Seen / Unseen: How humans and machines process our everyday photos

Christopher Seeds

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Design

University of Washington
2018

Committee:
Kristine Matthews
Audrey Desjardins

Program Authorized to Offer Degree:
Design
University of Washington

Abstract

Seen / Unseen: How humans and machines process our everyday photos

Christopher Seeds

Chair of the Supervisory Committee:
Associate Professor Kristine Matthews
School of Art, Art History and Design

This thesis researches the contrasting ways in which humans and machines view everyday photographs. Different ways of "seeing" photographs are explored through three unique prototypes. Each prototype looks at different capabilities that are unique to humans or to machines. The intention of this project is to create a heightened awareness of the importance of human involvement with machines.
Master of Design Thesis Documentation

Seen / Unseen: How humans and machines process our everyday photos

Christopher Seeds
University of Washington
Contents

Front Matter 4
Acknowledgments
Abstract
Preface

Introduction 8
The Discourses
Personal Motivation
Research Questions
Audience
Goals

Background 10
Related Work
Technology Involved

My Study 13
Overview
Research Structure
Findings

Design 17
Divergent Ideation
Convergence of Prototypes
Prototyping

Insights 23

Conclusion 24

Bibliography 25

Appendix 27
Acknowledgments

This thesis would not have been possible without the help of all the following people. I want to thank my committee member Audrey Desjardins. Our conversations guided, inspired and gave me the confidence to explore this project. I would like to thank my committee chair Kristine Matthews for all the guidance, insight and support throughout the process. The rest of the University of Washington Department of Art, Art History, and Design faculty deserve thanks as well for all the conversations and guidance.

This thesis would not have been possible if not for my fellow design graduate students Aubree M. Ball, Joe Costello, and Emma Teal Laukitis. I am fortunate and grateful that I’ve been on this journey with you all.

I also want to acknowledge and thank my brother Matthew Seeds for his thought-provoking conversations and wild ideas. My parents Laurel and Dale Seeds have played an instrumental role in my desire to pursue a graduate degree. You both taught me from a young age to be thoughtful and inquisitive.

My wife Magdalen Folkman deserves acknowledgment beyond words. Her support and patience guided me through the trying times of the thesis. I am eternally grateful.
Abstract

This thesis researches the contrasting ways in which humans and machines view everyday photographs. Different ways of “seeing” photographs are explored through three unique prototypes. Each prototype looks at different capabilities that are unique to humans or to machines. The intention of this project is to create a heightened awareness of the importance of human involvement with machines.

Preface

Simple everyday photographs have played a significant role in my life. Their relevance has been felt from both mundane moments to impactful points in my life. From catching up with family at holiday gatherings to paying respects to someone who has died, these candid images capture a photographer’s feeling at a point and place in time. I value a photo’s ability to share an experience and its power to capture a memory.

For me, this thesis has been about looking at my own everyday photography and how it is evolving based on the current technological landscape. I have found through this process that my relationship to technology is one that will be in a constant state of flux and reflections. My hope is that this thesis will inspire the reader to become more aware of their relationship with technology and spark curiosity to explore the unseen computational mechanisms at work behind our phones and other devices.
Introduction

The Discourses

From the earliest forms of photography to emerging digital technologies, there is a rich history of research centered on personal photography. Personal photography has been researched as a cultural phenomenon in the humanities and as a socio-technical process in human-computer interaction (HCI). This thesis sits in the middle of these two discourses.

Personal Motivation

The thesis was motivated by two experiences:
First, capturing everyday images and using them for prompting reminiscence. During the course of my life, photos have not only been memory artifact but also a memory trigger. This act of seeing a photo and engaging in reminiscence is something I have experienced continuously from a young age. The process of recollection triggered by the sight of an image is something that continuously fascinates me.

Second, I have observed how computational systems (Computer vision and machine learning) are playing a more significant role in everyday photo collections. Technology companies and their products have created computational systems that attempt to uncover the connotations that are embedded in our photos. Depending on the product, the output can be a simple slideshow of images, a stylized photo, an automatically generated photo album or even identifying images that could be deleted [Fig. 1].

In my opinion, the emergence of artificial intelligence, machine learning, and computer vision has provoked an opportunity for innovative research that explores how we maintain agency over machines within the space of our personal photos.
Research Questions

How can design be used to explore interactions with personal photo collections?

How can these human interactions guide the computational process of analyzing personal photos?

Audience

The intended audience for this project is other designers and technologists. My hope is that the research and designs will provoke further research and design explorations in addition to a heightened awareness of the importance of human involvement with machines.

Goals

My goal for the thesis is to explore new ways to interact with photos that utilize the unique characteristics of both humans and machines. I believe that we are at a turning point in how we interact with photos and that it is necessary to strike a balance of roles with machines that preserves the humanist value of images. This balance is necessary so that humans maintain agency over their memories.
Background

Since its invention 179 years ago, photography has continued to redefine itself, just like any other technology. In early years, photography was limited to a small number of people due to cost and complexity of the process. As camera technology became less expensive and development more straightforward, a new area of hobbyist photographers emerged in society. For years this was the norm until the birth of the digital camera.

The number of personal photos taken has rapidly increased in the past decade largely due to smartphones and decreased cost of storage. The purpose of photos has changed from purely a memory-capturing artifact to one that quickly communicates something. These new ephemeral photos take on a different role in our photo collections. These are the images of sticky notes, whiteboards, restaurant menus to name a few.

One by-product of the increasing number of photos taken is that large technology companies like Amazon, Apple, Google, and Microsoft are becoming the repository for our images. While there are benefits to storing all of your images in a central location, these companies now have access to very personal files. More recently companies are attempting to play a more active role in interpreting our images. These systems are grasping at the data [Fig. 2] that they have available to them and attempting to identify the connotations that only the photographer would have.

The involvement of machines in our personal photos is reducing the significance of the person who decided to capture an image in the first place. These computational systems shift humans into a passive role in the management of our photos. We blindly accept the values and priorities of the tech companies who quietly categorize, organize and prioritize our personal photos. Companies are deciding what images are worthy of becoming a photo album, or what images should be shared. Users of these systems are having their actions unintentionally directed by machines.
Related work

The core areas of research that are relevant to my thesis are photography as a cultural phenomenon and the intersection of human-computer interaction (HCI) and photography.

Photography as a cultural phenomenon

Personal photography has been the subject of research in the social sciences for many years. According to Walker, “The act of photography anticipates the future by ripping the appearance of a moment out of its time, creating a tangible image for the future of what will be the past.”¹ They go on to explain the role of photo albums for people, “Hence it is in their albums that ordinary people capture the complex memories, express the deep feelings, and bind together the fragmented experiences of modern private life into more or less coherent visual statements.”² This research has given me a grounding knowledge of how photography is a cultural phenomenon.

HCI and photography

The research areas of HCI and photography is rich with previous research. Nancy Van House writes, “... HCI literature (on photography) largely addresses issues related to technology design.”³ Her research and writing are centered on bridging the divide between social sciences and HCI research. This research also highlights the connection between technology and memory. Van House and Elizabeth F. Churchill write, “What is remembered individually and collectively depends in part on technologies of memory and socio-technical practices, which are changing radically.”⁴ They acknowledge that they do not want to sound alarmist, but there is a significance to these issues and I completely agree.

There has been inspiring research that explores this problem space. Most notably the work of Odom et al. explores the space of slow technology and novel ways of interacting with photos. This project sets a precedent of inquiry into the photo collection space that I am hoping to build on. The PhotoBox project is “…intended to be used over many years, which occasionally prints a randomly selected photo from the owner’s Flickr collection inside of a wooden chest.”⁵ This project inspired my own work because of its speculative nature and how it reimagined interactions with photos. Odom et al. also talks about the constraint of time as a new area for designers to explore. While my work does not have specific ties to the constraint of time, I was inspired to use other constraints that are typically used in HCI.

The goal of this thesis is to interpret previous social science and HCI research so to create a compelling investigation for future designers and researchers.

---

² 187
Later in the thesis, I will be explaining some technical research that I performed, but at this point, there are two specific computer science areas of expertise that should be explained.

Computer vision (CV) is the research area that is the directly involved in my research and problem space. According to the British Machine Vision Association and Society for Pattern Recognition, computer vision, “... is concerned with the automatic extraction, analysis, and understanding of useful information from a single image or a sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding.” Computer vision allows computational systems to “see” what is present in an image.

The other major computer science research area that has significant involvement is Machine learning (ML). Machine learning is a field of computer science that often uses statistical techniques to give computers the ability to “learn” (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed. ML uses the information that the CV process has processed and identifies patterns.

I acknowledge that both of these computer science areas are not my domain of expertise and not the focus of this thesis. However, it is important to mention these areas because they have implications for my research and it is my intent to have my design work grounded in accurate processes.

This thesis intends to investigate them from a design perspective.

---

My Study

Overview

Through my primary research I hoped to gain a better understanding of the following:

– current technology companies’ efforts in the photography space
– structure and narrative quality of images
– people’s relationships to the photos they take
– how people perceive automatically generated photo albums and videos
– the current landscape of computer vision products

In my study, I used a variety of methods to explore the above areas. This included semi-structured conversations in the consumer product area, interviews about users photo habits, and a technical survey of computer vision products.

Research Structure

Conversations

I had a semi-structured conversation with members of the Microsoft Photos team and a second semi-structured conversation the founders of Lifolio. Lifolio is a software product that hopes to collect a family’s digital and physical artifact in order to form a “life portfolio”. The goal with both of these conversations was to gain a better understanding of the current product offering, how they arrived there, and where future iterations are headed.

To gain a better understanding of the structure of narratives and the practice of storytelling, I had a semi-structured conversation with Andrew Tsao, Associate Professor of Drama at the University of Washington. His expertise is in the area of directing, playwriting, and screenwriting. Early in the thesis, I was thinking about designing a collaborative, both human and machine experience that builds photo-based narratives.
Interviews
In addition to the conversations, I also performed six semi-structured interviews with people who covered a broad range of user profiles related to photo collection use. The goal of these interviews was to gain insights into how other people interact with their photos and their feelings toward automatically generated photo albums. The participants ranged in age from 25 to 74 years-old. The intent of this wide age-range was to cover the entire spectrum of technology exposure, from the analog beginnings of photography moving on to the birth of digital photography and to the ubiquity of smartphone photography.

Technical and Product
The research into the technical and product spaces was focused on comprehending the current capabilities of computer vision technologies. In this research, I performed two experiments. The first one involved me selecting 11 images [Fig. 3] and uploading them into various computer vision platforms. These images were selected based on their variety of content. I utilized the public computer vision platforms of Amazon, ClarifAI, Google, IBM, and Microsoft. The goal of this experiment was to expose the current capabilities and shortcomings of each of the platforms.

In addition to the above software engineer focused products, I wanted to investigate the emerging product sector of artificial intelligence powered consumer cameras. Currently, this area is only inhabited by Google Clips. According to Google, “Google Clips [Fig. 4] is smart enough to recognize great expressions, lighting, and framing. So the camera captures beautiful, spontaneous images. And it gets smarter over time.”

The goal of experimenting with this products was to see how Google was approaching building a consumer product that was entirely controlled by machine learning and computer vision.

Findings

My studies led me to three specific areas that will impact my designs. Each of these areas provides actionable design guidance.

Data’s Role

Data that is created by computer vision processes will impact future interactions with personal photo collections. This data is the underlying thread that connects the individual photos of a collection together. The data that a computer vision process creates often takes the form of specific tags. These can be objects, people, locations, or conditions in the photo. This data, in addition to what is embedded in a photo (date-stamp, location), make up the very fabric of how our photo collections are organized. Misclassifications of tags or absent tags all together were very common in my studies. For example, an image of people posing for a photo at a holiday party yielded tags that identified the broad elements of an image, but lacked any specificity [Fig 5]. Whereas photos of nature and the outdoors were somewhat accurately described [Fig 6]. These discrepancies highlighted the fact that computer vision systems are still in their formative stages.

Fig 5: Google Cloud API label output

Fig 6: Google Cloud API label output
Mental Models
When designers are creating interactions for personal photos collections, they need to take into consideration the mental model of memory retrieval. In the interviews, four of my interviewees expressed that when they are looking for images they use adjacent chronological memories to orient their search. It emerged that this habit was formed based on the chronological structure of digital photo feeds on smartphones and the innate nature of personal memory. In most photo collection products, the search model is mapped existing data structures and there is no acknowledgment of the human mental model. Google Photos is the primary example of how a search method is build on the data structure and not the mental model of a person's memory.

Control and Awareness
When designing for personal photo collections, it is important to give users control and awareness of the computational process. In all cases, the interviewees expressed wanting to have more visibility and control of the process that was happening without their knowledge. These computations that choose and sort images for albums are very opaque. I would argue that not all processes need to be visible to the user, but there is a great opportunity to show the user feedback of their actions. I would say that making a process transparent and understandable gives the user the opportunity to comprehend the complex system that is at work.
Design

Divergent Ideation

The design process started with broad ideation that utilized a slider scale that identifies different variables and where the prototype sits between to two contrasting points. [Fig. 7] These early sketches were completed by hand and were limited to a short amount of time. The goal was to explore the broadest range of ideas without any self-restriction to a specific area. The research was used as a guide to illuminate the technical and cultural areas, but at this point in the design process, I attempted to explore the far reaches of the space.

Convergence of prototypes

Once I had explored around 30 different prototype ideas through sketching, I looked at the ideation more directly through the lens of research. In the initial prototypes were a notable amount of ideas that presented hypothetical future scenarios that critically intensified certain capabilities of photo products (i.e. an automatically generated photo album with images containing featured products). While I do acknowledge there is design value in exploring and elevating these subversive design scenarios, I wanted to be cognizant that the prototypes I developed provided an optimistic view of the future.

From the 30 prototype ideas, I aligned three unique prototypes that grew from combining different aspects of the original set. Each prototype looks at different capabilities that are unique to humans or to machines. The three final prototypes focus on the abstract and literal interpretations of images, the language of memory as an interface, and people and relationships as a filter.
Prototyping

The process of prototyping each of my designs starts with a simple experiment to explore the space and then moves to a conceptual higher-fidelity visualization of the idea. The first prototype takes a critical look at photo albums and the second and third prototype are speculative products. By keeping all three prototypes in the critical and speculative areas, my designs were able to take liberties that would not have been available in a consumer product. Examples of these liberties would include ubiquitous projectors and voice assistants.

Album Trio

The Album Trio was first explored by experiencing Google’s Photobooks. These physical photo albums are created entirely by a computational process. The photos are selected, arranged and a title is even given to that album. This passive way to create a photo album sparked my interest because of the background research that I had done in the space of how people arrange photo albums and the stories they are told in them. I was hesitant that an album created entirely by machines would have an emotional connection to me.

Once receiving the machine-produced album of a recent vacation and handling it in-person, I felt that the album itself was sufficient at representing the experience. There were, however, significant aspects of the narrative that were missing, which was to be expected. At this point, I took the same images in the same order that Google choose, and designed three different albums exploring this intersection of humans and machines in the photo album space.

Each of these albums presents the same photos. From what I can assume Google created the sequence of images automatically because of the high number of images taken in a specific geographic location. This was interpreted as a group of images worthy of an album.
Album A has captions written by me that give context to the image [Fig. 8]. Album B allows the viewer to build a narrative from their own interpretations of the images [Fig. 9]. Album C has images replaced by captions created by Microsoft’s CaptionBot [Fig. 10].

Fig 8: Album A with personal captions.

Fig 9: Album B with no captions.

Fig 10: Album C with images replaced by captions created by Microsoft’s CaptionBot.
The purpose of this prototype is to visually see the differences in how we as humans and inversely, machines construct meaning from images. By having all three of the albums side-by-side, a viewer is able to compare the differing narratives. One from the photographer’s point-of-view, the viewer’s own connection to the image, and the machine’s computational interpretation.

For me, this prototype really highlighted the limitations of how machines interpret images and how complex the deep human connections to images are. I believe that machines will never be able to fully make those connections that are rooted in the actual experience.

*Speech Reveal*

The concept for Speech Reveal was first tested by using an Amazon Alexa Skill to control the appearance of a website showing various photos. I built this very simple application that allowed a user to view different images by asking for them. This prototype was very limited in features but was a necessary step in validating the concept of language as an interface.

Often, the current process of searching for a photo is driven by the photo’s chronological position and the objects that are visible to a computer vision system. This prototype [Fig. 11] allows people to use spoken language and the recollection of a memory to search for a specific image or set of images. It relies on a natural mental process of recalling a moment to reveal images. When someone tries to verbally recount a moment captured in a photograph, images that match the description are shown. As the description becomes more detailed, the group of photos becomes more defined.
This prototype is making the assumption that it exists in a time of ubiquitous voice interactions and projection screens.

Through the data gathered through computer vision, the interface is available to show images based on what a user sees. The intent is that as the verbal recollections begin, certain images will trigger memories and in turn provoke more verbal recollections by completing a loop. The ebb and flow of the recollection are visualized through the change in images.

This prototype highlights the fact that machines have the capability to process and analyze more data than humans, but in the areas of photography and memory, humans make very nuanced connections that machines cannot observe. This balance of interplay between machine-capabilities and human-traits allows for innovative interactions.

Proximity Share

The Proximity Share prototype [Fig. 12] started out as an experiment with Bluetooth beacons and an iPhone application. The goal of the experiment was to simulate how a user’s movement in space towards another person (simulated by a beacon) would change what was visible on their smartphone screen. The experiment utilized a simple iOS application that changed screens depending on which beacon was closest. The outcome of the experiment was that physical movement in space that yielded change to the interface was compelling interactions.

![Still image from Proximity Share prototype video. View full video.](image-url)
At any moment, we have access to every photo we have taken. This prototype explores what happens when access to our photos is dictated by our proximity and personal connection to other people. A user’s photos are blurred until they are near another person that has similar images. The mutually similar images could include faces, locations, or objects. Some of the similarities seen through images may be straightforward and others may be only apparent to machines. Machines have the ability to analyze large amounts of images and see patterns and similarities between people.

The hope is that the barrier to access photographs will provide an incentive for people to create moments where connections between people are revealed through the changing visibility of images.

The prototype has exposed some interesting connections between physical proximity to others and how that can be the mechanism for an interface.
Insights

Computational systems have the potential to provide real value to humans in the space of personal photos and memory. Inversely, humans have intrinsic relationships to images that machines will never understand. Principles and value systems need to be defined because of these differences that in turn inform future interactions.

We experience the world around us before, during and after a photo is taken. We have a rich multi-sensory memory of that moment. A photo is often the catalyst for reminiscence of that moment, but that same photo may trigger another seemingly unrelated memory. These abstract connections only surface to the person that lived the experience. Our ability to move in our environment also plays a role in what triggers memories. In the third prototype, I begin to explore how the notion of spatial relationships to other people could be an interaction pattern. By having the user’s experience directly controlled by their proximity to other people, the user feels that they are knowingly in control.

Machines only view photographs from a data and visual standpoint. They attempt to identify the contents and decipher how that image fits into a larger collection of images. These computational systems try to force a connection even when the only relation is its chronological placement. This narrow view on photos was seen in the first prototype’s Album C where images were replaced by captions created through Microsoft's CaptionBot. By comparing the nuanced, personal captions that I wrote in Album A to the generated captions in Album C, it is clear that computational analysis of an image is still in its infancy and in my opinion, will never mirror the capabilities of a human.

We have been creating meaningful connections between images and ourselves for a long time. It has not been until recently that the number of images is exceeding our capacity to actively manage and consume. The value in computational systems assisting us with our photo collections is in their ability to make sense of a large number of images. These same systems also need to be restrained and given specific limits in how they formulate experiences and artifacts from our photos. In each of my prototypes there were specific moments when either the limitations of machines emerged or the strengths of humans, in my opinion, were superior. Prototype 1 highlighted the contrasting interpretations of images by machines and humans. In prototype 2 a human mental model of recollection was explored for a search interface. Prototype 3 presented a spatial proximity interaction that only humans could perform.
Conclusion

At the beginning of this project, I was very cautious about the involvement of machines with our personal photo collections. Previously, my opinion was that these machines made up of computational systems were overstepping their bounds and exerting too much control over our photos.

Through my research, studies, and prototypes, it became apparent that the data these systems create will play a significant role in how we interact with our photo collections. I maintain my argument that products like Google Photos are overstepping their bounds, specifically in the areas of album and video creation.

I believe there is value in computational systems and their ability to process and analyze very large numbers of photos. Without the machine’s involvement, we would be left to organize and manage vast libraries of images.

In my prototypes, I explored and exposed the complex relationships that are emerging between humans and machines. I took both a critical and speculative approach to defining the machine’s subordinate role. I acknowledge that this body of work is just beginning to expose the evolving human-machine relationship, but I believe that it is critical that humans are the primary influencer and maintain an active role with personal photo collection.


Appendix

Henry Gallery Installation Photos

Photo credit Mark Woods.

Photo credit Chona Kasinger.