Abstract

Despite recent advancements in automated image understanding, action recognition remains a task only reliably performed by humans. Elements of human interpretation, most notably human gaze patterns, can be harnessed to discriminate between images of different action classes and can shed light on human perception of these actions. A subset of images from the challenging PASCAL VOC 2012 Actions image set, which includes images of ten different action classes, was selected as a dataset and gaze data collected over the dataset were visualized. Illustrations of gaze transitions, clustering of fixations with a Gaussian mixture model and the creation of fixation density maps elucidate spatial, temporal and durational patterns in gaze. These patterns are then quantified to create novel gaze features, while visual features derived from a fixed-subregion iteration of the state-of-the-art method of regularized max pooling are obtained as a baseline for comparison. Two Support Vector Machine classifiers are trained, one with gaze features and the other with state-of-the-art features. Furthermore, gaze and the baseline were combined on the classifier level by summing the weighted confidence vector outputs of the classifiers and on the feature level by weighting baseline features by fixation density map gaze features. Confusion matrices computed for the gaze classifier revealed that intuitively similar action classes were frequently misclassified among each other, and four distinct class groups were identified. When classifiers were retrained to discriminate between these groups, performance of the gaze classifier improved significantly relative to that of the baseline classifier, demonstrating that the four class groups were behaviorally meaningful. Furthermore, the confidence vector combination method outperformed the baseline overall, illustrating how gaze can improve state of the art methods of automated action classification.