There's nothing like (more) data

Tatt-C Workshop at NIST

Daniel Manger, Fraunhofer IOSB, Karlsruhe, Germany June 8, 2015

- » IOSB Background: EU-Project
- » Submission details
- » Research todo list
- » Towards the dream dataset

(b)(4)
(b)(6)

Fraunhofer
IOSB

Image Retrieval at IOSB - Background

(b)(4)

(b)(4) & (b)(6)

[redacted]

EU-Project FAST-ID (2010-2013)

»**FAST** and efficient international disaster victim **Id**entification«

[redacted]

System Integration

(b)(4)

Secondary Identifiers

Main Identifiers

Fraunhofer IOSB

FASTID Image Retrieval: Find images of the very same tattoo

all -(b)(4)

[redacted]

Tattoo image of dead body

[redacted]

Tattoo images of missing persons

[redacted]

Match candidate

Fraunhofer IOSB

Complexity of Retrieval

Exact duplicate image retrieval – file hashes

e.g. Microsoft PhotoDNA (Bing, SkyDrive, Twitter, Facebook…)

Near duplicate image retrieval – global / local feature hashes

(b)(4)
(b)(6)

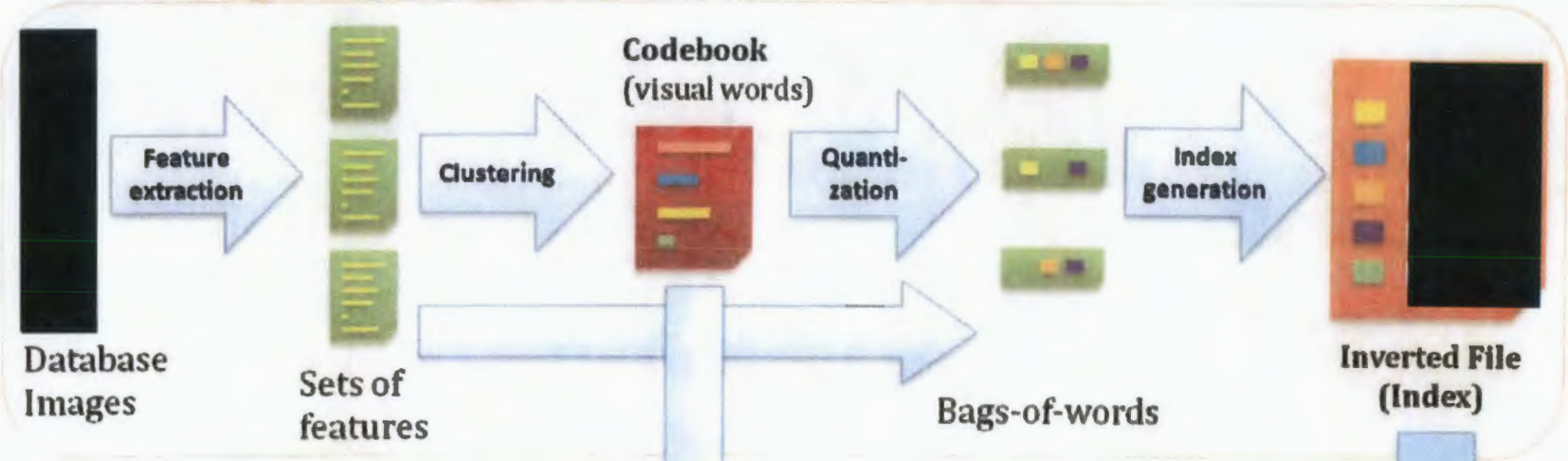
FASTID

Real world object retrieval – local feature quantization

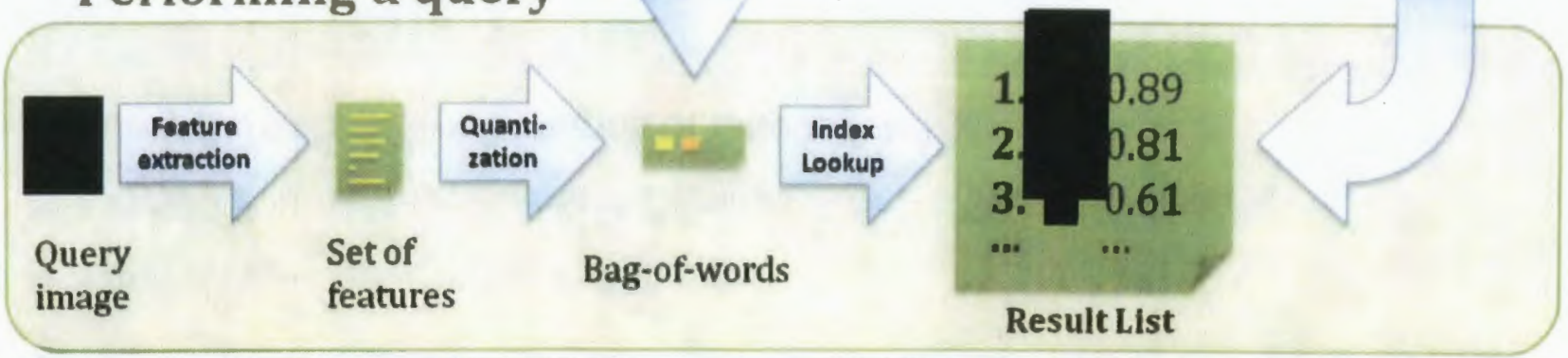
Fraunhofer IOSB

all - (b)(4)

Creating the Image Retrieval database



Performing a query

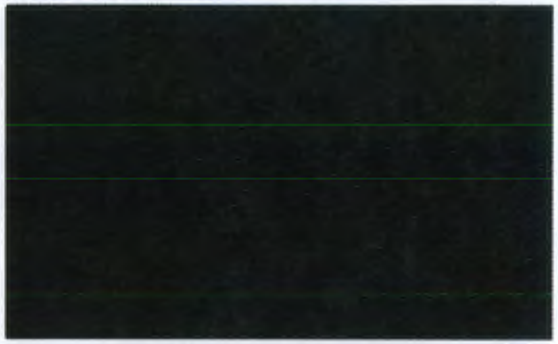


Fraunhofer IOSB

Image Retrieval: data, results

all (b)(4) & (b)(6)

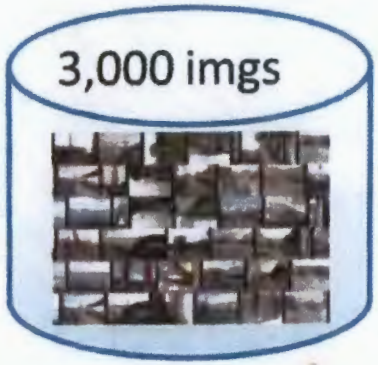
- 8,400 images from tattoodesign.com
- 330,000 images from Federal Criminal Police Office of Germany
- Query & Truth: 424 x 2 images of tattoos from www.wildcat.de



- Results (424 probes, Rank 1 acc.):
  - with 10k tattoo distractors: 90%
  - with 330k tattoo+bodymod distractors: 85%

Fraunhofer IOSB

Image Retrieval: Batch matching



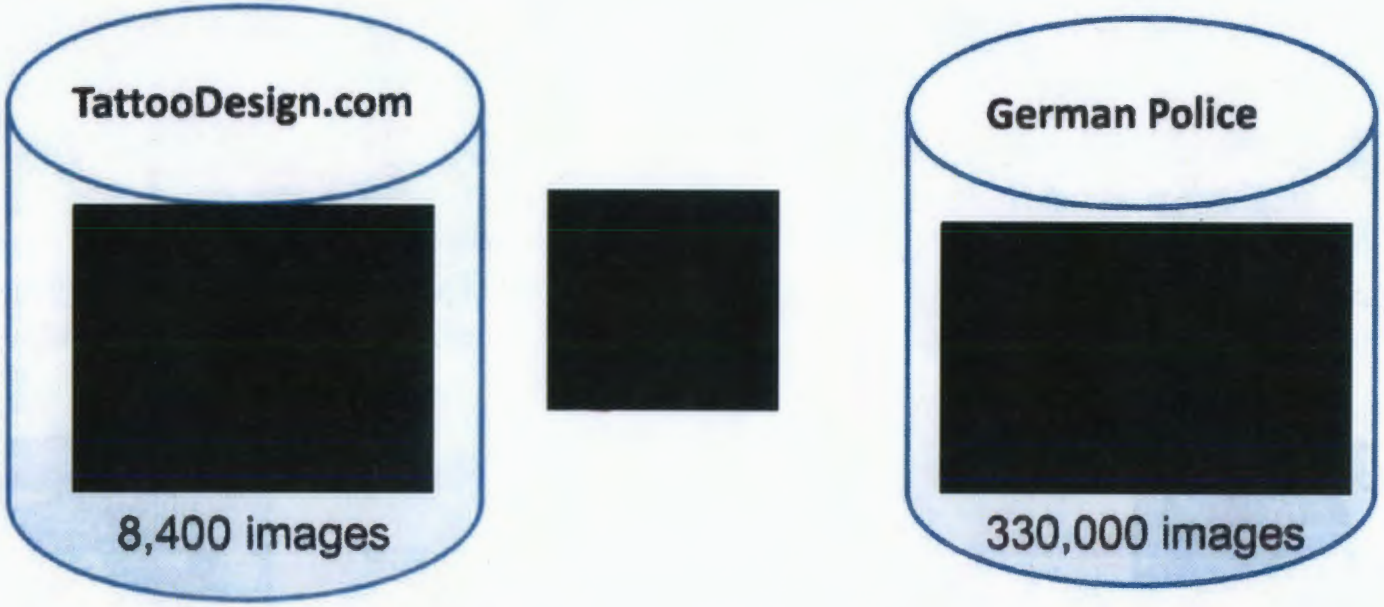
Matching of databases:
find objects / scenes which are in both databases

Fraunhofer IOSB

Image Retrieval: Batch matching

all - (b)(4)

- Task: find tattoos which are in both databases



© Fraunhofer IOSB

all - (b)(4) & (b)(6)

[redacted]

[redacted]

IOSB

all - (b)(4) & (b)(6)

[redacted]

all (b)(4) & (b)(6)

[redacted]

© Fraunhofer IOSB

[redacted]

all - (b)(4) & (b)(6)

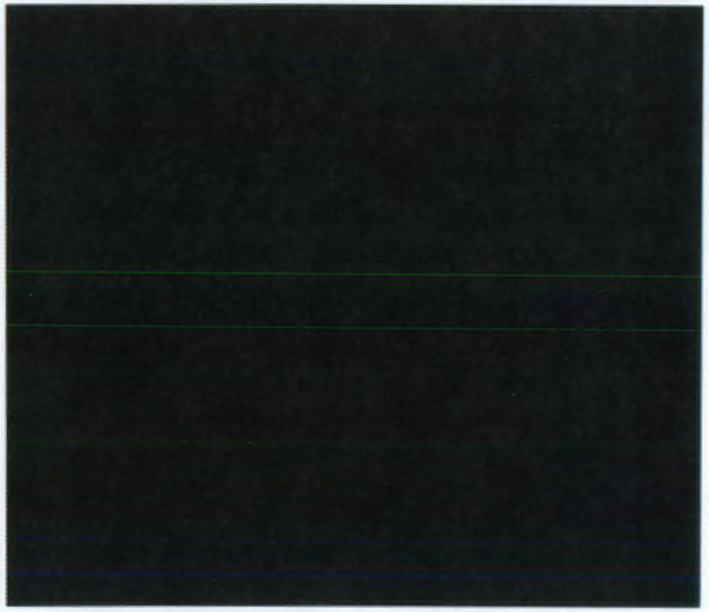
[redacted]

Fraunhofer
IOSB

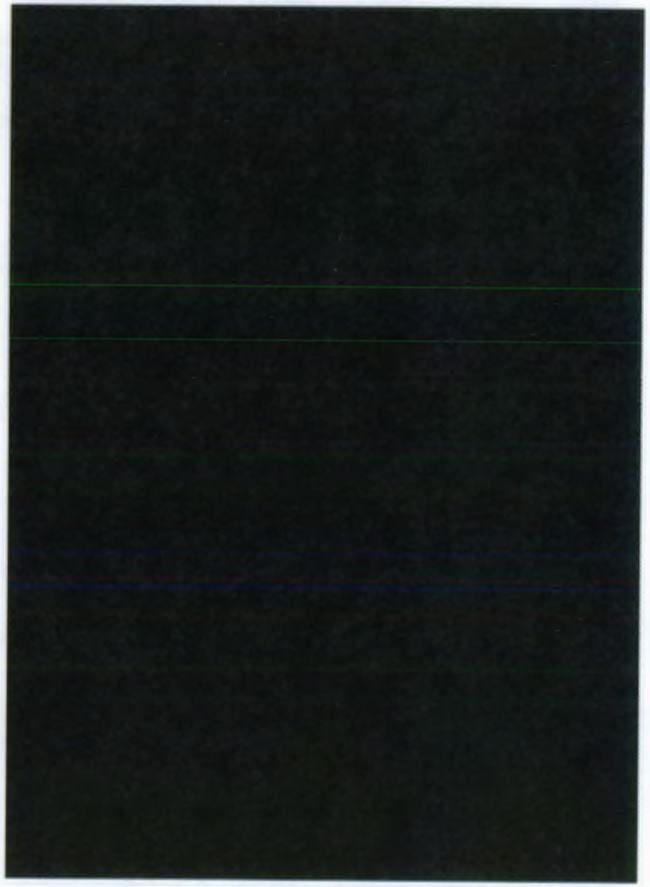
Limits

all - (b)(4) & (b)(6)

Query image



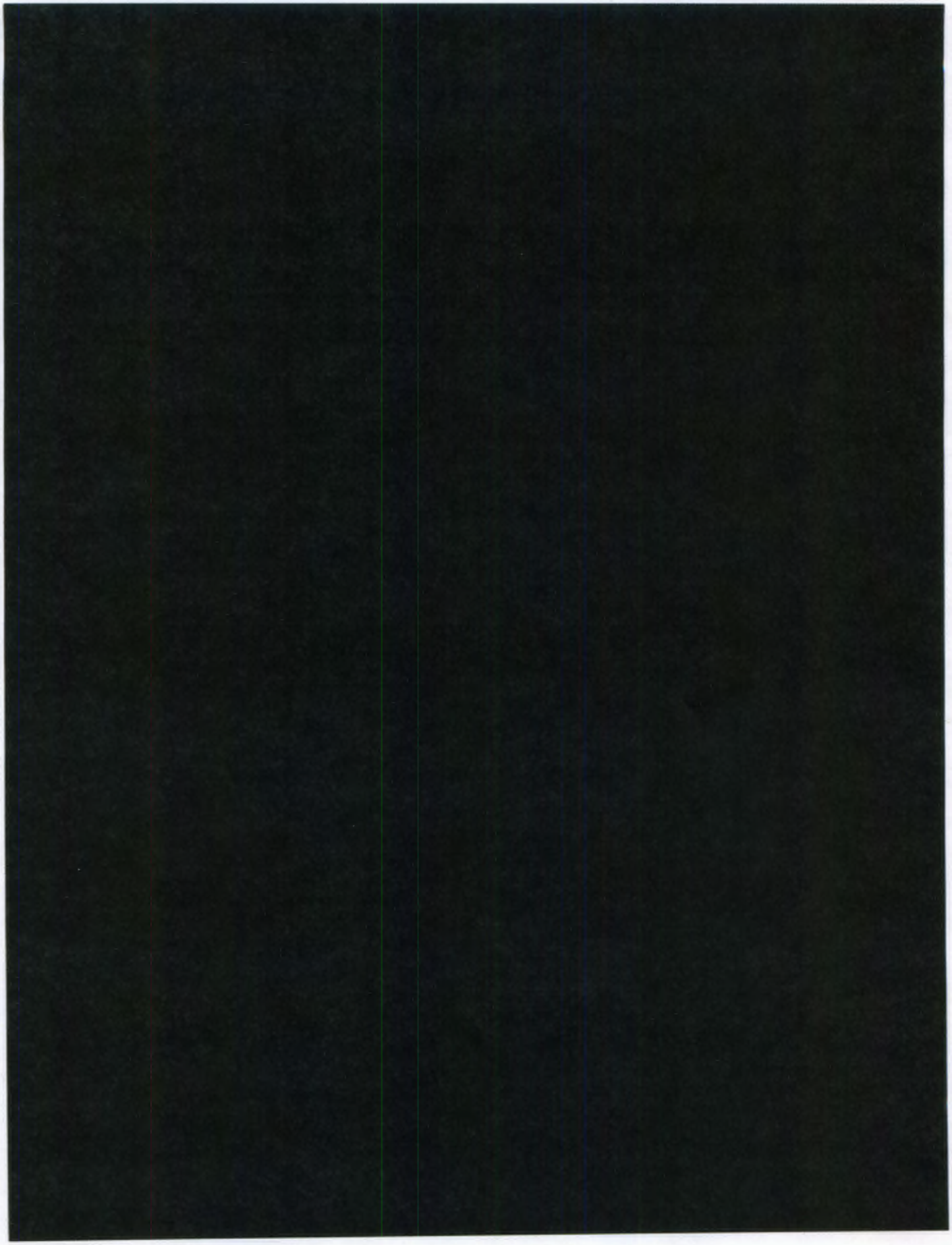
Most similar image of DB



© Fraunhofer IOSB

Fraunhofer
IOSB

(b)(4) & (b)(6)



[redacted]

[redacted]

Fraunhofer
IOSB

all (b)(4)

Submission details

[redacted]

ID-1 (97.5% rank1):

Features: SIFT features (Vlfeat lib), RootSIFT normalization.

Matching score using the Re-Ranking function:

- Matches: Lowe's matching distance ratio (2nd/1st > 1.5)
- Assess local plausibility of matching features using the positions and angles of the neighboring matching features.
- Score: Nb of plausible matches normalized by sum of features of both images

ROI-1 (96% rank1):

Matching score:

see ID-1 but without the local plausibility assessment since local context in cropped images is often too small

ID-1: missing images

Submission details (2)

all (b)(4)

ID-2 (96.8% rank1; 1 missing image compared to ID-1):

Features: SIFT features (Vlfeat lib), RootSIFT normalization.

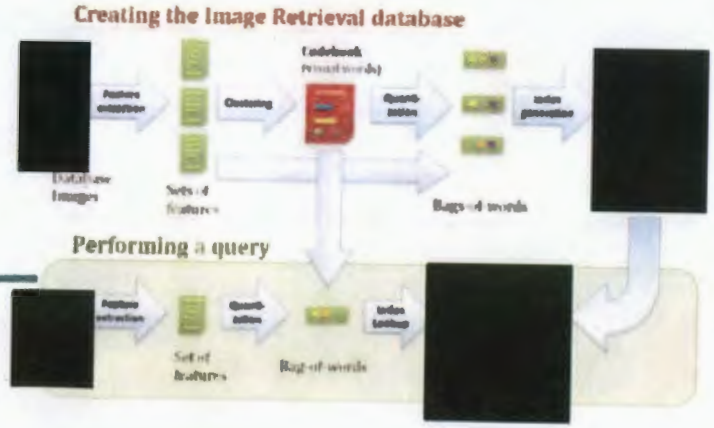
Codebook: hierarchical K-Means 7^6=117,649 Visual Words

Hamming Embedding, WGC (angle only), Multiple assignment to up to 3 features closer that 1.2*NN

Re-Ranking the first 500 images with the local plausibility score

Too small database to really challenge the quantization process

„Tuning space“: One more single image to find on rank1 is not enough to perform parameter tuning…



Research todo list

[redacted]

[redacted]

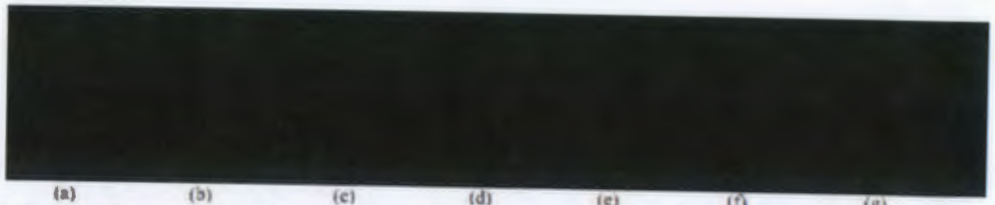
1. Early stages of tattooing when „edges still move“
2. Large tattoos on limbs: use features with an up to 90° difference to camera viewing angle
3. Prevent small blurry tattoos with very few features from getting lost in random noise
4. Too large differences in scale (hairs, pores, reflections from vessels as distractors)

© Fraunhofer IOSB

Fraunhofer IOSB

(b)(4)

Towards the dream dataset



(a) (b) (c) (d) (e) (f) (g)

Fig. 10. Examples of image transformations: (a) original and variations due to (b) blurring, (c) illumination change, (d) color component changes, (e) affine transformation, (f) aspect ratio change, and (g) rotation 1)

- The early approaches :
  - Too easy for algorithms, too unrealistic!
- From users in the internet (e.g. the 2x424 imgs from wildcat.de)
  - Better but laborious. Often, images are from the same camera, close in time, similar background etc.
- Police and user images
  - Realistic use case – different capturing setup guaranteed; not only „nice“ tattoos. But: Getting truth data very hard! (example Tattoo convention)
- Privacy and copyright issues
  - Imgs of german police were not considered personal data, however…

1) Scars, marks and tattoos (SMT): Soft biometric for suspect and victim identification - JE Lee, AK Jain, R Jin - Biometrics Symposium, 2008

Towards the dream dataset (2)

- **Quantity**:
  - Cross validation is not the panacea for overfitting in small datasets
  - >1,000 probes, for each topic on the research todo list at least 100 probes to exhaustively cover the aspects
  - Distractor images: >100,000 to examine Image Retrieval Frameworks
    - However, tattoos are not that unique any more in large databases
- **Metric**:
  - CMC can't reward algorithms which find all 4 truths instead of only one
    - „the probability that one or more correct matching image for a probe is observed within the top K ranks". One easy truth renders the 3 other difficult ones „useless"

(b)(4) & (b)(6)

Thank you!

[redacted]

Questions?

Fraunhofer
IOSB