Multimedia Video Detection

010Ex and 100Ex
- Kernel Subclass Discriminant Analysis for dimensionality reduction and fast linear SVM for training (KSDA+LSVM)
- Both DCNN-based and motion features

DCNN-based features
- GoogLeNet trained with 12988 ImageNet concepts
- The best FT network on the 345 TRECVID SIN concepts (the same used for the AVS task)
- GoogLeNet trained with 500 EventNet concepts
- GoogLeNet trained with 5055 ImageNet concepts and FT on 487 Sports-1M concepts
- GoogLeNet trained with Places 205 concepts

Motion features
- HOG, HOF, MBBx, MBHy

Video Representation
- Concatenate motion and DCNN-based feature vectors
- New feature vector in SIGIR13
- KSDA drastically reduces the feature vector dimensionality

Conclusion
- Plentiful DCNN-based features lead to better results

000Ex
- An improved version of our MED15 zero-example event detection framework is used
- Extensive visual concept pool: up to 14k concepts
- Pseudo-relevance feedback using KSDA+LSVM and retrieved videos from the MED16-EvalSub set

DCNN-based features
- Direct output of 5 DCNNs (from 010Ex and 100Ex)
- Direct output of GoogLeNet trained with 5055 ImageNet concepts

Conclusion
- Pseudo-relevance feedback has a significant impact to our performance (the relative improvement is 16.39%)
- Using a large number of visual concepts gives a boost to the performance compared with our MED15 submission

Instance Search

VERGE video search engine

VERGE retrieval and presentation modules

High Level Visual Concept Retrieval
- 346 TRECVID SIN concepts using pre-trained DCNNs & training with Linear SVMs
- 205 Places scene concepts using GoogLeNet for landscape recognition

Visual Similarity Search module
- Use of pre-trained DCNN on 5055 ImageNet categories & selection of last pooling layer for keyframe representation
- Nearest Neighbour search realized using Asymmetric Distance Computation

Face detection and Face Retrieval Module
- Face detection: detect facial landmarks using cascade of DCNNs
- Face feature extraction: extract DCNN descriptors using VGG-Deep-16 architecture & use of last FC layer as feature vector
- Construction of an IVFADC index for fast face retrieval

Run ID | MAP | Recall
--- | --- | ---
Run 1 | 0.114 | 1000/11197

Future work
- Combination of the different modules for improving the video search results

Supported by: