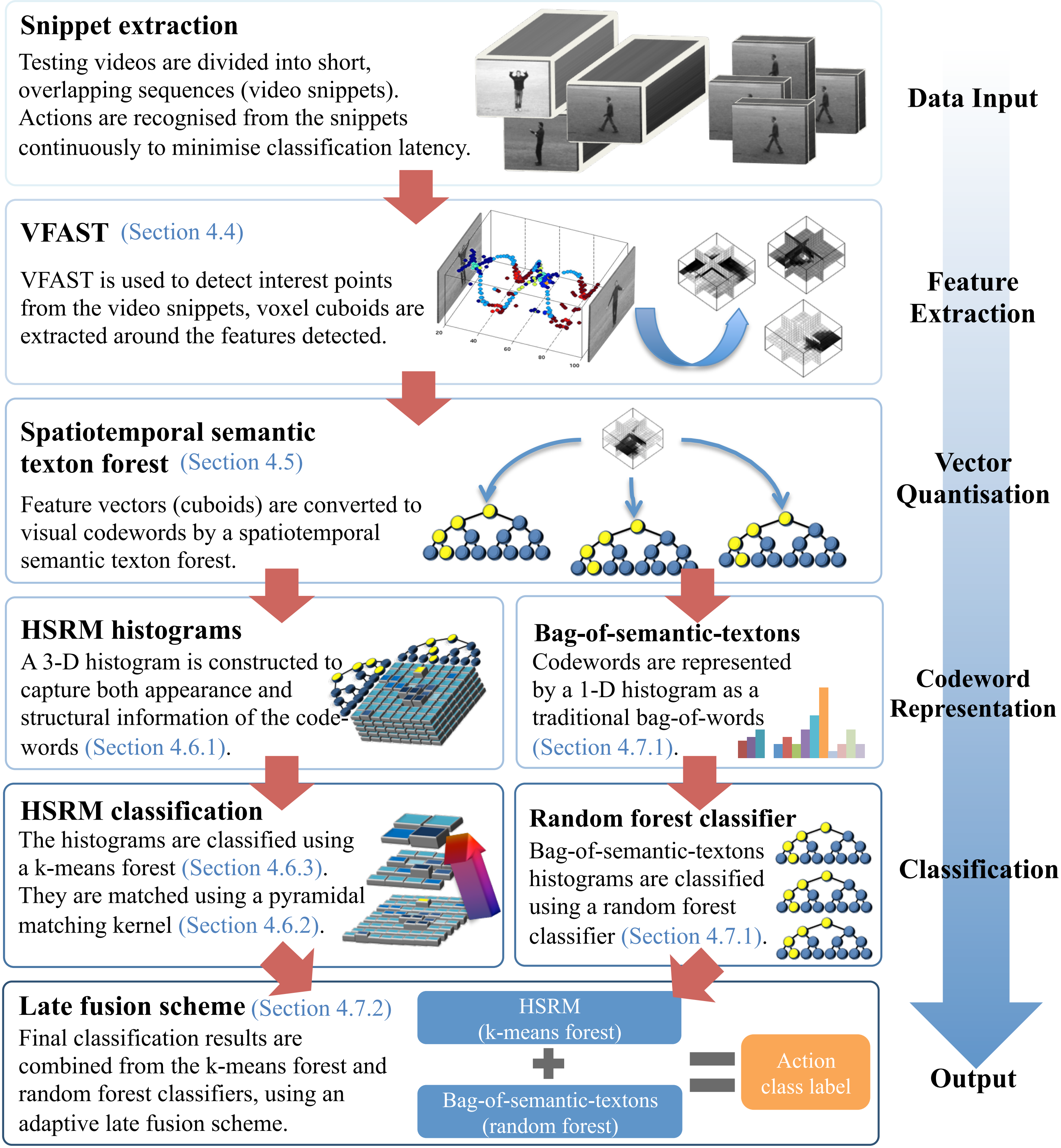

Snippet extraction

Testing videos are divided into short, overlapping sequences (video snippets). Actions are recognised from the snippets continuously to minimise classification latency.

VFAST (Section 4.4)

VFAST is used to detect interest points from the video snippets, voxel cuboids are extracted around the features detected.

Spatiotemporal semantic texton forest (Section 4.5)

Feature vectors (cuboids) are converted to visual codewords by a spatiotemporal semantic texton forest.

HSRM histograms

A 3-D histogram is constructed to capture both appearance and structural information of the codewords (Section 4.6.1).

Bag-of-semantic-textons

Codewords are represented by a 1-D histogram as a traditional bag-of-words (Section 4.7.1).

HSRM classification

The histograms are classified using a k-means forest (Section 4.6.3). They are matched using a pyramidal matching kernel (Section 4.6.2).

Random forest classifier

Bag-of-semantic-textons histograms are classified using a random forest classifier (Section 4.7.1).

Late fusion scheme (Section 4.7.2)

Final classification results are combined from the k-means forest and random forest classifiers, using an adaptive late fusion scheme.

HSRM (k-means forest) + Bag-of-semantic-textons (random forest) = Action class label