

混合粒子フィルタに基づく 視覚的共通パターン検出手法

青木 工太^{†1} 長橋 宏^{†1}

本稿では、1組の画像に共通して存在する視覚的なパターンを検出するための手法を提案する。粒子フィルタの枠組みを追跡器としてではなく、検出器として用いる。複数存在し得る共通パターンを検出するために、提案手法は混合粒子フィルタの枠組みを応用する。また、提案手法は与えられた1組の画像から得られる情報だけを用い、事前学習を必要としない。検出器は矩形領域の集合で表現され、対応する部分画像から特徴量を抽出する。特徴量間の距離に基づいて、各画像上の矩形領域間の類似度を計算する。粒子フィルタの枠組みに従って、共通パターンを覆うように矩形領域の中心座標が更新される。テスト画像に対する実験により、提案手法の有効性を検証する。

Common Visual Pattern Detection Based on Mixture Particle Filtering

KOTA AOKI ^{†1} and HIROSHI NAGAHASHI^{†1}

We present a novel approach to detect common visual patterns between image pairs. We use a particle filtering approach as a detector of possibly multiple similar visual patterns rather than as a visual tracker. To handle multiple instances, we adopt an algorithm of the mixture particle filter. Our method doesn't require a learning phase in advance but leverages only the information from given two images. A detector is a set of bounding boxes. A feature vector will be extracted from the image area covered by a bounding box and the similarity between two boxes can be calculated based on the distance among their corresponding features. The centers of bounding boxes should be updated to move toward the locations of common patterns in turn. In the experiments, we demonstrate the validity of our method on common visual pattern detection.



(a) (b)
図1 共通パターンを含む1組の画像の例 (Toys).

Fig. 1 Example of a pair of images containing common visual patterns (Toys).

1. はじめに

与えられた2枚以上の画像から視覚的な対応関係を見つけることは、コンピュータビジョンにおける基本的な課題の1つであり、例えば、画像分類¹⁾、構造が類似するシーン間の位置合わせ²⁾、映像の領域分割³⁾、オブジェクトの追跡⁴⁾など、多くの応用技術にとって必須の処理である。効果的な画像特徴量を記述する手法が提案され、同じシーンを複数の視点から撮影した画像や、構造が類似する複数の画像について、頑健な対応付けが可能になってきた⁵⁾⁻⁷⁾。一方で、図1に示した1組の画像のように、いくつかの共通する視覚的パターンあるいは、オブジェクトがそれぞれの画像内で異なる位置や姿勢で現れる場合、曖昧な特徴点あるいは、乱雑な背景の影響によって、対応関係を見つけることは容易でない⁸⁾。

本稿では、与えられた2枚の画像に共通して現れる視覚的なパターンを検出するための手法を提案する。提案手法の基本的なアルゴリズムは混合粒子フィルタ⁹⁾の枠組みに従うが、(複数オブジェクトの)視覚的な追跡手法としてではなく、共通パターンの検出手法として用いる。画像に現れる共通パターンに関する事前知識あるいは、事前に学習されたモデルを必要とせず、検出は入力画像から得られる特徴量のみに基づく。また、提案手法はそれぞれの画像内にいくつの共通パターンが存在するかを明示的に扱わず、複数のパターンを同時に検出する。

^{†1} 東京工業大学
Tokyo Institute of Technology

1.1 概要

提案手法では、粒子フィルタの枠組みにおける標本を矩形領域によって表現し、それぞれの画像上に多数配置する。矩形領域に対応する部分画像から特徴量を抽出し、特徴量間の距離に基づいて矩形領域間の類似度を計算する。標本の重みは類似度によって定義される。矩形領域の中心座標を標本に対する状態遷移確率分布によって更新すると、共通パターンを包含する可能性が高い標本ほど重みが大きくなる。

以下では、提案手法について概説する。与えられた1組の画像をそれぞれ \mathcal{I}, \mathcal{J} とし、画像上に配置される矩形領域を $R = (c_x, c_y, l_x, l_y)$ と表す。標本の状態変数は領域の中心座標によってベクトル $\mathbf{x} = [c_x, c_y]^T$ と表現する。矩形領域の幅 l_x と高さ l_y は固定とする。

追跡問題などに適用される標準的な粒子フィルタとは異なり、提案手法では、アルゴリズムの反復が時系列に相当する。反復ステップ t における標本集合は $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^M$ と表現される。ただし、 w は標本の重みを表し、 M は標本の総数である。また、反復回数を T とする。提案手法のアルゴリズムは以下ようになる。

- 1: **for** $t = 1$ **to** T **do**
- 2: **if** $t = 1$ **then**
- 3: それぞれの画像上において、標本の状態 \mathbf{x}_t を初期化する(3.1節)。
- 4: **else**
- 5: 遷移確率分布に従って、標本の状態 \mathbf{x}_t を更新する(3.2節)。
- 6: **end if**
- 7: それぞれの画像上において、標本集合をクラスタリングする(3.3節)。
- 8: 矩形領域に対応する部分画像から、特徴量を計算する(3.4節)。
- 9: 画像 \mathcal{I} (\mathcal{J})上の標本を参照して、画像 \mathcal{J} (\mathcal{I})上の標本の重み w_t を更新する(3.5節)。
- 10: **end for**

以下、2節では混合粒子フィルタ⁹⁾の枠組みを概説し、3節で共通パターン検出への適用および、提案手法の詳細について述べる。3.4節では、矩形領域間の類似度計算に用いる画像特徴量について述べる。4節では、テスト画像⁸⁾に対する実験結果から本手法を定量的に評価し、5節で結論を述べる。

1.2 関連研究

視覚的に共通するパターンを複数の画像から検出するという課題に対して関心が高まっている^{8),10)-12)}。Tan & Ngo¹⁰⁾は色の類似度によって領域分割された画像からグラフを構

築し、Earth Mover's Distanceに基づいて画像間に対応付けの手法を提案した。Yuan & Wu¹¹⁾は画像をランダムに分割し、部分画像間に対応付けによって、共通パターンが存在する確信度を表す投票マップの生成手法を提案した。Cho et al.⁸⁾は、共通パターンを構成する部分領域の拡大・縮小と分割・統合を反復するデータ駆動型のマルコフ連鎖モンテカルロ法を提案した。Liu & Yan¹²⁾は特徴点間の局所的な類似度および、特徴点の空間的な配置における類似度を考慮したグラフを構築し、強く結びついた部分グラフを探索することによって共通パターンを検出する手法を提案した。

類似する研究課題として、複数の画像に共通する領域を分割して切り出すという問題^{13),14)}(cosegmentation)あるいは、単一の画像から複数の同一オブジェクトを検出する問題¹⁵⁾と関連が深い。また、逐次モンテカルロ法を適用して、超音波画像から複数のオブジェクトを1つずつ検出する手法が提案された¹⁶⁾。

2. 混合粒子フィルタ

時間の経過とともに変化する事後確率分布を近似的に推定する枠組みとして、粒子フィルタあるいは、より一般的には、逐次モンテカルロ法¹⁷⁾が広く用いられており、コンピュータビジョンにおいては特に、オブジェクト(対象物)の追跡問題¹⁸⁾に適用されることが多い。事後確率分布 $p(\mathbf{x}_t|\mathbf{y}^t)$ はそれまでの観測すべて $\mathbf{y}^t = (\mathbf{y}_1, \dots, \mathbf{y}_t)$ について条件付けされたオブジェクトの状態変数 \mathbf{x}_t を記述し、再帰的なベイズ推定では、オブジェクトの状態がどのように遷移するかを表現する動的モデル $D(\mathbf{x}_t|\mathbf{x}_{t-1})$ および、観測データの下での状態(仮説)の尤度 $L(\mathbf{y}_t|\mathbf{x}_t)$ によって定義される。このとき、事後確率分布は次の2つの式によって計算される。

$$p(\mathbf{x}_t|\mathbf{y}^{t-1}) = \int D(\mathbf{x}_t|\mathbf{x}_{t-1})p(d\mathbf{x}_{t-1}|\mathbf{y}^{t-1}) \quad (1)$$

$$p(\mathbf{x}_t|\mathbf{y}^t) = \frac{L(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}^{t-1})}{\int L(\mathbf{y}_t|\mathbf{s}_t)p(d\mathbf{s}_t|\mathbf{y}^{t-1})} \quad (2)$$

一般的に、尤度の定義によって、事後確率分布の解析的な推定を実行することはできないため、粒子フィルタに基づく近似手法が適用される。すなわち、事後分布は標本(粒子)の集合および、対応する重みによって表現され、それらを確率的に伝搬することによって、各時刻の事後分布が近似表現される。

2.1 混合モデル

粒子フィルタの問題点として、確率分布における複数の最頻値を維持する能力が低いと

ということが挙げられる。複数のオブジェクトが存在する場合、または観測が曖昧であったり、雑音が含まれている場合など、事後確率分布が多峰性の分布となることは少なくない。Vermaak et al.⁹⁾ は次式のように、事後分布を明示的に混合モデルとして表現する枠組みを提案した。

$$p(\mathbf{x}_t|\mathbf{y}^t) = \sum_{k=1}^K \pi_{k,t} p_k(\mathbf{x}_t|\mathbf{y}^t) \quad (3)$$

ここで、 K は混合成分の数、 $\pi_{k,t}$ は成分分布 $p_k(\mathbf{x}_t|\mathbf{y}^t)$ に対する混合係数であり、 $\sum_{k=1}^K \pi_{k,t} = 1$ である。混合モデルを用いると、事後分布は標本集合によって次式のように近似される。

$$\bar{p}(\mathbf{x}_t|\mathbf{y}^t) = \sum_{k=1}^K \pi_{k,t} \sum_{i \in \mathcal{L}_k} w_t^{(i)} \delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t) \quad (4)$$

ここで、 $\mathcal{L}_k = \{i \in \{1, \dots, M\} | z_t^{(i)} = k\}$ は k 番目の混合成分に属している標本の識別番号の集合であり、 i 番目の標本が k 番目の混合成分に属す場合に $z_t^{(i)} = k$ と表すものとする。また、 $\delta_a(\cdot)$ は a に質量を持つディラック測度である。各混合成分 $k \in \{1, \dots, K\}$ について、標本の重みの和は $\sum_{i \in \mathcal{L}_k} w_t^{(i)} = 1$ である。

時刻 t における標本の集合を $\mathcal{P}_t = \{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^M$ と表す。事後分布 $p(\mathbf{x}_{t-1}|\mathbf{y}^{t-1})$ に従って近似的に分布する標本集合 \mathcal{P}_{t-1} が与えられたとき、以下のように、事後分布 $p(\mathbf{x}_t|\mathbf{y}^t)$ からの近似的な標本として \mathcal{P}_t が生成される。 k 番目の混合成分について考えると、標本集合 $\{\mathbf{x}_{t-1}^{(i)}, w_{t-1}^{(i)}\}_{i \in \mathcal{L}_k}$ は成分分布 $p_k(\mathbf{x}_{t-1}|\mathbf{y}^{t-1})$ から近似的に生成された標本であるとみなすことができ、各混合成分について独立に新しい標本を生成できる。すなわち、適切に選択された提案分布 $q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)}, \mathbf{y}_t)$ ($i \in \mathcal{L}_k$) から、新しい標本 $\mathbf{x}_t^{(i)}$ を生成するとき、対応する重みは次式によって計算される。

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j \in \mathcal{L}_k} \tilde{w}_t^{(j)}}, \quad \tilde{w}_t^{(i)} = \frac{w_{t-1}^{(i)} L(\mathbf{y}_t|\mathbf{x}_t^{(i)}) D(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)}, \mathbf{y}_t)} \quad (5)$$

このとき、新しい標本集合 $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i \in \mathcal{L}_k}$ は成分分布 $p_k(\mathbf{x}_t|\mathbf{y}^t)$ に従って近似的に分布する。また、混合係数は以下のように近似的に計算される。

$$\pi_{k,t} = \frac{\pi_{k,t-1} \tilde{w}_{k,t}}{\sum_{h=1}^K \pi_{h,t-1} \tilde{w}_{h,t}}, \quad \tilde{w}_{k,t} = \sum_{i \in \mathcal{L}_k} \tilde{w}_t^{(i)}. \quad (6)$$

2.2 混合成分の更新

理想的には、事後確率分布における1つの最頻値に対して、1つの混合成分が対応することが望ましい。すなわち、1つの混合成分について、共通パターン検出においては1つのパターンが対応し、追跡問題においては1つのオブジェクトが対応することが望ましい。実際

には、パターンやオブジェクトがいくつ存在するかが事前に分かっていることは稀である。

後述するように提案手法では、アルゴリズムの反復ステップごとに混合成分の数の変更 $K' \leftarrow K$ および、標本の再クラスタリングによる混合成分の構成の変更 $z_t'^{(i)} \leftarrow z_t^{(i)}$ を行う。このような処理によって混合モデルが変更される時、新しい混合係数 $\{\pi_{k,t}'\}_{k=1}^{K'}$ と標本の重み $\{w_t'^{(i)}\}_{i=1}^M$ は次のように更新される⁹⁾。

$$\pi_{k,t}' = \sum_{i \in \mathcal{L}_k'} \pi_{z_t'^{(i)},t} w_t^{(i)}, \quad w_t'^{(i)} = \frac{\pi_{z_t'^{(i)},t} w_t^{(i)}}{\pi_{z_t'^{(i)},t}} \quad (7)$$

ただし、この処理によって標本の状態変数 $\{\mathbf{x}_t^{(i)}\}_{i=1}^M$ が変わることはないことに注意する。

3. 共通パターン検出

本節では、共通パターン検出に混合粒子フィルタの枠組みを適用する方法について述べる。前述のように、標本の状態変数は対応する矩形領域の中心座標で表され、標本の集合は与えられたそれぞれの画像上に分布する。

3.1 初期化

与えられた画像の幅と高さをそれぞれ L_x, L_y とすると、幅と高さが l_x, l_y である矩形領域の中心座標 (c_x, c_y) は $l_x \leq c_x \leq L_x - l_x, l_y \leq c_y \leq L_y - l_y$ の範囲で一様分布に従って初期化される。本手法では矩形領域の大きさは固定とし、後述するすべての実験において $l_x = l_y = 51$ と設定する。

3.2 状態遷移

局所解に陥ることを回避するために、自己組織化画像¹⁹⁾などの競合学習法あるいは、擬似焼きなまし法²⁰⁾から着想を得た動的モデルを用いる。すなわち、次式のような単純な焼きなまし計画を導入した動的モデルによって標本の状態遷移を表す。

$$c_{x,t} \sim \mathcal{N}(c_{x,t-1}, \sigma_t^2), \quad c_{y,t} \sim \mathcal{N}(c_{y,t-1}, \sigma_t^2) \quad (8)$$

$$\sigma_t = \sigma_0 / (1 + \kappa(t/T)) \quad (9)$$

ここで、 $\mathcal{N}(\mu, \sigma^2)$ は平均 μ 、分散 σ^2 の正規分布である。提案手法では上式の動的モデルによって提案分布を定義するため、式(5)は次のようになる。

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j \in \mathcal{L}_k} \tilde{w}_t^{(j)}}, \quad \tilde{w}_t^{(i)} = w_{t-1}^{(i)} L(\mathbf{y}_t|\mathbf{x}_t^{(i)}) \quad (10)$$

実験では $\sigma_0 = 10$ および、 $\kappa = 3$ と設定する。このとき、 σ_t は図2に示すように変化する。標準偏差 σ_t に応じて位置座標の変化量が制御されることによって、反復処理の初期では標本の探索範囲は比較的大きく、後半になるにつれてその範囲は狭まり、収束する。

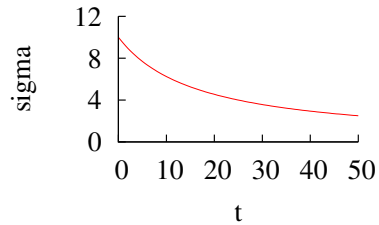


図 2 動的モデルの焼きなまし計画.

Fig. 2 Annealing schedule for dynamic model.

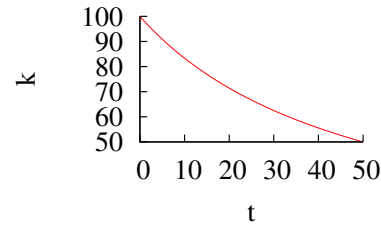


図 3 クラスタ数の焼きなまし計画.

Fig. 3 Annealing schedule for the number of clusters.

3.3 クラスタリング

標本集合は k 平均法を用いて、対応する矩形領域の中心座標についてクラスタリングする。標本は状態遷移によって画像上を移動するため、クラスタリングによってできるポロノイ図は時々刻々変化する。このときできるポロノイ図は画像特徴量を考慮していないため、画像の意味的な分割とは必ずしも対応しない。クラスタリングされた標本によって混合成分を構成し、混合モデルとして事後確率分布を近似することで分布の多峰性を維持することが目的であり、したがって、混合成分の数 K は (十分に大きければ) 1 組の画像に共通するパターンあるいは、領域の数と一致しなくてもよい。

本手法では、各ステップ t におけるクラスタ数 K_t を次式により決定する。

$$K_t = \lfloor \sigma_0 / (1 + \kappa(t/T)) \rfloor \quad (11)$$

これは動的モデルの式 (9) と同様の考え方によって定義されている。ただし、ここでのパラメータは $\sigma_0 = 100$ および、 $\kappa = 1$ と設定する (図 3)。クラスタの数を反復とともに減らしていくことにより、共通パターンである可能性を示す最頻値は徐々に絞り込まれる。

共通パターンの中でも特に顕著なものだけが検出されるという問題に対処するため、混合係数の値に上限 τ を設ける。これにより、矩形領域の類似度が極めて高いパターンだけではなく、一定の基準を超える類似パターンを検出できるようになる。

3.4 画像特徴量

画像特徴量として、SIFT⁷⁾ に基づく特徴量および、彩度によって重み付けられた色相ヒストグラム²¹⁾ を用いる。それぞれの特徴量を画素単位で密に計算し、標本に対応する矩形領域について累積する。この特徴量抽出は積分画像^{22),23)} をあらかじめ計算しておくことで

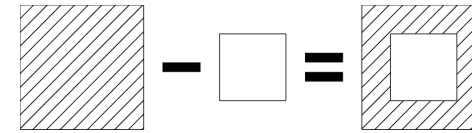


図 4 積分画像による特徴量の抽出.

Fig. 4 Feature extraction by using integral image.

効率的に実行できる。SIFT に基づく特徴量は 128 次元ベクトルであり、色相ヒストグラムは 24 ビンからなるものとする。ベクトルおよび、ヒストグラムの規格化は画素単位ではなく、矩形領域について行う。SIFT に基づく特徴量は文献 7) に従い、単位長ベクトルに規格化した後、しきい値処理をして再び規格化する。また、色相ヒストグラムは和が 1 になるように規格化する。

積分画像を用いて、矩形領域内で中心部分を重視した特徴量を抽出する²⁴⁾。図 4 に示すように、矩形領域 (l_x, l_y) 内で枠の部分 (b_x, b_y) に対応する特徴量は、矩形領域とその中心部分 $(l_x - 2b_x, l_y - 2b_y)$ のそれぞれに対応する特徴量の差分として計算されるため、係数を掛けて再び加算することで、中心部分と枠の部分の比重を制御できる。この処理を中心部分に向かって段階的に実行すれば、平滑化フィルタと同様の効果を得ることもできる。実験では、枠の大きさを $b_x = b_y = 10$ とし、枠の部分に対応する特徴量に対して係数 $\beta = 0.5$ を掛けて中心部分の特徴量に加算する。

3.5 尤度

一方の画像上における標本の尤度を計算するときには、他方の画像上における標本を参照する。ここでは、重みが更新される標本の状態変数を $\{\mathbf{x}^{(i)}\}_{i=1}^M$ とし、参照される標本を観測 $\{\mathbf{y}^{(j)}\}_{j=1}^M$ として説明する (図 5)。

標本 $\mathbf{x}^{(i)}$ に対応する矩形領域から抽出される SIFT に基づく特徴量および、色相ヒストグラムをそれぞれ $\mathbf{f}_1(\mathbf{x}^{(i)})$, $\mathbf{f}_2(\mathbf{x}^{(i)})$ とし、観測 $\mathbf{y}^{(j)}$ に対する特徴量との距離 $d_{i,j}$ を次式のように定義する。

$$d_{i,j} = (1 - \alpha)d_1(\mathbf{f}_1(\mathbf{x}^{(i)}), \mathbf{f}_1(\mathbf{y}^{(j)})) + \alpha d_2(\mathbf{f}_2(\mathbf{x}^{(i)}), \mathbf{f}_2(\mathbf{y}^{(j)})) \quad (12)$$

ここで、 $d_1(\cdot, \cdot)$ はベクトル間のユークリッド距離を表し、 $d_2(\cdot, \cdot)$ はヒストグラム間の χ^2 距離を表す。それぞれの距離は $d_1, d_2 \in [0, 1]$ であり、したがって $0 \leq \alpha \leq 1$ のとき上式の距離は $d_{i,j} \in [0, 1]$ となる。実験では、 $\alpha = 0.2$ と設定する。

標本 $\mathbf{x}^{(i)}$ の対数尤度 ℓ は参照画像上の標本との画像特徴量間の距離に基づいて、次式の

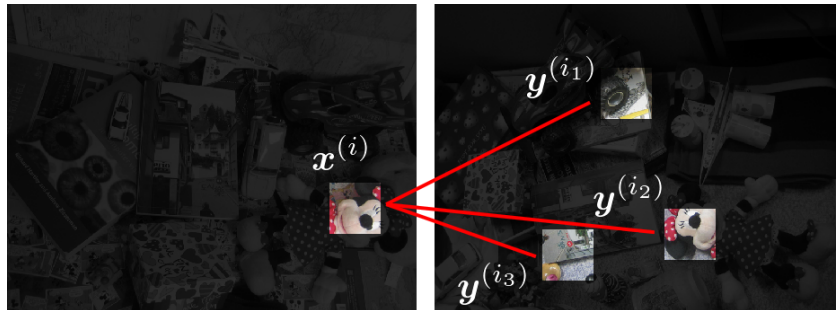


図5 標本に対応する矩形領域間の比較.

Fig. 5 Comparison between rectangular areas corresponding to samples on each image.

ように定義する.

$$\ell(\mathbf{x}^{(i)}) = 1 - \min_{j \in \{1, \dots, M\}} d_{i,j} \quad (13)$$

この定義では、標本 $\mathbf{x}^{(i)}$ が共通パターンを包含する場合でも、そのパターンを包含する参照画像上の標本 $\mathbf{y}^{(j)}$ が存在しなければ、そのステップでの尤度は低くなる。しかしながら、粒子フィルタの枠組みに従って重みが更新されると、過去の探索において尤度が高ければ、その標本は一定の重みを維持する。反復処理によって、共通パターンが存在する領域の付近にある標本は重みが大きくなり、また、それらがクラスタリングされ、混合成分を構成すると対応する混合係数も大きくなる。

与えられた画像のそれぞれにおいて、すべての標本から次式のような投票を行うことによって、共通パターンが存在する確信度を表す画素単位のマップが生成される。

$$\mathcal{M}(x, y) = \sum_{i=1}^M r(x, y; \mathbf{x}_T^{(i)}) \pi_{z_T^{(i)}, T} w_T^{(i)} \quad (14)$$

ここで、 $r_0 = \min(l_x, l_y)/2$ とすると、関数 r は次式ようになる。

$$r(x, y; \mathbf{x}) = \begin{cases} 1 - \frac{(x-c_x)^2 + (y-c_y)^2}{r_0^2} & \text{if } (x-c_x)^2 + (y-c_y)^2 \leq r_0^2 \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

すなわち、標本からの投票は対応する矩形領域内だけで有効であり、中心座標から離れるほど影響が小さくなる。

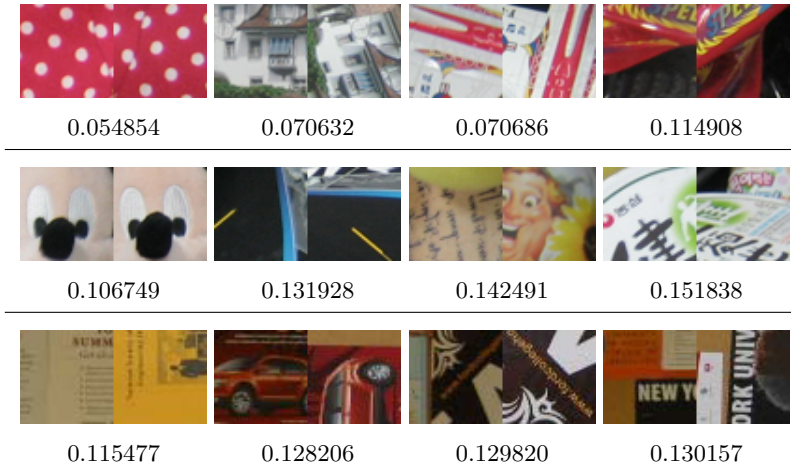


図6 矩形領域間の対応付け (数値は特徴量間の距離を示す).

Fig. 6 Matching among rectangular areas; values indicate the distance between image features.

4. 実験

実験には、先行研究 8) と同じテスト画像*1を用いた。まず予備実験の結果として、1組の画像のそれぞれにランダムに配置された矩形領域間の対応付けについて述べる。次に、提案手法による共通パターン検出の実験として、生成された確信度マップを示し、マップのしきい値処理によって切り出された領域と正解マスク画像との比較による検出精度の客観評価について述べる。

4.1 矩形領域間の対応付け

本稿では、SIFT⁷⁾に基づく特徴量および、色相ヒストグラム²¹⁾を画素単位で密に計算し、矩形領域について累積して画像特徴量とした。矩形領域間の類似度は式 (12) で定義される特徴量間の距離に基づく。対応付けの精度に関する実験を行った結果を図 6 に示す。ここでは、3組の画像それぞれについて、最も類似する矩形領域を対応付けた結果として、部分画像の組および、特徴量間の距離の例を示している。

1行目の画像 (**Toys**) については、それぞれの画像上に共通する各パターンの大きさが

*1 <http://cv.snu.ac.kr/~corecognition/> から入手できる。

ほぼ同じであり、回転による変形も少ないため、比較的精度良く対応付けできている。2行目の画像 (Mickey's) では、共通パターンのいくつかが大きく変形しており、それぞれ異なるパターンを包含する矩形領域間の類似度よりも小さく (特徴量間の距離が大き) くなってしまっている。特徴量間の距離計算におけるパラメータは $\alpha = 0.2$ と設定しており、これは部分画像の色情報よりもテクスチャ性を重視するため、3行目の画像 (Bulletins) では、1列目の組が最も類似する結果となった。

矩形領域を検出器とすると、部分的に共通パターンを包含する場合に背景部分を含むことになり、類似度が低くなる。アフィン変換による領域の規格化^{25),26)}によって、背景部分の影響を抑えられる可能性がある。また、式 (12) のように複数の特徴量間の距離を線形結合するのではなく、色情報を考慮した SIFT に基づく特徴量^{21),27)} などとの比較も必要である。

4.2 検出結果

図 1 に示した画像を入力として、提案手法により生成された確信度マップを図 7 に示す。画素の明るさによって、共通パターンが存在する可能性の高さが表現される。共通パターンが存在する領域の付近で特に明るさが増していることが分かる。その他の画像の組について実験を行った結果を図 9 に示す。

先行研究^{8), 11)}では、抽出結果の客観的な評価方法として、アルゴリズムによって出力された画像と正解画像 (ground truth) を次式の h_r と b_r という基準を用いて比較している。

$$h_r = \frac{|\text{GroundTruth} \cap \text{Result}|}{|\text{GroundTruth}|}, \quad b_r = \frac{|\text{Result}| - |\text{GroundTruth} \cap \text{Result}|}{|\text{Result}|} \quad (16)$$

提案手法は確率的な推定手法であるので、アルゴリズムを実行するたびに異なる結果が得られる。それぞれの画像の組について 5 回ずつ実験を行い、最も高い抽出精度となったときのしきい値に対する h_r と b_r の平均値を表 1 に示す。また、5 回の実験からそれぞれ真陽性と偽陽性を計算し、その平均値から描いた ROC 曲線を図 8 に示す。これらの評価は、共通するパターンに対応する領域を精確に切り出すという観点の尺度に基づいており、本手法による確信度マップの単純なしきい値処理による結果はあまり良くない。確率的な解釈に基づいたハフ変換による検出手法²⁸⁾など、より洗練された手法によって、確信度マップから共通パターンを特定することが必要である。また、前述した矩形領域間の対応付けに関する実験から、本稿で採用した画像特徴量では対応付けがうまくいかない領域があることが分かっている。特に Minnie's の画像では、色が比較的類似し、テクスチャ性の無いような領域を識別できないため、画像特徴量について検討が必要である。



図 7 共通パターンの確信度マップ (中段) および、共通パターン抽出結果 (下段).
Fig. 7 Confidence map for common visual patterns (middle) and segmentation result (bottom).

5. おわりに

本稿では、1組の画像に共通して存在する視覚的なパターンを検出する手法として、粒子フィルタの枠組みを検出アルゴリズムに適用した。特に、混合粒子フィルタを適用すること

表 1 h_r と b_r による抽出結果の客観的評価.

Table 1 Objective evaluation of segmentation based on h_r and b_r .

	Mickey's	Minnie's	Jigsaws	Toys	Books	Bulletins
h_r	.253/.158	.067/.764	.633/.664	.557/.644	.625/.713	.122/.174
b_r	.342/.488	.192/.436	.311/.306	.248/.233	.223/.166	.725/.540

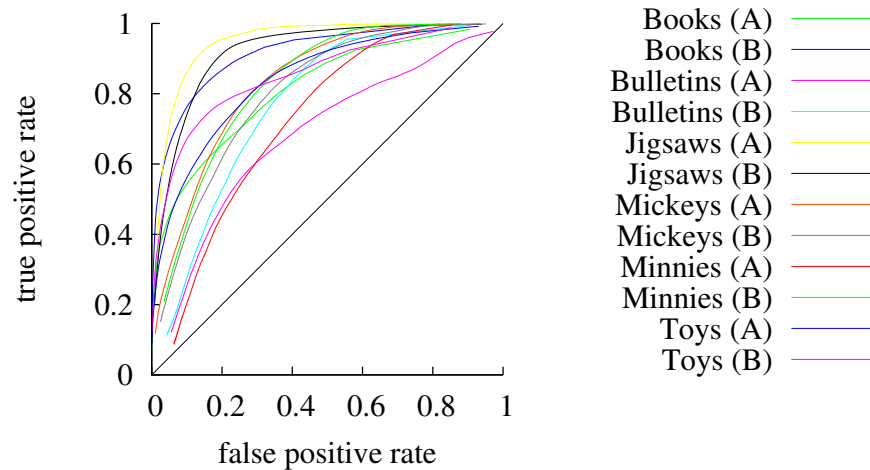


図 8 ROC 曲線.
Fig.8 ROC curves.

によって、複数存在しうる共通パターンあるいは、乱雑な背景や雑音に対して頑健な手法となる。

色情報を考慮した画像特徴量の様々な記述手法²¹⁾について比較したい。アフィン領域を検出する手法²⁵⁾や対応付けのための領域の規格化手法²⁶⁾との統合についても検討したい。また、1組の画像のそれぞれで共通パターンの大きさが異なる場合への対応として、マルチスケール表現が必要であると考えられる。

参考文献

- 1) Lee, Y.J. and Grauman, K.: Foreground Focus: Unsupervised Learning from Partially Matching Images, *IJCV*, Vol.85, No.2, pp.143–166 (2009).
- 2) Liu, C., Yuen, J. and Torralba, A.: Nonparametric Scene Parsing: Label Transfer via Dense Scene Alignment, *CVPR* (2009).
- 3) Brendel, W. and Todorovic, S.: Video Object Segmentation by Tracking Regions, *ICCV* (2009).
- 4) Hedau, V., Arora, H. and Ahuja, N.: Matching Images Under Unstable Segmentations, *CVPR* (2008).
- 5) Berg, A.C. and Malik, J.: Geometric Blur for Template Matching, *CVPR* (2001).
- 6) Oliva, A. and Torralba, A.: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope, *IJCV*, Vol.42, No.3, pp.145–175 (2001).
- 7) Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints, *IJCV*, Vol.60, No.2, pp.91–110 (2004).
- 8) Cho, M., Shin, Y.M. and Lee, K.M.: Co-recognition of Image Pairs by Data-Driven Monte Carlo Image Exploration, *ECCV* (2008).
- 9) Vermaak, J., Doucet, A. and Pérez, P.: Maintaining Multi-Modality through Mixture Tracking, *ICCV* (2003).
- 10) Tan, H.-K. and Ngo, C.-W.: Common Pattern Discovery using Earth Mover's Distance and Local Flow Maximization, *ICCV* (2005).
- 11) Yuan, J. and Wu, Y.: Spatial Random Partition for Common Visual Pattern Discovery, *ICCV* (2007).
- 12) Liu, H. and Yan, S.: Common Visual Pattern Discovery via Spatially Coherent Correspondences, *CVPR* (2010).
- 13) Rother, C., Kolmogorov, V., Minka, T. and Blake, A.: Cosegmentation of Image Pairs by Histogram Matching - Incorporating a Global Constraint into MRFs, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.993–1000 (2006).
- 14) Joulin, A., Bach, F. and Ponce, J.: Discriminative Clustering for Image Cosegmentation, *IEEE Conference on Computer Vision and Pattern Recognition* (2010).
- 15) Cho, M., Shin, Y.M. and Lee, K.M.: Unsupervised Detection and Segmentation of Identical Objects, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1617–1624 (2010).
- 16) Sofka, M., Zhang, J. and Zhou, S.K.: Multiple Object Detection by Sequential Monte Carlo and Hierarchical Detection Network, *CVPR* (2010).
- 17) Doucet, A., de Freitas, N. and Gordon, N.: *Sequential Monte Carlo Methods in*



図9 入力画像, 共通パターンの確信度マップおよび, 共通パターン抽出結果.

Fig. 9 Input pair of images, confidence map for common visual patterns, and segmentation result.

Practice, Springer-Verlag (2001).

- 18) Pérez, P., Hue, C., Vermaak, J. and Gangnet, M.: Color-Based Probabilistic Tracking, *ECCV* (2002).
- 19) Kohonen, T.K., Schroeder, M.R. and Huang, T.S.: *Self-Organizing Maps*, Springer-Verlag New York, Inc., 3rd edition (2001).
- 20) Geman, S. and Geman, D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE PAMI*, Vol.6, No.6, pp.721–741 (1984).
- 21) Van de Sande, K. E.A., Gevers, T. and Snoek, C. G.M.: Evaluating Color Descriptors for Object and Scene Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.32, No.9, pp.1582–1596 (2010).
- 22) Viola, P. and Jones, M.J.: Robust Real-Time Face Detection, *International Journal of Computer Vision*, Vol.57, No.2, pp.137–154 (2004).
- 23) Porikli, F.: Integral Histogram: A Fast Way to Extract Histograms in Cartesian Spaces, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.829–836 (2005).
- 24) Adam, A., Rivlin, E. and Shimshoni, I.: Robust Fragments-based Tracking using the Integral Histogram, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.798–805 (2006).

- 25) Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Van Gool, L.: A Comparison of Affine Region Detectors, *International Journal of Computer Vision*, Vol.65, No.1-2, pp.43–72 (2005).
- 26) Ferrari, V., Tuytelaars, T. and Van Gool, L.: Simultaneous Object Recognition and Segmentation from Single or Multiple Model Views, *International Journal of Computer Vision*, Vol.67, No.2, pp.159–188 (2006).
- 27) Fulkerson, B., Vedaldi, A. and Soatto, S.: Localizing Objects with Smart Dictionaries, *European Conference on Computer Vision*, Lecture Notes in Computer Science, Vol.5302, pp.179–192 (2008).
- 28) Barinova, O., Lempitsky, V. and Kohli, P.: On Detection of Multiple Object Instances using Hough Transforms, *CVPR* (2010).