The Southampton York Natural Scenes (SYNS) dataset

Wendy J. Adams
Centre for Vision and Cognition
University of Southampton, UK
W.Adams@soton.ac.uk

Alexander A. Muryy
Centre for Vision and Cognition
University of Southampton, UK
a.muryy@soton.ac.uk

Erich W. Graf
Centre for Vision and Cognition
University of Southampton, UK
E.W.Graf@soton.ac.uk

Arthur J. Lugtigheid
Centre for Vision and Cognition
University of Southampton, UK
lugtigheid@gmail.com

James H. Elder
Centre for Vision Research
York University, Toronto, Canada
jelder@yorku.ca

CCS Concepts

•**Computing methodologies** → *Scene understanding; 3D imaging;* •**Hardware** → *Scanners;* •**General and reference** → **Measurement;**

Humans are adept at estimating 3D scene geometry from a stereo image pair, or even from a single image. Computer vision algorithms are less good. Gaining traction on this problem requires a dataset that contains good quality images and ground truth data, and represents the complex and diverse scenes that we encounter. To this end we have developed the Southampton York Natural Scenes (SYNS) public dataset: https://syns.soton.ac.uk.

We randomly sampled scene locations within Hampshire from 25 diverse scene categories, both indoor (e.g. offices, cafes/bars, theatres) and outdoor (e.g. farms, wetlands, industrial, and residential areas), identified via the UK Landuse Dataset. Each scene is represented by 3 types of data (Figure 1). Device nodal points were coregistered and aligned to within 3mm at the average human eye level of 165cm. Residual rotation and translation differences between range and image data were estimated and corrected for in post-processing.

We estimated the surface normal at each LiDAR point using a 3D principle component analysis of a k-point neighbourhood (KNN) around the point, where k was optimized over the range k = [7,...,100] (Figure 2b). Degenerative elongated neighbourhoods with a ratio of eigenvalues λ_2 / λ_1 < 0.3 were rejected. An iterative method removed outlier points > 2.5 standard deviations from the estimated plane. We quantified the distribution of surface attitude across slant and tilt in ego-centric coordinates, i.e. relative to the view vector at each patch.

Overall, the joint distribution over slant and tilt is dominated by the ground plane, in both built and natural environments (Figure 2). We found a small peak at fronto-parallel, as predicted by the geometry of projection. Near the horizon, other regularities are also apparent, e.g. elevated probability density at ±90°tilt due to vertical surfaces (trees, buildings).

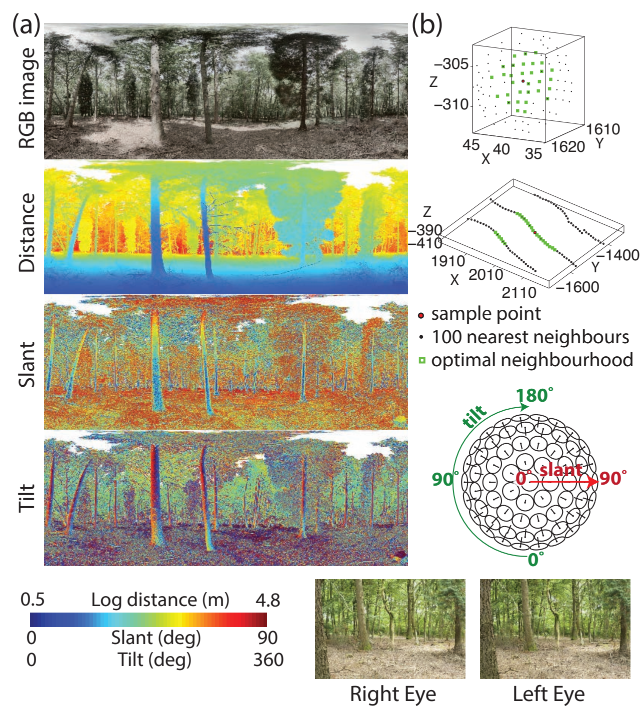

Figure 1: (a) Spherical, high-dynamic range images were captured with a Spherocam (26 F-stops, angular resolution 0.067°). A laser rangefinder (LiDAR) captured ground truth in the form of 3D point clouds with angular resolution of 0.036°, depth resolution 0.4-9mm. A custom stereo rig, with camera separation matched to the average human IPD, captured 18 stereo image pairs, tiling a 360°panorama (each image: 35°x 24°, resolution 0.006°). (b) Example neighbourhoods. Upper: Patch at small range. Lower: Patch on ground plane near horizon.

Joint analysis of 2D image and 3D scene data showed that

CVMP 2015 November 24-25, 2015, London, United Kingdom

DOI: http://dx.doi.org/10.1145/2824840.2824857

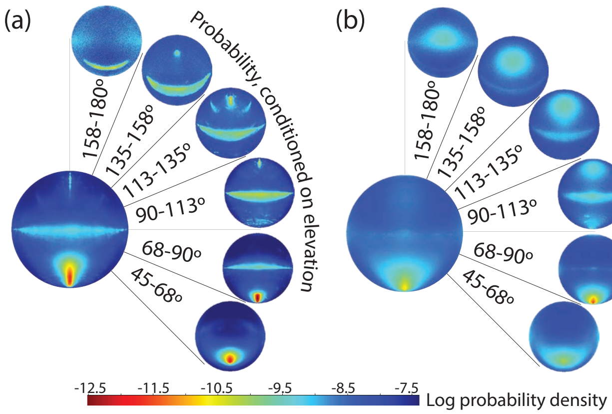

Figure 2: Surface attitude, shown in polar coordinates (see Figure 1b), averaged across (a) built environments (N=25) and (b) natural environments (N=19), for different elevations, and across all elevations (larger plots)

luminance is predicted by both distance and surface orientation: In both natural and built environments, distant surfaces tend to be darker. Slant, relative to horizontal, was a stronger predictor of luminance: horizontal surfaces tend to be brighter. In combination, distance and slant explain 22% of luminance variance.

We foresee a number of uses for the SYNS dataset: Regularities in image and scene structure may predict human perceptual biases. Computer vision algorithms for 3D reconstruction can be trained on and tested against complex, ecologically valid scenes.