# Session Overview
# Robotic Vision

Yoshiaki Shirai[1] and Bob Bolles[2]

[1] Osaka University
[2] SRI International

At the DARPA Grand Challenge in October 2005, laser range finders, especially the ones manufactured by SICK, were the predominant range sensors. Does that mean that stereo sensors are dead? No. It means that laser scanners satisfied the requirements of the Grand Challenge outdoor vehicle navigation application better than stereo. Stereo sensors, on the other hand, are the sensor of choice for several other applications, such as people monitoring and human-computer interfaces, because they are passive, relatively inexpensive, have no moving parts, and provide registered range and color data.

In this session, the authors present camera-based algorithms for computing 3-D descriptions of scenes. Two of the papers focus on stereo techniques and one on monocular constraints.

Sibley, Matthies, and Sukhatme describe two types of biases associated with stereo analysis, and then describe approaches that dramatically reduce their effects. In the case of triangulation-based range estimation, their new approach reduces the bias by an order of magnitude. To accomplish this reduction, they re-express stereo triangulation as a second order series expansion, taking into account the distribution of errors associated with image-based stereo matching and the calculation of range. The second type of bias occurs when several stereo measurements are combined, often over time, to improve the precision of the measured range values. To reduce this bias, they developed an iterative non-linear Gauss-Newton technique that focuses on image space measurements instead of directly averaging/filtering 3-D range values.

Blake et al describe a technique for combining stereo analysis, color analysis, and occlusion reasoning to segment an image into foreground, background, and occluded regions. Their approach uses a probabilistic model formulated as a Conditional Random Field that fuses prior information about the expected structures in a scene with stereo and color results. They apply their technique to a video conferencing application and show its effectiveness at precisely extracting the foreground people, which is the key to several enhancements, including automatic camera control, eye-gaze correction, and the insertion of virtual objects into the scene.

Delage, Lee, and Ng approach the problem of extracting a 3-D description of a scene quite differently than the other two groups in this session. They use generic knowledge about buildings and cameras to derive 3-D models of indoor scenes from monocular images. In particular, their technique uses such facts as floors are horizontal planes and walls are vertical planes to extract and interpret linear edge patterns as floors and walls. They use a Markov Random Field to label every pixel in an image as part of a surface or edge, and then apply an iterative 3-D reconstruction algorithm. Their approach, which includes an analysis of the vanishing points, locates a floor footprint first, and then fills in the walls.