VISUAL FEEDBACK BASED OBJECT DETECTION AND MANIPULATION

SHRADDHA CHAUDHARY

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2018

VISUAL FEEDBACK BASED OBJECT DETECTION AND MANIPULATION

by

SHRADDHA CHAUDHARY

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2018

Dedicated to

My Teachers and Family

Certificate

This is to certify that the thesis titled "**Visual Feedback based Object Detection and Manipulation**", submitted by **Ms. Shraddha Chaudhary** to the **Department of Electrical Engineering**, Indian Institute of Technology Delhi, for the award of **Doctor of Philosophy** is a record of bona-fide research work carried out by her under my guidance and supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

Prof. Sumantra Dutta Roy
Professor
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi - 110016, India.

Acknowledgements

Working on the Ph.D has been a wonderful and overwhelming experience. In any case, I am thankful to many people for making the time of working on my Ph.D. an forgettable experience. I would like to express my sincere gratitude to my advisor Prof. Sumantra Dutta Roy. This thesis would not have been possible without his guidance, critical review and encouragement. His intelligent ideas and comprehensive understanding helped me to establish the overall direction of the research, got me past many significant challenges and always motivated me to strive for excellence. I would like to express my sincere thanks to the Bhaba Atomic Research Centre (BARC) Mumbai for its financial aid during my Ph.D. This gave me the opportunity to work on the project Vision Guided Control of a Robotic Manipulator in the Programme for Autonomous Robotics Lab. Thanks are also due to the IRD section IIT Delhi, for processing the financial support to attend the conferences whenever required.

I thank the members of my thesis committee: Prof. S.D Joshi, Prof. Prem Kalra and Dr. Brejesh Lall for their insightful suggestions which encouraged me to widen my research from various perspectives. I am also grateful to all the professors at IIT Delhi from whom I have learned a lot during the courses.

I sincerely admire the contribution of my lab mates Mr.Riby Abraham Boby, Mr. Ravi Prakash Joshi, Mr. Aamir Hayat, Mr. Anil Patidar and Ms. Tanya Raghuvanshi for extending their unstinting support and sympathetic attitude during the course of this study. I feel lucky to have so many incredible friends around. I thank each and every one of them for their trust and friendship.

I acknowledge the significant contribution of my parents in law. Their patience and support

will always be my inspiration. I would like to express my heartiest gratitude to my children, Master Pratyush and Master Prithvi for their love, and my indebtedness towards my husband, Piyush Arora, for his indefatigable assistance. Their infallible love and support has always been my strength. My Parents, Avinash Kumar Chaudhary and Shanti Chudhary, has always been a reassuring rock solid support for all my endeavours.

I thank the High Performance Computing facility at IIT Delhi for providing a multi-GPU enabled cloud computing platform which made training of deep models possible.

Shraddha Chaudhary

Abstract

Object detection along with pose estimation remains a quintessential computer vision problem. By pose estimation, we imply inferring the orientation by taking a 2D image and 3D point cloud of the object. This problem is motivated by the use of vision-based robotic manipulators. These are the robotic arms which rely at least in part, on cameras to obtain images of their surroundings. The thesis aims to provide an automated solution to the pick-and-place problem. Along with this, we have attempted subproblems of this broad scope objective, such as generalized object detection and pose estimation for different geometrical shapes using deep learning techniques.

Regarding the methodology, the first contribution proposes an autonomous machine vision system which grasps a texture-less object from a single layer of cluttered featureless objects of varied sizes, rearranges it for proper placement and then places it at a proper position using computer vision. It contributes to a unique vision-based pose estimation algorithm, collision-free path planning and a dynamic changeover algorithm for final placement. This framework has a well-defined application for the assembly task in industries. The second contribution of this thesis is the use of a low-resolution 3-D scan to guide a localized image search, to optimally utilize information from multiple sensors. This information guides segmentation of the image for better object pose estimation, for a pick-and-drop. We apply this algorithm for emptying a bin where cylindrical texture-less objects are lying in multi-layered clutter. The third contribution models uncertainty of the autonomous system proposed in the previous contribution. Application of the pick-and-place task using a 5-DOF Kuka YouBot is demonstrated in the thesis. The fourth contribution widens the scope of the thesis by working on benchmarked

datasets and moving more towards a robust system, by taking into account different viewpoints and refining the search process. We formulate a novel object classification algorithm with the assistance of viewpoint using an SVM. The fifth part extends the previous task using a deep neural architecture. We show encouraging results for object recognition, object instance prediction and object pose estimation, with the help of an Inception-based convolution neural networks. This architecture requires only RGB images instead of RGB-D required by previous models. We show an improvement with respect to state-of-the-art object detection models: we achieve an accuracy of 89.8%. This is the first work to use an Inception architecture along with 1×1 filters, for object pose estimation.

The thesis presents architectural features, training procedures and shares futuristic recommendations. The proposed algorithms and their extensions, and techniques have been tested on publicly available *de facto* 'standard' datasets, such as CIFAR and SVHN. We show encouraging results of the same, indicating that the hybrid architecture can culminate in a very competitive and robust model for pose determination of generalized objects.

सार

पॉज़ अनुमान के साथ ऑब्जेक्ट डिटेक्शन एक उत्कृष्ट कंप्यूटर दृष्टि समस्या बनी हुई है। अनुमान लगाकर, हम ऑब्जेक्ट की 2-D छवि और 3-D पॉइंट क्लाउड ले कर ओरिएंटेशन का उल्लंघन करते हैं। यह समस्या दृष्टि-आधारित रोबोट मैनिपुलेटर्स के उपयोग से प्रेरित है। ये रोबोटिक हथियार हैं जो कैमरे पर अपने आसपास के चित्रों को प्राप्त करने के लिए कम से कम भाग में भरोसा करते हैं। थीसिस का लक्ष्य पिक-एंड-प्लेस समस्या के लिए एक स्वचालित समाधान प्रदान करना है। इसके साथ-साथ, हमने इस व्यापक दायरे के उद्देश्य की उप-समस्याओं का प्रयास किया है, जैसे कि सामान्यीकृत ऑब्जेक्ट डिटेक्शन और गहरी सीखने की तकनीक का उपयोग करके विभिन्न ज्यामितीय आकारों के लिए अनुमान लगाया गया है।

पद्धति के बारे में, पहला योगदान एक स्वायत्त मशीन दृष्टि प्रणाली का प्रस्ताव करता है जो विभिन्न आकारों के अव्यवस्थित फीचरलेस ऑब्जेक्ट्स की एक परत से एक बनावट-कम वस्तु को पकड़ता है, इसे उचित प्लेसमेंट के लिए पुनर्व्यवस्थित करता है और फिर इसे कंप्यूटर दृष्टि का उपयोग करके उचित स्थिति में रखता है। यह एक अद्वितीय दृष्टि-आधारित मुद्रा अनुमान एल्गोरिदम, टकराव मुक्त पथ योजना और अंतिम प्लेसमेंट के लिए गतिशील परिवर्तन एल्गोरिदम में योगदान देता है। इस ढांचे में उद्योगों में असेंबली कार्य के लिए एक अच्छी तरह से परिभाषित आवेदन है। इस थीसिस का दूसरा योगदान एकाधिक सेंसर से जानकारी का बेहतर उपयोग करने के लिए स्थानीयकृत छवि खोज को मार्गदर्शन करने के लिए निम्न-रिज़ॉल्यूशन 3-डी स्कैन का उपयोग है। यह जानकारी एक पिक-एंड-ड्रॉप के लिए बेहतर ऑब्जेक्ट पॉज़ अनुमान के लिए छवि का विभाजन करती है। हम एक बिन खाली करने के लिए इस एल्गोरिदम को लागू करते हैं जहां बहु-स्तरित अव्यवस्था में बेलनाकार बनावट-कम वस्तुएं झूठ बोल रही हैं। तीसरे योगदान मॉडल पिछले योगदान में प्रस्तावित स्वायत्त प्रणाली की अनिश्चितता।

5-डीओएफ(DOF) कुका यूबीओटी(YouBOT) का उपयोग करके पिक-एंड-प्लेस कार्य का उपयोग थीसिस में प्रदर्शित किया गया है। चौथा योगदान विभिन्न दृष्टिकोणों को ध्यान में रखते हुए और खोज प्रक्रिया को परिष्कृत करके, बेंचमार्क किए गए डेटासेट पर काम करके और एक मजबूत प्रणाली की ओर बढ़कर थीसिस के दायरे को बढ़ाता है। हम एक एसवीएम का उपयोग करके दृष्टिकोण की सहायता से एक उपन्यास ऑब्जेक्ट वर्गीकरण एल्गोरिदम तैयार करते हैं। पांचवां भाग एक गहरे तंत्रिका वास्तुकला का उपयोग कर पिछले कार्य को बढ़ाता है। हम एक प्रारंभ-आधारित संकल्प तंत्रिका नेटवर्क की सहायता से ऑब्जेक्ट मान्यता, ऑब्जेक्ट इंस्टेंस भविष्यवाणी और ऑब्जेक्ट पॉज़ अनुमान के लिए उत्साहजनक परिणाम दिखाते

हैं। इस आर्किटेक्चर के लिए पिछले मॉडल द्वारा आवश्यक आरजीबी-डी(RGB-D) की बजाय केवल आरजीबी(RGB) छवियों की आवश्यकता है। हम अत्याधुनिक ऑब्जेक्ट डिटेक्शन मॉडल के संबंध में सुधार दिखाते हैं: हम 89.8% की शुद्धता प्राप्त करते हैं। ऑब्जेक्ट पॉज़ अनुमान के लिए, 1X1 फ़िल्टर के साथ एक इंसेप्शन आर्किटेक्चर का उपयोग करने वाला यह पहला काम है।

थीसिस वास्तुशिल्प सुविधाओं, प्रशिक्षण प्रक्रियाओं और भविष्य की सिफारिशों के शेयर प्रस्तुत करता है। प्रस्तावित एल्गोरिदम और उनके एक्सटेंशन, और तकनीकों का परीक्षण सार्वजनिक रूप से उपलब्ध 'मानक' डेटासेट्स जैसे सीआईएफएआर(CIFAR) और एसवीएनएन(SVHN) पर किया गया है। हम इसके बारे में उत्साहजनक परिणाम दिखाते हैं, यह दर्शाते हुए कि हाइब्रिड आर्किटेक्चर सामान्यीकृत वस्तुओं के निर्धारण के लिए एक बहुत प्रतिस्पर्धी और मजबूत मॉडल में समाप्त हो सकता है।

Contents

List of Figures

List of Tables