

局所特徴の位置関係の学習に基づく 複雑背景下での把持状態の検出*

森岡 慎也 平本 康裕 島田 伸敬 松尾 直志 白井 良明 (立命館大学)

概要 把持状態の手は一部が隠れるため、画像による検出や姿勢推定が難しい。画像特徴と手首の位置、把持パターン固有の物体の出現位置の共起性を学習することで、把持状態の手指を検出する手法を提案する。人間が道具を使用しているシーンの中で多様な道具の持ち方を学習することで、既知の手形状がどの位置・向きの把持動作に対応する物体位置を予想する。複雑背景下での手首位置と物体位置を検出し、手と物体領域を切り出した結果を示す。

キーワード: 物体検出, 統計的学習, インタラクション解析

1 はじめに

人間の手が物体を握る、つまむなど様々な動作を認識するために、画像から手の検出や姿勢推定の課題がある²⁾。手の動作と物体の機能関係を用いて物体のカテゴリを認識する研究がある³⁾⁴⁾⁵⁾。しかし、把持状態では指や手の一部が隠れるために、検出や姿勢推定が困難である。そこで、本研究は手の部分特徴を学習することで、隠れている部分があっても把持状態を検出する。人間が道具を使用しているシーンを多様な道具の持ち方を学習することで、既知の手形状がどの位置にどの向きの把持動作に対応する物体がどこにありそうかを予想する。複雑背景下での手首位置と物体位置を検出し、手と物体から背景を切り出せることを実験で示す。

2 Randomized Trees を用いた手の局所特徴の位置関係に基づく手首位置検出

2.1 手首位置確率分布を出力する Randomized Trees モデルの学習

Randomized Trees⁶⁾ は多次元の特徴を入力としクラスラベルの確率分布を学習/出力することができる識別器である。学習時と異なるサイズ、向き、位置に手が現れても適切に確率分布を推定するために、学習画像中の個々の SURF 特徴のサイズ、向きを用いて手首位置の教師データを正規化して得られた相対位置を Randomized Trees に入力して学習を行う。しかし、Randomized Trees が出力する確率分布は離散的な確率分布なので、そのままでは空間座標のような連続分布を学習することはできない。そこで画像中の手首位置を離散的なグリッドに分割し、128 次元の SURF⁷⁾ 特徴を入力すると、その離散的な手首位置のラベルの確率分布を出力する Randomized Trees を構成する。具体的には正規化した相対位置をグリッドに切り、各グリッドに固有のクラスラベルを付与し、教師データの手首位置がどのクラスラベルのグリッドに入ったかを教師信号として学習を行う。

以下に手首位置分布の Randomized Trees の学習方法を詳述する。手の把持状態を再現した背景単純の学習画像を多数用意し、学習画像に教師信号である手首位置を与える (Fig.1-(a))。次に学習画像から SURF 特徴を抽出する (Fig.1-(b))。学習画像中の個々の SURF 特徴の位置・方向・スケールを基準とした局所座標系での手首の相対位置を求める。抽出された特徴を全てカ

バーできるように離散化したラベルのついた手首位置グリッドを設定し (Fig.1-(c))、個々の特徴に対する手首相対位置をラベル付けする。手首検出のために、付けるラベルは式 (1) で与えられ、SURF 特徴ベクトルからの手首の相対位置関係を表す固有番号とする。

$$C_{wrist} = (C_{wrist_1}, C_{wrist_2}, \dots, C_{wrist_n}) \quad (1)$$

SURF 特徴 f のスケールの大きさを s_f 、オリエンテーションを θ_f 、座標を x_f, y_f として、 j 番目のグリッド C_{wrist_j} に手首位置 (x_w, y_w) が属する条件は以下のよう

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{s_f} \text{Rot}(-\theta_f) \begin{pmatrix} x_w - x_f \\ y_w - y_f \end{pmatrix} \quad (2)$$

$$|u - u_c(j)| \leq S_{cell}/2 \quad (3)$$

$$|v - v_c(j)| \leq S_{cell}/2 \quad (4)$$

ただし、 $\text{Rot}(\theta)$ は画面内回転行列、 $u_c(j), v_c(j)$ は j 番目のグリッドの中心位置の座標、 S_{cell} 手首位置の量子化サイズ、グリッドのセルの u 軸方向の個数 n_u と v 軸方向の個数 n_v 、つまり、 C_{wrist} の個数 n は $n = n_u n_v$ である。作成した学習画像中の各 SURF 特徴 f ごとに、上の条件を満たすグリッドのクラスラベル j (すなわち教師データの手首位置が所属する相対位置座標) を求め、 f が持つ 128 次元の SURF 特徴量とクラスラベル j の組を Randomized Trees に入力して学習を行う。

把持状態の手画像については必ず手首位置の教師信号が与えられるので、SURF 特徴量と手首位置の関係が確率分布として Randomized Trees に学習される。しかし手以外の背景領域に対する SURF 特徴に対しては、手首位置が対応しない、すなわちその特徴が背景であることを学習させたい。そこで背景クラスラベル C_{back} を追加し、背景の SURF 特徴量には C_{back} を対応させて Randomized Trees に入力する。

$$C_{wrist} = (C_{wrist_1}, C_{wrist_2}, \dots, C_{wrist_n}, C_{back}) \quad (5)$$

2.2 手首位置確率分布の投票に基づく手首位置検出

検出時には入力画像から SURF 特徴を抽出し、前節で構築した Randomized Trees に入力すると、対応する手首位置ラベルの確率分布が得られる。得られた手首位置ラベル確率分布を画像空間に投票する。画像中の全ての SURF 特徴について投票を行い、投票が集中し

*本研究は第 16 回 画像の認識・理解シンポジウム MIRU2013 で発表したものである¹⁾。

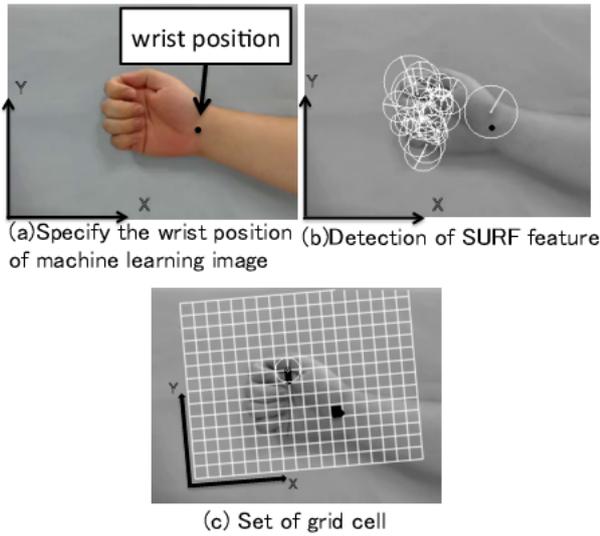


Fig. 1: Conversion of wrist position and positions of features to relative positional relationship class labels

ている位置を手首の位置とする.

SURF 特徴ベクトルを v , 手首位置確率分布を $P_{wrist}(C_{wrist}|v)$, 投票空間を $V_{wrist}(x, y)$ とする. 各 v に対して式 (2) を用いて投票空間の座標 (x, y) が所属するクラスラベル j を決定し, そのクラスラベル j の確率値 $P_{wrist}(C_{wrist_j}|v)$ を (x, y) の位置に加算することで投票を行う. ($V_{wrist}(x, y)$) において最も大きい投票数を得た座標 (x_{wrist}, y_{wrist})

$$(x_{wrist}, y_{wrist}) = \arg \max_{x, y} (V_{wrist}(x, y)) \quad (6)$$

を手首位置とする. 複数の極大位置を効率的に検出するためにここでは mean-shift を用いる. 画像空間中に等間隔に mean-shift のシードを配置し (Fig.2-(a)), 移動の収束地点をクラスタリングし (Fig.2-(b)), 各クラスタ中の最大投票数を持つ画素位置を手首位置の候補として検出する (Fig.2-(c)). (青丸が開始地点, 緑丸が軌跡, 赤丸が終了地点を表している.Fig.2-(a) 緑輪が第1手首位置候補, 青輪が第2手首位置候補, 茶色の領域が終了地点からの収束領域を表している. Fig.2-(b) 赤四角形が最終的な手首位置を表している.Fig.2-(c))

手領域に比べて背景の領域が広い場合, 背景の SURF 特徴からの投票が多数生成されてノイズ成分となり, 手首位置の決定の妨げとなる. そのため, 予め手らしい SURF 特徴とそれ以外の特徴の学習サンプルを用意しそれらを大雑把に識別する SVM⁸⁾ を構築する. その際明らかに手ではない特徴のみをふるい落とすように, SVM で手である確率が十分低いと判定された特徴を除外してから前述の投票を行う. Fig.3-(g) 中の赤色の四角形が検出された手首位置であり, 手首検出に成功しているのがわかる.

3 手特徴の見えと物体の位置関係の学習に基づく手・物体領域の抽出

物体特徴は画像中のどの場所に現れるかは分からないので, 物体のたまかな位置を推定する必要がある. 物体のたまかな位置である物体重心位置を Randomized Trees によって推定し, 手首-物体座標系を作成する. 手

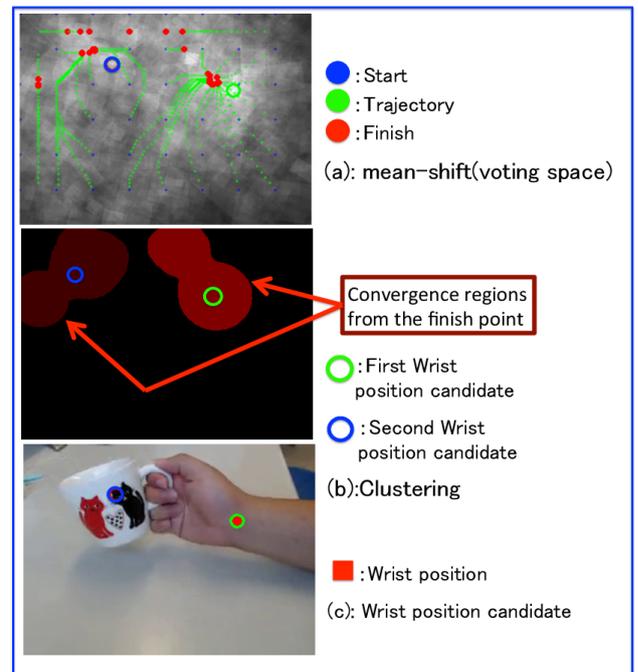


Fig. 2: Flow of wrist candidate detection

首-物体座標系における手特徴と物体特徴の One Class SVM⁹⁾ を構築することで, 手・物体領域の検出を行う.

3.1 特徴の手らしさに基づく手・物体特徴の分類

実際に物体を把持した画像を学習画像とし, 2章のアルゴリズムを用いて手首位置を検出する. ここでは物体把持状態の画像を対象とするので, 画像特徴を手に所属するものと物体に所属するものに分類する必要がある. そこで, 各 SURF 特徴が, 手首位置を決定する際にその検出位置への投票にどれだけ貢献したか (投票した確率値) をその特徴の手らしさ h として評価し (Fig.4(b)), 画像中の位置 (x, y) と h を使って分類する. 手らしさ h の算出方法は, v_{max} に投票した SURF 特徴のインデックスを i , 特徴 i がその場所に投票した値を P_{wrist}^i とすると, 式 (7) で表せる.

$$v_{max} = \sum_i P_{wrist}^i \quad (7)$$

P_{wrist}^i の値が大きいほど, 特徴 i は手首位置の決定に強く貢献したということになる. つまり, 各特徴の手らしさ h_i は式 (8) になる.

$$h_i = P_{wrist}^i \quad (8)$$

画像中の特徴位置とその手らしさの組 (x, y, h) を, K-means 法を用いて 2 クラス分類を行い, 各クラスタの中心 x_{m1} と x_{m2} の h 値を比較して大きい方を手クラス L_{hand} , 他方を物体クラス L_{obj} と判別する式 (9) .

$$\begin{cases} h_{m1} > h_{m2} \rightarrow \begin{cases} m1 \rightarrow m_{L_{hand}} \\ m2 \rightarrow m_{L_{obj}} \end{cases} \\ otherwise \rightarrow \begin{cases} m1 \rightarrow m_{L_{obj}} \\ m2 \rightarrow m_{L_{hand}} \end{cases} \end{cases} \quad (9)$$

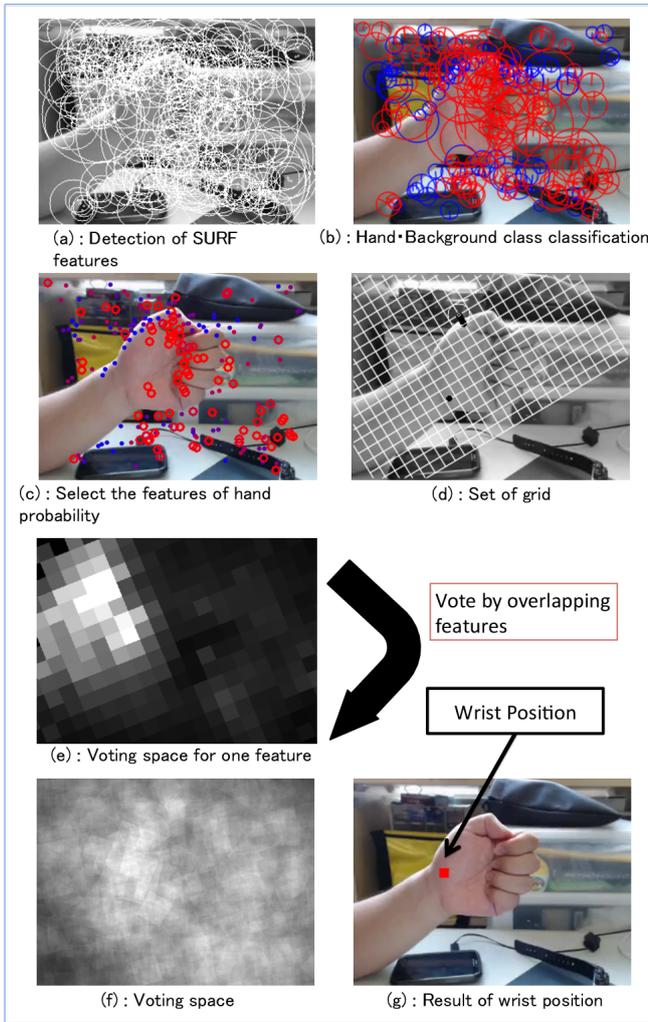


Fig. 3: Flow of based on votes from the probability distribution of wrist positions

各特徴 \mathbf{x}_i と重心ベクトル \mathbf{x}_m のユークリッド距離 $d(\mathbf{x}_i, \mathbf{x}_m)$ を用いて \mathbf{x}_i が手クラス L_{hand} か物体クラス L_{obj} かを判別する式 (10)

$$\begin{cases} d(\mathbf{x}_i, \mathbf{x}_{m_{L_{hand}}}) < d(\mathbf{x}_i, \mathbf{x}_{m_{L_{obj}}}) \rightarrow i \in L_{hand} \\ otherwise \rightarrow i \in L_{obj} \end{cases} \quad (10)$$

これより、手首位置と物体クラス特徴の重心位置 (Fig.4-(c)), 手首-物体座標系を生成する. Fig.4-(d) は手クラスに分類された SURF 特徴の有効範囲に基づいて手領域を表示したものである (色の濃いところほど手らしい).

3.2 物体位置確率分布を出力する Randomized Trees の学習

前節の処理によって、学習用物体把持サンプル画像を手と物体の各領域に分割することができる. 分割された学習用把持画像を用いると、そのような把持パターンにおいて手首位置と把持された物体の中心位置の関係を新たに学習することが可能となる.

そこで、前節で抽出した学習用物体把持サンプル画像中の物体重心位置と、同じ画像中の手の SURF 特徴の関係を 2.1 節と同様に Randomized Trees を使って学習し、局所特徴に対して相対的な物体位置予測器を構築

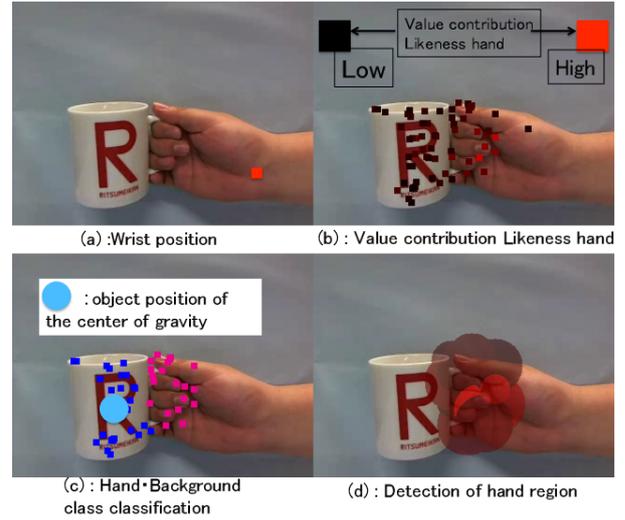


Fig. 4: Flow of Classification of hand and object features

する. 検出時には手首位置の検出をしたあと、手らしさ h が高い SURF 特徴に対してこの物体位置予測器の出力確率分布を画像空間に投票して物体重心位置を検出する. 検出結果例を Fig.5 に示す (赤四角が手首位置, 青四角が物体重心位置).

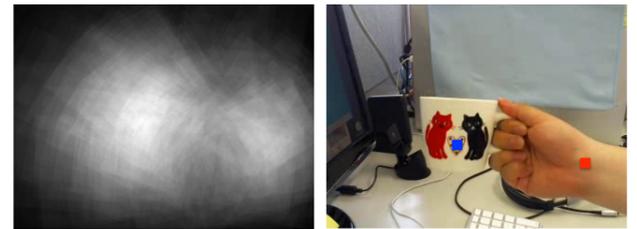


Fig. 5: Detection of center of gravity of objects.

3.3 手首-物体座標系に基づく手・物体特徴を識別する One Class SVM の学習

前節の学習器を獲得することによって、手の把持状態を入力すると画像中のどの位置に物体中心が現れやすいかを予測することができるようになる. すると、検出された手首位置と前説の物体中心位置予測から推定される手首-物体座標系を設定することができるので、その基準座標系において相対的にどこに把持された物体領域が出現するかを学習することで、物体領域と背景領域を分割することが可能になる.

そこで、3.1 節で定義した手首-物体座標系において、画像中の手と物体の領域範囲を推測するために、手特徴の One Class SVM, 物体特徴の One Class SVM をそれぞれ構築する. One Class SVM は正事例のみを用いて学習する 2 クラス識別器である. 手特徴については、手首-物体座標系における位置と手らしさの $\mathbf{x}' = (x, y, h)$ を入力に用いた識別器を学習する. 具体的に手クラス特徴識別器 $f_{hand}(\mathbf{x}')$ は式 (11) で表せる.

$$\begin{cases} f_{hand}(\mathbf{x}') > 0 \rightarrow \mathbf{x}' \in L_{hand} \\ otherwise \rightarrow \mathbf{x}' \notin L_{hand} \end{cases} \quad (11)$$

物体特徴については、一般に把持される物体の見えに共通する、「そのクラスの物体らしさ」が評価できないので、手首-物体座標系における画像中の2次元位置 $\mathbf{y}' = (x, y)$ のみを用いて識別器を学習する。具体的に物体クラス特徴識別器 $f_{obj}(\mathbf{y}')$ は式 (12) で表せる。

$$\begin{cases} f_{obj}(\mathbf{y}') > 0 \rightarrow \mathbf{y}' \in L_{cobj} \\ otherwise \rightarrow \mathbf{y}' \notin L_{cobj} \end{cases} \quad (12)$$

入力画像から \mathbf{x}' と \mathbf{y}' を算出する。把持形状1つに対して、物体は無数に存在する。つまり、 $f_{obj}(\mathbf{y}')$ よりも $f_{hand}(\mathbf{x}')$ の方が信頼できる可能性があるので、物体クラス候補認識は式 (13) のようにする。

$$\begin{cases} f_{obj}(\mathbf{y}') > 0 \cap f_{hand}(\mathbf{x}') \leq 0 \rightarrow \mathbf{y}' \in L_{cobj} \\ otherwise \rightarrow \mathbf{y}' \notin L_{cobj} \end{cases} \quad (13)$$

これら2つの識別器により手クラス l_{hand} , 物体候補クラス l_{obj} , 背景クラス l_{back} の識別関数 $f_l(\mathbf{x}', \mathbf{y}')$ は、式 (14) で表せる。

$$f_l(\mathbf{x}', \mathbf{y}') = \begin{cases} L_{hand} & f_{hand}(\mathbf{x}') > 0 \\ L_{cobj} & f_{obj}(\mathbf{y}') > 0 \cap f_{hand}(\mathbf{x}') \leq 0 \\ L_{back} & otherwise \end{cases} \quad (14)$$

手・物体・背景クラス（手・物体の双方で負となる特徴）の分類を行った結果例を Fig.6 に示す（ピンク四角形：手、青四角形：物体、緑点四角形：背景、赤四角形：手首位置）。

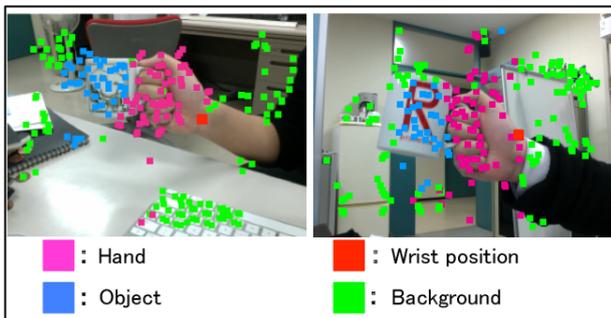


Fig. 6: Result of feature class classification

4 まとめ

手クラスの SURF 特徴と手首位置の関係を学習した手位置確率分布を出力する Randomized trees を構築し、画像中から手首位置検出を行った。手首位置検出結果から、SURF 特徴の手らしさを算出し、単純背景画像から手クラスと物体クラスの分類を行った。手クラス・物体クラスの分類結果から、物体の中心位置を予測する物体位置確率分布を出力し、手首-物体座標系を作成した。作成した手首-物体座標系において手と物体の領域範囲を推測するために、手特徴の One Class SVM, 物体特徴の One Class SVM をそれぞれ構築し、2つの識別器から複雑背景下による手・物体・背景領域の分割を行った。

今後の方針は、様々な物体がどのような把持形状をするのかという関係を学習し、それを基に物体検出を行う。

参考文献

- 1) 森岡慎也, 平本康裕, 島田伸敬, 松尾直志, 白井良明: “局所特徴の位置関係の学習に基づく複雑背景下での把持状態の検出”, 第16回画像の認識・理解シンポジウム, SS2-25, 2013
- 2) 川端聡, 日浦慎作, 佐藤宏介: “カーネル PCA による未知物体領域の高速抽出法”, 電子情報通信学会論文誌, 情報・システム J91-D(11), pp.2673-2683, 2008-11-01
- 3) 鎌倉矩子, “手のかたち 手の動き”, 医歯薬出版株式会社, May. 1989.
- 4) 笠原啓雅, 松崎淳, 島田伸敬, 田中弘美: “把持パターン画像の学習に基づく欠損画素復元と物体認識”, 第11回画像の認識・理解シンポジウム, IS2-5, pp.623-628, 2008
- 5) 浜田康志, 島田伸敬, 白井良明: “手話認識のための複雑背景下で高速に運動する手指の形状推定” 電子情報通信学会論文誌 D Vol.J90-D No.3, pp.617-627, 2007.
- 6) V. Lepetit and P. Fua: “Keypoint recognition using randomized trees”, Transactions on Pattern Analysis and Machine Intelligence, 28,9,pp.1465-1479
- 7) H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool: “SURF: Spedded Up Robust Features”, Computer Vision and Image Understanding, Vol.110,No.3,pp346-359(2008)
- 8) V.Vapnik: “The Nature of Statistical Learning Theory”, Springer, (1995).
- 9) B.Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson: “Estimating the support of a high-dimensional distribution.”, Neural Computation, 13(7):1443-1471,2001.