Research Statement

Yağız Aksoy

The view seen in the photograph above gives the impression to most of us that it is a space of tranquility, at a peaceful time, with a calming effect. If you alter the scene in simple ways, such as removing the lake from the view or changing the time of the day to be midday, these intense impressions of the scene simply become lost.

In the future, we will attempt to create artificial intelligence with higher-level capabilities such as empathizing with humans. On the way to true artificial consciousness, we will devise artificial intelligence that is capable of doing creative tasks such as storytelling, authoring a novel, or creating visual art, which requires high-level cognitive processes such as *being inspired*. In addition, if the intention behind a creative output is to induce, say, awe in humans, the machine will require an abstract understanding of the feeling of awe. It is very common for cognitive processes like intense feelings or inspiration to begin with a visual stimulus of particular significance. This intimate connection between high-level processes and visual stimuli is a strong incentive to study the cognitive effects of visual stimuli in humans on the way to generalizing or studying such concepts in the context of a true artificial intelligence.

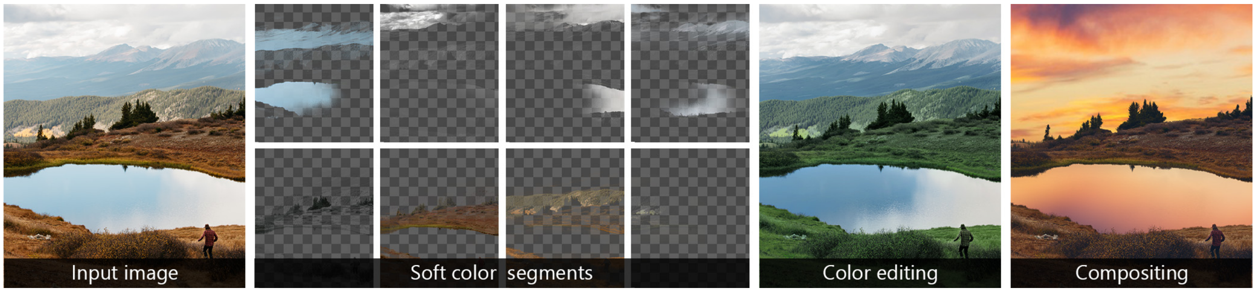
Towards this goal, we will eventually come closer to understanding how human consciousness works through the cooperation of many scientific and artistic disciplines. In terms of visual perception, we will start reasoning about questions such as why the photograph above represents tranquility for us, why sunsets in particular create awe, among many others. For such a level of machine understanding, the first thing we need is a better understanding of how this scene came together with the objects and their attributes in the view, how the scene is illuminated, and what kind of an environment they represent, a detailed *analysis* of the image. Moreover, in order to be able to freely alter the scene or create new such scenes for teaching itself higher level tasks, this machine will require the ability to re-create the scene, a precise *synthesis* of the scene. Bringing these tasks together will be stepping stone in the pursuit of the understanding our visual system and its role in consciousness. One of the most important challenges for visual computing on the way to artificial consciousness will be bridging two of its major sub-disciplines, computer vision and computer graphics.

My doctoral research is at the intersection of computer vision and computer graphics, focusing on generating different representations of photographs that allow easier analysis of color, illumination or semantics in images. My first research goal is to create stronger links between computer graphics and computer vision by studying illumination and materials in visual imagery, in order to be able to accurately and completely resynthesize a given scene. This will allow the use of computer graphics in cognitive and psychological research. I will build stronger connections between these fields in order to study deeper aspects of visual cognition in humans. In the end, my main research vision is to bring computer graphics to a point where it can become a major driver in the pursuit of a general artificial intelligence.

Past and Current Research

During my Ph.D., I have been a part of groups that individually focus on computer vision, computer graphics, and video processing. Simultaneously being a part of groups with diverse backgrounds and research interests allowed me to develop my own research agenda that can be positioned at the intersection of computer vision and computer graphics, as well as occasional use of methodologies in human-computer interaction. The focus of my doctoral research has been analyzing photographs through different lenses by breaking the images apart in terms of fundamental properties that underly the appearance and content of a scene. My research began with addressing real-world problems in movie post-production. I extended the theory behind the solutions to these problems to automatic and accurate decomposition of images into main colors, main objects and main sources of illumination.

Compositing, combining multiple images to produce a novel scene, is an important part of movie post-production and a labor-intensive process which demands specialized artists. For a realistic composition, targeted objects in a scene are segmented with a focus on accurate soft transitions between image regions. The soft segmentation can be done in controlled environments, as seen on the right. My Ph.D. started with this controlled setup, where I introduced an interactive approach [TOG16] that depends on a parametric representation of scene colors to segment out the green background. I generalized this application scenario to images without controlled backgrounds, a research problem referred to as natural matting. I introduced a graph-based approach in which the relationship between image regions are conceptualized as information flows [CVPR17] that represent how the information coming from the user input can be redirected and diffused to cover the whole image. My approach is easily generalizable to guided image filtering which was utilized to increase 3D modeling quality by one of my Master's students. These new approaches for interactive soft segmentation went on to become prototypes in collaboration with visual effects professionals under the Disney umbrella. More importantly, they paved the way to create generic image representations, which defines the core of my thesis research.



An important low-level cue that defines the overall look of a photograph is its main colors. These main scene colors, once identified, can represent each pixel in the image through linear combinations. The main challenge here is to represent different mixtures of colors in each pixel in terms of the main colors in the image. This is a vastly underconstrained problem, and the common solution was a global optimization which incorporates spatial smoothness. This results in prohibitively long computation times and spatial smoothness issues due to the iterative algorithms getting stuck in local minima. I deconstructed this problem into three main steps, where each step can be implemented with computational efficiency and a straightforward parallelization [TOG17]. The first step is through a novel energy formulation, sparse color unmixing, an extension of [TOG16], which estimates the linear weights and colors for each pixel independently. Other steps ensure spatial smoothness, and finally, the nonlinear constraints that are required to fully represent the image in terms of its colors. My approach gives decompositions that is accurate enough to realistically edit images unlike previous methods while taking orders of magnitude less time to compute. Currently, my students and I are working on using this representation to analyze out-of-focus blur for 3D reconstruction.

As a high-level decomposition of a photograph, I introduced semantic soft segmentation [TOG18], which brings together semantic segmentation, detection of objects in the scene, and natural matting in a single formulation for the first time. A data-driven method is typically required to bring in semantics, or objectness, to a system, but soft segmentation data is hard to collect, as well as the ground-truth being ill-defined in most cases. In the end, the solution I proposed makes use of a neural network to understand the semantics together with a spectral segmentation formulation for soft segmentation. In order to create a single elegant spectral decomposition formulation, I fuse the high-level information coming from the neural network with the low-level information on soft color transitions in the same graph structure. The neural network, instead of learning to generate the final segmentation, learns to produce semantic features that best fit the graph structure formulated. With a careful graph design that extends the information-flow concept [CVPR17], the formulation produces the soft segments with accurate boundaries between objects through the spectral analysis of the graph, effectively fusing the information from the neural network into our mathematical modeling of image segmentation. This work extends the potential of neural networks by using them to support our mathematical modeling of image formation where our current formulations come short, in this case towards a semantic understanding of the scene.

Finally, a physically realistic approach to image decomposition is through illumination, using the superposition of illumination principle. For instance, a flash photograph can be represented as the summation of the ambient illumination and the flash. Decomposing a flash photograph into these components is a challenging and underconstrained problem with many application scenarios such as single-image depth estimation using the special nature of flash. I introduced a novel crowd-sourcing methodology to collect a dataset of flash and ambient illumination pairs and used the dataset to train a deep network for illumination decomposition. To crowdsource a computational photography dataset, I devised an approach that uses the crowd as casual photographers who takes the photographs themselves to create the dataset for the first time. Using this dataset, I am currently researching how different combinations of flash and ambient illuminations effect perception of the final flash photograph.

Future Research

Research thrust I: Illumination and material analysis

In order to advance computer graphics towards a more complete physical understanding of photographs, I will study illumination and materials in digital imagery. My research will focus on extracting physically meaningful information about illumination and material properties through making use of our current modeling of image formation from computer graphics together with higher-level information extracted from the photograph through machine learning and computer vision. Following my approach to semantic soft segmentation [TOG18], my goal is to join data-driven methods with our physical modeling of how illumination and geometry interact with each other. Geometry, illumination, and materials all jointly contribute to the formation of a photograph, and developing a better understanding of any of them will benefit our modeling of others, with the ultimate goal being resynthesizing the whole scene in a rendering engine from a single photograph.

Research thrust II: Editing virtual worlds

My findings from the previous thrust will open new possibilities on how we understand and edit images. I will study the challenges that limit the perception of the virtual worlds by humans by freely editing and augmenting real scenes to study the individual effects of particular scene characteristics. One current problem is a limited feeling of presence in virtual worlds. By changing the structure of the scene, the illumination parameters and how specific materials appear, I will parameterize which aspects of a virtual world can be altered to make it a more welcoming space for humans.

Strongly connected analysis and synthesis systems also open up the field of next-generation image editing. Being able to alter the scene structure and illumination will be a valuable tool for amateur and professional content creators targeting a variety of media. I intend to discover new interactive techniques that enable editing digital imagery in a physically realistic way such as changing the intensity of individual light sources in a scene that will be useful for artists extending their media to include augmented reality and virtual reality devices. This richer representation will also enable fully digital movie studios, where the studios will be able to create visual effects easily in shots taken in the real world instead of controlled studio environments. This line of research will induce collaborations with research groups working on human-computer interaction.

Research thrust III: Towards understanding and simulating human perception

The next-generation scene editing capabilities have a lot of potential to create carefully controlled visual stimuli for cognitive and perceptual studies. This line of research will trigger increasing collaborations between computer graphics and cognitive sciences, neuroscience and psychology, as the researchers in these fields get access to increasingly useful tools from computer graphics. With these tools at hand, I will start exploring which attributes of photographs or scenes induce specific emotional responses in humans, as well as questions in spatial cognition, what aspects of the world around us affect our mood or decision making.

Answering these high-level questions about the effects of visual stimuli on humans and creating quantitative models with the help of physically realistic scene synthesis tools are important steps towards creating artificial systems that can imitate or simulate human responses. At this point, with collaborations with artificial intelligence researchers, I would like to start exploring the possibilities of a more capable AI in terms of visual cognition through computer graphics.

Conclusion

My research is at the intersection of computer graphics and computer vision, and my focus has been finding new ways to analyze and understand photographs. I will extend our understanding of visual stimuli in both machine perception and human vision, in order to develop a higher level of artificial intelligence. With this mission in mind, I believe computer graphics has the potential to make significant contributions to our understanding of visual perception, and this defines the primary goal for my future research.

References:

[TOG16] Yağız Aksoy, Tunç Ozan Aydın, Marc Pollefeys and Aljoša Smolíc, Interactive High-Quality Green-Screen Keying via Color Unmixing, ACM Transactions on Graphics, 2016 (Presented in SIGGRAPH 2017)

[TOG17] Yağız Aksoy, Tunç Ozan Aydın, Aljoša Smolíc and Marc Pollefeys, Unmixing-Based Soft Color Segmentation for Image Manipulation, ACM Transactions on Graphics, 2017 (Presented in SIGGRAPH 2017)

[CVPR17] Yağız Aksoy, Tunç Ozan Aydın and Marc Pollefeys, Designing Effective Inter- Pixel Information Flow for Natural Image Matting, Proc. CVPR, 2017

[TOG18] Yağız Aksoy, Tae-Hyun Oh, Sylvain Paris, Marc Pollefeys and Wojciech Matusik, Semantic Soft Segmentation, ACM Transactions on Graphics, 2018 (Proc. SIGGRAPH)

[ECCV18] Yağız Aksoy, Changil Kim, Petr Kelnhofer, Sylvain Paris, Mohamed Elgharib, Marc Pollefeys and Wojciech Matusik, A Dataset of Flash and Ambient Illumination Pairs from the Crowd, Proc. ECCV, 2018