Improvements of Face Detection and Recognition

Vladimír Míč
Jan Sedmidubský
Michal Batko
Pavel Zezula

Outline

- Basic terms
  - Face detection and recognition
  - Quality evaluation

- Face retrieval system

- Original approach

- Face detection improvement

- Face recognition improvement

- Future work

Face detection problem

- Goal is to put an ellipse on the place where a face is

- Recall: how many faces out of real faces were detected? (100 %)

- Precision: how many faces out of detected ones represent real faces? (66,7 %)

Face recognition problem

- In general, the ability to answer the question "is it the same person in these two images?"

- In practice, sorting faces according to the similarity with respect to a query

- Example: 21 photos of person "00003", 10-NN query

- Recall = 4 / 21 = 19 %,

- Precision = 40 %

Distance	Image
0	00003_941121_rd.jpg
8932	00003_941121_hr.jpg
9145	00003_941121_qr.jpg
9167	00003_941121_rc.jpg
9277	00750_941201_hr.jpg
9281	00765_941201_hr.jpg
9282	00972_960627_qr.jpg
9283	00695_941121_qr.jpg
9285	00772_941201_re.jpg
9286	00750_941205_hr.jpg

# Face retrieval system

- Detection and recognition of human faces

    - Only static images

- Web application

- Real time system

    - Huge number of photos

    - Demo with 1 000 000 pictures

    - Images in a low quality

    - Result – the most similar photos for an arbitrary query in a few seconds

Original approach – presumptions

- Extractor for MPEG7 descriptors

  - Extraction from an arbitrary picture

  - Meets metric properties

- More sophisticated software (software B)

  - Higher quality than MPEG7

  - No constraints (e.g., about metric properties)

Original approach – performance boost

- **Face detection** using software B

- Face **crops** from images

- MPEG7 descriptors from crops

- **Index** over MPEG7 descriptors

- **Candidates selection** using index

- **Overrank** of candidates according to the recognition

  results performed by software B
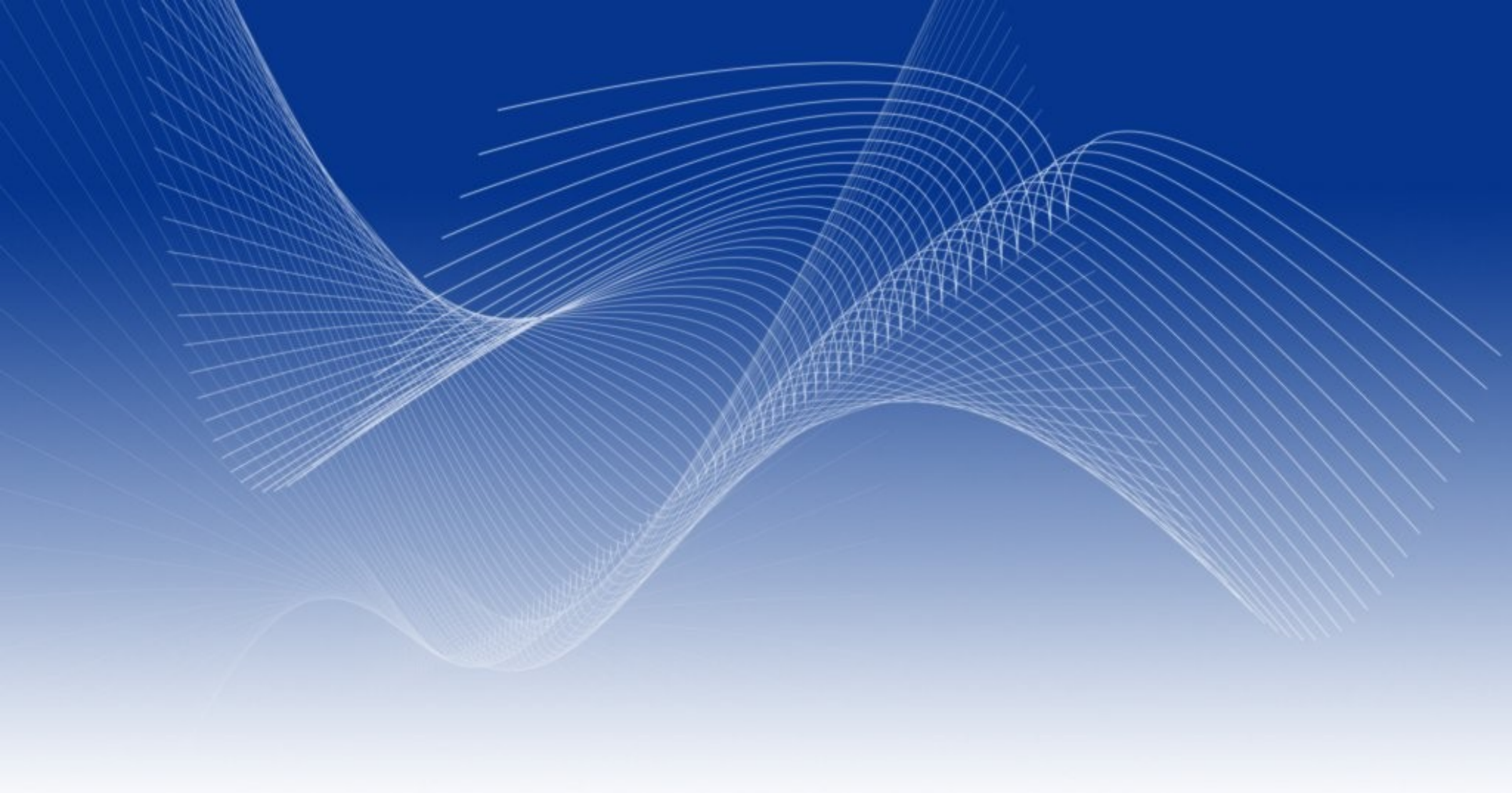
Face Detection Improvement

New approach

- Performed by 3 independent pieces of software

    - OpenCV (makes MPEG7 descriptors)

    - Luxand

    - Neurotechnology (Verilook)

    - Extraction made in parallel

- Detection: compliance of at least 2 out of 3 pieces

of software

    - Precision nearly 100 %

Face detection results

Name	Recall[1]	Precision[1]	Recall[2]	Precision[2]
Open CV (OCV)	55 %	89 %	92 %	86 %
Luxand	63 %	83 %	95 %	94 %
Neurotechnology (Verilook)	73 %***	84 %	100 %	96 %
Aggregated extractor	62 %	98 %	97 %	100 %

[1] 1260 small faces, low quality [2] 66 big faces, high quality

# Face Recognition Improvement

Aggregated face descriptor

- Aggregated face descriptor encapsulates

  - MPEG7 descriptor

  - Luxand descriptor

  - Verilook descriptor

- MPEG7 descriptor is always present

  - Can be added via crop made according to the Luxand
    or Verilook descriptor

- One of Luxand and Verilook descriptors may be missing

  - Not both

Face recognition principle

- Based on a distance function

  - How to measure a distance between two Aggregated descriptors?

- Missing encapsulated descriptors can't participate in distance calculation

- Distances between encapsulated objects

  - For each couple of Aggregated descriptors it's possible to get from 1 to 3 distances

- How to mix more distances in order to get the best recognition function?
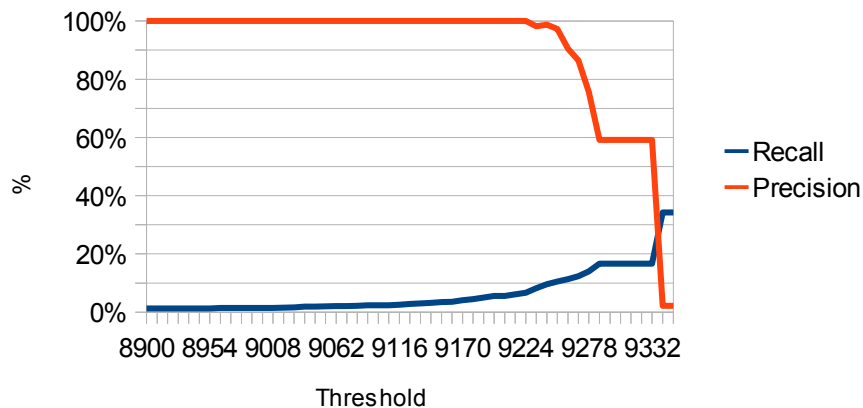
Distance quality measurements
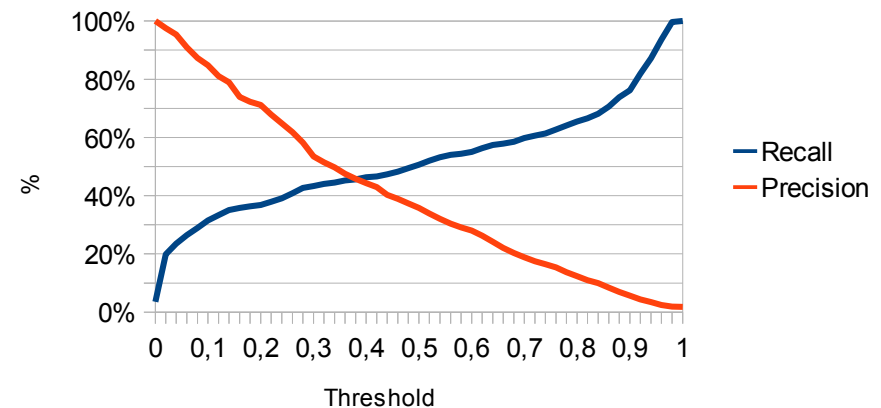

MPEG7 distance function
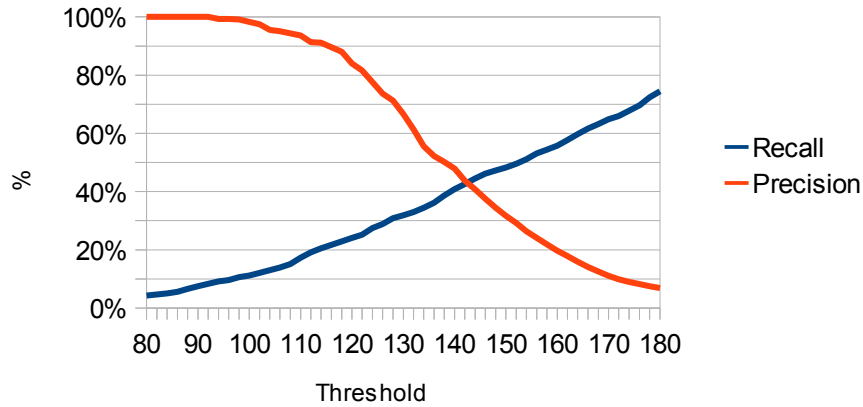
Luxand distance function
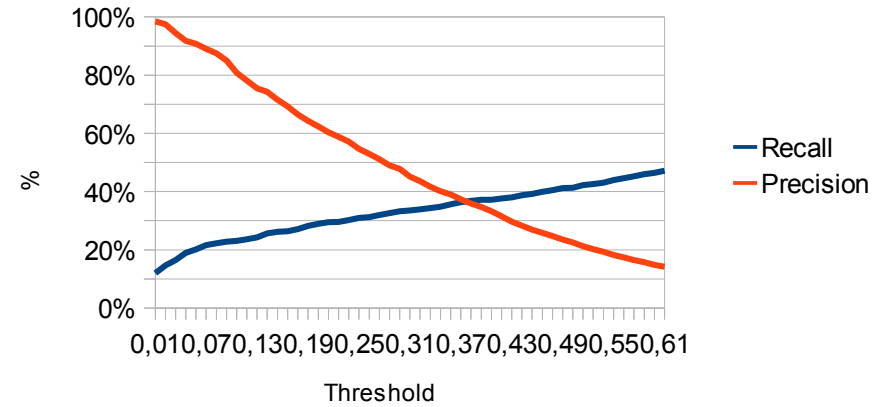
Verilook distance function

Aggregated distance function

Distance of encapsulated objects normalization

- Normalization according to the precision

- Normalized distance is within [0, 1]

- Original distance d on which the precision was p

  is normalized to a new distance $1 - p$

  - Quite accurate approximation is used

- Aggregated distance is equal to minimum of normalized

  partial distances

Face recognition summary

Name	Recall[1] on precision 85 %	Recall[1] on precision 95 %	Recall[3] on precision 85 %	Recall[3] on precision 95 %
Open CV (OCV)	24 %	14 %	8 %	3 %
Luxand	23 %	16 %	14 %	Not possible
Neurotechnology (Verilook)	12 %	11 %	53 %	51 %
Aggregated extractor	31 %	24 %	54 %	51 %

[1] 753 small images, low quality [3] 11 338 big images, high q.

Future work

- Index build on Aggregated descriptors using MPEG7 distance function

- Candidates selection

- Overrank according the aggregated distance function

- Prototype system is expected to run in a few days :-)

Thank you for your attention