

# 対応点集合類似度学習を用いた剛体・非剛体物体検出

金崎 朝子<sup>†</sup> Emanuele Rodolà<sup>††</sup> Daniel Cremers<sup>††</sup> 原田 達也<sup>†</sup>

<sup>†</sup> 東京大学大学院情報理工学系研究科 〒113-8656 東京都文京区本郷 7-3-1

<sup>††</sup> Computer Vision Group, TU Munich, Germany

E-mail: <sup>†</sup>{kanezaki,harada}@mi.t.u-tokyo.ac.jp, <sup>††</sup>emanuele.rodola@in.tum.de, <sup>†††</sup>cremers@tum.de

あらまし 入力三次元点群と検出対象物体の三次元点群との対応点を求め、類似度行列の二次割当問題を解いて誤対応点除去された対応点集合に対して、その正当性を評価する対応点集合類似度を導入し、パラメータを学習する手法を提案した。提案手法は、対応点集合類似度が、正解の対応点集合は高く、不正解の対応点集合は低くなるよう、二組の対応点ペアに対する重みを学習する。剛体物体検出実験および非剛体物体検出実験において、提案手法の有効性を確かめた。

キーワード 特徴点マッチング, グラフマッチング, 最適化, 勾配法, 三次元形状

## Learning Similarities for Rigid and Non-Rigid Object Detection

Asako KANEZAKI<sup>†</sup>, Emanuele RODOLÀ<sup>††</sup>, Daniel CREMERS<sup>††</sup>, and Tatsuya HARADA<sup>†</sup>

<sup>†</sup> Graduate School of Information Science and Technology, The University of Tokyo, Japan

<sup>††</sup> Computer Vision Group, TU Munich, Germany

E-mail: <sup>†</sup>{kanezaki,harada}@mi.t.u-tokyo.ac.jp, <sup>††</sup>emanuele.rodola@in.tum.de, <sup>†††</sup>cremers@tum.de

**Abstract** We propose an optimization method for estimating parameters in graph-theoretical formulations of the matching problem for object detection. Unlike several methods which optimize parameters for graph matching in a way to promote correct correspondences and to restrict wrong ones, our approach aims at improving performance in the more general task of object detection. In our formulation, similarity functions are adjusted so as to increase the overall similarity among a reference model and the observed target, and at the same time reduce the similarity among reference and "non-target" objects. We evaluate the proposed method in two challenging scenarios, demonstrating substantial improvements in both settings.

**Key words** keypoint matching, graph matching, optimization, gradient descent method, 3D shape

### 1. はじめに

知能システムが人間のように実世界を“見る”ことの実現は、古くからのコンピュータビジョン分野の大目標であった。近年ではコンシューマデブスセンサから環境のリアリスティックな三次元データを計測することが可能であるなど、身の回りの知能システムが豊富な視覚情報を容易に取得可能となってきている。しかしながら、これらの知能システムが実世界を“見る”-何がどこにあるかを知る-ことは依然として難解な課題である。

本稿では、物体検出に用いるスコア関数の最適化手法を提案する。物体検出の手段として、検出対象物体の参照形状データと実環境シーンの三次元データとの間で、点対点のマッチングを行う。このような方法は、画像内の各検出領域からの大域特徴を用いてスライディングボックス等を行う従来の物体検出と比較して、物体の姿勢変化や隠れに対してロバストである。参照形状データと環境の三次元データとの間で得られた対応点集

合の中から誤対応点を除去する方法として、対応点ペアの類似度の総和を評価する、いわゆるグラフマッチングの問題を解く。

具体的には、 $(X, d_X)$  と  $(Y, d_Y)$  を二つの距離空間（ただし  $X, Y \subset \mathbb{R}^m$ ）とし、 $C \subset X \times Y$  をこれらの間の対応点集合とすると、解くべきグラフマッチングの問題は下記の  $L_p$  正則化二次割当問題（Quadratic Assignment Problem, QAP）で表せる。

$$\max_x x^T A x \quad \text{s.t. } \|x\|_p = 1. \quad (1)$$

ただし  $x \in [0, 1]^{|C|}$  は各対応点の割り当てを表すベクトルであり、 $A$  は対応点ペア間の類似度を表す  $|C| \times |C|$  の対称行列である。端的に言えば、上式は二つの距離空間を比較したときの総合的な類似度を最大化する問題である。このような QAP 定式化は形状マッチングおよびグラフマッチング境界でよく知られており、特に  $L_p$  緩和は様々な場面で活用される。例えば、Leordeanu and Hebert [1] は  $p = 2$ , Rodolà et al. [2] は  $p = 1$  を使用した手法を提案した。より最近の手法として、 $p = 1$  と

$p = 2$  を混合したものも提案されている [3, 4] .  $A$  はある類似度関数  $\pi : C \times C \rightarrow [0, 1]$  を行列化したものであるが、アプリケーションによって様々な  $\pi$  を設計することができる . 例えば、変形物体を扱う場合はデータの非剛体変換 (例えば姿勢変化等) に対してなるべく不変な類似度関数を使うべきであり、この場合は、ユークリッド空間への形状の埋め込みに依らず一定な、例えば測地線距離等の内在的計量を扱う必要がある .

さて、物体検出を行うためには、検出対象の物体らしさを表すスコア関数を学習する必要がある . ここで、我々は幾何的な性質と外観の性質との中から、各物体の検出に特に有効となるものを高く評価し、活用する方法に焦点を当てる . すなわち、対応点ペアを評価する複数の類似度関数を用意し、これらに対する重みパラメータを物体毎に学習する手法を提案する .

本研究と同様に対応点集合の類似度を最大化する問題を QAP 定式化として解き、そのパラメータを学習した研究例は他にも多数存在する [5–9] . 例えば Caetano et al. [5] は QAP における正解の割当ベクトル  $x$  と推定された割当ベクトル  $\hat{x}$  の間のハミング損失を定義し、これを最小化することで、60 次元の Shape Context 記述子 [10] に対する重みを学習した . また、Leordeanu et al. [7] は教師なし学習の枠組みで、距離や角度等の計 7 次元の類似度に対する重みを学習した . これらに対し、本研究は点集合としての物体検出に用いるスコアの最適化という観点からパラメータを学習する点が新しい .

## 2. 手 法

本研究の目的は、ある観測された物体  $X$  と検出対象物体  $Y$  との類似度を表すスコア関数  $g : X \times Y \rightarrow \mathbb{R}$  のパラメータを学習することである . 提案手法は、学習サンプルとしてある形状データを観測する毎にパラメータを更新するオンライン学習のアプローチを取り、スコア関数の出力が、正解の対応点集合は高く、不正解の対応点集合は低くなるよう、二組の対応点ペアに対する重みを学習する . 以下、QAP を解き対応点集合を獲得することによる学習サンプルの準備、物体らしさのスコア  $g$  の定義、そしてパラメータの学習則について説明する .

### 2.1 学習サンプル

本研究では、類似度関数  $\pi$  を下記の形式で定義する .

$$\pi(c_i, c_j) = \exp[-s(c_i, c_j)]. \quad (2)$$

ここで、 $s : C \times C \rightarrow \mathbb{R}$  は二つの候補対応点ペア  $c_i$  と  $c_j$  の合致度を表す関数である .

学習用形状データとして、対象物体の参照用形状データ  $R$  と、対象物体の観測形状データ  $\{O_i^{(pos)}\}_{i=1, \dots, N}$ 、対象物体以外の観測形状データ  $\{O_j^{(neg)}\}_{j=1, \dots, N'}$  を用意する . はじめに、形状データ内の各点のプロパティとして何らかの局所記述子を計算し、これを局所記述子集合とする . 次に、対象物体のモデル形状データからサンプリングした  $K$  個の点について、観測形状データから局所記述子空間における  $k_{nn}$  個の最近傍点を探索する . こうして、各形状データのペア  $(R, O_i^{(pos)})$  と  $(R, O_j^{(neg)})$  に対して  $k_{nn}K$  個の対応点集合候補を得る . 対応点集合候補の中には、正解の対応点ペアと誤対応点ペアとが含まれている .

そこで、式 (1) の QAP を解いてある閾値以下となった  $x$  の要素に対応する対応点ペアを除去することで、誤対応点ペアの除去を行う . 本研究では特に、 $L_1$  ノルム正則化  $\|x\|_1 = 1$  を採用する手法 [2] を使用する . 本手法によれば、スパースではあるが正確な対応点集合を得ることができる . 以上により、 $R$  と  $O_i^{(pos)}$  ( $i = 1, \dots, N$ )、 $R$  と  $O_j^{(neg)}$  ( $j = 1, \dots, N'$ ) に関する選抜された対応点集合が得られる .

候補対応点ペアの合致度を表す関数  $s$  として、任意のものを使用することに注意されたい . 本稿では、剛体物体検出タスクにおいて、ユークリッド距離  $d_E(a, b)$  を用いて対応点ペア  $(a, b) \in X \times X$  の合致度を下記のとおりに定義する .

$$s_r((a, a'), (b, b')) \equiv |d_E(a, b) - d_E(a', b')|. \quad (3)$$

ただし  $(a, a'), (b, b') \in C$  である . 剛体物体の変換はユークリッド距離を保持するため、正解の対応点ペアについての出力は 0 になるとみせる . 同様に、非剛体物体検出タスクにおいては、内在的距離 (intrinsic metrics) を用いる . 具体的には、multi-scale diffusion (MD) metric [11]  $d_M(a, b)$  と commute-time (CT) metric [12]  $d_C(a, b)$  とを使用して、対応点ペア  $(a, b) \in X \times X$  の合致度を下記のとおりに定義する .

$$s_n((a, a'), (b, b')) \equiv \frac{s_M((a, a'), (b, b')) + s_C((a, a'), (b, b'))}{2}. \quad (4)$$

ただし  $s_M$  と  $s_C$  は式 (3) をそれぞれに対応する距離で置き換えたものである .

### 2.2 スコア関数

対応点集合類似度として、最も単純には、選抜されて (誤対応点除去を行って) 最終的に残った対応点集合中の全対応点ペアの類似度  $\pi(c_i, c_j)$  の総和を (残った) 対応点の総数  $M$  で正規化した下記の値が使用できる .

$$g_{<base>} \equiv \frac{2}{M(M-1)} \sum_{i=1}^M \sum_{j=i+1}^M \exp[-s(c_i, c_j)]. \quad (5)$$

本稿ではこれをベースラインのスコア関数と呼ぶ . この対応点集合類似度は各対応点ペアの類似度を平等に評価するが、対応点の持つ別のプロパティ (例えば色) によって区別して重み付けしたり、また、異なる類似度関数を統合する際の重みを適切に調整したりすることで、対応点集合類似度を改良することができる . 以下、我々は第一に Kinect センサ等で撮影した色つきの剛体物体検出 (RGBD) を行う場合と、第二に非剛体物体検出を行う場合の二つの場面を考え、それぞれについて対応点集合類似度を設計する .

#### 2.2.1 剛体物体検出 (RGBD)

剛体物体検出では、候補対応点ペアの合致度を表す関数としてユークリッド距離を用いるのが最も効果的であることは述べた . しかしながら、形状情報の他に色等の他の性質の情報が活用できる場合は、その情報によって対応点を区別し、検出対象物体にとって特に特徴的な対応点を高く評価する (すなわち、環境中にありふれた対応点を無視する) ことでより識別的なスコア関数を得ることができる . 提案手法は、対応点集合類似度が、正解の対応点集合は高く、不正解の対応点

集合は低くなるよう、二組の対応点ペアに対する重みを学習する。具体的には、まず Hue 値を  $k$  個のビンに分割し、参照画像中の各点の Hue 値がどのビンに所属するかを示すベクトルを  $\mathbf{h} \in \mathbb{R}^k$  とする。所属するビンの値は 1、それ以外のビンの値は 0 とする。そして、参照画像中の点  $i$  のベクトル  $\mathbf{h}_i$  と点  $j$  のベクトル  $\mathbf{h}_j$  について行列  $H \equiv \mathbf{h}_i \mathbf{h}_j^T$  を求め、さらに  $H'_{mn} = H_{mn} + H_{nm} (m \neq n)$ ,  $H'_{nn} = H_{nn}$  となる行列  $H'$  を計算する。そして  $H'$  の上三角成分 (対角成分を含む) を並べたベクトルを  $\mathbf{q}_{ij} \in \mathbb{R}^{k(k+1)/2}$  とする。本研究の目的は、このベクトル  $\mathbf{q}_{ij}$  に対する重み  $\mathbf{w} \in \mathbb{R}^{k(k+1)/2}$  を物体毎に最適化することである。提案する対応点集合類似度  $g_r(\mathbf{w})$  を下式で定義する。

$$g_r(\mathbf{w}) \equiv \frac{2}{M(M-1)} \sum_{i=1}^M \sum_{j=i+1}^M \left( 1 - \exp \left[ \frac{-\alpha \cdot \mathbf{w} \cdot \mathbf{q}_{ij}}{s_r(c_i, c_j) + \epsilon} \right] \right). \quad (6)$$

ここで、 $\alpha > 0$  は指数関数の出力が 0 から 1 に張り付かないよう調整する定数であり、 $\epsilon$  は指数関数の中身の分母が 0 になることを防ぐための小さな値の定数である。本研究では  $\alpha = 10^{-3}$ ,  $\epsilon = 10^{-20}$  とした。

### 2.2.2 非剛体物体検出

非剛体物体検出における重要な課題は、ある種類の変形に対する不変性を保持する適切な距離関数を如何に設計し、マッチング問題を定式化するかにある。例えば、測地線距離は等尺変形にはほぼ不変であるが、トポロジーの変化には大きく影響を受けるという問題があり、一方で、commute-time metrics [12] はトポロジー変化や大域的なスケール変化にはロバストであるが、局所的なスケール変化に対しては影響を受けがちである。本研究では、複数種類の内在的距離 (intrinsic metrics) の重み付き和を使用し、その重みを異なる変形のタイプ毎に適切に学習することを目的とする。ここでは特に、よく知られる内在的距離である multi-scale diffusion (MD) metric [11] と commute-time (CT) metric を使用する。これらの統合重みを  $\mathbf{w} \equiv [w_M \ w_C]^T \in \mathbb{R}^2$  とし、 $\mathbf{s}(c_i, c_j) \equiv [s_M(c_i, c_j) \ s_C(c_i, c_j)]^T \in \mathbb{R}^2$  とし、提案する対応点集合類似度  $g_n(\mathbf{w})$  を下式で定義する。

$$g_n(\mathbf{w}) \equiv \frac{2}{M(M-1)} \sum_{i=1}^M \sum_{j=i+1}^M \exp[-\mathbf{w} \cdot \mathbf{s}(c_i, c_j)]. \quad (7)$$

### 2.3 パラメータ学習

提案手法は学習データセットに対して、正解の対応点集合の  $g(\mathbf{w})$  が大きく、不正解の対応点集合の  $g(\mathbf{w})$  が小さくなるような  $\mathbf{w}$  を決定する (式 (6) と式 (7))。このために、SVM 等の識別器学習と同様に学習サンプルを適切に分離する超平面を考え、そのマージン内に学習サンプルが入ってしまう損失の量であるヒンジロスを最小化させることを考える。このためにはスコア関数の値域が  $-\infty$  から  $+\infty$  までの値を取る必要がある。そこで、本研究では対応点集合の正解らしさを評価するスコア関数を下式で定義する。

$$\begin{aligned} f(\mathbf{w}, b) &\equiv \text{logit}(g(\mathbf{w})) + b \\ &= \log(g(\mathbf{w})) - \log(1 - g(\mathbf{w})) + b. \end{aligned} \quad (8)$$

ただし  $b$  はオフセットであり、 $\mathbf{w}$  と共に学習する変数である。ここで正解対応点集合のラベルを  $y = 1$ 、不正解対応点集合のラベルを  $y = -1$  とすると、ヒンジロス関数は下記のとおりとなる。

$$l(\mathbf{w}, b; (f, y)) = \begin{cases} 0 & yf(\mathbf{w}, b) \geq 1, \\ 1 - yf(\mathbf{w}, b) & \text{otherwise.} \end{cases} \quad (9)$$

提案手法は  $\mathbf{w}$  の初期値を零ベクトルに近いベクトル  $\mathbf{w}_0 = (\epsilon', \dots, \epsilon')$ ,  $\epsilon' \sim 0$ ,  $b$  の初期値を  $b_0 = 0$  とし、対応点集合を一つ観測する度にこれらを更新する逐次学習を行う。  $t$  回めの更新後のパラメータを  $\mathbf{w}_t, b_t$  とすると、 $\mathbf{w}_{t+1}, b_{t+1}$  は下式で計算される。

$$\{\mathbf{w}_{t+1}, b_{t+1}\} = \arg \min_{\mathbf{w}, b} \frac{1}{2} (\|\mathbf{w} - \mathbf{w}_t\|^2 + \|b - b_t\|^2) \quad (10)$$

$$\text{s.t. } l(\mathbf{w}, b; (f_t, y_t)) = 0. \quad (11)$$

本式は閉形式である。解法を下記に示す。まず、 $y_t f_t(\mathbf{w}_t, b_t) \geq 1$  であれば、 $\mathbf{w}_{t+1} = \mathbf{w}_t, b_{t+1} = b_t$  となる。このため、以下、 $y_t f_t(\mathbf{w}_t, b_t) < 1$  の場合のみを考える。ラグランジュの未定乗数法を用いて

$$\begin{aligned} L(\mathbf{w}_t, b_t, \lambda) &= \frac{1}{2} \|\mathbf{w}_{t+1} - \mathbf{w}_t\|^2 + \frac{1}{2} \|b_{t+1} - b_t\|^2 \\ &\quad + \lambda(1 - y_t f_t(\mathbf{w}_t, b_t)). \end{aligned} \quad (12)$$

$\frac{\partial L}{\partial \mathbf{w}_t} = 0, \frac{\partial L}{\partial b_t} = 0$  より、下式が得られる。

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \lambda y_t \frac{\partial f_t(\mathbf{w}_t, b_t)}{\partial \mathbf{w}_t}, \quad (13)$$

$$b_{t+1} = b_t - \lambda y_t. \quad (14)$$

ここでは  $\frac{\partial}{\partial b_t} f_t(\mathbf{w}_t, b_t) = 1$  であることに注意されたい。上式を式 (12) に代入し、下式が得られる。

$$L(\lambda) = \frac{1}{2} \lambda^2 \left\| \frac{\partial f_t(\mathbf{w}_t, b_t)}{\partial \mathbf{w}_t} \right\|^2 + \frac{1}{2} \lambda^2 + \lambda(1 - y_t f_t(\mathbf{w}_t, b_t)). \quad (15)$$

ここで  $\frac{\partial L}{\partial \lambda} = 0$  より、下式が得られる。

$$\lambda = \frac{y_t f_t(\mathbf{w}_t, b_t) - 1}{\left\| \frac{\partial f_t(\mathbf{w}_t, b_t)}{\partial \mathbf{w}_t} \right\|^2 + 1}. \quad (16)$$

したがって、 $\mathbf{w}_{t+1}$  と  $b_{t+1}$  の更新式は下式となる。

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \frac{(y_t - f_t(\mathbf{w}_t, b_t))}{\left\| \frac{\partial f_t(\mathbf{w}_t, b_t)}{\partial \mathbf{w}_t} \right\|^2 + 1} \cdot \frac{\partial f_t(\mathbf{w}_t, b_t)}{\partial \mathbf{w}_t}, \quad (17)$$

$$b_{t+1} = b_t + \frac{(y_t - f_t(\mathbf{w}_t, b_t))}{\left\| \frac{\partial f_t(\mathbf{w}_t, b_t)}{\partial \mathbf{w}_t} \right\|^2 + 1}. \quad (18)$$

$f(\mathbf{w}, b)$  の勾配は下式のとおり計算できる。

$$\begin{aligned} \frac{\partial f(\mathbf{w}, b)}{\partial \mathbf{w}} &= \frac{\partial}{\partial \mathbf{w}} (\log(g(\mathbf{w})) - \log(1 - g(\mathbf{w}))) \\ &= \frac{1}{g(\mathbf{w})(1 - g(\mathbf{w}))} \frac{\partial g(\mathbf{w})}{\partial \mathbf{w}}. \end{aligned} \quad (19)$$

ここで、 $g_r(\mathbf{w})$  の勾配は下式である。

$$\frac{2}{M(M-1)} \sum_{i=1}^M \sum_{j=i+1}^M \left( \frac{\alpha \cdot \mathbf{q}_{ij}}{s_E(c_i, c_j) + \epsilon} \cdot \exp \left[ \frac{-\alpha \cdot \mathbf{w} \cdot \mathbf{q}_{ij}}{s_E(c_i, c_j) + \epsilon} \right] \right). \quad (20)$$

また,  $g_n(\mathbf{w})$  の勾配は下式である .

$$-\frac{2}{M(M-1)} \sum_{i=1}^M \sum_{j=i+1}^M \mathbf{s}(c_i, c_j) \cdot \exp[-\mathbf{w} \cdot \mathbf{s}(c_i, c_j)]. \quad (21)$$

以上の更新を全学習サンプルについて行う操作を 1 試行とし,  $T$  試行を行った後に学習を終了させる . なお, 途中でヒンジロスが十分に小さくなった場合はそこで学習を打ちきっても良い .

提案手法のパラメータ更新則は線形識別器のオンライン学習手法である Passive Aggressive (PA) [13] と関連が深い . PA はサンプル特徴量  $\mathbf{x}$  に対するスコア  $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$  に対するヒンジ損失を最小にする勾配を求め,  $\mathbf{w}$  のノルムを最小化する制約でパラメータを更新するものである . 実際のアルゴリズムとしては, 本研究の提案手法は上記の線形なスコア関数  $f$  を式 (8) で置き換えたものに等しい .

### 3. 実 験

#### 3.1 剛体物体検出 (RGBD)

Kinect センサで撮影した RGB-D 画像を用いて 10 個の物体の検出性能を検証する . 学習データセットとして, 各対象物体につき一つの RGB-D 参照画像と, 少し視点をずらして撮影した 9 枚の RGB-D 画像を用意し (図 1), 正解の対応点集合を求めた . また, 対象物体が写っていない RGB-D 画像を各対象物体毎に 70 枚用意し, 不正解の対応点集合を求めた . 不正解の対応点集合の一例を図 2 (a) に示す . 対応点集合は, 参照画像からランダムに 2,000 点を選択し, 各点について各学習用画像から RGB 値が最近傍となる 5 個の点を求め (計 10,000 対応点), [2] の手法で絞り込みを行うことで求めた . ここで, 最後に得られた割当ベクトル  $\mathbf{x}$  の二値化を行うが, その閾値は  $\mathbf{x}$  の最大要素に 0.5 をかけた値とした . 比較手法として, 参照画像と各学習画像から SIFT 特徴点を求め, 総当たり法でマッチングを行った後に [2] の手法で絞り込みを行うものも試す . なお, Hue 値の分割数を  $k = 3$  とした ( $k = 2, 4, 5$  も試したが, 結果は大きく変わらなかった .) また, 学習試行回数の上限  $T$  は 100 とした .

テスト RGB-D 画像 120 枚に対して, 同様に対応点集合を求め, 性能を評価した . 得られた正解の対応点集合の例を図 2 (b) に, 不正解の対応点集合の例を図 2 (c) に示す . テスト画像の背景は学習画像と異なり, 紛らわしい模様をついたテーブル上であり, チャレンジングなケースである . Precision-Recall カーブを図 3 に示す . 青線は SIFT 特徴点と式 (5) (ベースライン) の対応点集合類似度を用いた場合, 緑線は RGB 最近傍探索と式 (5) (ベースライン) の対応点集合類似度を用いた場合, そして赤線は RGB 最近傍探索と式 (6) (提案手法) の対応点集合類似度を用いた場合である . まず, テクスチャが多く平面的である対象物体 No. 1 と No. 5 を除いて, SIFT 特徴点を用いた場合の性能が低いことが分かる . そして, RGB 最近傍探索を用いた場合, 対象物体 No. 4 を除いて, 全ての物体について提案手法の性能がベースラインよりも高いことが分かる . 全対象物体の Average Precision の平均値は, SIFT 特徴点を用いた場合が 0.13, RGB 最近傍探索を用いた場合のベースラインが 0.27, そ

して提案手法が 0.31 であった .

#### 3.2 非剛体物体検出

本節では, 非剛体物体検出における提案手法の有効性を検証する . データセットとして, 物体形状検索のコンペティション Shape Retrieval Contest (SHREC'10) のうち, 非剛体物体の対応点探索タスクで用いられた SHREC'10 correspondence benchmark [14] を使用する . (図 4) . 本データセットは “human”, “dog”, “horse” の三物体クラスからなり, 各クラスは参照用のモデル形状データ個と, 様々な非剛体変形を施された形状データからなる . 非剛体変形として, 図 4 の二段め左から “isometry”, “topology”, “sampling”, “localscale”, “scale”, “holes”, “microholes”, “noise”, “shotnoise” の 9 種類の変形があり, それぞれの変形の度合いには 5 つのレベルが存在する . 本実験では “human”, “dog”, “horse” の三物体クラスのうちの 1 クラスを対象物体クラスとし, その他の二クラスに属する形状データを負例サンプルとして実験を行った . 正例サンプル, 負例サンプルともに, 9 種類の変形のうちの 1 種類のみを選択した . 学習サンプルとしては, 変形レベルが 1, 3, 5 のものを, それぞれにつきランダムに選んだ一種類のサンプル点群を使用

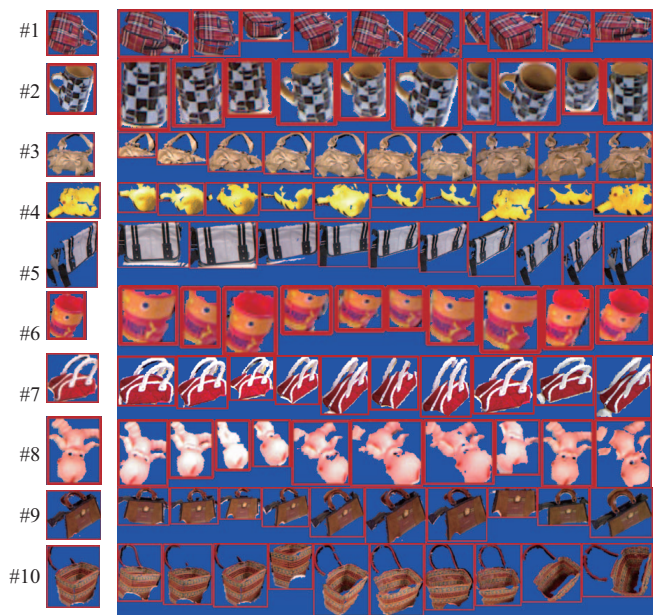
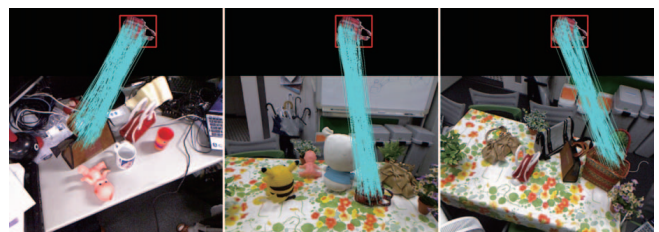


図 1 対象物体の参照画像と 9 枚の学習用 RGB-D 画像



(a) 学習サンプル (負例) (b) テストサンプル (正解) (c) テストサンプル (不正解)

図 2 対象物体 No. 1 の参照形状データと環境中の観測物体との間で得られた対応点集合の例 . (a) 学習データにおける正解の対応点集合 . (b) テストデータにおける正解の対応点集合 . (c) テストデータにおける不正解の対応点集合 .

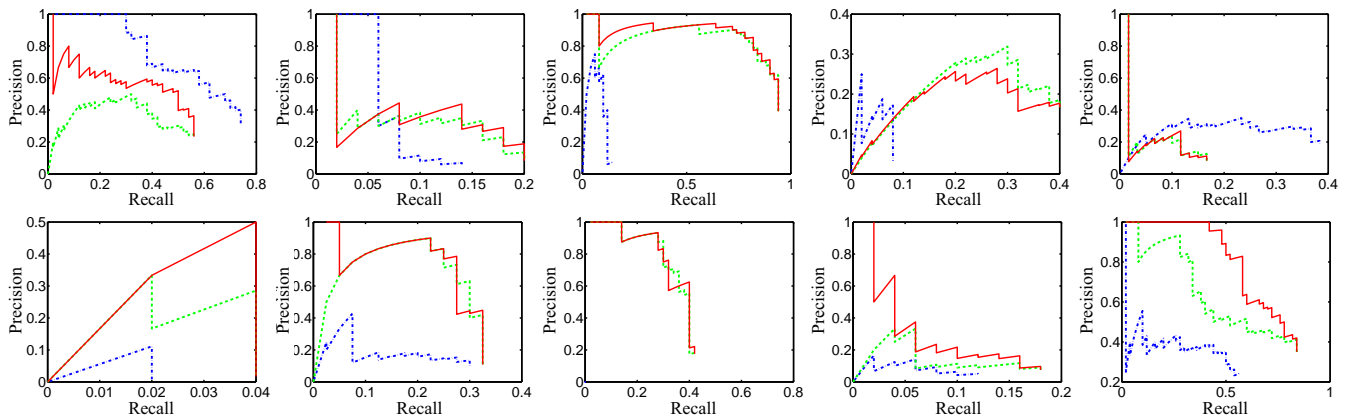


図3 Precision-Recall カーブ. 左上から順に, 対象物体 No. 1, ..., No. 10 である. 青線は SIFT 特徴点と式 (5) (ベースライン) の対応点集合類似度を用いた場合, 緑線は RGB 最近傍探索と式 (5) (ベースライン) の対応点集合類似度を用いた場合, 赤線は RGB 最近傍探索と式 (6) (提案手法) の対応点集合類似度を用いた場合である.

した. すなわち, 正例サンプルは 3 サンプル, 負例サンプルは  $6 (= 3 \times 2)$  サンプルである. 評価用サンプルとしては, 変形レベルが 2, 4 のものを, それぞれにつきランダムに選んだ 10 種類のサンプル点群を使用した. すなわち, 正例サンプルは 20 ( $= 2 \times 10$ ) サンプル, 負例サンプルは 40 ( $= 2 \times 2 \times 10$ ) サンプルである.

対応点集合獲得の方法について示す. まず, 各物体クラスのモデル形状データ上の点を, Farthest Point Sampling (FPS) [12] によって等間隔に 200 点サンプリングする. FPS はランダムに選んだ一点を初期値とし, 現在までに選んだ全ての点から最も遠い一点を追加するサンプリング手法である. このとき比較する二点間距離として任意の距離計量を使用することができる. 本実験では, ユークリッド距離を使用した. また, 初期値として選ぶ一点をランダムに変化させて FPS を 10 回繰り返すことで, 10 通りの異なるサンプリングを獲得した. 各サンプル点について, 非剛体物体マッチングに用いられる state-of-the-art の局所特徴である Scale-Invariant Heat Kernel Signatures (SI-HKS) [15] を抽出した. 各物体クラスのモデル形状データ  $R$  と正例サンプルの形状データ  $O_i^{(pos)}$  との対応点集合は, 人手で作成した (200 ペアの) Ground Truth を用いた. 一方で, 各物体クラスのモデル形状データと他の物体クラスの形状データ, あるいはどのクラスにも属さない負例サンプルの形状データとの対応点集合は, FPS で得られた各モデル形状データ上の 200 点に対して, SI-HKS の特徴空間上で各点につき 5 個の最近傍点をマッチング対象の形状データから選んで 1,000 ペアの候補対応点集合を作成し, 誤対応点除去を行うことで獲得した. Ground Truth の正例対応点集合と負例対応点集合の例を図 5 に示す.

Average Precision の値を表 1 に示す. 左から順に MD を単独で用いた場合 (“MD”), CT を単独で用いた場合 (“CT”), これらの類似度統合重みとして  $\{0.5, 0.5\}$  を使用した場合 (式 (5), “baseline”), そして提案手法で学習された類似度統合重みを使用した場合 (式 (7), “learned”) の結果を示している. Average Precision の平均値 (最下段) について, 提案手法で学習された

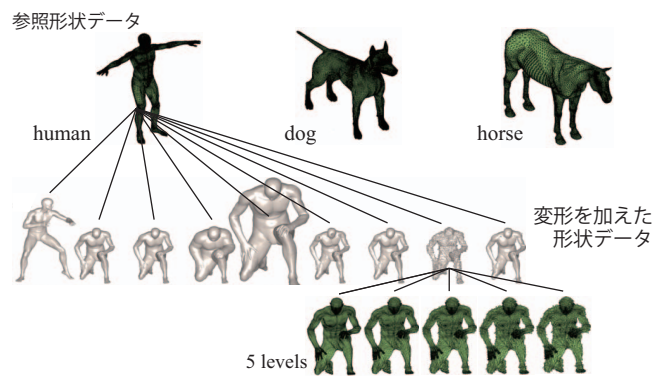


図4 SHREC'10 correspondence benchmark [14] の形状データ

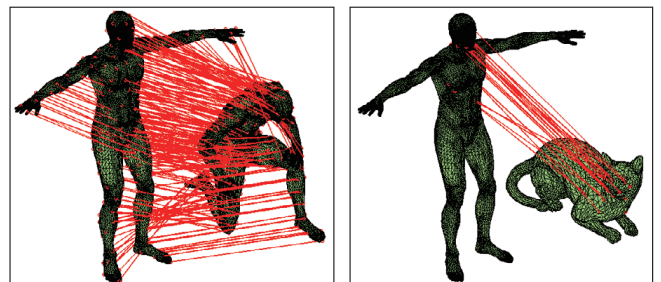


図5 正例の対応点集合 (左) と負例の対応点集合 (右)

類似度統合重みを使用した場合 “learned” の性能が最も高いことが示された.

図 6 は “scale” 変形クラスと “shotnoise” 変形クラスにおける類似度関数 (MD と CT) の学習された統合重みを示す. ここで, “scale” 変形クラスについては CT を用いた方が MD を用いる場合よりも性能が高く(注1), “shotnoise” 変形クラスについては MD を用いた方が CT を用いるよりも性能が高いことに注意されたい (表 1). これに対し, “scale” 変形クラスについては CT の重みが大きく, “shotnoise” 変形クラスについては MD の重みが大きくなっていることから, 提案手法によって適切な類

(注1): CT がスケール不変距離計量であるのに対して MD がスケール不変ではないことより自明の結果である.

表 1 非剛体物体検出実験結果 (Average precision)

Deform.	Target	Average Precision			
		MD	CT	baseline	learned
holes	human	1.00	0.94	1.00	1.00
holes	dog	0.69	0.94	0.75	1.00
holes	horse	1.00	0.96	1.00	1.00
isometry	human	1.00	1.00	1.00	1.00
isometry	dog	1.00	1.00	1.00	1.00
isometry	horse	1.00	1.00	1.00	1.00
microholes	human	1.00	1.00	1.00	1.00
microholes	dog	1.00	1.00	1.00	1.00
microholes	horse	1.00	1.00	1.00	1.00
noise	human	0.33	0.82	0.33	1.00
noise	dog	0.33	0.82	0.33	0.75
noise	horse	0.85	1.00	0.85	1.00
localscale	human	0.33	0.69	0.33	1.00
localscale	dog	0.33	0.52	0.36	1.00
localscale	horse	1.00	1.00	1.00	1.00
topology	human	0.33	0.55	0.34	1.00
topology	dog	0.68	1.00	0.82	1.00
topology	horse	1.00	1.00	1.00	1.00
sampling	human	1.00	1.00	1.00	1.00
sampling	dog	0.66	1.00	0.78	1.00
sampling	horse	1.00	1.00	1.00	1.00
scale	human	0.33	1.00	0.33	0.74
scale	dog	0.33	1.00	0.33	1.00
scale	horse	0.33	1.00	0.33	1.00
shotnoise	human	1.00	0.77	1.00	0.96
shotnoise	dog	0.85	0.75	0.77	1.00
shotnoise	horse	1.00	1.00	1.00	1.00
Average		0.76	0.92	0.77	0.98

似度統合重みが学習されていることが分かる。

#### 4. 結 論

本研究では、三次元点群を入力として検出対象物体の三次元点群との対応点を求め、対応点ペア類似度行列の二次割当問題を解いて得られた対応点集合に対して、その正当性を評価する対応点集合類似度を導入し、パラメータを物体毎に学習する手法を提案した。提案手法は、対応点集合類似度が、正解の対応点集合は高く、不正解の対応点集合は低くなるよう、二組の対応点ペアに対する重みを学習する。剛体物体検出実験では、全ての対応点ペアを均一に評価する場合（ベースライン）に対して、提案手法を用いた場合に約 15%の性能向上が確認できた。また、非剛体物体検出実験では複数種類の類似度関数の統合重みを適切に学習することを確認し、各類似度関数を単独で用いる場合、および均一な重みで統合する場合に対する優位性を示した。提案手法は対応点ペアを評価する任意の類似度関数に対して適用可能であり、とりわけ、三組以上の対応点ペアの類似度を用いるハイパーグラフマッチングへの応用が今後の展望である。

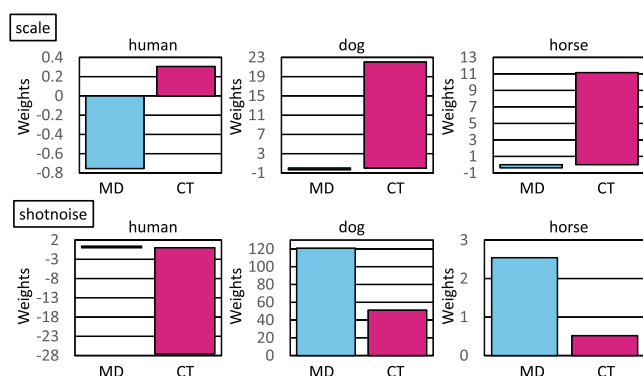


図 6 類似度関数 MD と CT に対する学習された統合重み。上段は“scale”変形クラスのサンプル群に対して学習されたものであり、下段は“shotnoise”変形クラスのサンプル群に対して学習されたものである。

#### 文 献

- [1] M. Leordeanu and M. Hebert: “A spectral technique for correspondence problems using pairwise constraints”, Proc. IEEE ICCV (2005).
- [2] E. Rodolà, A. Albarelli, F. Bergamasco and A. Torsello: “A scale independent selection process for 3d object recognition in cluttered scenes”, International Journal of Computer Vision (IJCV) - Special Issue on 3D Imaging, Processing and Modeling Techniques, **19**, (2012).
- [3] E. Rodolà, A. Torsello, T. Harada, Y. Kuniyoshi and D. Cremers: “Elastic net constraints for shape matching”, Proc. IEEE ICCV (2013).
- [4] E. Rodolà, T. Harada, Y. Kuniyoshi and D. Cremers: “Efficient shape matching using vector extrapolation”, Proc. BMVC (2013).
- [5] T. S. Caetano, J. J. McAuley, L. Cheng, Q. V. Le and A. J. Smola: “Learning graph matching”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), **31**, 6, pp. 1048–1058 (2009).
- [6] M. Leordeanu and M. Hebert: “Smoothing-based optimization”, Proc. IEEE CVPR (2008).
- [7] M. Leordeanu, R. Sukthankar and M. Hebert: “Unsupervised learning for graph matching”, International Journal of Computer Vision, **96**, 1, pp. 28–45 (2012).
- [8] M. Leordeanu, A. Zanfir and C. Sminchisescu: “Semi-supervised learning and optimization for hypergraph matching”, Proc. IEEE ICCV (2011).
- [9] D. Pachauri, M. Collins, V. Singh and R. Kondor: “Incorporating domain knowledge in matching problems via harmonic analysis”, Proc. ICML (2012).
- [10] S. Belongie, J. Malik and J. Puzicha: “Shape matching and object recognition using shape contexts”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), **24**, 4, pp. 509–522 (2002).
- [11] E. Rodolà, A. M. Bronstein, A. Albarelli, F. Bergamasco and A. Torsello: “A game-theoretic approach to deformable shape matching”, Proc. IEEE CVPR (2012).
- [12] A. M. Bronstein, M. M. Bronstein and R. Kimmel: “Numerical geometry of non-rigid shapes”, Springer (2008).
- [13] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz and Y. Singer: “Online passive-aggressive algorithms”, Machine Learning Research, **7**, pp. 551–585 (2006).
- [14] A. M. Bronstein, M. M. Bronstein, U. Castellani, A. Dubrovina, et al.: “Shrec 2010: robust correspondence benchmark”, Proc. EUROGRAPHICS Workshop on 3D Object Retrieval (3DOR) (2010).
- [15] M. M. Bronstein and I. Kokkinos: “Scale-invariant heat kernel signatures for non-rigid shape recognition”, Proc. IEEE CVPR (2010).