Figure 1: Additional example of object matching (ref to section 4.1 of the paper). The example shows the image used for learning (left side) and for recognition (right side) as well as the two matches established. 10 attentional fixations were used for learning and 15 fixations for recognition. The original resolution used for processing of the images is 1024x1536 pixels.
Figure 2: The 15 most salient patches of one indoor image shown merged as a cumulative image. The background (not selected parts of image) is shown slightly to provide context.
Figure 3: The 15 most salient patches of one indoor image shown merged as a cumulative image. The background (not selected parts of image) is shown slightly to provide context.
Figure 4: An example from the supplementary video provided. In the upper part the original 320x240 frame as recorded by the camera of the robot is shown and in the lower part the 3 most salient patches as used for learning and recognition are shown. We provide a MPEG1 encoded version of the video in the supplement. Additional versions (MPEG4, WMV) are provided on http://www.klab.caltech.edu/~urut/cvpr04.