JOANNEUM RESEARCH

DIGITAL

Institute for Information and Communication Technologies

www.joanneum.at

JOANNEUM RESEARCH and Vienna University of Technology at INS Task

Werner Bailer

TRECVID Workshop, Nov. 2010

JOANNEUM RESEARCH

Outline

- Approach
- Subsystems and features
- Fusion strategies
- Results
- Conclusion

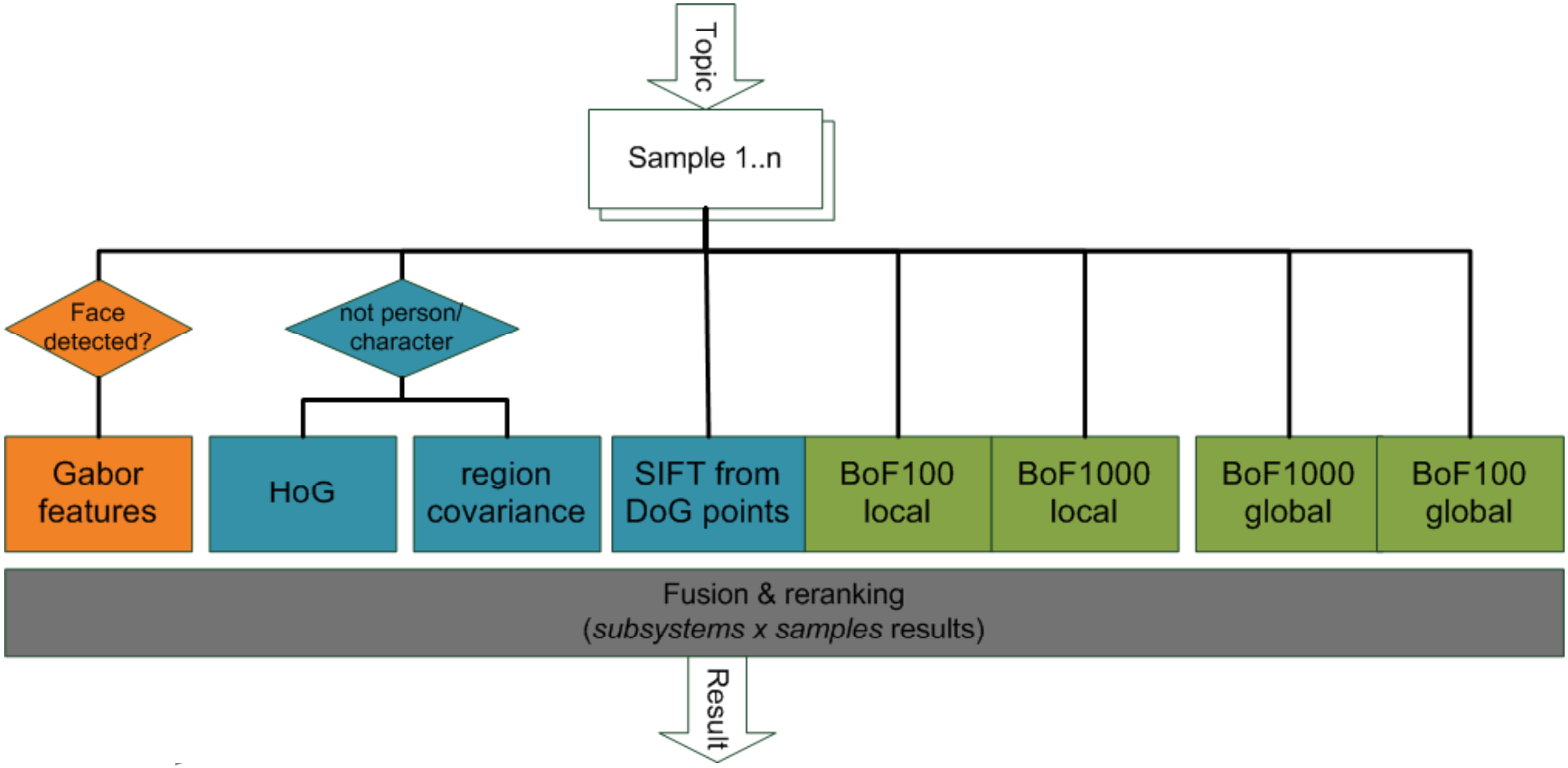
JOANNEUM RESEARCH

Approach

- fully automatic
- set of independent subsystems, using different features
- query each sample of a topic independently
- each subsystem returns a ranked result list for each sample
- research focus: fusion strategies

JOANNEUM RESEARCH

System Overview



JOANNEUM RESEARCH

Subsystems (1)

- Gabor feature
  - perform face detection (Viola-Jones)
  - if face detected, extract Gabor wavelet descriptor from face region
  - match against descriptors of all face regions in database
  - k-NN search
- Histogram of gradients
  - not used for person/character
  - descriptor with 36 bins (9 orientations, 4 cells)
  - cell layout is adapted to aspect ratio of query object: 2x2 or 1x4 cells
  - search window is shifted ¼ cell size
  - 3 scales: 1x, 1.5x and 2x initial size

JOANNEUM RESEARCH

Subsystems (2)

- Region covariance
  - covariance of rectangular region (can be determined efficiently using integral images)
  - from RGB and first-order derivatives of intensity
  - same cell sizes/scales as for HoG
- SIFT
  - from DoG points
  - matching: voting in a position histogram (1/10 of image size), report match for bins with 5+ votes
- Bag of visual features (BoF)
  - SIFT descriptors from DoG points and global
  - codebook sizes 100 and 1000 for both

JOANNEUM RESEARCH

Pre-computed features

- Pre-computed for database
  - face detection + Gabor descriptor
  - global SIFT extraction
  - BoF codebook generation
- At query time
  - interest point detection + SIFT extraction
  - HoG
  - Region covariance

JOANNEUM RESEARCH

Fusion strategies (1)

- Two simple methods, not making use of query samples
- Max-max
  - For each shot in the results, take maximum scope of all samples and features
- Top-k
  - For each feature, take for each shot the maximum of all samples
  - Rerank per feature
  - Take the top-k per feature (k=1000/no. features used)

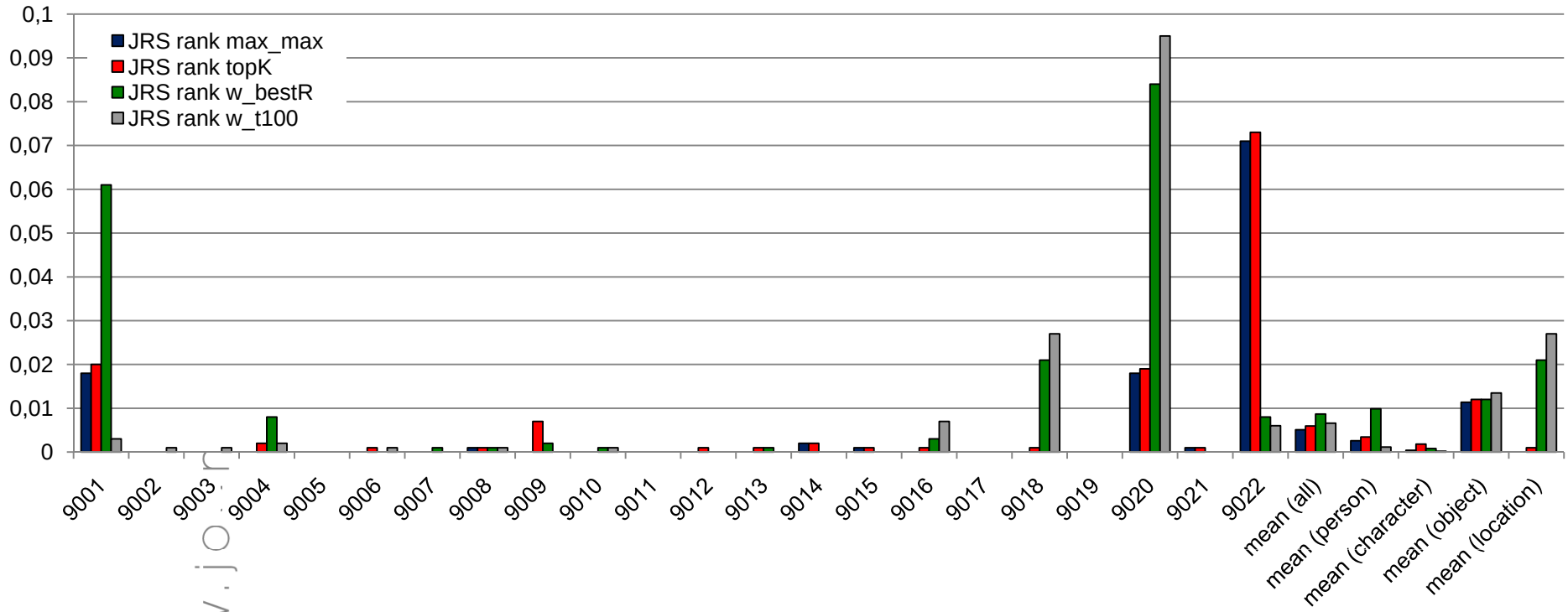
JOANNEUM RESEARCH

Fusion strategies (2)

- Two methods using query samples
  - idea: weight features by their relative performance
  - for each sample, determine where the other samples would be ranked in the result if they were in the database
- best rank
  - determine mean best rank over all samples for each feature
  - calculate feature weight as $w_{bestR}(f_i) = \frac{\max_{\forall f_j}(\bar{r}_j) - \bar{r}_i}{\sum_{\forall f_k} \max_{\forall f_j}(\bar{r}_j) - \bar{r}_k}$
- top 100
  - determine how many samples are in the top 100 results
  - calculate feature weight as $w_{t100}(f_i) = \frac{\bar{n}100_i}{\sum_{k=1}^{N} \bar{n}100_k}$

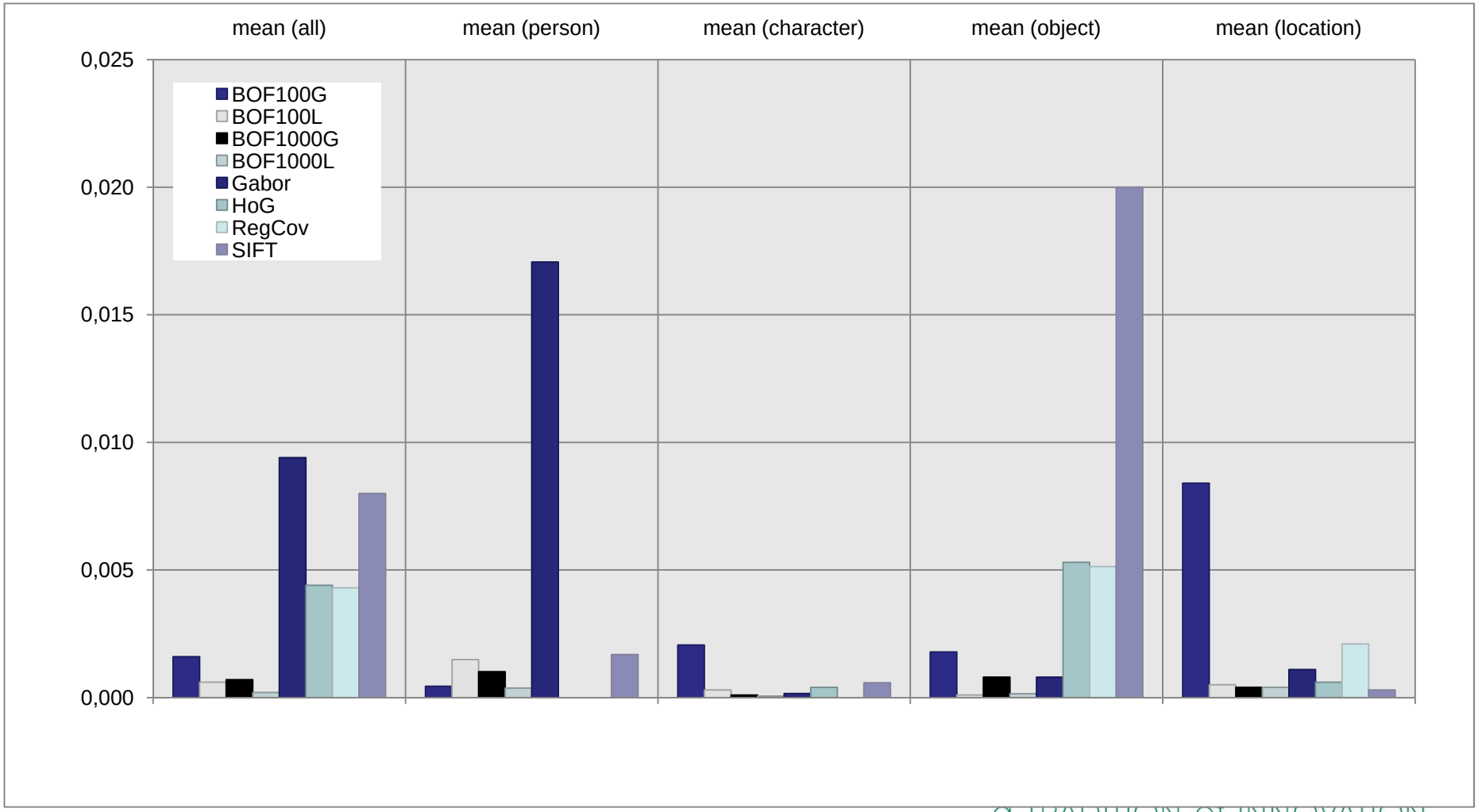
JOANNEUM RESEARCH

Results per topic/type



JOANNEUM RESEARCH

Results per feature



JOANNEUM RESEARCH

Conclusion (1)

- Task is difficult, results for automatic system poor
  - different sizes, lighting, perspectives, …
  - “needle in a haystack”: very few relevant results in a large set with many similar objects (e.g. pedestrian crossing, blinds)
- Features
  - as expected, our features perform best for object queries
  - better results could be possible for some of the features, but would make matching process more costly

JOANNEUM RESEARCH

Conclusion (2)

- Fusion methods
  - Overall, the fusion methods using information from query samples perform better
  - Only slight difference for object queries
- To fuse or not to fuse?
  - for person and object queries, a single feature outperforms the best fused results
  - few topics for the other query types, thus difficult to say if fusion is actually useful in these cases

JOANNEUM RESEARCH

The research leading to these results has received funding from the European Union's Seventh Framework Programme under the grant agreements no. FP7-215475, "2020 3D Media – Spatial Sound and Vision" (http://www.20203dmedia.eu/) and no. FP7-248138, "FascinatE – Format-Agnostic Script based INterAcTive Experience" (http://www.fascinate-project.eu/), as well as from the Austrian FIT-IT project "IV-ART – Intelligent Video Annotation and Retrieval Techniques".