Physically grounded image editing

Keywords. Graphics, vision, image editing, scene understanding, reconstruction, illumination

Introduction. Imagine the ability to manipulate images as if the 3D spatial layout of the scene were respected. Objects could be effortlessly placed into the scene so that they are lit properly, cast accurate shadows, and the position and scale of the objects would be consistent with the scene. An object could be removed and seamlessly replaced by the surfaces it occludes, relighting the scene as if the object were never there. The 3D information in the scene would allow for changing the viewpoint of a camera, turning one image into many. These concepts are all 3D phenomena, which makes them especially difficult to achieve using purely 2D manipulations.

This spatial information required to perform these manipulations is abundant in images as evidenced by a human's ability to grasp the layout of a scene from a single image. However, current editing software only allows 2D manipulations with no regard to the high level spatial information that is present when a person looks at a photograph. *I propose to extract 3D scene information from digital images to allow for seamless object insertion, removal, and relocation.* An amateur editor can use these techniques to achieve high quality interior design or animation renderings. Even for experts, inserting and removing objects with current image editing software is extremely tedious or even impossible.

Research Plan. High level scene information is present in images, but collecting such information automatically is an open problem. Recent advances in this field have created an opportunity to pursue and eventually solve this issue. Past research has achieved simple but compelling 3D reconstructions for single images, including recovering surface orientation and object depth [3]. Also, given multiple images, recovering scene geometry is well understood [2]. These previous works provide motivation for enhancing digital image editing, but no research has been done to combine and extend these results to create a powerful image manipulation tool.

I plan to draw from these seminal works and go beyond them to create novel approaches to image editing. I will create algorithms that allow for object removal, object insertion and changing the viewpoint in images. Once these methods have been realized, I will employ them in a new kind of image editor that will enable seamless 3D manipulations of a 2D image. Additionally, multiple images can provide more accurate scene information, which may improve the editor's functionality.

To *remove an object* from an image, the pixels that belong to an object must be known. A mechanism must also exist for replacing discarded pixels occluded by the object. Interactive image matting can determine which pixels belong to the foreground and background, but I will devise more intelligent selection methods using camera parameters and other 3D scene information. Similarly, I am confident that this additional image information can be used to robustly synthesize the textures behind removed objects by rectifying planar segments of an image into head-on views. Traditional texture synthesis implementations [1] will not suffice in generating these textures due to the perspective layout of these scenes.

A considerable obstacle of *inserting objects* is illumination. Lalonde et al. describes how to insert objects into images under very specific conditions [4], and I will extend these conclusions to develop a more general lighting model for inserted objects. I believe that using a local lighting model on a uniform grid may be one way to improve the lighting of inserted objects. Initially, I will focus on inserting objects with known 3D models, but eventually I will attempt to insert image-based models. I will also research methods to accurately cast shadows from inserted objects.

As an extension to these editing techniques, I will explore rendering slight *changes to the original*

viewpoint and *free space estimation*. Texture synthesis is a necessary component of viewpoint change for single images because occluded pixels in the original view must be synthesized in the new view. Also, knowing where an object can be physically placed in the room (i.e. the free space) will allow for higher quality results and an enhanced user experience.

To evaluate this research, I will employ a user study in which subjects will use current image editing software (e.g. Photoshop) as well as the proposed techniques to complete editing tasks. The aesthetics of the results, task workload, and timing can all be used to qualify the differences between these methods. Similar studies can be done for specific proposed tools. For example, previous synthesis mechanisms can be compared to the ones used in the framework.

In preparation for this proposal, I have completed work on automatically discovering a representation of a single image given a cuboid parameterization. With this parameterization, I can rectify the projected portions of the image into head-on views, which provides a coarse 3D reconstruction of the image. I have created a user interface combining these algorithms which will serve as a prototype for the proposed image editing techniques.

My previous knowledge of image segmentation and image processing will aid me in this research. This research is at the intersection of computer graphics and computer vision, and I will work with professors from both fields. I will collaborate with Prof. David Forsyth and Prof. Derek Hoiem, who have extensive experience on topics such as multi-view geometry, recovering scene layout, and illumination models. I will also work with Prof. John Hart on other topics of this proposal, such as object geometry and rendering. The NSF GRF will allow me the freedom to work with all of these professors without being constrained to a certain research topic. With their guidance, I am confident that I will achieve the goals I have proposed.

Intellectual Merit. This research will enhance and simplify current image editing methods by incorporating 3D information into the editing process. An image is brimming with information other than pixel data, and image editors will finally be able to harness this additional information to make better and easier edits. Tedious editing procedures will no longer be required to perform common edits, such as inserting and removing objects from a scene. The ability to insert illuminating objects into a scene (e.g. a lit candle) will also be possible given these techniques, a task that is nearly impossible to achieve with modern editing tools.

Broader Impact. The field of digital photography is vast and becoming increasingly popular, but novice photo editors can only perform uncomplicated edits, such as cropping. This work will provide seamless 3D editing mechanisms for all levels of users. During the course of this research, I will develop new techniques of inferring spatial layout from single images which will have applications in animation, household robotics, surveillance and security, and vehicle safety. To facilitate broader dissemination I will publicly release all data, code, and executables on a public website. The interface I have created has already proven to be an effective educational tool for perspective geometry, and I will release modified versions of the editor for teaching purposes.

References.

[1] Efros, A.; Leung, T. "Texture synthesis by non-parametric sampling." In *ICCV* 1999.

[2] Hartley, R. I., and Zisserman, A. *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press, 2004.

[3] Hoiem, D., Efros, A. A., and Hebert, M. "Recovering surface layout from an image." *IJCV*, 75(1):151-172, 2007.

[4] Lalonde, J.-F., Hoiem, D., Efros, A. A., Rother, C., Winn, J., and Criminisi, A. "Photo clip art." *ACM SIGGRAPH* 2007.