可能なのであると言える。

また、図 6 から、CH1 以外は Edge Pattern に比べより高い性能を示している。この結果は CH1 がその他に比べて、ハイライトピクセル割合しきい値が大きいため、Cell に出現する透明物体のハイライト特徴がそのしきい値を越えられず、特徴量として有効に働かなかったことが原因と考えられる。よって、小さいハイライ

トピック割合しきい値を設定した Cell/Patch ハイライト特徴量の併用により、認識性能を更に改善できる。

4.4 検出評価用の画像テストセット

検出性能の評価を行うために、教師画像データセット作成時とは別に撮影された 16 枚の机上シーン画像を用意した。テストセット全体で計 48 個の透明物体が含まれている。加えて、各テスト画像に真の透明物体領域を設定し、検出の正答とする。真の透明物体領域の大きさは教師画像データのサイズと同じである。

4.5 検出性能評価法

透明物体の検出はテスト画像に対して、検出窓を走査して行う。検出窓の大きさは教師画像データのサイズと同じである。線形 SVM により得られる検出窓領域の出力値をしきい値 $\theta = 0, 0.1, \dots, 1.0$ で 2 値化し、検出窓領域を 2 クラスへ分類する。

各テスト画像へ事前に設定された真の透明物体領域 B_{gt} と SVM によって透明物体と分類された検出窓の領域 B_p との重なり率 α_o 式 (2) が、一定の重なり率しきい値 α_{th} を越える場合、正しい検出とみなす。 α_{th} は検出時にユーザが設定する。

$$\alpha_o = \frac{\text{area}(B_p \cap B_{gt})}{\text{area}(B_p)} \quad (2)$$

検出性能の評価には、以下の評価値を用いる。

- Detection Precision $p_\theta^d = TP_\theta^d / (TP_\theta^d + FP_\theta^d)$
- Detection Recall $r_\theta^d = TP_\theta^d / gt_{all}^d$

$TP_\theta^d, TN_\theta^d, FP_\theta^d$ はそれぞれ、しきい値 θ 時の検出窓の正誤判定により得られる True Positive, True Negative, False Positive の数を示す。 gt_{all}^d はテストセット全体の真の透明物体領域の総数である。

4.6 評価結果

4.3 節から透明物体毎に各手法の中で最も Precision Mean の高い設定を一つずつ選び、検出に用いる。具体的には、以下に示す設定を選んだ。

- ワイングラス: Edge Pattern, CH3, PH3
- タンブラー: Edge Pattern, CH3, PH1

また、同じデータセットに対して Bag of Words モデルによる特徴量で学習を行わせた SVM の検出結果も併記する。Bag of Words の単語には SIFT 特徴量を採用し、k-means 法によるクラスタリングで語彙を作成した。

検出窓の正誤判定に用いる重なり率しきい値は $\alpha_{th} = 0.75$ とした。

各設定の $\theta = 0, 0.1, \dots, 1.0$ における Precision および Recall を Precision/Recall 曲線としてプロットし、図 8, 図 9 に示す。

検出性能評価について考察する。まず、図 8, 図 9 から、Bag of Words モデルに比べ、エッジパターン特徴量を用いた場合の方が高い性能を示している。エッジパターン特徴量は局所領域 Patch 毎に記述するため、透明物体の形状に応じて各局所領域に出現しやすい特定のエッジパターンが存在することを考慮しており、更に確率的トピックモデルによって表現されるエッジパターンは局所領域に出現する共通のパターンを表現しているため、背景の変化にロバストである。そのため、背景が変化した場合でもより高い検出性能を示すことができる。

次に図 8 を見ると、PH は Edge Pattern に比べ、Precision/Recall 共に向上している。またワイングラスの検出において、CH は Edge

Fig. 8 Wineglass detection result

Fig. 9 Tumbler detection result

Pattern に比べ、Precision が低下する代わりに Recall が上昇している。これは Cell が Patch より小さな局所領域であるため、姿勢変化や照明条件変化に伴うハイライト領域の位置変化に弱いことが原因であると考えられる。

一方で図 9 を見ると、タンブラー検出において CH および PH は、Edge Pattern に比べて性能が低下する。これはタンブラーのハイライト領域パターンが特徴的ではないために、かえって透明物体と背景との分類を困難にしまったと考えられる。よって、透明物体の形状に強く依存するハイライト特徴の利用は、透明物体の種類に応じて決める必要がある。

5. おわりに

本研究では、確率的トピックモデル学習に基づくエッジベースの透明物体認識手法を、複数の姿勢を持つ机上の透明物体認識に適用した。さらに、物体のハイライト領域を記述するハイライト特徴量を定義し、従来の特徴量と組み合わせることにより、認識性能が向上することを示した。

今後の課題として、様々なスケール、姿勢での透明物体認識のための改良や、ロボットによる操作のための位置推定手法の開発などがある。

文 献

- [1] C. Choi and H.I. Christensen. 3d textureless object detection and tracking: An edge-based approach. In *IROS-2012*, pp. 3877–3884.
- [2] C. Wang et al. Simultaneous image classification and annotation. In *CVPR-2009*, pp. 1903–1910.
- [3] M. Fritz et al. An additive latent feature model for transparent object recognition. In *NIPS-2009*, pp. 558–566.
- [4] M. Osadchy et al. Using specularities for recognition. In *ICCV-2003*, pp. 1512–1519.
- [5] Z. et al. Lei. Transparent object detection using color image and laser reflectance image for mobile manipulator. In *ROBIO-2011*, pp. 1–7.
- [6] K. McHenry. Finding glass. In *CVPR-2005*, Vol. 2, pp. 973–979.