Content-based Video Genre Classification Using Multiple Cues

Hazım Kemal Ekenel, Tomas Semela, Rainer Stiefelhagen
ekenel@kit.edu

ACM Multimedia 2010 Workshop: AIEMPro10

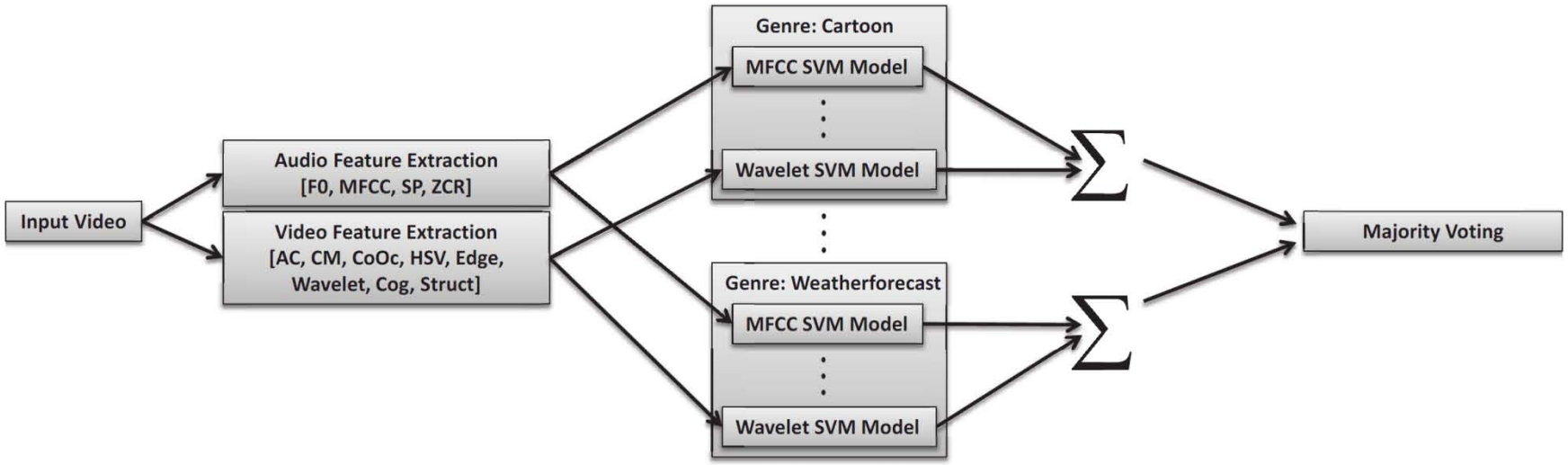
Institute of Anthropomatics, Department of Computer Science

Introduction

- Automatic video genre classification
  - An important task in multimedia indexing
  - The genre of the processed video could give cues about the available high-level features in the video

- A shared feature extraction framework for genre classification and high-level feature detection
  - Saves processing time
  - Genre and high-level feature information can be used to improve each system's performance further

KIT System



- Several feature extractors
- SVM classification
- Fusion of results with majority voting

KIT Feature components

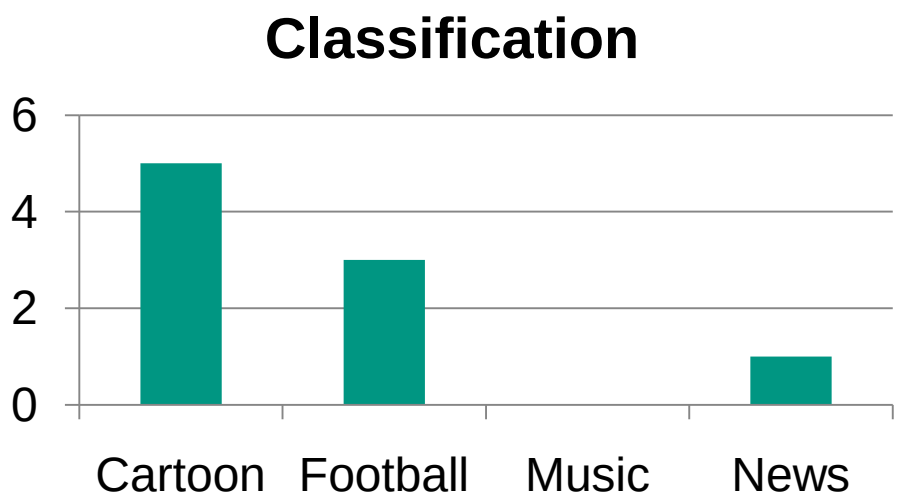
- Six low level visual features
  - Color descriptors
    - Color moments
    - HSV histogram
    - Color auto correlogram
  - Texture
    - Wavelet texture
    - Edge histogram
    - Co-occurrence texture
- Structural features
  - Structural-syntactic information, e.g. relationships between frames, shots and scenes (shot boundary detection)

KIT Feature components

- Cognitive features
  - Cognitive information extracted using a face detector
  - Average number of faces per frame
  - Distribution and location of the faces in the frame
- Four low level aural features
  - Computed on a mono-channel uncompressed audio signal
  - Using a 320 samples window and hamming window function
  - Mel-frequency cepstral coefficients (MFCCs)
    - 8th order
  - Fundamental Frequency
  - Signal Energy
  - Zero Crossing Rate

Classification

- Training of one support vector machine (SVM) for each genre and each feature
  - Binary classification (one vs. all others)
  - Using radial basis function (RBF) kernel
  - Cross-validation scheme to determine optimal function parameters
  - Fusion in decision level: majority voting



Predicted Genre: Cartoon

Evaluation datasets

Italian TV program		Youtube		Quaero 2009	
110 hours of complete TV programs		Complete programs and small clips (lower quality)		Complete programs	
Mainly same programs → less variety		Many different programs per genre		4 genres, same programs	
Resolution: 352x288		Resolution: 352x288		Resolution: 284x288	
Genre	# videos		# videos		# videos
Cartoon	27	Cartoon	60	Cartoon	-
Commercial	58	Commercial	60	Commercial	-
Football	22	Football	10	Football	3
Music	7	Music	60	Music	15
News	49	News	60	News	30
Talk show	39	Talk show	60	Talk show	27
Weather	60	Weather	60	Weather	-

Evaluation results

- 6-fold cross validation on Italian and Youtube data
- 3-fold cross validation on Quaero data

Dataset	Class. Rate
Italy	99.6%
Italy (low quality)	99.2%
Youtube	92.4%
Quaero	99 %

Confusion matrices

Italian data

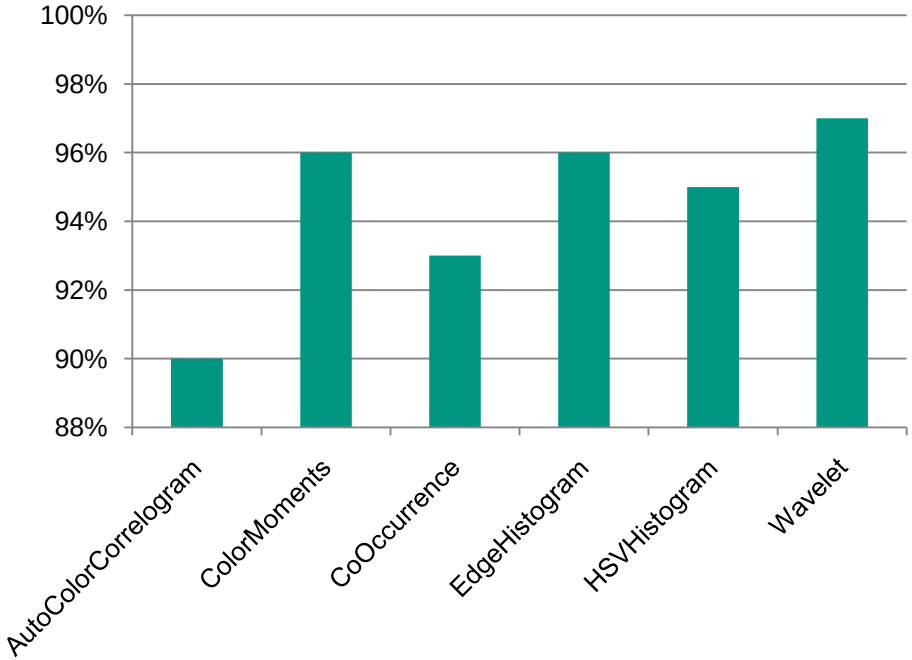
	Ca	Co	Fo	Mu	Ne	Ta	We
Ca	100	0	0	0	0	0	0
Co	0	100	0	0	0	0	0
Fo	0	0	100	0	0	0	0
Mu	0	14.2	0	85.7	0	0	0
Ne	0	0	0	0	100	0	0
Ta	0	0	0	0	0	100	0
We	0	0	0	0	0	0	100

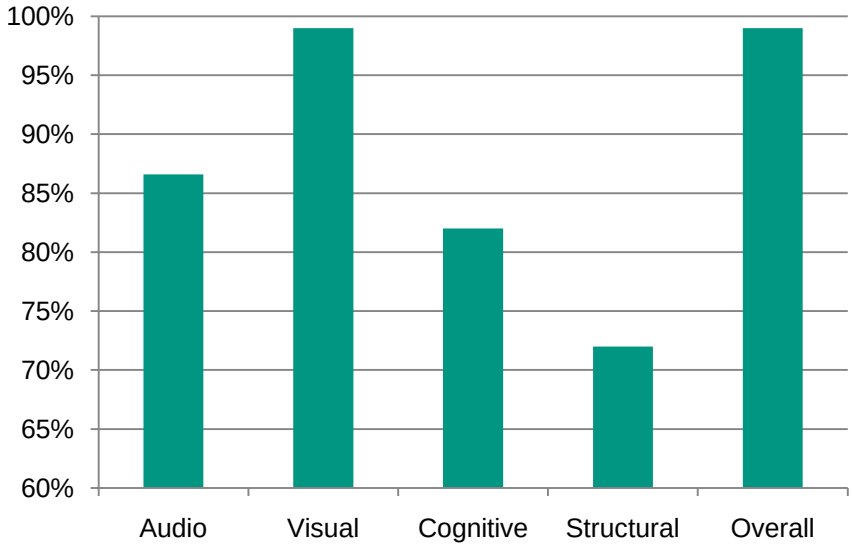
Youtube data

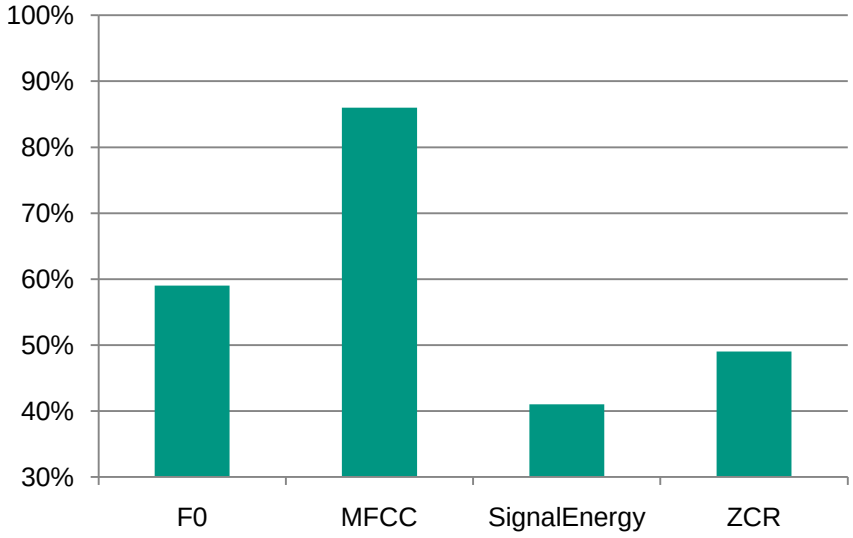
	Ca	Co	Fo	Mu	Ne	Ta	We
Ca	100	0	0	0	0	0	0
Co	0	95.0	3.3	1.6	0	0	0
Fo	0	10	80	10	0	0	0
Mu	5	8.3	0	85.0	0	1.6	0
Ne	0	1.6	0	0	90.0	6.6	1.6
Ta	0	0	0	5	1.6	93.3	0
We	3.3	0	0	0	0	3.3	93.3

Contribution of individual features

- Italian TV program
  - Visual best results
    - 99% alone
    - Wavelet best feature
  - Audio MFCCs attain 86%

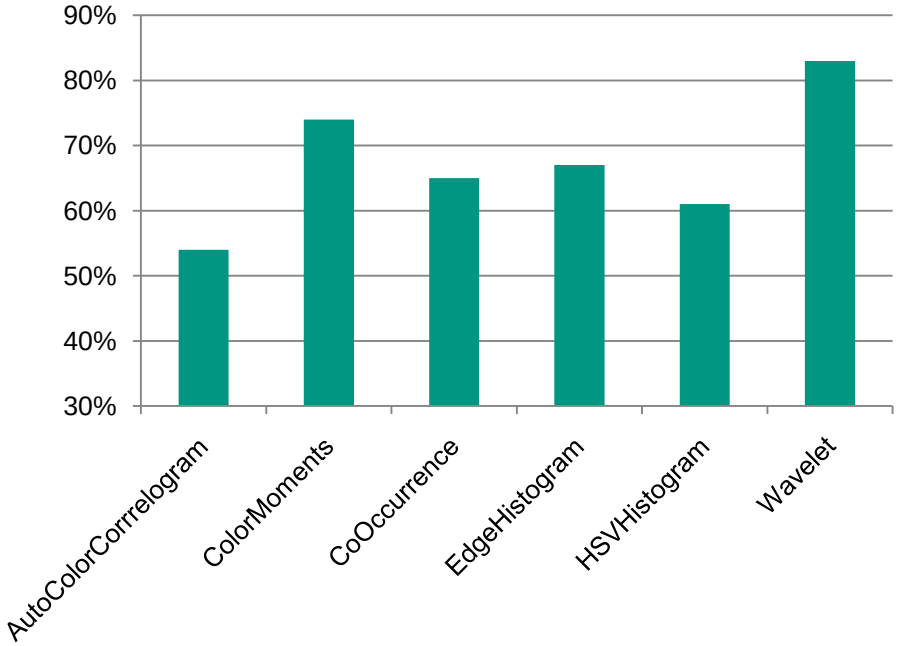


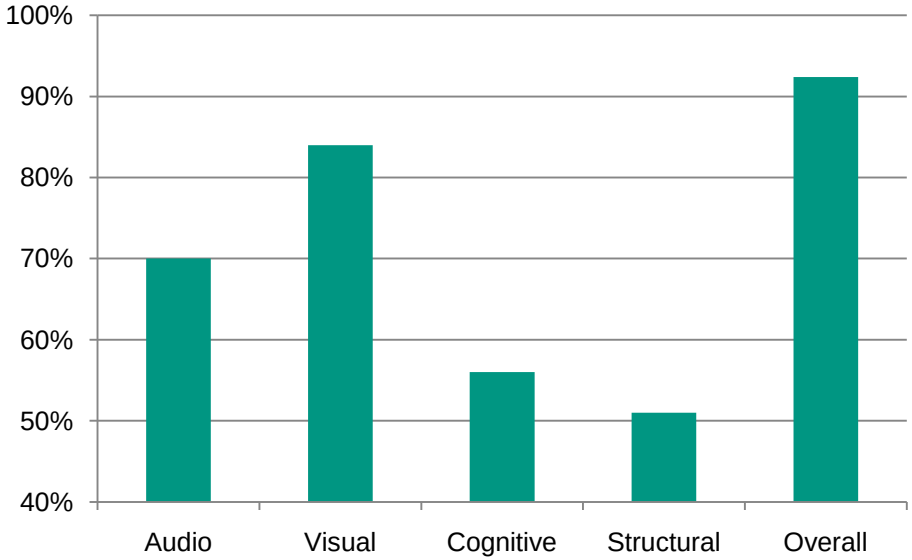


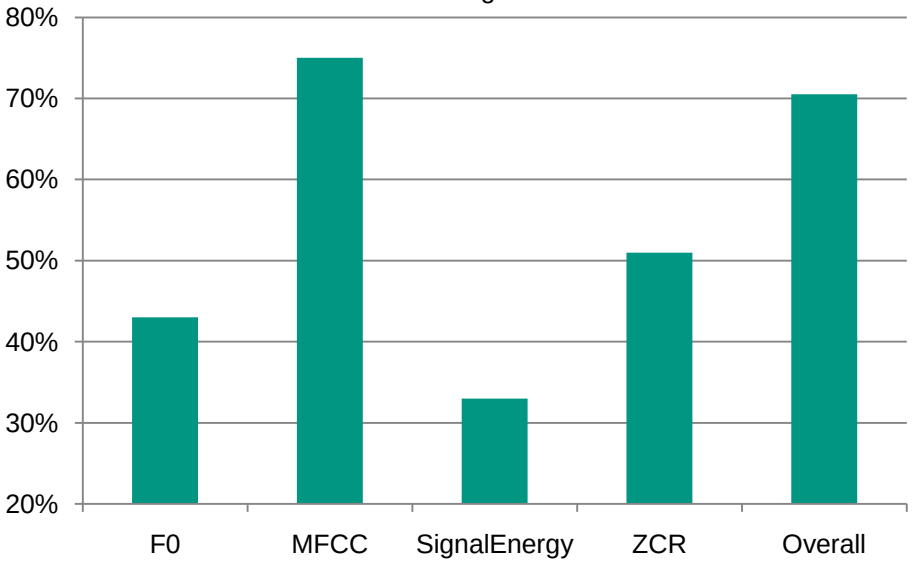


Contribution of individual features

- Youtube data
  - Visual features best 84%
    - Color moments and Wavelet
  - Audio second
    - MFCCs over 70%







Conclusion & Feature Work

- New visual features and use of SVM classifiers provide robustness
  - Visual features with color moments and wavelet texture work best
  - MFCC is the best audio feature
- Video quality has less impact on prediction, while variety in program degenerates the performance more (see Youtube results)
- More genres and more variety in later evaluations
  - Quaero 2010: Cartoon, Commercial, Documentary, Magazine, Movie, News, Show-Games, Talk show, Traffic, Weather
  - Youtube Categories: Entertainment (Comedy, Film, Music, Games), News, People, Sports
- Investigate temporal fusion techniques
- Utilizing document information (title, tags, description) for web videos

Thank you for your attention

Questions ??