

3次元空間における物体の形状と見えを 記述した特徴量の評価実験

○近藤直明 大石修士 岩下友美 倉爪亮(九州大学)

1. はじめに

近年、RGB-D カメラやレーザスキャナなど、様々な距離センサが開発され、実世界での3次元形状計測は身近な技術になりつつある。特に、Kinect(Microsoft)やXtion(ASUS)などのRGB-D カメラは、リアルタイムに物体の形状とテクスチャ画像を計測することができるため、計測された物体の識別などに利用されている[1]。本研究では、RGB-D カメラを用いて物体の距離画像とテクスチャ画像を取得し、それらの画像から3DGHOG[2]を抽出することで、物体識別を行う。

Tangら[3]は、物体の形状を表現するための特徴量HONV(Histogram of Oriented Normal Vector)を提案している。評価実験では、RGB-D カメラにより得られた距離画像にHONVを適用し、約91%の識別率を実現している。しかし、Tangらの手法のように、距離センサから得られる形状情報のみに基づいて識別を行う場合、異なる種類であるが形状の類似した物体を識別することは困難になる。この場合、物体の形状情報に加え、物体表面の色や模様などの“見え”の情報を用いた識別が有効となる。そこで我々は、距離画像とテクスチャ画像から、物体の形状と見えを3次元空間における勾配方向の分布として表現する特徴量3DGHOG(3D Geometry based Histogram of Oriented Gradients)を開発した[2]。本稿ではRGB-D カメラから得られた距離画像とテクスチャ画像の例として公開されているRGB-D データセット[4]に対し、提案した3DGHOGの評価実験を行い、その結果について考察する。

2. 提案手法

本章では、3DGHOGのアルゴリズムについて述べる。3DGHOGを用いた物体の識別手順を以下に示す。

1. RGB-D カメラを用いて距離画像とテクスチャ画像を計測する(図3(a))。
2. 見えの特徴を記述するために、テクスチャ画像からHOG(2.1節)を計算する(図3(b))。
3. 形状特徴を抽出するために、距離画像からHONV(2.2節)を計算する(図3(c))。
4. 距離画像から抽出したHONVに従い、HOG特徴量を3次元空間に投影し、3DGHOG(2.3節)を計算する(図3(d))。
5. SVM(Support Vector Machine)を用いて上記の特徴量に基づき物体を識別する。

2.1 HOG(Histogram of Oriented Gradients)

HOGは、Dalalら[5]により開発された特徴記述子で、局所領域における輝度の勾配方向の分布として物体の見えを記述する。



(a) 距離画像 (b) テクスチャ画像

図1 距離画像とテクスチャ画像

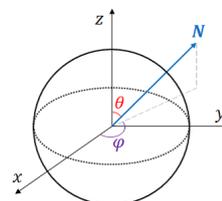


図2 法線ベクトル

HOGを抽出するためには、まず入力画像の各ピクセルにおいて輝度の勾配方向と勾配強度を計算する。次に、入力画像を $k \times k (= N_c)$ 個の小領域(セル)に分割し、各セルにおいて勾配方向に関するヒストグラムを作成する。その際、勾配方向は離散化され、 N_α 個のビンで表される。最後に、 $q \times q (= S_b)$ のセルを含むブロック領域において勾配ヒストグラムを正規化することで、HOG特徴量を抽出する。ブロック正規化を行うことで、HOGは照明変化に対して頑健な特性を持つ。

2.2 HONV(Histogram of Oriented Normal Vector)

HONVは、Tangら[3]により提案された特徴量で、物体の3次元形状を局所領域における法線方向の分布として表現している。

まず距離画像の各ピクセル $p = (x, y)$ において法線方向を計算する。法線 N は式(1)のように計算できる。

$$\mathbf{N} = \begin{pmatrix} \frac{\partial d(x,y)}{\partial x} \\ \frac{\partial d(x,y)}{\partial y} \\ 1 \end{pmatrix} \quad (1)$$

ここで、 $d(x, y)$ はピクセル p における距離値を表す。また、計算された法線方向は、図2に示すように、天頂角 θ と方位角 φ で表される。

$$\theta = \tan^{-1} \sqrt{\left(\frac{\partial d(x,y)}{\partial x}\right)^2 + \left(\frac{\partial d(x,y)}{\partial y}\right)^2} \quad (2)$$

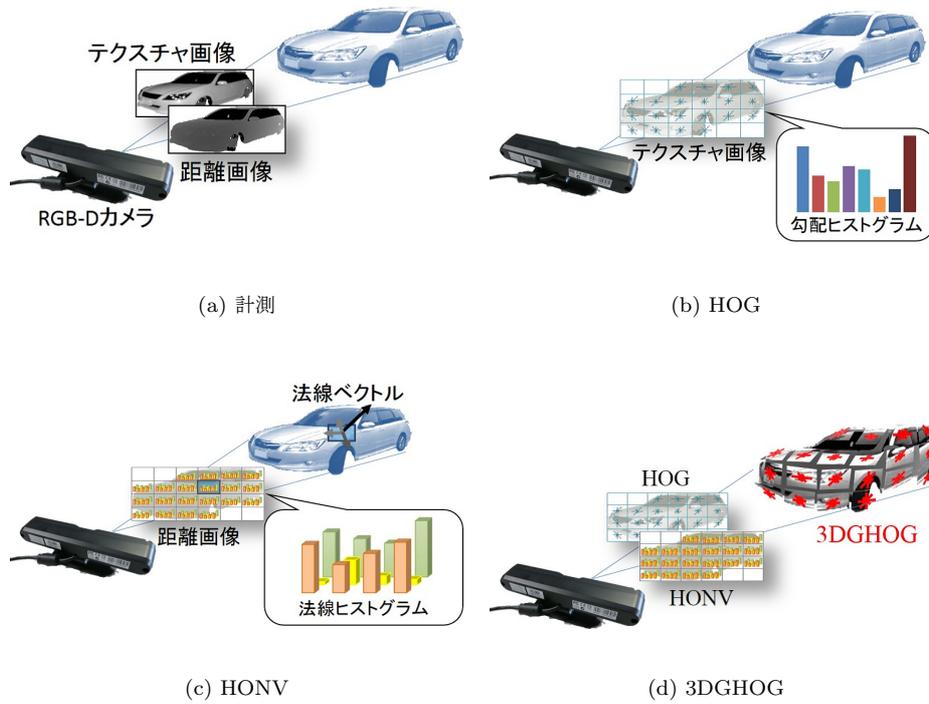


図3 特徴抽出

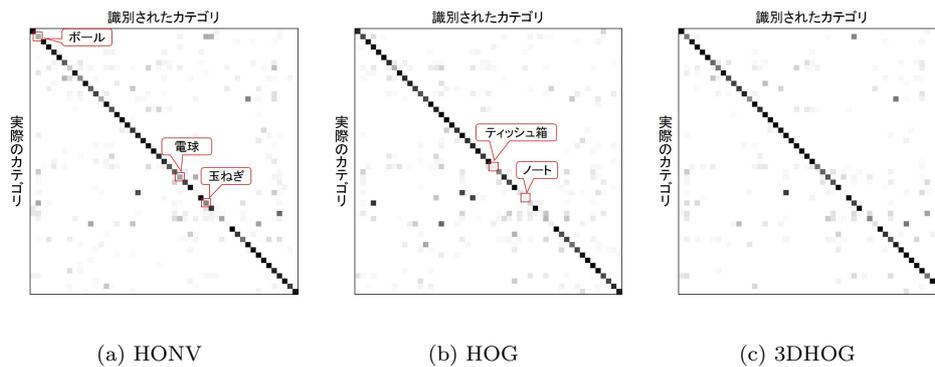


図4 識別結果

$$\varphi = \tan^{-1} \left(\frac{\partial d(x, y)}{\partial y} / \frac{\partial d(x, y)}{\partial x} \right) \quad (3)$$

法線方向を計算後、距離画像を $k \times k (= N_c)$ 個のセルに分割する。次に、分割されたセル内において、法線方向を離散化し、 $N_\theta \times N_\varphi$ 個のビンに投票を行うことでヒストグラムを作成する。最後に、HOGと同様、 $q \times q (= S_b)$ のセルを含むブロック領域においてヒストグラムを正規化し、HONV 特徴量を抽出する。

2.3 3DGHOG

まず、物体の見えを記述するためにテクスチャ画像から HOG を抽出する。その後、距離画像から抽出した HONV に従って HOG を 3次元空間に投影する。HONV は各セルにおける法線分布を持っている。そのため、テクスチャ画像の法線 N_i (画像に垂直なベクトル) を HONV の法線 $N_{\theta, \varphi}$ に回転させることで、対応するセルの 3次元空間における見えを抽出する。3DGHOG のアルゴリズムをアルゴリズム 1 に示す。

3. 実験

本節では、RGB-D データセット [4] を用いた 3DGHOG の評価実験について述べる。RGB-D データセットには 51 カテゴリの物体の距離画像とテクスチャ画像が含まれている。特徴量のパラメータは、画像の分割数 $N_c = 8 \times 8$ 、ブロックサイズ $S_b = 2 \times 2$ 、HOG のビン数 $N_\alpha = 9$ 、HONV のビン数 $N_\theta = 15$ 、 $N_\varphi = 15$ とした。また、特徴量には、{HONV}、{HOG}、{3DGHOG} の 3 種類の組み合わせを用い識別を行った。各組み合わせの次元数は、{{HONV}、{HOG}、{3DGHOG}} = {28224, 1764, 28224} となる。識別実験では、各カテゴリからランダムに 1 つずつデータを選び、それらのデータをテストデータ、残りのデータを学習データとして、性能を評価した。

図 4 と表 1 に識別結果及び正答率を示す。図 4 は、図の左上から右下に向かう対角線の輝度値が正答率に対応しており、その輝度値が小さいほど高い正答率を表す。RGB-D データセットには、ボールや電球、玉ねぎなど

Algorithm 1: 3DGHOG**Input:**

- Range and reflectance images
- The number of cells $N_c (= k \times k)$
- Block size $S_b (= q \times q)$
- The number of bins for HOG N_α
- The number of bins for HONV N_θ, N_φ

Output:

- Three-dimensional HOG which contains:
 - 1) 3dgradArr - array of discretized 3D gradient vectors
 - 2) 3dgradStr - array of the strength of each 3D gradient

Definition:

- A HOG structure contains:
 - 1) gradArr - array of discretized gradient vectors
 - 2) gradStr - array of the strength of each gradient
- A HONV structure contains:
 - 1) normArr - array of discretized normal vectors
 - 2) normStr - array of the strength of each normal vector

The algorithm:

- 1) divide range and reflectance images into small cells
- 2) create HOG[N_c] and HONV[N_c] structures for each cell
- 3) **for each cell c do**
 - A) calculate HOG[c] from the reflectance image
 - B) calculate HONV[c] from the range image
 - C) **for each bin $u_{\theta, \varphi}$ in HONV[c] do**
 - a) calculate a projection matrix A from N_i to HONV[c].normArr[$u_{\theta, \varphi}$]
 - b) **for each bin u_α in HOG[c] do**
 - i) calculate 3D edge by multiplying the matrix A and HOG[c].gradArr[u_α]
 - ii) create 3DGHOG histogram by casting a weighted vote according to HOG[c].gradStr[u_α] \times HONV[c].normStr[$u_{\theta, \varphi}$] for the corresponding bin
- 4) normalize 3DGHOG at each block

の形状が非常に似ている物体がいくつか含まれており、物体の形状を記述する HONV のみを用いた場合、特にそれらの正答率は低い値を示した。一方、3DGHOG を用いることで、物体の形状だけでなく見えも抽出することができるため、形状の類似した物体でも識別性能

表 1 正答率

	HONV	HOG	3DGHOG
正答率	74.8 \pm 7.8%	73.7 \pm 8.7%	79.8 \pm 6.5%

が向上した。また、物体によって表面の模様が大きく異なるティッシュ箱やノートなどの物体に対し、HOG のみを用いた場合、正答率は低下したが、物体の形状を考慮して見えの情報を記述した 3DGHOG を用いることで識別性能が向上した。

4. まとめ

本稿では、一般に公開されている RGB-D データセット [4] を用いて、3DGHOG の評価実験を行った。3DGHOG とは、距離画像から得られる法線情報に基づき、テクスチャ画像から得られる HOG を 3 次元空間に投影し、物体の形状に応じた見えを表現する手法である。これにより、物体の形状と見えを組み合わせることで捉えることができ、従来手法よりも高い識別性能を示すことを確認した。

ところで、画像を用いた物体識別では、計測を行う角度によって物体の見え方が変化する。このため、視点方向に依存しない特徴量を用いることで、識別性能の向上が見込まれる。そこで今後は、3DGHOG に視点不変性を導入し、識別性能の向上を計る。

謝辞

本研究の一部は文部科学省科学研究費補助金基盤研究 (A) (課題番号 26249029) の支援を受けた。

参考文献

- [1] M. Blum, J. T. Springenberg, J. Wülfing, and M. Riedmiller, "A Learned Feature Descriptor for Object Recognition in RGB-D Data", in IEEE International Conference on Robotics and Automation, pp.1298-1303, 2012.
- [2] 近藤 直明, 大石 修士, 岩下 友美, 倉爪 亮, "レーザスキャナによる 3 次元形状と見えに基づくセマンティックラベリング", 日本機械学会ロボティクスメカトロニクス講演会 2014, 1P1-V05, 2014.
- [3] S. Tang, X. Wang, X. Lv, T. X. Han, J. Keller, Z. He, M. Skubic, and S. Lao, "Histogram of oriented normal vectors for object recognition with a depth sensor", in The 11th Asian Conference on Computer Vision, pp.525-538, 2012.
- [4] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view rgb-d object dataset", in IEEE International Conference on Robotics and Automation, pp.1817-1824, 2011.
- [5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", in Computer Vision and Pattern Recognition, Vol.1, pp.886-893, 2005.