Aligning Videos In Space and Time

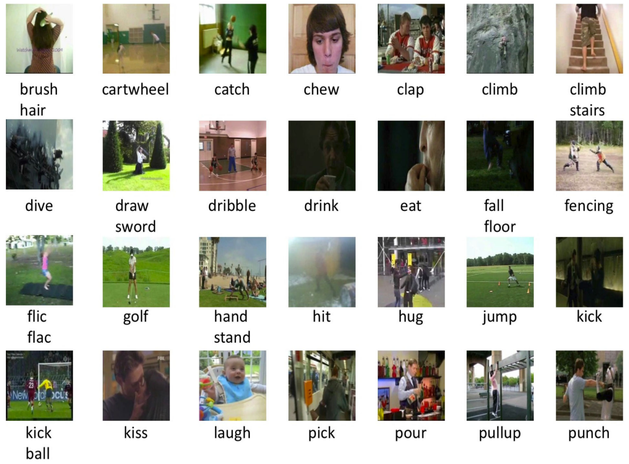
Senthil Purushwalkam[1], Tian Ye[1], Saurabh Gupta[3], Abhinav Gupta[1,2]

[1]Carnegie Mellon University
[2]Facebook AI Research
[3]University of Illinois Urbana-Champaign (UIUC)

European Conference on Computer Vision (ECCV), 2020

Short Presentation

Video Understanding in Computer Vision

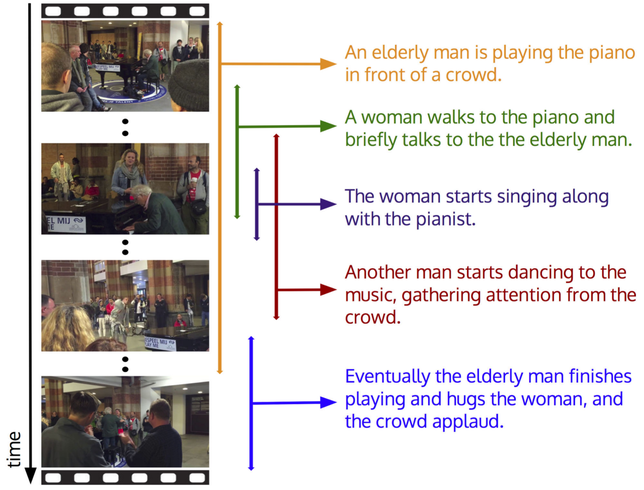


Action Recognition

Classification of Videos into Predefined Action Categories

Action Detection

Localizing Predefined Actions Temporally in Videos



Video Captioning

Generating Textual Descriptions for Videos

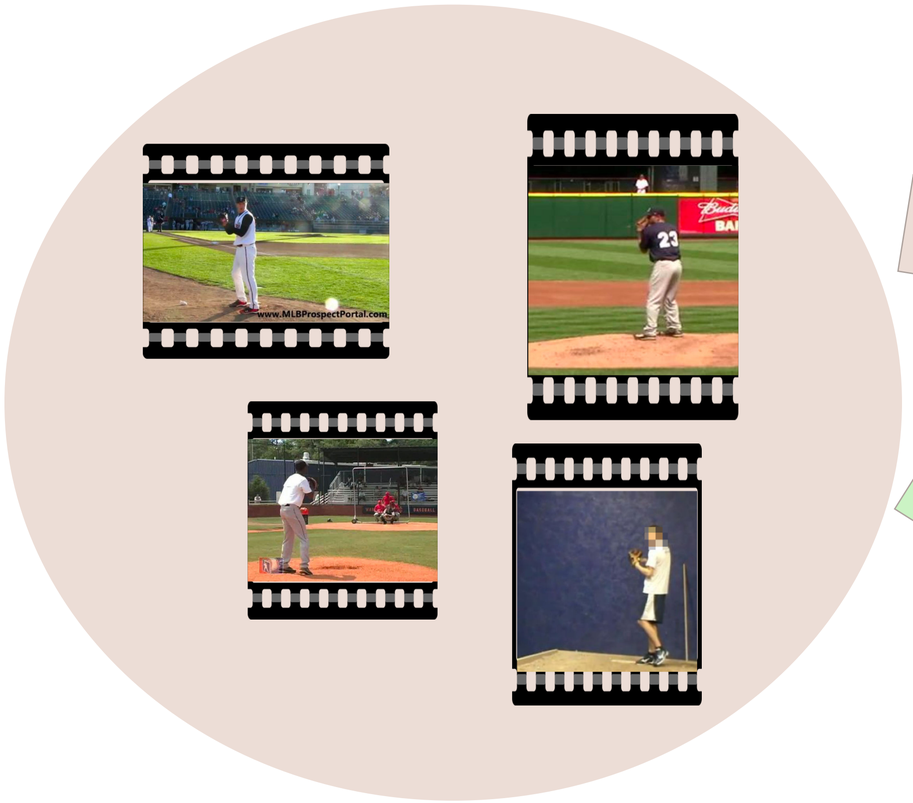
- Coarse Understanding of Videos

- Data Collection is Not Scalable to Denser Annotations

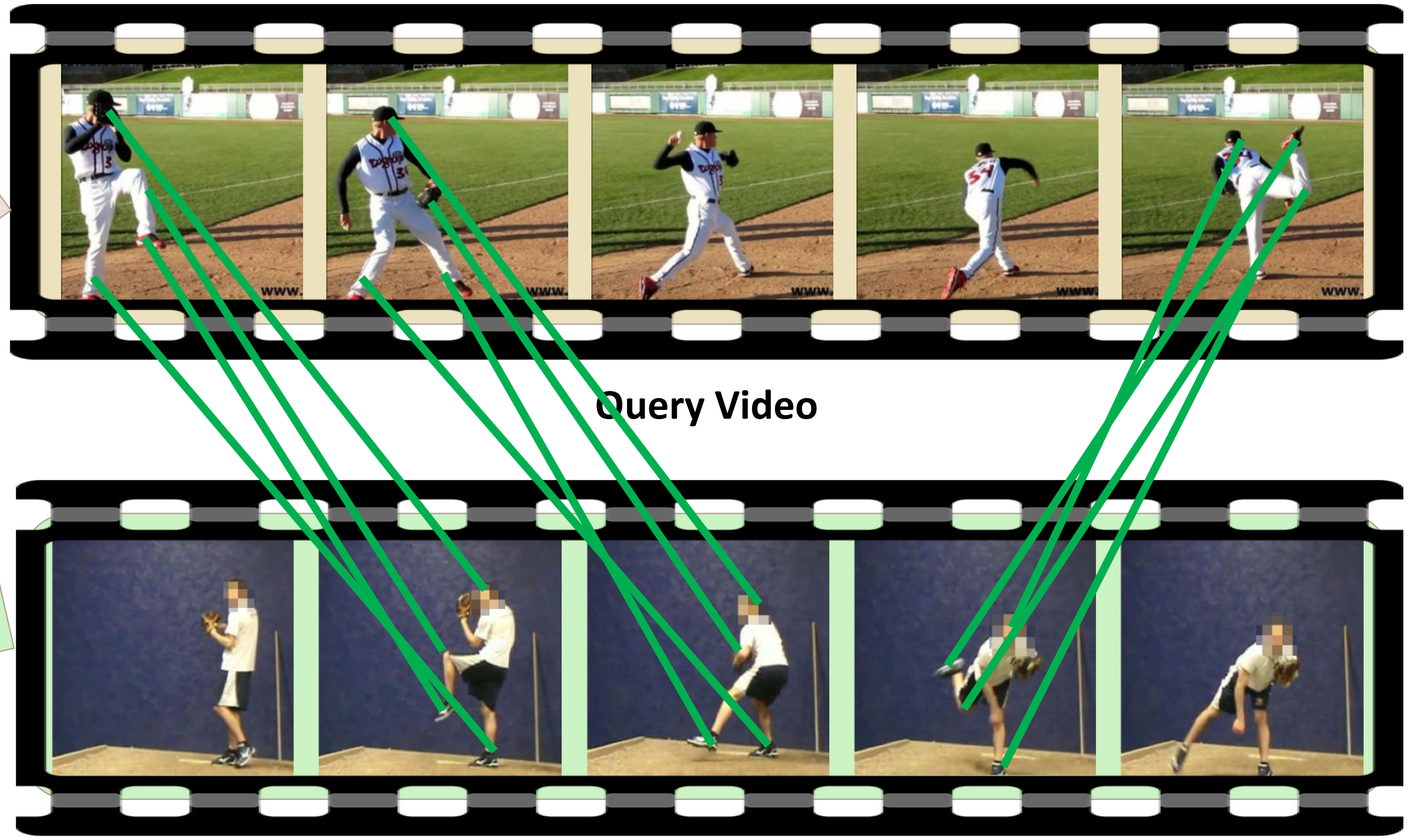
Figures taken from Kuehne et al. ICCV 2011; Krishna et al. ICCV 2017; Xiong et al. arXiv 2017.

Video Understanding via Association

Ask not “what is this?”, ask “what is this like”.
-Moshe Bar



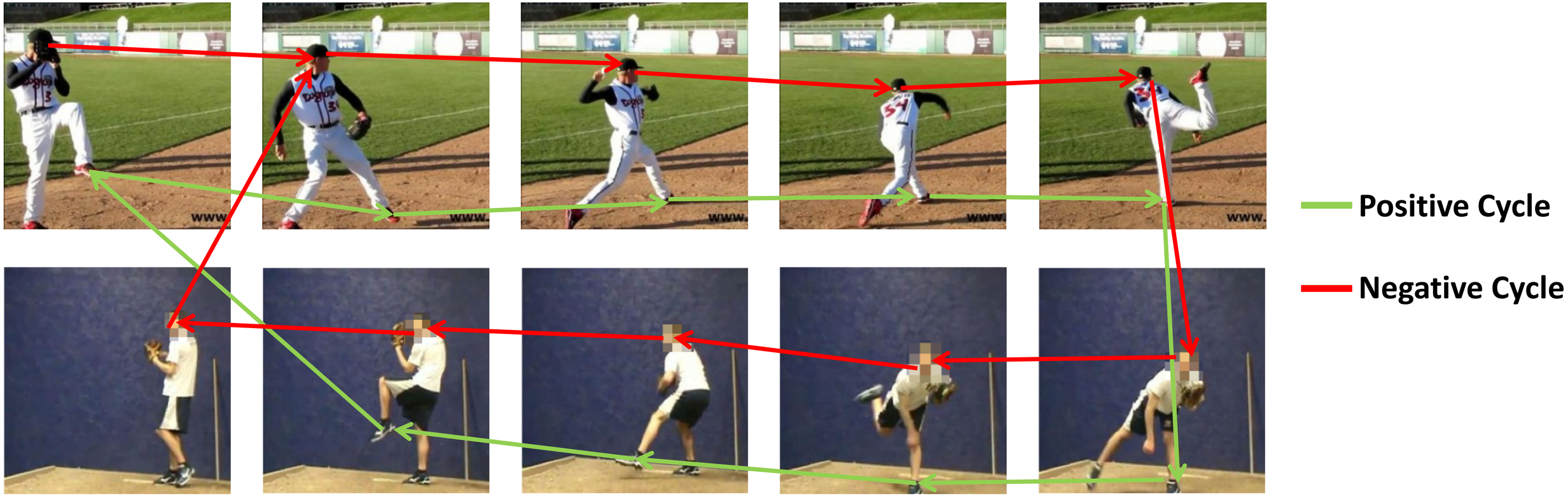
Baseball Bowling



Query Video

Retrieved Video

Unsupervised Learning of Spatio-Temporal Associations



Qualitative Results

Learned representation can effectively ***spatio-temporally align videos***

Nearest neighbor using learned representation shows that it ***encodes object states and appearance***

Thank You for Listening!

Project Webpage: http://www.cs.cmu.edu/~spurushw/publication/alignvideos/