JRS-VUT at TRECVID 2012: Semantic Indexing and Instance Search

Werner Bailer[1], Robert Sorschag[2], Harald Stiegler[1], Christoph Wiedner[1]

[1]JOANNEUM RESEARCH
Forschungsgesellschaft mbH
DIGITAL
Institute for Information and
Communication Technologies

Steyrergasse 17
8010 Graz, Austria

Tel. +43 316 876-1218
Fax +43 316 876-1191

werner.bailer@joanneum.at

digital@joanneum.at
www.joanneum.at/digital

[2]Vienna University of Technology
Interactive Media Systems Group

Semantic Indexing (SIN)

Features

- MPEG-7 features: Color Layout and EdgeHistogram
- Color-SIFT
  - 300 densely sampled image regions from 3 different scales
  - 384 dimensional Color-SIFT
  - bag-of-features histogram with 100 bins
  - global and local (2x2, 1x3, 3x1, 3x3 blocks)
- extracted from reference key frames and more densely samples key frames based on visual activity

Classifier

- SVM
- Multiple kernel learning (MKL) with equence-based kernels
- L1-norm MKL, using Shogun Toolbox
- Longest common subsequence (LCS) kernel, feature vector sequence per shot
- 35 subkernels:
  - 7 features (MPEG-7 Color Layout and Edge Histogram, 5 spatial configurations of Color-SIFT BoF histograms)
  - LCS kernel with $\vartheta_{sim} = \{0.10, 0.30, 0.50, 0.70, 0.90\}$

Results

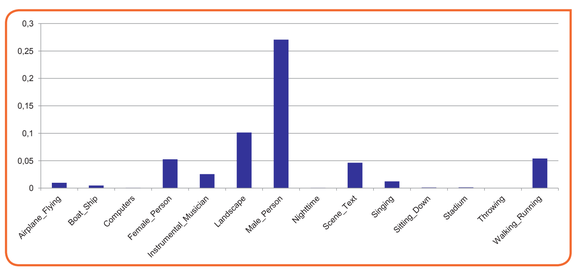

Figure 1: Results of the SIN light run (infAP).

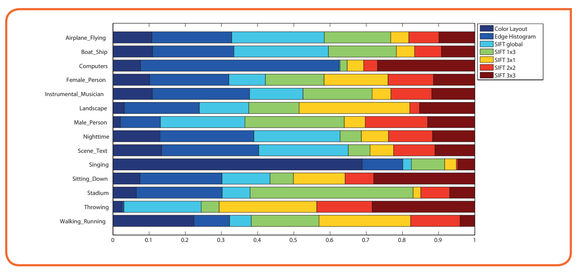

Figure 2: Weights of the subkernels of the MKL problem by different values of the similarity threshold ϑ_{sim} of the LCSS kernel.

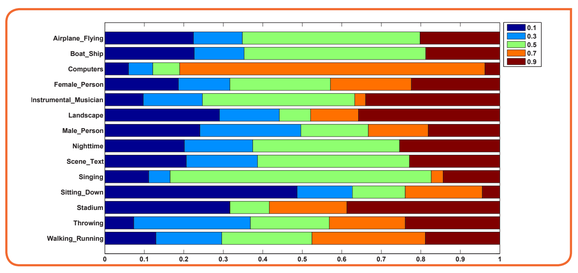

Figure 3: Weights of the subkernels of the MKL problem by different features.

Instance Search (INS)

Two Approaches

- Color-SIFT bag-of-features (cf. SIN)
  - 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

Results

- Indexed Color-SIFT
  - 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 in terest (e.g., low resolution query samples)
  - issues with not sufficiently discriminative feature points
- 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

Runs

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

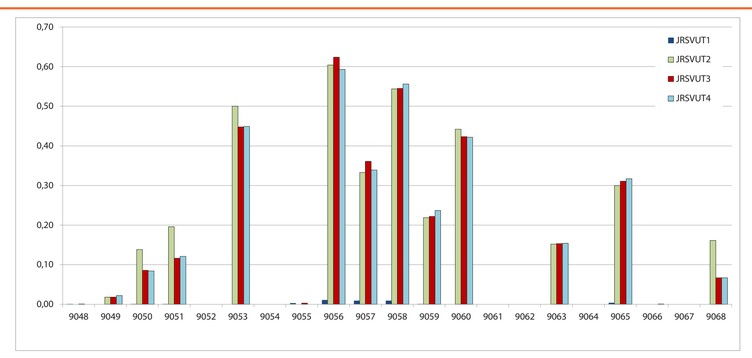

Figure 4: Average precision of the four INS runs.

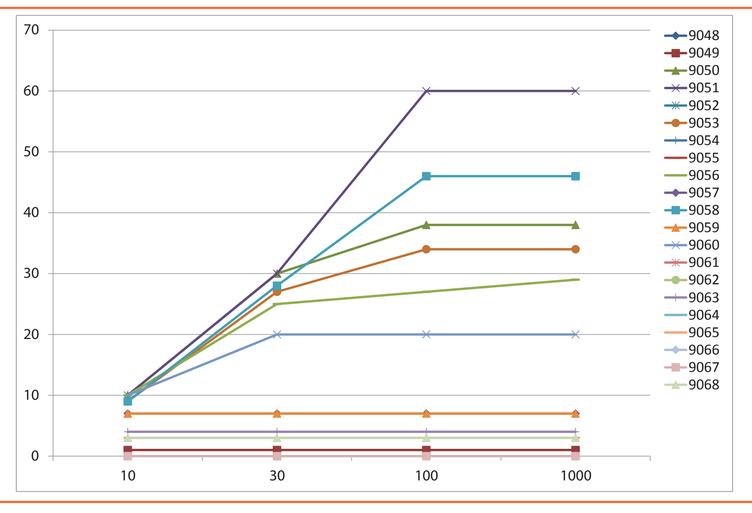

Figure 5: Number of hits found for run JRSVUT2 at rank 10, 30, 100 and 1000.

The research leading to these results has received funding from the European Union's Seventh Framework Programme under the grant agreements no. FP7-248138, "FascinatE – Format-Agnostic SCript-based INterAcTive Experience" (http://www.fascinate-project.eu/) and no. FP7-287532, "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".

diggos312012