TECHNISCHE UNIVERSITÄT CHEMNITZ

Faculty of Computer Science
Junior Professorship Media Computing & Chair Media Informatics
Marc Ritter, Manuel Heinzig, Robert Herms, Stefan Kahl, Daniel Richter, Robert Manthey, and Maximilian Eibl

Technische Universität Chemnitz at TRECVID Instance Search 2014

Introduction

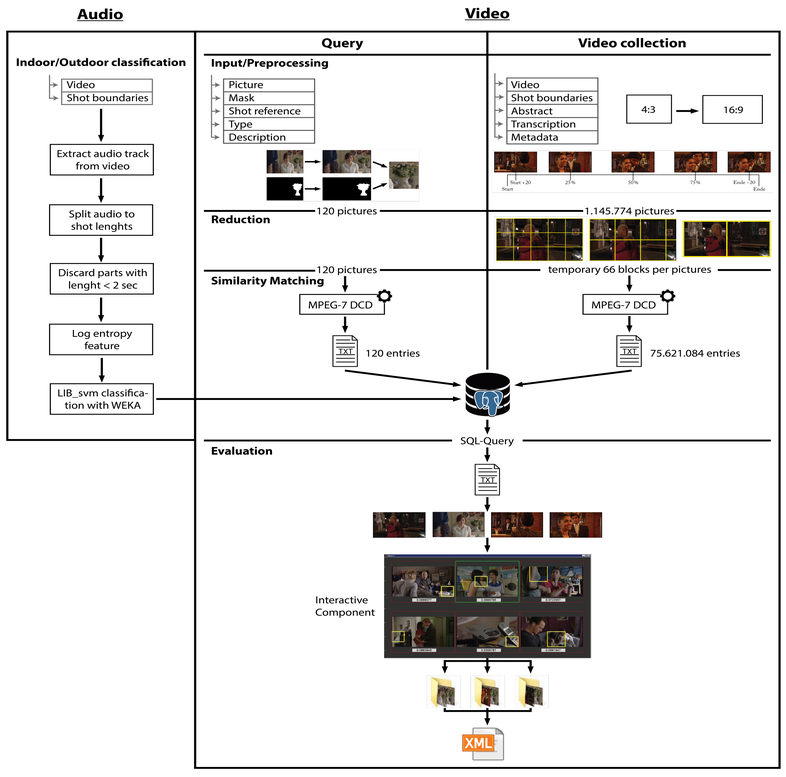
Motivation

- Automatic processing of large amount of video databases
- Object detection in images or videos
- Instance search in the framewok of the Text REtrieval Conference Video (TRECVid) 2014

Methods of Evaluation

1. Interactive run: MPEG-7 dominant color, audio-based indoor/outdoor detection and a semantic shot composition
2. Interactive run: Probabilistic Run-length weighted Neighborhood Algorithm (PRNA) combined with a semantic shot composition based on the advanced dominant color descriptor
1. Automatic run: PRNA in combination with the semantic shot composition
2. Automatic run: PRNA with a shot composition approach for similar shots within a specific environment using basic MPEG-7 descriptors

System Architecture



Algorithmic Approaches

Visual Feature Extraction

- Features extraction from full images as well as from the rectangular cells of a grid structure
- Extraction of three different scaled grids with resolutions of 240x144, 168x96 and 96x48 pixel
- Shifting the grid for half the size of a subsection once in vertical as well as in horizontal direction
- A total of 66 different blocks at multiple resolutions for each image and features extraction per block
- Extraction of a maximum of 8 dominating colors from a given image by the dominant color descriptor

Probabilistic Run-length weighted Neighborhood Algorithm (PRNA)

- Longer shots have a higher probability of containing a searched instance than shorter shots
- The probability of containing the searched instance in neighborhood shots is higher than shots in a far temporal distance
- Requirement to prior knowledge about some shots containing a searched instance (four shots of each topic)

Indoor/Outdoor Detection on Audio

- Low background noise in most of the indoor shots
- The indoor/outdoor classification of shots is based on the log-energy feature
- The indoor/outdoor classification on scene-level depends on the number of indoor shots
- If one shot meets the indoor requirements, all other shots in the same scene are indoor

Shot Composition and Advanced Dominant Color (ADC)

- ADC for finding semantic linked shots in direct temporal connection
- Grouping related shots to one scene
- Related shots which were filmed at the same location have a quite similar overall coloring
- ADC is based on one dominant color which is extracted from an image by k-means clustering
- 27 distinct final colors

Corpus

Database

- 464 hours video data at 768x576 pixel in 4:3 format
- 30 topics with 4 different occurrences in the corpus yielding to 120 query pictures
- Using the PostgreSQL as central database repository to process large amounts of data
- Calculation of distances between descriptors of the query and each block of the video corpus directly in the database via user defined functions and storage to avoid a repeated recalculation

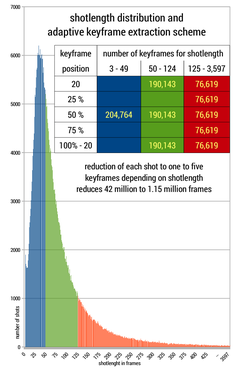
Preprocessing

- Splitting the video collection into the 471,526 shots
- Reduction image size to 456x256 pixel
- Creation of an audio version at 16 kHz and 16 bit PCM format
- Extraction of keyframes from each shot with a trivial selection scheme

Approachs

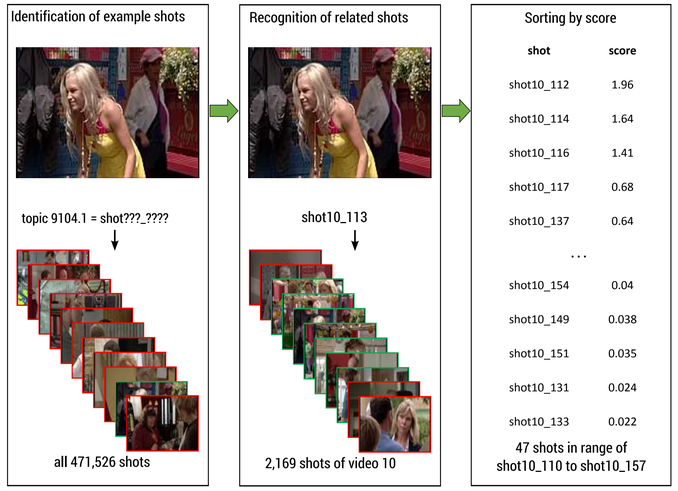
Interactive Runs

- Preprocessing of query data and video collection
- Extraction of visual descriptors at the position of the requested object
- Distinction between indoor and outdoor scenes using audio analysis
- Subdividing each frame by a multi-scale grid and feature extraction from each cell
- Calculation of distances between descriptors and an incoming query
- Evaluation with a GUI to assess the absence or presence of an object



Automatic Runs

- Using the PRNA
- Combination of PRNA with a shot composition and advanced dominant color



Results

Interactive Runs

1. Interactive run: Mean Average Precision (MAP) = 0.037
2. Interactive run: MAP = 0.034

Automatic Runs

1. Automatic run: MAP = 0.015
2. Automatic run: MAP = 0.017

Mediocre results placing us in the lower-middle tier in comparison to other participants