

# 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

Semantic labeling based on 3D structure and appearance information with a laser scanner

○ 近藤 直明 (九大)      大石 修士 (九大)  
岩下 友美 (九大)      正 倉爪 亮 (九大)

Naoaki KONDO, Kyushu University, kondou@irvs.ait.kyushu-u.ac.jp  
Shuji OISHI, Kyushu University  
Yumi IWASHITA, Kyushu University  
Ryo KURAZUME, Kyushu University

This paper presents a new object classification method using range and reflectance data obtained by a laser scanner. Several laser scanners produce not only range data but also reflectance which shows the power of reflected laser as a side product of range data. The proposed method utilizes both information for object classification. Since laser reflectance contains appearance information, the proposed method classifies objects based on their shapes and appearances from range and reflectance data. Moreover, we extend the conventional Histogram of Oriented Gradients (HOG) so that it couples geometrical and appearance information more tightly. We show the validity of the proposed method through experiments in real scene.

**Key Words:** Reflectance image, Laser scanner, Object recognition

## 1 はじめに

人と共存する生活支援ロボットにとって、ロボット周囲の環境情報の獲得は重要な技術の一つである。特に、レーザスキャナなどの距離センサから得られる3次元幾何情報は、豊富な情報量や高い計測精度から周囲環境の認識に有用であり、ロボットによる自動地図作成や経路計画等の分野で広く用いられている。さらに、距離センサやカメラから得られた情報に基づき、物体の識別を行うことで物体や環境種別などの高次の空間情報をアノテーションする”セマンティックラベリング”は、ロボットが安全かつ効率的に行動したり、物体のラベル付き地図を獲得するための判断材料として利用することができる。

形状情報に基づく物体のセマンティックラベリングとして、Zhuら [1] は、レーザスキャナから得られた距離画像に基づいた物体識別手法を提案している。入力された距離画像に対して Felzenszwalb and Huttenlocher アルゴリズム [2] を適用することで、物体のセグメンテーションを行う。その後、各物体から、物体のサイズや法線の分散などの特徴量を抽出し、SVM と CART により木、人、建物などの6カテゴリに対する識別実験を行い、提案手法の有効性を確認している。Xiongら [3] は、3次元点群データを k-means++ [4] により小領域に分割し、隣接した領域のコンテキスト情報に基づいた物体識別を行っている。また、グラフィカルモデルを用いる代わりに多階層の識別を行うことにより計算時間を減らしている。実験では、異なるシステムにより計測された各データに対して高い識別結果を示している。

一方で、形状情報に加えてカメラなどから得られるテクスチャ情報を用いることも識別性能を向上させるための有効な手段である。Posnerら [5] は、レーザスキャナとデジタルカメラからそれぞれ得られる幾何特徴と見えの特徴を組み合わせるセマンティックラベリングを実現している。まず、3次元点群データを平面パッチに分割し、各パッチを色情報に基づきさらに分割を行う。これにより、見えの類似したパッチに幾何構造を与える。最後に、bag-of-words 識別器を用いて物体識別を行う。形状情報とテクスチャ情報を同時に用いることで高い識別率を実現できるが、一般的にこれらの情報は、異なる視点から計測されるため、画像同士の位置合わせの精度が識別性能に悪影響を与える可能性がある。さらに、同じ場所であっても照明条件により画像が変化して



(a) Range image



(b) Reflectance image

**Fig.1** Range and reflectance images

しまう問題もある。

そこで本稿では、カメラから得られるテクスチャ画像の代わりに、レーザスキャナの距離計測時に副産物として得られるレーザー光の反射光強度情報を用いて物体の見えの情報を取得し、形状と見えに基づいた物体識別を行う。さらに、HOG 特徴量 [6] を拡張し、物体の見えを3次元空間で表現できる”3DHOG”を提案する。

## 2 提案手法

### 2.1 リフレクタンス画像

レーザスキャナのような time-of-flight センサは、センサから発射されたレーザー光が物体表面で反射し、再びセンサへ戻ってくるまでの時間を計測することで、物体までの距離値を取得する。レーザスキャナから得られた距離画像を図 1(a) に示す。一方、光学式距離センサの多くは、レーザーの反射光強度 (リフレクタンス値) を計測できる。リフレクタンス値を濃淡値として表示したリフレクタンス画像を図 1(b) に示す。さらに、距離画像とリフレクタンス画像は、同じ光学システムにより計測されるため、2つの画像は厳密に位置合わせが行われた画像となる。

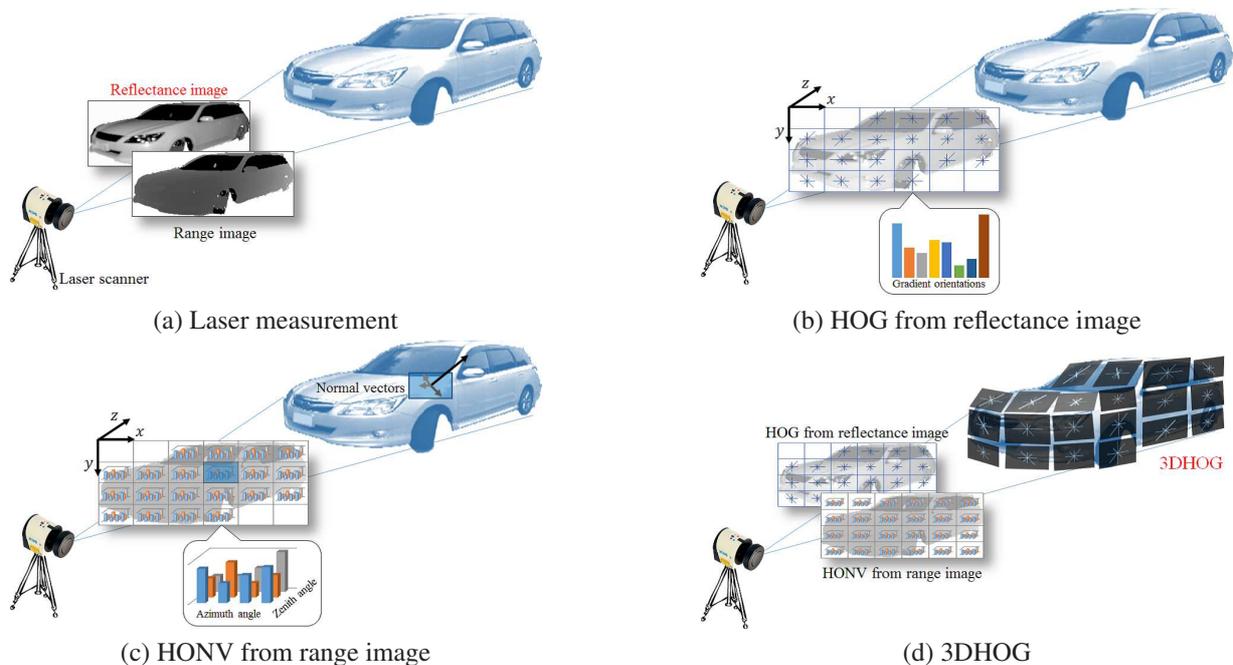


Fig.2 Feature extraction

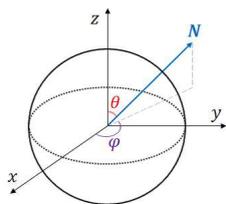


Fig.3 Representation of the direction of a normal vector by Zenith angle  $\theta$  and the azimuth angle  $\varphi$

## 2.2 識別手順

我々は、距離画像とリフレクタンス画像を用いた2種類の物体識別手法を提案する(図2(a))。提案手法の概要を以下に示す。

### 手法1: HOGによる見えの記述

1. レーザスキャナにより距離値とリフレクタンス値を計測する。
2. 計測された距離値とリフレクタンス値を各ピクセルの輝度値とする距離画像とリフレクタンス画像を作成する。
3. 形状特徴を抽出するために、距離画像からHONV(2.4節)を計算する。
4. 見えの特徴を記述するために、リフレクタンス画像からHOG(2.3節)を計算する。
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。

### 手法2: 3次元空間における見えの記述

物体の見えを3次元空間で表現するために、識別(手法1の5)を行う前に以下の処理を加える。

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、3DHOG(2.5節)を計算する。

## 2.3 HOGによる見えの記述

HOG(Histogram of Oriented Gradients)は、歩行者検出のためにDalalら[6]が開発した特徴記述子で、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できる(図2(b))。

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 $k \times k$ 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、 $N_\alpha$ 個のビンで表される。最後に、 $q \times q$ 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。ブロック正規化を重ねて行うことで、HOGは照明変化や小さな変形に対して頑健な特性を持つ。

## 2.4 HONVによる法線抽出

HONV(Histogram of Oriented Normal Vector)は、距離センサから得られた物体の形状特徴を記述するために、Tangら[7]により提案された(図2(c))。HONVは、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ため、まず距離画像の各ピクセル $p = (x, y)$ において法線を計算する必要がある。法線 $N$ は式(1)のように計算できる。

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

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3に示すように、天頂角 $\theta$ と方位角 $\varphi$ で表される。

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

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

法線方向を計算後、HONVはHOGと同じ手順で計算できる。まず、距離画像を $k \times k$ 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。

## 2.5 3DHOG

本節では、レーザスキャナから得られた点群データを用いた物体識別のための新たな特徴量について述べる。近年、SIFT[8]

や SURF[9], HOG[6] といった画像による物体識別のための特徴量が開発されている。これらの特徴量は、2次元空間での物体検出・識別において優れた性能を示している。しかし、実世界の物体は、見えだけでなく3次元形状を有しており、物体の2次元特徴のみでの表現には限界がある。

そこで我々は、レーザスキャナから得られる距離画像とリフレクタンス画像を組み合わせることで、物体の見えを3次元空間における勾配方向の分布として表現する(図2(d))。距離画像とリフレクタンス画像は厳密に位置合わせされているため、まず物体の見えを記述するためにリフレクタンス画像から 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、リフレクタンス画像の法線  $N_i$  を HONV の法線  $N_{\theta, \phi}$  に回転させることで、対応するセルにおける HOG を物体表面に投影できる。ここで、法線  $N_i$  は (0,0,1) とする(図2(b))。提案する 3DHOG のアルゴリズムを以下に示す。

#### Algorithm 1: Three-dimensional HOG

##### 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  $N_\alpha$
- The number of bins for HONV  $N_\theta, N_\phi$

##### 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  $u_{\theta, \phi}$  in HONV[ $c$ ] do**
    - a) calculate a projection matrix A from  $N_i$  to HONV[ $c$ ].normArr[ $u_{\theta, \phi}$ ]
    - b) **for each bin  $u_\alpha$  in HOG[ $c$ ] do**
      - i) calculate 3D edge by multiplying the matrix A and HOG[ $c$ ].gradArr[ $u_\alpha$ ]
      - ii) create 3DHOG histogram by casting a weighted vote according to HOG[ $c$ ].gradStr[ $u_\alpha$ ]  $\times$  HONV[ $c$ ].normStr[ $u_{\theta, \phi}$ ] for the corresponding bin 3DHOG[ $c$ ].3dgradStr[ $u'_{\theta, \phi}$ ]
- 4) normalize 3DHOG at each block

## 2.6 識別

距離画像とリフレクタンス画像から抽出した特徴ベクトルを特徴量とし、線形 SVM(Support Vector Machine) を用いて識別を行う。その後、対象物体のカテゴリを決定し、対応するラベルを3次元点群に付加する。

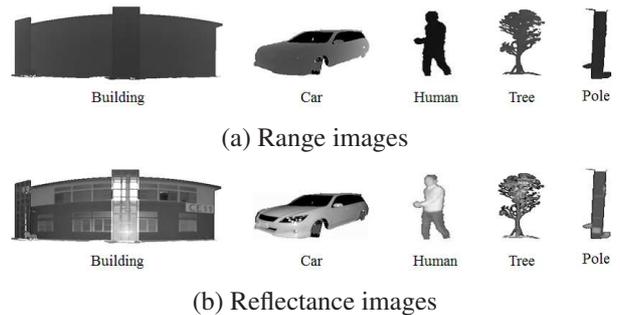


Fig.4 Categories of objects in environmental maps

Table 1 Classification results of semantic labeling

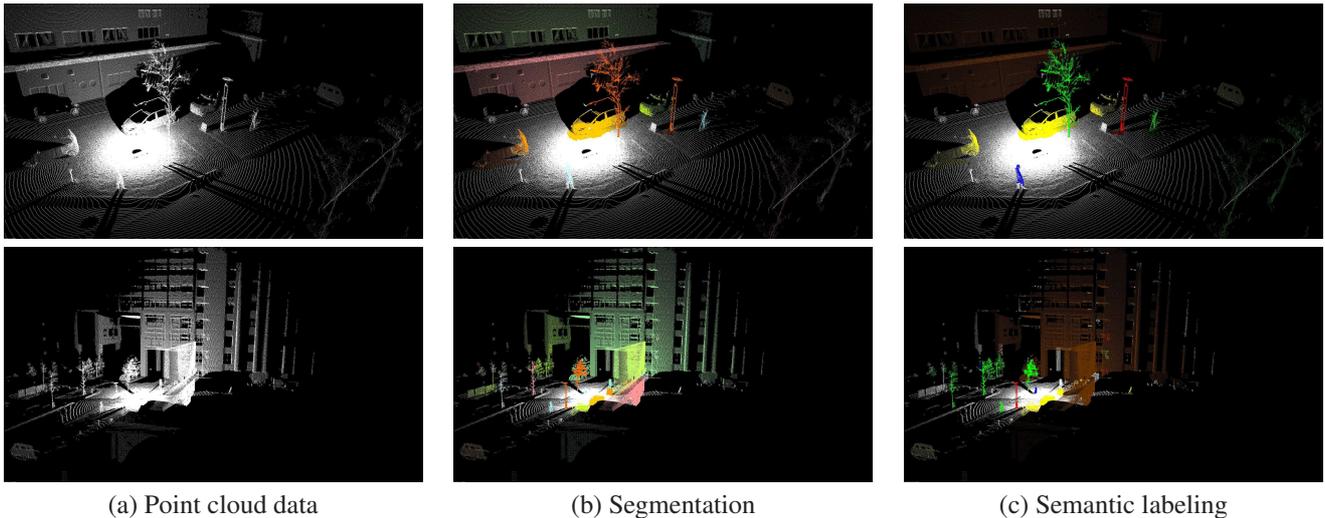
	HONV					
	Building	Car	Human	Tree	Pole	Total
Images	103	104	124	131	28	490
Detected	104	103	123	135	25	
Correct	94	97	115	124	23	453
Precision	90.4%	94.2%	93.5%	91.9%	92.0%	
Recall	91.3%	93.3%	92.7%	94.7%	82.1%	92.4%
	HONV and HOG (Proposed method 1)					
	Building	Car	Human	Tree	Pole	Total
Images	103	104	124	131	28	490
Detected	104	101	125	132	28	
Correct	98	98	121	127	26	470
Precision	94.2%	97.0%	96.8%	96.2%	92.9%	
Recall	95.1%	94.2%	97.6%	96.9%	92.9%	95.9%
	HONV and 3DHOG (Proposed method 2)					
	Building	Car	Human	Tree	Pole	Total
Images	103	104	124	131	28	490
Detected	102	104	129	127	28	
Correct	100	101	124	127	27	479
Precision	98.0%	97.1%	96.1%	100%	96.4%	
Recall	97.1%	97.1%	100%	96.9%	96.4%	97.8%
	3DHOG					
	Building	Car	Human	Tree	Pole	Total
Images	103	104	124	131	28	490
Detected	106	101	124	131	28	
Correct	103	99	123	128	27	480
Precision	97.2%	98.0%	99.2%	97.7%	96.4%	
Recall	100%	95.2%	99.2%	97.7%	96.4%	98.0%

## 3 実験

本節では、提案手法の有効性を示すための実験について述べる。計測では、回転テーブルに設置したレーザスキャナ(LMS151, SICK)を用い、760×2250画素の画像を取得する。識別対象の物体のカテゴリは建物、車、人、木、ポールの5つとした(図4)。特徴量のパラメータは、 $N_c = 8 \times 8, S_b = 4 \times 4, N_\alpha = 18, N_\theta = 9, N_\phi = 9$ とする。また、特徴量には、{HONV}, {HONV,HOG}, {HONV,3DHOG}, {3DHOG}の4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}, {HONV,HOG}, {HONV,3DHOG}, {3DHOG}}={32400, 39600, 64800, 32400}となる。識別実験では、leave-one-out cross-validationを繰り返し適用し、性能を評価した。

### 3.1 実験手順

物体から特徴量を抽出するためには、複数の物体が含まれている点群データに対してセグメンテーションを行う必要がある。セグメンテーションでは、まず入力データから地面領域を除去する。その後、距離値や法線方向の連続性に基づき、識別対象の物体領域を抽出する。計測点群、およびこの方法によりセグメンテーションされた結果を図5(a), (b)に示す。



**Fig.5** Semantic labeling using 3DHOG (proposed method 2) : (a) Point cloud data obtained with a laser scanner and a rotation table. (b) Segmentation result applying the enhanced radially bounded nearest neighbor (c) Semantic labeling result using the proposed method 2. Note that each object in the scene is colored according to the classification results; Building:Brown, Car:Yellow, Human:Blue, Tree:Green, Pole:Red, White:The ground and under-segmented clusters.

### 3.2 実験結果

セマンティックラベリングの実験結果を表 1 に示す。この表から HONV の高い識別性能は確認できるが、建物とポールなど、法線分布は似ているが、異なるカテゴリの物体の識別性能は低下している。一方、HONV と HOG を用いた場合の識別性能は、HONV のみを用いた場合の識別性能よりも高い性能を示している。さらに、HONV と 3DHOG を用いた場合と、3DHOG のみを用いた場合では、組み合わせの中で最も高い識別率を示している。HONV と 3DHOG を用いて識別を行い、カテゴリのラベルを付けた結果を図 5(c) に示す。表 1 と図 5 から、距離画像とリフレクタンス画像を組み合わせた 3DHOG は、非常に高い識別性能を有することが分かる。

### 4 まとめ

本稿では、単一のレーザスキャナから得られた距離情報とリフレクタンス情報を組み合わせた新たな物体識別手法を提案した。レーザスキャナの距離計測の副産物として得られる反射光強度を利用することで、見えの情報を抽出することができる。さらに、距離画像とリフレクタンス画像は厳密に位置合わせされている。そのため、距離画像から得られる法線情報に基づき、リフレクタンス画像から得られる HOG を 3 次元空間に投影することで、物体の見えを 3 次元空間で記述する 3DHOG を開発した。また、計測により得られた距離画像とリフレクタンス画像を用いた識別実験を行い、リフレクタンス情報を用いることで、より高い識別性能を示すことを確認した。

提案手法では物体の 3 次元テクスチャを記述するためにリフレクタンス画像を用いているが、距離画像とカメラから得られるテクスチャ画像が完全に位置合わせされている場合、提案手法を適用することができる。そこで今後は、RGB-D データセット [10] に提案手法を適用し、3DHOG の性能を評価する予定である。

### References

[1] X. Zhu, H. Zhao, Y. Liu, Y. Zhao, and H. Zha, "Segmentation and classification of range image from an intelligent vehicle in urban environment", in The 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010.

[2] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation", Int. J. Comput. Vision, vol. 59, no. 2, pp. 167-181, sep 2004. [Online]. Available: <http://dx.doi.org/10.1023/B:VISI.0000022288.19776.77>

[3] X. Xiong, D. Munoz, J. A. D. Bagnell, and M. Hebert, "3-d scene analysis via sequenced predictions over points and regions", in IEEE International Conference on Robotics and Automation (ICRA), May 2011.

[4] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding", in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, ser. SODA '07. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007, pp. 1027-1035. [Online].

[5] I. Posner, M. Cummins, and P. Newman, "Fast probabilistic labeling of city maps", June 2008.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", in Computer Vision and Pattern Recognition, vol. Vol.1, 2005.

[7] 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 (ACCV 2012), 2012.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", Int. J. Comput. Vision, vol. 60, no. 2, pp. 91-110, nov 2004.

[9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speed-up robust features (surf)", Comput. Vis. Image Underst., vol. 110, no. 3, pp. 346-359, jun 2008. [Online]. Available: <http://dx.doi.org/10.1016/j.cviu.2007.09.014>

[10] 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 (ICRA). IEEE, 2011, pp. 1817-1824.