Neural RGB→D Sensing : Depth and Uncertainty from a video camera

Summary

1. The goal is depth sensing from a video stream.
2. Pipeline consists of 3 parts :
   - D-Net : estimates Depth Probability Volume (DPV) for each input frame.
   - K-Net : helps to integrate DPVs over time.
   - R-Net : improves spatial resolution of DPVs.

DPV

1. DPV is denoted as $p(d; u, v)$: the prob. pixel (u,v) has depth d.
2. $d \in [d_{min}, d_{max}]$ discretized into $N = 64$ planes.
3. Depth Estimate $\hat{d}(u,v) = \sum_{d_{min}}^{d_{max}} p(d; (u,v)) \cdot d$
4. Confidence $\hat{C}(u,v) = p(\hat{d}; (u,v))$ — omit for rest of paper