Springer Tracts in Advanced Robotics

Volume 135

The Springer Tracts in Advanced Robotics (STAR) publish new developments and advances in the fields of robotics research, rapidly and informally but with a high quality. The intent is to cover all the technical contents, applications, and multidisciplinary aspects of robotics, embedded in the fields of Mechanical Engineering, Computer Science, Electrical Engineering, Mechatronics, Control, and Life Sciences, as well as the methodologies behind them. Within the scope of the series are monographs, lecture notes, selected contributions from specialized conferences and workshops, as well as selected PhD theses.

Special offer: For all clients with a print standing order we offer free access to the electronic volumes of the Series published in the current year.

Indexed by DBLP, Compendex, EI-Compendex, SCOPUS, Zentralblatt Math, Ulrich's, MathSciNet, Current Mathematical Publications, Mathematical Reviews, MetaPress and Springerlink.

More information about this series at http://www.springer.com/series/5208

Pascal Meißner

Indoor Scene Recognition by 3-D Object Search

For Robot Programming by Demonstration

Springer

Pascal Meißner
IAR-IPR
Karlsruhe Institute of Technology
Karlsruhe, Germany

Dissertation approved by the KIT Department of Informatics. Oral examination on July 6th, 2018 at Karlsruhe Institute of Technology (KIT)

ISSN 1610-7438 ISSN 1610-742X (electronic)
Springer Tracts in Advanced Robotics
ISBN 978-3-030-31851-2 ISBN 978-3-030-31852-9 (eBook)
https://doi.org/10.1007/978-3-030-31852-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Carlo Bourlet—Professor at CNAM, Paris, France—, a role model in dedication and determination.

Foreword by Rüdiger Dillmann

Today's artificial intelligence systems can be found on simple embedded systems as much as in cloud-computing data centers and already have a huge impact on both economic growth and the structures of our societies. In daily life, coexistence and cooperation between humans and increasingly intelligent machines have become a reality. A concept which could play an important role in further advancing this man–machine symbiosis is the so-called anthropomorphism. Applied to robotics, it fosters the development of humanoid robots which are provided with anthropomorphic skills so as to interact with their human counterparts. However, how to reach the level of intuitiveness and richness in interpersonal communication is still an open question.

Roboticists around the world are considering a large variety of modalities and algorithms for programming robots in the most natural manner, for instance, through voice commands, gestures or even physical demonstrations of everyday tasks. Their common goal is to overcome the explicit modeling of daily human–robot interaction by expert programmers. This book is part of these attempts in the sense that it contributes to the vast field of Robot Programming by Demonstration (PbD). The main goal of PbD research is to enable humans to teach robots real-world tasks through physical demonstrations. By this means, a future autonomous robot should be able to select actions according to their adequacy within the environment conditions it encounters. However, it will not be sufficient to just program actions through demonstrations. Capabilities for deciding whether or not an action is appropriate in a given situation, are also required. In order to decide this, characteristic models of scenes to be expected have to be available in turn.

The absence of scene models that are suitable for that purpose has been the impulse for the author to develop his active-vision-based approach for recognizing scenes, which he fully introduces in this book. Derived from the Implicit Shape Model, his novel scene representation models both which objects occur in a scene and how they co-occur in terms of the 6-DoF spatial relations they are engaged in. This scene representation—named Implicit Shape Model (ISM) trees—can be learnt from the very demonstrations recorded for task learning through PbD. His method for recognizing scenes with this representation favors a modular approach,

starting off from results of third-party object-localization algorithms instead of trying to recognize scenes from raw image data. While contradicting current computer vision trends, proceeding in such a modular manner has already shown strong results in other contexts. Meißner's approach allows for the precise modeling of 3-D relationships, a requirement specific to scene modeling which is to precede the execution of manipulation tasks. The author's partly symbolic representation also offers large generalization capabilities, yet avoiding symbol-grounding issues due to its minimalistic design. Other issues the author deals within this book, include how to model spatial relations in indoor scenes as well as how to assess deviations between expected and actual layouts of scenes. His work also addresses the question which of the spatial restrictions in a scene should be considered and which ones should be left out. This important yet neglected question is directly linked to the complexity of scene recognition and thus, indirectly, to that of decision-making for robots.

The author designed his scene representation with the goal of enabling efficient object search. Beyond pure scene recognition, he proposes an active scene recognition system for mobile robots in this book. The system integrates the scene recognition algorithms he developed with a novel algorithm for guiding the focus-of-attention of such a mobile robot. Active scene recognition on the basis of object search fills an important gap in the overall PbD workflow. But well beyond PbD, the introduction of ISM trees is an important step forward towards developing service robots which reliably and robustly operate in dynamic household scenarios in accordance with the principle of explainable artificial intelligence.

Karlsruhe, Germany Rüdiger Dillmann
July 2019

Foreword by Bruno Siciliano

At the dawn of the century's third decade, robotics is reaching an elevated level of maturity and continues to benefit from the advances and innovations in its enabling technologies. These all are contributing to an unprecedented effort to bringing robots to human environment in hospitals and homes, factories and schools; in the field for robots fighting fires, making goods and products, picking fruits and watering the farmland, saving time and lives. Robots today hold the promise for making a considerable impact on a wide range of real-world applications from industrial manufacturing to health care, transportation, and exploration of the deep space and sea. Tomorrow, robots will become pervasive and touch upon many aspects of modern life.

The *Springer Tracts in Advanced Robotics (STAR)* is devoted to bringing to the research community the latest advances in the robotics field on the basis of their significance and quality. Through a wide and timely dissemination of critical research developments in robotics, our objective with this series is to promote more exchanges and collaborations among the researchers in the community and contribute to further advancements in this rapidly growing field.

The monograph by Pascal Meissner is based on the author's doctoral thesis. It focuses on Robot Programming by Demonstration (PbD) to enable humans to teach robots real-world tasks through physical demonstrations. The concept of Implicit Shape Model (ISM) trees is introduced to derive scene representation models in terms of the spatial relations among the objects to be manipulated. Then, an optimization algorithm for Active Scene Recognition (ASR) allows embedding canonical scene recognition in a decision-making system to select best camera views for 3-D object localization.

Rich of experiments in a setup mimicking a kitchen, the results demonstrate the good performance of ISM trees as scene classifiers for a large number of object arrangements. A very fine addition to the STAR series!

Naples, Italy
July 2019

Bruno Siciliano
STAR Editor

Preface

While it is the purpose of this thesis to convey the most important findings of my Ph.D. research, I want to take this preface as an opportunity to report on the very nature of doing doctoral studies as I got to know it. While some may argue that finding the right institution and getting admitted there is the main challenge for a graduate—from my point of view, the former is a matter of personality, while a good recipe for the latter is to carry out one's studies at a lab of one's own choice as continuously as possible—I think that being a researcher is a major challenge that has little in common with succeeding in one's university studies. From my own experience, numerous of my colleagues experienced disappointment and frustration while being Ph.D. candidates, even though working under—in my view—good conditions. I suggest that this kind of issues results from misconceptions of what it actually means to do doctoral studies. As an attempt to clarify this at least for my field in Germany, I want to draw an analogy between being a Ph.D. candidate and an entrepreneur on the basis of Long et al. from 1983. More precisely, I propose that Ph.D. students consider themselves as being entrepreneurs. According to Long et al., a first defining aspect of entrepreneurship is self-employment. While the colleagues at my lab and myself were employees in the public service, I still think that this attribute applied to us, e.g. because we were continuously expected to come up with new research challenges on our own. Far beyond our mere interest in technology, it was essential to have the ambition to discover research questions as well as to develop and present appropriate answers. In the sense of my entrepreneur metaphor, we had to figure out promising business opportunities, to develop offers and to sell them. In my opinion, the fact that research findings are mostly attributed to individuals is closely linked to the self-employment in academia and is thus an indication of it. For example, Nobel Prizes are to this day awarded to an individual and not to collaborative achievements. The impact of findings from Ph.D. research is commonly regarded as a good measure for the achievement they represent. If one considers a publication which contains such findings, an offer and the authors as its supplier, the impact can be equated with the benefit Ph.D. students can strive for. As entrepreneurs, Ph.D. candidates should, therefore, keep the actual purpose

of their endeavor in mind—maximizing impact through appropriately publishing relevant results.

With no further proof, I claim that publications are offered on a market where their authors compete with others. This means that working in academia coincides with facing highly competitive situations many graduates may be confronted with for the first time. Maximizing benefit is only possible if one has good knowledge of the market he participates in and is permanently adapting to it. In concrete terms, one should carefully assess which conferences or journals are best suited for his results regarding their thematic focus and reputation. Besides, one should present his findings in a way that they are as easily accessible as possible to his reviewers and to other potential readers. When designing a publication, it is indispensable to adopt their perspective in terms of, for instance, their knowledge on the topic in question, their possible associations with the employed vocabulary and the time they are willing to invest in order to understand the publication. Assessing a market furthermore goes along with estimating how many competitors one has on a research problem and who they are. A Ph.D. candidate should ask himself whether it makes more sense for him to work on a popular topic with a large community but presumably under time pressure and with considerable risks that his results may be overlooked? Or does he prefer to look for a niche, with the consequence that little exchange will be possible or that the relevance of either the problem he addresses or the solution he proposes may be challenged as such? Another question one might have to answer in the second case is whether the state of the art is advanced enough for him to generate substantial results in the short time span of doctoral studies.

Returning to the economics perspective, it seems obvious that investments have to be undertaken in order to create benefits. Whether and to which degree investing one's lifetime pays off as scientific impact is highly speculative. Besides activity on markets, considerable uncertainty is another attribute of entrepreneurship, thus supporting the analogy I make. The last aspect of research entrepreneurship I want to address is management and how to optimize its cost–benefit ratio. At my lab, managing not just applied to ourselves but also to the undergraduates we supervised. Depending on our strengths, i.e. on whether one of us made better progress working alone or in a team with contributing undergraduates, the additional resources provided by these undergraduates for solving problems outweighed the costs of attracting and supervising them. In my opinion, negative experiences from supervising students often originate in ignoring that supervision represents an investment into obtaining contributions to research problems or to tasks of lesser scientific benefit. Being an investment, supervision has to be treated accordingly. Of course, attracting and supervising undergraduates are fields prone to optimization—especially since they come along with participating in a market. Acquisition can be optimized by thinking about how, where and when to make offers that match the interests of undergraduates, i.e. their thematic interests—one shouldn't underestimate the importance of trends, their insecurities and their call for reliability. Optimizing supervision equals optimizing the outcome of the time that both the supervisor and the undergraduates invest. Of course, this happens under the constraint that the quality of supervision is kept up—in my view, the foremost priority

for any Ph.D. candidate as soon as he starts supervising. We tried to optimize our efforts in supervising with various concepts such as chaining fixed-length appointments, undergraduates working together on greater problems and experiments, undergraduates supervising each other, groupware-supported supervision or the usage of development frameworks such as Scrum. What proved to be essential to us was not only relying on the aforementioned mechanistic approaches but also taking into account the specific traits of each individual undergraduate in order to adapt their respective tasks, working conditions as well as our leadership style during his stay at our lab.

Provided sufficient expertise as well as the toughness and perseverance to remain focused on obtaining research findings—despite the numerous encountered distractions and interruptions—I am convinced that anyone who can identify with being a research entrepreneur can find his fulfillment in my field. To conclude, I wish everyone a hopefully insightful and maybe even enjoyable read of this thesis.

This book is equivalent to the Ph.D. thesis I submitted under the title "Indoor Scene Recognition by 3-D Object Search for Robot Programming by Demonstration" to the KIT Department of Informatics. I defended this thesis at Karlsruhe Institute of Technology (KIT) on July 6th, 2018. The source code for all contributions of this approved thesis is freely available under https://github.com/asr-ros.

Karlsruhe, Germany Pascal Meißner
August 2018

Acknowledgements

My sincere gratitude goes to Dr. Stefan Gächter Toya, Prof. John K. Tsotsos, Dr. Robert Eidenberger and Prof. Antonio Torralba for inspiring me with their research. They laid the foundations for the contributions of my thesis.

I am very grateful to my advisor Prof. Rüdiger Dillmann for putting his trust in me while I pursued my doctoral studies. I particularly thank him for supporting my vision while granting me complete freedom in defining and implementing it. Moreover, I would like to thank Prof. Michael Beetz for the interesting conversations about my research problems, we had. My special thanks go to Prof. Torsten Kröger for his tremendous support towards the end of my doctoral studies.

My deepest gratitude goes to my mentor Dr. Sven R. Schmidt-Rohr. First as my supervisor, then as a colleague, he provided decisive support in word and deed throughout highs and lows. I also thank him for broadening my horizon in unexpected directions with his compelling enthusiasm and strategic foresight. Many thanks to Dr. Rainer Jäkel for his expert advice as well as for his friendly, calm and consistently helpful manner. My thanks additionally go to Dr. Martin Lösch for being such a committed leader of our research group at the beginning of my doctoral studies.

My gratitude goes to my student co-workers Tobias Allgeyer, Florian Aumann-Cleres, Jocelyn Borella, Souheil Dehmani, Benny Fuhry, Nikolai Gaßner, Joachim Gehrung, Fabian Hanselmann, Heinrich Heizmann, Florian Heller, Robin Hutmacher, David Kahles, Oliver Karrenbauer, Daniel Kleinert, Felix Marek, Matthias Mayr, Jonas Mehlhaus, Sebastian Münzner, Trung Nguyen, Reno Reckling, Ralf Schleicher, Patrick Schlosser, Patrick Stöckle, Daniel Stroh, Jeremias Trautmann, Richard Weiss and Valerij Wittenbeck for spending countless days and nights in my two labs and joining me in struggling with both hard- and software.

Special thanks go to Armin Dürr—former owner of the “1001 Computer” store in the small town of Bretten—for introducing me to the world of IT. I conclude by thanking my family, in particular, Antje Lossin as well as Corinne and Jürgen Meißner, for providing great assistance through the eventful years of my doctoral studies.

Contents