Super-Resolved Texture Tracking

Frank Dellaert

Problem: Two important tasks in many computer vision applications are motion estimation and tracking of objects in video-streams. Scenarios where this is particularly difficult are those where the motion is fast, noise levels are high, and the computation needs to happen in real time. An example of such a domain is *mobile robotics*. In particular, three mobile robot scenarios under investigation at CMU each display typical challenges. Indoor robots are not that fast, but operate in changing and noisy environments. Autonomous vehicles operate at high speeds, and although more predictable than people in a building, perceiving and avoiding other cars presents significant perceptual challenges. Finally, an autonomous helicopter has perhaps a more predictable environment, but it must operate under high speed and cope with high noise levels.

Impact: Deducing scene motion or ego-motion from an image sequence has applications ranging from image stabilization in camcorders to enabling an autonomous landing approach in aircraft. Tracking the motion of objects in a scene finds applications in environments as diverse as the factory floor and operating rooms. Any approaches that advance the level of accuracy and robustness previously attainable while at the same time maintaining reasonable computational demands will have a large impact in a large number of application domains. It is my hope that the approach I developed, *Super-Resolved Texture Tracking* (see below), will become a standard tool in the arsenal of applied computer vision.

State of the Art: To cope with fast motion and high noise levels, previous approaches used recursive estimation techniques to optimally integrate all available measurements *over time*, typically using a *Kalman filtering* approach. Unfortunately, not all information available in the video-stream is used, as, to the best of our knowledge, all current approaches extract sparse features from the images to use as the measurements. The reasons are twofold: (a) the cost of using complete or partial images as measurements is assumed to be too great to achieve real time performance, and (b) it is not immediately clear how to integrate image based measurements or how to predict them from the state estimate, as can easily be done for discrete features.

Image-based approaches to motion estimation, on the other hand, use all the information available *in the image*, but do not employ recursive estimation techniques to integrate those measurements over time. Presumably, it is deemed infeasible to formulate a state space representation that can accurately predict the images, nor is it clear how such a state would be updated and maintained over time. However, unlike feature-based approaches, image-based techniques *do* use all of the available information in one image.

Approach: The method I propose, *Super-Resolved Texture Tracking* [1, 2], is an attempt at using *all* information available in the video-stream, *both in space and in time*, yielding unprecedented accuracy and robustness. As with the current state of the art in feature-based motion estimation, a Kalman filter is used to formalize the problem as a recursive state estimation problem. However, to be able to use the whole image as our measurement vector, we incorporate a *texture map* into the system state, modeling the texture present on the surfaces that we are tracking (see Figure 1). As the measurement model, we use *texture mapping*, a technique from computer graphics that is normally used to render realistically looking surfaces.

The novel combination of a Kalman filter with texture mapping yields some unique advantages. In particular, the estimated texture map can be kept at an arbitrary resolution. Thus, if we keep it at a *higher* resolution than the source images themselves, our method can produce *super-resolved* texture estimates as more image measurements are taken. However, the texture map can also be kept at a *lower* resolution while still maintaining accurate tracking. In addition, since we can predict entire images, deviations from the

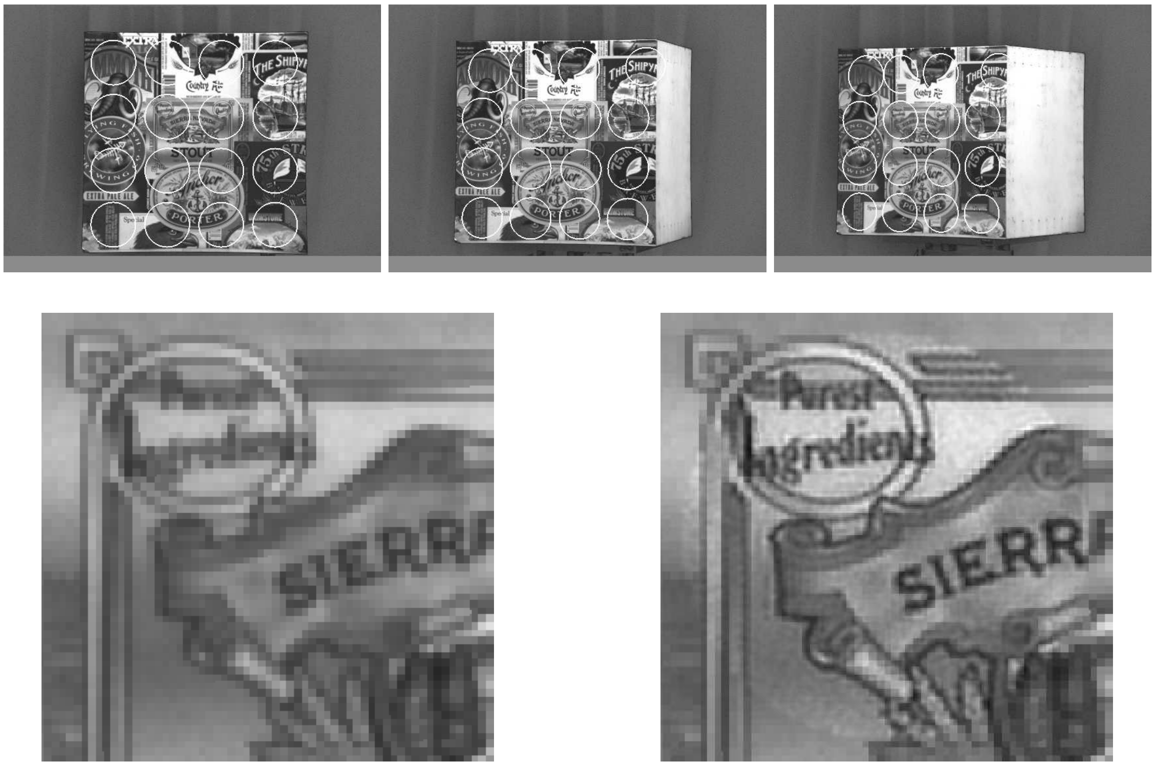

Figure 1: **Top**: The texture based trackers I developed perform motion estimation in 3D by 'sticking' to the textured surface of an object. In the figure, you can see 16 *stickers* tracking the textured face of a cube in parallel. The complete sequence is available for viewing on the web at URL `http://www.cs.cmu.edu/~dellaert/research/patches.html`. **Bottom**: By tracking a surface over time, one can *super-resolve* the texture present on a given surface. In the figure, you can see the original image resolution at the left, and the texture estimate of one 'sticker' after 20 frames into the sequence. As you can see, the previously unreadable words 'Purest Ingredients' are now readable inside the super-resolved circle.

prediction enables us to see what objects are incompatible with the expectations formed using our internal model. As an example, this could allow us to detect independently moving objects such as cars or people in a known environment.

Future Work: There are no important difficulties in extending this approach to non-planar surface models. Future work will investigate *arbitrary* surface representations, and how their parameters could be estimated from the image sequence along with the texture. In addition I would like to investigate the simultaneous recovery of camera parameters in uncalibrated scenarios. Finally, I am planning to apply approach towards several hitherto unsolved problem domains in mobile robotics.

Selected References:

[1] Frank Dellaert, Chuck Thorpe, and Sebastian Thrun. Super-resolved tracking of planar surface patches. In *submitted to IROS 98*, 1998.

[2] Frank Dellaert, Sebastian Thrun, and Chuck Thorpe. Jacobian images of super-resolved texture maps for model-based motion estimation and tracking. In *submitted to WACV 98*, 1998.