

## 階層的姿勢探索木を用いた単眼カメラからの高速3次元物体位置姿勢認識

小西 嘉典<sup>†a)</sup> 半澤 雄希<sup>†</sup> 川出 雅人<sup>†</sup> 橋本 学<sup>††</sup>

Fast 6D Pose Estimation from a Monocular Image Using Hierarchical Pose Trees

Yoshinori KONISHI<sup>†a)</sup>, Yuki HANZAWA<sup>†</sup>, Masato KAWADE<sup>†</sup>,  
and Manabu HASHIMOTO<sup>††</sup>

あらまし 単眼カメラ画像からテクスチャレス物体の3次元位置姿勢を高速に認識する手法として、テンプレートマッチングに基づく手法がこれまで多く提案されてきた。しかし認識対象となる3次元姿勢範囲が広い場合に、照合に用いるテンプレートの数が膨大になり処理速度が低下するという課題があった。この課題に対して本研究では、透視投影に基づく累積勾配方向特徴量(PCOF)と階層的姿勢探索木(HPT)の二つの手法を提案する。PCOFは対象物体の3次元CADを様々な視点から見た2次元投影画像を用いて特徴抽出を行っており、3次元姿勢変化による対象物体の見えの変化に対する許容性と複雑背景に対する頑健性とを兼ね備えている。HPTは各視点において作成された大量のテンプレートに対して、類似度に基づいたクラスタリングとテンプレートの低解像度化を行うことで構築されており、数万個の3次元姿勢候補の中から高速に絞り込み探索を行うことが可能である。9種類の金属部品を様々な方向から撮影したデータセットを用いて評価実験を行い、PCOFとHPTを組み合わせた提案手法が3次元物体位置姿勢認識の高速性・正確性両面において既存手法を上回ることを確認した。

キーワード 3次元物体位置姿勢認識, テクスチャレス物体, テンプレートマッチング, 累積勾配方向特徴量, 階層的姿勢探索木

## 1. ま え が き

産業用や家庭用など様々な用途のロボット向け画像処理において、把持のための3次元物体位置姿勢認識は最も重要な技術の一つである。近年ではKinectなど低価格の3次元センサが普及し、学術研究において3次元位置姿勢認識に使用されることも増えてきた[1], [2]。しかし3次元センサの場合、信頼性や耐久性が要求される産業向けでは高価格である、サイズが大きく取り付ける場所が限定される、計測可能な環境や対象物体が限定的である、キャリブレーションが煩雑であるなど実用上の制約や課題も多い。このため実

アプリケーションにおいては単眼カメラが好んで使用されることも多く、単眼カメラから物体の3次元位置姿勢を高速・高精度に認識することは未だに重要な課題の一つである。

工業製品・部品や生活用品の多くは物体の表面に文字や模様がほとんどないテクスチャレス物体であり、認識の手がかりとなるような特徴が少ない。このためSIFT [3] や SURF [4] などのキーポイント検出と局所特徴量マッチングに基づく手法をテクスチャレス物体の3次元位置姿勢認識に適用することは難しい。そこで物体を様々な方向から見たときの2次元投影画像全体をモデルテンプレートとして用いるテンプレートマッチングに基づく手法がこれまで多く用いられてきた[5]~[10]。しかし認識対象となる3次元姿勢範囲が広い場合には照合に用いるテンプレートの数が膨大になり、認識処理速度が遅くなるという課題があった。

本研究ではそういった課題に対し、位置姿勢認識精度を低下させることなく、テンプレート数を減らして処理を高速化する手法を提案する。提案手法は大き

<sup>†</sup> オムロン株式会社技術・知財本部, 木津川市  
Technology and Intellectual Property H.Q., OMRON Corporation, 9-1 Kizugawadai, Kizugawa-shi, 619-0283 Japan

<sup>††</sup> 中京大学工学部, 名古屋市中京区  
School of Engineering, Chukyo University, 101-2 Yagoto Hommachi, Showa-ku, Nagoya-shi, 466-8666 Japan

a) E-mail: ykoni@ari.ncl.omron.co.jp

DOI:10.14923/transinfj.2016IUP0001

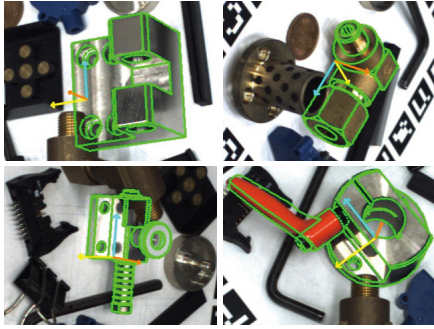


図 1 提案手法による単眼カメラ画像上での 3 次元物体位置姿勢認識結果例。複雑背景下かつ部分的な遮蔽が生じる条件においても、テクスチャがほとんど無く光沢のある金属物体の 3 次元位置姿勢を高速に認識可能である。処理時間は CPU シングルコア使用時で平均 150 ミリ秒

Fig. 1 Our new template based algorithm can estimate 6D pose of texture-less and shiny objects from a monocular image which contains cluttered backgrounds and partial occlusions. It takes an average of approximately 150 ms on a single CPU core.

く二つの要素からなり、一つはより広い 3 次元姿勢範囲の見える変化に対応した透視投影に基づく累積勾配方向特徴量 (PCOF: Perspectively Cumulated Orientation Feature) である。もう一つは、3 次元位置姿勢認識処理を効率化する階層的姿勢探索木 (HPT: Hierarchical Pose Trees) である。これら二つの提案手法を組み合わせることにより、単眼カメラからの高速・高精度な 3 次元物体位置姿勢認識を実現する (図 1)。

本論文では以下、2. で関連研究について概説した後、3. で PCOF と HPT 及び両者を用いた 3 次元物体位置姿勢認識手法について説明を行う。更に 4. で提案手法の有効性を示す実験についての説明と実験結果に対する考察を行い、5. において結論を述べる。

## 2. 関連研究

本章では単眼カメラからの 3 次元物体位置姿勢認識に関する既存研究の概説 (2.1) に加え、提案手法である PCOF 及び HPT と関連の深いテクスチャレス物体に適用可能な画像特徴量に関する既存研究 (2.2) と探索木を用いた効率的な位置姿勢認識に関する既存研究 (2.3) についても概説する。

### 2.1 単眼カメラからの 3 次元物体位置姿勢認識

本節では主に単眼カメラからの 3 次元物体位置姿勢認識に関する既存研究について概説するが、3 次元セ

ンサ及び 3 次元特徴量を用いた研究であっても 2 次元特徴量に置き換えが可能な研究については言及する。3 次元物体位置姿勢認識に関する研究は 1980 年代から数多く行われてきたが、当初は物体領域全体をモデルとして用いるテンプレートマッチングに基づく手法が主流であった [5]~[7]。2000 年代初頭からは局所特徴量マッチングから得られる 2 次元と 3 次元の対応点に基づいて 3 次元位置姿勢を認識する手法が増加していった [11]~[13]。こういった手法は物体の姿勢変化による見えの変化に対して頑健である点や物体の種類数や画像解像度が増加しても処理時間が線形的に増加しない点などの長を備えているが、表面にテクスチャが豊富にある物体にしか適用できないという制約がある。テクスチャレス物体にも適用範囲を広げるためにエッジや線分を局所特徴量として用いる手法も提案されているが [14], [15]、そういった特徴量はあまりに単純すぎて弁別性が低く、複雑背景下では局所特徴量マッチングの誤対応が頻発するという課題がある。

テンプレートマッチングと並んで長い歴史をもつ投票に基づく手法も 3 次元物体位置姿勢認識に適用された例は多く、局所特徴量 [16] や random fern [17] による投票、Hough forests [2] や coordinate regressions [18], [19] など様々な手法が提案されてきた。投票に基づく手法には物体の種類数や画像解像度が増加しても処理時間が線形的に増加しないという長があるものの、3 次元位置姿勢認識に適用する場合には投票空間が広大になり、投票結果の最頻値の探索に長い処理時間を要してしまうという課題がある。

近年では CNN を 3 次元物体位置姿勢認識に適用した研究も見られるが [20]~[22]、GPU を使用した場合でも認識に数秒要するため、限られた計算機環境において高速処理が要求されるロボット向けアプリケーションには適していないと考えられる。

テクスチャレス物体の 3 次元位置姿勢認識に対しては、テンプレートマッチングに基づく手法によって実用的な正確性・高速性を実現できることが示されてきた [8], [9], [23]。Hinterstoisser ら [1], [10] は、量子化された輝度勾配方向特徴量と特徴量のメモリ配置最適化に基づく LINE-2D 及び面法線特徴量を加えた LINE-MOD を提案し、複雑背景下においてもテクスチャレス物体の 3 次元位置姿勢を高速に認識できることを示した。対象物体の種類数が多い場合や認識対象の 3 次元姿勢範囲が広い場合に処理速度が低下する課題に対しては、対象物体と背景を識別する 2 クラス

識別器 [24] やテンプレート照合処理を高速化するハッシュテーブル [25], [26] による正確性・高速性の向上が図られている。しかし識別器の学習には認識環境ごとの背景学習サンプル収集が必要である点や、ハッシュテーブル作成時の特徴量次元削減によって認識精度が低下する点などの短所があり、課題の解決には至っていない。

## 2.2 テクスチャレス物体に適用可能な特徴量

単眼カメラ画像からの 3 次元物体位置姿勢認識においては、使用する画像特徴量が認識性能に大きな影響を与える。テクスチャレス物体に対応するためにエッジ特徴量を用いた研究例もあるが、複雑背景下での誤認識を低減させるためには前処理としてセグメンテーション処理 [23] が必要であったり、照明条件の異なる複数枚の撮影画像が必要であったり [8] するなど、高速な認識処理は実現できていない。

一方で輝度勾配方向ベクトル [27] や量子化輝度勾配方向 [28] が複雑背景や照明変動に対して頑健かつ高速な照合処理が可能であることが示されてきたが、物体の見えがわずかでも変化した場合に照合スコアが大きく低下する課題も指摘されてきた。この課題に対して、一定領域内の主要な方向のみ照合に用いる [29]、照合に際して位置許容範囲を設ける [10] 等の手法が提案されてきたが、物体の見えの変化だけでなく背景部分における照合条件も緩和されるため誤認識の増加が懸念される。小西ら [30] は対象物体の 2 次元姿勢変化（平行移動・平面内回転・スケール）による見えの変化のみを許容する累積勾配方向特徴量（COF: Cumulative Orientation Feature）を提案したが、COF では物体の 3 次元姿勢変化は考慮されておらず、3 次元姿勢変化に起因する見えの変化が生じた際には認識性能が低下すると考えられる。

## 2.3 探索木による位置姿勢認識

テンプレートマッチングに基づく手法においてはモデルテンプレートのデータ構造や探索手法も性能に大きな影響を与える。多数の候補の中から最も類似したものを効率的に探索するための主要な手法として探索木があり、これまでに物体認識のための近似最近傍探索 [31]～[33] や、物体と姿勢の同時認識 [34]、2 次元あるいは 3 次元の物体位置姿勢認識 [35], [36] に用いられてきた。更に Ulrich ら [9] は、こういった探索木による効率的な姿勢探索に画像ピラミッドを用いた粗密探索 [37] を加え、画像内の位置探索も同時に効率化した手法を提案した。Ulrich らの手法では近傍の視点を統

合、階層化することで探索木を構成している。しかしテクスチャレス物体のように物体の輪郭周辺にのみ特徴が存在する場合には、視点位置は離れていても作成したテンプレートの特徴量ベクトル同士は類似するといったことが頻繁に発生する。こういった場合 Ulrich らの手法で作成した探索木には類似したテンプレートが多数含まれてしまい、探索処理が冗長になると考えられる。

## 3. 提案手法

提案手法は、3 次元姿勢変化により生じる対象物体の見えの変化のみを許容する透視投影に基づく累積勾配方向特徴量（PCOF, 3.1）と、2 次元的な見えの類似度のみに基づいて階層的に統合された階層的姿勢探索木（HPT, 3.2）から構成される。更に 3.3 において、PCOF のテンプレート群から作成した HPT を用いて物体の 3 次元位置姿勢を認識する方法について説明を行う。なお、これら提案手法の一部については技術情報誌の紹介記事 [38] において発表済みである。

### 3.1 透視投影に基づく累積勾配方向特徴量（PCOF）

PCOF は小西ら [30] の累積勾配方向特徴量（COF）を発展させた特徴量であり、COF を包含する関係にある。COF は物体の 2 次元姿勢変化による見えの変化のみを考慮した特徴量であり、ベルトコンベア上を流れる物体の位置姿勢認識といったタスクへの適用が想定される。一方で PCOF は、3 次元 CAD から生成した様々な 3 次元姿勢の投影画像を用いて特徴抽出を行い、3 次元姿勢変化による見えの変化に対する照合許容度を向上させた特徴量である。このため PCOF は、箱の中にバラ積みされた物体の位置姿勢認識といったタスクへの適用が想定される。なお COF では平行移動パラメータに対しても  $\pm 1$  画素以内の微小な変動を加えているが、認識時に画像内を走査することからこのような変動付加は不要であると考え PCOF では行っていない。

テクスチャレス物体の表面には認識の手がかりとなるような特徴がほとんどなく、認識のためには物体の輪郭や稜線といった形状に関する特徴量が重要である。そこで PCOF 及び COF では形状特徴を捉えつつ照合時の多少の位置ずれを許容するため、輝度勾配方向ヒストグラムを用いた特徴抽出を行っている。輝度勾配方向ヒストグラムは人体検出などで優れた検出性能を示してきた HOG 特徴量 [39] でも用いられているが、



PCOF では位置姿勢の認識分解能を向上させるため、HOG 特徴量のように一定領域内に一つのヒストグラムではなく 1 画素ごとに一つのヒストグラムを作成し特徴抽出を行う。更に処理速度向上のため、DOT 特徴量 [29] と同様に主要な勾配方向の有無のみを記述した 8 桁の二進数特徴量を勾配方向ヒストグラムから抽出し、論理演算を用いた高速照合を可能にしている。

ここからはテクスチャレス物体である L 字ホルダを例として、PCOF の算出方法について説明する。まず対象物体を中心とした球面上 (図 2(a)) に仮想の視点をランダムに設定し、その視点から見た 2 次元投影画像を多数生成する。視点位置を決定するためのパラメータは  $x \cdot y$  軸周りの回転角度、物体までの距離、カメラ光軸周りの回転角度の 4 種類であり、一定範囲内において発生させた一様乱数により決定する。ここで発生させる乱数の範囲は一つのテンプレートで対応できる物体の見えの変動範囲内にとどめる必要があり、本研究では最適範囲を実験的に決定し、 $x \cdot y$  軸周り回転角度  $\pm 12$  度以内、物体までの距離  $\pm 40\text{mm}$  以内、光軸周り回転角度  $\pm 7.5$  度以内とした。z 軸上 680mm の距離、光軸周り回転角度 0 度に配置した視点を中心とし、この変動範囲内のパラメータから生成された

投影画像例を図 2(b) に示した。ただし左上の画像は変動範囲の中心である z 軸上の視点からの投影画像である。

投影画像を生成する際には物体の輪郭部分の特徴量を抽出しやすくするため、隣り合うメッシュのなす角度が一定以上の場合に色を変えて描画する (図 2(b))。本研究ではこの角度のしきい値を 30 度とした。

以上の手順によって生成した  $N$  枚の投影画像を用い、小西ら [30] の記述に従って累積勾配方向特徴量 (COF) の抽出を行う。まず生成した全ての画像に対して Sobel フィルタを適用し輝度勾配ベクトルを算出する。作成した勾配ベクトル画像の一例について、勾配方向に応じて色づけした結果を図 3(a) に示した。

次にこれらの勾配ベクトル画像群から画素ごとに勾配方向ヒストグラムを作成する。画像群共通座標系の各画素に八つのビンをもつヒストグラムを用意し、全画像の勾配ベクトルを図 3(b) のように 8 方向に量子化した後、対応する画素のヒストグラムに投票を行う。この際勾配ベクトル方向に近い二つのビンに対して (図 3(b) の例では 5 と 6) 合計 1 となる頻度値の加算を行い、勾配ベクトルをもたない画素については何も加算しない。その結果、頻度の合計値が最大で  $N$  となるような勾配方向ヒストグラムが各画素において得られる。なお隣り合う二つのビンに加算する頻度値は、各ビンの中心角度と勾配ベクトル角度との差分の比に基づいて合計が 1 となるように算出する。このため各ビンに加算される値は小数となる。

最後に各画素のヒストグラムにおいて頻度しきい値 ( $Th$ ) 以上の勾配方向のみを選択し、対応するビットを 1 にした 8 桁の二進数を累積勾配方向特徴量として抽出する。また頻度の大きな勾配方向は安定して出現する主要な特徴量であると考えられるため、ヒストグ

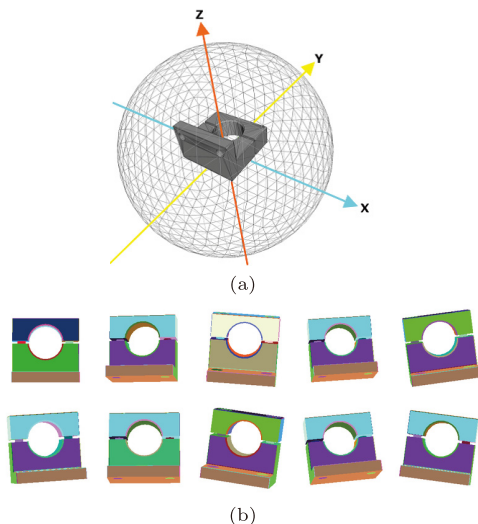


図 2 (a) L 字ホルダの 3 次元 CAD、及び物体の座標軸と視点を配置する球。 (b) z 軸上の視点を中心としてランダムに配置した視点からの投影画像例

Fig. 2 (a) 3D CAD of L-Holder, its coordinate axes and a sphere for viewpoint sampling. (b) Examples of projection images from randomized viewpoints around the viewpoint on z-axis.



図 3 (a) z 軸上にある視点からの投影画像 (図 2(b) 左上) から抽出した輝度勾配方向。 (b) 明暗反転を無視した輝度勾配方向の量子化

Fig. 3 (a) Colored gradient directions of the upper-left image in Fig. 2(b). (b) Quantization of gradient directions disregarding their polarities.

ラムの最大頻度値を照合時に使用する特徴量の重みとして抽出する。最大頻度がしきい値以下の場合は安定的な特徴量が得られなかったとし、その画素を照合には用いないこととする。

z 軸上 680mm の距離にある視点を中心として生成した投影画像 (図 2(b)) から算出した累積勾配方向特徴量 (ori) と特徴量の重み (w) を、任意に選択した四つの画素について図 4 に示した。画像の生成枚数 ( $N$ ) 及び頻度に対するしきい値 ( $Th$ ) は実験的に決定し [30], それぞれ  $N = 1000$ ,  $Th = 120$  とした。PCOF 抽出処理の結果, (2) や (3) のような直線や円弧上の画素においては少数の勾配方向に投票が集中し、重みの大きな特徴量が抽出される傾向が見られた。一方で点 (1) や (4) のような物体の角や複雑な構造上にある画素においては、多数の勾配方向に投票が分散し重みの小さい特徴量が抽出されるあるいは最大頻度がしきい値以下となり照合に使用されないといった傾向が見られた。

重みがゼロとなった画素を除く  $n$  個の PCOF からなるテンプレート  $T$  は次のように表され、

$$T : \{x_i, y_i, ori_i, w_i | i = 1, \dots, n\}, \quad (1)$$

入力画像中の座標  $(x, y)$  における照合スコアは次式で算出する。

$$score(x, y) = \frac{\sum_{i=1}^n \delta_k(ori_i^T(x+x_i, y+y_i) \in ori_i^T)}{\sum_{i=1}^n w_i}. \quad (2)$$

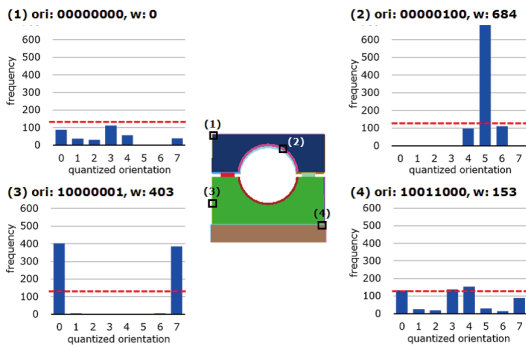


図 4 任意に選択した四つの画素において算出された勾配方向ヒストグラム及び累積勾配方向特徴量 (ori) とその重み (w)。ヒストグラム上の点線は特徴抽出に用いたしきい値

Fig. 4 Examples of the orientation histograms, binary features (ori) and their weights (w) on arbitrarily selected four pixels. Red dotted lines show the threshold for feature extraction.

上式では入力画像の量子化勾配方向 ( $ori^I$ ) がテンプレートの PCOF ( $ori^T$ ) に含まれている場合に重み ( $w$ ) が照合スコアに加算される。また式 (2) のデルタ関数は式 (3) のようにビット積 (記号  $\wedge$ ) で高速に演算可能であり, CPU 固有の SIMD 命令を用いることで照合演算の更なる高速化を図ることもできる。

$$\delta_i(ori^I \in ori^T) = \begin{cases} w_i & \text{if } ori^I \wedge ori^T > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

### 3.2 階層的姿勢探索木 (HPT)

3.1 で説明した PCOF からなるテンプレートは、投影画像生成時に用いたパラメータ範囲内の物体の見えの変化に対応可能である (本研究では  $x \cdot y$  軸周り  $\pm 12$  度以内, カメラ距離  $\pm 40$ mm 以内, 光軸周り  $\pm 7.5$  度以内)。更に広い範囲の 3 次元姿勢変化に対応するため, 図 2(a) のように球面上に 642 点の頂点をほぼ等間隔に設定し各頂点を中心として PCOF テンプレートを作成する (隣り合う頂点は  $x \cdot y$  軸周り約 8 度間隔)。更に各テンプレートが対応する 3 次元姿勢範囲に重なりが生じように、物体までの距離は 30mm 刻み, 光軸周り回転角度は 5 度刻みでテンプレートを作成した。

テクスチャレス物体から抽出される 2 次元特徴量は輪郭部分に集中しており、視点位置が離れている (3 次元姿勢が大きく異なる) 場合でもテンプレートの特徴ベクトル同士が類似していることが多々ある。この傾向は粗密探索のために低解像度で作成したテンプレートにおいて特に顕著である。こういった特性を利用して効率的な 3 次元位置姿勢認識を行うため, Ulrich ら [9] のような近傍の視点のみの統合と階層化ではなく、視点位置を考慮せずテンプレート間の類似度のみに基づいて統合と階層化を行う階層的姿勢探索木 (HPT) を提案する。

HPT 作成アルゴリズムは, Algorithm 1 に示したように入力として PCOF テンプレートとテンプレート作成時の勾配方向ヒストグラムを用い, クラスタリング, 統合, 低解像度化の三つの主要な処理から構成される。まず全ての PCOF テンプレートに対してテンプレート間の照合スコア (式 (2)) に基づいたクラスタリングを行う。クラスタリング手法にはベイズ情報量規準に基づいて最適なクラスタ数を決定する X-means 法 [40] を用いた。次に同じクラスタに分類

**Algorithm 1** 階層的姿勢探索木 (HPT) の作成

**Require:** a number of PCOF templates  $T$  and their orientation histograms  $H$

**Ensure:** hierarchical pose trees (HPT)

```

 $T'_0 \leftarrow T$ 
 $H'_0 \leftarrow H$ 
 $i \leftarrow 1$ 
loop
   $C_i \leftarrow$  cluster the templates in  $T'_{i-1}$ 
  for each cluster  $C_{ij}$  do
     $H_{ij} \leftarrow$  add histograms at each pixel of  $H'_{i-1} \in C_{ij}$ 
     $H_{ij} \leftarrow$  normalize histograms  $H_{ij}$ 
     $T_{ij} \leftarrow$  thresholding  $H_{ij}$  and extract new binary features and weights
  end for
  for each  $T_{ij}$  and  $H_{ij}$  do
     $H'_{ij} \leftarrow$  add histograms of nearby  $2 \times 2$  pixels
     $H'_{ij} \leftarrow$  normalize histograms  $H'_{ij}$ 
     $T'_{ij} \leftarrow$  thresholding  $H'_{ij}$  and extract new binary features and weights
  end for
   $N'_i \leftarrow$  minimum number of feature points in  $T'_i$ 
  if  $N'_i < N_{min}$  then
    print
  else
     $i \leftarrow i + 1$ 
  end if
end loop

```

されたテンプレートに対し、各画素において勾配方向ヒストグラムの加算と正規化、しきい値処理を行って新たな二進数特徴量と重みを抽出し、テンプレートの統合を行う。最後に近傍  $2 \times 2$  画素の勾配方向ヒストグラムの加算と正規化を行って解像度を4分の1に削減し、しきい値処理により低解像度テンプレートの特徴抽出を行う。この手順をテンプレートに含まれる特徴量の数がしきい値 ( $N_{min}$ ) 以下になるまで繰り返す、HPT を作成する。

図2のL字ホルダを用い、4.2の実験2と同じ3次元姿勢範囲 ( $x \cdot y$  軸周り  $\pm 60$  度以内、カメラまでの距離  $660\text{mm} - 800\text{mm}$ 、光軸周り  $\pm 180$  度以内) で作成した HPT の一部を図5に示した。この場合、最初に作成される PCOF テンプレートの総数は 73,800 個になる (視点数 205、距離の種類数 5、光軸周り回転角度数 72 の乗算)。これらの PCOF テンプレートが1回目のクラスタリング、統合、低解像度化処理によって 23,115 個になり、更に2回目 4,269 個、3回目 233 個と減少した後 HPT 作成処理が停止した。なお、本研究では  $N_{min}$  を 50 とした。

**3.3 3次元物体位置姿勢認識**

3次元位置姿勢認識時は、まず入力画像の画像ピラ

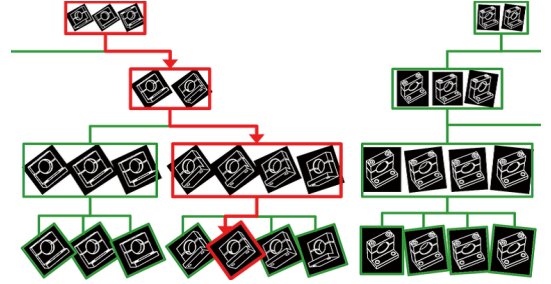


図5 L字ホルダの HPT の一部分。最下層が最初で作成した PCOF テンプレートであり、これらのテンプレートのクラスタリングと統合、低解像度化を繰り返して下から上へと姿勢探索木を構築する。認識時には赤色の矢印のように3次元姿勢を絞り込みながら探索を行う。

Fig. 5 Part of hierarchical pose trees of L-Holder are shown. The bottom templates are originally created PCOF templates and the tree structures are built in a bottom-up way by clustering, integrating and downscaling. In estimation of object pose, HPT is traced from top to bottom along the red arrow.

ミッドを作成した後、各階層において量子化勾配方向特徴量を算出する。その後最上位の階層において HPT のルートテンプレートを用いてラスタスキャンを行い、式(2)に基づいて照合スコアを算出する。照合スコアが探索しきい値以上の位置姿勢認識結果については、一つ下の高解像度の階層において対応するテンプレートを用いて照合を行う。これを繰り返しながら徐々に解像度の高い画像上においてより詳細な位置姿勢を特定していく (図5の赤矢印)、非極大値抑制処理を行った後に入力画像上での検出位置と照合に用いたテンプレートが特定される。この結果から3次元CAD上の3次元座標と入力画像上での2次元座標との対応関係を得ることができ、PnP問題を解くことで物体の3次元位置姿勢を算出する [41]。

**4. 性能評価実験及び考察**

提案手法の性能評価のため2種類の実験を行った。実験1ではPCOFの3次元姿勢変化による対象物体の見えの変化に対する許容性と複雑背景に対する頑健性について、実験2ではPCOFとHPTを組み合わせた3次元物体位置姿勢認識手法の正確性及び高速性についてそれぞれ性能評価及び既存手法との比較を行った。



## 4.1 実験 1

### 4.1.1 実験設定

実験 1 では図 7 に示した 9 種類の金属部品（テクスチャレス物体）を用い、4 種類の輝度勾配方向特徴量の認識性能評価を行った。台上に置いた対象物体の周囲に無関係な物体を乱雑に配置し、3.1 で説明に用いた PCOF 抽出時の視点設定範囲（ $z$  軸上 680mm の距離，光軸周り回転角度 0 度を中心として， $x \cdot y$  軸周り回転角度  $\pm 12$  度以内，物体までの距離  $\pm 40$ mm 以内，光軸周り回転角度  $\pm 7.5$  度以内）内において単眼カメラをランダムに移動させながら撮影を行い，物体ごとに約 100 枚の画像を取得した。L 字ホルダのテスト画像例を図 6 に示した。



図 6 実験 1 で用いた L 字ホルダのテスト画像例

Fig. 6 Example images of L-Holder used in experiment 1.

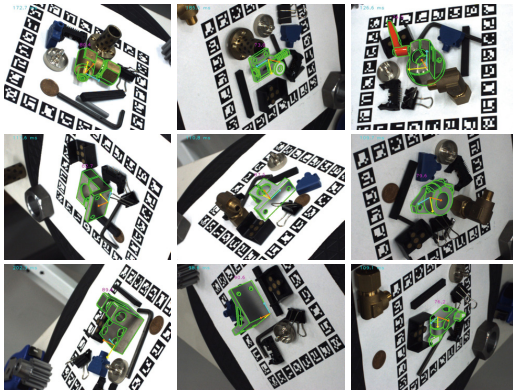


図 7 実験 2 で用いた 9 種類の金属部品データセットの画像例（実験 1 においても同じ金属部品を使用）。上段左からコネクタ，サイドクランプ，ストッパ，中段左から L 字ホルダ，T 字ホルダ，フランジ，下段左からヒンジベース，ブラケット，支柱クランプ。提案手法の 3 次元位置姿勢認識結果に基づき，物体のエッジ点と座標軸を描画した。

Fig. 7 The example images of our dataset consisting of 9 metallic parts in experiment 2 (Same objects are used in experiment 1). Top: Connector, SideClamp, and Stopper. Middle: L-Holder, T-Holder, and Flange. Bottom: HingeBase, Bracket, and PoleClamp. The edges and coordinate axes of the target objects are drawn on the images based on the estimated 6D pose by our proposed method.

比較を行った特徴量は正規化勾配ベクトル特徴量 [27]，拡散勾配方向特徴量 [10]，累積勾配方向特徴量 (COF) [30]，PCOF (提案手法) の 4 種類であった。PCOF 以外の 3 手法では， $z$  軸上距離 680mm，光軸周り 0 度の視点における投影画像（L 字ホルダの場合は図 2(b) 左上の画像）をモデル画像として用いた。

### 4.1.2 結果と考察

テスト画像において 4 種類の特徴量を使用したモデルテンプレートとの照合を行い，対象物体領域（前景）における最大スコアとそれ以外の領域（背景）における最大スコア及びそれらの差分について対象物体毎と全対象物体の平均値を表 1 に示した。物体位置姿勢認識処理においては照合スコアに対してしきい値を設定し，スコアがしきい値以上であった位置姿勢を認識結果として出力する。このため前景と背景における最大スコアの差が大きいくほど設定可能なしきい値範囲が広く，認識性能の高い特徴量であると言える。

正規化勾配ベクトル特徴量は輝度勾配方向の単位ベクトルで表され，単位ベクトル同士の内積和を画素数で除した値を照合スコアとして使用する。照明変動や複雑背景に対して頑健な特徴量ではあるが，対象物体自身の見えの変化については特に考慮されていない。このため対象物体の 3 次元姿勢変化が生じた場合に前景スコアが大きく低下して前景-背景スコア差が小さくなっており（表 1），4 種類の特徴量の中で最も認識性能が低いことが分かる。

拡散勾配方向特徴量は特徴量の照合位置ずれに一定の許容範囲を設けた特徴量である。テスト画像において算出した量子化勾配方向特徴量について， $\pm 4$  画素内にある周辺特徴量とのビット和演算による拡散処理を行い，モデル画像から抽出した量子化勾配方向特徴量と照合を行った。表 1 から，正規化勾配ベクトルと比較すると 3 次元姿勢変化による物体自身の見えの変化に対する許容度は向上しているものの，COF や PCOF と比較すると前景-背景スコア差は小さく，等方的な照合位置ずれの許容では不十分であると考えられる。

COF は対象物体の 2 次元幾何学的変換（平行移動，平面内回転，スケール）による見えの変化に対する許容度を高めた特徴量であり，小西ら [30] の記述に従ってモデル画像からの COF 抽出及びテスト画像から算出した量子化勾配方向特徴量との重み付き照合を行った。表 1 から，COF は背景領域のスコアを低く抑えたまま物体自身の見えの変化を許容できているが，3

表 1 実験 1 における前景 (FG), 背景 (BG) の最大スコアとその差分 (FG-BG) の, 対象物体毎及び全対象物体の平均値

Table 1 The mean values of maximum scores at FG and BG, and the differences between them in experiment 1.

	正規化勾配ベクトル			拡散勾配方向			COF			PCOF (ours)		
	FG	BG	FG-BG	FG	BG	FG-BG	FG	BG	FG-BG	FG	BG	FG-BG
コネクタ	0.567	0.504	0.064	0.614	0.491	0.122	0.616	0.492	0.125	0.707	0.559	<b>0.148</b>
サイドクランプ	0.535	0.486	0.049	0.584	0.469	0.115	0.443	0.325	0.118	0.547	0.403	<b>0.144</b>
ストッパ	0.507	0.421	0.087	0.581	0.400	0.181	0.707	0.360	0.346	0.800	0.398	<b>0.402</b>
L 字ホルダ	0.610	0.492	0.118	0.739	0.479	0.260	0.742	0.374	0.368	0.874	0.440	<b>0.434</b>
T 字ホルダ	0.610	0.470	0.140	0.773	0.466	0.307	0.751	0.378	0.373	0.879	0.480	<b>0.399</b>
フランジ	0.596	0.510	0.086	0.670	0.470	0.200	0.732	0.316	0.416	0.852	0.406	<b>0.446</b>
ヒンジベース	0.548	0.487	0.061	0.658	0.512	0.146	0.606	0.466	0.140	0.703	0.530	<b>0.172</b>
ブラケット	0.501	0.442	0.059	0.528	0.469	0.059	0.702	0.397	0.305	0.921	0.578	<b>0.343</b>
支柱クランプ	0.542	0.468	0.075	0.619	0.418	0.201	0.719	0.325	0.393	0.803	0.386	<b>0.417</b>
平均	0.557	0.475	0.082	0.641	0.464	0.177	0.669	0.381	0.287	0.787	0.464	<b>0.323</b>

次元姿勢変化を考慮していないため PCOF と比較すると前景-背景スコア差が小さく認識性能が低いと考えられる。

PCOF は 3.1 で説明した手順により特徴量と重みを抽出し, テスト画像から抽出した量子化勾配方向特徴量との重み付き照合処理を行った。表 1 から PCOF は前景-背景スコア差が最も大きく, 複雑背景に対する頑健性と 3 次元姿勢変化による対象物体の見えの変化に対する許容性とを兼ね備えていることが分かる。

## 4.2 実験 2

### 4.2.1 実験設定

実験 2 では実験 1 と同じ 9 種類の金属部品を様々なカメラ位置・角度から撮影した画像データセットを用い, 3 次元位置姿勢認識の正確性及び高速性の評価を行った。複雑背景下で部分遮蔽が生じるように対象物体の周囲に無関係な物体を乱雑に配置し,  $x \cdot y$  軸周り  $\pm 60$  度以内, カメラ光軸周り  $\pm 180$  度以内, 物体中心までの距離 660mm - 800mm の範囲内において様々な位置・角度に単眼カメラを設置して撮影を行った。画像の解像度は  $640 \times 480$ px で, 各物体について約 500 枚の画像を取得した。対象物体の 3 次元位置姿勢正解値については, 物体と同一平面上に配置した AR マーカーを認識することで算出した [42]。正解値と認識結果の誤差が,  $x \cdot y$  軸方向距離 10mm 以内,  $z$  軸方向距離 40mm 以内,  $x \cdot y$  軸周り回転角度 10 度以内,  $z$  軸周り回転角度 7.5 度以内の場合に認識成功とした。図 7 に 9 種類の物体それぞれについてデータセットの画像例を示した。

提案手法の他に, Ulrich らの手法 [9], LINE-2D [10], COF [30] の評価を行い, 認識成功率及び処理時間の比較を行った。Ulrich らの手法

については HALCON11 (MvTEC 社) 内の関数「find\_shape\_model\_3d」を, LINE-2D については OpenCV 2.4.11 に実装されているプログラムを使用した。提案手法と Ulrich らの手法については, モデルテンプレート作成に各物体の 3 次元 CAD を使用した。COF と LINE-2D については, データセットの視点範囲をカバーするために必要な 73,800 枚の投影画像をモデル画像として用いた。全ての処理は同一 PC (CPU: Core i7 3.4GHz, OS: Windows7 64bit) 上にて, CPU コアを一つだけ使用して行った。

### 4.2.2 3 次元位置姿勢認識成功率

図 8 に四つの手法による 3 次元位置姿勢認識結果を示した。縦軸は認識成功率, 横軸は画像 1 枚あたりの誤検出数 (FPPI: False Positives Per Image) であり, 照合スコアに適用するしきい値を様々な値に変更することで得られたデータをプロットした。このグラフにおいては左上にプロットされるほど認識正確性が高いことを示しており, PCOF と HPT を組み合わせた我々の提案手法が既存手法と比較して高い認識正確性を有することが分かる。

提案手法の認識成功率が高かったのは, 表 1 で示したように PCOF は複雑背景に対する頑健性と 3 次元姿勢変化による対象物体の見えの変化に対する許容性とを両立した特徴量であるためだと考えられる。特に COF や LINE-2D においてはテンプレート作成を行う視点の数を PCOF と同数に設定したため, 特徴量の認識性能差がそのまま 3 次元位置姿勢認識の成功率の差として現れたのであろう。

Ulrich らの手法で用いられている正規化勾配ベクトル特徴量は 3 次元姿勢変化による見えの変化に対する許容度は最も低い (表 1), 各視点が対応する姿



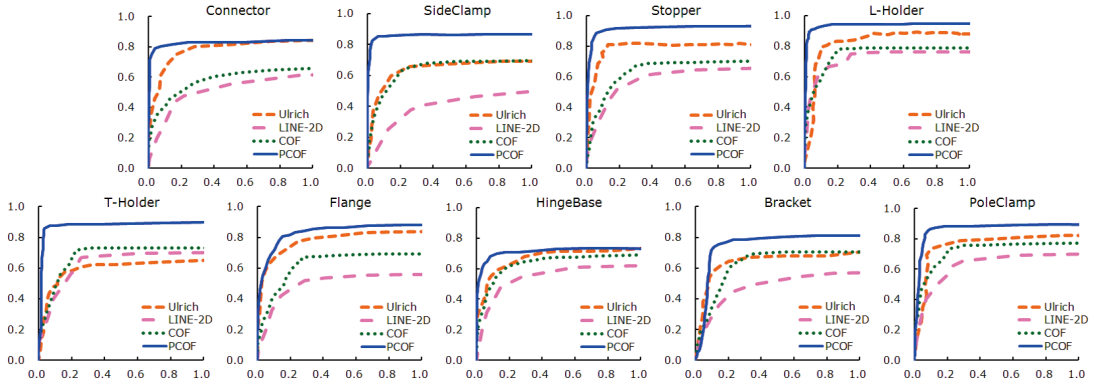


図 8 実験 2 における 3 次元位置姿勢認識成功率 (縦軸) と画像 1 枚あたり誤認識数 (横軸) のグラフ

Fig. 8 The relation between success rate of correctly estimated 6D pose (vertical axis) and false positives per image (FPPI, horizontal axis) in experiment 2 are presented.

表 2 実験 2 における L 字ホルダのモデルテンプレート作成時の視点数と各階層のテンプレート数

Table 2 The number of viewpoints and templates of L-Holder at each level.

	視点数	階層 1	階層 2	階層 3
Ulrich ら [9]	73,800	39,835	8,937	1,024
Ulrich ら (視点位置非依存)	73,800	39,245	8,010	435
提案手法	73,800	23,115	4,269	233

勢範囲において照合スコア低下が生じないようにテンプレート作成時に視点位置の設定を行っている。表 2 に、Ulrich らの手法と提案手法とで L 字ホルダのテンプレート作成を行った場合の、視点数と各階層のテンプレート数を示した。この表から Ulrich らの手法では、3 次元姿勢変化による照合スコアの低下を防ぐためにより多くのテンプレートを照合に用いていることが分かる。しかしながら一定姿勢範囲内を代表視点のみから作成したテンプレートで照合を行うため、対応している姿勢範囲周辺部分では多少なりとも認識正確性が低下すると考えられる。

提案手法において照合に用いるテンプレート数が少ない理由は二つあり、一つ目は PCOF テンプレートの対応可能な 3 次元姿勢範囲が広いこと、二つ目は視点位置に関係なくテンプレート間の類似度 (照合スコア) のみに基づいてクラスタリングを行っていることである。この二つ目の理由である類似度のみに基づいたクラスタリングを、Ulrich らの手法に導入したときのテンプレート数を表 2 中段に示した。この結果から、類似度のみに基づいたクラスタリングは階層 3 な

表 3 実験 2 における FPPI が 0.5 のときの物体毎及び全対象物体の平均認識処理時間 (ミリ秒)

Table 3 The averaged processing times (ms) for pose estimation when FPPI is 0.5 in experiment 2 are presented.

	Ulrich	LINE-2D	COF	PCOF
コネクタ	964.1	375.8	1258.5	<b>167.1</b>
サイドクランプ	2724.4	383.2	1387.5	<b>220.4</b>
ストッパ	2703.0	345.7	1149.9	<b>129.9</b>
L 字ホルダ	963.8	357.1	1015.8	<b>122.6</b>
T 字ホルダ	912.2	376.3	1140.1	<b>137.5</b>
フランジ	973.0	390.5	1238.1	<b>137.4</b>
ヒンジベース	1137.1	348.9	1124.6	<b>226.1</b>
ブラケット	792.4	358.5	961.4	<b>127.1</b>
支柱クランプ	1439.0	375.9	1320.1	<b>137.4</b>
平均	1401.0	368.0	1177.3	<b>156.2</b>

ど低解像度の階層においてテンプレート削減効果を発揮することが分かる。一方で階層 1 や階層 2 など高解像度の階層においては、一つ目の理由である PCOF テンプレートの 3 次元姿勢対応範囲の広さによるテンプレート数削減効果の方が優勢であると分かる。

#### 4.2.3 認識処理時間

FPPI が 0.5 のときの認識処理時間 (ミリ秒) を表 3 に示した。提案手法では二進数特徴量同士の照合を論理演算によって高速に行っている上、HPT によって少ない数のテンプレートで効率的に位置姿勢認識を行っており、既存手法と比較して 2-20 倍程度高速であるという結果が得られた。LINE-2D と COF も二進数特徴量の高速照合演算を用いているが、全ての視点について照合処理を行っており照合に用いるテンプレート数が多いため、提案手法と比較すると認識処理が遅

い、ただし LINE-2D は、照合時に参照する特徴量がメモリ上において連続するようあらかじめ配置を変えておき、メモリ参照のみで照合スコアを算出している。このため COF や Ulrich らの手法と比較すると認識処理は高速であった。

4.2.2 で述べたように、Ulrich らの手法で用いている正規化勾配ベクトル特徴量は一つのテンプレートで対応可能な 3 次元姿勢範囲が狭い上、近傍の視点のみ統合を行うために 3 次元位置姿勢認識に必要な各階層のテンプレート数が提案手法と比較して 2-4 倍程度多い。更に照合スコア演算においては、正規化勾配ベクトル特徴量は浮動小数点数の乗算が必要であるのに対して、提案手法では二進数特徴量同士の論理演算を用いて高速に実施可能である。このような理由から、Ulrich らの手法は提案手法と比較して認識処理速度が 5 倍以上遅かったと考えられる。

提案手法である HPT は 3 次元位置姿勢認識処理の高速化を目的としている。HPT を作成せずに 73,800 個の PCOF テンプレート全てを用いて照合処理を行った場合の認識成功率と処理時間を表 4 に示した。HPT を使用することによって認識処理がおよそ 1,000 倍高速になっているが、認識成功率が 3.5% 低下している。これは正解位置姿勢の上位層（低解像度階層）における照合スコアが低いために、それより下位層における探索が打ち切られたためだと考えられる。この精度低下に関しては、上位層における探索しきい値を調整することで軽減することが可能である。

#### 4.2.4 設定パラメータが認識性能に及ぼす影響

提案手法には、認識精度・速度に大きな影響を与えるパラメータとして PCOF 抽出時の視点位置変動範囲と HPT 作成時のテンプレート作成間隔の二つが挙げられる。ただし HPT では 3 次元姿勢の認識対象範囲をもれなくカバーする必要があるため、テンプレート作成間隔は PCOF の視点位置変動範囲に基づいて決定すべきである。つまり視点変動範囲が狭い場合はテンプレート作成間隔も狭くする必要があり、逆に視

点変動範囲が広い場合はテンプレート作成間隔を広くすることができる。

またこれらのパラメータは認識精度と速度のトレードオフを生む。具体的には視点位置変動範囲が狭い場合は照合される見えの変動範囲も小さく、複雑背景下においても誤認識が少ないため認識精度は向上するが、認識に必要なテンプレート数が増加するため処理速度が低下する。一方で変動範囲が広い場合は多様な見えとの照合を許容するために複雑背景下での誤認識が増加する危険性は高まるが、必要なテンプレート数は少ないため処理速度は向上する。

実験 2 のテンプレート作成間隔をおよそ 2 分の 1 と 2 倍にした場合の L 字ホルダの認識成功率と処理時間を表 5 に示した。PCOF の視点位置変動範囲はテンプレート作成間隔に応じて調整を行った。表の中段が実験 2 で用いたテンプレート作成間隔であるが、間隔を変化させることにより上で述べたような認識精度と速度との間に生じるトレードオフを確認できる。

#### 4.2.5 物体の形状が認識精度に及ぼす影響

PCOF ではテンプレート作成を行う代表視点以外の視点についても認識精度が低下しないよう、3 次元 CAD を用いて代表視点以外の 3 次元姿勢の見えを多数生成、統合して特徴抽出を行っている。このため代表視点からの見えのみを用いてテンプレートを作成する従来手法と比較した場合、わずかな視点移動で見えが大きく変わるような物体、例えばカメラ奥行き方向に高さをもつ物体や突起部をもつ物体に対して大きな認識精度向上が見込まれる。実験 2 で使用した 9 種類の物体の中では、高さがある物体としてヒンジベースやブラケット、突起部をもつ物体としてサイドクランプやストッパが挙げられるが、ヒンジベースを除く三つの物体では従来手法と比較して大きな精度向上が確認できた（図 8）。

一方でヒンジベースでは PCOF による精度向上幅が小さかったが、これは他の物体と比較して表面の光沢度が高く、表面光沢によって生じる反射や映り込み

表 4 HPT 有り、無しそれぞれの場合の L 字ホルダの認識成功率 (FPPI = 0.5) 及び処理時間 (ミリ秒)

Table 4 Success rate (FPPI = 0.5) and processing time (ms) for L-Holder with and without HPT.

	認識成功率	処理時間 (ms)
HPT 有り	0.942	122.6
HPT 無し	0.977	126727.7

表 5 テンプレート作成間隔を変えた場合の L 字ホルダの認識成功率 (FPPI = 0.5) 及び処理時間 (ミリ秒)

Table 5 Success rate (FPPI = 0.5) and processing time (ms) for L-Holder using three different numbers of viewpoints.

間隔 (視点数)	認識成功率	処理時間 (ms)
約 15 度 (54 点)	0.803	76.7
約 8 度 (205 点)	0.942	122.6
約 4 度 (835 点)	0.963	151.2



図 9 カメラから物体までの距離を変え、L 字ホルダが画像の左上端に映るように 3 次元 CAD の投影を行った画像。左からカメラと物体の距離が 680mm, 340mm, 170mm

Fig.9 2D projection images of L-Holder which are rendered at the left-upper part of the images from 3 different distances (left: 680mm, center: 340mm, right: 170mm) are presented.

によって特徴量照合の誤りが多く発生したためだと考えられる。このように認識精度は物体表面の材質によっても大きく左右されるため、実験 2 の結果を対象物体の形状だけから論じることは難しい。

本研究では L 字ホルダを用いて PCOF 作成時の視点位置変動範囲及び HPT のテンプレート作成間隔を決定したが、これらの値を物体ごとに最適化することも可能である。その場合、高さのある物体や突起部をもつ物体は視点をわずかに移動させただけで見えが大きく変わる可能性があるため、そうでない物体と比較して最適な変動範囲及び作成間隔は狭いと考えられる。

#### 4.3 射影ひずみが認識精度に及ぼす影響

PCOF 抽出時に 3 次元 CAD から 2 次元投影画像を生成する際、カメラの光軸は常に物体中心に向けられている。このため生成される画像では常に物体が画像中心にあり、物体が画像端にある場合の見え方については考慮されていない。このため画像端に映った物体に射影ひずみが生じた場合、PCOF では 3 次元位置姿勢を正しく認識できない可能性がある。射影ひずみが認識精度に与える影響を検証するため、カメラから物体までの距離を変えて物体が画像の左上端に描画されるよう投影を行い、図 9 に示した。

本論文の実験設定ではカメラと物体の距離は 680mm であり、この場合画像端での物体の見え方は画像中心での見え方とほとんど変わらない（図 9 左）。しかし半分の距離（340mm）から見た場合は物体の左上部分に微小なひずみが認められ（図 9 中央）、更に半分の距離（170mm）では大きなひずみが現れることが分かる（図 9 右）。このことから焦点距離が短いカメラを用いて対象物体を近くから撮影した場合、画像の端に映った物体はひずんでしまい PCOF では 3 次元位置姿勢を認識できないと考えられる。ただし本論文の実験設定を用いた場合、画像端に映った物体の射影ひず



図 10 画像端に対象物体が映っている場合の提案手法による 3 次元位置姿勢認識結果例

Fig.10 Example images of 6D pose estimation results when the objects are around the corners of images.

みは無視できる程度であり、図 10 に示したように提案手法を用いて物体の 3 次元位置姿勢を正しく認識することが可能である。

なお物体とカメラの距離が近く射影ひずみによる影響が無視できない場合においても、画像の領域ごとに PCOF テンプレートと HPT を作成し、認識時には画像領域に応じて照合に用いる HPT を切り替えることで認識精度や速度を損なうことなく射影ひずみに対応可能であると考えられる。

## 5. む す び

本研究では、3 次元姿勢変化による対象物体の見えの変化に対する許容性と複雑背景に対する頑健性とを両立させた透視投影に基づく累積勾配方向特徴量 (PCOF) を提案し、実画像を用いた照合実験によってその有効性を確認した。更にテンプレート間の類似度のみに基づく視点位置に依存しない階層的姿勢探索木 (HPT) も併せて提案し、独自に取得した 9 種類のテクスチャレス物体画像データセットを用いて性能評価実験を行った結果、PCOF と HPT を組み合わせた提案手法が既存手法と比較してより高速・高精度に物体の 3 次元位置姿勢を認識できることを示した。

## 文 献

- [1] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G.R. Bradski, K. Konolige, and N. Navab, "Model based training, detection and pose estimation of textureless 3D objects in heavily cluttered scenes," Proc. Asian Conf. Comput. Vision, pp.548–562, 2012.
- [2] A. Tejani, D. Tang, R. Kouskouridas, and T.K. Kim, "Latent-class Hough forests for 3D object detection and pose estimation," Proc. European Conf. Comput. Vision, pp.462–477, 2014.
- [3] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol.60, no.2, pp.91–110, 2004.
- [4] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded up robust features (SURF)," Comput. Vis. Image Understand., vol.110, no.3, pp.346–359, 2008.



- [5] W. Grimson and B. Huttenlocher, "On the verification of hypothesized matches in model-based recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.13, no.12, pp.1201–1213, 1991.
- [6] S. Lanser, O. Munkelt, and C. Zierl, "Robust video-based object recognition using CAD models," *Intelligent Autonomous Systems IAS-4*, pp.529–536, 1995.
- [7] C.M. Cyr and B.B. Kimia, "A similarity-based aspect-graph approach to 3D object recognition," *Int. J. Comput. Vis.*, vol.57, no.1, pp.5–22, 2004.
- [8] M.Y. Liu, O. Tuzel, A. Veeraraghavan, Y. Taguchi, T. Marks, and R. Chellappa, "Fast object localization and pose estimation in heavy clutter for robotic bin picking," *Int. J. Robot. Res.*, vol.31, no.8, pp.951–973, 2012.
- [9] M. Ulrich, C. Wiedemann, and C. Steger, "Combining scale-space and similarity-based aspect graphs for fast 3D object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.34, no.10, pp.1902–1914, 2012.
- [10] S. Hinterstoisser, C. Cagniard, S. Ilic, P. Sturm, N. Navab, P. Fua, and V. Lepetit, "Gradient response maps for real-time detection of textureless objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.34, no.5, pp.876–888, 2012.
- [11] D.G. Lowe, "Local feature view clustering for 3D object recognition," *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, vol.1, pp.682–688, 2001.
- [12] V. Lepetit, J. Pilet, and P. Fua, "Point matching as a classification problem for fast and robust object pose estimation," *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, vol.2, pp.244–250, 2004.
- [13] H. Najafi, Y. Genc, and N. Navab, "Fusion of 3D and appearance models for fast object detection and pose estimation," *Proc. Asian Conf. Comput. Vision*, pp.415–426, 2006.
- [14] P. David and D. DeMenthon, "Object recognition in high clutter images using line features," *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, vol.2, pp.1581–1588, 2005.
- [15] D. Damen, P. Bunnun, A. Calway, and W. Mayol-cuevas, "Real-time learning and detection of 3D texture-less objects: A scalable approach," *Proc. British Mach. Vision Conf.*, 2012.
- [16] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3D object recognition," *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp.998–1005, 2010.
- [17] J. Rodrigues, J.S. Kim, M. Furukawa, J. Xavier, P. Aguiar, and T. Kanade, "6D pose estimation of textureless shiny objects using random ferns for bin-picking," *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pp.3334–3341, 2012.
- [18] E. Brachmann, A. Krull, F. Michel, S. Gumhold, J. Shotton, and C. Rother, "Learning 6D object pose estimation using 3D object coordinates," *Proc. European Conf. Comput. Vision*, pp.536–551, 2014.
- [19] E. Brachmann, F. Michel, A. Krull, M.Y. Yang, S. Gumhold, and C. Rother, "Uncertainty-driven 6D pose estimation of objects and scenes from a single RGB image," *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016.
- [20] P. Wohlhart and V. Lepetit, "Learning descriptors for object recognition and 3D pose estimation," *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp.3109–3118, 2015.
- [21] A. Crivellaro, M. Rad, Y. Verdie, K.M. Yi, P. Fua, and V. Lepetit, "A novel representation of parts for accurate 3D object detection and tracking in monocular images," *Proc. IEEE Int. Conf. Comput. Vision*, pp.4391–4399, 2015.
- [22] A. Krull, E. Brachmann, F. Michel, M.Y. Yang, S. Gumhold, and C. Rother, "Learning analysis-by-synthesis for 6D pose estimation in RGB-D images," *Proc. IEEE Int. Conf. Comput. Vision*, pp.954–962, 2015.
- [23] M. Zhu, K. Derpanis, Y. Yang, S. Brahmabhatt, M. Zhang, C. Phillips, M. Lecce, and K. Daniilidis, "Single image 3D object detection and pose estimation for grasping," *Proc. IEEE Int. Conf. Robot. Automat.*, pp.3936–3943, 2014.
- [24] R. Rios-Cabrera and T. Tuytelaars, "Discriminatively trained templates for 3D object detection: A real time scalable approach," *Proc. IEEE Int. Conf. Comput. Vision*, pp.2048–2055, 2013.
- [25] W. Kehl, F. Tombari, N. Navab, S. Ilic, and V. Lepetit, "Hashmod: A hashing method for scalable 3D object detection," *Proc. British Mach. Vision Conf.*, 2015.
- [26] T. Hodan, X. Zabulis, M. Lourakis, S. Obdrzalek, and J. Matas, "Detection and fine 3D pose estimation of texture-less objects in RGB-D images," *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pp.4421–4428, 2015.
- [27] C. Steger, "Occlusion, clutter, and illumination invariant object recognition," *Int. Archives Photogrammetry Remote Sensing*, vol.34, pp.345–350, 2002.
- [28] F. Ullah and S. Kaneko, "Using orientation codes for rotation-invariant template matching," *Pattern Recognit.*, vol.37, no.2, pp.201–209, 2004.
- [29] S. Hinterstoisser, V. Lepetit, S. Ilic, P. Fua, and N. Navab, "Dominant orientation templates for real-time detection of texture-less objects," *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp.2257–2264, 2010.
- [30] 小西嘉典, 井尻善久, 川出雅人, 橋本 学, "累積勾配方向特徴量を用いたテクスチャレス物体検出," *信学論 (D)*, vol.J99-D, no.8, pp.689–698, Aug. 2016.
- [31] D. Nister and H. Stewenius, "Scalable recognition

- with a vocabulary tree,” Proc. IEEE Conf. Comput. Vision Pattern Recognit., pp.2161–2168, 2006.
- [32] C. Silpa-Anan and R. Hartley, “Optimised kd-trees for fast image descriptor matching,” Proc. IEEE Conf. Comput. Vision Pattern Recognit., pp.1–8, 2008.
- [33] M. Muja and D. Lowe, “Scalable nearest neighbor algorithms for high dimensional data,” IEEE Trans. Pattern Anal. Mach. Intell., vol.36, no.11, pp.2227–2240, 2014.
- [34] K. Lai, L. Bo, X. Ren, and D. Fox, “A scalable tree-based approach for joint object and pose recognition,” Proc. AAAI Conf. Artificial Intell., pp.1474–1480, 2011.
- [35] D.M. Gavrila, “A bayesian, exemplar-based approach to hierarchical shape matching,” IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.8, pp.1408–1421, 2007.
- [36] B. Stenger, A. Thayananthan, P.H.S. Torr, and R. Cipolla, “Hand pose estimation using hierarchical detection,” Proc. European Conf. Comput. Vision Workshop on HCI, pp.105–116, 2004.
- [37] G. Borgefors, “Hierarchical chamfer matching: A parametric edge matching algorithm,” IEEE Trans. Pattern Anal. Mach. Intell., vol.10, no.6, pp.849–865, 1988.
- [38] 小西嘉典, 半澤雄希, 川出雅人, 橋本 学, “階層的統合モデルを用いた単眼カメラからの高速 3 次元位置・姿勢認識,” 映像情報インダストリアル 3 月号, pp.81–86, 2016.
- [39] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” Proc. IEEE Conf. Comput. Vision Pattern Recognit., pp.886–893, 2005.
- [40] D. Pelleg and A. Moore, “X-means: Extending k-means with efficient estimation of the number of clusters,” Proc. Int. Conf. Mach. Learning, pp.727–734, 2000.
- [41] R.I. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Second ed., Cambridge University Press, 2004.
- [42] S. Garrido-Jurado, R. Muñoz Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez, “Automatic generation and detection of highly reliable fiducial markers under occlusion,” Pattern Recognit., vol.47, no.6, pp.2280–2292, 2014.

(平成 28 年 9 月 29 日受付, 29 年 2 月 3 日再受付,  
5 月 2 日早期公開)



小西 嘉典 (正員)

2000 年京都大学総合人間学部卒, 2002 年京都大学人間・環境学研究科博士前期課程修了, 2005 年京都大学人間・環境学研究科博士後期課程指導認定退学, 2007 年オムロン(株)入社, 顔画像認識技術, 映像監視技術, 産業用視覚センサの研究開発に従事, 2014 年度画像センシングシンポジウム優秀学術賞, 2015 年度ビジョン技術の実利用ワークショップ小田原賞受賞.



半澤 雄希

2010 年東北大学大学院情報科学研究科博士前期課程修了, 同年オムロン(株)入社, 顔画像認識技術, 産業用視覚センサ, 車載画像処理の研究開発に従事, 2015 年度ビジョン技術の実利用ワークショップ小田原賞受賞.



川出 雅人 (正員)

1981 年名古屋工業大学・計測工学科卒, 同年立石電機(株)(現オムロン)入社, 1990 年–1994 年米・カーネギーメロン大学計算機科学科・訪問研究員, 人・顔画像処理, FA 画像処理, 車載・社会インフラ画像処理の研究・開発に従事, 2009 年画像センシングシンポジウム高木賞受賞, 本会及び IEEE 各会員.



橋本 学 (正員)

1987 年大阪大学大学院工学研究科(溶接工学専攻)前期課程修了, 同年三菱電機(株)入社, 生産システム研究所, 先端技術総合研究所等にてロボットビジョン, パターン認識, ヒューマン認識などの研究に従事, 2008 年より中京大学情報理工学部教授, 2013 年より同工学部教授, 1998 年度日本ロボット学会実用化技術賞, 1999 年科学技術庁注目発明表彰, 2012 年度画像センシングシンポジウム優秀学術賞, 2015 年度 MIRU インタラクティブ発表賞, ビジョン技術の実利用ワークショップ小田原賞受賞.