ShARc: Shape and Appearance Recognition for Person Identification In-the-wild
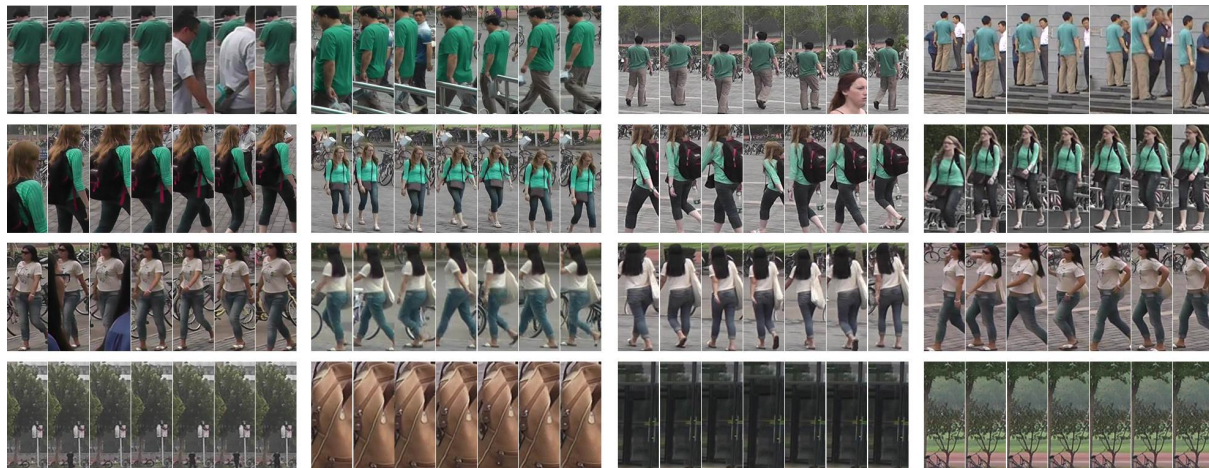
Haidong Zhu, Wanrong Zheng, Zhaoheng Zheng and Ram Nevatia
University of Southern California

Task Definition

- Person identification in-the-wild:
    - Recognize a person's identity with a period of video

Zheng, Liang, et al. "Mars: A video benchmark for large-scale person re-identification." *ECCV*, 2016.

Challenges and Motivation

- Challenges
    - Different activities and outfits
    - A single modality can handle some of the cases, but not all of them
- Motivation
    - We separate the recognition to pose/ shape with appearance, combining strengths from different modalities
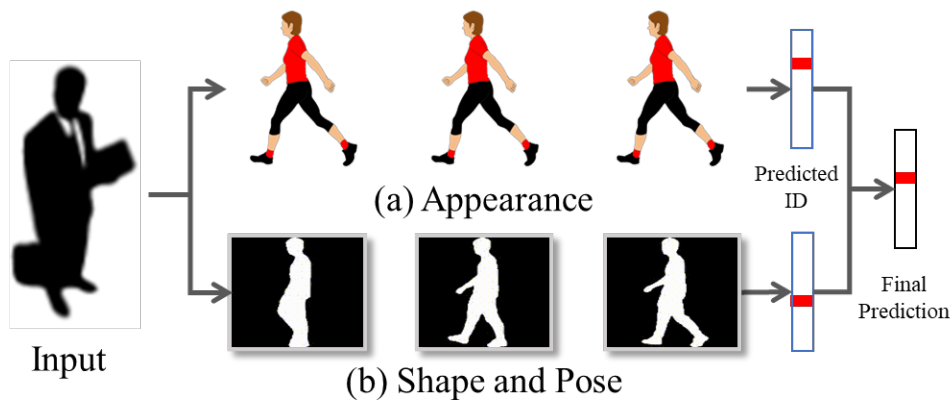

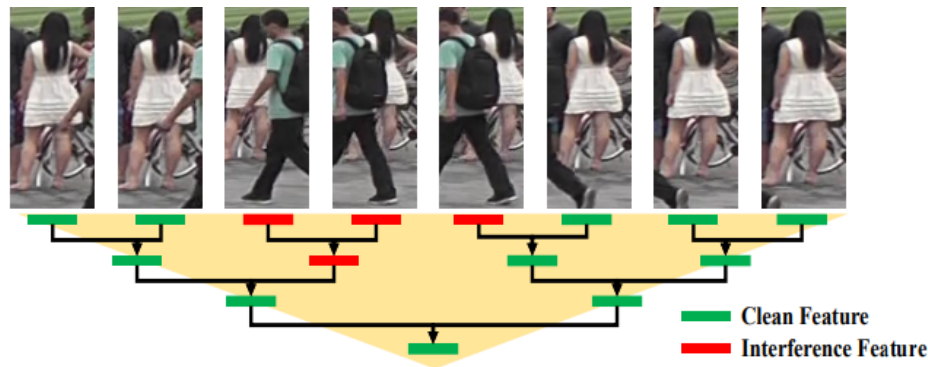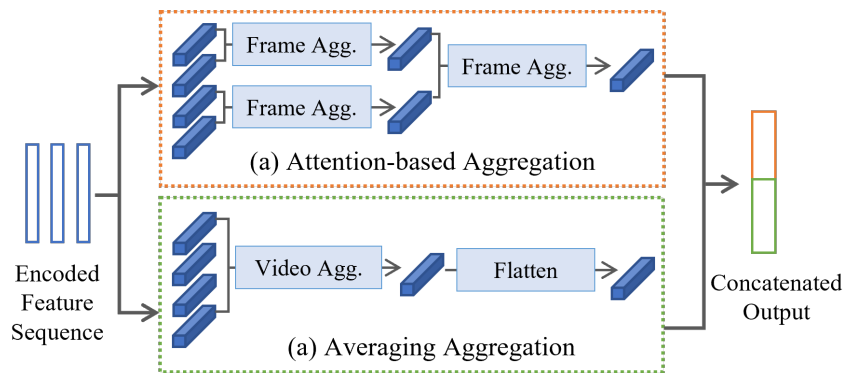| Gallery Frame	Standing Videos	Different Clothing	Turbulence & Occlusion
Gait		✓	✗
Body shape	✓	✗	✓
Appearance	✓	✗	✗
Ours	✓	✓	✓

Method Overview

- Two branch design for shape and appearance recognition
  - Appearance branch - appearance of the frames in the video
  - Shape and pose branch - gait motion and body shape recognition
- Predicted scores are aggregated after two branches

(a) Appearance

(b) Shape and Pose

Input

Predicted ID

Final Prediction

Methods

- Appearance-based recognition (AAE)

  - Attention-based aggregation for one and two consecutive frames [1]

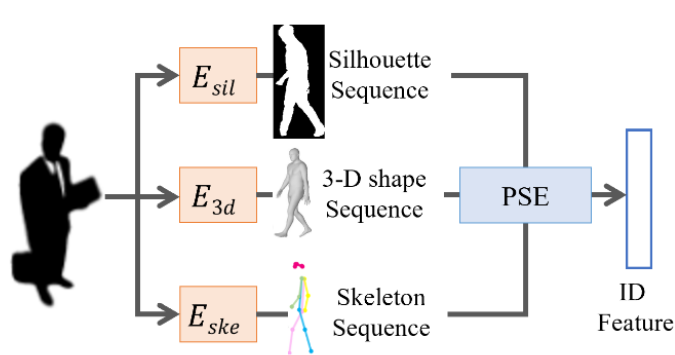 - Averaging of the appearance for all the frames in the video

(a) Attention-based Aggregation

(a) Averaging Aggregation

Encoded Feature Sequence

Frame Agg.

Frame Agg.

Frame Agg.

Video Agg.

Flatten

Concatenated Output

Clean Feature

Interference Feature

[1] Yingquan Wang, et al. "Pyramid spatial-temporal aggregation for video-based person re-identification." In ICCV, pp. 12026-12035. 2021.

USC Viterbi
School of Engineering
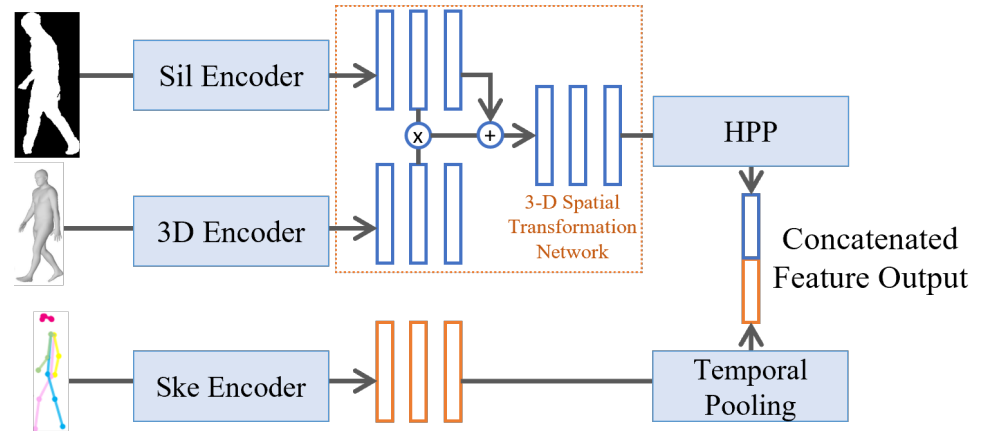
University of Southern California

Methods

- Pose and shape based recognition (PSE)
  - Three encoders for silhouettes, 3-D body shape and skeletons
  - Silhouettes and body shape are framewise combined
  - Skeletons are combined after temporal pooling

(a) Pose and Shape Encoder (PSE)

Experiments and Results

- Datasets:
    - CCVID
        - 75 IDs for training, 151 for inference
    - MEVID
        - 104 IDs for training, 54 for inference
    - BRIAR
        - 407 IDs for training, 642 for inference
- Metrics
    - Accuracy for all three and mAP for MEVID and CCVID

Performance on CCVID

Method	General		Clothes Changes	
	Rank 1	mAP	Rank 1	mAP
GaitNet	62.6	56.5	57.7	49.0
GaitSet	81.9	79.2	71.0	62.1
CAL	82.6	81.3	81.7	79.6
ShARc	**89.8**	**90.2**	**84.7**	**85.2**

Performance on MEVID and BRIAR

MEVID

Method	Rank 1	Rank 20
PSTA	46.2	77.8
ARGL	48.4	77.9
Attn-CL	42.1	73.1
Attn-CL+RR	46.5	71.8
CAL	52.5	80.7
ShARc	**59.5**	**82.9**

BRIAR

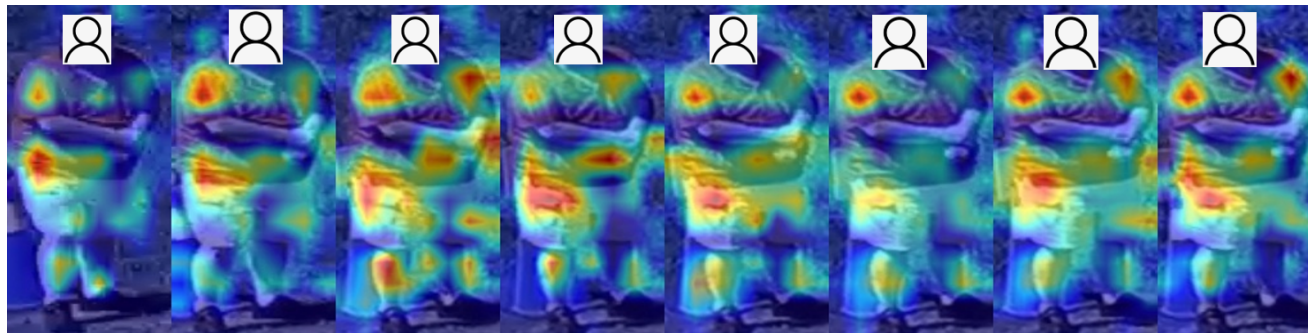| Method	Rank 1	Rank 20
GaitGL	15.6	45.6
GaitRef	17.7	50.2
PSTA	33.6	67.3
Attn-CL+RR	27.6	61.8
CAL	34.9	71.4
ShARc	**41.1**	**83.0**

Attention Visualization

- Visualization of attention produced by the model for different activities

Walking

Standing

Thank you!