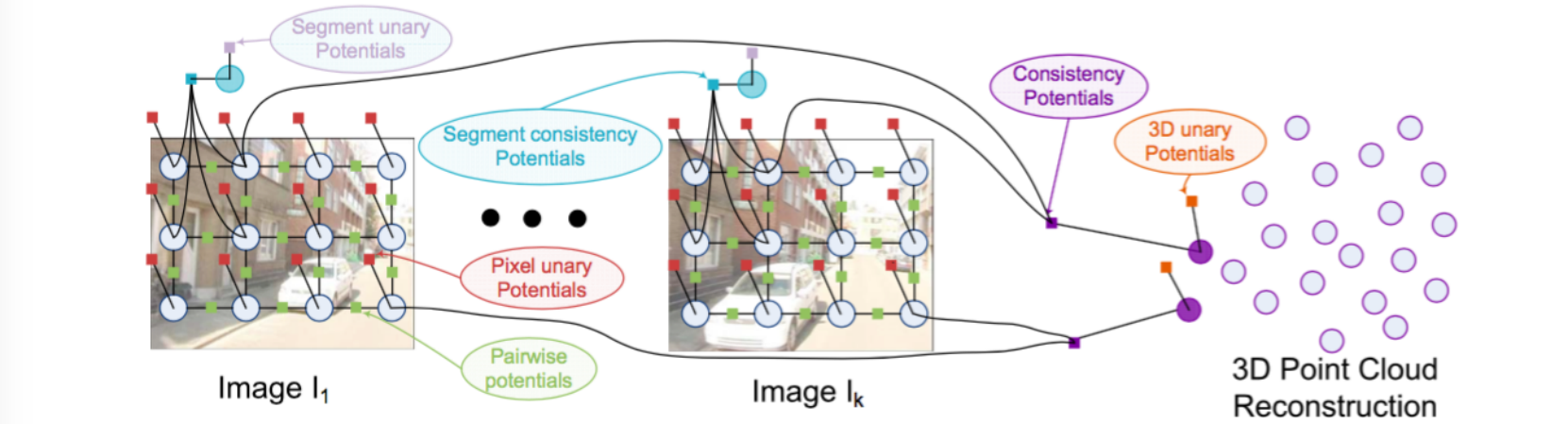
Joint 2D-3D Temporally Consistent Semantic Segmentation of Street Scenes

G. Floros, B. Leibe (CVPR 2012)



- Proposes a novel Conditional Random Field (CRF) formulation for the semantic scene labeling problem which is able to enforce temporal consistency between consecutive video frames and take advantage of the 3D scene geometry to improve segmentation quality
- Uses 3D scene reconstruction as a means to temporally couple the individual image segmentations, allowing information flow from 3D geometry to the 2D image space
- Details:
  - Optimizes the semantic labels in a temporal window around the image we are interested in
  - Augments the higher-order cliques of the CRF with the sets of pixels that are projections of the same 3D point in different images
  - Since these new higher-order cliques contain different projections of the same 3D point, the labels of the pixels inside the clique should be consistent
  - Forms a grouping constraint on these pixels
- Evaluates on Leuven and City stereo dataset