World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:17, No:08, 2023

# The Application of Video Segmentation Methods for the Purpose of Action Detection in Videos

**Authors :** Nassima Noufail, Sara Bouhali

**Abstract :** In this work, we develop a semi-supervised solution for the purpose of action detection in videos and propose an efficient algorithm for video segmentation. The approach is divided into video segmentation, feature extraction, and classification. In the first part, a video is segmented into clips, and we used the K-means algorithm for this segmentation; our goal is to find groups based on similarity in the video. The application of k-means clustering into all the frames is time-consuming; therefore, we started by the identification of transition frames where the scene in the video changes significantly, and then we applied K-means clustering into these transition frames. We used two image filters, the gaussian filter and the Laplacian of Gaussian. Each filter extracts a set of features from the frames. The Gaussian filter blurs the image and omits the higher frequencies, and the Laplacian of gaussian detects regions of rapid intensity changes; we then used this vector of filter responses as an input to our k-means algorithm. The output is a set of cluster centers. Each video frame pixel is then mapped to the nearest cluster center and painted with a corresponding color to form a visual map. The resulting visual map had similar pixels grouped. We then computed a cluster score indicating how clusters are near each other and plotted a signal representing frame number vs. clustering score. Our hypothesis was that the evolution of the signal would not change if semantically related events were happening in the scene. We marked the breakpoints at which the root mean square level of the signal changes significantly, and each breakpoint is an indication of the beginning of a new video segment. In the second part, for each segment from part 1, we randomly selected a 16-frame clip, then we extracted spatiotemporal features using convolutional 3D network C3D for every 16 frames using a pre-trained model. The C3D final output is a 512-feature vector dimension; hence we used principal component analysis (PCA) for dimensionality reduction. The final part is the classification. The C3D feature vectors are used as input to a multi-class linear support vector machine (SVM) for the training model, and we used a multi-classifier to detect the action. We evaluated our experiment on the UCF101 dataset, which consists of 101 human action categories, and we achieved an accuracy that outperforms the state of art by 1.2%.

**Keywords :** video segmentation, action detection, classification, Kmeans, C3D
**Conference Title :** ICPRCV 2023 : International Conference on Pattern Recognition and Computer Vision
**Conference Location :** Paris, France
**Conference Dates :** August 24-25, 2023