Hand-Object Sense: A Hand-held Object Recognition System Based on RGB-D Information

Xiong Lv, Shuqiang Jiang, Luis Herranz, Shuang Wang
Key Laboratory of Intelligent Information Processing, Institute of Computing Technology Chinese Academy of Sciences, Beijing 100190, China
{xiong.lv,luis.herranz,shuang.wang}@vipl.ict.ac.cn; sqjiang@ict.ac.cn

ABSTRACT

Hand-held objects play an important role in human-human and human-machine interaction. It can be used as a reference for understanding user intentions or user requirements. In this technical demonstration, we introduce an object recognition system called Hand-Object Sense that can automatically recognize the object held by user. This system first detects and segments the hand-held object by exploiting skeleton information combined with depth information. Second, in the object recognition stage, this system exploits features computed in different ways and fuses them to improve the recognition accuracy. Our system can recognize objects in real-time and have a good tolerance to angle and scale transformation. Furthermore, it has a good generalization capability for unknown objects.

Categories and Subject Descriptors

I.4.6 [Image Processing and Computer Vision]: Segmentation—***Region growing, partitioning; Edge and feature detection***; I.5.4 [Pattern Recognition]: Application**—***Computer vision***

Keywords

RGB-D, hand-held object recognition, feature fusion

1. INTRODUCTION

Object recognition has many applications in human-machine interaction, information retrieval, mobile visual assistance, video surveillance etc. Specifically, hand-held object recognition is a special and important case of object recognition as it can not only help machine obtain more information about the user but also know some implicit intentions or interests of the user. Therefore, we propose a hand-held object recognition system called Hand-Object Sense.

Traditional object recognition based on RGB image meets the following problems: 1) clutter background makes it difficult to segment the object; 2) the same objects may have various visual appearances; 3) objects from different classes may look similar. With the availability of inexpensive RGB-D devices (e.g. Kinect), depth information can be exploited to alleviate these problems. Nevertheless, how to precisely locate and segment the object from background and extract representative and discriminative features to describe the object is still a challenging problem.

This system is developed based on our previous work[1], which takes advantage of RGB-D devices and uses specific segmentation technique based on skeleton and depth information to find hand-held objects. Thus it can effectively eliminate the background. Then a rich object representation is obtained by fusing different kind of features.

Hand-Object Sense can segment and recognize objects held by the user in real-time and simultaneously recognize the objects held in both hands. The system can allow the user rotate and move the objects freely.

MM'15, October 26–30, 2015, Brisbane, Australia.
ACM 978-1-4503-3459-4/15/10.
DOI: http://dx.doi.org/10.1145/2733373.2807990.

2. HAND-HELD OBJECT RECOGNITION SYSTEM

2.1 Framework

We now briefly describe the Hand-Object Sense framework. The details of the algorithm were elaborated in our previous work [1]. The block diagram of our framework is shown in Figure 1. The main steps are object segmentation, feature extraction, feature fusion and object recognition.

2.1.1 Hand-held Object Segmentation

First, we use the hand position located by Kinect API as the seed. Then the object mask is obtained using a region-growing algorithm. This algorithm examines the neighbors of the points in the seed set and includes them in the seed set when they are at a similar depth as the hand. Finally, we use the seed set to get the object region and its corresponding depth region.

2.1.2 Feature Extraction

When the RGB and depth regions of the object are acquired, we extract features from them. Point cloud is built by using RGB and depth information. In order to have a comprehensive and representative description of the object. We use three features together to represent object. They are Ensemble Shape Functions (ESF), Circular Color Cubic Higher-order Local Auto-Correlation descriptor (C^3-HALC) and Global Radius-based Surface Descriptor (GRSD).

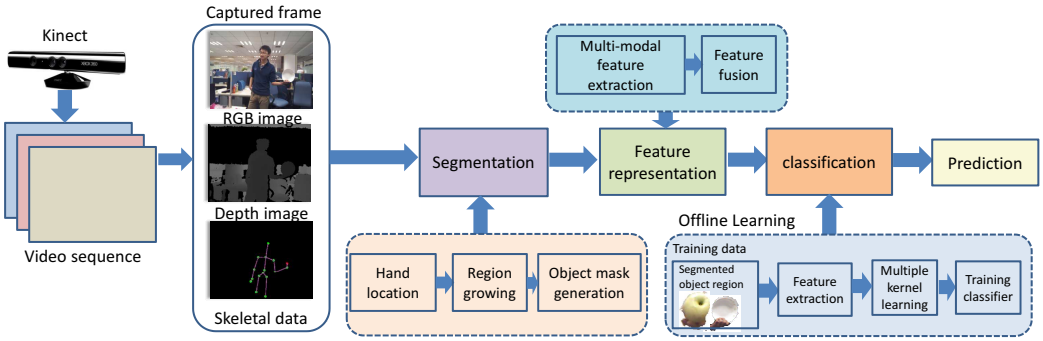

Figure 1: Hand-Object Sense Framework

2.1.3 Feature Fusion and Recognition

Different fusion methods lead to a very different classification performance. We consider two operations to fuse features: concatenation and multiple kernel learning (MKL). For MKL, we implement Gaussian kernels and polynomial kernels. After fusion, we train an SVM to predict the object category.

For each hand, we use the same procedure to predict the object category. The SVM and MKL are trained on our Hand-held object dataset(HOD) dataset[1], which contains 16 daily categories.

2.2 System Architecture

In order to effectively executes segmenting and recognizing simultaneously and ensure that our system can be implemented in real-time. We take advantage of parallel processing, Hand-Object Sense is implemented with a multi-thread architecture. Each thread runs as a separate module. In this way, we divide Hand-Object Sense into four modules: UI, segmentation, feature extraction and object recognition. Each of them is ran with a thread.

We use a voting scheme combining the predictions obtained from several video frames to improve the prediction accuracy. The voting window is 10, which means that we use the majority of the predicted label in 10 frames as the current label.

In human-machine interaction, we simultaneously use two hands, which requires our system to recognize the objects in both hands. We utilize the skeletal data to locate both hands and repeat the above operations for each of them.

2.3 Evaluation

Object recognition can be categorized into instance level and category level object recognition[2]. For instance-level object recognition, it indicates recognizing a concrete object, category-level object recognition indicates recognizing a basic concept. Thus, category-level object recognition has a better generalization capability for new objects. This demonstration focuses on category-level object recognition.

HOD dataset[1] was used to evaluate the performance of our Hand-Object Sense. We trained our classifier and MKL on the HOD dataset. The experiment was performed on a laptop with Intel Core i5 3.10GHz processor and 16GB DDR3, running Windows7. HOD contains 16 categories and each category includes 4 instances. We used all the 16 categories to train our system.

We invited 5 subjects to test our system with 16 categories. In the test, the 16 kinds of testing objects are the same category with the training data but different instances, user can hold object in single hand or both hands. It validates the generalization capability and robustness of our system. The average test accuracy is 65.96% and the recognition time is in real-time.

3. CONCLUSION AND FUTURE WORK

In this demo, we have presented Hand-Object Sense, a hand-held object recognition system. We employ a region-growing algorithm to segment objects and feature fusion for recognition. Hand-Object Sense works stably with high accuracy and scalability.

In the future, we will add more daily categories to make recognition more extensive. More features (e.g. deep features) and fusing method will be explored to make the recognition more robust. Furthermore, we will investigate incremental learning and more applications of our system and deploy the system into intelligent interaction platform.

Acknowledgement

This work was supported in part by the National Basic Research 973 Program of China under Grant No. 2012CB316400, the National Natural Science Foundation of China under Grant Nos. 61322212 and 61450110446, the National High Technology Research and Development 863 Program of China under Grant No. 2014AA015202, and the Chinese Academy of Sciences Fellowships for Young International Scientists under Grant No. 2011Y1GB05. This work is also funded by Lenovo Outstanding Young Scientists Program (LOYS).

4. REFERENCES

[1] Xiong Lv, Shuqiang Jiang, Luis Herranz, and Shuang Wang. Rgb-d hand-held object recognition based on heterogeneous feature fusion. *Journal of Computer Science and Technology*, 30(2):340–352, 2015.

[2] Shuang Wang and Shuqiang Jiang. Instre: A new benchmark for instance-level object retrieval and recognition. *ACM Trans. Multimedia Comput. Commun. Appl.*, 11(3):37:1–37:21, February 2015.