Modeling Geometric-Temporal Context With Directional Pyramid Co-Occurrence for Action Recognition

Abstract:

In this paper, we present a new geometric-temporal representation for visual action recognition based on local spatio-temporal features. First, we propose a modified covariance descriptor under the log-Euclidean Riemannian metric to represent the spatio-temporal cuboids detected in the video sequences. Compared with previously proposed covariance descriptors, our descriptor can be measured and clustered in Euclidian space. Second, to capture the geometric-temporal contextual information, we construct a directional pyramid co-occurrence matrix (DPCM) to describe the spatio-temporal distribution of the vector-quantized local feature descriptors extracted from a video. DPCM characterizes the co-occurrence statistics of local features as well as the spatio-temporal positional relationships among the concurrent features. These statistics provide strong descriptive power for action recognition. To use DPCM for action recognition, we propose a directional pyramid co-occurrence matching kernel to measure the similarity of videos. The proposed method achieves the state-of-the-art performance and improves on the recognition performance of the bag-of-visual-words (BOVWs) models by a large margin on six public data sets. For example, on the KTH data set, it achieves 98.78% accuracy while the BOVW approach only achieves 88.06%. On both Weizmann and UCF CIL data sets, the highest possible accuracy of 100% is achieved.