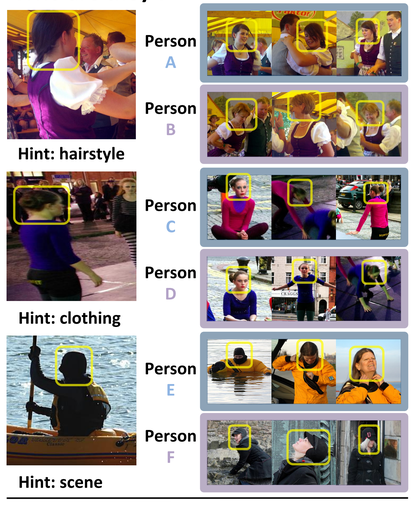
Person Recognition in Personal Photo Collections

Seong Joon Oh, Rodrigo Benenson, Mario Fritz, Bernt Schiele
Max Planck Institute for Informatics, Saarbrücken, Germany

SANTIAGO, CHILE DECEMBER 11-18, 2015 ICCV15 International Conference on Computer Vision

goo.gl/DKuhlY

Who are they ?



Problem & Contributions

Problem

Benchmark	Pose	Time	Input
Face recognition[2]	Frontal	Years	Head
Re-identification[3]	Upright	Hours	Full body
Person recognition[1]	**Diverse**	**Years**	**Full image**

Contributions

1. New state of the art person recogniser.
2. Analysis of different cues.
3. New challenging setup.

Dataset

Person In Photo Albums (PIPA) [1]

- 37,107 Flickr images (CC licensed).
- Head box + identity annotations.

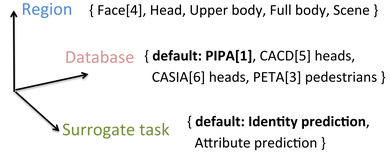
Evaluation protocol [1]

- How well can you recognise a person, having seen ~10 training examples per person?

Approach

naeil: state of the art person recogniser

- Crop five different regions.
- Prepare a CNN pretrained on ImageNet.
- Finetune the network on **PIPA** with **identity** labels.
- Alternatively, finetune on a different database with a different surrogate task.

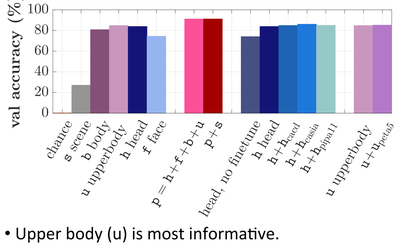


Recognition result

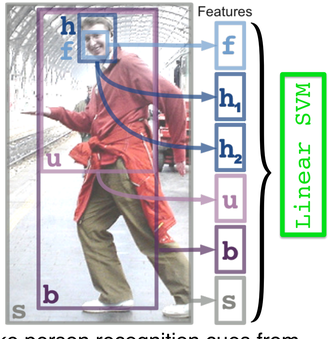
Method	Accuracy
Chance level	0.17 %
h_{rgb}	33.77 %
PIPER[1]	83.05 %
h+b	83.36 %
naeil (17 cues)	**86.78** %

- h_{rgb} (RGB as feature) works.
- State of the art result.
- Compared to PIPER[1], no pose estimation or specialised face feature is used.

Analysis of cues



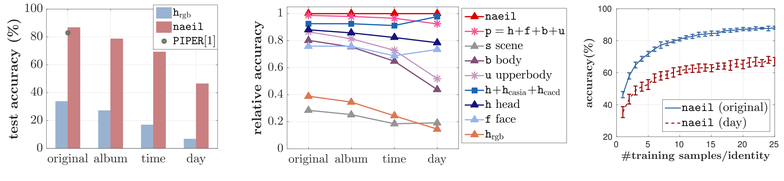
- Upper body (u) is most informative.
- Finetuning gives ~10 pp gain.
- More data helps (h_{cacd}, h_{casia}).
- Attribute prediction task helps (h_{pipa11}, h_{peta5}).



Take person recognition cues from different regions of image.

New challenging experimental setups for recognising a person across time.

New splits & Analysis



New splits proposed

Code & new splits at goo.gl/DKuhlY

Split	Description
Original	• As proposed by [1].
Album	• Separate based on photo albums to which they were uploaded.
Time	• Use photo-taken-date EXIF metadata to separate examples.
Day	• Manually separate instances.

- Motivation: RGB feature gives 33.77% → Many nearly identical training and test examples.

Do we need more examples per person?

- No. Performance nearly saturates after ~20 training examples.
- Better features are needed.

Which cues are helpful across time?

- Face and head cues are effective across time.
- Body and scene cues become weaker.
- Head cue with additional data (h_{cacd}, h_{casia}) covers almost all of the performance by naeil in the day split.

References

[1] Zhang et al. Beyond frontal faces: improving person recognition using multiple cues. CVPR'15.
[2] Huang et al. Labeled faces in the wild, a database for studying face recognition in unconstrained environments. UMASS Tech report 2007.
[3] Deng et al. Pedestrian attribute recognition at far distance. ACM'14.
[4] Mathias et al. Face detection without bells and whistles. ECCV'14.
[5] Chen et al. Cross-age reference coding for age-invariant face recognition and retrieval. ECCV'14.
[6] Yi et al. Learning face representation from scratch. arXiv'14.

Qualitative results

Original split

Day split