Improvements of Face Detection and Recognition

Combination of existing methods

Bc. Vladimír Míč
FI MU

Outline

- Basic terms
  - Face detection and recognition
  - Quality evaluation
- Avaliable software
- Face detection improvement
- Aggregated descriptors
- Face recognition improvement

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
- Similarity expressed via distance function
- Example: 21 photos of person “00003”, 10-NN query
- Recall = 4 / 21 = 19 %,
- Precision = 40 %

Distance	Person id
0	00003
8932	00003
9145	00003
9167	00003
9277	00750
9281	00765
9282	00972
9283	00695
9285	00772
9286	00750

Available software

- OpenCV
  - Opensource, MPEG7 descriptors, metric properties
  - Can make descriptor from an arbitrary picture
- Luxand, Neurotechnology (Verilook)
  - Commercial software, own descriptors
  - Recognition only of faces detected by its own

Detection improvement

- Based on compliance of pieces of software
  - Compliance means sufficient overlay of detected areas
- Recommended variant: compliance of at least two out of all pieces of software
  - Precision nearly 100 %
- Aggregated descriptor
  - Holds several descriptors of the same face

Aggregated face descriptor – our variant

- Encapsulated objects
  - MPEG7 descriptor
  - Luxand descriptor
  - Verilook descriptor
- MPEG7 descriptor is always present
  - Can be added via crop made according to the Luxand or Verilook descriptor

Face detection results

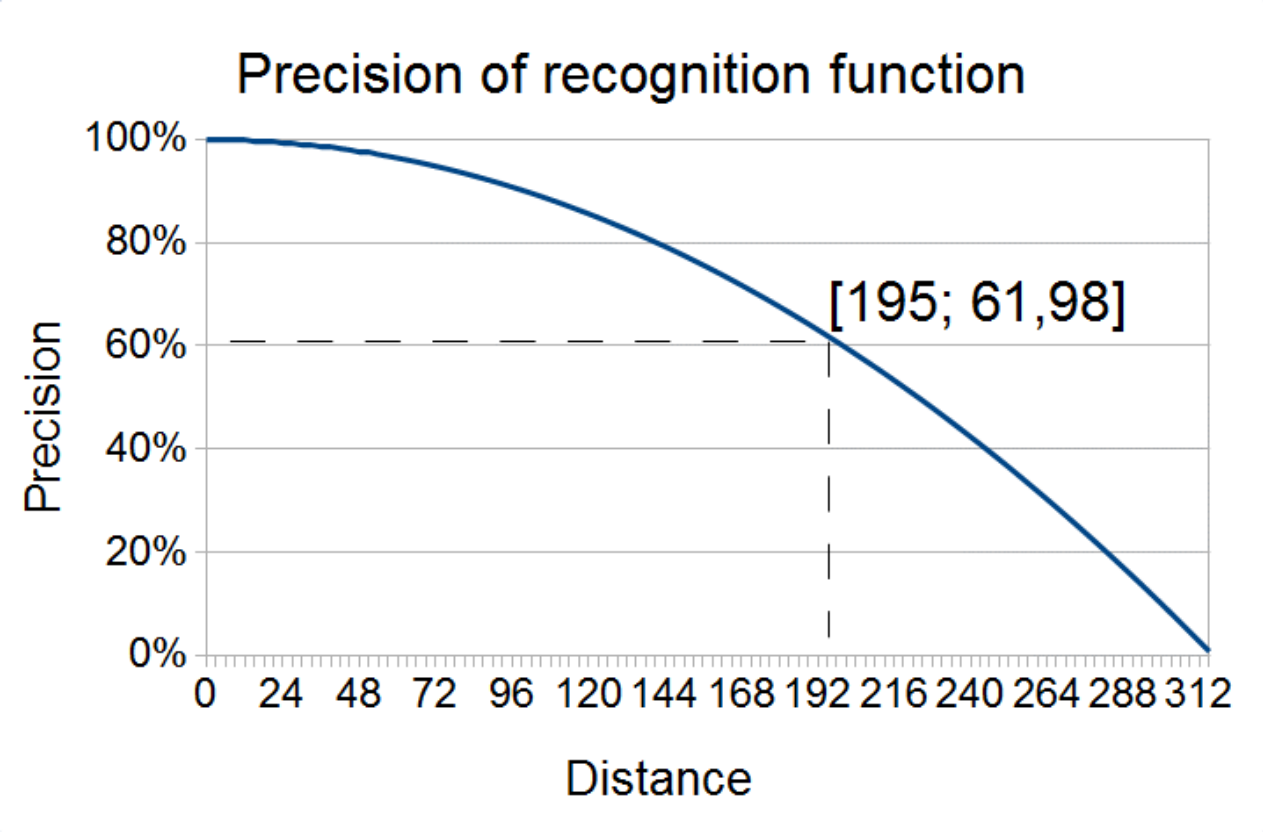
Name	Recall[1]	Precision[1]	Recall[2]	Precision[2]
Open CV (OCV)	55 %	89 %	92 %	86 %
Luxand	63 %	83 %	95 %	94 %
Neurotechnology (Verilook)	73 %	84 %	100 %	96 %
Aggregated extractor	62 %	98 %	97 %	100 %

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

Recognition improvement

- Distance between Aggregated descriptors
  - Combination of several distances between encapsulated objects (partial distances)
  - Some partial distances may be missing
- Normalization of each partial distance by a precision
- Use minimal value of normalized partial distances

Distance normalization



- Precision is measured
- Need of training sample data

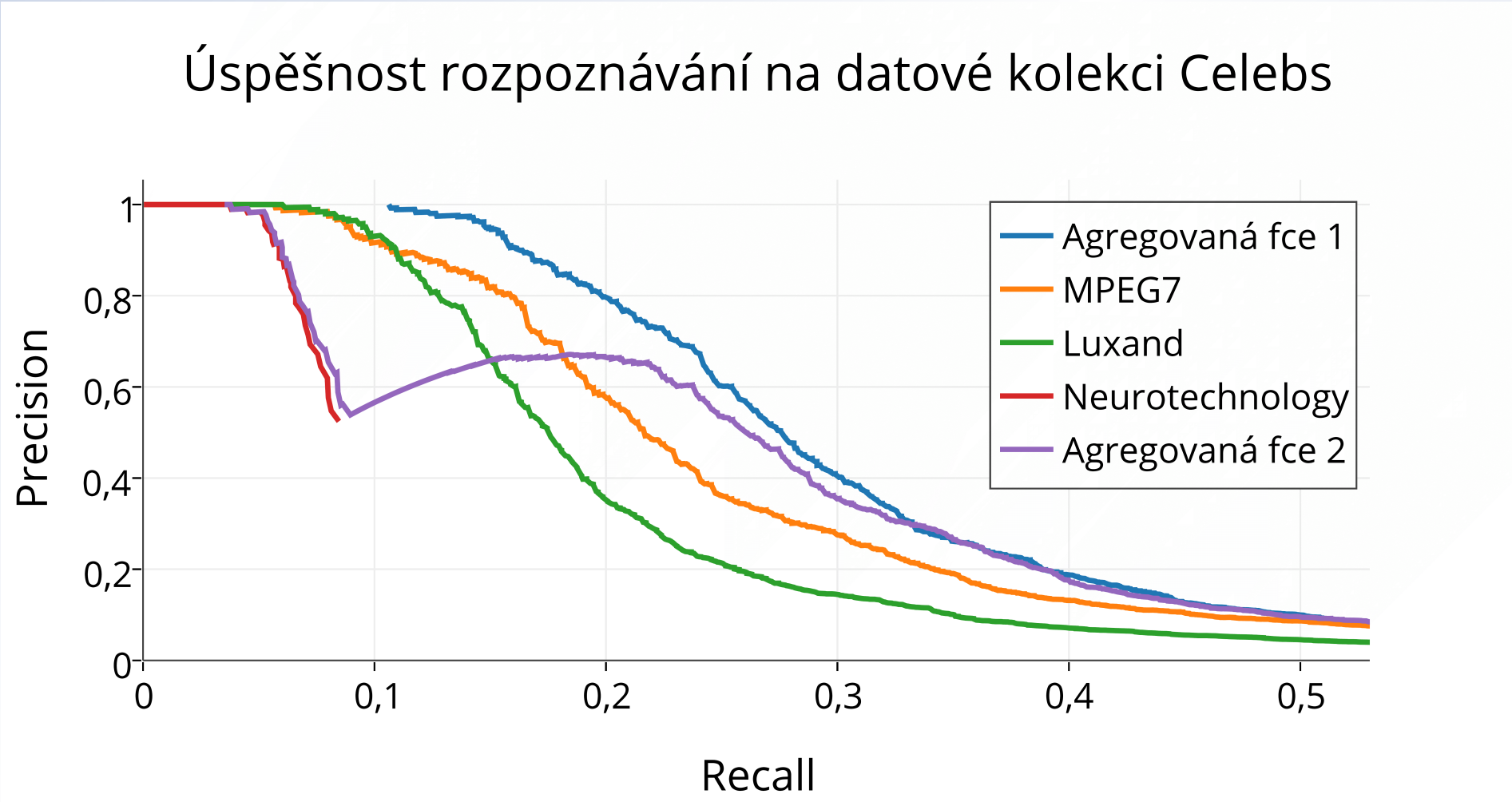
- Normalization function: norm(*d*) = 1 – *p*
- Illustration:

  distance 195 is normalized to 1 – 0,6198 = 0,3802

Training data problem

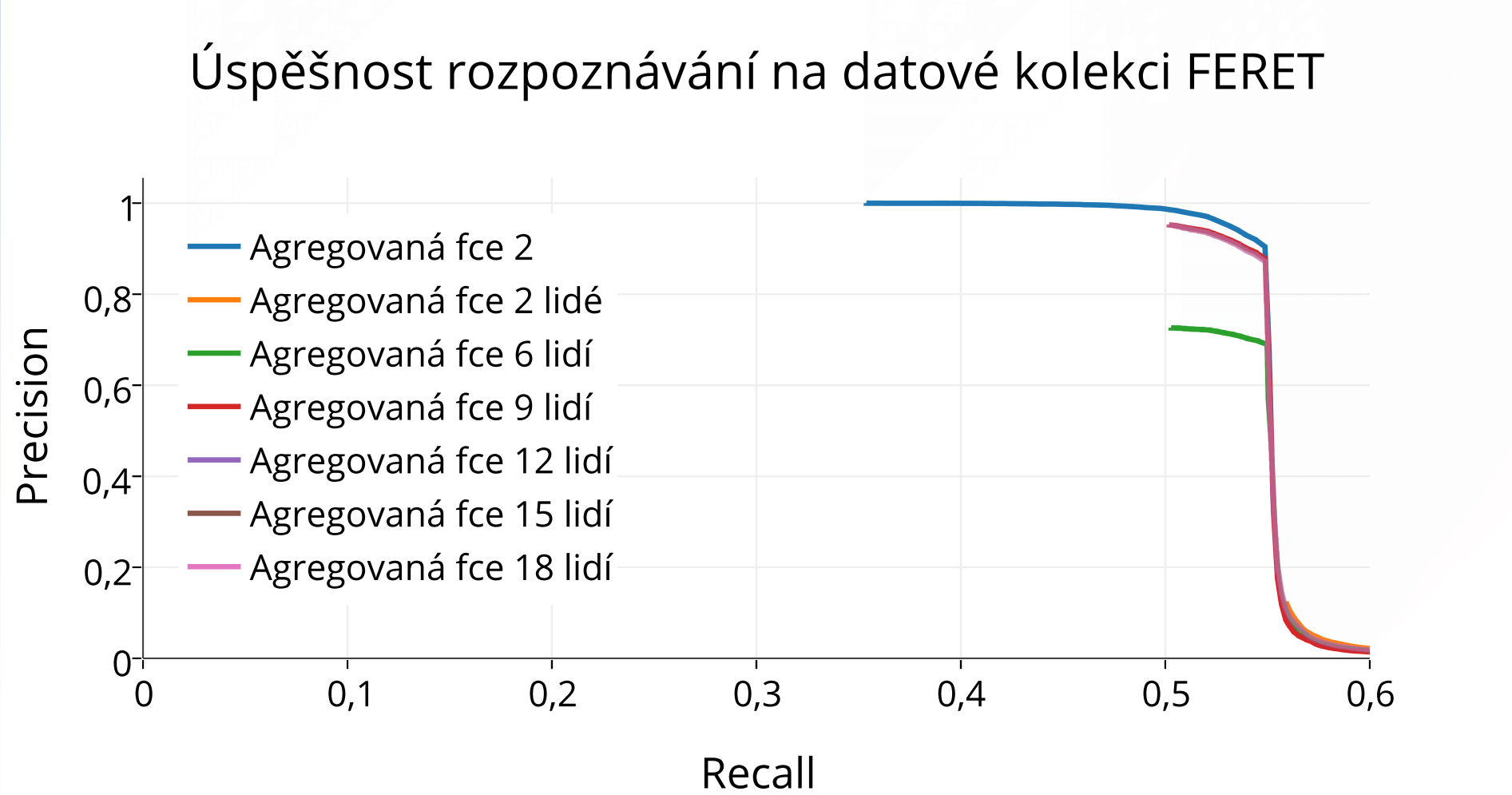
- Cons:
  - Normalization strongly depends on a dataset
  - It's suitable to provide own training sample
  - User must identify faces (abstract id, name, …)
- Pros:
  - High precision of face detection may be used
    - Face id in file path for images with one face
  - Training data may be very small to provide good results

Real testing sample – 753 small images



Results with small training datasets

- FERET: 998 people, 6 379 faces (red: 9 people, 224 faces)



Performance boost

- Index build on Aggregated descriptors using MPEG7 distance function (needs metric properties)
- Candidates selection (logarithmic complexity)
- Overrank according the aggregated distance function

Thank you for your attention