JOANNEUM RESEARCH DIGITAL

DIGITAL - Institute for Information and Communication Technologies

JOANNEUM RESEARCH and Vienna University of Technology at INS Task

Werner Bailer, Robert Sorschag, Harald Stiegler, Christoph Wiedner

TRECVID Workshop, Nov. 2012

Outline

- Approaches & Runs
- Results & Discussion
- Using Google image search for alternative examples
- Conclusion

abstract

The JRS-VUT team applied two approaches with quite complementary properties: One with preprocessing and indexing (based on a bag-of-words approach using ColorSIFT), and very fast query times (at most one minute), and the other doing no preprocessing at all, but doing SIFT matching at query time, implemented on the GPU (but still significant query times). The fast approach turned out to be not discriminative enough for this type of queries, performing worse than on the TRECVID 2011 data set. The matching at query time performs quite well overall, providing few but mostly correct results for many queries. The performance varies among the different queries, depending significantly on the size and structure of the query samples. We also experimented with fusing the results, by ranking those found with high confidence using matching at query time on top, and filling the list with the list obtained from the fast method. The fused results improve the results for the majority of the queries. Nonetheless the mean AP overall queries is slightly lower, due to the fact that a number of correct results from the matching at query time are discarded for single queries.

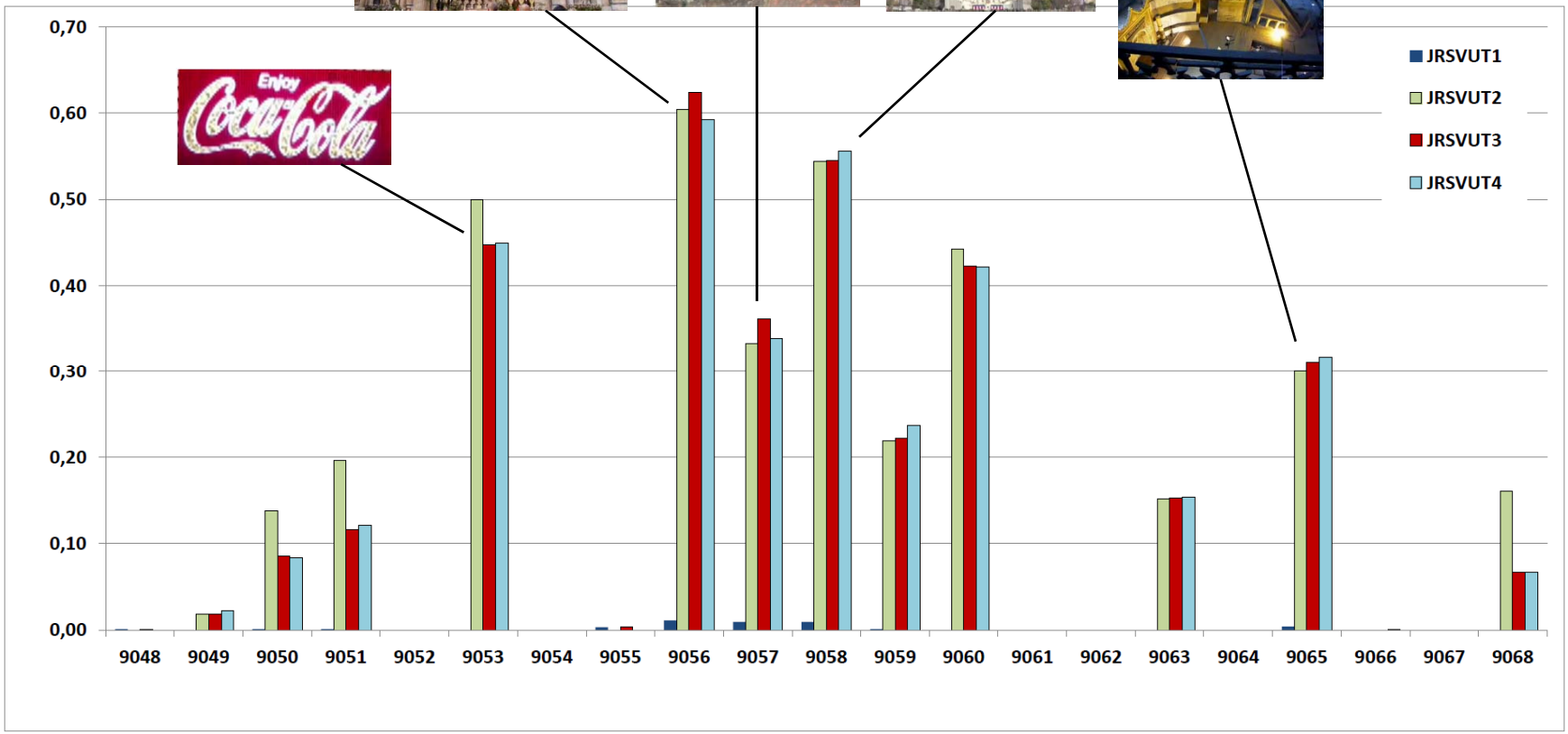
Two Approaches

- ColorSIFT BoW
  - index descriptors from database in advance
  - very short query times (<1min)
- SIFT matching at query time
  - no preprocessing
  - extract and match SIFT descriptors extracted from DoG points at query time
  - GPU accelerated matching, but still time consuming (~10 hours/query)

Runs

- JRSVUT1: only indexed ColorSIFT, from densely sampled points
- JRSVUT2: only SIFT matching at query time
- JRSVUT3: fusion of top results of SIFT matching at query time and indexed ColorSIFT from densely sampled points
- JRSVUT4: fusion of top results of SIFT matching at query time and the indexed ColorSIFT from DoG points.
- Fusion method for runs 3 and 4
  - observation: top SIFT matches are very reliable
  - estimate threshold for SIFT results: score with the steepest gradient at the lower third of score values

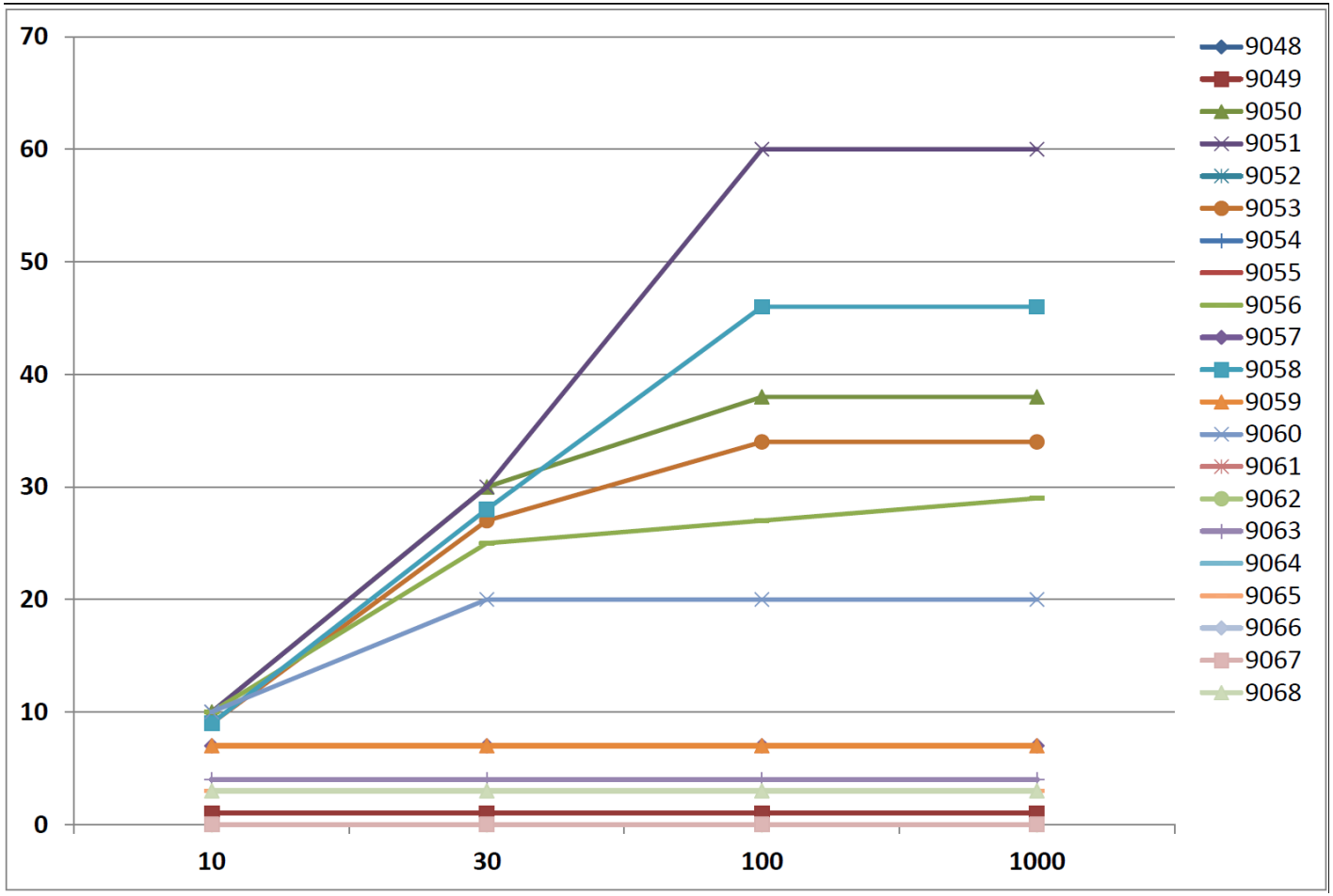
Results (1)


0,70
0,60
0,50
0,40
0,30
0,20
0,10
0,00
9048
9049
9050
9051
9052
9053
9054
9055
9056
9057
9058
9059
9060
9061
9062
9063
9064
9065
9066
9067
9068
JRSVUT1
JRSVUT2
JRSVUT3
JRSVUT4
Enjoy
Coca-Cola

Results (2)

- Indexed ColorSIFT
  - fast
  - not discriminative enough
  - performed worse than a very similar approach used for INS 2011
- SIFT at query time
  - results for queries 9053, 9057 and 9058 are at or close to overall best result.
  - issues with very low number of reliable interest (e.g., low resolution query samples)
  - issues with not sufficiently discriminative feature points

Number of hits found for run JRSVUT2 at rank 10, 30, 100 and 1000



Results (3)

- Person query had surprisingly good results, but only because many occurrences are very similar
- Fusion
  - for 9 queries, one or both fused runs yield slightly better average precision
  - for 5 queries, too many correct results dropped (threshold too high, ranking not discriminative around threshold)

SIFT issues (details)

- Using only fields improved results
  - higher spread of scores between true and false matches
- Low # interest points
  - Mercedes star (9048) and London underground logo (9052): query samples with very low resolution.
  - Pepsi logo (9061) few interest points even on samples with higher resolution
- not sufficiently discriminative feature points
  - Stonehenge (9054) and Hoover Dam (9066) queries
  - many of the extracted feature descriptors have lots of matches

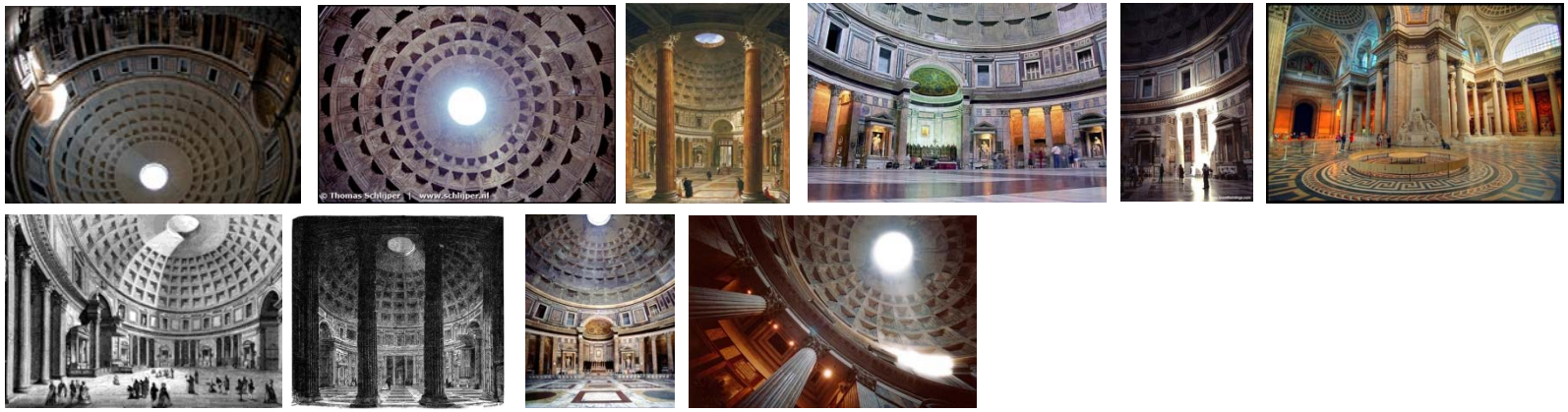
SIFT issues (details)

- both of those issues
  - Empire State building (9064)
  - Sears Tower (9055)
- MacDonald’s logo (9067) is special case
  - samples contain differently illuminated versions of the logo.
  - descriptors rather capture the lighted areas on dark
  - background: many shots with different light effects




Example images using Google image search

- Use text of search topic as query for Google images search (under quotes), top 10 images
- Pantheon interior



  - similar to provided examples (except for 2 engravings), similar performance (slightly worse)

Example images using Google image search

- Mercedes star

- MacDonald‘s arches

- most are graphical/synthetic, only small improvement

Conclusion

- Our bag-of-words approach was not discriminative enough
- SIFT matching a query time
  - performed very well for some queries
  - stricter verification did not help, as most problems due to low number of feature points
  - Example images from Google image search yield similar results, for “good” queries ranking is different, so fusion of the two could further a little further improvement
  - Slow despite matching on GPU, but still room for optimisation

TOSCA-MP

FFG

The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreements no. 248138, "FascinatE – Format-Agnostic SCript-based INterAcTive Experience" (http://www.fascinate-project.eu/) and no. 287532, "TOSCA-MP – Task-oriented search and content annotation for media production" (http://www.tosca-mp.eu), as well as from the Austrian FIT-IT project "IV-ART – Intelligent Video Annotation and Retrieval Techniques".