ARTISTIC VISION: PROVIDING CONTEXTUAL GUIDANCE FOR CAPTURE-TIME DECISIONS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jane L. E
March 2021

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

James Landay, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Pat Hanrahan, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Maneesh Agrawala

Approved for the Stanford University Committee on Graduate Studies.

Stacey F. Bent, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

To learn photography is to become more intentional about the creative choices you make with your camera. Many of these creative choices happen in real time during the capture process, as the photographer takes in the scene around them and navigates a space of so many possibilities and uncertainties. However, today's resources for learning photography, such as books, classes, and example photos, are largely disconnected from the capture process. Photographers are therefore faced with the task of navigating, in real-time, a seemingly infinite space of possible creative choices while relying on a disconnected space of learning resources that can feel both inaccessible and overwhelming in the moment. The primary insight of my research is that real-time contextual guidance, embedded directly in the camera, can make accessing relevant parts of this wealth of information more approachable and actionable. The feedback assists in cutting through the noise of endless possibilities and focuses photographers' attention on targeted, meaningful creative choices.

My dissertation presents a set of capture-time interfaces that provide real-time contextual guidance. This guidance takes the form of light touch cues presented as automatically generated visual overlays, where each overlay is designed to focus on a specific photographic concept. Each interface's goal is to understand what an expert might be noticing in considering the targeted photographic concept and to, via an annotation overlay, direct a novice user's awareness in a similar manner. In designing this real-time contextual guidance, I take inspiration from photographers' current practice of directing attention through manually drawing annotations onto photos. Today, this practice is mostly restricted to post-hoc feedback used to point out specific decisions or potential mistakes that the artist made. I develop algorithmic

approaches designed to understand conceptually relevant aspects of the scene that the photographer is viewing. These algorithms generate annotations that are displayed in the camera in real time. The annotations can move beyond explaining why a specific decision was made, towards helping the photographer become aware of artistic choices that *could* be made, providing guidance while encouraging creativity and exploration. Through the overlays, we hope to help novices train their eye to see in the way that experts do.

Specifically, I present in-camera guidance interfaces tackling three important photographic concepts: portrait lighting, composition, and decluttering. The portrait lighting tool helps users be more aware of the available lighting styles and reorient their subject to best achieve the lighting style of their choice. The composition guidance tool makes users more aware of the current composition by highlighting lines in a composition grid that are most relevant to the camera view. The decluttering tool increases users' awareness of clutter that would draw attention away from the main story of the image by abstracting the camera view to outline edges around the subject(s) or the image borders. For each interface, I describe my process for designing a novice-interpretable visualization and how it captures context relevant to the target concept. I then evaluate each interface by asking novice photographers to take photos with these tools while focusing on their target concept.

Together, these tools and their evaluations demonstrate that such awareness-based visual guidance camera interfaces can help people be more intentional about their artistic choices. By making users more aware of possible options and mistakes, the interfaces introduced in this dissertation encourage users to explore the space in a more informed manner. In this way, the tools presented in my dissertation help users become more confident in their ability to achieve their artistic goals.

To 妈妈

为了我您放弃了自己的PhD。
30年后，我终于帮您补上了一个。

♡ 简简

Acknowledgements

It's hard to believe that I'm actually sitting here writing the acknowledgements section of my dissertation. Wow... I actually made it... here. But I would be lying if I said that this wasn't an incredibly tough journey. Many times along this path, it was only with the support of my advisors, my friends, and my family that I was able to keep going. So I want to start by just saying thank you to all of you that have been there for me these past many years. Words have never been my strong suit, so here I'll begin to try to express my appreciation for all of you who have helped me on this journey and who have made this dissertation possible, but I emphasize that the appreciation goes much further.

I'd like to start by expressing how grateful I am for the constant support of my advisors, James Landay and Pat Hanrahan. I'm so grateful for the relationship that I have with both of them. Even more than research advisors, they were my life advisors for the past many years.

> When I first approached James as a first-year PhD student, I had no background in HCI. Yet with zero hesitation, he chose to take a risk on me as a rotation student, and shortly after I became his first Stanford advisee. James very quickly became someone I knew I could be vulnerable in front of and could confide in with any problems, research or non-research related. Perhaps something that contributed to this was TAing for him twice, once in China for a several week summer workshop where the class basically spent all of our time together for the duration of the course (thank you for giving me that very special opportunity). Even in this past quarter as he tried to reduce his number of Zoom meetings, he

kept my weekly full hour meetings to make sure he was available to support me as I finished my PhD, despite most of those meetings being about my personal life rather than actual research progress. Thank you James for being someone who remembers small anecdotes, e.g., that egg tarts are one of my favorite childhood snacks, and finds ways to use that information to cheer me up—on multiple occasions he has brought me several as a much needed pick-me-up, including a box not long ago to fuel my dissertation writing.

Pat has a way of lighting up when he's genuinely excited about an idea (or a particularly beautiful pixel or art project). I will forever remember when this happened during a meeting I had with him; and at that moment, I finally felt like I had landed on a topic worth pursuing for my thesis. Pat also recognized my teaching and mentorship skills early on and gave me an opportunity to teach with him, even though I was unfamiliar with some to the content. Despite my doubts in my ability to TA the class, Pat always found ways to reassure me and make me feel more valuable to the class than I could recognize on my own. Thank you Pat for constantly reminding me to be more confident in myself and in my work.

From watching both of you, I've learned so much about the type of person I want to be. I don't know what I did to deserve your time and trust, but your unwavering support and willingness to invest in me has helped me to become the person I am today. I hope that someday, I can have that kind of impact on someone's career and life.

In addition to my advisors, I've been incredibly lucky to have the support and guidance of my committee members and many amazing mentors throughout my PhD.

Maneesh Agrawala stepped in when I was particularly lost during my third year, in my pursuit of the ever-elusive "thesis topic." We had been informally chatting during his visits to my office to talk to his advisees, and we realized that the topics I was brainstorming somewhat aligned with his research interests. He offered to meet with me weekly to brainstorm project ideas, and since then I've

been lucky to have his guidance in shaping the direction of my work. He's taught me so much about how to write research papers, down to making figures using consistent fonts and arrow shapes. Thank you Maneesh for going out of your way to give me yet another lab to belong to.

I met Ge Wang for the first time during our meeting to discuss the workshop in China (that he would be co-teaching with James, and I would be TAing). We immediately bonded over our shared Princeton memories and love of food. Ge's enthusiasm for sharing food (and drink) helped to create a sense of community in the class that was reminiscent of my favorite college memories. Thank you Ge for believing in my work and my vision, and for sharing your love of food and photography with me. As the one who is usually behind the camera, that summer was one of the most documented experiences of my life.

Scott Klemmer, thank you for believing in me, and for the time and effort you've already invested in me. Thank you for being so understanding and supportive as I struggle to wrap up my dissertation. I look forward to learning and growing with you and your lab in the coming years.

Thank you to my many amazing industry mentors who have found ways to continue to support me past my internships.

Thank you Jose Echevarria for being an amazingly supportive mentor with an incredible artistic flair. Thank you for being someone who shares my vision so completely that we were able to complete each other's sentences on our first call. And thank you for being someone who is willing to suddenly hop on a call with me at 7pm to calm my nerves before a weekend of user studies.

Thank you Mira Dontcheva for always being available to meet me over meals or coffee at conferences or during my Seattle visits. Thank you for being such a constant source of support and encouragement despite not ever having formally been my research mentor.

Thank you Neel Joshi for giving me the validation I needed, that I indeed had the potential to be a successful researcher, at the time when I needed it most.

Thank you Merrie Morris for going out of your way to champion me and my work, and for making known to me many of the most influential opportunities I've had during my PhD.

I'm grateful to have been part of the community that is the Stanford HCI & Graphics lab for the past many years. Thank you to everyone in the lab who has volunteered hours and hours of time to help me with my research and provide me with invaluable feedback throughout the course of my PhD.

This dissertation would not have been possible without the help of my main collaborator, Ohad Fried. Thank you Ohad for being a constant source of guidance, reassurance and friendship—as well as snark. Thanks to you, I'll always remember to judge myself for overly large PDF files and overthinking minor details. I hope you figure out the body roll and shimmy someday.

Thank you to my officemates (Dan Fu, Jingyi Li, Mitchell Gordon, Sean Liu, Anh Truong, and Kesler Tanner), and honorary officemates (David Yao, Griffin Dietz, Parastoo Abtahi, Will Crichton, Teguh Hofstee, Ante Qu, Evan Strasnick, Ranjay Krishna, Jacob Ritchie, Michelle Park, and Arjun Arora), for sharing in the little daily ups and downs of research, for being people I could be weird and silly in front of, but also for being people I could have extended conversations with about serious topics. I'll miss our whiteboard of Pokémon sketches, our office Zootopia costumes, and matching Pokémon shirts.

Thank you to Ed Quigley, Winnie Lin, Mikey Bowser, Minjae Lee, Matthew Cong, and the rest of the Fedkiw crew for providing me with a space to go to when I needed a break from work, for the many coffee breaks, and for the countless Tobang trips. Thank you Winnie and Ed for being my thesis fairies. Thank you for all of the surprise treats, and for making sure my fridge was stocked when I didn't have the energy to buy groceries.

Thank you Mitchell Gordon for everything you've done to make it possible for me to take steps towards the finish line. You've found countless ways to support me in this endeavour, from bringing me surprise chocolate milk, to actively

cowriting with me, to finding ways to motivate me to work, to being willing to take walks with me to improve my mood (sometimes at a moment's notice). I'll miss biking with you to work despite not at all being on your way, and I'll miss our long walks around campus discussing life or studying quals readings.

Thank you Griffin Dietz and Parastoo Abtahi for both being always willing to jump in and help me with my work (and life decisions) in any way possible. Thank you both for being a constant source of company and support during the pandemic through regular outdoor park hangouts. I am grateful for our many adventures together throughout the PhD from birthday dinners to an extremely food-filled Seattle weekend (oops.)

Thank you Griffin for always lighting up my day with constant bits of joy, whether it was taking photos of flowers or making animal impressions in the hall, just for me to see.

Thank you Parastoo for the most perfect gifts—the necklace of a camera with a Mickey head, and the custom-assembled spherification kit. Thank you for creating the perfect relaxation setup, lying on a hammock while enjoying homemade tiramisu.

Thank you Kesler Tanner, my first 3B PhD friend who I actually saw outside of the office. Thank you and Jane for welcoming me into your home to meet Fenton when he was just a week old, for the dinners and playground outings and the many cute surprises, from Minnie cupcakes and beautifully handwritten cards to surprise baby visits to the office.

Thank you Teguh Hofstee for being my carpool buddy and punching bag. Thank you for going with me on day trip excursions to food and to explore, but mostly to food, and I guess of course to tea. Thank you for being there to quietly listen to and deal with all of my rants, and to help me with all of life's most important decisions from which cocktail to order next to which outfit to wear for a dance cover. And thank you for always making sure to document my most unflattering moments.

Thank you Will Crichton for the many hours you spent helping me draft thesis statements and revise my defense motivation, repeatedly. Thank you for the many chats over coffee (including when I was home in NJ), and for encouraging me to learn the harmony to “Lucky.”

Thank you Jingyi Li, my fellow custom button enthusiast, for reading a early draft of my dissertation intro, and for being very willing to just sit in a box and take photos on a random day in the office.

Thank you Dan Fu, my fellow dancer in the office. Thank you for the many dance video breaks and for patiently teaching me at social dance and salsa events.

Thank you to Sean Liu and Ranjay Krishna who both happen to be my almost birthday buddies, and also happen to be two of my Brown Institute All Hands Manhattan exploration buddies.

Thank you Sean for the impromptu photoshoot, walking along the High Line. Thank you for always being excited to talk about KPop covers and for being so enthusiastic about showing up to and recording my dance performances.

Thank you Ranjay for coordinating the fun night of NYC bar hopping and karaoke-ing, culminating in pickle backs and Halal Guys. I’ll always remember our many drinking adventures, also with Cathy Yuan and Matthew Cong—for the drinks, but mostly for the company and conversation.

Thank you Ross Daly for the numerous coffee chats/walks, and more recently the beer shares. I will forever remember the night at Varnish that ended with a group of us singing on a hill in LA overlooking the city lights.

Thank you Ante Qu, my fellow Pirate, Princetonian, Microsoftee, and Stanford PhD. Thank you for the Nerf guns that have provided endless entertainment (and distraction) in our office and through the halls of 3B.

A special thanks to Anh Truong for keeping me sane in the days leading up to deadlines, for staying up with me many nights in a row to help me with making figures and assembling paper submissions. You were with me through some of the toughest times in the PhD, and I really don't know how I would've done it without you. You were one of my first friends in graduate school, and I feel so lucky to have met you when I did. I can't wait for future opportunities to continue traveling the world with you.

And thank you Matthew Cong, my student buddy and first collaborator at Stanford, and so so much more. Thank you for everything you did behind the scenes to encourage me in my work and my hobbies, and for helping me manage my stress as I worked towards completing my PhD (especially in these past few days). Thank you for always making sure I had access to my favorite foods and for taking me to Disneyland more times than I can count. Finally, thank you for helping fulfill all of my dissertation requests, down to the heart.

During my time at Stanford, the dance community always made it feel more like home. I'm happy and thankful to have found another dance family in XTRM and am grateful for all of our performances together, memories that I'll treasure and rewatch for years to come.

Thank you to Aloy Makalinao for welcoming me into the community and encouraging me to take part.

Thank you to Arjun Arora for the constant encouragement, and especially for giving me the much needed nudge to get back into dance after a several year hiatus to focus on research.

Thank you to Mayee Chen for being my fellow CS PhD, Triple 8 alum, and XTRM dance friend. I was so excited to meet you during admit weekend, and am so glad we've gotten a chance to become closer friends.

Thank you to everyone in XTRM who believed in me and patiently taught me despite my slow choreo retention, especially Emily Wu and So Young Lee.

My thesis was absolutely a team effort. To all of my friends from all walks of life, thank you for being my team. From traveling for food to long talks over cocktails to coworking sessions and beyond, I really couldn't have done it without all of your love and encouragement.

Thank you to my college friends for continuing to be my unwavering support system.

Thank you to my college roommates, Alex Kim and Catherine Hsieh. You've forever defined what it means to be a "roommate" in my mind.

Thank you Alex for being my coworking buddy leading up to my defense, and for providing me with your fairy godmotherly love and encouragement as I pushed to finish my PhD. Thank you for being my tea date buddy and for appreciating the roastytoasty with me. Thank you for being there and sharing in all of life's moments with me, despite being across the country.

Thank you Catherine for making NYC always feel like a home, and for continuing to be my roommate during my (previously) frequent visits. As I work to finish this dissertation, I'm recalling the night that you stayed up and kept me company while I finished writing my college senior thesis, all while sewing a bag for my group assignment.

Thank you Steve Kuei for being my fellow grad student in our college friend group. Thank you for fueling me with a mountain of yummy surprise pastries for working on my defense, for pushing me to grow as a dancer, for providing the background soundtrack to my work these past few months, and for actually editing... my entire dissertation...

Thank you Seung Nam and Rose Nguyen for supporting me during the first moment when graduate school almost pushed me to a breaking point. I'll never forget the day during the first year of my PhD when you drove down together from Berkeley to bring me a beautiful pink treat and a lovely accompanying card that somehow made everything feel OK.

Thank you Seung for always being late, but ready to judge me. But also, thanks for supporting me from the very beginning of my grad school journey by keeping me company as I worked on applications back in Seattle.

Thank you Rose for being someone who knows me well enough to translate my strange sounds and gestures into fluent and eloquent complete sentences, and for communicating my thoughts better than I ever could. Thank you as well for supporting me in my current journey of self-reflection and discovery.

Thank you Calvin for being my "preschool buddy" and basically a brother (and for storing my stuff at your parents' home for way too long... sorry). Thank you for being there for me during all the most difficult and confusing parts of the PhD.

Thank you Laura Herman. I'll forever be grateful for the cold email and your excitement to hear of a fellow Princeton/Adobe/HCI friend. After our very first in-person hangout, I knew we would be great friends. Thank you for being the supportive friend that you are and always finding ways to cheer me on as you live your nomadic life.

Thank you to Tiff Liu, Neal Bennett, Maple Zhang, Angela Wu, and the rest of CBCB for the many game nights and karaoke sessions, the Overcooked and takeout meals, the group trips to Seattle and Big Sur, and for introducing me to Draeger's chocolate cake.

Thank you to my high school friends for supporting me from near and afar, for the many Friendsgivings and New Years parties. Thank you Alson Wu, Charlene Seto Kung, Chris Wu, Cyndie Hu, Dorky Ahn, Gary Chung, Hansoo Kim, and Joe Guan for being a group that I've been able to rely on for almost two decades now.

Thank you to Chris and Vivian for being my quarantine family, and for giving me something to look forward to each week, and feeding me soooo much delicious food and beer (and even saving me some when I couldn't attend in person).

Thank you Dorky for being your dorky self. It's been an incredible journey growing up with you these past few years, being able to share in each others' struggles and successes. Here's to more randomly exploring Japantown and taking lots of Snow selfies.

Thank you Cyndie for the support from afar. Thank you for cards from Japan and for hopping on an impromptu catch-up call despite being so many time zones away. And thank you for always being there to listen and to reassure me even in the most difficult of times.

Thank you Charlene for being my fellow EIC in yearbook back in high school. The InDesign skills that we gained in that time has become invaluable in my research. I look forward to more adventures exploring random art spaces and eating cheesy pasta with you in San Diego.

The pandemic has been an especially difficult time, and I'll be forever grateful to everyone who went out of their way to bring bits of joy into my daily life. Thank you to the many friends who still found ways to support me during this time. From food deliveries, to surprise packages and random little gifts, to virtual coffee dates, outdoor socially distanced hangouts, or even just the occasional text to check-in, each thoughtful gesture meant so much.

As I worked towards graduation, it was so hard to stay motivated during this time. It really meant so much knowing I'd be able to see so many familiar and friendly faces at my defense, and to have something that genuinely excited me to look forward to. So really, thank you so much to everyone who showed up at my virtual defense and made it an incredibly memorable and special experience.

Thank you Tricia Ngoon for being my thesis buddy. Thank you for all the spirited discussions at our weekly "work" sessions that I so look forward to, covering topics from cute dinosaurs, to chubby babies, to progress updates, to research philosophies—it's all about that adaptive guidance, am I right?

And a special mention to Bryan Chiem (his coffees are probably responsible for fueling a good chunk of the text of this dissertation). Thank you Bryan for your carefully crafted and practiced coffee experience. Tono Coffee Project became the first public space I felt safe enough to regularly visit in pandemic times, and brought a much needed bit of normalcy back into my life.

Thanks to my family for always supporting me in my academic as well as creative endeavors. As a family, we took a trip to NYC on the coldest day of winter that year because I wanted to take photos of benches at Central Park for my photography class. Then again for my research, they spent many hours in the NJ winter wandering campus with me to take portraits with different types of lighting for my paper. Thank you to my parents who continue to find ways to support me even from a distance, including by giving me feedback on a draft of my defense talk (despite my dad being in China), as well as on drafts of my dissertation. I truly would not be where I am without you.

Finally, I want to thank a mentor who has a very special place in my heart.

Floraine Berthouzoz passed away while I was interning with her at Adobe after my first year. Despite having only known her for a short while, she's had an immeasurable impact on my life. A few days into the internship, I remember calling my parents and excitedly telling them that I found a mentor that I was really hoping to work with for the rest of my PhD. And while she hasn't been able to physically be there, even today, sometimes when I feel like giving up, thinking about her and how much she believed in me, gives me the strength to keep going. I wish I could've had more time with you, to share in our love of dance and chocolate, and to celebrate our birthday together. Thank you Floraine, for everything you did for me. You are missed.

Thank you Marie-Paule Berthouzoz for welcoming me into your home and for showing us around your beautiful hometown. It was incredible to have an opportunity to get to know you (and I hope to be able to come visit again soon), and to see all of the traits I admired about Floriane in you. Thank you for

telling stories of her (and your) adventures and allowing us to share in those memories. Thank you for your constant support and for making the time to attend my defense. Your presence meant the world to me.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Photography has become embedded in our daily lives. It is an essential part of how we save memories, as well as how we communicate. With the increasing prevalence of digital cameras, more and more people are interested in learning to become creative photographers. Ample resources exist from which one can learn the artistic aspects of photography. Experienced photographers have shared their craft with the world, producing a collection of exemplar photos of almost every scenario imaginable. They've additionally taken the time to produce books and videos to carefully document their processes and their knowledge of photographic principles. However, this information can be overwhelming for novices and is difficult to access in the moment while one takes a photo.

In-person photography classes often include a critique portion, where the student can receive feedback on their captured photographs. The contextual nature of the feedback provided is incredibly helpful for the student to learn how to apply artistic photographic concepts to new scenarios. However, these classes face two challenges. First, few learners actually have access to them. Second, the delayed post-hoc nature of this feedback loop means that the photographer cannot explore creative choices in real-time. In many cases, students can reflect on their creative choices, but not change them, as they often cannot recreate the originally photographed scenario. Inspired by these critique sessions, the work in this dissertation aims to solve both of these challenges by making feedback more accessible and by further closing this feedback loop by providing automated feedback at capture-time. This real-time feedback allows the artist more opportunity to physically move around in the scene to discover approaches that incorporate and address the feedback. Specifically, I design capture-time interfaces that provide live feedback in the form of visual annotations that guide the users' attention and awareness. Therefore throughout the document, I use the words feedback and guidance interchangeably.

My approach to providing capture-time feedback is inspired by an existing photography practice—when describing the artistic structure of a photo, photographers will often manually draw annotations on the image to direct the viewer's attention to specific elements of the photo. Based on this existing practice, I propose providing feedback as interactive overlays on the camera view. In this dissertation, I propose three new capture-time camera interfaces that each provide contextual feedback to aid users in learning to consider a single, specific photographic element at a time: portrait lighting, composition, and decluttering. The design of these interfaces is informed by a set of design goals synthesized from formative surveys and interviews I conducted with participants with varying levels of photography experience. I describe the development of each interface in the context of these design goals, and then present results from summative user studies to demonstrate that such capture-time interfaces allow users to maintain creative freedom while producing photos that they believe are higher quality.

1.1 Photography

In-camera guidance tools are more desirable now than ever. Over time, the nature of photography itself has changed such that it has become central to our everyday communication. However, while photography is becoming increasingly important to our lives, there are currently limited tools for assisting a user in growing as a photographer. In this section, I describe the factors that have allowed photography to become so widespread, the challenges of pursuing photography as an art, and some of the benefits and limitations to current photography learning methods.

1.1.1 Prevalence of Cameras

People have long enjoyed taking photos and, in particular, documenting memories through photographs. In the past, this meant hiring an expert to bring their camera to a specific location to take photos. The introduction of Kodak's snapshot camera in 1883 provided a more portable device for documenting personal experiences.

The advent of digital cameras further transformed personal photography—while still needing to carry a designated device for photography, individuals became more likely to carry cameras around regularly, rather than just on vacation or to special events [177, 284]. Furthermore, in addition to this changed materiality of photography from film to digital [177], in recent years a large population has gained access to relatively high quality cameras directly built into their phones that travel around with them throughout the day. Cameras have become one of the most important features of a phone [139, 226], and are rapidly improving in quality.

These phone cameras have made way for a new form of more casual and serendipitous photography. People no longer have to spend a fortune on a separate device (and multiple lenses and possibly film), spend hours learning the knobs and dials of a complex camera, and lug around the heavy equipment to capture photos. The digitization of photos and easy access to digital storage essentially removes the cost of each individual photo, enabling casual photographers to feel less limited in experimenting with different shots [164, 177, 284]. The number of photos taken per year tripled from 2010 to 2015, with 75% of those captured using a phone in 2015 as compared to only 40% in 2010 [129]. People are taking and sharing more photos, and as a result, more and more people are interested in improving their photography [52, 129, 164, 177, 284, 298, 314]. However, there is still limited access to effective opportunities for learning photography.

1.1.2 Photography Challenges

> *"The simplicity of photography lies in the fact that it is very easy to make a picture. The staggering complexity of it lies in the fact that a thousand other pictures of the same subject would have been equally easy."*

– *Szarkowski* [273]

Unlike many other art forms, there is no "blank canvas" in photography. The act of taking a photo is incredibly easy—any image seen through the camera viewfinder has the potential to be a candidate for the final photo. However, for the same reason, "doing photography" is incredibly challenging. For a novice who is interested in photography, but has limited training and equipment, the prospect of trying to take a good photo can be somewhat daunting [27, 74]. As personal photography becomes integrated into our daily lives, it increasingly becomes a means of identity formation and communication rather than simply memory preservation [284]. The significant role that photos play in our lives makes the task of taking a good photo increasingly important. But how does one transition from taking casual pictures to actually "doing photography"? The increased popularity and significance of photography further adds to the intimidation of trying to take the step of crafting a photo [27, 32, 74].

> *"... photographers' viewfinders are never empty given any amount of light... Photographers, however, are essentially engaged in a subtractive process, one of taking away or distilling. They select from the entire universe available to them."* – *Barrett* [27]

In the moment of capturing a new photo, there are numerous considerations to manage simultaneously. The camera is impartial to the objects within its field of view; it is up to the photographer to determine what is important and how to frame the image to communicate its importance, and to do so "aesthetically." In describing how photography is a different form of "picture-making" from other additive art forms like drawing or painting, Barrett describes this "selectivity" as a unique characteristic of photography. Photographs are highly selected instants in time, whereas painting involves the artist gradually adding strokes one at a time to the final image. Amongst everything in their view to choose from, photographers need to make a long sequence of selections to arrive at a single photo—they need to pick a subject to capture, an angle to capture it from, how much it should fill the image, what to keep in the background, etc. Rather than doing all of the selecting required to narrow down to a specific image in the moment, photographers will usually capture hundreds if not thousands of versions of a photo to select a final image from later on [27]. This process of selecting can be rather overwhelming for a novice trying to learn photography.

1.1.3 Learning Photography

What are today's methods for trying to learn photography? At the highest level, the resources for learning photography can be broken down into approximately two categories: static resources such as books or video tutorials that one can use to self-learn, and traditional art practice classes where learning is guided by an instructor. We will walk through the benefits and limitations of each of these approaches.

Static Resources

Photographers have produced (and continue to produce) volumes of content across various media and platforms that are readily available to a wide population. This content can range from books that teach basic photography principles [110, 111, 138, 169, 273], blog posts that describe interpretations of existing artwork [200, 243], or even YouTube channels and Instagram profiles that document photographer's personal workflows behind the scenes. The benefit of such resources is that one can gradually absorb vast amounts of information and inspiration through this content and can refer to it whenever it is needed. Experts have employed a range of methods to help capture and share their knowledge through words and annotated examples (see Figure 1.1), and so almost any information one might want to learn is likely captured in many formats.

Nonetheless, there is still a gap between having access to this information and being able to actively use it. When trying to capture a particular photo, it can be hard to determine how to apply the information learned; in order to execute, one needs to remember/find every concept to consider in order to first determine what is relevant to their situation, and then understand how to apply it. Novices often will not have the prior knowledge necessary to be able to know what questions to ask in searching for the relevant resources [210].

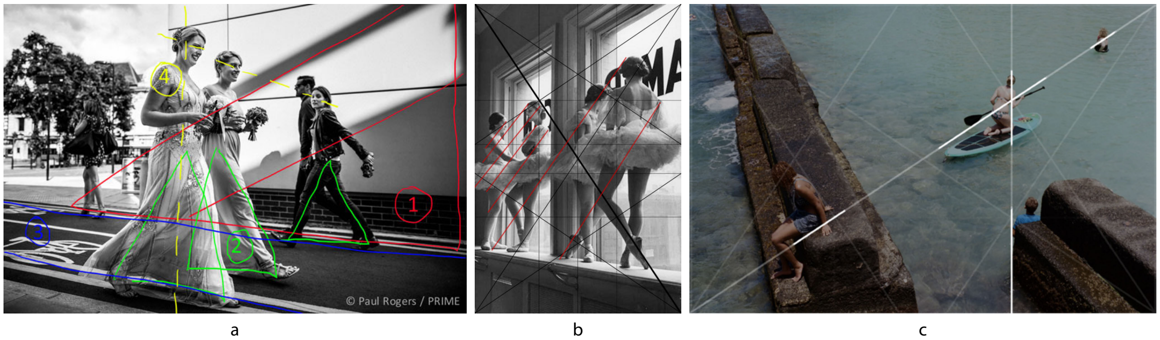


Figure 1.1: Photographers provided these manual annotations alongside text to aid in the understanding of these photos' compositions. Left to right: (a) Rogers highlights his use of the dress, shadows, and gaze to create triangles and diagonals in the scene [243]; (b) Marelli describes Alfred Eisenstaedt's use of the strong diagonal through the ballerina's leg (bottom right) along with repeated diagonals connecting through the ballerinas' bodies to make this image more about the ballerinas than the window frame [200]; (c) Glover highlights the alignments between the people and structures that form the overall composition [111]. Photos by Paul Rogers, Alfred Eisenstaedt, and Tavis Leaf Glover.

However most importantly, since this content is often static, it therefore needs to rely on describing these concepts in the context of existing examples. In practice, it can be difficult to take these abstract concepts and adapt them to the context of a new photo. In the education literature, this is referred to as transfer—it can be challenging to bridge the gap when transferring knowledge you've learned in one context to a completely new context [13, 216].

Traditional Classes

Traditional art practice classes on the other hand, are often more effective for addressing this knowledge gap, but are also harder to make broadly available as they require the time and involvement of expert photographers/instructors. One unique aspect of these classes is that they focus heavily on in-class critiques—in which the teacher and other classmates provide feedback directly in the context of the student's work [87, 92, 95]. Commonly this feedback translates abstract photographic concepts into descriptions of how they relate to the student's work, often taking the form of annotations directly in the context of the students' work (e.g., sketched or air-drawn as gestures by the teacher). These annotations can sometimes be used to highlight characteristics of the photo or to point out mistakes to fix (see Figure 1.1). These concrete visualizations of abstract concepts can feel more defined and actionable, narrowing the knowledge gap [48]. This form of feedback reduces the burden on the student both to find the relevant concepts to consider or questions to ask, as well as how to apply them [48,210].

However, the teacher's instruction and feedback is still a scarce resource—they'll usually only provide feedback during class-time, and their attention is spread across the many students in the class. Therefore, each student will usually only get feedback on a select subset of their work. The class may occasionally go on photo walks together, where the student can receive more immediate feedback from the teacher and from peers. Nonetheless, it is infeasible for teachers to be around to give feedback on each and every photo as each student is out in the field capturing them. Thus, when the feedback is provided, the student is often already more committed to the image. This can make receiving feedback more intimidating, but also more difficult to apply. The student has already left the location where this photo was taken, and so the effort required to capture new images that apply most feedback at this point is significant. It would be preferable to receive this feedback in the moment when the student is taking the photo and able to explore new ideas with the context of the feedback within the environment [13].

This is somewhat unique to photography due to the immediacy of this art form as compared to other slower art forms like sketching, painting, or sculpting, where much of the art process is happening in the classroom. In these classes, teachers walk around and do in fact give feedback in the moment as the student progresses through each step of creating their art piece [27]. How might we enable this type of contextual, real-time feedback in the moment for photography?

1.2 In-Camera Guidance

As researchers develop more computational methods for understanding and generating visuals, I wondered if it might be possible to leverage some of this work to help those trying to better learn about photography. In particular, I explored the possibility of using these computational methods to provide photography guidance directly in the camera. In doing so, I considered ways in which such interfaces might help address some of the limitations of traditional photography learning methods, while also maintaining some of their strengths.

One immediate benefit to having in-camera guidance is having the feedback directly on the device capturing the scene. The nature of the guidance being given in-situ has implications for both the potential feedback that can be given as well as how a user can respond to the feedback. It means that in providing feedback, the guidance can incorporate additional knowledge from the photographic environment. It also means that the user is able to consider this feedback while physically in the environment trying to capture the scene and can immediately make adjustments accordingly.

1.2.1 Context is Readily Available

As we saw from traditional photography teaching methods (see Section 1.1.3), context is helpful for reducing the knowledge transfer gap. For static materials, photographers try to teach concepts directly in the context of existing photos as examples of the more abstract ideas. For in-person classes, teachers use a combination of such examples as well as critiques of the students' own artwork. Providing photographic guidance directly in the camera enables access to more context surrounding the photo. It enables the use of this context to anchor suggestions (e.g., searching for relevant static resources to present to the student). It also enables the use of this context to generate new adaptive guidance in the context of the current image.

Being directly integrated into the camera, the guidance can take advantage of information captured at the time of taking the photograph. The guidance can consider not only the context of the exact photograph, but can also easily have access to preceding frames from the camera viewfinder, or even have additional sensors for tracking other environmental factors such as location, time, weather, lighting, crowdedness, or camera movement. The availability of such context naturally enables the guidance to adapt to the scene, such as to help determine the relevant topics for guidance, examples to show, or style of feedback.

We have seen the benefits of being shown examples early on in the creative process for improving overall creativity [170]. Using computational methods increases access to a growing library of images on the internet rather than the few selected by the author of a given book or blog post. We've also seen the benefits of individualized and contextualized feedback in computer-assisted intelligent tutoring systems for subjects like arithmetic or geography [225, 285]. Researchers have also started looking at providing this type of adaptive guidance for creative domains, including photo editing [98, 168]. We want to leverage the additional context available at capture-time to provide such adaptive photographic guidance in the camera.

1.2.2 In-the-Moment Feedback

In addition to being able to design computational guidance to incorporate additional context, another advantage to receiving feedback in the camera is the immediacy of the feedback. Relying on classes or static resources involves waiting to be in class to receive feedback or searching through resources to find the relevant information. Providing photographers with artistic feedback directly in the camera viewport, allows them to use this information in the moment as they make decisions about their photos, thus further closing the feedback loop.

By moving feedback into the camera, photographers can respond to the guidance while the image in question is still in frame. This means the photographer can make adjustments to the image either by moving the camera or objects in the scene in response to the provided guidance. Schön describes a practice of reflection-in-action where a skilled practitioner works through an open-ended problem by repeatedly experimenting with different options and reflecting on these choices to inform future actions [247]. In this way, we hope that this in-the-moment feedback can also better support a reflective practice. By helping the photographer better understand their artistic choices and encouraging them to see their images from different angles, it can assist the photographer in engaging in active internal conversations with regards to their photographic decisions [123, 248, 274].

Additionally, design thinking teaches that having feedback in-the-moment is crucial for encouraging creativity through iteration and exploration [79, 80, 149, 233, 253, 254]. Specifically, it encourages exploration through rapidly prototyping drastically different ideas or solutions, and receiving feedback on these earlier prototypes. In photography, this suggests that it would be better to iterate in the camera, trying out different ideas rapidly and immediately getting feedback, rather than waiting to perfect a single idea later in post through editing.

Once the photo has entered the editing room, the photographer is much more committed to the image. The changes that are possible are limited—the photographer can no longer benefit from the ability to interact with and move around in the environment and immediately respond to the result. Thus, we aim to provide feedback while the photographer is still at the scene and able to interact with the physical space being captured in the camera.

1.2.3 Learning Through Physical Movement

Photography is a spatially oriented and physical activity. A photograph is a documentation of how someone sees and frames a physical space. It involves the photographer walking around a space, finding an object to photograph, and making sense of the object and the story as they frame the image through the camera. It is a dialogue between the physical environment, the photographer's vision, and the camera's gaze [177]. In-camera guidance can contribute information to this conversation that the photographer can immediately apply to the current photo. As we just saw, there are numerous practical benefits to receiving feedback on photos while still physically at the scene, simply due to the ease of manipulating the camera as well as physical objects in a photograph.

However, perhaps more importantly, this encouragement of immediate response through physical actions promotes active learning and thinking through doing, by enabling the mind (thinking) and body (doing) to work together to develop an understanding of the image and the scene [43, 166, 283]. A lot of how photographers interact with cameras involves "epistemic action" rather than "pragmatic action," that is, they'll make constant adjustments to explore the scene rather than necessarily to accomplish a specific task in mind [165]. People might manipulate the camera position, the light in the environment, or the objects in the scene to quickly test out options. Instead of trying to imagine these different scenarios, the physical movement can facilitate the mental work—the photographer can quickly better understand the interplay between the different factors in the image and how changing them influences the overall scene and the final image [131]. Providing additional information through in-camera guidance can further encourage this kind of active and explorative learning.

1.3 Artistic Vision

In particular, we are inspired by the fact that experts have trained their eyes to "see" in a certain way [87, 169, 201, 246]. As they look at an image, they know how to scan the scene to make sure they are aware of potential different ideas, or mistakes. As a result, they can be very intentional about their different artistic choices despite the wide range of options presented in front of them.

Novices, on the other hand, might not know where to put their attention or what to consider when framing a photo. We wondered, if we could help novices see in these different ways as well, would they then be able to have access to the range of artistic choices that experts have available to them?

The inspiration behind this dissertation's title comes from this goal of providing these users with interfaces that encourage them to see the image in the different ways as an expert would and to be aware of different aspects of the photo, in order to encourage them to develop their personal "artistic vision."

1.3.1 Visual Annotations

Specifically in this work, we approach this goal of helping develop "artistic vision" through designing annotations to overlay on the camera image, similar to those that professional photographers manually produce (Figure 1.1), that aim to guide the awareness and attention of the user. While the photographers will still describe the concepts and the relevant components in text, these annotations are visually anchored within the image, and hence can more quickly and accurately direct the readers' gaze to the right spots in the image.

I focus on visual annotations that provide an external and visible representation of the image because of their ability to help store and amplify cognition [94, 128, 271]. Suwa and Tversky discuss how these types of diagrams and sketches are helpful to designers for reducing working memory load because they capture elements of the image that must otherwise be kept in mind [271]. They also described how they can promote discovery of spatial relationships between objects in the scene. In general, these diagrams are helpful at focusing attention on specific aspects of the scene that otherwise might be easy to miss. In fact, looking specifically at directing attention through gaze, researchers have found that subtly directing the gaze of a novice to follow an expert radiologist's scanpath can significantly improve the novice's ability to identify abnormalities [24, 204, 264]. We wondered if we could use annotations to guide novice photographers to examine photos in ways that more closely matched expert photographers.

Part of what makes this focusing of attention helpful is that it makes the viewer more aware of specific regions of the image to consider. Famed tennis coach Timothy Gallwey pioneered a light-touch strategy of "awareness" prompts that cue learners' focus to a particular aspect of the process [102, 181]. Remarkably—and importantly for our approach—such awareness cueing can yield impressive learning gains, even outside the coach's domain. At one point, Gallwey (possessing almost no musical knowledge) coaches a tuba player to improved outcomes through careful elicitation of goals and awareness cueing. Following this principle, it is reasonable to believe that this direction of awareness can be effective without concrete guidance on how to address the concern.

Additionally, once the photographer has an idea of the image they'd like to achieve, such visualizations can assist in the step of interpreting the image. For instance, when I see a rule of thirds grid overlaid on my photo, it is much easier to see if objects align with the thirds lines than to try to imagine it on my own [223]. Expert photographers have trained their eye to be able to quickly consider a number of high level photographic concepts almost simultaneously, and to be able to seamlessly transition between them when considering how to frame an image [169]. A big part of this training is their understanding of these concepts. We aim to design visualizations that can assist novices in this process of training their eye.

1.3.2 Designing for Novices

In this work, we design for a specific type of user who in interested in developing this aristic eye. Thus, when we refer to novices, we use the definition from Davis and Moar [73]—defining novices as non-experts who have limited training, but have an interest in working towards developing expertise. By targeting novice photographers, we are referring to mini-C (personal) creativity, focusing on the process aspects of creativity and self-evaluation [158].

While we hope that our interfaces can also benefit experts, we primarily target novices because experts often already have a strong grasp of their own interpretation of these concepts and know how to actively apply them to their photography—they have already developed their own "artistic vision."

More broadly, a difference between experts and novices in any field is that experts have developed a deeper understanding of the problems at hand, enabling them to construct higher level groupings and abstract representations of the problem [48, 65]. Through training and experience, experts develop abstractions in their perception [9]. These abstractions can help guide their attention without the aid of visual annotations.

Novices, on the other hand, tend to only understand surface level features and have trouble understanding their relationships. We hope that through these visualized annotations, we can help novices to see their photos with expert-level abstractions, and ultimately help them to better understand how these surface features fit into the picture with respect to different photography concepts.

1.4 Thesis Statement and Contributions

Therefore in this dissertation, we explore the design of in-camera feedback interfaces that interactively display computationally-generated annotations as guidance to help train a novice user to develop their "artistic vision." Guidance should not be prescriptive, but should inspire the user to be aware of the artistic choices that they are making. Therefore, in this work we focus on the exploratory stages of photography rather than the refinement of a specific idea. Since we are interested in understanding how this might fit in a novice's creative process, we focus on personal creativity and satisfaction of quality in our user evaluations.

My thesis statement is:

> **Contextual capture-time camera interfaces that visualize key artistic concepts improve self-assessed creativity and photo quality.**

In my dissertation, I demonstrate this idea through three camera interfaces that each focus on a different photographic concept: portrait lighting, composition, and decluttering. Specifically in my work, I contribute new camera guidance interfaces implemented as prototype capture-time tools. For each tool, I design novice-interpretable visual overlays to represent the considerations for a key artistic concept, and realize them through real-time algorithms for generating these visualizations. The design of each interface visualization leverages an understanding of what an expert might be noticing in considering the targeted photographic concept, and aims to direct a novice user's awareness in a similar manner.

The specific contributions of the work in this dissertation include:

Concepts and Techniques

- A formative survey and interviews for understanding current photography practices and tools, as well as participants' photography learning experiences. [Chapter 3]

- Based on these formative studies, I developed a set of design goals for designing capture-time interfaces. [Section 3.4]

Artifacts

- A **portrait lighting** tool that guides users to select a lighting style and reorient their subject to best achieve the lighting style of their choice. This visualization aims to help users be more aware of the available lighting styles as well as the light in the scene. [Chapter 4]
- A **composition** guidance tool that highlights the lines in a composition grid that are most relevant to the current camera view, and a crowdsourced dataset of such composition annotations. This visualization aims to make users more aware of areas of visual attention that are influencing the overall photo composition. [Chapter 5]
- A **decluttering** tool that abstracts the camera view to outline edges around the subject(s) or the image borders. This visualization aims to increase the users' awareness of potential clutter that may draw attention away from the main story of the image. [Chapter 6]

Experimental Results

- A user evaluation comparing the **portrait lighting** tool to guidance showing a static set of portrait lighting styles or no guidance. The study shows that the tool allows users to capture more well-lit portraits, and reduces the cognitive load for novices to achieve their desired portrait lighting styles. [Section 4.5.3]
- A user evaluation comparing the **composition** guidance tool to a static composition grid. The study shows that the tool is indeed helpful for capturing better composed images, and enables participants to feel more confident in their ability to compose photos. Additionally, user evaluation that shows that using static composition guidance to compose photos similarly helps participants feel more confident and creative than using no guidance. [Section 5.5]
- A user evaluation comparing the **decluttering** tool to a currently employed method of viewing the camera as grayscale to emphasize contrast and clutter.

The study shows that the tool was more effective in making users feel confident in their abilities to address decluttering principles. [Section 6.5]

1.5 Dissertation Outline

In this dissertation, I explore the context around designing capture-time guidance. Specifically, I focus on using computational methods to enable new camera interactions and visualizations. I present the design and evaluation of three such camera interfaces.

Chapter 2 gives an overview of the related literature in tools for creativity support and the role that computation can play in providing creative guidance. It then dives into the range of work for improving photography ranging from automated assistance in the editing stages to a few existing interfaces for interactive guidance in the camera.

In Chapter 3, I describe our formative studies with participants of varying photography experience and summarize our findings. I then describe how the learnings from our formative studies informed our design goals for in-camera guidance interfaces.

Chapters 4-6 then go on to describe our three interfaces for portrait lighting, composition, and decluttering, respectively. For each tool, I present the process of designing the interface, the algorithm for generating the visual annotation, and the user evaluation. For each interface, I then close by describing how it is informed by our design goals (from Chapter 3). Through these chapters, I walk through how these interfaces might help a photographer in the process of capturing a portrait as shown in Figure 1.2:

- Chapter 4 (lighting): The photographer notices that the lighting is very dark on her subject's face. The interface shows a set of lighting styles that are attainable in the current environment. The photographer scrolls through the choices and picks one they like, and the tool assists the photographer in reorienting the subject to achieve the desired lighting style.
- Chapter 5 (composition): Next, the photographer composes the subject relative to the objects in the background. The image is currently centered as shown by the overlay. The photographer instead wants to follow the rule of thirds, and shifts the camera to place the subject at the top left thirds.

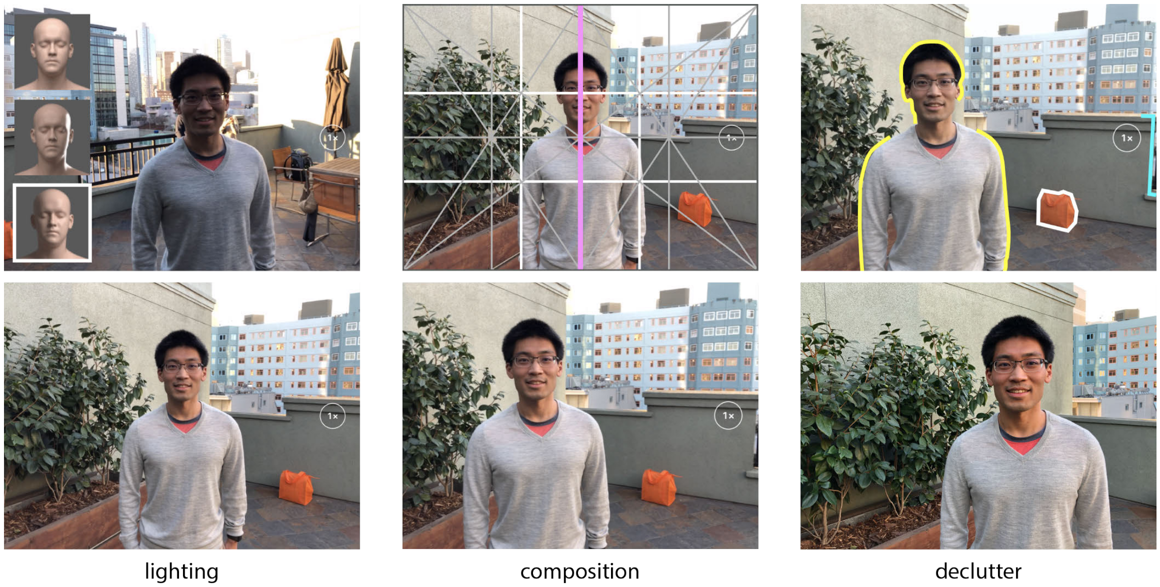


Figure 1.2: These photos illustrate a walkthrough of the sequence of steps a photographer might take to capture a portrait, and how our tools might assist at each stage. The top row shows the images at the beginning of each stage (**lighting**, **composition**, and **declutter**) with a mockup of our tools' interfaces overlaid. The bottom row shows the end state for each stage once the photographer has decided how to address the concept at hand.

- Chapter 6 (declutter): However, the photographer notices that the interface highlights unwanted clutter in the background—a sharp corner right along the right border, and a bright orange bag. They adjust the composition to place the subject at the top right thirds instead to remove the distracting clutter, and the photographer takes the final photo.

Finally, Chapters 7 and 8 close the dissertation with a discussion of future work in photography guidance and how to apply the learnings from this work more broadly in creativity support.

An important value throughout this work is that computational guidance should assist the user in being more aware of the artistic choices they are making, rather than impose specific artistic preferences. This dissertation begins to explore the potential of computational methods for providing this type of support, specifically in the domain of photography, with the hopes that these types of interfaces can enable and inspire more people to pursue interests in photography. I additionally hope that it can inspire further work in photography and other creative domains that similarly encourage creative learning through increased awareness, and developing “artistic vision.”

Chapter 2

Related Work

Researchers have long been interested in building creativity support tools that help users produce artistic work. Prior research has leveraged computational methods to build such tools for a variety of creative domains, including photography. Computation can assist users in creative work in many ways, especially in reducing tedium, helping to generate ideas, and bridging the gulf of execution. Depending on the user's needs, the tool may need to integrate into the user's artistic practice in different ways to assist in inspiring and producing novel content.

In this chapter, I group this work into two approaches through which computational methods can be used to assist users in their creative work: (1) helping with **execution** of their existing creative intentions, and (2) providing them with **guidance** towards a new idea. For each approach, I describe the existing work broadly across domains of authoring tools, and then specifically in image editing, and finally in the camera.

As described previously, a value in our work is to allow the user to make artistic choices. In doing so, we make the conscious decision to use computation to assist in the user's awareness and attention, but not to actively assist the user in determining the steps to achieve their final result. Here, I describe each section of the related work as a spectrum based on whether the creative ideas come primarily from the system or the user. I start with work that aims to automate more of the process of the artistic task, reducing the effort/work from the user, and make my way towards systems that aim to assist the user in generating ideas, but leave the work of actually generating

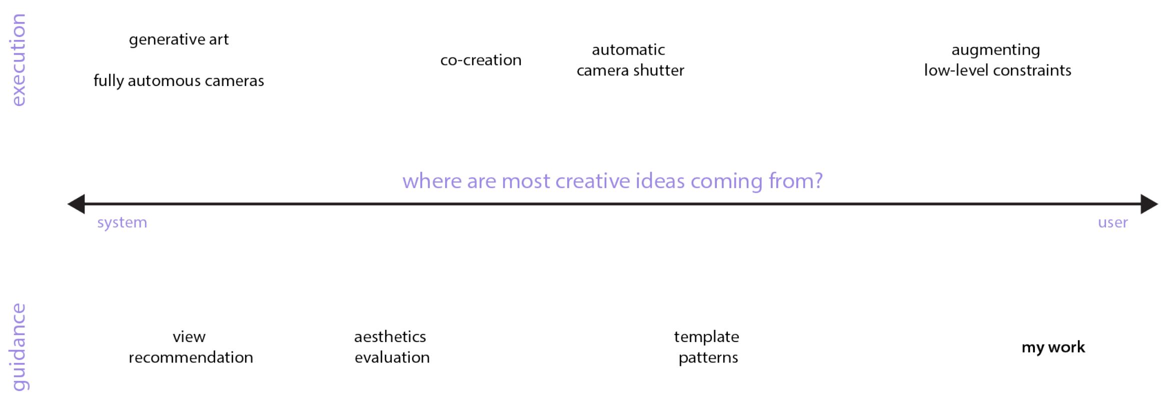


Figure 2.1: The creativity support tools described in this chapter belong to either the approach of assisting with artistic **execution** (top) or **guidance** (bottom).

and producing the ideas to the user.

Figure 2.1 visually displays some of the styles of work described in this section in the context of the spectrum for each approach. For **execution**, this ranges from automatically generating the work given a very high level user prompt to reducing the tedium of refining a piece to fit very specific user constraints. For **guidance**, this ranges from guiding the user to a specific idea to guiding the user to discover new ideas.

For **execution**, tools exist along the spectrum of giving the user varying levels of creative control. However, few tools fall towards the rightmost end of the **guidance** spectrum (see Figure 2.1). Our work is designed to fit in this space, with the goal of helping users to develop the "artistic vision" to interpret and evaluate their own work in a way that makes them explore potential ideas so as to be more intentional about their artistic choices. I describe how my work of designing capture-time guidance through visual annotations explores a missing piece in this space, and why that is important for the goal of helping novices train their "artistic vision."

2.1 Assisting in Artistic Execution

Computational tools can be helpful in assisting users in executing on their artistic intentions. Users' intentions can vary from very high level to very specific, low level details, depending on the level of involvement that they want in the creative process. The growth of algorithms that can help to improve or create art is enabling researchers to develop systems that allow users with a wide array of goals for creative work to produce their intended artistic output.

2.1.1 Authoring Tools

Researchers have extensively studied artists' workflows to understand what tools they can design to assist in specific tasks at different stages of the creative process. Users' creative goals vary drastically, and thus the expected role of computation does as well. Here, I describe several creativity support tools that explore the range of roles—from taking very high level user intentions and producing a complete artistic work, to collaboratively contributing to a piece, to executing on low level constraints posed by the user.

Generative Art

In particular, advances in AI are rapidly increasing the pool of algorithms that can be leveraged in these tools. On one end of the spectrum are generative tools that require minimal creative experience or effort to allow users to produce something creative. Some of this work takes very high level inputs. For instance, tools that can output a poem given a topic [108], options for a compound icon given a text description [317], a novel sequence of dance choreography [68], a coloring for a 2D pattern [189], or a song of a given style [120]. This style of work allows users to steer the output with some lower level inputs such as specific notes or rhythms. Other work in this area gives the user more control over the creative output by generating a range of solutions across the available parameters of the design space. Matejka et al. present a system that uses simulation to automatically generate many possible 3D structures that satisfy the user-provided structural constraints. It further facilitates user control over the final output by providing interactive support for the user to search over the space of possible results [203]. Thus, while the art is completely computer generated, it captures detailed, low level artistic intentions of the user.

Collaborative Co-Creation

A growing area of work in this space enables a dialogue between the user and the system to co-create new content. Merrell et al. present a furniture layout tool that suggests functional and visually pleasing layouts. Users can manipulate the layouts based on their personal preferences, and the system will iteratively update the suggestions to satisfy the constraints, while also considering interior design guidelines [205]. More recently, it has been increasingly possible for AI to engage in these types of more collaborative processes of co-creation. As a result, researchers are exploring different roles of AI in human-AI co-creation to understand how much initiative the system can take, and they are finding that different users have different desired interactions with their AI collaborators [71, 119, 227].

Augmenting Low-Level Constraints

Finally, for users who want to maintain full creative control, automation can still assist in the more tedious aspects of an artist's work. This also further allows them to focus their time more on the creative aspects of the task. Due to the tedious nature of video editing, many tools exist to help with different aspects of the editing process to enable users to focus on the creative storytelling. Researchers have designed systems to help with each stage of editing, from organizing footage to generate an initial rough cut [174, 282], to providing feedback on an edit [228], to fine tuning details in the final video [39]. Many of these tools leverage the fact that computers can process the mass amounts of data that is video, and provide the user with additional information and structure (such as transcripts or faces) that can reduce the time they need to spend repeatedly watching the video content.

In addition to the ability to process large amounts of data, there are other advantages of computation that make them well-suited to relieving some of the tedium in creative work. One line of work enables designers to use constraints to specify relationships in direct manipulation vector graphics tools and therefore can greatly reduce tedious manual manipulations such as maintaining alignment and spacing [38, 156, 157] (or even negative-space strokes for resizing logos [38]), propagating changes across related elements [146, 305], or duplicating equally spaced copies of an object along a path [146].

Similarly, a number of animation tools use simulation to realize the user's goals temporally across frames. Draco enables users to animate illustrations by specifying elements to animate and a path along which to animate [159]. This system enables artists that primarily have experience creating static visual art to be able to bring motion to their work. Cong et al. present an artist directed facial simulation system that adapts physically-based simulations to respect the artist's animation intentions [66]. These and many other works use the strengths of computation to reduce tedium in artistic work. Such tools generally assume the user wants some level of artistic control and has the skill to execute and communicate their intentions. However, they still enable users to extend their artistic repertoire. For example, many of these tools allow users who were only familiar with 2D/still art to produce moving animations [72, 159, 302], or someone with no experience with fabrics to design custom patterns for clothing [30].

Finally, another line of work focuses on involves interpreting the imperfect manual inputs of a user to predict (and ideally execute) the user's intentions. Tools like Pegasus and Fluid Sketch take the user's freehand stroke and "beautify" them using geometric constraints [16, 142]. Some commercial tools, like Adobe Illustrator, include features to allow users to apply constraints to their drawing path to produce cleanly executed strokes free of slight uncertainties in the artist's hand. Such tools enable users to refine the execution of their artistic intentions without being limited by their manual control and skill.

2.1.2 Image Editing

Specifically in the domain of image manipulation, rapid advances in computer vision and graphics techniques enable a wide range of image editing through relatively low effort interactions.

Automated Image Processing Algorithms

In fact, Isola et al. make it possible to generate a relatively photo-realistic image without even capturing new content. Once trained on a dataset of related existing photos, the user can generate new photos just through sketching/painting its structure [33, 145]. These algorithms can also be used to make a wide range of edits to aesthetically improve an image after it is captured. Examples include automatically cropping a photo to improve its composition [295], removing automatically identified distractors [100], segmenting out the foreground to change the background [251], or transferring the style, color, or lighting from another image [178, 252, 257]. However, much of the related work in this area focuses on automated workflows, which can somewhat detach users from the creative experience of taking a photo.

Interactive Image Editing

As with general authoring tools, researchers have also explored how computational tools can assist users in executing on the more tedious aspects of image manipulation. Agarwala et al. present an interactive system for creating photo montages [10]. Similar in interaction to Isola et al. [145], the user specifies the overall structure of what region is taken from which image, and the algorithm automatically determines how to merge these components in the most realistic manner. SelPh is another more interactive system that reduces tedium over time [168]. As the user uses the tool to make edits, the system learns the user's lower level photo editing preferences (e.g., brightness or contrast) to automatically edit similar photos. This system relies on a variety of image analysis algorithms to compute the distance between images to determine if it can predict the user's edits based on past selections.

The variety of post-processing methods available to users allows them to quickly explore a wide range of options for their image that they otherwise might not know how to produce. Nevertheless when focusing on image editing, the options are still limited by the already fixed content of the image (or other existing images). As more of this work in the graphics and vision community becomes interactive, I am excited about the increased opportunities to use and build on these methods to design interactive guidance interfaces for photography.

2.1.3 Camera

Within the camera, tools for artistic execution generally involve assistance in either triggering a camera shutter (or starting/stopping the recording for video) or positioning the camera (or executing a camera path), or both.

Fully Autonomous Cameras

On the one hand are systems where the camera, often mounted on and controlled by some form of robot, chooses the path and composition of the camera and also actuates the camera shutter. Examples include autonomous robot photographers that automatically capture well-composed images of an event or environment [50, 311], or drone cameras that capture cinematic compositions and calculate camera paths to transition between them [104, 152]. Other tools position cameras with more practical goals, such as framing people speaking in video conferences [291], or capturing the best viewpoints for computer vision tasks [242, 281]. Similarly, such tools also exist for placing cameras within virtual environments [17, 103, 125].

Automatic Camera Shutter

Other systems for handheld cameras put the control of camera placement into the hands of the users. SenseCam, for instance, is a camera worn around the users neck that automatically captures a digital record of the wearer's day to aid in memory recollection [130]. However, here the role of the user is solely to move the camera to different physical environments to document relevant memories rather than to frame interesting images. While the user isn't directly framing images, awareness of the camera can still influence the wearer's behavior, e.g., not staying still for too long to avoid boring sequences of photos, or making sure not to obstruct the camera view [191]. Most such tools do involve the user framing images based on their artistic goals. For example, a number of commercial cameras provide support for using smile detection to trigger the camera shutter [261]. In this case, the user composes the image and initiates the capture. The role of the camera is solely to reduce the uncertainties in the scene to hopefully better capture the user's artistic intentions.

In-Camera Editing

Computational photography techniques have also made it possible to develop a set of tools that preview the results of image processing algorithms in the camera [2]. In fact, many of these have become built-in features in camera phones. Some are previewed interactively as the user moves the camera in space such as basic color adjustment filters, portrait mode, or studio lighting [290]. Others, such as HDR or low-light, are processed once the photo is captured and previewed in the camera roll [124]. These features enable users to consider these post-processing actions directly during capture. Baek et al. further explored the WYSIWYG interaction of editing images directly in the viewfinder by extending the range of edits and allowing users to select local regions on which to apply edits [23]. Editing in the camera allows users to make more informed choices in the camera as they can see a result closer to their final output.

Camera Operation Assistance

Finally, another direction of work in this space is to enable users to capture shots that they otherwise would not be able to easily capture. For some users, the physical aspect of operating a camera, such as actuating the camera button and/or positioning the camera, are limiting factors to their photographic ability. Not being able to perform the physical actions required to capture a photo can greatly hurt a user's confidence in their photography abilities [37, 212]. Researchers are exploring how to assist users with visual or motor impairments in positioning the camera to frame the image they envision [115, 286, 287], and capturing the image without shake [212].

Another form of camera operation challenge is operating more involved cameras, e.g., drone cameras. Joubert et al. present a system where users interactively design a drone camera path in a virtual environment [153]. The tool communicates if the flight path is physically feasible for a drone to fly and outputs commands for a drone to follow a flight path autonomously. Such tools enable users to capture drone footage with minimal experience in manually flying a drone. Note that using this system, the design process is somewhat of a feedback loop—the user provides a potential flight path, gets feedback on path feasibility from the system, and this feedback causes the user to reconsider the flight path and potentially make some artistic judgements based on this information, and then gets more feedback on the updated path from the system. Our work similarly aims to provide feedback that inspires the user to explore their artistic options.

2.2 Providing Artistic Guidance

The previous section described tools that assisted users in executing their high and low-level artistic goals. However to do so, either the system or the user had to determine how to execute on these intentions. Here I describe systems that assist users in achieving these goals by guiding them towards the steps to achieve a specific result, or to explore the options and discovery of new ideas.

The practice of photography involves exploring a physical space through the lens of a camera. Thus in this chapter, I focus on artistic work that involves exploring a creative space, rather than recreating an existing artwork. For the latter space of work, there is a concept of "ground truth" and the process is about learning to accurately reproduce that truth. For instance, I do not discuss methods for learning a choreographed dance, but might discuss tools for creating dance choreography. I also will not be describing tools for drawing/painting a physically accurate or realistic scene, but might touch on tools for inspiring more abstract or stylized drawings.

2.2.1 Authoring Tools

Researchers have shown that guidance can be useful for training artistic practices where there is a measure of accuracy, or a distance from the "ground truth" (such as how correctly proportioned and realistic a sketch of a face is). For example, sketching is a widely explored domain for this type of guidance. These tools rely on having fixed procedures in order to be able to guide users through step-by-step instructions and provide feedback based on performance [78]. They can assist by automating steps of the procedure such as generating guiding construction lines [141] or visualizing correct shading regions [306].

However, for a lot of artistic work where the focus is generating creative new ideas, there are not nearly as many guidance tools. One possible reason is there is not really a correct answer or process to guide users towards. Some tools that do exist primarily focus on inspiring the user by showing examples [170, 175, 272], or suggesting changes or tutorials [98, 99]. Ngoon et al. provide guidance by adding structure to an otherwise open-ended creative task of giving creative feedback—good creative feedback should be specific, actionable, and justified—and users are made aware of the quality of their feedback based on these attributes and guided to improve their feedback [220]. Metaphoria generates possible suggestions for word connections to inspire ideas for extended metaphors. By helping users generate suggestions, the tool makes it easier for the user to quickly consider and explore different metaphorical connections [107]. In my work, I'm interested in trying to guide people in training the their eye to actively evaluate their artistic decisions.

2.2.2 Image Editing

The most prevalent form of guidance in commercial tools for editing photos, is overlaying a composition grid to assist in cropping or straightening the photo. Tools such as Adobe Photoshop give the user some additional creative control over this guidance by providing a few alternative composition grid options. Seeing these grid lines can make users more aware of alignment (and misalignment) in their image to help them improve. By making it easier for the user to evaluate their image within the context of these grids, the user can more quickly explore and assess various compositions to achieve a desired composition.

Such feature-rich image editing tools can be daunting for novices to approach—both due to not knowing how they want to edit their photo (and what is achievable in the software) as well as not knowing how to use the software. Fraser et al. design a panel with searchable action suggestions to make it easier to find relevant edit actions [98].

Seeing previews of these possible actions make exploring the breadth of editing options more approachable [275]. Berthouzoz et al. further enable automatic generation of photo manipulation tutorials that can be transferred to new images [113]. Such tools can help bridge the gap described earlier in Section 1.1.3 of transferring knowledge learned through one working example to a new context.

2.2.3 Camera

Additionally, a variety of guided photography interfaces do exist. Many of these use existing knowledge and data from photography, such as expert practices or previously captured photos, to provide guidance on how to capture a better photo or to present patterns to emulate. My research aims to further understand how such guidance can impact the quality of the photographs captured by users, as well as their ability to develop their own artistic preferences.

View Recommendations and Quality Assessment

A number of these photography guidance tools guide users by recommending a new viewpoint. These methods rely on a range of aesthetics evaluation algorithms to help make their suggestions [167, 239, 267]. Some such tools rely on crowdsourced preferences [186], datasets of publicly available images and social metadata [235–237], aesthetics guidelines in photography literature and practice [207, 215, 268], to suggest viewpoints that the system believes are good (e.g., popular, unique, or aesthetically pleasing). Many such tools guide the user to specific “better” options. They have employed a number of interaction techniques for communicating their guidance to the users ranging from navigational directions and nudges [20, 55, 186, 207], highlighting potential errors [55], suggesting crops or zooms [199, 268], or displaying preferred view proposals [199, 237]. Others more passively guide the user towards the system’s preferences by displaying an aesthetics score on the camera view [192, 199].

Most of these tools have an internal model of aesthetics that are mostly hidden from the user, and apply this model without providing much explanation. A few systems are beginning to address this by learning a user preference model [199, 268]. SmartEye additionally communicates how well a suggestion matches the user's preference through a confidence score [199]. However, by imposing a specific suggestion from the tool, these types of interactions tend to limit the amount of exploration the user does and claims some of the user's creative flexibility.

Template Patterns

Another direction that researchers have explored, in particular for video, is suggesting patterned templates for the user to capture to assist in overall narrative structure [22, 162]. Adams et al. observes that these pre-templated methods can present a barrier to exploration, especially as they become more specific [5, 269]. They instead ask the user for information about the scene while situated in the event that they are trying to capture, and use this context to provide shot suggestions. They additionally note that by the act of having to explicitly answer questions about the setting and contents of the image, the system has the potential to make the user more aware of environmental details that might otherwise have been overlooked.

Collaborative Guidance

This awareness that Adams et al. describes that comes from the user communicating ideas and intention to the tool, can be similar to collaboratively working with another human to take photos [150, 299]. In fact, we are inspired by the idea of having an expert looking over a novice's shoulder as they are trying to capture a photo. Rather than tell the novice how to move the camera, the expert points out interesting things in the image or in the environment, such as lights, shadows, colors, and textures, to make the novice aware of these contexts to consider while framing their photo.

Existing In-Camera Tools

Commercial cameras already provide a range of interactive guidance overlaid directly in the camera viewport. Some examples include the use of zebra stripes to highlight areas that are overexposed, highlighting sharp edges to signal regions that are in focus (focus peaking), or displaying horizontal and vertical leveling. However, the focus of much of this guidance is to help the user determine the appropriate camera settings. Given this information, the user can more quickly assess their performance across the numerous camera parameters that they can be adjusting in the moment. The light meter, for example, attempts to evaluate if the current image is properly exposed and displays this information to the user on a negative (underexposed) to positive (overexposed) scale, where being at 0 means the algorithm believes the image is properly exposed. The photographer uses this information to adjust the camera to achieve better overall lighting. However, also note that given this feedback, a photographer can intentionally over or underexpose the image based on their personal preferences, or unique circumstances with their current image. Nonetheless, providing this representation of light to the user allows them to more easily consider the overall lighting, while still focusing on other aspects of the image. We take inspiration from these existing tools to design our guidance interfaces.

2.2.4 Summary

Looking back at Figure 2.1, notice how systems towards the left enable users to create artistic work with minimal expression of their personal artistic preferences. With a simple high level idea, the system supports the user in determining the steps to produce output that achieves this high level goal. As we move towards the right, users are required to put more effort into making artistic choices throughout the creative process. In order to help train "artistic vision," users not only need to be making these decisions, they need to be actively making these decisions. In this way, the process of understanding their choices in the context of various artistic principles can help them develop their artistic preferences.

Chapter 3

Formative Studies

> *"I've found that shooting lots of pictures, experimenting, and reviewing results has been the best method [for learning photography]."*
>
> – *survey participant*

To inform the design of our camera interfaces, I conducted a survey to understand people's current photographic process and their learning experiences, including what was effective about the different learning methods and whether any of it could be helpful at capture-time. We additionally did nine in-depth interviews with experienced photographers about their current practices and tools. I supplemented these with my own experience in photography and giving feedback to many students about their still photography and videography camera work.

3.1 Photography Practice Survey

Our first question involved understanding if people wanted in-camera guidance at all, and if so, what type of feedback would be helpful or not to their current photography practice.

Much of the material of this chapter is as it appears in *"Adaptive Photographic Composition Guidance."* by Jane L. E, Ohad Fried, Cynthia Liu, Jianming Zhang, Radomír Měch, Jose Echevarria, Pat Hanrahan & James A. Landay, published in the Proceedings of CHI '20.

To answer this question, we designed a survey asking about people's existing photography practice and past learning experiences. We surveyed adults with an interest in photography through a number of Adobe mailing lists for related interest groups. The survey was voluntary and no compensation was given. We received a total of 127 responses from participants (74 male, 51 female, 1 non-binary, 1 preferred not to say), 19 to 63 years old ($\mu = 34$, $\sigma = 11$), with a range of photography experience.[1]

3.1.1 Photography Practice Survey Results

Many of our participants had had some sort of photography learning experience (only 23 had never used any resources such as classes, videos, or books to learn photography)—several of the ones who had taken in-person classes (22 of 45) mentioned the effectiveness of the *"direct, immediate feedback in the moment"* (P7), that those provided. Specifically, many mention the benefits of the feedback being *"individualized"* (P111) and *"hands-on"* (P73) for making immediate adjustments and correcting mistakes. Without this guidance, it can be difficult to *"transfer knowledge to other conditions"* (P10). Thus we saw benefit in pursuing the direction of trying to provide contextually-based in-camera guidance, so users can make better creative decisions during capture.

When asked why they would or would not use capture-time guidance in their camera, many mentioned concern with an app being too *"distracting"* (P51), *"disruptive"* (P53), or *"intrusive"* (P93). This was due to either concern about missing a moment (P12), intruding on others' time when in a group (P14), or feeling less artistic freedom (P86). Guidance should thus prioritize being minimally distracting and more about providing suggestions than insisting on specific artistic choices.

Our survey respondents supported the idea of providing suggestions for larger changes in the camera. We coded their responses to an open-ended question asking what they want to improve in their photography for mentions of popular photography concepts. Many explicitly mentioned composition (18); the other most popular concepts included camera settings (24) and lighting (22). Additionally, some responses

[1] Materials for the study can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

mention higher level goals such as being more creative or having more "professional"-looking photos (11). On a Likert scale from 1 (strongly disagree) to 5 (strongly agree), participants rated their editing practices as significantly more around small framing adjustments (Mdn = 4, IQR = 3-5) than substantial cropping to change the image's composition (Mdn = 3, IQR = 2-4) [Wilcoxon signed-rank test [304] $V = 2217$, $p < .001$], which further supports the idea of providing composition guidance at capture time to reduce the need for more drastic changes while editing later.

Overall, people have a preference for getting feedback on the framing of their current photo (Mdn = 4, IQR = 3-4) rather than thinking of possible photos to take (Mdn = 3, IQR = 2-4) [Wilcoxon signed-rank test $V = 2183$, $p < .001$]. Thus, we should additionally focus our guidance on helping people refine their current image's composition. Of our survey respondents, 89% were willing to spend up to 5 seconds on capture-time guidance to get a high-quality result.

3.2 Experienced Photographer Interviews

To go into more depth on what types of guidance could be helpful in photography practices, we interviewed nine experienced photographers about their current photography practices and tools. All had formal training in photography. Five additionally had teaching experience, with two actively teaching as their profession. Participants were compensated $15 for their time.[2] Interviews were structured around the following questions:

- Describe your typical photography process(es).
- What photography tools do you use and what guidance does it provide?
- Would in-camera guidance be helpful for you?

3.2.1 Experienced Photographer Interview Results

Here I report the results of the interviews in two parts: their current process and tools they actively use, followed by reports of and reactions to guidance.

[2] Materials for the study can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

Process & Tools

In describing their own photography processes, these photographers recounted both a process of searching for good compositions, as well as pre-composing and waiting for a shot—the latter was mentioned as particularly important in street photography. One described this waiting as a *"necessary tension"* in street photography (P5).

Six of the interviewees expressed that they already consistently use overlays of sorts such as focus dots/focus peaking, light meters, levels, composition grids (primarily the rule of thirds), or zebra stripes (highlighting of overexposed regions). For the purposes of planning, they even mentioned using a number of additional tools such as apps for determining sun location (and when golden hour will be), or scouting Google Maps for a sense of the area. This suggests that photographers are okay with some amount of their camera view being obstructed when that information is useful to their overall ability to take better photographs. Additionally they felt that these did not overly limit their overall creative flexibility, and even want additional information to help inform their artistic choices.

The light meter, for example, attempts to evaluate if the current image is properly exposed. The photographer uses this information to adjust the camera to achieve better overall lighting. However, also note that a photographer can choose to intentionally over or underexpose the image based on this feedback. P6 takes post-processing into consideration while shooting, and intentionally underexposes photos because *"Nikon is good at capturing details in the dark areas."* In fact, photographers will often quickly take a set of photos with different light meter readings. They are aware of limitations of the technology and consider that in their process, *"light meter is a single number for the entire image, what is it metering on? I'm not too worried about zero-ing it out. I'll usually just take shot, look at it, and adjust accordingly"* (P0). They instead use their intuition and then quickly look back at photos they've captured to make adjustments in the moment. This can be thought of as a quick succession of iterating to generate a number of prototypes. The photographer can then quickly review these "prototypes" and then further iterate to achieve the image they want.

This process of reviewing and iterating in the field is also common process that all participants mentioned. When possible, they'll also spend additional time to more carefully review their photos to evaluate and refine to avoid noticing "mistakes" after. P4 focuses on still life photography, so she spends a lot of time carefully composing the physical objects in the scene, *"trying trying"* to create an image that she likes. She mentioned even analyzing the histogram between photos (and to supplement the light meter reading), but still missing things she wished she'd noticed at capture time. She describes *"shoot[ing] an image with camera, and then after having spent some time without seeing the image, when seeing on the computer again, I start to analyze and will sometimes still notice something I don't like about the image."* (P4). Sometimes they are exposure mistakes, sometimes she'll use Photoshop to move the objects after the fact, but she said instead she *"should spend the time in situ when taking the photos."*

However in other situations where the photographer isn't *"directing the scene,"* P4 describes that in these moments there is *"no time to mediate, [you] need to shoot shoot shoot."* In photography, a moment is often fleeting. However, as suggested by her quote, even in these moments, it is important to try to capture multiple shots of the image. One photographer emphasizes the importance to *"shoot whatever is happening immediately"* (P1) before making too many adjustments or framing to prepare for it to possibly happen again. While artistic choices are important, some of these can be decided in post. Thus the priority is to catch the moment, but once that is guaranteed, the next priority is to catch it from multiple perspectives. The same photographer describes that for *"static images, more time can be spent exploring around the subject, just looking through the viewfinder for what compositions seem interesting. I'll take multiple so that I can decide after the fact"* (P1).

In fact, in discussing how he teaches his college photography courses, P8 says he tells students to *"always take more than one picture, from different angles—sometime 10 pictures, or maybe a whole roll of one particular scene, to make sure they consider different ways of framing their subject matter"*. He also emphasizes that he *"tries to understand students' interests. [You] don't want to project your own interests on them"*. Therefore he just tries to encourage students' creativity through frameworks, *"limiting them to a single space and see[ing] how they can do interesting documentation within those parameters"* (P8). These methods strongly parallel some tools in design thinking to encourage creativity in brainstorming and prototyping [233].

Both interviewees who formally teach photography mentioned teaching their students the general principles, but encouraging them to employ them in their own ways. Through their own experiences photographing, each of these photographers have established some of their own approaches to being unblocked when capturing a scene. We describe a few here:

- P4 mentioned that *"with more and more experience, [she] can reduce the time needed for composing"*. She describes developing *"intuition for how objects play together"* and having learned from her mistakes and seeing many of her own images.
- P2 describes his two styles corresponding to either using an ultra wide or telephoto lens, with clear patterns to focus on for each. For ultra wide, he looks for diagonals and vanishing points and might tilt the camera to create the image he wants. For telephoto on the other hand, he won't tilt the camera and focuses more on framing (e.g., rule of thirds vs centering, the relationship between the foreground and the background). He wants a *"clearly articulated palette"* and *"removes any unnecessary complexity that detracts from the immersiveness."* He picks lenses depending on the subject and meaning of the image.
- P3 captures a lot of sport photography. While there is little time to think, he does experiment with different angles. Specifically he describes the balance of capturing at eye level as being more intimate (but also having more distractions), whereas capturing from above increases chances at capturing the moment.
- P6 describes focusing on light for his portraiture: *"The first thing to look for is light, if light is interesting and has some sort of mood associated with it, I base everything else around that mood."* He quantifies his process as *"20% identifying light, 40% trying things until something works, and finally 40% perfecting the particular feel, composition, and pose."*

We take inspiration from these existing methods and tools to design guidance interfaces that can similarly encourage this rapid prototyping/iteration cycle.

Guidance

When asked what guidance might be helpful for them, some (5) expressed interest themselves in something that might give them new perspectives, like an *"experienced photographer on my shoulder saying try this, try that"* (P7). In general, these experienced photographers were very open to having any feedback that might help them try out different ideas in taking a photo in order to have more choices to pick from when going back to edit in the future.

Some suggested that they are able to use guidance as needed while maintaining creative freedom to break the rules and follow their intuition. For example, explicitly disregarding the composition grid: *"[I] might have a rule-of-thirds overlay, but [I] don't follow it super closely. I gravitate towards the bottom two eyes in the rule-of-thirds..."* (P1).

However, others mentioned conforming to a certain style, whether it be due to the prominence of the rule of thirds overlays, realizing an unexpected theme across many photos, or just having a *"tendency to view the world in a specific way, but someone else might be different, always open to try a different type of shot or idea"* (P6). P0 mentioned that *"rule-of-thirds becomes default that I'm conforming to because that's what's being overlaid"* and therefore appreciating the potential to have guidance that is more dynamic, e.g., even providing random composition guidelines to swipe through and try.

Participants proposed a number of ideas in which the tool could provide suggestions for them to consider. P4 described a *"semi-adapting suggester of shots. It could size up the emotion/affect of the current image, and suggest ways to frame it."* However, he emphasized that he'd want a *"more subtle, friendly tool."* In particular, it could provide some *"sensible"* suggestions, but he wants to be able to use the feedback to learn more about his own preferences. They also suggested concrete ways in which guidance could assist in helping them execute on specific ideas: e.g., identifying vanishing points and diagonals (P4), detecting visual flow (P4), assisting in centering (P6), identifying when objects are almost aligned (P7), proposing pose adjustments (P7), or pointing out imbalance in the photo (P7).

From our interviews, we learned that even experienced photographers could benefit from feedback that encourages them to try new ideas. In certain scenarios, they need to capture a shot immediately and thus rely on their instincts to quickly frame the photos, while in other scenarios they are willing to spend more time and devote both screen real-estate and their attention to a tool that helps them achieve a higher quality image.

3.3 Evaluating Novice Camera Work

As we saw in our interviews with experienced photographers, many noted the importance of lighting and composition to their photographic processes. These topics also came up in our survey as something people wanted to improve on. As a final part of our formative studies, I was interested in observing what mistakes novices actually make to understand what topics might be good targets for my guidance interfaces.

One assignment in our introductory Human-Computer Interaction course (CS147) at Stanford is to produce a concept video. There is no visual arts prerequisite for such a class, so for many students this is their first time putting together such a video, from scripting, to capturing, and finally editing. The professor presents a few guidelines on how to make a good compelling video, touching both on the storytelling and the visual appeal aspects, and students are provided with examples of past videos.

I thought this was a good sample for understanding the mistakes novices commonly make when they are trying to produce "good quality" content. While the output was video and not still photography, many of the visual and storytelling concepts are still relevant and both can benefit from feedback in the camera. I watched each video from the course (46 videos in total) and annotated them with feedback that I would provide for improving the video quality. I then grouped each piece of feedback into categories to find common themes.

For context, I took three photography courses during college and have continued to pursue my photography hobby informally. Additionally, I previously took and subsequently was Head Teaching Assistant of CS147 and therefore have prior experience both working on and grading this assignment.

Three of the most common themes that surfaced in this feedback were suggestions to consider lighting, composition, and removing clutter from the scene. Therefore, these became the photography concepts targeted by the camera interfaces presented in this dissertation.

3.4 In-Camera Guidance Design Goals

Based on our formative survey and interviews, we came up with three design goals for our photography guidance:

- **Context-Aware.** Guidance should adapt to the current scene and appear overlaid on the viewfinder.
- **Encourage Exploration.** It should help photographers discover new ideas while executing existing intentions.
- **Maintain Flexibility.** It should not restrict photographers from pursuing other creative choices.

Context-Aware

It can be difficult to apply an abstract photographic concept. Experts have experience doing so in many contexts and have developed patterns and established styles that they actively seek (methods/approaches that they use). Providing concrete suggestions and feedback in the context of the current image can make applying these abstract concepts approachable for non-experts. This includes both understanding how to apply a concept more broadly as well as refining the execution of an idea.

Many survey participants expressed that they appreciated the individualized and immediate feedback of in-person classes or even going on a photo walk with more experienced friends. This enables reviewing and adjusting while still in the context of the current photo.

Encourage Exploration

It can be hard to come up with an initial idea, just as it can be easy to fixate on perfecting a specific photo idea. Encouraging exploration can help with both. However, it can be difficult to know how to explore. Experts know to take many different shots of any given subject and have a set of "knobs" and "dials" in mind to manipulate to generate drastically different ideas, and yet still expressed interest in guidance that could help them generate more ideas to explore. Making the design space easier to explore and making different options apparent can make it easier to come up with new ideas to try—especially for non-experts.

Both survey participants and experts were interested in new ideas or perspectives, in particular suggestions for bigger changes, as they preferred to only make smaller framing adjustments while editing.

Maintain Flexibility

Our goal is to assist the user in their creative process. Thus, the interface should allow users to have creative agency. It should help users better achieve their own artistic intentions, and not distract them by providing restrictive guidance.

Both survey participants and experts also expressed concern that the interface might be too distracting and disruptive. Survey participants were concerned that it would slow down their photographic process. This was especially important in situations with time pressure, such as when capturing a fleeting moment or when traveling with a group. Experts similarly worried about flexibility, but put more emphasis on the need to maintain creative freedom.

Chapter 4

Portrait Lighting

For portrait lighting, the important context is the lighting in the environment. To encourage exploration, we want to make sure users are aware of lighting style options in the context of their current environment, and that they have flexibility to choose between these different options. Looking at this segment of our walkthrough in Figure 4.1, after being presenting with a number of lighting style options, the photographer moves away from an unintentionally backlit image to achieve a more desirable lighting style.

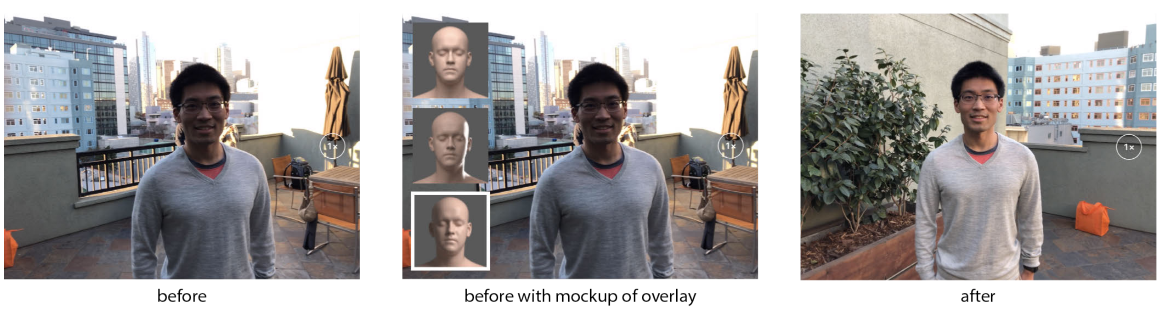


Figure 4.1: Here we see photos from our walkthrough (Figure 1.2) from **before** and **after** the photographer considered lighting. In the middle we show a **mockup** of what the photographer might see upon launching our tool. With our lighting tool, a novice can choose amongst a number of lighting styles, and reorient to achieve their desired style.

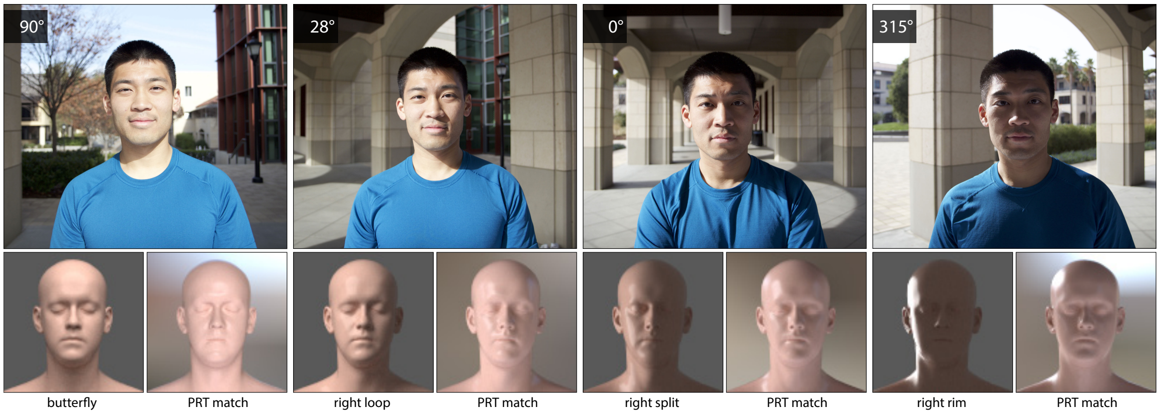


Figure 4.2: In a fixed lighting environment like this *arch walkway*, photographers can produce many different lighting styles (e.g., **butterfly**, **right loop**, **right split**, and **right rim**) just by rotating the subject in place without changing their location. Given an HDR environment map from a 360 camera at some initial orientation (Figure 4.3) and a target lighting style (bottom left), our tool automatically identifies the optimal angle for reorienting the subject to match the desired lighting—e.g., 90° for butterfly lighting. We use a precomputed radiance transfer-based method on a generic head, skin, and camera model for efficiently optimizing lighting orientation and for visualizing the best orientation **match** (bottom right).

4.1 Introduction

> *"Posing, location, rapport, camera angle... are all important. However, that said, the lighting matters even more. We can do everything else beautifully, but if our lighting is bad, our portrait will be bad. It is that simple."*
>
> – *Light Science & Magic* [138]

In portrait photography, lighting is one of the most important elements for establishing the overall look and mood of the image [42, 138]. Professional portrait photographers working in a studio typically place one or more lights around the subject to carefully control the distribution of bright and dark regions on the subject's face. Outside

Much of the material of this chapter is as it appears in *"Optimizing Portrait Lighting at Capture-Time Using a 360 Camera as a Light Probe."* by Jane L. E, Ohad Fried, and Maneesh Agrawala, published in the Proceedings of UIST '19. For more information, see the project page: `http://graphics.stanford.edu/projects/portraitlighting/`.

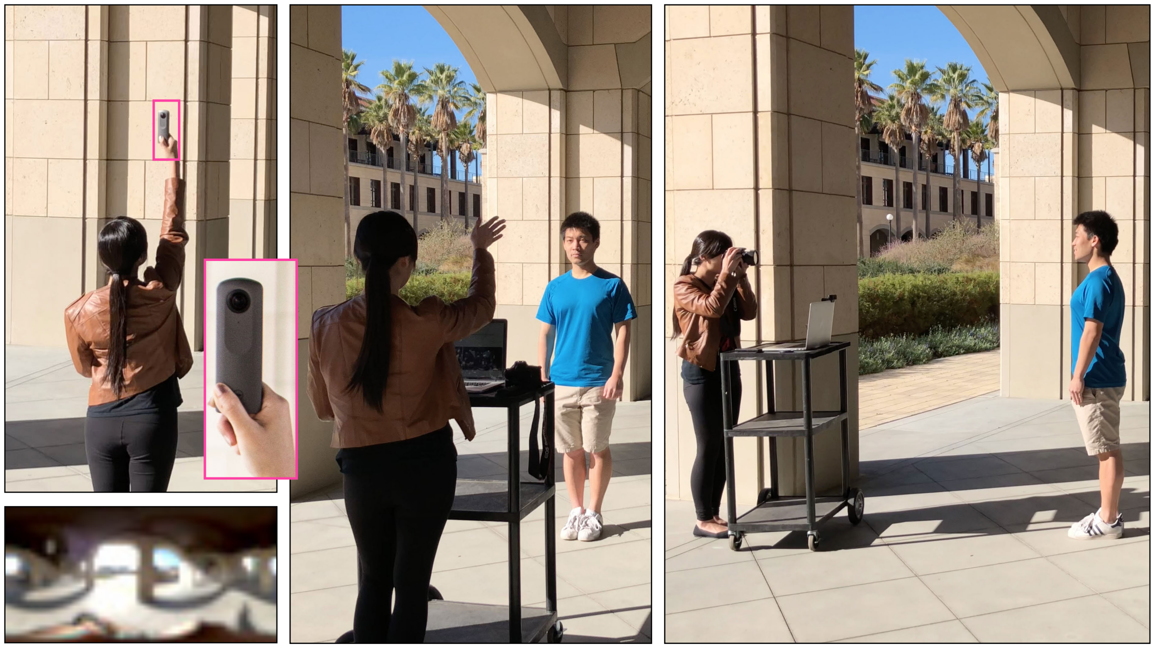


Figure 4.3: Photographers first capture an HDR environment map at the expected location of the subject's head using a 360 camera (left). The subject then moves to the location, and our tool suggests how far to rotate the subject to achieve different lighting styles (middle). After reorientation, the photographer shoots the final portrait using a primary portrait camera (right). Note that our approach also allows for selfies, where the subject is the photographer.

the studio, where they cannot control light placement, professional photographers instead focus on orienting the subject with respect to the lights in the environment to adjust the distribution of light falling on the face. While the fixed light placement makes it impossible to achieve all the different looks that are possible in a studio, good photographers can often produce several distinctly different appearances just by reorienting the subject (Figure 4.2). Yet, most casual photographers primarily focus on maintaining even brightness on the face [186] (or forget to adjust for lighting), rarely considering all the variations in facial appearance that are possible via reorientation.

In this chapter, we present a capture-time tool that suggests how portrait photographers should orient their subject within an environment to best achieve a user-specified target facial appearance. Photographers start by using an off-the-shelf 360 camera as a light probe to capture an HDR environment map of the light in a scene (Figure 4.3).

They then select a target facial appearance from a pre-designed gallery of common studio lighting styles that the tool determines are most feasible in the current environment. Photographers can optionally paint weights to adjust the distribution of bright and dark regions in the target appearance. Our tool then tells the photographer how far the subject should rotate (clockwise or counterclockwise) to produce the best match to the target appearance and provides real-time feedback showing how close the current camera view is to the view at the target orientation. If the photographer has a secondary source of light (e.g., a phone screen or a flash), our tool can also suggest how to orient it about the subject's face to further improve the match to the target lighting. After reorientation, the photographer can shoot the portrait using any available primary camera (e.g., a phone or DSLR).

Our prototype implementation runs on a laptop, uses a 360 camera to capture the environment, and a webcamera to determine the current view to aid in reorientation. But as 360 cameras become more widespread, we envision that they will become a standard tool for capturing environment maps and that all of our prototype hardware would be built into the primary portrait camera. Already it is possible to buy a 360 camera that attaches to a cell phone [143]. Thus, we have designed our implementation to be computationally lightweight enough to easily run on a phone or a DSLR, while remaining performant enough to be used at capture-time.

The key idea of our approach is to use an efficient precomputed radiance transfer (PRT) method [44, 218, 258] with a generic head, skin, and camera model to compute the appearance of the face under different orientations of the environment. We show how to formulate a target facial appearance as a weighting function in the image domain, and pre-integrate it against a light transport matrix for the generic head to significantly accelerate the search for the orientation of the lighting environment that best achieves the target appearance. We also introduce a method for optimally placing a secondary light source in the environment to best achieve a target facial appearance. Our prototype implementation running in Matlab takes 0.19 milliseconds (5263 fps) to compute the optimal orientation.

In summary, in this chapter we present:

- an **algorithm** for computing the optimal orientation for achieving a given lighting style in the current environment,
- an **interactive capture-time tool** that previews a gallery of portrait lighting styles at their optimal orientations and provides reorientation guidance to help the user achieve their selected lighting style, and
- a **user evaluation** that shows that the tool is very useful and reduces the mental effort required to produce well-lit portraits as compared to using a static lighting style gallery or no guidance.

We additionally contribute:

- an **algorithm** that guides users in positioning a secondary light source to capture a desired lighting style, and
- an **interface** that allows users to customize lighting style designs through painting regions of light and dark.

4.2 Related Work

For an in-depth discussion of related work on capture-time guidance, see Chapter 2. Here we focus on work specific to portrait lighting.

Human faces are one of the most popular subjects in photography. As a result, researchers have developed a number of techniques for enhancing portraits via perspective correction [101], transferring makeup [116, 279], improving attractiveness [182], transfiguring appearance based on Internet photos [160], and bringing still portraits to life [18]. Lighting design is another well studied problem, and here we focus on the subset of these methods that are most related to our work on portrait lighting.

4.2.1 Computational Image Relighting

Debevec et al. [76] were the first to introduce the approach of capturing multiple images of a face (or scene) from the same viewpoint, but under different lighting conditions, and then letting users composite these basis images to produce the image under

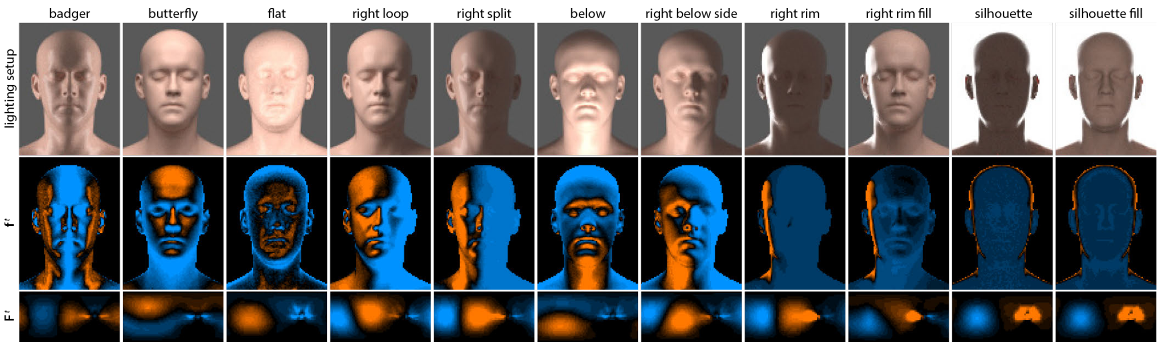


Figure 4.4: Our tool includes a set of 16 **lighting style setups** (top), selected from commonly used studio lighting styles to represent the diversity of placement of the main light source(s). Only 11 styles are shown here as each named style that starts with the word "right" also has a "left" version in our tool. We render these in PBRT [232] using the generic head and skin model, manually placing area light sources around the head based on studio lighting setups shown in portrait photography books [42, 138]. We convert a lighting style into a target facial appearance weighting function $\mathbf{f}^t$ by rescaling it to lie between $[0, 1]$ and then shifting it to put the mean brightness of the facial pixels at 0 (middle). We then pre-integrate the weighting function against the light transport matrix $\mathbf{T}$ to form the pre-integrated target in the lighting domain $\mathbf{F}^t$. Orange represents positive values where the image should be brighter and blue represents negative values where the image should be darker.

novel illuminations. Extensions to this approach let users specify higher level design objectives (e.g., emphasize contour, remove shadows), and then automatically find the set of input images that should be composited to achieve the desired result [10,12,46,47]. However, all of these methods require tens or hundreds of input images under varying lighting conditions and some require specialized lighting hardware, which can make it difficult for subjects to remain still throughout the capture process.

Others have focused on computational relighting using a smaller number of input images. Quotient image methods [190,202,229,241,300] require two images of the same face under different lighting conditions A and B and compute the ratio of the pixel value. Then, given an image of a new face under lighting condition A, these methods can generate the new face under lighting condition B. Portrait style transfer [252] focuses on transferring image statistics from a user-chosen style exemplar to an input portrait. While this method does not focus on relighting, the resulting changes to the image usually affect local contrast and can appear to change the overall illumination falling on the face.

While these computational relighting methods are powerful, they are designed to post-process input images after capture, and all require pixel-level alignment between the faces in each input image to ensure artifact-free results. Moreover, some of them can generate unrealistic lighting as they may fuse together pixels or image statistics from portions of different images. In contrast, we focus on directing the photographer at capture-time to shoot the single photograph that best achieves the desired lighting. Therefore, our approach cannot suffer from pixel-level artifacts or unrealistic lighting.

4.2.2 Automated Lighting Design for Synthetic Scenes

Lighting plays a central role in the perception of scenes. Based on this fact, researchers have developed lighting optimization techniques designed to place highlights and shadows [161, 230], to enhance shape [114, 176, 245, 250, 288], and to improve materials perception [44] in synthetic scenes. Our work is inspired by these methods, but focuses on optimizing the lighting for depicting faces in real world environments.

4.2.3 Capture-time Lighting Optimization

A few research groups have developed capture-time lighting optimization techniques that focus on adding light to a scene at capture-time. Adelsberger et al. [8] develop custom flash hardware that uses a depth image to spatially adapt the illumination so that the scene is evenly lit despite variations in depth. Srikanth et al. [265] automatically position rim lights around the subject using a robotic light-carrying drone to create dramatic contour lighting. Murmann et al. [214] develop custom bounce flash hardware that automatically reorients the flash to bounce off nearby surfaces to ensure that the subject's face is well lit. While these projects are similar to our work, they require customized hardware (e.g., depth sensors, drones, reorientation motors) making them inaccessible to photographers today. In contrast, our approach focuses on using off-the-shelf hardware so that anyone with a 360 camera and primary portrait camera could use it today. Li and Vogel [186] present a smartphone application that can guide users towards an evenly lit portrait at a fixed location. Their work focuses on scenes with one main directional light source whereas we account for all lighting in the scene.

4.3 Portrait Lighting Optimization

Our lighting optimization tool takes two inputs: (1) an HDR environment map, representing the illumination falling on the subject's face, and (2) a user-specified target facial appearance that represents the desired distribution of bright and dark regions on the subject's face. The tool then suggests how to reorient the subject with respect to the environment to best produce the target appearance.

Portrait photography books [42, 136, 138, 194, 211] suggest that because skin is a diffuse reflector, photographers should focus on the broad, low-frequency distributions of bright and dark regions on the face rather than on high-frequency highlights and details. They describe studio lighting setups (placement and orientation) that produce different looks by changing the bright and dark facial regions (Figure 4.4 top row). Outside the studio, they suggest orienting the subject with respect to the lights in the environment to similarly control facial lighting. In short, the orientation of the face with respect to the environment has a much greater impact on facial appearance due to lighting, than either the individual differences in facial features (e.g., geometry or skin tone), or perspective effects from the camera, and a low-resolution environment map is enough to capture the light falling on the subject.

These guidelines allow us to use three key approximations in rendering a suitable proxy for evaluating the lighting on the face: we use (1) a generic head and skin model, (2) a camera with a default field of view placed at a fixed distance from the head, and render it with (3) a significantly downsampled version of our input HDR environment map. Because the scene and camera models are fixed, we can adapt Bousseau et al.'s [44] pre-computed radiance transfer approach to our problem and efficiently find the orientation of the face that best matches the target lighting. In the next four subsections, we first summarize how we adapt Bousseau et al.'s lighting optimization approach (Section 4.3.1), then describe how the target facial appearances are specified for our problem (Section 4.3.2), next show how our approach can be extended to place a secondary light source (Section 4.3.3), and finally provide implementation details (Section 4.3.4).

4.3.1 Lighting Optimization with Pre-Integration

Bousseau et al.'s [44] lighting optimization approach builds on precomputed radiance transfer techniques [218, 258]. In matrix form, the image $\mathbf{B}$ of the scene lit by an environment map $\mathbf{L}$ is given by

$$\mathbf{B} = \mathbf{T}\mathbf{L}. \tag{4.1}$$

Both $\mathbf{B}$ and $\mathbf{L}$ are vectors, of the image and environment respectively, while $\mathbf{T}$ is the light transport matrix that represents how light from the environment map is transported through the scene to form the image. The columns of $\mathbf{T}$ are each a rendered image of the scene as lit by a single pixel of the environment map.

Suppose $f(\mathbf{x})$ is a weighting function that specifies the target facial appearance at each image pixel $\mathbf{x}$, using positive values where the face should be brighter and negative values where the face should be darker (Figure 4.4). Then we can define an image quality metric as the integral over all the pixels in the image of the pixel-wise product of the weight function and the image. In matrix form, the image quality metric C is

$$C = \mathbf{f}^t\mathbf{B} = \mathbf{f}^t\mathbf{T}\mathbf{L}. \tag{4.2}$$

Once the target weighting function $\mathbf{f}$ is fixed, we can pre-integrate it over the image domain

$$\mathbf{F}^t = \mathbf{f}^t\mathbf{T}, \tag{4.3}$$

and the image quality metric becomes

$$C = \mathbf{F}^t\mathbf{L}. \tag{4.4}$$

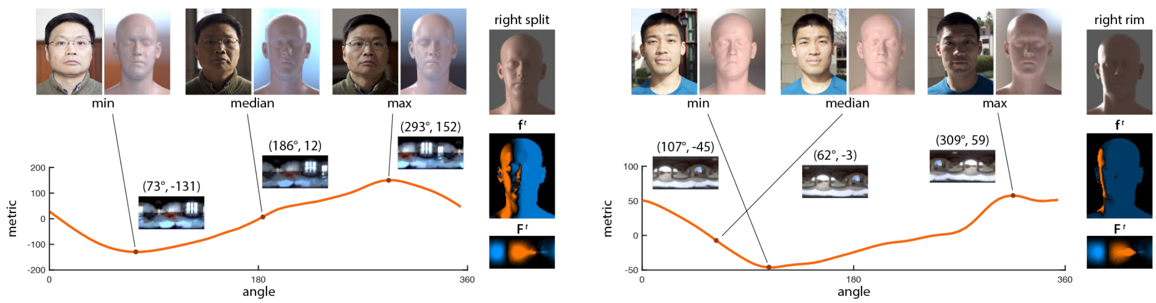


Figure 4.5: The image quality **metric** varies as we rotate the environment. **Maximizing** the metric produces a good match (both photo and PRT render) to the target facial appearance and the bright/dark areas in the environment map match those in the pre-integrated target weighting function $\mathbf{F}^t$. At the **median** and **minimum** metric values the matches are not as good.

The pre-integration allows us to efficiently compute the image quality C under any lighting $\mathbf{L}$ as a dot product of two vectors that are the size of the environment map.

In order to compute the optimal orientation for the lighting $\mathbf{L}$, we find the rotation $\mathbf{R}$ that maximizes

$$C(\mathbf{R}) \quad = \quad \mathbf{F}^t \, \mathbf{R}(\mathbf{L}). \tag{4.5}$$

In practice, since most subjects are standing or sitting up vertically with respect to the environment, we limit the search to a set of 1D rotation angles about the vertical axis (azimuth).

4.3.2 Specifying Target Facial Appearance

The target facial appearance weighting function $f(\mathbf{x})$ specifies which pixels in the final image should be brighter (positive values) and which should be darker (negative values). A weight value of $f(\mathbf{x}) = 0$ implies that pixel $\mathbf{x}$ should be ignored when evaluating the image quality metric C.

Our lighting optimization tool provides a built-in set of such weighting functions that are based on common studio lighting styles (Figure 4.4). Suppose $\mathcal{T}$ is a rendered image of our generic head model in one of these lighting styles. We mask out the pixels falling outside of the rendered face (i.e., background pixels) as they are irrelevant to the target facial appearance. We then linearly rescale the remaining pixels of $\mathcal{T}$ to lie between $[0, 1]$. Finally we set

$$f(\mathbf{x}) = \begin{cases} \tilde{\mathcal{T}}(\mathbf{x}) - mean(\tilde{\mathcal{T}}) & \text{if } \mathbf{x} \in \text{ Face} \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

where $\tilde{\mathcal{T}}$ is the masked and re-scaled version of $\mathcal{T}$, and $mean(\tilde{\mathcal{T}})$ is the mean pixel value of $\tilde{\mathcal{T}}$ across the facial pixels. By setting our target weighting function $f(\mathbf{x})$ in this way, our optimization tries to match the brightness at each point on the face relative to the average brightness of the face. In practice, we have found that this approach generally finds a good match in terms of the overall distribution of bright and dark regions on face and is relatively robust to differences in the total amount of light in the environment—i.e., we can use the same target in both dark and bright settings. For additional control, our tool lets users manually adjust the built-in target weighting functions using a painting interface as described in Section 4.4.2.

4.3.3 Placing an Additional Light Source

Photographers may sometimes have an additional light source (e.g., a phone screen or a flash unit) that they can manually orient around the face to further improve the match to the target lighting. We can adapt Equation 4.5 to compute the optimal orientation for this additional light source as follows

$$C(\mathbf{R_{add}}, \mathbf{R}) \;\; = \;\; \mathbf{F}^t \left(\mathbf{R_{add}}(\mathbf{L_{add}}) \odot \mathbf{R}(\mathbf{L})\right), \tag{4.7}$$

where $\mathbf{L_{add}}$ is an environment map capturing just the additional light source with an accompanying alpha mask set to 1 within the light source and 0 outside it, and $\odot$ is the *over* image compositing operator. Our optimization searches over the space of rotations for the additional light source $\mathbf{R_{add}}$ as well as the primary scene lighting $\mathbf{R}$ and considers the quality metric C for the composite of these two environment maps. We use the over operator under the assumption that the additional light source is closer to the face and therefore occludes the scene lighting behind it. Thus, our optimization finds the pair of orientations that best produce the target appearance. Note that because the additional light source is often handheld, we search a 2D band of rotation angles that vary both the light source's azimuth and altitude angles.

4.3.4 Implementation

Our optimization algorithm requires an environment map $\mathbf{L}$ and a target facial appearance weighting function $f(\mathbf{x})$ as input. While current 360 cameras can generate high resolution (e.g. 4K) environment maps, we have found that resolutions of 64×32 for $\mathbf{L}$ and 100×100 for the image are sufficient for our optimization (see Section 4.5.2). Additional resolution in the lighting or the image makes little visual difference in the appearance of the rendered head model, as the subsurface skin material we use is relatively diffuse.

We precompute the transport matrix $\mathbf{T}$ using the PBRT raytracer [232] treating the generic PBRT head with the built-in subsurface scattering skin material model (`kdsubsurface`) as our scene. Each column of $\mathbf{T}$ is an image of the scene as lit by a single pixel of the environment map $\mathbf{L}$. For our image and lighting resolutions, serially computing $\mathbf{T}$ on a single machine takes 17.1 hours. The resulting three-channel RGB transport matrix $\mathbf{T}$ is 246 MB uncompressed, and after precomputing $\mathbf{T}$, we can generate a color rendering of the scene under any lighting $\mathbf{L}$ using Equation 4.1 in 13 milliseconds (77 fps). In practice however, we only use the three-channel transport matrix to render the visualization of the lighting on the generic head model at the optimal orientation which we call the *PRT render match* (Figures 4.2 and 4.5). For all other computations, since the distributions of bright and dark regions matter more than color, we use a grayscale transport matrix $\mathbf{T}$ of size 81.9 MB uncompressed and also convert the lighting $\mathbf{L}$ to grayscale.

Pre-integrating the target weighting function $\mathbf{f}^{\,t}$ against the single-channel transport matrix $\mathbf{T}$ to compute $\mathbf{F}^t$ as in Equation 4.3 takes 3.5 milliseconds (286 fps). After this pre-integration, computing our image quality metric is very fast as it requires computing a dot product between two vectors $\mathbf{F}^t$ and $\mathbf{L}$ and takes 0.003 milliseconds (333333 fps). Given the pre-integrated target weighting function $\mathbf{F}^t$, we can identify the lighting rotation $\mathbf{R}$ that maximizes our image quality metric C using Equation 4.5 (Figure 4.5). We have found that it is sufficient to compute the image quality metric C at 5.625° rotation increments about the vertical axis. Finer sampling is unnecessary in practice as photographers usually only reorient to within a few degrees of the target orientation (Section 4.5.3). Computing the optimal orientation for the environment lighting takes 0.19 milliseconds (5263 fps) for a single lighting target and is linearly related to the number of rotation increments we sample (in our case 64).

To optimize placement of an additional light source, we sample a 2D band of rotation angles ranging from 0° to 360° at 5.625° increments around the vertical axis (azimuth) and ranging from $-67.5°$ to $67.5°$ about the horizontal axis (altitude) at 16.875° increments. Computing the orientation for the environment as well as an additional light source, then takes 110 milliseconds (9 fps) for a single target and is again linearly related to the number of rotation increments we sample (38864 in our case; 64 for the environment times 64×9 for the additional light). We have experimented with a greedy approach in which we first find the optimal rotation for the environment, keep that rotation fixed, and then search for the optimal rotation for the additional light source. This approach significantly reduces the number of rotation samples ($640 = 64 + 64 \times 9$), and therefore our overall cost to 1.9 milliseconds (526 fps). While it generally produces similar results to the full optimization, they are not always the same. In this chapter, we report all results using the full optimization and leave it to future work to further improve the efficiency of placing an additional light source.

All of our timings are computed using our Matlab implementation running on a MacBook Pro laptop (2017) with a 3.1 GHz Intel Core i7 processor. We expect a natively compiled implementation running on a GPU or even a CPU would run faster and could easily run on modern smartphones.

4.4 Interaction

For a photographer, using our portrait lighting optimization tool involves four steps: (1) capturing an HDR environment map of the location, (2) specifying a target facial

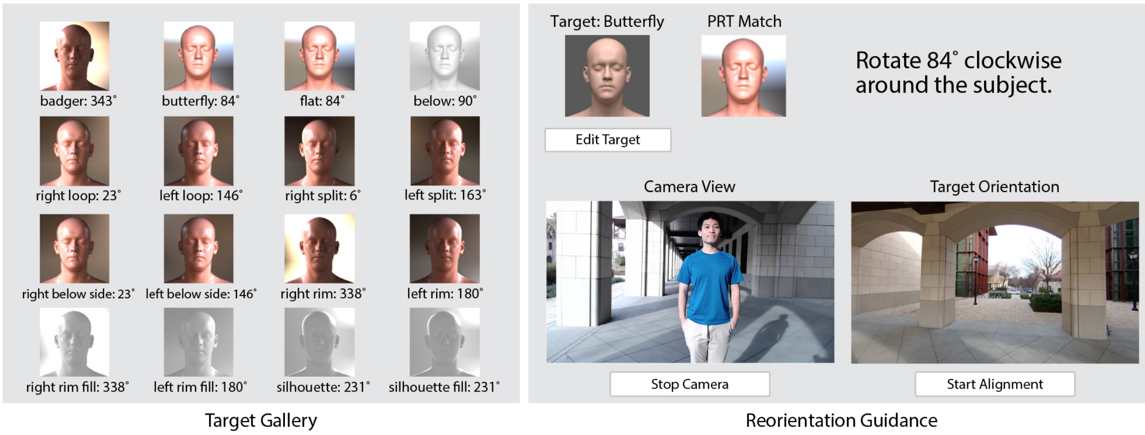


Figure 4.6: Upon loading an HDR environment map, the interface of our portrait lighting tool computes the optimal orientation for each target lighting style (Figure 4.4) and displays a **gallery of target appearances** (left). The tool grays out *unattainable* targets. Selecting a target brings up the **reorientation guidance** screen, which displays how far the photographer should rotate clockwise or counterclockwise about the subject (right). It also shows the background at the *target orientation* and the *camera view* from a webcamera at the current location of the photographer. As the photographer gets close to the target orientation, clicking the *Start Alignment* button overlays the current view on the target orientation view.

appearance, (3) following our tool's guidance to reorient self and subject, and (4) taking the final portrait.[1] Note that our tool can be used without modification to take selfies in which the photographer is the subject. We detail how the photographer interacts with our tool in the first three of these steps.

4.4.1 Step 1: Capture Environment Map

The photographer starts by capturing an HDR environment map in the location they would like the subject's face to appear in the final photograph using a 360 camera in bracketed exposure mode (we use a Ricoh Theta V). The photographer should ensure that nothing is obstructing the camera's view of the light sources in the environment. For handheld capture, we recommend that the photographer kneel under the 360 camera to minimize occlusions.

[1]Video demonstration of the tool can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

In bracketing mode, we take 3 exposures at fixed aperture and ISO—one exposed at the auto-exposure settings, one two stops brighter, and one two stops darker. The Ricoh takes about 10 seconds to capture the 3 images and outputs gamma-corrected sRGB at 5376×2688 resolution in JPG format. We provide a pre-processing tool that applies Ebner's [84] linearization algorithm via Matlab's `rgb2lin` function and then uses Reinhard et al.'s [240] algorithm via Matlab's `makehdr` function to convert the exposures into a single HDR environment map. Finally, we downsample the environment map to 64×32 resolution for use in our optimization procedure. This pre-processing takes about 16.2 seconds.

While the HDR capture, transfer to the laptop, and pre-processing can take a few minutes in our prototype implementation, the vast majority of the time is spent in the transfer. We believe that in the fully integrated tool we envision, where the tool runs on a single device that captures the environment at low resolution, this time would be significantly reduced (Section 4.7). To place an additional light source the photographer must a priori capture an HDR environment map of the additional light ($\mathbf{L_{add}}$ in Equation 4.7) in a dark (ideally blackout) room, using the same procedure and camera settings as used to capture the scene. We set the light (a phone screen) about one foot from the 360 camera in our experiments. In general we expect photographers would choose this offset distance based on the power of the light source and how close to the face they are willing to place the additional light. The photographer must capture a new HDR environment map for each additional light source and offset distance, but once captured they can be reused in any setting. In this case, the main environment map should be captured with matching camera exposure settings.

4.4.2 Step 2: Specify Target Facial Appearance

Our optimization tool includes a pre-designed gallery of target facial appearance weighting functions that are based on common studio lighting styles (Figure 4.4). Because these targets are known in advance, we pre-integrate them against the transport matrix $\mathbf{T}$ a priori. As soon as the photographer loads the environment map, our tool applies the optimization procedure of Section 4.3.1 to identify the optimal

Figure 4.7: Portraits captured in different environments using our tools to achieve a variety of target lighting styles. These lighting styles typically contain a strong primary light source in front of and above the face (butterfly) or to the side of the face (loop, split, rim). Our tool is able to emphasize the evenness of **butterfly** lighting and the key triangle in **loop** lighting. **Rim** lighting reproduces the strong highlight at the contour of the face in well-lit environments, but in the darker *library* and *quad night* scenes, the light is not quite strong enough to produce a bright contour.

orientation angle for each target and displays all of these results gallery of target appearances (Figure 4.6).

Not all target appearances are achievable in every environment. In such cases, the image quality metric is relatively low for all orientations and the quality of the maximizing orientation is close to the mean quality across all orientations. Therefore, our tool computes the maximum image quality across all orientations and if it is less than an absolute threshold ϵ, our tool marks the target appearance as *unattainable*. We empirically set $\epsilon = 45$. In the gallery of target appearances, all unattainable targets are grayed out. Additionally, photographers can manually adjust any of the targets, including those that are grayed out, using a painting interface (Figure 4.8). Our tool re-computes the pre-integration of the adjusted target weighting $\mathbf{f}^t$ against the transport matrix $\mathbf{T}$, and then runs the optimization using the adjusted $\mathbf{F}^t$.

4.4.3 Step 3: Follow Guidance to Reorient Self and Subject

Once the photographer chooses a specific target facial appearance, our lighting optimization tool tells the photographer how far to rotate (clockwise or counterclockwise) about the subject to achieve the desired appearance (Figure 4.6). In addition to the rotation angle, it provides the current view from the primary camera, and also displays the portion of the environment that should be visible after the rotation. The goal of the photographer is to match these two views. To further aid the reorientation process, our tool optionally provides real-time feedback showing how well the current and desired views match, by overlaying the current view onto the desired view after transformation by the best-fit homography between them (Figure 4.9). If the photographer is placing an additional light source, our tool next describes how to orient the light with respect to the subject's face as a pair of rotation angles (azimuth and altitude) for the center of the light, and provides a schematic diagram illustrating the relative orientation between the light and the face (Figure 4.12) after the subject has been oriented with respect to the environment. In practice, either the photographer, the subject, a third person or a tripod needs to hold the additional light source in place.

4.4.4 Interaction Details

In this section, we describe a couple features of our tool's interface in further detail. These are optional steps a photographer may choose to take after capturing the environment map and selecting a target lighting style. These features are designed to help the user better match their desired lighting style, either by adjusting the target based on their personal preferences and/or by aligned more accurately to the angle proposed by the tool.

Painting. If photographers have a specific lighting style in mind they can manually adjust the provided targets as they please using our painting interface (Figure 4.8). Users express additional light regions by painting with the orange brush and dark regions with the blue brush. After such adjustment, our tool re-computes the pre-integration of the adjusted target weighting $\mathbf{f}^t$ against the transport matrix $\mathbf{T}$, and then re-runs the optimization using the adjusted $\mathbf{F}^t$ to determine the optimal orientation

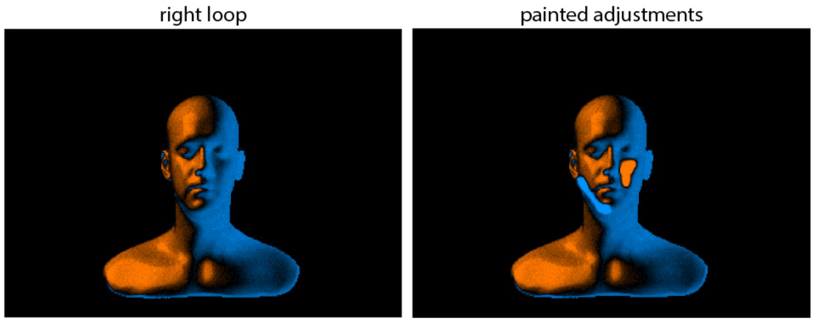


Figure 4.8: Photographers can manually adjust a built-in target appearance weighting function (left) using a painting interface that locally increases or decreases the weighting using a radial brush with Gaussian falloff (right). The positive paint stroke on the cheek (orange) tells the tool to find lighting that emphasizes the brightness of the key triangle under the eye, while the negative paint stroke (blue) along the bottom of the face tells the tool to look for lighting that provides extra contrast along the contour of the chin by making that region dark.

for the customized target facial appearance.

Alignment. Another feature that provides users with more precise control is the alignment guidance – optional real-time feedback to further aid in the reorientation process. Upon arriving close to the target orientation, the user can turn on this alignment guidance. The interface will compute the best-fit homography relating the current camera view and the target orientation view using SURF feature points [35] and MLESAC [280] (Figure 4.9). This homography is displayed to the user by overlaying the current camera view onto the desired view after transformation. When the images are aligned, the transformed current camera view will be vertically centered on the target view orientation.

4.5 Results

We have tested our portrait lighting tool in a variety of locations with many different subjects—e.g., of varying skin tones, hair styles, facial hair/accessories, etc. In each case, we captured the optimal orientation as reported by our tool for each of our built-in targets, including those that the interface suggested were unattainable in

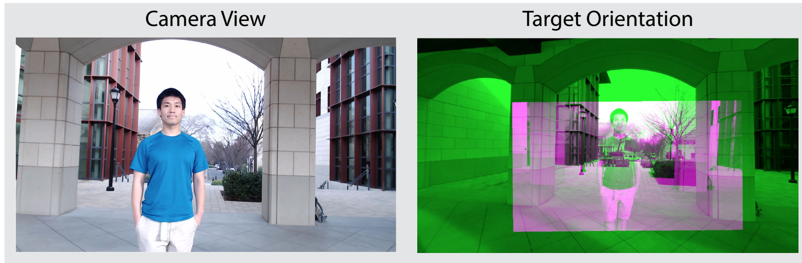


Figure 4.9: To aid the reorientation process our tool can compute the best fit homography between the current **camera view** and the view at the **target orientation**. It then overlays current view (pink) on the target view (green).

the environment. Figures 4.2, 4.7, and 4.10 show many of these results.[2] These examples include a range of different lighting conditions—e.g., indoors with windows and lamps, and outdoors during the day and night with sun, overcast, streetlights, or partly covered.

As shown in Figures 4.2 and 4.7, our tool is usually able to find good matches for the butterfly, loop (left and right), and split (left and right) lighting styles as these include a strong primary light source located above and either in front of the face (butterfly) or to one side of it (loop, split). In many of these cases, the strongest light is the sun either directly illuminating the face or passing through a window; but in some indoor and nighttime scenes, such as *library* and *quad night*, the primary light is due to lamps. Our tool is generally able to find orientations that emphasize the key triangle on the subject's cheek characteristic of loop lighting (i.e., in left loop, the cheek under the subject's right eye is bright relative to the right side of the face and vice versa for right loop), and reproduce the evenness of butterfly lighting. Rim lighting places a strong primary light slightly behind the subject's head to one side. In most of these environments, our tool finds a good orientation match for this lighting style (*arch walkway, gothic walkway, apartment, balcony*), but in darker environments like *library* and *quad night*, the primary light is not always strong enough to brightly illuminate the contour of the face.

[2]For more results, see our interactive visualization showing the image quality metric graph for each scene and lighting style: `http://graphics.stanford.edu/projects/portraitlighting/supplemental_vis/index.html`.

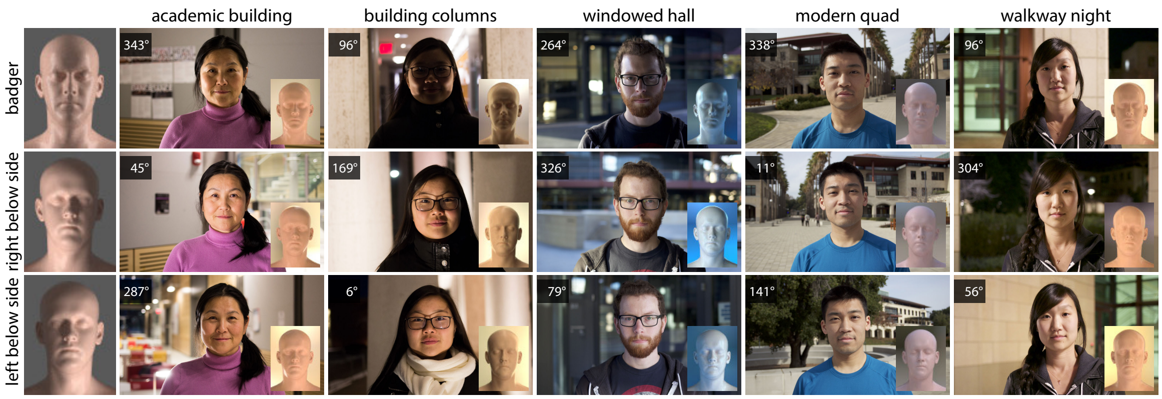


Figure 4.10: Some lighting styles are difficult to achieve in most fixed light environments. Badger requires two main lights facing one another with a dark center while the below lighting styles require strong illumination from below. However, our tool can sometimes find orientations that match such difficult target styles as we see here for the *academic building, columns night*, and *windowed hall*. For the other environments, *modern quad* and *walkway night*, it can only match the **badger** lighting on one side of the face and the **below side** lighting appears more like side lighting (like loop, which is from slightly above the face) than lighting from below.

Lighting styles that require two relatively strong lights facing one another (badger) or that require lighting from below (below, below side) are more difficult to achieve in most fixed lighting environments. Nevertheless as shown in Figure 4.10, our tool can sometimes find orientations that match these styles as we see for *academic building, columns night*, and *windowed hall*. In other cases like *modern quad* and *walkway night*, it matches the badger lighting on only one side of the face since there is only one strong main light. Similarly, in these two scenes, for the below lighting styles our tool suggests an orientation in which the subjects are lit a bit more from the side than from below.

Alternatively, users can use an additional light source to better achieve targets that require light from multiple directions. For example, Figure 4.12 shows how our tool suggests placing a phone screen to capture targets that are unattainable in the fixed lighting of the environment. In the target combining loop and rim lighting (top row), the additional light generates the bright rim highlight on the right contour of the face opposite the side lit by the primary light in the environment. In the badger example (bottom row), the additional light acts as a second key light brightening the left side of the face and approximately mirroring the primary light in the environment.

4.5.1 Additional Results

Here, we present some additional results generated using our tool to showcase its flexibility in achieving more diverse target lighting styles. Using the painting interface, a user can modify targets as well as potentially create more complex targets. We use an additional light source to create some of these varied targets that can be harder to achieve.

Painting. Figure 4.11 shows how users can adjust a target lighting style using our painting interface (Section 4.4.2). The painted strokes specify parts of the face that should be emphasized. Our tool re-computes the optimal orientation to brighten the the parts of the face painted with orange (positive) strokes and darken the parts of the face painted with blue (negative) strokes. In this case the user modifies the right split style to increase or decrease brightness in the center of the face and shift the location of the transition from bright to dark.

Additional Light. Some targets require specific relationships between lights in the environment that are unachievable. Figure 4.13 shows how users can use a mobile additional light source to capture some of these targets that otherwise would not be attainable in the fixed environment (Section 4.4.3). In these examples, the additional light is used to provide light in separate regions on the face than were lit by the environment. In addition to making more targets attainable, the light can also be used to enhance lighting styles by providing additional brightness in a dark environment, as shown for the left loop target in the fourth row of the figure. Finally, it can be used to adjust lighting in scenarios with a somewhat fixed subject orientation (e.g., with a selected background) to better achieve a desirable lighting style, such as the right loop lighting in the last row.

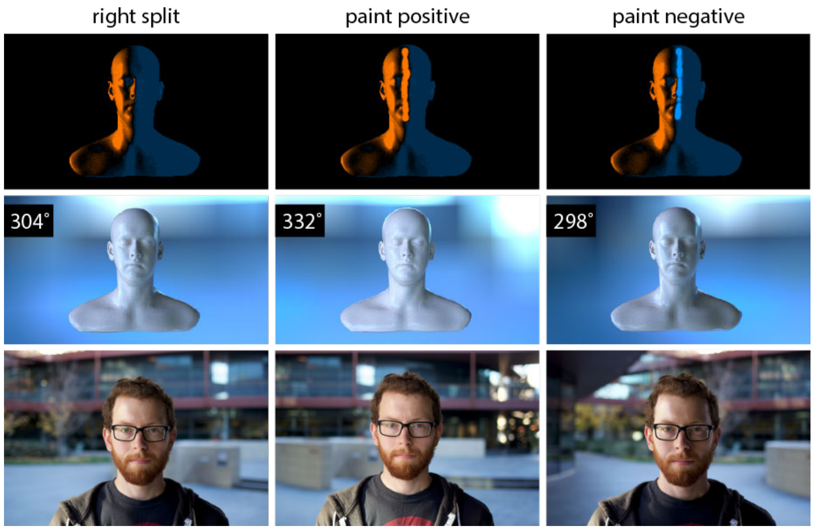


Figure 4.11: Initially, the built-in **right split** target lighting style makes one side of the face bright and the other dark (left). With the **positive** paint stroke (orange), our tool makes the center of the face brighter by shifting the orientation by 28° (middle). With the **negative** paint stroke (blue), it finds an orientation that makes the center of the face darker.

4.5.2 Algorithmic Evaluation

To evaluate the robustness of our optimization algorithm, we consider how several parameters affect the image quality metric and the resulting optimal orientation angle.

Resolution of Lighting and Rendered Image. To check how the resolution of the lighting and the image affect the image quality metric, we varied the lighting and image resolutions and generated image quality graphs for 192 (scene, target lighting) pairs as in Figure 4.14 (12 scenes and 16 lighting targets). For each such graph, we compute the difference between the optimal angle that maximizes image quality, and the corresponding optimal angle for the highest resolution lighting or image respectively. Averaged across all 192 (scene, target lighting) pairs, we obtain the following differences in optimal angles for lighting: 32×16 (diff: 6.15°), 64×32 (diff: 1.9°), 128×64 (diff: 0°); and for images: 50×50 (diff: 4.95°), 100×100 (diff: 2.02°), 200×200 (diff: 0°). While the differences in angles are relatively small (within a few degrees) on close inspection it is possible to see small differences in some of

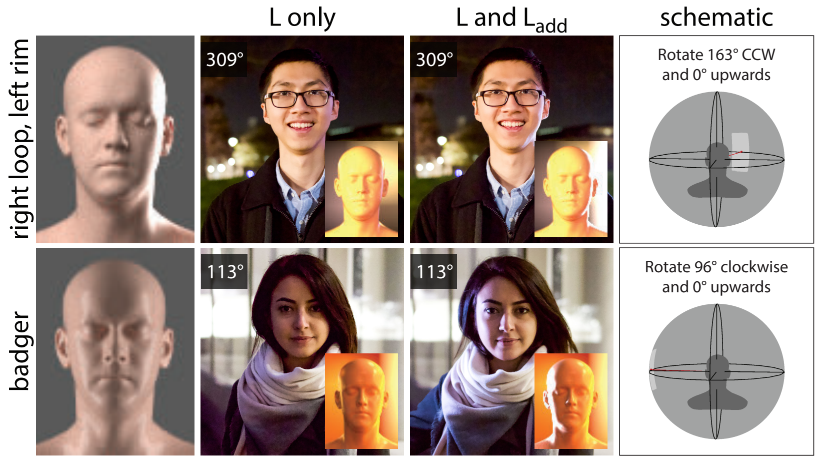


Figure 4.12: Our tool can suggest how to place a phone screen to better achieve lighting targets such as these that are not easily attainable in the fixed lighting environment. For each of these portraits, the environment provides light on one side of the face (**L only**), and the additional light is able to provide the light on the other side of the face (**L and $\mathbf{L_{add}}$**). The **schematic** shows how our interface conveys the optimal orientation for the additional light source relative to the face after orienting the subject to the environment.

the renderings. Overall, we find that a lighting resolution of 64×32 and an image resolution of 100×100 provides minimal loss in quality, while maintaining a relatively small sized transport matrix.

Distance Between 360 Camera and Subject's Head. We checked the robustness of our algorithm to using lighting captured by a 360 camera at various distances from the subject's head. As shown in Figure 4.15, lighting taken from 1 to 2 feet away from the subject produce similar image quality metric graphs, with maxima that are relatively close to one another. We find that the optimal angles differ from the optimal angle for the centered lighting by an average of 6.43° across all other lighting locations and lighting targets. These results suggest that as long as the lighting is relatively far away from the viewer, the environment map does not need to be taken exactly at the location of the subject.

Width of Optimal Image Quality Peak. Our optimization is designed to find the peak in the image quality metric graph. For each such peak, we compute the range of

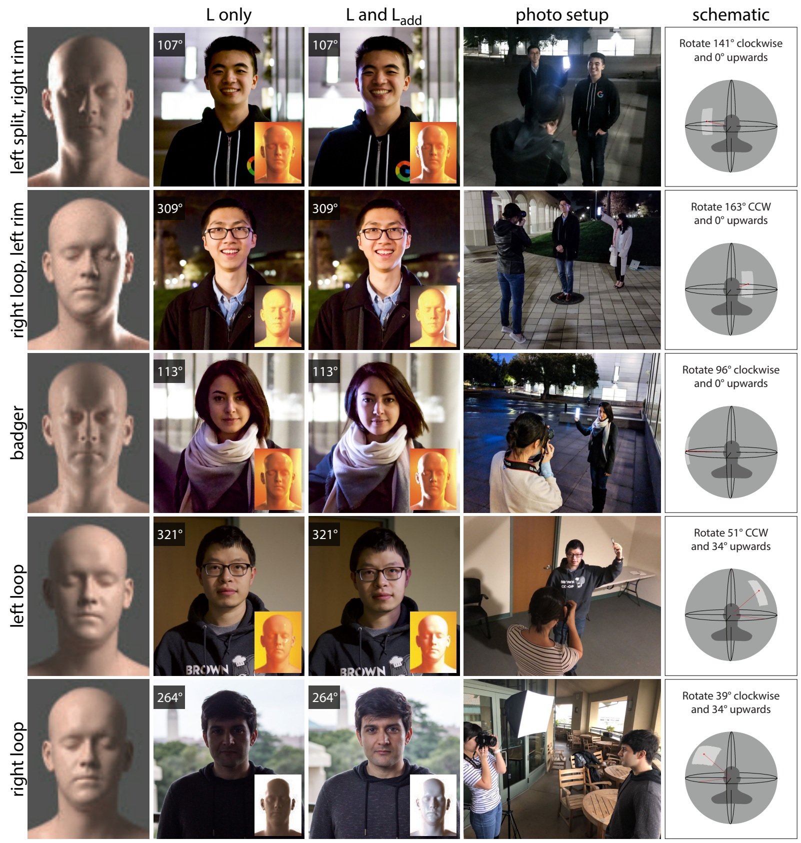


Figure 4.13: Our tool can suggest how to place an additional light source (a phone screen for all examples other than the last row, which uses a light box) to better achieve lighting targets, particularly for those that are not easily attainable in the fixed lighting environment. For each of the portraits with light on both sides of the face (top 3 rows), the environment provides light on one side of the face (**L only**), and the additional light is able to provide the light on the other side of the face (**L and $\mathbf{L_{add}}$**). The **photo setup** shows a third person perspective of how each of these portraits was achieved. For **left loop**, the additional light is used to further enhance the brightness of the lighter region of the face. For **right loop** (bottom), we have a desired background in the environment. Thus, the additional light is positioned relative to the fixed subject orientation to generate the desired lighting style. The **schematic** shows how our interface conveys the optimal orientation for the additional light source relative to the face after orienting the subject to the environment.

angles on either side of the peak for which the image quality metric remains within 3% of the maximum value. Averaged across all of our 192 (scene, target lighting) pairs, we find that this range is $17.05°\sigma = 9.73°$, or $\pm 8.5°$ on either side of the peak. As long as

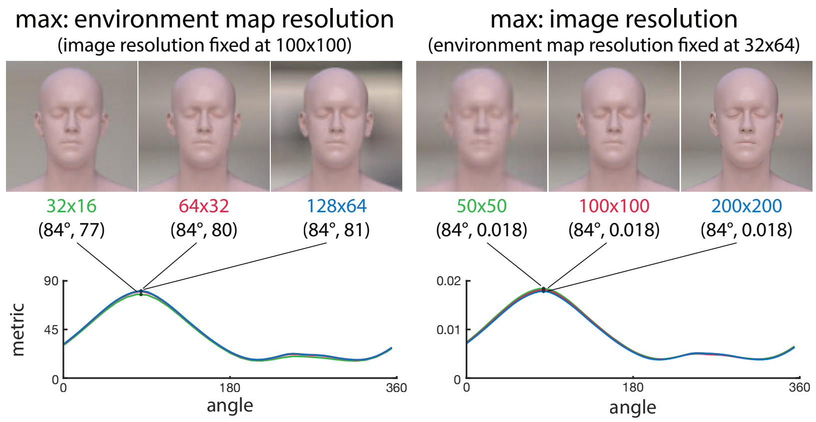


Figure 4.14: Effects of varying environment map and image resolution on **image quality metric** graphs for the *modern quad* scene and *butterfly* target. We have tested 3 lighting **environment map resolutions**: 32×16 (green), 64×32 (red), 128×64 (blue), and 3 **image resolutions**: 50×50 (green), 100×100 (red), and 200×200 (blue). The corresponding image quality graphs sit on top of one another indicating very little difference between them (readers should zoom in to see the red and green curves). It is possible to see some small differences in the renderings, however for this (scene, lighting target) pair the optimal angles are the same across all the resolutions.

photographers are within this range of the optimal orientation computed by our tool, they will produce an image with lighting that is very similar to the rendered target image (within 3% in terms of the image quality metric). As we show in Section 4.5.3, we find that users typically match the optimal orientation to within $\pm 9°$ of the peak.

4.5.3 User Evaluation

To better understand how our tool helps users, we ran a user study with 28 participants (13 male, 15 female), 19 to 31 years old ($\mu = 25$). In this group, 20 self-identified as novice and 8 as experienced at photography. We separate novice and experienced participants in this case because portrait lighting specifically is a topic where awareness varies drastically across experience levels. Many novices will have almost no context on how to think about lighting a face, while non-experts with a little more experience will likely know to pay attention to the lighting on the face and might know a few lighting styles, but rarely will have the experience that experts have of setting up

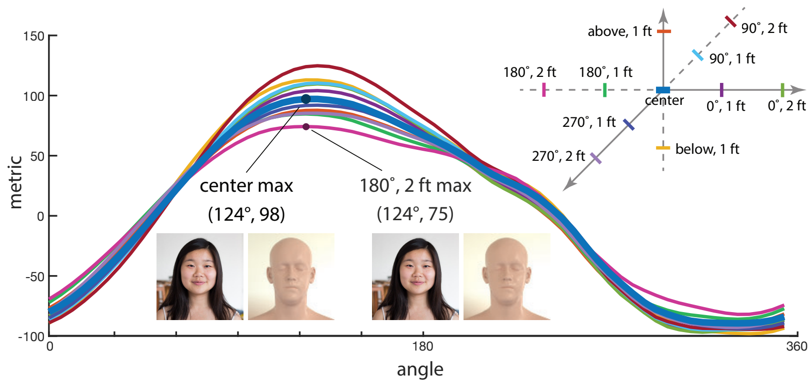


Figure 4.15: Effects of varying the position of lighting with respect to the subject's head on **image quality metric** graphs for the *art studio* scene and *left loop* target. We vary the position of the 360 camera around the subject's head (located at the **center** of coordinate frame as in inset). Despite the differences in the graphs, in this case there is no difference in the optimal angle for the centered lighting versus the optimal angle for the lighting **2 feet away at 180°**. We find that the optimal angles differ from the optimal angle for the centered lighting by an average of 6.43° across all other lighting locations and lighting targets.

studio lighting styles. They still are of interest in our work because developing our tool can assist in training their "artistic vision" and enabling them to see the many lighting possibilities available in a given location (see our definition in Section 1.3.2).

Participants were asked to complete the task of taking a well-lit portrait in three conditions: (1) First, we asked them to position and orient their subject with no guidance. (2) Next, we next showed them images of 16 pre-defined studio lighting styles (Figure 4.4 top row) and asked them to use one as reference and take another well-lit portrait matching it. (3) Finally, we asked them to use our interface to select an attainable target lighting style and capture it.

Following each condition, participants were asked to complete surveys with a number of Likert questions (on a 7-point scale) about their experience using the tool and evaluating their photo. We also asked them to answer NASA-TLX questions (0 to 100) [122] to measure their overall cognitive load. After all 3 conditions, we asked them to evaluate the accuracy of each of their photos (2nd and 3rd condition) in achieving their target lighting style and how easy it was to reorient to capture this photo. We ended the study with open-ended interviews asking about what they liked/disliked about the tool and how the interaction influenced their thought process as they took photos. Participants were compensated $15 for their time.[3]

After each task, we asked participants to rate the usefulness of the method they used to produce a well-lit portrait on a Likert scale running from 1-strongly disagree to 7-strongly agree. We found a significant difference (Friedman test) in usefulness between the three tasks [$\chi^2(2) = 18.6$, $p < 0.001$]. All three pairwise comparisons (Wilcoxon signed-rank test) were significant; in particular, participants rated our tool in task 3 as more useful ($\mu = 6.5$, $\sigma = 0.7$) than simply seeing a reference lighting style image as in task 2 ($\mu = 5.5$, $\sigma = 1.1$, $p < 0.025$) or having no guidance as in task 1 ($\mu = 4.5$, $\sigma = 1.2$, $p < 0.005$). We similarly asked them to rate how well-lit the images were in each condition (Likert) and found a significant difference [$\chi^2(2) = 7.0$, $p < 0.03$]. The images produced using our interface were more well-lit ($\mu = 6.3$, $\sigma = 0.6$) than using only a reference target ($\mu = 5.5$, $\sigma = 1.3$, $p < 0.025$) or having no guidance ($\mu = 5.5$, $\sigma = 1.3$, $p < 0.025$).

In a post-study interview, our novice participants consistently said that our tool reduced the mental effort required to produce well-lit portraits. They attributed this reduction to the separation between choosing the target lighting style and the guidance to achieve the chosen style. One wrote, *"The thinking was concentrated in the stage when I chose the target... After I chose what I wanted, the tool made it pretty simple to get [it]"* (P7). Evaluating based on NASA-TLX (excluding physical demand) to further assess differences in cognitive load between tasks 2 and 3 for novice participants, we found a significant difference [Wilcoxon $V(2) = 64.5$, $p < 0.05$], suggesting that our tool makes it mentally easier to achieve desired lighting styles compared to simply using a reference target image. Directly asking the novices to rate the ease of orienting the subject (Likert), we also found a significant difference between using a reference target in task 2 ($\mu = 3.8$, $\sigma = 1.6$) and using our interface in task 3 ($\mu = 6.1$, $\sigma = 0.9$) [$V(2) = 153$, $p < .001$].

We checked how well all the participants matched the orientation angle proposed by our tool in task 3, by comparing the background of their portraits to the corresponding

[3] Materials for the study can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

environment map. We found that they were able to reorient to within an average of 9° ($\sigma = 7.5$), on either side of the target orientation. When manually orienting the subject for task 2, participants were only able to orient to within an average of 49° ($\sigma = 50.1$) [$t(27) = 4.3$, $p < .001$]. In this case, only 3 of 20 novice participants and 5 of 8 of our more experienced participants were able to reorient to within 10° of the optimal target orientation.

Finally, interviews with our experienced photographers revealed that our interface made them more confident and deliberate in their lighting choices. They said they were able to achieve better lighting *"without as much mental demand"* (P5). They similarly appreciated that *"it separated the choice of lighting from the rest of the photographic process"* (P14), encouraging them to *"branch out"* (P14) and *"explore"* (P2). A professional New York Times photographer said that our tool would also be helpful in scenarios where he has limited time in a location with a subject.

4.5.4 Experienced Versus Novice Photographers

Our user study participants ranged in photography experience—including 20 novice photographers, and 8 experienced photographers. We discuss some differences in how experienced versus novice photographers rated our tool.

Overall Performance

Our paper reports the usefulness of methods used in each task aggregated across all participants. The breakdown of ratings between novice and experienced participants in Table 4.1 shows ratings were similar for task 2 and 3; the main difference is in task 1 where participants are provided no guidance. As expected, experienced photographers have a more concrete existing process for capturing a well-lit portrait with no external guidance. Similarly, the experienced participants rated the quality of their images produced in each task higher—all scores across all tasks reflected that they somewhat to strongly agreed (5-7 on the Likert scale) that their resulting portraits were well-lit.

Finally, Table 4.2 shows the breakdown for how accurately novice versus experienced participants were able to match the optimal orientation angles as proposed by our tool. Experienced participants were able to achieve much higher orientation accuracy than novices in both task 2 and task 3. Provided with just a target lighting style in task 2, experienced participants were able to manually determine orientations similar to those computed by our tool. Additionally in task 3, provided with the reorientation guidance, they were able to more accurately determine when they had achieved the optimal angle.

Qualitative Feedback

Our post-study interviews provided insight into how the tool impacted the capture process of novice versus experience participants differently.

Novice. Many novice participants said that the interface increased their awareness and intentionality with regards to lighting. Many mentioned having *"no idea what I was looking for"* (P27) in the first photo, but feeling more confident when taking the third. One described that this increased confidence was due to the tool feeling as if they had *"a sort of professional in the loop"* (P13). One wrote *"I liked that it made me think beyond just the content of the photo, and also pay attention to where the light sources are... It worked really well too, which makes me feel like I can take much more dramatic and varied photos in a limited space"* (P26). Several commented on the usefulness of the gallery of target facial appearances as it made them realize they *"had so many drastically different [lighting] options"* (P17).

Experienced. Our experienced participants expressed that the tool made them more confident and deliberate in their lighting. They said that it would be especially useful to them in complex or formal lighting environments where they needed more precise control to achieve a specific lighting style. In particular, P2 said that *"I would be satisfied with just the indication that some styles of lighting are achievable and some are not (that would also remind me of different styles that I might not be thinking about)."*

They described a number of ways in which the tool adjusted how they thought about the task of capturing well-lit portraits; specifically, experienced participants appreciated that the tool provided scaffolding to help organize their photo shoots.

measure	population	task 1	task 2	task 3	$\chi^2(2)$	p
method:	all	4.5, $\sigma = 0.3$	5.5, $\sigma = 0.3$	6.5, $\sigma = 0.2$	18.6	$< .001$
useful	novice	4.2, $\sigma = 1.4$	5.6, $\sigma = 1.1$	6.5, $\sigma = 0.7$		
	experienced	5.0, $\sigma = 0.7$	5.4, $\sigma = 1.1$	6.4, $\sigma = 0.9$		
photo:	all	5.5, $\sigma = 0.3$	5.5, $\sigma = 0.3$	6.3, $\sigma = 0.1$	7.0	$< .03$
well-lit	novice	5.3, $\sigma = 1.4$	5.3, $\sigma = 1.5$	6.2, $\sigma = 0.6$		
	experienced	6.2, $\sigma = 0.8$	6.2, $\sigma = 0.4$	6.6, $\sigma = 0.5$		

Table 4.1: Breakdown of results: user assessment of methods and resulting photos for each task. We found a significant difference (Friedman test) between the three tasks, as well as pairwise across tasks (Wilcoxon signed-rank test), both in terms of usefulness and how well-lit the resulting images were in each condition.

measure	population	task 2	task 3	t	p
accuracy:	all	48.8°, $\sigma = 50.1°$	9.3°, $\sigma = 7.5°$	4.3	$< .001$
orientation	novice	64.3°, $\sigma = 51.7°$	10.2°, $\sigma = 8.2°$		
	experienced	10.0°, $\sigma = 6.5°$	6.9°, $\sigma = 5.1°$		

Table 4.2: Breakdown of results: measurement of user accuracy in achieving the optimal orientation angle in task 2 versus task 3. We found a significant difference between the two tasks (Paired t-test).

P16 said: *"I would use this for shooting portraits of other people so that we could be efficient. Instead of trying lots of positions hoping to get good lighting, I could use the tool to identify where and how to position the subject for the desired lighting."*

Another expressed enjoying how it changed his creative process, *"I was really surprised at how fun the tool was to use and how creative I felt using it. It really helped me see the world in a different way"* (P14). He described that by first selecting a lighting style, he used that as a creative constraint to push his creative exploration in other aspects of capturing the portraits.

4.6 Limitations and Future Work

While our approach for optimizing portrait lighting at capture-time can often suggest how to achieve a variety of lighting styles, it does have a few limitations that suggest directions for future work.

4.6.1 Fixed Lighting

The main assumption of our approach is that lighting in the environment is fixed. As noted in Figure 4.10, some environments may not contain lights at the locations necessary to achieve certain lighting styles. However, our tool does mark lighting styles as unattainable in such situations so that the photographer is aware of the problem. Our tool can also help the photographer place an additional light source to achieve the target lighting. However this requires that either the photographer, the subject, a third person or a tripod hold the additional light in place.

4.6.2 Colored Lighting

We perform our optimizations using a single-channel grayscale transport matrix $\mathbf{T}$ that captures the brightness in the scene, but cannot differentiate between colored lights. However, in environments with multi-colored lights (e.g. on stage, at amusement parks, etc.) or strongly illuminated reflective colored surfaces (e.g. painted walls), it could be useful to optimize distribution of colored regions on the face—i.e., find the orientation that casts a cool blue shade or a warm red glow on the face [42]. This would require using the full-color transport matrix $\mathbf{T}$ in our optimization as well as three-channel target weighting functions.

4.6.3 Head and Skin Models

Using generic head and skin reflectance models allows our tool to use an efficient PRT-based approach to compute the optimal lighting orientation. As a result, our approach cannot account for individual differences in the geometric shape of the face or reflectance properties of the skin (due to skin tone, skin type, facial hair, etc.). We argue that the overall distribution of light and dark regions on the face is less affected by such individual differences. Nevertheless, these can have some effect on the appearance and shape of light and dark regions. We also do not consider how hair can occlude the light falling on the face. One approach may be to build a subject-specific head model using recent image-based facial geometry acquisition methods [18, 101] or RGB-D cameras. But even given the geometry, efficiently optimizing the lighting orientation without using our PRT-based approach is an open challenge.

4.6.4 Background Aesthetics

Our tool primarily focuses on matching the bright and dark regions on the face to a user-specified target. It does not consider other aesthetic criteria such as the background of the image. However in many locations, the background in some orientations is much nicer than in others. While automatically identifying "nice" backgrounds is an open problem, given such an algorithm, we could potentially incorporate it into our tool by limiting our space of rotations to only the region containing "nice" backgrounds. We have started experimenting with a clutter detection approach to identify "nice" clutter-free background regions in our environment maps.

4.6.5 Optimizing Subject Position

Our tool assumes distant lighting and only considers the orientation of the subject relative to the lighting environment. In many locations, changing the subject's position, even by a few feet, can significantly alter the light falling on their face, especially when there are nearby occluders. Optimizing position in addition to orientation would enable much more control over the lighting, but likely require more significant modeling of geometry and light transport in the environment.

4.6.6 Alternative Capture Approach

An alternative approach for producing well-lit portraits might involve first recording a video centered on the subject's face as the subject rotates in place by 360 degrees, and then analyzing the video to find the frame that best matches each target lighting style. Our experience is that capturing such video at high-quality (without motion blur, while maintaining a nice facial expression, etc.) is very difficult. Nevertheless, assuming it is possible to capture such a video at high-quality, an open direction for future work is to develop an automatic algorithm for matching frames of the video to a target lighting style.

4.7 Discussion

Our goals in designing a capture-time lighting optimizing tool were to provide a computationally simple and efficient technique that could be used with off-the-shelf hardware today and could be fully incorporated into a single camera system in the future. Our prototype implementation achieves these goals using two complementary ideas. The first idea is to treat a 360 camera as a light probe that can quickly capture the lighting in the environment. We believe that such light probes could eventually be built into cameras as just another sensor akin to light meters and focus sensors today. The second idea is to efficiently find the optimal lighting orientation by evaluating the appearance of a generic face model under different rotations of the captured lighting using a PRT-based approach. In fact, we show that evaluating each orientation requires computing a single dot product between vectors that are the size of the environment map, and that a low-resolution environment map of size 64×32 is sufficient for optimizing portrait lighting. One of the main implications of our work is that incorporating even a very low-resolution 360 light probe into a camera system would be useful for our application. Moreover, because our optimization is extremely efficient, requiring 0.19 milliseconds, it is well suited for running on a single integrated device. This would further eliminate the cumbersome switching between operating a 360 camera and a laptop and a primary portrait camera as in our prototype.

A concern when using PRT-based techniques is that the size of the transport matrix $\mathbf{T}$ (246 MB uncompressed in our case) can be prohibitive for devices with limited memory. However, if the target appearance weighting functions $\mathbf{f}^t$ are known a priori, we only have to store the pre-integrated lighting domain vectors $\mathbf{F}^t$ to evaluate the image quality, and it is unnecessary to store $\mathbf{T}$. In our application, we use $\mathbf{T}$ to render the PRT matches for the optimal orientation angles we identify for each target. These visualizations can help the user understand what the face will look like in the target orientation, but they are not essential for using our system. We also use the transport matrix $\mathbf{T}$ to pre-integrate the target weighting function when the user manually adjusts it via our painting interface. In practice however, we believe most casual users would focus on the built-in set of target lighting styles and would not need the additional control afforded by the painting interface. Thus, it may be possible to offer most of the benefits of our approach without the transport matrix in the end-user interface.

4.8 Chapter Summary

Lighting is a crucial element of portrait photography. But choosing how to orient a subject with respect to the environment has traditionally required paying careful attention to the available light, as well as understanding of how that light might fall on the face to produce different looks. We have demonstrated a capture-time tool that uses a 360 camera as a light probe and suggests how to orient the subject with respect the to surrounding environment to best achieve a user-specified target appearance. We believe that such a tool can make it easier for photographers to consider a variety of different looks that may be possible in an environment.

4.8.1 Design Goals

We close this chapter with a summary of how the design of this interface fits within our design goals (Section 3.4).

Context-Aware

For portrait lighting, the important context is the lighting on the subjects face. We capture this context using a 360 camera as an HDR image. Given our insight that portrait lighting is about relative regions of dark and light, we were able to generate a range of diverse studio lighting setups virtually using a generic 3D model of a head. By representing portrait lighting as a discrete set of lighting styles and corresponding reorientation angles, we were able to allow participants to consider varied lighting styles in the context of the current scene [275].

Encourage Exploration

Users are presented with the gallery of lighting styles, showing a wide range of lighting options at once. By default, the interface shows them a virtual rendering of the lighting styles as they can be best achieved through reorienting the subject in the current environmental conditions. This allows them to easily scan the gallery to preview what the end result might look like. Using the tool, participants were able to quickly explore through a gallery of lighting styles, pick their desired style, and ultimately achieve results that they believed were of better quality.

Maintain Flexibility

The user has flexibility to choose between the lighting styles in the tool. For additional flexibility, we also allow customization of the target lighting styles through a painting interface. For a more experienced user, one can also imagine customizing the tool to display and optimize over a completely different set of lighting styles based on personal preference.

Chapter 5

Composition

For composition, the context to consider is the relative positioning of objects within the frame. Exploring composition can involve moving around the camera to try out different angles and framing, as well as moving the objects in frame. Users should have creative flexibility in choosing whatever compositions they find appealing. At this stage of our walkthrough (see Figure 5.1), the photographer is made aware of the centeredness of this composition and chooses instead to compose using the rule of thirds.

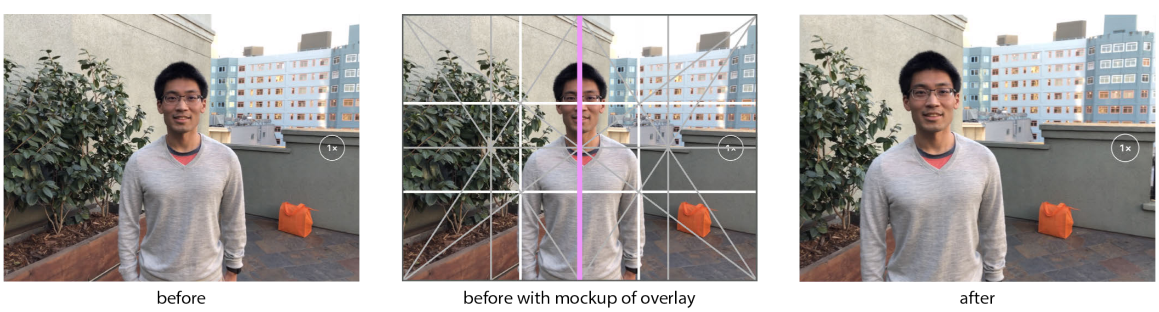


Figure 5.1: Here we see photos from our walkthrough (Figure 1.2) from **before** and **after** the photographer considered composition. In the middle we show a **mockup** of what the photographer might see upon launching our tool. The composition tool highlights the centered composition, and helps confirm alignment to a new composition.

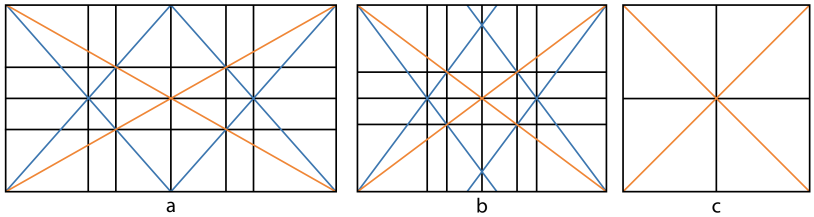


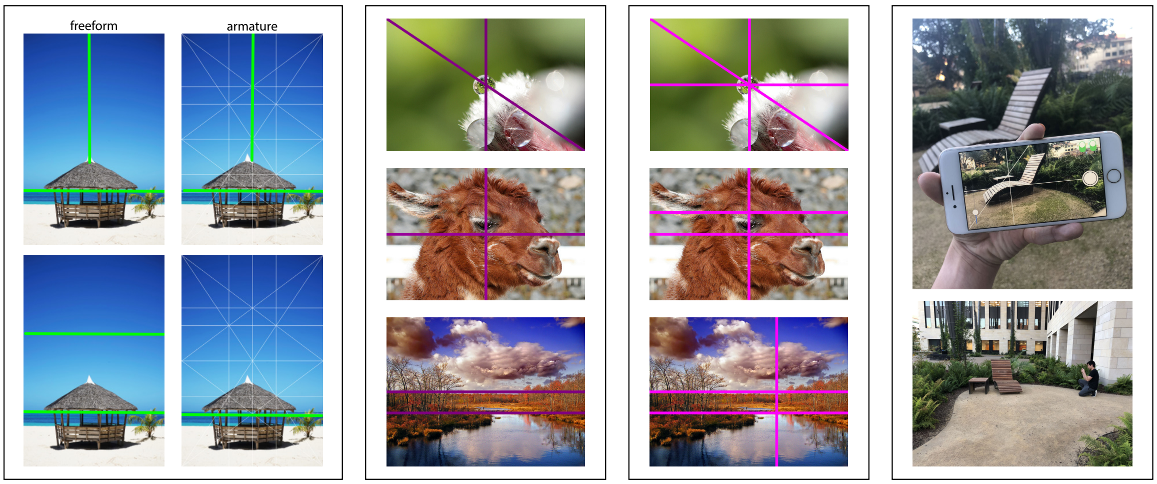
Figure 5.2: The armatures are designed around main diagonals (orange) and shorter, reciprocal diagonals (blue). Horizontal and vertical lines extend from intersection points of these diagonals. Our Mechanical Turk task presents three forms of armatures for different aspect ratios: **(a)** *wider/taller than root 2*: based on proportional design—reciprocal diagonals always connect to center resulting in horizontal/vertical lines at thirds and quarters, **(b)** *between root 2 and square*: equivalent to (a) at a specific aspect ratio, the reciprocal diagonals are perpendicular to the longer (main) diagonals, **(c)** *square*: (b) collapses to this set of lines.

5.1 Introduction

An important artistic aspect of photography is composition—the relationships between the different elements in an image and their properties (i.e., size, position, color, texture, etc.). Composition is used to direct the viewer's gaze through a photo by establishing visual hierarchies and rhythm [41, 110, 111, 223]. In this chapter, we focus on the spatial relationships of elements within the frame. These spatial relationships have been studied over the centuries in fields like painting, photography, cinema, and graphic design. They are often explained through the use of grids of lines, with popular examples like the rule of thirds or the golden ratio. These grids rely on the assumption that aligning important visual elements along their lines and intersections will tend to generate images that are more visually appealing to the human eye [109, 121, 223].

A common grid used by photographers like Henri Bresson-Cartier and Annie Leibovitz [110, 121, 147] is the harmonic armature (Figure 5.2), which dates back to Pythagoras [109, 121]. While some experienced photographers do not need to see

Much of the material of this chapter is as it appears in *"Adaptive Photographic Composition Guidance."* by Jane L. E, Ohad Fried, Cynthia Liu, Jianming Zhang, Radomír Měch, Jose Echevarria, Pat Hanrahan & James A. Landay, published in the Proceedings of CHI '20. For more information, see the project page: `http://graphics.stanford.edu/projects/adaptivearmatures/`.



experienced photographer annotations | novice annotations | adaptive armatures | interactive guidance

Figure 5.3: To design our interactive composition guidance interface, we were interested in better understanding people's ability to recognize composition and to annotate them on a composition grid. Left to right: We collected annotations from both **experienced** photographers as well as **novices** on Mechanical Turk. Inspired by these results, we developed an algorithm for heuristically computing these lines, or **adaptive armatures**. We display these adaptive armatures as an overlay in an in-camera composition **interactive guidance** tool and study how it impacts how people take photos. Photos by Sharon Mollerus, yugoQ, and Nicholas_T. Creative Commons Attribution License.

and strictly follow such grids, others directly place overlays on their camera view for guidance [67, 112]. Many cameras provide the option of directly overlaying a variety of different grids on the viewport. In any case, not all the lines in a grid are meaningful for a given scene, so it is up to the photographer to determine which ones to use and how. While at first glance human observers seem to intuitively perceive relationships between the elements in an image and an overlaid grid, it actually takes practice to find and focus on the more relevant ones. Books, blog posts, etc. often illustrate composition with composition grids overlaid and individual lines highlighted to emphasize certain aspects of the artist's composition choices [200]. However, this is an offline process—photographers manually annotate lines to better explain existing images.

Inspired by this behavior of manually highlighting lines in existing images, we propose a novel in-camera guidance that aims to automatically highlight the relevant lines to the composition of the current camera image (see Figure 5.3).

In this chapter, we present:

- an **algorithm** for automatically identifying adaptive armatures based on image saliency,
- an **interactive in-camera app** that shows these adaptive armatures to users as composition guidance, and
- a **user evaluation** of this app that shows that participants do in fact prefer the composition of photos they took using the adaptive guidance versus using static composition guidance.

We additionally contribute:

- a **dataset of crowdsourced composition annotations** that capture the common knowledge of what people believe are the most relevant lines to an image's composition within the constraints of the harmonic armature grid, and
- a **user evaluation** comparing the use of static composition guidance versus no guidance that shows participants are more confident in their compositional ability and feel more creative using the guidance.

5.2 Related Work

For an in-depth discussion of related work on capture-time guidance, see Chapter 2. Here we focus on work specific to composition. Since a number of guidance interfaces exist specifically for photographic composition, we discuss the most relevant systems despite being covered already earlier in the dissertation.

5.2.1 Composition Analysis

Previous work has focused on low-level image components that affect photographic composition. Zhou et al. [318] estimate vanishing points and use them to retrieve similar images. Lee et al. [179] detect dominant geometric elements and use convolutional neural networks to classify photographic composition rules in outdoor scenes. He et al. [126] find and highlight triangles on images to create compositional awareness and promote creativity. Other works classify an image to retrieve similar high quality examples for inspiration [185, 310, 316]. We focus on the placement of salient features with respect to composition grids, rather than specific image components or types of composition [313].

5.2.2 Composition Enhancement

Although several image recomposition methods have been proposed [144], cropping is still a very straightforward way to improve composition. Some have focused on explicit attention and aesthetics models [60, 292], but the most popular approach is to learn it directly from examples from professional photographers or user annotated samples [26, 61, 62, 117, 183, 187, 296, 309]. This approach tends to work well for automated workflows, although performance may vary depending on the dataset used for benchmarks [56]. Fang and Zhang [89] trained a network to find good crops in 360° images. Zeng et al. [312] proposed grid anchors to reduce the number of crop candidates to evaluate. Example-based cropping has been proposed for general images [58, 188], and portraits [315]. Rather than automatically finding good crops, we instead try to help users achieve good compositions actively by changing their point of view during capture.

5.2.3 Capture-time Guidance for Composition

Modern cameras come with some visual aids to help the photographer make technical and creative decisions during capture. In fact, one common feature is overlaying a static composition grid. It is commonly accepted as being useful guidance, but there is little formal evidence of this fact. In this chapter, we also take a step towards understanding how users use these static composition grids at capture-time (Section 5.5.2).

There is additionally a set of tools that provide composition guidance either by suggesting crops or by guiding the user to move the camera to better achieve a desired composition. Mitarai et al. [208] detect relevant elements in the current frame to classify the composition and show overlays providing specific guidance on how to refine them, limiting the user's creative freedom. Lujun et al. [197] perform image assessment and cropping suggestions; Ma et al. [199] use a view proposal network to suggest different crops, plus an interface that learns user preferences for future proposals. We are interested in composition guidance that creates awareness about different creative options, while providing a mental model for the user to reason about them. These interfaces provide limited feedback for building that mental model.

Closely related to our work, Xu et al. [308] devised a 3-camera system that provides real-time instructional feedback to users, so that they can learn to compose better portraits using the rule of thirds. For that, they compute a measure of the alignment between the regions of interest and the rule-of-thirds grid, and show arrows on the viewfinder for the users to follow to improve the alignment. Apart from focusing on composition beyond portraits, our guidance also uses composition grids, but aims for a less constrained user experience: instead of using arrows to make the user follow the criteria of the algorithm, our system highlights potentially interesting compositions using adaptive armatures. Our user experience is then closer to the *smart guides* implemented in many design tools [206]: when salient elements show interesting spatial relationships between them, such relationships are highlighted for the user, so they can decide whether to follow and refine them or not.

5.3 Composition Guidance

Our first question was how to design in-camera composition guidance. Many cameras provide options for a number of composition grid overlays, but are these effective for achieving well-composed images?

As we learned in our formative interviews (see Section 3), even experienced photographers tend to restrict themselves to options proposed in front of them—e.g., following the rule of thirds guidelines if those are visible on the camera screen. Thus, we were interested in using a more versatile grid, the harmonic armature, that allows for more compositional diversity. The harmonic armature provides different sets of orthogonal and diagonal lines. Diagonal lines are usually employed for dynamic symmetry [121], while from their intersections, the rule of thirds and the golden ratio grids emerge.

As noted in the chapter introduction, photographers currently manually highlight subsets of lines on existing images to describe how the lines are relevant to the image composition; and they instinctively do the same during capture, as surfaced in our formative interviews. We wanted to be able to provide similar feedback in real-time, in the camera. However, we first wanted to understand whether through training, experienced photographers develop a consistent view of which lines are relevant, and whether novice users are similarly able to produce such annotations.

5.3.1 Experienced Photographer Annotations

To understand whether experienced photographers had a consistent view of what lines described a composition, we ran a study where we had 8 participants annotate photos. First, participants were asked *"For each image, please freeform draw a set of lines that describe the composition of the physical elements of the image. Explain why you drew those lines."* Next, the same images are displayed one at a time with the harmonic armature overlaid (see Figure 5.2b). Participants were asked to perform the same task, but to only select lines from the armature. For this task, we chose 5 images of varying complexity in composition for all participants to annotate. We additionally asked them to bring in 5-6 of their own photos to annotate. We asked them to choose photos that they believed had particularly nice/interesting composition.[1]

Experienced Photographer Annotation Results

In watching these experienced photographers annotate images, we learned that experienced photographers are are able to recognize lines in image composition whether they are drawing freeform or using a composition grid. Given the constraints of a composition grid, they were also generally comfortable with selecting lines that

[1] Materials for the study can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

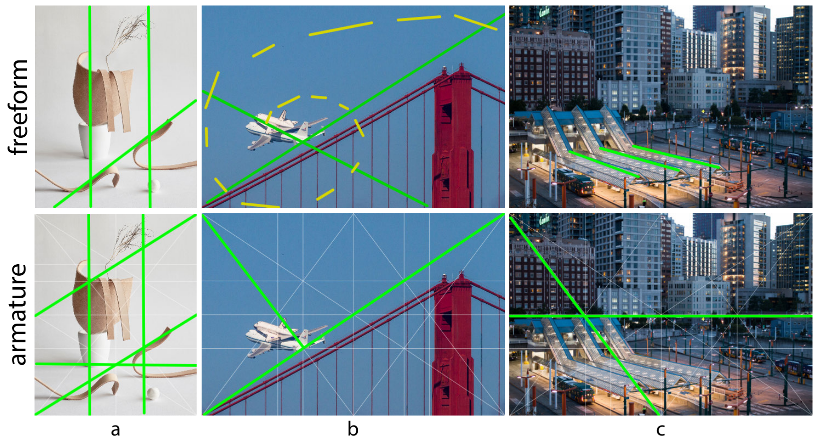


Figure 5.4: Comparing experienced photographers' **freeform annotations** with **armature-based annotations**: **(a)** participant selects nearest lines to freeform lines that does not perfectly exist in the armature, **(b)** participant draws golden spiral to represent the location of a salient object in the freeform annotation, finds nearest lines to other freeform lines in the armature that also highlight the object location, and **(c)** participant selects new alignments that were surfaced by overlaying the armature.

might not be perfectly aligned with features in the image to approximately represent these features, as shown in Figure 5.4a. For images with prominent distinct elements, participants did frequently draw lines through or bordering these elements to highlight alignment of elements.

However, while we saw some consistent line annotations, we also saw many unique annotations. The found that this inconsistency is likely due to the differences in our participants' artistic styles and their views on photographic composition. Line-based annotations are important for a variety of concepts related to composition, and experienced photographers already have many of these engrained in their mind. Despite the instruction to focus on composition of physical elements in the image, in addition to these alignments, participants annotated leading lines, strong hard edges, soft "edges" between regions (e.g., of different color), etc.

Nonetheless, many of these participants found the annotations to be a useful and fun exercise. In fact, while doing these studies, many participants indicated that they enjoyed the activity of identifying the lines within the fixed composition grid. Compared to drawing freeform, they noticed more unexpected alignments (both at intersection points and along edges) that they otherwise would not have seen. Some even mentioned possibly adjusting their images to align with lines that they highlighted rather than their current composition (we later saw this behavior in our user studies, see Sections 5.6). While these adjustments were not always deemed improvements, this suggests that even for experts who already have training in envisioning composition grids in their mind, searching for these alignments can change their perspectives in ways that might impact their photography.

5.3.2 Mechanical Turk Annotation

Our previous study confirmed that experienced photographers could identify relevant line alignments from a composition grid to a given image, but more important to our design was whether or not novices could easily interpret such annotations. We wondered if novices could also use the same mental model and leverage a composition grid to describe the composition of an image. In addition, with more annotations, we wondered if we would be able to find some consistency in their annotations to suggest a set of the "most relevant" lines to an image's composition. To collect the data to answer this question, we crowdsourced annotations on Mechanical Turk.

Mechanical Turk Task

We constructed a dataset of 500 photos for our Mechanical Turk studies. The MIRFLICKR-25000 dataset includes 25000 images from Flickr that are open under the Creative Commons license and selected based on their interestingness rating [137]. From these, we selected the 500 most aesthetically pleasing images according to Kong et al.'s photo aesthetics ranking network [167]. We removed those with extreme aspect ratios ($< .5, > 2$) or with inappropriate content, and replaced them with images from the next 100 most aesthetic images.

We created a Mechanical Turk task in which users were shown 10 of these images at random and instructed to *"Please select 1-4 lines that best describe the composition of the physical elements in the photo."* The interface was an image with the composition grid overlaid. Users used their mouse to hover over and click to select/deselect a line.

The first time a worker does a task, they are presented with a short description of the task and of composition. To avoid the task feeling too subjective and risky, we provide three example image annotations along with short explanations. Finally workers are given a short interactive tutorial on how to use the annotation interface. To complete the tutorial, users are required to select lines that match the example selection.

After the trial task, the first two annotations are easier tasks with simple compositions as further training. For these images, they are given the choice to see sample annotations and explanations. We used these tasks as validation to remove careless annotators. Per annotation task, workers would annotate 12 images (2 validation, 10 for dataset). We estimated this to take around 4 minutes, and paid $1 per task to match the minimum wage of $15/hr and used Fair Work to allow workers to inform us if the task took longer than anticipated [301].[2]

Mechanical Turk Annotation Results

We collected a total of 1004 annotation task responses across 582 workers, totaling 11961 non-validation image annotations (an average of 22 annotations per image). From these annotations, we filtered out task responses where the worker was unable to get both validation tasks correct—we counted a task as "correct" if they selected one of the lines provided in the sample annotation. This resulted in the removal of 218 tasks, for a resulting set of 9133 image annotations (~17 annotations per image).

From this dataset, we aimed to determine a set of "most relevant" composition lines for each image as our ground truth annotation. Images vary in how well they can be represented by a composition grid, therefore also in consistency in annotations. To select our ground truth lines, for a given image, we defined the score of each line as the percentage of worker annotations that included that line. After an initial batch of annotations, we empirically determined a score threshold of 0.4 for whether a line should be selected as one of the ground truth lines—this resulted in an average number of lines selected as ground truth being close to the average number of lines selected per

[2]Materials for the study can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

image in the annotations. Given this threshold, 10 images had no ground truth lines. These were also removed from the dataset, for a final set of 8966 image annotations (~17 annotations per image). For this final dataset, the average number of ground truth lines per image is 1.82 and the average number of lines annotated per image is 2.17.

With a relatively high threshold of 0.4, we see that these average numbers of lines in the ground truth annotations are quite close to those in the workers' initial annotations, meaning most lines are reaching this score threshold. Additionally, only 10 images resulted in no lines above this threshold. This suggests that there is consistency in the notion of what are the "most relevant" lines in a composition grid of a given image and that novices can perceive and annotate this relationship.

Our Mechanical Turk task also included an optional text box for feedback on the task. We received a lot of positive feedback on the task. Many workers mentioned they found the task fun (30), enjoyable (19), and/or interesting (29), and would like to keep doing more of these tasks (9). In particular, one mentioned it being helpful for their learning: *"I am working on my photography skills and this exercise was helpful for me to better understand this concept even if I wasn't entirely helpful to you."*

5.3.3 Annotation Insights

From our experienced and novice composition grid-based annotations, we learned:

- **Performing annotations can be useful.** Experienced photographers described noticing new alignments. Novice photographers found them helpful for understanding composition. Both groups found the task enjoyable.

- **Annotations are approachable.** Both experienced and novice photographers were able to perform the annotation task with reasonable consistency. Thus, we chose to pursue this path for our in-camera composition guidance.

5.4 Adaptive Armatures

Since our formative studies suggested that there exists a relatively consistent notion of prominent lines for a given image, we set out to automatically detect a set of relevant lines to provide these annotations in camera. We call this method *adaptive armatures*.

5.4.1 Heuristic Algorithm

Composition can be posed as aligning visually important elements in an image with the lines/intersection points of a composition grid [308, 313]. We use saliency to represent this visual importance. Thus, the adaptive armature should capture the set of lines that best align with the saliency map. Given an image, we score the candidate lines for the image armature such that the higher the score of a line, the better the elements of visual importance in the image are aligned to that line, and the more "relevant" they are to the image's composition.

For a given image, we compute an attention-based saliency map using Apple's built-in beta Vision library [15] (but any performant saliency estimation method could work). This map is used to vote for lines in the full armature grid. For each point $\mathbf{p}$ in the saliency map with saliency value $\mathbf{S}$, its contribution to the score of a composition line $\mathbf{L}$ is:

$$score = \mathbf{S} \cdot Gaussian(distance(\mathbf{p}, \mathbf{L})). \tag{5.1}$$

for a Gaussian kernel with a size of $1/10$ of the image's longer dimension and a sigma of $1/4$ of the kernel size. The score per line is then normalized by the length of the line. We select the top 3 scoring lines as our resulting heuristic annotation (see Figure 5.5), or *adaptive armature*.

5.4.2 Mobile Implementation

Our app runs on an iPhone 7 running the iOS 13.0 public beta. We built a basic camera app where saliency and the adaptive annotations are constantly computed in the background for the current camera image. The app has a camera shutter button and flashes the screen white when a photo is taken. For the purposes of our

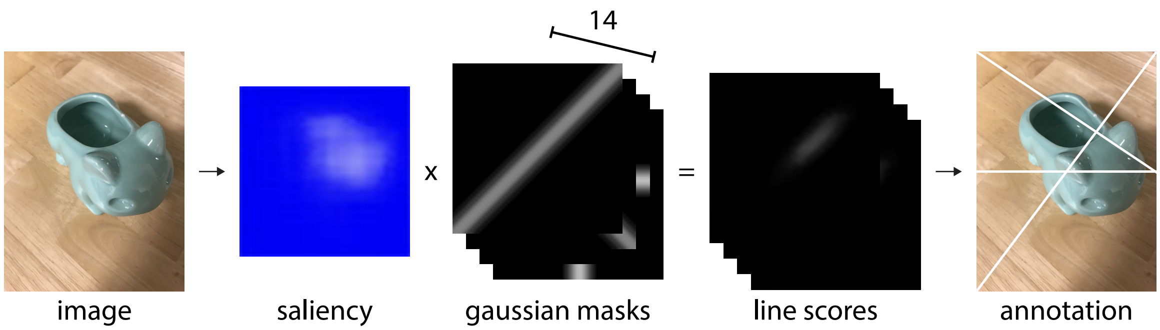


Figure 5.5: Computing the line scores to determine the *adaptive armature*. We compute the **saliency** map of the **image** and multiply that by a **Gaussian mask** per line. The sum of the resulting image is the **line's score**. The top 3 scoring lines become the adaptive armature.

studies it has no other camera functionality to encourage participants to focus solely on composition.

To reduce computation time, we precompute the Gaussian maps per line and store them as images. The score per line can then be computed by simply multiplying the saliency image by the corresponding precomputed Gaussian map and computing the sum [308]. To maintain an interactive experience, we compute line scores at discrete 0.1 second intervals. Since users need time to process the highlighted annotations, this interval strikes a balance between perceptually-constant visual update and non-distracting digestible user experience.[3]

5.4.3 Mechanical Turk versus Heuristic

We wanted to evaluate our heuristic results against the ground truth we obtained through Mechanical Turk annotations. For each image, we computed the average number of lines at the intersection of heuristic annotation and the ground truth lines divided by the number of lines at the union of these two sets of lines. The average across images of this metric is 0.38. For reference, the average is 0.11 for a random sampling of 10000 annotations—to sample these annotations, a number of

[3]Video demonstration of the tool can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

lines is uniformly sampled from [1, 4] (matching the instructions for the Mechanical Turk workers) and then each line is selected uniformly. This metric for individual worker annotations as compared to the ground truth lines is 0.47. It is reasonable and expected that this metric is lower for our heuristic method as the ground truth is computed based on the worker annotations.

5.5 Study Design Iteration

We conducted two small pilot studies and a formative user study with our tool as a part of an iterative design process. These studies helped us answer some questions about different potential interaction conditions of an in-camera composition guidance tool as well as refine our interface for our summative user study, which followed. Participants for all versions of the study were compensated $15 for their time.

5.5.1 Pilot Studies

Preliminary Pilot Study: No Guidance and Adaptive Lines

We ran a pilot study (n = 3) to evaluate our initial study design comparing two interface conditions: first no guidance, and then our adaptive composition guidance.

Participants were provided with a sheet of paper describing some basic composition guidelines to guarantee that they had a baseline perspective on how to consider composition. They were told both by the experimenter and on the sheet that they are not required to follow these guidelines. Participants were then asked to complete 6 photo tasks, focusing on composition. The first two participants did the same 3 tasks twice, once for each tool; the final did 6 different tasks, 3 per tool. After each condition, participants filled out a survey asking a variety of Likert questions (on a 7-point scale) related to their experience, and measuring the Creativity Support Index (CSI) (0 to 100) [63]. At the end of the study, they were asked to select their favorite photo per task and rate them on composition. Finally, they answered open-ended questions about what they liked/disliked about the tool and how it influenced their photo capture process. Studies were screen recorded and participants were asked to

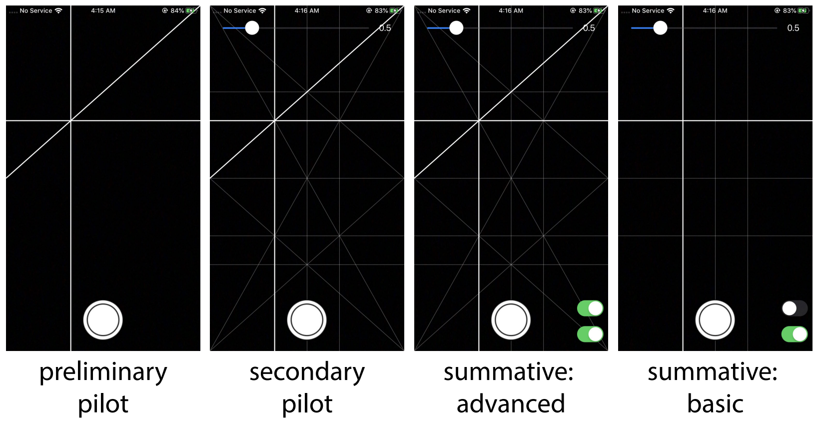


Figure 5.6: Iterations of the adaptive composition guidance. **Preliminary Pilot**: displays only the adaptive selected lines, updating every 0.1 sec. **Secondary Pilot**: displays non-selected lines at 0.3 opacity, slider allows user adjustment of update speed (defaults to 0.5 sec). **Summative**: iteration after pilot studies, top toggle switches between **advanced** and **basic** modes (just center and thirds, only two selected lines), bottom toggle turns composition visualization on or off.

think aloud.

Preliminary Pilot Study Interface

The no guidance interface had a single interactive button for photo capture and no other interface elements displayed on the camera view. The adaptive interface additionally dynamically displayed the 3 adaptive armature lines. A line was only visible when it was selected, other lines in the grid were transparent (see Figure 5.6).

Preliminary Pilot Study Results

We learned that repeating the same photo tasks resulted in learning effects, such that the second time around, participants had already explored the space of options and knew what photo they liked and wanted to take. They ended up using our tool primarily to refine the shot they already had in mind from the first condition. For the third pilot, we picked a second location and designed 3 tasks in the two locations that were comparable based on scale and environment (see Figure 5.7).

Participants gave us the feedback that in using our tool, without the context of

Figure 5.7: Corresponding photo tasks at the two study **locations**. Tasks are designed to be similar in subject and complexity and to increase in scale: **(1)** small object on a solid surface, **(2)** chair/table structure, and **(3)** facades of a building.

the composition grid, the lines that appeared did not seem to have any particular structure. Additionally, the lines would appear and disappear too quickly for them to react effectively, causing frustration in using the tool. Due to the feedback from this preliminary pilot, we chose to keep the full composition grid always visible, but faintly to avoid over-cluttering the camera view (see Figure 5.6). We added a slider that allowed users to adjust the speed of the line updates from 0.1 seconds to 3 seconds per update (at intervals of 0.5 seconds). We set the slider default to 0.5 second updates—it appeared to still be relatively responsive to movement of the camera, while at slower speeds the updates felt laggy.

Additionally, participants expressed that the adaptive highlighting helped them come up with ideas that they previously would not have thought of. We realized that this could be due to a number of factors, seeing lines of any sort overlaid on the camera view, or the adaptive nature of the lines. We hence decided to try two additional conditions: showing a static composition grid, and randomly selected lines to highlight rather than those chosen by our heuristic algorithm.

Secondary Pilot Study: 4 Conditions

We next ran another pilot study (n = 6), this time testing out all 4 conditions: no guidance, static composition guidance, random adaptive composition guidance, and heuristic adaptive composition guidance.

Participants were again provided with a sheet of paper describing some basic composition guidelines and told that they were not required to follow these guidelines. We used 4 different locations, with 3 comparable tasks per location (see Figure 5.7). All participants saw the tool with no guidance, followed by static guidance, and then half saw the algorithmic adaptive guidance first while the other half saw the random adaptive guidance. Locations were manually approximately counterbalanced. To reduce study length, we did no intermediate surveys, and participants were interviewed at the end about what they liked/disliked about the 4 guidance interfaces and how they influenced their photo capture process. Studies were screen recorded and participants were asked to think aloud.

Secondary Pilot Study Interface

The no guidance interface had a single interactive button for photo capture on the camera view. The static composition guidance additionally displayed the full composition grid at 0.3 opacity overlaid on the camera view. Our adaptive interface additionally dynamically displayed the 3 lines from the current annotation overlaid on the camera view. These lines were either randomly or algorithmically selected. The adaptive interface provides the user with a slider control to adjust the speed of annotation updates (see Figure 5.6).

Secondary Pilot Study Results

Some participants could not differentiate between the heuristic and random adaptive guidance, but those who could, expressed that the random highlighting was distracting and hard to ignore. P2 described it as *"unusable,"* and P1 said that at first, the random highlighting *"reduced trust"* in the tool. While this participant expressed later that the random highlighting allowed for more creative discovery due to the more drastic changes in suggestion, we chose not to pursue this condition further due to the otherwise negative response.

Two participants expressed that the composition grid was too complicated given their limited compositional knowledge. In particular, they were unable to use the diagonal lines due to a lack of understanding of how they are commonly used in composition. Some (3) also expressed that having a better mental model of how the adaptive highlighting algorithm selected lines would allow them to better use the tool.

5.5.2 Formative Study: No Guidance and Static Lines

The secondary pilot helped us eliminate the random highlighting condition, resulting in 3 conditions to consider: no guidance, static composition guidance, and our adaptive composition guidance. Many cameras provide the option to overlay static composition grids of different types. However, there is little understanding of how/if these help users compose better photos or if they limit users' creativity. We were curious to better understand how such static composition overlays impact how users capture photos and how that differs from how they capture photos without an overlaid composition grid, and chose to run another formative study with two experimental conditions: no guidance and static composition guidance.

We recruited 12 participants (7 male, 5 female), 22 to 36 years old ($\mu = 27$), for this formative user study. All participants went to two locations to complete 6 total photo tasks, and we counterbalanced locations and conditions. All participants performed the same tasks, conditions permitting (as we were using a public space, there were a few (5) occasions where one task was slightly adjusted due to other people occupying the space). For each task, we provided specific instructions on what should be the subject of the image. We also chose to constrain the space in which participants were allowed to move around the photo subjects. These were only stated when participants tried to walk outside of the boundaries. The experimenter brought the participant to the same starting location for each task, and tasks were always completed in the same order at a given location.

Participants were again provided with a sheet of paper describing some basic composition guidelines and told that they were not required to follow these guidelines. They were told they will be completing 2 sets of 3 photo tasks and to focus on composition—for each task, they should have a photo that they believe is well-composed. After each condition, participants filled out a survey asking them to rate their confidence in their ability to capture well-composed photos, and measuring the Creativity Support Index (CSI) [63]. At the end of the study, they were additionally asked to select their favorite photo per task and rate their composition (Likert 1-7). Finally, they were interviewed about what they liked/disliked about the tool and how it influenced their photo capture process. Studies were screen recorded and participants were asked to think aloud.

Formative Study Interface

The no guidance interface had a single interactive button for photo capture on the camera view. The static composition guidance additionally displayed the full composition grid at 0.3 opacity overlaid on the camera view.

Formative Study Results

We found that while the static composition guidance didn't improve users' opinions on the composition of their photos, it made participants feel more confident in their ability to compose photos (Mdn = 5, IQR = 4.75-5.25), versus no guidance (Mdn = 3.5, IQR = 2.75-4) [Wilcoxon signed-rank $V = 0$, $p = .005$]. We saw this boost in confidence in our interviews as well. P5 described that having the *"guidelines [made] it much easier to have the 'correct' composition."* P9 felt *"more comfortable taking fewer photos"* while subconsciously holding *"myself to a higher standard."* This was because the lines allowed the participant to precisely refine alignments before taking the photo. P10 described the lines as being *"useful because it gives you greater confidence that you are taking a good photo"* because they provided a concrete way of describing why it was better.

The static guidance also had a higher CSI than no guidance, suggesting that it provides better support for creativity [$V = 13$, $p = .04$]. During the study, P10 (who

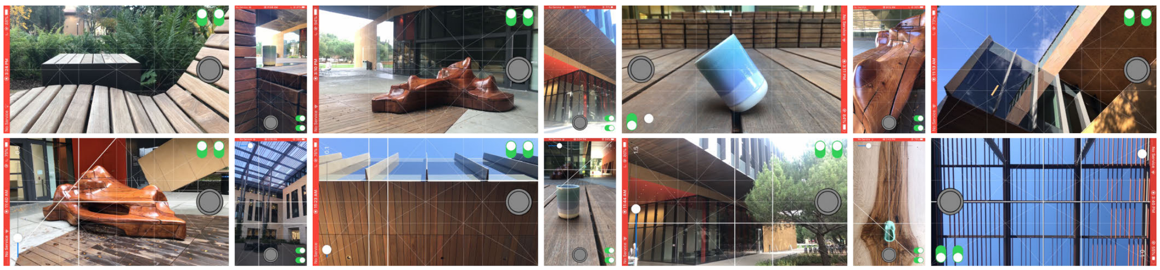


Figure 5.8: Some participant photos from summative user study. These are shown with the overlays that the participants saw while doing the user study corresponding to the two conditions: static guidance (top), and adaptive guidance (bottom). Participants often found grid lines to align to elements or edges in the image, but also sometimes used them as looser guidelines for leveling the image, splitting the image into regions (e.g., thirds), or occasionally even disregarded the grid lines.

first used the tool with no guidance) expressed that *"I already feel more creative"* while completing the first task using the guidance. The individual factors that significantly improved were Exploration, Results Worth Effort, and Enjoyment (see Table 5.1).

However as in our secondary pilot study, we again heard feedback that the overlaid grid was too complex, *"there were so many [lines] and I only knew how to use a few of them"* (P5).

5.6 Summative Evaluation

With our insights on the benefits of static composition guidance, we wondered how the dynamic nature of our adaptive armatures might similarly or differently influence how participants composed and captured photos.

5.6.1 Summative Study: Static and Adaptive Lines

We recruited another 24 participants (12 male, 12 female), 22 to 32 years old ($\mu = 26$), for our summative user study. These participants experienced two experimental conditions: static composition guidance and heuristic adaptive composition guidance.

Our summative study procedure mostly matched that of our formative study with the following changes. Since we repeatedly heard that complexity of the grid and

	Formative		Summative	
	no guidance	static	static	adaptive
	M (SD)	M (SD)	M (SD)	M (SD)
Exploration	8.8 (2.9)	*11.9 (3.3)	13.3 (4.4)	**13.8 (4.4)
Expressiveness	10.9 (3.9)	11.4 (1.9)	13.0 (4.3)	13.0 (4.5)
Results Worth Effort	10.4 (2.0)	*13.2 (1.7)	13.7 (4.3)	**14.3 (4.0)
Immersion	10.5 (3.2)	9.3 (4.6)	9.7 (5.3)	9.3 (6.1)
Enjoyment	7.9 (3.1)	*12.5 (3.1)	14.0 (5.0)	**13.8 (5.2)
Collaboration	5.2 (4.5)	6.1 (4.5)	5.5 (4.6)	6.1 (4.1)
Overall	**49.7 (11.0)**	***59.1 (10.7)**	**64.8 (20.1)**	****65.4 (20.0)**

* significant improvement over no guidance using Wilcoxon signed-rank test (within subjects)
** significant improvement over no guidance using Wilcoxon-Mann-Whitney test
(only correlation as this is between two different study populations)

Table 5.1: CSI breakdown for each study (overall score: 0 to 100, individual factors: 0 to 20). Factors are ordered in order of importance based on pairwise comparisons (orders matched for the two studies).

the sheer number of lines was overwhelming, we added a short tutorial of the app that described interface elements and explained the proportions in the composition grid and showed more examples that use lines other than the rule of thirds. This was shown immediately after the composition guidelines to all participants. We also wrote another tutorial that explained the adaptive algorithm at a high level due to the suggestion that having a better mental model of the tool would help participants use the tool in a more informed manner. This was shown to participants in addition to the description of the static composition grid before the algorithmic adaptive highlighting condition.[4]

Summative Study Interface

In our final tool, to further address the complexity of the composition grid, both the static and adaptive interfaces have toggles to switch between basic (centers and rule of thirds) and advanced modes (full armature), and to turn the lines on and off. The static composition guidance displays the full composition grid at 0.3 opacity overlaid on the camera view. The adaptive interface additionally dynamically highlights the 3 adaptive armature lines (2 lines in basic mode) by increasing the line stroke and

[4]Materials for the study can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

opacity of the selected lines. The adaptive interface provides the user with a slider control to adjust the speed of annotation updates (see Figure 5.6).

Summative Study Results

We found that users believed their photos were more well-composed when using the adaptive composition guidelines (Mdn = 5.5, IQR = 5-6) compared to the photos taken using the static composition guidance (Mdn = 5, IQR = 4-6) [Wilcoxon signed-rank test $V = 247.5$, $p = .003$]. In particular, the adaptive tool helped improve self-assessed composition of task 3 photos (static: Mdn = 4, IQR = 4-5; adaptive: Mdn = 5, IQR = 5-6) [$V = 24$, $p = .02$], arguably the most complex task due to the scale and number of lines naturally present in buildings. This was also reflected in participant interviews—when asked when they would be most likely to use the tool, many (11) specifically mentioned more complex scenes like landscapes, architecture, and with multiple objects, etc.

While we did not see a significant difference in the CSI (or for any individual factors) for the adaptive guidance over static guidance, qualitative feedback during interviews suggested that it allowed for more creativity due to its encouragement of exploration due to the generation of more ideas (15). P7 said the tool had a *"whimsical"* quality that made interacting with the tool more fun—the tool *"made me more experimental, more immersed, willing to try more things,"* whereas using the static guidance was more about coming up with an idea and just achieving it. P23 described that the *"bright lines help you think about specific lines, making it more helpful than static lines because it gives you ideas."* P0 noted that using the adaptive tool, *"I noticed more perspectives that I wasn't aware of or hadn't thought of."* In fact, P1 expressed usually relying too heavily on a specific type of aesthetics, and liked that the tool *"would allow me to explore more by challenging me to try putting salient objects in other locations."* P11 noted the *"diagonal lines gave you a chance to try different angles from traditional capture—angles that traditionally might look weird, given the perspective of the lines, instead look like a nice novel way of looking at the scene."*

Nonetheless, a few (3) did note that the grids made them focus more on aligning rather than coming up with new creative ideas. Photos from the study (Figure 5.8) show that with both static and adaptive guidance, participants were often inclined to find an alignment between the image and the grid lines. However, in some cases the image was not served by the specific composition grid and they disregarded it.

We did not see a significant difference in the participants' confidence in composing images, but again we saw support in our interviews. The tool provided *"validation, while helping reduce small motor changes"* (P1), served as good *"secondhand confirmation that their photo is actually nice"* (P5), and *"reinforced notions that I already had about composition"* (P14). P6 felt like the tool was *"guiding me,"* making the capture of well-composed images less stressful.

Comparing No Guidance and Adaptive Lines

Our summative study did not directly compare no guidance and adaptive lines. Thus results here compare our adaptive interface results from the summative study to the no guidance results from our formative study—it is to be noted that these results can only suggest correlation as we are comparing between different populations, neither of which saw both conditions. Here we found similar results to static composition guidance versus no guidance from our formative study.

The adaptive guidance increased participants' confidence in their ability to take well-composed photos (no guidance: Mdn = 3.5, IQR = 2.75-4; adaptive: Mdn = 6, IQR = 5-6) [Wilcoxon-Mann-Whitney $W = 26.5$, $p < .001$]. Compared to no guidance, adaptive guidance also had higher CSI with the same creativity factors significantly improving as well: Exploration, Results Worth Effort, and Enjoyment (see Table 5.1).

User Scenarios

We found some common ways in which our participants used the static and adaptive guidance.[5] Here, we walk through two user scenarios that detail some of these behaviors:

[5]Videos of specific examples from the studies can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

> While traveling, a user wants to capture an artistic photo of a unique landmark. She pulls up the tool on her camera and scans the space looking for interesting compositions.
>
> ***Static Guidance.*** The user checks all grid lines as she scans the space. She eventually notices two vertical thirds lines nicely aligning with parts of the structure and decides to try out this idea. She slightly adjusts the camera to place a foreground object at the bottom right thirds intersection point. She shifts the camera horizontally/vertically, placing the object at different intersection points to test out a few compositions. She picks one, refines the alignment of the verticals, and takes a photo. She has captured this idea and continues scanning the space looking for different ideas.
>
> ***Adaptive Guidance.*** The user follows the highlighted grid lines as she scans the space. Two vertical thirds lines light up, nicely aligning with parts of the structure and she decides she wants to try out this idea. As she adjusts the camera to place a foreground object at the bottom right thirds intersection point, she notices a different line highlight. After taking a photo capturing her current idea, she decides to try aligning to this newly highlighted line, and takes another photo. Again, a new line highlights, drawing her attention to a new alignment with the structure that she had not previously considered, giving her yet another idea.

We saw that overlaying a composition grid allowed participants to more easily explore a range of composition ideas because it turned the abstract concept of coming up with a new "composition" into a more concrete task of selecting elements in the scene to align (or ultimately misalign). We further saw that unexpected highlights helped focus the participants' attention on potential new composition ideas to explore. Figure 5.9 shows examples of how specific grid lines (highlighted or not) gave participants composition ideas to try. For Figure 5.9a, P16 mentioned the highlight actually inspired her to try out a composition idea that she otherwise would not have thought of: *"I want the side of the bench to align with the diagonal that was highlighted. I actually didn't see it fit*

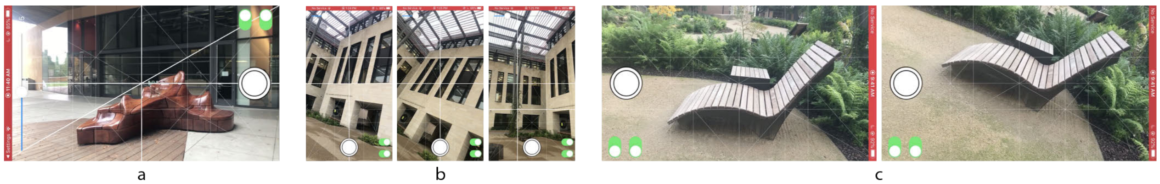


Figure 5.9: Example frames of participants using alignment to the composition grid to explore a range of compositions. Left to right: **(a)** This participant places the structure in the bottom right triangle after seeing the middle diagonal highlight. **(b)** This participant tests several ways of aligning lines in the building structure (window/edge) with the composition grid lines. **(c)** This participant switches back and forth between aligning the bench with the two diagonals.

into this corner until the line on the screen showed up."

While unexpected highlights were useful for exploring, we saw that expected highlights helped boost participants' confidence in their current composition. We saw several participants note that they were happy with their compositions because the adaptive armature matched their expectations: *"Ok I see the lines that I want lighting up, so this looks good."* (P14) and *"Oh yea, it's really nice because the lines go through the Bulbasaur"* (P22). These quotes further confirm that novices are indeed able to interpret the adaptive armature representations and use them to understand composition, as they were beginning to predict the expected behavior of the algorithm based on their composition choices.

5.7 Limitations and Future Work

While our adaptive armatures provided much promise for helping users explore compositions, here we describe some interesting directions for future work to further build on this method of dynamically highlighting composition.

5.7.1 Heuristic Algorithm

So far we have built our system around the idea of composition guidance solely based on salient regions. As the end goal of composition is to achieve some sort of visual balance (or lack of), it would be interesting to take into account additional high-level cues [154]. We explored the use of edge detection in our algorithm, but this didn't improve our consistency and made understanding the result more difficult. Thus, we chose to go with a simpler mental model for participants so as to better study the interaction in this initial implementation. As we did observe participants aligning to edges in many cases, we would be interested in further pursuing this direction in future iterations.

5.7.2 Interface Customization

Participants expressed interest in having more control over the set of composition grid lines. For example, some mentioned wanting to reflect personal stylistic preferences, while others mentioned setting a specific template overlay to constrain or match a given composition. This is a feature that would be straightforward to integrate via interactive selection of lines, or automatic extraction of templates from images using our current heuristics. These suggestions reinforce the usefulness of this mental model and the engagement it creates as an interactive tool.

5.7.3 Additional Applications

Apart from in-camera guidance, we believe our adaptive composition lines could be useful for interactive cropping tools. Likewise, appropriate saliency models [51] coupled with other grid systems [213] could provide a similar experience in graphic design tools. Other domains we are interested in include composition-based image retrieval, and other perceptually meaningful analyses of images.

5.8 Chapter Summary

Composition is an important aspect of photography. The use of composition grids is common for teaching composition. Many cameras also provide support for overlaying a range of composition grids. However, there is little support for whether composition grids are a mental model that can easily be grasped and leveraged by novices. We found that novices were relatively consistent in annotating how they perceived composition with respect to such grids. We also found that having such guidance on a camera can help users feel more confident and creative.

We explored composition guidance by creating a new kind of photographic composition guidance around our idea of adaptive armatures. We found that adaptive armatures support users both in exploring new composition ideas as well as in refining composition at capture time, aiding them in producing photos that they believe are better composed.

5.8.1 Design Goals

We close this chapter with a summary of how the design of this interface fits within our design goals (Section 3.4).

Context-Aware

When photographers describe composition, it is often with regards to relative placement of objects of strong visual attention in the image. This closely aligns with the definition of saliency, and we therefore use a gaze-based saliency to capture the relevant context of the image. Through our adaptive armatures, we aim to reflect the composition of the current camera image through highlights in the composition grid.

Encourage Exploration

By showing grid lines, we hoped that users could more easily think about possible composition options, and that the highlights especially, might help them understand their current composition or consider new composition options, while still attending to other aspects of the photo. These high-level grid representations made it easier to explore the space of possible compositions and also highlighted new composition ideas for participants to try [271].

Maintain Flexibility

Finally, the highlights are meant to represent the user's current composition choices, rather than impose a new composition. Users have flexibility in how they respond; they can choose to better align to highlighted lines, align with different lines, or intentionally misalign elements relative to the grid. However, by choosing the harmonic armature as the underlying grid, we were hoping to be able to support alignment to a wider range of compositions.

Chapter 6

Declutter

For decluttering, the important context to consider is what will draw the viewers' attention. Exploring how to declutter an image involves considering how elements of the image contribute to the overall story, and what might focus or distract. Creators should have the flexibility to decide how they want to present the story. As demonstrated in our walkthrough (see Figure 6.1), changes made to address unwanted clutter can also often influence composition (and vice versa). The photographer still wants to use the rule of thirds, but upon noticing the distracting clutter, switches to placing the subject at the top right thirds to remove the clutter from view.

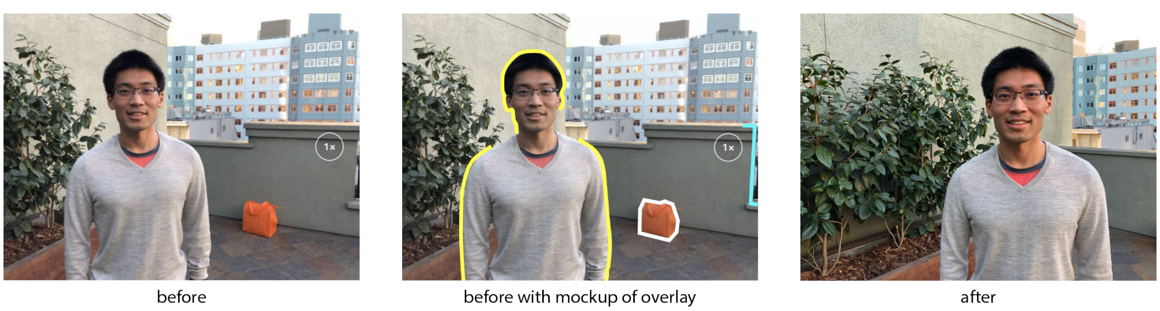


Figure 6.1: Here we see photos from our walkthrough (Figure 1.2) from **before** and **after** the photographer considered composition. In the middle we show a **mockup** of what the photographer might see upon launching our tool. The decluttering tool further emphasizes the orange bag as well as highlights the edge of the wall on the right border of the image, encouraging the photographer to recompose the image to remove this clutter.

6.1 Introduction

Photographers employ a wide range of visual techniques to communicate their intended story or narrative to a viewer [36,41,151,234]. The "story" in this sense of the word is often very open. It's more about capturing a moment or invoking an emotion, perhaps a sense of tension or inspiration—in general, it is just about capturing what caught the photographer's eye, e.g., a subject performing an action in their current environment. In fact, it is desirable for the photo to leave a gap in the narrative such that aspects are left up to the viewer's imagination. Two important factors to telling a clear story are picking a strong subject, and keeping visual distractions to a minimum.

Amateurs frequently make the mistake of taking too few photos in the moment, relying on editing to improve their photos. However, changes that can be made at the editing stages are limited, and often many mistakes cannot be fixed at all without returning to the photo location. Having unwanted background clutter is one such mistake that can be harder to fix without more significant changes, and can be particularly frustrating as they distract from the main focus of the image. While still in the scene, the photographer has many more options to reframe the image to clarify the subject and remove distractors, helping to improve these images and better direct viewers' eyes.

Experienced photographers tend to capture many photos of a given scene: they know how to consider different options, e.g., composition, lighting, or pose, and recognize the challenges of not having the option to physically move the camera or elements in the image upon editing. We see similarities in this photographic process and the design process—photographers are essentially iterating on their design (or image) in the camera as they consider these different aspects of photography and storytelling.

Designers have long known the benefits of quickly testing many ideas. Bill Buxton describes the benefits of ideating through sketching due to its flexibility "enabl[ing]

Much of the material of this chapter is from the unpublished manuscript, *"Dynamic Guidance for Decluttering Photographic Compositions."* by Jane L. E, Kevin Y. Zhai, Jose Echevarria, Ohad Fried, Pat Hanrahan & James A. Landay.

ideas to be explored quickly and cheaply" [49]. Designers use sketching to externalize ideas that are still vague in their minds. In doing so, they can spot potential unexpected issues and refine or be spurred to explore new ideas based on these observations [173, 271]. Similarly while taking photos, it is cheap to iterate and generate more "prototypes" or photos in the moment. The photographer can test out ideas just by moving the camera around in space, or by taking photos and immediately reviewing them, and similarly notice any issues like unexpected clutter. Upon leaving the location, these photos become "high fidelity prototypes"—the photographer is more committed to these photos as potential changes are limited and more expensive.

One significant difference between sketching ideas and trying them out in the camera is this level of fidelity. At any point in time, a photo has complete detail. Depth of field can enable some blurring of the background, but each pixel of the image is pigmented, every object in the frame is captured. A sketch on the other hand is a selective representation, an abstracted view of the idea or concept being explored. Removing these low-level details allow the designer to more quickly explore a broad range of high-level concepts. Specifically in photography, it can be easy for a photographer to be too focused on the primary subject of a photo and miss objects immediately surrounding it in the background. Abstracting the image can help them view the image as a whole rather than focus on perfecting the subject's pose or expression.

We are interested in designing camera interfaces that can encourage users to incorporate these exploratory stages of the design process into their photographic process. In particular, we wondered how we might be able to bring some of these benefits of sketching to photography to promote this behavior of intentional exploration, and if that might help users notice unexpected mistakes, such as unwanted clutter in their photos. We describe our process towards designing an abstracted annotation of a photo and study how that influences how users address capturing decluttered photos.

Specifically in this chapter, we present:

- our **design process** for determining what an abstraction overlay for decluttering might look like,

- an **algorithm** for visually annotating potential clutter by highlighting relevant edges around salient objects and around the image border,
- an **interactive in-camera app** that shows this edge highlighting overlay to users as decluttering guidance, and
- a **user evaluation** comparing this overlay to a grayscale overlay, a baseline method that many photographers currently employ for decluttering photos. The evaluation shows that the tool was helpful for making users more confident in their ability to take a clear and uncluttered photo.

6.2 Related Work

For an in-depth discussion of related work on capture-time guidance, see Chapter 2. Here we focus on work specific to decluttering and abstraction. Specifically, we contextualize our work of designing an abstraction-based camera overlay within the most relevant work in image manipulation and camera guidance.

6.2.1 Image Abstraction and Simplification

Graphics researchers have taken a non-photorealistic rendering lens on designing algorithms to create stylized image abstractions [77, 155, 303]. Many have taken the approach of trying to realize them as approximate image illustrations with strokes and colored regions, employing a number of edge detection algorithms for extracting lines for the strokes [31, 53], and smoothing filters or superpixel algorithms for quantizing the color [1, 93, 278]. DeCarlo and Santella consider visual perception through tracking eye movement to determine a structural hierarchy of the image for understanding the importance of different lines in the simplification process [77].

The goal of this line of work tends to be to serve one of two purposes, generating an artistic result, or reducing data for visual communication [155]. Both are useful for our goal of designing visual overlays that are friendly for users to process interactively. Additionally, while these methods focus on generating output from existing media, Winnemöller et al.'s method performs at interactive speeds and produces temporal results, making it appropriate for interactive use on a live camera feed [303].

6.2.2 Image Decluttering

Image processing methods also enable removing clutter using a variety of post-processing techniques. A photographer can choose to drag a slider to incrementally remove automatically detected distractors [100], or select a region with clutter to be automatically filled based on the surrounding content [19]. A photographer can also choose to segment out the foreground of a portrait to place it on a dramatic black or custom backdrop [251], a feature also available directly in-camera on some commercial phones (i.e., "Stage Light" in iPhone's Portrait mode). Nonetheless, these methods aren't intended to train the user to be more aware of background clutter while they are taking photos. Additionally, they all suffer from potentially jarring artifacts since they rely on pixel manipulation rather than adjusting the real physical objects in the image.

6.2.3 Capture-time Guidance

As discussed in Chapter 2, a range of contextual in-camera guidance exists both in research and in commercial cameras. Here I will describe a visualization feature that is most relevant to our concept of abstracting the image for the purposes of decluttering. Focus peaking helps users assess the quality of their image by helping them determine if their intended focal region is indeed in focus [7, 266]. Specifically, it does this by highlighting edges that are in focus in a bright color, which it determines by finding areas of high contrast. This type of feature is mostly reserved for cameras with physically interchangeable lenses rather than point-and-shoots or phone cameras. This is because the purpose of this visualization is to help quickly refine the focus especially when using lenses with a shallow depth of field, resulting in a very narrow area of focus [7]. Rather than providing assistance to improve accuracy and refinement, we are interested in understanding how similar overlays can be used for promoting creative exploration and understanding of how to apply higher level photography principles.

6.3 Abstraction Design Process

In this section, we describe the steps we took to design our abstraction interface for decluttering images. Initially we hypothesized that abstracting an image could help evenly spread the photographer's attention across the image, effectively drawing the photographer's attention away from the details of the main subject to other areas of the image.

Through our design process, we aimed to answer two questions: (1) will an abstracted visualization be effective in encouraging the user to see parts of the image outside of the main subject? If so, how does this make them change how they capture the scene? and (2) what is the right visualization to use for a photo "abstraction" to best invoke this type of awareness?

6.3.1 Wizard-of-Oz Prototype

To test out the concept of abstraction guidance, we started with a low-fidelity Wizard-of-Oz (WoZ) prototype where experimenters manually drew abstraction overlays. These abstraction overlays took shape as rough outlines of the objects in the scene, mostly approximated by basic 2D geometric primitives (see Figure 6.2 for a few examples).

We informally tested our low-fidelity prototype with 19 participants. Two experimenters were involved in the prototype testing, and when possible, both were present. We ran the study using the iPhone's default Camera and Photos apps. Participants were asked to stage a scene of a person (for convenience, this was often one of the experimenters) interacting with an object of their choice, in order to have a more direct concept of "story" in the photos. We then handed them the phone to frame and take a photo of the scene. After capturing a first photo, one of the two experimenters drew the abstraction in dry erase marker(s) on a transparency and taped it to the phone. The experimenter would return the phone and ask the participant to review the photo with the overlay. The participant would then be given an opportunity to take a new photo of the scene or stick with their current one. We would then apply a new transparency and draw the corresponding abstraction overlay for the new photo

Figure 6.2: Two pairs of photos (top) from two participants from our WoZ prototype: the **initial** photo and the **final** photo upon seeing the photo with the transparency overlay (below). Left to right: **(a)** The participant notices the clutter on the bookshelves, and rotates the camera and changes perspective to focus attention more on the action of the subject entering the office. **(b)** The participant notices the clutter on the tables and wall in the background, and tilts the camera angle down to focus attention on the action of the subject writing on the paper.

for the participant to review. While we didn't ask if they wanted to take another photo after this review step, occasionally (6) participants would ask if they could take a third photo and we would repeat the process above, presenting them with a third overlay. Participation was voluntary and no compensation was given.

Prototype Insights

From our WoZ prototype studies, we saw promising signs that the participants noticed high level changes to make to their photograph. It particular, it showed promise as assistance for decluttering. In all but one case, participants noticed clutter in the background upon seeing their photo with the abstracted overlay. In the one case where the participant did not make adjustments to address clutter, a whiteboard mostly filled up the background of the image, and was involved in the action in the image and therefore did not contribute to cluttering the image, but instead helped tell the intended story.

Participants tended to either adjust the camera position/angle (17) such that the clutter was no longer in frame, or move the clutter (3) out of the scene (or a mix of both). Figure 6.2 shows two such examples. In both cases, in removing clutter, the participant is also more intentional about framing the elements that are key to the story aspects of the image.

Moving the camera instead of the objects in the scene was often more practical because the objects could not be moved. Participants came up with a range of tactics for adjusting the camera to address clutter. In one situation the participant positioned the subject's head to block the unwanted clutter. Several (5) even discovered that zooming in by moving closer to the subject allows them to more easily remove/block clutter and focus on the subject, a technique often taught in photography resources. In doing so, participants additionally changed the overall compositions of the images. In a few cases, they chose to change orientation from portrait to landscape (6) to achieve a decluttered composition that they preferred. One participant noted that the first overlay made her realize that she hadn't achieved the composition she intended to because she was focused on other aspects of staging the scene. This initial photo featured the subject standing at the center of a portrait photo, soda in hand. In her second photo, she focused on capturing the intended composition, shifting the camera to the left to frame a less centered composition. Upon reviewing this second image and overlay, she again reconsidered this intended composition and rotated the camera to landscape to further emphasize the off-centeredness of the composition.

Note that while these participants' behaviors matched what we hoped for, there were limitations to our study design such that they cannot be directly mapped to how a user might respond to seeing this style of overlay interactively in the camera. The user takes a photo before seeing the overlay, and the process of reviewing (with or without) the overlay can influence the photographic process. To test the interactive experience, we needed to find a way to automatically generate these overlays.

6.3.2 Visualization Designs

Given the observations from our low-fidelity prototype, we were motivated to continue with this concept and move onto the step of answering the second question: What is an "abstraction" of a photo? What should such a visualization look like, and how might we implement them? In particular, we saw that the abstraction overlay seemed most helpful for the purposes of noticing unwanted clutter in an image, so we decided that we would target designing an abstraction overlay that provides decluttering guidance.

Decluttering Principles

To answer our questions on how to design an abstraction overlay for decluttering, we first wanted to understand how photographers think about directing the viewer's attention for effective storytelling [41, 110, 244].

Contrast is key to directing attention [110]. Our eyes are drawn to regions of high contrast. For the most contrast, the photographer should place light objects on a dark background, or dark objects on a light background. It is good practice to have this contrast around the subject as it will help make the subject distinct from the background. The contrast will clarify the story and declutter the overall image. In this chapter, we will refer to this as **subject-background separation** (SBS). In art, this is more commonly referred to as the figure-ground relationship, a Gestalt Psychology principle [110]. We use subject-background separation in our work to make it easier to relate the concept to more familiar terms.

On the other hand, contrast in other regions especially the border of the image, will distract, causing the eye to be attracted away from the focal subject. In particular, contrast near the border of the image can draw the viewer's attention outwards rather than within the image—we call this **image border flicker** (IBF) [110]. Again, this is better known as edge flicker, but we chose to specify "image border" to differentiate it from the term "edge," as "edge" is often used in computer science with regards to edge detection and identifying segments of sharp discontinuities [31, 53].

Given these principles, we wondered what annotation methods photographers currently used for highlighting clutter. Photographers recommend a number of

Figure 6.3: A few methods employed by photographers for highlighting contrast as applied on this painting by Emily Friant (left). Left to right: Original painting; A blurred and higher contrast version of the image, a representation of what it might look like to squint at the image; A grayscale version to focus on contrast without aspects of color; Glover's outline annotation such that areas where the contrast is possibly too low between the subject and background are shown as gaps [110]. Painting by Emily Friant.

methods to be able to more easily see the contrast in an image (see Figure 6.3). These range from squinting at the image to better focus on the contrast with low-level details blurred, to viewing the image in grayscale to better focus on contrast in the absence of color, to explicitly outlining boundaries along which there is clear contrast between the subject and background [41, 110, 244].

Figure 6.4 shows two examples of images that do not satisfy the decluttering principles, along with approximations of Glover's suggested outlining to emphasize contrast around the subject [110]. We note that while this overlay is helpful for identifying potential issues along the subject-background boundary, it does not help to draw any attention towards potential clutter to address along image borders.

6.3.3 Abstraction Visualization Options

Inspired by this idea of using outlines to highlight contrast and the lack of contrast, we hoped to recreate this outlining as an overlay directly in the camera, also extending it to contrast along the image borders (see Figure 6.4). We looked at different approaches

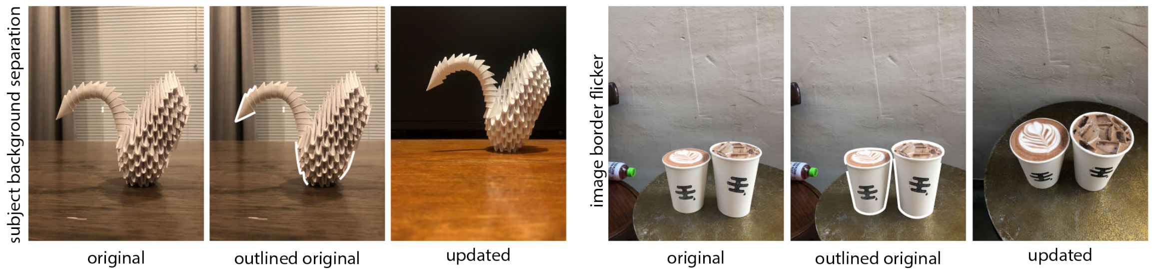


Figure 6.4: Two sets of example images addressing the two decluttering principles. For each set, we show the **original** photo that has poor **SBS** (left) or distracting **IBF** (right), the **outlined original** with the subject outlined where there is good contrast with the background, and an **updated** photo where the decluttering principle is addressed. On the left, the texture and color of the origami swan blends in with the blinds in the background, creating unclear SBS in the **original** photo. However, against the darker backdrop in the **updated** photo, the outline of the subject becomes much clearer. On the right, the bottle and dark corner of the chair against the light wall along the left border of the **original** photo cause distracting IBF. Adjusting the camera angle removes the distractors from frame in the **updated** photo, and even improves the SBS.

to executing on this concept as one potential direction to pursue for our abstraction overlay. This concept leads us to two components: a method of line drawing to determine the potential outline, and a method for considering location context relative to the subject/image frame to determine which lines in the image are relevant to the decluttering principles (e.g., if they are along the subject-background boundary, along the image border, or neither).

We additionally were inspired by a line of research in non-photorealistic rendering to generate stylized image abstractions [77, 155, 303]. Specifically, since we knew our goal would be to display the overlay interactively in the camera, we focused on Winnemöller et al.'s real-time and temporally coherent implementation of image abstraction. These papers break down the abstraction process into two components as well: a method of color flattening to smooth out detailed regions, and a method of line drawing.

Combining components from these two abstraction ideas, we have three components in total to consider: line drawing, location context, and color flattening. We additionally give ourselves line color (for the purposes of color-coding lines based on

Figure 6.5: Overlay options as presented to design survey participants for videos. Alongside the video with the set of overlays, we showed still frames of the beginning, middle, and end of the video with the same overlays. This video clip aims to show an example of someone noticing clutter in the background (balloons and bike in the top image) and choosing a new background for the photo. In the middle row, there is a highlight on the right from the light that is removed for the final image on the bottom.

their location context), and image darkening (to provide more contrast against the line drawings) as parameters to consider in designing abstraction overlays.

Adjusting these parameters, we designed a range of potential overlay proposals, and narrowed it down to the 6 in Figure 6.5 to study further:

(a) color flattening + line drawing (in black)

(b) color flattening

(c) line drawing (in white)

(d) darkened image + line drawing (color-coded: yellow is subject, white is background, cyan is image border)

(e) d without lines within subject

(f) e with white lines removed

Note: overlay (a) features the full image abstraction from Winnemöller et al. and (b) is just the color flattening (or region smoothing) component.

Design Survey

We ran a Qualtrics survey with 29 participants to try to understand if these overlay visualizations were interpretable by novice photographers, and if there were strong preferences between the overlay options.

The survey consisted of 12 pieces of visual media (8 photos, 4 videos) shown with the set of different abstractions overlaid. For a single photo, participants would see a row of images, first the original photo and then the 6 overlays. For a video, participants would see a concatenated video including the original video side-by-side with the set of overlays. The video would automatically loop, but participants also had controls to pause and play. Along with the videos would be 3 rows of still frames from the video to allow participants to compare the overlays across a few frames. We chose to also include a few videos to show the process of making adjustments while framing a photo, to mimic how the overlay might appear if being used in the camera. Figure 6.5 shows a set of these overlays for a video as they were presented in the survey.

Participants were presented with a small amount of training describing the two decluttering principles (see Section 6.3.2). They are also informed of our overall goal—to determine which overlay is best for recognizing the two decluttering guidelines. For each of the 12 photos/videos, participants were asked:

- Which overlay helps you to best determine if the image has good **subject-background separation**? (If you don't have a preference between a few similar overlays, feel free to pick up to 3)
- Which overlay helps you to best determine if the image has good **image border flicker**? (If you don't have a preference between a few similar overlays, feel free

overlay	SBS	IBF	total (weighted total)	% (weighted %)
a	83	29	112 (69.5)	10.4% (10.0%)
b	100	38	138 (102.5)	12.8% (14.7%)
c	112	59	171 (116.0)	15.9% (16.7%)
d	87	142	229 (137.2)	21.3% (19.7%)
e	64	155	219 (124.7)	20.4% (17.9%)
f	73	134	207 (147.2)	19.2% (21.1%)
overall	519	557	1076 (696)	100% (100%)

Table 6.1: Breakdown of design survey results: count for SBS, IBF, and total per overlay and overall. Since participants were allowed to pick up to 3 overlay choices per question, the weighted totals are computed by dividing a vote for an overlay by the total number of choices selected for that question.

to pick up to 3)

- (Optional) You will be asked to explain your preferences at the end of the survey, but feel free to explain any specific thoughts you have based on this image/video here.

Finally at the end of the survey, we asked participants to summarize their choices: "Please provide a brief explanation for your choices—why did you find these overlays most helpful? Are there specific characteristics of the overlays that you like (e.g., line drawing, color flattening, image darkening, or color)?" We hoped to gain some understanding of whether or not they had a general feeling of which overlays were helpful (maybe for different scenarios and considerations), and why. Participation was voluntary and no compensation was given.[1]

Design Survey Insights

Table 6.1 presents the results from our design survey. The results do not show a clear-cut "best" overlay design. However, overall (d)–(f) were more popular at 21.3%, 20.4%, and 19.2% of total overlay selections made, respectively. These each had the darkened image with selective color-coded line drawing based on their location context in the image. Counts show that the participants found (d)–(f) especially helpful for

[1] Materials for the study can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

identifying IBF, whereas (a)–(c) were overall more helpful for identifying SBS.

Both of these conclusions were also supported by the qualitative feedback from participants' general impressions. Participants were quite thorough in their explanations of their interpretations of the components of these proposed overlays. One participant very clearly summarized the potential reasoning for this: *"if the most important thing for subject background separation is the contrast, then the overlays should try to mute lines as much as possible. If the subject is still well separated, then you know you have the contrast. For edge flicker, it almost seems like the opposite. If there's no contrast, then your eye doesn't notice objects on the border. So if an overlay draws outlines of those objects or lights them up, then they're easier to notice"* (P10).

Overall, almost all (23) participants expressed interest in some form of line drawing. Participants noted that the edges helped to define objects (18) and that they were helpful for noticing edges around the image border (5), supporting the observation shown by the overall counts on (d)–(f) being more useful for addressing IBF. Therefore we decided line drawing had to be part of our final overlay design. Additionally, some are already somewhat familiar with selective outlining of edges from focus peaking features in commercial cameras, which selectively highlight edges in the image that are in focus [7, 266].

Many (10) participants also mentioned that the darkening of the image was particularly helpful for seeing the lines due to the contrast, but with the caveat that it made the original image harder to see. Participants (4) mentioned that color-coding was helpful, particularly accentuated in videos (P29). However, they also noted that outlining everything can be noisy (9), and especially did not want too many lines within the subject (6).

Further discussing the latter observation that (a)–(c) were more helpful for SBS, several (8) participants specifically noted that the color flattening was helpful for noticing SBS. They mentioned the smoothing of colors being helpful *"because it essentially simplifies the image and makes it easier to analyze the major color contrasts"* (P27) and noting that it helped in being able to *"dissociate the images from what I expect to see into what the colors actually are"* (P24). However, a few (3) noted that this flattening caused blurring that made the subject unclear. Both color flattening and image darkening made the image less clear, so we wanted our final overlay to only have one of these components. Since many participants specifically noted that the darkening was crucial for benefiting from the line drawing, we chose the image darkening component over color flattening for our abstraction overlay implementation.

Given the qualitative feedback, (f) appeared to be the best candidate for our overlay—it included the color-coded line drawings to help define the subject and possible distractors along the borders, while being less noisy than showing all outlines. However, since (d) was the most popular based on overall counts, we decided to go with a hybrid approach for our final abstraction overlay, enabling participants to switch back and forth between showing all lines including those within the subject and background for additional context and hiding these extra lines for minimal distraction.

6.4 Implementation

Results of both steps of our design prototyping process made us hopeful of the potential of an abstraction-based overlay. However, we had yet to try these overlays in an interactive manner. To do so, we needed an implementation that would run interactively on a phone. In this section, I describe our final interface along with the implementation and algorithm that I implemented to enable it.

6.4.1 Interaction

Based on our learnings from the design survey (Section 6.3.3), we chose to go with a design inspired by a combination of overlay options (d) and (f). Thus, our final abstraction overlay is a context-aware line drawing.

To implement this, our final camera tool has 3 layers: the camera view, a black layer of varying degrees of opacity, and the outlines layer. As shown in Figure 6.6, by default the black layer is opaque (opacity is 1.0) and outlines are color-coded and dependent on their location context, meaning only the edges most relevant to our two decluttering principles (Section 6.3.2), subject-background separation and image border flicker, are visible. As a reminder, the lines are color-coded such that lines

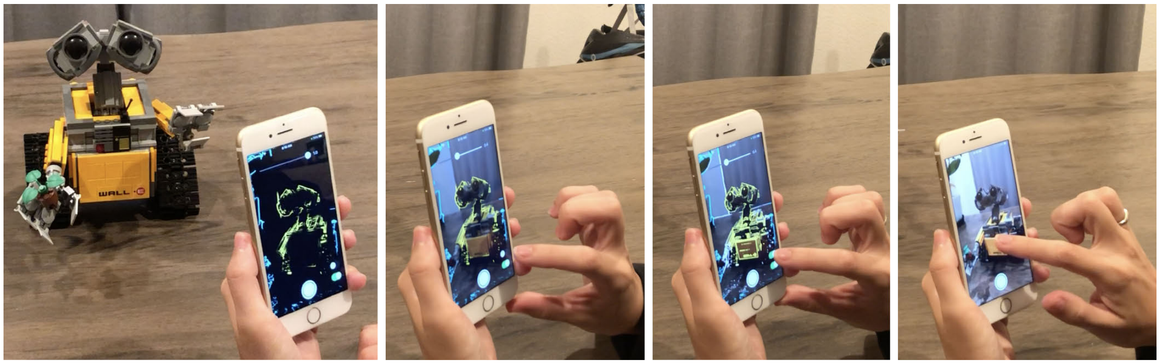


Figure 6.6: Here we show a few states of the decluttering visualization on the phone. Left to right: The **default visualization** shows the edges around the detected subject border as well as around the image border over an opaque black background. The user can choose to **reduce the opacity** of the background (here the user adjusts the opacity from 1.0 to 0.4) to show some of the camera view. The user can also toggle on **all edges** to see edges within the subject and background as well. Finally, the user can either use the bottom toggle or touch anywhere on the screen to **hide the visualization** completely.

within and immediately around the subject are yellow, lines along the image border are cyan, and remaining lines in the background are white.

Lines are color-coded to help users more easily interpret the edges. A solid and defined yellow outline of the subject would mean that there is likely good SBS. Gaps along this edge might signal a lack of contrast—this could mean that the subject is blending into the background, or there are objects directly around the subject that are interfering with the clarity of the subject border. The presence of many cyan edges suggests there might be noise and clutter near the image border that could also take attention away from the subject. The additional white lines help to complement either yellow or cyan lines. These are especially helpful if the main subject is incorrectly identified by the saliency algorithm.

In the default state, the app shows the yellow edges around just the subject and cyan edges around the image borders over the solid black background. The user can adjust the slider at the top to adjust the opacity of the black layer, bringing in more or less of the image color. They can use the toggle to bring in the remaining edges in the subject and background. They can also turn off the visualization by holding a finger down anywhere on the screen or using the toggle at the bottom right. We found in the composition study (see Section 5.5) that even though we provided such a toggle, people rarely used it to turn the visualization completely off. We therefore added the ability to more easily switch back and forth between showing and hiding the visualization by touching anywhere on the screen.

We choose the most abstracted form of the overlay (solid black, minimal edges) as the default as we imagine users starting in a more exploratory stage. As the user refines the image, they can reduce the opacity or bring in more edges to draw attention to lower level details.

6.4.2 Mobile Implementation

We built our overlay tool as a modification on top of the iOS app that we implemented for the composition project (see Section 5.4.2). We maintained the same base app—a basic camera app with just a camera shutter button and no other functionality to allow participants to focus solely on creating decluttered compositions during the user study. To generate the overlays, the edge detections (for line drawings) and saliency maps (for determining location context) of the current camera image are continuously computed in the background.[2] This app requires iOS 13.0 or higher.

6.4.3 Context-Aware Line Drawing Algorithm

Figure 6.7 walks through our algorithm for generating our abstraction overlay. Given an image, our tool detects edges throughout the image for the line drawing [14]. However, we want to be able to determine the relevant context in order to focus on edges related to SBS and IBF. In particular, we need to identify a border around the subject. We realized we could estimate the image subjects using object-based saliency maps [15]. We used both algorithms off the shelf. For determining appropriate parameters for the edge detection [14], we iteratively tested parameters to achieve a

[2]Video demonstration of the tool can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

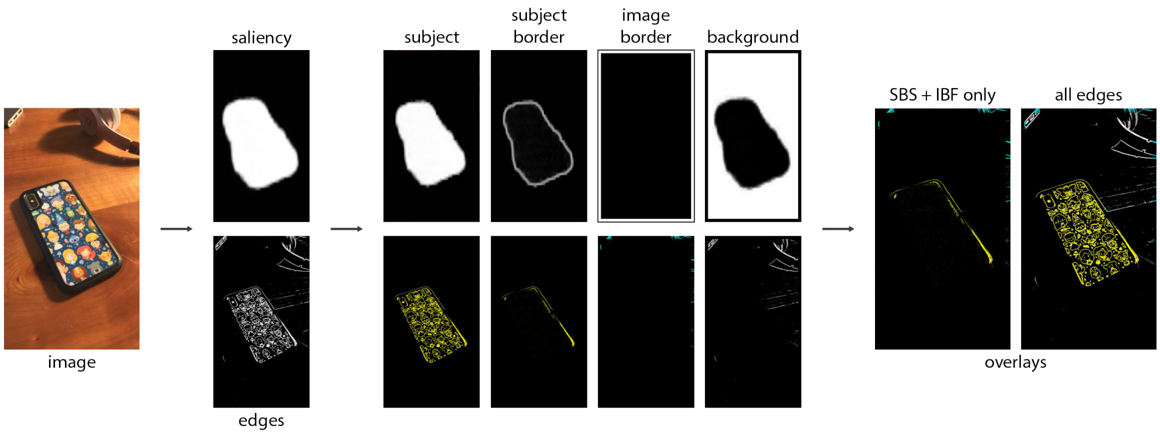


Figure 6.7: The breakdown of components of our abstraction overlay algorithm. On the left we have the original **image**. Next, we show results of computing the **edges** for the overall line drawing, and the object-based **saliency** map for location context. The next set of images are the intermediates for generating the overlay. On the top, we have the masks based on location context, and on the bottom, the edges masked and color-coded appropriately based on the context: yellow for **subject** and **subject border**, cyan for **image border**, and white for **background**. At the far right we have the two versions of the final abstraction **overlay**: showing the subset of edges relevant to SBS and IBF, and then showing the full color-coded line drawing.

balance between having enough definition in the edges and having too much noise. These are different from those we used for composition (as described in Section 5.4.2) as they try to identify entire objects rather than general regions of visual attention.

Given this saliency map, we segment the image into regions describing the subject, subject border, image border, and remaining background. These segmentations are used to classify and color code the edges: yellow designates edges within and around the subject, cyan for edges around the image border, and white for the remaining background edges. We merge these to form two edge-based overlays: one showing all edges color-coded, and the other only showing the relevant edges around the subject and image borders.

Figure 6.8: Participants were asked to choose their own subjects for each task. Here are a few example participant photos at each **task** scale: **(1) small**, **(2) medium**, and **(3) large**.

6.5 User Evaluation

We wanted to study how users would react to our abstraction guidance tool. We conducted a small formative pilot study to inform our summative user study design. In particular, in running our pilot, we also wanted to understand if a no guidance interface was a reasonable baseline to compare our tool against. Participants for both the pilot and larger summative study were compensated $15 for their time.

6.5.1 Study Procedure

Due to the COVID-19 pandemic, we had to run our studies remotely. From our composition study (see Section 5.5), we liked how the difference in scale resulted in a range of complexity while composing. To modify this task design to work for remote studies, we decided to keep the structure of having 3 tasks per condition at different scales (small, medium, and large), but instead of specifying the subjects, we asked that participants choose their own subjects to photograph. Thus, similar to the summative user study for the composition guidance, participants completed 6 photo tasks: 3 at each of 2 locations of their choosing, using the baseline condition at one location and the abstraction guidance at the second. Figure 6.8 shows a few example images to demonstrate the scale of images chosen by the participants. Since the locations were all different, here we just counterbalanced condition to avoid biases from learning effects.

We ran these user studies over Zoom, asking the participant to adjust the webcam when possible to keep their photographing within view. Additionally to prep for the study, we confirmed beforehand that the participant had a phone running iOS 13 or higher (or found a way to drop off a device in a socially distant manner outdoors). We distributed the app using TestFlight. After getting the participant's consent, we started by walking through having them install TestFlight and subsequently our guidance tool. As additional preparation, we also stepped through using the iPhone screen recording functionality with the microphone to ensure sound was also captured as we asked participants to think aloud as they captured photos for each task.

Participants were provided with a document describing the two decluttering principles (see Section 6.3.2 as training for how to think about decluttering their photo compositions). They were told that following these guidelines would be helpful in telling a clear story in their photos, but also that these guidelines are just to provide some possible perspectives to consider and that participants are by no means required to follow them (e.g., if communicating their story involved intentionally having the subject blend into the background they should do so). After reading the document, we asked that the participants briefly describe the principles in their own words to confirm understanding. For each photo task, we asked participants to focus on the overall clarity of the story. We encouraged them to explore the process of framing the image, but to limit each task to around 1-2 minutes. We quickly walked all participants through the basic (no guidance) camera app. If they were using the tool condition first, they were additionally provided with a brief tutorial describing the overlay.

After each condition, participants were asked to complete surveys with a number of Likert questions (on a 7-point scale) about their experience using the tool along with the Creativity Support Index (CSI) questions (0 to 100) [63]. Following both conditions, they were asked to favorite a single photo per task (for a total of 6), and asked to rate these based on each of the two decluttering principles as well as whether or not they liked the photo in general. We ended the study with open-ended interviews asking about what they liked/disliked about the tool and how the interaction influenced their thought process as they took photos. Again these aspects of the study matched that of the composition study, with questions adapted to highlight decluttering principles rather than composition.[3]

6.5.2 Pilot Study

We ran a pilot study (n = 5) to test this study design in a remote setting. Participants (3 male, 2 female) were 23 to 31 years old ($\mu = 29$). For these pilots, we compared our tool to a no guidance baseline condition.

Pilot Study Results

We found that overall, this study design worked reasonably well in the Zoom environment. However, even with just 5 participants we were seeing significant results suggesting that our tool was preferable to no guidance for these photo tasks. In terms of the tools influence on the process, participants felt more confident in their ability to capture a clear photo using the abstraction overlay (Mdn = 5, IQR = 5-6), versus no guidance (Mdn = 4, IQR = 4-4) [Wilcoxon signed-rank test $V = 0$, $p < .05$].

When evaluating their favorited photos per task, participants did believe that the photos captured using our tool had better subject-background separation (Mdn = 6, IQR = 4-5), than those captured using the no guidance baseline (Mdn = 4, IQR = 2-5) [$V = 0$, $p = .003$]. Participants found that the tool helped them achieve these more clear photos (no guidance: Mdn = 4, IQR = 2-4; tool: Mdn = 5, IQR = 5-5) [$V = 0$, $p = .001$]. They also liked their photos more when using the guidance tool (no guidance: Mdn = 5, IQR = 3-5; tool: Mdn = 6, IQR = 5-7) [$V = 0$, $p = .003$].

Thus, even though we did not find significant changes in CSI, we decided that we should compare against a baseline that provided a little more assistance. As we described earlier, photographers will sometimes use a grayscale display in their current practice, to help emphasize contrast in order to consider overall clarity and decluttering. We therefore use grayscale as our baseline condition. It is an active method employed by photographers and thus in some ways does encourage novices to see the image in the ways that expert photographers do.

[3] Materials for the study can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

6.5.3 Summative Evaluation

We ran remote studies over Zoom with 18 participants (6 male, 10 female), 24 to 32 years old ($\mu = 29$), to understand if the tool would help users declutter photos, and if users felt creative while using the tool. Users experienced two different conditions of the tool: a baseline grayscale overlay, and our tool highlighting edges along the subject and image borders.

Summative Study Results

Again we saw that the tool made the participants more confident in their ability to address the decluttering principles of subject-background separation and image border flicker (Mdn = 6, IQR = 5-7), versus no guidance (Mdn = 5, IQR = 4-6) [$V = 8$, $p = .03$]. For each favorited photo per task, we asked participants to self-assess them based on subject-background separation and image border flicker. Though we actually did not see a significant improvement in overall self-assessed quality in terms of these principles, participants did believe that the tool was helpful for the task of capturing clear and decluttered images (Mdn = 6, IQR = 5-6), versus no guidance (Mdn = 4, IQR = 4-6) [$V = 134$, $p = .003$].

Therefore we see that while participants are more confident in their ability to take clear photos, and found the tool helpful for achieving their favorite resulting photos, they didn't necessarily find that their photos were better with regard to their personal preferences or the decluttering principles.

Summative Study Discussion

We also did not find significant differences in CSI, but did find support for the increased confidence and descriptions of how the tool encouraged participants to explore more in the qualitative feedback.[4]

Confidence. Participants described feeling like they could take fewer photos, because they could be more confident in each photo they took: *"Usually when I take photos,*

[4] Videos of specific examples from the studies can be found in the Stanford Digital Repository: `https://purl.stanford.edu/dg109bb1678` [83].

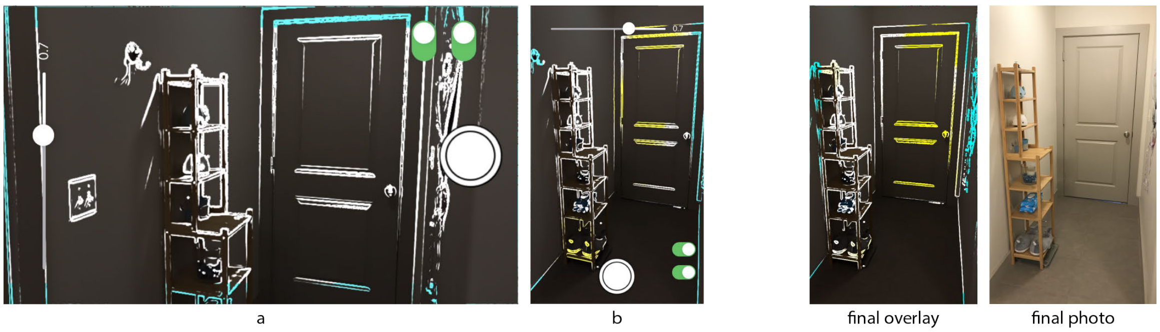


Figure 6.9: Frames from a participant using the tool—on the left are a set of intermediate steps, on the right is the **final overlay** that the participant saw as they took a photo and the resulting **final photo**. Left to right: **(a)** This participant frames a large scale landscape photo of this area of the room, but notices a lot of extra clutter, **(b)** upon switching to portrait to better focus on the region of interest, the hook becomes distracting clutter right along the top left image border. The participant makes small adjustments to refine the framing to remove this clutter for the final image.

I take a ton at once, but didn't do that here. I didn't need to because I was being so precise. I noticed myself reconsidering the composition more: e.g. should I have these things in the edges?" (P15).

For example in Figure 6.9, this participant (P7) refines the camera angle until the edges in the overlay look the way she wants. She is watching the hook on the top left, making sure it doesn't end up in her shot. Thus upon capturing the final photo, she has intentionally refined some details that were brought to her attention by the overlay, and is therefore more confident that this photo achieves her goals.

Exploration. In addition to encouraging more confidence, we found that the tool further encouraged creativity through exploring the space in new ways: *"It was really helpful with how I take photos because normally it's more just snap and done. This one was more like, can I move things out of the background, can I move the subject to frame it to not have a distracting background? Another thing that I don't normally do is pivot the camera and usually just move within a flat plane"* (P12).

Another participant described that the external representation provided by the interface assisted in the process of exploring the scene and quickly evaluating different

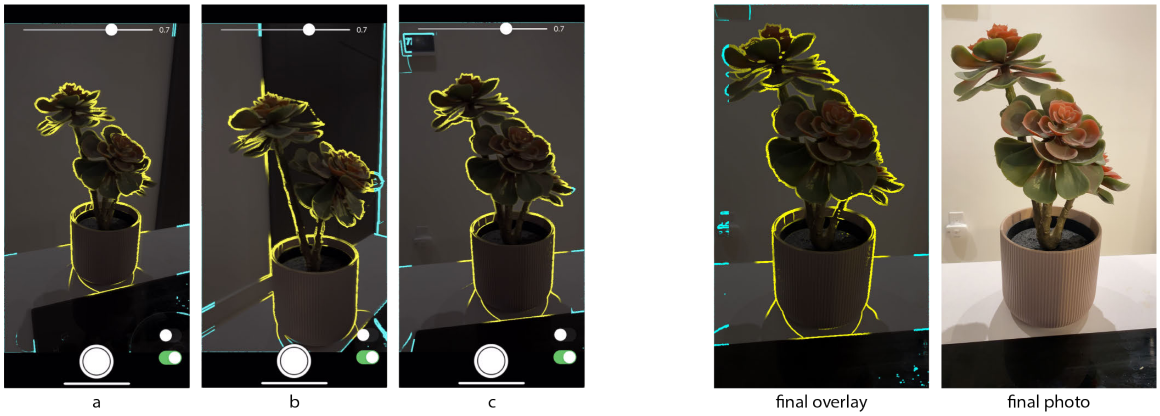


Figure 6.10: Frames from a participant using the tool—on the left are a set of intermediate steps, on the right is the **final overlay** that the participant saw as they took a photo and the resulting **final photo**. Left to right: **(a)** The participant frames a photo of this plant on the counter, but noticed that the strong edge at the bottom of the pot provides extra unwanted contrast and separation. **(b)** Attempting to move this edge out of frame, the participant tries to reframe the image with a different background, but doesn't like the dark wall in the background either. **(c)** The participant turns the camera towards the other direction to continue exploring the background, but notices the thermostat highlighted as it enters the frame at the top left. She keeps the initial edge, but frames a tighter shot of the plant to remove the high contrast thermostat screen from the background.

options: *"It caused me to experiment more… didn't see it as a rule that I needed to minimize lines, but the tool made it easy to move around and check by that metric, how 'good' it was"* (P9).

Figure 6.10 shows how an unexpected edge highlight encourages the participant to explore different backgrounds and compositions. This participant (P15) notices an unwanted edge at the bottom of the plant from the stove top, so she tries out different camera angles to remove the stove, but notices the dark panel in the background, which she doesn't want. She turns the camera back towards the white wall, but notices the high contrast object at the top left entering her shot. In this case, the participant actually ends up using the same background as she started with and still includes the edge that she identified as clutter. However, notice how through this process, this choice became much more intentional as she was aware of the alternative options she had in this space and decided that this best suited her preferences.

Subject Identification. An interaction that we didn't expect that emerged from the studies was that participants used the tool's ability to identify subjects to assess how someone might view their photos.

One participant (P1) observed that the tool jumps back and forth in its highlighting of a subject, She interprets this to mean that there is no clear subject in her photo, and confirms that this is consistent with her own perception. Therefore she tries to find a different way to capture this room with a more clear subject.

On the other hand, another participant (P18) had a slightly different interpretation of the lack of a consistent "subject" or in this case, no identified subject. He decided that since the large scale task shouldn't have a single focal subject, the lack of yellow aligned with the expectations of this goal.

For either of these interpretations, the tool has given the participant another perspective from which to consider what it means to clearly capture a subject. As a result, the participant has more clearly considered the concept of telling the story of their subject in this space.

6.6 Limitations and Future Work

In studying our abstraction interface design, we discovered insights as well as limitations that suggest directions for future work in exploring the potential of abstraction for encouraging rapid iteration in the camera.

6.6.1 Other Abstraction Design

Our design survey (Section 6.3.3) provided us with a lot of interesting insight on how participants might imagine an abstraction overlay to assist with decluttering. For example, we saw the potential benefits of color flattening especially for SBS. In fact, a few participants observed that the goals differed between finding clear separation between the subject and background and finding distracting constrast in the background/along the image border. In the first case, contrast is good—so it should be subdued to guarantee the contrast is still clear without details. In the second case, contrast is bad—so it should be emphasized to make sure the photographer realizes it's presence and has a chance to remove it from the frame. Thus, participants actually suggested using a mixed visualization where regions around the subject used color flattening and distractors near the edge are highlighted through line drawings.

6.6.2 Abstraction for Composition?

Additionally, we found in our low-fidelity prototype that the abstraction also encouraged participants to be more aware of composition. What might our abstraction overlay look like if our focus were instead exploring composition overall rather than decluttering? We would be interested in running a similar design survey asking participants to select overlays based on which best assisted in understanding the overall guidance. How might such an interface compare to our adaptive armatures from Chapter 5?

6.6.3 Algorithmic Implementation

In designing our overlays, we were somewhat limited in the methods that we used in order to produce something that could be computed interactively. It could be interesting to experiment more thoroughly with different types of edge detection to see which would best match what humans actually perceive as "noise"—e.g., in our user studies, we often found that textures like carpet ended up appearing as a lot of noise.

If we remove the requirement of our camera guidance running interactively, there are further approaches that can be considered both for identifying objects in the scene for location context and for generating line drawings. We can imagine instead of the interactive guidance, having an overlay that is shown upon reviewing previously captured photos (similar to our low-fidelity prototype interaction). These algorithms can run for a little longer without having too much of an impact on the overall interaction. The tool can also perform calculations in the background as the user is focused on taking a series of photos. For example, Fried et al. [100] take a non-edge based approach to finding distractors that better captures what humans annotate as being distractors than most edge detection algorithms. There also are a range of segmentation algorithms that could provide additional location context.

6.7 Chapter Summary

Much of the effort photographers put into designing a photo is to help more clearly communicate their intended story. Clutter in an image can greatly detract from the power of an image both in terms of effective communication and in terms of visual appeal. However, it can be easy for photographers to miss this clutter in the moment while paying attention to details in the subject. Inspired by the use of sketching in design to capture higher level structure, our goal was to bring some of the benefits of the abstraction in a sketch-like representation to photography. In this chapter, we walked through the design process behind our abstraction overlay and guidance tool. We demonstrated this capture-time tool and how it encourages users to explore creative options to address the concept of decluttering.

6.7.1 Design Goals

We close this chapter with a summary of how the design of this interface fits within our design goals (Section 3.4).

Context-Aware

For our abstraction overlay, the two pieces of context captured are (1) the line drawings, in the form of edge detection, and (2) the location context through object-based saliency for identifying subjects and proximity to the border of the image. Our edge-based overlay distilled the concept of decluttering down to the principles of subject-background separation and image border flicker specifically in the context of the current image, and helped draw focus away from the subject itself to these surrounding regions.

Encourage Exploration

By drawing attention to the subject-background relationship and the image borders, we hope this might encourage users to consider these concepts when exploring different ways of framing their subject. We found that our visualization helped participants quickly explore different backgrounds and angles for capturing a subject, and ultimately feel more confident in their photos.

Maintain Flexibility

Users can use the information provided by the overlay as they choose, and can actively decide whether or not highlighted objects should be in the final image. The tool allowed for flexible interpretation of the edges. In many cases we saw participants notice highlighted edges and consciously decide that they felt like the potential noise was acceptable to keep.

Chapter 7

Future Work

This dissertation has demonstrated how visualizing key artistic concepts as interactive overlays in the camera can allow novice users to see their image in new ways. Through the design of three capture-time guidance interfaces, I have shown how such visualizations can help novices train their eye to see as experts do. The tools encourage them to explore the space creatively and to be more confident in the photos that they capture. In this chapter, I propose some potential future directions inspired by insights and challenges that emerged from this research.

7.1 Exploring Styles of Photography Guidance

In this dissertation I focus on providing photographic guidance at capture-time through visual annotations that train the user's "artistic vision." These design decisions were motivated by the design goals (see Section 3.4) that we synthesized from our formative studies. However, there is a wide spectrum of methods for providing computational assistance for photography that has yet to be explored, much of which might similarly satisfy the design goals.

7.1.1 Higher Level Photography Goals

In the tools I described in this dissertation, the goal of each interface design was to target a specific photographic concept and help the user to adopt the thought process that a professional photographer might follow in considering that specific concept. However, the current, highly-crafted approach is too labor-intensive to practically scale this design process broadly across photography. Also, this current design requires users to seek out a specific-domain tutor, which doesn't scale on the user side either.

Additionally, such tools don't provide any assistance with the step of determining which concept is relevant to consider at a given point in time. How might we design an interface that combines these tools, and understands when to guide users to focus on one photography concept versus another? And especially when these visualizations are combined into a single application, how do we ensure that the interface itself isn't overwhelming?

Through our studies, we found that it's hard to predict the balance between being overly distracting/noisy and possibly being too familiar. During our declutter studies (see Chapter 6), participants noted that when using the grayscale overlay it was easy to forget the task at hand of capturing a decluttered image, and sometimes even forgot that they were capturing a color photo and not the grayscale overlay they saw. One possible interpretation of this is that the overlay should have some of its own movement/change that happens independently of the photo in order to guide attention and awareness. One way to achieve this is to have the overlay components move in a different manner or at a different pace from the overall image—e.g., the structure of the composition grid (see Chapter 5) made movements of the highlights occur in discrete jumps even though the camera movement was continuous.

Another interesting question that comes up here is the interpretability of the guidance, and the overall trust between the user and the system. Particularly because we are interested in training the user to develop their own artistic opinions in the context of these photographic concepts, we designed algorithms and visualized their behavior in ways that we hoped the user could quickly and actively understand—movement in the overlay should be based on the context of the image and in a way that is somewhat predictable to the user. In doing so, we made some choices to use simpler algorithmic approaches to prioritize the interpretability aspect. However for a tool that integrates multiple concepts, this becomes more challenging. Even if each individual visualization is easily interpretable, it can be challenging to keep track of multiple annotation visualizations. It is possible that such visualizations are not scalable in this way, or that they just need to be supported by another form of feedback communication.

On the other end of the spectrum of interpretability, we found that even just the dynamic nature of such annotations alone can provide some benefit. As demonstrated by the random highlighting of the armature lines (see Chapter 5) or the participants' reactions to the decluttering when it couldn't identify the subject (see Chapter 6), participants seem to be able to find ways to respond to this seemingly meaningless guidance. In the first case, we saw participants use the briefly appearing highlights to inspire more spontaneous composition ideas. Depending on the step of the photographic process that the user is in, as well as the user's experience, different levels of interpretability may be beneficial. In the early exploratory stages, the serendipity of less interpretable algorithms might help generate more diverse ideas and encourage the user to not satisfice too early [221]. As the user hones in on a single idea, it might be helpful to have more targeted and controlled feedback for refinement.

7.1.2 More Directed Guidance

In this dissertation, I have described design choices to deliberately avoid allowing the system to express its opinion on how the final image should look. This approach required a certain type of user population; they had to be interested in creative growth and be willing to invest effort into exploring their creative ideas. However even for such users, the lack of instruction in the presence of increased awareness can make the task overall more stressful, a balance we tried to strike with visualizations that provide some structure and with some initial training on photography principles. Developing and training a personal artistic style and opinion is a slow process; one approach could be to provide more directed guidance when the user is first learning a photographic principle, and then gradually reduce the direction as the user becomes more familiar with the concept.

In fact, through our user studies, we heard from participants that a little more guidance from the system in how to respond to the feedback could be helpful. For instance, if all portrait lighting styles are available (see Chapter 4), how does one pick between the options? Or, given an unexpected highlight composition grid line in the adaptive armature (see Chapter 5), should one intentionally incorporate it into their composition?

In particular, I noticed that decluttering (see Chapter 6) seemed maybe more difficult for people to address than lighting or composition, perhaps because it is a more new and unfamiliar concept. As a result, participants still made "mistakes" even while being aware of specific clutter to consider and carefully refining their photos (e.g., describing wanting to remove an object from the image that does end up partially appearing in the final image). Considering the decluttering principles, they also are somewhat unique in that there is a more defined good and bad (unless the photographer is choosing to intentionally blend the subject into the background or have a very cluttered background). The creativity in decluttering comes at the stage of determining how to address the clutter. Therefore I wonder if it is reasonable to have a slightly more "opinionated" interface without being prescriptive (at least during training) as is suggested by one of our design survey participants: *"Making it very clear about the distinction between whether the overlay is used to indicate if the two properties are good, bad, or both would be very helpful"* (P18).

In this case, the opinion is with regards to evaluation of the current image. The user is very directly made aware of potential issues that they possibly should address, but the system leaves the process of execution to the user. Thus, the final result still creatively belongs to the user. This is very different from many existing guidance interfaces that provide navigational direction towards executing a specific goal imposed by the system [186,207], and is more similar to guidance in commercial cameras like the light meter or interactions that focus on displaying an aesthetics quality score [192,199] or highlighting regions of potential concern (e.g., subject is too centered) [55].

Overall, we still want to use the advantage of computation to be contextual and adaptive. Nonetheless such guidance can still take on a range of tones in terms of level of prescriptiveness. For example, the tool can simply call out "your background has a lot of clutter, consider moving." A participant (P12) from our decluttering design survey suggested having an adaptively appearing overlay, along with accompanying descriptive text: *"hey user, you have very little subject-background separation, but we darkened your background a bit to make your subject pop, and also adjusted tones and lighting..."*

7.1.3 Beyond Visual Annotations

Note that these suggestions start to present a different interaction modality as well. These different forms present their own benefits and limitations. For instance, the use of annotations enables the feedback to be directly embedded in the image, but this also means the visuals obstruct some of the view of the actual image. Any additional navigation guidance further adds to the visual clutter. Text, on the other hand, can still be relatively concise and precise, especially in more instructive guidance and be presented both visually and through audio.

Researchers have explored combining audio and visual cues for guiding users with visual impairments towards their intended photo [115, 286, 287]. These systems often focus on the more practical application of communication, such as making sure the desired content is clearly within frame. Haptics also present promise as a mode to interactively communicating artistic feedback, particularly through tactile visual interfaces [148]. It would be interesting to explore the use of these modalities for communicating feedback on artistic concepts to guide users' exploration of the space.

Expanding the range of modalities may also enable different approaches to assisting in photography. Rather than focus on a specific topic within photography (i.e., lighting, composition, and declutter) as I did in my dissertation, it might be interesting to provide guidance towards a specific practice that is taught or trained, e.g., remember to take more photos, try moving the camera angle, or try moving the camera closer. These actions can influence a wide range of photographic concepts and can possibly focus more directly on the physicality of the photographic process. As mentioned above, the use of multiple modalities can also make it easier to transition between visual overlays for different photographic concepts.

7.2 Achieving Creative Intentions

In a number of our user studies, we heard participants mention that the annotations provided by the guidance tools could provide good scaffolding for communicating their intentions to another person (e.g., someone helping them take a photo). Participants described possibly describing their goals in the photo based on the composition grid, e.g., please place my head at this intersection and this building edge along this line, or based on the decluttering highlights, please adjust the camera such that there is a clear yellow outline around me. In this way, the tools could help users achieve their artistic intentions even when they are not behind the camera. Here I explore other types of interaction paradigms that could leverage the annotations as tools for describing the user's intentions.

7.2.1 Adapting to User Intentions

Since my work focuses primarily on the exploration aspects of photography, for the most part, I tried to limit interactions that might cause the user to focus on a single idea too early. In fact, this is a slight downside to the portrait lighting interface design (see Chapter 4). In order to get to the reorientation guidance, the user selects a single style to focus on from the gallery of portrait lighting options. For ease of implementation, in our prototype we provide a drop-down to easily navigate back to the gallery, but in the ideal situation the gallery would always be visible to the user. However, note that this interaction design is more helpful for refining the portrait lighting.

As noted above, the portrait lighting tool had a two-pronged workflow that provided different guidance for exploring the space of lighting options (a gallery previewing a range of styles), and then achieving a specific style that the user selected (reorientation and alignment guidance). The act of selecting a single portrait lighting style can help the user focus their attention on refining their scene to best capture that lighting style.

In this case, the user has communicated a part of their intention, and the system adapts to better guide them to achieve that intention.

As with the systems for collaboratively co-creating with an AI [71, 119, 205, 227], how might we design interfaces that adapt with the user's decisions? How might we enable the photographer to indicate their intentions to the tool (and how might that tool consider the feedback from the user)? For example, one could imagine the user selecting a subset of lines from the armature for composition guidance (see Chapter 5). This selection should be viewed as constraints on what the tool shows to the user. Once a selection is made, the user has communicated not only the specific artistic constraint, but also progress in the creative process—the user is moving out of exploration and into refinement. The interface can adapt to this change as well, towards helping the user achieve the specific intention they have communicated. Specifically for achieving the desired composition, one could imagine the interface focusing more on communicating how well-aligned the current image is to the target composition. The camera can guide the user to adjust alignment of elements in the photo, or even suggest adding/removing objects to better match the target image template.

7.2.2 Image Retrieval and Editing

Here we explore the use of annotations for steps in the photographic process post-capture, specifically curation and editing. For images captured with our guidance tools (or any type of interactive feedback), it can be reasonable to expect that the feedback provided to the user at the moment of capture can be stored as metadata. This additional information can be incredibly powerful for helping the photographer capture context on their intentions at capture time during the editing phase (at which point that information might otherwise be long gone from their memory). Researchers have designed tools specifically for quickly logging videos with contextual annotations to help speed up the culling process [282].

These annotations can be helpful just as additional context passively present as the photographer reviews photos. However, one can also imagine using the properties of the annotations to search through libraries of photos. For example, a photographer could search for every instance of a specific portrait lighting style or of a specific composition. A photographer could also generate a high level visual summary of their photos library: "these are my most common lighting styles and compositions, and I tend to struggle to notice clutter immediately behind a subject's head."

If comparing a set of near duplicates of photos captured with the same intention, the photographer could imagine searching the set for the one that best aligned with a target composition, and then even have the image automatically straightened/cropped to best match that target. Or while the user is interactively editing, rather than showing a static composition grid as is currently common in commercial tools, the software could adaptively highlight lines to represent the current composition.

7.3 Creativity Evaluation

One challenge we faced throughout the work in this dissertation, was determining how to find the appropriate quantitative and qualitative metrics for evaluating this style of work.

7.3.1 Quantifying Goodness of Photos

In all of the evaluations of the systems I built as described in this dissertation, I relied on the self-assessment of photo quality. Additionally, in these evaluations, we've asked participants to focus their attention on different photographic concepts independently. We've yet to consider how these photographic concepts play together to impact the overall aesthetic quality. For instance, how much do small variations in things like lighting, composition, and clutter effect the overall goodness of the image?

An interesting approach to answering this question might be to produce datasets of example images that vary across parameters within the space of these different concepts. With this dataset, we could imagine doing some A/B testing to see whether there is significant change in how people rate these different versions of the photo. Li and Vogel [186] begin to explore this question for portraiture. Using virtually generated portraits, they test a range of lighting styles, compositions, and camera distances to determine the look of an ideal portrait and use this to inform their selfie guidance. Additionally, a number of papers have started to explore automatically assessing the aesthetic quality of photos with consideration of the features/attributes that influence their quality [167, 267]. An interesting next step for building on this work might be to combine these methods to systematically understand the impact of different photographic concepts.

A number of interesting additional questions arise from this line of work. Can we use such methods to capture individuals' aesthetics preference models? If so, it could be interesting to capture the preference models of different artists to be able to evaluate how likely they are to find a photo visually appealing. Tying this back to this dissertation, how might these preference models inform the personalization of feedback provided by in-camera guidance interfaces? Could this enable providing guidance modeled after the user's personal preferences or even to mimic the style of feedback that different artists might provide?

7.3.2 Measuring the Creative Process and Learning

It would also be interesting to understand how we might come up with better quantitative metrics for evaluating the type of awareness-based interface proposed in this dissertation. Using existing evaluation tools such as the Creativity Support Index to measure creativity, or NASA-TLX to measure cognitive load is fairly attractive, but especially ambiguous in our case due to the subtle, "light-touch" nature of the guidance. For our capture-time guidance, these metrics often didn't provide much definitive signal. With regards to NASA-TLX, it often wasn't clear what was a "good" result for the tool—the tools' goals were simultaneously to increase awareness of new ideas to explore (which approximately maps to reducing cognitive load), while also trying to increase awareness to encourage the user to make more intentional choices (which in contrast, approximately maps to increasing cognitive load). Thus in our work, we only used it to measure the difficulty of the more "execution"-based step of the portrait lighting project (see Chapter 4), the ease of reorientation, once the more creative and exploratory step of choosing a desired lighting style was complete.

Additionally, the interesting impacts of these interfaces were more in how they influenced change in the users' photographic processes and workflows. We capture some of this through participants' think-aloud and post-study interviews, however it would be interesting to be able to more directly measure these changes. For example, could we maybe use the motions of the camera or the users' gaze to measure behavior change in the photographic process?

Another question to consider is, how might we measure creative growth and learning? In particular, since our tools are designed to train "artistic vision" in novices, measuring growth over time is particularly relevant. The short-term nature of our studies meant that participants were introduced to what were often relatively unfamiliar photographic concepts and immediately asked to apply this newly gained knowledge. Moreover, the evaluation of photo quality was also self-assessed, meaning they additionally had to evaluate with respect to this new concept. Running these studies over time would better allow us to understand if these visual annotations served as effective training and allowed the users to better grasp these photographic concepts over time. They would additionally improve the reliability of their evaluation abilities—in particular, we could even see if their evaluation possibly changed or converged over time. If we were able to actually measure and interpret gaze, through longitudinal studies, we could potentially even begin to see if novices in fact started to follow gaze paths that were structurally more similar to that of expert photographers—an evaluation method that approaches the idea of directly measuring the development of "artistic vision."

Chapter 8

Conclusion

Experts see differently than novices [9, 48, 65]. In my work, I hope to help novices train their eye to see in the way that experts do, enabling novices to be more intentional in their artistic choices. My primary insight is that providing real-time annotations can make accessing the kind of feedback that experts might provide more approachable and actionable, while encouraging creativity and exploration. This has the potential to help novices develop "artistic vision" as they capture photos.

In my dissertation, I developed three capture-time interfaces that aim to demonstrate this insight, using computation to provide real-time contextual guidance. This chapter summarizes the effects that each of the capture-time interfaces had on novice photography. I then highlight the key contributions of my dissertation and conclude with ideas of how to extend our insights beyond photography.

8.1 Effects of Capture-Time Guidance on Novice Photography

In this dissertation, I present three capture-time interfaces that provide guidance for portrait lighting, composition, and decluttering. Here I briefly describe each interface and how they encouraged awareness with respect to their target photographic concept, and how that influenced novice participants' photographic experiences.

The portrait lighting tool (see Chapter 4) helps users be more aware of the available lighting styles and reorient their subject to best achieve the lighting style of their choice. Participants noted being able to consider a wider range of lighting styles for their portraits that they previously weren't aware of. Participants found our tool to be useful for the task of producing well-lit portraits, and were able to use the tool to capture photos they believed to have better lighting.

The composition guidance tool (see Chapter 5) makes users more aware of the current composition by highlighting lines in a composition grid that are most relevant to the camera view. These highlights were able to inspire new composition ideas, while also providing confirmation that made participants feel more confident in the compositions they ultimately chose. Using the tool, participants were able to capture photos they believed were better composed.

Finally, the decluttering tool (see Chapter 6) increases users' awareness of clutter that would draw attention away from the main story of the image by abstracting the camera view to outline edges around the subject(s) or the image borders. Participants found the tool helpful for the task of capturing clear and decluttered images. We again saw that the increased awareness helped make participants more confident in their ability to address the concept at hand, in this case decluttering. It also encouraged participants to explore the scene to address concerns with distracting clutter.

A common thread amongst these interfaces is the use of computation to generate visualizations for the user to interpret rather than the system prescribing an action. These visualizations provide the user with an alternative context-aware representation of the image that helps focus their awareness on particular aspects of the image in relation to the different photographic concepts. We also find that these representations help to break down the search space into a number of discrete options, making it less intimidating to explore the space of possible options. Through our user studies, we saw that participants were indeed aware of the photographic concept that the individual tools encouraged them to focus on, and that their system-rooted observations led to making more intentional artistic decisions.

8.2 Contributions

This dissertation explores an approach to helping novices train their "artistic vision." This approach is exemplified as three different capture-time interfaces each focusing on a key photographic concept. Through these proposed contextual capture-time interfaces, I demonstrate that visualizing key artistic concepts improves self-assessed creativity and photo quality. This section summarizes the specific research contributions of this dissertation in three areas: algorithms and techniques, artifacts, and experimental results.

Concepts and Techniques. To inform the design of capture-time interfaces, I conducted a number of formative studies to better understand current photography practices and tools, and how in-camera guidance could be of assistance in the user's existing workflows. From these studies, we synthesized a set of three design goals: guidance should be **context-aware**, it should **encourage exploration**, and it should allow users to **maintain creative flexibility**. We apply these design goals in the design of the interfaces presented in this dissertation.

Artifacts. In this dissertation, I presented three artifacts—each in the form of an in-camera guidance interface that tackles a specific photographic concept. Informed by a set of design goals, these interfaces strive to not be prescriptive in their guidance in order to encourage creativity and exploration. Inspired by photographers' current practice of directing attention through manually drawing annotations onto photos, these interfaces aim to take a light-touch approach of providing targeted awareness cues through automatically generated annotation overlays. For each tool, I contribute designs for novice-interpretable visual overlays to represent the considerations for a key artistic concept, and real-time algorithms for interactively generating these visualizations.

Experimental Results. I additionally contribute user evaluations to demonstrate that these awareness-based guidance interfaces can help novices feel more creative while taking higher quality photos. In particular, we observed ways in which each of these tools encourages participants to explore new ideas in the context of the target photographic concepts, ultimately making them more confident in their ability to consider these concepts in their photography.

8.3 "Artistic Vision" Beyond Photography

While I specifically targeted photography in this dissertation, I am generally excited about applying some of the design goals and ideas we used here for photography, to other creative domains. Can we similarly help develop "artistic vision" for different types of creativity, such as graphic design or fabrication? For any context, there are aspects that experts have trained themselves to be aware of that might be harder for a novice to recognize. I am excited to design computational feedback and guidance that helps increase awareness of the decisions that need to be made in the creative process such that users can explore the creative space in a more intentional and informed manner. While we are seeing the success of using computational methods to generate creative work, in my work I am motivated by the idea of using computational guidance to coach humans in developing their creative skills. My hope is that this awareness-based approach is a step towards leveraging computation to help users pursue their artistic interests.

Bibliography

[1] Radhakrishna Achanta and Sabine Susstrunk. Superpixels and polygons using simple non-iterative clustering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[2] Andrew Adams, Eino-Ville Talvala, Sung Hee Park, David E Jacobs, Boris Ajdin, Natasha Gelfand, Jennifer Dolson, Daniel Vaquero, Jongmin Baek, Marius Tico, et al. The frankencamera: an experimental platform for computational photography. In *ACM SIGGRAPH 2010 papers*, pages 1–12. ACM New York, NY, USA, 2010.

[3] Brett Adams and Svetha Venkatesh. Director in your pocket: holistic help for the hapless home videographer. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 460–463, 2004.

[4] Brett Adams and Svetha Venkatesh. Dynamic shot suggestion filtering for home video based on user performance. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 363–366, 2005.

[5] Brett Adams and Svetha Venkatesh. Situated event bootstrapping and capture guidance for automated home movie authoring. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 754–763, 2005.

[6] Brett Adams and Svetha Venkatesh. An embedded suggestive interface for making home videos. In *2006 IEEE International Conference on Multimedia and Expo*, pages 2025–2028. IEEE, 2006.

[7] Eric Adams. How to use focus peaking to take better photos. https://www.gearpatrol.com/tech/a330370/how-to-use-focus-peaking-photography/, Feb 2021.

[8] Rolf Adelsberger, Remo Ziegler, Marc Levoy, and Markus Gross. *Spatially adaptive photographic flash*. ETH, Department of Computer Science, 2008.

[9] Beth Adelson. Problem solving and the development of abstract categories in programming languages. *Memory & cognition*, 9(4):422–433, 1981.

[10] Aseem Agarwala, Mira Dontcheva, Maneesh Agrawala, Steven Drucker, Alex Colburn, Brian Curless, David Salesin, and Michael Cohen. Interactive digital photomontage. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 294–302. ACM, 2004.

[11] Maneesh Agrawala, Wilmot Li, and Floraine Berthouzoz. Design principles for visual communication. *Communications of the ACM*, 54(4):60–69, 2011.

[12] David Akers, Frank Losasso, Jeff Klingner, Maneesh Agrawala, John Rick, and Pat Hanrahan. Conveying shape and features with image-based relighting. In *Proceedings of the 14th IEEE Visualization 2003 (VIS'03)*, page 46. IEEE Computer Society, 2003.

[13] Susan A Ambrose, Michael W Bridges, Michele DiPietro, Marsha C Lovett, and Marie K Norman. *How learning works: Seven research-based principles for smart teaching*. John Wiley & Sons, 2010.

[14] Apple. Cilineoverlay. https://developer.apple.com/documentation/coreimage/cilineoverlay.

[15] Apple. Framework: Vision. https://developer.apple.com/documentation/vision, 2019.

[16] James Arvo and Kevin Novins. Fluid sketches: continuous recognition and morphing of simple hand-drawn shapes. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 73–80, 2000.

[17] Jackie Assa, Lior Wolf, and Daniel Cohen-Or. The virtual director: a correlation-based online viewing of human motion. In *Computer Graphics Forum*, volume 29, pages 595–604. Wiley Online Library, 2010.

[18] Hadar Averbuch-Elor, Daniel Cohen-Or, Johannes Kopf, and Michael F. Cohen. Bringing portraits to life. *ACM Trans. Graph.*, 36(6):196:1–196:13, November 2017.

[19] Shai Avidan and Ariel Shamir. Seam carving for content-aware image resizing. In *ACM SIGGRAPH 2007 papers*, pages 10–es. 2007.

[20] Soonmin Bae, Aseem Agarwala, and Frédo Durand. Computational rephotography. *ACM Trans. Graph.*, 29(3):24–1, 2010.

[21] Soonmin Bae, Aseem Agarwala, and Frédo Durand. Computational rephotography. *ACM Trans. Graph.*, 29(3):24:1–24:15, July 2010.

[22] Ronald Baecker, Alan J Rosenthal, Naomi Friedlander, Eric Smith, and Andrew Cohen. A multimedia system for authoring motion pictures. In *Proceedings of the Fourth ACM international Conference on Multimedia*, pages 31–42, 1997.

[23] Jongmin Baek, Dawid Pająk, Kihwan Kim, Kari Pulli, and Marc Levoy. Wysiwyg computational photography via viewfinder editing. *ACM Transactions on Graphics (TOG)*, 32(6):1–10, 2013.

[24] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. Subtle Gaze Direction. *ACM Transactions on Graphics*, 28(4):100:1–100:14, 2009.

[25] Saeideh Bakhshi, David Shamma, Lyndon Kennedy, and Eric Gilbert. Why we filter our photos and how it impacts engagement. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 9, 2015.

[26] S. Banerjee and B. L. Evans. In-camera automation of photographic composition rules. *IEEE Transactions on Image Processing*, 16(7):1807–1820, July 2007.

[27] Terry Barrett. Teaching about photography: Selectivity, instantaneity, and credibility. *Art Education*, 39(3):12–15, 1986.

[28] Terry Barrett. Studio critiques of student art: As they are, as they could be with mentoring. *Theory into Practice*, 39(1):29–35, 2000.

[29] Barbara Barry and Glorianna Davenport. Documenting life: Videography and common sense. In *2003 International Conference on Multimedia and Expo. ICME'03. Proceedings (Cat. No. 03TH8698)*, volume 2, pages II–197. IEEE, 2003.

[30] Aric Bartle, Alla Sheffer, Vladimir G Kim, Danny M Kaufman, Nicholas Vining, and Floraine Berthouzoz. Physics-driven pattern adjustment for direct 3d garment editing. *ACM Trans. Graph.*, 35(4):50–1, 2016.

[31] Mitra Basu. Gaussian-based edge-detection methods-a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 32(3):252–260, 2002.

[32] Nino Batista, Follow Senior Writer Nino Batista's work is regularly published in lifestyle, and men's publications all over the world. Photography is difficult. `https://fstoppers.com/editorial/photography-difficult-55962`, Jan 2015.

[33] David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B Tenenbaum, William T Freeman, and Antonio Torralba. Gan dissection: Visualizing and understanding generative adversarial networks. *arXiv preprint arXiv:1811.10597*, 2018.

[34] Patrick Baudisch, Desney Tan, Drew Steedly, Eric Rudolph, Matt Uyttendaele, Chris Pal, and Richard Szeliski. Panoramic viewfinder: providing a real-time preview to help users avoid flaws in panoramic pictures. In *Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*, pages 1–10. Computer-Human Interaction Special Interest Group (CHISIG) of Australia, 2005.

[35] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *Computer vision–ECCV 2006*, pages 404–417, 2006.

[36] Finn Beales and Alex Strohl. *The Photography Storytelling Workshop: A five-step guide to creating unforgettable photographs*. White Lion Publishing, 2020.

[37] Cynthia L Bennett, Jane L. E, Martez E Mott, Edward Cutrell, and Meredith Ringel Morris. How teens with visual impairments take, edit, and share photos on social media. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–12, 2018.

[38] Gilbert Louis Bernstein and Wilmot Li. Lillicon: Using transient widgets to create scale variations of icons. *ACM Transactions on Graphics (TOG)*, 34(4):1–11, 2015.

[39] Floraine Berthouzoz, Wilmot Li, and Maneesh Agrawala. Tools for placing cuts and transitions in interview video. *ACM Transactions on Graphics (TOG)*, 31(4):1–8, 2012.

[40] Jim Blinn. Where am i? what am i looking at?(cinematography). *IEEE Computer Graphics and Applications*, 8(4):76–81, 1988.

[41] Bruce Block. *The Visual Story: Creating the Visual Structure of Film, TV and Digital Media*. Routledge, 2013.

[42] Edward S Bomback. *Manual of photographic lighting*. Fountain Press, Limited, 1971.

[43] Charles C Bonwell and James A Eison. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports.* ERIC, 1991.

[44] Adrien Bousseau, Emmanuelle Chapoulie, Ravi Ramamoorthi, and Maneesh Agrawala. Optimizing environment maps for material depiction. In *Computer graphics forum*, volume 30, pages 1171–1180. Wiley Online Library, 2011.

[45] Ivaylo Boyadzhiev, Kavita Bala, Sylvain Paris, and Frédo Durand. User-guided white balance for mixed lighting conditions. *ACM Trans. Graph.*, 31(6), November 2012.

[46] Ivaylo Boyadzhiev, Jiawen Chen, Sylvain Paris, and Kavita Bala. Do-it-yourself lighting design for product videography. In *Computational Photography (ICCP), 2016 IEEE International Conference on*, pages 1–9. IEEE, 2016.

[47] Ivaylo Boyadzhiev, Sylvain Paris, and Kavita Bala. User-assisted image compositing for photographic lighting. *ACM Trans. Graph.*, 32(4):36–1, 2013.

[48] John Bransford, John D Bransford, Ann L Brown, and Rodney R Cocking. *How people learn: Brain, mind, experience, and school*. National Academies Press, 1999.

[49] Bill Buxton. *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann, 2010.

[50] Zachary Byers, Michael Dixon, William D Smart, and Cindy M Grimm. Say cheese! experiences with a robot photographer. *AI magazine*, 25(3):37–37, 2004.

[51] Zoya Bylinskii, Nam Wook Kim, Peter O'Donovan, Sami Alsheikh, Spandan Madan, Hanspeter Pfister, Fredo Durand, Bryan Russell, and Aaron Hertzmann. Learning visual importance for graphic designs and data visualizations. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pages 57–69, New York, NY, USA, 2017. ACM.

[52] Caroline Cakebread. People will take 1.2 trillion digital photos this year - thanks to smartphones. `https://www.businessinsider.com/12-trillion-photos-to-be-taken-in-2017-thanks-to-smartphones-chart-2017-8`, Aug 2017.

[53] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[54] Stuart Card, JD Mackinlay, and B Shneiderman. Information visualization. *Human-computer interaction: Design issues, solutions, and applications*, 181, 2009.

[55] Scott Carter, John Adcock, John Doherty, and Stacy Branham. Nudgecam: Toward targeted, higher quality media capture. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 615–618, 2010.

[56] Luigi Celona, Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. Autocropping: A closer look at benchmark datasets. In Elisa Ricci, Samuel Rota Bulò, Cees Snoek, Oswald Lanz, Stefano Messelodi, and Nicu Sebe, editors, *Image Analysis and Processing – ICIAP 2019*, pages 315–325, Cham, 2019. Springer International Publishing.

[57] Caroline Chan, Shiry Ginosar, Tinghui Zhou, and Alexei A Efros. Everybody dance now. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5933–5942, 2019.

[58] Hui-Tang Chang, Yu-Chiang Frank Wang, and Ming-Syan Chen. Transfer in photography composition. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 957–960, New York, NY, USA, 2014. ACM.

[59] Huarong Chen, Bin Wang, Tianxiang Pan, Liwang Zhou, and Hua Zeng. Cropnet: Real-time thumbnailing. In *Proceedings of the 26th ACM International Conference on Multimedia*, MM '18, pages 81–89, New York, NY, USA, 2018. ACM.

[60] J. Chen, G. Bai, S. Liang, and Z. Li. Automatic image cropping: A computational complexity study. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 507–515, June 2016.

[61] Yi-Ling Chen, Jan Klopp, Min Sun, Shao-Yi Chien, and Kwan-Liu Ma. Learning to compose with professional photographs on the web. *CoRR*, abs/1702.00503, 2017.

[62] Bin Cheng, Bingbing Ni, Shuicheng Yan, and Qi Tian. Learning to photograph. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, pages 291–300, New York, NY, USA, 2010. ACM.

[63] Erin Cherry and Celine Latulipe. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(4):21, 2014.

[64] Ewen Cheslack-Postava, Nolan Goodnight, Ren Ng, Ravi Ramamoorthi, and Greg Humphreys. 4d compression and relighting with high-resolution light transport matrices. In *Proceedings of the 2007 Symposium on Interactive 3D Graphics and Games*, I3D '07, pages 81–88, New York, NY, USA, 2007. ACM.

[65] Michelene TH Chi, Paul J Feltovich, and Robert Glaser. Categorization and representation of physics problems by experts and novices. *Cognitive science*, 5(2):121–152, 1981.

[66] Matthew Cong, Kiran S Bhat, and Ronald Fedkiw. Art-directed muscle simulation for high-end facial animation. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 119–127, 2016.

[67] James W Cowman. Dynamic symmetry art. `https://www.dynamicsymmetryart.com/`, 2019.

[68] Luka Crnkovic-Friis and Louise Crnkovic-Friis. Generative choreography using deep learning. *arXiv preprint arXiv:1605.06921*, 2016.

[69] Wayne W Daniel. Kruskal–wallis one-way analysis of variance by ranks. *Applied Nonparametric Statistics*, pages 226–234, 1990.

[70] Deanna P Dannels and Kelly Norris Martin. Critiquing critiques: A genre analysis of feedback across novice to expert design studios. *Journal of Business and Technical Communication*, 22(2):135–159, 2008.

[71] Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko. Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pages 196–207, 2016.

[72] Richard C Davis, Brien Colwell, and James A Landay. K-sketch: a'kinetic'sketch pad for novice animators. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 413–422, 2008.

[73] Stephen Boyd Davis and Magnus Moar. The amateur creator. In *Proceedings of the 5th conference on Creativity & cognition*, pages 158–165, 2005.

[74] Michael D Day. Artist-teacher: A problematic model for art education. *Journal of Aesthetic education*, 20(4):38–42, 1986.

[75] Nicola De Pisapia, Francesca Bacci, Danielle Parrott, and David Melcher. Brain networks for visual creativity: a functional connectivity study of planning a visual artwork. *Scientific reports*, 6(1):1–11, 2016.

[76] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proc. of SIGGRAPH*, pages 145–156, 2000.

[77] Doug DeCarlo and Anthony Santella. Stylization and abstraction of photographs. *ACM transactions on graphics (TOG)*, 21(3):769–776, 2002.

[78] Daniel Dixon, Manoj Prasad, and Tracy Hammond. icandraw: Using sketch recognition and corrective feedback to assist a user in drawing human faces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 897–906, 2010.

[79] Steven P Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L Schwartz, and Scott R Klemmer. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 17(4):1–24, 2010.

[80] Hugh Dubberly. How do you design. *A compendium of models*, 2004.

[81] Jane L. E, Ohad Fried, and Maneesh Agrawala. Optimizing portrait lighting at capture-time using a 360 camera as a light probe. In *UIST*, pages 221–232, 2019.

[82] Jane L. E, Ohad Fried, Jingwan Lu, Jianming Zhang, Radomír Měch, Jose Echevarria, Pat Hanrahan, and James A Landay. Adaptive photographic composition guidance. In *CHI*, pages 1–13, 2020.

[83] Jane L. E, James A. Landay, and Pat Hanrahan. Data from "artistic vision: Providing contextual guidance for capture-time decisions". `https://purl.stanford.edu/dg109bb1678`.

[84] Marc Ebner. *Color constancy*, volume 6. John Wiley & Sons, 2007.

[85] Ernest Edmonds, Greg Turner, and Linda Candy. Approaches to interactive art systems. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 113–117, 2004.

[86] Elmar Eisemann and Frédo Durand. Flash photography enhancement via intrinsic relighting. *ACM transactions on graphics (TOG)*, 23(3):673–678, 2004.

[87] Elliot W Eisner. Educating artistic vision. 1972.

[88] James Elkins. *Why art cannot be taught: A handbook for art students*. University of Illinois Press, 2001.

[89] Hui Fang and Meng Zhang. Creatism: A deep-learning photographer capable of creating professional work. *CoRR*, abs/1707.03491, 2017.

[90] Naihui Fang, Haoran Xie, and Takeo Igarashi. Selfie guidance system in good head postures. In *Joint Proceedings of the ACM IUI 2018 Workshops co-located with the 23rd ACM Conference on Intelligent User Interfaces (ACM IUI 2018), Tokyo, Japan, March 11, 2018.*, 2018.

[91] Farshid Farhat, Mohammad Mahdi Kamani, and James Z. Wang. CAPTAIN: comprehensive composition assistance for photo taking. *CoRR*, abs/1811.04184, 2018.

[92] Edmund Burke Feldman. *Practical art criticism*. Prentice Hall Englewood Cliffs, NJ, 1994.

[93] Adam Finkelstein and David H Salesin. Multiresolution curves. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 261–268, 1994.

[94] Jonathan Fish and Stephen Scrivener. Amplifying the mind's eye: sketching and visual cognition. *Leonardo*, 23(1):117–126, 1990.

[95] Sebastien Fitch. Art critiques: A guide. *Studies in Art Education*, 57(2):185–187, 2016.

[96] James Fogarty, Desney Tan, Ashish Kapoor, and Simon Winder. Cueflik: interactive concept learning in image search. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 29–38, 2008.

[97] Juliano Franz, Anderson Maciel, and Luciana Nedel. Instant rephotography. In David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris, editors, *Human-Computer Interaction – INTERACT 2019*, pages 208–227, Cham, 2019. Springer International Publishing.

[98] C Ailie Fraser, Mira Dontcheva, Holger Winnemöller, Sheryl Ehrlich, and Scott Klemmer. Discoveryspace: suggesting actions in complex software. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pages 1221–1232, 2016.

[99] C Ailie Fraser, Tricia J Ngoon, Mira Dontcheva, and Scott Klemmer. Replay: contextually presenting learning videos across software applications. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[100] Ohad Fried, Eli Shechtman, Dan B Goldman, and Adam Finkelstein. Finding distractors in images. In *Proceedings of the IEEE Conference on Computer Vision and pattern Recognition*, pages 1703–1712, 2015.

[101] Ohad Fried, Eli Shechtman, Dan B Goldman, and Adam Finkelstein. Perspective-aware manipulation of portrait photos. *ACM Transactions on Graphics (TOG)*, 35(4):128, 2016.

[102] W. Timothy Gallwey, Zach Kleiman, and Pete Carroll. *The Inner Game of Tennis: The Classic Guide to the Mental Side of Peak Performance*. Random House Trade Paperbacks, revised edition, 1997.

[103] Quentin Galvane, Marc Christie, Rémi Ronfard, Chen-Kim Lim, and Marie-Paule Cani. Steering behaviors for autonomous cameras. In *Proceedings of Motion on Games*, pages 93–102. 2013.

[104] Quentin Galvane, Christophe Lino, Marc Christie, Julien Fleureau, Fabien Servant, Fran¸ois-louis Tariolle, and Philippe Guillotel. Directing cinematographic drones. *ACM Transactions on Graphics (TOG)*, 37(3):1–18, 2018.

[105] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. Learning to predict indoor illumination from a single image. *arXiv preprint arXiv:1704.00090*, 2017.

[106] Jim Gemmell, Aleks Aris, and Roger Lueder. Telling stories with mylifebits. In *2005 IEEE International Conference on Multimedia and Expo*, pages 1536–1539. IEEE, 2005.

[107] Katy Ilonka Gero and Lydia B Chilton. Metaphoria: An algorithmic companion for metaphor creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[108] Marjan Ghazvininejad, Xing Shi, Yejin Choi, and Kevin Knight. Generating topical poetry. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1183–1191, 2016.

[109] Matila Costiescu Ghyka. *The Geometry of Art and Life*. Courier Corporation, 1977.

[110] Tavis Leaf Glover. *Canon of Design: Mastering Artistic Composition*. Tavis Leaf Glover, 2014.

[111] Tavis Leaf Glover. *Photography Composition & Design: A Fun Approach for Serious Artists*. Tavis Leaf Glover, 2016.

[112] Tavis Leaf Glover. Dynamic symmetry grids: Unity, movement, rhythm & strength. `https://ipoxstudios.com/dynamic-symmetry-grids-for-photographers-and-painters/`, 2019.

[113] Floraine Grabler, Maneesh Agrawala, Wilmot Li, Mira Dontcheva, and Takeo Igarashi. Generating photo manipulation tutorials by demonstration. In *ACM SIGGRAPH 2009 papers*, pages 1–9. 2009.

[114] Stefan Gumhold. Maximum entropy light source placement. In *Visualization, 2002. VIS 2002. IEEE*, pages 275–282. IEEE, 2002.

[115] Anhong Guo, Xiang'Anthony' Chen, Haoran Qi, Samuel White, Suman Ghosh, Chieko Asakawa, and Jeffrey P Bigham. Vizlens: A robust and interactive screen reader for interfaces in the real world. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 651–664, 2016.

[116] Dong Guo and Terence Sim. Digital face makeup by example. In *Computer Vision and Pattern Recognition*, pages 73–79. IEEE, 2009.

[117] G. Guo, H. Wang, C. Shen, Y. Yan, and H. M. Liao. Automatic image cropping for visual aesthetic enhancement using deep neural networks and cascaded regression. *IEEE Transactions on Multimedia*, 20(8):2073–2085, Aug 2018.

[118] Philip J Guo. Codeopticon: Real-time, one-to-many human tutoring for computer programming. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 599–608, 2015.

[119] Matthew Guzdial, Nicholas Liao, Jonathan Chen, Shao-Yu Chen, Shukan Shah, Vishwa Shah, Joshua Reno, Gillian Smith, and Mark O Riedl. Friend, collaborator, student, manager: How design of an ai-driven game level editor affects creators. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–13, 2019.

[120] Gaëtan Hadjeres, François Pachet, and Frank Nielsen. Deepbach: a steerable model for bach chorales generation. In *International Conference on Machine Learning*, pages 1362–1371. PMLR, 2017.

[121] Jay Hambidge. *The Elements of Dynamic Symmetry*. Courier Corporation, 1967.

[122] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[123] Björn Hartmann, Scott R Klemmer, Michael Bernstein, Leith Abdulla, Brandon Burr, Avi Robinson-Mosher, and Jennifer Gee. Reflective physical prototyping through integrated design, test, and analysis. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, pages 299–308, 2006.

[124] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (TOG)*, 35(6):1–12, 2016.

[125] Li-wei He, Michael F Cohen, and David H Salesin. The virtual cinematographer: a paradigm for automatic real-time camera control and directing. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 217–224, 1996.

[126] S. He, Z. Zhou, F. Farhat, and J. Z. Wang. Discovering triangles in portraits for supporting photographic creation. *IEEE Transactions on Multimedia*, 20(2):496–508, Feb 2018.

[127] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 203–212, 2010.

[128] Mary Hegarty, Bernd Meyer, and N Hari Narayanan. *Diagrammatic Representation and Inference: Second International Conference, Diagrams 2002 Callaway Gardens, GA, USA, April 18-20, 2002 Proceedings*, volume 2317. Springer, 2003.

[129] Stephen Heyman. Photos, photos everywhere. `https://www.nytimes.com/2015/07/23/arts/international/photos-photos-everywhere.html`, Jul 2015.

[130] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. Sensecam: A retrospective memory aid. In *International Conference on Ubiquitous Computing*, pages 177–193. Springer, 2006.

[131] James Hollan, Edwin Hutchins, and David Kirsh. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(2):174–196, 2000.

[132] Jim Hollan and Scott Stornetta. Beyond being there. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 119–125, 1992.

[133] Kristina Höök, Phoebe Sengers, and Gerd Andersson. Sense and sensibility: evaluation and interactive art. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 241–248, 2003.

[134] Eric Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 159–166, 1999.

[135] Chuan-Shen Hu, Yi-Tsung Hsieh, Hsiao-Wei Lin, and Mei-Chen Yeh. Virtual portraitist: An intelligent tool for taking well-posed selfies. *ACM Trans. Multimedia Comput. Commun. Appl.*, 15(1s):12:1–12:17, January 2019.

[136] Barry Huggins. *Creative photoshop lighting techniques*. Lark Books, 2005.

[137] Mark J. Huiskes and Michael S. Lew. The mir flickr retrieval evaluation. In *MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval*, New York, NY, USA, 2008. ACM.

[138] Fil Hunter, Steven Biver, and Paul Fuqua. *Light Science & Magic: An Introduction to Photographic Lighting*. CRC Press, 2015.

[139] Hayato Huseman. The most important features in a phone - ranked by the ac staff. `https://www.androidcentral.com/android-central-staffs-most-important-features-phone`, May 2020.

[140] Edwin L Hutchins, James D Hollan, and Donald A Norman. Direct manipulation interfaces. *Human–computer interaction*, 1(4):311–338, 1985.

[141] Emmanuel Iarussi, Adrien Bousseau, and Theophanis Tsandilas. The drawing assistant: Automated drawing guidance and feedback from photographs. In *ACM Symposium on User Interface Software and Technology (UIST)*. ACM, 2013.

[142] Takeo Igarashi, Satoshi Matsuoka, Sachiko Kawachiya, and Hidehiko Tanaka. Interactive beautification: A technique for rapid geometric design. In *ACM SIGGRAPH 2007 courses*, pages 18–es. 2007.

[143] Insta360. Insta360 nano, 2018.

[144] Md Baharul Islam, Wong Lai-Kuan, and Wong Chee-Onn. A survey of aesthetics-driven image recomposition. *Multimedia Tools and Applications*, 76(7):9517–9542, Apr 2017.

[145] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[146] Jennifer Jacobs, Sumit Gogia, Radomír Měch, and Joel R Brandt. Supporting expressive procedural art creation through direct manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6330–6341, 2017.

[147] Michel Jacobs. *The Art of Composition: A Simple Application of Dynamic Symmetry*. Doubleday, Page & Company, 1926.

[148] Sungjune Jang, Lawrence H Kim, Kesler Tanner, Hiroshi Ishii, and Sean Follmer. Haptic edge display for mobile tactile interaction. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 3706–3716, 2016.

[149] David G Jansson and Steven M Smith. Design fixation. *Design studies*, 12(1):3–11, 1991.

[150] Pradthana Jarusriboonchai, Thomas Olsson, Sus Lundgren Lyckvi, and Kaisa Väänänen. Let's take photos together: exploring asymmetrical interaction abilities on mobile camera phones. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 529–540, 2016.

[151] About JasenkaG. The 6 fundamentals of storytelling through photography. `https://www.lightstalking.com/storytelling-through-photography/`, Feb 2020.

[152] Niels Joubert, Jane L. E, Dan B Goldman, Floraine Berthouzoz, Mike Roberts, James A Landay, Pat Hanrahan, et al. Towards a drone cinematographer: Guiding quadrotor cameras using visual composition principles. *arXiv preprint arXiv:1610.01691*, 2016.

[153] Niels Joubert, Mike Roberts, Anh Truong, Floraine Berthouzoz, and Pat Hanrahan. An interactive tool for designing quadrotor camera shots. *ACM Transactions on Graphics (TOG)*, 34(6):1–11, 2015.

[154] Baris Kandemir, Zihan Zhou, Jia Li, and James Z. Wang. Beyond saliency: Assessing visual balance with high-level cues. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, Thematic Workshops '17, pages 26–34, New York, NY, USA, 2017. ACM.

[155] Henry Kang, Seungyong Lee, and Charles K Chui. Flow-based image abstraction. *IEEE transactions on visualization and computer graphics*, 15(1):62–76, 2008.

[156] Solange Karsenty, James A Landay, and Chris Weikart. Inferring graphical constraints with rockit. In *People and Computers VII, Proceedings of HCI '92*, pages 137–153, 1992.

[157] Solange Karsenty, Chris Weikart, and James A Landay. Inferring graphical constraints with rockit. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, page 531, 1993.

[158] James C Kaufman and Ronald A Beghetto. Beyond big and little: The four c model of creativity. *Review of general psychology*, 13(1):1–12, 2009.

[159] Rubaiat Habib Kazi, Fanny Chevalier, Tovi Grossman, Shengdong Zhao, and George Fitzmaurice. Draco: bringing life to illustrations with kinetic textures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 351–360, 2014.

[160] Ira Kemelmacher-Shlizerman. Transfiguring portraits. *ACM Transactions on Graphics (TOG)*, 35(4):94, 2016.

[161] William B Kerr and Fabio Pellacini. Toward evaluating lighting design interface paradigms for novice users. In *ACM Trans. on Graphics (TOG)*, volume 28, page 26. ACM, 2009.

[162] Joy Kim, Mira Dontcheva, Wilmot Li, Michael S Bernstein, and Daniela Steinsapir. Motif: Supporting novice creativity through expert patterns. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1211–1220, 2015.

[163] Minju Kim and Jungjin Lee. Picme: Interactive visual guidance for taking requested photo composition. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 395:1–395:12, New York, NY, USA, 2019. ACM.

[164] Tim Kindberg, Mirjana Spasojevic, Rowanne Fleck, and Abigail Sellen. The ubiquitous camera: An in-depth study of camera phone use. *IEEE Pervasive Computing*, 4(2):42–50, 2005.

[165] David Kirsh and Paul Maglio. On distinguishing epistemic from pragmatic action. *Cognitive science*, 18(4):513–549, 1994.

[166] Scott R Klemmer, Björn Hartmann, and Leila Takayama. How bodies matter: five themes for interaction design. In *Proceedings of the 6th conference on Designing Interactive systems*, pages 140–149, 2006.

[167] Shu Kong, Xiaohui Shen, Zhe Lin, Radomír Měch, and Charles Fowlkes. Photo aesthetics ranking network with attributes and content adaptation. In *ECCV*, 2016.

[168] Yuki Koyama, Daisuke Sakamoto, and Takeo Igarashi. Selph: Progressive learning and support of manual photo color enhancement. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2520–2532, 2016.

[169] Bert Krages. *Photography: the art of composition*. Simon and Schuster, 2012.

[170] Chinmay Kulkarni, Steven P Dow, and Scott R Klemmer. Early and repeated exposure to examples improves creative work. In *Design thinking research*, pages 49–62. Springer, 2014.

[171] M. Kumano, K. Uehara, and Y. Ariki. Online training-oriented video shooting navigation system based on real-time camerawork evaluation. In *2006 IEEE International Conference on Multimedia and Expo*, pages 1281–1284, July 2006.

[172] Matthias Kümmerer, Thomas S. A. Wallis, and Matthias Bethge. Deepgaze II: reading fixations from deep features trained on object recognition. *CoRR*, abs/1610.01563, 2016.

[173] James A Landay and Brad A Myers. Interactive sketching for the early stages of user interface design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 43–50, 1995.

[174] Mackenzie Leake, Abe Davis, Anh Truong, and Maneesh Agrawala. Computational video editing for dialogue-driven scenes. *ACM Trans. Graph.*, 36(4):130–1, 2017.

[175] Brian Lee, Savil Srivastava, Ranjitha Kumar, Ronen Brafman, and Scott R Klemmer. Designing with interactive example galleries. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2257–2266, 2010.

[176] Chang Ha Lee, Xuejun Hao, and Amitabh Varshney. Geometry-dependent lighting. *IEEE TVCG*, 12(2):197–207, 2006.

[177] Dong-Hoo Lee. Digital cameras, personal photography and the reconfiguration of spatial experiences. *The Information Society*, 26(4):266–275, 2010.

[178] Joon-Young Lee, Kalyan Sunkavalli, Zhe Lin, Xiaohui Shen, and In So Kweon. Automatic content-aware color and tone stylization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2470–2478, 2016.

[179] Jun-Tae Lee, Han-Ul Kim, Chul Lee, and Chang-Su Kim. Photographic composition classification and dominant geometric element detection for outdoor scenes. *Journal of Visual Communication and Image Representation*, 55:91 – 105, 2018.

[180] Christian Lessig, Tyler de Witt, and Eugene Fiume. Efficient and accurate rotation of finite spherical harmonics expansions. *Journal of Computational Physics*, 231(2):243–250, 2012.

[181] Michael Lewis. Against the rules with michael lewis. `https://atrpodcast.com/`.

[182] Tommer Leyvand, Daniel Cohen-Or, Gideon Dror, and Dani Lischinski. Data-driven enhancement of facial attractiveness. In *ACM Transactions on Graphics (TOG)*, volume 27, page 38. ACM, 2008.

[183] D. Li, H. Wu, J. Zhang, and K. Huang. Fast a3rl: Aesthetics-aware adversarial reinforcement learning for image cropping. *IEEE Transactions on Image Processing*, 28(10):5105–5120, Oct 2019.

[184] Debang Li, Huikai Wu, Junge Zhang, and Kaiqi Huang. A2-rl: Aesthetics aware reinforcement learning for image cropping. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[185] Jia Li, Lei Yao, and James Z. Wang. *Photo Composition Feedback and Enhancement*, pages 113–144. Springer International Publishing, Cham, 2015.

[186] Qifan Li and Daniel Vogel. Guided selfies using models of portrait aesthetics. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, pages 179–190, New York, NY, USA, 2017. ACM.

[187] Z. Li and X. Zhang. Collaborative deep reinforcement learning for image cropping. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 254–259, July 2019.

[188] Y. Liang, X. Wang, S. Zhang, S. Hu, and S. Liu. Photorecomposer: Interactive photo recomposition by cropping. *IEEE Transactions on Visualization and Computer Graphics*, 24(10):2728–2742, Oct 2018.

[189] Sharon Lin, Daniel Ritchie, Matthew Fisher, and Pat Hanrahan. Probabilistic color-by-numbers: Suggesting pattern colorizations using factor graphs. *ACM Transactions on Graphics (TOG)*, 32(4):1–12, 2013.

[190] Zicheng Liu, Ying Shan, and Zhengyou Zhang. Expressive expression mapping with ratio images. In *Proc. of SIGGRAPH*, pages 271–276. ACM, 2001.

[191] Sara Ljungblad. Passive photography from a creative perspective: "if i would just shoot the same thing for seven days, it's like... what's the point?". In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 829–838, 2009.

[192] Kuo-Yen Lo, Keng-Hao Liu, and Chu-Song Chen. Intelligent photographing interface with on-device aesthetic quality assessment. In *Asian Conference on Computer Vision*, pages 533–544. Springer, 2012.

[193] Kuo-Yen Lo, Keng-Hao Liu, and Chu-Song Chen. Intelligent photographing interface with on-device aesthetic quality assessment. In Jong-Il Park and Junmo Kim, editors, *Computer Vision - ACCV 2012 Workshops*, pages 533–544, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[194] Barbara London, Jim Stone, and John Upton. *Photography*. Prentice Hall, 2011.

[195] Peng Lu, Hao Zhang, Xujun Peng, and Xiaofu Jin. An end-to-end neural network for image cropping by learning composition from aesthetic photos. *CoRR*, abs/1907.01432, 2019.

[196] Peng Lu, Hao Zhang, XuJun Peng, and Xiang Peng. Aesthetic guided deep regression network for image cropping. *Signal Processing: Image Communication*, 77:1 – 10, 2019.

[197] Chen Lujun, Yao Hongxun, Sun Xiaoshuai, and Zhang Hongming. Real-time viewfinder composition assessment and recommendation to mobile photographing. In Weisi Lin, Dong Xu, Anthony Ho, Jianxin Wu, Ying He, Jianfei Cai, Mohan Kankanhalli, and Ming-Ting Sun, editors, *Advances in Multimedia Information*

Processing – PCM 2012, pages 707–714, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[198] S. Ma, Y. Fan, and C. W. Chen. Finding your spot: A photography suggestion system for placing human in the scene. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 556–560, Oct 2014.

[199] Shuai Ma, Zijun Wei, Feng Tian, Xiangmin Fan, Jianming Zhang, Xiaohui Shen, Zhe Lin, Jin Huang, Radomír Měch, Dimitris Samaras, and Hongan Wang. Smarteye: Assisting instant photo taking via integrating user preference with deep view proposal network. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 471:1–471:12, New York, NY, USA, 2019. ACM.

[200] Adam Marelli. The great compositions of photographer alfred eisenstaedt. `http://petapixel.com/2013/04/15/the-great-compositions-of-photographer-alfred-eisenstaedt/`, 2013.

[201] Michael Marmor and James Ravin. *The artist's eyes.* Harry N. Abrams, 2009.

[202] Stephen R Marschner and Donald P Greenberg. Inverse lighting for photography. In *Color and Imaging Conference*, volume 1997, pages 262–265. Society for Imaging Science and Technology, 1997.

[203] Justin Matejka, Michael Glueck, Erin Bradner, Ali Hashemi, Tovi Grossman, and George Fitzmaurice. Dream lens: Exploration and visualization of large-scale generative design datasets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[204] Ann McNamara, Reynold Bailey, and Cindy Grimm. Search Task Performance Using Subtle Gaze Direction with the Presence of Distractions. *ACM Transactions on Applied Perception*, 6(3):17:1–17:19, 2009.

[205] Paul Merrell, Eric Schkufza, Zeyang Li, Maneesh Agrawala, and Vladlen Koltun. Interactive furniture layout using interior design guidelines. *ACM transactions on graphics (TOG)*, 30(4):1–10, 2011.

[206] Microsoft. PowerPoint 2013 Smart Guides. `https://www.youtube.com/watch?v=fouyvhVtxWg`, 2013.

[207] Hiroko Mitarai, Yoshihiro Itamiya, and Atsuo Yoshitaka. Interactive photographic shooting assistance based on composition and saliency. In *International Conference on Computational Science and Its Applications*, pages 348–363. Springer, 2013.

[208] Hiroko Mitarai, Yoshihiro Itamiya, and Atsuo Yoshitaka. Interactive photographic shooting assistance based on composition and saliency. In Beniamino Murgante, Sanjay Misra, Maurizio Carlini, Carmelo M. Torre, Hong-Quang Nguyen, David Taniar, Bernady O. Apduhan, and Osvaldo Gervasi, editors, *Computational Science and Its Applications – ICCSA 2013*, pages 348–363, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[209] Hiroko Mitarai and Atsuo Yoshitaka. Interactive video cam system for emotive video production. In *2011 IEEE International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2011.

[210] Naomi Miyake and Donald A Norman. To ask a question, one must know enough to know what is not known. *Journal of verbal learning and verbal behavior*, 18(3):357–364, 1979.

[211] Dave Montizambert. *Creative lighting techniques for studio photographers*. Amherst Media, Inc, 2003.

[212] Martez E Mott, Jane L. E, Cynthia L Bennett, Edward Cutrell, and Meredith Ringel Morris. Understanding the accessibility of smartphone photography for people with motor impairments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[213] J. Müller-Brockmann. *Grid Systems in Graphic Design: A Visual Communication Manual for Graphic Designers, Typographers and Three Dimensional Designers*. Visual communication books. Niggli, 1996.

[214] Lukas Murmann, Abe Davis, Jan Kautz, and Frédo Durand. Computational bounce flash for indoor portraits. *ACM Transactions on Graphics (TOG)*, 35(6):190, 2016.

[215] Frank Nack, Chitra Dorai, and Svetha Venkatesh. Computational media aesthetics: Finding meaning beautiful. *IEEE multimedia*, 8(4):10–12, 2001.

[216] DC. National Academy of Sciences-National Research Council, Washington. *How people learn: Brain, mind, experience, and school*. ERIC Clearinghouse, 1999.

[217] Markus Neuhäuser. Wilcoxon–mann–whitney test. In *International encyclopedia of statistical science*, pages 1656–1658. Springer, 2011.

[218] Ren Ng, Ravi Ramamoorthi, and Pat Hanrahan. All-frequency shadows using non-linear wavelet lighting approximation. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 376–381. ACM, 2003.

[219] Ren Ng, Ravi Ramamoorthi, and Pat Hanrahan. Triple product wavelet integrals for all-frequency relighting. *ACM Transactions on Graphics (TOG)*, 23(3):477–487, 2004.

[220] Tricia J Ngoon, C Ailie Fraser, Ariel S Weingarten, Mira Dontcheva, and Scott Klemmer. Interactive guidance techniques for improving creative feedback. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2018.

[221] Tricia J Ngoon, Caren M Walker, and Scott Klemmer. The dark side of satisficing: Setting the temperature of creative thinking. In *Proceedings of the 2019 on Creativity and Cognition*, pages 591–596. 2019.

[222] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. Vremiere: in-headset virtual reality video editing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5428–5438, 2017.

[223] Tony Northrup. *Tony Northrup's DSLR Book: How to Create Stunning Digital Photography.* Tony Northrup, 2014.

[224] Derek Nowrouzezahrai, Patricio Simari, and Eugene Fiume. Sparse zonal harmonic factorization for efficient sh rotation. *ACM Transactions on Graphics (TOG)*, 31(3):23, 2012.

[225] Hyacinth S Nwana. Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, 4(4):251–277, 1990.

[226] O. O2. The most important smartphone features ranked. `https://www.engadget.com/2017-07-10-most-important-smartphone-features-ranked.html`, Feb 2020.

[227] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[228] Amy Pavel, Dan B Goldman, Björn Hartmann, and Maneesh Agrawala. Vidcrit: video-based asynchronous video review. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 517–528, 2016.

[229] Pieter Peers, Naoki Tamura, Wojciech Matusik, and Paul Debevec. Post-production facial performance relighting using reflectance transfer. In *ACM Transactions on Graphics (TOG)*, volume 26, page 52. ACM, 2007.

[230] Fabio Pellacini. envylight: an interface for editing natural illumination. *ACM Transactions on Graphics (TOG)*, 29(4):34, 2010.

[231] Georg Petschnigg, Maneesh Agrawala, Hugues Hoppe, Richard Szeliski, Michael Cohen, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph.*, 23(3):664–672, August 2004.

[232] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2016.

[233] Hasso Plattner, Christoph Meinel, and Ulrich Weinberg. *Design-thinking*. Springer, 2009.

[234] Pedro Quintela. How to create storytelling photography that will captivate your audience. `https://www.photoblog.com/learn/storytelling-photography-tips/`, May 2019.

[235] Y. S. Rawat and M. S. Kankanhalli. Clicksmart: A context-aware viewpoint recommendation system for mobile photography. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(1):149–158, Jan 2017.

[236] Yogesh Singh Rawat. Real-time assistance in multimedia capture using social media. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 641–644, 2015.

[237] Yogesh Singh Rawat and Mohan S. Kankanhalli. Context-aware photography learning for smart mobile devices. *ACM Trans. Multimedia Comput. Commun. Appl.*, 12(1s):19:1–19:24, October 2015.

[238] Byron Reeves and Clifford Ivar Nass. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press, 1996.

[239] Katharina Reinecke and Krzysztof Z Gajos. Quantifying visual preferences around the world. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 11–20, 2014.

[240] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[241] Tammy Riklin-Raviv and Amnon Shashua. The quotient image: Class based recognition and synthesis under varying illumination conditions. In *Computer Vision and Pattern Recognition*, volume 2, pages 566–571. IEEE, 1999.

[242] Mike Roberts, Debadeepta Dey, Anh Truong, Sudipta Sinha, Shital Shah, Ashish Kapoor, Pat Hanrahan, and Neel Joshi. Submodular trajectory optimization for aerial 3d scanning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5324–5333, 2017.

[243] Paul Rogers. A look at composition in documentary photography. `https://petapixel.com/2015/03/23/a-look-at-composition-in-documentary-photography/`, Mar 2015.

[244] Fred Ross. *The Art of Composition: A Simple Approach for Creating Masterful Designs*. The Art of Composition.

[245] Szymon Rusinkiewicz, Michael Burns, and Doug DeCarlo. Exaggerated shading for depicting shape and detail. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 1199–1205. ACM, 2006.

[246] Jerry Saltz. *How to be an Artist*. Penguin, 2020.

[247] Donald Schön. The reflective practitioner. *New York*, 1083, 1938.

[248] Donald A Schön. *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions.* Jossey-Bass, 1987.

[249] Adrian Secord, Jingwan Lu, Adam Finkelstein, Manish Singh, and Andrew Nealen. Perceptual models of viewpoint preference. *ACM Transactions on Graphics (TOG)*, 30(5):1–12, 2011.

[250] Ram Shacked and Dani Lischinski. Automatic lighting design using a perceptual quality metric. In *Computer graphics forum*, volume 20, pages 215–227. Wiley Online Library, 2001.

[251] Xiaoyong Shen, Aaron Hertzmann, Jiaya Jia, Sylvain Paris, Brian Price, Eli Shechtman, and Ian Sachs. Automatic portrait segmentation for image stylization. In *Computer Graphics Forum*, volume 35, pages 93–102. Wiley Online Library, 2016.

[252] YiChang Shih, Sylvain Paris, Connelly Barnes, William T Freeman, and Frédo Durand. Style transfer for headshot portraits. 2014.

[253] Ben Shneiderman. Creativity support tools. *Communications of the ACM*, 45(10):116–120, 2002.

[254] Ben Shneiderman. Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM*, 50(12):20–32, 2007.

[255] Ben Shneiderman and Pattie Maes. Direct manipulation vs. interface agents. *interactions*, 4(6):42–61, 1997.

[256] Ben Shneiderman and Catherine Plaisant. *Designing the user interface: Strategies for effective human-computer interaction*. Pearson Education India, 2010.

[257] Zhixin Shu, Sunil Hadap, Eli Shechtman, Kalyan Sunkavalli, Sylvain Paris, and Dimitris Samaras. Portrait lighting transfer using a mass transport approach. *ACM Transactions on Graphics (TOG)*, 36(4):1, 2017.

[258] Peter-Pike Sloan, Jan Kautz, and John Snyder. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 527–536. ACM, 2002.

[259] Noah Snavely, Steven M Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM Siggraph 2006 Papers*, pages 835–846. 2006.

[260] Robert L. Solso. Brain activities in a skilled versus a novice artist: An fmri study. *Leonardo*, 34(1):31–34, 2001.

[261] Sony. Help guide | smile/face detect. `https://helpguide.sony.net/gbmig/45349331/v1/en/contents/TP0000518244.html`.

[262] Robert Spence. *Information visualization*, volume 1. Springer, 2001.

[263] sRGB Wikipedia Page. srgb — wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=SRGB&oldid=816935716, 2017.

[264] Srinivas Sridharan, Reynold Bailey, Ann McNamara, and Cindy Grimm. Subtle Gaze Manipulation for Improved Mammography Training. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 75–82, New York, NY, USA, 2012. ACM.

[265] Manohar Srikanth, Kavita Bala, and Frédo Durand. Computational rim illumination with aerial robots. In *Proceedings of the Workshop on Computational Aesthetics*, pages 57–66. ACM, 2014.

[266] Shawn C. Steiner. What is focus peaking? https://www.bhphotovideo.com/explora/photography/tips-and-solutions/what-focus-peaking, Oct 2015.

[267] Hsiao-Hang Su, Tse-Wei Chen, Chieh-Chi Kao, Winston H Hsu, and Shao-Yi Chien. Scenic photo quality assessment with bag of aesthetics-preserving features. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 1213–1216, 2011.

[268] Hsiao-Hang Su, Tse-Wei Chen, Chieh-Chi Kao, Winston H Hsu, and Shao-Yi Chien. Preference-aware view recommendation system for scenic photos based on bag-of-aesthetics-preserving features. *IEEE Transactions on Multimedia*, 14(3):833–843, 2012.

[269] Lucy A Suchman. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press, 1987.

[270] Kalyan Sunkavalli, Fabiano Romeiro, Wojciech Matusik, Todd Zickler, and Hanspeter Pfister. What do color changes reveal about an outdoor scene? In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[271] Masaki Suwa and Barbara Tversky. External representations contribute to the dynamic construction of ideas. In *International Conference on Theory and Application of Diagrams*, pages 341–343. Springer, 2002.

[272] Amanda Swearngin, Mira Dontcheva, Wilmot Li, Joel Brandt, Morgan Dixon, and Andrew J Ko. Rewire: Interface design assistance from examples. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[273] John Szarkowski. *Looking at photographs: 100 Pictures from the Collection of the Museum of Modern Art*. New York Graphic Society, Greenwich, Conn., 1973.

[274] Michael Terry and Elizabeth D Mynatt. Recognizing creative needs in user interface design. In *Proceedings of the 4th conference on Creativity & cognition*, pages 38–44, 2002.

[275] Michael Terry and Elizabeth D Mynatt. Side views: persistent, on-demand previews for open-ended tasks. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, pages 71–80, 2002.

[276] Inc. The MathWorks. makehdr. `https://www.mathworks.com/help/images/ref/makehdr.html`, 2007-2014.

[277] Inc. The MathWorks. rgb2lin. `https://www.mathworks.com/help/images/ref/rgb2lin.html`, 2007-2014.

[278] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*, pages 839–846. IEEE, 1998.

[279] Wai-Shun Tong, Chi-Keung Tang, Michael S Brown, and Ying-Qing Xu. Example-based cosmetic transfer. In *Computer Graphics and Applications, 2007. PG'07. 15th Pacific Conference on*, pages 211–218. IEEE, 2007.

[280] Philip HS Torr and Andrew Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, 2000.

[281] Bill Triggs and Christian Laugier. Automatic camera placement for robot vision tasks. In *Proceedings of 1995 Ieee International Conference on Robotics and Automation*, volume 2, pages 1732–1737. IEEE, 1995.

[282] Anh Truong, Floraine Berthouzoz, Wilmot Li, and Maneesh Agrawala. Quickcut: An interactive tool for editing narrated video. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 497–507, 2016.

[283] Barbara Tversky. *Mind in motion: How action shapes thought*. Hachette UK, 2019.

[284] José Van Dijck. Digital photography: communication, identity, memory. *Visual communication*, 7(1):57–76, 2008.

[285] Kurt VanLehn. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4):197–221, 2011.

[286] Marynel Vázquez and Aaron Steinfeld. Helping visually impaired users properly aim a camera. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, pages 95–102, 2012.

[287] Marynel Vázquez and Aaron Steinfeld. An assisted photography framework to help visually impaired users properly aim a camera. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(5):1–29, 2014.

[288] Romain Vergne, Romain Pacanowski, Pascal Barla, Xavier Granier, and Christophe Schlick. Light warping for enhanced surface depiction. In *ACM Transactions on Graphics (TOG)*, volume 28, page 25. ACM, 2009.

[289] Luis Von Ahn, Shiry Ginosar, Mihir Kedia, and Manuel Blum. Improving image search with phetch. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 4, pages IV–1209. IEEE, 2007.

[290] Neal Wadhwa, Rahul Garg, David E Jacobs, Bryan E Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T Barron, Yael Pritch, and Marc Levoy. Synthetic depth-of-field with a single-camera mobile phone. *ACM Transactions on Graphics (ToG)*, 37(4):1–13, 2018.

[291] Hong Wang and Peter Chu. Voice source localization for automatic camera pointing system in videoconferencing. In *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 187–190. IEEE, 1997.

[292] W. Wang, J. Shen, and H. Ling. A deep network solution for attention and aesthetics aware photo cropping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1531–1544, July 2019.

[293] Yinting Wang, Mingli Song, Dacheng Tao, Yong Rui, Jiajun Bu, Ah Chung Tsoi, Shaojie Zhuo, and Ping Tan. Where2stand: A human position recommendation system for souvenir photography. *ACM Trans. Intell. Syst. Technol.*, 7(1):9:1–9:22, October 2015.

[294] Colin Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2019.

[295] Zijun Wei, Jianming Zhang, Xiaohui Shen, Zhe Lin, Radomír Mech, Minh Hoai, and Dimitris Samaras. Good view hunting: Learning photo composition from dense view pairs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5437–5446, 2018.

[296] Zijun Wei, Jianming Zhang, Xiaohui Shen, Zhe Lin, Radomír Měch, Minh Hoai, and Dimitris Samaras. Good view hunting: Learning photo composition from dense view pairs. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[297] Mark Weiser. The computer for the 21st century. *ACM SIGMOBILE mobile computing and communications review*, 3(3):3–11, 1999.

[298] Liz Wells. *Photography: a critical introduction*. Routledge, 2015.

[299] James Wen and Ayça Ünlüer. Redefining the fundamentals of photography with cooperative photography. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, pages 37–47, 2015.

[300] Zhen Wen, Zicheng Liu, and Thomas S Huang. Face relighting with radiance environment maps. In *Computer Vision and Pattern Recognition*, volume 2, pages II–158. IEEE, 2003.

[301] Mark E Whiting, Grant Hugh, and Michael S Bernstein. Fair work: Crowd work minimum wage with one line of code. 2019.

[302] Nora S Willett, Rubaiat Habib Kazi, Michael Chen, George Fitzmaurice, Adam Finkelstein, and Tovi Grossman. A mixed-initiative interface for animating static pictures. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pages 649–661, 2018.

[303] Holger Winnemöller, Sven C Olsen, and Bruce Gooch. Real-time video abstraction. *ACM Transactions On Graphics (TOG)*, 25(3):1221–1226, 2006.

[304] RF Woolson. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pages 1–3, 2007.

[305] Haijun Xia, Bruno Araujo, Tovi Grossman, and Daniel Wigdor. Object-oriented drawing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4610–4621, 2016.

[306] Jun Xie, Aaron Hertzmann, Wilmot Li, and Holger Winnemöller. Portraitsketch: Face sketching assistance for novices. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 407–417, 2014.

[307] Pengfei Xu, Hongxun Yao, Rongrong Ji, Xian-Ming Liu, and Xiaoshuai Sun. Where should i stand? learning based human position recommendation for mobile photographing. *Multimedia Tools and Applications*, 69(1):3–29, Mar 2014.

[308] Yan Xu, Joshua Ratcliff, James Scovell, Gheric Speiginer, and Ronald Azuma. Real-time guidance camera interface to enhance photo aesthetic quality. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1183–1186, New York, NY, USA, 2015. ACM.

[309] Jianzhou Yan, Stephen Lin, Sing Bing Kang, and Xiaoou Tang. Change-based image cropping with exclusion and compositional features. *International Journal of Computer Vision*, 114(1):74–87, Aug 2015.

[310] Lei Yao, Poonam Suryanarayan, Mu Qiao, James Z. Wang, and Jia Li. Oscar: On-site composition and aesthetics feedback through exemplars for photographers. *International Journal of Computer Vision*, 96(3):353–383, Feb 2012.

[311] Manfredas Zabarauskas and Stephen Cameron. Luke: An autonomous robot photographer. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1809–1815. IEEE, 2014.

[312] Hui Zeng, Lida Li, Zisheng Cao, and Lei Zhang. Reliable and efficient image cropping: A grid anchor based approach. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[313] F. Zhang, M. Wang, and S. Hu. Aesthetic image enhancement by dependence-aware object recomposition. *IEEE Transactions on Multimedia*, 15(7):1480–1490, Nov 2013.

[314] Michael Zhang. There are now 8x more people taking pictures than 10 years ago. `https://petapixel.com/2015/12/15/there-are-now-8x-more-people-taking-pictures-than-10-years-ago/`, Dec 2015.

[315] X. Zhang, Z. Li, M. Constable, K. L. Chan, Z. Tang, and G. Tang. Pose-based composition improvement for portrait photographs. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(3):653–668, March 2019.

[316] Y. Zhang, X. Sun, H. Yao, L. Qin, and Q. Huang. Aesthetic composition represetation for portrait photographing recommendation. In *2012 19th IEEE International Conference on Image Processing*, pages 2753–2756, Sep. 2012.

[317] Nanxuan Zhao, Nam Wook Kim, Laura Mariah Herman, Hanspeter Pfister, Rynson WH Lau, Jose Echevarria, and Zoya Bylinskii. Iconate: Automatic compound icon generation and ideation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[318] Z. Zhou, F. Farhat, and J. Z. Wang. Detecting dominant vanishing points in natural scenes with application to composition-sensitive image retrieval. *IEEE Transactions on Multimedia*, 19(12):2651–2665, Dec 2017.