Efficient & Effective Image-Based Localization

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von Diplom-Informatiker
Torsten Sattler
aus Düsseldorf, Deutschland

Berichter: Prof. Dr. Leif Kobbelt
Prof. Dr. Bastian Leibe
Prof. Dr. Marc Pollefeys

Tag der mündlichen Prüfung: 25.10.2013

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar.

Selected Topics in Computer Graphics

herausgegeben von
Prof. Dr. Leif Kobbelt
Lehrstuhl für Informatik 8
Computergraphik & Multimedia
RWTH Aachen University

Band 11

Torsten Sattler

Efficient & Effective Image-Based Localization

Shaker Verlag
Aachen 2014

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: D 82 (Diss. RWTH Aachen University, 2013)

Printed in Germany.

ISBN 978-3-8440-2743-3
ISSN 1861-2660

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • e-mail: info@shaker.de

Abstract

The problem of image-based localization is the problem of accurately determining the position and orientation from which a novel photo was taken relative to a 3D representation of the scene. It is encountered in many interesting applications such as pedestrian or robot navigation, Augmented Reality, or Structure-from-Motion, creating a strong need for algorithms solving the image-based localization problem. In this thesis, we therefore present solutions to this problem that are both effective and efficient, *i.e.*, we propose methods that can localize novel query images taken under a wide range of viewing conditions while requiring only a small amount of processing time.

We assume that the 3D scene representation is obtained by using Structure-from-Motion techniques to reconstruct the environment from a set of photos. As a result, we can associate each 3D point with multiple image descriptors modeling the local appearance of the scene around this point. We can then obtain 2D-3D correspondences between 2D feature points in the query image and 3D scene points in the model by solving a descriptor matching problem. These 2D-3D matches can in turn be used to estimate the camera position of the query image, *i.e.*, the position and orientation from which it was taken. The main difficulty of descriptor matching lies in the sheer size of the problem, since our models contain millions of 3D points while thousands of features are found in our query images. As a major contribution, we show that the resulting descriptor matching problem can still be solved very efficiently using prioritized search. We propose a prioritization scheme that is easy to implement, yet can be expected to perform close to optimal in practice. By combining our prioritization with a novel active search step that is able to discover additional matches, we are able to derive an image-based localization approach that achieves or surpasses state-of-the-art effectiveness while offering the fastest run-times published so far.

Analyzing such direct matching methods, we demonstrate that their major advantage, namely their ability to identify a set of high-quality matches, also prevents their scalability to larger datasets. Consequently, we also consider image retrieval methods for image-based localization since they are inherently more scalable. As a second major contribution, we identify the algorithmic factors preventing image retrieval methods to achieve the same effectiveness as our original system and propose a modification that is able to close the gap in effectiveness without sacrificing scalability.

Zusammenfassung

Das Ziel von bildbasierten Lokalisierungsverfahren ist es, für ein gegebenes Fotos die Position und Ausrichtung der dazugehörigen Kamera relativ zu einem 3D Szenenmodel zu bestimmen. Das entsprechende Problem der bildbasierten Lokalisierung findet dabei viele praktische Anwendungen, wie z.B. Fußgängernavigation, Augmented Reality und Structure-from-Motion. In dieser Arbeit stellen wir effektive und effiziente Ansätze zur Lösung dieses Problems vor, d.h., wir präsentieren Verfahren welche die Position und Orientierung der Kamera für eine große Bandbreite von Blickpunkten und Beleuchtungsbedingungen in kurzer Zeit berechnen können.

Im folgenden gehen wir davon aus, dass das 3D Szenenmodell durch eine Structure-from-Motion Rekonstruktion der Umgebung aus einer Menge von Bilder erzeugt wurde. Dies erlaubt es uns jedem 3D Punkt mehrere Featuredeskriptoren zuzuweisen, welche das Aussehen der Szene um diesen Punkt herum beschreiben. Folglich können wir 2D-3D Korrespondenzen zwischen Featurepunkten im Anfragebild und 3D Punkten im Modell mit Hilfe der dazugehörigen Deskriptoren bestimmen. Diese Korrespondenzen erlauben es uns wiederum die Position und Ausrichtung der Anfragekamera zu berechnen. Die Hauptschwierigkeit beim Deskriptorenvergleich liegt dabei in der Größe des betrachteten Problems da unsere Szenenmodelle mehrere Millionen 3D Punkte enthalten während tausende von Featuren in den Anfragebildern gefunden werden. Als ein Hauptbeitrag dieser Arbeit zeigen wir, dass selbst solche großen Vergleichsprobleme immer noch effizient mittels prioritisierten Suchverfahren gelöst werden können. Wir stellen dabei ein einfach umzusetzendes Prioritisierungsverfahren vor, welches in der Praxis trotzdem eine nahezu optimale Lösung darstellt. Wir verbinden dabei unsere Prioritisierungsstrategie mit einem neuen Ansatz der aktiv nach weiteren Korrespondenzen sucht. Das resultierende Verfahren zur bildbasierten Lokalisierung erreicht dabei die schnellsten Laufzeiten die bisher veröffentlicht wurden während es andere Verfahren in Effektivität erreicht oder sogar übertrifft.

Wir zeigen außerdem, dass die große Stärke dieser Klasse von Verfahren, ihre Fähigkeit qualitativ hochwertige Korrespondenzen zu finden, gleichzeitig deren Anwendbarkeit auf beliebig große Datensätze verhindert. Im letzten Teil der Arbeit beschäftigen wir uns daher mit besser skalierenden Ansätzen und zeigen wie diese Skalierbarkeit mit Effizienz und Effektivität in Einklang gebracht werden kann.

Acknowledgments

I want to thank my parents for their never-ending love and support. Without them, none of what I have achieved would have been possible.

I thank my advisors, Leif Kobbelt and Bastian Leibe, for providing this opportunity and a nurturing environment in which I could grow both academically and as a person. I am very grateful for all the (technical) discussions and the guidance and help they offered me over all the years. They provided direction when I got lost and they are the reason I fell in love with Computer Vision and Computer Graphics.

I thank all my colleagues at the Computer Graphics and Computer Vision groups, Alex, Alexander, Arne, Darko, David B., David S., Dennis, Dominik, Ellen, Esther, Georgios, H.C., Henrik, Jan, Johannes, Jun, Lars, Lucas, Marcel, Martin, Michael, Mike, Ming, Patrick, Robin, Robert, Sven, Tobias, Volker, and Wolfgang, for the great times that I had over all this years. I will never forget the kart sessions with Darko and Arne, the weekly Friday lunch, and especially the "City Reconstruction" meetings. A special thanks goes to Jan for the excellent technical support and keeping up with me when I again occupied most of the hard disk space or copied massive amounts of data over the internal network. A very special "thank you" goes to Dennis Mitzel for all the fun we had at the conferences we visited together. I am even more thankful towards Tobias Weyand for all our discussions, listening to and improving on my ideas, and helping with collecting datasets.

I want to thank Ole for being the best HiWi ever.

Finally, I want to thank all of my friends, Ann, Bianca, Bert & Manu, Claudia, Dirk & Laura, Elias, Eugen, Henrik, Jan, Jana & Manish, Jan-Thorsten, Mehmet, Melanie, Micha, Michel, Nadine, Oliver, Robert, Sara & Torsten, just to name a few, for their support and patients with me, especially as I too often stayed at work instead of meeting them. They are a major reason I enjoyed my time in Aachen and I will always remember the times we had.

This thesis is dedicated to my brother.

Contents