Overview of the proposed method:

- **Goal**
  - Recognize people and their actions in video.
  - Use weak supervision derived from video scripts.
  - Address challenges of:
    - temporal localization,
    - spatial localization,
    - visual variability.

- **Contributions**
  - Joint model for weakly-supervised learning of actions and actors.
  - Solution in terms of quadratic problem with linear constraints.
  - Improved results for action and face recognition.

- **Overview**
  - Input: video sequence with associated script.
  - Output: tracks of people with names and action labels.

- **Discriminative Clustering**

- **Constraints**

- **Optimization**
  - Quadratic cost under quadratic constraints
  - Quadratic cost under linear constraints if we fix one of the variables

- **Results**
  - Results on actions
  - Comparison with state of the art ([2], [3])

**Experiments on two movies: Casablanca, American Beauty**

- **Features**
  - Detect Faces (4)
  - Group Faces into tracks
  - Locate Facial Landmarks
  - Warp to average location
  - Re-locate landmarks
  - Extract ista
  - Win-Hen kernel ([1])

**Examples**