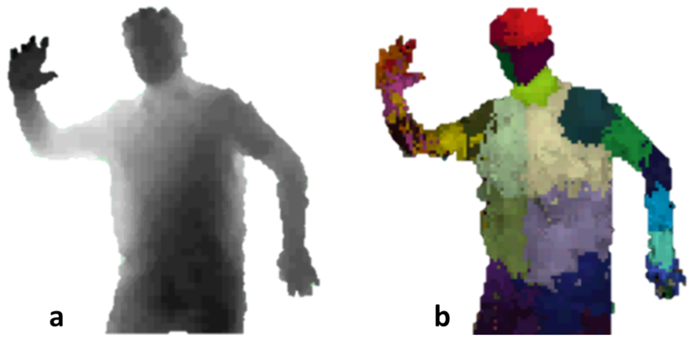

Fig. 3.16: **Classification forests in Kinect for XBox 360. (a)** An input depth frame with background removed. **(b)** The body part classification posterior. Different colours corresponding to different body parts, out of 31 different classes.

wish to estimate the posterior $p(c|\mathbf{v})$. Visual features are simple depth comparisons between pairs of pixel locations. So, for pixel $\mathbf{p}$ its feature vector $\mathbf{v} = (x_1, \ldots, x_i, \ldots, x_d) \in \mathbb{R}^d$ is a collection of depth differences:

$$x_i = J(\mathbf{p}) - J\left(\mathbf{p} + \frac{\mathbf{r}_i}{J(\mathbf{p})}\right) \tag{3.2}$$

where $J(.)$ denotes a pixel depth in mm (distance from camera plane). The 2D vector $\mathbf{r}_i$ denotes a displacement from the reference point $\mathbf{p}$ (see fig. 3.15c). Since for each pixel we can look around at an infinite number of possible displacements ($\forall\ \mathbf{r} \in \mathbb{R}^2$) we have $d = \infty$.

During training we are given a large number of pixel-wise labelled training image pairs as in fig 3.15b. Training happens by maximizing the information gain for discrete distributions (3.1). For a split node j its parameters are

$$\boldsymbol{\theta}_j = (\mathbf{r}_j, \tau_j)$$

with $\mathbf{r}_j$ a randomly chosen displacement. The quantity τ_j is a learned scalar threshold. If $d = \infty$ then also the whole set of possible split parameters has infinite cardinality, *i.e.* $|\mathcal{T}| = \infty$.

An axis-aligned weak learner model is used here with the node split