MULTI-POSE FACE RECOGNITION FOR PERSON RETRIEVAL IN CAMERA NETWORKS[1]

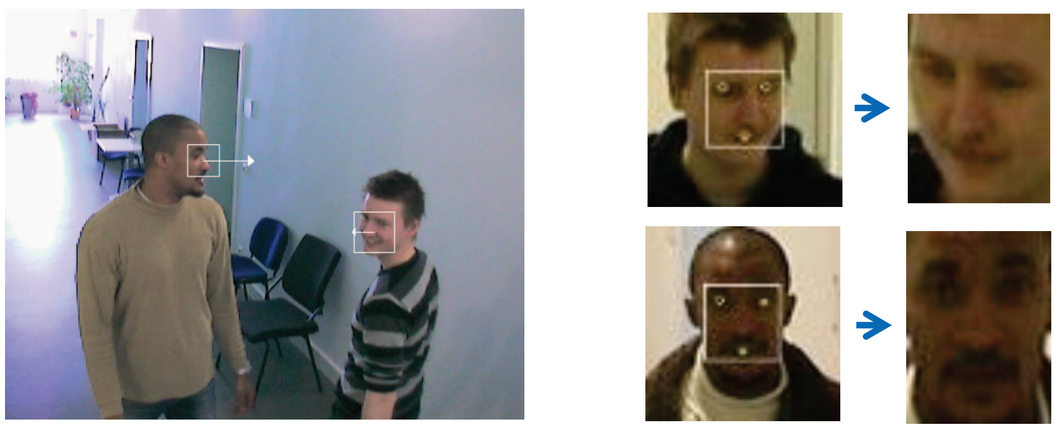
Bäuml M., Bernardin K., Ekenel H., Stiefelhagen R.

Institute for Anthropomatics, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

contact: baeuml@kit.edu http://cvhci.anthropomatik.kit.edu

Abstract We study the use of facial appearance features for the re-identification of persons using distributed camera networks in a realistic surveillance scenario. In contrast to features commonly used for person re-identification, such as whole body appearance, facial features offer the advantage of remaining stable over much larger intervals of time. The challenge in using faces for such applications, apart from low captured face resolutions, is that their appearance across camera sightings is largely influenced by lighting and viewing pose. Here, a number of techniques to address these problems are presented and evaluated on a database of surveillance-type recordings. A system for online capture and interactive retrieval is presented that allows to search for sightings of particular persons in the video database. Evaluation results are presented on surveillance data recorded with four cameras over several days.

Face Tracking and Alignment



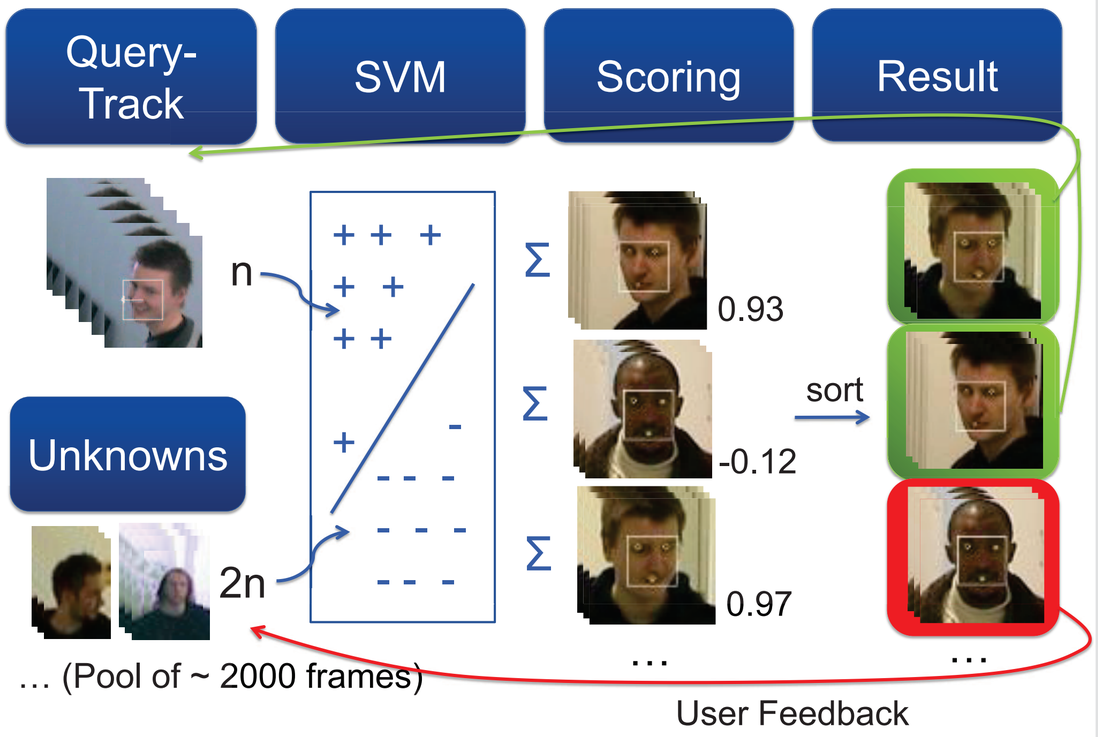
We perform detector-based face tracking to extract face images from videos and connect them to longer tracks. Multiple detectors for different yaw angles are combined using a particle filter to track faces and head-pose robustly over a large range image conditions. We explicitly include the head-pose in the particles' state:

$$\boldsymbol{x} = \{x, y, s, \alpha\}$$

For each frame, we detect eyes and mouth, and also use the head pose information from the tracker to warp the face to a frontal, normalized pose.

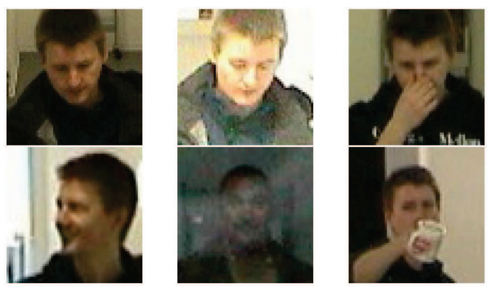
Face Retrieval

For a given query track, we train a support vector machine (SVM) using local DCT-based facial features [2]. We compute a matching score for each track by fusing the individual frame scores. User feedback can be easily incorporated.



Track score: $$s_{track} = \frac{1}{N}\sum_{i=1}^{N} d_{SVM}(\boldsymbol{x_i})$$

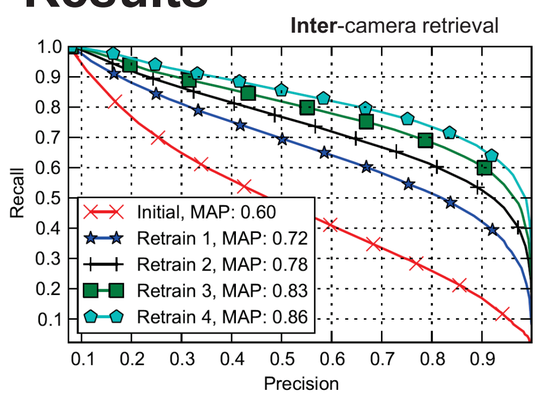
Dataset

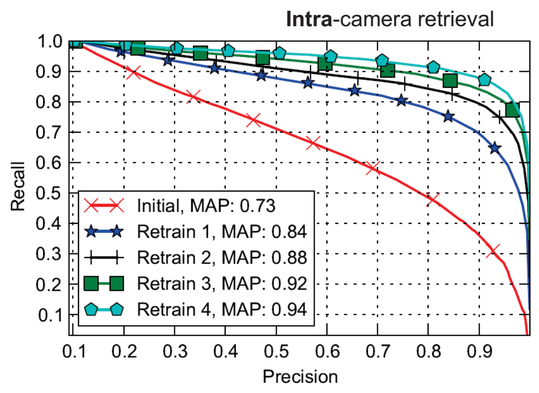


	# tracks	# persons	# persons with ≥ 10 tracks (# total tracks)
Cam 1	67	10	1 (24)
Cam 2	492	70	10 (348)
Cam 3	139	28	3 (49)
Cam 4	99	23	5 (64)
Total	797	92	18 (604)

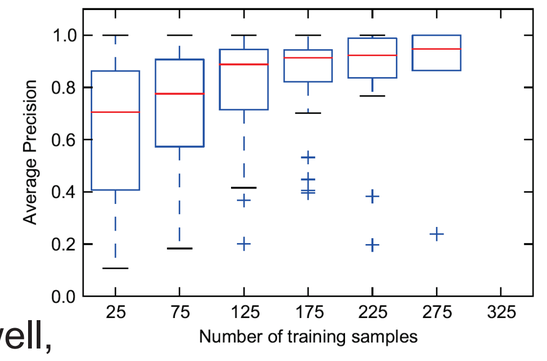
Our evaluation dataset consists of a total of 797 tracks with 92 different persons recorded in our lab. The dataset contains many challenging conditions, including bad lighting, low resolution, occlusions and non-frontal poses.

Results





We evaluate our approach on the tasks of inter- and intra-camera person retrieval. We report results in terms of Precision/Recall and Mean Average Precision. Our approach can leverage additional training data very well, underlining the benefits of a robust tracker and the inclusions of user feedback.



References

[1] M. Bäuml, K. Bernardin, M. Fischer, H. Ekenel and R. Stiefelhagen. *Multi-Pose Face Recognition for Person Retrieval in Camera Networks.* AVSS 2010, Boston

[2] H. Ekenel and R. Stiefelhagen. *Analysis of Local Appearance-Based Face Recognition: Effects of Feature Selection and Feature Normalization*. CVPR Workshop 2006