Improving Concept Detection by Utilizing Temporal Relationships

Khurram Soomro
University of Central Florida

Christian Weigandt
High Point University

Levi Smith
University of Central Florida

University of Central Florida

UCF Center for Research in Computer Vision

I. Introduction

- We aim to improve concept (action) detection using the temporal dependence from one concept to the next, which we define as *causality*
- The task is to temporally locate concepts in a long video and identify them
- We implement this using a Markov chain-inspired graph-based approach
- We have applied our method to both:
  1. A self-annotated soccer dataset
  2. The 25 events of the TRECVID dataset

II. Datasets

1. Soccer Dataset

- 14 different concepts

1. Arguing
2. Attempt at Goal
3. Card (Red or Yellow)
4. Celebration
5. Corner
6. Foul
7. Free Kick
8. Goal
9. Goal Kick
10. Injury
11. Offside
12. Penalty
13. Substitution
14. Throw In

2. TRECVID Dataset (EC11 and EC12)

- 25 events and 185 different concepts

- **EC11 Events**

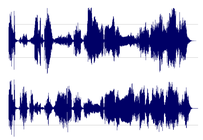
1. Attempting a Board Trick
2. Feeding an Animal
3. Fishing
4. Wedding
5. Woodworking
6. Birthday Party
7. Changing a Vehicle Tire
8. Flash Mob Gathering
9. Getting a Vehicle Unstuck
10. Grooming an Animal
11. Making a Sandwich
12. Parade
13. Parkour
14. Repairing an Appliance
15. Working on a Sewing Project

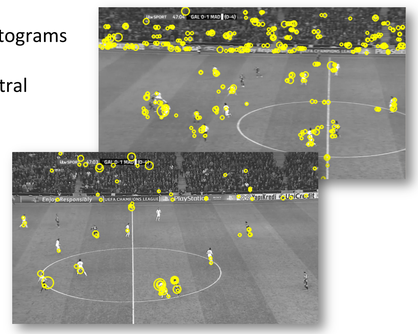
- **EC12 Events**

1. Attempting a Bike Trick
2. Cleaning an Appliance
3. Dog Show
4. Giving Directions to a Location
5. Marriage Proposal
6. Renovating a Home
7. Rock Climbing
8. Town Hall Meeting
9. Winning a Race (No Vehicle)
10. Working on a Metal Crafts Project

III. Features for Soccer Dataset

- STIP – Space Time Interest Points
- MBH – Motion Boundary Histograms
- MFCC – Mel-Frequency Cepstral Coefficients



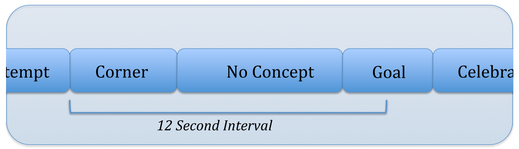


IV. Causality

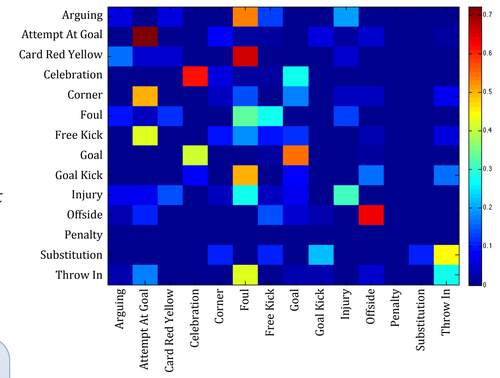
We define the causality between two concepts, A and B, by the equation:

$$p_t(B|A) = \frac{|B \cap A|_t}{|A|}$$

- $|B \cap A|_t$ is defined as the number of times where B is the first concept that occurs within t seconds of A
- $|A|$ is defined as the number of times that concept A occurred



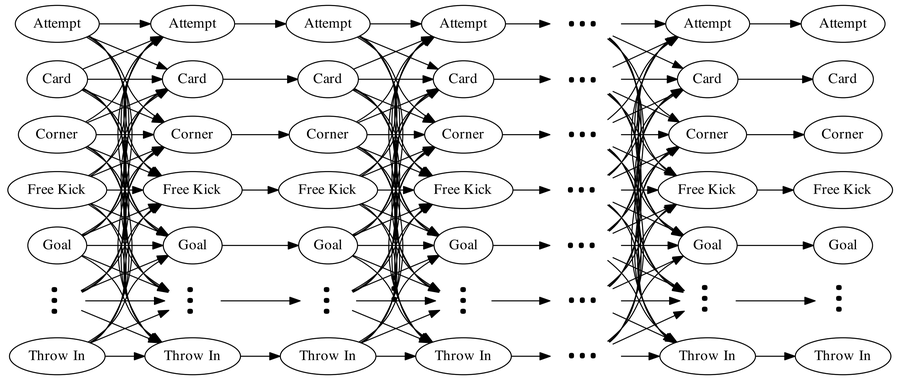
Causality Matrix for Soccer Dataset *



In this example, corner would be said to have *caused* goal, and the probability of a corner leading to a goal would increase.

* normalized across rows to highlight relationships

V. Graphical Model



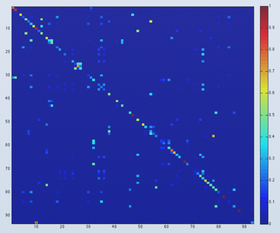
Matrix Representation

$$\begin{bmatrix} P_{K,1} \\ \vdots \\ P_{K,N} \end{bmatrix} = \left(\begin{bmatrix} C_{1,1} & \dots & C_{1,N} \\ \vdots & \ddots & \vdots \\ C_{N,1} & \dots & C_{N,N} \end{bmatrix}^T \begin{bmatrix} P_{K-1,1} \\ \vdots \\ P_{K-1,N} \end{bmatrix} \right) \cdot \begin{bmatrix} S_{K,1} \\ \vdots \\ S_{K,N} \end{bmatrix}$$

$P_{(i,a)}$: Updated probability that clip i contains concept a

$S_{(i,a)}$: SVM probability that clip i contains concept a

$C_{(a,b)}$: Probability that concept a *causes* concept b

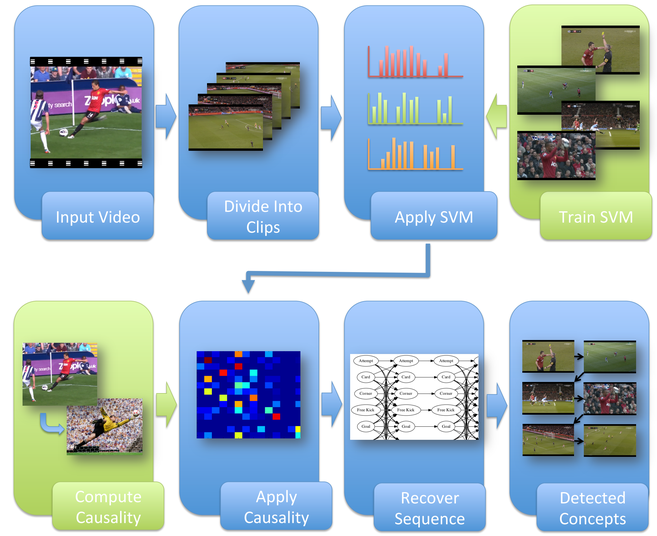


An example causality matrix taken from the EC11 dataset

Equation for a Single Concept

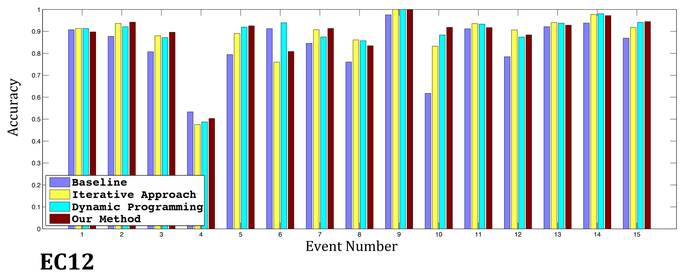
$$P_{(K,1)} = (P_{(K-1,1)}C_{1,1} + P_{(K-1,2)}C_{2,1} + \dots + P_{(K-1,N)}C_{N,1}) * S_{(K,1)}$$

VI. Process

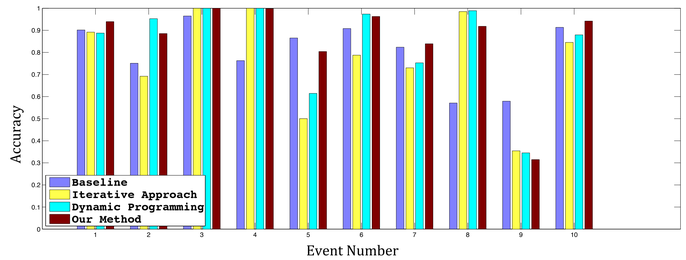


VII. Results

EC11



EC12



Average Accuracies

	Baseline	Iterative	Dynamic	Our Method
EC11	83.04%	87.56%	88.89%	88.57%
EC12	80.40%	77.86%	83.94%	86.04%
Soccer	21.77%	22.32%	21.48%	24.65%

VII. Conclusion

Temporal relationships between concepts can be utilized to increase the accuracy of concept detection across an event. We see improvements on 22 of the 25 events using one of the above methods (our method, iterative approach, dynamic programming). Each event has a unique set of concepts that may have a different causal timespan between them. The best timespan for each event was determined empirically and used to attain the best results possible for each individual event. The improvement using temporal relationships may vary depending on the strength of the causality between concepts in an event.

Special thanks to: Dr. Mubarak Shah, Dr. Niels Lobo, Amir Roshan Zamir, Afshin Dehghan, Shayan Modiri Assari, the CRCV staff, and the NSF