3D Object Recognition Based on ADAPTIVE-SCALE and SPCA-ALM in Cluttered Scenes

Dan XU, Xu-Zhi WANG, Wang-Gen WAN, Xiang-Jie LI
School of Communication & Information Engineering, Institute of Smart City, Shanghai University, Shanghai, China
1172039852@qq.com, wanwg@staff.shu.edu.cn, wangxzw@shu.edu.cn, 1364616657@qq.com

Abstract—In this paper a novel 3D object recognition method which can improve the recognition accuracy of object recognition in the cluttered scenes was proposed. The proposed method use the adaptive-scale to detect the keypoint (ASDK) of 3D object in the cluttered scenes, it use the algorithm of Sparse Principal Component Analysis Augmented Lagrangian Method (SPCA-ALM) to extract the feature of object, the algorithm of SPCA-ALM has a good performance in the high dimensional due to the Spares PCA, and the ALM can raise the speed of the SPCA. The experiment shows that the proposed method can decrease the time of 3D object recognition and improve the recognition accuracy.

1. Introduction

3D object recognition has been a widely used in the computer vision during the past several decades. Many significances works have been done, such as automatic navigation, remote sensing, mobile manipulation, intelligent surveillance, automatic assembly, robotics, biometric analysis and so on. In the past years, 2D Image recognition has been extensively investigated. But with the development of technology people are more interested in 3D scenes, 3D scenes have several advantages for object recognition due to its range images. For example, (i) Features extracted from range images are commonly not affected by scale, rotation and illumination. (ii) Compare to 2D images, it has more geometrical information from range images. Nowadays, Kinect can catch the range images, and it have the potential to overcome many difficulties which faced by 2D images in the context of object recognition. Existing 3D object recognition approaches can be divided into two main categories: the point cloud data [18]and rgb-d data. Lai K, Bo L, Ren X, et a proposed a large-scale hierarchical multi-view rgb-d object dataset [10]. Filliat D, Battiesti E, Bazeille S, et al use RGBD object recognition and visual texture classification for indoor semantic mapping [16]. In literature [1], the method of 3D object recognition in cluttered scenes with local surface features was proposed. However, those methods didn’t solve the problem that the accuracy of object recognition is relatively low in the cluttered scenes.

In this article, we proposed a novel method which can solve above problem. The proposed method firstly use the point cloud data to recognition the object in the cluttered scenes, and using adaptive-scale to detection keypoint, then using ASDK SPCA-ALM to select informative object features, third, we descript the feature and match the feature, finally, we make hypothesis generation and hypothesis verification.

Figure 1. The flow chart of Object Recognition

DOI: 10.1051/itmconf/20171201022
2. Object Recognition

This part is made by four steps as follows: Firstly, in Section A we should use adaptive-scale detection to get the keypoint; Then in Section B we mainly study about the informative feature selection; in Section C introduces the feature description and feature Match, Finally, in Section D we describes Hypothesis Generation and Hypothesis Verification. Figure.1 shows the flow chart of Object Recognition [17].

2.1. Adaptive-scale Keypoint Detection ASDK

3D object recognition have several steps, Keypoint detection is the first step and is also the most important step. Keypoint detection included surface mesh decimation and sparse sampling. Though several keypoint detection methods were proposed, the keypoint detected can’t meet the demand of experimental. Because when detecting the keypoint, they do not or seldom consider the difference of information points. Hence, we should according to their distinctiveness to detect keypoints. Keypoint detection methods can be broadly divided into two categories based on whether the scale is predetermined or adaptively detected: the one is fixed-scale keypoint detection methods and the other is adaptive-scale keypoint detection methods. As shown in Figure.2 is the adaptive-scale keypoint detection method, (a) is 3D Viewer and (b) is Range Image. Fixed-scale keypoint detection is detected the keypoint in a space which has a fixed scale. Adaptive-scale keypoint detection is to build a scale-space, after the scale-space was built in the range image, select extreme specificity measures for points in space and scale communities as keypoints. According to the information of the keypoint to change the scale-space. Then, the location and scale of the keypoints can be detected.

2.2. ASDK SPCA-ALM

Feature extraction is the second step of 3D object Recognition, in this paper we use the ASDK SPCA-ALM to extract the feature, firstly, detecting keypoint and then extracting the feature around it.

PCA method which can decompose the data matrix A’s empirical covariance matrix eigenvalue to compute the PCs, in general, the entries of the corresponding PVs of PCA are dense and nonzero. But in some conditions, the PV needs to be obtained which can use a combination of several nonzero variables to describe the maximum variance and it can increase the interpretability of these data. Sparse principal component analysis can determine the hidden information in the targeted data, which involving the physical interpretation of the coordinate axis. Thus, we computed an empirical covariance matrix for every object category of the database to recognize the visual words by Sparse PCA.

In the past decades, Sparse PCA has become a very popular research topic. Most widely used methods including SPCA, SCGLASS and SLRA, these methods intended to find modified PVs with sparse entries. Nevertheless, all the above method need to solve nonconvex objective functions. In order to overcome the drawback. In this paper we use ASDK to detect the keypoint and use Augmented Lagrangian Method (ALM) to improve the speed of SPCA, ASDK SPCA-ALM was rapid convergent and speed in $\ell$ - minimization. And it can solve the semidefinite relaxation form of Sparse PCA using ALM.

![Figure 2. Adaptive-scale keypoint detection](image)
The matrix variable $X = XX^T$. We rewrite this problem as:

$$\min_{U \in S^n} \lambda_{\max}(\Sigma + U) : -\rho \leq U_{ij} \leq \rho$$  \hspace{1cm} (3)

The maximum eigenvalue function $\lambda_{\max}(\cdot)$ can be approached by a smooth, once uniform objective:

$$f_\mu(U) = \mu \log \left( \text{Tr} \exp \left( \frac{\Sigma + U}{\mu} \right) \right)$$  \hspace{1cm} (4)

$$\nabla f_\mu(U) = \exp \left( \frac{\Sigma + U}{\mu} \right) / \text{Tr} \exp \left( \frac{\Sigma + U}{\mu} \right)$$  \hspace{1cm} (5)

The $\mu = \varepsilon / 2 \log(n)$ produces an $\varepsilon$- approximate solution. The formula (3) can be rewritten as:

$$\min_{U} f_\mu(U) : -\rho \leq U_{ij} \leq \rho$$  \hspace{1cm} (6)

Based on the formulation above established, then we use ALM to increase the speed of the ASDK SPCA. ALM is to eliminate constraints and add a penalty term to the cost function that describe a high cost to impossible points. We can rewrite an unconstrained form for the box-constraint convexity problem of (6):

$$F(U,Y) = \min_U \left\{ f_\mu(U) + \sum_{1 \leq i,j \leq n} P\left(U_{ij},Y_{ij},c\right) \right\}$$  \hspace{1cm} (7)

The $Y_{ij}, 1 \leq i, j \leq n$ is Lagrange variable, and $c$ is the variable which represent penalty severity.

$$P(\mu, y, c) = \begin{cases} \frac{y^2}{2c} & \text{otherwise}, \\ y(\mu - y) + \frac{c}{2}(\mu - y)^2 & \text{if } y - \frac{\mu}{c} \leq \mu, \\ y(\mu + y) + \frac{c}{2}(\mu + y)^2 & \text{if } y - \frac{\mu}{c} \leq \mu. \end{cases}$$  \hspace{1cm} (8)

It is worth to notice that we should solve the unconstrained minimization problem in (7) when we iteration of the outer loop of the algorithm, and this method has no closed-form solution. Thus, the first order gradient technique is used to obtain the best value. After getting the minimized of augmented Lagrange function, we use the rule to update the Lagrange multipliers $Y$.

We can get the primal variable by the gradient in (5), once the algorithm is convergent, i.e., $X^k = \nabla f_\mu(U^k)$. As the leading eigenvector of $X^k$ the sparse principal component can be recovered.

### 2.3. Feature Description and Match

The geometric information of the local surface can be extracted after detecting the keypoint, and it can be encoded into a feature descriptor. Recently the proposed methods can be classified into three categories: signature based, histogram based, and transform based methods. The core issue of the local feature description [16] are invariance and robust. When descriptor feature, we need to consider the invariance of perspective, scale, and rotation in shape recognition and object retrieval. And we also need to consider the invariant of shape of feature descriptor.

![Figure 3. feature match in the local RF, (left) the model feature. (right) the scene feature](image)

How to choose an appropriate matching strategy when searching the model feature library, the efficiency is an important issue. The simplest approach is to use a brute-force search, which matches a scene feature with all model features. The computational complexity of this approach is $O(N^2)$, the $N$ is the number of model features. A faster alternative is to adopt an appropriate data structure or indexing method. Hough transform is one of the basic methods of image processing which is widely used in the geometry image processing, there are many improvements proposed. The most basic Hough transform is detected from the black and white images straight line. As shown below: We make a Hof elections for the scene and the model in global coordinate, and then converted to a local coordinate, if the parameters get from the Hough election in the local coordinate are equal, that means election hypothesis is correct, otherwise the Hof elections is wrong. In the Fig.3 $y_{i,o}^m$ is the Hough variables of the model under the global
coordinates, $V_i^M$ is the Hough variables of model under the local coordinates, $F_j^M$ is the Feature points in the model library, $F_j^S$ is the Feature points in the scene, $C^M$ is the reference point, $R_{GL}^M$ is the conversion between Local coordinate and global coordinate, $V_i^S$ is the Hough variables of scene Under the local coordinates, $V_i^G$ is the Hough variables of scene Under the global coordinates. In Figure 4 shows the result of feature match, the chairs from scene match to the model library chair, the yellow chairs represent objects in the scene.

$$V_{i,G}^M = C^M - F_i^M$$ (10)

$$V_{i,L}^M = V_i^S$$ (11)

$$V_{i,L}^M = R_{GL}^M \cdot V_i^G$$ (12)

$$V_{i,G}^S = R_{LG}^S \cdot V_i^G + F_j^S$$ (13)

Figure 4. The chairs from scene match to the model library chair. The yellow chairs represent objects in the scene.

2.4. Hypothesis Generation and Verification

Hypothesis generation have two tasks: get potential candidate models in the scene and generate hypotheses for each candidate model conversion. Feature matching sometimes wrong, the point in the scene [9] and model to match the other point, which appears outliers, as shown in Fig 5, where C is the reference point, a point on the model left and right for the scene the point where the green dot on the spot did not match the model, became the outliers, the green dotted line represents a successful match, a red dotted line represents the match fails.

Figure 5. Generate hypotheses for each candidate model conversion. (left) the keypoint from model. (Right) the keypoint from scene.

The purpose of Hypothesis verification is to find the correct hypothesis from the false assumption that inside the existing authentication methods into separate authentication methods and global validation methods[11], hypothesis testing is independent of the assumptions used method of converting the objects in the scene aligned with the candidate models. Global hypothesis testing methods are as follows: 1. All the assumptions xi as a binary set x; 2. Geometrical Cues (goodness of fit, Multiple Assignment, randomness) Construction cost function, 3. optimization of the cost function (Simulated Annealing)

3. Experimental Results

In this paper, we use a laser scanner to get the point cloud data, then we use ASDK SPCA-ALM to extract feature among the keypoint, at the same time SPCA-ALM use to extract the feature from the same object and compare the two methods. Figure 6(a) shows that the estimation error increased with the dimension, when the dimension attach 300 the Error of the two methods are equal, then ASDK SPCA-ALM will have a smaller error rate compared with SPCA-ALM. while Figure 6(b) compares the speed of two methods, the chart shows that with the increased sharply ASDK SPCA-ALM spend less time. And in Figure 7 is the final recognition result of the chair from the model.

Conclusion

In order to get a better recognition rate of the object in the cluttered scenes, this paper presents a novel method which use Adaptive-scale Detection to get keypoint and use Sparse PCA to extract feature, then use Augmented Lagrangian Method to get the best value of recognition time and a better recognition accuracy. The experiment shows that the proposed method can decrease the time of 3D object recognition and improve the recognition accuracy. In the future, we will consider the combination of convolutional neural network and recurrent neural network to identify
object, currently machine learning is a hot research field. If we consider the machine learning method, I believe that it will get a better recognition performance.

Acknowledgement

This research was partially supported by the National Nature Science Foundation of China (No.61373084) and the Innovation Program of Shanghai Municipal Education Commission (No.14YZ011).

References

[1] Guo Y, Bennamoun M, Sohel F, et al. 3D Object Recognition in Cluttered Scenes with Local Surface Features:A Survey[J]. IEEE, 2014.

[2] Lai K, Bo L, Ren X, et al. Detection-based object labeling in 3D scenes[C]/Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012: 1330-1337.

[3] Bo L, Ren X, Fox D. Depth kernel descriptors for object recognition[C]/Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on. IEEE, 2011: 821-826.

[4] A. d’Aspremont, L. El Ghaoui, M. Jordan, and G. Lanckriet. A direct formulation for Sparse PCA using semidefinite programming. SIAM Rev., 2007.

[5] Lai K, Bo L, Fox D. Unsupervised feature learning for 3d scene labeling[C]/Robotics and Automation (ICRA), 2014 IEEE International Conference on. IEEE, 2014: 3050-3057.

[6] Guo Y, Sohel F, Bennamoun M, et al. A novel local surface feature for 3D object recognition under clutter and occlusion[J]. Information Sciences, 2015, 293: 196-213.

[7] An Y, Li Z, Shao C. Feature Extraction from 3D Point Cloud Data Based on Discrete Curves[J]. Mathematical Problems in Engineering, 2013.

[8] Lai K, Bo L, Ren X, et al. Sparse distance learning for object recognition combining rgb and depth information [C]/Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011: 4007-4013.

[9] Othman A, El Ghoul O. A novel approach for 3D head segmentation and facial feature points extraction [C]/Electrical Engineering and Software Applications (ICEESA), 2013 International Conference on. IEEE, 2013: 1-6.

[10] Lai K, Bo L, Ren X, et al. A large-scale hierarchical multi-view rgb-d object dataset[C]/Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011: 1817-1824.

[11] Gomes R B, da Silva B M F, de Medeiros Rocha L K, et al. Efficient 3D object recognition using foveated point clouds[J]. Computers & Graphics, 2013, 37(5): 496-508.

[12] Bariya P, Nishino K. Scale-Hierarchical 3D Object Recognition in Cluttered Scenes[J]. 2010, 119(5):1657-1664.

[13] Blum M, Springenberg J T, Wulffing J, et al. A learned feature descriptor for object recognition in rgb-d data[C]/Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012: 1298-1303.

[14] Lai K, Bo L, Ren X, et al. RGB-D object recognition: Features, algorithms, and a large scale benchmark [M]/Consumer Depth Cameras for Computer Vision. Springer London, 2013: 167-192.

[15] Silberman N, Hoiem D, Kohli P, et al. Indoor segmentation and support inference from RGBD images[M]/Computer Vision–ECCV 2012. Springer Berlin Heidelberg, 2012: 746-760.

[16] Filliat D, Battesti E, Bazeille S, et al. RGBD object recognition and visual texture classification for indoor semantic mapping[C]/Technologies for Practical Robot Applications (TePRA), 2012 IEEE International Conference on. IEEE, 2012: 127-132.

[17] Naikal N, Yang A Y, Sastry S S. Informative feature selection for object recognition via sparse PCA[C]/2011 International Conference on Computer Vision. IEEE, 2011: 818-825.

[18] Guan P, Huang J, Anvar A, et al. Multi-view 3D object recognition from a point cloud and change detection: U.S. Patent 9,619,691[P]. 2017-4-11.
Figure 6. A result of ASDK SPCA-ALM and SPCA-ALM

Figure 7. Final recognition out of the chair from the model