Automation and standardization of semantic video annotations for large-scale empirical film studies

SWIB 2018

Henning Agt-Rickauer / Christian Hentschel / Harald Sack
Hasso Plattner Institute, University of Potsdam, Germany
Analyzing Audio-Visual Rhetorics of Affect

- empirical research on audio-visual rhetorics by means of film analysis
  - film scientist from FU Berlin
  - computer scientists from HPI, Université de Nantes

- guiding research question/project goals:
  - How do audio-visual images shape emotional attitudes towards certain topics?
  - identifying an initial set of audio-visual rhetorical figures (typology)
  - developing computational methods for the study of audio-visual rhetorics

- subject matter:
  - feature films, documentaries and tv news reports on the global financial crisis (2007-), total: >100h
Motivation

- identification, localization and classification of audio-visual staging patterns
- many annotations necessary for a scientific and holistic understanding of a movie
- technological requirements
  a. consistent data management
  b. support for semi-automatic annotation data generation
Linked Open Data - consistent data management
AdA Ontology - Motivation

eMAEX annotation routine
- Film-analytical method
- Systematic: categories, types, values
- ...but not machine-readable

Free annotations
- Natural language
- Typos
- Synonyms (medium shot vs. waist shot)
- Spelling (colour range vs. color range)

Goal
- Reusable, explicit vocabulary with film-analytical concepts, terms and descriptions
- Accessible on the Web
- Integrate into video annotation software Advene
Unique identifiers for domain-specific concepts and terms

- Uniform Resource Identifier (URI)
  - http://ada.filmontology.org/resource/2018/09/25/AnnotationType/FieldSize

- Store information and make it retrievable
- encoded with RDF
AdA Ontology - Vocabulary Visualization Demo

Annotation Vocabulary
- 9 Annotation Level
- 78 Annotation Types
- 435 Annotation Values

Download at https://github.com/ProjectAdA/public

http://ada.filmontology.org/ontoviz/

Automation and standardization of semantic video annotations for large-scale empirical film studies

Christian Hentschel
AdA Ontology - Example Annotation

“And this is wrong!”

“I’m late for a meeting.”

Body Language
Emotion: tensioned

Camera Movement
Type: tracking shot

Camera Movement Speed: fast → slow → static

Light Contrast: high

Chart 8
example: Company Men
- More than 24,000 annotations, mostly manual

Goal
- Publish this valuable data by means of Linked Data

How
- Advene RDF Export
- AdA Ontology Data Model
- W3C Web Annotation Standard, Media Fragments URI

Make Linked Data Usable
- Visual Analysis
- Queries

<table>
<thead>
<tr>
<th>annotate</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://ada.filmontology.org/resource/2018/09/25/AnnotationType/ColourRange">http://ada.filmontology.org/resource/2018/09/25/AnnotationType/ColourRange</a></td>
<td>3041</td>
</tr>
<tr>
<td><a href="http://ada.filmontology.org/resource/2018/09/25/AnnotationType/ImageContent">http://ada.filmontology.org/resource/2018/09/25/AnnotationType/ImageContent</a></td>
<td>1383</td>
</tr>
<tr>
<td><a href="http://ada.filmontology.org/resource/2018/09/25/AnnotationType/DominantMovementDirection">http://ada.filmontology.org/resource/2018/09/25/AnnotationType/DominantMovementDirection</a></td>
<td>1288</td>
</tr>
<tr>
<td><a href="http://ada.filmontology.org/resource/2018/09/25/AnnotationType/FieldSize">http://ada.filmontology.org/resource/2018/09/25/AnnotationType/FieldSize</a></td>
<td>1086</td>
</tr>
<tr>
<td><a href="http://ada.filmontology.org/resource/2018/09/25/AnnotationType/CameraAngle">http://ada.filmontology.org/resource/2018/09/25/AnnotationType/CameraAngle</a></td>
<td>1051</td>
</tr>
<tr>
<td><a href="http://ada.filmontology.org/resource/2018/09/25/AnnotationType/CameraMovementType">http://ada.filmontology.org/resource/2018/09/25/AnnotationType/CameraMovementType</a></td>
<td>1019</td>
</tr>
<tr>
<td><a href="http://ada.filmontology.org/resource/2018/09/25/AnnotationType/ImageIntrinsicMovement">http://ada.filmontology.org/resource/2018/09/25/AnnotationType/ImageIntrinsicMovement</a></td>
<td>937</td>
</tr>
<tr>
<td><a href="http://ada.filmontology.org/resource/2018/09/25/AnnotationType/CameraAngleVerticalPositioning">http://ada.filmontology.org/resource/2018/09/25/AnnotationType/CameraAngleVerticalPositioning</a></td>
<td>881</td>
</tr>
</tbody>
</table>
Motivation
■ Huge amount of annotations
■ How to find interesting parts / patterns?

Goals
■ Search and retrieve segments with same characteristics
■ Within a movie and across movies

Annotation Query

Movie 1

BodyLanguageIntensity: 5
ImageContent: Group

Movie 2

BodyLanguageIntensity: 5
ImageContent: Group

Automation and standardization of semantic video annotations for large-scale empirical film studies
Christian Hentschel
Annotation Query - Demo

Automation and standardization of semantic video annotations for large-scale empirical film studies

Christian Hentschel

Chart 11

http://ada.filmontology.org/annotations/
Automated Multimedia Analysis - support for semi-automatic annotation data generation
Automated Multimedia Analysis

- huge amounts of annotations
  - Company Men: more than 24,000
  - labor intense: 3 mins of video $\rightarrow$ 10-12h of manual annotation
  - error-prone

- make a computer able to summarize the contents of video
  - (to some syntactical extend)
  - by extracting low-level features
  - increase the speed of video annotation

- two modalities:
  - audio stream
  - video stream
Automated Multimedia Analysis

Examples:
- Montage/ShotDuration
- ImageComposition/ColourRange
- Language/DiscourseText
Automated Multimedia Analysis

Montage/Shot Duration
Duration of a shot. A Shot of a film is a perceivable continuous image and is bound by a discontinuation of the whole composition.
Automated Multimedia Analysis

<table>
<thead>
<tr>
<th>Segment</th>
<th>Structural Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenes</td>
<td></td>
</tr>
<tr>
<td>Shots</td>
<td></td>
</tr>
<tr>
<td>Subshots</td>
<td></td>
</tr>
<tr>
<td>Frames</td>
<td></td>
</tr>
<tr>
<td>Keyframes</td>
<td></td>
</tr>
</tbody>
</table>
Automated Multimedia Analysis

- Example: Shot-Detection
  - Uses differences in consecutive images to identify discontinuities
    - idea: high visual redundancy in video stream

- Type of cuts:
  - hard-cuts
  - soft-cuts (fade-in, fade-out, wipe)
  - should be robust to artifacts (e.g., dropouts)
Automated Multimedia Analysis

ImageComposition/ColorRange

Simplified notation of the color range that is used in a sequence. For the purpose of comparability colors have to be picked from a reduced set of colors.
Automated Multimedia Analysis

- quantize all colors in a shot according to their most similar color from palette
  - compute Euclidean distance between color values of palette and frames
  - find NN
Automated Multimedia Analysis

- quantize according to CIE L*a*b*
  - color model according to human perception
  - separates chroma from lightness
  - Euclidean distance between color values similar to perceived color differences

black, white, wheat1, gold, saddlebrown, khaki, blue

'black': 0.63,
'dimgrey': 0.21,
'saddlebrown': 0.06,
'silver': 0.05
Automated Multimedia Analysis

Language/DialogueText

Dialogue is a transcription of understandable, spoken language that is dominant within the film. This is usually dialogue from protagonists, off-commentary, but also chorus. Nonverbal utterances (e.g. laughing, coughing, stuttering) will not be transcribed in this basic version.
Automatic Multimedia Analysis

- Automatic Speech Recognition (ASR)
  - subtitles?
Automated Multimedia Analysis - ASR

- based on supervised machine learning
  - requires (large) corpus of manually transcribed speech

- 2 stage approach
  1. acoustic model
     - convolutional neural network that transcribes utterances to letters
     - trained on ~1000 hours of audiobook recordings (LibriSpeech)
  2. language model
     - domain specific mapping of letters to words
     - based on word/letter co-occurrences
we review and some five hundred projects and programmes focusing on those with significant marketing opportunities song everything home contribute immediately to profitability selecting thirty set as promising to teach it growth program setting aside the rest for future consideration as taking more than miss mclarry you were talking earlier about fiscal two thousand eleven and a good job of convincing us that the credit markets frozen your sales revenue in two thousand ten love grey overs but can you talk about two thousand eleven what sort of a percentage increase you anticipate talk your people do our people who are you suggesting that you are expecting any gross in your division extent arm suggesting that we face increased foreign competition and a difficult credit market for large capital expenditures like no growth and two thousand eleven i am confident that while shipbuilding will remain challenge
Thank you for your attention!

Henning Agt-Rickauer, Christian Hentschel, Harald Sack

Hasso Plattner Institute, University of Potsdam, Germany