The Non-Photorealistic Camera: Automatic Stylization With Multi-Flash Imaging

Ramesh Raskar * MERL | Kar-Han Tan† MERL | Rogerio S. Feris‡ UCSB | Jingyi Yu§ MIT | Matthew Turk¶ UCSB

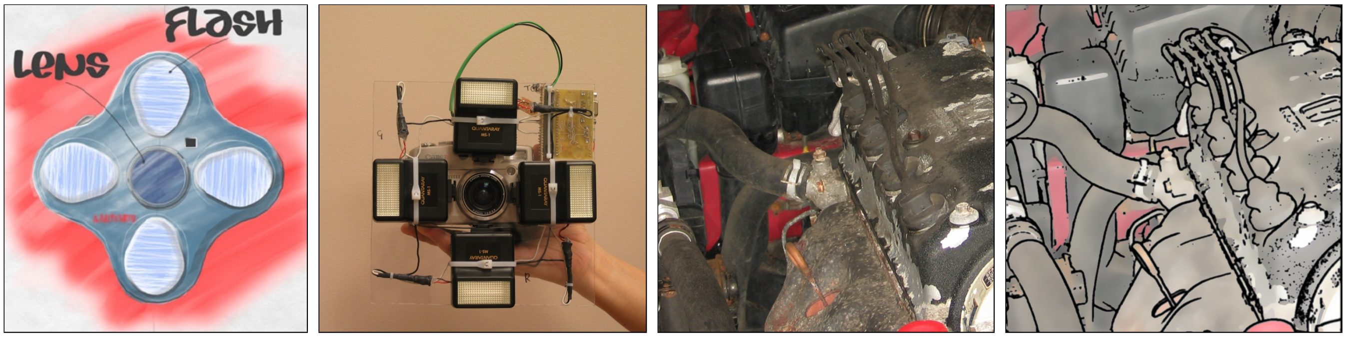

Figure 1: *(a) Conceptual Sketch and (b) Prototype of Multi-Flash Camera. (c) Photo of a car engine and (d) Stylized rendering highlighting shape. Notice the four spark plugs and the dip-stick which are now clearly visible*

1 Introduction

While photographs are the *de facto* visual medium for depicting reality, for some scenarios it is hard to produce pictures that convey clearly the 3D structure of a scene to the human eye. Consider imaging a white piece of paper with a white background. A traditional camera will record a mostly white image, and the shape of the paper will be lost or difficult to perceive. Our goal is to create enhanced images and video that make it easy for the viewer to understand the relative depth of the objects in the scenes depicted. This non-photorealistic camera is inspired by techniques used by skilled artists and digital illustrators to make images comprehensible: accentuating important features and reducing visual clutter.

A multi-flash camera uses strategically positioned flashes to cast shadows along silhouettes in the scene, which can then be reliably detected. The detected silhouettes can then be rendered in cartoon style or as technical illustrations. This overcomes the need for per-frame photo editing or 3D scanning of environments and allows automatic stylization of real world images.

2 Vision

Non-photorealistic rendering (NPR) techniques aim to outline the shapes of objects, highlight the moving parts to illustrate action, and reduce visual clutter such as shadows and texture details. Traditionally, detecting silhouettes and depth edges in a real scene is a very challenging task. Intensity edges are different from depth edges, so a simple intensity-edge detection in photo editors will not create the same quality of shape boundaries that we generate. We show applications in generating stylized images, but the captured depth edges can be used in many other applications. Effects such as depth-of-field effect during post-processing and synthetic aperture using camera array and screen matting for virtual sets (with arbitrary background) require high-quality signed (local foreground vs. background) depth edges.

Many artists have shown a wonderful array of work in stylizing images of videos using photo editors or rotoscoping ("Waking Life", "Avenue Amy"). Our goal is to supply digital artists with powerful scene-feature detection tools, so that tedious manual marking can be reduced or eliminated, allowing artists to focus on the creative aspects of rotoscoping-based animation.

While we believe that our camera will be a useful tool for professional artists and photographers, we also expect it to enable casual users to easily create stylized imagery. We also aim to demonstrate and create awareness of the use of silhouettes beyond stylized imaging in other applications.

3 Innovations

When a rich 3D model of the scene is available, rendering subsets of view-dependent contours is a relatively well-understood task. Extending this approach to real scenes by first creating 3D scene models, however, remains difficult.

This project shows that it is possible to bypass conventional 3D geometry acquisition and directly create stylized renderings from images and video. In the place of expensive, elaborate equipment for multi-camera stereo or laser-based methods, we propose using a camera with a simple extension: multiple strategically positioned flashes. Rather than estimating the full 3D coordinates of points in the scene, and then looking for depth discontinuities, our technique reduces the general 3D problem of depth-edge recovery to one of 2D image processing. This also makes it realistic and feasible to implement our method using camera components available today.

Our approach is based on taking successive photos of a scene, each with a different light source close to and around the camera's center of projection. We use the location of the shadows abutting depth discontinuities as a robust cue to create a depth-edge map in both static and dynamic scenes. We produce enhanced images and videos that more clearly convey the 3D structure of the imaged scene. The depth-edge map can also be used to produce other types of non-photorealistic or artistic renderings.

*raskar@merl.com
†tan@merl.com
‡rferis@cs.ucsb.edu
§jingyi@graphics.csail.mit.edu
¶mturk@cs.ucsb.edu