Abstract

Analyzing video and getting information from it is a growing field in computer-vision. The differential feature of a video compared to an image is motion. A video is a very high dimensional data and contains information of the environment, which is important for applications such as navigation, surveillance and video indexing. Therefore, extracting a compact representation of a video is required which can be used for various applications. There are various methods for estimating the motion from a video. One of the popular methods for estimating motion is optical flow.

The main theme of this work is to separate motions of different objects, such as background and foreground, by computing higher order derivatives of the motion vector. We proposed a method to compute differential motion on optical flow, which captures the motion of moving objects with respect to their potentially non-stationary backgrounds. Two methods based on curl and divergence for computing differential motion are proposed. The curl of optical flow is used to determine the interest points in video and extract features around it. We demonstrate the robustness of the proposed curl-based detector under common video transformations. We also proposed a descriptor that captures the location of the interest point with respect to neighboring interest points, which contains important information that state-of-the-art descriptors do not capture. The combination of the proposed detector and descriptor in a bag-of-features action recognition framework tested competitively with state-of-the-art methods.

In the second method, divergence is used for computing differential motion maps. We project differential motion maps on Cartesian planes and interpolate them to a fixed size. Even after this projection, the dimension of the feature vector was very high.

Based on the insight that contributing features of multiple moving objects in a video are likely to amalgamate additively on our representation, we show that semi-supervised non-negative matrix factorization performs better than other techniques to reduce the feature dimension. We demonstrate that optical flow, which captures absolute motion, is inadequate to capture discriminating features due to its inability to capture relative motion between an object and its background when the camera is in motion. The method to capture differential motion proposed by us is more effective than optical flow.

Recently, Convolutional Neural Networks (CNN) have been producing state-of-the-art results in classifying images of objects, complex events, and scenes. We have also proposed a method to detect abnormal events for human group activities using CNN. Our main contribution is to develop a strategy that learns with very few videos by isolating the action and by using supervised learning. First, we subtract the background of each frame by modeling each pixel as a mixture of Gaussians(MoG) to concatenate the higher order learning only on the foreground. Next, features are extracted from each frame using a convolutional neural network (CNN) that is trained to classify between normal and abnormal frames. These feature vectors are fed into long short term memory (LSTM) network to learn the long-term dependencies between frames. The LSTM is also trained to classify abnormal frames, while extracting the temporal features of the frames. Finally, we classify the frames as abnormal or normal depending on the output of a linear SVM, whose inputs are the features computed by the LSTM.

Evaluation of proposed methods is done on the popular benchmark datasets for action recognition task.