DEPTH SENSOR BASED OBJECT DETECTION

USING SURFACE CURVATURE

A Thesis presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

YANG LIU

Dr. Tony X. Han, Thesis Supervisor

MAY 2014

The undersigned, appointed by the Dean of the Graduate School, have examined the thesis entitled

DEPTH SENSOR BASED OBJECT DETECTION

USING SURFACE CURVATURE

presented by Yang Liu,

a candidate for the degree of Master of Science and hereby certify that, in their opinion, it is worthy of acceptance.

_____

Dr. Tony X. Han

_____

Dr. Zhihai He

_____

Dr. Jianlin Cheng

ACKNOWLEDGMENTS

I would like to thank my lab mates Shuai Tang, Guang Chen, Yan Li, Miao Sun and Guang (Arthur) Chen who have been helping me with my research a lot. I also would love to thank my parents living in China who has been supporting my study financially, spiritually and emotionally.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

ABSTRACT

An object detection system finds objects from an image or video sequence of the real world. The good performance of object detection has been largely driven by the development of well-established robust feature sets. By using conventional color images as input, researchers have achieved major success. Recent dramatic advances in depth imaging technology triggered significant attention to revisit object detection problems using depth images as input.

Using depth information, we propose a feature, Histogram of Oriented Curvature (HOC), designed specifically to capture local surface shape for object detection with depth sensor. We form the HOC feature as a concatenation of the local histograms of Gaussian curvature and mean curvature. The linear Support Vector Machine (SVM) is employed for the object detection task in this work. We evaluate our proposed HOC feature on two widely used datasets and compare the results with other well-known object detection methods applied on both RGB images and depth images. Our experimental results show that the proposed HOC feature generally outperform the HOG and HOGD features in object detection task, and can achieve similar or higher results compared with the state-of-the-art depth descriptor HONV on some object categories.

Chapter 1

Introduction

An object detection system finds objects from an image or video sequence of the real world. While humans perform object detection effortlessly and instantaneously, algorithmic description of this task for implementation on computers has been very difficult.

The good performance of object detection has been largely driven by the development of well established robust feature sets. By using conventional color images as input, researchers have achieved significant success. And compared with color images, depth maps have their certain advantages. For one, depth maps reflect pure geometry and shape cues, which can be more discriminative than color and texture under many circumstances when it comes to the task of object recognition. Moreover, depth images are insensitive to changes in lighting conditions, which can provide us with more robust feature sets.

Although depth sensors have been available for many decades, their applications have been limited due to the high cost and complexity of operation. However, recent dramatic advances in depth imaging technology has reduced the cost of a depth sensor to a consumer price point with the launch of the Kinect [1, 2], triggered significant attention to revisit object detection and activity recognition problems using depth

images as input instead of color images. Recent works [3, 4, 5, 6, 7, 8, 9, 10] have achieved significant improvement.

Since the depth image can be considered as a gray level intensity image, traditional computer vision algorithms can be applied on the depth image. However, we won't attain optimal results due to the waste of full advantages of depth information. That's why we need to develop algorithms exclusively designed for depth images to achieve better results.

We propose a feature, Histogram of Oriented Curvature (HOC), designed specifically to capture local 3D geometric characteristics for object detection with depth sensor. We aim to recognize the object category by discriminate the surface of an object using surface curvature. From the 3D coordinate of a surface point $\big(x, y, h(x, y)\big)$, where $h(x, y)$ is the depth value acquired by the depth sensor, we can describe the surface shape by using Gaussian curvature and mean curvature. We form the HOC feature as a concatenation of the local histograms of Gaussian curvature and mean curvature.

The linear Support Vector Machine (SVM) is employed for the object detection task in this work. We use the SVM light [11,12] package for all the experiments.

We evaluate our proposed HOC feature on two widely used datasets: RGB-D Object Dataset [3] and NYU-Depth Dataset V2 [13]. For the RGB-D Object Dataset, we compare our results on 6 object categories with 3 other methods: HOG which is applied on RGB images, HOGD which is HOG being applied on depth images and HONV. Overall, our algorithm achieve better results than HOG and HOGD, while not

as good as HONV. For the NYU-Depth Dataset, we compare our results on 5 object categories with the results of HONV, and the two feature sets achieve similar results.

The remainder of this thesis is organized as follows. Chapter 2 discusses some previous related work. Chapter 3 introduces the HOC feature for object detection. Chapter 4 shows our experimental results and analysis. Summary of this work is given in Chapter 5.

Chapter 2

Related Work

There are extensive work for object detection, some of which are significant for general object detection problems mainly focused on intensity images. The SIFT feature [14] has been one of the most popular features for object detection and image retrieval problems due to its scale invariant property. The local gradient based feature HOG [15], has been extensively applied among computer vision tasks and achieved great performance. Local texture based features such as Local Binary Pattern (LBP) is also widely used as complementary information to gradients [16, 17].

The emergence of the Kinect sensor has provided researchers opportunities to pursue new algorithms applied on depth images. Some of the approaches adapt the existing techniques previously applied on intensity images, or combine the RGB and depth information to improve the performance. There are also some new techniques proposed in recent work which designed specifically for depth images. Bo et al. [7] develop a set of kernel features on depth images that model size, 3D shape, and depth edges in a unified framework, which has been demonstrated to have much better performance than spin images [18]. The Relational Depth Similarity Features (RDSF) [5] calculates features derived from a similarity of depth histograms that represents the relationship between two local regions using the Bhattacharyya distance [19]. Lai et al. [9] tried to address the joint object category and instance recognition problem in

the context of RGB-D (depth) cameras using instance distance learning. Jamie et al. [4] proposed a real-time pose recognition method using the Kinect. Cai et al. [20] proposed a regularized maximum likelihood deformable model algorithm for 3D face tracking. There is also some work on 3D modeling in indoor environments [21, 22, 23] which takes advantage of the depth information. Xia et al. [6] proposed a model based approach which detects humans using a 2-D head contour model and a 3-D head surface model.

A large-scale 3D object dataset, the RGB-D Object Dataset, was announced as a benchmark dataset [3] for algorithms based on depth information. It has been used to evaluate recent approaches such as Depth Kernel Descriptors (DKD) [7], Hierarchical Kernel Descriptors (HKDES) [8] and Instance Distance Learning (IDL) [9].

Another 3D object dataset, the NYU-Depth Dataset V2 [13] was recently contributed. It has been main used for pixel wise scene segmentation. For object detection, Wang et al. [24] achieved very good results by implementing a structured Hough voting method for detecting objects with heavy occlusion in indoor environments.

Local tangent has been studied for object recognition. Yu and Zhang [25] extends the Local Coordinate Coding (LCC, [26]) algorithm by exploring local tangent directions. The HONV [27] feature represent an object in 3D space as a distribution of the normal vector orientation. Xu and Xu [28] proposes an object description and recognition approach based on the relationship between the arc length and tangent orientation of object contours.

Surface curvature has also been widely used for object detection task. Monroy et al. [29] presented an approach that directly uses curvature cues in a discriminative way to perform object recognition. Deng et al. [30] presented a new structure-based interest region detector called Principal Curvature-Based Regions (PCBR) which they use for object class recognition.

Chapter 3

Histogram of Oriented Surface Curvature

3.1 The Depth Image

The traditional RGB camera captures the RGB intensity value at each pixel position (x, y). Using a Kinect sensor, the distance between the pixel position and the Kinect sensor $h(x, y)$, is obtained and quantized into 11-bit digits. The depth information captured by a depth sensor is usually called the depth image. The depth image h is described as $S(x, y) = [x, y, h(x, y)]^T$.

For the intensity image, the pixel intensity value and the pixel position have completely different physical meanings. For the depth image, however, both the pixel value of the depth image and the pixel position describe the 3D position of a surface point.

3.2 Surface Normal

Consider a parametric surface, $\mathbf{S}(u,v) = [x(u,v), y(u,v), z(u,v)]^T$, and a point P on $\mathbf{S}$. Assume that all first and second derivatives of $\mathbf{S}$ with respect to u and v exist at P. The tangent plane at P is identified by the two vectors $\mathbf{S}_u = \partial\mathbf{S}/\partial u$ and $\mathbf{S}_v = \partial\mathbf{S}/\partial v$.

The surface normal is the unit normal vector to the tangent plane of $\mathbf{S}$ at P; that is,

$$\mathbf{n}_s(P) = \frac{\mathbf{S}_u \times \mathbf{S}_v}{\|\mathbf{S}_u \times \mathbf{S}_v\|}.$$

The depth image h correspond to the parameterization $S(x,y) = \left[x, y, h(x,y)\right]^T$. In this case,

$$\mathbf{n}_s(P) = \frac{(-h_x, -h_y, 1)}{\sqrt{1 + h_x^2 + h_y^2}}$$

where $h_s = \partial h / \partial s$.

3.3 Surface Curvature

Curvatures make useful shape descriptors as they are invariant to viewpoint and parameterization. We now want to extend the notion of curvature of a curve to define the curvature of a surface.

For a parametric curve $\alpha(t) = (x(t), y(t))$ with t as a parameter, the curvature is given at each point by

$$k(t) = \frac{x'y'' - x''y'}{\left((x')^2 + (y')^2\right)^{\frac{3}{2}}}$$

where $\alpha' = d\alpha / dt$. For the common parameterization $\alpha(x) = \left[x, y(x)\right]^T$, the curvature becomes

$$k(t)=\frac{y''}{\left(1+(y')^2\right)^{\frac{3}{2}}}$$

Consider a parametric surface $S(u,v)$, and a point P on S. Assume that all first and second derivatives of S with respect to u,v exist at P. We define surface curvatures in four steps.

Step 1: Normal Curvature of a Curve on S. Consider a curve C on S going through P. We define the normal curvature of C at P as

$$k_n = k\cos\theta$$

where k is the curvature of C at P, and θ is the angle formed by the surface normal at P, $\mathbf{n}_s(P)$, with the curve normal $\mathbf{n}_C(P)$.

Step 2: Normal Curvature along a Direction. It can be proven the k_n does not depend on the particular curve C chosen, but only on the tangent of C at P, identified by the unit vector $\mathbf{d}$. This enables us to speak of normal curvature along a direction. For the sake of visualization, we can choose C as the planar curve obtained by intersecting S with a plane through P and containing both $\mathbf{d}$ and $\mathbf{n}_s(P)$. Obviously, C is a cross-section of S along $\mathbf{d}$, and describes the surface shape along that direction. In this case, $k_n = k$.

Step 3: Principal Curvatures and Directions. We could now describe the local shape of S at P by taking the normal curvatures at P in all directions. This is totally impractical, but fortunately unnecessary. Assume we know the maximum and minimum normal curvatures at P, k_1 and k_2 respectively, called principal

curvatures, and the corresponding directions, $\mathbf{d}_1$ and $\mathbf{d}_2$, called principal directions. It can be proven that

- The principal directions are always orthogonal, as shown in Figure 3.1;
- The normal curvature along any direction, $\mathbf{v} = (\cos\beta, \sin\beta)$, where β is the angle from $\mathbf{d}_1$ to $\mathbf{d}$, can be computed through Euler's formula:

$$k_n = k_1 \cos^2 \beta + k_2 \sin^2 \beta;$$

- Consequently, the local shape of the surface is completely specified by the principal curvatures and direction.

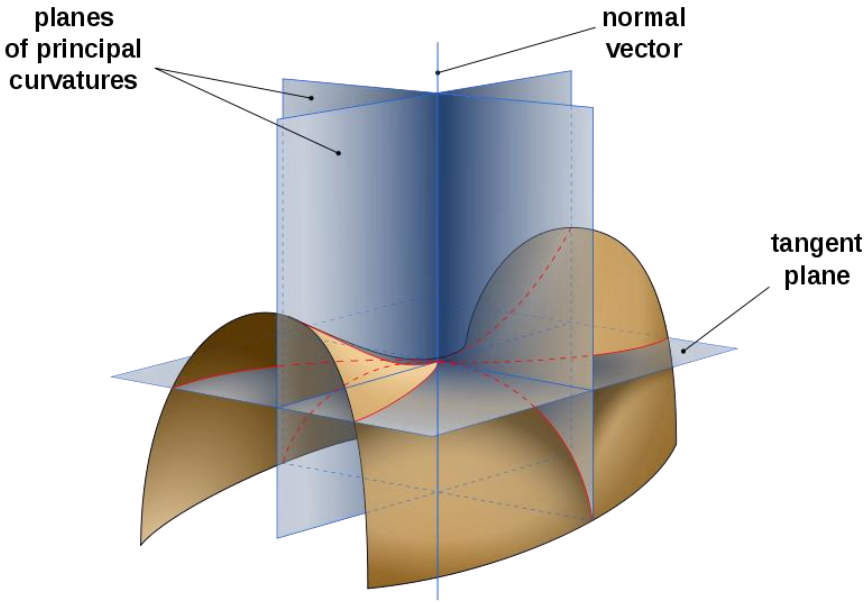

Figure 3.1 Surface with normal planes in directions of principal curvatures

Step 4: Classifying Local Shape. Finally, the shape classification is achieved by defining two further quantities, the mean curvature H, and the Gaussian curvature K:

$$H = \frac{k_1 + k_2}{2}$$

$$K = k_1 k_2$$

The Gaussian curvature measures how fast the surface moves away from the tangent plane around P, and in this sense is an extension of the 1-D curvature k.

3.4 The HOC Feature

Based on previous derivation, here is how to compute H, K from a depth image h in the form of $S(x,y)=\left[x,y,h(x,y)\right]^T$

$$K=\frac{h_{xx}h_{yy}-h_{xy}^2}{\left(1+h_x^2+h_y^2\right)^2}$$

$$2H=\frac{\left(1+h_x^2\right)h_{yy}-2h_xh_yh_{xy}+\left(1+h_y^2\right)h_{xx}}{\left(1+h_x^2+h_y^2\right)^{3/2}}$$

Following most of the histogram-based features, we first divide the detection window into non-overlapped m by n cells. For each cell, the Gaussian curvature and mean curvature at each pixel is quantized and voted into a 2D histogram of H and K. Assuming I bins and J bins are used for H and K respectively, the process produces a $I\times J$ dimensional feature vector for each cell. To avoid boundary effects, the orientation value is spread into multiple neighborhood bins using bilinear interpolation. A 2D Gaussian smoothing process is employed over adjacent cell histograms to relieve the spatial boundary effects. The final feature representation of the detection window is obtained by concatenating the HOC feature from each cell.

Unfortunately, we have to note that the input depth image contains noise, and this distorts the numerical estimates of derivatives and curvatures; thus, noise smoothing is required.

Chapter 4

Experiment

4.1 Implementation

For the object detection task, we take a traditional sliding window approach. We tried different cell sizes for each detection window, i.e. 4×4 pixel cells, 8×8 and 16×16, which all provide us with similar results. $L1$ norm is used for histogram normalization in each cell.

4.2 The Support Vector Machine

The linear Support Vector Machine (SVM) is employed for the object detection task in this work. We use the SVM light [35] package for all the experiments.

The support vector machine is a maximum margin classifier for the two-class classification problem

$$y(\mathbf{x}) = w^T \phi(\mathbf{x}) + b$$

The training data set comprises N input vectors $\mathbf{x}_1,...,\mathbf{x}_N$, with corresponding target values $t_1,...,t_N$ where $t_n \in \{-1,1\}$, and new data points $\mathbf{x}$ are classified according to the sign of $y(\mathbf{x})$.

During the training process, a detection model is generated based on the known positive and negative instances to be used in the following detection process.

4.3 Noise Issue

A Kinect depth sensor can only obtain the distance between the pixel position and the device as a 11-bit digits, and as we need to compute second-order derivative to get surface curvature, the depth information may not be able to provide us with accurate enough data. In this kind of situation, any image noise will seriously affect the performance of the algorithm. In order to reduce or eliminate the effect of noise issue, median filter and bilateral filter are applied prior to the computation, and we use Gaussian Smoothing before each round of derivation.

4.4 Datasets

The proposed HOC feature are evaluated for object detection task on two datasets: the RGB-D Object Dataset [3] and the NYU-Depth Dataset V2 [13].

4.4.1 RGB-D Object Dataset

The RGB-D Object Dataset is a large dataset of 300 common household objects. The objects are organized into 51 categories. This dataset was recorded using a Kinect style 3D camera that records synchronized and aligned 640x480 RGB and depth images at 30 Hz. Each object was placed on a turntable and video sequences were captured for one whole rotation. For each object, there are 3 video sequences, each recorded with the camera mounted at a different height so that the object is viewed

from different angles with the horizon. It contains 250,000 RGB and depth images from various viewpoints. The testing set for the detection task contains 8 clips of video in 8 different scenes.

In this work, we evaluate our proposed HOC feature on 6 object categories: bowl, cap, cereal box, coffee mug, flashlight and soda can.

4.4.2 NYU-Depth Dataset V2

The NYU-Depth Dataset V2 is comprised of video sequences from a variety of indoor scenes as recorded by both the RGB and Depth cameras from the Microsoft Kinect. It contains 1449 densely labeled pairs of aligned RGB and depth images, 795 of which are used as training set, while the other 654 for test purpose.

In this work, we evaluate our proposed HOC feature on 5 object categories: bed, sofa, door, table and chair.

4.5 Results

When coming to the task of object detection, we need to consider the factor of multiple aspect ratios for one category of object. A single model with fixed aspect ratio is not sufficient for detecting the object under all conditions. Thus, we use multiple components for each object category for our experiments. For RGB-D Object Dataset, we use 2 components for each object category. And for NYU-Depth Dataset, due to its variety and complexity of the scenes, we use 4 components for each object

category. During the training process, each positive instance will be assigned to the corresponding component according to its aspect ratio.

As discussed in section 3.4, the Gaussian curvature and mean curvature at each pixel is quantized and voted into a 2D histogram of H and K, which contain I bins and J bins respectively. The bin number I and J of the 2D histogram will affect the performance substantially. Too few bins would reduce the discriminative ability of HOC while too many bins would make it too sensitive. During the experiments, we find out that different object category requires different bin number setting to get the best performance. In this chapter, we choose show only the best result obtained for each category.

The detection performance is evaluated using the same criteria as the PASCAL VOC Challenge [54]. To be considered as a correct detection, the overlapping area between the predicted bounding box and the ground truth bounding box must exceed 0.5. If multiple bounding boxes have been detected around the same ground truth, only the one with the highest score is considered as a true detection.

4.5.1 Experimental Results on RGB-D Object Dataset

We use the precision-recall curve and Average Precision (AP) for performance comparison. We compare four methods on the RGB-D Dataset: HOG which is applied on RGB images, HOGD which is HOG being applied on depth images, HONV and our proposed HOC feature. The precision-recall curves are presented in Figure 4.1. The average precision achieved with different features are presented in Table 4.1.

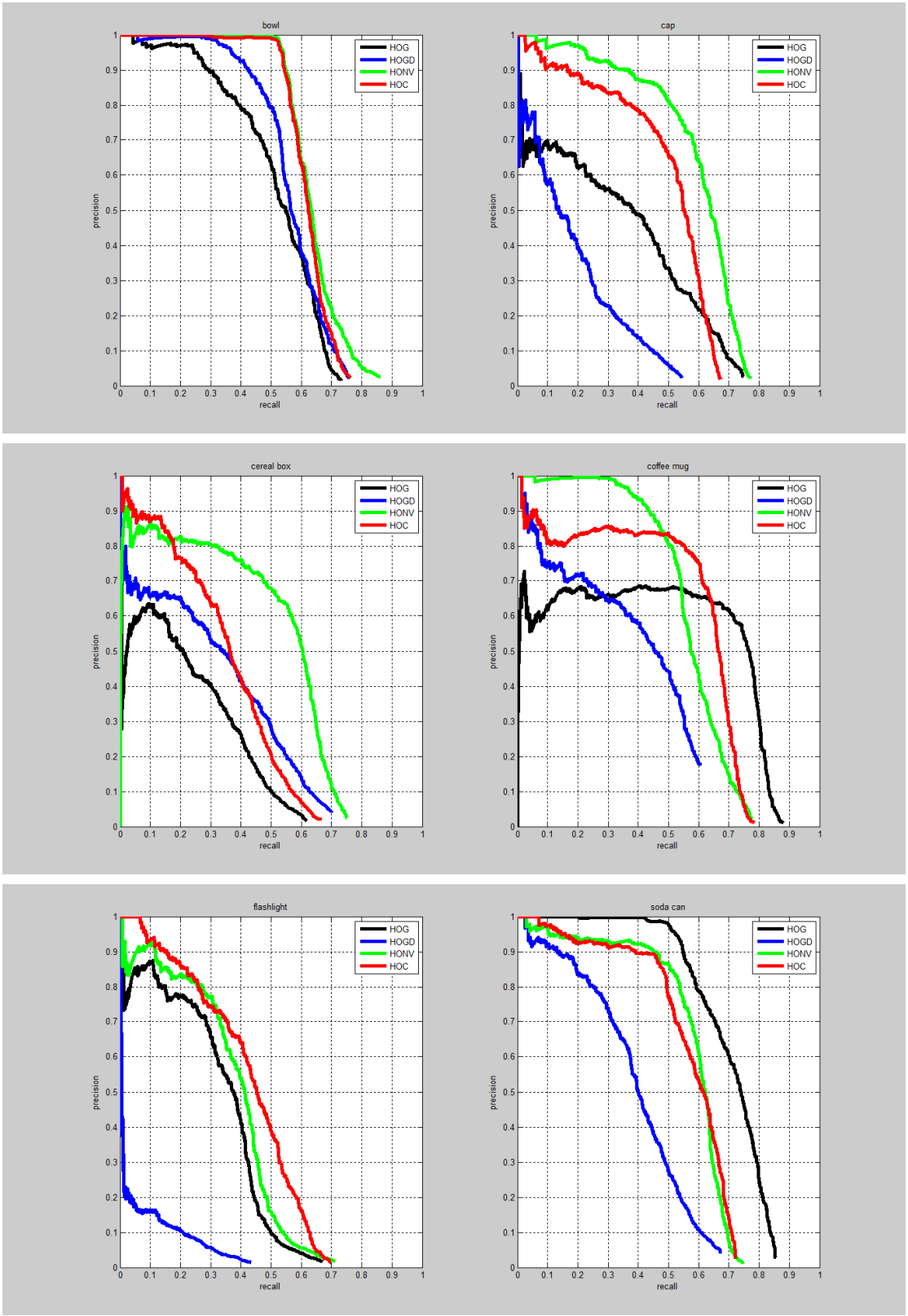

Figure 4.1 Precision-recall curves on the RGB-D Object Dataset

	HOG	HOGD	HONV	HOC
bowl	51.6	56.8	**64.3**	**62.3**
cap	33.3	17.9	**59.3**	**48.7**
cereal box	21.4	30.6	**49.2**	**34.6**
coffee mug	54.1	37.0	**57.9**	**56.6**
flashlight	32.0	4.7	36.7	**42.2**
soda can	**71.0**	39.0	**57.6**	56.8
AVERAGE	43.9	31.0	**54.2**	**50.2**

Table 4.1 Summarization of AP (%) on the RGB-D Object Dataset

From the above figure and table, we can find that generally our proposed HOC feature outperforms HOG and HOGD, while not performs as good as HONV.

One of the strong suits of the HOC feature over the HONV feature is that surface curvature is invariant to viewpoint. Since objects such as bowls, coffee mugs and soda can are symmetric from most viewpoints, the HOC feature doesn't have obvious advantages over the HONV feature. Considering the significant noise issue for the HOC feature, it is acceptable that the performance is worse than that of the HONV feature.

Figure 4.2 Example result obtained by HOC on the RGB-D Object Dataset

From the above image, we can see our method successfully captures some objects correctly in this sample image. One problem is that as “bowl” and “cap” have very similar surface curvature, they can often be categorized wrong. In this image, the algorithm fails to detect some small objects, such as “soda can” when they are relatively far from the depth sensor; in this kind of situation, the HOG feature can perform better as the depth sensor is unable to produce sufficient resolution to extract discriminative depth features.

4.5.2 Experimental Results on NYU-Depth Dataset V2

Since NYU-Depth dataset is originally for pixel wise scene segmentation and provides pixel-level labels, some instances are heavily occluded and in low resolution. Previous work [33] did an excellent precise labeling for this dataset and achieved state-of-the-art results, but unfortunately their label is not available publicly until now,

which makes our results not comparable with theirs. Thus, under exactly the same experiment settings, we compare the performance of two methods on this dataset: HONV and our proposed HOC feature. The precision-recall curves are presented in Figure 4.3. The average precision achieved with different features are presented in Table 4.2.

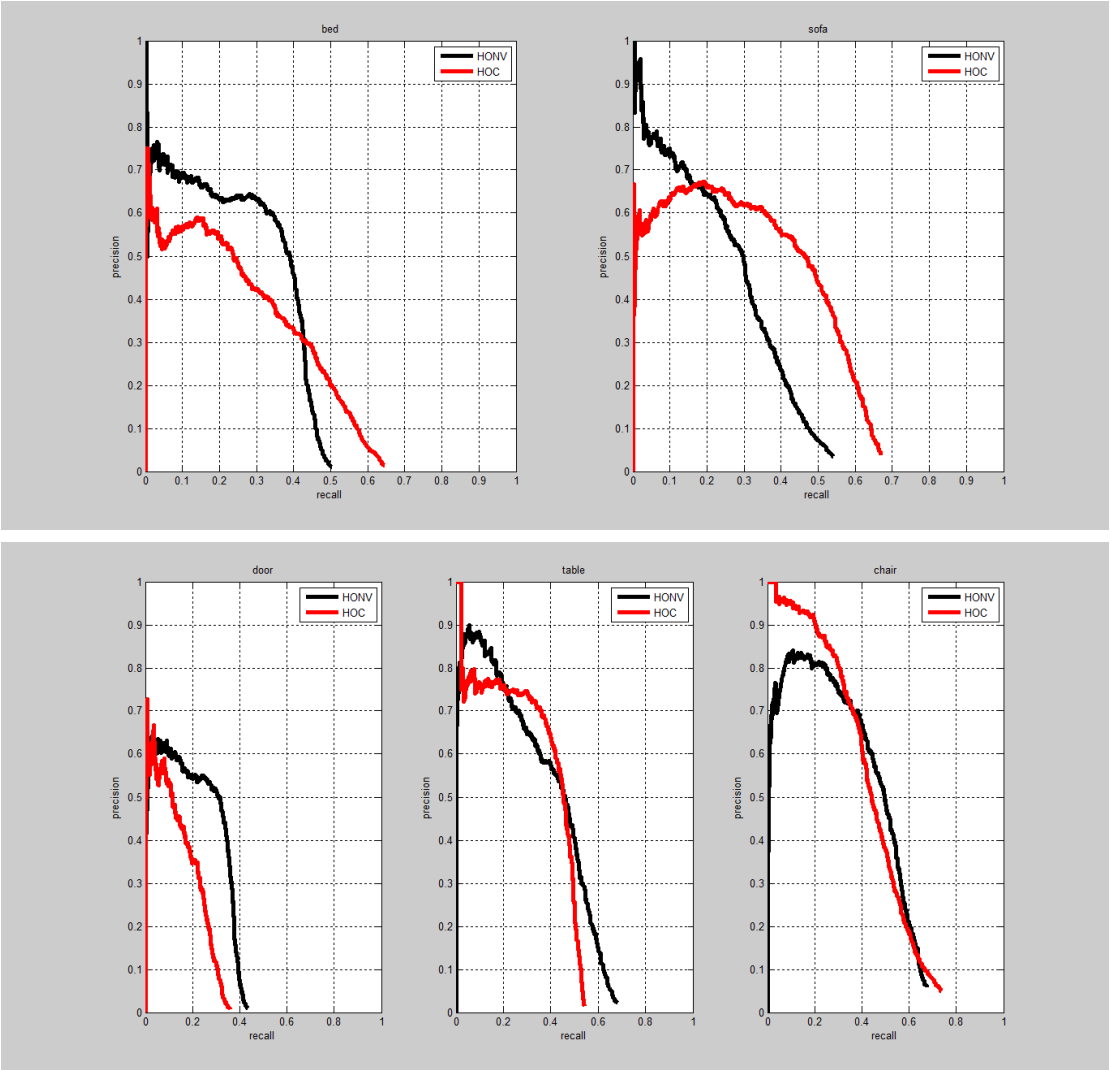

Figure 4.3 Precision-recall curves on the NYU-Depth Dataset V2

	HONV	HOC
bed	**27.6**	24.1
sofa	25.7	**34.0**
door	**20.6**	12.7
table	**38.0**	35.5
chair	40.7	**43.7**
AVERAGE	**30.5**	30

Table 4.2 Summarization of AP (%) on the NYU-Depth Object Dataset V2

From the above figure and table, we can see that the average performance on these five object categories of the HONV and HOC features are very close. As the image quality and labeling information for the NYU-Depth Dataset are not quite satisfied, the comparison of the results may not denote significance.

Figure 4.4 Example result obtained by HOC on the NYU-Depth Object Dataset V2

Since NYU-Depth dataset is originally for pixel wise scene segmentation and provides pixel-level labels, some instances are heavily occluded and in low resolution. That's the main reason why in the above image our method only successfully detect "sofa" and the real "door". Some objects are even hard to recognize for human eyes if they are not placed in an indoor scene.

Chapter 5

Summary

In this work, we proposed a novel feature for object detection using depth information: Histogram of Oriented Curvature (HOC). The HOC feature captures the local surface of an object to describe its 3D shape. Our experimental results show that the proposed HOC feature generally outperform the HOG and HOGD features in object detection task. And They achieve higher average precision than the state-of-the-art depth descriptor HONV for some of the object categories.

In the future, we will continue optimizing the HOC feature by trying different methods of capturing local surface curvature, and further eliminating the effect of noise. We will also explore the performance of the HOC feature on more tasks such as action recognition.

BIBLIOGRAPHY

[1] Microsoft Corp. http://www.xbox.com/en-US/kinect.

[2] PrimeSense Corp. http://www.primesense.com/.

[3] K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view rgb-d object dataset. In International Conference on Robotics and Automation, 2011.

[4] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[5] Sho Ikemura and Hironobu Fujiyoshi. Real-time human detection using relational depth similarity features. In The Tenth Asian Conference on Computer Vision, 2010.

[6] L. Xia, C.-C. Chen, and J. K. Aggarwal. Human detection using depth information by kinect. In Workshop on Human Activity Understanding from 3D Data in conjunction with IEEE Conference on Computer Vision and Pattern Recognition (HAU3D), 2011.

[7] L. Bo, X. Ren, and D. Fox. Depth kernel descriptors for object recognition. In IEEE/RSJ International Conference on Intelligent Robots and Systems, 2011.

[8] L. Bo, K. Lai, X. Ren, and D. Fox. Object recognition with hierarchical kernel descriptors. In IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[9] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Sparse distance learning for object recognition combining rgb and depth information. In International Conference on Robotics and Automation, 2011.

[10] Stone E and Skubic M. Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In Pervasive Health Conference, pages 71-77, 2011.

[11] Corinna Cortes and V. Vapnik. Support-vector networks. In Machine Learning, 1995.

[12] Thorsten Joachims. Making large-scale svm learning practical. LS8-Report 24, Universit at Dortmund, LS VIII-Report, 1998.

[13] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. In European Conference on Computer Vision, 2012.

[14] David G. Lowe. Object recognition from local scale-invariant features. In IEEE International Conference on Computer Vision, 1999.

[15] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In IEEE Conference on Computer Vision and Pattern Recognition, pages 886-893, 2005.

[16] Timo Ahonen, Abdenour Hadid, and Matti Pietikinen. Face recognition with local binary patterns. In ECCV, pages 469-481, 2004.

[17] Xiaoyu Wang, Tony X. Han, and Shuicheng Yan. An hog-lbp human detector with partial occlusion handling. In IEEE International Conference on Computer Vision, 2009.

[18] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. TPAMI, 1999.

[19] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. Bulletin of the Calcutta Mathematical Society, 35:99–109, 1943.

[20] Q. Cai, D. Gallup, C. Zhang, and Z. Zhang. 3d deformable face tracking with a commodity depth camera. In European Conference on Computer Vision, 2010.

[21] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. Rgbd mapping: Using depth cameras for dense 3d modeling of indoor environments. In RGB-D: Advanced Reasoning with Depth Cameras Workshop in conjunction with RSS, 2010.

[22] H. Du, P. Henry, X. Ren, M. Cheng, D. B. Goldman, S. M. Seitz, and D. Fox. Interactive 3d modeling of indoor environments with a consumer depth camera. In ACM International Conference on Ubiquitous Computing, 2011.

[23] E. Herbst, X. Ren, and D. Fox. Rgb-d object discovery via multiscene analysis. In IEEE/RSJ International Conference on Intelligent Robots and Systems, 2011.

[24] T. Wang, X. He, and N. Barnes. Learning structured hough voting for joint object detection and occlusion reasoning. In CVPR, 2013.

[25] Kai Yu and Tong Zhang. Improved local coordinate coding using local tangents. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), pages 1215-1222, 2010.

[26] Kai Yu, Tong Zhang, and Yihong Gong. Nonlinear learning using local coordinate coding. In Advances in Neural Information Processing Systems 22, pages 2223-2231. 2009.

[27] 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. ACCV'12.

[28] Dong Xu and Wenli Xu. Description and recognition of object contours using arc length and tangent orientation. Pattern Recognition Letters, pages 855-864, 2005.

[29] Antonio Monroy, Angela Eigenstetter, and Björn Ommer. Beyond straight lines—object detection using curvature. Image Processing (ICIP), 2011 18th IEEE International Conference on. IEEE, 2011.

[30] Hongli Deng, et al. Principal curvature-based region detector for object recognition. Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007.