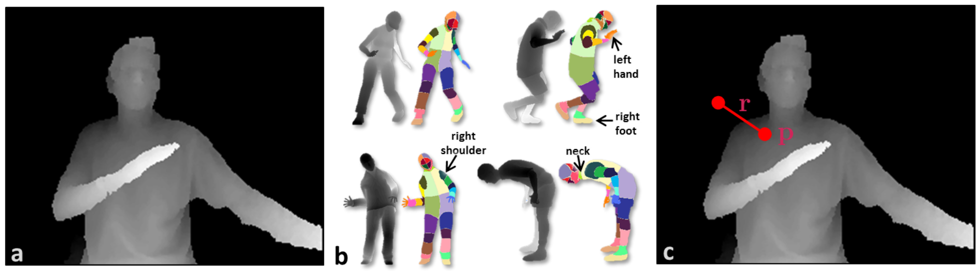

Fig. 3.15: **Classification forests in Microsoft Kinect for XBox 360. (a)** An input frame as acquired by the Kinect depth camera. **(b)** Synthetically generated ground-truth labeling of 31 different body parts [82]. **(c)** One of the many features of a "reference" point $\mathbf{p}$. Given $\mathbf{p}$ computing the feature amounts to looking up the depth at a "probe" position $\mathbf{p} + \mathbf{r}$ and comparing it with the depth of $\mathbf{p}$.

recently they have been adapted to work with multiple classes. Figure 3.14c shows how the sequentiality of the one-v-all SVM approach may lead to asymmetries which are not really justified by the training data.

3.6 Human body tracking in Microsoft Kinect for XBox 360

This section describes the application of classification forests for the real-time tracking of humans, as employed in the Microsoft Kinect gaming system [100]. Here we present a summary of the algorithm in [82] and show how the forest employed within is readily interpreted as an instantiation of our generic decision forest model.

Given a depth image such as the one shown in fig. 3.15a we wish to say which body part each pixel belongs to. This is a typical job for a classification forest. In this application there are thirtyone different body part classes: $c \in \{\texttt{left hand}, \texttt{right hand}, \texttt{head}, \texttt{l. shoulder}, \texttt{r. shoulder}, \cdots\}$. The unit of computation is a single pixel in position $\mathbf{p} \in \mathbb{R}^2$ and with associated feature vector $\mathbf{v}(\mathbf{p}) \in \mathbb{R}^d$.

During testing, given a pixel $\mathbf{p}$ in a previously unseen test image we