



**Fig. 3** A visualization of our spatial histogram feature  $d_{ij}$ . We consider the location of the center of window  $j$  with respect to a coordinate frame defined by window  $i$ , denoted by the thickly outlined box. The dashed and dotted rectangles represent regions over which the center of window  $j$  are binned. The relative location of  $j$  must either be **far** or **near**. For near windows, we consider **above**, **ontop**, **below**, and symmetric **next-to** bins as shown. To allow our model to reproduce the behavior of baseline modules that perform NMS with a criteria of 50% relative overlap, we also include a binary overlap feature. This makes  $d_{ij}$  a 7 dimensional sparse binary vector.