PeR-ViS: Person Retrieval in Video Surveillance using Semantic Description

Parshwa Shah Arpit Garg Vandit Gajjar

School of Engineering and Applied Science, Ahmedabad University

School of Computer Science, The University of Adelaide

Motivation

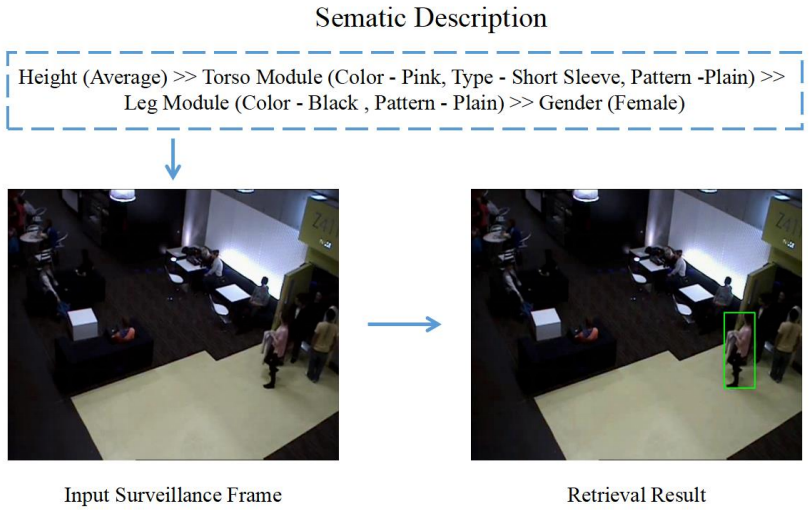
- Given the image query to the network, it finds the similarity between the query image and that surveillance image. The most identical person than retrieved from the surveillance footage according to the similarity score. However, this problem requires at least one image as a query for the network, which has a major limitation in practice.

- In certain cases, such as a lost person, there might be only a description provided of the person(s) appearance – gender, cloth types, etc. Here, the semantic description can be used accurately to describe the person(s) appearance.

Idea

- Instead of using an image query, in our work, we study the problem of person retrieval in video surveillance with a semantic description.

- The core aim is to use Computer Vision to fully automate the person retrieval task in challenging conditions (i.e., Background Clutter, Occlusion, etc.) by using limited semantic descriptors.



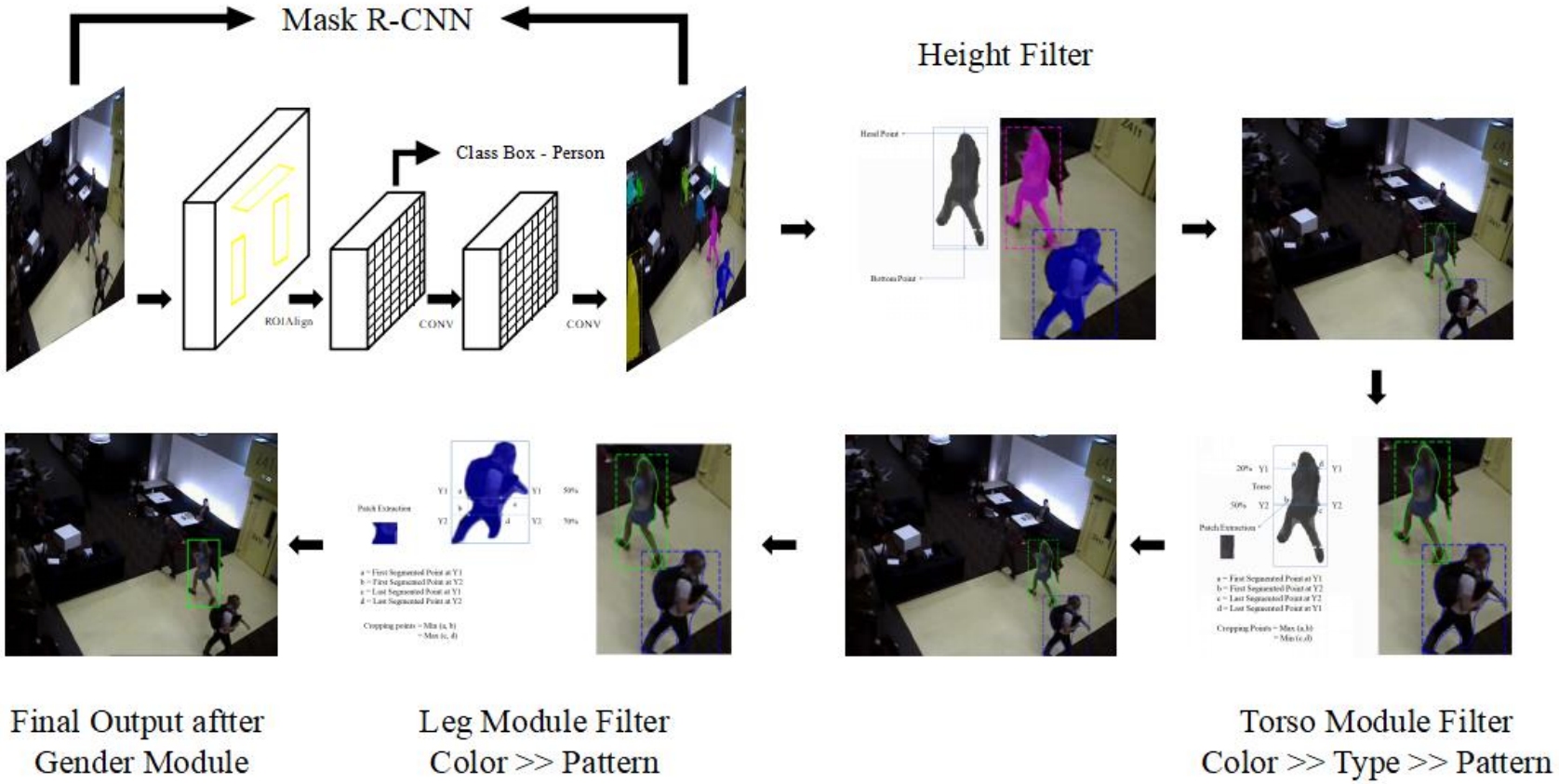
Our approach – PeR-ViS

- We first apply Mask R-CNN to the input image for a person(s) accurate detection and instance segmentation.

- The detected person(s) then fed into the linear filter, which ultimately narrows down the search space and leaves only the target at the end.

- The filtering sequence follows the given order: Height; Torso cloth color, type, pattern; Leg cloth color, pattern; and Gender

Our approach – PeR-ViS



(Best viewed in color and magnification.)

Experimentation

- Accuracy results on the validation set of different semantic descriptors based on the above hyper-parameter setting.

Descriptor	Validation Accuracy
Torso Color	81.29%
Torso Type	79.14%
Torso Pattern	76.5%
Leg Color	71.52%
Leg Pattern	72.5%
Gender	77.79%

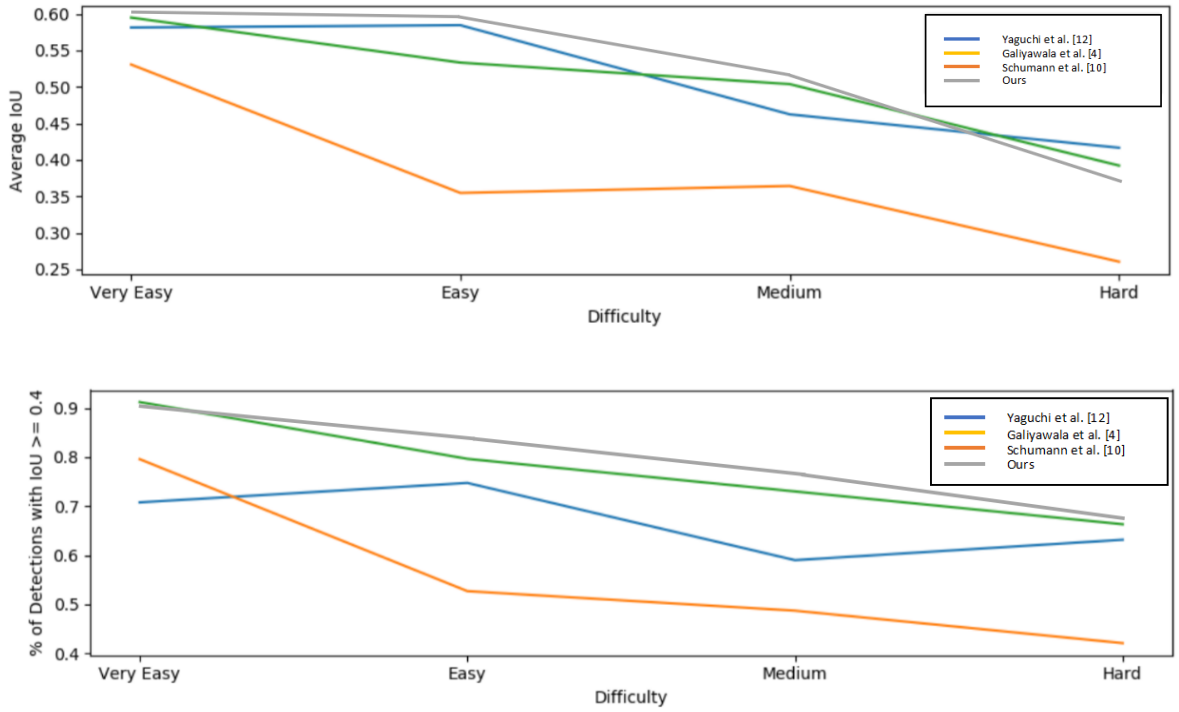
Experimentation

- Overall IoU of different methods on the test set.

Approach	**Average IoU**	**%w** $IoU > 0.4$
Baseline [6]	0.290	0.669
Galiyawala et al. [9]	0.363	0.522
Schumann et al. [26]	0.503	0.759
Yaguchi et al. [35]	0.418	-
Yaguchi et al. [35]	0.462	-
Yaguchi et al. [35]	0.511	0.669
Ours	**0.566**	**0.792**

Experimentation

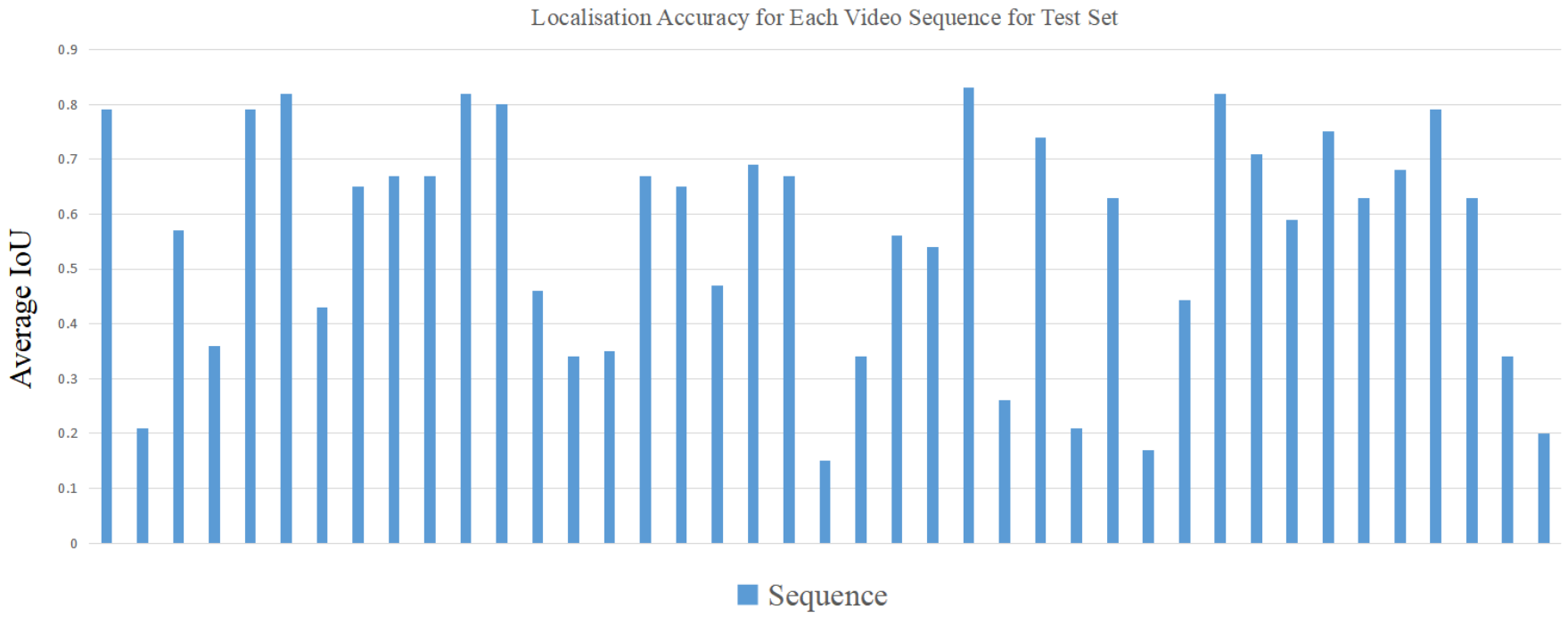
- Performance broken down by sequence difficulty and compared with Average IoU and by sequence difficulty and compared with %w IoU > 0.4.



(Best viewed in color and magnification.)

Experimentation

- Localization accuracy for each video sequence for test set – 41 sequences.



(Best viewed in color and magnification.)

Ablation Study Choice of Classification Network:

- The following table shows the IoU result on 5 video sequences (Sequence Number 4, 13, 21, 23, and 28) using different network architecture in our approach.

	Very Easy	Easy	Easy	Medium	Hard
AlexNet	0.567	0.534	0.523	0.325	0.183
VGG-16	0.641	0.621	0.615	0.336	0.237
ResNet-152	0.742	0.712	0.64	0.492	0.36
DenseNet-161	0.762	0.733	0.642	0.582	0.461

Results

- True positive cases of person retrieval using semantic description

Person retrieved using height and torso module filter

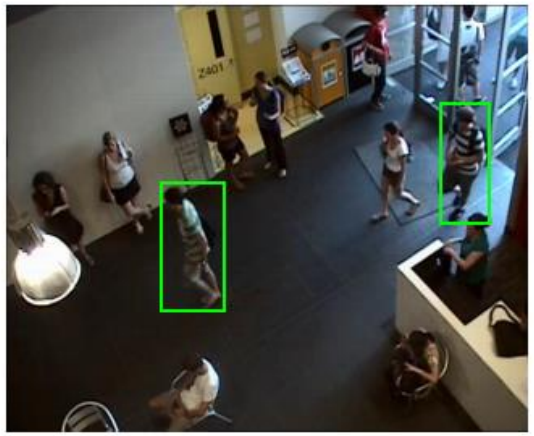
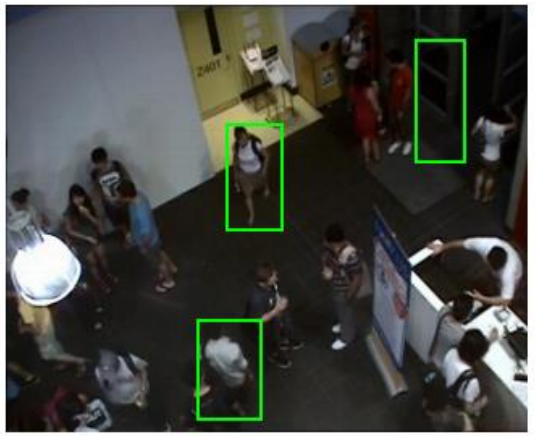
Person retrieved using height, torso, leg and gender module filter

(Best viewed in color and magnification.)

Results

- False negative cases of person retrieval.



(Best viewed in color and magnification.)

Conclusion

- The major benefit of this filtering sequence is that Height, and Torso Color, Type, and Patterns are easily differentiable compared to other descriptors, where the heavy crowd is present.

- Instance segmentation allows precise height estimation and accurate color patch extraction from the torso and leg.

- The possible future work will now focus on how to improve the results by incorporating human pose estimation and other descriptors and investigate the architecture for generalization in more real-world scenarios.

PeR-ViS: Person Retrieval in Video Surveillance using Semantic Description

Parshwa Shah, Arpit Garg, Vandit Gajjar

https://parshwa1999.github.io/PeR-ViS/