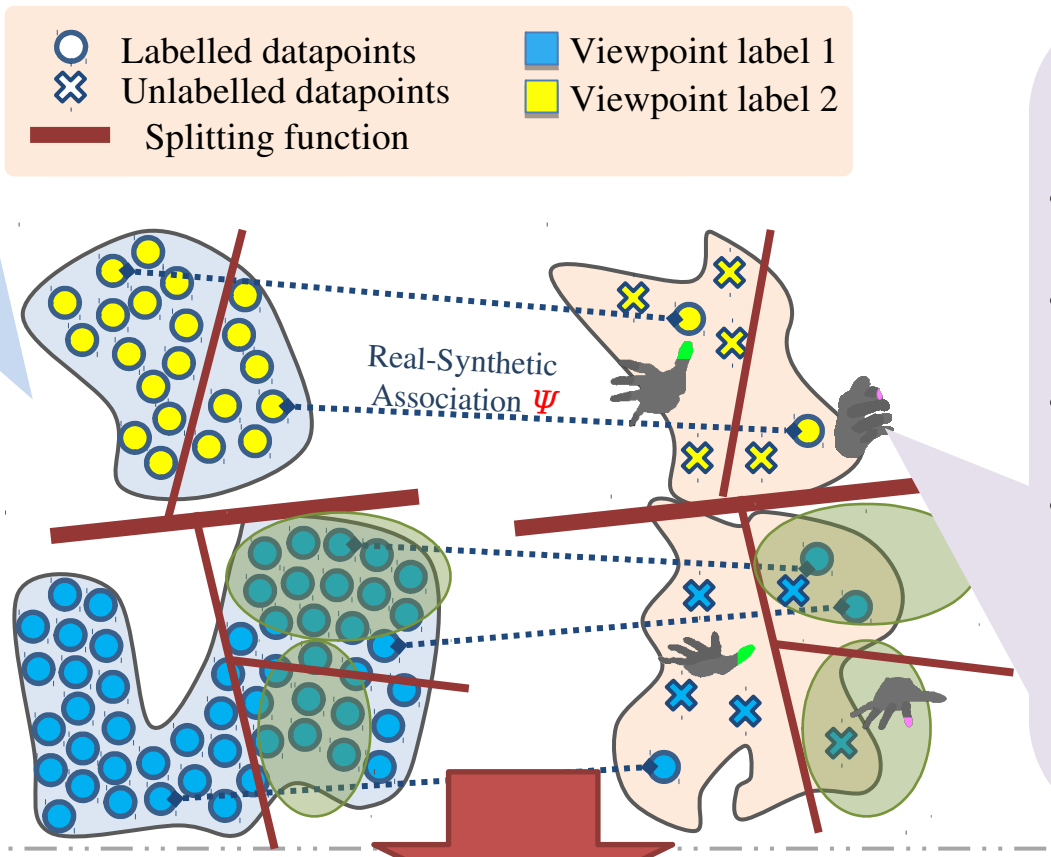
1. Training Dataset *D* (Section 6.4.1)

Synthetic data *S*
(Source space)

- Synthetic depth images are generated by an articulated hand model.
- All synthetic data are clean and automatically labelled.
- Synthetic data are not affected by noise and occlusion. A synthetic instance looks very different from its real counterpart. It is infeasible to train a pose estimator using synthetic data only.
- An efficient and cheap method to generate training data.



Real data *R*
(Source space)

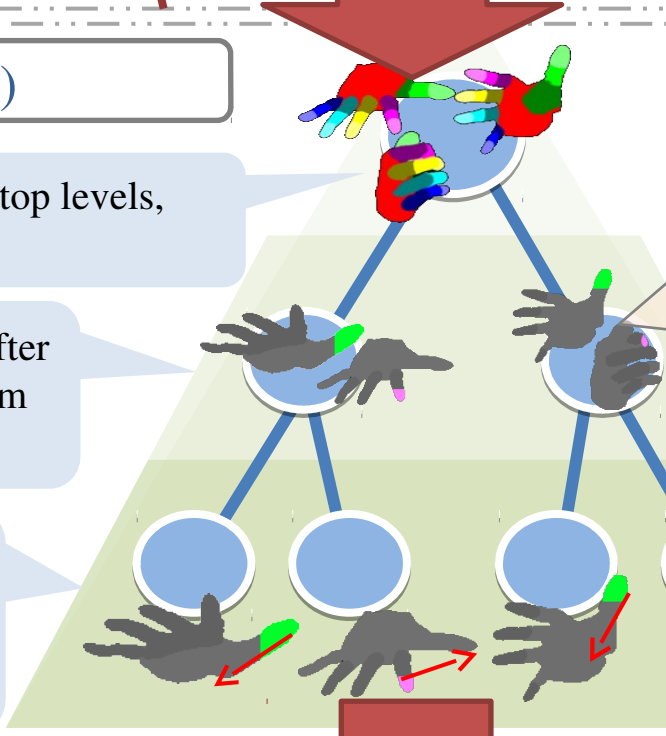
- Real depth images are captured from a Kinect sensor.
- They are affected by sampling noise and self-occlusions.
- Labelling is expensive, so the dataset is sparsely labelled.
- Some real instances are associated with their corresponding synthetic instance through ***Ψ***.

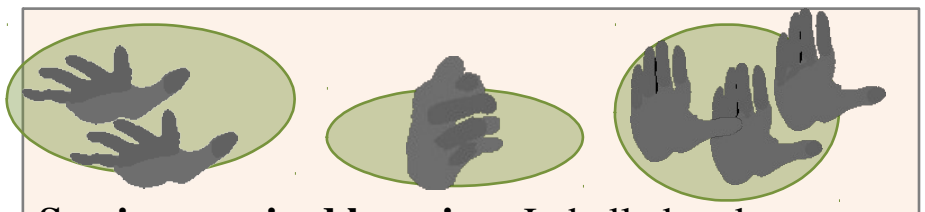
2. STR Forest (Section 6.4.2)

Viewpoint Classification is first performed at the top levels, controlled by the viewpoint term Q_a .

Joint Classification is performed at mid levels. After viewpoint classification, the joint classification term Q_p determines joint labels of each pixel.

Regression is performed at bottom levels. To describe the distribution of real hand pose data, nodes are optimised for data compactness via Q_v and Q_u towards the bottom levels.





Semi-supervised learning: Labelled and unlabelled data are clustered via Q_u, by comparing appearances of patches.

Transductive learning: The real-synthetic fusion are learned by the transductive term Q_t throughout the whole forest.

3. Data-driven joint refinement (Section 6.4.3)

The STF forest does not consider the physiological structure of human hands. As a result, a data-driven approach is presented to rectify incorrect joint locations. In addition, occluded joints are recovered by matching with a dataset of synthetic hand poses.