ROBUST SCENE GEOMETRY RECOVERY FROM EGO-CAMERAS FOR VISUAL NAVIGATION

SUVAM PATRA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

SEPTEMBER 2019

ROBUST SCENE GEOMETRY RECOVERY FROM EGO-CAMERAS FOR VISUAL NAVIGATION

by

SUVAM PATRA

Department of Computer Science and Engineering

Submitted

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

to the

Indian Institute of Technology Delhi
Semptember 2019

Certificate

This is to certify that the thesis titled **Robust Scene Geometry Recovery from Ego-cameras for Visual Navigation**, being submitted by **Mr. Suvam Patra** for the award of **Doctor of Philosophy** in Computer Science and Engineering, is a record of bona-fide work carried out by him under our guidance and supervision at the Computer Science and Engineering Department, Indian Institute of Technology Delhi. To the best of our knowledge, the work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

Subhashis Banerjee
Professor
Department of Computer Science
and Engineering
Indian Institute of Technology Delhi
New Delhi 110 016

Prem K. Kalra
Professor
Department of Computer Science
and Engineering
Indian Institute of Technology Delhi
New Delhi 110 016

Chetan Arora
Associate Professor
Department of Computer Science
and Engineering
Indian Institute of Technology Delhi
New Delhi 110 016

DEDICATED TO

My Parents and My Teachers.

Acknowledgments

I would like to sincerely thank all my teachers who guided me in my professional and educational growth and motivated me towards research. First and foremost, I would like to express my sincere gratitude towards my supervisors Prof. Subhashis Banerjee, Prof. Prem Kumar Kalra and Prof. Chetan Arora for their continuous guidance. I was fortunate enough to get research directions from them whenever I got stuck. Prof. Banerjee was always enthusiastic and there to help me whenever I needed help. He gave me the freedom to work and explore my problems on my own.

My co-supervisor Prof. Prem Kumar Kalra has always been there to hear me out patiently and help me out with his advice for solving my research problems and also sometimes for my personal development. He also helped me to improve my communication skills and my technical writing.

I would like to sincerely thank my co-supervisor Prof. Chetan Arora who helped me throughout my doctoral journey. His advice and guidance was to the point and precise. It helped me immensely in understanding my research problems and tackle them wisely. His timely responses and suggestions contributed highly to the critical milestones in my research.

I would also like to thank my research committee members Prof. Subodh Kumar, Prof. Naveen Garg and Prof. Sumantra Dutta Roy for their valuable reviews in all my research demonstrations.

I would like to express my gratitude towards Prof. Venu Madhav Govindu of Indian Institute of Science Bangalore for his valuable guidance and expert suggestions throughout my research work.

I would like to thank my dearest friend and my father who always motivated me to pursue higher education and research. I would like to express my gratitude towards my

mother and my brother who were always by my side during my most difficult times.

I would like to specially thank my friend and mentor Brojeshwar Bhowmick for his strong support and guidance which helped me to enhance my knowledge in computer vision. I would like to thank Britty Baby for helping me with my communication and technical writing skills. I would also like to thank my friends and colleagues Syamantak Das, Shibashis Guha, Dinesh Khandelwal, Shashank Sharma, Kinshuk Sarabhai, Chinmay Narayan, Gayathri Ananthanarayanan and Ankit Anand for their help and support.

I would also like to thank my friends and lab mates Kartikeya Gupta and Shashank Yadav who helped with my research work; and special thanks to Pranjal Maheshwari for his joint work entitled "A Joint 3D-2D based Method for Free Space Detection on Roads". I would like to thank the computer science department staff especially Mrs. Rekha Rathi, Mr. Hemant Singh, Mr. Surendra Negi, Mr K. R. Kaushik and Mr. Abhishek Sharma for their assistance.

Last but not the least, I would like to express my immense gratitude towards the almighty for his blessings and grace which helped me throughout my journey.

Suvam Patra

Abstract

In this thesis, we address the challenges in typical motion profiles of ego-cameras and design systems that utilize the camera motion to robustly estimate the location of a wearer or a vehicle and map its environment for navigation. In an ego-centric video, the camera views the same scene multiple times as the wearer's head sweeps back and forth. We use this specific motion profile to perform short visual loop closures aligned with the wearer's footsteps and use it to correct the pose inaccuracies arising out of wild camera motions [*Patra et al., WACV'17*]. Accordingly, we first propose a framework for robust camera pose estimation and use this framework successfully for solving different ego applications such as EgoSampling, Hyperlapse, Gaze Fixation, Temporal Segmentation and Activity Classification where state-of-the-art (SOTA) visual odometry (VO) methods have been reported to fail [1, 2, 3, 4]. This pose estimation framework estimates camera poses relative to their adjacent keyframes and lacks the sense of absoluteness in position and depth necessary for autonomous navigation. We thus improve our pose estimation method to propose a robust structure and pose estimation method for solving both the absolute positions and the environment map of the wearer.

We observe that incremental structure from motion (SfM) algorithms employed in most current VO methods in the presence of unreliable pose and 3D estimates from ego-centric videos often generate drift and eventually lead to failures. We address this issue by stabilizing the camera poses first using 2D techniques such as motion averaging

and then compute 3D structure using bundle adjustment [*Patra et al., WACV'19*]. Additionally, the use of domain knowledge from camera motion profile (e.g., local loop closures) aids the robustness of the proposed algorithm. We validate the accuracy of the estimated poses on different publicly available VO datasets. Further, visual navigation requires reliable estimates of camera position and structure, and hence we use our algorithm to aid visual navigation using ego-cameras mounted on cars.

The existing free space detection methods for autonomous navigation either involve 2D road segmentation techniques or 3D structure estimation techniques for identifying navigable spaces in front of the vehicle. We initially propose an improved 2D road segmentation method for providing strong road detection priors for unmarked roads, under varying illumination conditions [*Yadav and Patra et al., ICIP'17*]. The proposed conditional random field (CRF) based method uses SegNet [5] for modelling road texture and color lines model [6] for modeling illumination variations. This method can generalize across datasets and provide better road segmentation when compared to SOTA 2D road segmentation methods. However, 2D road segmentation methods do not provide the depth estimates for detecting navigable free space and fail on cases of uneven textures due to shadows, potholes, road restoration, etc. The use of 3D information becomes necessary to overcome these texture based classification failures. We propose an improved method for free space detection that uses a joint 3D/2D based CRF formulation using the generated higher level 3D road priors from our proposed structure and pose estimation algorithm and 2D priors from SegNet [*Patra et al., WACV'18*]. Both of these cues complement each other along with illumination invariance from the color lines model to create a comprehensive model for free space detection on roads. Experiments show the superiority of the proposed approach over SOTA. This work contributes to an extensive study on ego-cameras and provides insights for further research on camera pose and structure estimation using challenging motion profiles.

सारांश

इस थीसिस में, हम विशिष्ट गति प्रोफाइल में चुनौतियों का समाधान अहंकार कैमरा के उपयोग से करते हैं और सिस्टम डिजाइन करते हैं कि कैमरे की गति का उपयोग करने के लिए एक पहनने वाले या एक वाहन के स्थान का मजबूती से अनुमान लगाने और इसके नक्शे नेविगेशन के लिए वातावरण प्रस्तुत करते हैं । एक अहंकार केंद्रित वीडियो में, कैमरा विचार एक ही दृश्य कई बार पहनने वाले के सिर वापस आगे और पीछे होने के कारण दिखता है । हम लघु दृश्य पाश प्रदर्शन करने के लिए इस विशिष्ट गति प्रोफ़ाइल का उपयोग करने के लिए लूप क्लोजर के पहनने वाले के नक्शेकदम के साथ गठबंधन किया और इससे जंगली कैमरा गति से उत्पन्न होने वाली अशुद्धियों को हटाया ।

तदनुसार, हम पहले मजबूत कैमरे के लिए एक रूपरेखा का प्रस्ताव आकलन मुद्रा और इस ढांचे को हल करने के लिए सफलतापूर्वक उपयोग इस तरह के विभिन्न अहंकार अनुप्रयोगों में करते हैं जहां राज्य के अत्याधुनिक (SOTA) दृश्य ओडोमेट्री (वी.ओ.) तरीकों की सूचना दी गई है विफल होने का । यह मुद्रा आकलन ढांचे का पोसेस के उनके आसन्न कीफरम्स के साथ का रिश्ता के अनुमान करता है । हम इस प्रकार में सुधार एक मजबूत संरचना और मुद्रा का प्रस्ताव करने के लिए हमारी मुद्रा आकलन विधि निरपेक्ष पदों और दोनों को हल करने के लिए आकलन विधि पहनने का पर्यावरण नक्शा करते हैं । हम देखते हैं कि वृद्धिशील संरचना गति से (एसएफएम) एल्गोरिदम में सबसे वर्तमान वी.ओ. तरीकों में कार्यरत अर्वकी वीडियो से अविश्वसनीय मुद्रा और 3 डी अनुमान की उपस्थिति अक्सर बहाव उत्पन्न करते हैं । हम इस पते कैमरा स्थिर करके मुद्दा बन गया है पहले का उपयोग कर 2 डी तकनीक जैसे गति औसत v और फिर बंडल समायोजन का उपयोग कर 3 डी संरचना की गणना करते हैं । साथ ही, से डोमेन ज्ञान का उपयोग कैमरा गति प्रोफ़ाइल की एल्गोरिथ्म की मजबूती के लिए प्रस्तावित करते हैं । हम अनुमानित एल्गोरिथ्म की सटीकता को मान्य विभिन्न सार्वजनिक रूप से उपलब्ध वीओ डेटासेट पर करते हैं । इसके अलावा, दृश्य नेविगेशन कैमरा स्थिति और संरचना के विश्वसनीय अनुमान की आवश्यकता है, और इसलिए हम हमारे एल्गोरिथम का उपयोग गाड़ी के ऊपर स्तिथ अहंकारी कैमरे के मदद से चलने वाली दिश्य चलन के ऊपर करते है ।

स्वयक्त नवीगेशन के लिए मौजूदा मुक्त स्तल अन्वेषण तरीके या तो 2 डी विभाजन या फिर 3 डी सहजता अन्वेषण के ऊपर निरफर है । हम शुरू में एक बेहतर 2 डी सड़क का प्रस्ताव के लिए मजबूत सड़क का पता लगाने पूर्व प्रदान करने के लिए विभाजन विधि अचिह्नित सड़कें, अलग-अलग रोशनी की स्थिति में प्रस्तावित सशर्त यादृच्छिक फ़ील्ड (CRF) आधारित विधि के लिए मॉडलिंग सड़क बनावट और रंग लाइनों मॉडल के लिए SegNet का उपयोग करते है । यह विधि भर हर दतासेट के ऊपर सामान्यीकरण कर सकते हैं । और बेहतर सड़क विभाजन तकनीक प्रदान करते हैं । हालाकि 2 डी विभाजन तरीके मुक्त स्टल को ढूंडने के लिए गहराए अनुमान को नहीं प्रदान करता है और असामान बनावटों पे विफल होता है । हम इस के लिए एक बेहतर विधि का प्रस्ताव मुक्त अंतरिक्ष का पता लगाने कि एक संयुक्त 3 डी / 2 डी आधारित सीआरएफ का उपयोग कर तैयार का उपयोग करता है और हमारे प्रस्तावित ढांचे से उत्पन्न उच्च स्तर 3 डी सड़क पूर्व और आकलन एल्गोरिथ्म और सेगनेट पे आधारित तरीके को प्रस्तावित करते हैं । इन संकेतों के दोनों के साथ एक दूसरे के पूरक रंग लाइनों मॉडल से प्रसरण में रोशनी एक बनाने के लिए

सड़कों पर मुक्त अंतरिक्ष का पता लगाने के लिए व्यापक मॉडल प्रस्तावित करते हैं । प्रयोग दिखाते हैं की प्रस्तावित दृष्टिकोण की श्रेष्ठता SOTA के ऊपर दिखाते हैं । यह काम अहंकार कैमरा पर एक व्यापक अध्ययन करने के लिए योगदान देता है और कैमरा मुद्रा और संरचना आकलन का उपयोग पर आगे अनुसंधान के लिए चुनौतीपूर्ण गति प्रोफाइल अंतर्दृष्टि प्रदान करता है ।

Contents

List of Figures

List of Tables