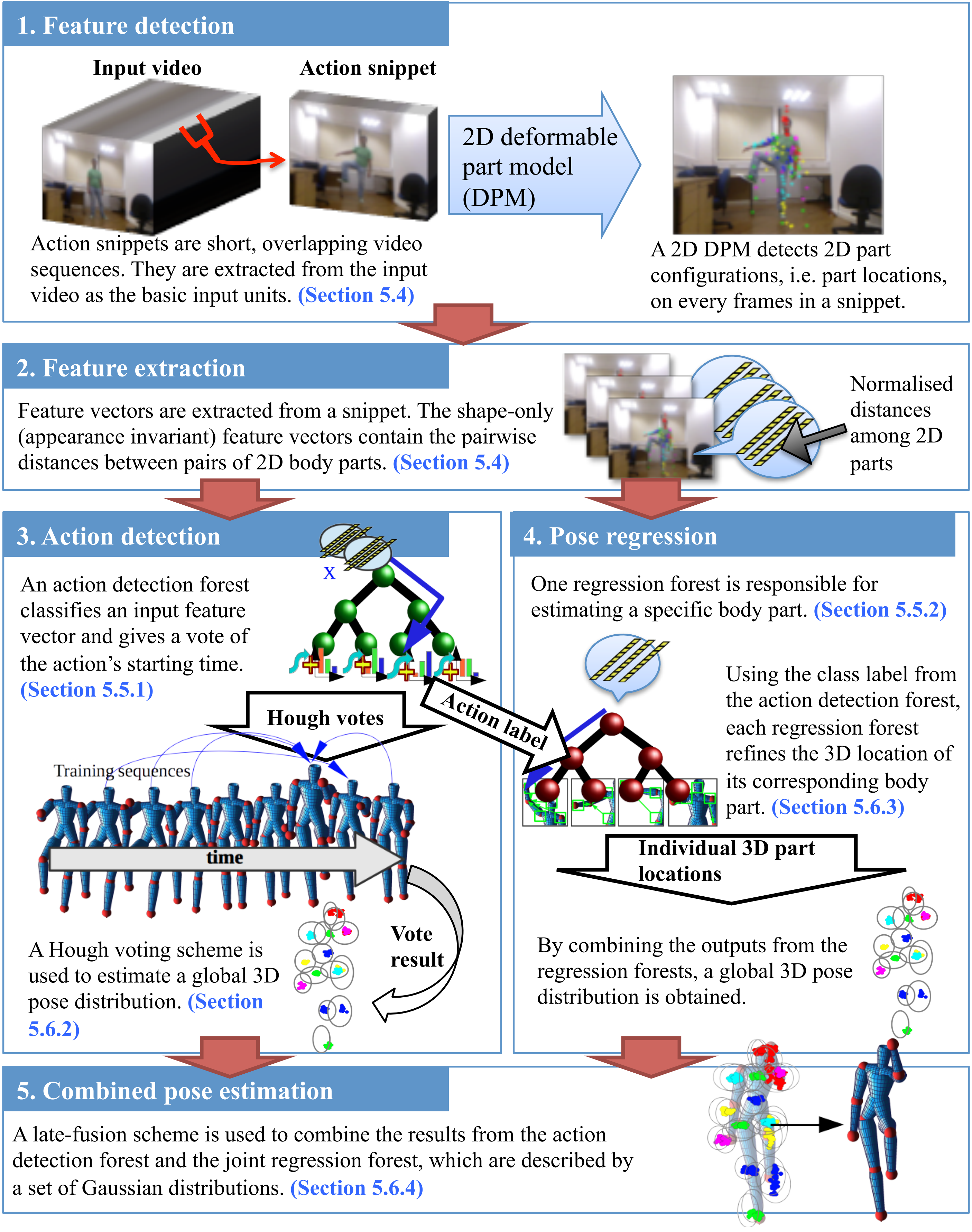

## 1. Feature detection

Action snippets are short, overlapping video sequences. They are extracted from the input video as the basic input units. (Section 5.4)

A 2D DPM detects 2D part configurations, i.e. part locations, on every frames in a snippet.

## 2. Feature extraction

Feature vectors are extracted from a snippet. The shape-only (appearance invariant) feature vectors contain the pairwise distances between pairs of 2D body parts. (Section 5.4)

## 3. Action detection

An action detection forest classifies an input feature vector and gives a vote of the action's starting time. (Section 5.5.1)

A Hough voting scheme is used to estimate a global 3D pose distribution. (Section 5.6.2)

## 4. Pose regression

One regression forest is responsible for estimating a specific body part. (Section 5.5.2)

Using the class label from the action detection forest, each regression forest refines the 3D location of its corresponding body part. (Section 5.6.3)

By combining the outputs from the regression forests, a global 3D pose distribution is obtained.

## 5. Combined pose estimation

A late-fusion scheme is used to combine the results from the action detection forest and the joint regression forest, which are described by a set of Gaussian distributions. (Section 5.6.4)