

Action Recognition Using Nonnegative Action Component Representation and Sparse Basis Selection

Abstract:

In this paper, we propose using high-level action units to represent human actions in videos and, based on such units, a novel sparse model is developed for human action recognition. There are three interconnected components in our approach. First, we propose a new context-aware spatial-temporal descriptor, named locally weighted word context, to improve the discriminability of the traditionally used local spatial-temporal descriptors. Second, from the statistics of the context-aware descriptors, we learn action units using the graph regularized nonnegative matrix factorization, which leads to a part-based representation and encodes the geometrical information. These units effectively bridge the semantic gap in action recognition. Third, we propose a sparse model based on a joint $l_{2,1}$ -norm to preserve the representative items and suppress noise in the action units. Intuitively, when learning the dictionary for action representation, the sparse model captures the fact that actions from the same class share similar units. The proposed approach is evaluated on several publicly available data sets. The experimental results and analysis clearly demonstrate the effectiveness of the proposed approach.