Distant Viewing

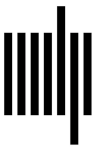
Computational Exploration of Digital Images

By: Taylor Arnold, Lauren Tilton

Citation:

Distant Viewing: Computational Exploration of Digital Images
By: Taylor Arnold, Lauren Tilton
DOI: 10.7551/mitpress/14046.001.0001
ISBN (electronic): 9780262375160
Publisher: The MIT Press
Published: 2023

The open access edition of this book was made possible by generous funding and support from the author



The MIT Press

Introduction

In the fall of 2010, we began working on our first collaboration. In what would eventually become the public digital project *Photogrammar*, we set out to build an interactive space that allowed visitors to visualize the more than 170,000 digitized Farm Security Administration / Office of War Information (FSA-OWI) photographs produced by the US government between 1935 and 1944. Our collaboration served as an ideal mixture of our interests. Taylor was a graduate student in a statistics department with a research focus on exploring and visualizing complex datasets. Lauren was a graduate student in American Studies with a concentration in public humanities with a focus on twentieth-century American film and photography. Combining the FSA-OWI collection's rich metadata and meticulously digitized public-domain images to create a publicly accessible interface overlaid perfectly with both of our areas of research. A proof of concept developed in Laura Wexler's public humanities graduate seminar turned into a full-fledged digital humanities public project thanks to the National Endowment for the Humanities Office of Digital Humanities.

Our work became a popular public project that would welcome millions of visitors and encourage numerous extensions and revisions. We created interactive visualizations that allowed exploration of almost all the collection's available metadata. Visitors could view one of several interactive maps, follow the journeys of individual photographers, search by themes, and explore the photographic captions. The critical context and the contribution of these elements to visitors' understanding of the FSA-OWI collection should not be understated. However, something seemed missing. We were working with an extensive collection of documentary photography, and it was ultimately the photographs that drew us and others to this collection.

Our work, however, facilitated an aggregative analysis of every element *but* the photographs. The images were only accessible by looking at them individually, with no way to search by visual themes or identify objects and people present within the frame. There was a disconnect between the main objects of interest and the affordances provided by our work.

The absence of image-based methods in our initial iterations of *Photogrammar* was driven by a scarcity of readily available tools, not a lack of interest. Digital images are challenging to work with computationally, for reasons that we interrogate in the following chapters. The best available methods we could find performed poorly on historic black-and-white photographs. Face detection methods missed more faces than they found, failing to find faces that were not at a particular angle and unable to detect anyone wearing a hat. Algorithms for detecting objects in images were more likely to produce comically bizarre predictions than usable information. Methods for aggregating based on dominant colors fared better but were not well suited to our collection of predominantly black-and-white photographs. Illustrations of these predictions were a consistent element of the earliest talks we gave on the work. Our favorite example came from the photograph featured in chapters 1 and 2 of a shepherd riding a horse in a field next to his sheepdog. Even though it was in vivid color and contained three distinct objects within the lexicon of popular computer vision algorithms, most failed to identify any element of the image correctly. The experience led us to ask more questions about exactly how these algorithms were built and for whom. We kept asking ourselves: what ways of seeing are these built to view and what if we thought differently about why we should use them.

By 2016 the landscape of available tools for working computationally with images had undergone a dramatic expansion. Software libraries such as darknet (2016), TensorFlow (2015), keras (2015), PyTorch (2016), and Detectron (2017) suddenly provided out-of-the-box access to increasingly accurate and powerful computer vision algorithms.[1] Scholars working with digital collections of images began to use this new set of approaches. Applications appeared in venues such as the Culture Analytics program at UCLA's Institute for Pure and Applied Mathematics, workshops held by the special interest group for audiovisual material within the Association of Digital Humanities Organizations (ADHO), and articles in the newly created *International Journal for Digital Art History*. Our work shifted as well.

Presentations that previously ended by critiquing algorithmic results were replaced with forward-looking examples of how computer vision was helping us reimagine the FSA-OWI collection by providing approaches for a visual search and discovery interface. Rather than relying solely on existing computer vision algorithms, we began to customize and build algorithms that viewed in the ways that furthered our areas of interest.

Our excitement about the improvements in computer vision algorithms was tempered by our prior experiences that had highlighted the comparative difficulty of training computers to understand digital images. The tools seemed to be producing helpful information, but what features of the images continued to be lost through their algorithmic transformation? Many of these new tools were created or sponsored by large corporations and government entities. What are the implications of aligning our analyses with the interests of these organizations? Software for data exploration and visualization was not built around the study of digital images. How can our exploratory methods catch up with the new methods in computer vision? Numerous scholars in media and visual culture studies—such as John Berger, David Bordwell, Lisa Cartwright, Stuart Hall, Lev Manovich, Lisa Nakamura, Leigh Raiford, Marita Sturken, and Laura Wexler[2]—have stressed the importance of thinking carefully about how images are created, circulated, and interpreted. When applying complex computational approaches to the study of digital images, it is as vital as ever to consider these questions. To enable the careful and critical computational exploration of digitized visual collections, we need a cohesive theory for how computer vision creates meaning and a methodological specificity that takes into account the intricacies of digital images as a form and format.

In this text, we present a theory and methodological description of what we refer to as *distant viewing*, the application of computer vision methods to the computational analysis of digital images. Our goal is to offer a constructive and generative critique of computer vision that focuses on enabling fruitful applications. To the best of our knowledge, this text is the first book-length treatment that approaches the application of computer vision to the study of visual messages as its own object of study. The distinction here is important because our approach allows for a critical understanding of the possibilities and limitations of existing computer vision techniques. It also provides a framework for a reflexive understanding of computer vision as a way of circulating and producing knowledge.

The focus of distant viewing on *digital* images is a pragmatic one, resulting from the fact that the application of computer vision requires machine-readable inputs. However, this does not limit our objects of study to born-digital materials. Distant viewing can be applied to digitized collections originally produced in almost any medium. For example, we can apply our approach to digitized collections of photographs, photographic negatives, newspapers, comics, and posters. We can also work with digital images of material culture, something we return to in chapter 6. Distant viewing is also not limited to still images; it can be used to study collections of objects from media such as television, film, and video games. An example of distant viewing applied to a pair of television series is illustrated, for example, in chapter 5. In most cases, when one is applying computer vision to a digital image, we argue that this is distant viewing.

Our terminology is motivated by the concept of *distant reading* from the field of computational literary studies. The specific meaning and importance of the term *distant reading* has been extensively discussed; it is not our goal to make specific connections or proclamation within these debates.[3] Rather, our terminology signals a general interest in adapting the computational literary studies approach of applying computational and statistical techniques to large corpora in the service of humanistic research questions.[4] While certainly not without their critics, these approaches have opened exciting new lines of scholarship.[5] Our terminology also signals a departure from the textual focus of literary studies. The process of interpreting a visual message is semiologically and phenomenologically different from the act of reading a text, which we theorize in chapter 1. As we will explore in the following chapters, these differences lead to important changes in the way that we can apply and interpret the results of computational analyses.

In the tradition of visual culture studies and computer vision as well as our history of collaboration, we take a transdisciplinary perspective to our work. Both of us were trained in interdisciplinary fields that taught us the power of thinking across boundaries of disciplines and fields. We primarily draw from and engage with scholarship from film and media studies, visual semiotics, digital humanities, information science, computer science, and data science. The text's structure and focus are designed to be legible and useful to audiences coming from any of these varied perspectives.

The first two chapters establish our main theoretical and methodological claims about distant viewing. Chapter 1 begins by investigating what it

means to say that computer vision "understands" visual inputs. We draw from information science and semiotics to illustrate why the way that digital images convey information necessitates a different approach. Specifically, we see that this process involves creating annotations that capture some, though never all nor ever perfectly, of the information present in the images. We conclude the chapter by showing how the process of annotation can be seen as a machine-mediated way of viewing images; this leads us to understand how existing scholarship in media studies shapes the application of computer vision. In chapter 2, we engage with the methodological aspects of working with computer vision annotations. We investigate how standard approaches used in data science to explore data must be adjusted when working with computer vision, resulting in four phases of analysis. Namely, we must *annotate* our collection using computer vision algorithms, then *organize* the annotations and metadata, *explore* the data and our research questions, and finally *communicate* the results. The first two chapters engage in a close analysis of a single image, the FSA-OWI photograph of a shepherd mentioned above. We hope to model how computational analyses should also help highlight, rather than supplant, the close reading of individual images.

Chapters 3 through 6 present the use of distant viewing within four different application domains. As readers move from chapter to chapter, the complexity of the computer vision models build. Each chapter is structured around the first three phases of the distant viewing method described in chapter 2: annotate, organize, and explore; the fourth phase, communication, is this book. After establishing a research question, we start by understanding one or more annotations provided by computer vision algorithms, organize other metadata attached to the collection, and finish by conducting an exploration of the organized data. Along the way, we discuss the limitations of these algorithms as we think carefully about exactly what these computer vision algorithms view, and do not view. Chapter 3 investigates the use of color in movie posters and its relationship to genre. We see how distant viewing can address complex research questions even when using relatively low-level annotations. In chapter 4, we apply a region segmentation algorithm to the photographs from the FSA-OWI archive. This chapter shows how computer vision annotations can both support and supplant the organizational logic of the archive. We illustrate in chapter 5 how distant viewing can also be used with moving images. We see how formal

film elements can be applied to study issues of gender and power within a pair of network-era sitcoms. Finally, in chapter 6, we apply distant viewing to a collection of images from a large encyclopedic museum to see how computer vision can open digital collections through public interfaces.

Our excitement about the possibilities for the computational analysis of collections of digital images has been shared by many other research groups. Some of the earliest examples come from the manual annotation of film and television metrics by Barry Salt, Gunars Civjans, Yuri Tsivian, and Jeremy Butler.[6] Recently, the journal *Digital Humanities Quarterly* (DHQ) sponsored a special issue focused on film and video analysis in 2020, with articles describing projects such as Barbara Flueckiger's FilmColors project and Masson et al.'s *Sensory Moving Image Archive project* (SEMIA).[7] Along with Stefania Scagliola and Jasmijn Van Gorp, we edited another DHQ special issue titled "AudioVisual Data in DH" with over twenty research articles from a wide variety of disciplines and nation contexts.[8] Interest has also expanded from the growing field of digital art history, which has had several special issues, conferences, and a new journal that have included a significant amount of computational work. Notably, in the first issue of the *International Journal for Digital Art History*, K. Bender made the first use of the term "distant viewing" within his study of the iconography of Aphrodite/Venus.[9] Numerous other exciting research papers have been published in other journals, such as Nanne Van Noord, Ella Hendriks, and Eric Postma's study of artistic style and Laure Thompson and David Mimno's analysis of the study of Dadaism.[10] We hope that our work in this book further enables and encourages more developments in these and other areas.

The book is designed to be read and used. Along with being open access, the text is organized such that the chapters should be readable in any order. One reader might be interested in the theory and then an application. Another reader might interested in a particular application and therefore wish to start with one of the applications before engaging with the more theoretical opening chapters. Many of the results in the following chapters are presented as tables. We chose to communicate results using numeric tables because of the limitations of other visualization types within the existing print form, such as the lack of interactivity and color. Other ways of visualizing these results are also given in the supplementary materials.

While making novel contributions to the fields of data science and digital humanities, we have avoided superfluous technical jargon.[11] There is

significant translation work to do when talking across fields and forging transdisciplinary scholarship. We aimed for a writing style that is inclusive yet precise, while the footnotes provide more technical descriptions. A glossary of common terms, particularly for terms that may be used differently in different communities, is included at the end of the text to aid in this process. In addition, we have published datasets, code, and many additional visualizations under an open-source license that replicate and further explore the applications described in the text. All of these can be viewed and downloaded on the book's accompanying website, found here:

https://distantviewing.org/book

Finally, we have developed the Distant Viewing Toolkit, open-source software made possible through generous funding from the National Endowment for the Humanities and the Mellon Foundation, that puts theory and method into practice. Information on how to install a current version of the Distant Viewing Toolkit can also be found at the link above.

By theorizing and offering a method, our approach of distant viewing participates in the call for a more careful use of algorithms in our society. When we understand computer vision as a way of seeing, we are then accountable to the histories of vision and the ways we train algorithms to see, look, and view. We are also accountable for what they do not see, look, and view. We have had a plethora of conversations with colleagues who attest to the neutrality of algorithms and resist ideas of algorithms as a technology of vision and mode of communication inculcated in social and cultural pasts, presents, and futures. Distant viewing challenges such claims and calls on us to ask each time a computer vision algorithm looks at an image what is this algorithm viewing, mislabeling, and missing as well as why did we design this algorithm to view in this way. By doing so, we can more carefully engage with the computational analysis of digital images. Now, let's go distant viewing.

The open access edition of this book was made possible by generous funding from the National Endowment for the Humanities and the University of Richmond.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Arnold, Taylor, author. | Tilton, Lauren, author.
Title: Distant viewing : computational exploration of digital images / Taylor Arnold and Lauren Tilton.
Description: Cambridge, Massachusetts : The MIT Press, [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022052202 (print) | LCCN 2022052203 (ebook) | ISBN 9780262546133 (paperback) | ISBN 9780262375177 (epub) | ISBN 9780262375160 (pdf)
Subjects: LCSH: Computer vision. | Image data mining. | Image processing—Digital techniques. | Visual sociology—Technique.
Classification: LCC TA1634 .A76 2023 (print) | LCC TA1634 (ebook) | DDC 006.4/2—dc23/eng/20230111
LC record available at https://lccn.loc.gov/2022052202
LC ebook record available at https://lccn.loc.gov/2022052203