EE368 PROJECT PROPOSAL: DEPTH EXTRACTION FROM CAMERAPHONE IMAGES

JAKE MATLICK AND VINITH MISRA

The extraction of depth information from 2D images or 2D video can serve many purposes, ranging from adjustable focus and depth of field effects to postconversion of 2D image/video data into 3D. We intend to construct a simple proof-of-concept implementation of depth extraction on images and video collected from a DROID phone.

Specifically, we aim to do the following:

(1) A moving smartphone captures a video of a static scene (no objects in the scene are moving — just the camera).
(2) Features are identified and tracked across these images.
(3) From the relative movement of the features, their relative distances from the camera are computed and displayed as a depth map superimposed on the video.

The project may be partitioned into three phases: one for each of the above.

**Phase 1:** Identifying features in static images. There are effectively two options for us here. The first is to constrain the scene to consist of only a few objects at clearly separate depths and seek to separate each of these objects within the image. Alternatively, we may seek to identify distinctive features in the image [Bay et al.], such as corners [Harris and Stephens], edges, and unusually dark or bright points. This latter approach is frequently used in the related problem of match-moving, wherein one seeks to extract a camera’s path of movement from the video it has captured of a static scene. The benefit of this technique is that it makes no restrictive assumptions about the scene (besides it being static) and is likely to be more robust.

**Phase 2:** Tracking features between video frames. In order to identify the depth of each feature, we must be able to estimate a path of movement for it between frames. Assuming the camera is moving relatively slowly, each feature should remain relatively similar between frames and simple template-matching [Brunelli, 2009] techniques could work adequately well.

**Phase 3:** Converting the paths taken by the features into depth values. While it is certainly possible to perform this operation when arbitrary camera movement is permitted [Triggs, 1996], we intend to start by focusing on the case where the camera only moves in the depth-direction (towards
and away from the center of the scene). The closer objects are, the more they will scale in size.

References: