NEW

Problem Setting

- **Goal**: Given two videos, discover common events in an unsupervised fashion.
- We name this problem Temporal Commonality Discovery (TCD).

- **TCD is difficult**
  1. No prior knowledge (what, where, and how many) on commonalities is provided.
  2. Exhaustive search approaches (e.g., sliding window) are computationally prohibitive.

- **Problem Interpretation**
  - A rectangle \( R \) indicates a solution.
  - Search over rectangle sets \( R \) to find most possible matches.

Differences with Previous Work

- **Difference between TCD and ESS [1] / STBB [2]**
  - Different learning framework
  - Unsupervised vs. Supervised
  - New bounding functions for TCD

- **Difference between TCD and [3]**
  - Different objective: Commonality Discovery vs. Temporal Clustering

The Proposed TCD Method

- **Problem Formulation**: \( f(A_1, A_2) = \min \{ \frac{1}{n} \sum_{i=1}^{n} d(i) \} \) subject to \( n \geq \epsilon \).
- **Optimization** by Branch and Bound (B&B)

- **Derive Bounding functions**
  1. Bounding L1 distance:
     \[ d(i) = \max_{j=1}^{n} \left( \sum_{k=0}^{j} \left( A_{1,k} - A_{2,k} \right) \right) \]
  2. Bounding L2 distance:
     \[ d(i) = \max_{j=1}^{n} \left( \sum_{k=0}^{j} \left( A_{1,k} - A_{2,k} \right)^2 \right)^{1/2} \]
  3. Bounding L\( \infty \) distance:
     \[ d(i) = \max_{j=1}^{n} \left( \sum_{k=0}^{j} \left( A_{1,k} - A_{2,k} \right)^{\epsilon/\epsilon} \right) \]

- **An example** for two synthetic time series:

Common Facial Events Discovery

- **Evaluation performed** on the RU-FACS database [3,4].
- **Facial feature representations**
  - Left leg
  - Left right leg
  - Left right arm
  - Appearance

Multiple Common Motions Discovery

- **Evaluation performed** on the CMU Mocap [6].

Search Video by Video

- Slight modification leads to efficient video indexing
  \[ \min_{(A_1, A_2)} d(A_1, A_2) \] where \( d \) is \( \leq \).
- **Example on the Hollywood dataset [7]**