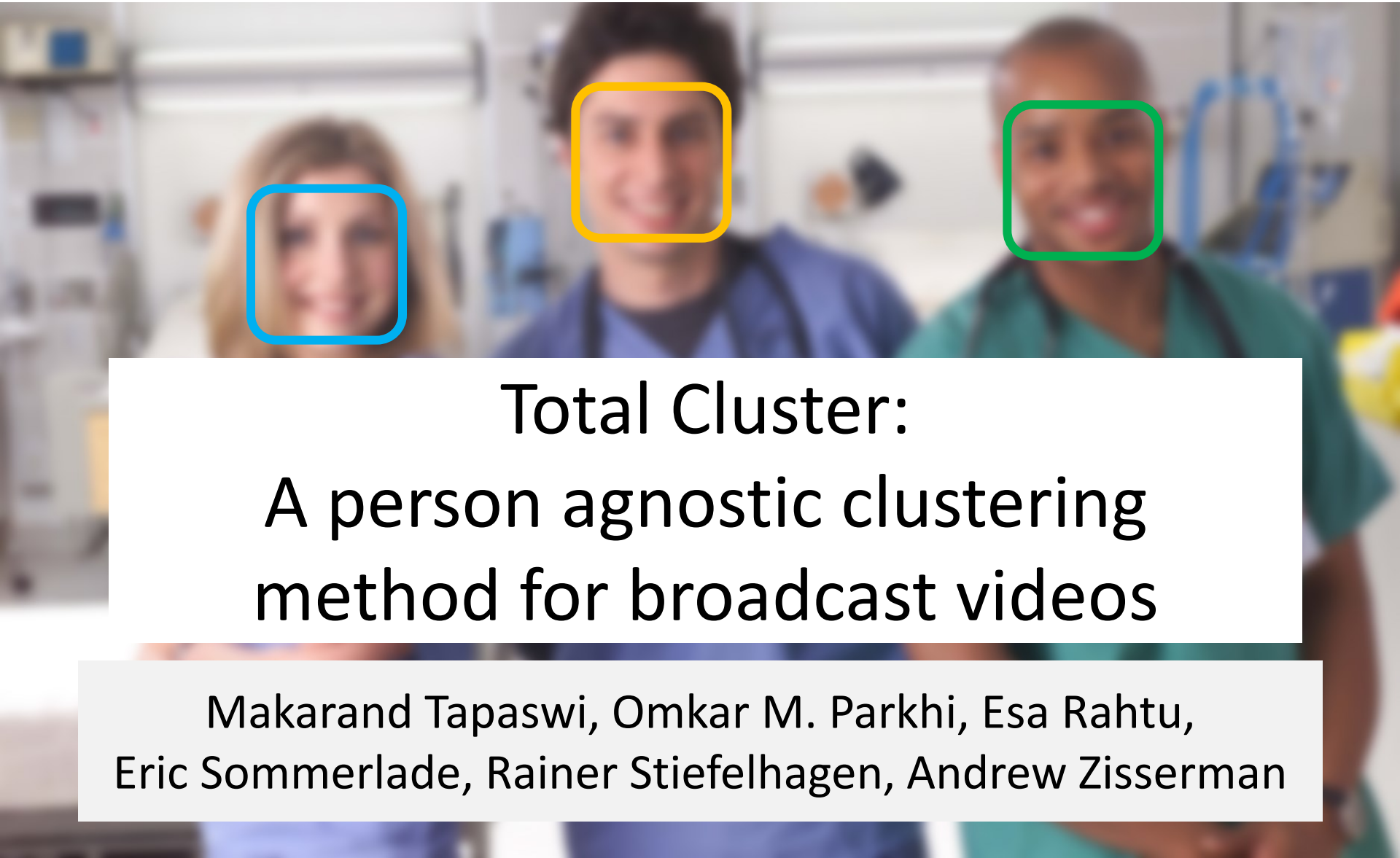
Karlsruhe Institute of Technology

UNIVERSITY of OULU
OULUN YLIOPISTO

UNIVERSITY OF OXFORD



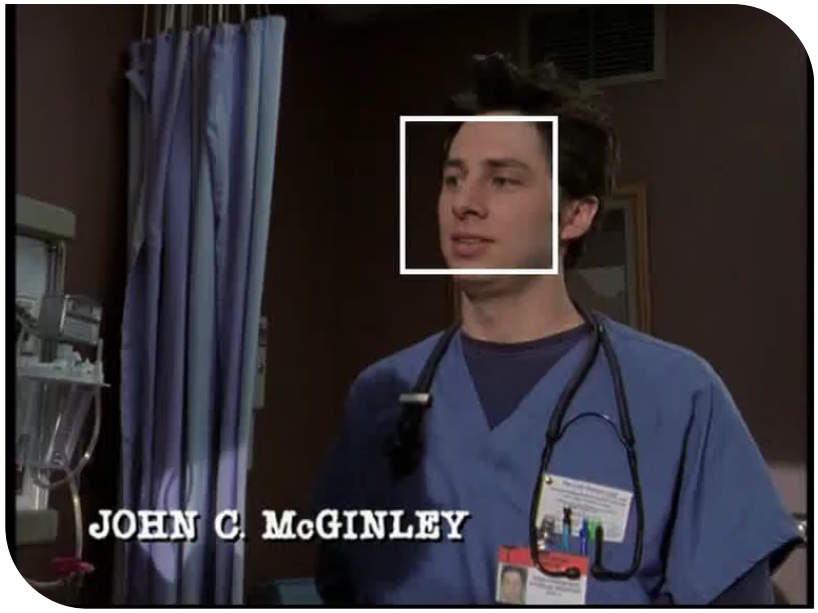
Total Cluster: A person agnostic clustering method for broadcast videos

Makarand Tapaswi, Omkar M. Parkhi, Esa Rahtu,
Eric Sommerlade, Rainer Stiefelhagen, Andrew Zisserman

Face track clustering

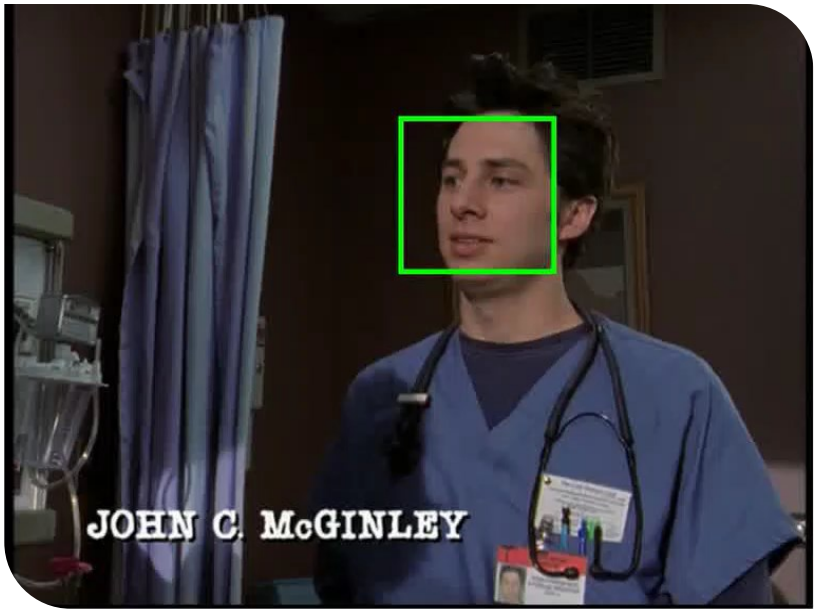
Input

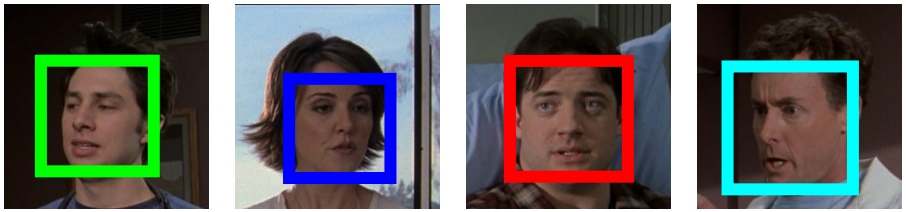
Video with face tracks



Output

Face track clusters



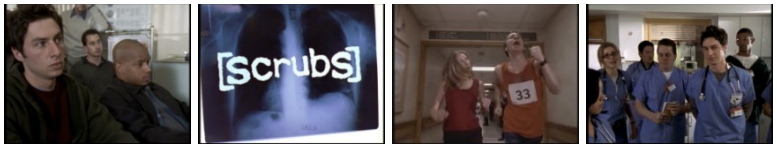


Why cluster face tracks?

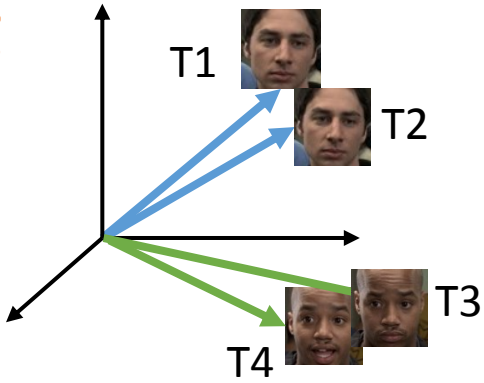
- Large number of videos
- Good basis for manual or automatic labelling
- Facilitates video retrieval or summarization

Overview

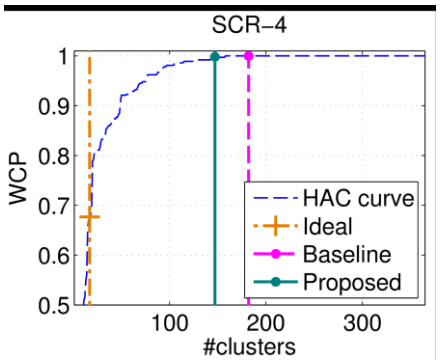
- **Editing structure in videos**
  - Shots, threads and scenes



- **Using editing structure for clustering**
  - Negative pairs
  - Scene level clustering
  - Episode level clustering



- **Dataset and evaluation**
  - Scrubs, Buffy
  - Weighted clustering purity
  - Clustering results



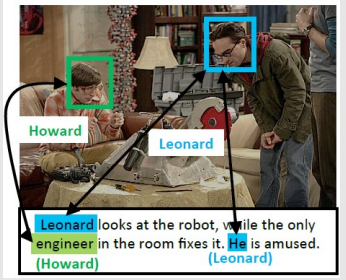
Related work

Person Identification in TV series

(Everingham et al., 2006)

(Cour et al., 2009)

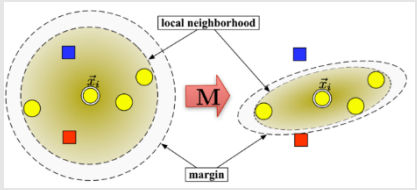
(Bojanowski et al., 2013)



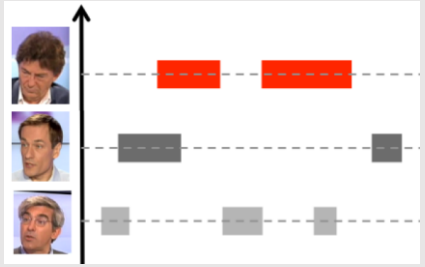
(Ramanathan et al., 2014)

Face Clustering

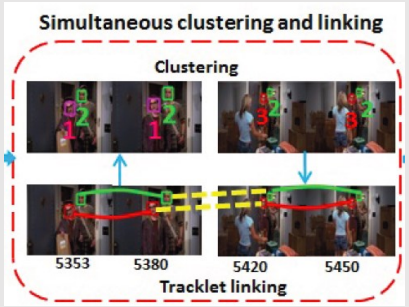
(Ramanan et al., 2007)



(Cinbis et al., 2011)



(Khoury et al., 2013)



(Wu et al., 2013, 2013b)

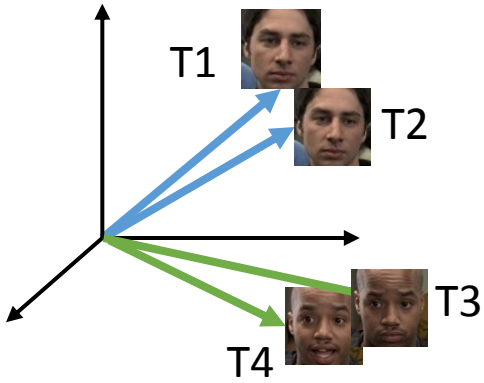
Novelty of the current work

- Error free clustering
- Utilize video-editing structure
- Use state-of-the-art face track descriptor

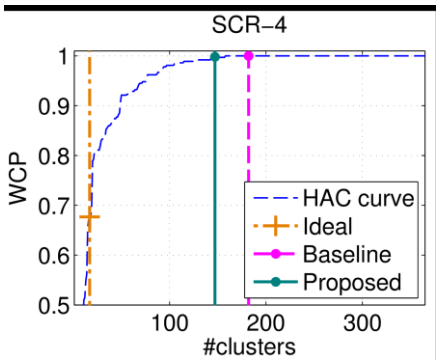
Overview

- **Editing structure in videos**
  - Shots, threads and scenes

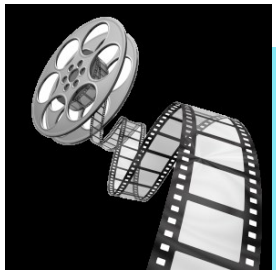
- Using editing structure for clustering
  - Negative pairs
  - Scene level clustering
  - Episode level clustering

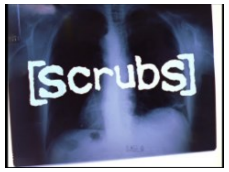


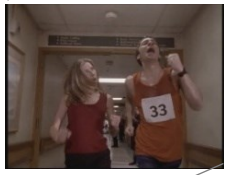
- Dataset and evaluation
  - Scrubs, Buffy
  - Weighted clustering purity
  - Clustering results



Video editing overview







• • •

Episode

Scenes

Threads

Shots

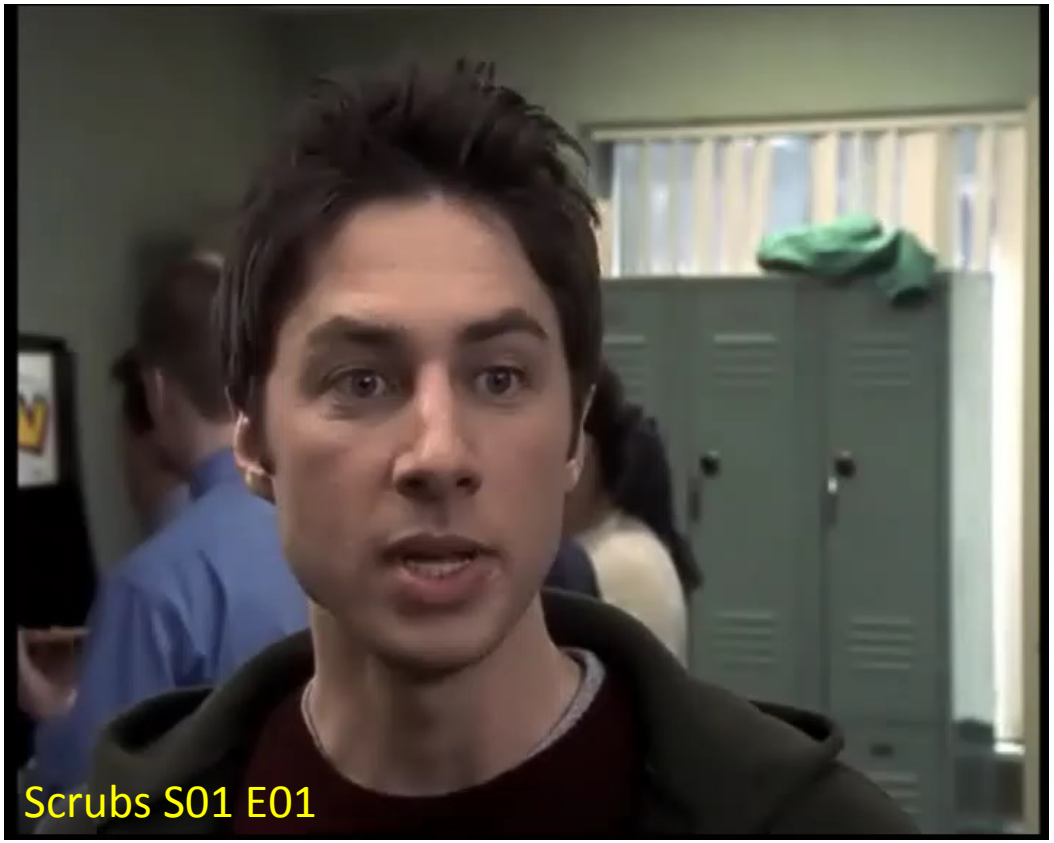
Shots and Threads

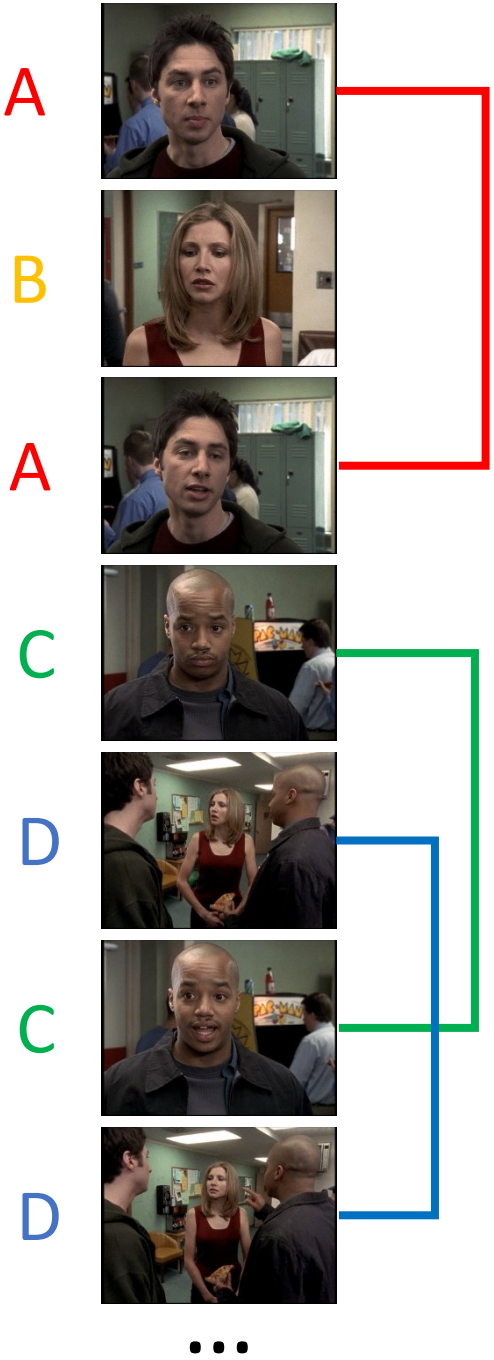
shot *(n.)*

A sequence of frames taken from the same camera without interruption

thread *(n.)*

A set of shots that share a theme with the same viewing angle and content





Scenes

scene *(n.)*

A sequence of shots in a video with the same characters, at the same location and time

Scene 4





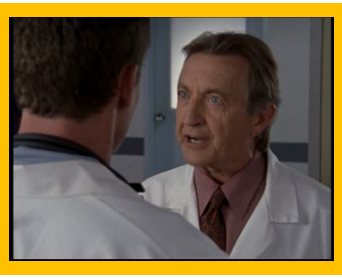
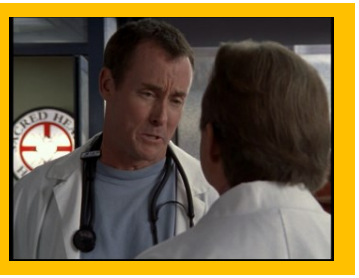
Scene 5





Scene 6



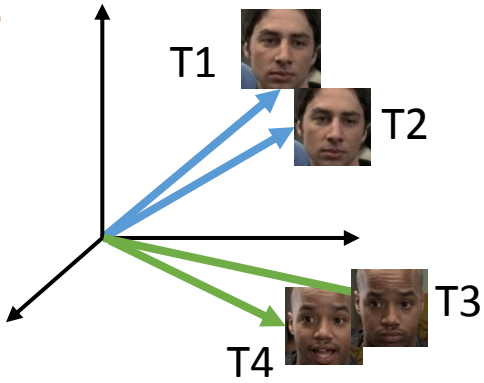
Scene boundary detection (Tapaswi et al. 2014)

- Maximize within-scene visual similarity
- Don't break shot threads

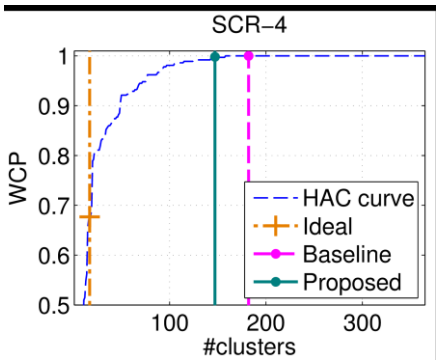
Overview

- Editing structure in videos
  - Shots, threads and scenes

- **Using editing structure for clustering**
  - Face tracking
  - Negative pairs
  - Scene level clustering
  - Episode level clustering



- Dataset and evaluation
  - Scrubs, Buffy
  - Weighted clustering purity
  - Clustering results



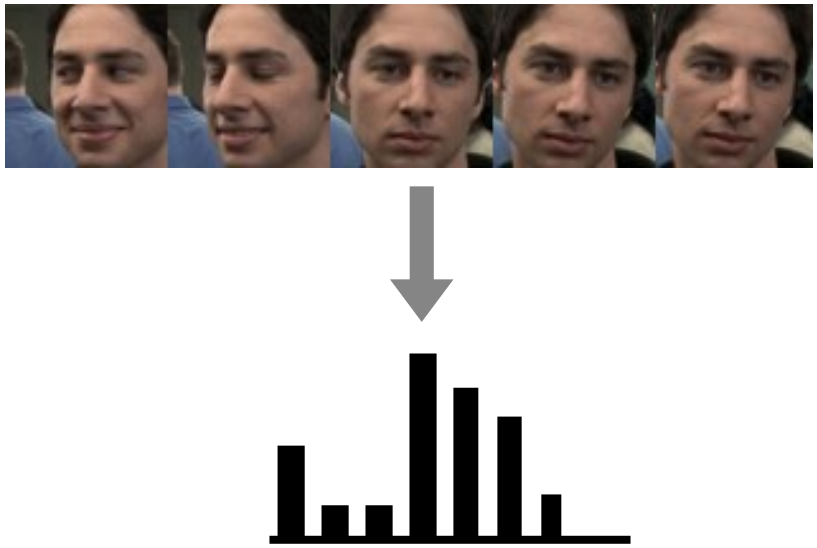
Face tracks

Face tracking

- frontal+profile VJ detector
- multi-pose head detector
- KLT tracking
- false positive removal

Face track descriptor

(Parkhi et al. 2014)



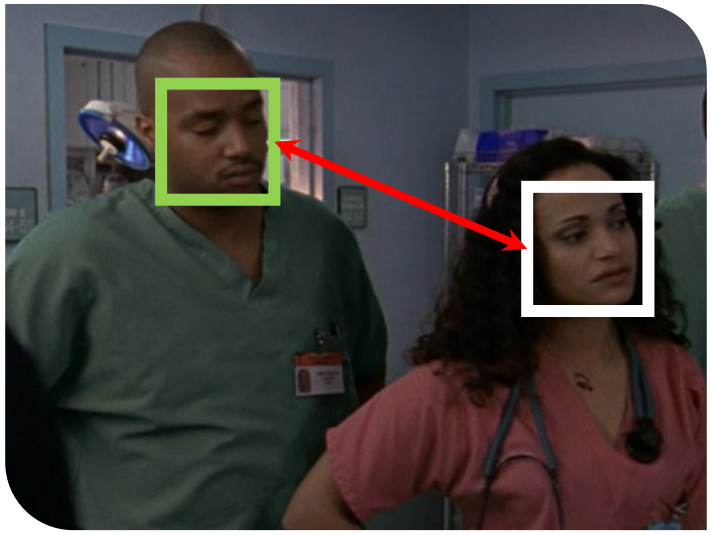
- dense SIFT
- Fisher encoding
- discriminative dimensionality reduction

Negative constraints

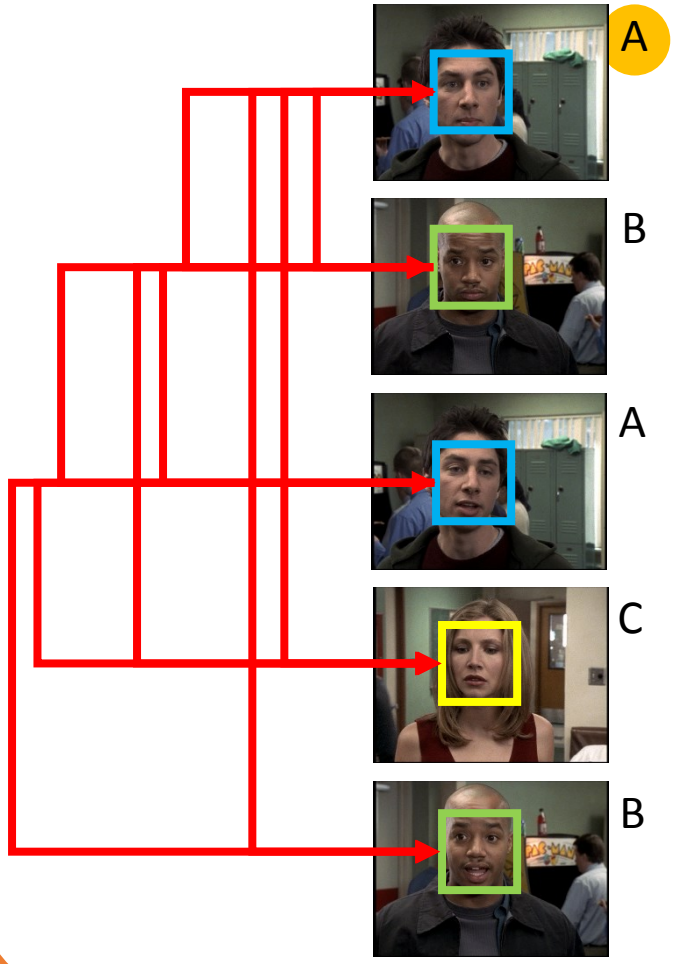
do *NOT* merge these pairs!

Tracks in the same frame

(Cinbis et al. 2011, Wu et al. 2013)

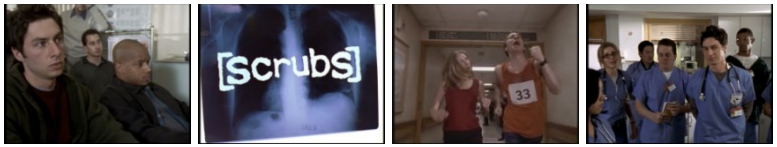


Tracks in a thread

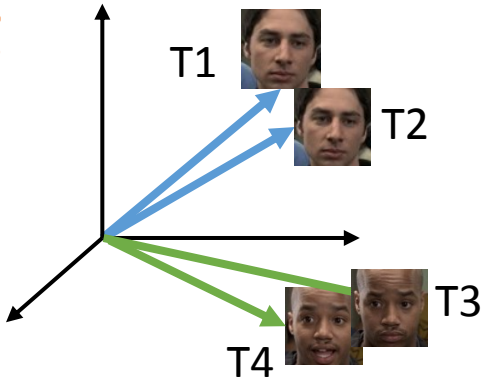


Overview

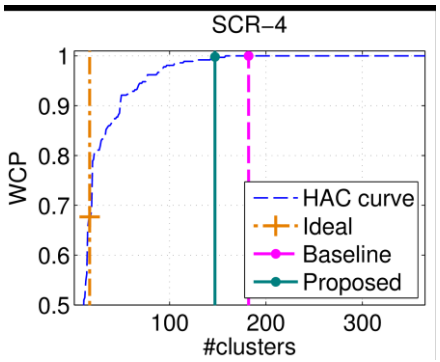
- Editing structure in videos
  - Shots, threads and scenes



- **Using editing structure for clustering**
  - Negative pairs
  - **Scene level clustering**
  - Episode level clustering



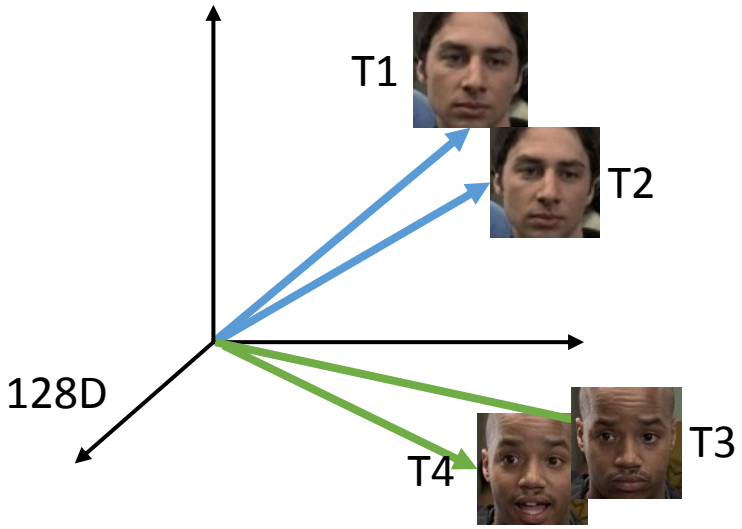
- Dataset and evaluation
  - Scrubs, Buffy
  - Weighted clustering purity
  - Clustering results

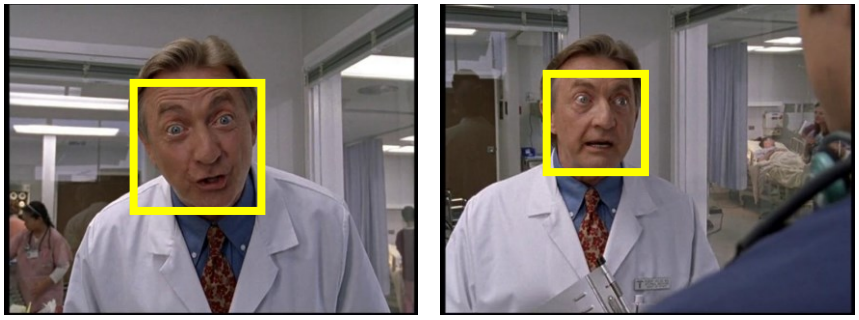


Within scene clustering

for tracks not in negative pairs

Strong face similarity

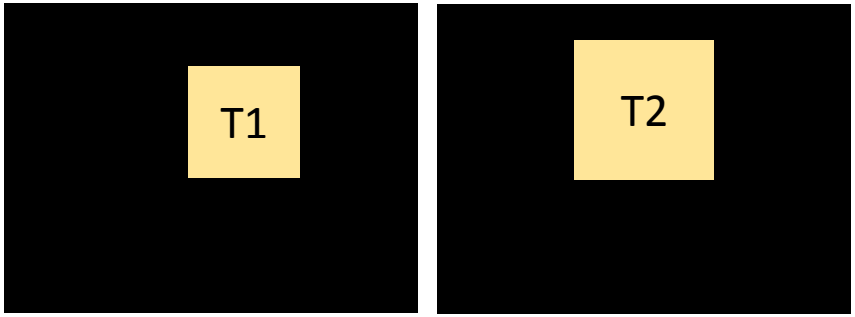




Dr. Kelso

In a thread,

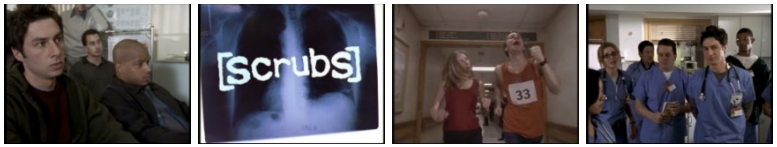
- area overlap
- relaxed face similarity



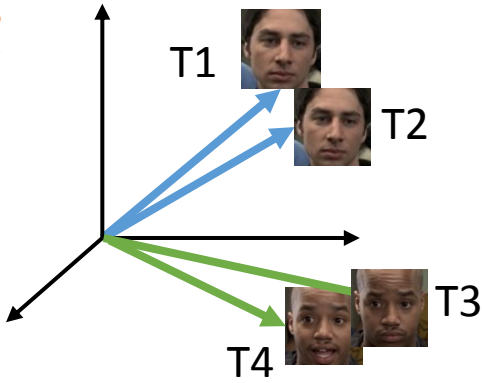
Turk

Overview

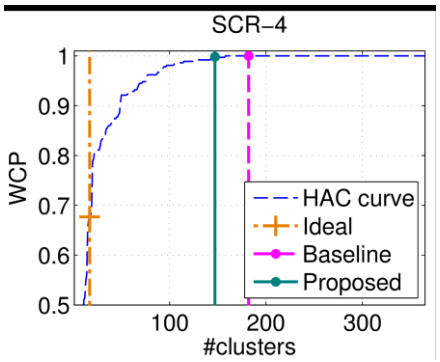
- Editing structure in videos
  - Shots, threads and scenes



- **Using editing structure for clustering**
  - Negative pairs
  - Scene level clustering
  - **Episode level clustering**

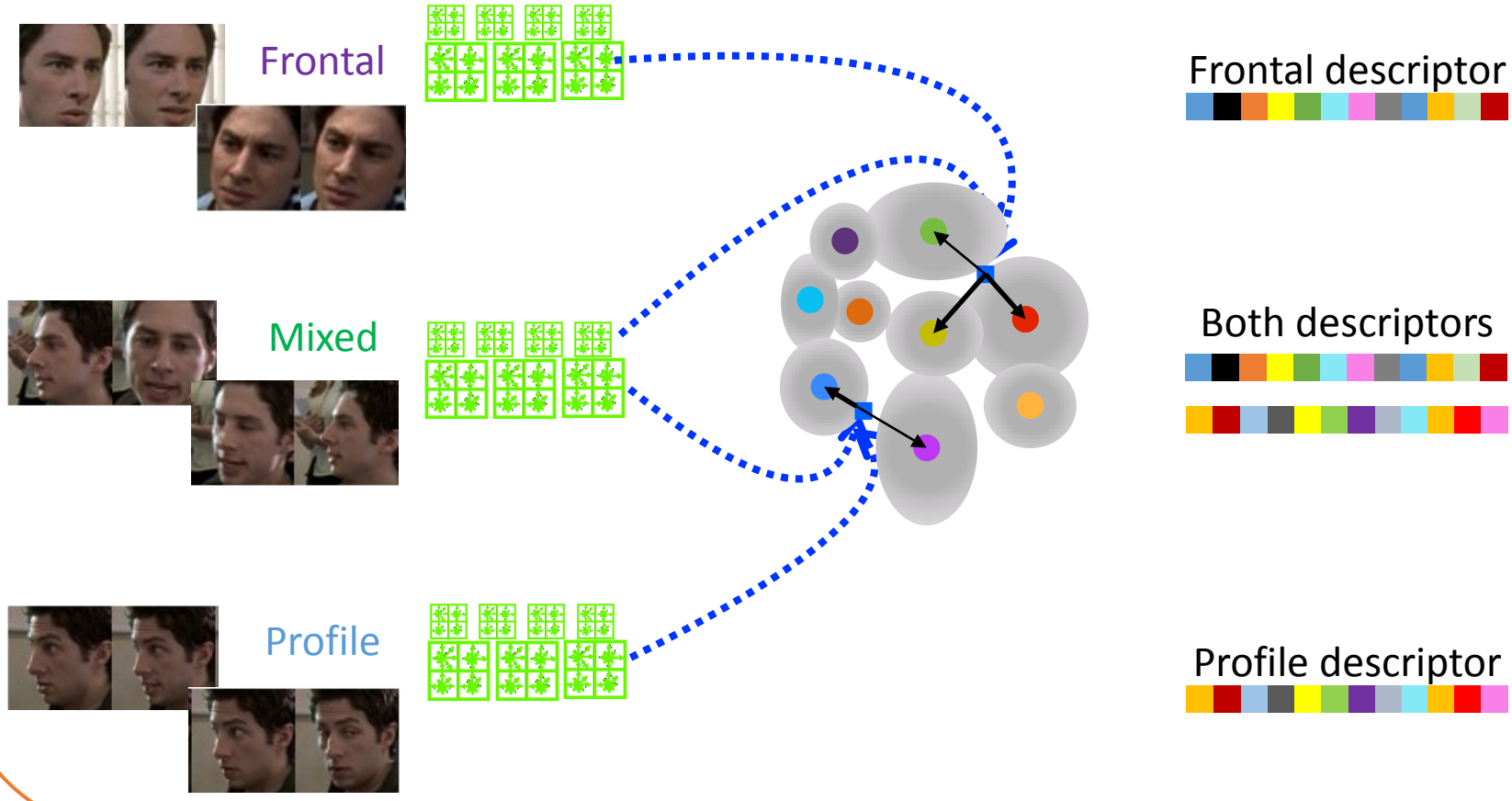


- Dataset and evaluation
  - Scrubs, Buffy
  - Weighted clustering purity
  - Clustering results



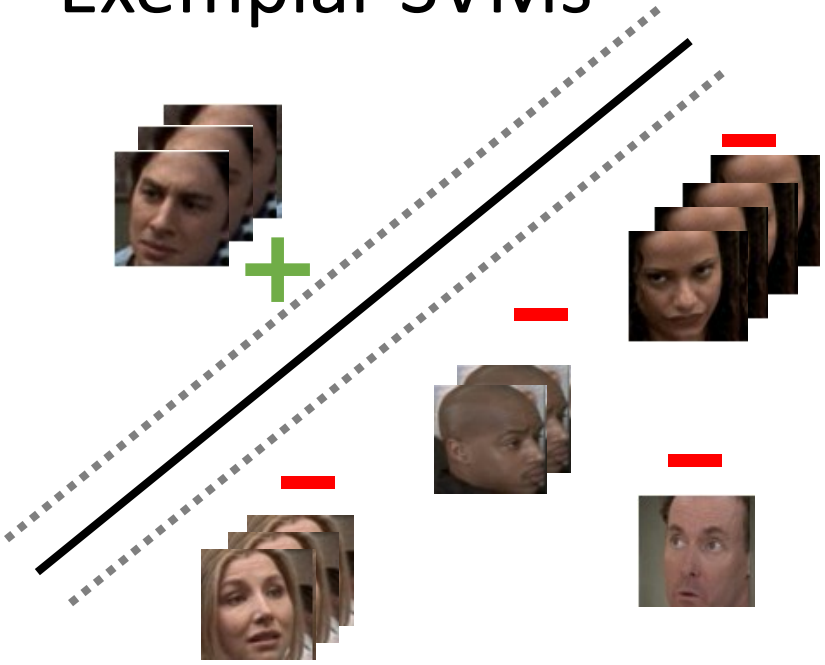
Cluster descriptor

compare frontal vs. frontal, profile vs. profile



Full episode clustering

Exemplar SVMs



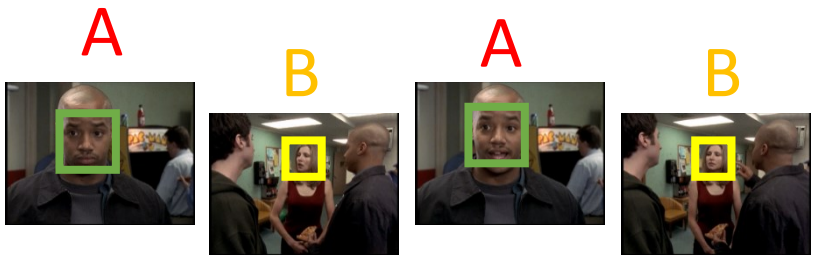
- one positive cluster
- against negative clusters
  + YouTube Faces

Cluster across scenes

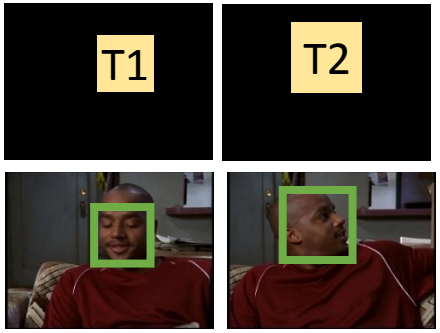
- score clusters with e-SVM
- merge high-scoring

Recap

- Video editing structure

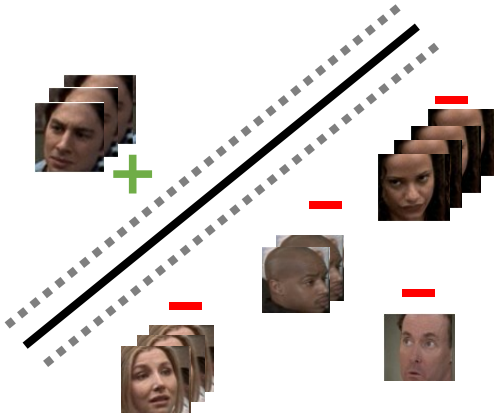


- Do not merge track pairs



Turk

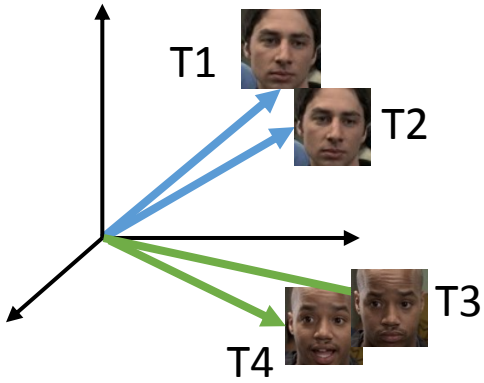
- Scene level clustering
- Episode level clustering



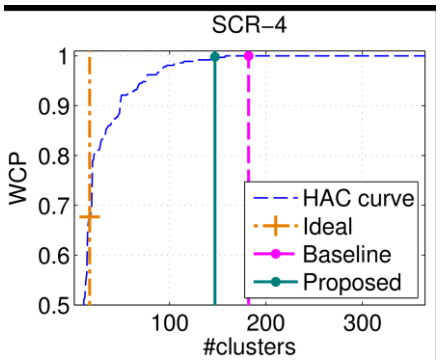
Overview

- Editing structure in videos
  - Shots, threads and scenes

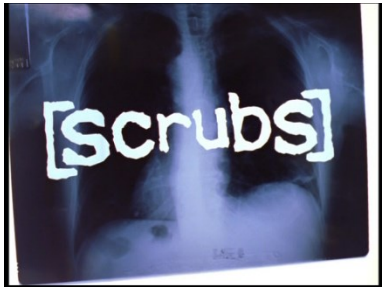
- Using editing structure for clustering
  - Negative pairs
  - Scene level clustering
  - Episode level clustering



- **Dataset and evaluation**
  - Scrubs, Buffy
  - Weighted clustering purity
  - Clustering results



Data set



- episodes 1..5, 23
- ~20 minutes each
- sitcom, interns at a hospital

- episodes 1..6
- ~40 minutes each
- supernatural drama, action

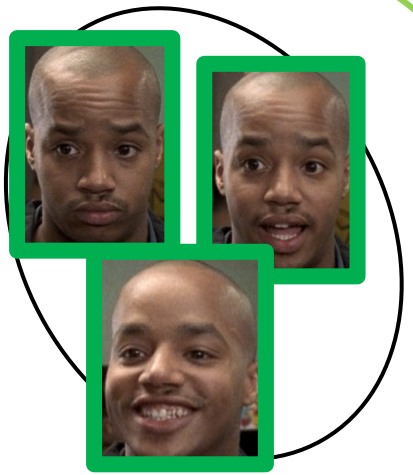
average episode		SCRUBS	BUFFY
# named characters		16	15
# tracks		412	756
# tracks in thread		280	515
# do not merge pairs	in shots	214	440
	in threads	851	1891

Evaluation metrics

Goal: minimize #clusters, don't make errors!

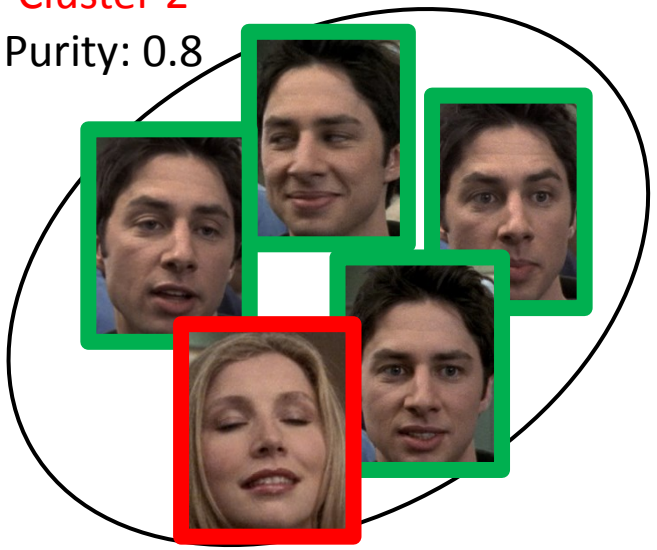
WCP: *weighted clustering purity*

- Cluster 1: Purity $p_1 = 3/3$
- Cluster 2: Purity $p_2 = 4/5$
- WCP: $\frac{1}{N}\sum_c n_c p_c = 0.875$

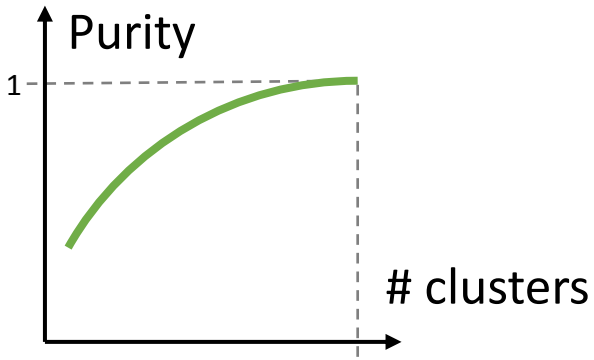


Cluster 1
Purity: 1

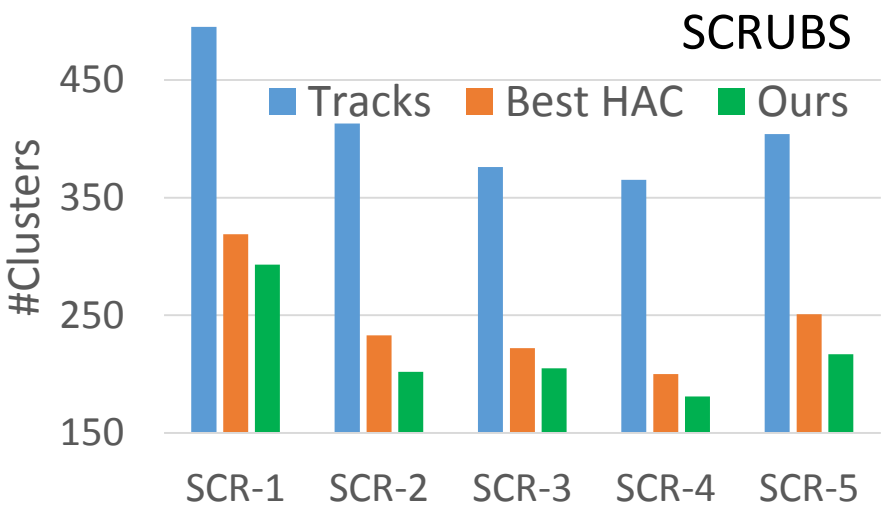
Cluster 2
Purity: 0.8

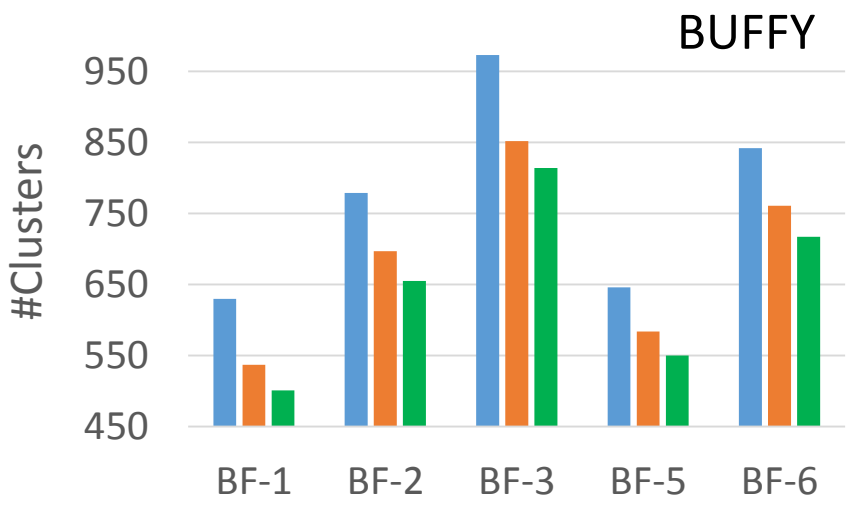


Trade-off WCP vs. #clusters

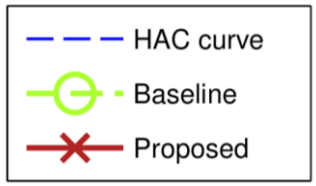


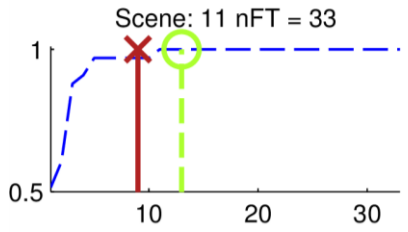
Within scene clustering

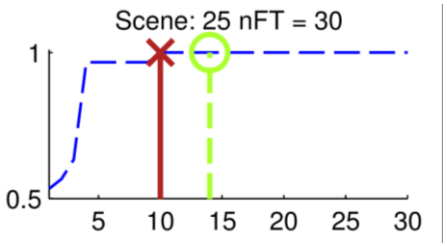




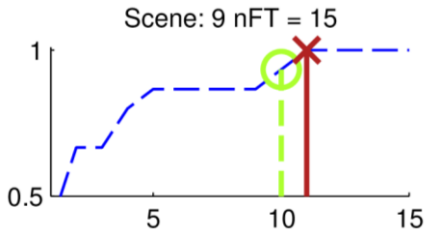
Lower number of clusters @ purity = 1

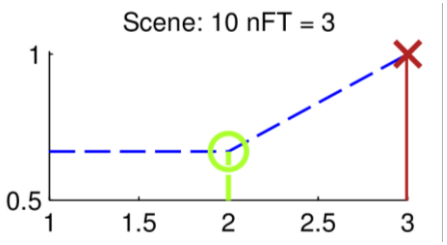




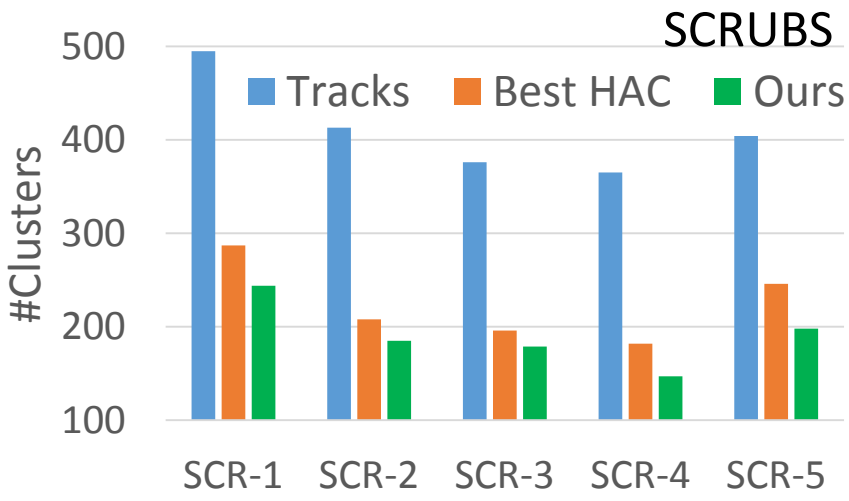


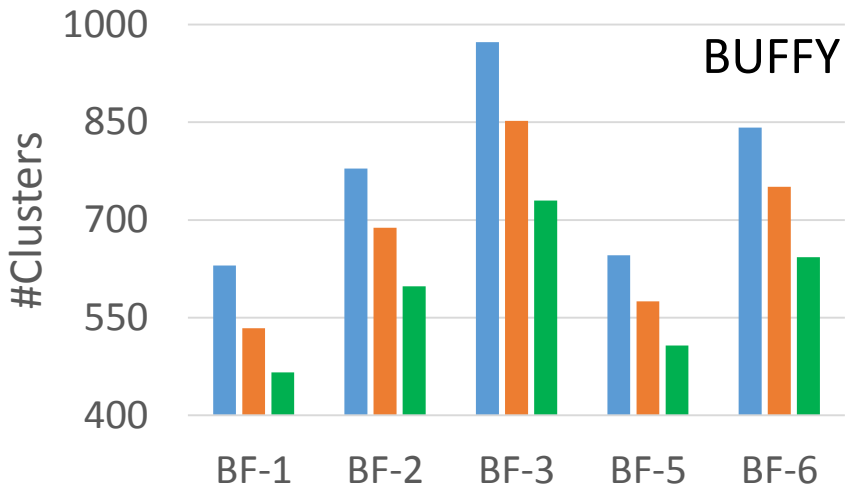
Prevent errors





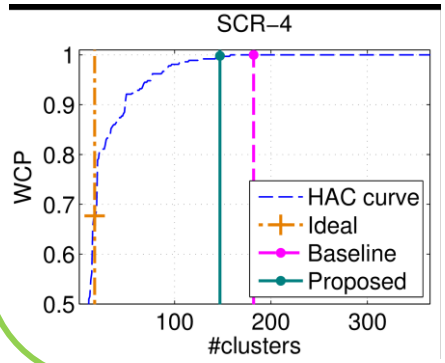
Full episode clustering

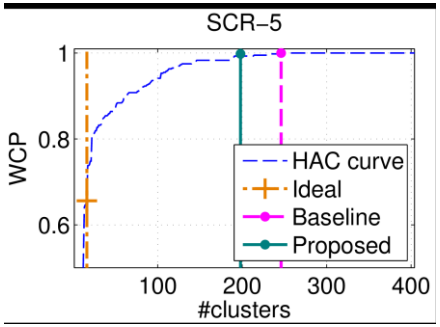


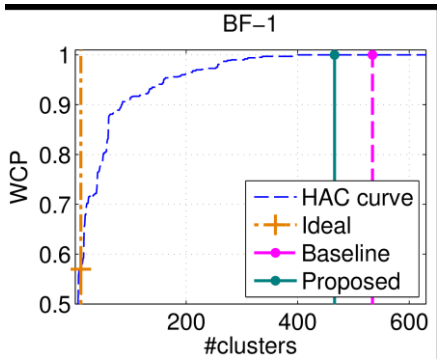


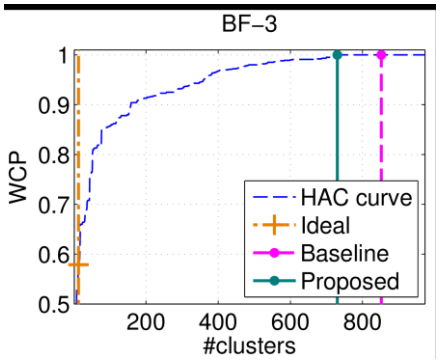
Full episode clustering

- hard to choose threshold for HAC
- proposed method reduces #clusters @ purity → 1

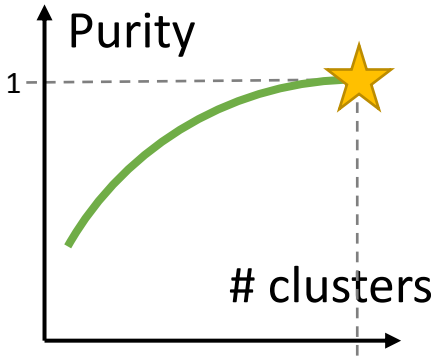




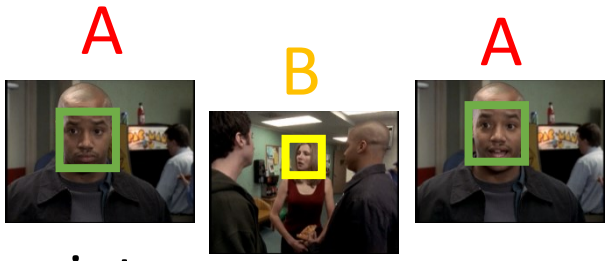




Summary



- Minimize #clusters at very high purity
  - Affords manual and automatic annotation



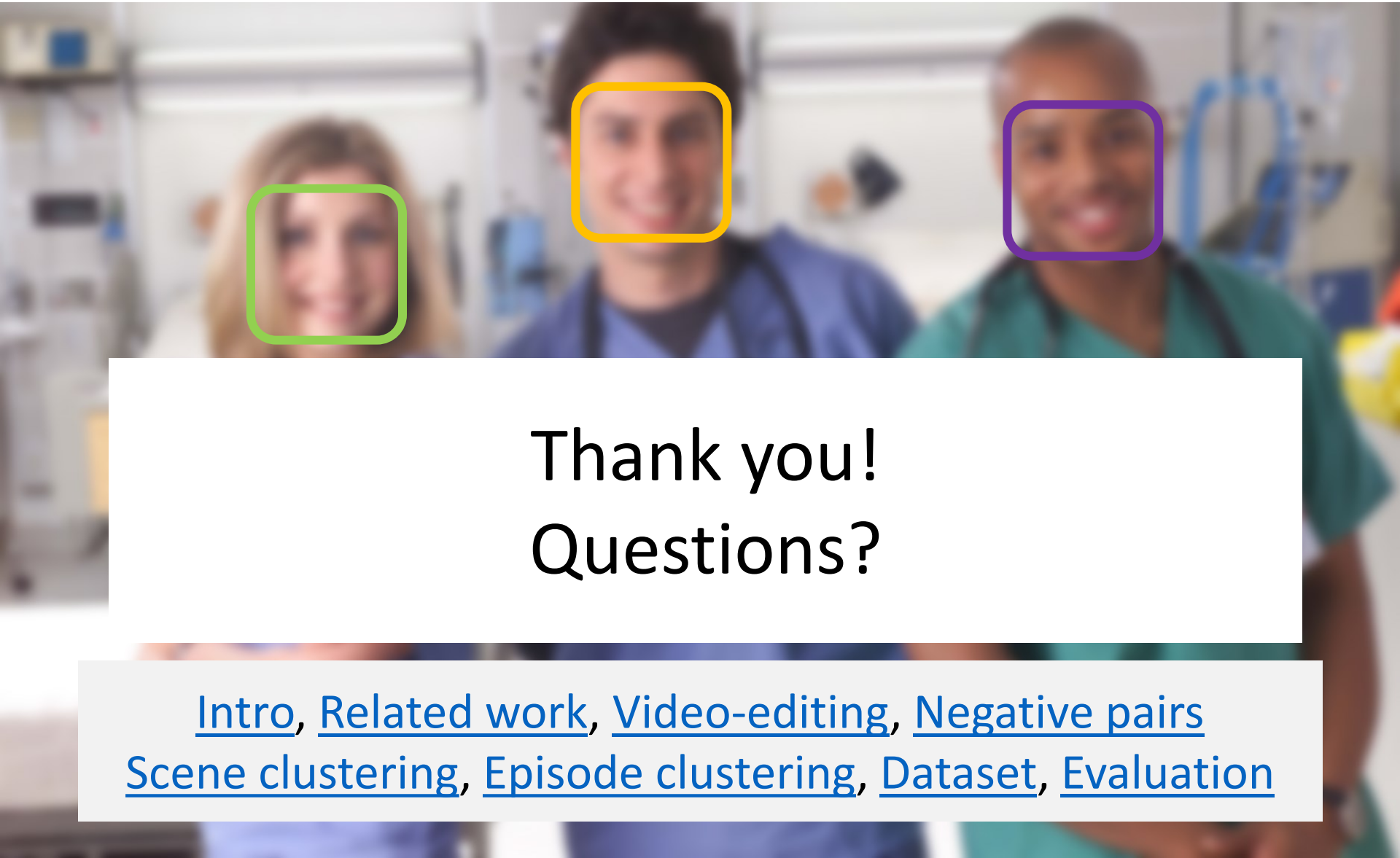
- Leverage video editing structure
  - Face tracks are *not* independent data points
  - 4x number of do not merge track pairs

- Stage-wise clustering
  - Small number of characters in one scene
  - At purity 1, #clusters is about half #tracks

Karlsruhe Institute of Technology

UNIVERSITY of OULU
OULUN YLIOPISTO

UNIVERSITY OF OXFORD



Thank you!
Questions?

Intro, Related work, Video-editing, Negative pairs
Scene clustering, Episode clustering, Dataset, Evaluation