

Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Invariant Hough random ferns for RGB-D-based object detection

Xiaoping Lou
Mingli Dong
Jun Wang
Peng Sun
Yimin Lin

Invariant Hough random ferns for RGB-D-based object detection

Xiaoping Lou,^{a,b} Mingli Dong,^{a,b,*} Jun Wang,^{a,b} Peng Sun,^{a,b} and Yimin Lin^{a,b}

^aBeijing Information Science & Technology University, Beijing Key Laboratory of Optoelectronic Measurement Technology, Beijing 100192, China

^bBeijing Information Science & Technology University, Beijing Engineering Research Center of Optoelectronic Information and Instrument, Beijing 100192, China

Abstract. This paper studies the challenging problem of object detection using rich image and depth features. An invariant Hough random ferns framework for RGB-D images is proposed here, which primarily consists of a rotation-invariant RGB-D local binary feature, random ferns classifier training, Hough mapping and voting, searches for the maxima, and back projection. In comparison with traditional three-dimensional local feature extraction techniques, this method is effective in reducing the amount of computation required for feature extraction and matching. Moreover, the detection results showed that the proposed method is robust against rotation and scale variations, changes in illumination, and part-occlusions. The authors believe that this method will facilitate the use of perception in fields such as robotics. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.55.9.091403](https://doi.org/10.1117/1.OE.55.9.091403)]

Keywords: object detection; RGB-D; invariant Hough random ferns.

Paper 151642SS received Nov. 21, 2015; accepted for publication Feb. 22, 2016; published online Mar. 11, 2016.

1 Introduction

In the field of computer vision, object detection has long been a challenging and important task. State-of-the-art algorithms deliver satisfactory results for two-dimensional images. Nonetheless, these methods suffer from limited variation and cluttered backgrounds.¹ Highly accurate RGB-D cameras that have recently been developed can easily provide high-quality three-dimensional (3-D) information (color and depth information).² Objects can thus be examined by acquiring color and depth information together, which is better than using only raw color images to learn feature representations.³

In order to solve the problems of rotation and scale, part-occlusions, and nonrigid transformations, recent RGB-D object detectors have featured the following tools:⁴ feature extraction using a rotation-invariant descriptor,⁵ part-based coding scheme using the generalized Hough transform,⁶ feature matching using machine learning frameworks,⁷ and so on.

A number of researchers have paid attention to object representation through depth images in order to improve detection performance, such as the incorporation of information regarding the shape and spatial geometry of objects.⁸ A representative sparse feature is the integration of 3-D coordinates with the color fast feature.⁵ However, the computational cost of local feature extraction and matching increases with the number of classifiers. Hence, this paper involves the computation of a fast RGB-D local binary feature (LBF) in polar coordinates, which have yielded remarkable results for object categorization under challenging conditions such as rotation variation and cluttered backgrounds. This is because it is easier to rotate the coordinates of the descriptor at a certain polar angle relative to the patch orientation in a polar coordinate system.

The generalized Hough transform has been successfully adapted to the problem of part-based object detection because it is robust against partial occlusion and slightly deformed shapes.⁹ Moreover, it is tolerant to noise, and can find multiple occurrences of a shape in the same processing pass. Its main disadvantage is that it requires a considerable amount of storage and extensive computation. However, it has been reported that Hough voting efficiency during object categorization can be improved using a highly efficient classifier.¹⁰

With regard to invariant Hough random ferns (IHRF),¹¹ this paper applies a random ferns classifier (RFC)¹² to a Hough transform to improve search speed and reduce the need for a large storage space for data. Furthermore, the Hough voting is performed in rotation-invariant Hough space, since each support point shares a stable polar angle and scalable displacement related to the center of the relevant object.

This paper is structured as follows: the framework for RGB-D object detection is presented in Sec. 2. Experimental results, including a comparison of the proposed method with state-of-the-art techniques, are provided in Sec. 3. The contributions of this paper and ideas for future research are discussed in Sec. 4.

2 Methodology

This section describes the procedure for RGB-D object detection based on IHRF. Figure 1 outlines the procedure for this approach.

Figure 1 formulates the rotation-invariant and multiscale object detection problem as a probabilistic Hough voting procedure. For this example, the IHRF is trained on images of coffee cups (color and depth) obtained from the RGB-D Object Dataset.¹³ Some positive [Fig. 1(a-1)] and negative [Fig. 1(a-4)] samples including color and depth images were provided for training. The depth value of the positive samples in their modeling center should then be recorded

*Address all correspondence to: Mingli Dong, E-mail: dongml@sina.com

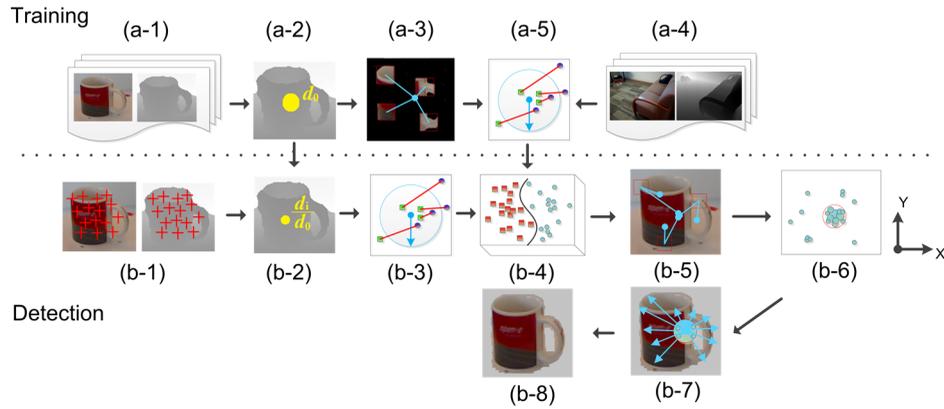


Fig. 1 An overview of training, represented by “a,” and detection, represented by “b.” (a-1) Positive samples. (a-2) Depth value recording. (a-3) Local coding. (a-4) Negative samples. (a-5) Feature extraction. (b-1) Corner detection. (b-2) Scale transform. (b-3) Feature extraction. (b-4) RFC. (b-5) Hough voting. (b-6) Finding local maxima in 2-D Hough space. (b-7) Back projection. (b-8) Detection results.

as d_0 [Fig. 1(a-2)]. For the positive image, the system extracts a large number of scanning windows within the color and depth images, and forms the local coding [Fig. 1(a-3)]. Following this, the rotation-invariant LBF [Fig. 1(a-5)] is extracted and used to train the RFC [Fig. 1(b-4)]. When presented with the image for detection, the system extracts a large number of interest points [Fig. 1(b-1)] and computes the scale value of the scanning windows, which is equal to the depth value d_i at the time divided by the original depth value d_0 [Fig. 1(b-2)]. The rotation-invariant LBF is then extracted in these scanning windows [Fig. 1(b-3)] and used for instance recognition [Fig. 1(b-4)]. The local scanning windows then cast probabilistic votes containing the locations of the centroid of the objects, which are collected in the voting space [Fig. 1(b-5)]. As a visualization of this space in Fig. 1(b-6) shows, the system searches for local maxima in the voting space and returns the correct detection as the strongest hypothesis. By back-projecting the contributing votes [Fig. 1(b-7)], the system retrieves support for the hypothesis in the image [Fig. 1(b-8)], and roughly separates the location of the object from the background. All the key steps are described in detail in subsequent sections.

2.1 Rotation-Invariant RGB-D Local Binary Feature

Random fern descriptors, also called LBF, consist of some logical pairwise comparisons of the intensity or gradient levels of randomly selected pixels in images.¹⁴ However, such comparisons are not robust against rotation and scale variations because each pairwise pixel is randomly generated offline while remaining fixed in runtime. Therefore, a rotation-invariant descriptor with a high degree of stability in RGB-D images is defined as

$$f(i, j, n) = \begin{cases} 1 & I_n(x_i, y_i) > I_n(x_j, y_j) \\ 0 & I_n(x_i, y_i) \leq I_n(x_j, y_j) \end{cases},$$

$$d(i, j) = \begin{cases} 1 & I_D(x_i, y_i) > I_D(x_j, y_j) \\ 0 & I_D(x_i, y_i) \leq I_D(x_j, y_j) \end{cases}, \quad (1)$$

where $I_n(x, y)$ and $I_D(x, y)$ are the n 'th ($n \in [1, 16]$) feature channel¹⁵ obtained from a color image and the n 'th depth

channel obtained from the depth image, respectively, with both centered at pixel locations x and y . (x_i, y_i) and (x_j, y_j) are random pairwise pixel locations, and each comparison returns 0 or 1. In general, the pairwise pixels S that are chosen map an image patch to a 2^{2S} -dimensional space of binary descriptors in each fern. According to Eq. (1), RGB-D LBF can be computed as

$$F_m = \{f_1, f_2, \dots, f_S, d_1, d_2, \dots, d_S\}_2, \quad (2)$$

where F_m is the m 'th fern and f (or d) is the i 'th binary feature. Therefore, the entire set of random ferns can be denoted by $F = \{F_1, F_2, \dots, F_K\}$. A trade off between performance and memory can be made by changing the number of ferns K and their sizes S .

To achieve orientation invariance, the pairwise pixels used in Eq. (1) can be calculated by the polar coordinates

$$\begin{cases} x_1 = R_1 \cos(\theta_1 + \theta_m) \\ y_1 = R_1 \sin(\theta_1 + \theta_m) \end{cases}, \quad \begin{cases} x_2 = R_2 \cos(\theta_2 + \theta_m) \\ y_2 = R_2 \sin(\theta_2 + \theta_m) \end{cases}, \quad (3)$$

where gradient orientations are rotated relative to the maximum gradient orientation (MGO) θ_m , and the polar coordinates R_i and θ_i can be converted to Cartesian coordinates x_i and y_i , respectively, by using the trigonometric functions sine and cosine.

Note that the fixed pole is located at the center of the image, and the fixed polar axis has the same direction as the MGO. This allows pairwise pixels to be matched correctly under arbitrary orientation changes between the two images. Therefore, by assigning a consistent orientation to each LBF based on local image properties, the rotation-invariant RGB-D LBF can be represented simply relative to this orientation and, therefore, be rendered invariant to image rotation. Furthermore, by reserving typical features and reducing redundancy features, satisfactory generalization performance and training efficiency of the classifier are guaranteed. Figures 2(a) and 2(b) show the results of rotation-invariant RGB-D LBF on color and depth images.

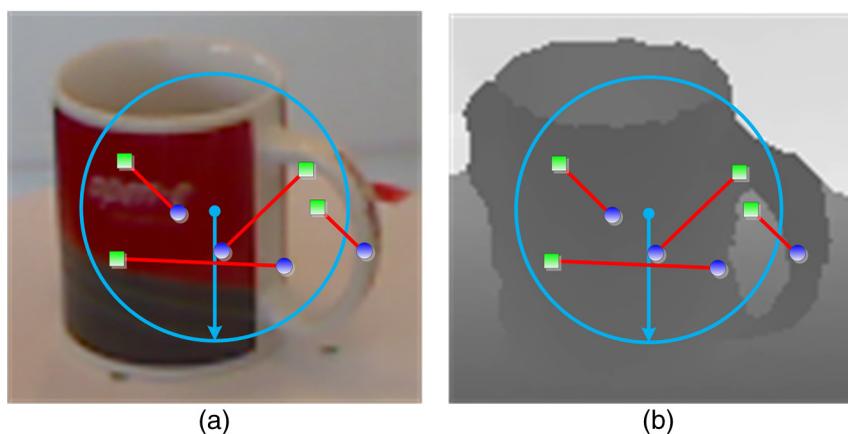


Fig. 2 The results of rotation-invariant RGB-D LBF on (a) color and (b) depth images.

2.2 Training the Random Ferns Classifier

Random ferns are of great interest in computer vision because of their speed, parallelization characteristics, and robustness against noisy training data. They are used for various tasks, such as keypoint recognition and image classification.¹² When they are applied to a large number of input vectors of the same class C , the output of each fern is a frequency distribution histogram, which is shown in Fig. 3. In the histogram, the horizontal axis represents a 2^S -dimensional space of binary descriptors and the vertical axis represents the number of times the binary code appears in class C , also called the class conditional probability $p(F_i/C)$, where $i \in [1, K]$.

Random ferns replace trees in random forests (RFs) with nonhierarchical ferns and pool their answers in a naive Bayesian manner to yield better results and improve classification rates in terms of the number of classes. As discussed in Sec. 2.1, the set of $F(\theta_{mt})$ located in a local patch with the MGO θ_{mt} is regarded as a class. Thus, a randomly selected patch detected in another image is assigned to the most likely class \hat{C} by calculating the posterior probability

$$\hat{C} = \underset{k}{\operatorname{argmax}} \prod_{L=1}^K p(F_L(\theta_{mt})/C_k). \quad (4)$$

For a given a test input, one can simply apply the binary representations to account for the ferns and look up the corresponding probability distribution over the class label, as shown in Fig. 3. Finally, the RFC selects the class with the highest posterior probability as the categorized result.

RFC is a remarkable classification algorithm that randomly selects and trains a collection of ferns. In this way, classifying new inputs involves only simple lookup operations.

2.3 Probabilistic Voting on Hough Space and Back Projection

We refer to the implicit shape model (ISM),¹⁶ which is a well-known approach based on the generalized Hough transform technique. During training, the ISM learns a model of the distribution of spatial occurrences of local patches with respect to the object's center. During testing, this learned model is used to cast probabilistic votes regarding the location of the object's center through the generalized Hough transform. The ISM is represented as

$$\text{ISM}(C) = [p(F/C), \theta_{mt}, d], \quad (5)$$

where θ_{mt} is the MGO of the local patch and d is the displacement vector from the center of the object to that of a local patch in its polar coordinate system. As a result, each fern in the IHRF consists of the ISM of each local patch belonging to object class C . Note that the size of an object used for training can be represented by a scale factor $s = 1$. For negative instances, the IHRF simply record their own class labels and pseudo-displacements.

This allows the classifier to exploit the available training data more efficiently because image patches representing the same object but in a different configuration (i.e., rotated or scaled) can be considered representations of the same type of information. During classification, the IHRF use the depth value d_i/d_0 to classify multiple scaled versions of the

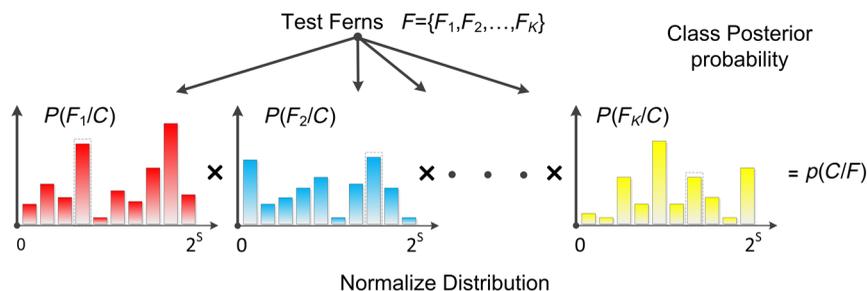


Fig. 3 Classification using an RFC, where “ \times ” is the symbol for multiplication.

image, which results in lower complexity. To integrate votes coming from the scanning grid pyramid of the input image Ω , they are accumulated into the Hough image H

$$\begin{aligned} H &= p(C/\Omega) = \sum_{Y \in \Omega} p[C/F(\theta_{\text{md}}, Y)] \\ &= \frac{1}{KN} \sum_{Y \in \Omega} \sum_{L=1}^K \sum_{d \in D} p[F_L(\theta_{\text{md}})/C_k] \delta_d \\ &\quad \times (Y_x - X_x - X_d, Y_y - X_y - Y_d), \end{aligned} \quad (6)$$

where N is the number of displacement vectors and D represents all displacement vectors. X is an object position used for the Hough vote, and (X_d, Y_d) is the connection vector relative to the position (Y_x, Y_y) at the given time. The subscripts x and y indicate the image position in the x and y directions, respectively. As a result, the value $p(\hat{C}/\Omega)$ serves as a confidence measure for hypothesis \hat{C} . After all the votes are cast, a global search for the local maxima yields the position of the center of the object as a nonparametric probability density estimate.

In addition to their voting capabilities pertaining to the hypothesis, the IHRF can be applied in reverse to detect the positions of their support. The location of a local maximum encodes scale, hypothesis \hat{C} , and its ISM of the object. More specifically, given the local maximum of a hypothesis at S_m , the support for this hypothesis \hat{C} is defined as the sample set

$$S_m = \bigcup_l \{\theta_{\text{md}}^l, \text{ISM}_l(\hat{C})\} = \bigcup_l \{\theta_{\text{md}}^l, p^l(F/\hat{C}), \theta_{\text{mt}}^l, d^l\}, \quad (7)$$

which contains the patch entries of all local samples l that have voted for the center S_m . By using their corresponding voting vectors d and MGO θ_{md} , IHRF back-projects the

original position of samples l onto the image space. In this way, a sparse point set of positions supposedly belonging to the object that had voted for the center position S_m is obtained.

3 Experiments and Results

This experimental section evaluates our method's performance and compares it with that of state-of-the-art approaches. It applies these methods to the challenging RGB-D Object Dataset¹⁷ and adheres to the experimental protocols and detection accuracy criteria established for each of the datasets in previous works. All experiments were conducted on a standard 3.2 GHz PC with 2 GB of RAM. For IHRF, the settings were as follows: the RFC consisted of $K = 10$ ferns, and picked $S = 13$ pairwise pixels for RGB-D LBF. Note that while using more ferns achieves higher recognition rates, it also requires more memory to store the distributions and results in a higher computational cost.¹² So, the Fern size of 10 and 13 features used for each Fern have proved to be a good compromise in our recent experiments.¹¹ The RGB-D Object Dataset¹⁷ is a large dataset of 300 common household objects, eight of which were used for training and detection, as shown in Fig. 4.

The goal of the experiment was to assess the accuracy of instance recognition using our method. According to the leave-sequence-out method,¹⁷ the first experiment involved training on the video sequences of each object, where the camera was mounted at 30 deg and 60 deg above the horizon, and evaluations on a 45-deg video sequence. The eight RGB-D objects shown here formed the largest multiview dataset, where both RGB and depth images were provided for each view. Therefore, this part tested whether combining RGB and depth is helpful when well-segmented or cropped images are available.

The recall-precision curve (RPC; see Fig. 5) was generated by changing the probability threshold on the vote strength of the hypothesis. Table 1 lists a performance

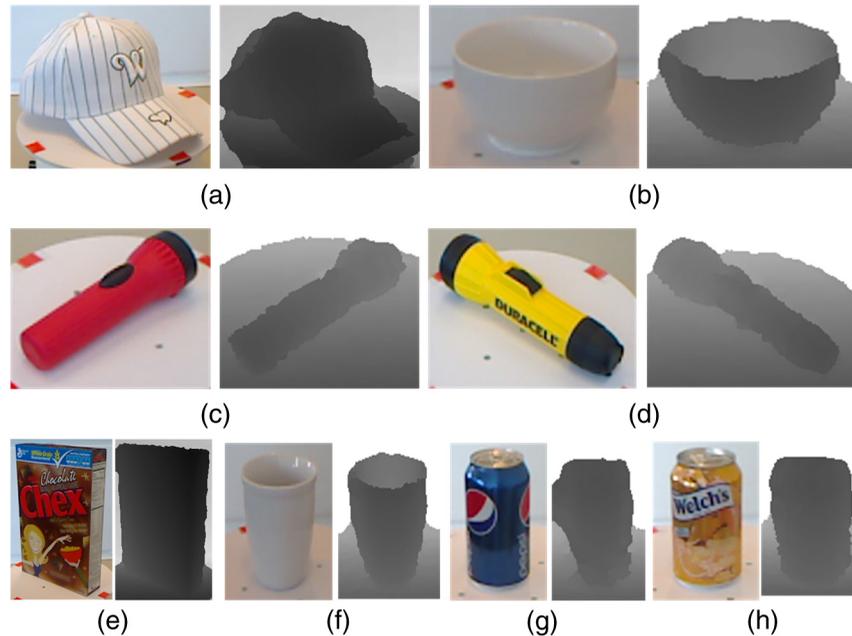


Fig. 4 Eight RGB-D objects used for training and detection. (a) Cap_1. (b) Bowl_4. (c) Flashlight_2. (d) Flashlight_5. (e) Cereal_box_2. (f) Coffee_mug_5. (g) Soda_can_1. (h) Soda_can_6.

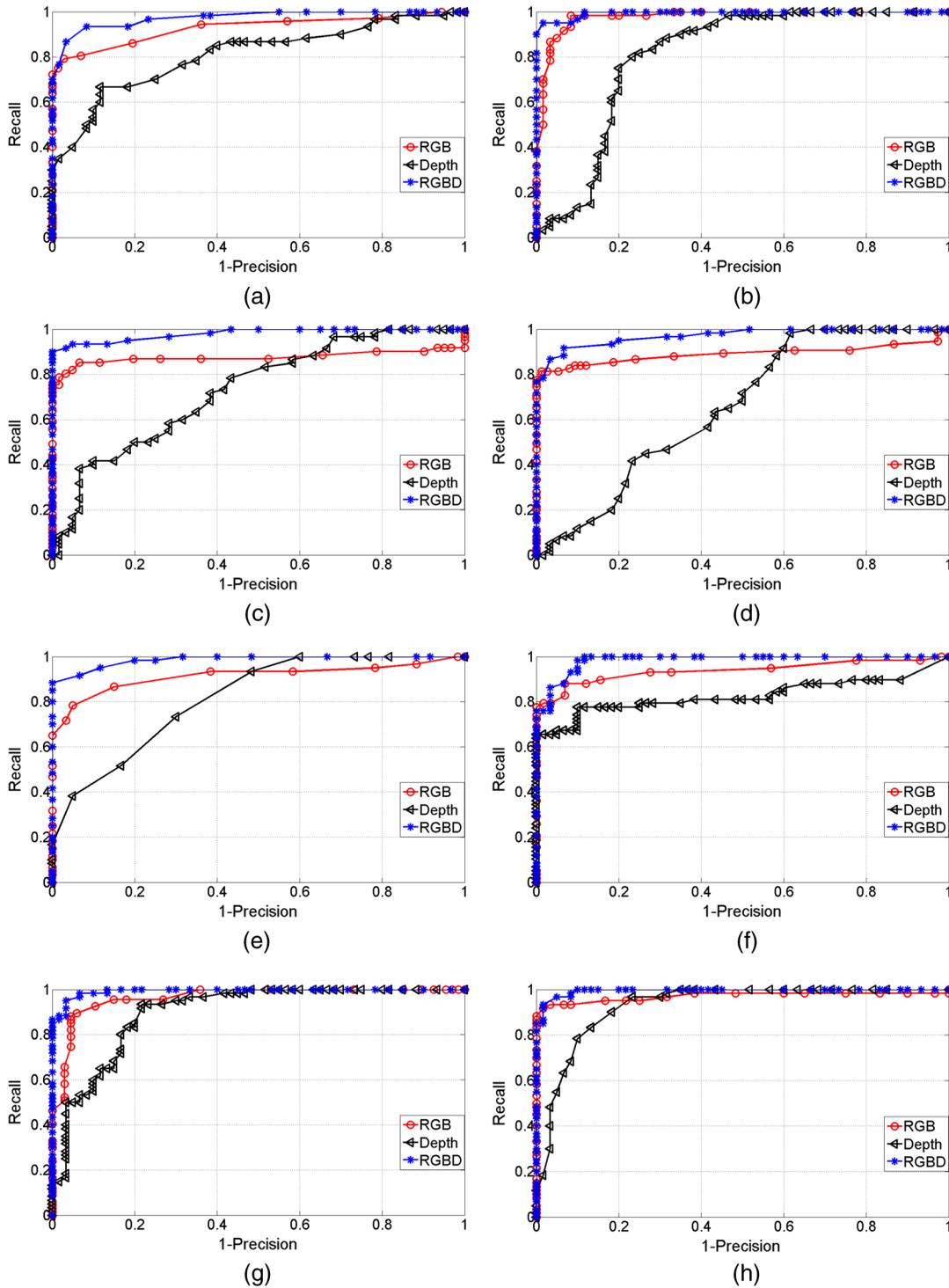


Fig. 5 RPC on eight RGB-D objects. All curves were generated by RGB, depth and RGB-D images, respectively. (a) Cap_1. (b) Bowl_4. (c) Flashlight_2. (d) Flashlight_5. (e) Cereal_box_2. (f) Coffee_mug_5. (g) Soda_can_1. (h) Soda_can_6.

comparison with the recognition results obtained by RGB, depth, and RGB-D images, respectively.

As shown in Fig. 5 and Table 1, RGB images attained an 87.4% EER, which was better than that for depth images (73.5%). This means that RGB images are more useful than depth images for instance-level recognition. This result showed that objects can easily use different textures and colors to distinguish among one another. The RGB-D approach

achieved an impressive 93.7% EER for the objects, outperforming both RGB and depth. Hence, the most significant conclusion is that combining RGB and depth images yields better performance. The leave-sequence-out evaluation was much more challenging and showed that combining shape and visual features significantly improves accuracy.

The second part of the experiment related to a set of quantitative experiments comparing our method with

Table 1 Performance of RGB, depth, and RGB-D images on eight objects in terms of recall-precision equal error rate (EER: %).

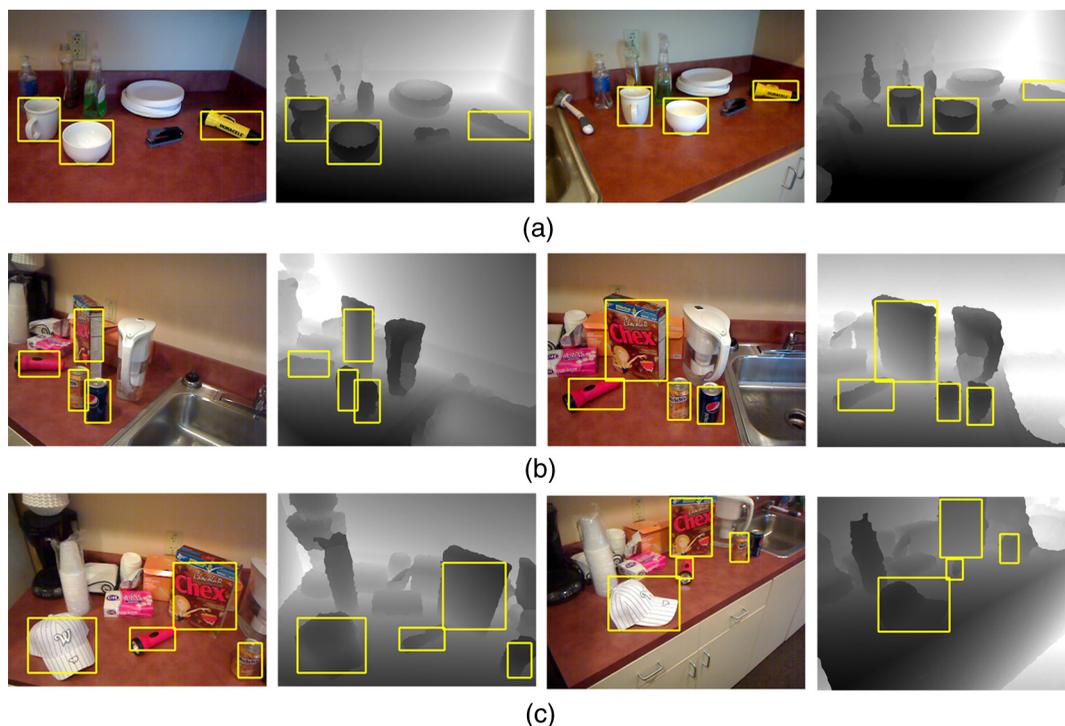
Objects	RGB	Depth	RGB-D
Cap_1	85.1	72.6	93.3
Bowl_4	88.3	80.0	95.0
Flashlight_2	85.3	63.3	93.2
Flashlight_5	81.3	56.6	91.6
Cereal_box_2	86.7	70.1	91.6
Coffee_mug_5	87.9	77.5	93.1
Soda_can_1	91.5	83.3	95.6
Soda_can_6	93.3	84.6	96.6
Means	87.4	73.5	93.7

Table 2 Comparison of different methods on the eight datasets using RGB, depth, and RGB-D images (EER: %).

Methods	RGB	Depth	RGB-D
LinSVM ¹⁷	59.3	32.3	73.9
RF ¹⁷	59.9	45.5	73.1
kSVM ¹⁷	60.7	46.2	74.8
RGB-D kernel descriptors ¹⁸	90.8	54.7	91.2
Hierarchical matching pursuit ¹⁹	92.1	51.7	92.8
Proposed method	87.4	73.5	93.7

other relevant algorithms, including linear support vector machine (LinSVM),¹⁷ RF,¹⁷ Gaussian kernel SVM (kSVM),¹⁷ RGB-D kernel descriptors,¹⁸ and hierarchical matching pursuit.¹⁹ Table 2 shows a comparison between the EER obtained using our method with IHRF and the results of other methods on the RGB-D dataset. All EER results were consistent with the first set of conclusions, whereby RGB-D images outperformed RGB and depth images regardless of classification technique. As can be seen from the results, our method attained impressive detection results for RGB-D images with an EER performance of 93.7%, which presents an improvement over results of past approaches. This situation is the same as the depth's EER 73.5% since the use of rotation-invariant LBF. Note that the EER of RGB images was 87.4%, which is slightly less than the corresponding values for the fourth and the fifth methods. This was because the RFC was $K = 13$, which refers to a trade off between performance, and memory and computational cost. If the value of K increases, performance improves efficiently.

Some examples of multi-instance detection are shown in Fig. 6. The results show that the RGB-D-based IHRF not only detects the object despite partial occlusion but also can often even deal with rotation variations and large perspectival changes. In Table 2, the first four methods all take approximately more than 2.5 s to label each scene from the RGB-D dataset, and the running time for hierarchical matching pursuit is 0.5 s. However, our method only required 0.4 s (no less than 200×150 pixel resolution) for object recognition. For conventional methods, the required amounts of memory increase linearly with the number of samples for each class. By contrast changing joint probability for features in each Fern does not result in increasing memory usage.


Fig. 6 The detection results on RGB-D datasets including the (a) first, (b) second, and (c) third scene.

4 Conclusions

This paper proposed a 3-D object detection method using RGB-D images based on IHRF. It relies on a rotation-invariant LBF based on RFCs that can cast probabilistic votes within the Hough transform framework. Experiment results show that such RGB-D-based IHRF can be efficiently used to detect instances of classes in large, challenging images with an accuracy that is superior to that of previous methods, and achieves the best results. This approach also allows for an efficient implementation in terms of time and space in comparison with related techniques. In the future, the authors intend to use the recognition results to improve the precision of object segmentation.

Acknowledgments

This work was supported by the National Natural Science Foundation of PR China (Grant No. 51475047), the Project of Construction of Innovative Teams, the Teacher Career Development for Universities and Colleges under Beijing Municipality (Grant No. IDHT20130518), and the Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT, Grant No. IRT1212).

References

1. I. K. Chen et al., "An integrated system for object tracking, detection, and online learning with real-time RGB-D video," in *Proc. 2014 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP '14)*, Vol. 5, pp. 6558–6562 (2014).
2. L. C. Caron, D. Filliat, and A. Geppert, "Neural network fusion of color, depth and location for object instance recognition on a mobile robot," in *Computer Vision—ECCV 2014 Workshops*, Springer International Publishing (2014).
3. K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Comput. Vision Image Understanding* **101**(1), 1–15 (2006).
4. T. Nakashika et al., "3D object recognition based on LLC using depth spatial pyramid," in *Proc. 22nd IEEE Int. Conf. on Pattern Recognition (ICPR '14)*, pp. 4224–4228 (2014).
5. P. Henry et al., "RGB-D mapping: using depth cameras for dense 3-D modeling of indoor environments," in *12th Int. Symp. on Experimental Robotics (ISER '10)*, pp. 477–491 (2010).
6. A. Srikantha and J. Gall, "Hough-based object detection with grouped features," in *IEEE Int. Conf. on Image Processing (ICIP)*, pp. 1653–1657 (2014).
7. S. Gupta et al., "Learning rich features from RGB-D images for object detection and segmentation," in *Computer Vision—ECCV 2014*, D. Fleet et al., Eds., Springer International Publishing, Switzerland (2014).
8. L. Bo, X. Ren, and D. Fox, "Learning hierarchical sparse features for RGB-D object recognition," *Int. J. Rob. Res.* **33**(4), 581–599 (2014).
9. R. Okada, "Discriminative generalized Hough transform for object detection," in *12th IEEE Int. Conf. on Computer Vision (ICCV '09)*, pp. 2000–2005 (2009).
10. M. Godec, P. M. Roth, and H. Bischof, "Hough-based tracking of non-rigid objects," in *Proc. Int. Conf. on Computer Vision (ICCV '11)*, pp. 81–88 (2011).
11. Y. Lin et al., "Invariant Hough random ferns for object detection and tracking," *Math. Probl. Eng.* **2014**, 1–20 (2014).
12. M. Ozuysal et al., "Fast keypoint recognition using random ferns," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(3), 448–461 (2010).
13. A. Janoch et al., "A category-level 3D object dataset: putting the Kinect to work," in *Consumer Depth Cameras for Computer Vision*, A. Fossati et al., Eds., pp. 141–165, Springer (2013).
14. M. Ozuysal, V. Lepetit, and P. Fua, "Fast keypoint recognition in ten lines of code," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8 (2007).
15. J. Gall and V. Lempitsky, "Class-specific Hough forests for object detection," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1022–1029 (2009).
16. B. Leibe and B. Schiele, "Interleaving object categorization and segmentation," in *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, H. I. Christensen and H. H. Nagel, Eds., pp. 145–161, Springer (2006).
17. K. Lai et al., "A large-scale hierarchical multi-view RGB-D object dataset," in *IEEE Int. Conf. on Robotics and Automation (ICRA '11)* (2011).
18. L. Bo, X. Ren, and D. Fox, "Depth kernel descriptors for object recognition," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS '11)*, pp. 821–826 (2011).
19. L. Bo, X. Ren, and D. Fox, "Unsupervised feature learning for RGB-D-based object recognition," in *Proc. of the 13th Int. Symp. on Experimental Robotics*, J. P. Desai et al., Eds., pp. 387–402 (2013).

Xiaoping Lou achieved her master's degree at Beihang University in 1998. Now, she is a professor at Beijing Information Science and Technology University. Her main research interests focus on machine vision, optical-electrical test technology, and so on.

Mingli Dong received her master's degree at HeFei University of Technology in 1989 and earned her PhD at Beijing Institute of Technology in 2009. Now, she is a professor at Beijing Information Science and Technology University. Her main research interests focus on machine vision, precise optical-electrical test technology, and so on.

Jun Wang received his PhD at Beijing University of Posts and Telecommunications in 2007. Now, he is an associate professor at Beijing Information Science and Technology University. His research areas include machine vision and vision metrology.

Peng Sun is a PhD student at the Institute of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications. His research interests include image analysis, pattern recognition, robot vision, and industrial photogrammetry.

Yimin Lin received his PhD at Beijing University of Posts and Telecommunications in 2014. His research areas include machine vision and pattern recognition.