ITI-CERTH participation to TRECVID 2015

Foteini Markatopoulou1,2, Anastasia Ioinnidou1, Christos Tzelepis1,2, Theodoros Mironidis1, Damiano Galanopoulos1, Stavros Arestis-Chartamplias1, Nikiforos Pittaras1, Konstantinos Avgerinakis1, Nikolaos Gkalelis1, Anastasia Mourtzidou1, Stefanos Vrochidis1, Vasilios Mezaris1, Ioannis Kompatsiaris1, Ioannis Patras2

1Information Technologies Institute (ITI), Centre for Research and Technology Hellas (CERTH), Thessaloniki, Greece
2Queen Mary University of London, Mile End Campus, UK

m.markatopoulou, ioananas, tzelepis, mironid, dgalanop, stav_ares, npittaras, kaofgeri, gkalelis, moumtzid, stefanos, bmezaris, ikom) @iti.gr

Semantic Indexing

- System overview: 1) Features based on pre-trained and fine-tuned deep convolutional neural networks (DCNN); 2) Dimensionality reduction using Kernel Subclass Discriminant Analysis (KSDA); 3) Cascade of classifiers using local and DCNN-based descriptors; 4) Multi-task learning (MTL) using the logistic-lasso algorithm; 5) Two-layer stacking architecture

- System A: Cascade of four stages: 11 visual descriptors, one LSVM per descriptor (3xORB, 3xSIFT, 3xSURF, 2xDCNN-based)
- System B: Cascade as in A, with features reduced to a much lower dimension using KSDA
- System C: One MTL model for each of the reduced descriptors in B; per-concept late fusion of the MTL models’ scores
- System D: Late fusion of 7 DCNN-based descriptors, using 5 LR models per descriptor; resulting in 35 LR models per concept

Multimedia Event Detection

In 010Ex and 100Ex:
- **KSDA+LSVM**: (c-1KDALSVM used visual, motion, fc7+fc8 DCNN descriptors); achieved better results; 2nd best among all competitors on 100Ex
- **RDKSVM**: (c-2RDKSVM used fc8, c-3RDKSVM used fc7+fc8); exploits near-miss videos; performance limited by descriptors

In 000Ex:
- 3 ELMs*6 CLMs*4 mat. op. = 72 detectors;
- Runs c-1, c-3 use our sole best; c-2, c-4 use top-10 average.

Surveillance Event Detection

Activity detection system stages:
- Offline Stage: Extraction of Dense Trajectories + HOGoHOF descriptors; Fisher encoding; LSVM model training
- Online Stage: a 50-frame sliding window along with pre-trained models for localizing the desired events inside test videos

5 Events of interest:
- Embrace, PeopleMeet, Pointing, PersonRuns, and PeopleSplitUp

Instance Search

- Participation with VERGE interactive video search engine
- VERGE retrieval and presentation modules
  - High Level Visual Concept Retrieval (Semantic Indexing)
  - Object-based Visual Search Module
    - Fast Hessian detector + SIFT descriptor for local feature extraction
    - BoW model: 2-layer Visual Vocabulary (100K) construction using Repeated Bisecting K-Means
    - Inverted Index implementation using Apache Lucene search engine library
    - Capability of querying the system either with whole frames or any cropped part of them
    - Support of Multiple Queries
    - Similarity Score: Borda Count based on tf-idf weights and ranking positions in the retrieved list
    - Best performance with fusion of baseline BoW + Saliency-based BoW models

KSDA demo software (for GPU): http://mklab.it.gr/project/gpu-agsda
Online demo for video concept detection: http://multimedia2.iti.gr/onlinetvideoanalysis/service/start.html

Supported by: