The goal of my research is to develop computer vision, machine learning, and graphics techniques to understand, analyze, and visualize imagery depicting changes over time. Advances in this direction have applications within computer vision and in other contexts ranging from cinematic experiences to large-scale environmental monitoring.

The tools for capturing and sharing photos and videos have exploded to the point of ubiquity. Internet photo collections of popular landmarks now contain over a decade's worth of photographs; thousands of timelapse scenes are available on video sharing websites; live cameras stream real-time videos of many scenes 24 hours a day; high-resolution images of an entire face of the earth are available at 10-minute intervals from state-of-the-art weather satellites. These data sources contain a stunning amount of information, but extracting it remains challenging.

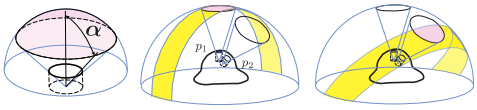
Understanding how scenes change over time is challenging because of the many factors affecting a scene's appearance. My research aims to untangle, analyze, and visualize changes over time due to illumination and motion. For example, given an unstructured collection of photos of a scene, can we determine the illumination conditions in each photo? How does this illumination relate to capture time? Given a long video of a scene, can we isolate and visualize the scene's appearance changes over time? Can we decompose and quantify appearance changes by frequency and cause (i.e., illumination vs motion)? Can we isolate long-term appearance changes, such as material weathering or plant growth, from short-term phenomena?

These questions have exciting applicability in art (e.g., timelapse cinematography), visualization (e.g., video summarization), pervasive sensing and environmental monitoring. Tackling these challenges requires bringing together traditional computer vision techniques based on physical models of image formation, modern deep learning techniques that bring semantic insight, and methods from computer graphics for compositing and visualization. After discussing my thesis work, I describe several projects for future work in which I hope to involve undergraduate researchers.

Thesis work

Illumination in outdoor photo collections

My early PhD work focused on analyzing illumination and determining capture time in outdoor photo collections. Estimating physical properties of outdoor scenes from unstructured photo collections requires reasoning about the complex interactions among geometry, materials, and illumination. Inverse rendering approaches attempt to jointly model all of these factors, but often produce disappointing results, either by falling into local minima or using overly simplistic models to achieve tractable optimization. In contrast, my work on illumination in photo collections factored out geometry and materials by using 3D reconstructions. In [1], I used models of sun/sky illumination to estimate the statistics of natural illumination, which can then be matched to the statistics observed across many images in a photo collection. Analyzing illumination statistics required accounting for the effects of position on the earth, weather conditions, and varying cameras.



Outdoor illumination statistics depend on local occlusion, surface normal, and geographic location.

In [2], I took another approach to understanding illumination by focusing on shadows: by looking at ratios of illumination between different points in a scene, my method estimates shadow labels

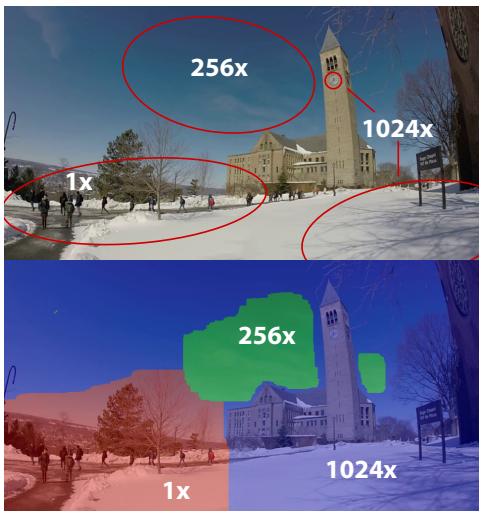
at each point in the scene. Given shadow labels and reconstructed geometry, simple techniques can then be used to infer the time of day under sunlight, which provides a straightforward path to leveraging physically-based illumination models. Modeling outdoor scenes is a challenging endeavor, and the key utility of this work was to provide robust illumination information that can be used to bootstrap further progress in disentangling and modeling how a scene's geometry, materials, and illumination explain its appearance.

Motion segmentation in video

When studying the (typically) long-term illumination changes in scenes, it is helpful to ignore moving objects; in photo collections, we do this using 3D reconstruction techniques that reconstruct only scene elements that appear in many photos over time. In video data, however, motion often explains the majority of appearance changes. A key step on the path to understanding motion in videos is to separate moving objects from static backgrounds. While general motion estimation remains challenging after decades of study, motion of rigid backgrounds can be modeled using quite simple techniques. In my recent video segmentation paper [3] I proposed to model the motion (due to a moving camera) of a rigid scene background in order to segment moving foreground objects. My approach achieves state-of-the-art results in motion-based segmentation with simpler and more efficient techniques.

Timelapse scene summarization

Timelapse video represents a compelling medium for visualizing changes that occur in a scene over longer time-spans not observable in real-time. However, many scenes have multiple interesting phenomena that occur at different timescales; for example, clouds drift through the sky at one rate, while the sun crosses the sky at a much slower rate. In an ongoing project, I aim to create novel visualizations that show phenomena occurring at several timescales in a single output video. Starting with a real-time input video (e.g., 30fps), I begin by speeding the video up by different rates and using optical flow to identify dynamic scene elements in each timescale at a pixel-granular level. I then use a compositing technique based on graph cut optimization that selects the source timescale for each output pixel based on its movement speed while maintaining a smooth, seamless output. The key impact of this work is to enable a new form of "scene summary", qualitatively illustrating all the types of dynamic phenomena happening in a scene in a short clip.



Top: people, clouds, and shadows are naturally viewed at different timelapse speed. Bottom: the proposed technique automatically selects and seamlessly blends different scene content to concisely summarize the activity in the scene at different timescales.

Future Work

One exciting direction for further research is to leverage and adapt deep learning techniques for temporal analysis

and visualization settings. Below, I propose several projects in this direction that are well-suited to undergraduates. The applications are accessible and numerous, and I would be excited to see the direction of these projects be driven by student interest. The background knowledge and setup time necessary to get involved is relatively minimal. For projects in deep learning, popular frameworks enable fast-paced experimentation, and the expertise gained is widely applicable to a multitude of domains, both in industry and academia.

Edge-aware deep feature upsampling

Convolutional neural networks have taken computer vision by storm, beginning with their sudden dominance in image recognition problems: convolutional architectures are able to learn surprisingly good feature spaces for classification tasks. These architectures aggregate context and extract successively higher-level feature vectors, but discard spatial detail in the process. For this reason, it remains challenging to use neural networks for per-pixel tasks that are useful in graphics and visualization applications, such as segmentation, depth estimation, and matting. The dominant solution to this problem is to add a series of layers (the "decoder") that mirror the downsampling part of the network (the "encoder") by increasing the spatial dimensionality and reducing the channel depth. State-of-the-art outputs have low visual quality, even when their numerical performance is good.

To address visual quality in neural network outputs, I plan to draw on insights from edge-aware upsampling. In particular, there are two interesting questions to answer:

1. Can edge-aware upsampling be applied to feature maps to improve the visual quality of predictions?
2. Can current neural network architectures be trained to make predictions in some alternate domain that is better suited to visual quality, then transform the output back to image space?

The first question can be answered by experimenting with using edge-aware upsampling techniques, such as bilateral or guided upsampling, in place of the upsampling operators used in current decoder architectures. The second question involves making predictions in some representation that is used (implicitly or explicitly) by an edge-aware filtering technique, then deterministically transforming the result back into pixel space. This representation could be a bilateral grid (as in the bilateral filter) or coefficients of a patch-wise locally affine model (as in the guided filter).

Semantic smoothness

Another underexplored way to leverage semantics in graphics outputs is to incorporate information from semantic features into conventional techniques. Image compositing techniques using graph cuts typically enforce smoothness based on color, for example by encouraging seams along image edges where they will be least obvious. However, this reasoning is often too low-level: a color edge between shirt and pants might be selected to have a seam, and the output contains only half a person. Incorporating semantic features could allow for smoothness terms that encourage seams only along image edges that also correspond to semantic changes.

This idea has broad applicability and could improve and enable a wide range of applications, including the video segmentation and timelapse projects discussed above. Good segmentation and semantically-aware compositing could enable a variety of new visualizations that layer moving objects

on backgrounds, such as turning panning videos into panoramic videos with moving foreground objects. It could also vastly improve image and video editing capabilities, giving users intuitive control over much higher-level elements of an image or video than are currently available. For example, a semantically-aware "magic wand" Photoshop tool could select pixels of an image not only with similar color but also similar semantic features as the selected point in the image.

Timelapse data for illumination understanding

Timelapse videos also have the potential to help us better model illumination changes over time; particularly if moving objects can be ignored, timelapse data could provide a rich source of ground truth for learning to predict or simulate different illumination conditions; for example, an interesting learning-based application (also well-suited for undergraduate researchers) would be to use timelapses as training data to learn to predict image timestamps.

Casual capture for motion timelapse

A long-term research goal that ties together and motivates much of my research agenda is enabling casual capture of motion timelapse videos, which are currently limited to professional photographers with expensive motion stage equipment. A user would capture a static timelapse and a short video with camera motion along the desired path for the timelapse. Simulating a motion timelapse would require accurately modeling the static geometry of the scene, the motion of dynamic objects, and the effects of changing illumination on the appearance over time.

References

[1] Daniel Hauagge, Scott Wehrwein, Paul Upchurch, Kavita Bala, and Noah Snavely. Reasoning about photo collections using models of outdoor illumination. In *British Machine Vision Conference*, 2014.

[2] Scott Wehrwein, Kavita Bala, and Noah Snavely. Shadow detection and sun direction in photo collections. In *International Conference on 3D Vision*, 2015.

[3] Scott Wehrwein and Richard Szeliski. Video segmentation with background motion models. In *British Machine Vision Conference*, 2017.