Computer Vision Metrics

Survey, Taxonomy, and Analysis

Scott Krig

Apress open

Computer Vision Metrics: Survey, Taxonomy, and Analysis

Scott Krig

ISBN-13 (pbk): 978-1-4302-5929-9

ISBN-13 (electronic): 978-1-4302-5930-5

President and Publisher: Paul Manning
Lead Editors: Steve Weiss (Apress); Stuart Douglas (Intel)
Technical Reviewers: Sanjay Addicam and Shahzad Malik
Coordinating Editor: Melissa Maldonado
Cover Designer: Anna Ishchenko

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com.

For information on translations, please e-mail rights@apress.com, or visit www.apress.com.

About ApressOpen

What Is ApressOpen?

- ApressOpen is an open-access book program that publishes high-quality technical and business information.
- ApressOpen eBooks are available for global, free, noncommercial use.
- ApressOpen eBooks are available in PDF, ePub, and Mobi formats.
- The user friendly ApressOpen free eBook license is presented on the copyright page of this book.

Thanks to my wife Janie, family, and parents for being part of my life.

Contents at a Glance

Contents

About the Author

Scott Krig is a pioneer in computer imaging, computer vision, and graphics visualization. He founded Krig Research in 1988 (`krigresearch.com`), providing the world's first imaging and vision systems based on high-performance engineering workstations, super-computers, and dedicated imaging hardware, serving customers worldwide in 25 countries. Scott has provided imaging and vision solutions around the globe, and has worked closely with many industries, including aerospace, military, intelligence, law enforcement, government research, and academic organizations.

More recently, Scott has worked for major corporations and startups serving commercial markets, solving problems in the areas of computer vision, imaging, graphics, visualization, robotics, process control, industrial automation, computer security, cryptography, and consumer applications of imaging and machine vision to PCs, laptops, mobile phones, and tablets. Most recently, Scott provided direction for Intel Corporation in the area of depth-sensing and computer vision methods for embedded systems and mobile platforms.

Scott is the author of many patent applications worldwide in the areas of embedded systems, imaging, computer vision, DRM, and computer security, and studied at Stanford.

Scott also enjoys acoustic guitar design and lutherie work, particularly 12-string acoustic guitars, as well as acoustic guitar composition and performance.

Acknowledgments

This book would not be as well thought out without the early technical feedback, conversations, and observations on very rough materials by Vadim Pizarevsky of ITSEEZ, who also is a major force behind the OpenCV foundation. Vadim brings vast and quantitative expertise in computer vision across a wide range of application domains. Thanks, Vadim.

Special thanks also go to Stuart Douglas at Intel Press for the commission to write this book, and for introductions to people at Apress. Also, special thanks to the key editors at Apress, including Melissa Maldonado, Mark Powers, Jeffrey Pepper, Steve Weiss, Robert Hutchinson, James Markham, and Carole Berglie for making this book a reality, and for adding value through the editorial process.

Introduction

Dirt. This is a jar of dirt.

Yes.

. . . Is the jar of dirt going to help?

If you don't want it, give it back.

—*Pirates Of The Carribean*, Jack Sparrow and Tia Dalma

This work focuses on a slice through the field - Computer Vision Metrics – from the view of feature description metrics, or how to describe, compute and design the macro-features and micro-features that make up larger objects in images. The focus is on the pixel-side of the vision pipeline, rather than the back-end training, classification, machine learning and matching stages. This book is suitable for reference, higher-level courses, and self-directed study in computer vision. The book is aimed at someone already familiar with computer vision and image processing; however, even those new to the field will find good introductions to the key concepts at a high level, via the ample illustrations and summary tables.

I view computer vision as a mathematical artform and its researchers and practitioners as artists. So, this book is more like a tour through an art gallery rather than a technical or scientific treatise. Observations are provided, interesting questions are raised, a vision taxonomy is suggested to draw a conceptual map of the field, and references are provided to dig deeper. This book is like an attempt to draw a map of the world centered around feature metrics, inaccurate and fuzzy as the map may be, with the hope that others will be inspired to expand the level of detail in their own way, better than what I, or even a few people, can accomplish alone. If I could have found a similar book covering this particular slice of subject matter, I would not have taken on the project to write this book.

What is not in the Book

Readers looking for computer vision "'how-to"' source code examples, tutorial discussions, performance analysis, and short-cuts will not find them here, and instead should consult the well-regarded http://opencv.org library resources, including many fine books, online resources, source code examples, and several blogs. There is nothing better than OpenCV for the hands-on practitioner. For this reason, this book steers a clear path around duplication of the "how-to" materials already provided by the OpenCV community and elsewhere, and instead provides a counterpoint discussion, including a comprehensive survey, analysis and taxonomy of methods. Also, do not expect all computer vision topics to be covered deeply with proofs and performance analysis, since the bibliography references cover these matters quite well: for example, machine learning, training and classification methods are only lightly introduced, since the focus here is on the feature metrics.

In summary, this book is about the feature metrics, showing "'what'" methods practitioners are using, with detailed observations and analysis of "'why'" those methods work, with a bias towards raising questions via observations rather than providing too many answers. I like the questions best because good questions lead to many good answers, and each answer is often pregnant with more good questions…

This book is aimed at a survey level, with a taxonomy and analysis, so no detailed examples of individual use-cases or horse races between methods are included. However, much detail is provided in over 540+ bibliographic references to dig deeper into practical matters. Additionally, some "'how-to'" and "'hands-on'" resources are provided in Appendix C. And a little 'perfunctory' source code accompanying parts of this book is available online, for Appendix A covering the interest point detector evaluations for the synthetic interest point alphabets introduced in Chapter 7; and in Appendix D for extended SDM metrics covered in Chapter 3.

What is in the Book

Specifically, Chapter 1 provides preamble on 2d image formation and 3d depth imaging, and Chapter 2 promotes intelligent image pre-processing to enhance feature description. Chapters 3 through 6 form the core discussion on feature description, with an emphasis on local features. Global and regional metrics are covered in Chapter 3, feature descriptor concepts in Chapter 4, a vision taxonomy is suggested in Chapter 5, and local feature description is covered in Chapter 6. Ground truth data is covered in Chapter 7, and Chapter 8 discusses hypothetical vision pipelines and hypothetical optimizations from an engineering perspective, as a set of exercises to tie vision concepts together into real systems (coursework assignments can be designed to implement and improve the hypothetical examples in Chapter 8). A set of synthetic interest point alphabets is developed in Chapter 7, and ten common detectors are run against those alphabets, with the results provided in Appendix A. It is difficult to cleanly partition all topics in image processing and computer vision, so there is some overlap in the chapters. Also, there are many hybrids used in practice, so there's inevitable overlap in the Chapter 5 vision taxonomy, and creativity always arrives on the horizon to find new and unexpected ways of using old methods. However, the taxonomy is a starting point and helped to guide the organization of the book.

Therefore, the main goal has been to survey and understand the range of methods used to describe features, without passing judgment on which methods are better. Some history is presented to describe why certain methods were developed, and what properties of invariance or performance were the goals, and we leave the claims to be proven by others, since "how" each method is implemented determines performance and accuracy, and "what" each method is tested against in terms of ground truth data really tells the rest of the story. If we can glean good ideas from the work of others, that is a measure of the success of their work.

Scope

For brevity's sake, I exclude a deep treatment of selected topics not directly related to the computer vision metrics themselves; this is an unusual approach, since computer vision discussions typically include a wider range of topics. Specifically, the topics not covered deeply here include statistical and machine learning, classification and training, feature database construction and optimization, and searching and sorting. Bibliography references are provided instead. Distance functions are discussed, since they are directly linked to the feature metric. (A future edition of this book may contain a deep dive into the statistical and machine learning side of computer vision, but not now.)

Terminology Caveat

Sometimes terminology in the literature does not agree when describing similar concepts. So in some cases, terminology is adopted in this work that is not standardized across independent research communities. In fact, some new and nonstandard terminology may be introduced here, possibly because the author is unaware of better existing terminology (perhaps some of the terminology introduced in this work will become standardized). Terminology divergence is most pronounced with regard to mathematical topics like clustering, regression, group distance, and error minimization, as well as for computer vision topics like keypoints, interest points, anchor points, and the like. The author recognizes that one is reluctant to change terminology, since so many concepts are learned based on the terminology. I recall a friend of mine, Homer Mead, chief engineer for the lunar rover and AWACS radar at Boeing, who sub-consciously refused to convert from using the older term condenser to use the newer term capacitor.

Inspiration comes from several sources, mostly the opportunity of pioneering: there is always some lack of clarity, structure and organization in any new field as the boundaries expand, so in this vast field the opportunity to explore is compelling: to map out structure and pathways of knowledge that others may follow to find new fields of study, create better markers along the way, and extend the pathways farther.

The inspiration for this book has come from conversations with a wide range of people over the years. Where did it all start? It began at Boeing in the early 1980s, while I was still in college. I was introduced to computer graphics in the Advanced Development Research labs where I worked, when the first computer-shaded 3D renderings of the space shuttle were made in raster form. At that time, mainly vector graphics machines were being used, like Evans & Sutherland Picture Systems, and eventually a BARCO frame buffer was added to the lab, and advanced raster computer renderings of shaded images from graphics models were pioneered by Jeff Lane and his group, as well as Loren Carpenter. Fractals, NURBS, and A-buffer techniques were a few of the methods developed in the labs, and the math of computer graphics, such as bi-cubic patches and bi-quintic patches, scared me away from graphics initially. But I was attracted to single pixels in the BARCO frame buffer, one pixel and line and frame at a time, since they seemed so intuitive and obvious. I initially pursued imaging and computer vision rather than all the computer graphics and associated math. However, it turned out that the computer vision and image processing math was far more diverse and equally complex anyway. Since then I have also spent considerable time in computer graphics. Back in the mid-1980s, Don Snow, my boss, who was co-founder and VP of research at Pacific

Western Systems and later at Applied Precision, asked me to analyze the View-PRB fixed-function hardware unit for pattern recognition to use for automatic wafer probing (in case we needed to build something like it ourselves) to locate patterns on wafers and align the machine for probing. Correlation was used for pattern matching, with a scale-space search method we termed "super-pixels." The matching rate was four 32x32 patches per second over NTSC with sub-pixel accuracy, and I computed position, rotation, and offsets to align the wafer prober stage to prepare for wafer probing; we called this auto-align. I designed a pattern recognition servo system to locate the patterns with rotational accuracy of a few micro-radians, and positional accuracy of a fraction of a micron. In the later 1980s, I went to work for Mentor Graphics, and after several years I left the corporate R&D group reporting to the president Gerry Langeler to start a company, Krig Research, to focus on computer vision and imaging for high-end military and research customers based on expensive and now extinct workstations (SGI, Apollo, Sun… gone, all gone now…), and I have stayed interested ever since. Many things have changed in our industry; the software seems to all be free, and the hardware or SOC is almost free as well, so I am not sure how anyone can make any money at this anymore.

More recently, others have also provided inspiration. Thanks to Paul Rosin for synthetic images and organizational ideas. Thanks to Yann LeCun for providing key references into deep learning and convolutional networks, and thanks to Shree Nayar for permission to use a few images, and continuing to provide the computer vision community with inspiration via the Cave Research projects. And thanks to Luciano Oviedo for vast coverage of industry activity and strategy about where it is all going, and lively discussions.

Others, too many to list, have also added to my journey. And even though the conversations have sometimes been brief, or even virtual via email or SKYPE in many cases, the influence of their work and thinking has remained, so special thanks are due to several people who have provided comments to the manuscript or book outline, contributed images, or just plain inspiration they may not realize. Thank you to Rahul Suthankar, Alexandre Alahi for use of images and discussions; Steve Seitz, Bryan Russel, Liefeng Bo, and Xiaofeng Ren for deep-dive discussions about RGB-D computer vision and other research topics; Gutemberg Guerra-filho, Harsha Viswana, Dale Hitt, Joshua Gleason, Noah Snavely, Daniel Scharstein, Thomas Salmon, Richard Baraniuk, Carl Vodrick, Hervé Jégou, and Andrew Richardson; and also thanks for many interesting discussions on computer vision topics with several folks at Intel including Ofri Weschler, Hong Jiang, Andy Kuzma, Michael Jeronimo, Eli Turiel, and many others whom I have failed to mention.

Summary

In summary, my goal is to survey the methods people are using for feature description—the key metrics generated—and make it easier for anyone to understand the methods in practice, and how to evaluate the methods using the vision taxonomy and robustness criteria to get the results they are looking for, and find areas for extending the state of the art. And after hearing all the feedback from the first version of this work, I hope to create a second version that is even better.

Scott Krig
Anno Domini 2014