Scene Separation & Data Selection: Temporal Segmentation Algorithm for Real-time Video Stream Analysis

Yuelin Xin[1,2] Zihan Zhou[1] Yuxuan Xia[1]

[1]SWJTU-Leeds Joint School, CS
Southwest Jiaotong University

[2]School of Computing
University of Leeds

Spatio-Temporal Reasoning and Learning, 2022

Outline

Introduction

- The problem (Background & What we want to achieve)
- Our motivation (Why not neural networks?)

Remark

Scene separation is a problem in which we want to separate a video stream into different scenes. **A scene** is defined as a group of similar-looking frames that are temporally adjacent to each other.

The problem

- **Background**: real-time video stream interpretation, including video semantics / video accessibility / surveillance footage auto-interpretation, etc.
- **Difficulties**: algorithms do not see video as a continuous stream of images, but as discrete frames.

Figure 1: Video semantics.

The problem

- **The traditional approach**: 3D CNNs
- **What's missing**: algorithms do not see video as a continuous stream of images, but as discrete frames.

Our motivation

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$, and let

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i}^{n} X_i$$

denote their mean. Then as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$.