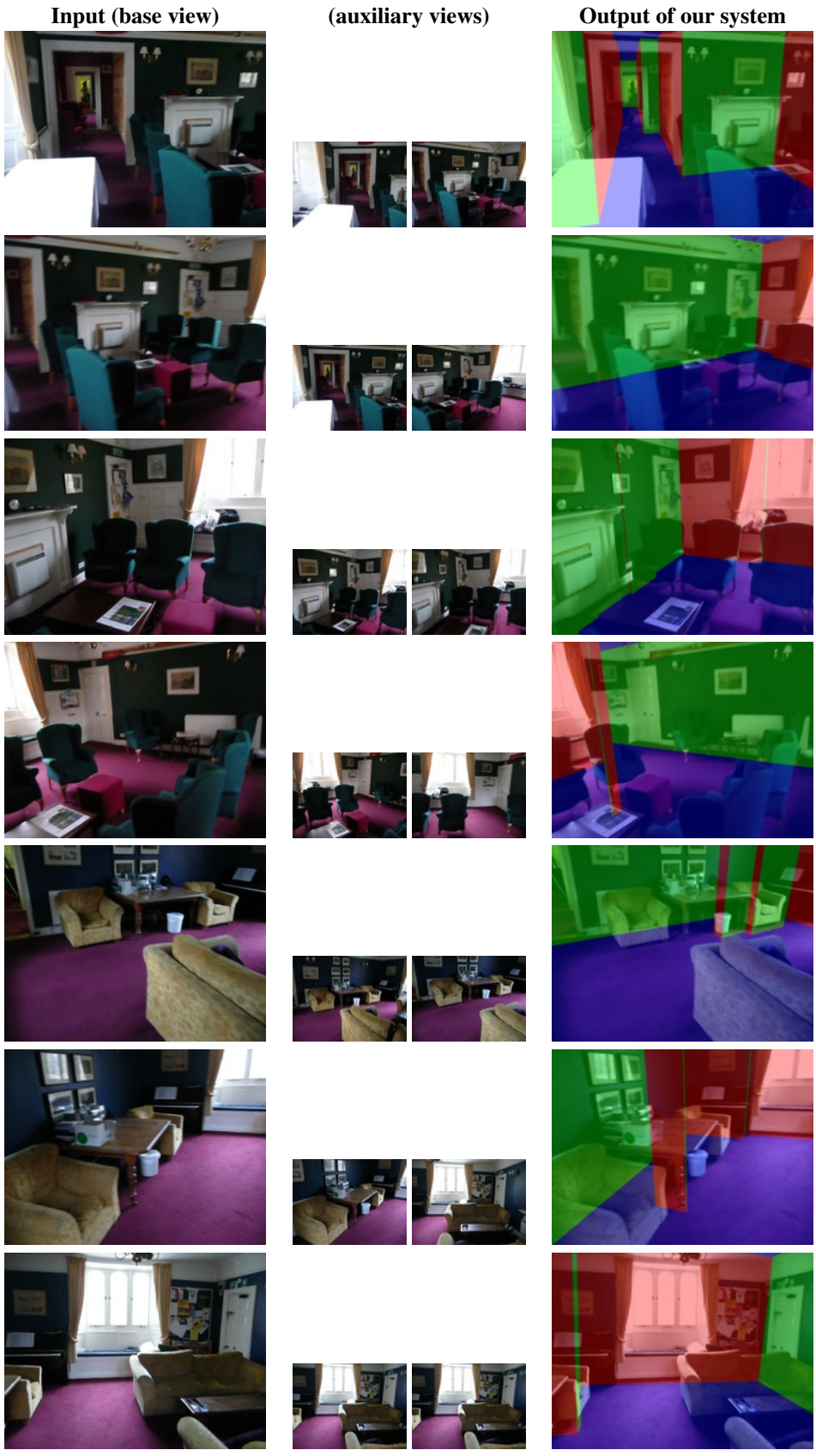
Addendum to Manhattan Scene Understanding Using Monocular, Stereo, and 3D Features

Alex Flint, David Murray, and Ian Reid
Active Vision Laboratory
Oxford University, UK
{alexf,dwm,ian}@robots.ox.ac.uk

A. Additional Results

In this section we showcase further examples of models generated by our system. In the table below, the left panel shows the base input view I_0, the middle panel shows auxiliary views used for photoconsistency calculations, and the right panel shows the MAP model M inferred by our system. In all our experiments we used two image auxiliary images per base image, which were sampled one second before and one second after the base image in the video sequence. For further details of parameter settings see the main paper.


Input (base view)
(auxiliary views)
Output of our system


Input (base view)
(auxiliary views)
Output of our system

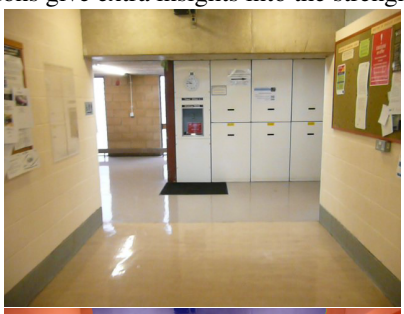

Input (base view)
(auxiliary views)
Output of our system


Input (base view)
(auxiliary views)
Output of our system


Input (base view)
(auxiliary views)
Output of our system

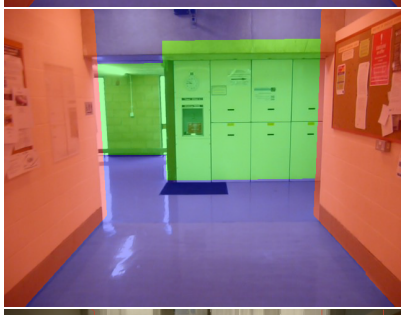
B. Payoffs

In this section we provide some visualizations of the payoff matrices discussed in the main paper. We hope these visualizations give extra insights into the strenghts and weaknesses of each sensor modality.

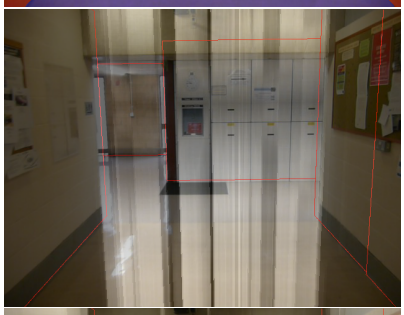


The raw image provided as input to our system.

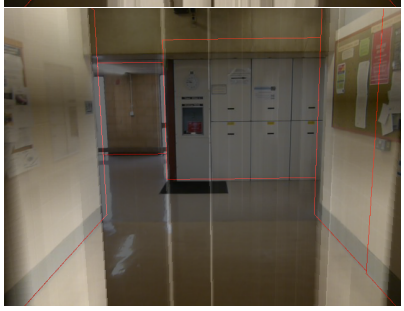
The ground truth segmentation for this image. Horizontal surfaces are shaded blue. Vertical surfaces are shaded red and green. We will refer to these as the "red" and "green" orientations.



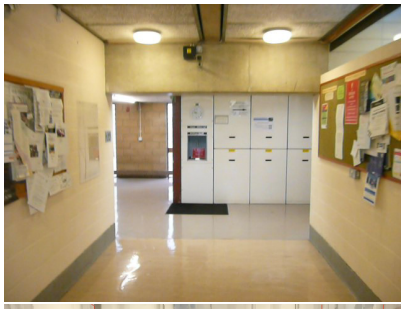
The MAP indoor manhattan model M output by our system for this input.

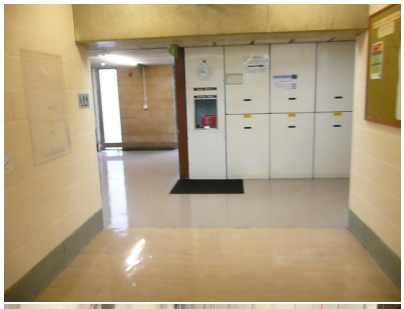


Payoffs π_{mono} derived from monocular image features, for the "green" orientation. Pixels of higher intensity correspond to larger values in the payoff matrix. The MAP model is shown in wireframe using red lines. Intuitively, the optimization over models can be thought of as finding the minimal cost path through the payoff matrix, where higher intensity pixels correspond to lower costs. This is only a rough picture, however; the real optimization situation is more complex since models are penalized for each additional corner.



As above, for the "red" surface orientation.





Auxiliary images used for stereo photoconsistency. In our experiments we used two image auxiliary images for each base image, which were sampled one second before and one second after the base image in the video sequence.

Payoffs π_{stereo} corresponding to the auxiliary images above. Each pixel represents the photoconsistency score for a wall segment with floor/wall (or ceiling/wall) intersection y_x at that pixel. Notice the repeated "pizza slice" patterns in which one tip of the triangle is located at the floor/wall intersection.

Here we show the structure–from–motion point cloud. The points are shown projected into the image, but the system has access to their 3D locations. Notice how the points are not uniformly distributed in the image.

Here we show the structure–from–motion point cloud projected onto the floor and ceiling planes, which were recovered as a seperate step as described in the main paper. The red dots show the original 3D point cloud and the blue dots show the projections onto the floor and ceiling.

This shows the component of the payoffs π_{3D} intended to provide a bias towards models that explain the observed 3D points. This is the component corresponding to $t = \mathsf{ON}$. Each bright spot corresponds to the projection of a 3D point onto the floor or ceiling plane.

This shows the component of the payoffs π_{3D} intended to penalize walls that occlude observed 3D points. This corresponds to the case that $t \in \{\mathsf{IN}, \mathsf{OUT}\}$. Notice that for each 3D point, payoffs are assigned to walls that pass between the floor and ceiling projection of that point. Such walls are precisely those which do *not* occlude the point.

Joint payoff matrix π_{joint}.