Multimedia Databases
Media - Text, Video, Audio

Prof. (FH) PD Dr. Mario Döller

Table of Contents

Media – Text, Video, Audio

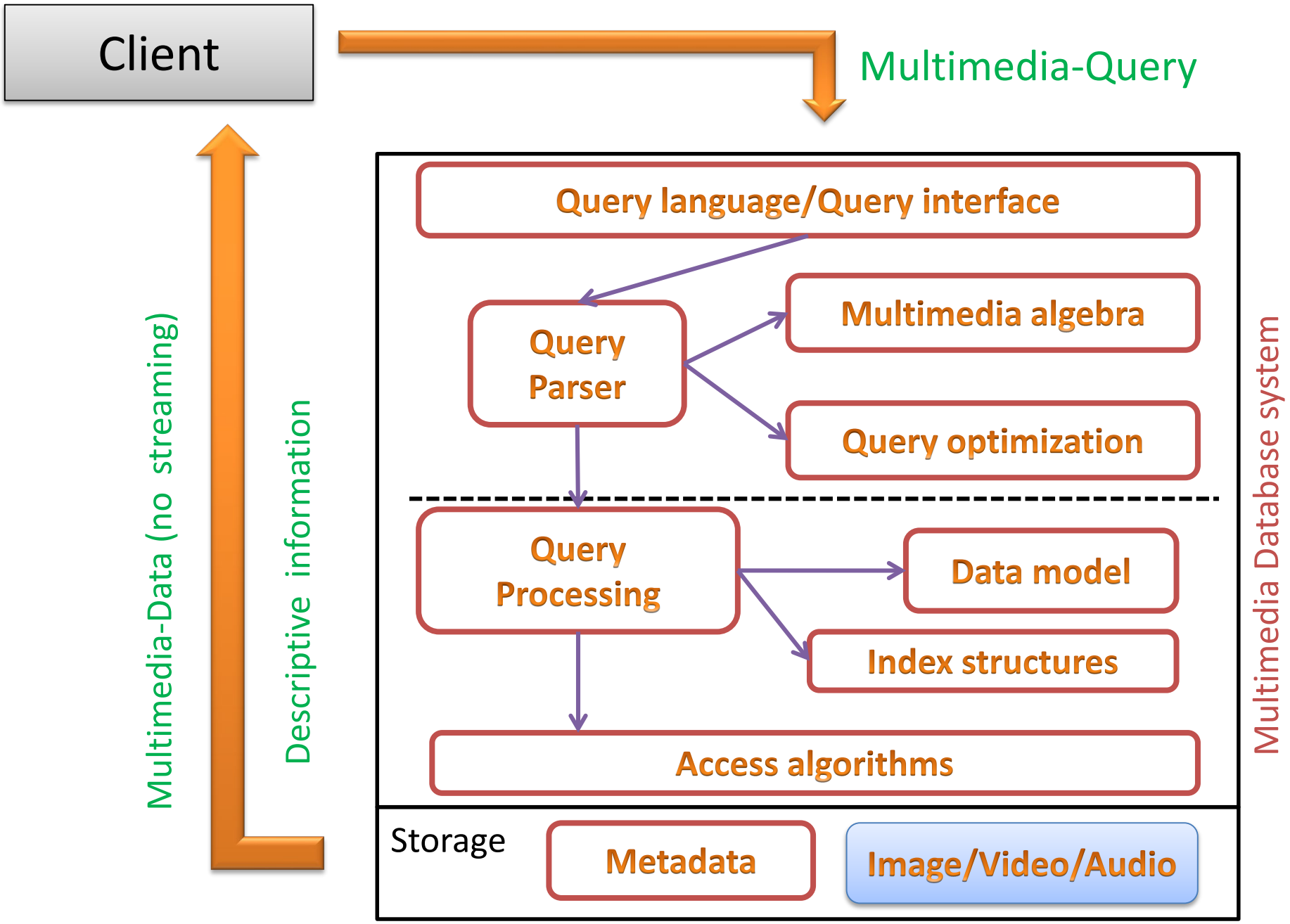
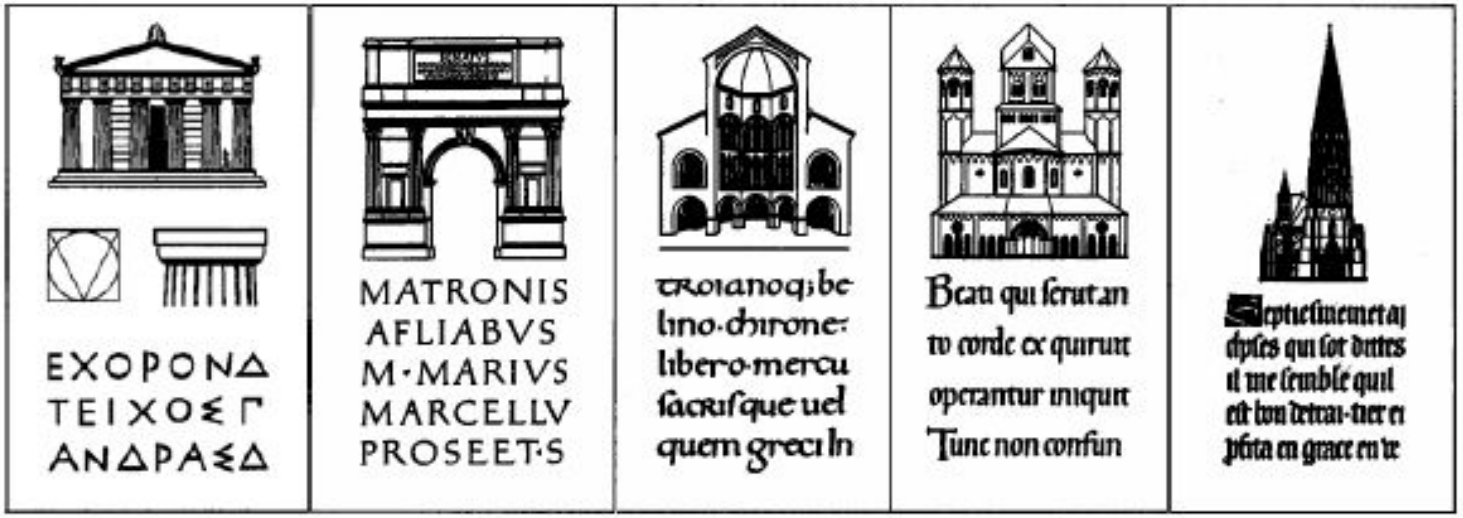

Client
Multimedia-Query
Multimedia-Data (no streaming)
Descriptive information
Query language/Query interface
Query Parser
Multimedia algebra
Query optimization
Query Processing
Data model
Index structures
Access algorithms
Storage
Metadata
Image/Video/Audio
Multimedia Database system

Table of Contents

Media – Text, Video, Audio

History of Characters



Greek	Roman	Carolingian	Romanic	Gothic
500 b.c.	100	800	1100	1250
Geometric forms	Capitalis Quadrata	Small letter fonts		Textura

Typographic Measurements

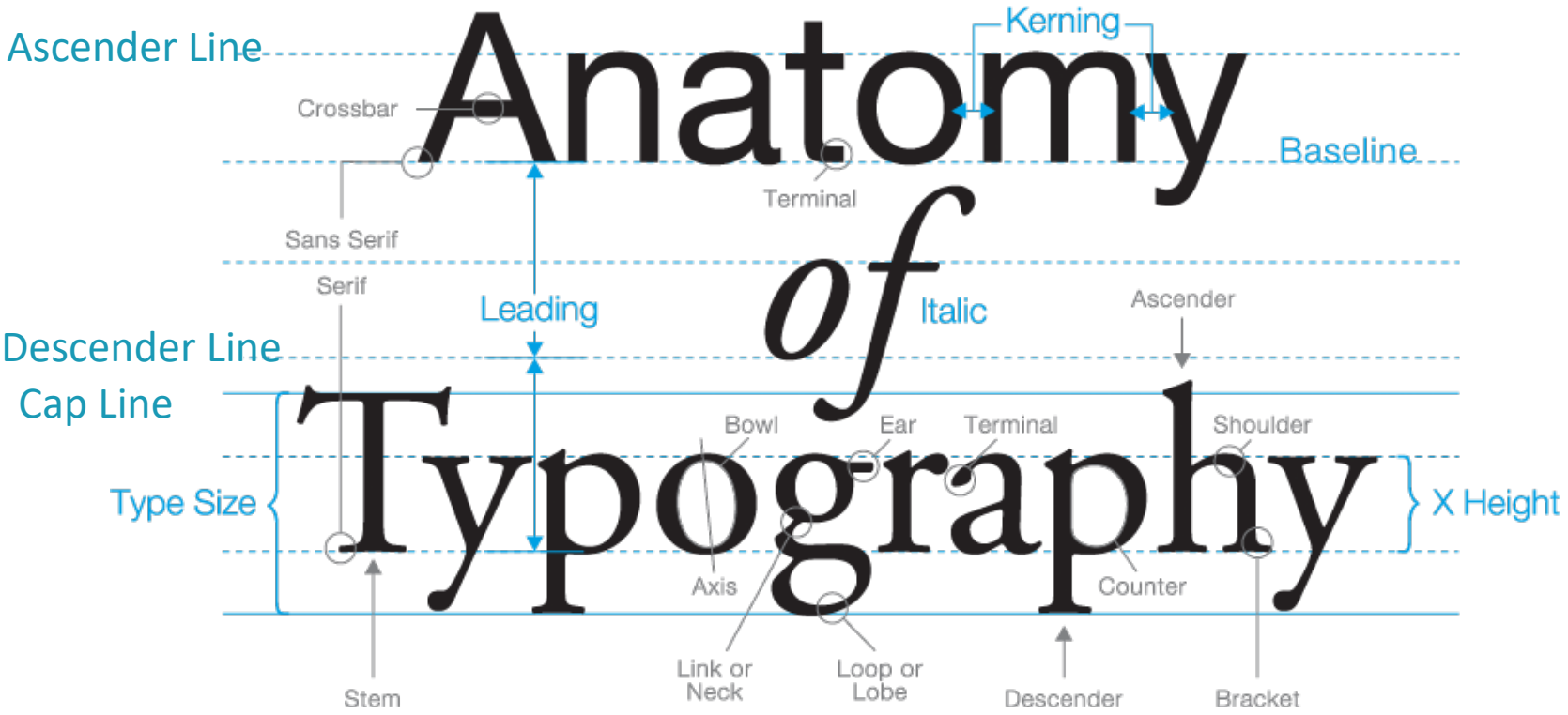


Image Source: http://typostrate.com/inspirations/typography-specials/the-anatomy-of-type/

- Font Size usually measured in Big Points (pt, bp), where 1pt=1/72 inch = 0.3527 m
- Relative Measures
  - 1 ex = height of glyph
  - 1 em = widht of glyph

Distinguishing Features of Characters

- Glyph width:
  - Monospaced (e.g. I M) vs. proportional (e.g. I M)

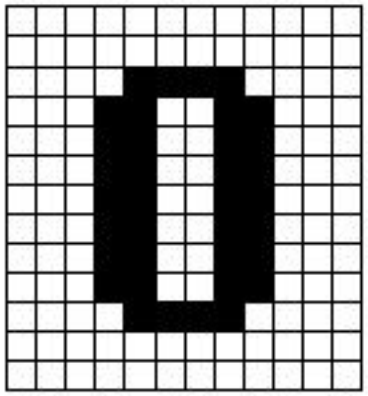
- Serif/Sans-serif

- Combination of glyphs
  - Kerning
  - Ligatures

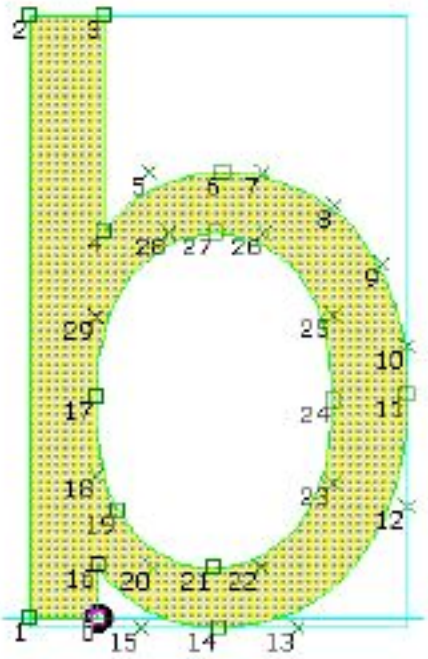
$AE \rightarrow Æ$ $ij \rightarrow ij$
$ae \rightarrow æ$ $st \rightarrow st$
$OE \rightarrow Œ$ $ft \rightarrow ft$
$oe \rightarrow œ$ $et \rightarrow \&$
$ff \rightarrow ff$ $fs \rightarrow ß$
$fi \rightarrow fi$ $ffi \rightarrow ffi$

Character Representation

Bitmapped fonts



Outlined fonts



Bitmap Representation

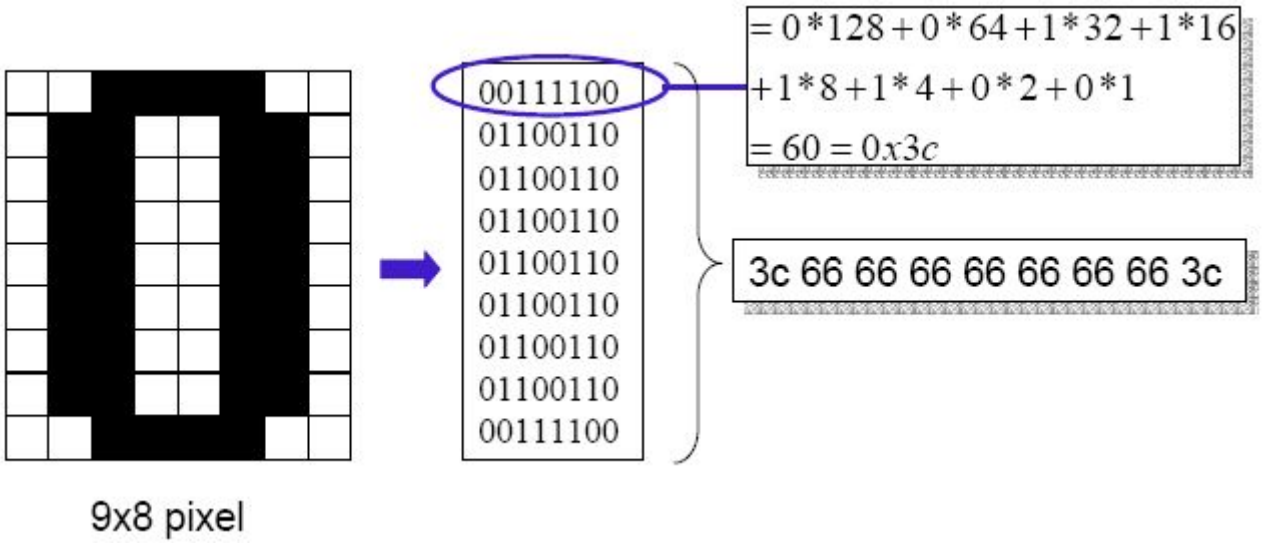


Table of Contents

Media – Text, Video, Audio

Media type Text

- REPRESENTATION
  - ASCII
  - ISO character sets
  - Markup text
  - Structured text
  - Hypertext

- OPERATIONS
  - Character and string operations
  - Editing
  - Formating
  - Pattern recognition and search
  - Sorting
  - Compression
  - Encryption
  - Language specific operations

Representation of Text

- ASCII (American Standard Code for Information Interchange)
  - 7-Bit-Code, American Standard Code for Information Interchange, 1 Bit free for special characters
  - ISO character sets
  - Extension of ASCII to accentuated characters, e.g. Latin-1
- Markup text
  - Separation of form and content
  - Special tags to specify the structure of presentation
- Structured text
  - Processing oriented representation based on suitable data structures,
  - e.g. trees
- Hypertext
  - Graph oriented combination of information units
  - Nodes, edges, links

e.g. XML

Character set ISO Latin-1

- Only printable characters; the line number defines the higher half of the byte, the column number the lower half

	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-a	-b	-c	-d	-e	-f	
0-																	0-
1-																	1-
2-		!	"	#	$	%	&	'	(	)	*	+	,	-	.	/	2-
3-	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?	3-
4-	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	4-
5-	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_	5-
6-	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	6-
7-	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~		7-
8-																	8-
9-																	9-
a-		¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬		®	¯	a-
b-	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿	b-
c-	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï	c-
d-	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß	d-
e-	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï	e-
f-	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ	f-
	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-a	-b	-c	-d	-e	-f	

Operations for Text

- Character operations
  - e.g. comparison ’a’ < ’b’
- String operations
  - e.g. comparison ’abc’ < ’abd’
  - Concatenation
    ’abc’ + ’def’ → ’abcdef’
- Editing
  - Modification of content and form
  - Copy, insert, remove, etc.
- Formatting
  - Assignment of layout information
  - Editor & Formatter → WYSIWYG
  - Page description languages→
    Bitmap-oriented display (Postscript)
- Pattern recognition and search
  - e.g. search and replace
  - Regular expressions
- Sorting
- Compression
  - e.g. Huffman-Code (cf. JPEG),
    Reduction of 1/2 -1/3
- Encryption
  - Important for data transmission
  - DES (Data Encryption Standard),
    Public-Key cryptography

Example of compression : UTF

- Unicode Characters (e.g. used in Java) are coded on 16 Bit.
- For storing **U**nicode **T**ransformation **F**ormat (UTF-8):
  - Enables to transform in a defined way 2 Byte long UNICODE-Characters into 1, 2 or 3 Byte long single characters
  - UTF-8 coding:

for UNICODE

From	To	Byte	Representation
\u0000	\u007F	1	0nnnnnnn
\u0080	\u07FF	2	110nnnnn 10nnnnnn
\u0800	\uFFFF	3	1110nnnn 10nnnnnn 10nnnnnn

  - In addition, two-bytes representing the length are stored at the beginning of each UTF-8 string.
- UTF-8 coding enables to save space when storing texts written in the most common languages.
- All ASCII characters are coded with a single byte.

Table of Contents

Media – Text, Video, Audio

Extensible Markup Language (XML)

- Developed by the World Wide Web Consortium (www.w3.org) since1996
- Extensible Markup Language
- Goal: ensure two main properties:
  - Flexibility and performance of the Standard Generalized Markup Language (SGML),
  - Wide acceptance of HTML
  - ⇒ XML is defined as a subset of SGML.
- XML = markup language for the representation of structured data.
- XML is platform-independent.
- XML is an open standard, which enable the encoded documents to self-describe their structure.
- XML is thus an „ideal medium“ for the exchange of documents or data between users or software.

Syntax of XML documents

- An XML document is composed of two parts:
  - An optional Prolog (XML-Declaration; marked in green below), which contains information about:
    - The XML version,
    - The character set and
    - An optional Document Type Definition – more on that later
  - The Body of the document = actual content of the dokument

```
<?xml version="1.0" encoding="ISO-8859-1"?>

<song>
  <title>Aber bitte mit Sahne</title>
  <author type="music">Udo Jürgens</author>
  …
</song>
```

Character coding with 8 Bit (ASCII + special characters of west-European languages)

Elements (Structural Elements, XML Elements)

- Elements = basic units of XML.
  - Elements may contain other elements, text or comments.
- An element is delimited by tags.
  - A tag consists of chevrons ('>', '<') enclosing a keyword.
  - Each element is defined between an opening and an enclosing tag.
  - Contrarily to HTML, an opening tag must be matched by a closing tag.
  - Only exception: an empty element, may be written using an abbreviated notation.
- The following table shows examples of tags:

Description	Tags
An element with start and end tags	`<title>`Aber bitte mit Sahne`</title>`
An empty element with start and end tags	`<melody></melody>`
An empty element in abbreviated notation	`<melody/>`

Well-formed XML

- Contrarily to HTML, XML document must fulfill a hard criterion in order to be validated (e.g. by a parser).
- XML documents that are not „well-formed“ must be rejected by users or software.
- An XML document is well-formed if it respects the following conditions:
  - An XML document may only have a single root-element.
  - All subsequent elements must be connected to the root-element in a single hierarchical tree structure.
  - An XML document must not contain invalid characters. This condition refers in particular to the character set specified in the prolog.

Validation of XML-documents

- Well-formedness is an important criterion when verifying XML documents
  - It guarantees that the document is conform to the general syntactic rules of the Extensible Markup Language.

- More advanced criterion: validity
  - Means that the structure of the XML document is evaluated for conformity to the corresponding Document Type Definition.
  - If no DTD was specified for an XML document, it cannot be validated.
  - Only well-formed XML documents may be validated.
  - (an XML-Schema may also be used instead of a DTD)

XML-Schema

- Up to this point, DTDs were used to define specific markup languages
- However, DTDs have a number of drawbacks:
  - Limited capability to define the structure and content of elements.
    - E.g. not possible to assign data types to specific elements.
  - Limited capabilities to specify the cardinalities of the elements.
  - The syntax for DTD definition itself is not fully XML compliant!
- XML-Schema offers advanced functionalities to specify the structure of the elements as well as their data types.
- More information: cf. www.w3.org

Representation of XML documents

- XML documents only contain structured data!
- Contrarily to HTML, they provide no information about the presentation of the data.

- Data-centric vs. Document-centric view

- The Extensible Stylesheets Language (XSL) comes here into play:
  - This language is itself based on XML; it enables to transform an XML document into another XML document.
  - To this end, an XSL document specifies a number of transformation rules.

- Thanks to the clear separation of presentation and data, an XSL-Stylesheet enables to transform an XML document:
  - in another XHTML document,
  - in a WAP document
  - or even in a PDF document (with additional formatting steps)

Linking and querying

- As in HTML, an XML document may contain references or links to other documents or to a section of the same document.
  - The description of links in XML is based on the Xlink specification.

- Several specifications enable to select elements from XML docs (i.e. querying):
  - The most basic is XPath (see path expressions in XSL).
  - Enables to select single elements based on their hierarchy, their attribute values or their content.
  - The hierarchy of the searched elements are specified as regular expressions: „find elements that are located inside a specific element“.

- More advanced language: XML Query – described by the W3C as follows:
  - „The mission of the XML Query working group is to provide flexible query facilities to extract data from real and virtual documents on the Web, therefore finally providing the needed interaction between the web world and the database world. Ultimately, collections of XML files will be accessed like databases.“

Table of Contents

Media – Text, Video, Audio

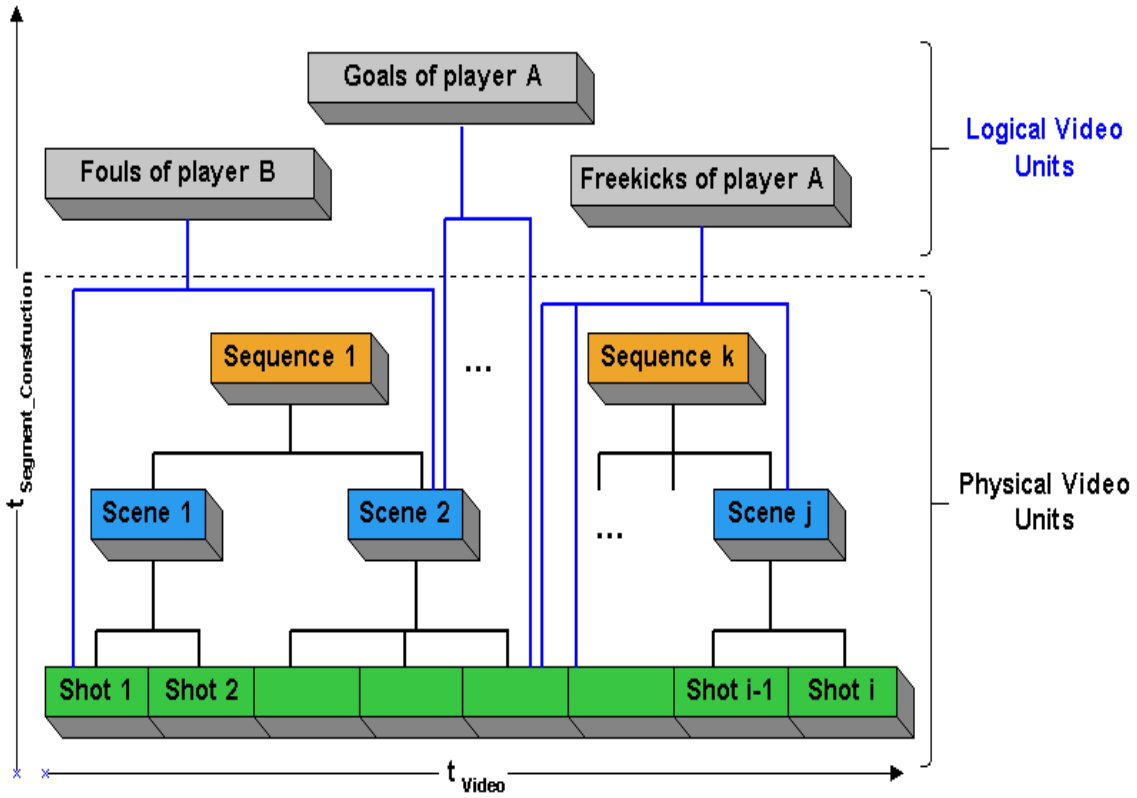
Video hierarchy I

- Digital videos are composed of the following elements:
  - Frame
  - Shots
  - Shot boundaries
  - Scenes

Video hierarchy II

- Partition of a video stream in physical and logical video segments

- The higher the level of a video unit, the higher the amount of necessary semantic information
- Logical units are based on semantic content



Video hierarchy

Shot

- Characteristics
  - General Shot
    - The object is relatively far away from the camera.
  - Medium Shot
    - The object is quite close to the camera.
  - Close Up
    - The object is very close, it almost represents the whole image.

- Types
  - Static
    - The camera does not move for the duration of the shot!
  - Dynamic
    - The camera position changes during the shot: zoom, panning, etc.

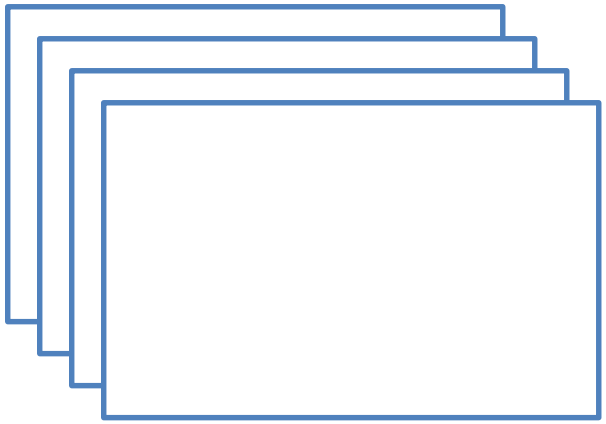
Video hierarchy

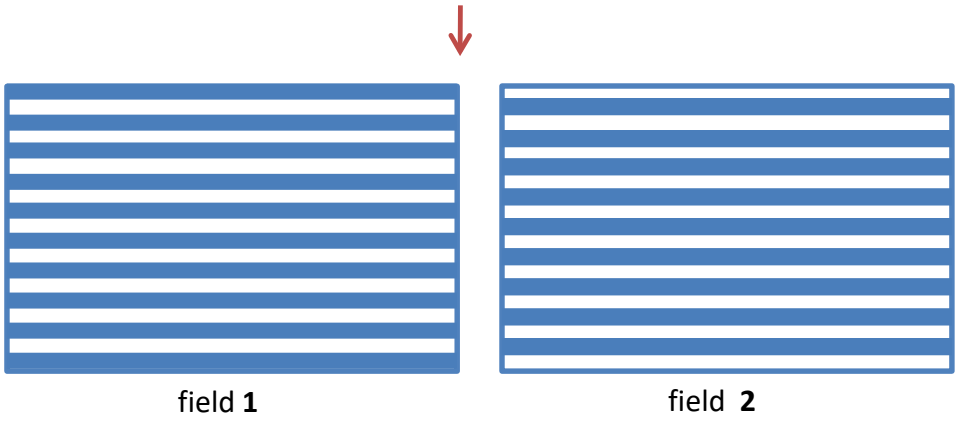
Scene

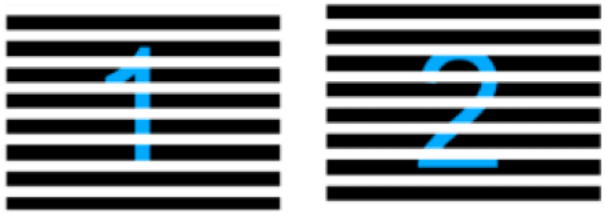
- Scene (Scenario)
  - Sequence of shots that are coherent in the time and space dimensions with respect to the real or represented world.
  - Scene detection is a subjective task
    - Depends on cultural background, professional training, intuition, etc.
- Screenplay
  - Describes all scenes, their dialogs as well as the setup of the cameras.

Video Frame Rate

- A video is a sequence of frames
- It has a specific resolution and must be played with a specific frame rate.
  - e.g. 25 fps, 30 fps (29.97), 50 fps
  - progressive (p) or interlaced (i)
  - Interlaced Frames are also called fields







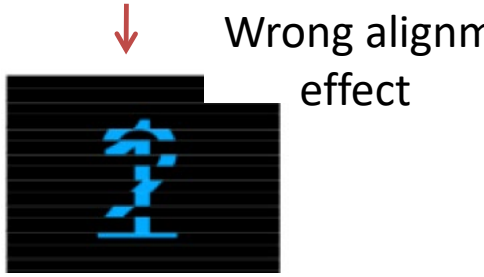


Image Source: https://wolfcrow.com/blog/understanding-terminology-progressive-frames-interlaced-frames-and-the-field-rate/

Usual Resolutions

- **576i** (digital equivalent to PAL)
  - 720x**576** Pixel, 50 half-images per second
  - Video-DVD, DVB-S, DVB-T (PAL e.g. UHF/VHF)
- **720p** (HDTV)
  - 1280x**720** Pixel, 60 images per second
  - Blu-ray video, DVB-S2 (*HD Ready*):
    - e.g. ORF HD, ARD/ZDF HD
- **1080p/i** (HDTV)
  - 1920x**1080** Pixel, 24/50/60 images (p) or 50/60 half-images (i) per second
  - Blu-ray video, DVB-S2 (*Full HD*):
    - e.g. Servus TV HD, Astra HD
- **2K**, **4K** (Digital Cinema)
  - 2048x1080 Pixel and 4096x2160 Pixel, 24 images per second
  - Digital cine films
- **4320p** (UHDTV)
  - 7680x**4320** Pixel, 60 images per second
- **8K UHD** (7680 × 4320 progressive scan)

Table of Contents

Media – Text, Video, Audio

Shot Segmentation

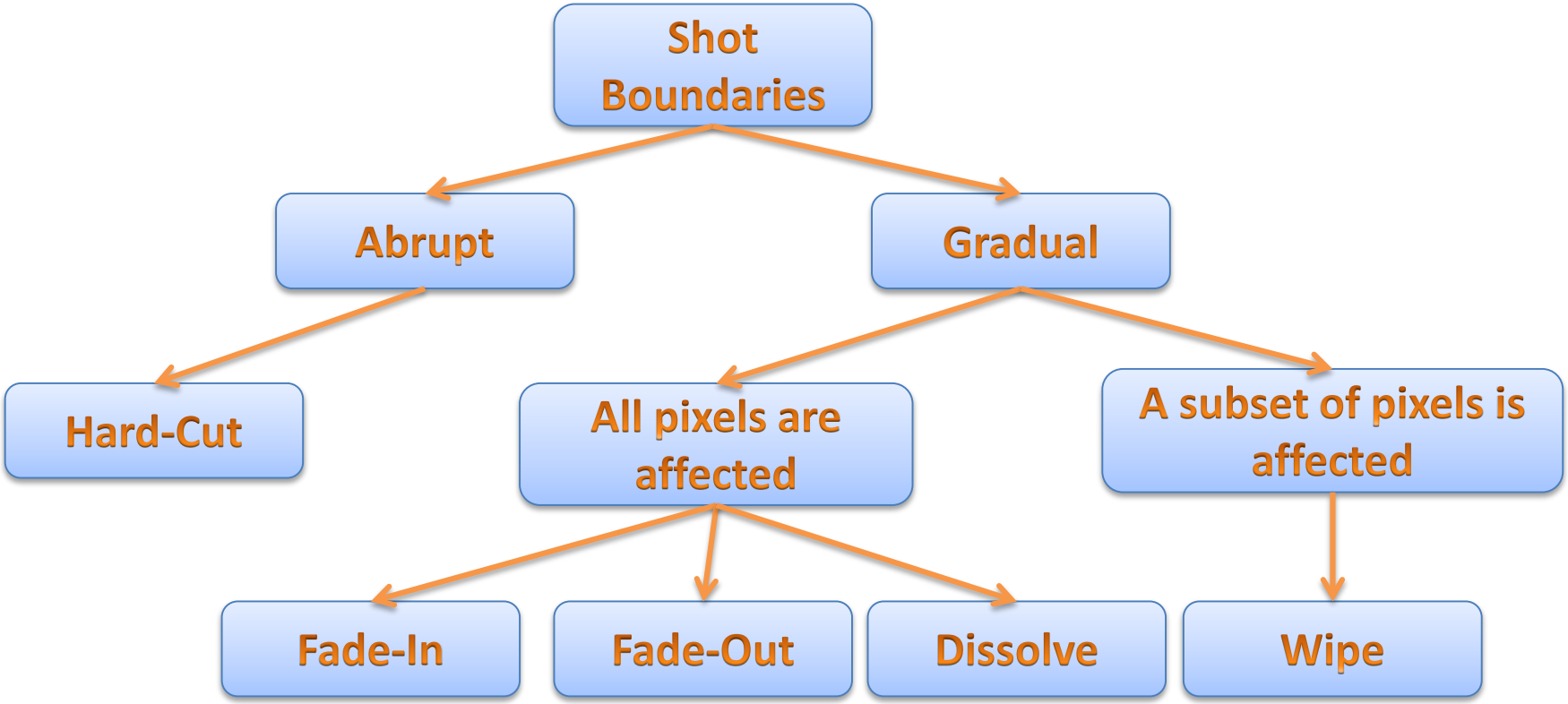
- Challenges
  - Object motion
    - Ex.: A person moves within a camera shot, …
  - Camera motion
    - Ex.: Zoom, panning, …
  - Lighting changes
    - Camera blitz, illumination, …
  - Different types of shot boundaries
    - Ex.: Fade, Dissolve, …

- Preventing *False positives*
  - Adjust threshold values
  - Empiric restrictions
    - Ex.: A shot must last more than 100 frames.

Shot Segmentation

Overview

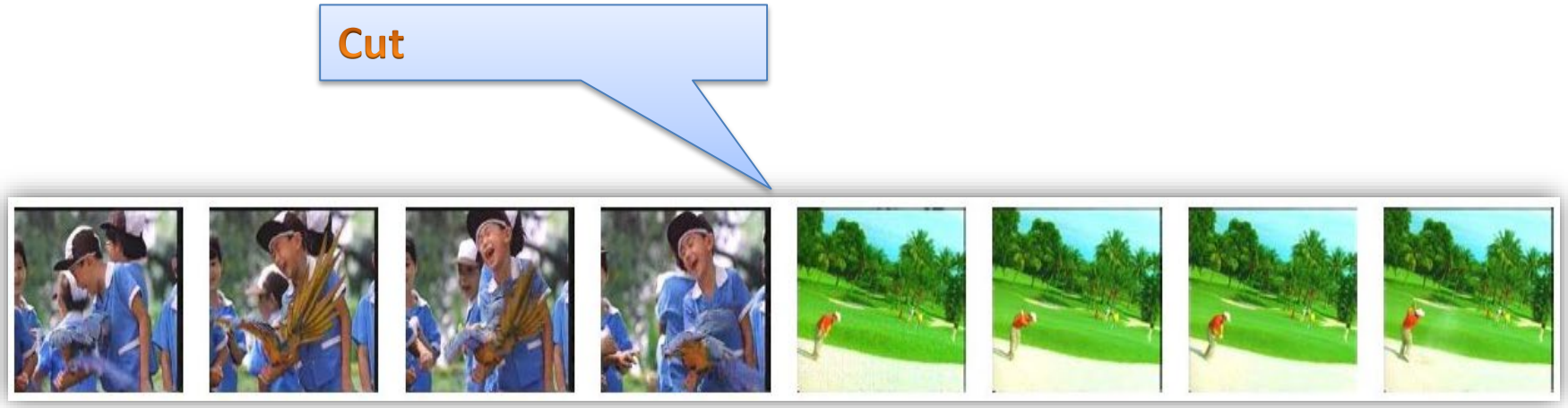
- Classification according to spatial/temporal properties



Shot Segmentation

Shot types

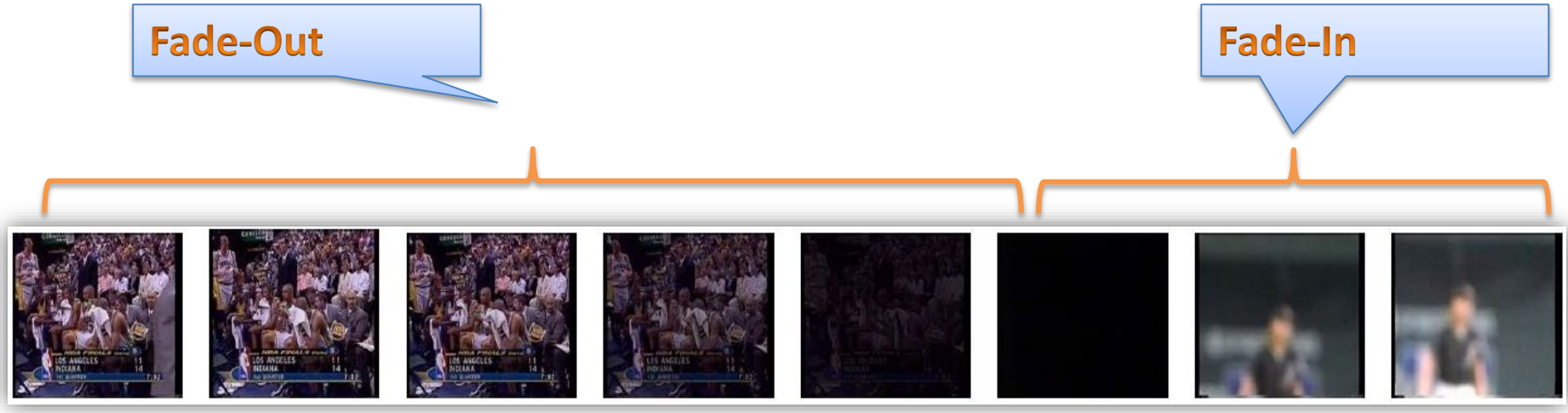
- Hard-Cut
  - Sudden change of the whole image content



Shot-Segmentation

Shot Types

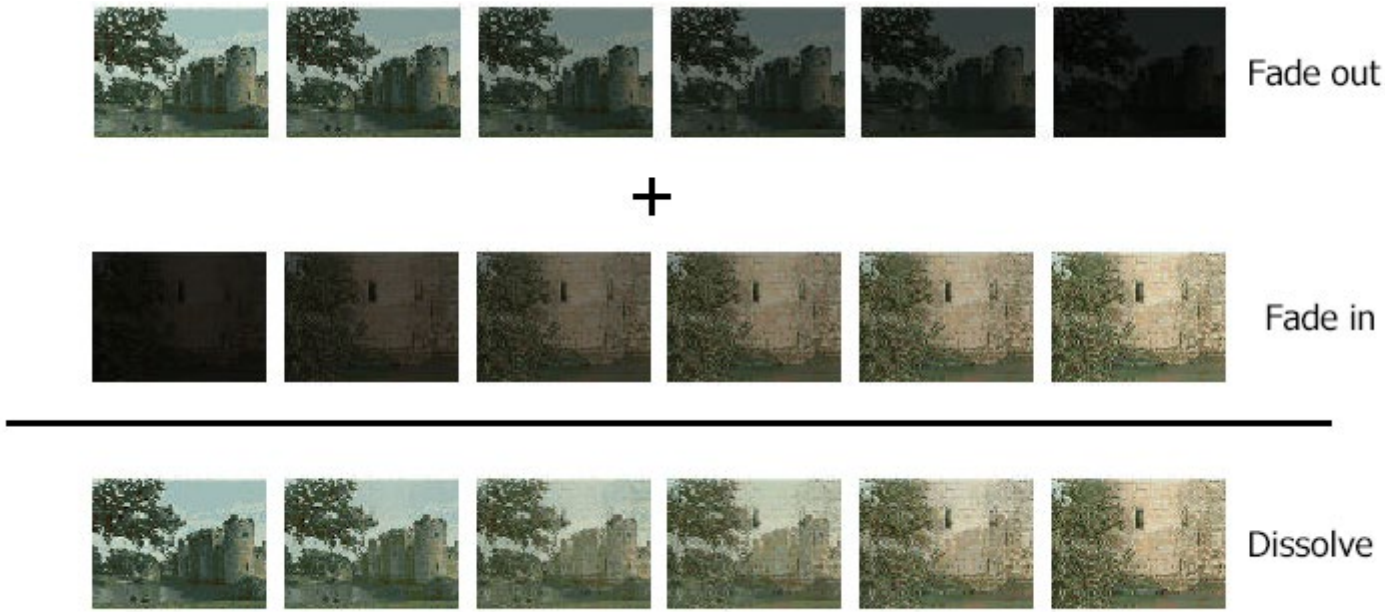
- Fade-In
  - Change from a black image to a complete image
- Fade-Out
  - Change from a complete image to black



Shot-Segmentation

Shot Types

- Dissolve
  - Mix of two shot sequences where the first one is faded out while the second one is faded in.



Shot-Segmentation

Shot Types

- Wipe
  - A new shot "pushes away" the previous one through a horizontal/vertical/form-based movement.

Vertical motion

Shot-Segmentation

Spatio-temporal properties

Types of shot boundary		Cut	Wipe	Dissolve	Fade
Spatial	Content change in the whole frame	yes	no	no	yes
	Content change in a subset of pixels	no	yes	no	no
	Brightness change to black	no	no	no	yes
Temporal	Sudden change of content between consecutive frames	yes	no	no	no
	Slow change of content over several frames	no	yes	yes	yes

Evaluation of Shot Detection Algorithms

- Recall
  - Ratio of the number of detected Shots (S_D) to the total number of shots (detected S_D and undetected S_M)

$$R = \frac{S_D}{S_D + S_M}$$

- Precision
  - Proportion of correctly detected shots (S_D corresponds to the number of detected shots and S_F to the wrongly detected shots (*false positives*)).

$$P = \frac{S_D}{S_D + S_F}$$

Shot detection methods

Classification

- Uncompressed domain
  - Video must first be decompressed
  - The methods make use of raw data
    - Pixel-based
    - Histogram-based
    - …
- Compressed domain
  - The methods use properties of the compressed data (e.g.: motion vectors, etc.)
    - Macro-block-based
    - Motion vectors
    - …
- Many methods are listed at:
  - http://www.visionbib.com/bibliography/applicat821.html

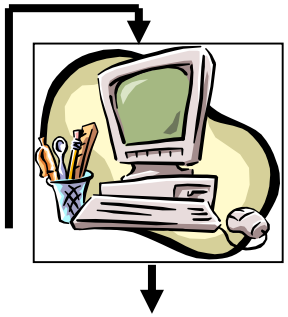
Shot Detection Methods

Color histogram

- The color distribution of each frame gets extracted
- The distributions are compared with those of each neighbor frame
- If the difference lies above a given threshold
  - -> Shot

Color histogram computed

Comparison with previous Color histogram



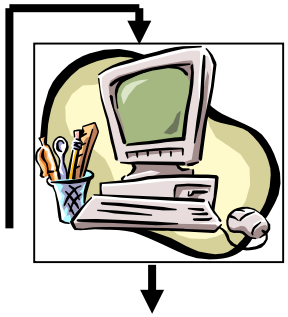
Difference > Threshold => Shot change

Shot Detection Methods

Edge Detection

- Reduction of the frame colors to greyscale
- Application of an edge detection algorithm
  - Canny Edge Detector, Sobel, etc.
- Differences are computed
  - Difference > Threshold=> Shot change

Comparison with the edges of the previous frame



Difference > Threshold => Shot change

Shot Detection Methods

Macroblock (compressed domain)

- Method based on MPEG compression
- Frame divided in fixed regions => Macroblocks
- Types of macroblocks
  - I: coded independently from other macroblocks.
  - P: does not encode the region, but rather a motion vector and an error block wrt. the previous frame.
  - B: same as P, but motion vector and error block encoding done wrt. to the previous and next frame

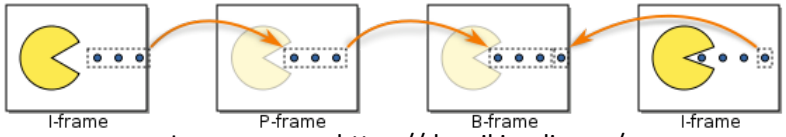
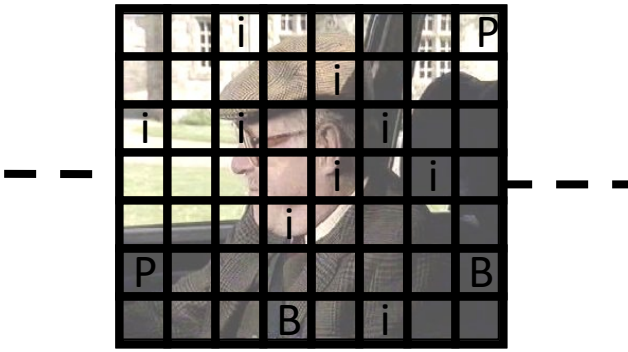
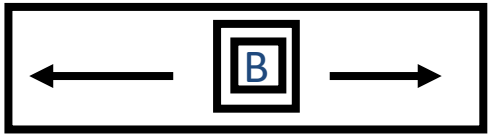


Image source: https://de.wikipedia.org/

Frame mit Makroblocks





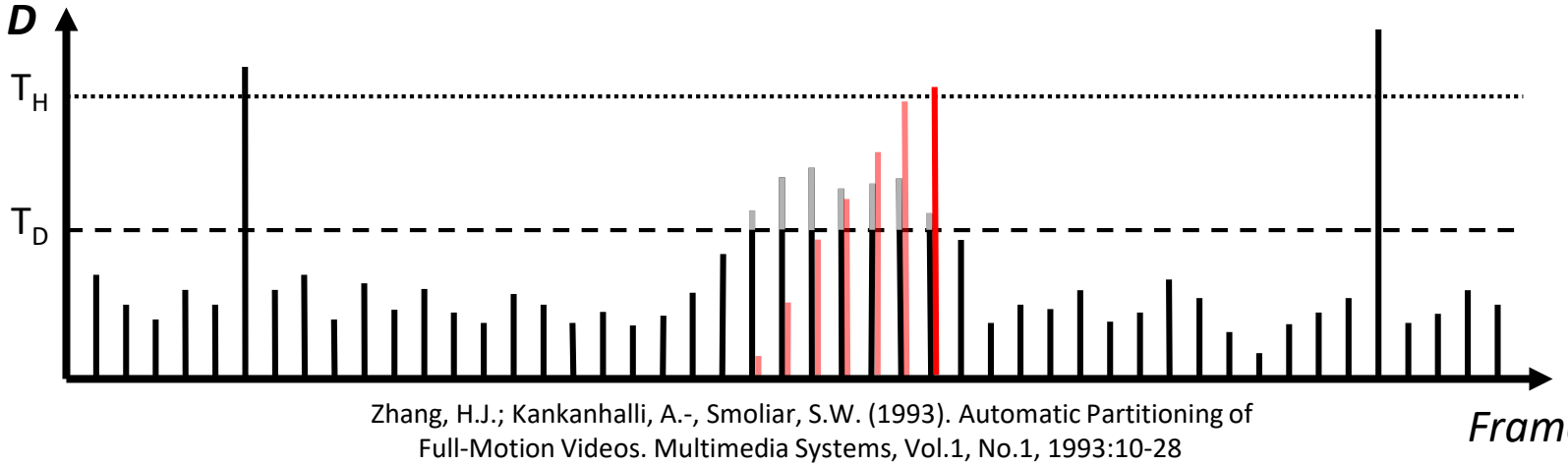
- Shot boundaries are identified by the apparition of a specific amount of macrobock types in a frame.

Shot Detection Methods

Gradual means: *twin-comparison*

- Comparison 1: Difference between two consecutive frames
- Comparison 2: Cumulative difference over a sequence of frames
- Difference value:

$$D_{cut} = \sum_{x,y} \left| I(x,y,t) - I(x,y,t+1) \right|$$



Zhang, H.J.; Kankanhalli, A.-, Smoliar, S.W. (1993). Automatic Partitioning of Full-Motion Videos. Multimedia Systems, Vol.1, No.1, 1993:10-28

Evaluation of the methods

- Average Precision over an 8 hours video
  - Color Histogram 90.4
  - Edge Detection 90.0
  - Macroblock 87.4
- Average Recall value over an 8 hours video
  - Color Histogram 78.9
  - Edge Detection 70.2
  - Macroblock 75.3

- Programs with lowest Recall:
  - Home & Away (Australian soap)
  - Cooking program

Figures from: Paul Browne
Centre for Digital Video Processing
Dublin City University

Table of Contents

Media – Text, Video, Audio

Key Frame

- Properties:
  - Should be representative of the content of the shot
  - In the compressed domain, an I-Frame to summarize a video

- Extraction methods
  - Optimal algorithm
    - Compares each frame with all the others of a Scene/Shot.
    - The frame with the lowest difference to all others is selected.
    - Very computationally intensive, unusable for most applications.

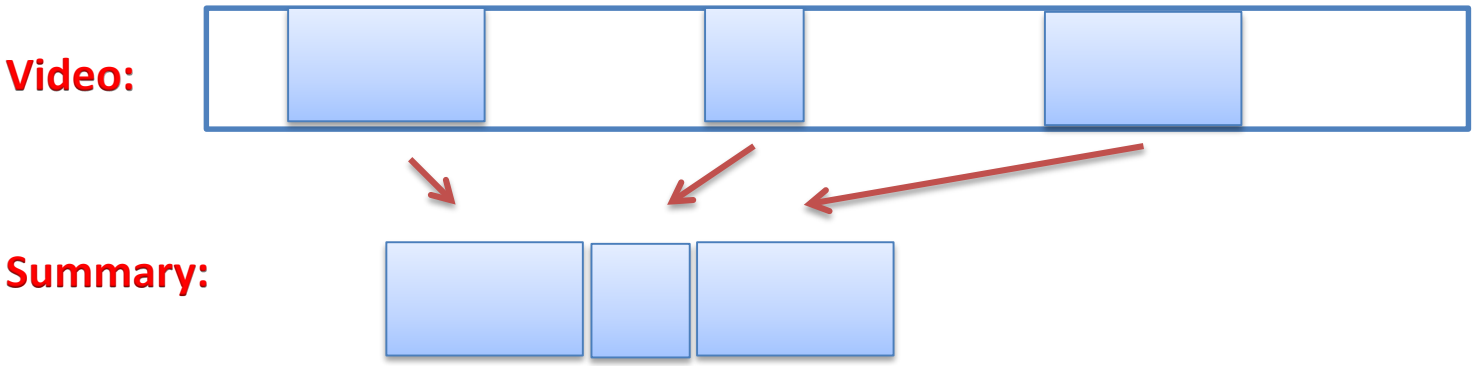
Key Frame

Extraction Methods

- Selection of the first frame of a shot/scene as key frame
  - Basically a logical choice since all the other frames of the shot/scene should actually be a continuous extension of/based on that frame.
  - However, the first frame is not necessarily the best and most meaningful representation of a shot/scene.
- Select the frame with the *most complex* or *clearest* content.
  - e.g.: Frames containing text, frames containing easily recognized persons, etc.

Video summarization

- Goal of video summarization:
  - Give a quick overview of the content of a video to a user



- Requirements of video summarization:
  - Conciseness => the summary cannot bet longer than the Video
  - Content representation=> Temporal and semantic aspects of the content must be preserved
  - Coherence

Video summarization

Classification

- Types
  - Static and dynamic methods

- Applications
  - Video Skimming
  - Video Story Board

- Two classes of video summarization
  - Independent
    - Creates a preview representative of the whole content.
    - Depends only on a time limit *L*
  - Dependent
    - A user has specific preferences wrt. the content (e.g.: specific person, specific event, specific time domain, etc.)

Table of Contents

Media – Text, Video, Audio

Video Formats

Differenciation: Format - Codec

- A file format represents a kind of „architecture"
- Special notation about the way a file is structured (programs that read or create header-data, storage of audio-data, etc.)
- Examples of video formats: MPG, VOB, AVI, ASF
- Different compression methods (Codecs) possible
- A file with .AVI extension may also contain an MPEG-1 encoded video

Video formats

AVI (Audio Video Interleave) I

- Special case of RIFF (Resource Interchange File Format), a Multimedia format created for Windows 3.1
- Can store audio and video information in such a way that they can be reconstructed
- An AVI file consists of chunks (nested data structures), which can themselves contain sub-chunks

Video formats

AVI (Audio Video Interleave) II

- Created by Microsoft as a consistent format for the playback of short video clips (1992/93)

- Original specifications:
  - maximum Resolution: 160 x 120 Pixel
  - Frame rate: 15 fps

- Keyframe-Technique: each 12th to 17th image (depending on image content) stored fully (keyframe *k*), whereas only the differences to the previous frame are stored for all frames located between *k* and the next keyframe

Video formats

AVI (Audio Video Interleave) III

- Part of "Video for Windows„: free driver
- Gained popularity because video editing systems (Miro/ Pinnacle, Fast) and related software supported AVI by default
- Only provides a framework for various encoding algorithms (bsp. Cinepak, Intel Indeo, DivX)

Video formats

AVI (Audio Video Interleave) - Drawbacks

- AVI is a true data format, i.e., an AVI-Clip can only be played if the whole file is available, contrarily to Video - Streaming.
- Still used by semi-professional video capture cards despite its age and many problems.
- Poor support of subtitles (in DVDs, as a workaround subtitles are represented as images).
- Not for all audio formats.

Video formats

MPEG / MP4

- Motion Picture Experts Group: international committee that develops Standards for motion picture coding

- In order to guarantee its applicability to the largest set of software, the MPEG-standard only specifies a data model for compressing motion pictures and audio signals. That enables MPEG to remain independent from the large set of available computer platforms.

- 3 main coding standards: MPEG-1, MPEG-2, MPEG-4

Video formats

MOV/ QuickTime

- Introduced by Apple in 1991 as a system extension for its Macintosh computer line
- Ported to x86-based PCs and a number of operating systems (Windows, OS/2 and Unix derivatives)
- In functionality and quality, QuickTime was much above Microsoft's AVI between 1993 and 1995
- The data format is extensible and Track-based.
  - The Tracks are composed of different content types: Audio, Video, Flash, HTML and more.
- New digital media technologies can be easily integrated by creating new track types.

Video formats

MOV/ QuickTime

- Modularity: Application software can communicate with QuickTime over a standard API (Plugin)
- Contains important codecs: MPEG & JPEG supported
- Development environment / multimedia platform
- Composed among others of: QuickTime Player (Media player), QuickTime Pro (editing) and QuickTime Streaming Server (Streaming over LAN or Internet)

Video formats
WebM

- WebM is an open standard for the distribution of media files over the Internet.
- Provided by Google as free software under a BSD-license.
- Based on the Video-Codec VP8 (developed by On2 Technologies) and the Vorbis Audio Codec
- Available as free codec for the HTMLS Video tag

Overview of container formats

Typische Kombinationen

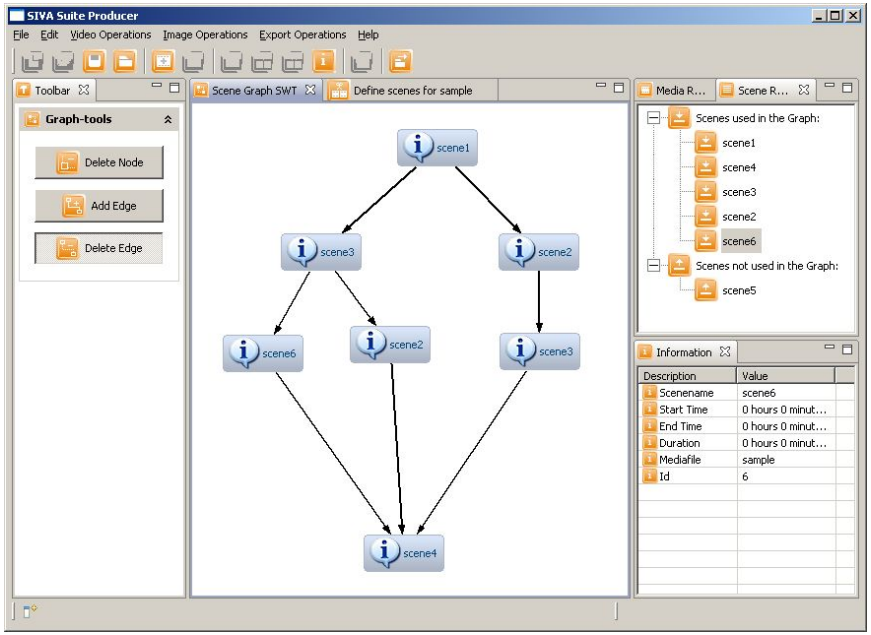
Container	Name	Videocodecs	Audiocodecs
3GP/3GP2	3rd Generation Mobile	H.263, MPEG-4, H.264	AMR-NB/WB, (HE-)AAC
AVI	Audio Video Interleave	MPEG-4, DV, MJPEG, Indeo, Cinepak	MP3, MP2, (AD)PCM, AC3
ASF	Advanced Streaming/Systems Format	Windows Media Video, VC-1, MS MPEG4 v3	Windows Media Audio
DIVX	DivX Media Format	DivX	MP3, AC3, PCM
DV	Digital Video	DV	PCM
DVR-MS	Microsoft Digital Video Recording	MPEG-2	MP2, AC3
EVO	Enhanced VOB	H.264, VC-1, MPEG-2	(E)AC3, DTS (HD), PCM
FLV	Flashvideo	H.263, VP6, H.264	MP3, AAC, ADPCM
M2TS/MTS	MPEG-2 Transport Stream (192 Byte)	H.264, VC-1, MPEG-2	(E)AC3, DTS (HD), PCM
MKV	Matroska	H.264, MPEG-4	MP3, AC3
MP4	MPEG-4	MPEG-4, H.264	AAC
MPG	MPEG Program Stream	MPEG-1, MPEG-2	MP2
MOV	QuickTime Movie	H.264, MPEG-4, MPEG-1, MJPEG, Sorenson Video	MP3, AC3, PCM
OGM	Ogg Media	Ogg Theora, Xvid	Ogg Vorbis, MP3, AC3
PS	MPEG-2 Program Stream	MPEG-2	MP2, AC3, DTS, PCM
RM(VB)	Real Media (variable Bitrate)	Real Video	Real Audio, AAC
TS/TP/TRP/PVR/VDR	MPEG-2 Transport Stream (188 Byte)	MPEG-2, H.264	MP2, AC3
VOB	Video Object	MPEG-2	AC3, DTS, PCM, MP2
WMV	Windows Media Video	Windows Media Video, VC-1	Windows Media Audio

Table of Contents

Media – Text, Video, Audio

Interactive videos

- An interactive video is a digitally enhanced form of video material, which provides viewers with additional information about the content and lets them control the flow of video content in particular through advanced navigation options.
- Arrangement
  - Scene Graphs



Example project

- SIVA Project:

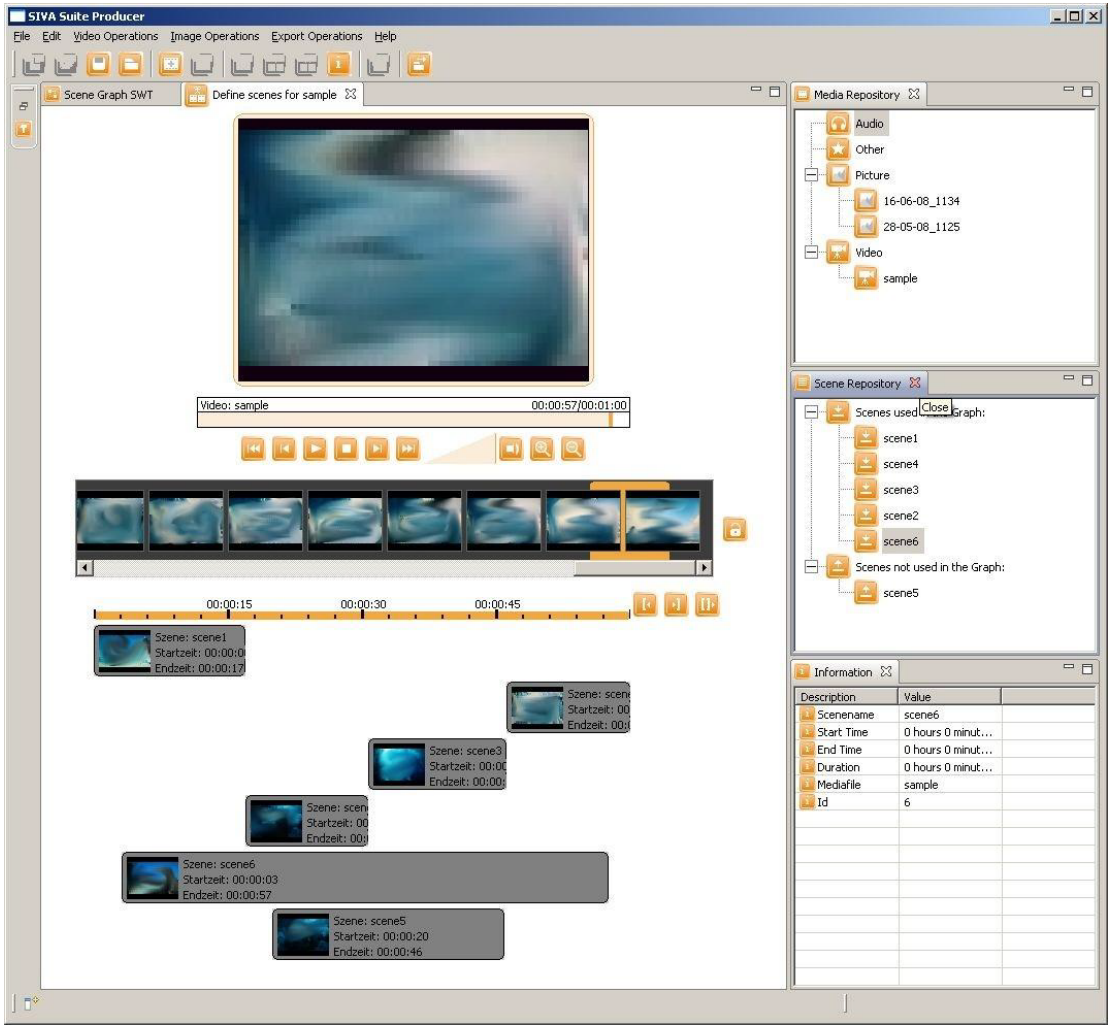


Table of Contents

Media – Text, Video, Audio

Audio coding - Basics

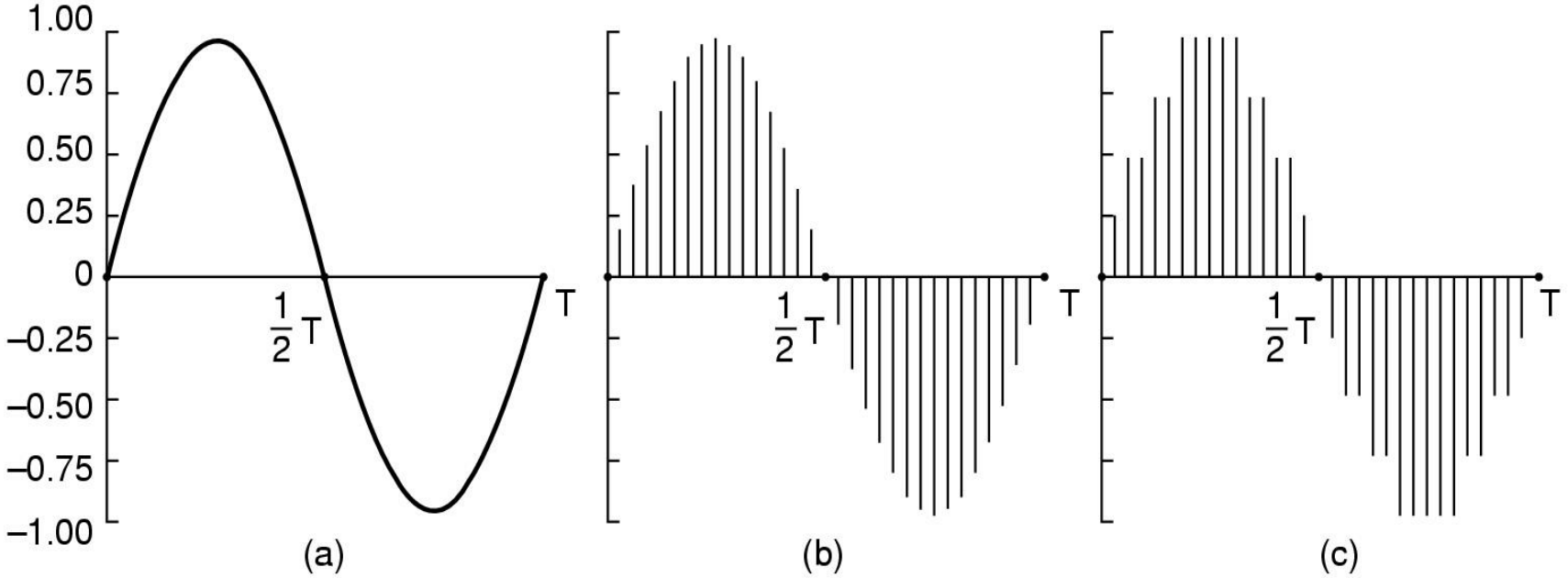
- Audio (acoustic) waves
  - One-dimensional, acoustic (pressure) wave
  - Causes vibration of the eardrum (human perception) or of a microphone
- Frequency spectrum of the human ear
  - 20 – 20.000 Hz (20 KHz)
  - Quasi logarithmic perception: the ratio of amplitudes A and B is expressed by: $dB = 20 \log_{10} (A/B)$

Very low pressure (20 μPascal)	0 dB
Conversation	50 – 60 dB
Heavy traffic	80 dB
Rock band	120 dB
Pain threshold	130 dB

Analog-Digital Trasnformation (ADC)

- Sampling of the audio waves each ΔT seconds
  - If the acoustic wave is a linear superposition of undisturbed sine waves with max. frequency *f* :
  - Sampling rate= *Shannon and Nyquist* Theorem
    - Uniform sampling of rate of at least 2*f, in order to be able to reconstruct the original signal without information loss.
  - e.g. CD are sampled at 44.1 KHz ⍰ 2 * 20 KHz
- Quantization
  - Precision of the digital sampling process depends on the number of bits
  - *Quantization error*
    - Error due to the finite number of available bits/sample

Audio coding example



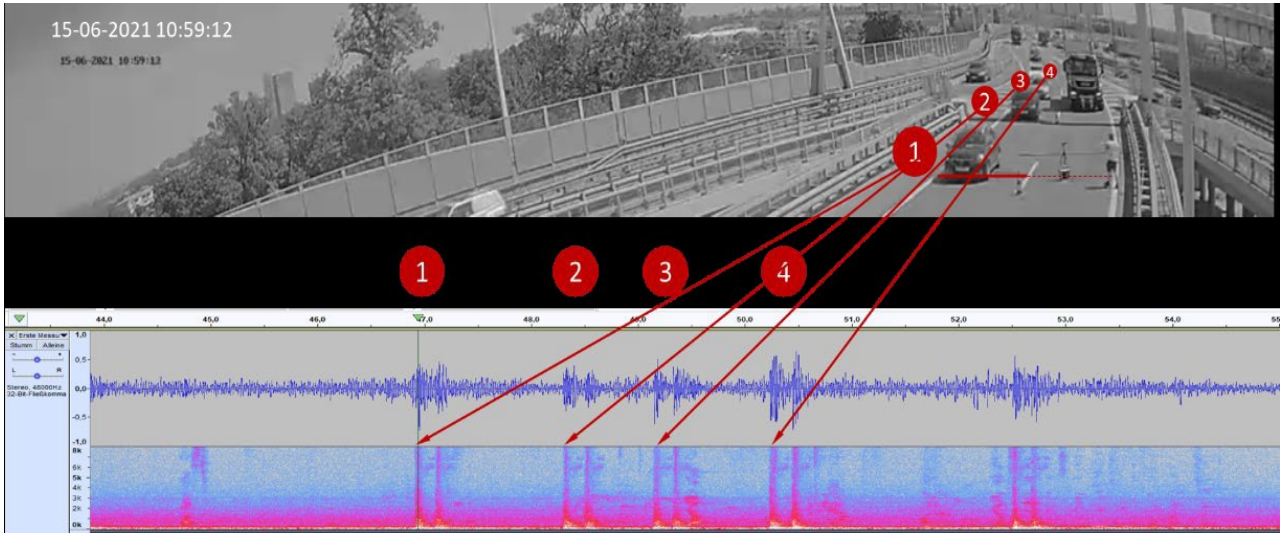
a) Sinus wave

b) Sampling of the sinus wave

c) Quantization of the sampled data to 4 Bits

Vehicle Classification based on Audio Signal

- Research project in cooperation with Bernard Technology
- Audio Sensor installed under a Bridge and test data has been recorded in combination with synchronized video data
- Evaluated Random Forest [1] and XGBoost [2] Method which resulted in 92% accuracy (XGBoost)



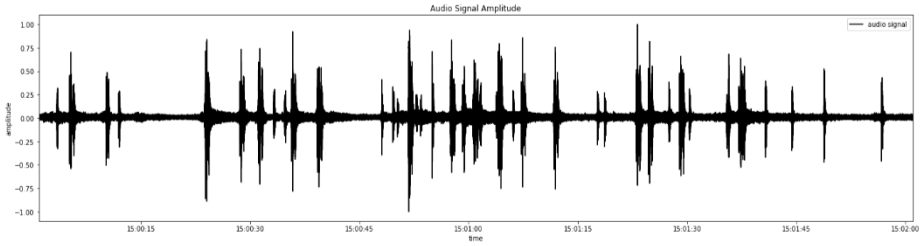
[1] Leo Breiman, Random forests. In Machine Learning, 45(1), 5-32, Springer.

[2] Tianqi Chen, Carlos Guestrin, XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016, DOI: 10.1145/2939672.2939785

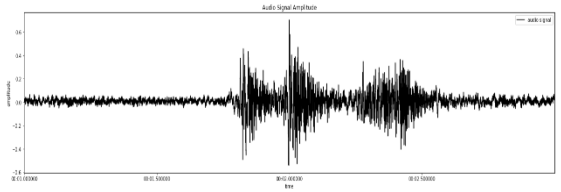
Vehicle Classification based on Audio Signal

- Creation of Ground Truth according to synchronized video data and Label Studio (https://labelstud.io/)

Audio-File



Peak Detection & Snippet Creation



Label Studio Implementation

2. Edit Labeling config

```
  </Labels>

  <AudioPlus name="audio" value="$audio"/>
  <View style="padding: 10px 20px; margin-top: 2em; box-shadow: 2px 2px 8px
        visibleWhen="region-selected">
    <Header value="Provide Transcription" />
    <TextArea name="transcription" toName="audio"
              rows="2" editable="true" perRegion="true"
              required="true" />
  </View>
  <View style="padding: 10px 20px; margin-top: 2em; box-shadow: 2px 2px 8px
        visibleWhen="region-selected">
    <Header value="Select Gender" />
    <Choices name="gender" toName="audio"
             perRegion="true" required="true">
      <Choice value="Male" />
      <Choice value="Female" />
    </Choices>
  </View>

  <View style="width: 100%; display: block">
    <Header value="Select region after creation to go next"/>
  </View>

 </View>
</View>
```

Start typing in the config, and you can quickly preview the labeling interface. At the bottom of the page, you have live serialization updates of what Label Studio expects as an input and what it gives you as a result of your labeling work.

3. Inspect Interface preview

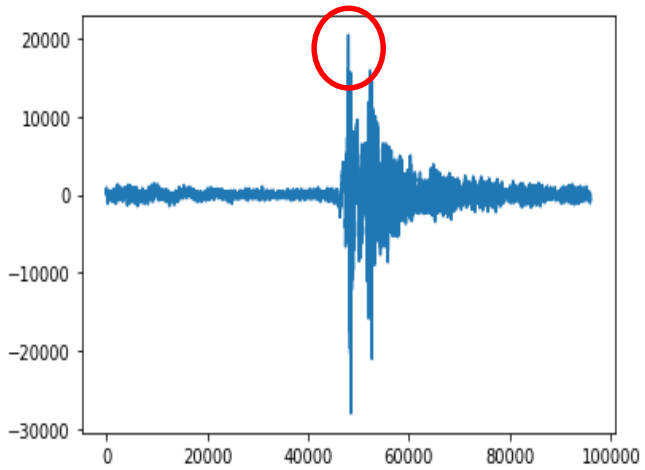
Labled Data


```
[
    {
        "original_length": 3.774375,
        "value": {
            "start": 0,
            "end": 0.6394089226973688,
            "labels": [
                "Speaker 1"
            ]
        },
        "id": "wavesurfer_8aaohk6v8a",
        "from_name": "label",
        "to_name": "audio",
        "type": "labels"
    }
]
```

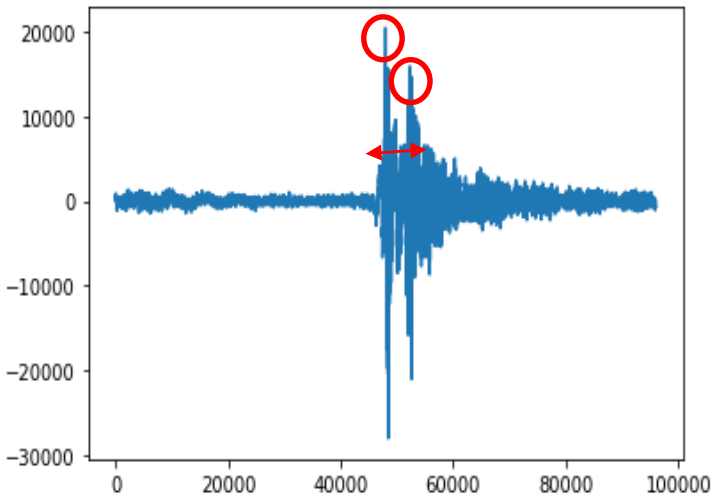
Vehicle Classification based on Audio Signal

- Extraction of statistical features by using SciPy (https://scipy.org/) for training.

Max Peak:



Temporal distance between peaks



The End