A PIPELINE FOR LARGE-SCALE STRUCTURE-FROM-MOTION

BROJESHWAR BHOWMICK

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

JULY 2018

A PIPELINE FOR LARGE-SCALE STRUCTURE-FROM-MOTION

by

BROJESHWAR BHOWMICK

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
July 2018

Certificate

This is to certify that the thesis titled **A Pipeline for Large-scale Structure-from-Motion** being submitted by **Mr. Brojeshwar Bhowmick** 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

DEDICATED TO

My Parents and My Teachers.

Acknowledgments

I would like to thank all my teachers who guided me in my academic life and equipped me to take the doctoral training. First of all, I wish to express my sincere gratitude to my supervisor Prof. Subhashis Banerjee for his guidance, suggestions and support towards my research career. I have been fortunate enough to have a guide who gave me the freedom to explore on my own, and at the same time guide me to recover whenever I stumbled. I will always be thankful to him for his advice, guidance and helping me in my growth as a computer vision researcher.

My co-advisor, Prof. Prem. K. Kalra, has been always there to listen and give advice. I am grateful to him for the discussions that helped me sort out the problems regarding technology development, data acquisition etc. Prof. Kalra kindly arranged my visit to Hampi, at Karnataka from where I was able to collect data required for my research. He also guided me to visit Max Planck Institute for Informatics as an intern where I have learnt new techniques on motion capture using computer graphics. He always inspired me to improve my knowledge in both vision and graphics.

My sincere thanks to my research committee members, Prof. Subodh Kumar, Prof. K. K. Biswas, and Prof Santanu Chaudhury for their valuable suggestions from the proposal to thesis completion.

I would also like to extend my sincere thanks to Dr. Venu Madhav Govindu from Indian Institute of Science, Bengaluru, India for his suggestions which helped me in my thesis work.

I would like to thank my friends from IIT Delhi - Suvam Patra, Shibhasis Guha Syamantak Das, Kinshuk Sarabhai, Dinesh Khandelwal and Avishek Chatterjee from Indian Institute of Science, Bengaluru for their cooperation and support. Special thanks to Suvam who helped me a lot including allowing me to stay in his hostel room when I had no place to stay in Delhi. I would also like to thank the staff members of Computer Science & Engineering department especially our lab instructors Mr. K. R. Kaushik and Mr. Surendra S. Negi; office staff members Renu Bhatia, Rekha Rathi, Abhisek Sharma and Hemant Singh for their help.

I would also like to thank my colleague in TCS, Mr. Aniruddha Sinha and Dr. Arpan Pal for their support and encouragement in pursuing PhD at IIT Delhi. I would also thank TCS for my travel and stay at Delhi.

A special gratitude and love goes to my family for their support. I thank my parents for their love, support and sacrifices. Special thanks to my mother who supports me despite of many obstacles after demise of my father. My sister Shibani and her husband Bappaditya also deserve a special mention for their invaluable supports.

Finally, I would like to thank my wife Gargi for her support and sharing the hard work of day-to-day life during the most difficult times when I was not in home. Her patience, love, and encouragement is always a powerful source of inspiration and energy to me. I also want to convey my love and affection to my cute little son Aritra whose playful and naughty nature always keeps me happy and stress free.

Brojeshwar Bhowmick

Abstract

In this thesis we address the problem of large-scale Structure-from-Motion (SfM) using a large set of images. The images required for Structure-from-Motion can be acquired from a site or can be downloaded from websites like *Flickr* [28]. Apart from visualization, 3D structure and camera locations obtained from SfM is also useful for a variety of applications like autonomous navigation of robots, augmented reality etc.

In Structure-from-Motion, we recover both 3D structure and camera motion parameters by minimizing the reprojection error through bundle adjustment. Incremental SfM uses bundle adjustment by seeding the reconstruction using two images and then adds cameras incrementally into the existing reconstruction. Despite significant advances in incremental SfM techniques, such pipelines suffer from several problems. Apart from requiring significant computational power to solve the large-scale computations involved, such techniques often fail to correctly reconstruct a scene when the accumulated error due to drift is large or produce partial reconstruction when the number of 3D to 2D correspondences are insufficient for adding new cameras into the existing reconstruction. Moreover, such methods are precisely dependent on the initial seed pair.

As opposed to incremental methods, global SfM methods compute 3D structure and camera motions for all the images in a set simultaneously. Such methods require a good initialization for all the camera parameters and the 3D structure for bundle adjustment. To this end, these methods use motion averaging techniques to solve for global camera

motions from pairwise epipolar geometry estimates followed by the multi-view triangulation methods to compute the initial 3D structure. Then global bundle adjustment refines the initial 3D structure and the camera motion parameters simultaneously by minimizing reprojection errors. However, obtaining good initialization for global camera motion parameters using only pairwise epipolar relationships for large number of images ($\sim$ 30K images) is difficult due to noise present in pairwise epipolar geometry estimates. Thus, incremental methods and global approaches define two prototypical extreme solutions towards the large-scale SfM with their own set of problems.

In this thesis, we propose a divide and conquer based Structure-from-Motion pipeline where the entire corpus of images is partitioned into clusters such that the images within a cluster are related to each other by dint of their viewing common parts of the 3D scene. Such an approach drastically reduces the computational load for matching image pairs without compromising accuracy. It also makes the resulting SfM problems well-conditioned. We solve the SfM problem in each cluster using motion averaging followed by global bundle adjustment. As each cluster is reconstructed independently, they may not share common 3D points to merge two reconstructions in a common frame of reference. Hence, we also propose a method for registering these individual reconstructions using the available epipolar relationships only that connect images across clusters. Our 3D registration or alignment technique works well even when the available camera-point relationships are poorly conditioned. We also selectively apply a step of global bundle adjustment on the combined clusters after successive merging. The overall result is a robust, accurate and efficient pipeline for large-scale SfM. We present extensive experimental results that demonstrate these attributes of our pipeline on a number of large-scale, real-world publicly available datasets and compare our method with the state-of-the-art. [1]

[1] The code is available at https://github.com/IITD-COMPUTER-VISION-GROUP/lsfm

अब्स्त्रक्त

इस सिद्धांत में हम बड़े पैमाने पर स्ट्रक्चर-फ्रॉम -मोशन (एसएफएम) की समस्या का समाधान करते हैं छवियों के एक बड़ा सेट से। स्ट्रक्चर-फ्रॉम -मोशन के लिए आवश्यक छवियों को अधिग्रहित किया जा सकता हैसाइट से या फ़्लिकर जैसी वेबसाइटों से डाउनलोड किया जा सकता है। विज़ुअलाइज़ेशन के अलावा, एसएफएम से प्राप्त 3-D संरचना और कैमरा स्थान भी विविधता के लिए उपयोगी है रोबोटों की स्वायत्त नेविगेशन, बढ़ी हुई वास्तविकता आदि जैसे अनुप्रयोगों के लिए ।

स्ट्रक्चर-फ्रॉम -मोशन में, हम दोनों 3D संरचना और कैमरा गति पैरामीटर पुनर्प्राप्त करते हैं बंडल समायोजन के माध्यम से पुन: प्रक्षेपण त्रुटि को कम करके। वृद्धिशील एसएफएम दो छवियों का उपयोग करके पुनर्निर्माण बीजिंग द्वारा बंडल समायोजन का उपयोग करता है और फिर जोड़ता है मौजूदा पुनर्निर्माण में कैमरे वृद्धिशील रूप से। साइनि कैंट अग्रिम के बावजूद वृद्धिशील एसएफएम तकनीकों, ऐसी पाइपलाइनों कई समस्याओं से। आवश्यकता के अलावा बड़े पैमाने पर गणनाओं को हल करने के लिए साइनि कैंट कम्प्यूटेशनल पावर, संचित त्रुटि के दौरान ऐसी तकनीकें अक्सर दृश्य के पुनर्निर्माण में असफल हो जाती हैं बहाव के कारण 3D से 2D डी पत्राचार की संख्या होने पर आंशिक पुनर्निर्माण का उत्पादन होता है मौजूदा पुनर्निर्माण में नए कैमरे जोड़ने के लिए अपर्याप्त हैं। इसके अलावा, ऐसी विधियां प्रारंभिक बीज जोड़ी पर निर्भर करती हैं।

वृद्धिशील तरीकों के विपरीत, वैश्विक एसएफएम विधियों की गणना 3 डी संरचना और एक साथ एक सेट में सभी छवियों के लिए कैमरा गति। इस तरह के तरीकों को एक अच्छा की आवश्यकता है बंडल समायोजन के लिए सभी कैमरा पैरामीटर और 3 डी संरचना के लिए प्रारंभिकरण। इस अंत में, ये विधियां वैश्विक कैमरे के लिए हल करने के लिए गति औसत तकनीकों का उपयोग करती हैं जोड़ीदार एपिप्लोर ज्यामिति अनुमानों से गति बहु-दृश्य त्रिभुज के बाद होती है शुरुआती 3 डी संरचना की गणना करने के तरीके। फिर वैश्विक बंडल समायोजन प्रारंभिक 3 डी संरचना और कैमरा गति पैरामीटर एक साथ फिर से रेप्रॉजेक्शन त्रुटियों को कम करना। हालांकि, वैश्विक कैमरे के लिए अच्छी शुरुआत प्राप्त करना बड़ी संख्या के लिए केवल पैरविसे एपीपोलर संबंधों का उपयोग कर गति पैरामीटर छवियों (~30K छवियां) पंथ पैरविसे एपीपोलर ज्यामिति में मौजूद शोर के कारण है अनुमान। इस प्रकार, वृद्धिशील

तरीकों और वैश्विक दृष्टिकोण डी दो प्रोटोटाइपिकल बड़े पैमाने पर एसएफएम की समस्याओं के अपने सेट के साथ चरम समाधान।

इस थीसिस में, हम एक विभाजन का प्रस्ताव देते हैं और आधारित संरचना-से-मोशन पाइपलाइन पर आधारित होते हैं जहां छवियों का पूरा कॉर्प क्लस्टर में विभाजित होता है जैसे कि छवियांएक क्लस्टर एक दूसरे से 3 डी के सामान्य भागों को देखने के संकेत से संबंधित हैं दृश्य। इस तरह के दृष्टिकोण मिलान छवि के लिए कम्प्यूटेशनल लोड को काफी हद तक कम कर देता है सटीकता समझौता किए बिना जोड़े। यह परिणामस्वरूप एसएफएम समस्याओं को भी बनाता है अच्छी तरह से वातानुकूलित। हम गति औसत का उपयोग कर प्रत्येक क्लस्टर में एसएफएम समस्या को हल करते हैं वैश्विक बंडल समायोजन के बाद। चूंकि प्रत्येक समूह को स्वतंत्र रूप से पुनर्निर्मित किया जाता है, वे आम तौर पर दो पुनर्निर्माणों को मर्ज करने के लिए सामान्य 3 डी बिंदु साझा नहीं कर सकते हैं सम्बन्ध का दायरा। इसलिए, हम इन व्यक्तिगत पुनर्निर्माणों को पंजीकृत करने के लिए एक विधि भी प्रस्तावित करते हैं उपलब्ध एपिपलर संबंधों का उपयोग केवल छवियों को जोड़ता है समूहों। हमारे 3 डी पंजीकरण या संरेखण तकनीक उपलब्ध होने पर भी अच्छी तरह से काम करती है कैमरा-बिंदु संबंध खराब स्थिति में हैं। हम चुनिंदा रूप से एक कदम भी लागू करते हैं लगातार विलय के बाद संयुक्त समूहों पर वैश्विक बंडल समायोजन का। समग्र परिणाम बड़े पैमाने पर एसएफएम के लिए एक मजबूत, सटीक और ई? सीआईटी पाइपलाइन है। हम उपस्थित है व्यापक प्रयोगात्मक परिणाम जो हमारी पाइपलाइन के इन विशेषताओं को प्रदर्शित करते हैं बड़े पैमाने पर, असली दुनिया सार्वजनिक रूप से उपलब्ध डेटासेट की संख्या और हमारी विधि की तुलना करें अत्याधुनिक के साथ।

Contents

List of Figures

List of Tables