Mapping objects in 3D videos using a single camera

Luca Manzari – manzari@kth.se

A mandatory introduction

In the era of camera-enabled smartphones and self-driving cars, computer vision is a hot topic. Advancements in the field allowed for the use of cameras in non-contact measurement systems in place of strain gauges and accelerometers. In order to estimate depth – just like humans need two eyes! – two cameras are needed.

Stereopsis, the science of 3-dimensional perception, deals with mapping points between the real world and the cameras image planes. Other techniques take then care of the object recognition in between different images, in order to know what went where.

In many applications (e.g. microscopy, macro photography) only one camera is available, and no additional means of estimating depth can be used.

The main idea to explore

In the field of visual arts, photographers use many shallow-depth-of-field shots and stack them together to get an extended-focus image.

Figure 1: Focus bracket of a Tachinid fly. The rightmost image results from the composition of the two other images. Source: Wikipedia

What if the depth information could then be extracted by the shallow-depth-of-field shots? If the outcome of a phenomenon is deterministic, such phenomenon could be video recorded many times, each time focusing at a different plane. The in-focus parts of each image would then be detected, and the geometry of the object could be reconstructed in 3D for the whole observation period.
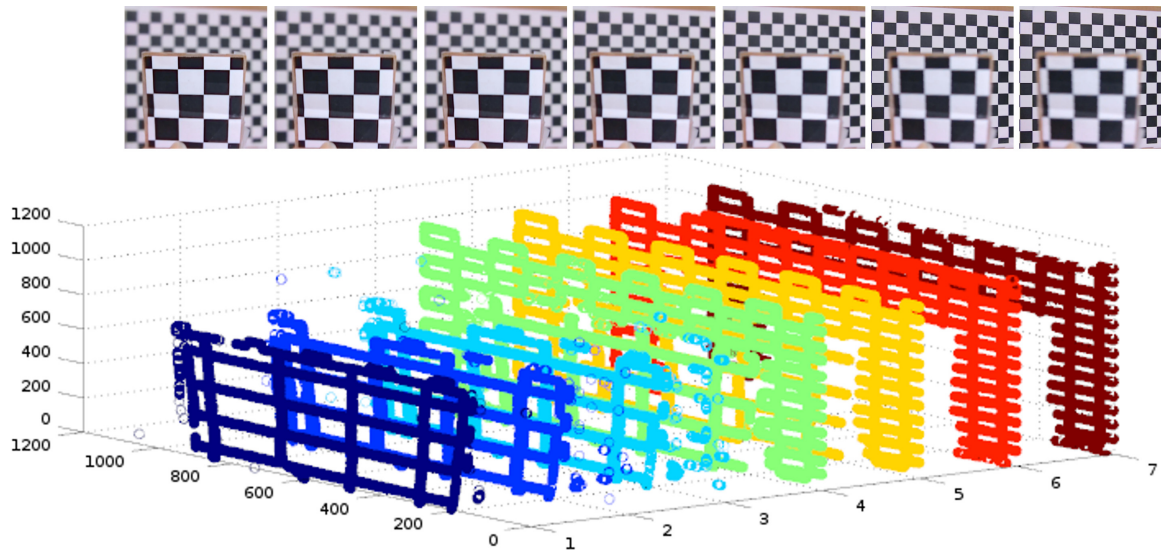

Figure 2: Same scene focused at different planes, and extraction of the points that are in focus. Own work.