Unsupervised Learning of Procedures from Tutorial Videos

Amlan Sengupta
Aparna Krishnan
Arunima S

SSN College of Engineering, Chennai

October 31, 2020

Introduction

1. Tutorial videos are a popular format of online learning today.
2. These videos are extremely useful. However, they are not always accessible:
   - No access to the Internet or not having enough bandwidth to stream videos
   - Time constraints that prevent the learner from watching the entire video
3. Textual instructions are a viable alternative.
4. The goal is to create an efficient, reliable system that can translate a video into a textual set of step-by-step instructions.

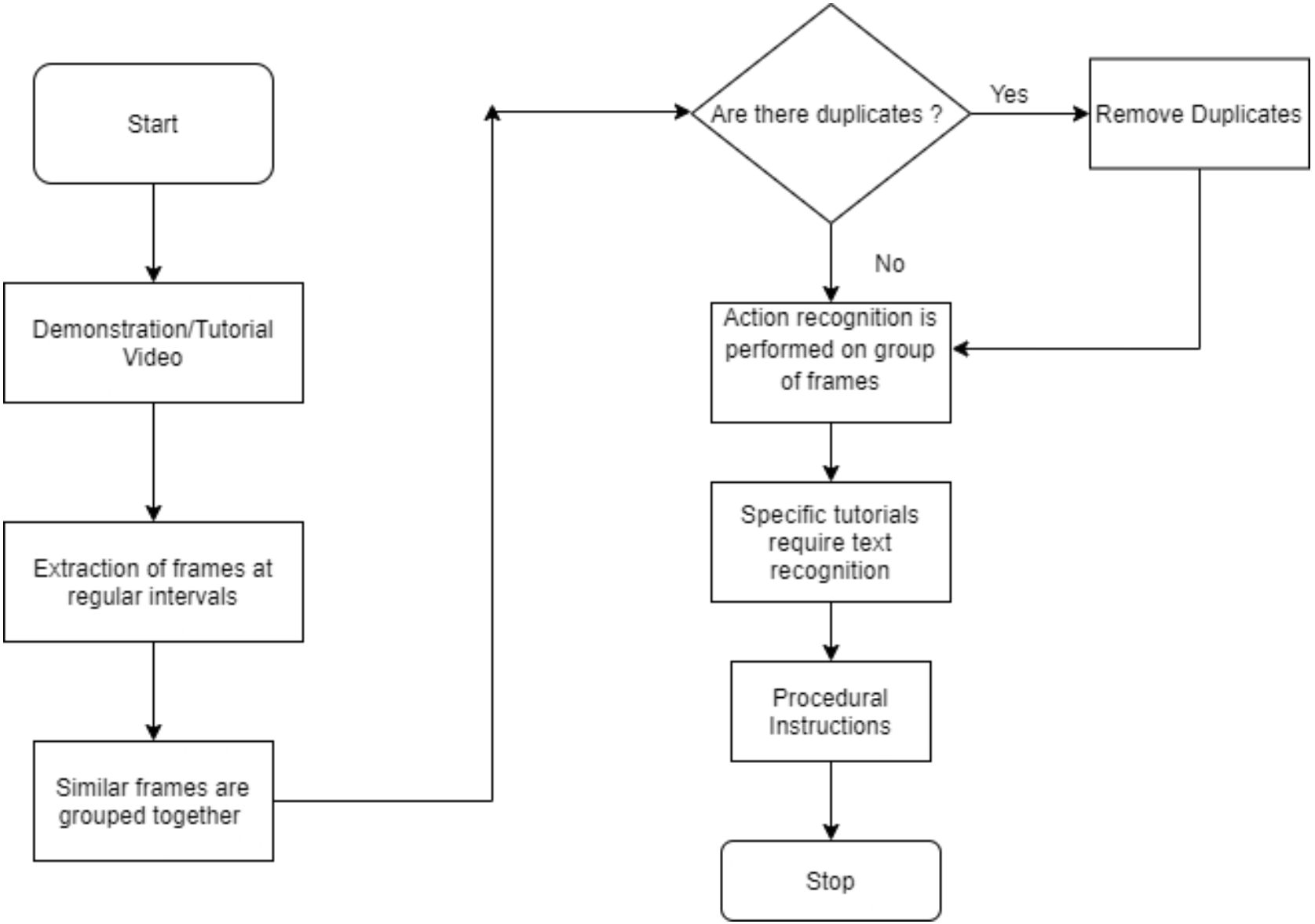
Papers from related fields and references :

- Unsupervised Learning of Procedures from Demonstration Videos
- Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words
- Improved Spatio-temporal Salient Feature Detection for Action Recognition
- Unsupervised learning from video with deep neural embeddings
- Unsupervised learning from narrated instruction videos
- CompILE: Compositional imitation learning and execution

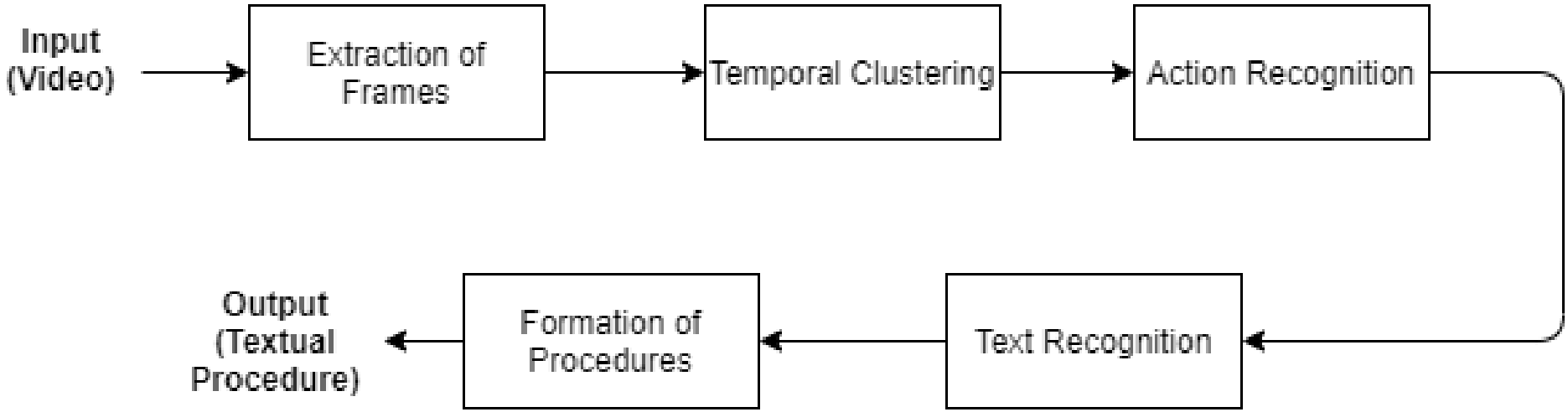
Salient features of our project :

- Past works use only one methodology: temporal clustering OR action recognition.
- Our project combines these methodologies.
- This will achieve better grouping of similar videos.
- Text recognition will also cover academic tutorial videos.

Proposed System



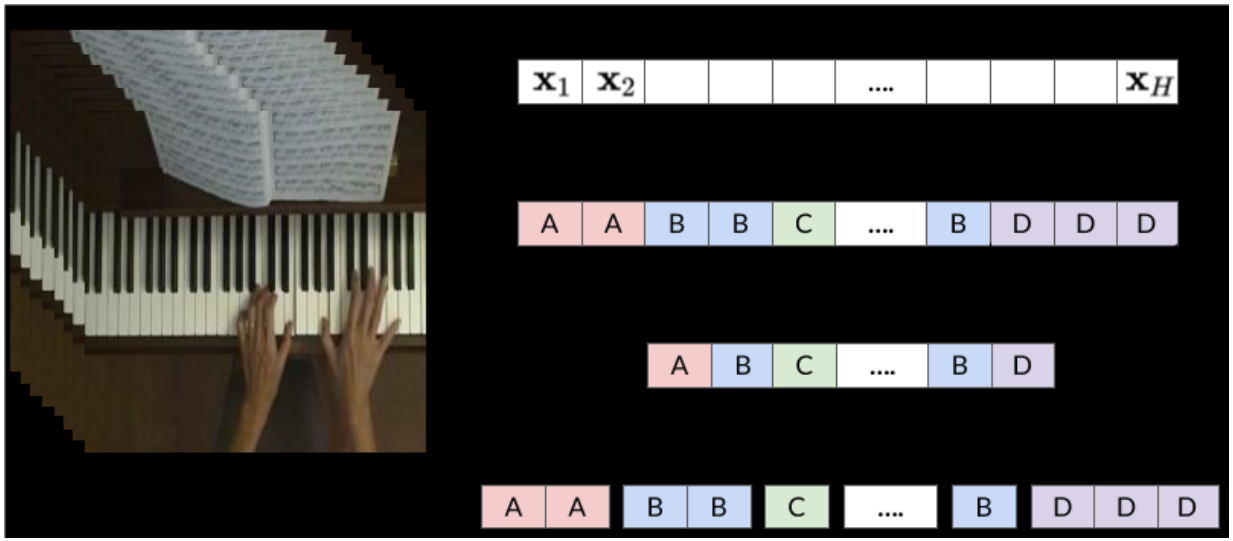
Modules Split-up



Extraction of Frames

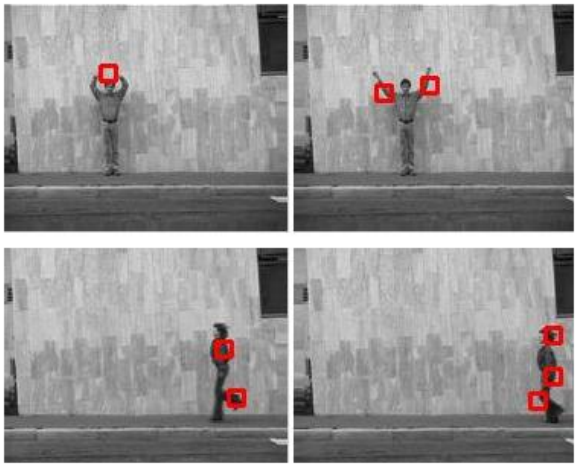
Extracting each frame from the tutorial video using OpenCV.

Temporal Clusters



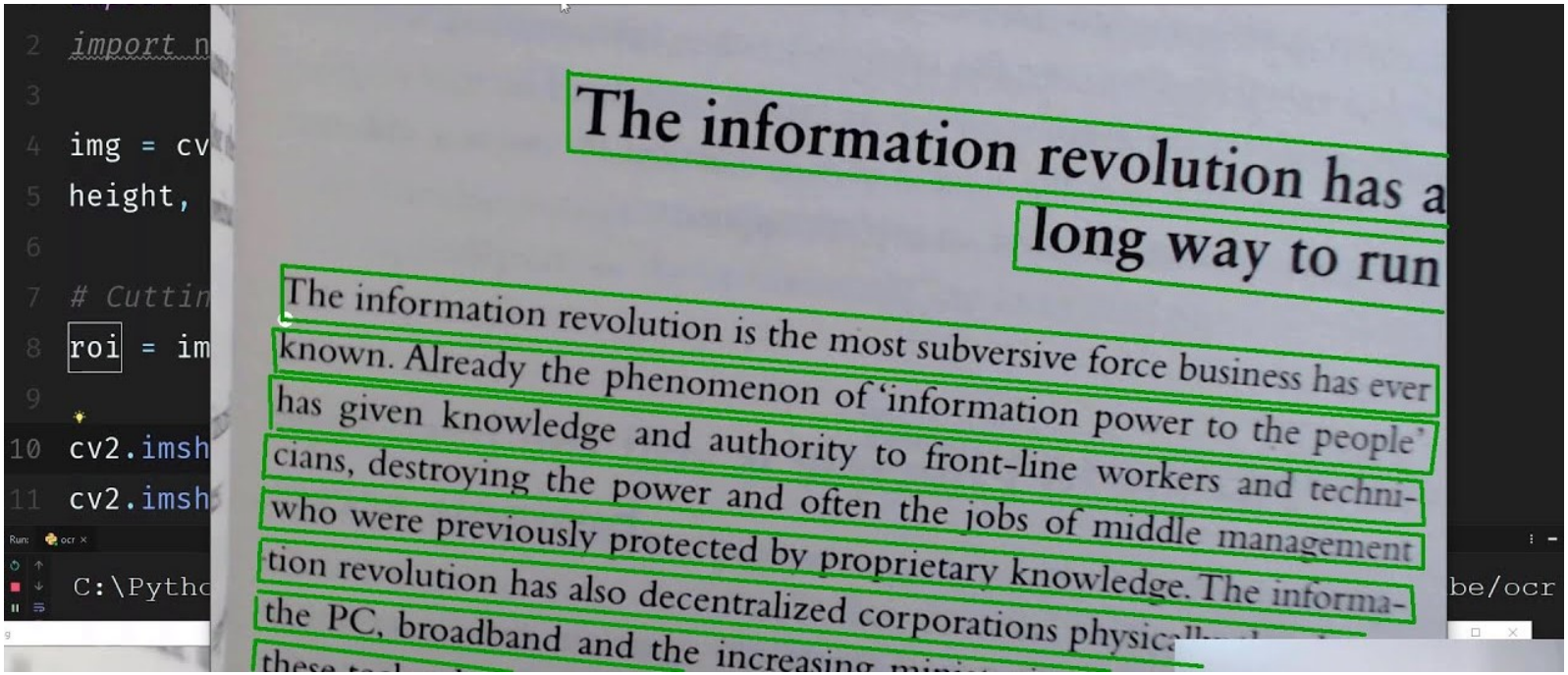
Factorization of multiple time series into a set of non-overlapping segments that belong to k temporal clusters.

Action Recognition



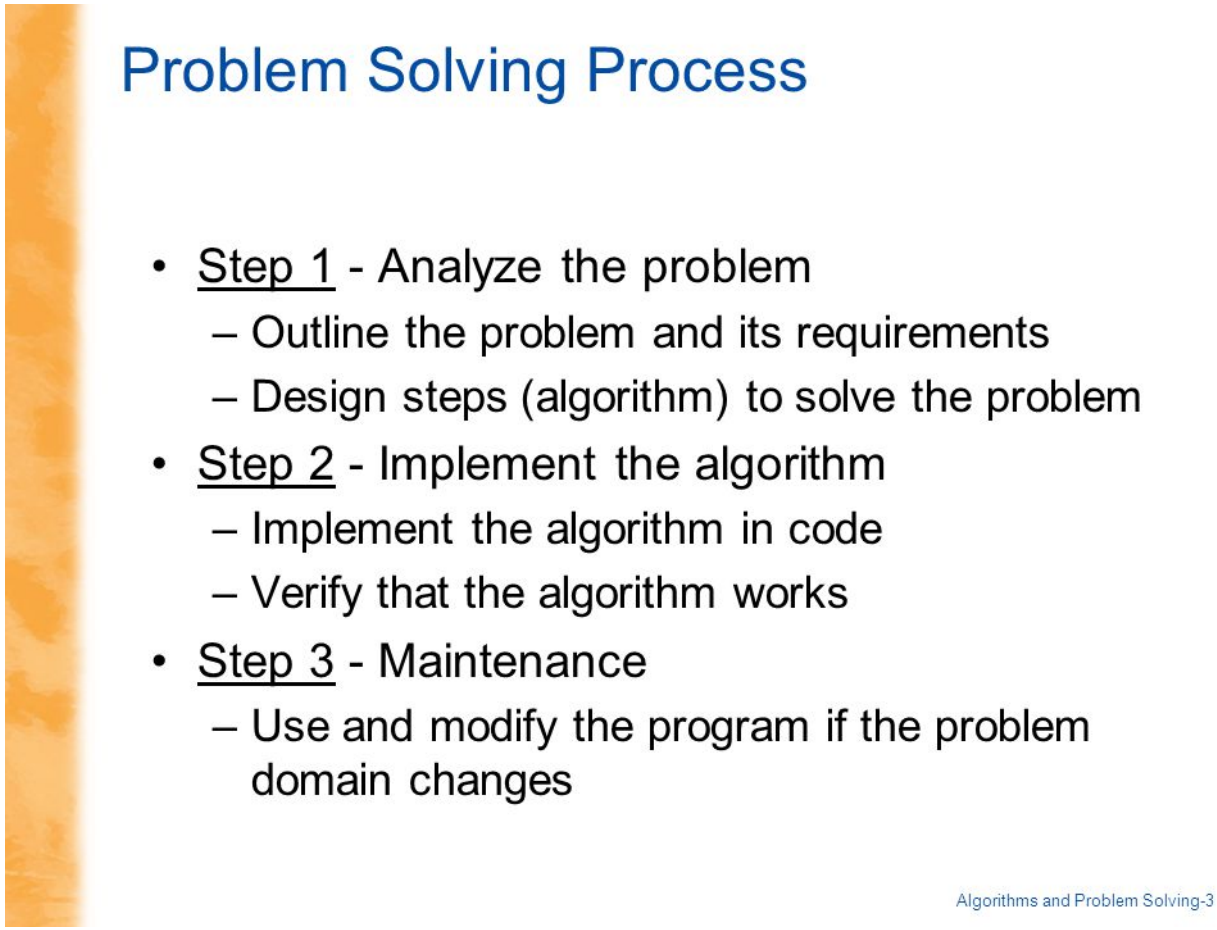
Identification of different actions from video clips (a sequence of 2D frames) or images using Video Instance Embedding (VIE) framework. Actions may or may not be performed throughout the video.

Text Recognition



The electronic or mechanical conversion of images of typed, handwritten or printed text into digital text using OCR package from python.

Desired Output - Procedural Instructions



The final output of the system will be clear, step-by-step textual instructions.

Project Timeline

Months	1	2	3	4	5	6
Planning						
Extraction Of Frames						
Temporal Clustering						
Action Recognition						
Text Recognition						
Formation of Procedure						
Testing						
Delivery						

Thank You