Semantic Person Retrieval in Surveillance Using Soft Biometrics

Michael Halstead, Simon Denman, Clinton Fookes, YingLi Tian, Mark Nixon

November 27th, 2018

QUT a university for the real world
UNIVERSITY OF Southampton
The City College of New York

Table of contents

Introduction

In Brief:

- Unique challenge proposal - first of it's kind.
- Task 1 related to person re-id, Task 2 more of a realistic surveillance task.
- Five participants across two tasks:
  - Low participation but promising for the first attempt;
  - Four participants in Task 1;
  - Three participants in Task 2.

Task 1 Performance

- Participants had access to the entire 520 subjects for training, including:
  - RGB images;
  - Parsed images;
  - Semantic query.
- Evaluation was performed on 196 images:
  - RGB images only (no parsed images);
  - Blind semantic queries (image name witheld).
- No restrictions placed on methodologies.

Task 1 Performance

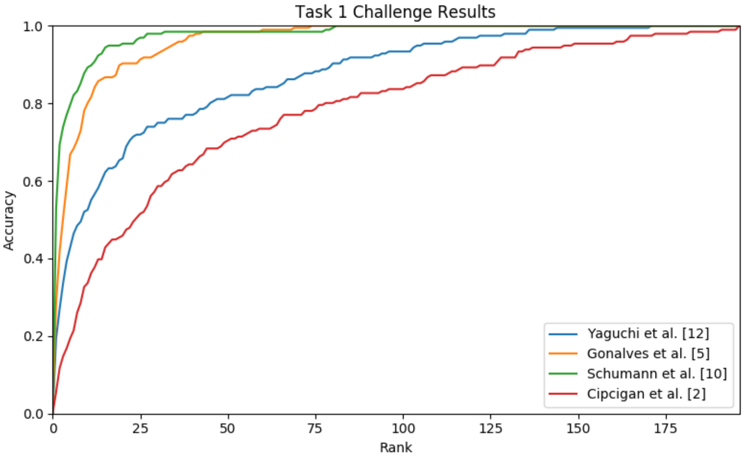
- Unsurprisingly all techniques based around deep learning.
- Cipcigan et. al:
  - Used semantic segmentation to locate body parts;
  - Hand selected features and classifiers used on appropriate regions;
  - Only technique to not be solely deep learning.
- Gonalves et. al, Schumann et. al, and Yaguchi et. al:
  - Similar pipeline;
  - Trained DCNN (four for Schumman) adapted to attribute prediction.

Task 1 Performance

- Pedestrian classification:
  - Gonalves and Schumann pass images directly to DCNN to predict attributes;
  - Schumann uses an ensemble which predicts better than any single network;
  - Yaguchi first segment pedestrian from bg:
    - Local and global classifiers with results averaged.
- Gonalves, Schumann, and Yaguchi classify attributes jointly.
  - Yaguchi uses Hamming distance;
  - Gonalves and Schumann investigate distances:
    - Gonalves - productory distance;
    - Schumann - euclidean;
    - Investigation of distances shows that delaying hard decision enforced by Hamming is beneficial.

Task 1 Performance

The results of Task 1 - CMC curve and rank performance.



Task 1 Performance

Approach	Rank 1	Rank 5	Rank 10	Rank 25
Yaguchi [12]	0.194	0.429	0.526	0.719
Gonalves [5]	0.286	0.668	0.801	0.913
Schumann [10]	**0.531**	**0.796**	**0.893**	**0.969**
Cipcigan [2]	0.056	0.194	0.337	0.515

Table: Summary of Performance for Task 1

Task 1 Performance

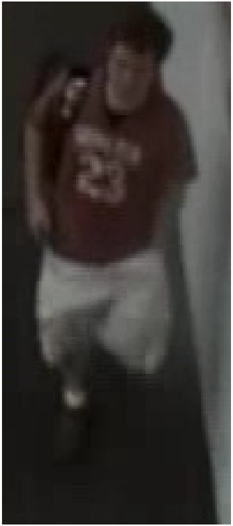
Overall performance was very good.

- Impressively at Rank 25 all participants $> 50\%$.
- Schumann and Gonalves $> 90\%$.
- Some subjects proved difficult for all, average rank of > 60.

Task 1 Performance: Difficult Subjects

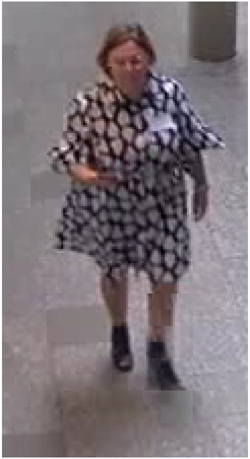
- Subject 20, average rank 72.
- Query: Male, long sleeve black, white and grey checked shirt, long black and grey irregular patterned pants, no luggage.
- Erroneous or ambiguous annotations.
  - No luggage?
- Ambiguity in clothing colour.

Task 1 Performance: Difficult Subjects



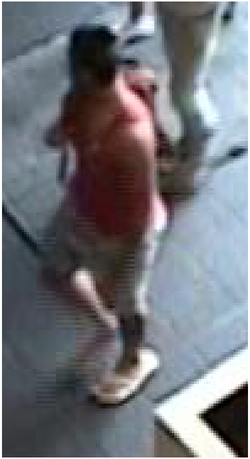
- Subject 111, average rank 66.
- Query: Male, short sleeved brown and white shirt, grey shorts, with luggage.
- Poor lighting conditions.
- Ambiguity in clothing colour.

Task 1 Performance: Difficult Subjects



- Subject 125, average rank 60.
- Query: Female, short sleeved while and brown irregular patterned dress, with no luggage.
- Ambiguity in clothing texture.

Task 1 Performance: Difficult Subjects



- Subject 138, average rank 66.
- Query: Male, long sleeved pink shirt, grey and white shorts, no luggage.
- Ambiguity in clothing annotations.
  - No luggage?
  - No sleeves or long sleeves?
- Illumination.

Task 1 Winner

CONGRATULATIONS

Winner of Task 1

A. Schumann and A. Specker.
Attribute-based person retrieval and search in video sequences.

Task 2 Performance

- Training set contained 110 subjects in video.
  - Tsai [2] method of camera callibration.
  - Full soft biometric signature and key body locations.
- Evaluation set contained 41 subjects in video.
  - Camera callibrations.
  - Full soft biometric signature, no included body locations.
- No restrictions placed on methodologies.

Task 2 Performance

Overview of techniques:

- Schumann et. al and Yaguchi et. al build on Task 1 techniques.
- Both deploy a DCNN to detect pedestrians.
- Schumann incorporate pedestrian tracking.
- Galiyawala et. al deploy a DCNN for pedestrian detection.
  - Cascade of classifiers.
  - Whittles down pedestrians.
  - Use a subset of traits.

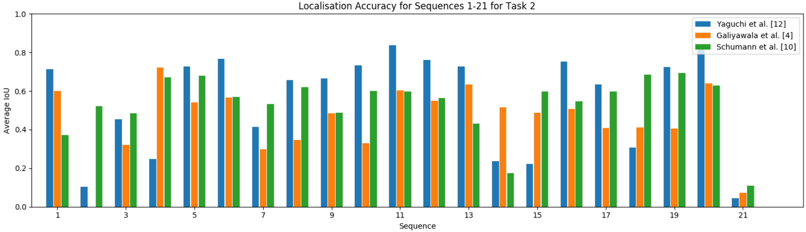
Task 2 Performance

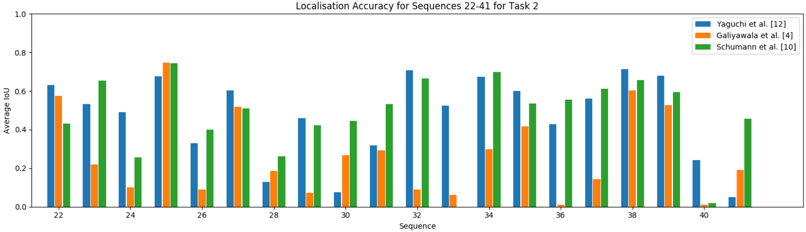
Results based on Intersection over Union.

Approach	Average IoU	% w IoU > 0.4
Yaguchi [12]	**0.511**	0.669
Galiyawala [4]	0.363	0.522
Schumann [10]	0.503	**0.759**
Baseline [1]	0.290	0.493

Task 2 Performance

All sequence performance.





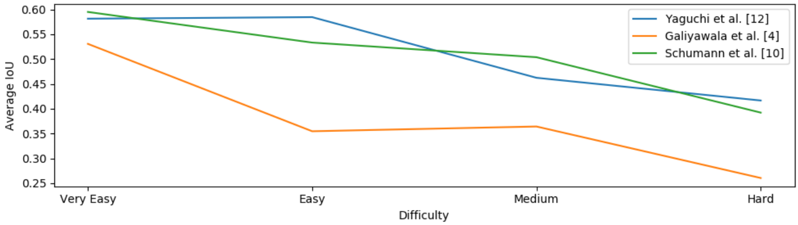
Task 2 Performance

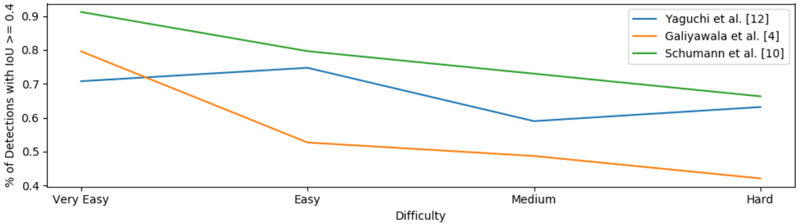
Scenes were split into difficulty levels.

- Complicating factors considered
  - similar subjects (i.e. partial match to query) present
  - occlusion of target
  - heavy crowding
- Very Easy: No complications, sparse scenes, subject clearly visible
- Easy: No complications, scene contains multiple people, but the target is clearly distinct
- Medium: One complicating factor
- Hard: Two (or more) complicating factors

Task 2 Performance

Performance based on difficulty of scene.





Task 2 Performance

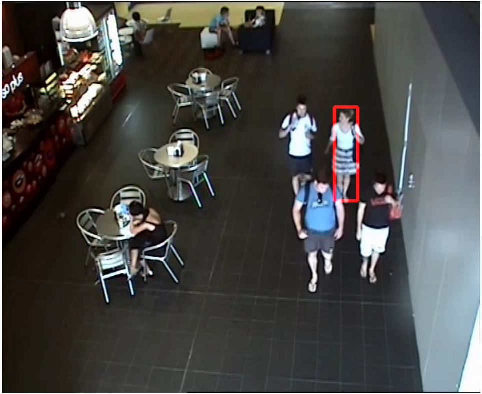
Observations

- Performance decreases with difficulty.
- All techniques perform well on "Very Easy".
- Schumann and Yaguchi similar IoUs.
- Detection > 0.4 Schumann performs best.
  - Indicates possible tracking advantages - smoothing.

Task 2 Performance

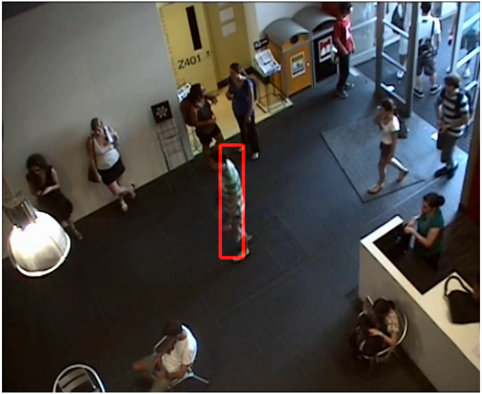
Difficult subjects

- Sequence 20 (Hard).
- Limited frames with full appearance (occlusion).
- Similar appearance in other pedestrians.



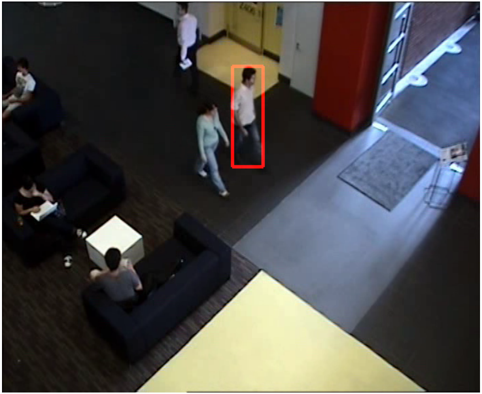
Task 2 Performance

- Sequence 27 (Hard).
- High level of crowding.
- Illumination issues.
- Subjects with similar appearance (striped shirt).



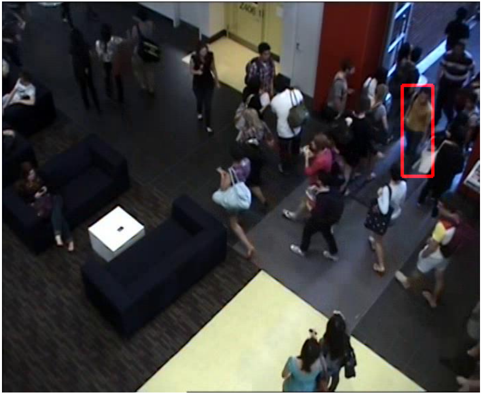
Task 2 Performance

- Sequence 39 (Easy).
- Complications due to high appearance similarity.



Task 2 Performance

- Sequence 40 (Hard).
- Crowding.



Task 2 Winner

CONGRATULATIONS

Winner of Task 2

T. Yaguchi and M. Nixon. Transfer learning based approach for semantic person retrieval.

Overall Observations

- Deep learning wins again.
  - Semantic segmentation, which seems intuitively useful, did not help performance.
  - Creates issues for explainable results. Do we know what the models are learning? Could results be used as evidence?
- Some annotation was not used the way we intended:
  - Pose for Task 1 was included for completeness, but perhaps simplified the task too much.

Overall Observations

- Annotation is hard!
- Ambiguous annotations pose a challenge:
  - Annotating consistently is difficult, ideally we need multiple annotators, from different background to counter bias
  - Annotation errors impact system performance too
  - Could uncertainty/confidence be captured in the labels somehow?
- Comparative labels and ranking would likely lead to further gains:
  - i.e. 'Taller than' or 'shorter than' rather than just a height value
  - Annotation difficulties persist however

Thank You to the Participants

- Emil Barbuta Cipcigan
- Luiz Eduardo Lima Coelho
- Matheus Diniz
- Vandit J. Gajjar
- Hiren Galiyawala
- Gabriel R. Gonalves
- Antonio C. Nazare
- Mark Nixon
- Mehul Raval
- Arne Schumann
- William R. Schwartz
- Kenil S. Shah
- Andreas Specker
- Takuya Yaguchi

Final Notes

- Both datasets publicly available for research purposes:
  - `https://data.researchdatafinder.qut.edu.au/dataset/saivt-semantic-person1`
- Metrics, evaluation tools and test set ground truth available on GitHub:
  - `https://github.com/simondenman/SemanticSearchChallengeAVSS18`
- Spread the word on semantic search - bridge the gap.

Bibliography I

S. Denman, M. Halstead, C. Fookes, and S. Sridharan.
Searching for people using semantic soft biometric descriptions.
Pattern Recognition Letters, 68:306–315, 2015.

R. Y. Tsai.
An efficient and accurate camera calibration technique for 3d machine vision.
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 364–374, 1986.