

Feature Extraction Tools

Robbie De Sutter

Medianet Vlaanderen Workshop 8

Overview

- ◆ What are Feature Extraction Tools & Why are they “cost-efficient”?
- ◆ (Lot’s of) Examples & (Some) Demo’s
- ◆ In Depth: Face Recognition
- ◆ Some (Commercial) Available Tools
- ◆ Ready for Production? (aka Issues / Problems / Bottlenecks)
- ◆ EBU to the Rescue
- ◆ Conclusions

What

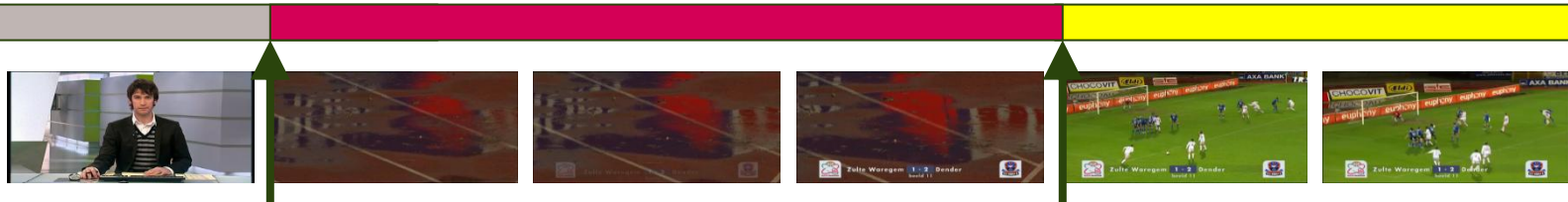
- ◆ Type of Information Extraction tool “*whose goal is to **automatically extract structure information** from unstructured **machine-readable documents**”*”
 - Such as video files, audio files
- ◆ Analyze the bits & bytes of media data
- ◆ Generate information about the content
- ◆ E.g., shot-segmentation, face detection, face recognition, speech-to-text, text-recognition, audio classification, scene segmentation, ...

Why

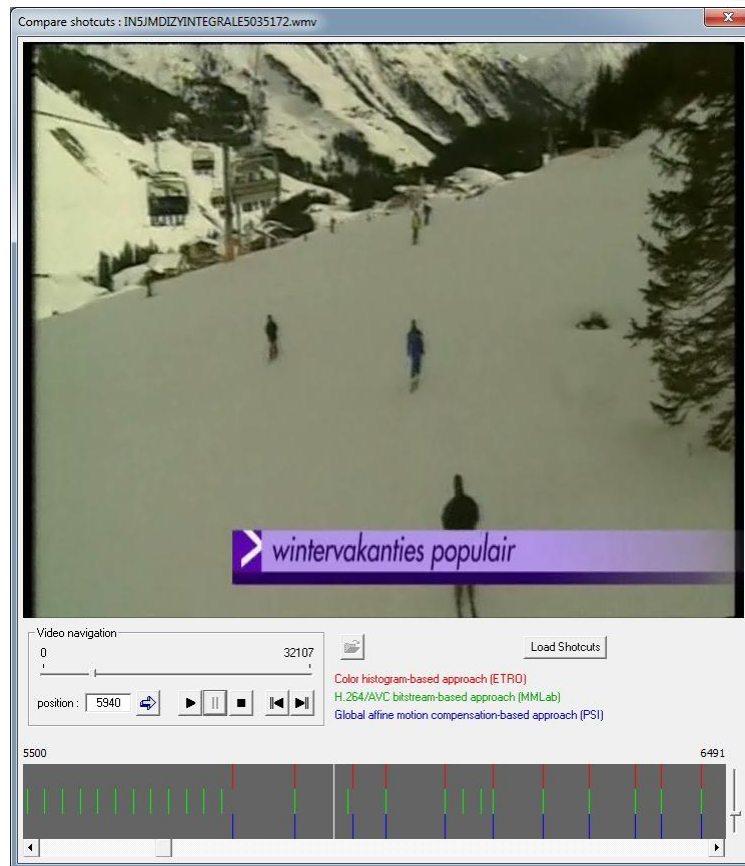
- ◆ Speeds up the annotation process
- ◆ Reliefs archivist of repetitive, boring tasks
- ◆ Freeing up time to add more high-value information about the content
- ◆ Automated, Scalable, Pipe-line concatenation
- ◆ Integration in production process: anywhere
 - Information is available when desired/required
 - Not at the end of the production chain (classical archival task)

Example – Shot cut detection

- ◆ Split the video stream in atomic units, corresponding to a continuously moving camera
- ◆ Distinguish between *abrupt* and *smooth* shot cuts
- ◆ Different methods
 - Using color histograms
 - Using affine motion compensation
 - Using motion estimation within the compressed domain



Demo – Shot cut detection



Courtesy:
KULeuven – ESAT
VUB – ETRO
UGent – MMLab

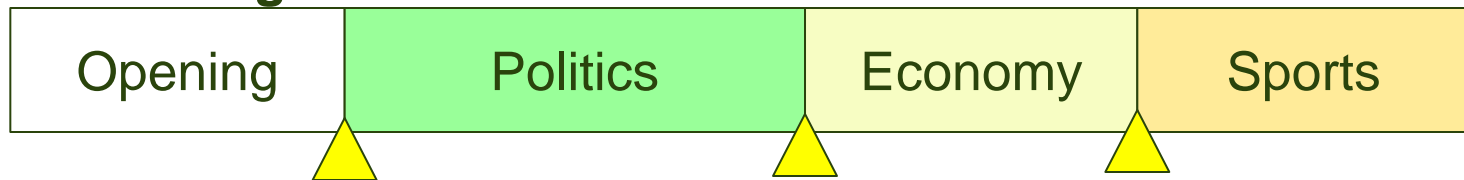
Example – Scene segmentation



Example – Subject Classification

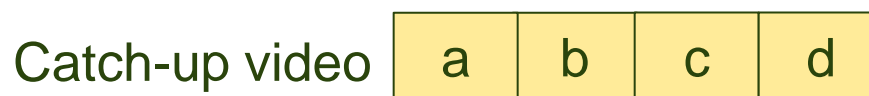
- ◆ Important for structuring content database
- ◆ In the case of News Program
 - Find story boundary
 - Mainly based on texts, sometimes by image characteristics
 - Categorize each story
 - Use machine learning, i.e., SVM etc.

News Program



Example – Content Summarization

- ◆ Automatic Summarization is an essential function.
 - The Catch-up (Digest) contains important results/facts.
 - Trailer is another summary type not containing results.
 - Useful for managing huge amount of raw materials (footage).
- ◆ Detection of important portion & redundant portion
- ◆ Generate target summary



Example – Audio Classification

- ◆ Basic operation like shot detection in visual
 - “silence,” “speech,” “music,” “noise”
- ◆ Detect audio portion
- ◆ Classify them into several category

TV Program



music

speech

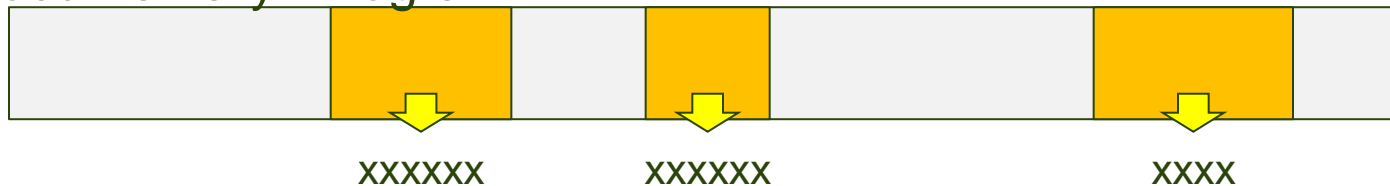
noise

The rest portions are silent portions.

Example – ASR / Speech-to-Text

- ◆ ASR = automated speech recognition aka Speech-to-Text
- ◆ Transcribing voice to text
- ◆ Requires segmented audio (only “speech” parts)

Documentary Program



- ◆ **Demo:** Nuance Dragon Natural Speaking

Example – ASR (CC) + NLP

- ◆ Results of ASR (automated speech recognition) or closed captions are the most viable resource from which semantic information can be extracted.
- ◆ Combining with NLP (Natural Language Processing) may provide more precise meaning.

Documentary Program



ASR(CC) >

xxxx xx xxx xxxxxxxx xx xxxx xxxx xx xxxxx xxxxxx xx



Animal appears

Animal eats

Animal natural enemy

Example – OCR in Video

- ◆ Text in Video is useful information as indices
 - It appears at some positions with various fonts and duration.
 - Mainly person name, organization name, place name, time etc.



- Detection area
- Recognized characters
 - > “Tengu-dake”
- Classify into category
 - > Object > Mountain

Example – Face detection

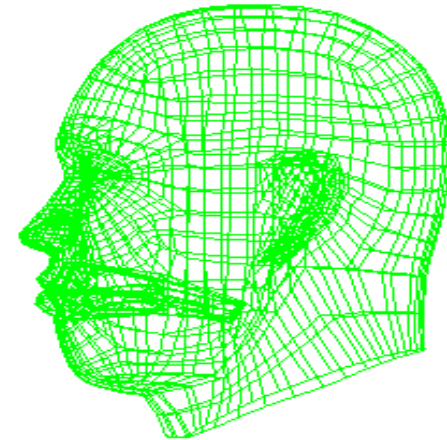


Courtesy – KULeuven - ESAT

Example – Face recognition



Courtesy – KULeuven - ESAT



Example – Compound Application



Logo = "één"

Face = "Goedele Devroy"

Logo = "vrt"

OCR = "Live", "Brussel",
"Goedele Devroy"



Music

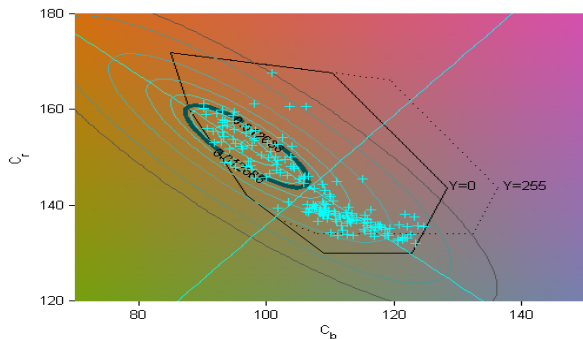
Speech

"Ik sta in Brussel te wachten maar..."

Overview

- ◆ What are Feature Extraction Tools & Why are they “cost-efficient”?
- ◆ (Lot’s of) Examples & (Some) Demo’s
- ◆ In Depth: Face Recognition
- ◆ Some (Commercial) Available Tools
- ◆ Ready for Production? (aka Issues / Problems / Bottlenecks)
- ◆ EBU to the Rescue
- ◆ Conclusions

Face Detection



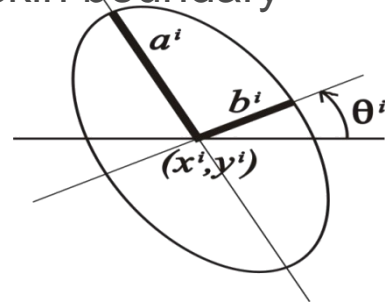
Face candidates selection:

- Candidate regions have skin color \rightarrow region-based skin segmentation
- Personalized chrominance skin boundary



Verification based on cues:

- Shape of ellipse
- Ellipse-filling percentage
- Gray-tone smoothness
- Corners of the facial features



Face detection results

◆ 92% good face detections

- Comparable to state-of-the-art *face learning* of 'Viola & Jones' obtains 93%
- adaptation to lighting conditions and personal face looks



Courtesy – KULeuven - ESAT

Face recognition

- ◆ Based on a 3D morphable model
- ◆ 3D model is fitted to 2D image
- ◆ Shape and texture parameters used as face descriptor
- ◆ Robust to
 - viewpoint changes,
 - illumination changes,
 - partial occlusions.



Courtesy – KULeuven – ESAT

Face recognition



Courtesy – KULeuven - ESAT

Face recognition



Courtesy – KULeuven - ESAT

Face recognition



Courtesy – KULeuven - ESAT

Face recognition



Courtesy – KULeuven - ESAT

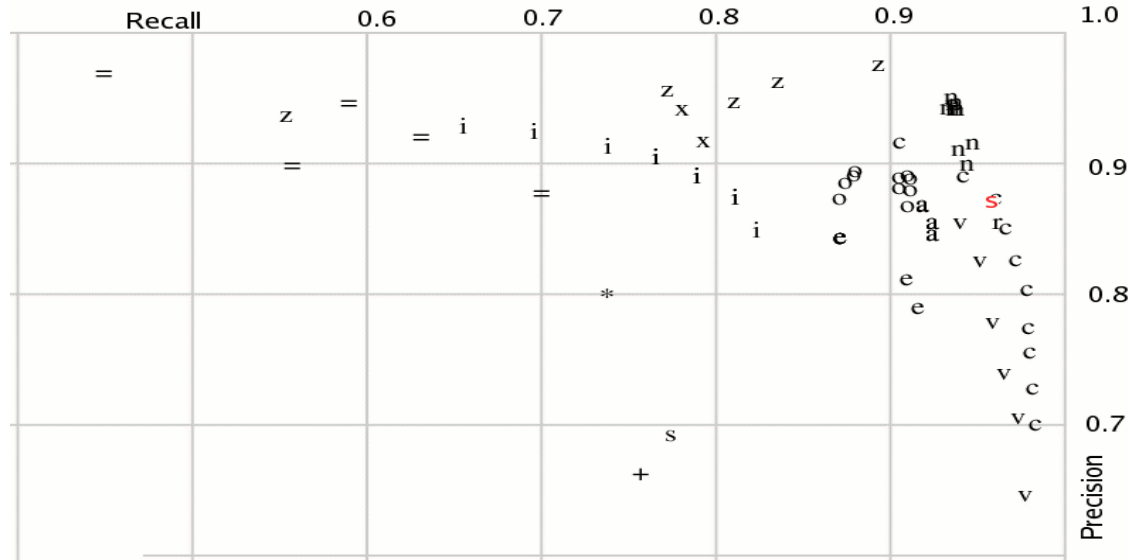
Face recognition



Courtesy – KULeuven - ESAT

Evaluation

- ◆ Recall: $R = \kappa(\text{correct}) / (\kappa(\text{correct}) + \kappa(\text{missed}))$
- ◆ Precision: $P = \kappa(\text{correct}) / (\kappa(\text{correct}) + \kappa(\text{incorrect}))$



Where can I “buy” them?

- ◆ Mostly academic tools
 - Gap-closing initiatives: EBU EC-M SCAIE, NIST TRECVID
- ◆ Some commercial tools starting to appear
- ◆ BUT small companies with limited solutions
 - “black box” approach
 - Limited w.r.t. input formats (usually no broadcaster formats)
 - Isolated Islands
 - Not pipe-line / workflow friendly (no SOA)
 - Not using/incorporating results from other tools

Commercial Available Tools

◆ Cluster Media.Labs

- ◆ Spin-off
- ◆ Speaker recognition
- ◆ Music search & retrieval (“query by humming”)

◆ Pictron

- ◆ Scene detection and relevant key frame extraction
- ◆ Text extraction (subtitles) from video
- ◆ Speech recognition and face recognition

Commercial Available Tools

◆ Autonomy

- Virage
- Audiovisual analysis tools
- OCR + capturing of subtitles from teletext
- Face recognition



◆ Memnon

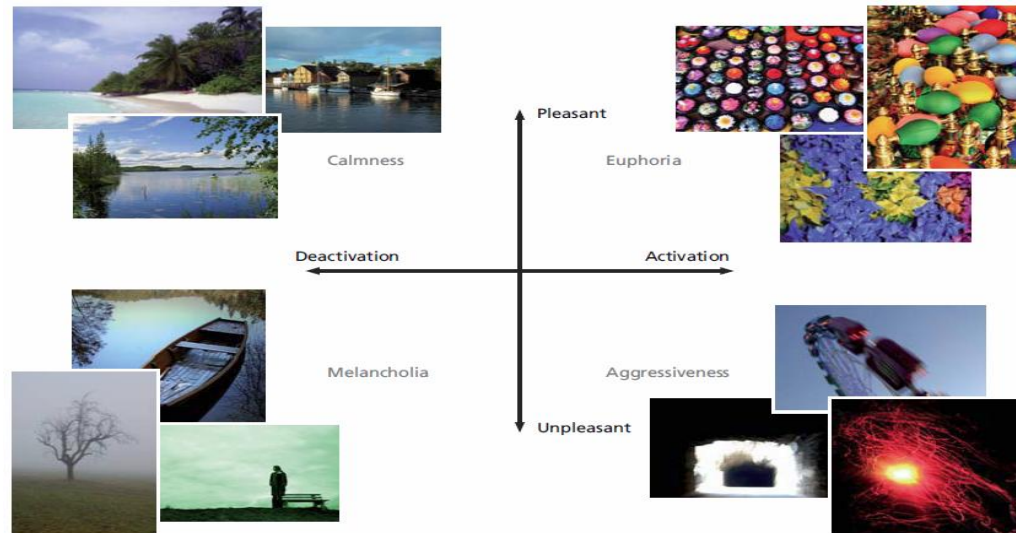
- Audio analysis tools
- Core component: Nuance
- Speech recognition (supports Flemish)
- Audio segmentation (speech/music/silence), ...



Commercial Available Tools

◆ Fraunhofer

- Mood-player
 - GenreID music analysis: automatic genre recognition
 - Mood detection (audio & video)



Commercial Available Tools

◆ iPharro

- Spin-off Fraunhofer Institute
- Copy-detection tools via Fingerprinting
 - “a video fingerprint is an electronic summary of all of the unique characteristics of a piece of video.”
 - “this includes, but is not limited to, looking at the colors and screen layout and how this changes over time.”
- Use case:
 - Commercial identification / “did my ad aired on time?”
 - Televised DRM, content monitoring
 - Internet Tracking and Monitoring
 - Archive or Database Search / Organize



Overview

- ◆ What are Feature Extraction Tools & Why are they “cost-efficient”?
- ◆ (Lot’s of) Examples & (Some) Demo’s
- ◆ In Depth: Face Recognition
- ◆ Some (Commercial) Available Tools
- ◆ Ready for Production? (aka Issues / Problems / Bottlenecks)
- ◆ EBU to the Rescue
- ◆ Conclusions

Ready for Production?

- ◆ Currently mainly academic tools;
Some (expensive/limited) commercial tools available
- ◆ Mainly “low-level” feature extractions
- ◆ Tools not optimized for the requirements of broadcasters
- ◆ How to evaluate quality of the tools?
 - How reliable is a specific tool from a specific vendor?
 - How to train the tool & required effort?
- ◆ How to integrate tools in the production workflow?

EBU to the Rescue

- ◆ EBU Workgroup EC-M SCAIE (formerly P/SCAIE)
 - *Study and evaluation of Content Analysis-based Automatic Information Extraction tools in media production*
 - Since 2007
 - RAI, VRT, IRT, DR, NHK, VGTRK, ERT
- ◆ Subgroup of EBU EC-M
 - Expert Community on Metadata
 - <http://tech.ebu.ch/groups/ecm>
 - Open for **EVERYONE** (even non-EBU-members!)



Goals EC-M SCAIE

- ◆ Platform broadcasters – (academic) researchers
 - Proclaim the real broadcaster requirements
 - Request for Technologies
 - AIEMPro workshops (3rd edition: ACM Multimedia 2010 – Italy)
 - Provide genuine broadcast quality media material for test / evaluation
 - Media Exchange Platform (aka Mammie)
- ◆ Design Data Model to Store FE Tool Results
 - Slimmed down MPEG-7 profile – ISO standard
 - Interface – Interoperability – Pipe-lining
- ◆ Provide Test Guidelines

Media Exchange Library

- ◆ Objective: collection of relevant material w.r.t. relevant applications
 - Target: to have several hours of material in good quality for each of the identified applications, possibly coming from a wide group of members
- ◆ Ground truth annotation
 - to be performed/submitted by EBU members or by the public
- ◆ Public access, conditioned to scientific use
 - An usage contract is available and agreed by EC-M SCAIE members
 - Several users registered to download the content
- ◆ Currently:
 - 2.7 TB audiovisual material (online)
 - High res & Browse proxy

Media Exchange Library

EBU EC-M Scaie Media Exchange Platform

Create New
Log Out
Help

Explore all

Free search

search

Filter by Contributor


Filter by Task Set


Filter by Available Ground Truth


- ITEM SEGMENTATION 48
- SPEECHRECOGNITION 45
- FACEDETECTION 3


Showing items 1 - 10 of 48 found matching *.* and matching (DS_memberOfCollection:"ItemSegmentation")

Click an item's thumbnail to play the content or click the title to view the details

JSB20030710.mpg

Added by NHK on 06 November 2009

JSB20030612.mpg

Added by NHK on 06 November 2009

JSB20030605.mpg

Added by NHK on 06 November 2009

JSB20030731.mpg


EBU EC-M Scaie Media Exchange Platform

Create New
Log Out
Help

Explore all

Free search

search

Filter by Contributor

- NHK 44
- VRT 4


Filter by Task Set

Filter by Available Ground Truth

- ITEM SEGMENTATION 4
- FACEDETECTION 3
- SPEECHRECOGNITION 1

Antartica
Added by you on 06 November 2009
[delete item](#) | [back to overview](#)

ID: ifip:049ad660-ad10-012c-e5fd-00145e2a38b6



contributor:
VRT

coverage:
(click to enter coverage)

creator:
(click to enter creator)

date:
2007-12-15



format:
(click to enter format)

language:
Nederlands

relation:
(click to enter relation)

Ground Truth

Below you see an overview of the ECM Scaie tasks. To the left, you can add or remove an item from the train or test set of a particular task. To the right, you can add or edit the ground truth data of the various tasks.

 ITEM SEGMENTATION use for train: remove use for test: add	delete source edit Ground Truth Metadata
 FACE DETECTION use for train: add use for test: add	No Ground Truth available for this task. add

<http://media.ibbt.be>

Conclusions

- ◆ Automation can help in reducing the related costs
- ◆ Research is active and moving since several years
 - Impressive results are available, but consistency is lacking
- ◆ Broadcaster requirements are not yet fully understood by researchers
- ◆ EBU helps
- ◆ You too can help:
 - join the EC-M group
 - follow the EC-M SCAIE activities
- ◆ It's time to think on how these tools can be embedded in our production infrastructures

References

◆ EC-M SCAIE web pages on tech.ebu.ch

- <http://tech.ebu.ch/groups/ecm>
- <http://tech.ebu.ch/groups/pscaie>
- http://tech.ebu.ch/pscaie_rfts



JOIN THIS GROUP

◆ Media Exchange Library / MAMMIE system

- <http://media.ibbt.be/>

◆ AIEMPro workshop series

- 2010: <http://www.crit.rai.it/attivita/AIEMPro10/index.html>
- 2009: <http://wiamis2009.qmul.net/program.php#6>
- 2008: <http://www.crit.rai.it/AIEMPro08/>