Digital Asset Management

2008 Compact Version

张宏鑫 zhx@cad.zju.edu.cn

What is DAM?

大不自多
海纳江河

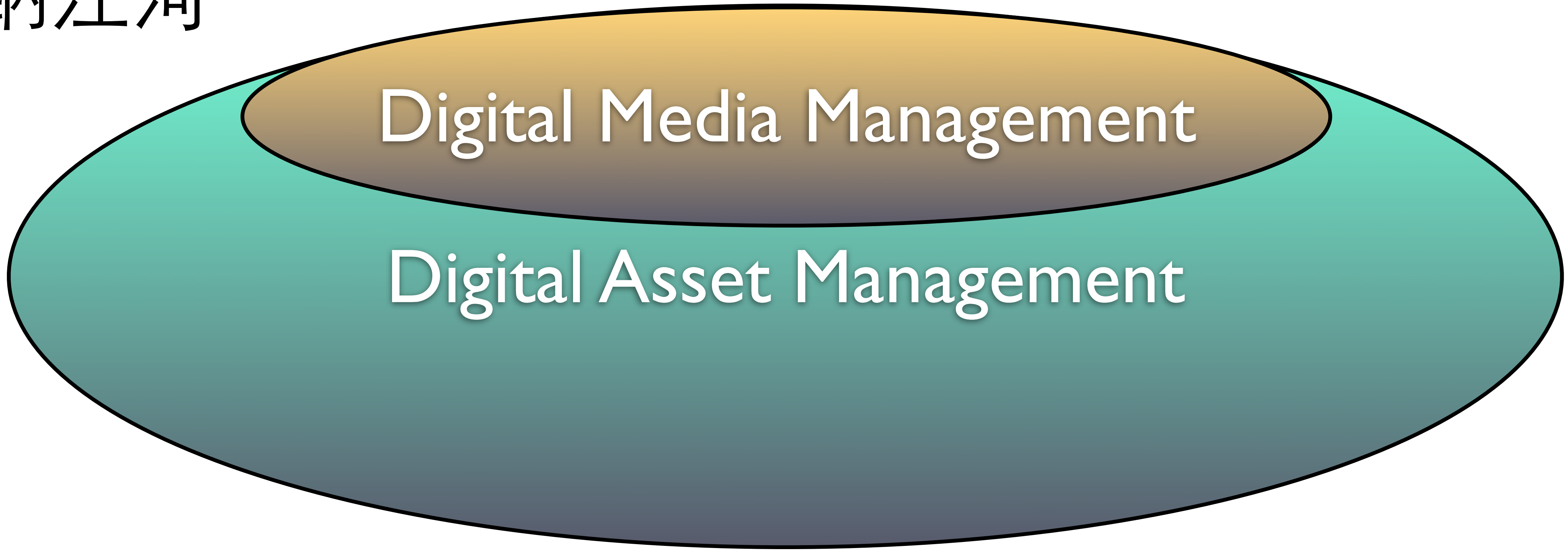
What is DAM?

大不自多
海纳江河

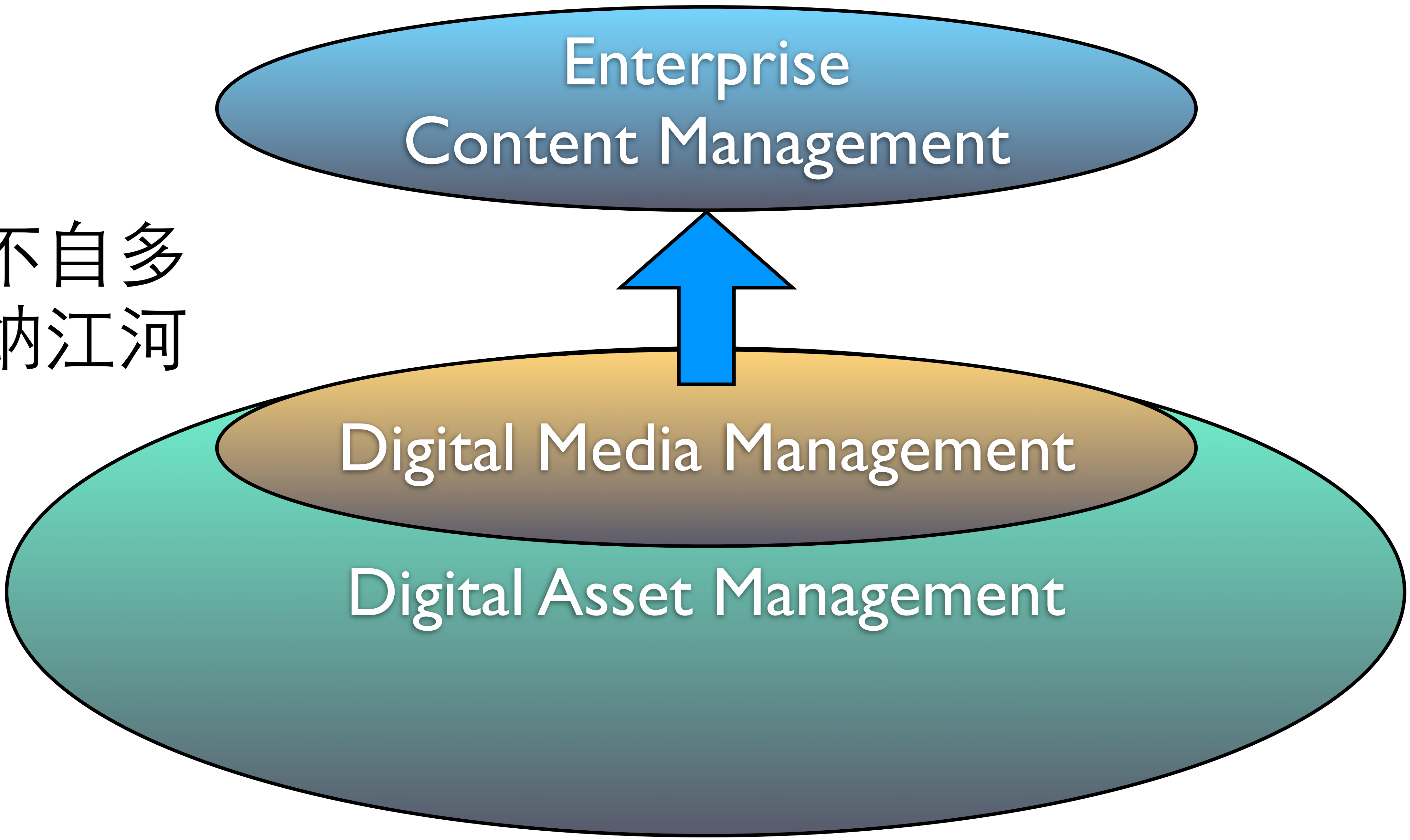
Digital Asset Management

What is DAM?

大不自多
海纳江河



What is DAM?



大不自多
海纳江河

What is Digital Asset Management?

- Tool for organizing, storing and retrieving content in digital format
  - downloading, renaming, backing up, rating, grouping, archiving, optimizing, maintaining, thinning, and exporting ...
- Includes:
  - text, video, images, movies, sound, and 3D content

What Can DAM Do for You ?

What Can DAM Do for You ?

- Catalog large numbers of formats

What Can DAM Do for You ?

- Catalog large numbers of formats
- Create a visual category using thumbnails

What Can DAM Do for You ?

- Catalog large numbers of formats
- Create a visual category using thumbnails
- Add keywords, data fields

What Can DAM Do for You ?

- Catalog large numbers of formats
- Create a visual category using thumbnails
- Add keywords, data fields
- All fields can be searched

What Can DAM Do for You ?

- Catalog large numbers of formats
- Create a visual category using thumbnails
- Add keywords, data fields
- All fields can be searched
- Select images for an electronic gallery - specific lecture topics

What Can DAM Do for You ?

- Catalog large numbers of formats
- Create a visual category using thumbnails
- Add keywords, data fields
- All fields can be searched
- Select images for an electronic gallery - specific lecture topics
- Share over the internet

What is Digital Asset?

Media file format

- Image
- Video
- Audio
- Document
- Graphics

特点与标准?

兼总条贯，知至知终

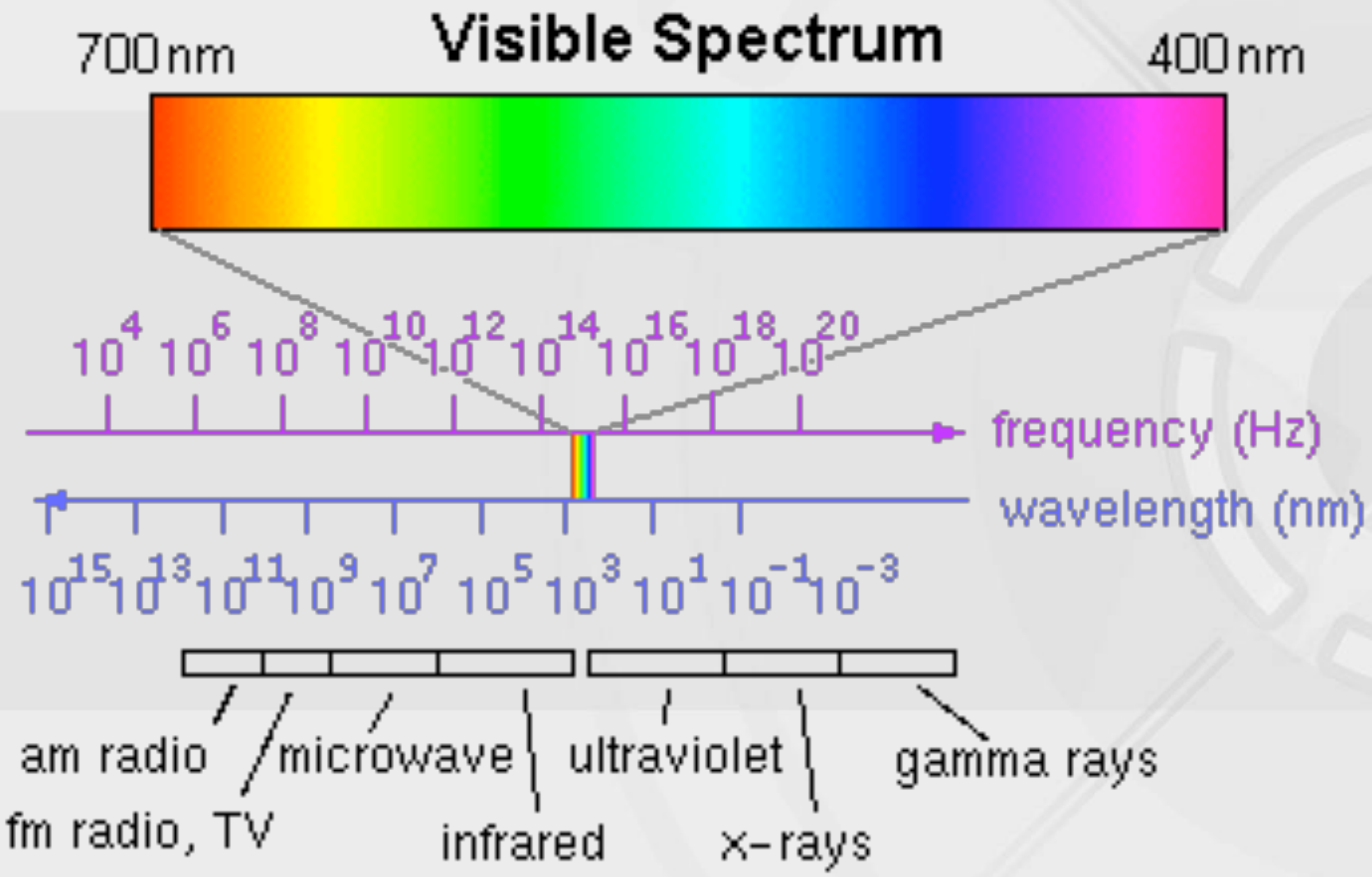
Digital media data types

Table. File format used in Macromedia Director

File import					File export		Native
Image	Palette	Sound	Video	Animation	Image	Video	
BMP GIF, JPG, PICT, PNG, PNT, PSD, TGA, TIFF, WMF	PAL ACT	AIFF AU MP3 WAV	AVI MOV	DIR FLA FLC FLI GIF PPT	BMP	AVI MOV	DIR DXR EXE

Color systems and color models

- RGB
- XYZ
- Lab
- YUV
- HSV



RGB vs. XYZ

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{b_{21}} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Since 1931

YUV color spaces

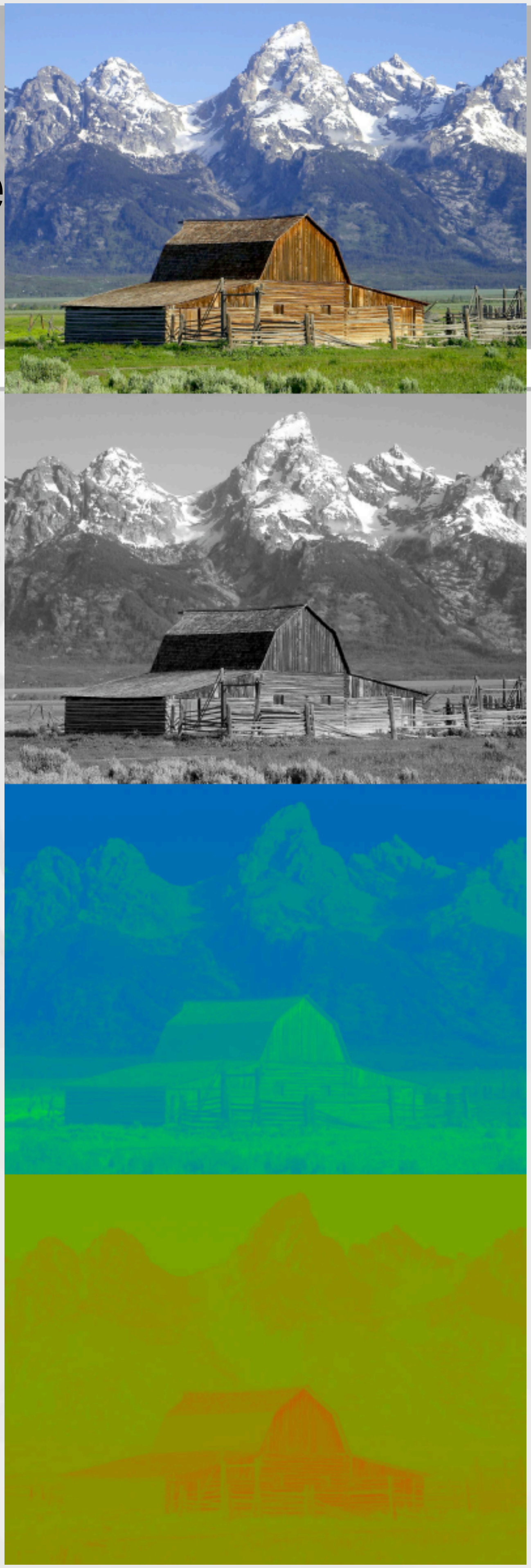
- used in most video capture system
- PAL television system

Image

Y

U

V



YUV color spaces

- used in most video capture system
- PAL television system

$$\begin{bmatrix} Y' \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.14713 & -0.28886 & 0.436 \\ 0.615 & -0.51499 & -0.10001 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Image

Y

U

V

Representation of digital images

- An image can be viewed as a ***N×M*** vector matrix

- Grayscale image
- Color image
- Palette

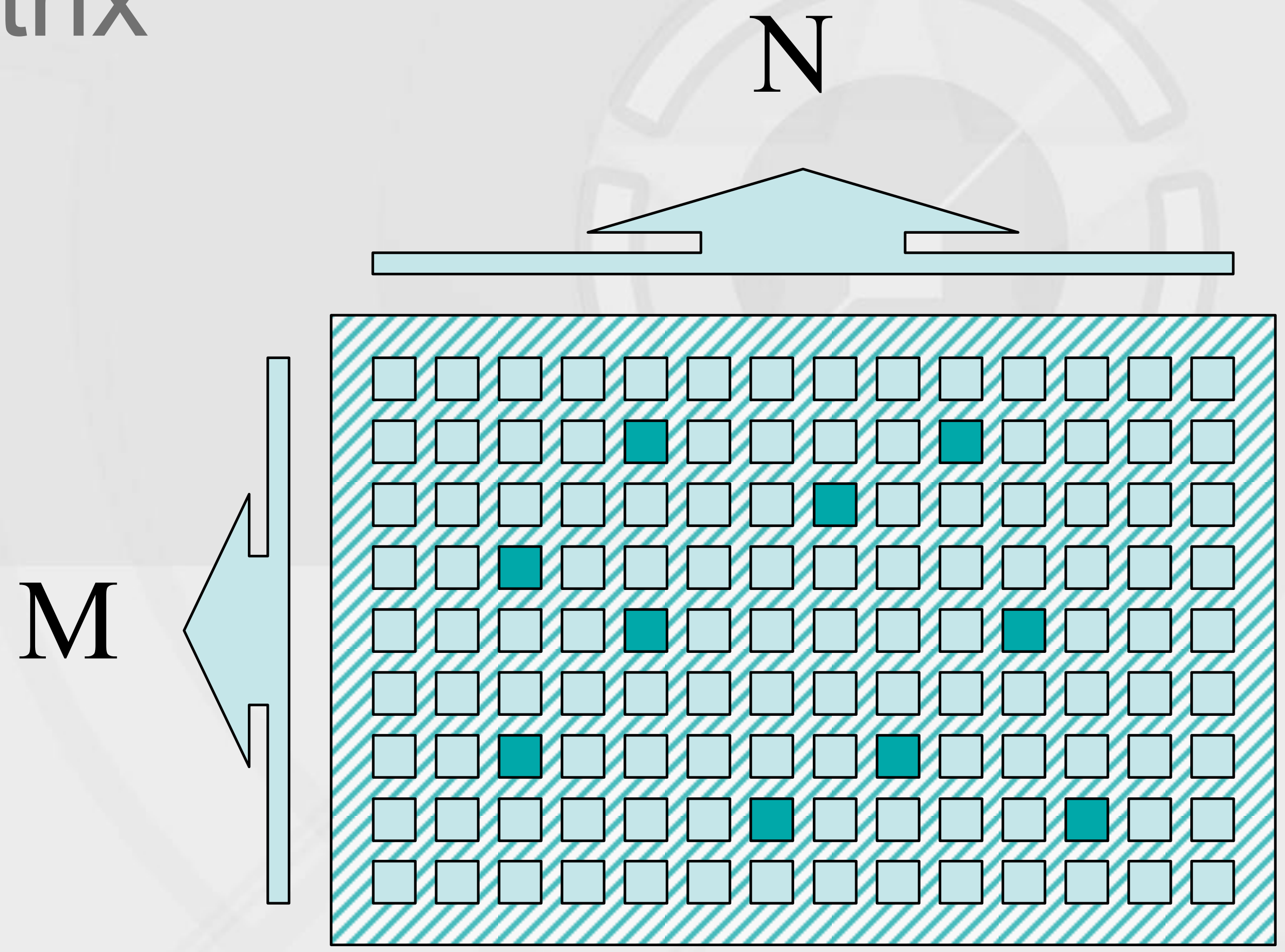


Image resolutions

Dimensions	MEGA pixels	Name	Comments
640x480	0.3	VGA	
720x576	0.4	CCIR 601 DV PAL	PAL DV, and PAL DVDs
768x576	0.4	CCIR 601 PAL full	PAL with square sampling grid ratio
800x600	0.4	SVGA	
1024x768	0.8	XGA	
1280x960	1.2		
1600x1200	2.1	UXGA	
1920x1080	2.1	1080 HDTV	high resolution digital TV format
2048x1536	3.1	2K	Typically used for digital effects in feature films.

Rep of Images

- Most used color images
  - –24bit RGB
  - –Red/Green/Blue each channel has 256 degrees of grayscale
  - –Can represent 2^24 = 16,777,216 types of color

Image compression methods

- lossless compression
- lossy compression

Lossless image compression methods

- Based on information theory
- General encoding methods
  - –RLC (Run-Length Coding)
  - –VLC (Variable-Length Coding)
  - –Dictionary Coding
  - –Arithmetic Coding

Lossy image compression methods

- Quantization
- Transform coding
  - –Discrete Cosine Transform
  - –Discrete Wavelet Transform
  - –Karhune-Loeve Transform (Principle component analysis)

Image compression standards

- JPEG
  - –Joint picture encoding group
  - –Discrete Cosine Transform
- JPEG 2000
  - –newer standard
  - –Discrete Wavelet Transform

DNG: Digital Negative

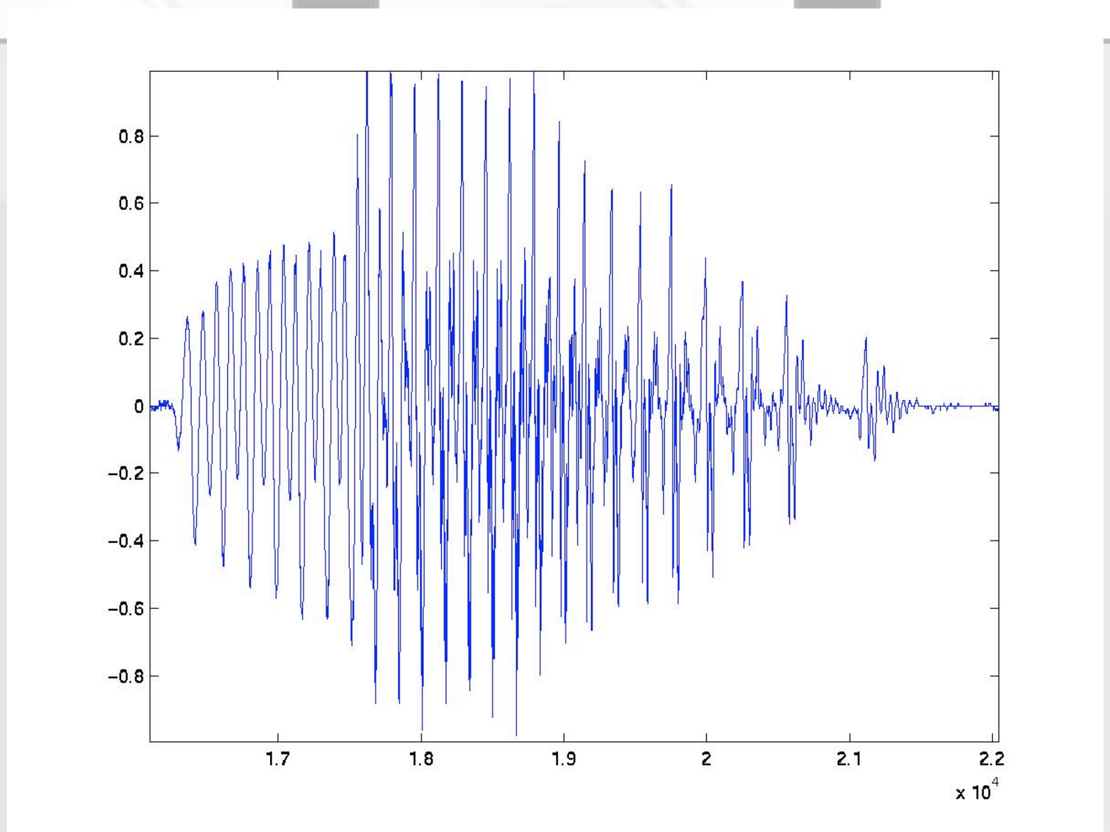
- a royalty free RAW image format
- design by Adobe
- based on TIFF/EP
- mandates use of metadata

Summary – Essential factors of image storage

- Resolution
- Compression rate
  - 1bpp,2bpp, …
  - Compression methods
- Color representation
  - RGB, YUV, Lab …

Digitalized audio / sound

- What is sound?
  - Knowing from ear?!?
  - Sound wave ?!?



- Digitalization
  - Analog signal → digital signal
  - Quantization

Bit rate and bit

- a kind of energy wave.
- a continuous function of wave amplitude
  - Sequence is related to the X axis (the time line).
  - Amplitude is related to the Y axis.

Higher coding rate and quantization rate, better sampling quality

- discretely sampled during the digital coding period
  - Bit rate: number of samples obtained in one second
    - The highest frequency ~ 20kHz.
    - 40k samples per second (Nyquest theorem)
    - The bit rate of CD is 44.1kHz
  - Quantization rate: must be the power of 2.
    - The quantization rate of audio CD is normally 16bit.

Audio compression: lossless vs. lossy

- There is no absolute looseness coding schemes!
  - According to the definitions of bit rate and quantization rate, audio coding can only approximate to the natural sound signal as much as possible.
  - Comparing with natural signal, all coding schemes are lossy.

- Related looseness scheme: PCM
  - PCM can reach the highest preserving level.
  - widely applied in raw data saving and music data, e.g. CD、DVD and WAV files.
  - PCM is viewed as a looseness coding scheme. How, PCM only approximate to the raw data.
  - Comparing with the PCM coding method，we usually put MP3 coding methods into the lossy audio encoding methods.

PCM coding

- PCM - **P**ulse **C**ode **M**odulation

- PCM coding
  - Advantage： good play back quality.
  - Shortage: large storage space.

- Audio CD mainly leverage the PCM coding scheme. One piece of CD can store 72 minutes music.

PCM audio stream bit-rate

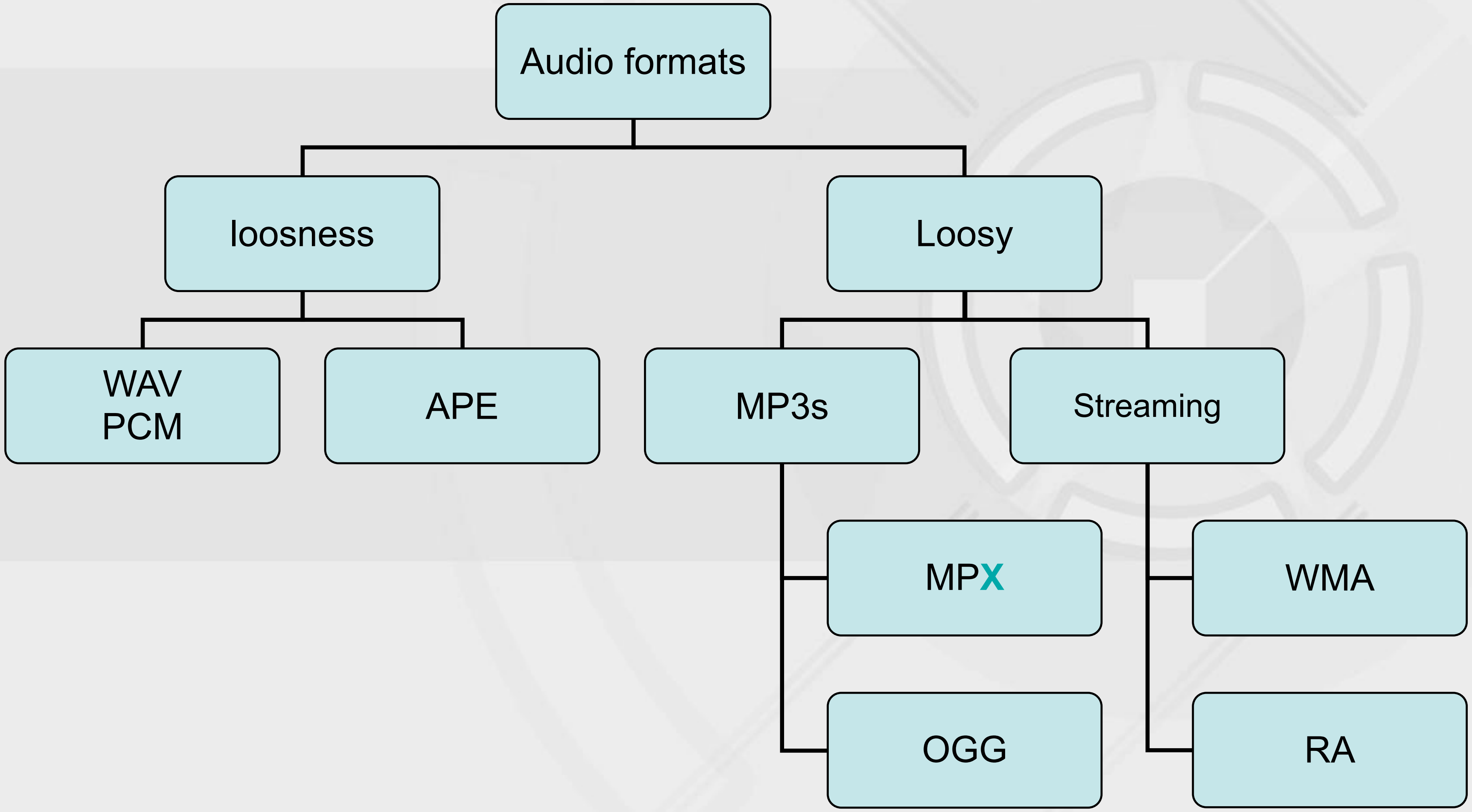
- Formula
  - Bit rate × Quantization rate × number of sound channels (bps)。

- EXAMPLE：
- WAV file: bit rate 44.1KHz，quantization rate 16bit，stereo sound.
  - Coding rate: 44.1K×16×2 =1411.2 Kbps.
  - 128K MP3 ~ 1411.2 K bits per second
  - also called data width, similar to the concept of band width used in network transfer.

  - Data speed: transferred bytes per second, = Bit rate / 8. In this example, the speed is 176.4KB/s.
  - It takes space of 176.4KB per second. Recording 1 minute music requires 10.34M.

Common audio formats

- WAV
- MP3
- WMA
- RA
- OGG
- APE



什么是MIDI

- MIDI (Musical Instrument Digital Interface即乐器数字化接口) is an international standard for general interface.
  - It provides a set of standard interface for transferring data among different types of devices. MIDI devices shall precisely send MIDI messages.

- Wildly use in music creation, game background music and ring tone of mobile phones.

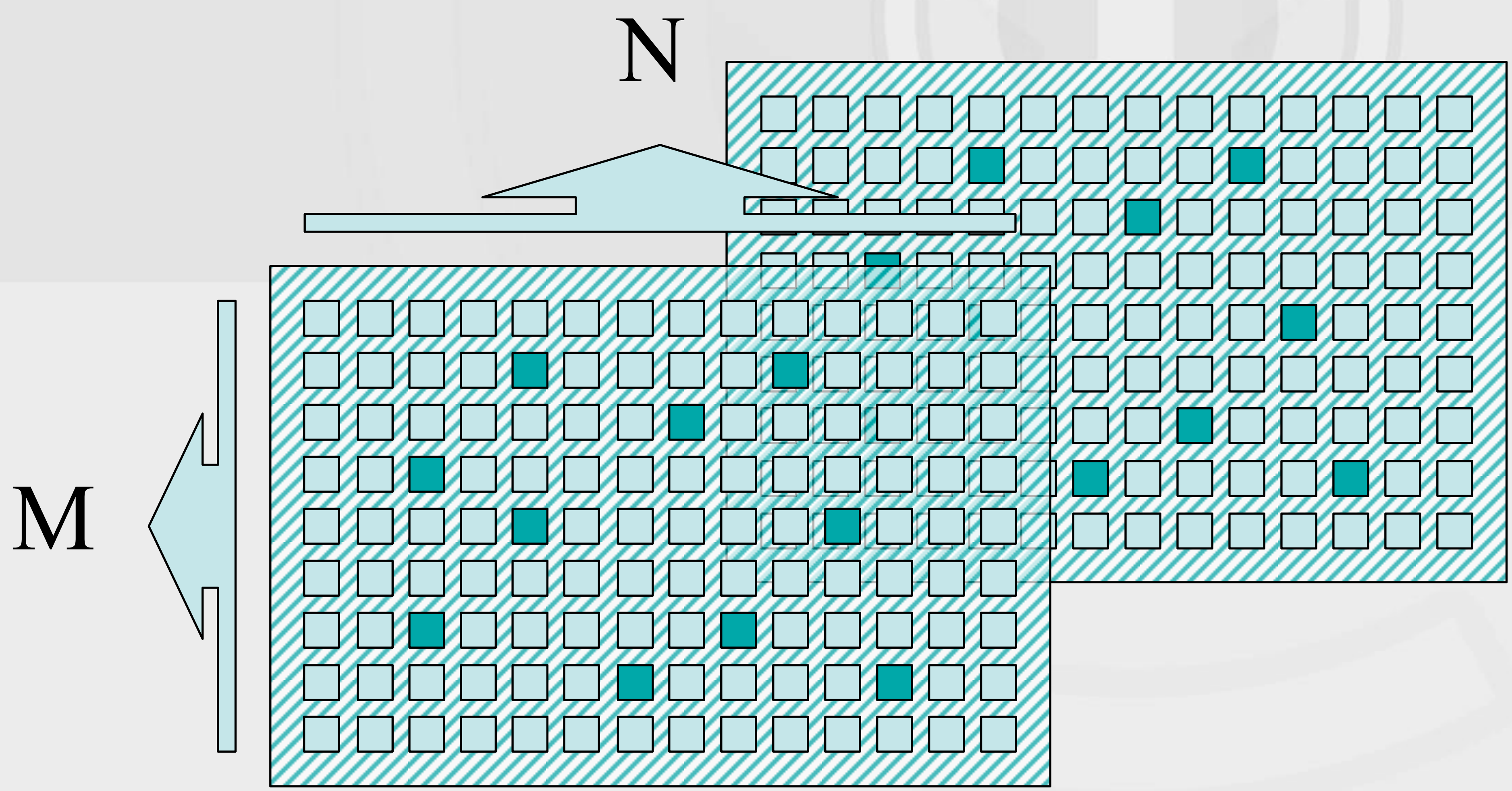
MIDI概况

- MIDI is type of description language.
  - Different directly record digitalized sound signal
  - Only record ‘events’ that how instruments make sound.
  - Small storage size.

- Three elements of MIDI
  - Synthesizer
    - Generate sound and can control the length, height, strength and other features of sound.
  - Sequencer
    - Devices or software that store and modify MIDI information.
  - MIDI device
    - Do not generate any sound but a sequence of MIDI commands.
    - E.g. MIDI keyboard, MIDI harp, MIDI guitar, and MIDI violin, etc.

Representations of video

- Sequence of images？！？
  - –Can be viewed as a 3-dimentional matrix
  - –But it is only 50％ correct



Common video formats

- AVI (Microsoft, Divx, …)
  - –avi, wmv, asf
- RM (Realplayer)
  - –rm, rmvb
- MOV (Quicktime)
  - –mov
- MPEG
  - –MPEG-1, MPEG-2, MPEG-4 …

http://www.bigbuckbunny.org/index.php/download/

Video compression standards

- MPEG standards
  - Audio/Video compression, storage and play back standards
  - MPEG-1: VCD
  - MPEG-2: broadcast TV, e.g., DVD、HDTV etc.
  - MPEG-3: replaced by MPEG-2
  - MPEG-4: network video transfer，stream media
  - MPEG-7:
  - MPEG-21:

- ITU-T H.26x series

Video compression standards

- ITU-T H.26x series
  - Mainly used in video communication applications
  - Now it has H.261，H.262，H.263，H.264

  - ISDN network based H.320 standards
    - the video compression part: H.261, H.262 and H.263

  - LAN network based H.323
  - PSTN network based H.324
    - the video compression part: H.261 and H.263

MPEG概况

- MPEG = **M**otion **P**icture **E**xpert **G**roup

- **ISO/IEC JTC1/SC29**
  - WG11:Motion Picture Experts Group (MPEG)
  - WG10:Joint Photographic Experts Group (JPEG)
  - WG7: Computer Graphics Experts Group (CGEG)
  - WG9: Joint Bi-level Image coding experts Group (JBIG)
  - WG12: Multimedia and Hypermedia information coding Experts Group (MHEG)

MPEG概况

- MPEG-1,2 standards were started at 1988
  - 需求 [Requirement]
  - 系统 [System]
  - 视频 [Video]
  - 音频 [Audio]
  - 实现 [Implementation]
  - 测试 [Testing]
- Newest MPEG standards: MPEG-4, MPEG-7, MPEG-21

MPEG-1 Standard ISO/IEC 11172-2 (1991)

"Coding of moving pictures and associated audio for digital storage media"

- **Video**
  - optimized for bit rates around 1.5 Mbit/s
  - originally optimized for SIF picture format,
  - but not limited to it:
    - [**NTSC based**] : 352x240 pixels at 30 frames/sec
    - [**PAL based**] : 352x288 pixels at 25 frames/sec
  - progressive frames only
    - no direct provision for interlaced video applications, such as broadcast television

MPEG-1 Standard ISO/IEC 11172-2 (1991)

- Audio
  - joint stereo audio coding at 192 kbit/s (layer 2)
- System
  - mainly designed for error-free digital storage media
  - multiplexing of audio, video and data
- Applications
  - CD-I, digital multimedia, and
  - video database (e.g. video-on-demand)

MPEG-2 Standard ISO/IEC 13818-2 (1994)

- Video
  - 2-15 or 16-80 Mbit/s bit rate (target bit rate: 4…9 Mbit/sec)
  - TV and HDTV picture formats
  - Supports interlaced material

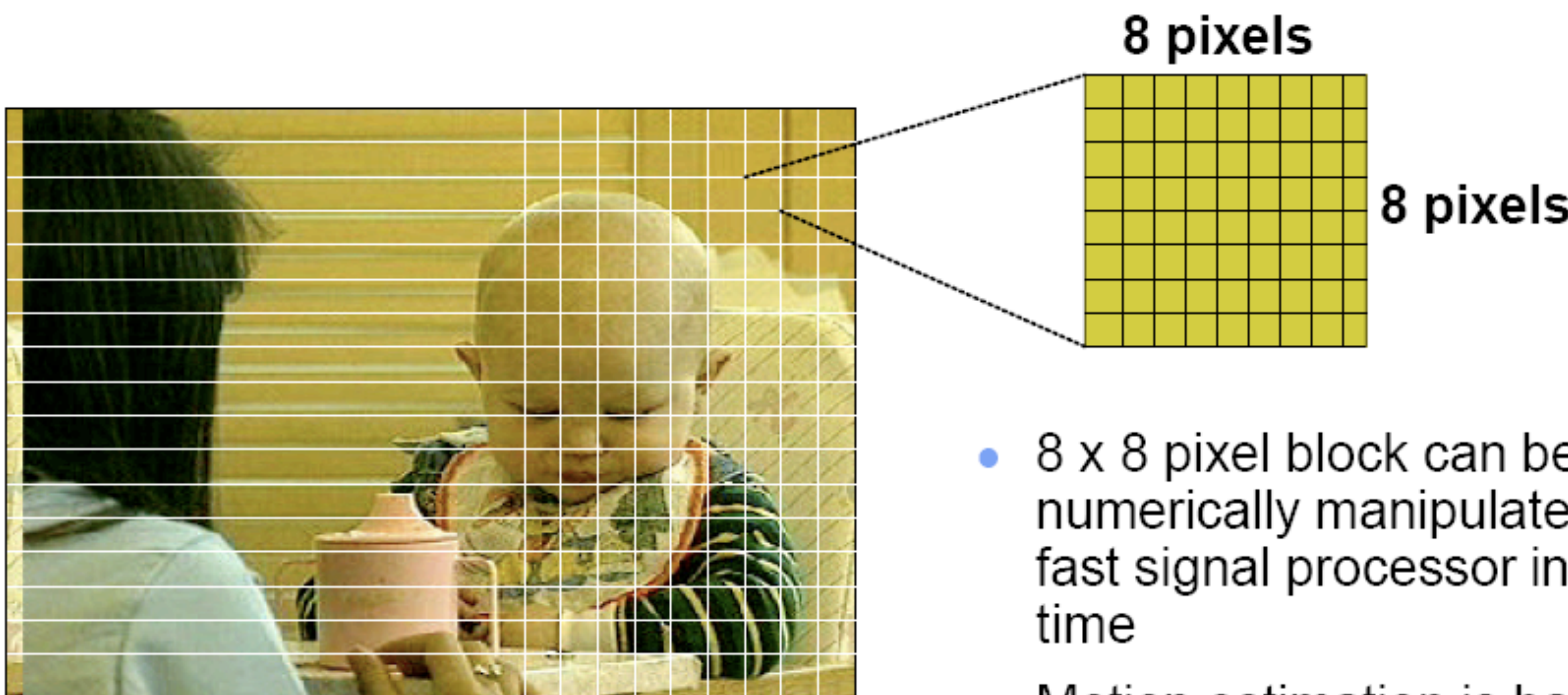
  - MPEG-2 consists of *profiles* and *levels*
    - Main Profile, Main Level (MP@ML)
      - 720x480 resolution video at 30 frames/sec
      - < 15 Mbit/sec (typical ~4 Mbit/sec)
      - for NTSC video

    - Main Profile, High Level (MP@HL)
      - 1920x1152 resolution video at 30 frames/sec
      - < 80 Mbit/sec (typical ~15 Mbit/sec)
      - HDTV

MPEG-2 Standard ISO/IEC 13818-2 (1994)

- Audio
  - compatible multichannel extension of MPEG-1 audio
- System
  - video, audio and data multiplexing defines tow presentations:
    - ***Program Stream*** for applications using near error free media
    - ***Transport Stream*** for more error prone channels
- Applications
  - satellite, cable, and terrestrial broadcasting,
  - digital networks, and
  - digital VCR

MPEG compression is based on 8 x 8 pixel block processing



- 8 x 8 pixel block can be numerically manipulated by fast signal processor in real time
- Motion estimation is based on comparing the blocks between series of pictures

MPEG: only compress moving parts

new picture

-

previous picture

=

difference

Encoder

Decoder

difference

+

previous picture

=

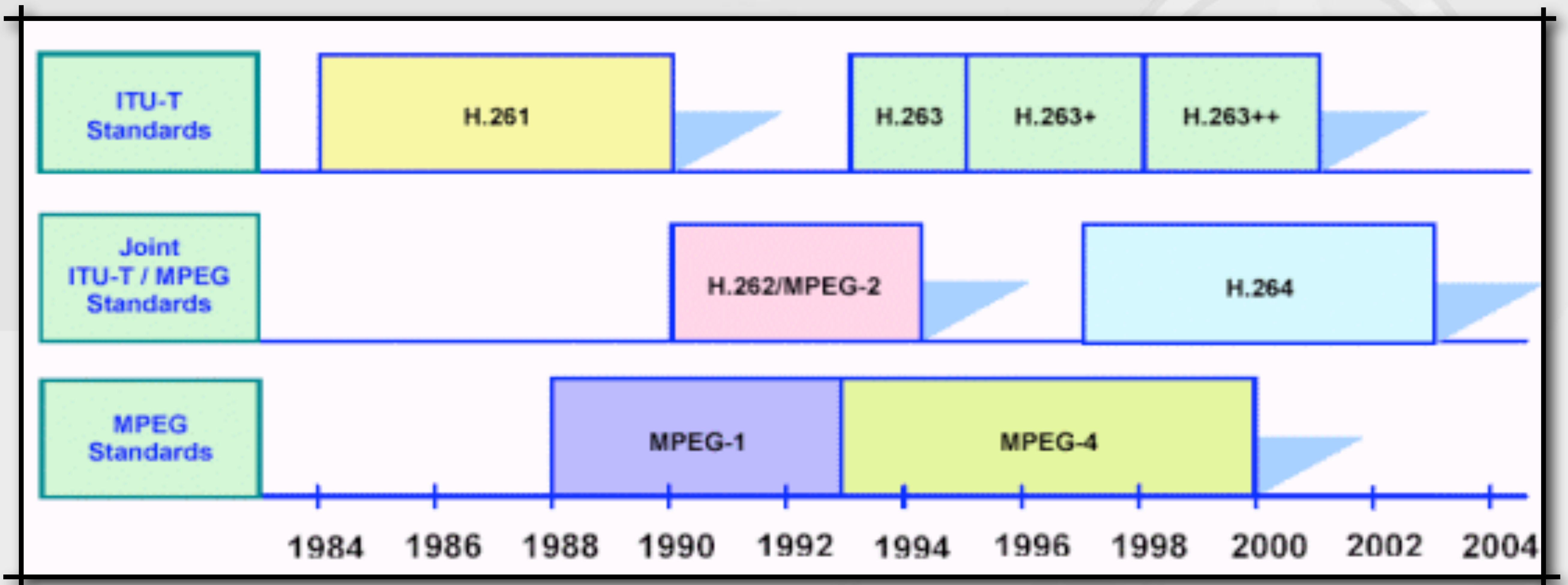
new picture

Overview of H.264

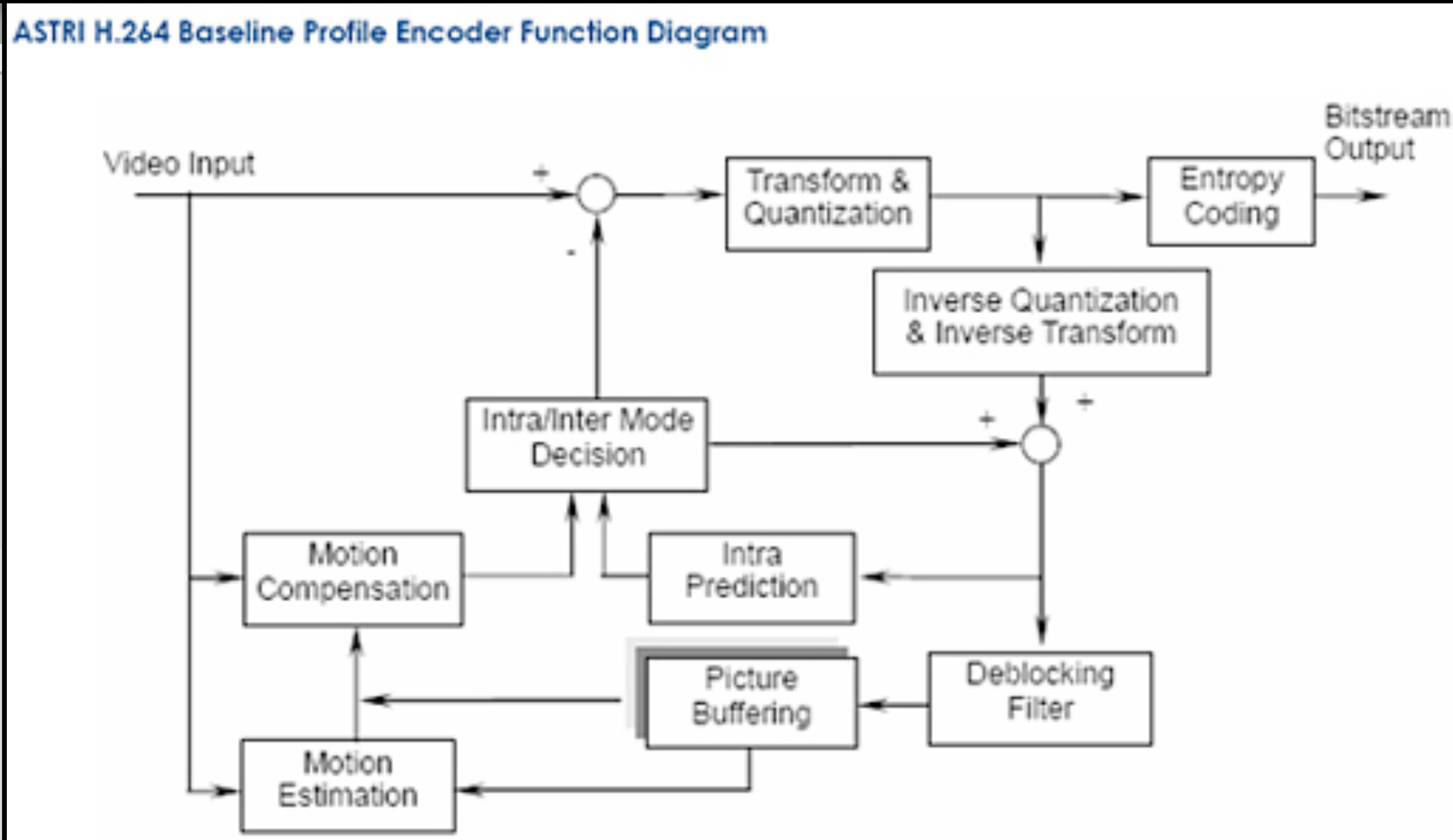
- JVT（Joint Video Team）
  - founded on December 2001, Pattaya Thailand.
  - video coding specialists from ITU-T and ISO, the two international standards organizations
  - goal: define a new video coding standards to achieve high compression rate, high image quality, good network adaptive coding frame.

- H.264: A new video compression standard
  - accepted by ITU-T
  - accepted by ISO
    - called AVC (Advanced Video Coding) standard
    - as the 10th part of MPEG-4

Major history of digital video standard



H.264 coding principle



H.264的主要技术特点

1. 4类DCT整数变换以及相应的量化方法

2. 7种宏块预测模式
   - 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4
   - 运动估计和补偿更加精确

3. 多参考帧
4. 帧内预测
5. 改进的去块效应滤波器（Deblocking filter）
6. 增强的熵编码方法
   - UVLC（Universal VLC）、CAVLC（Context adaptive VLC）和CABAC

7. 1/4像素插值
8. 宏块级逐行、隔行自适应编码MBAFF

Advantages and shortages of H.264

- ☑ High compression rate
  - In the same image quality, H.264 can be compressed as size of
    - 36% of MPEG-2， 61% of MPEG-4， 51% of H.263
  - Low bit stream, high quality
- ☑ High error correctness rate
  - H.264 provides necessary tools to solve the error coding problem in unstable network environments
- ☑ Network adaptation
  - H.264 provides Network Adaptation Layer so as to make files of H.264c can be easily transferred in different network environments.
- ☐ High computation price
  - In the same image quality, H.264 is twice of MPEG-2 in computation complexity.

Overview of HTML

- Hypertext Markup Language
  - Developed by Tim Berners-Lee.
    - lightweight markup language vs. complex SGML.
  - Based on pure text format

  - Rich abilities to display multimedia information.
    - Later added tags to support image and videos.

  - HTML 3.2 => HTML 4.0 => HTML 5.0
    - Different browser has their own display effects.

Overview of all HTML elements

Reference: http://htmlhelp.com/reference/wilbur/overview.html

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
   <title>Apple中国</title>
   <meta http-equiv="content-type" content="text/html;
charset=gb2312">
   ...
</head>

<body>
<!-- Tag for Activity Group: General, Activity: Apple China -
Homepage -->
...
</body>
</html>
```

http://www.apple.com.cn/

Overview of all HTML elements

Reference: http://htmlhelp.com/reference/wilbur/overview.html

Head → {

Body → {

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
   <title>Apple中国</title>
   <meta http-equiv="content-type" content="text/html;
charset=gb2312">
   ...
</head>

<body>
<!-- Tag for Activity Group: General, Activity: Apple China -
Homepage -->
...
</body>
</html>
```

http://www.apple.com.cn/

Overview of HTML - Head elements

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
     <title>Apple中国</title>
     <meta http-equiv="content-type" content="text/html; charset=gb2312">
     ...
     <link rel="home" href="http://www.apple.com/">
     ...
     <script src="http://images.apple.com/global/scripts/lib/prototype.js" type="text/javascript" charset="utf-8">
     </script>
     ...

     <style type="text/css" media="all">
     ...
     #billboard { width: 1100px; margin: 0 auto 15px; overflow: hidden; position: relative; }
     #ticker { margin-bottom: 15px; }
     #homefooter { margin: 60px auto 50px; }
     ...
     </style>
</head>
```

- TITLE - Document title
- ISINDEX - Primitive search
- META - Meta-information
- LINK - Site structure
- BASE - Document location
- SCRIPT - Inline script
- STYLE - Style information

Overview of XML

- Extensible Markup Language
  - Aim at data searching

  –Similar to HTML
    - More restrict grammar checking
    - User defined tags to describe data structure
    - Flexible data displaying schemes
    - Cross-platform, language and application independent
    - DTD and XML Schema.

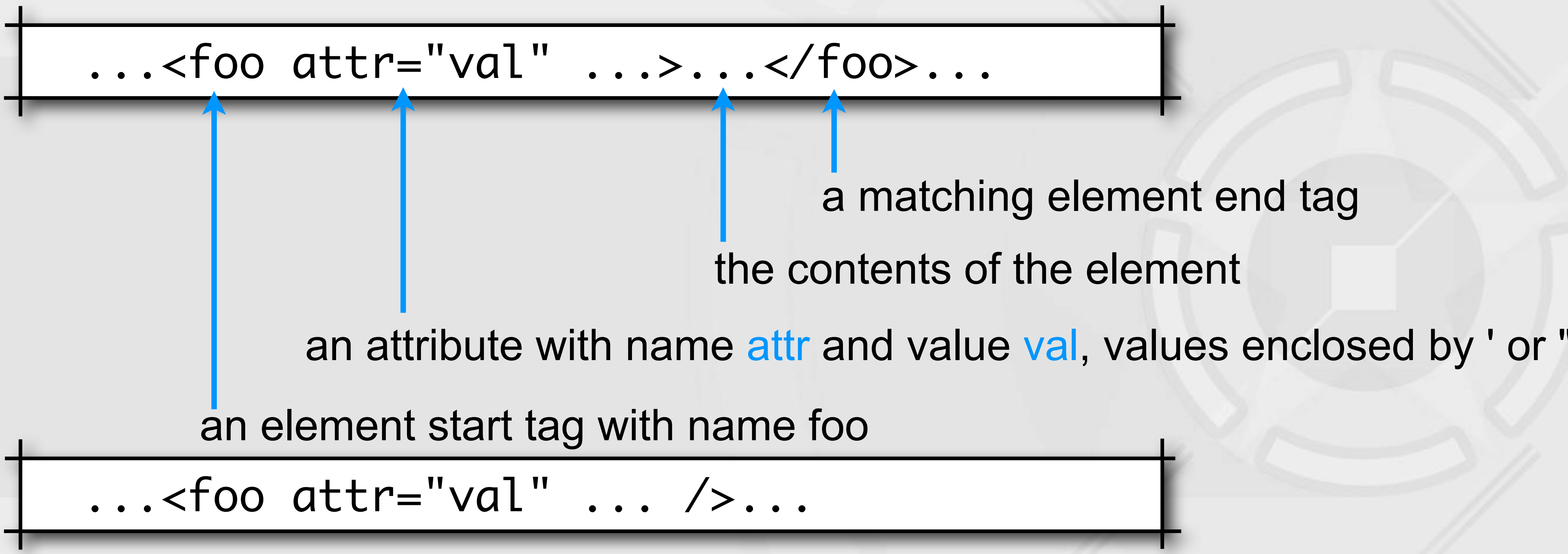
- http://www.brics.dk/~amoeller/XML/overview.html

HTML v.s. XML

```
<h1>Rhubarb Cobbler</h1>
<h2>Maggie.Herrick@bbs.mhv.net</h2>
<h3>Wed, 14 Jun 95</h3>

Rhubarb Cobbler made with bananas as the main sweetener.
It was delicious.  Basicly it was
  <table>
  <tr><td> 2 1/2 cups <td> diced rhubarb
  <tr><td> 2 tablespoons <td> sugar
  <tr><td> 2 <td> fairly ripe bananas
  <tr><td> 1/4 teaspoon <td> cinnamon
  <tr><td> dash of <td> nutmeg
  </table>
Combine all and use as cobbler, pie, or crisp.
Related recipes: <a href="#GardenQuiche">Garden Quiche</a>
```

```
<recipe id="117" category="dessert">
  <title>Rhubarb Cobbler</title>
  <author><email>Maggie.Herrick@bbs.mhv.net</email></author>
  <date>Wed, 14 Jun 95</date>

  <description>
    Rhubarb Cobbler made with bananas as the main sweetener.
    It was delicious.
  </description>

  <ingredients>
    <item><amount>2 1/2 cups</amount><type>diced rhubarb</type></item>
    <item><amount>2 tablespoons</amount><type>sugar</type></item>
    <item><amount>2</amount><type>fairly ripe bananas</type></item>
    <item><amount>1/4 teaspoon</amount><type>cinnamon</type></item>
    <item><amount>dash of</amount><type>nutmeg</type></item>
  </ingredients>

  <preparation>
    Combine all and use as cobbler, pie, or crisp.
  </preparation>

  <related url="#GardenQuiche">Garden Quiche</related>
</recipe>
```

A conceptual view of XML



XML documents as text with markup

A conceptual view of XML

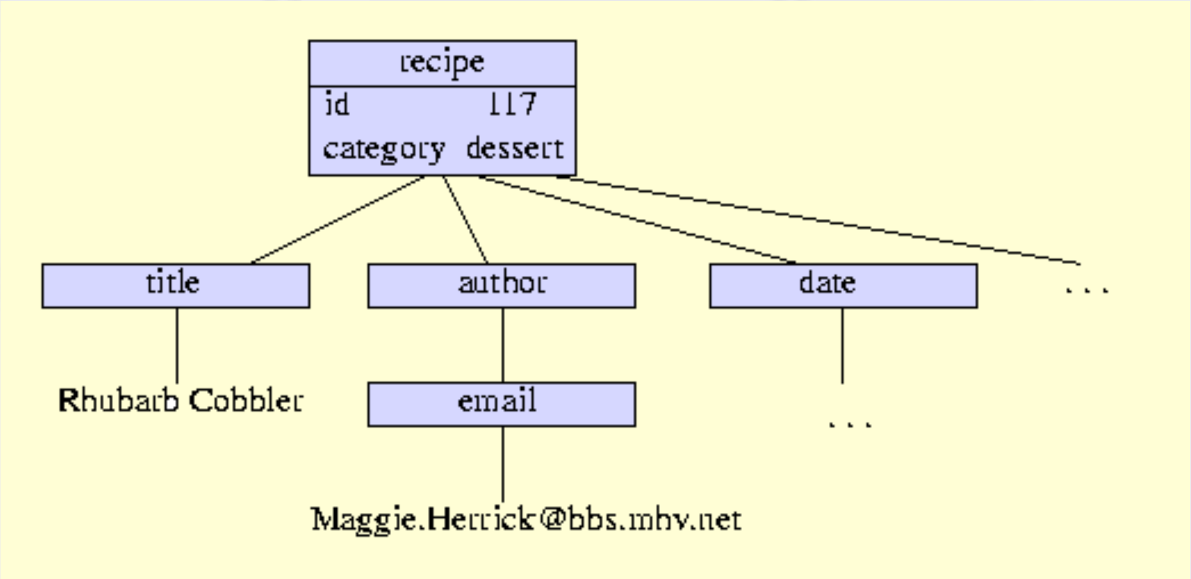
- An XML document is a (Unicode) text with markup tags and other meta-information.
- An XML document must be well-formed:
  - start and end tags must match
  - element tags must be properly nested
  - + some more subtle syntactical requirements

- XML is case sensitive!
- Special characters can be escaped using Unicode character references:
  - < and < both yield <

A conceptual view of XML

- An XML document is an ordered, labeled tree:
  - character data leaf nodes contain the actual data (text strings)
    - usually, character data nodes must be non-empty and non-adjacent to other character data nodes
  - elements nodes, are each labeled with
    - a name (often called the element type), and
    - a set of attributes, each consisting of a name and a value,
    - and these nodes can have child nodes

XML documents as labeled trees



A conceptual view of XML

- XML trees may contain other kinds of leaf nodes:
  - processing instructions - annotations for various processors
  - comments - as in programming languages
  - document type declaration

XML documents as labeled trees

- The XML vision offers:
  - common extensions to the core XML specification
    - a namespace mechanism, document inclusion, etc.
  - schemas
    - grammars to define classes of documents
  - linking between documents
    - a generalization of HTML anchors and links
  - addressing parts of read-only documents
    - flexible and robust pointers into documents
  - transformation
    - conversion from one document class to another
  - querying
    extraction of information, generalizing relational databases

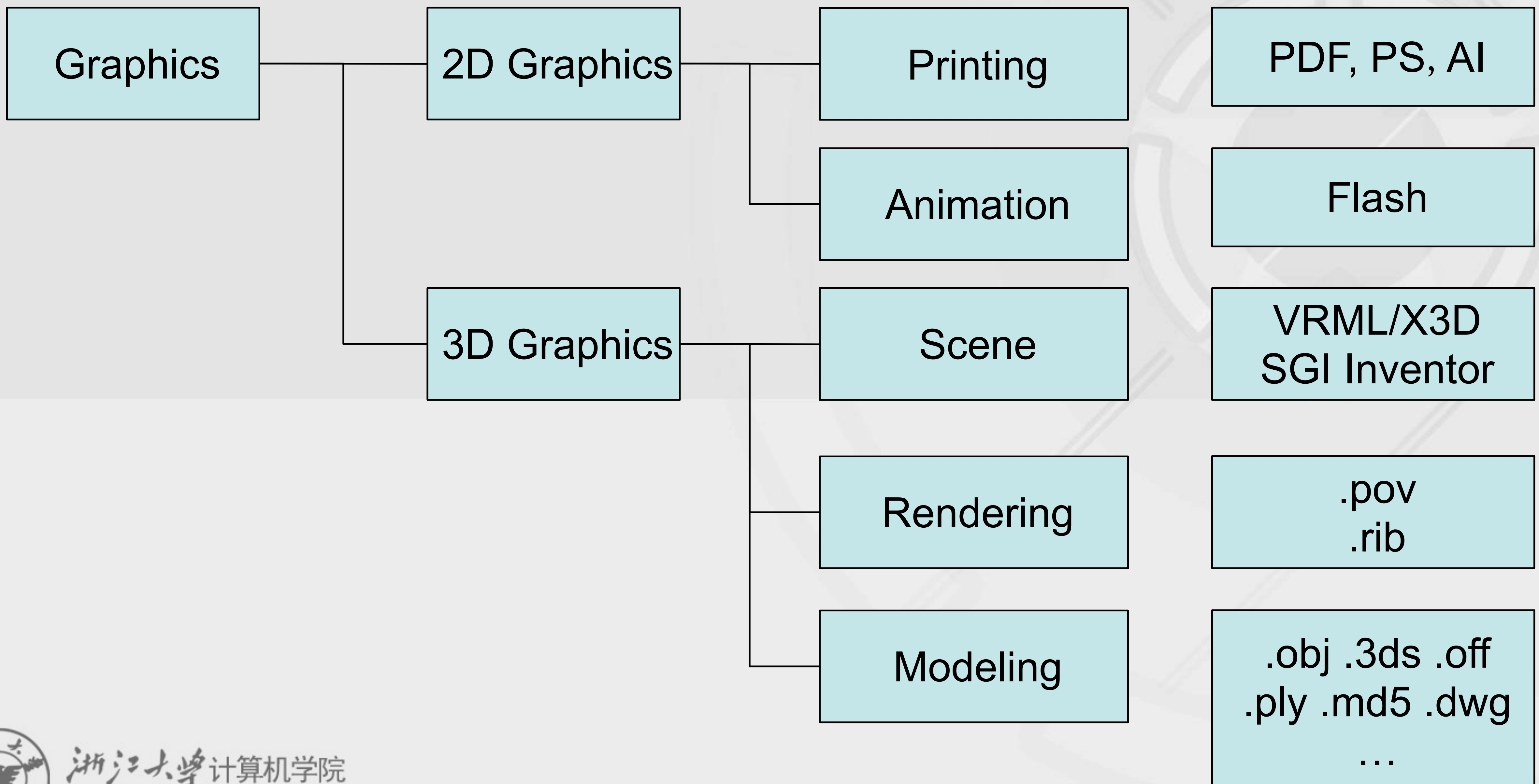
To use XML

- Define your XML language
  - use XML Schema to define its syntax
- Exploit the generic XML tools
  - XSLT and XQuery processors
- As a generic protocols, and the generic programming frameworks
  - DOM or SAX to build application tools

Graphics ≠ Images

- Representation ability
  - Graphics are usually described in vectors which can provide arbitrary precision
  - Images are usually sampled in fragments/pixels which can only provide limited precision
- Application area
  - Graphics are mainly applied in CAD, model design, computer animation, system simulation and printing.
  - Images are mainly used for photo display and image processing etc.

Classification of different graphics formats



Elements of 3D graphics format

- Global scene description
  - Parameters of light and camera, other system configurations
- Geometric model description
  - Curves and surfaces
    - Line, plane, quadratic surface, spline …
  - Mesh surfaces ＝ vertex coordinates＋ topology connectivity
  - Texture coordinates, normals
- Material description
  - Reflectance model, texture image
- Animation description
  - Skeleton model …

Overview of X3D

- X3D [Extensible 3D] is an international standard of 3D graphics. It defines how to integrate and access interactive 3D content in a multimedia environment.

- The former of X3D is VRML which is established on 1998 as a network graphics ISO standard (ISO/IEC14772).

- X3D decompose scene descriptions of VRML97 into components. Therefore it is very convenient to extend original VRML functions by adding new components.

Multimedia database

- MM content organization
- MM database system architecture
- MM system service model
- Multimedia Data Storage
- Multimedia application

无曰已是，无曰虽真，
靡革匪因，靡故匪新，何以新之，开物前民

What's Multimedia?

- Multi => Many

- Media:
  - A means to distribute and represent information: Text, graphics, pictures, voice, sound and music..
    - Perception media (how do humans perceive information?)
      - Audio/visual media
    - Representation media (how is information encoded?)
      - ASCII, JPG, MPEG, PAL.
    - Presentation media (medium used for output/input)
      - Input/output media (keyboards, papers)
    - Storage media (Where is information stored?)
      - Magnetic disk, optical disk

What is Multimedia?

- Serious definitions:

  "From a user's perspective - multimedia enables computer information to be represented through audio, video, text, images, graphics and animation."

  "Multimedia is defined as an interactive computer-mediated presentation that included at least two of the following elements: text, sound, still graphic images, motion graphics and animation."

What is Multimedia?

- Multimedia involves **Many Media**
  - **THE MEDIA DOMAIN**
- Multimedia involves **Computers**
  - **THE SYSTEMS DOMAIN**
- Multimedia enhances the **presentation and communication of information**
  - **THE APPLICATION DOMAIN**

Classification of Media Types

- Media types can be divided into two groups:
  - Temporal (Continuous media)
    - Time or more exactly time-dependency between information items, is part of the information itself.
    - dynamic, time-based, continuous
    - e.g., audio, video, music, animation
  - Non-temporal (Discrete media)
    - Time is not part of the semantics of the media.
    - static, non-time-based, discrete
    - e.g., text, graphics, images

Challenges about managing MM data

- Huge Size
- Quality of Service (QoS)
- Synchronization
- Content-Based Retrieval

Characteristics of MM Data
-- challenges about managing MM data

- ***Huge Size***
  - multimedia objects are large in size (compared with traditional alphanumerical data) and are not readily accommodated by 'old' DBMSs
  - E.g.,
    - a video – each second contains, say, 30 frames; each frame may require, depending on video quality, several megabytes of storage.
    - a color picture of 1280 x 960 pixels using 32-bit color requires about 5MB of memory

≈ 4,915,200 = 1280 x 960 x 4 *bytes*

- Needs fast and powerful processors
- Large storage capacities
- Multimedia data have to move very quickly through the different components of the computer

Characteristics of MM Data

- ***Similarity-based Search***
  - Unlike traditional ‘exact search’ in relational database, users usually ask for similar objects based on their contents
    - Finding an image with similar color
  - A multimedia object may contain multiple features
    - Video may contain text, image, audio, etc.
    - Similarity search may be based on multiple features, i.e., integrating content and semantic features

Characteristics of MM Data
-- challenges about managing MM data

- ***Quality of Service (QoS)***
  - multimedia applications differ from traditional DB applications with respect to **performance requirement**
  - multimedia applications, in general, require **high throughput and a constant delivery** of information
  - e.g, real-time requirement for audio

  note that lost audio samples are perceived much sooner by a human user than lost video frames:

  the ear is more "sensitive" than the eye!

Characteristics of MM Data
-- challenges about managing MM data

- ***Synchronization***
  - multimedia objects: composed of several components,
  - require synchronization of them
    - e.g., a film consists of moving pictures, speech, and subtitles
    - *It'll be odd if the lip movements of actors/actresses do not synchronize precisely with their voices and with the text of the subtitles.*

- Research results have shown that
  - **video/audio** or **video/video** synchronization is **less critical**, expressed in time constraints, than **audio/audio** synchronization!

Multimedia equation

Multimedia equation

Multimedia = presentation + context

Multimedia equation

Multimedia = presentation + context

presentation: sensory, aesthetic part （美学）

Multimedia equation

Multimedia = presentation + context

presentation: sensory, aesthetic part （美学）

context = convergence + information + architecture

- convergence = data +platform + distribution
- information = storage and retrieval
- architecture = compression + components + connectivity

Metadata Model

- Metadata => **data about data**
  - forms an essential part of any database
    - providing **descriptive data** about each stored object, and
    - is the key to **organizing** and **managing data** objects
  - critical for describing essential aspects of content:
    - main topics, author, language, publication, etc.
    - events, scenes, objects, times, places, etc.
    - rights, packaging, access control, content adaptation, …

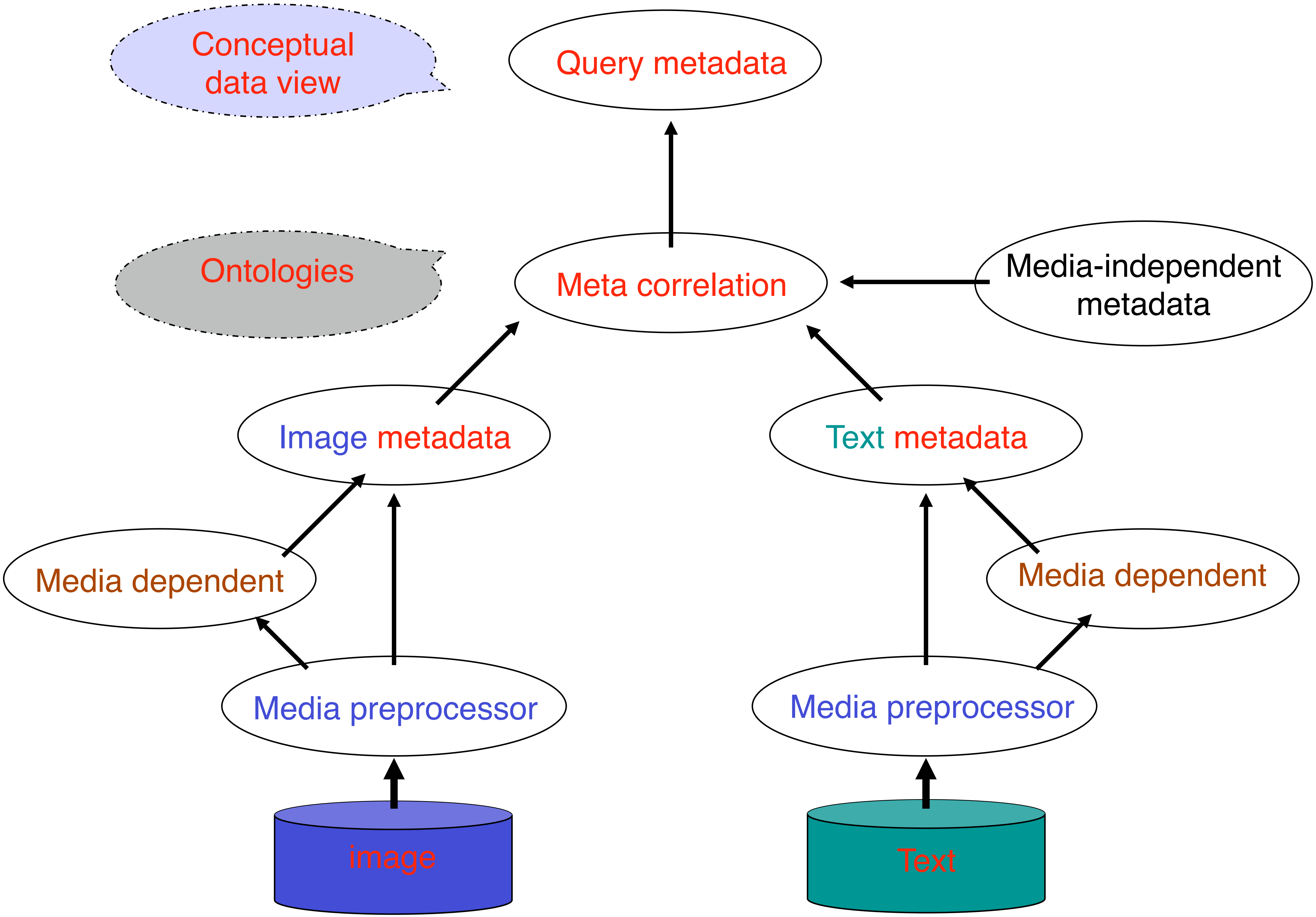
Metadata Model

- **Purposes** of metadata:
  - **Administrative**
    - managing and administrating the data collection process
  - **Descriptive**
    - describing and identifying for retrieval purpose, creating indices
  - **Preservation**
    - managing data refreshing and migration
  - **Technical**
    - formats, compression, scaling, encryption, authentication and security
  - **Usage**
    - users, their level and type of use, user tracking, versioning (e.g., a high resolution version and corresponding thumbnail).

Metadata Model

- Conformity with **open metadata standard** will be a vital:
  - Faster design and implementation
  - Interoperability with broad field of competitive standards-based tools and systems
  - Leveraging of rich set of standards-based technologies for critical functions
    - e.g., content extraction, advanced search, and personalization

The "role" of metadata in query processing:



Classifying Metadata

Classification of metadata can be:

1. Specific to the **media involved**
2. Specific to the **processing**
3. **Content** specific metadata

Image object	Text object	Video
Image capture	title	time based
Image storage	author	play rate
Caption	abstract	camera motion
Genre		camera lighting
Period	Full text indices	
Subjects		
Photographer		
IP rights		
Texture		

Sample Metadata

Metadata Classification

Metadata can be classified as:

- **Content dependent** (e.g., face features; used in CBR)
- **Content-descriptive** (used in TBR)
  1. Domain-independent metadata: independent of the application or subject topic
  2. Domain-dependent metadata: specific to the application area
- **Content-independent** (e.g., photographer’s name; used in ABR)

Metadata Classification

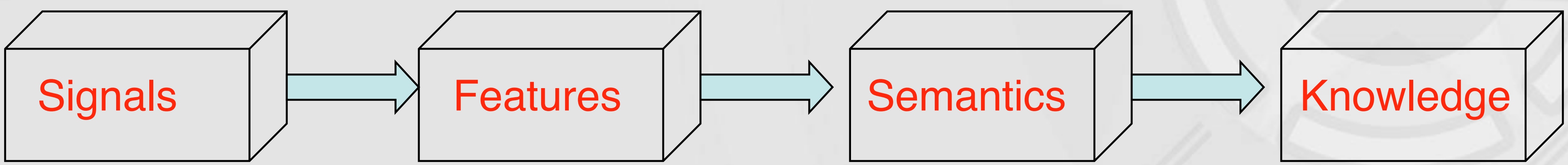
Media	Content independent	Content descriptive	Content dependent
Text	status, location, date of update components	keywords, formats, categories, language	subtopic boundary word image spotting
speech	start, end time location confidence of word recognition	speakers	speech recognition speaker recognition prosodic cues change of meaning
Image	creator title date	keywords, formats	feature selection image features (e.g., histogram, segmentation)
Video	product title data distributor	camera shot action distance close-up	shot boundary frame features (e.g., histogram, motion lighting level, height)

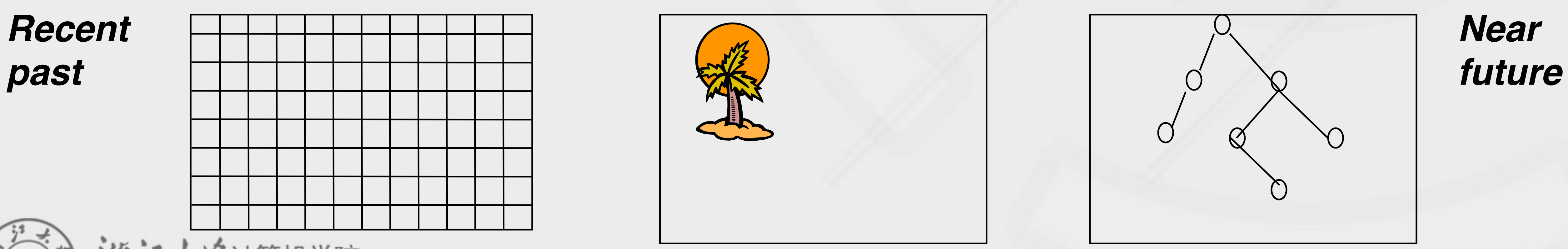
Domain-independent Metadata Standards

- The Dublin Core Metadata set
  http://purl.org/metadata/dublin_core
  - Originally for resource description records of **online libraries** over Internet
  - version 1.1
    - broaden to other media with a link to the ISO/IEC 11179 standard
      - Each Dublin Core element is defined using a set of 10 attributes from the ISO/IEC 11179
      - Six of them are common to all the Dublin Core element (3-5, 7-9)
    - 15 metadata elements (the Dublin Core) has been proposed
      - which are suggested to be the minimum number of metadata elements to support retrieval of a document-like object (DLO) in a networked environment

Domain-independent Metadata Standards

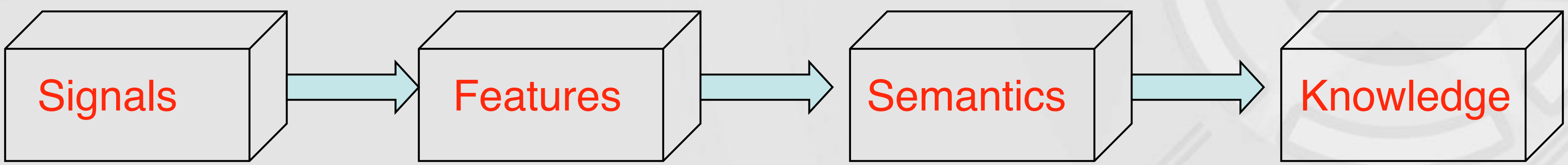
- MPEG series
  - Moving Picture Experts Group (MPEG) since 1998
  - responsible for developing standards of the coded representation of moving pictures and associated audio

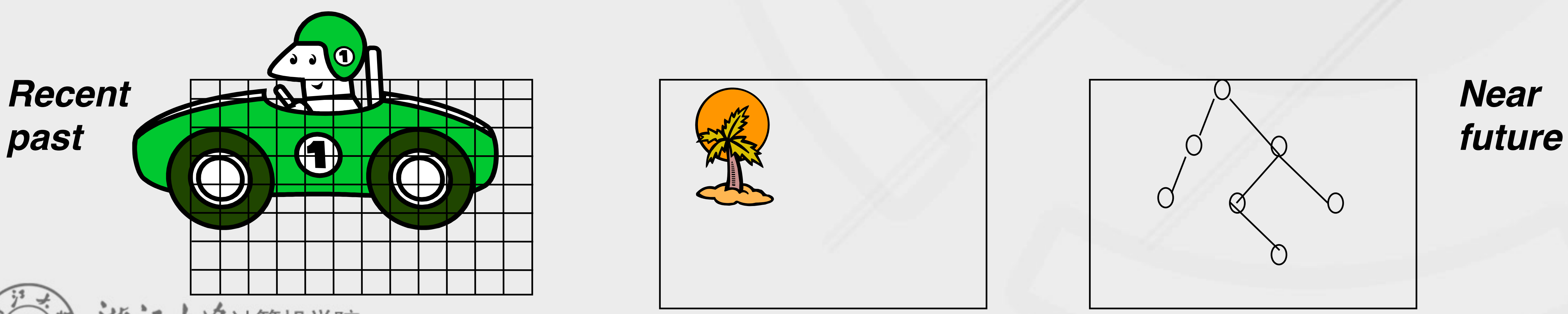




Domain-independent Metadata Standards

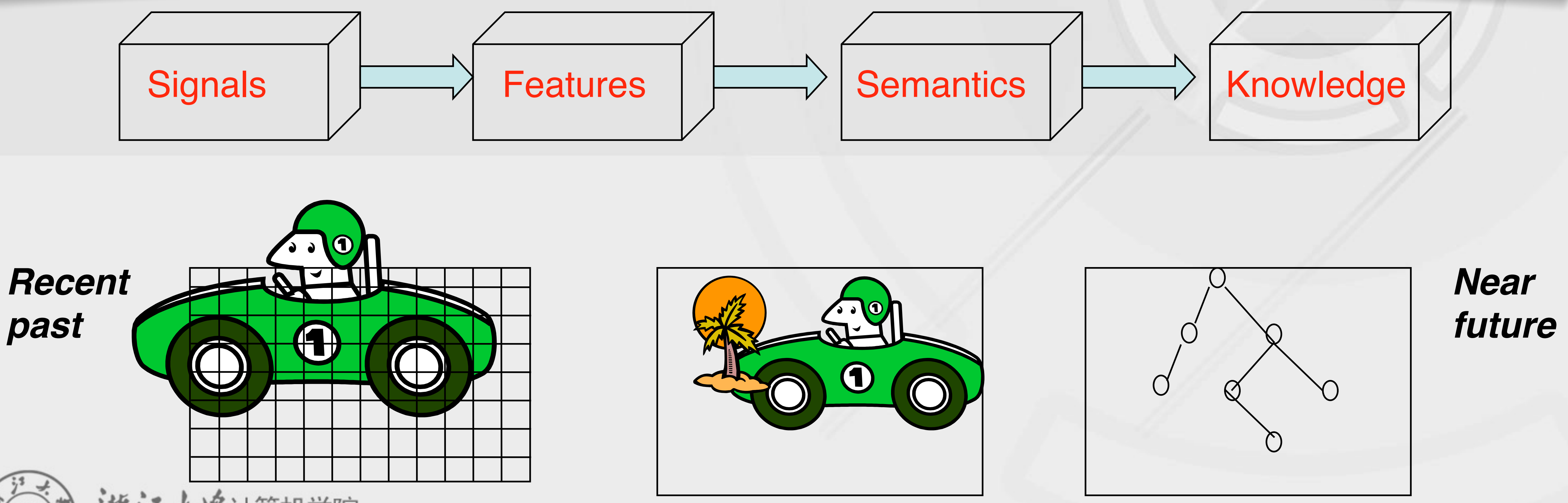
- MPEG series
  - Moving Picture Experts Group (MPEG) since 1998
  - responsible for developing standards of the coded representation of moving pictures and associated audio





Domain-independent Metadata Standards

- MPEG series
  - Moving Picture Experts Group (MPEG) since 1998
  - responsible for developing standards of the coded representation of moving pictures and associated audio



Domain-independent Metadata Standards

Applications			
MPEG-1,-2,-4	MPEG-4,-7	MPEG-7	MPEG-21
Video storage **Broadband, streaming video delivery**	**CBR** **Multimedia filtering** **Content adaptation**	**Semantic-based retrieval and filtering** **Intelligent media services (iTV)**	**Multimedia framework** **e-Commerce**
Problems and Innovations			
Compression coding communications	Similarity search object- & feature- based coding	Modeling & classifying, personalization, summarization	Media mining, decision support

MPEG-1,-2 ,

MPEG-4 ,

MPEG-7 ,

MPEG-21 ,

MPEG-7

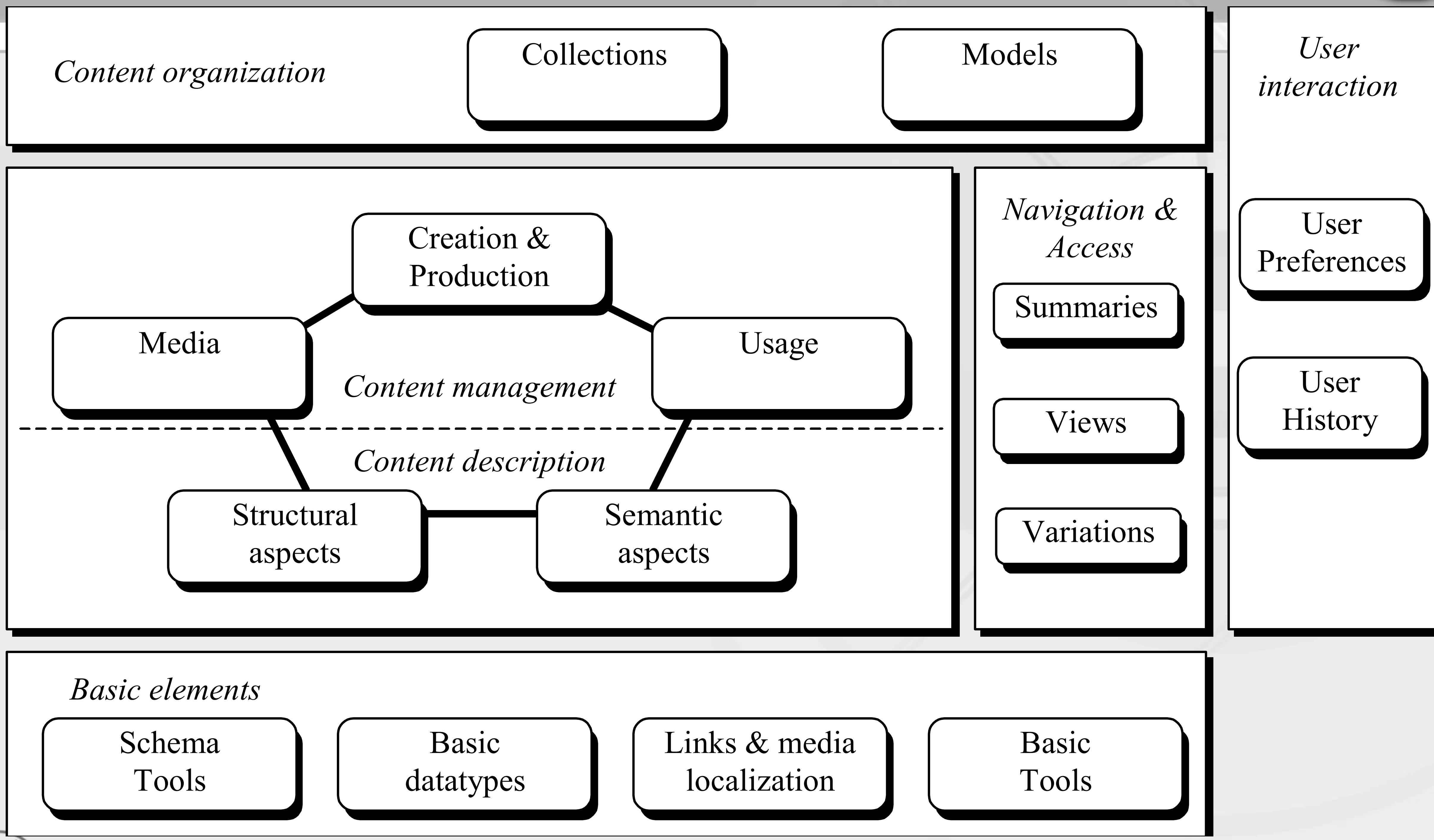
- ***Multimedia Content Description Interface***
  - Representation of information **about** the content
    - still pictures, graphics, 3D models, audio, speech, video & their combination
  - Goal:
    - to **support efficient search** for multimedia content using **standardized descriptions**
    - desirable to use textual information for the descriptions

</description>
<xml>
<resource>

MPEG-7

Set of description tools	Functionality
Media	Description of the storage media: typical features include the storage format, the encoding of the multimedia content, the identification of the media. Note that several instances of storage media for the same multimedia content can be described.
Creation & Production	Meta information describing the creation and production of the content: typical features include title, creator, classification, purpose of the creation, etc. This information is most of the time author generated since it cannot be extracted from the content.
Usage	Meta information related to the usage of the content: typical features involve rights holders, access right, publication, and financial information. This information may very likely be subject to change during the lifetime of the multimedia content.
Structural aspects	Description of the multimedia content from the viewpoint of its structure: the description is structured around segments that represent physical spatial, temporal or spatial-temporal components of the multimedia content. Each segment may be described by signal-based features (color, texture, shape, motion, and audio features) and some elementary semantic information.
Semantic aspects	Description of the multimedia content from the viewpoint of its semantic and conceptual notions. It relies on the notions of objects, events, abstract notions and their relationship.

MPEG-7


Content organization
Collections
Models
User interaction
Navigation & Access
Summaries
Views
Variations
User Preferences
User History
Creation & Production
Media
Usage
Content management
Content description
Structural aspects
Semantic aspects
Basic elements
Schema Tools
Basic datatypes
Links & media localization
Basic Tools

MPEG-7 Standard Elements

- **Descriptors** (Ds)
  - describe features, attributes, or groups of attributes of MM content
- **Description Schemes** (DSs)
  - a DS specifies the structure and semantics of the components (which may be other DSs, Ds, or datatypes)
- **Datatypes**
- **Classification Schemes** (CS):
  - lists of defined terms and meanings
- **System Tools**
- **Extensibility**
  - e.g., new DS's and D's; registration authority for CS

Multimedia Database System

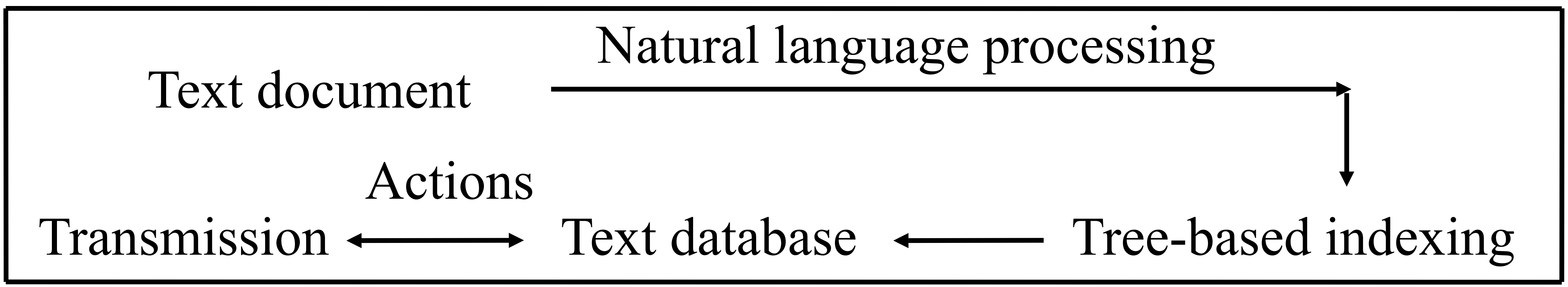
- **Multimedia database** v.s. **text database**
  - **Temporal data**: Requires temporal modeling
  - **Huge amount of data**: Compression helps get around this.
  - Data is **not easily indicative** of the information
  - Requires a lot of **pre-processing** in order to store data efficiently:
    - PCA, feature extraction and segmentation
  - **Novel Query mechanisms**
  - **Hypermedia**: The ability to interactively move around in the data.

How to Build Multimedia Database Systems?

How to build text database? *Yahoo, Google*

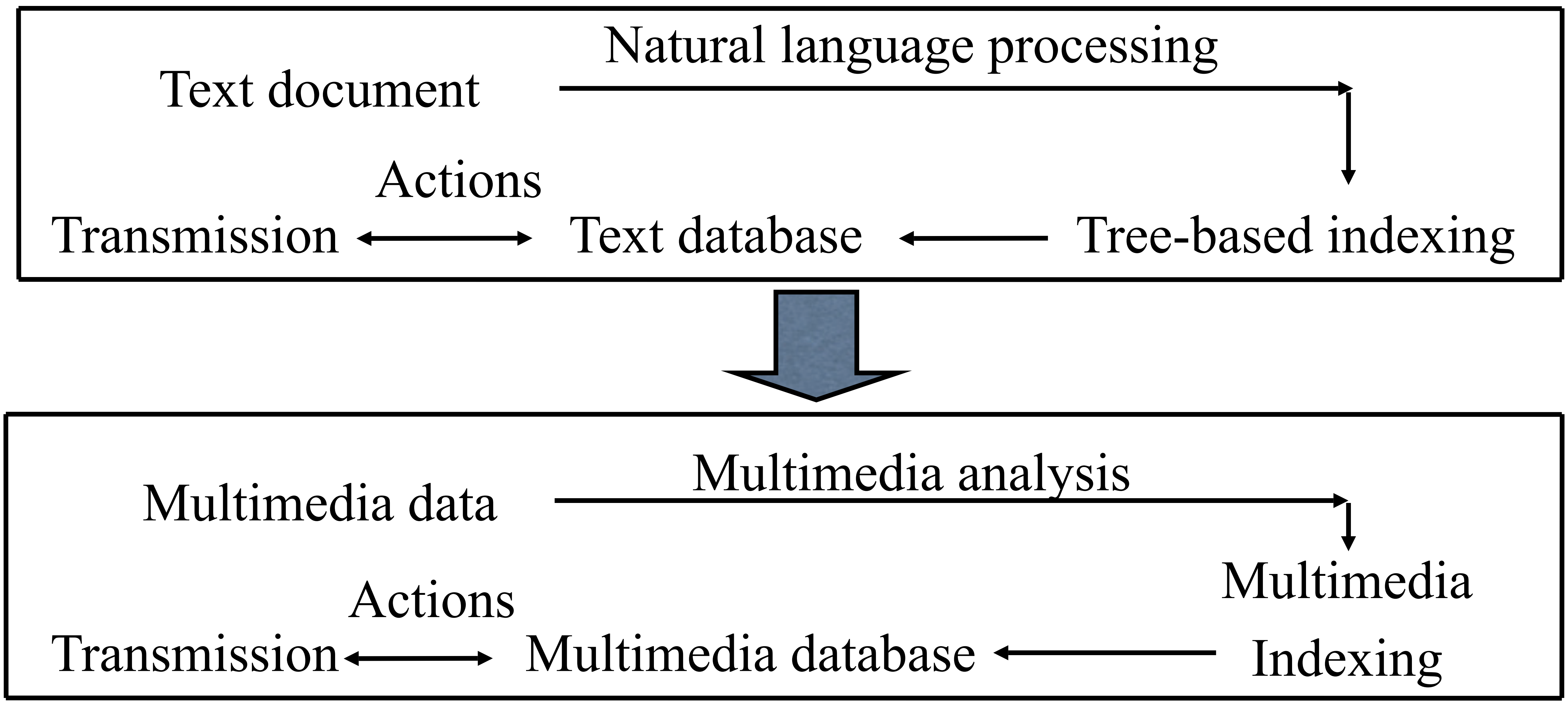
How to Build Multimedia Database Systems?

How to build text database? *Yahoo, Google*

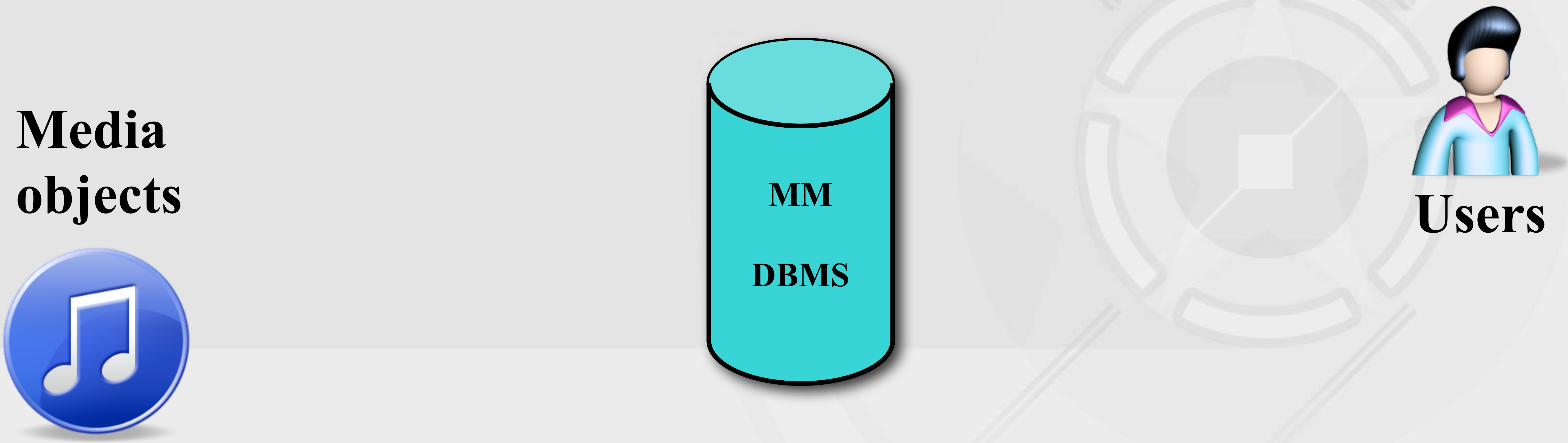


How to Build Multimedia Database Systems?

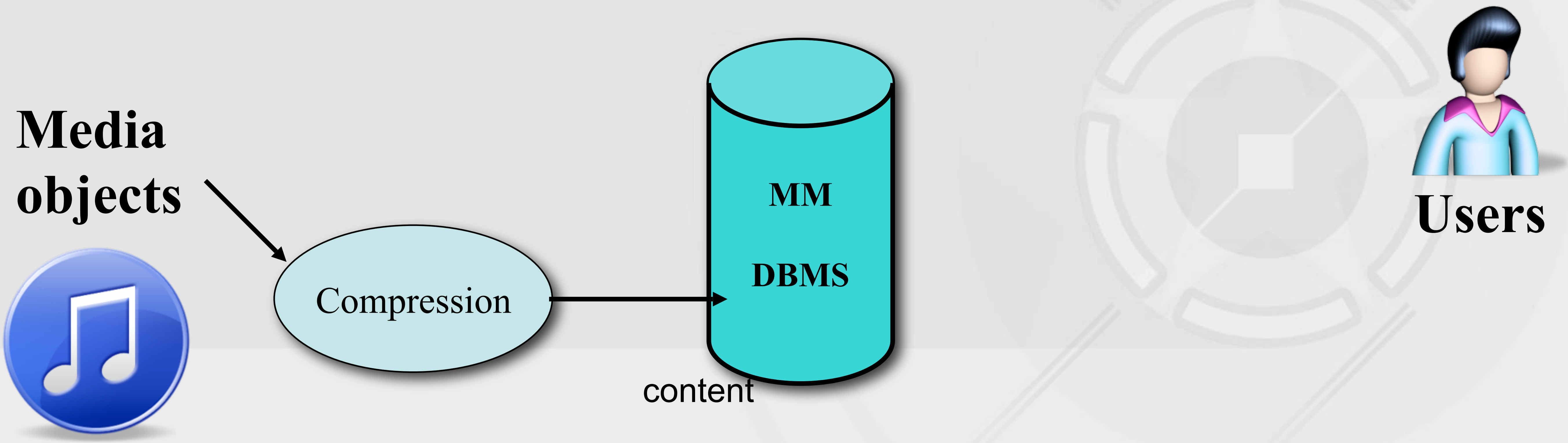
How to build text database? *Yahoo, Google*



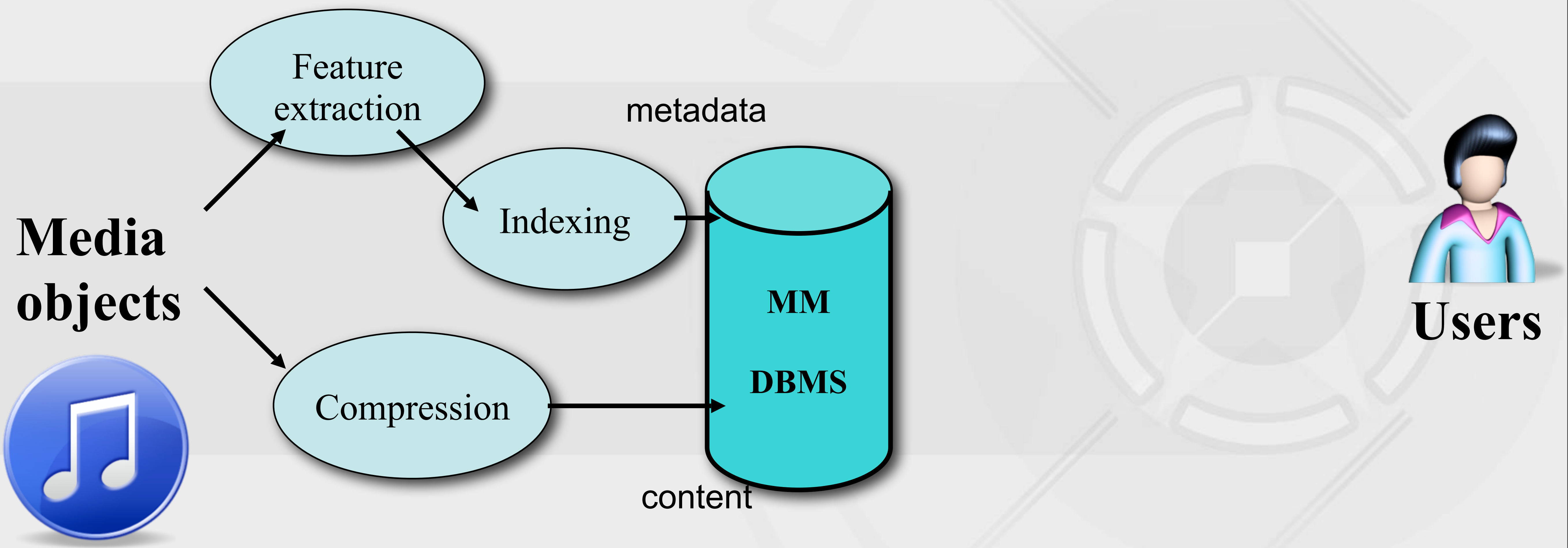
A Generic Architecture of MMDBMS



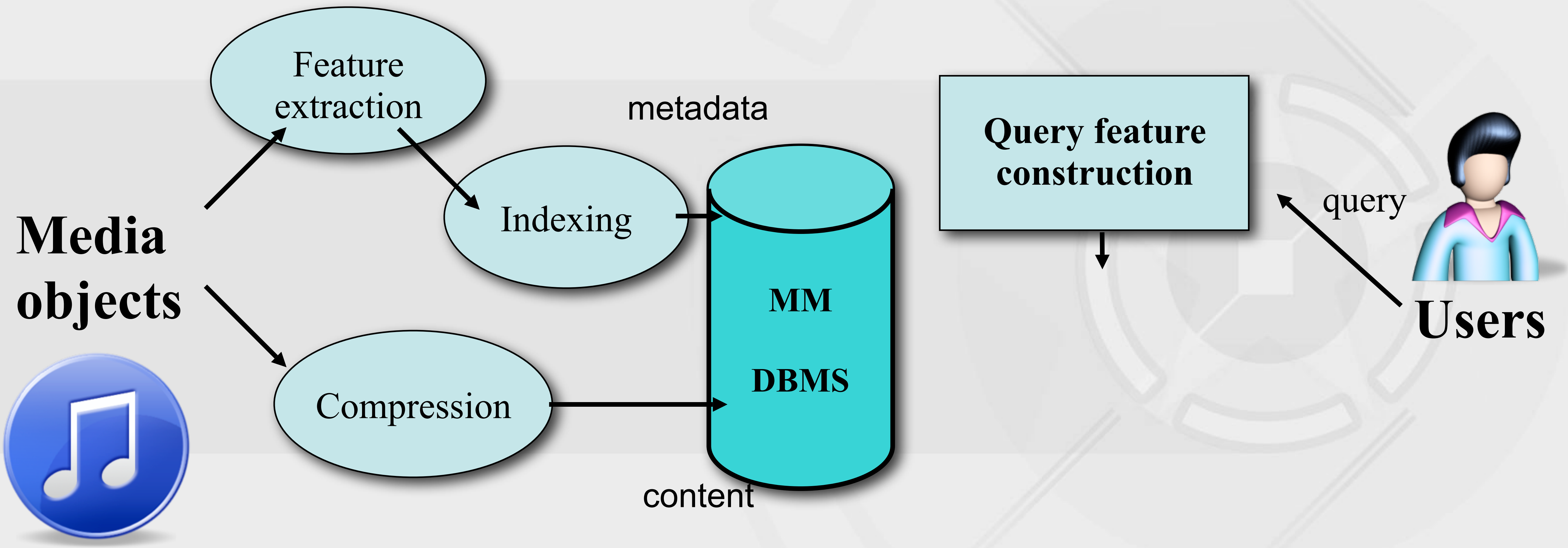
A Generic Architecture of MMDBMS



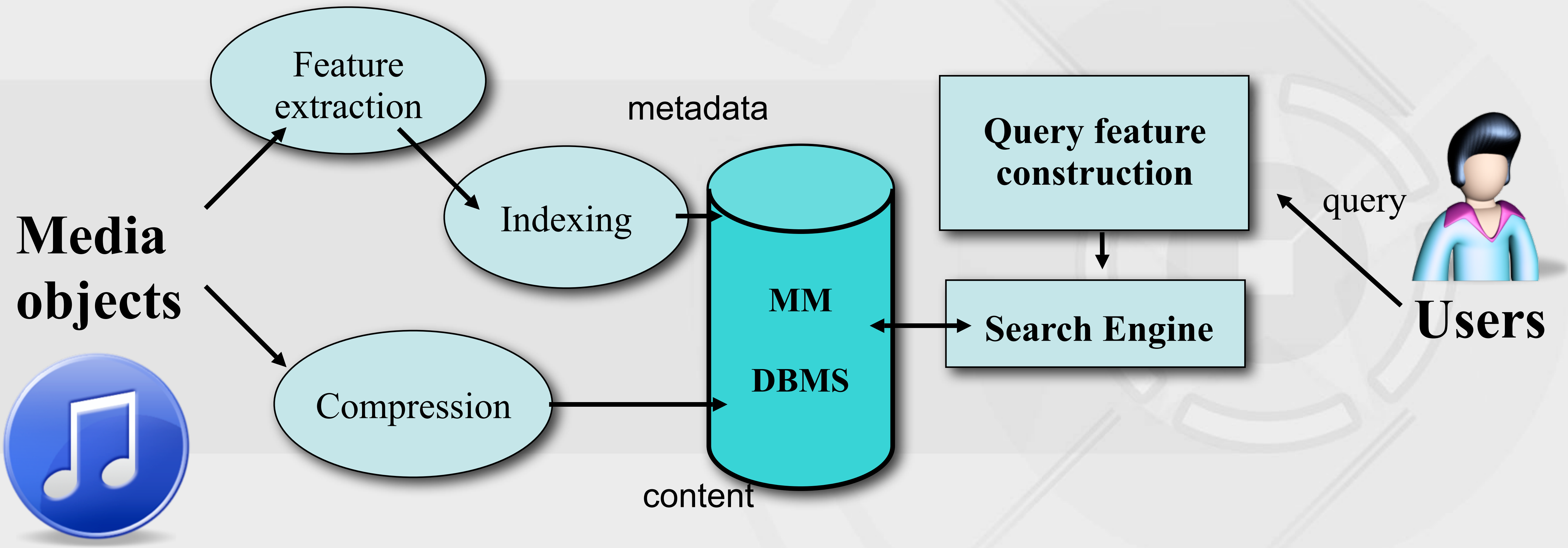
A Generic Architecture of MMDBMS



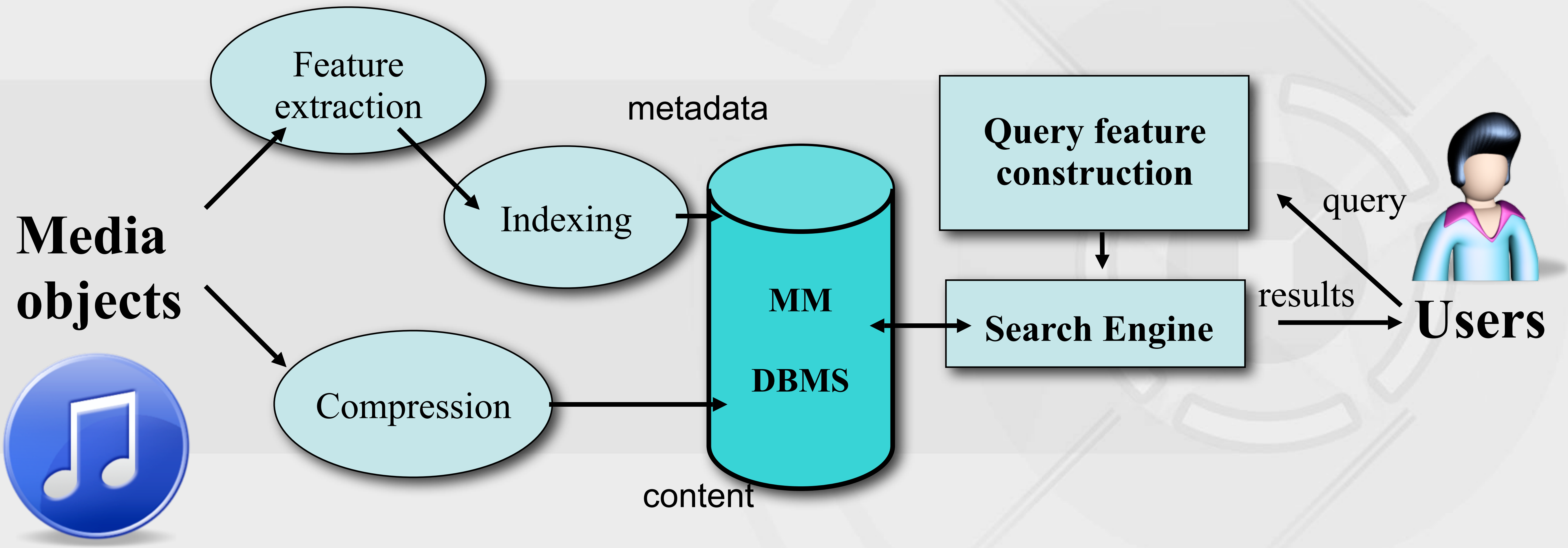
A Generic Architecture of MMDBMS



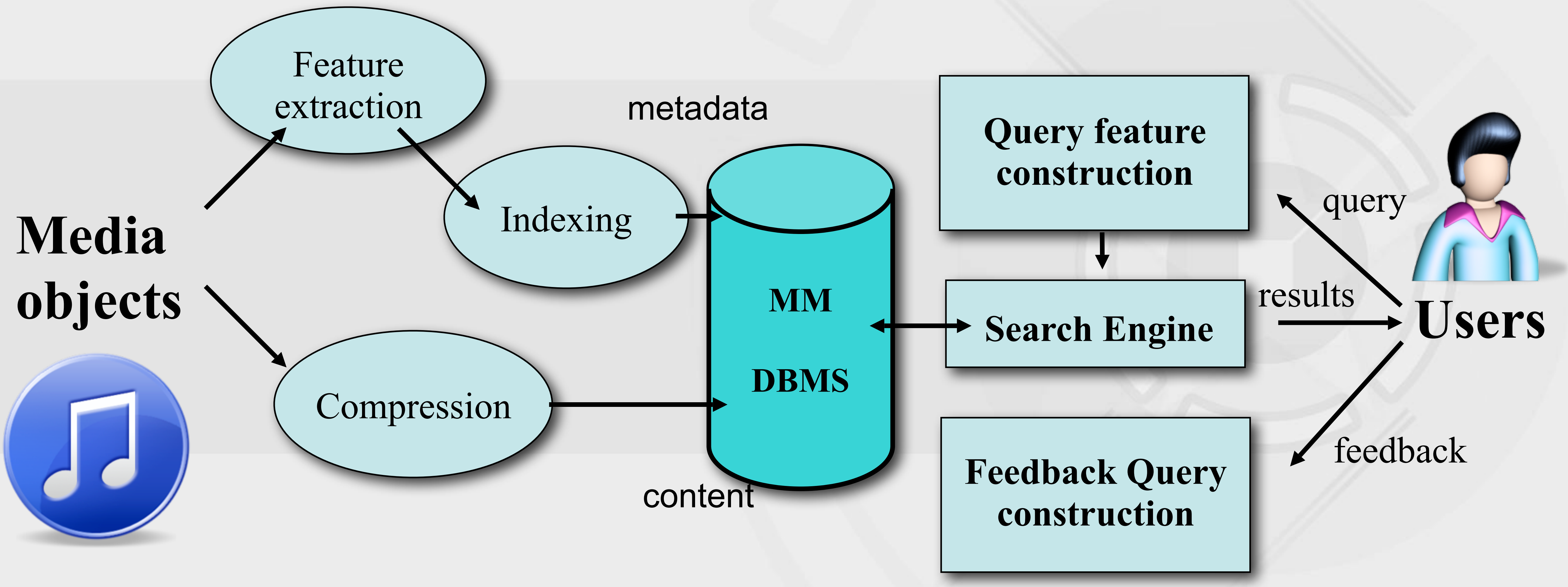
A Generic Architecture of MMDBMS



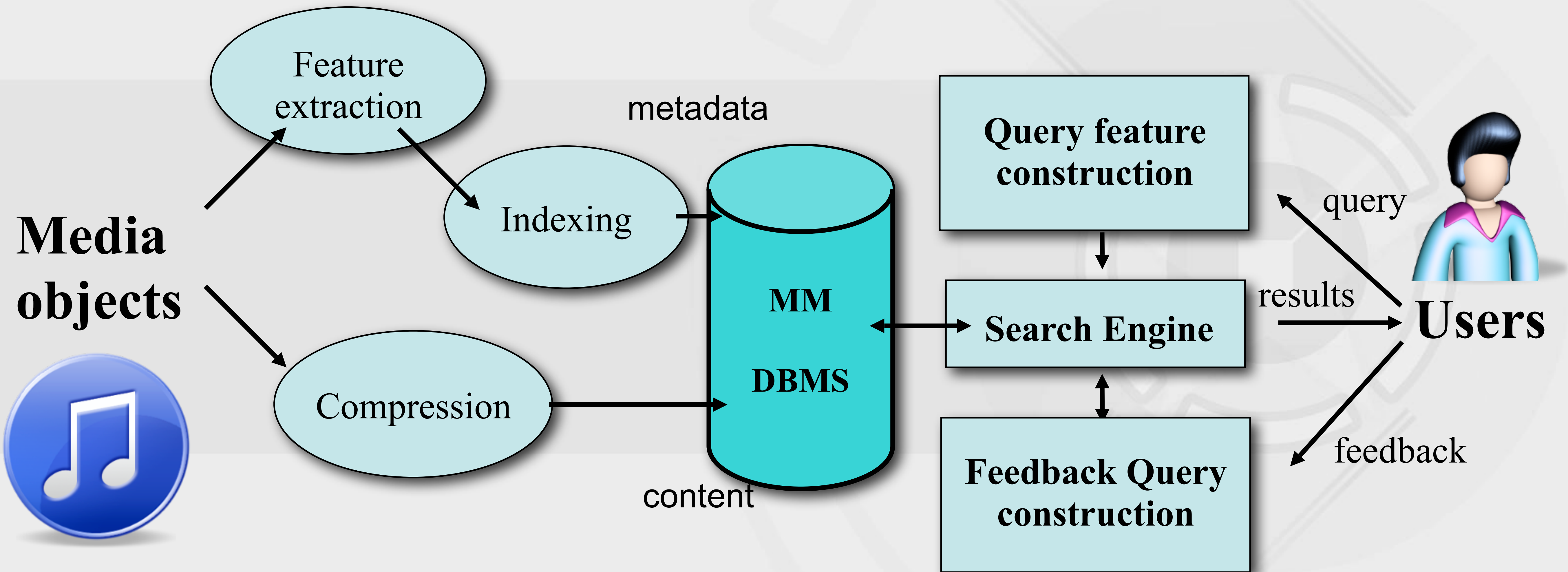
A Generic Architecture of MMDBMS



A Generic Architecture of MMDBMS



A Generic Architecture of MMDBMS



MMDBMS Development

Major steps in developing MMDBMS

1. **Media acquisition:**
   - collect media data from various sources, such as WWW, CD, TV, etc.
2. **Media processing:**
   - extract media representations and their features, including noise filtering, rending, etc.
3. **Media storage:**
   - store the data and their features in the system based on application requirement.
4. **Media organization:**
   - organize the features for retrieval. i.e., indexing the features with effective structures.
5. **Media query processing:**
   - Accommodated with indexing structure, efficient search algorithm with similarity function should be designed.

Multimedia Data Storage

- Storage Requirements
- RAID Technology
- Optical Storage Technology

Requirements of MM Information

- Storage and Bandwidth Requirement
  - measured in bytes or Mbytes for storage
  - measured in bits/s or Mbits/s for bandwidth
    - An image 480 x 600 (24 bits per pixel),
      - 864k bytes (without compression).
      - To transmit it within 2 sec => 3.456Mb/s.
    - 1GB Hard-disk
      - 1.5 hr. of CD-audio or
      - 36 seconds of TV quality video
      - require 800s to be transferred (10Mbits/s network).

RAID Redundant Arrays of Inexpensive Disks

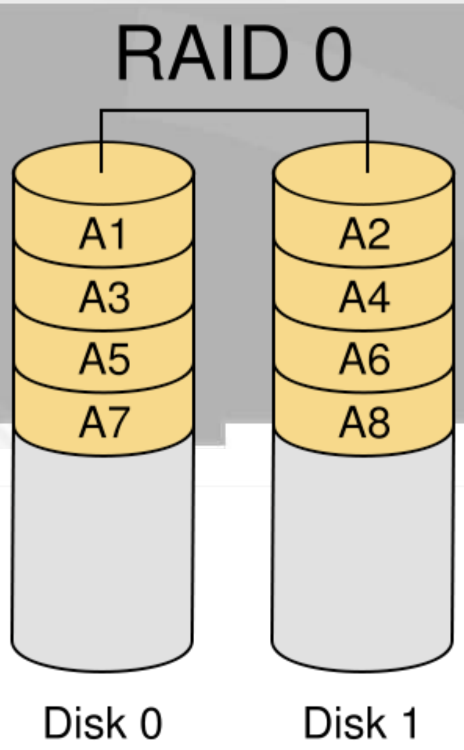
- **By definition RAID has three attributes:**
  - a set of disk drives viewed by the user as one or more logical drives
  - data is distributed across the set of drives in a pre-defined manner
  - redundant capacity or data reconstruction capability is added, in order to recover data in the event of a disk failure

- **Objectives of RAID**
  - Hot backup of disk systems (as in mirroring)
  - Large volume storage at lower cost
  - Higher performance at lower cost
  - Ease of data recovery (fault tolerance)
  - High MTBF (mean time between failure)

Different Levels of RAID

- **Eight discrete levels of RAID functionality**
  - Level 0 - disk striping
  - Level 1 - disk mirroring
  - Level 2 - bit interleaving and Hamming Error Correction (HEC) parity
  - Level 3 - bit interleaving and XOR parity
  - Level 4 - block interleaving with XOR parity
  - Level 5 - block interleaving with parity distribution
  - Level 6 - Fault tolerant system
  - Level 7 - Heterogeneous system
- Data is spread across the drives in units of **512 bytes** called **segments**. **Multiple segments form a block.**

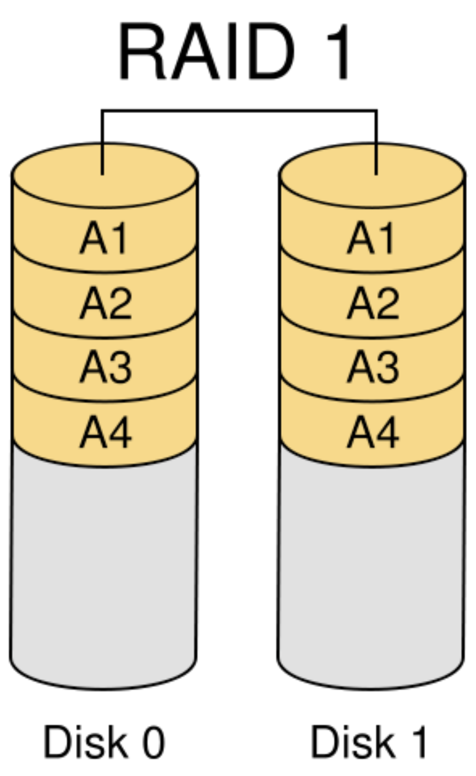
RAID Level 0 - Disk Striping



- To improve performance by overlapping disk reads and writes
- Multiple drives connected to a single disk controller
- Data is striped to spread segments of data across multiple drives in block sizes ranging from 1 to 64 Kbytes
- Disk striping provides a higher transfer rate for write and retrieve block of data
- Typical application: database applications
- Drawbacks:
  - If one drive fails, the whole drive system fails
  - Does not offer any data redundancy, no fault tolerance

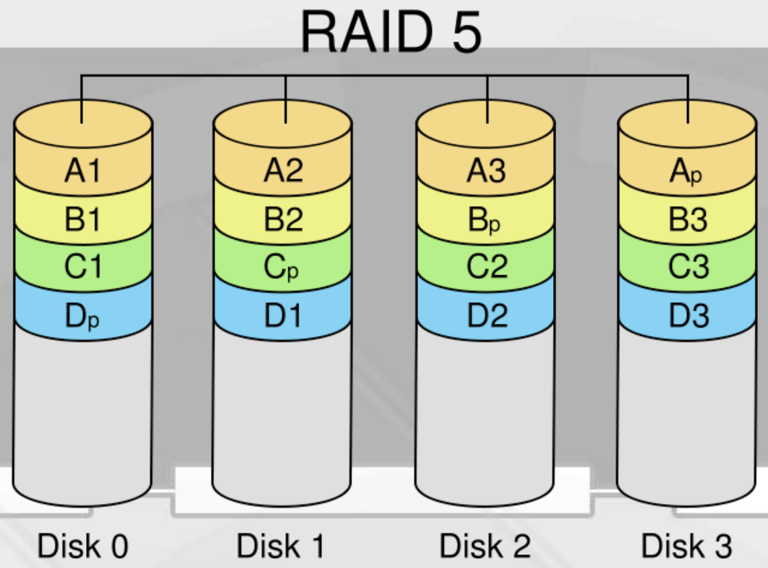
RAID Level 1 - Disk Mirroring

- Each main drive has a **mirror drive**
- Two copies of every file will write to two separate drives
  complete redundancy
- Performance:
  * Disk write : take almost twice time
  * Disk read : can be speed up by overlapping seeks
- Typical use:
  * in file servers provides backup in the event of disk failure
- Duplexing:
  * Use two separate controllers
  * The second controller enhances both fault tolerance and performance
  * Separate controllers allow parallel writes and parallel reads



RAID Level 5

- Block Interleaving with Parity Distribution



RAID Level 5

- Block Interleaving with Parity Distribution

- RAID 5 is organized in a similar fashion to RAID 4 but avoids the bottleneck encountered in RAID 4.
- It does not use a dedicated parity drive
- Parity data is interspersed in the data stream and spread across multiple drives.
- Block of data falling within the specified block size requires only a single I/O access.
- Block of data are stored on a different drive, multiple concurrent block-sized accesses can be initiated.
- Good for database applications in which **most I/O occurs randomly and in small chunks.**
- Drawbacks: **high cost and low performance** for large block sizes objects such as audio and video.

Data Storage

- The strategy adopted for data storage will depend on the **storage technology, storage design,** and **the nature of data itself**.
- Any storage has the following parameters:
  - Storage capacity
  - Standard operations of Read and Write
  - Unit of transfer for Read and Write
  - Physical organization of storage units
  - Read-Write heads, Cylinders per Disc, Tracks per Cylinder, and Sectors per Track
  - Read time and seek time
- Of the storage technologies that are available as computer peripherals, the optical medium is the most popular in the multimedia context.

Digital Rights Management

- DRM and multimedia
- Protecting Digital Intellectual Property
  - Encryption
  - Watermarking
- General architecture

念哉典学，思睿贯通

DRM

- **Set of technologies that enable content owners to specify and control:**
  - the access they want to give consumers and
  - the conditions under which it is given.
- **It includes:**
  - **Persistent Protection:**
    - Technology for protecting files via encryption and allowing access to them only after the entity desiring access has had its identity authenticated and its rights to that specific type of access verified
  - **Business rights:**
    - Capability of associating business rights with a content by contract, e.g. author's rights to an article or musician's rights to a song
  - **Access tracking:**
    - Capability of tracking access to and operations on content
  - **Rights licensing:**
    - Capability of defining specific rights to content and making them available by contract

DRM Functional Architecture

- **IP Asset Creation and Capture Module**
  - Rights Validation to ensure content being created includes the rights to do so
  - Rights Creation to allow rights to be assigned to new content
  - Rights Workflow to allow for content to be processed through series of workflow steps
- **IP Asset Management Module**
- **IP Asset Usage Module**

DRM Functional Architecture

- **IP Asset Creation and Capture Module**
- **IP Asset Management Module**
  - Repository functions to enable the access/retrieval of content in potentially distributed databases and the access/retrieval of metadata
  - Trading functions to enable assignment of licenses to parties who have traded agreements for rights over content, including payments from licensees to rights holders (e.g., royalty payments)
- **IP Asset Usage Module**
  - Permissions Management to enable usage environment to honor the rights associated with the content, e.g., if user only has the right to view the document, then printing will not be allowed
  - Tracking Management to enable monitoring of usage of content where such tracking is part of the agreed to license conditions, e.g., user has license to play video ten times

Interested Players in DRM

- **Government Agencies**
  - Interested in controlled viewing and sharing of highly secure and confidential documents, audio and video data.
  - “Need to know basis”
- **Private Corporations**
  - Want to limit the sharing of their proprietary information
  - Track accesses and any modifications made to it.
  - E.g. news agencies like Reuters
- **Owners of commercial content**
  - Content owners, artists, and publishers want to gain revenue through sale and promotions
  - Concerned about protecting their copyrighted works from illegal use

Interested Players in DRM (cont.)

- **Intermediaries (service providers, content distributors etc.)**
  - Concerned about minimizing costs of providing services
  - Cautious about protecting themselves from lawsuits over illegal distribution
- **Producers of end user equipment (PCs, players, etc.)**
  - Concerned about minimizing design and production costs
  - Unwilling to pay for features that only some users need
- **End users**
  - Interested in immediate access to desired content
  - Want to use the content conveniently

The Division

- **The content development industry:**

  (e.g., the recording industry and the movie studios)

  – the need for immediate DRM solutions that stop all unauthorized copying and distribution.

- **The IT industry:**

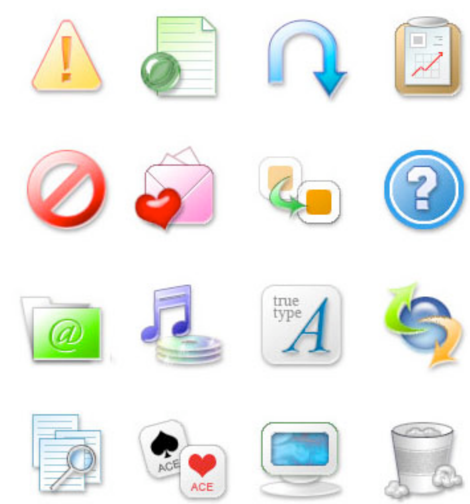
  – DRM solutions should support the concept of "fair use,"

    - allows consumers to make copies of some types of copyrighted content for their own personal use.

Fair Use

- Copyright principle based on **the belief**:
  - the public is entitled to freely use **portions of copyrighted materials** for purposes of **commentary and criticism**.
- Unfortunately, if the copyright owner disagrees with your fair use interpretation, the dispute will have to be resolved by courts or arbitration.
- The four factors for measuring fair use:
  - the purpose and character of your use
  - the nature of the copyrighted work
  - the amount and substantiality of the portion taken, and
  - the effect of the use upon the potential market.

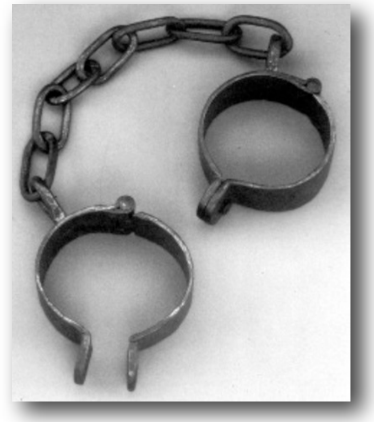
- rights holder
- end customer



create/use

User ⟷ Content

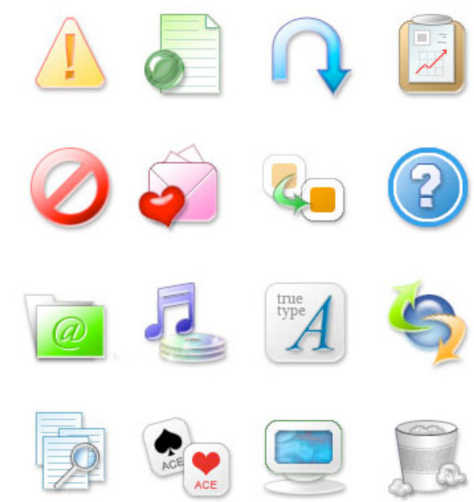
DRM basic Model



Rights

own

- rights holder
- end customer

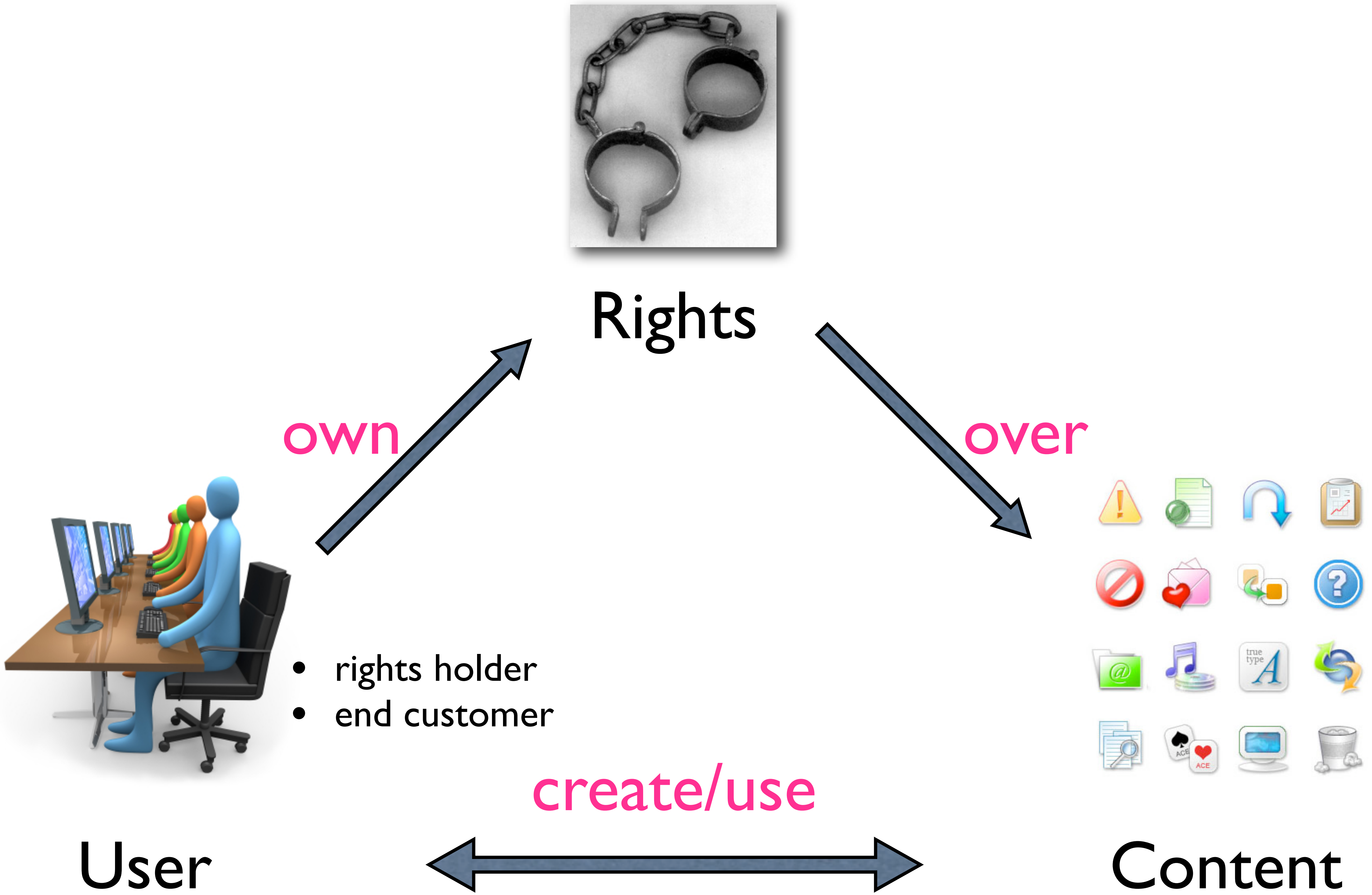


create/use

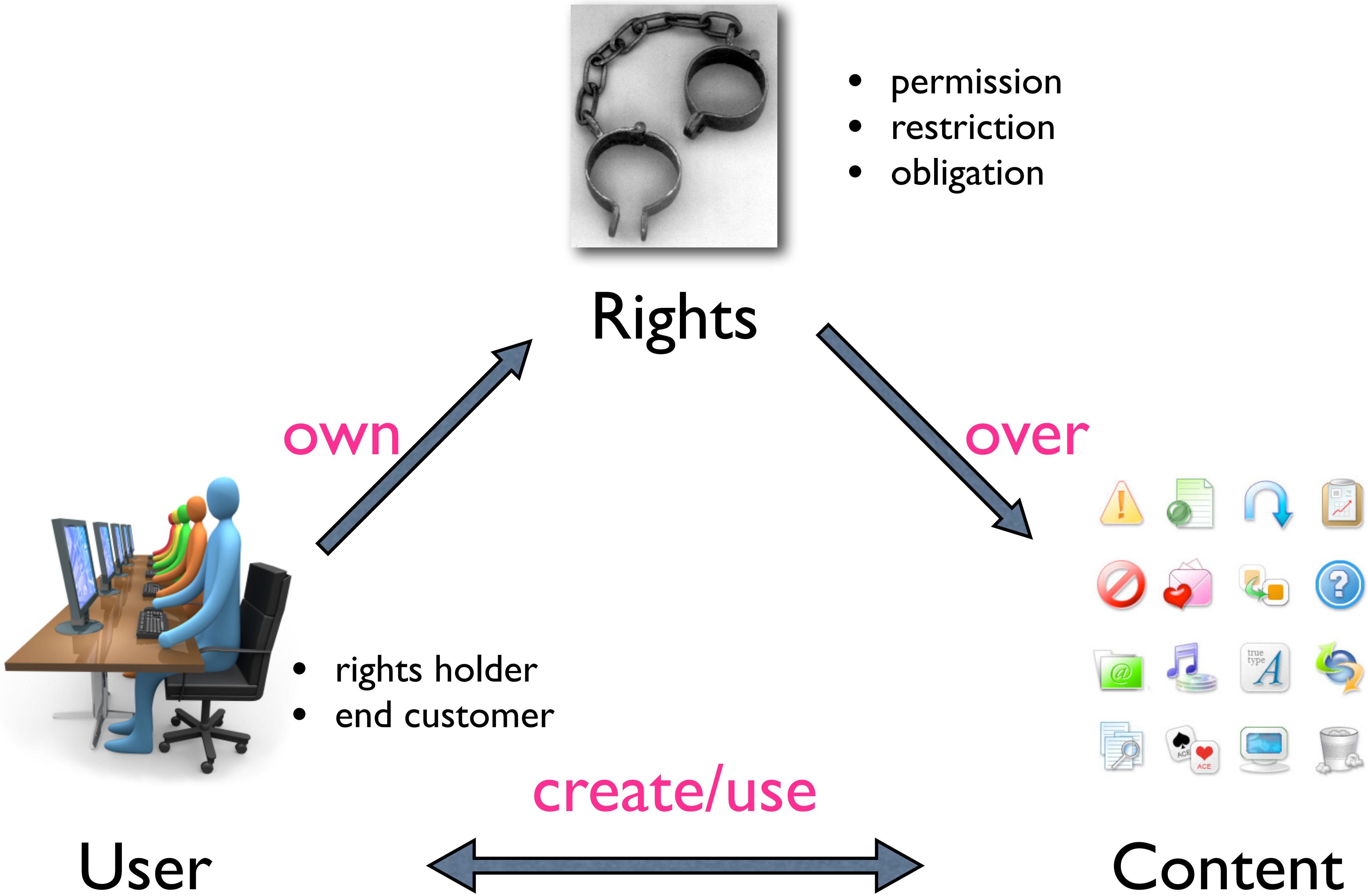
User

Content

DRM basic Model

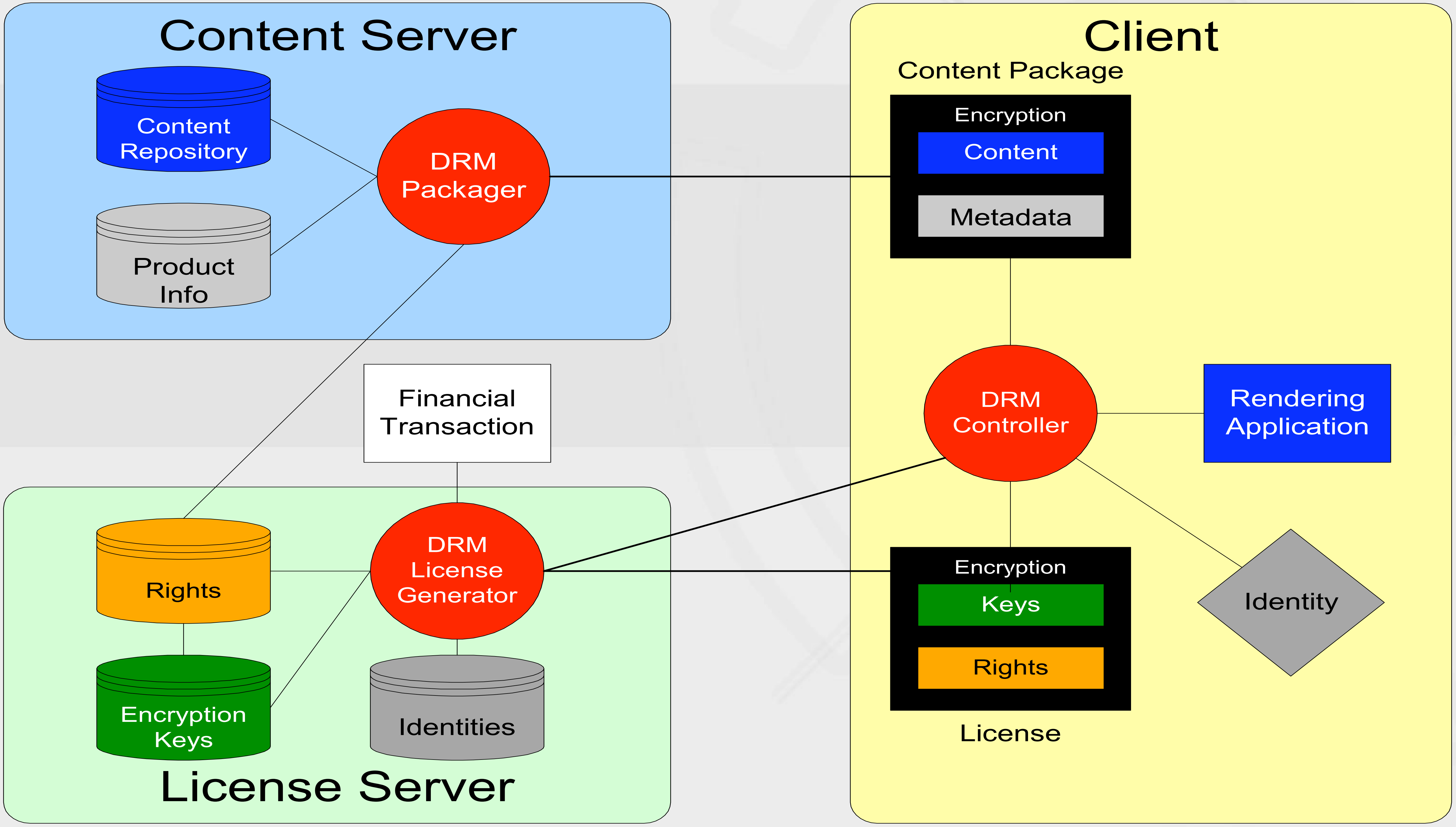


DRM basic Model



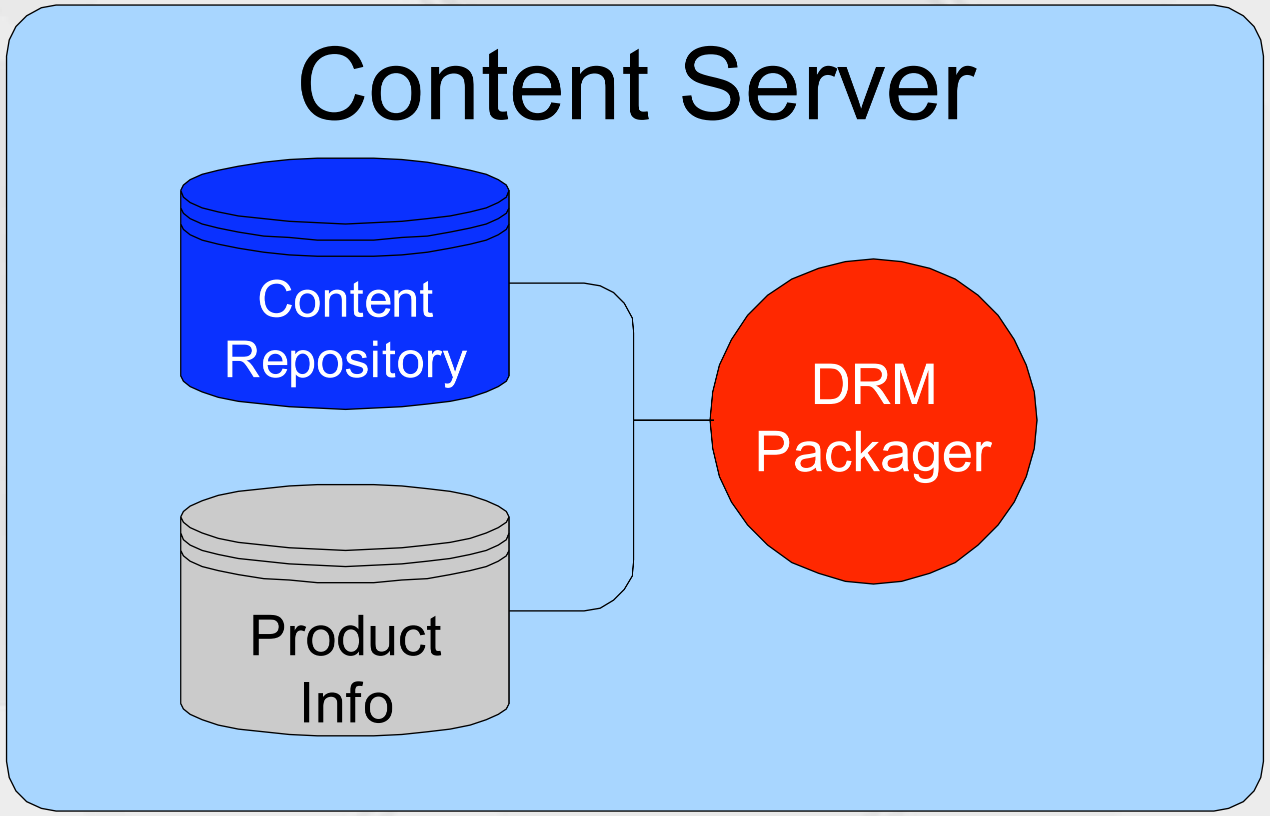
DRM basic Model

DRM Reference Architecture



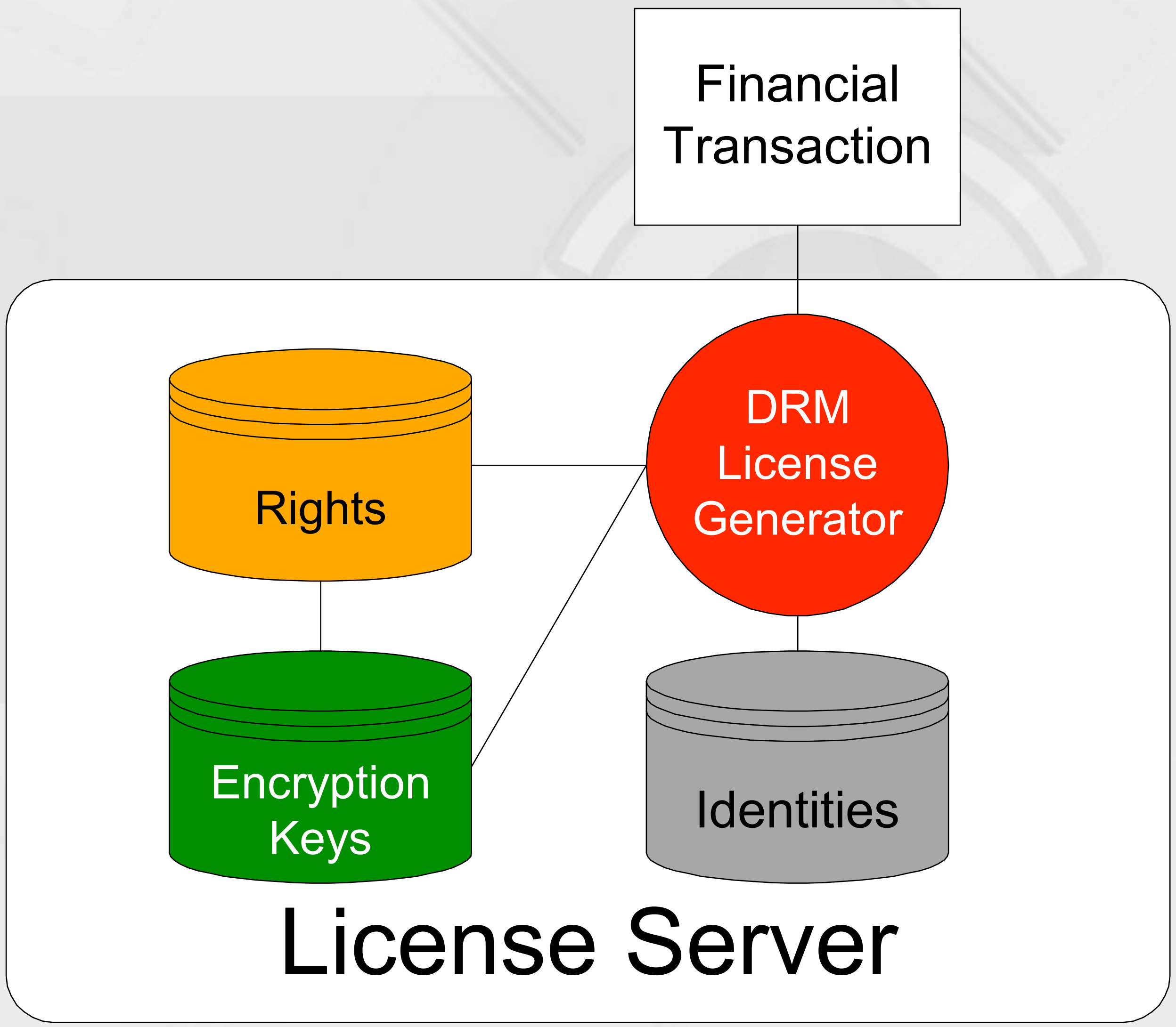
Content Server

- Content Repository
  - Content Management system
  - Digital Asset Management system
  - File server
- Product Info
  - **Rights**
  - Product metadata
- DRM Packager
  - Packages content with metadata
  - Encrypts



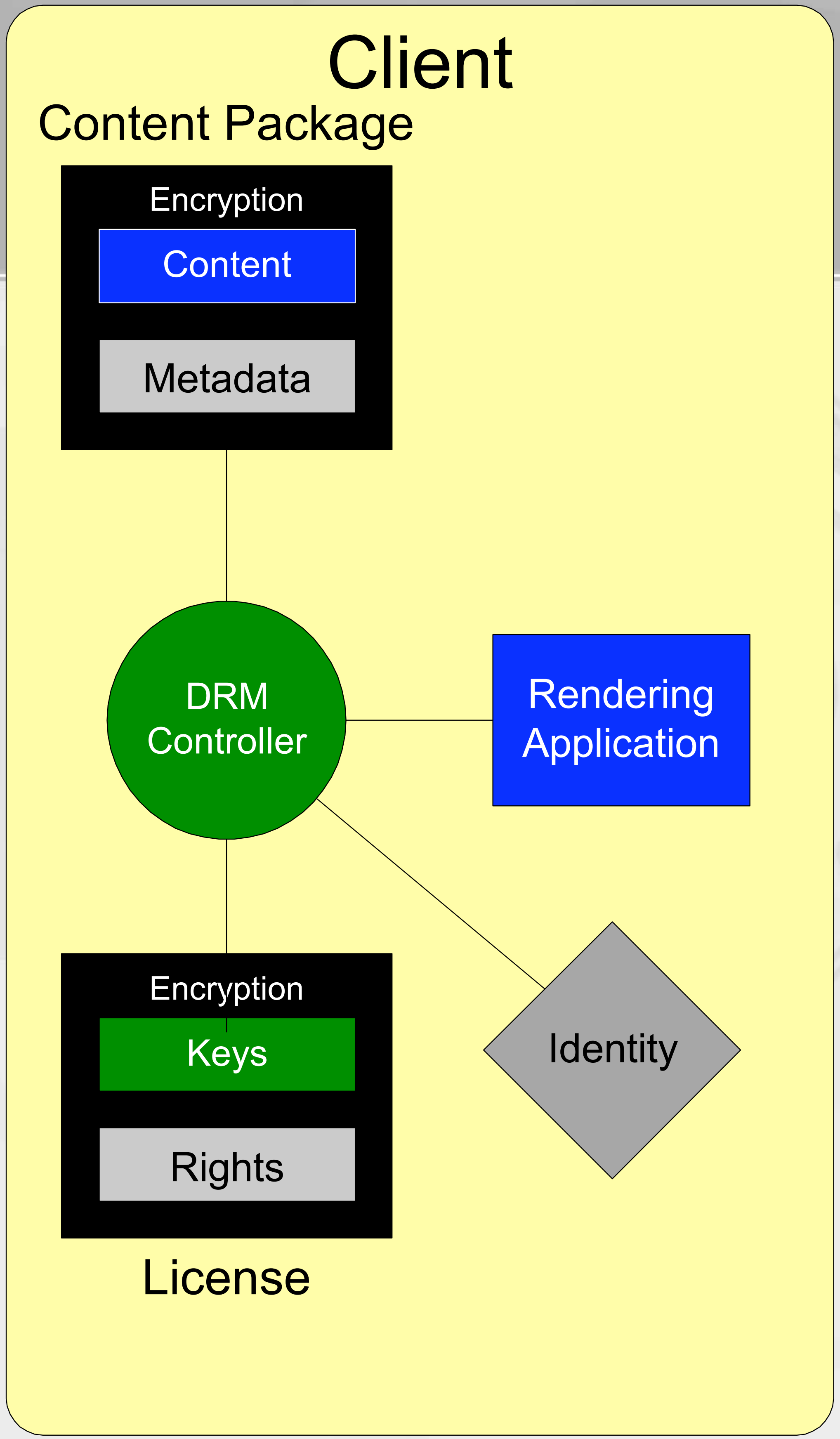
License Server

- Encryption key repository
- User identity database
  - Usernames
  - Machine IDs
- DRM License Generator



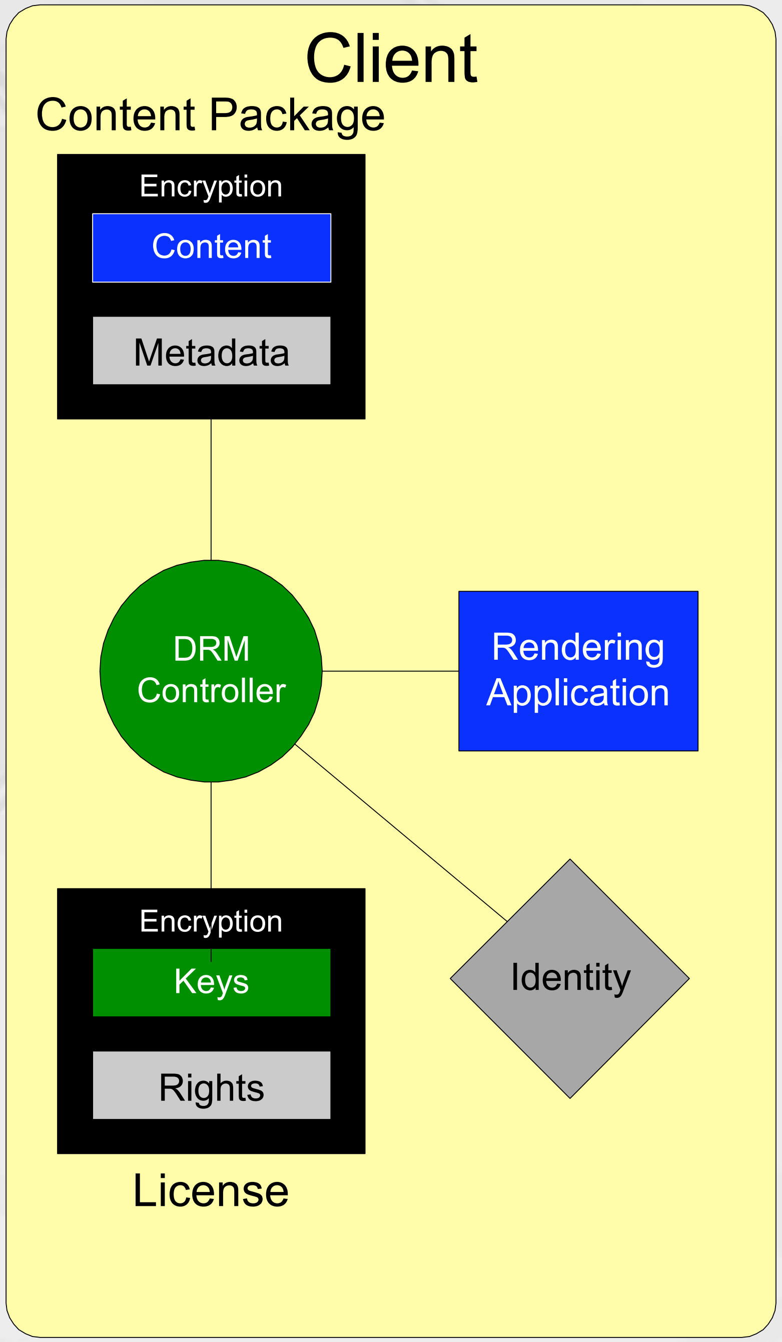
Client

- DRM Controller
  - Nerve center of process
- Rendering application
- Content packages
- Licenses
- Identity



Processes - User Initiation

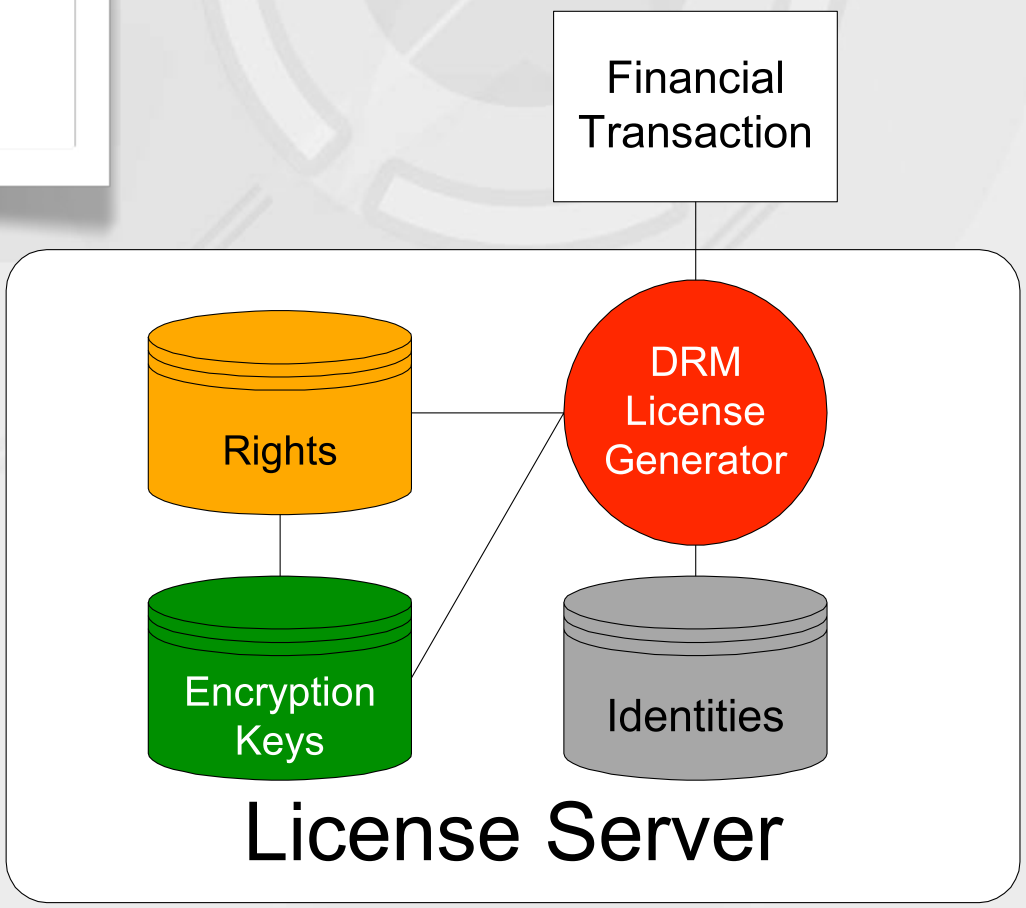
- User obtains content package
- User requests operation
  - view, play
- DRM controller collects info
  - Content
  - Identity
  - Requested rights
- DRM controller:
  - license generator



Processes - License Generation

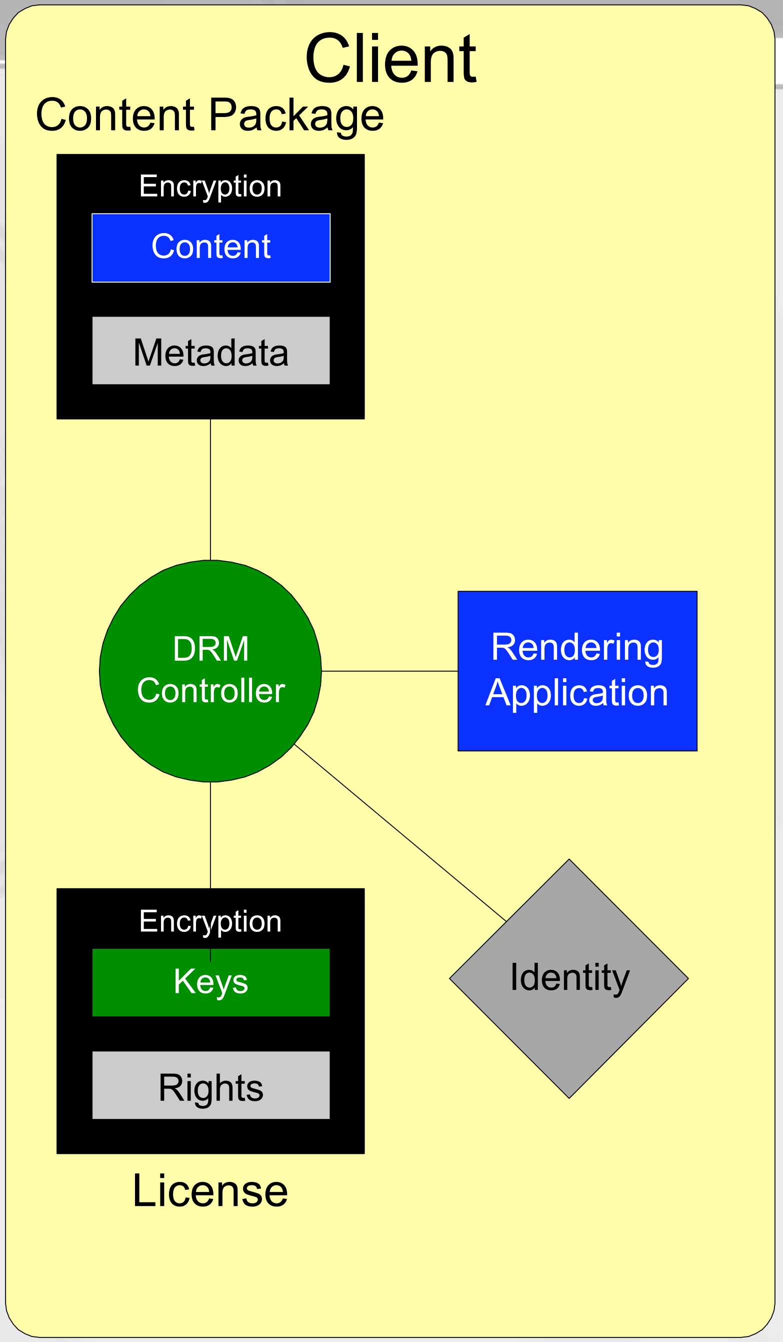
DRM License Generator…

- Checks content & identity
- Obtains keys from key repository
- Creates & sends license to client
- Generates financial transaction, where necessary

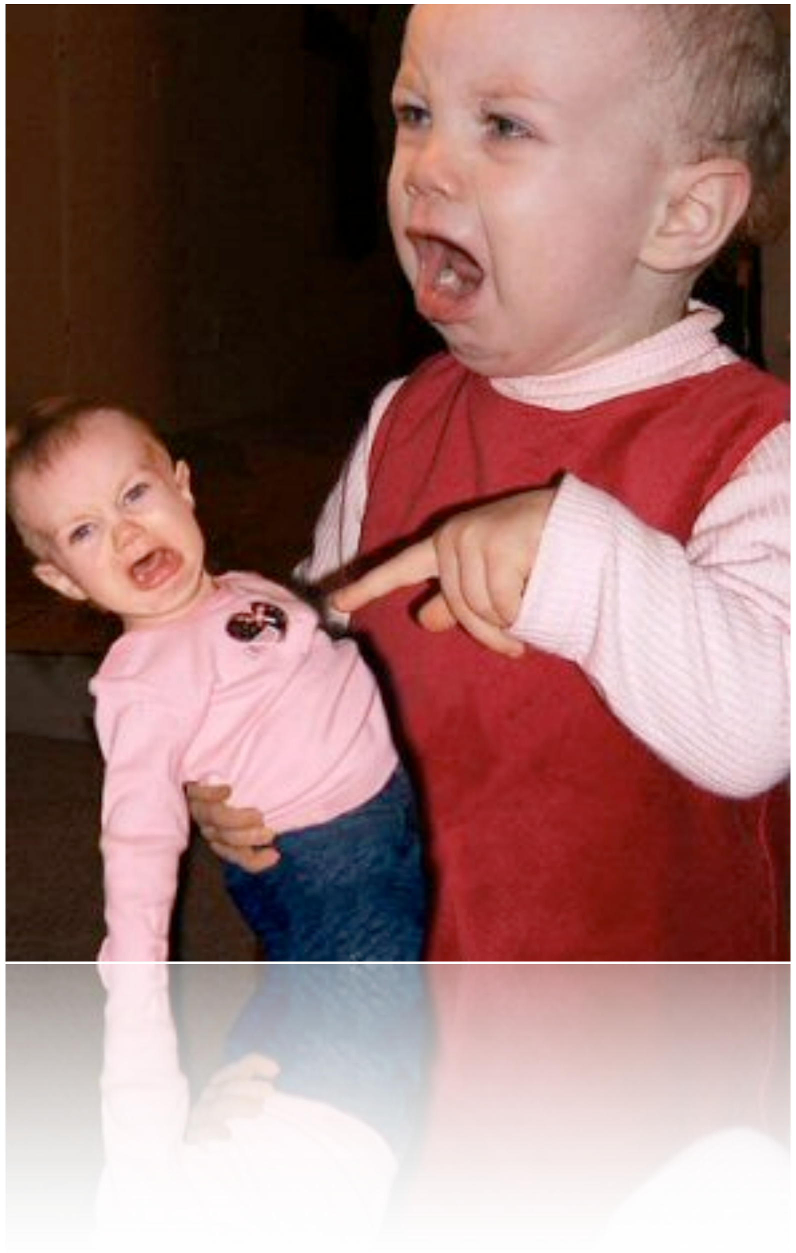


Processes - User Completion

- DRM Controller…
  - Receives license
  - Extracts keys from license
  - Decrypts content
  - Generates financial transaction, where necessary
  - Hands content to rendering application
- Rendering application plays content



What were the results of Sony's attempt at DRM?



What were the results of Sony's attempt at DRM?

Apple's FairPlay Technology

- DRM for iTunes
  - playing, recording, and sharing of files
- Moves beyond "protection only"
  - allows media to be shared among devices
  - allows others to listen to (but not copy) music
  - allows music to be burned to an audio CD, which loses the DRM protection

How FairPlay Works

- iTunes uses encrypted MP4 audio files
- Acquire decryption key by trying to play song
  - player generates a unique ID
  - sends this ID to the iTunes server
  - if there are less than N authorizations in your account, the server responds with decryption key
- The decryption key itself is encrypted so cannot be given to another machine

Protecting Digital Intellectual Property

- Understand technical background of Napster, DVD DeCSS, Sklyarov cases
- Representing and distributing digital IP
  - Compression
  - Distribution: CD’s and DVD’s
  - Distribution via the Internet
- Protecting digital IP
  - Encryption basics
  - DVD CSS, and how it was cracked
  - eBooks, and how they were cracked
- Social and legal issues
- Discussion: “Napsterizing” other industries

IP Protection Tactics

- Prevent Copying through technological means
  - Disallow copying in the first place
  - Use combination of encryption and trusted clients
- Prevent copying through legal means (DMCA)
  - Usually targeted at large scale piracy
  - Why is this hard to enforce against individuals?
- Track copies
  - if illegal copy detected, punish those who make/distribute them
  - Find the original owner of illegal copies
  - Uses digital watermarking

Protecting Digital Intellectual Property

- Preventing Copying With Encryption
  - 加密
- Preventing Copying With Watermarking
  - 水印

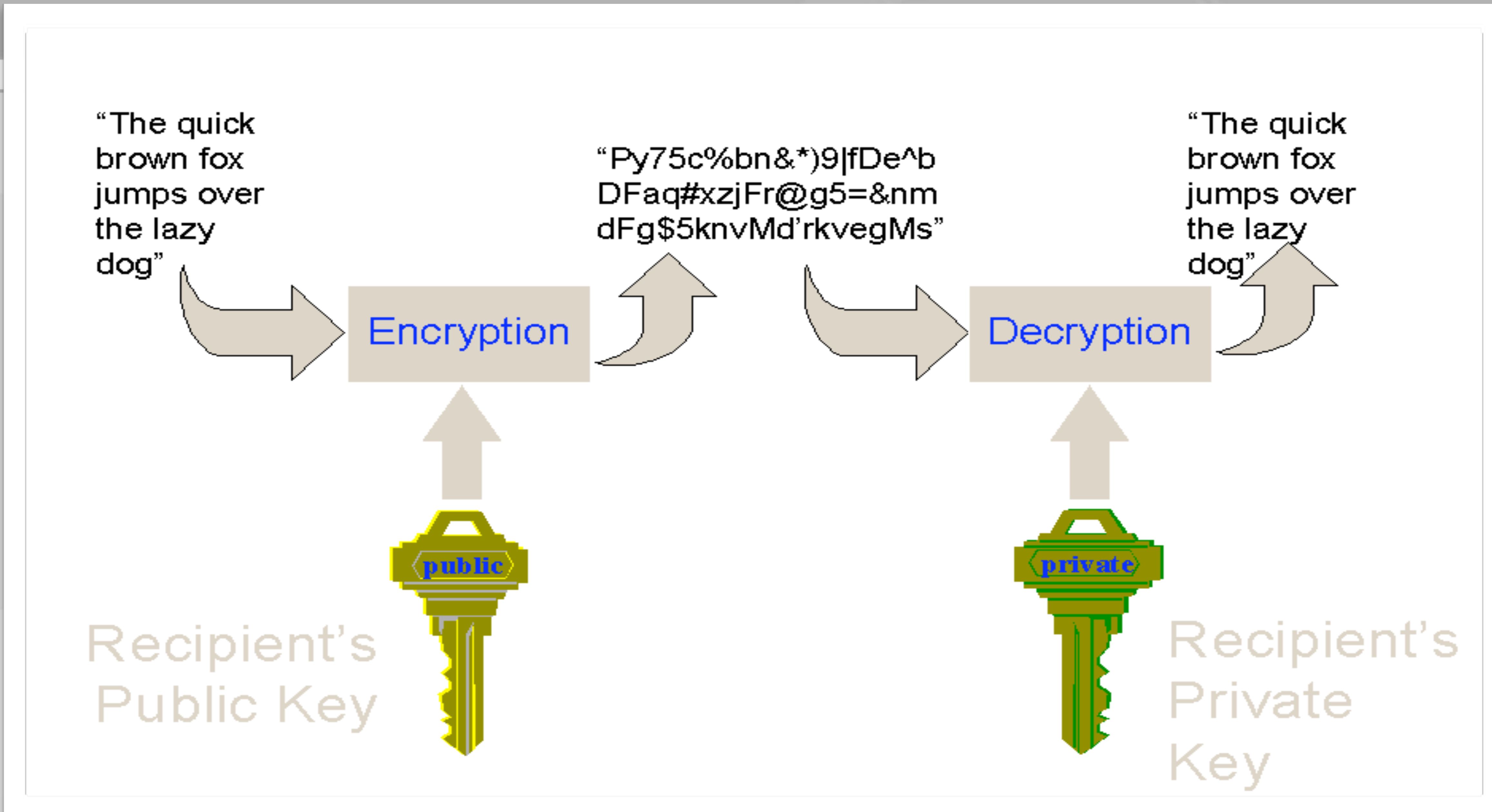
Preventing Copying With Encryption (加密)

- Encryption is the scrambling of a message
  - Simple one is Caesar encryption
  - To decrypt (decode) message, you need one or more *Keys*
  - Also need an encryption *algorithm*, that specifies how to apply the key to the message to produce the scrambled message
- Symmetric key crypto: same key used for encrypt/decrypt
- Public key (we'll talk about the details later…):
  - Keys come in matched pairs: one encrypts, other decrypts
  - Given one key, you cannot deduce the other

Encryption

- RSA
- DES
- MD5

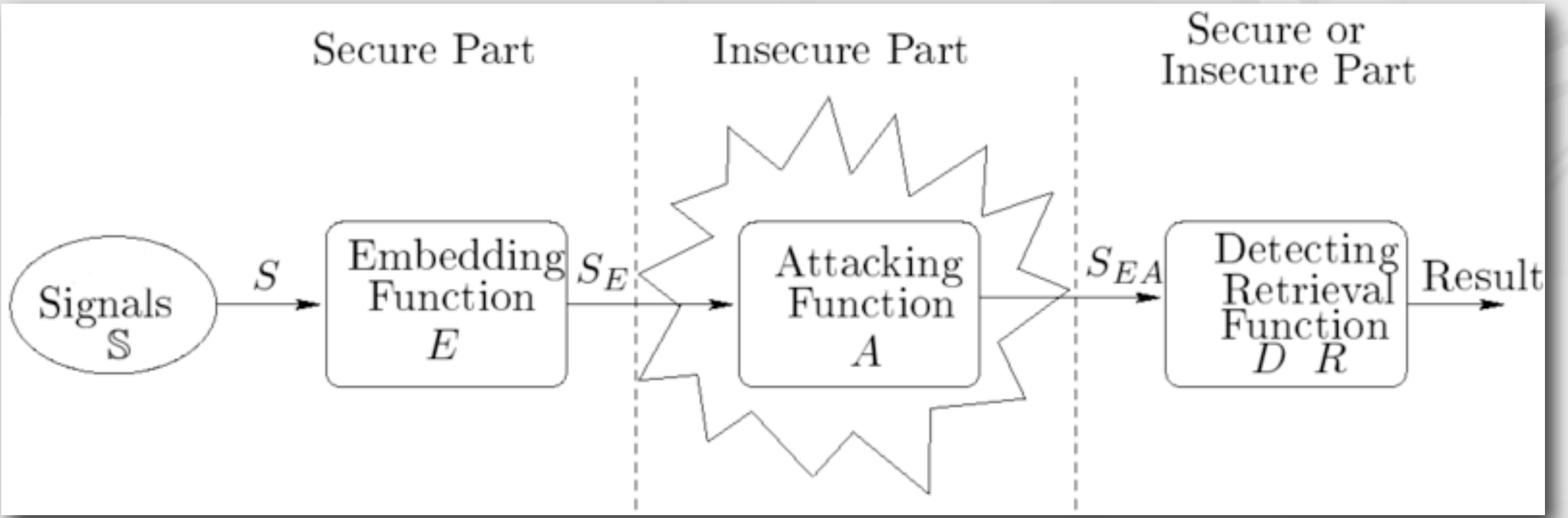
Basic Idea of Cryptography



Think of encryption key as sealing an envelope, and decryption key as unsealing it.

Preventing Copying With Watermarking (水印)

- digital art
- 票据防伪
- 数据隐藏
- 隐蔽通讯



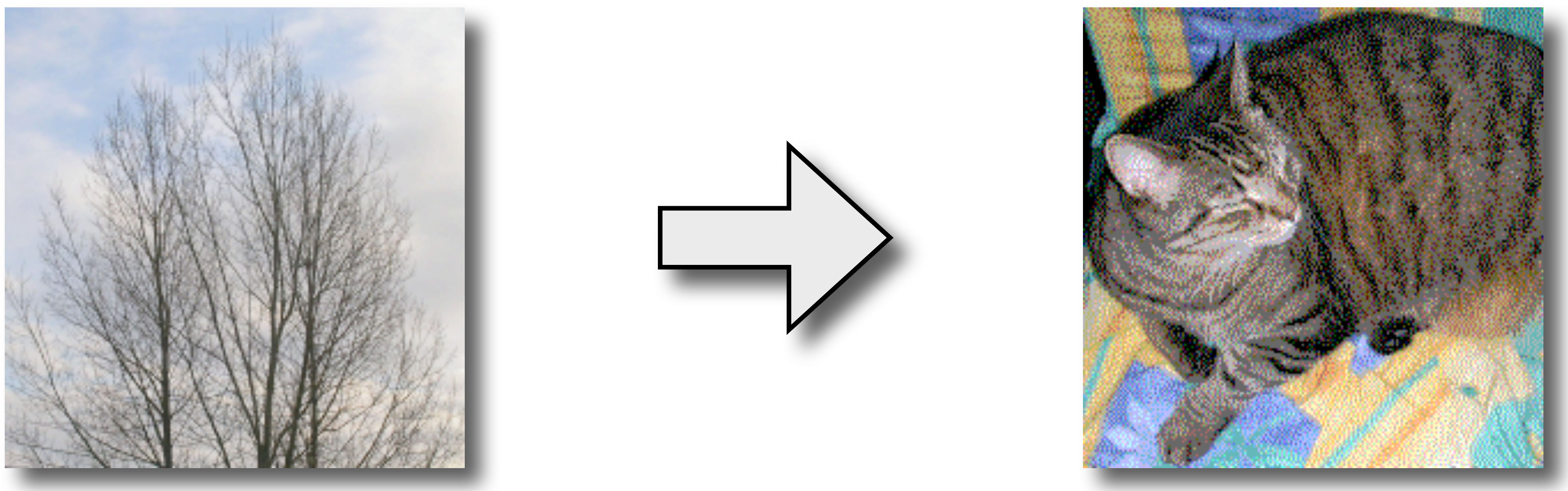
Stenography

Stenography

Stenography

1. removing all but the last 2 bits of each color component
2. X 85

Stenography



1. removing all but the last 2 bits of each color component
2. X 85

Digital Watermark

- Invisible ink on multimedia data
  - image
  - video
  - music
  - graphics

Digital Watermark

- Image

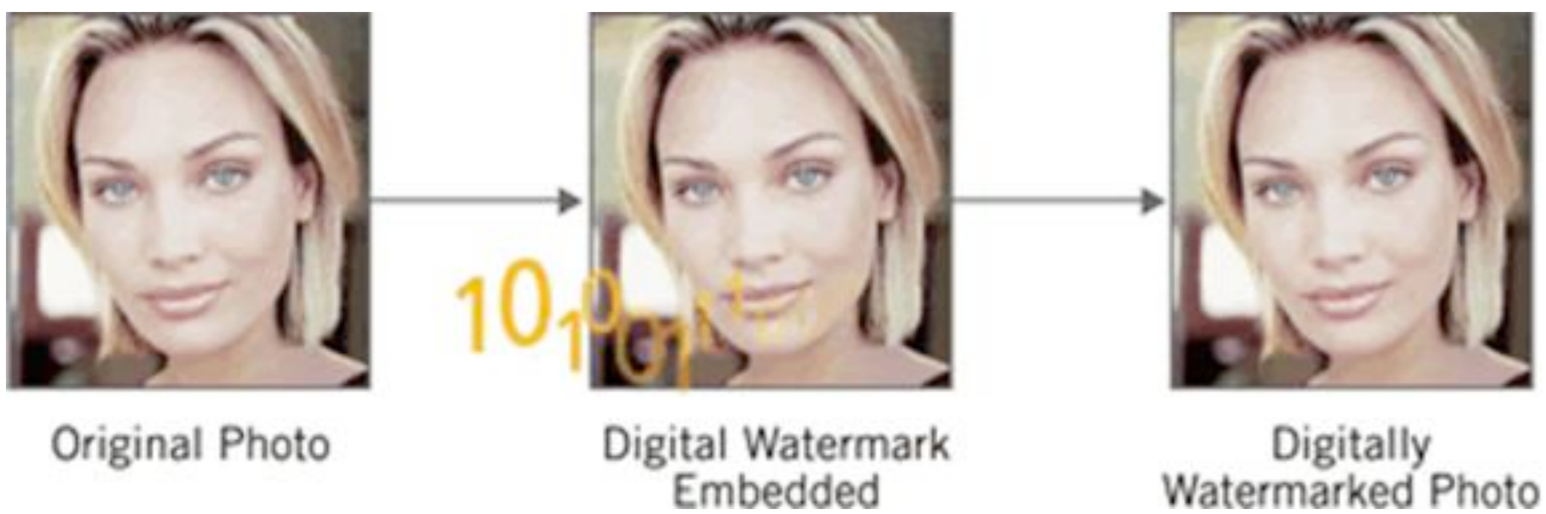
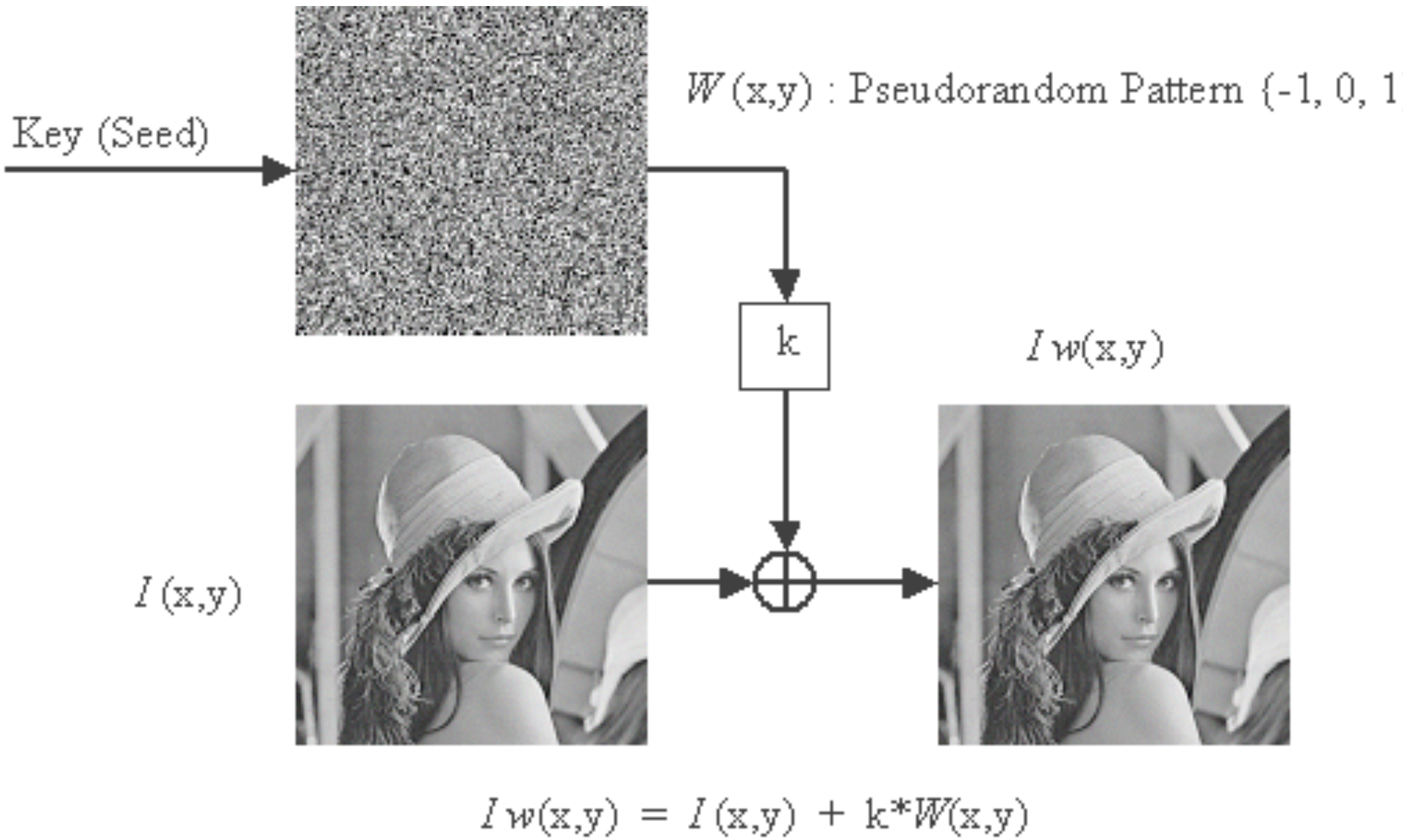
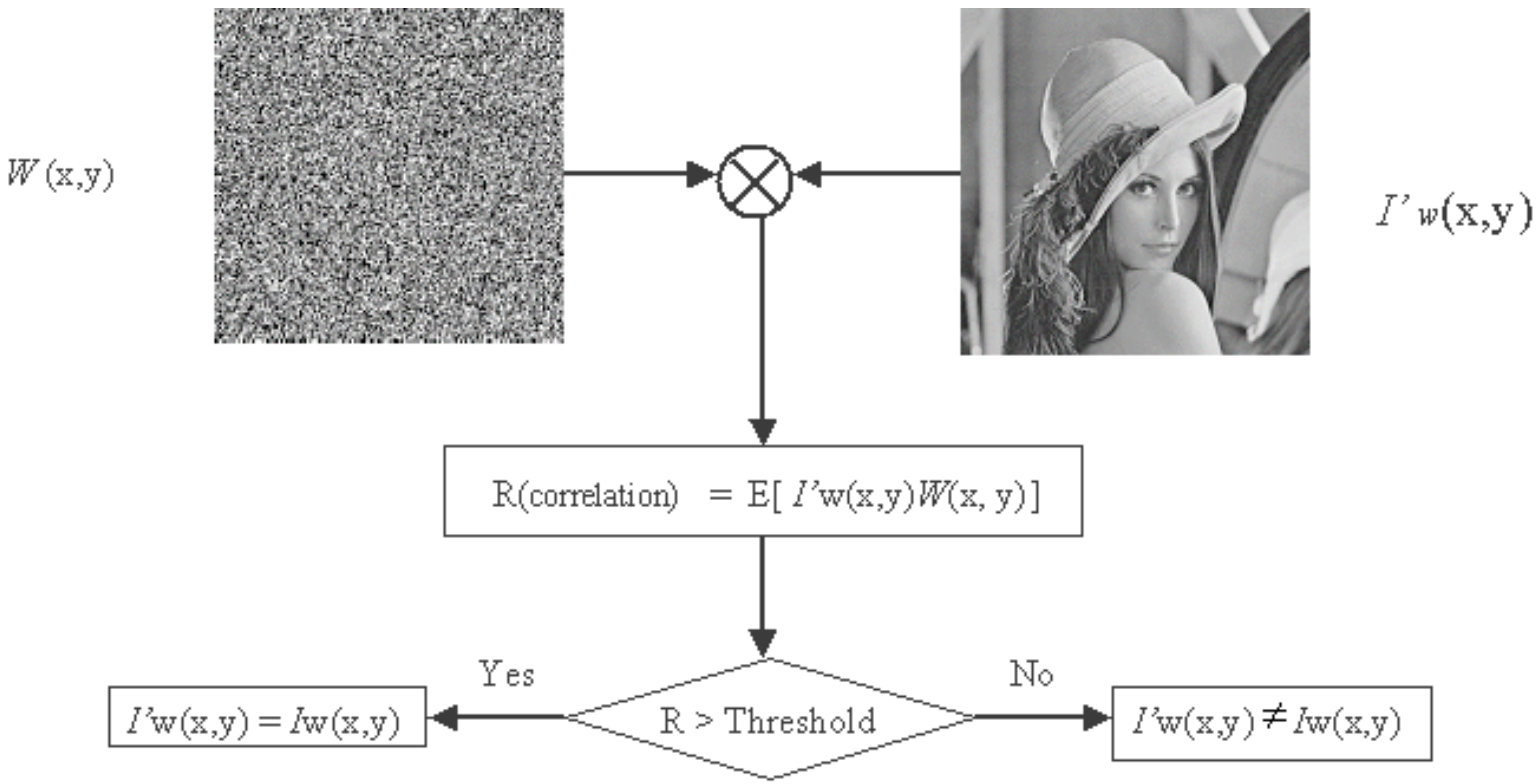


Image watermarking



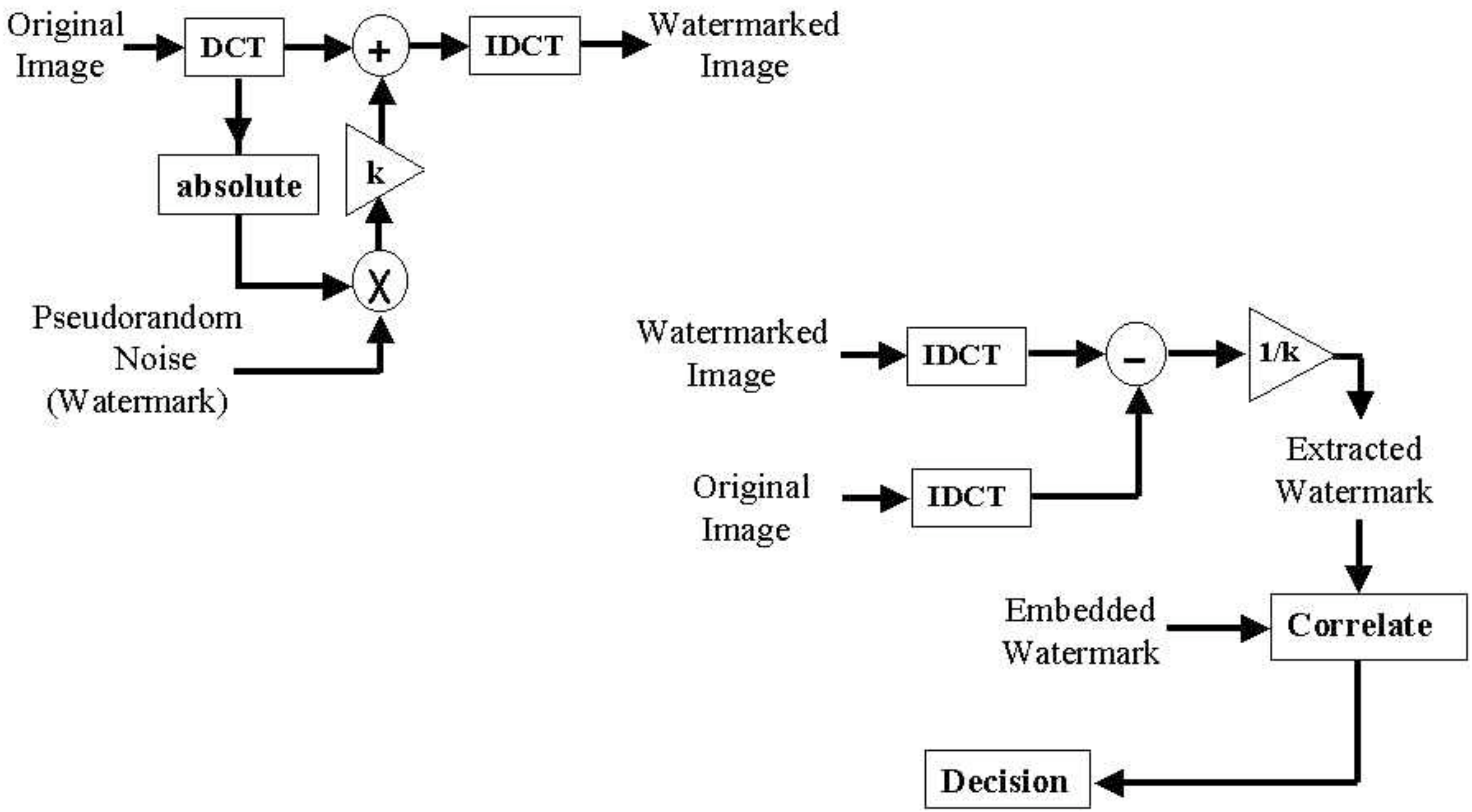
embedding

Image watermarking



detecting

DCT based algorithm

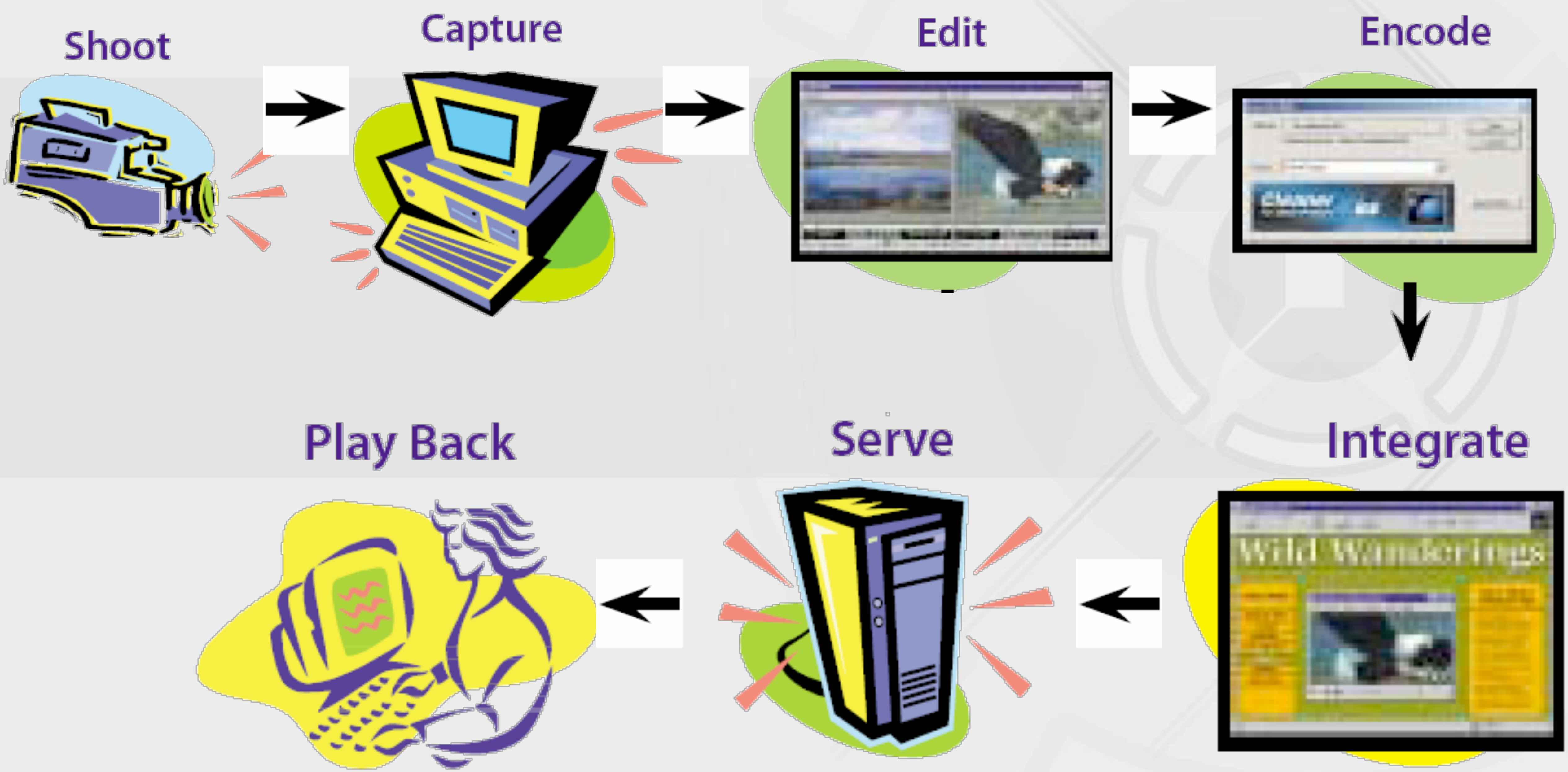


Streaming multimedia

- Framework: how it works?
- Protocol: standard solutions

成章乃达，若金之在熔

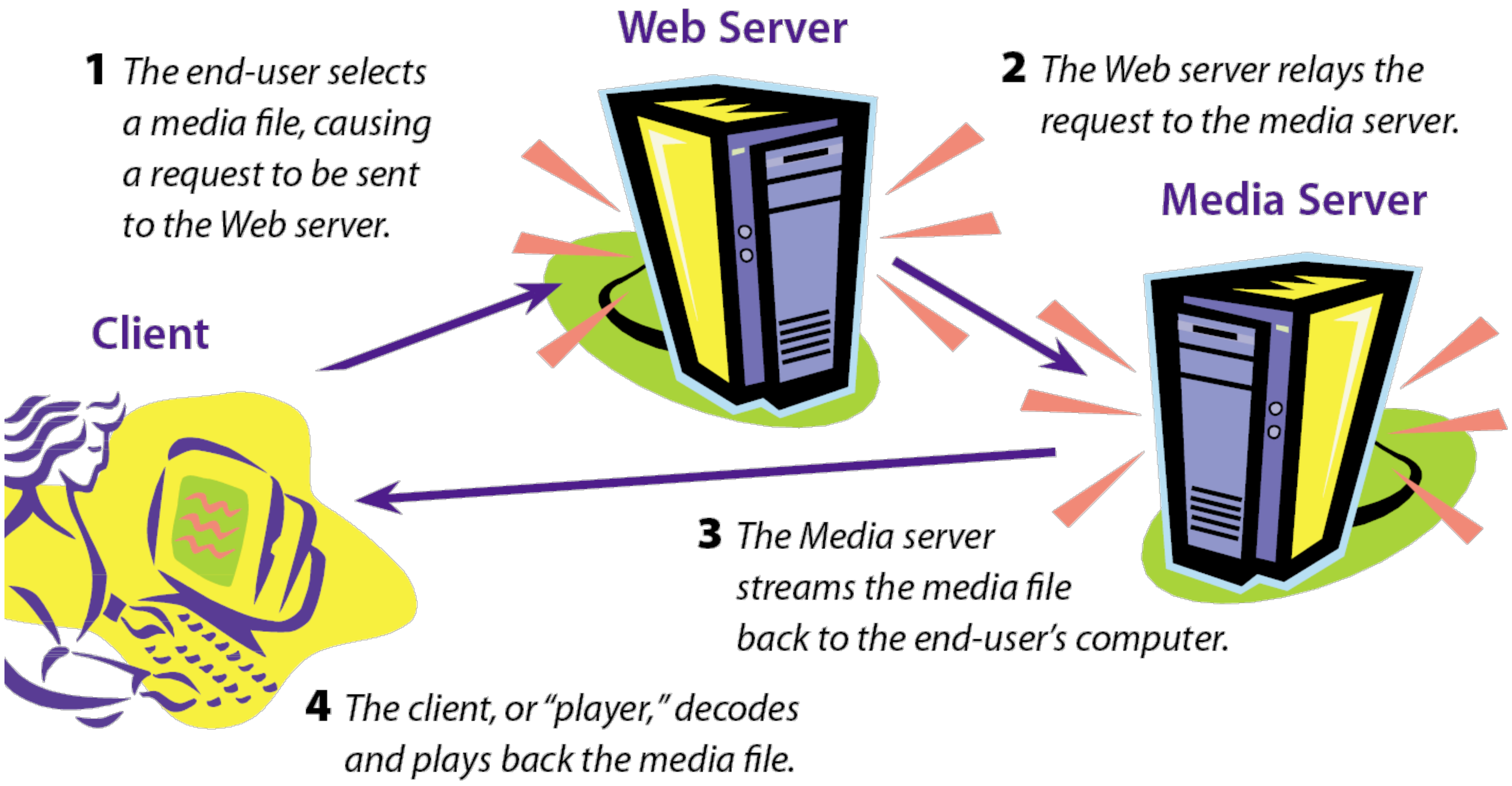
Workflow of Streaming Media



The primary characteristics of "streaming media"

- Three primary characteristics combine to define streaming media
  - Streaming media technology enables real-time or on-demand access to multimedia content via the Internet or an intranet.
  - Streaming media is transmitted by a media server application, and is processed and played back by a client player application, as it is received.
  - A streamed file is received, processed, and played simultaneously and immediately, leaving behind no residual copy of the content on the receiving device.

HOW DOES STREAMING WORK?

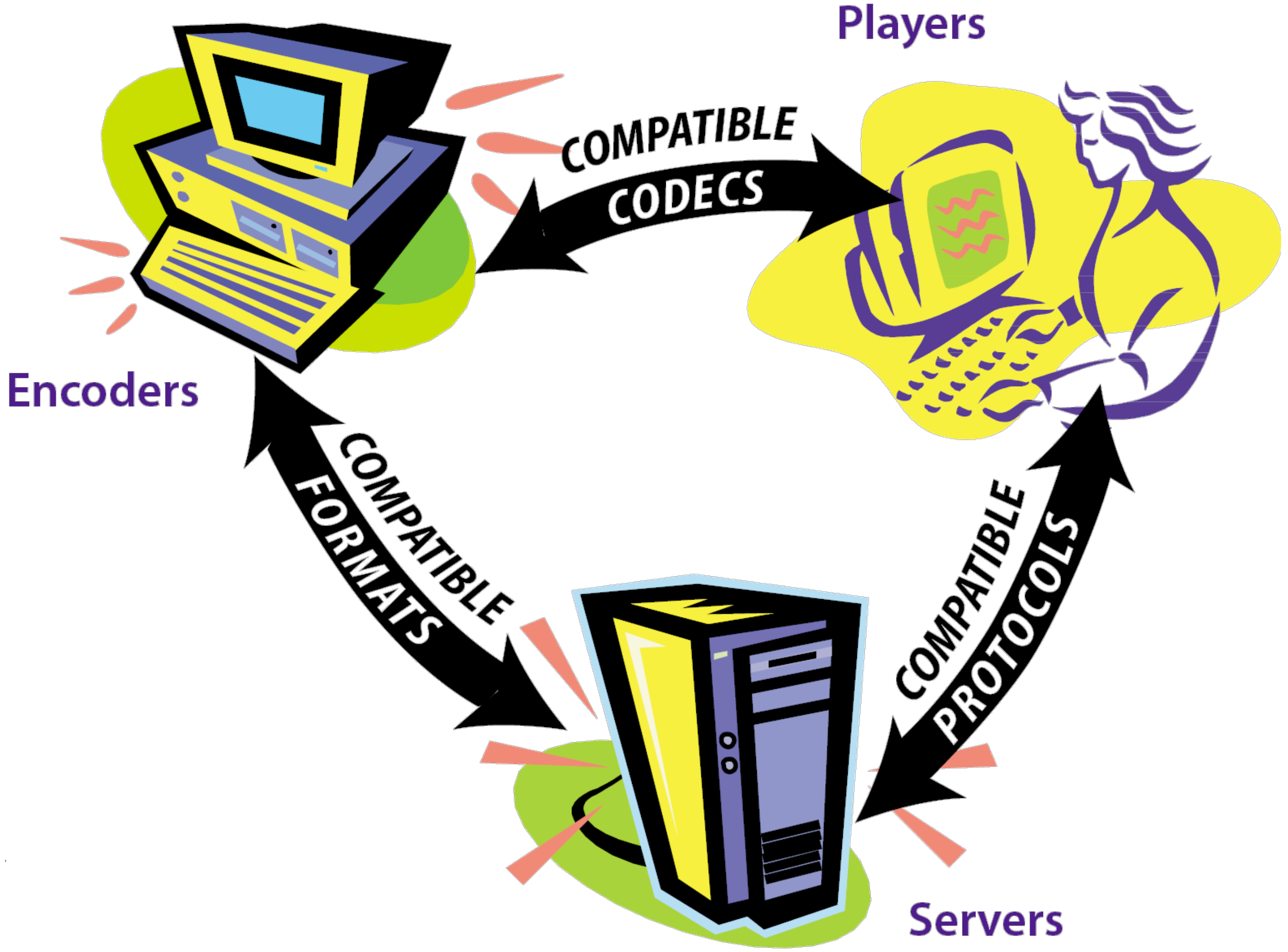


WHERE DO STREAMS COME FROM?

- Streaming media architectures.
  - Streaming media architectures are comprised of
    - encoding and transmission methods,
    - server software, and
    - players (client software).
  - The three most popular streaming media architectures
    - RealMedia,
    - Windows Media, and
    - QuickTime.

It is all interrelated

- In a streaming architecture, everything must be compatible.



Streaming media formats

Architecture	Native Formats	Streaming Media File Extensions
QuickTime	QuickTime Format	.mov (sometimes .qt or .qti)
RealMedia	RealMedia Format	.rm
Windows Media	Advanced Streaming Format or Windows Media Video/Audio	.asf, .wmv, .wma

- **MPEG standard**
  - Windows Media Video v1 is a derivative of the MPEG-4 codec, which has been renamed to avoid confusion.
  - QuickTime 5 is the first full implementation of MPEG-4 for streaming media.

Streaming ...

- Progressive streaming transport (PST)
  - use HTTP
  - no jump

- Real-time streaming transport
  - Real server (Real-time streaming protocol, RTSP)
  - Windows Media server (M$ media server, MMS)
  - Quicktime server

Unicast and Multicast

- When media is streamed over the Internet, it may be either **unicast** or **multicast**.

  – ***Unicast:*** A unicast is a one-to-one "narrowcast," in which each end-user gets a separate stream—even if they are experiencing the same media simultaneously.

  – ***Multicast:*** Multicasting, or "IP multicasting," is also considered a narrowcast strategy, and it is designed to conserve both server processing capacity and bandwidth.

Streaming bandwidth and storage

- Streaming media storage size is calculated from streaming bandwidth and length of the media with the following formula:

$$\textit{storage size (in megabytes)}$$
$$= \textit{length (in seconds)} \cdot \textit{bit rate (in kbit/s)} / 8{,}388.608$$

➢ 1 megabyte = 8 * 1,048,576 bits = 8,388.608 kilobits

Streaming bandwidth and storage

- Real world example:
  - One hour of video encoded at 300 kbit/s (this is a typical broadband video for 2005 and it's usually encoded in a 320×240 pixels window size) will be:
    - (3,600 s · 300 kbit/s) / 8,388.608 = 128.7 MB of storage
  - Assume the file is stored on a server for on-demand streaming. If this stream is viewed by 1,000 people, you would need
    - 300 kbit/s · 1,000 = 300,000 kbit/s = 300 Mbit/s of bandwidth

QoS: Quality of Service

- A defined measure of performance in a data communications system
- Refers to the control mechanisms in the network software that make the actual determination of which packets have priority
  - provide different priority to different users or data flows, or
  - guarantee a certain level of performance to a data flow in accordance with requests from the application program

IETF IP QoS Efforts

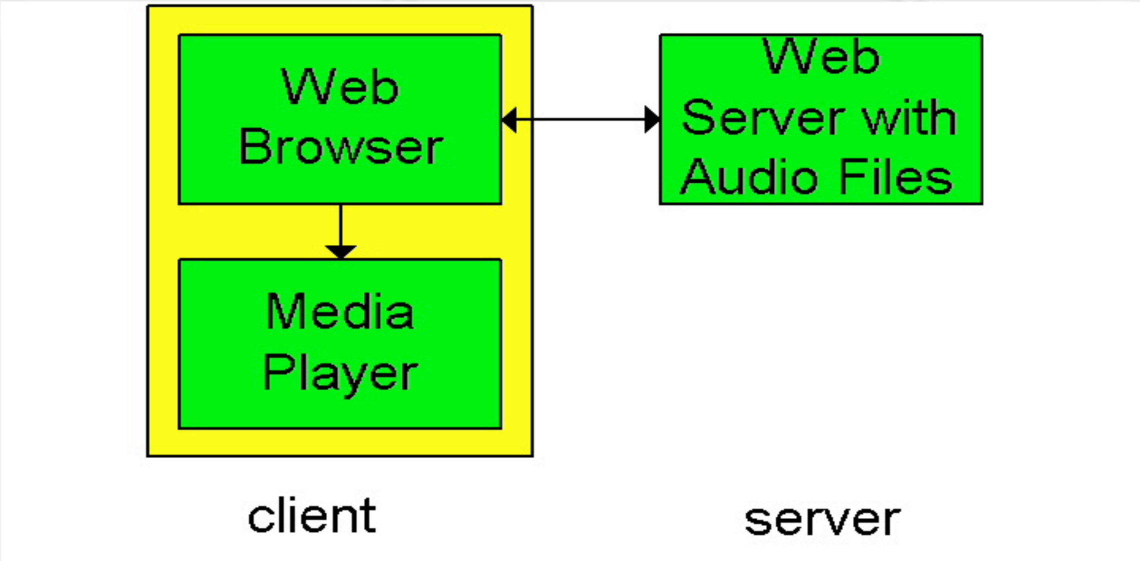
- Policy based IP QoS Solutions
  - Integrated Services (RSVP protocol): flow based
  - Differentiated Services (DiffServ byte settings): packet based
  - Multi-Protocol Label Switching (MPLS): flow +packet based
- IP Multicast and Anycast
- IPv6 QoS Support

Streaming from Web server (1)

- Audio and video files stored in Web servers

naïve approach

- browser requests file with HTTP request message
- Web server sends file in HTTP response message
- content-type header line indicates an audio/video encoding
- browser launches media player, and passes file to media player
- media player renders file

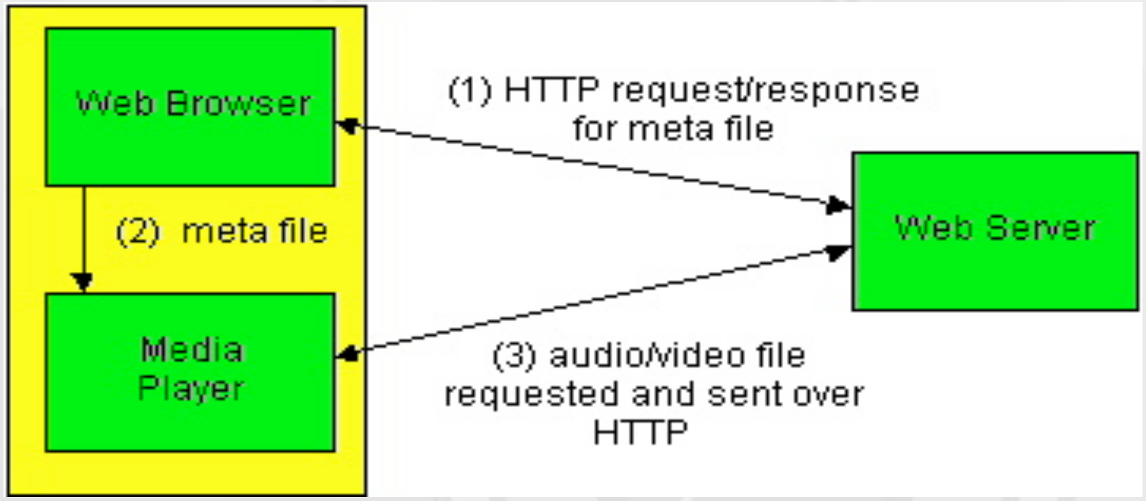


- Major drawback: media player interacts with server through intermediary of a Web browser

Streaming from Web server (2)

Alternative: set up connection between server and player

- Web browser requests and receives a **meta file** (a file describing the object) instead of receiving the file itself;
- Content-type header indicates specific audio/video application
- Browser launches media player and passes it the meta file
- Player sets up a TCP connection with server and sends HTTP request.



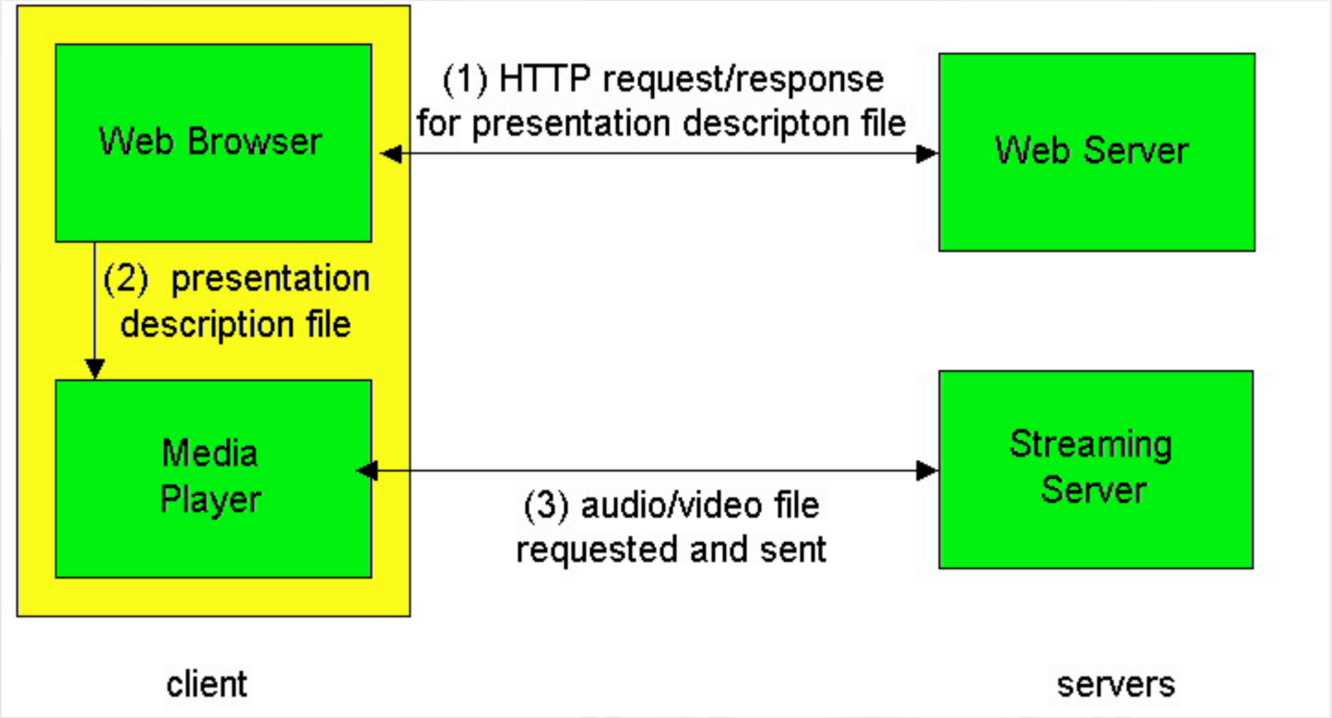
Some concerns:

- Media player communicates over HTTP, which is not designed with pause, ff, rwnd commands
- May want to stream over UDP

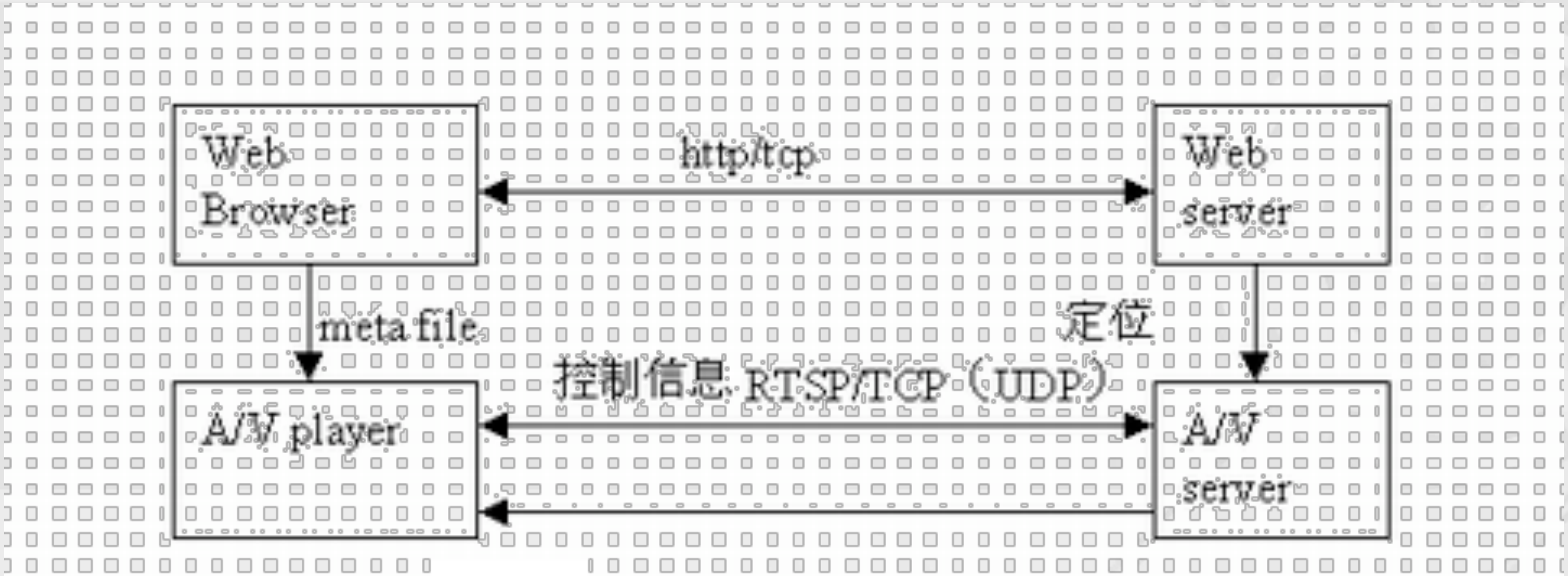
浙江大学计算机学院
数字媒体与网络技术

Streaming from a streaming server

- This architecture allows for non-HTTP protocol between server and media player
- Can also use UDP instead of TCP.



Streaming from a streaming server



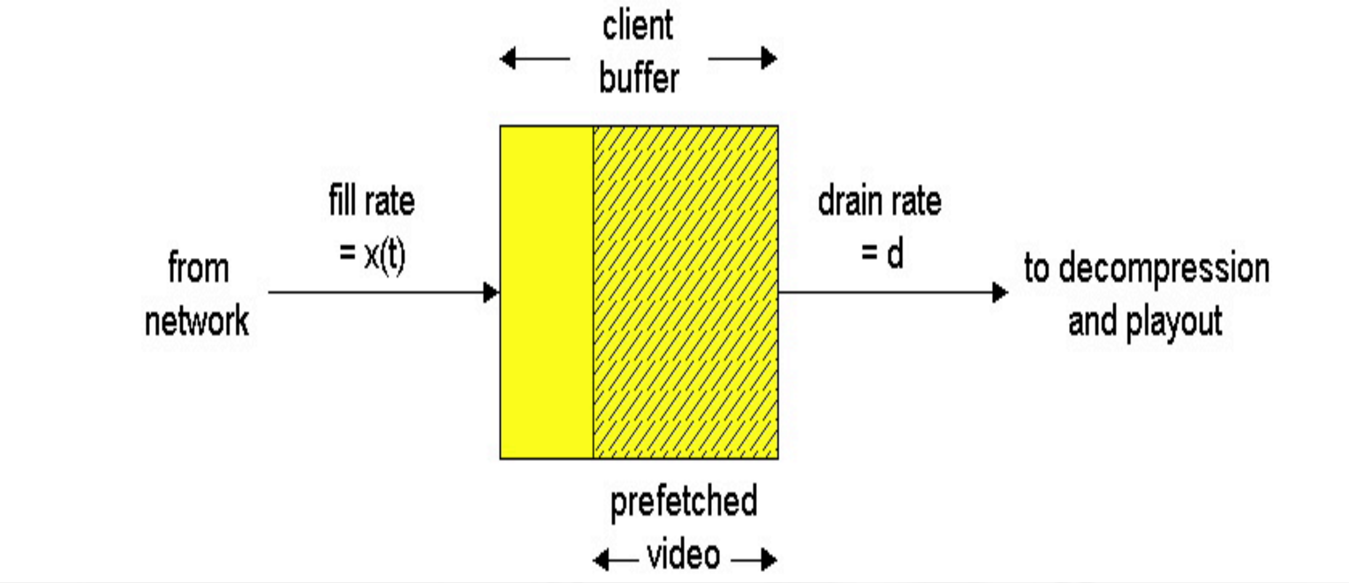
流式传输基本原理

Options when using a streaming server

- Send at constant rate over UDP. To mitigate the effects of jitter, buffer and delay playback for 1-10 s. Transmit rate = d, the encoded rate. Fill rate x(t) equals d except when there is loss.

- Use TCP, and send at maximum possible rate under TCP; TCP retransmits when error is encountered; x(t) now fluctuates, and can become much larger than d. Player can use a much large buffer to smooth delivery rate of TCP.

-



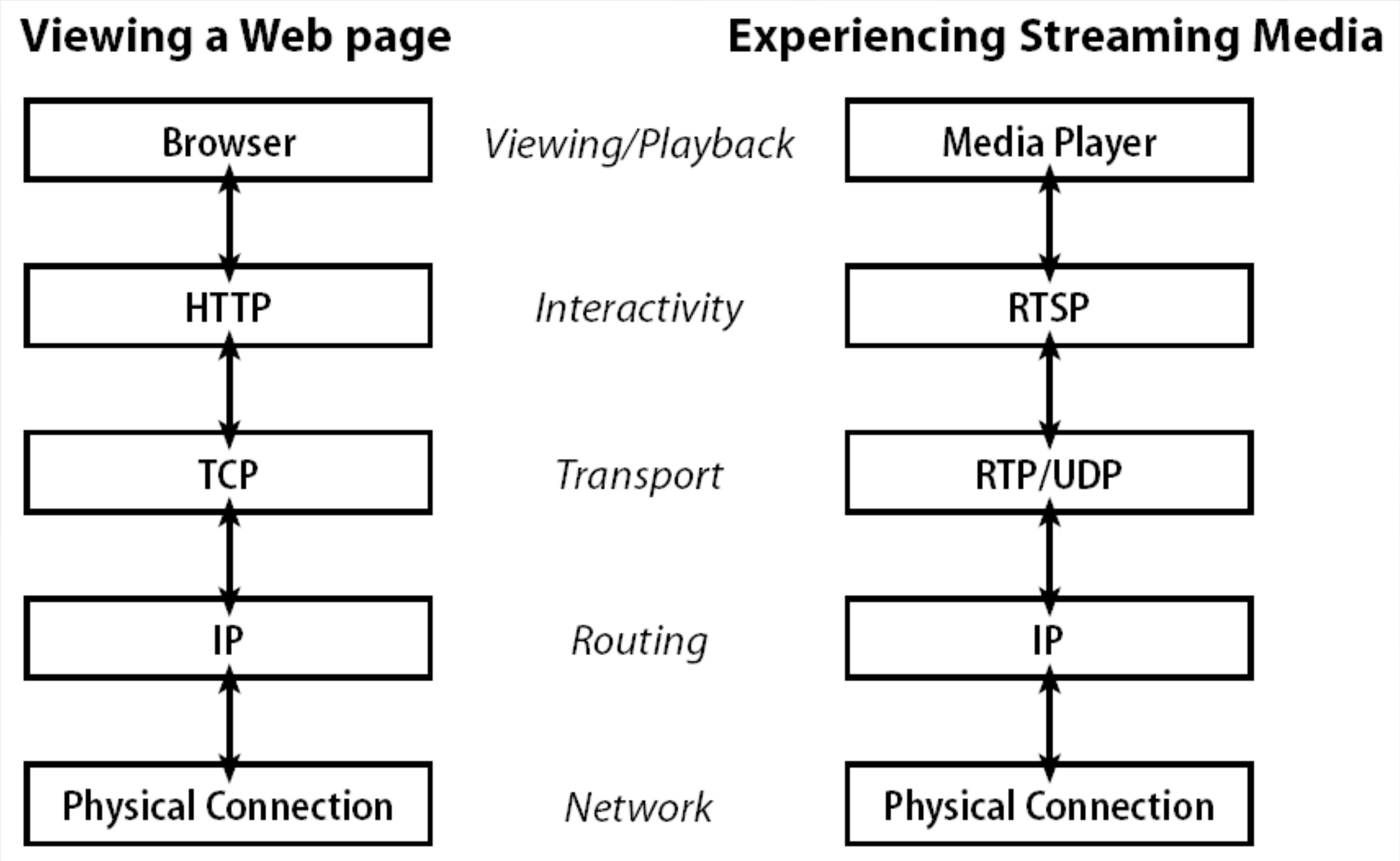
Streaming Protocols

- **RTP (Real-time Transport Protocol)**
  - similar to **HTTP** and to **FTP**—protocols used by Web servers
  - However, **HTTP** and **FTP** cannot be used for true streaming
    - both layered on top of **TCP**
  - layered on top of **UDP**
  - RTP enables a one-way stream, transmitting media from the server to the client

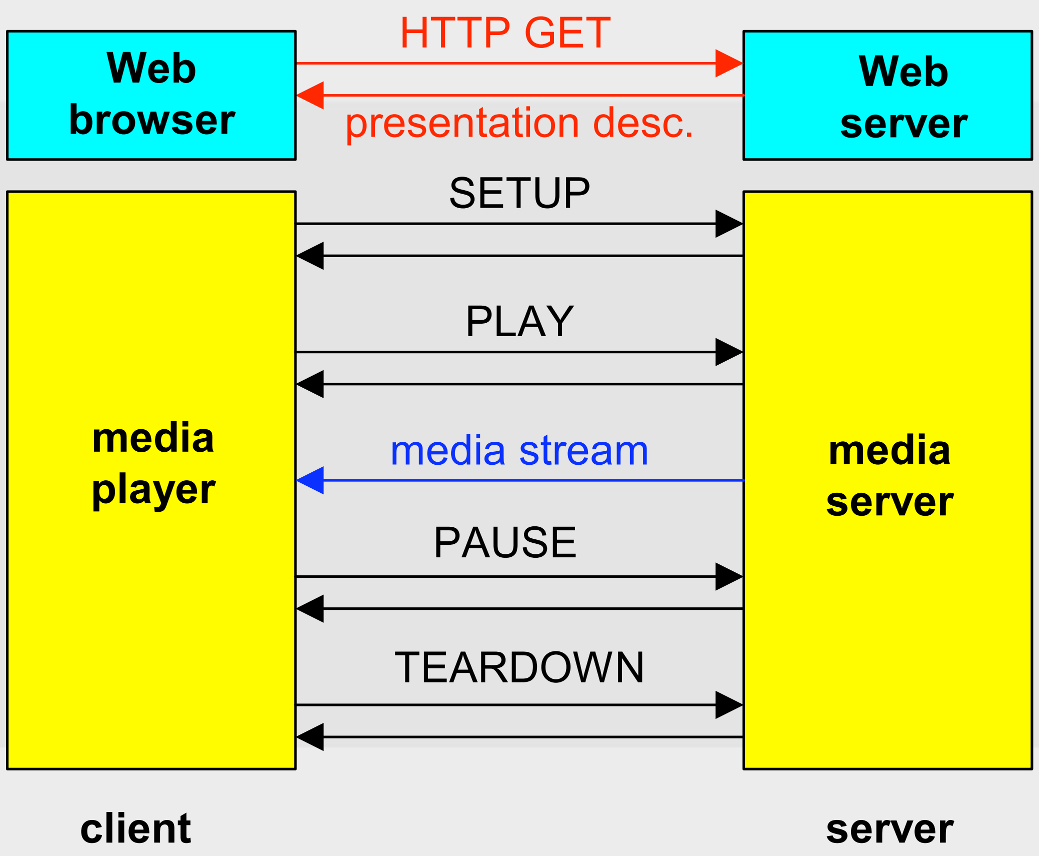
- **RTSP (Real-time Streaming Protocol)**
  - a two-way protocol which uses TCP to communicate, and
  - usually layered on top of RTP

- **RTCP (Real-time Transport Control Protocol)**

Streaming Protocols



RTSP initiates and controls delivery



- Client obtains a description of the multimedia presentation, which can consist of several media streams.
- The browser invokes media player (helper application) based on the content type of the presentation description.
- Presentation description includes references to media streams, using the URL method rtsp://
- Player sends RTSP SETUP request; server sends RTSP SETUP response.
- Player sends RTSP PLAY request; server sends RTSP PLAY response.
- Media server pumps media stream.
- Player sends RTSP PAUSE request; server sends RTSP PAUSE response.
- Player sends RTSP TEARDOWN request; server sends RTSP TEARDOWN response.

RTSP: exchange example

- C: SETUP rtsp://audio.example.com/twister/audio RTSP/1.0
- Transport: rtp/udp; compression; port=3056; mode=PLAY

- S: RTSP/1.0 200 1 OK
- Session 4231

- C: PLAY rtsp://audio.example.com/twister/audio.en/lofi RTSP/1.0
- Session: 4231
- Range: npt=0-

- C: PAUSE rtsp://audio.example.com/twister/audio.en/lofi RTSP/1.0
- Session: 4231
- Range: npt=37

- C: TEARDOWN rtsp://audio.example.com/twister/audio.en/lofi RTSP/1.0
- Session: 4231

Multimedia retrieval

- Why do we need content based retrieval?
- Main difficulties
  - Feature extraction
  - Similarity measure
- Main ideas of different searching methods

习坎示教，始见经纶

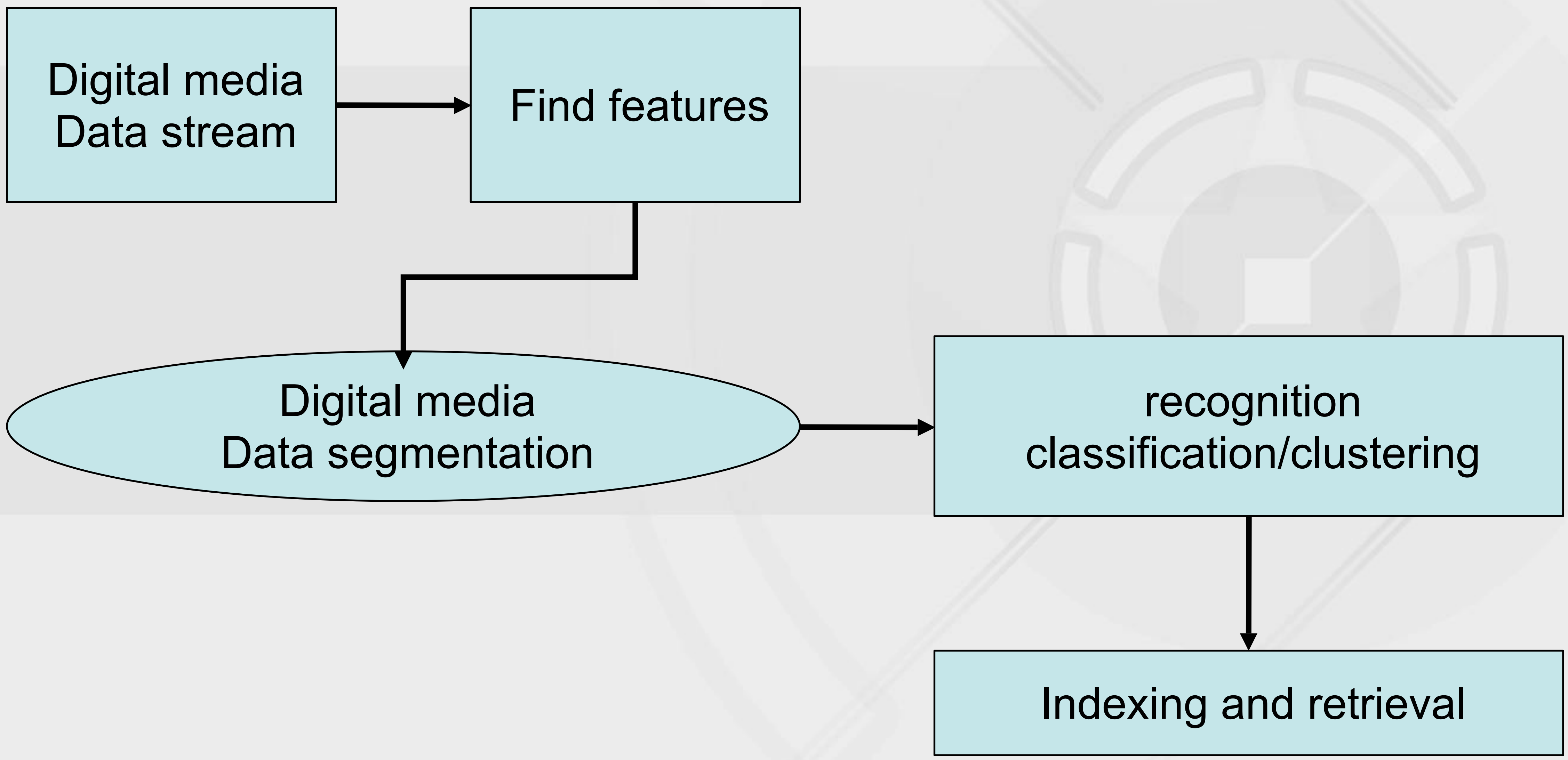
Origin of digital media retrieval

- IR (Information retrieval)
  - To retrieve information that users want based on some keys or hints

Main methods of digital media retrieval

- **Text-based** digital media retrieval
  - –Boolean model
  - –Clustering model
  - –Vector model
  - –Probability model
- **Content-based** digital media retrieval

The workflow of digital media analysis and retrieval



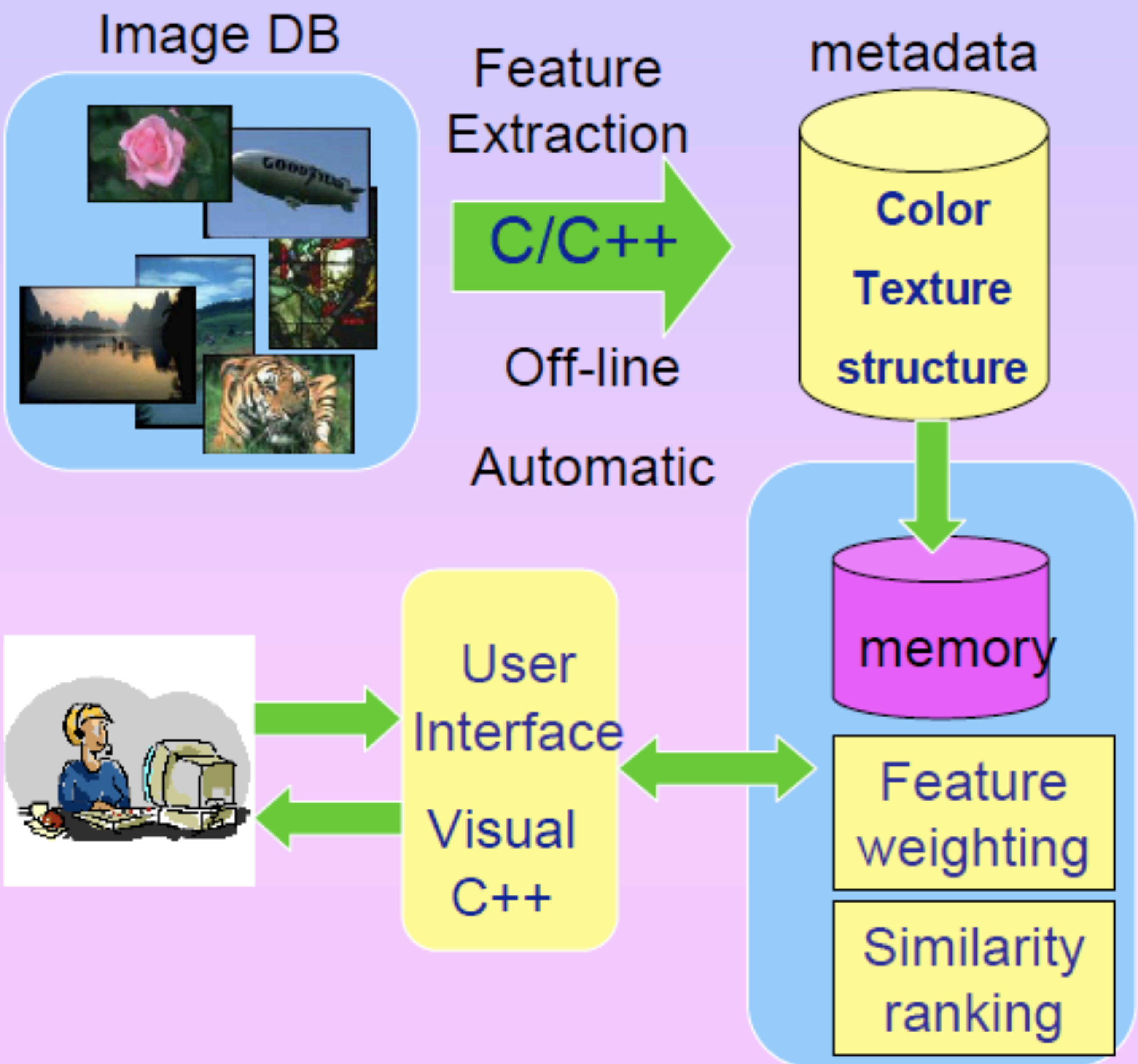
Content-based digital media retrieval

- Content-based image retrieval
- Content-based video retrieval
- Content-based audio retrieval
- Content-based graphics retrieval

- Merging and analysis of multiple media
- Development and challenging

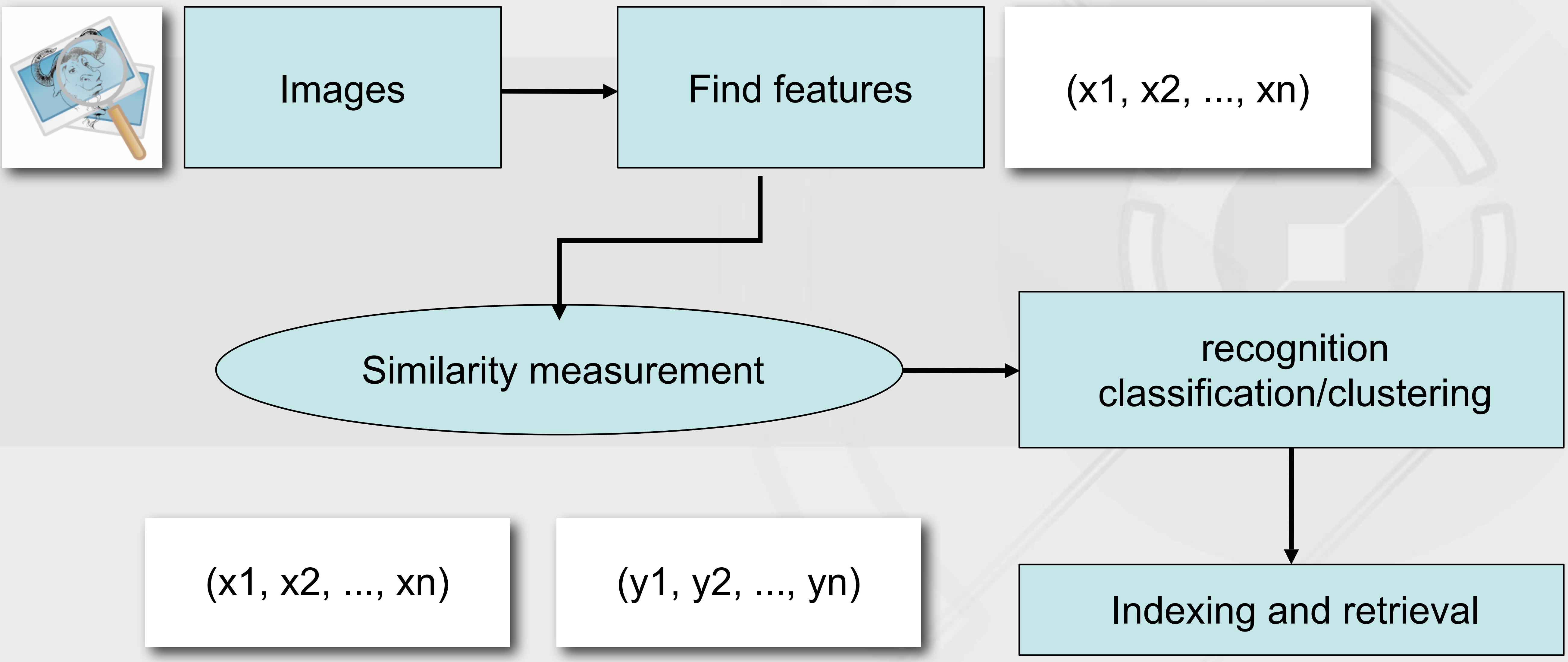
Multimedia Information Retrieval

- **Content-based Image Retrieval**



- **Color**
  - ✓Color histogram
  - ✓Color moments
  - ✓Color correlogram
- **Texture**
  - ✓Tamura texture
  - ✓Co-occurrence matrices
  - ✓Gabor features
  - ✓Wavelet moments
- **Shape**
  - ✓Fourier descriptor
- **Structure**
  - ✓Edge-based features

Workflow of CBIR



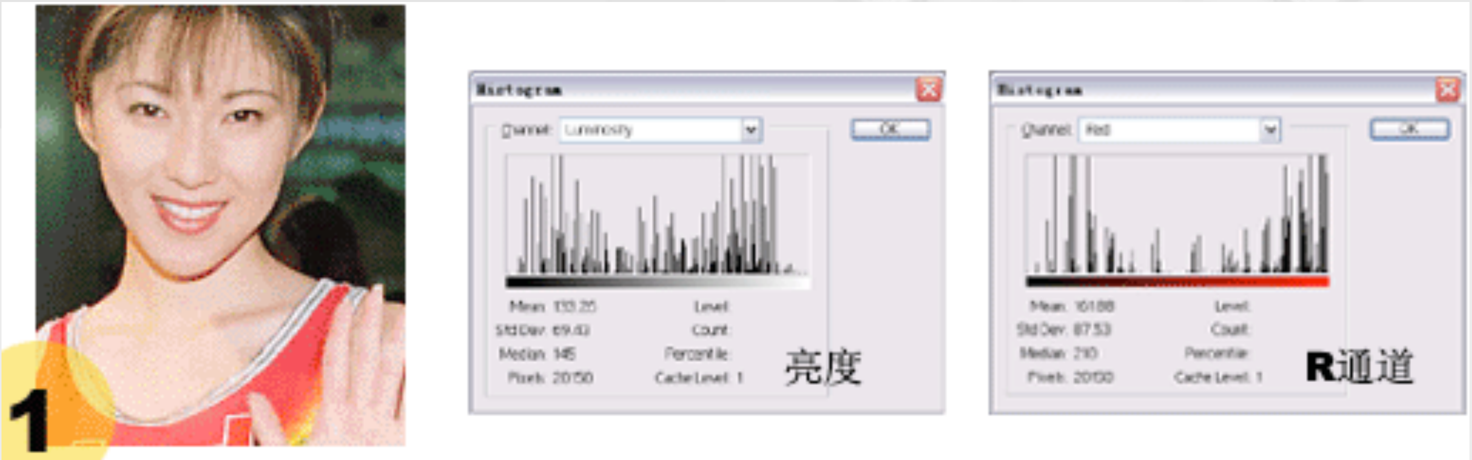
Features of image

- Finding out features of image is a key step of image retrieval
  - Image-based retrieval usually need to pre-construct feature database of images for retrieval

- Major image features:
  - Color features
  - Texture features
  - Shape features
  - Space relation features

Color features of image

- Color feature is a most widely used vision feature. It is mainly used to analyze color distributions in an image, including:
  - Color histogram
  - Color moments
  - Color set
  - Color clustering vectors
  - Color relation graph



图像的颜色矩 (color moments)

- Color moments are global statistical features of an image, which are proposed by Stricker and Orengo.
  - First order moment (mean) $$\mu_i = \frac{1}{n}\sum_{j=1}^{n} I_{ij}$$
  - Second order moment (variance) $$\sigma_i^2 = \frac{1}{n}\sum_{j=1}^{n} (I_{ij} - \mu_i)^2$$
  - Third order moment (skewness) $$s_i^3 = \frac{1}{n}\sum_{j=1}^{n} (I_{ij} - \mu_i)^3$$
- Color moments are always applied with other image features for efficiently shrinking seeking ranges.

color moments: example

1	3	6	3	1
3	6	8	6	3
6	8	10	8	6
3	6	8	6	3
1	3	6	3	1

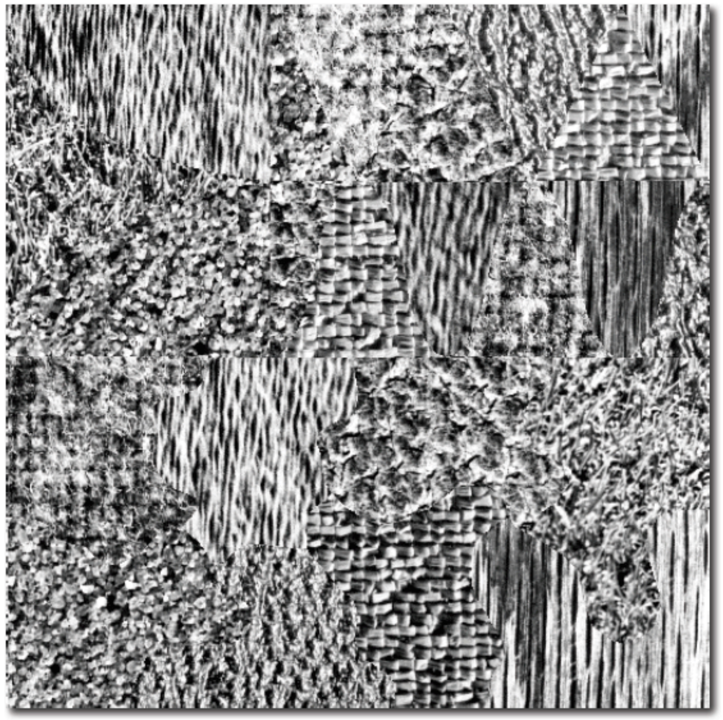
mean =4.72

variance =6.52

skewness =2.34

Image texture features

- Texture features are such vision features employed to measure homogeneous phenomenon in images. They are
  - independent to color or illuminance,
  - and are intrinsic features of object surfaces.



- Major texture features
  - Tamura texture features
  - Self-regression texture model
  - Transform based texture features
    - DWT，DFT，Garbor filter bank
  - others

Image texture features

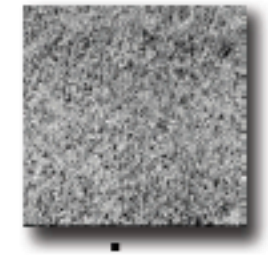
- Texture features are such vision features employed to measure homogeneous phenomenon in images. They are
  - independent to color or illuminance,
  - and are intrinsic features of object surfaces.

- Major texture features
  - Tamura texture features
  - Self-regression texture model
  - Transform based texture features
    - DWT，DFT，Garbor filter bank
  - others

Tamura texture features

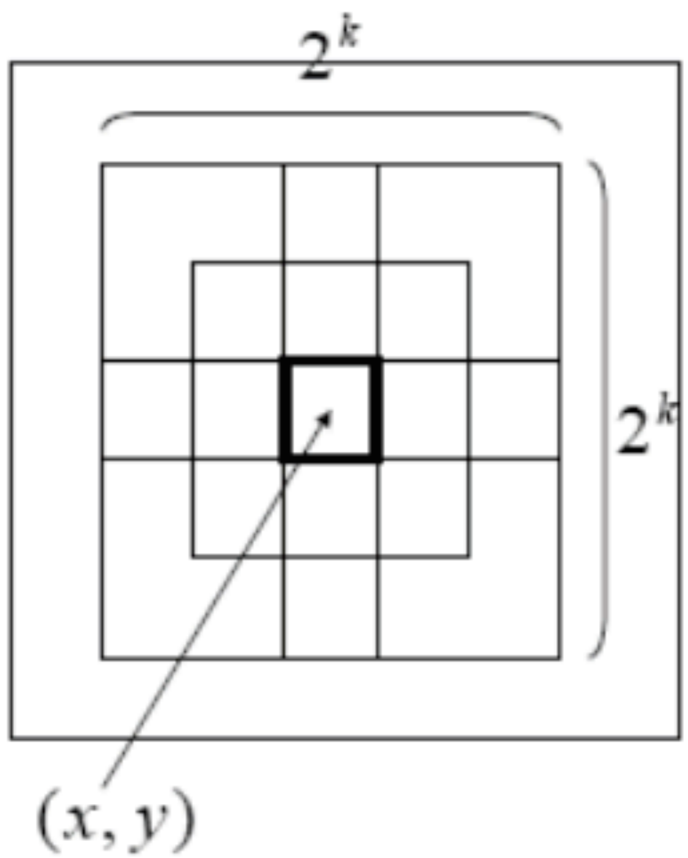
- a set of texture feature representation based on the psychology research results on human vision cognition of textures:
  - coarseness (粗糙度)
  - contrast (对比度)
  - directionality (方向度)
  - linelikeness (线像度)
  - regularity (规整度)
  - roughness (粗略度)

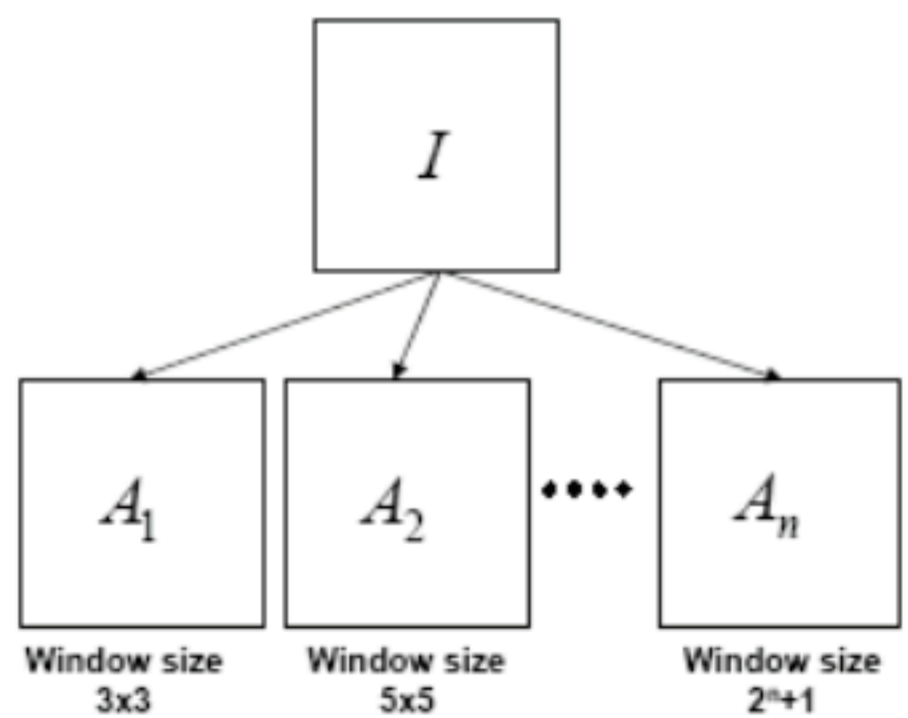
Tamura – Coarseness



- Goal
  - Pick a large size as best when coarse texture is present, or a small size when only fine texture
- Step 1: Compute averages at different scales at every points

$$A_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} I(i,j) / 2^{2k}$$

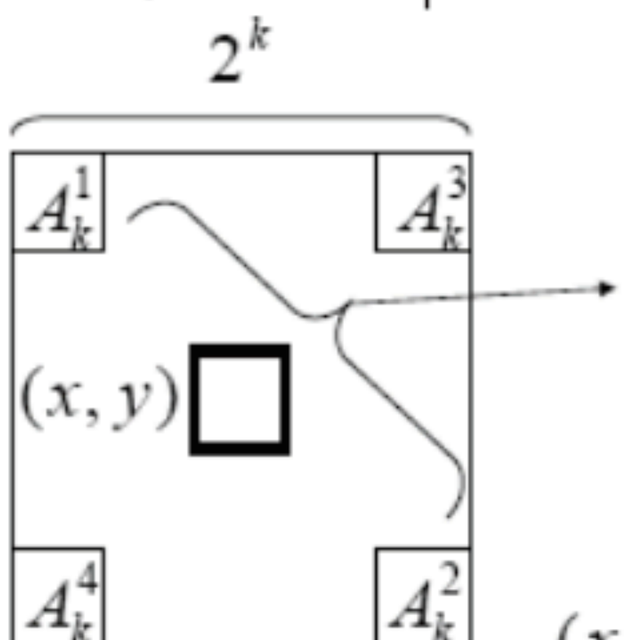




Tamura – Coarseness (cont.)

- Step 2: compute neighborhood difference at each scale on opposite sides of different directions

$$E_{k,h}(x,y) = \left| A_k(x-2^{k-1},y) - A_k(x+2^{k-1},y) \right|$$



$$E_{k,a}(x,y) = \left| A_k^{\ 1} - A_k^{\ 2} \right|$$

$$E_{k,b}(x,y) = \left| A_k^{\ 3} - A_k^{\ 4} \right|$$

$$(x,y) \longrightarrow \{E_{1,a}, E_{1,b}, E_{2,a}, E_{2,b}, \ldots, E_{n,a}, E_{n,b}\}$$

Tamura – Coarseness (cont.)

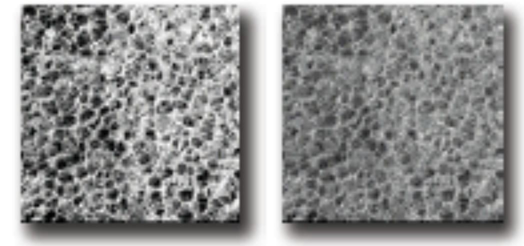
- Step 3: select the scale with the largest variation

$$S_{max}(x,y) = 2^k \quad / \quad E_k = max\{E_1, E_2, \ldots, E_L\}$$

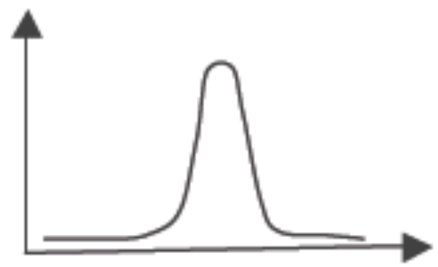
- Step 4: compute the coarseness

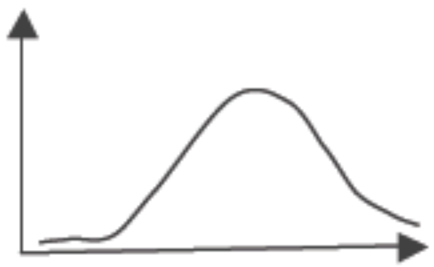
$$M_{crs} = \frac{1}{n \times m} \sum_{i}^{n} \sum_{j}^{m} S_{max}(i,j)$$

Tamura – Contrast

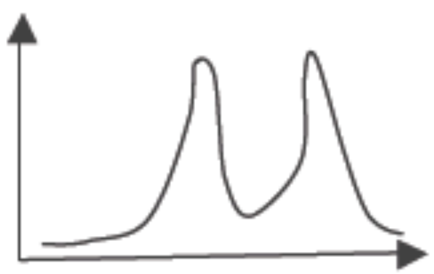


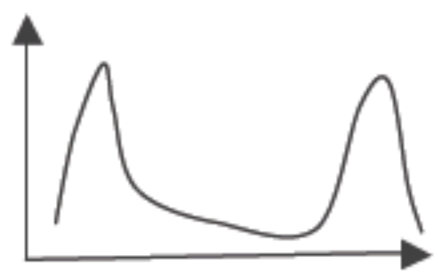
- Gaussian-like histogram distribution → low contrast





- Histogram polarization. Is it Gaussian? How many peaks it has? Where they are?





- Polarization can be estimated by the kurtosis (曲率度)

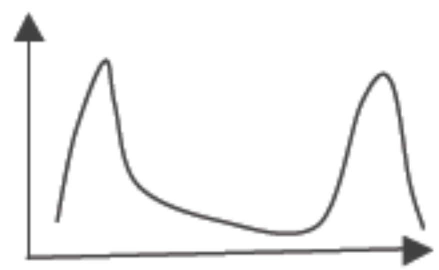
$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

$$\mu_4 = E[I^4(x, y)]$$

$$\sigma^4 = E[(I(x, y) - \mu)^4]$$

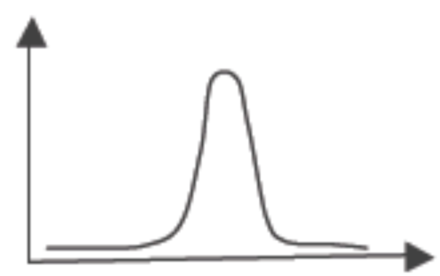
Tamura – Contrast (cont.)

$$\alpha_4 = \frac{\mu_4}{\sigma^4} \downarrow$$



distribution with two separate peaks

$$\alpha_4 = \frac{\mu_4}{\sigma^4} \uparrow$$

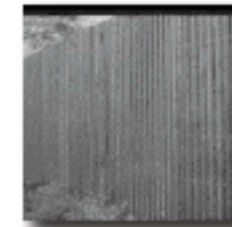


unimodal distribution

- Contrast estimate is given by:

$$M_{contrast} = \frac{\sigma}{(\alpha_4)^{\frac{1}{4}}}$$

Tamura – Orientation

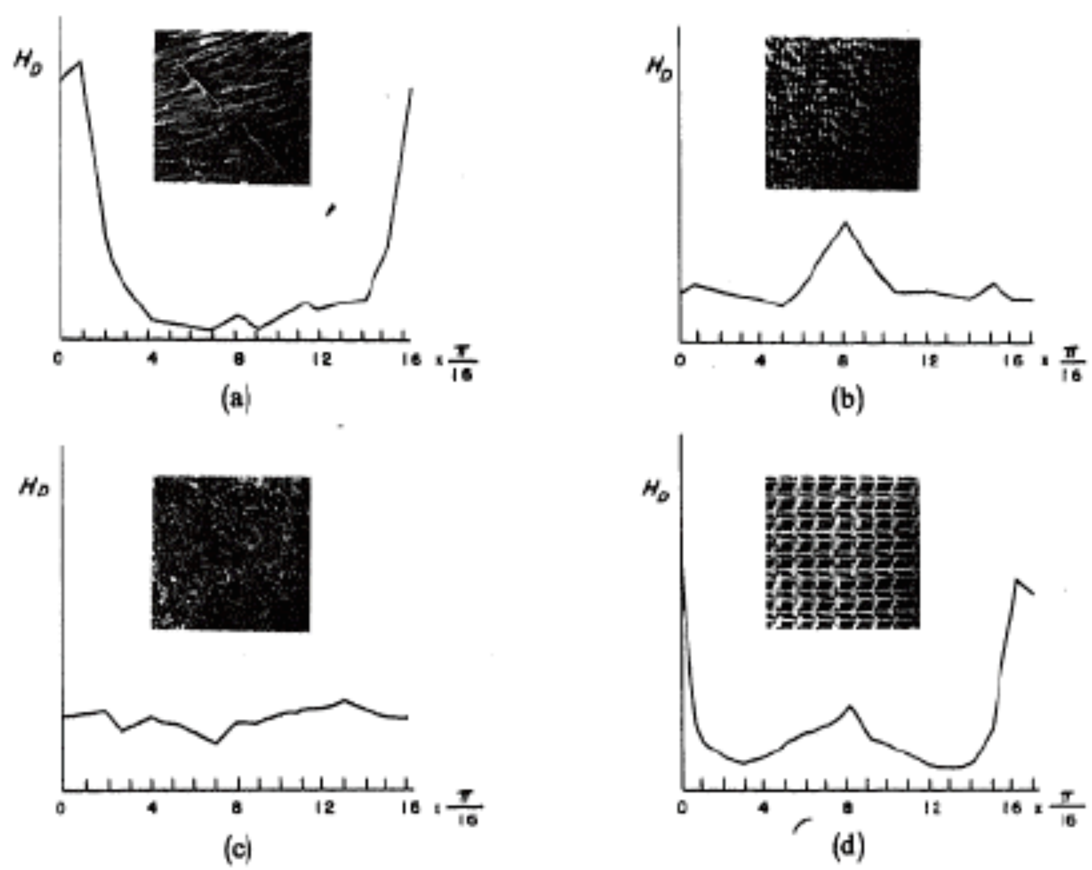


- Building the histogram of local edges at different orientations $H_D(k)$
  - By deriving the edge magnitude at X and Y directions

$$\theta = \mathrm{tg}^{-1}(\nabla_V / \nabla_H) + \frac{\pi}{2}$$

$$|\nabla G| = \left(|\nabla_V| + |\nabla_H|\right)/2$$

$$\nabla_V = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \qquad \nabla_H = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix}$$



Tamura – Orientation (cont.)

- Compute the estimate from the sharpness of the peaks
  - By summing the second moments around each peak
    e.g., flat histogram
    → large 2nd moment (variance)
    → small orientation

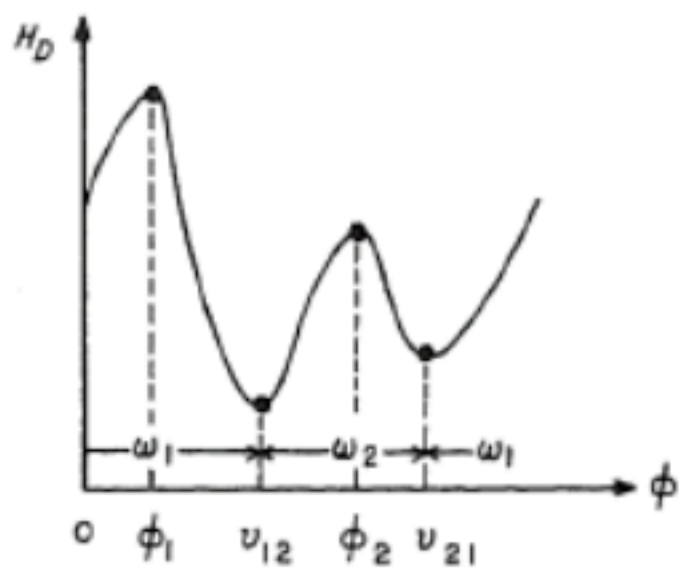
$$M_{orient} = 1 - r \cdot n_p \cdot \sum_{p}^{n_p} \sum_{\phi \in w_p} (\phi - \phi_p)^2 \cdot H_D(\phi)$$

$n_p =$ *Number of peaks*

$\phi_p =$ *Position of peak, p, in* H_D

$w_p =$ *Points in peak p*

$r =$ *Normalisation factor*

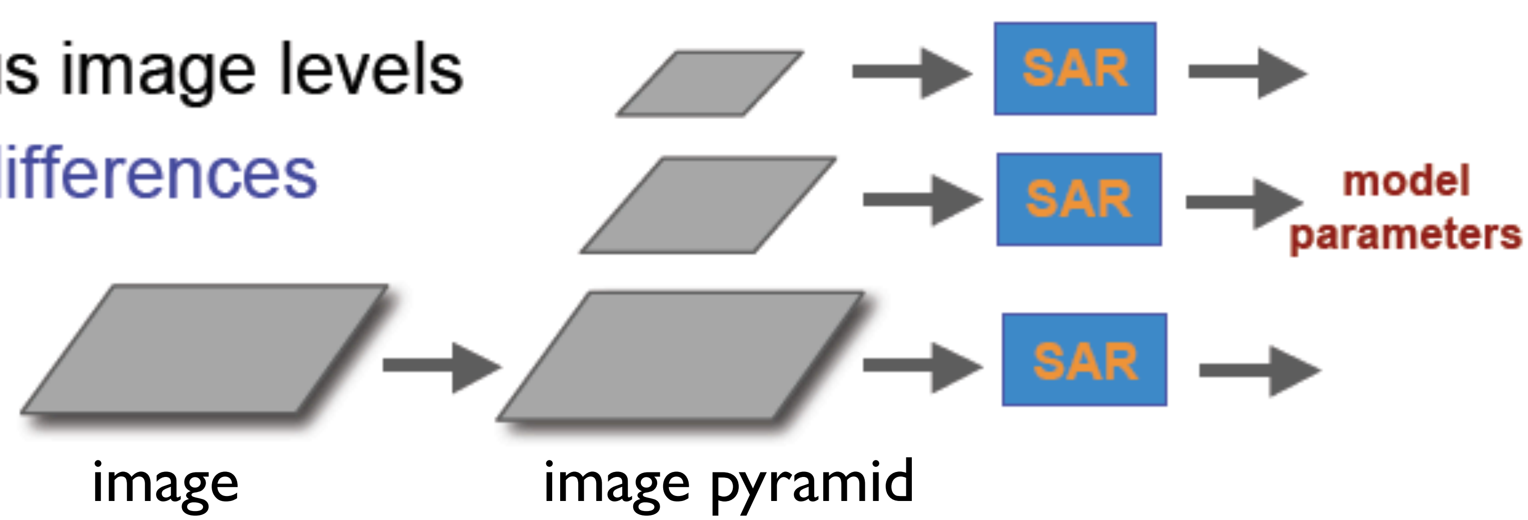


(MR)SAR

[Mao'92]

- Each pixel is a random variable whose value is estimated from its neighboring pixels + noise
  - A kid of Markov Random Field model

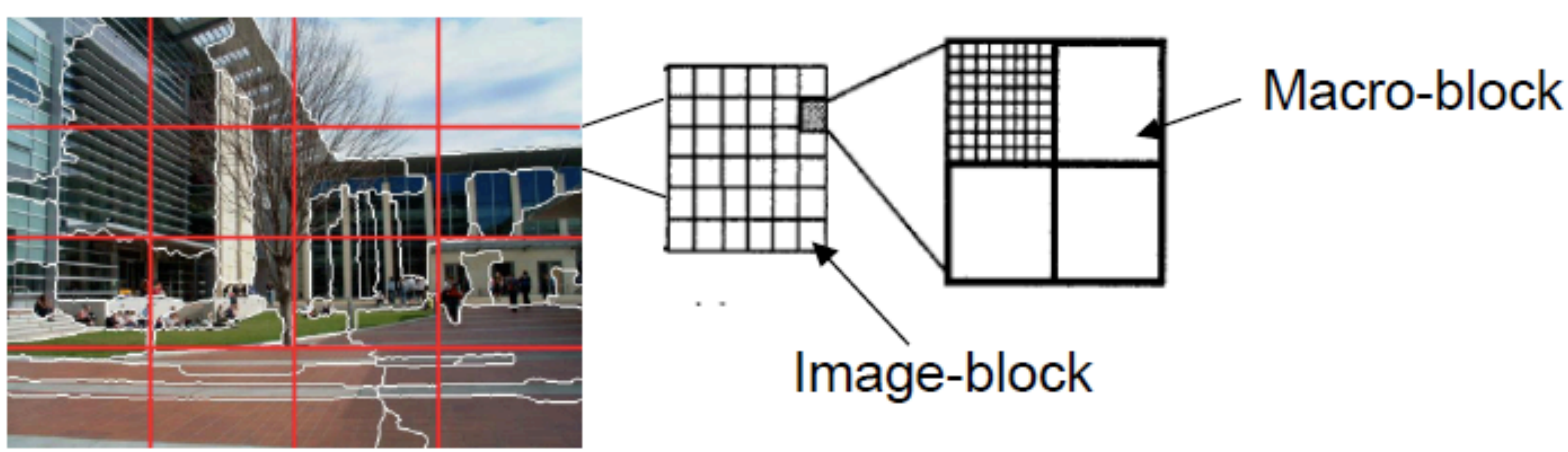
- **SAR** Model (Simultaneous Autoregressive)
  - Describes each pixel in terms of its neighboring pixels.
- **MRSAR** Model (MultiResolution SAR)
  - Describing granularities by representing textures at variety of resolutions
  - SAR applied at various image levels
  - Metric → parameter differences



Edge Histogram

- Edge histogram (EHD)
- Captures the spatial distribution of the edge in six statues: 0º, 45º, 90º, 135º, non direction and no edge.
  - Utilizing the filters

  - Global EHD of an image: Concatenating 16 sub EHDs into a 96 bins
  - Local EHD of a segment
    - Grouping the edge histogram of the image-blocks fallen into the segment



Frequency Domain Features

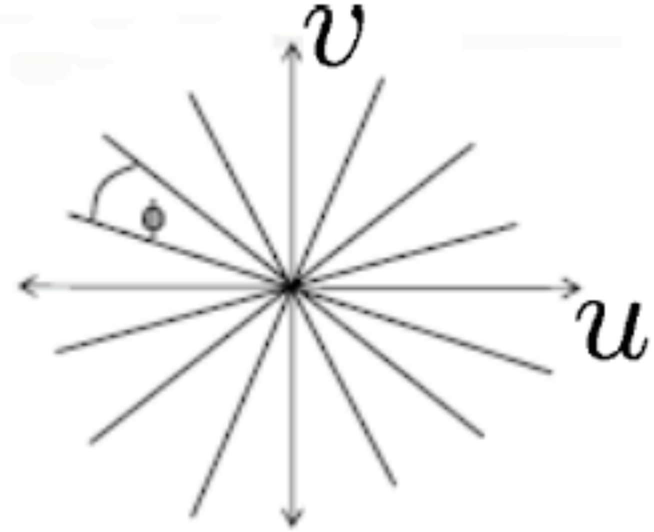
Fourier domain energy distribution

- Angular features (directionality)

$$V^{(a)}_{\theta_1\theta_2} = \int\int |F(u,v)|^2 du dv$$

where,

$$\theta_1 \leq \tan^{-1}[\frac{v}{u}] \leq \theta_2$$

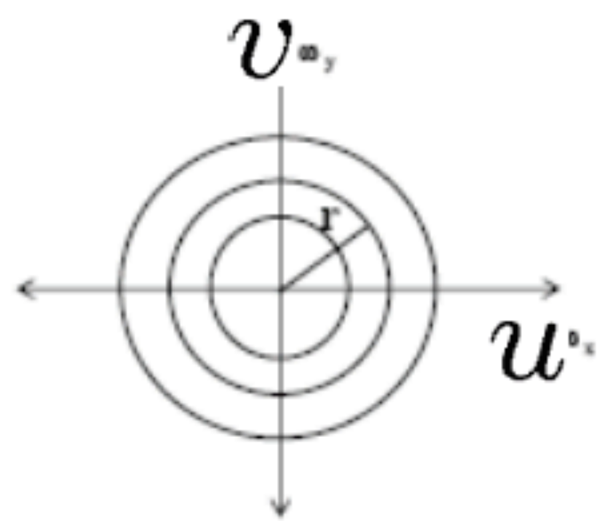


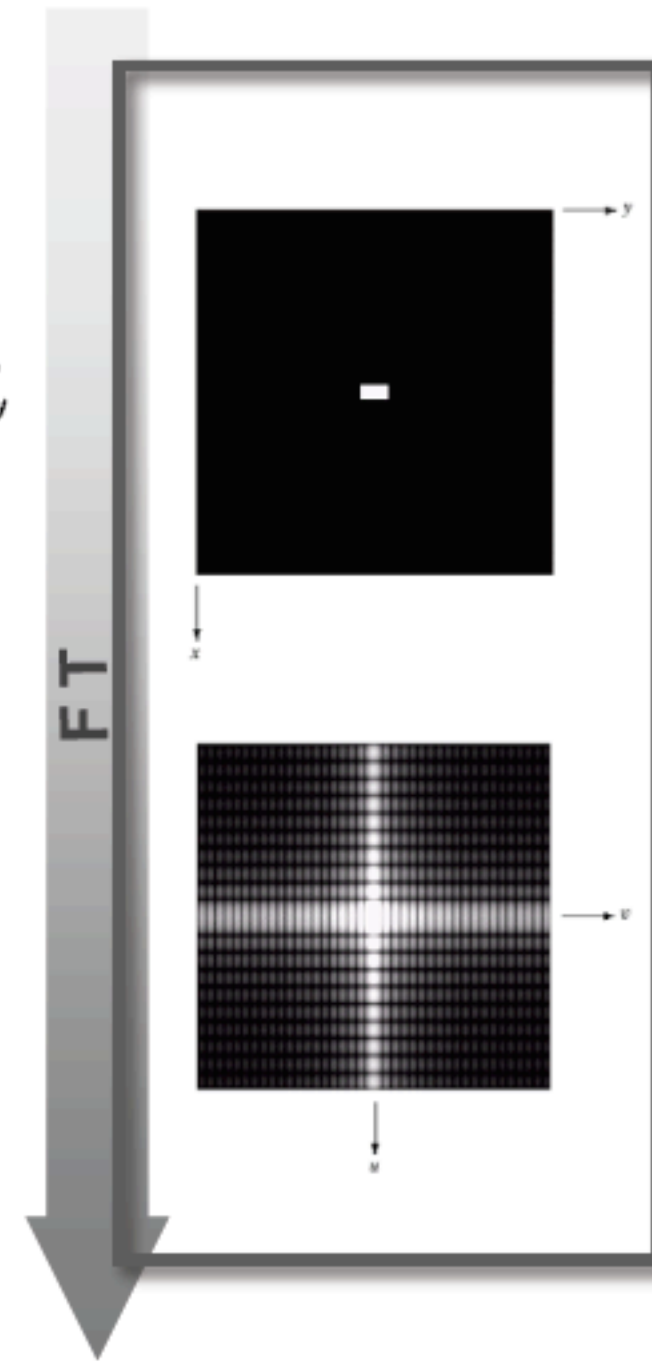
- Radial features (coarseness)

$$V^{(r)}_{r_1 r_2} = \int\int |F(u,v)|^2 du dv$$

where,

$$r_1 \leq u^2 + v^2 < r_2$$

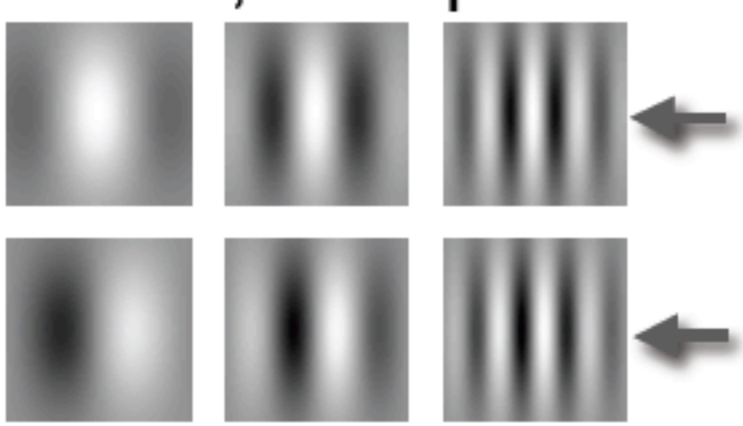




Uniform division may not be the best!!

Gabor Texture

- Fourier coefficients depend on the entire image (Global) → we lose spatial information
- Objective: local spatial frequency analysis
- Gabor kernels: looks like Fourier basis multiplied by a Gaussian
  - The product of a symmetric (even) Gaussian with an oriented sinusoid
  - Gabor filters come in pairs: symmetric and anti-symmetric (odd)
  - Each pair recover symmetric and anti-symmetric components in a particular direction
  - (k_x, k_y): the spatial frequency to which the filter responds strongly
  - σ : the scale of the filter. When σ = infinity, similar to FT
- We need to apply a number of Gabor filters are different scales, orientations, and spatial frequencies

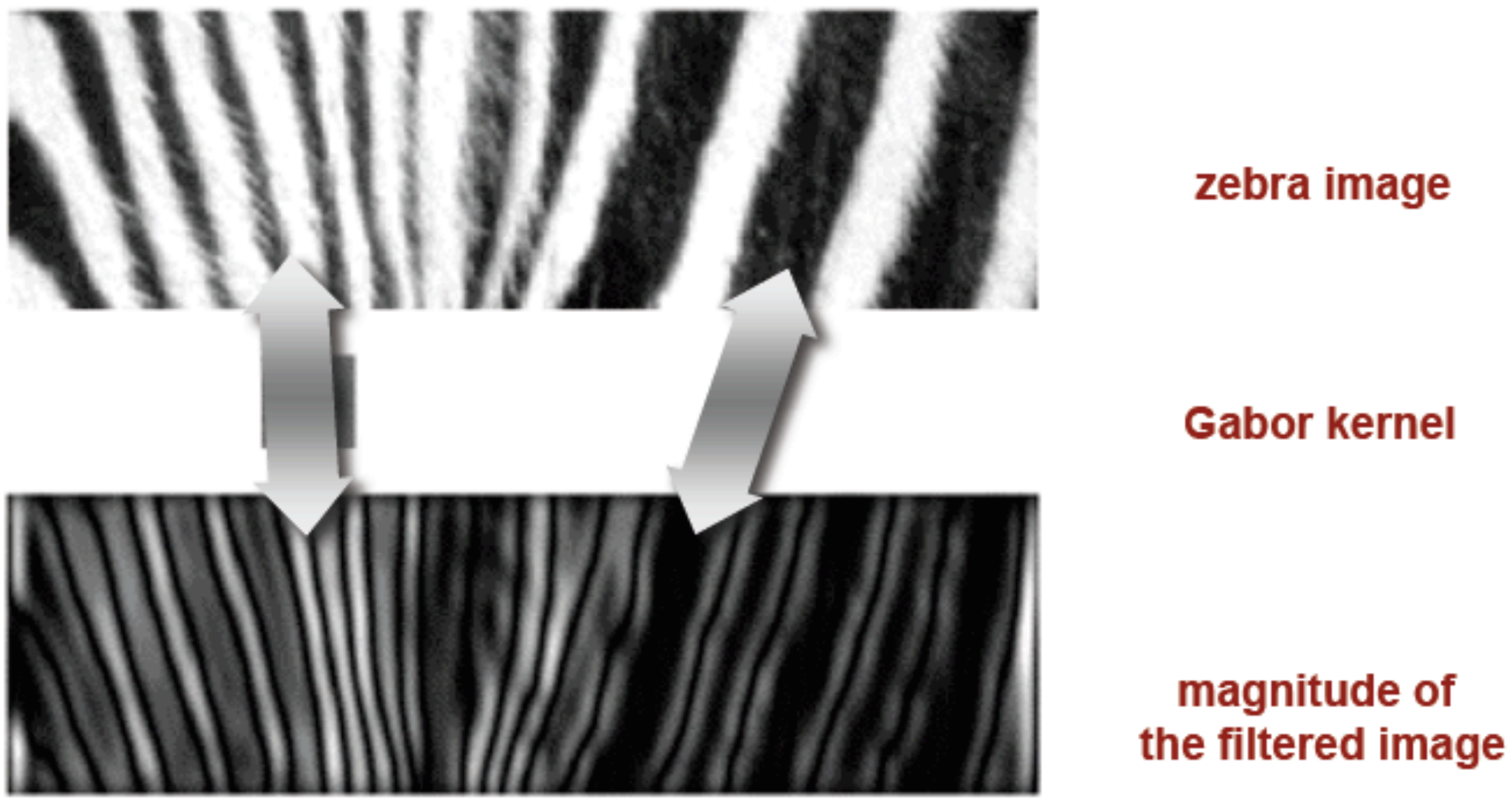


$$G_{symmetric}(x,y) = \cos(k_x x + k_y y) \exp -\frac{x^2+y^2}{2\sigma^2}$$

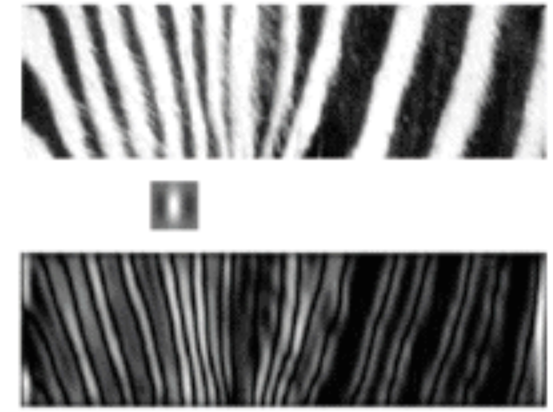
$$G_{anti-symmetric}(x,y) = \sin(k_x x + k_y y) \exp -\frac{x^2+y^2}{2\sigma^2}$$

Example – Gabor Kernel

- Zebra stripes at different scales and orientations and convolved with the Gabor kernel
- The response falls off when the stripes are larger or smaller
- The response is large when the spatial frequency of the bars roughly matches the windowed by the Gaussian in the Gabor kernel
- Local spatial frequency analysis



Gabor Texture (cont.)



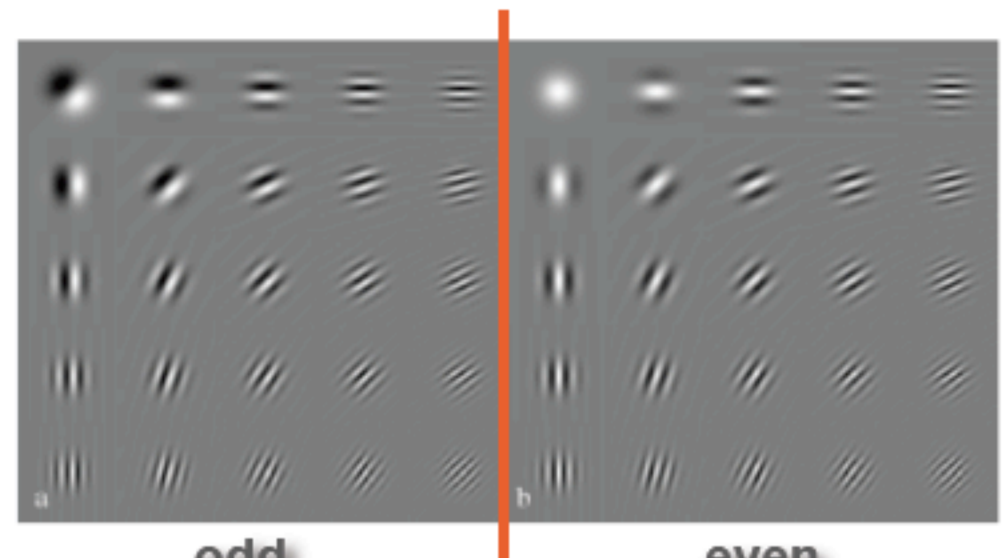
- Image $I(x,y)$ convoluted with Gabor filters h_{mn} (totally M x N)

$$W_{mn}(x,y) = \int I(x_1,y_1)h_{mn}(x-x_1,y-y_1)dx_1dy_1$$

- Using first and 2nd moments for each scale and orientations

$$\mu_{mn} = \int\int |W_{mn}(x,y)|dxdy$$

$$\sigma_{mn} = \sqrt{\int\int (|W_{mn}(x,y)| - \mu_{mn})^2 dxdy}$$



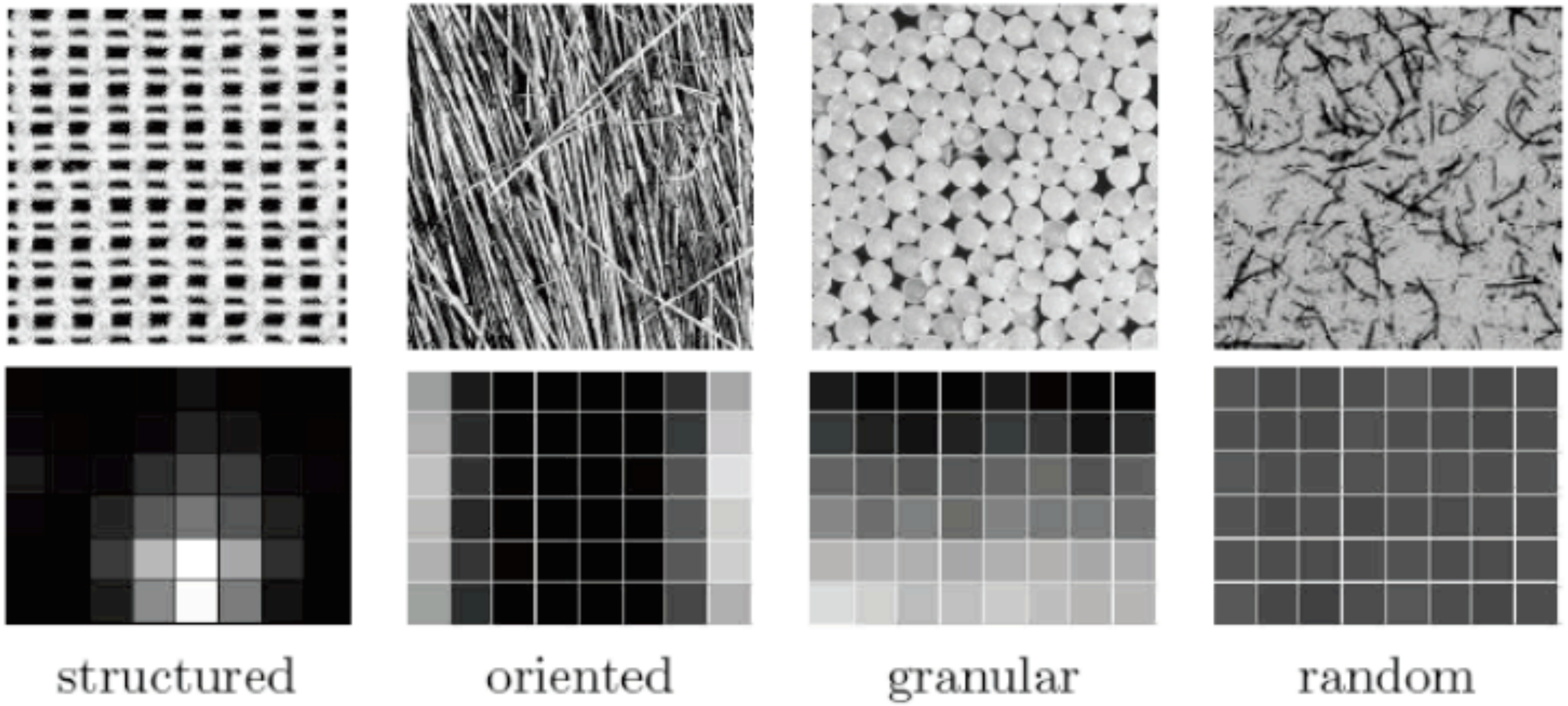
odd even

Gabor kernels

- Features: e.g., 4 scales, 6 orientations → 48 dimensions

$$\bar{v} = [\mu_{00}, \sigma_{00}, \mu_{01}, \ldots, \mu_{35}, \sigma_{35}]$$

Gabor Texture (cont.)



- Arranging the mean energy in a 2D form
  - structured: localized pattern
  - oriented (or directional): column pattern
  - granular: row pattern
  - random: random pattern

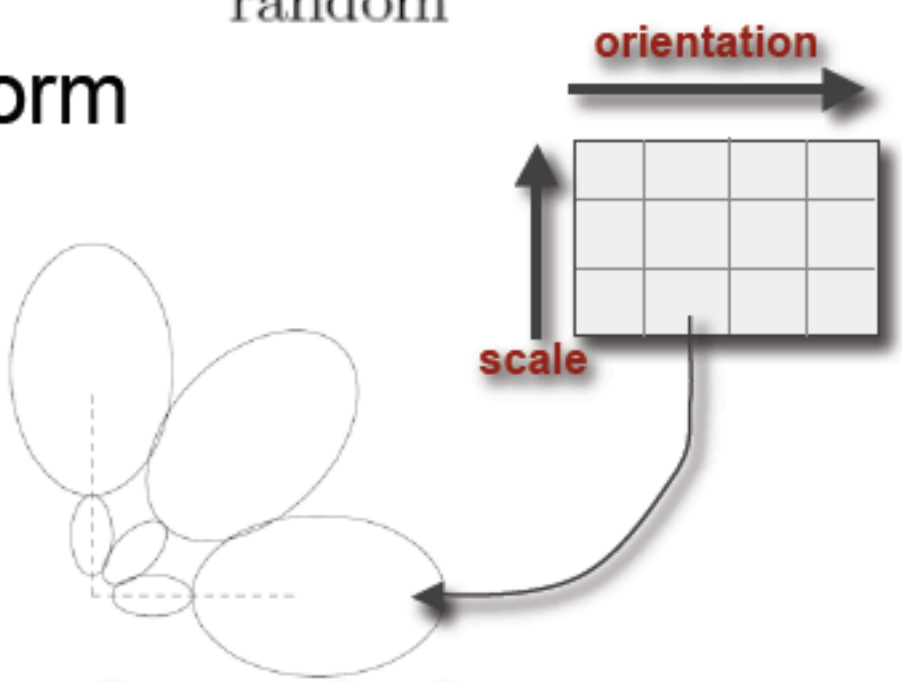


Image shape features

- Shape features are computed out based on object segments or regions, mainly including
  - contour features
  - and regions features.

- Typical approaches include
  - Fourier shape description
  - Moment invariants

Dimensional reduction for image features

In image retrieval system, increasing feature dimension can enhance precision of retrieval greatly. However, high feature dimension will lead to high computation cost. Hence it is important to reduce the redundant in feature data.

- Image feature space reduction
  - Linear dimensional reduction techniques: PCA …
  - Nonlinear dimensional reduction techniques: Isomap, LLE …
  - Clustering based feature reduction methods
- High-dimensional feature indexing
  - Database oriented high-dimensional data indexing
    - Bucketing grouping searching techniques，K-d tree, R tree …
  - Clustering methods
  - SOM

Image similarities

- How to measure similarity of different images base on features?
  - Image features always form into a fixed-length feature vector.
  - The similarity therefore can be measure by
    - Euclidian distance
    - Histogram intersection
    - Quadratic distance
    - Mahalanobis distance (马氏距离)
    - Non-geometrical similarity

Similarity and distance

- Similarity: ↑
  - e.g.

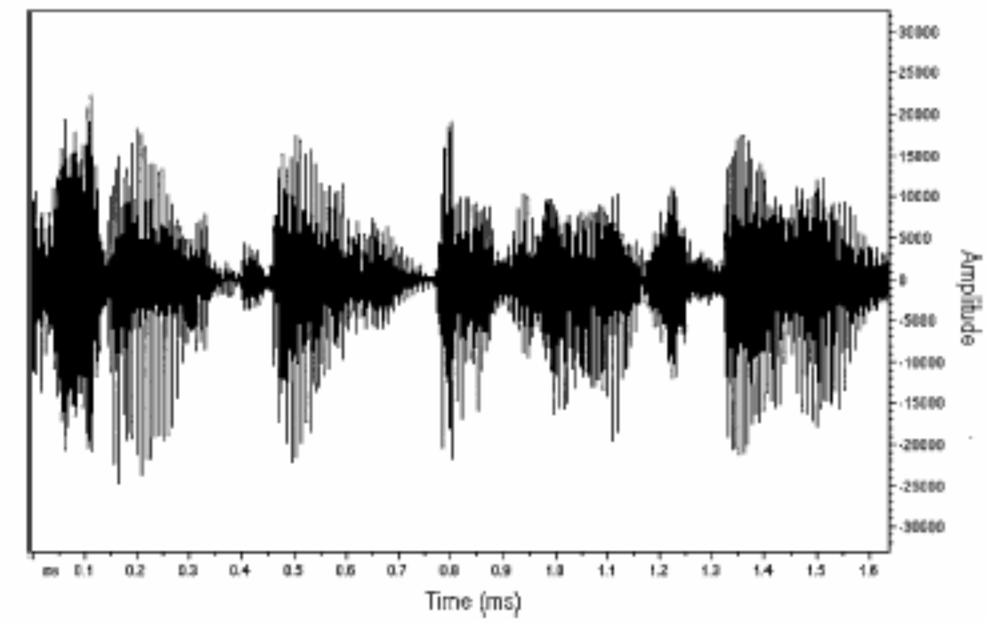
$$S(a,b) = \frac{1}{dist(a,b)+1}$$

- Distance: ↓
  - e.g. Euclidian Distance

$$dist(a,b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 ... + (a_n - b_n)^2}$$

Main Audio Features

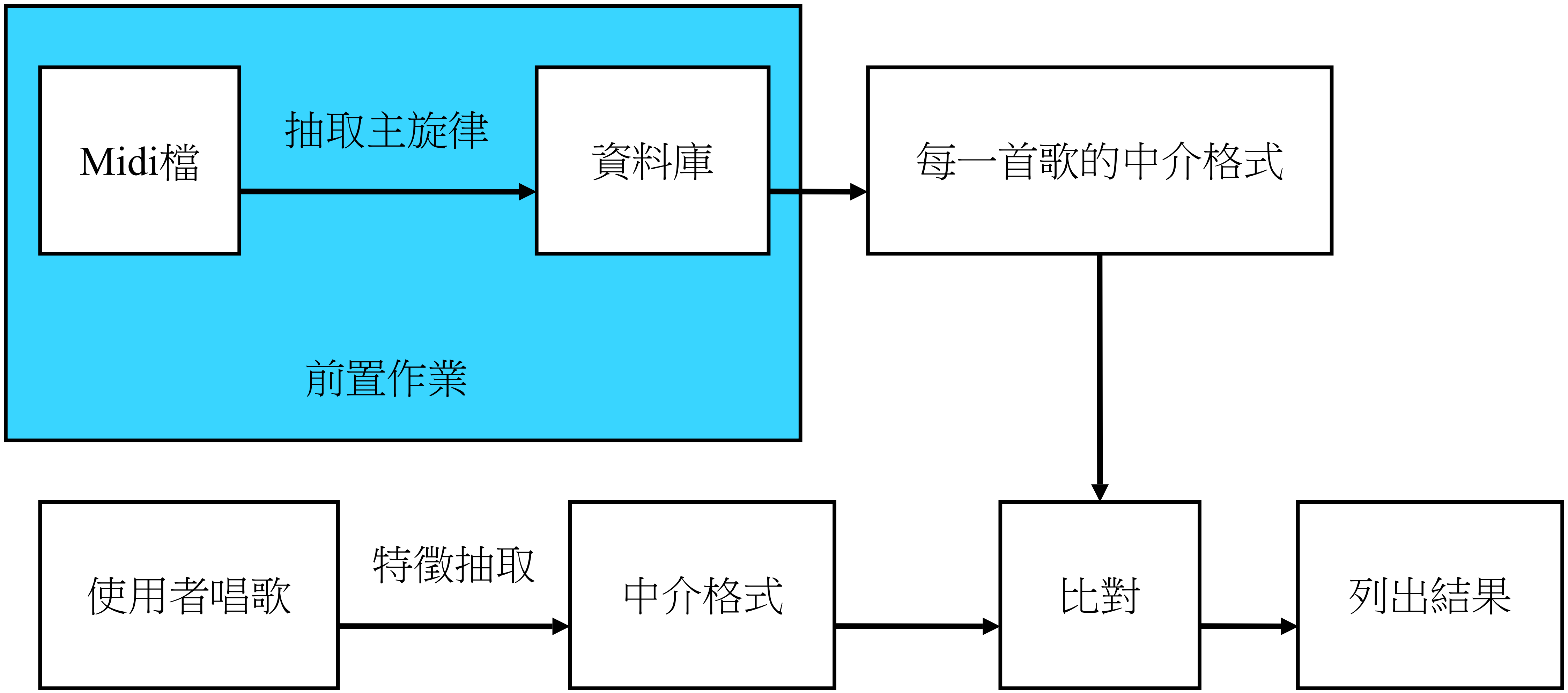
- Time-Domain Features
  - Average Energy
  - Zero Crossing Rate
  - Silence Ratio
- Frequency-Domain Features
  - Sound Spectrum
  - Bandwidth
  - Energy Distribution
  - Harmonicity
  - Pitch
- Spectrogram



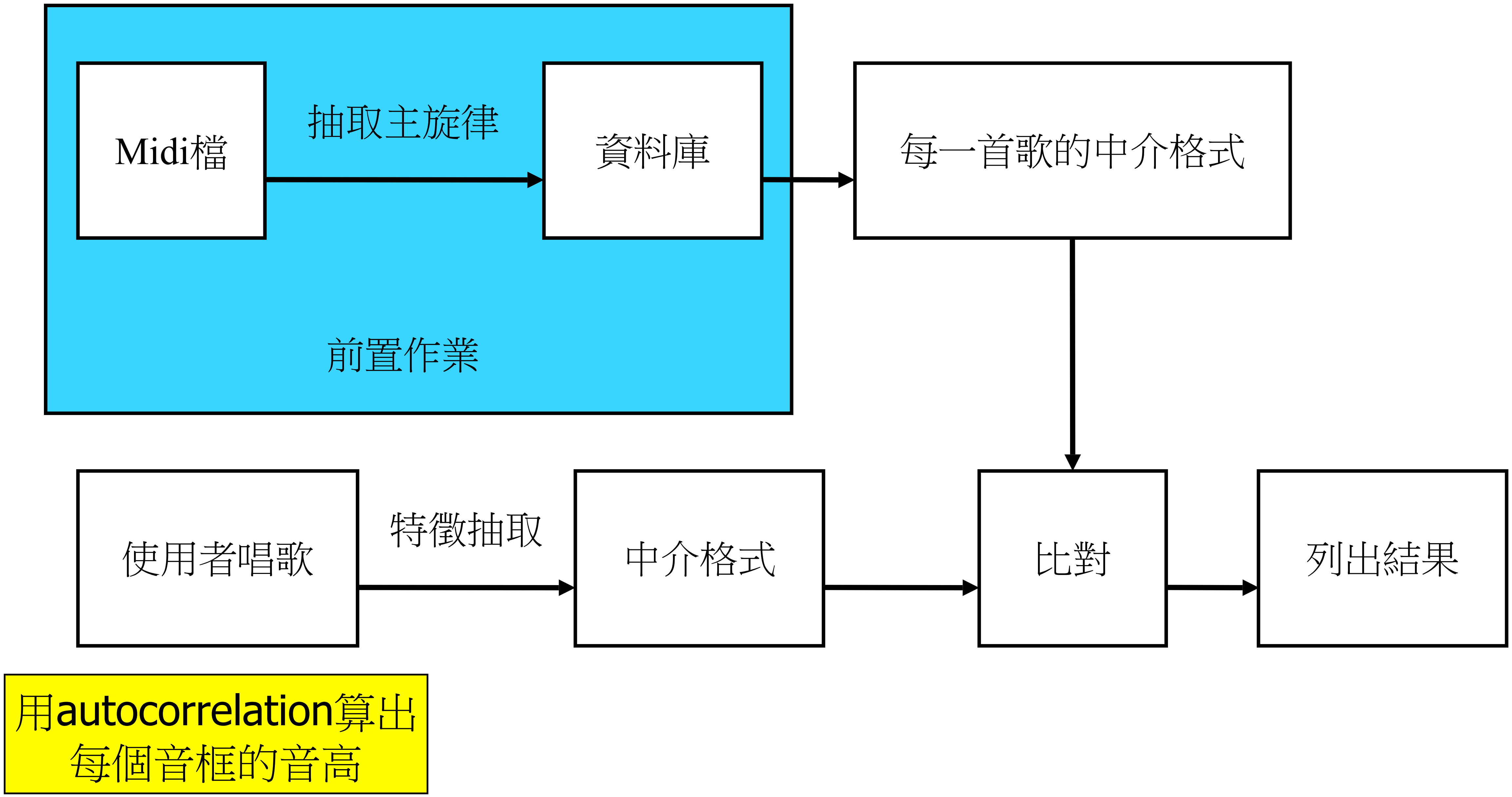
Content based music retrieval

- 說明
  - 用聲音的內容為根據，做音樂的檢索
- 目的
  - 讓使用者可以用自然的方式點選歌曲
- 困難
  - 使用者的節奏（tempo）快慢不同、拍子不準、音調（key）高低不同
  - 若允許使用者從歌的任意處唱，計算量很大

CBMR系統流程圖

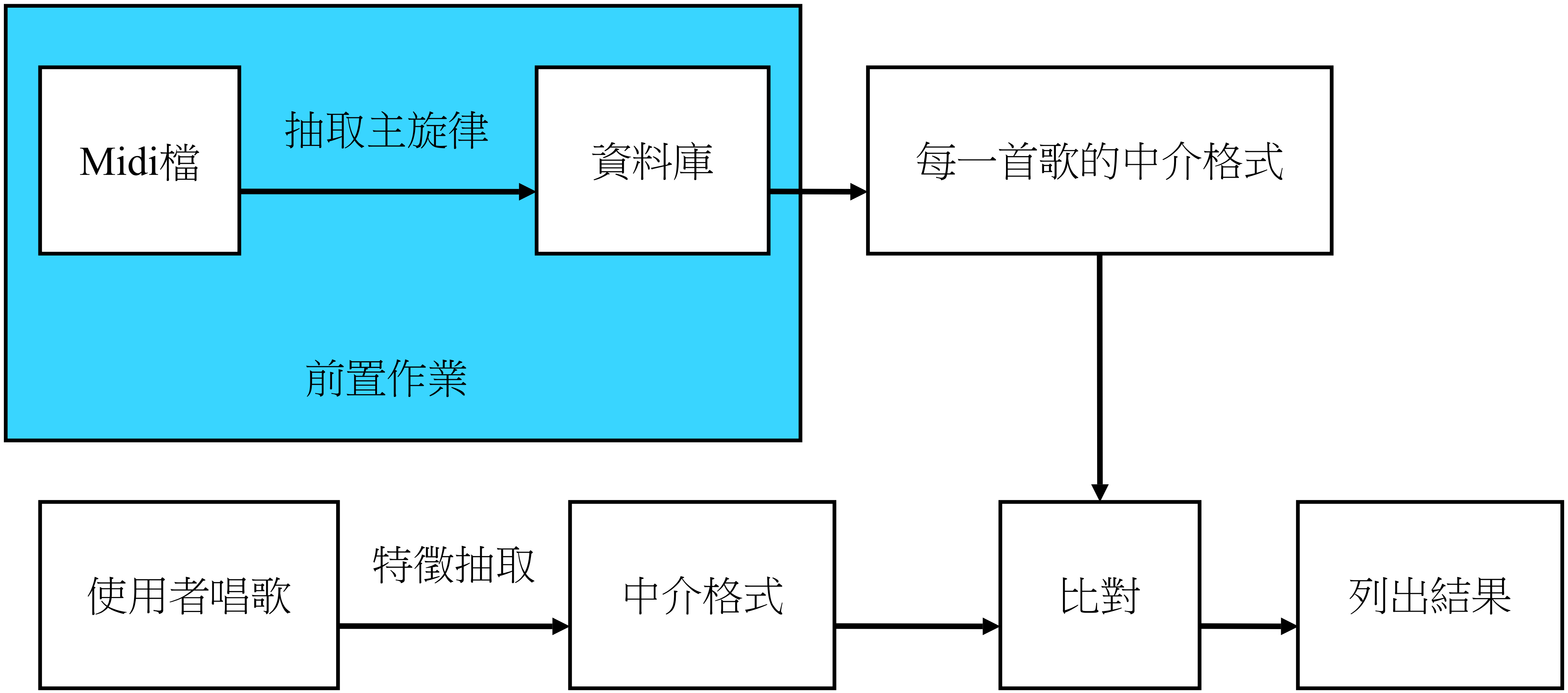


CBMR系統流程圖

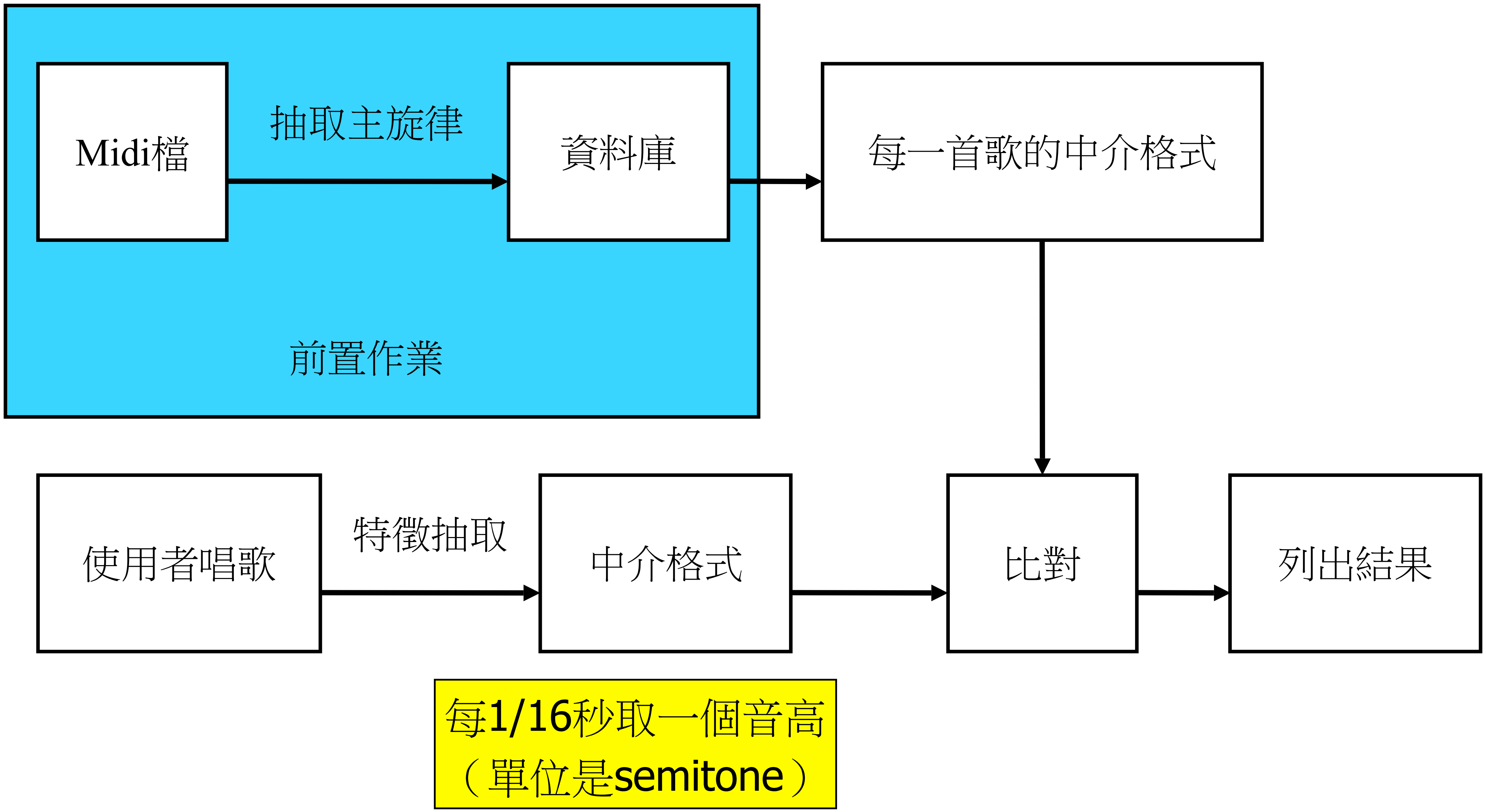


抽取主旋律
Midi檔
資料庫
前置作業
每一首歌的中介格式
特徵抽取
使用者唱歌
中介格式
比對
列出結果
用autocorrelation算出
每個音框的音高

CBMR系統流程圖

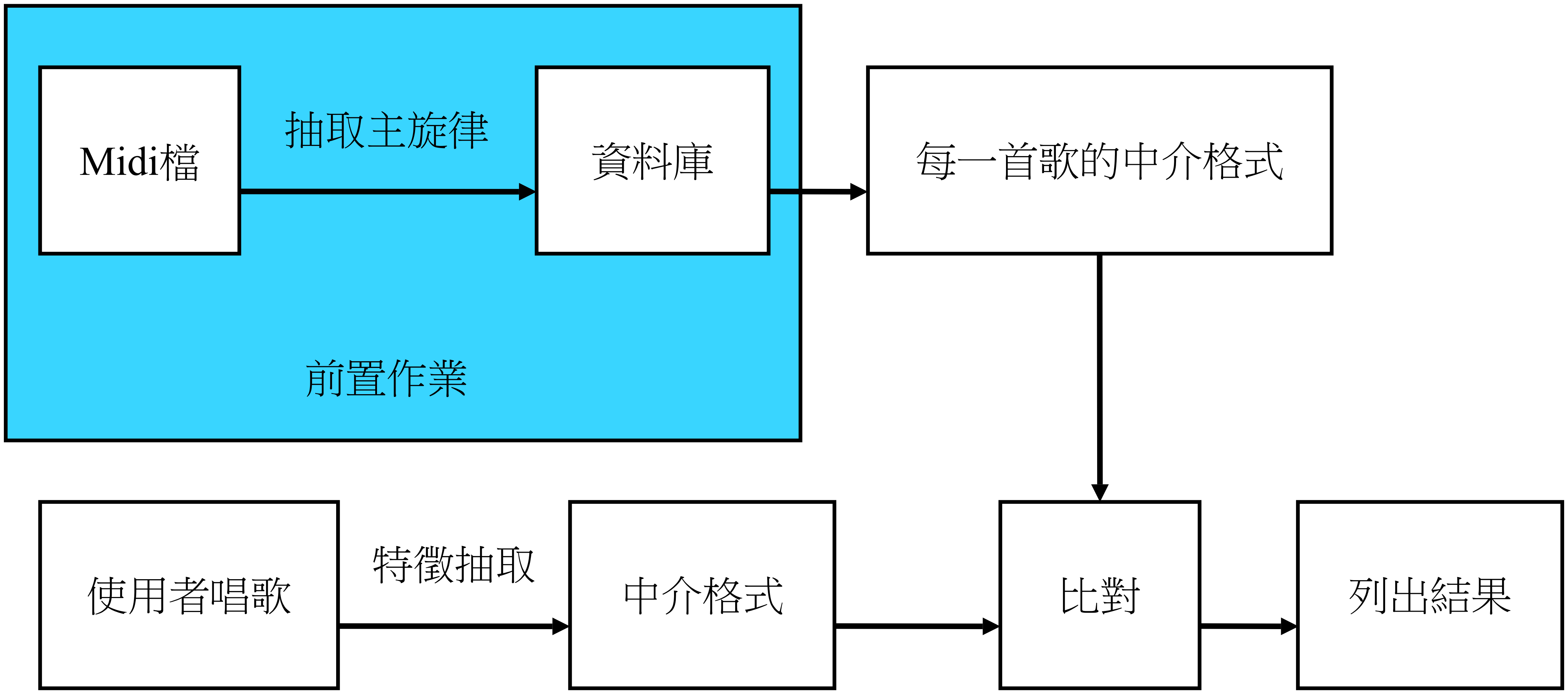


CBMR系統流程圖

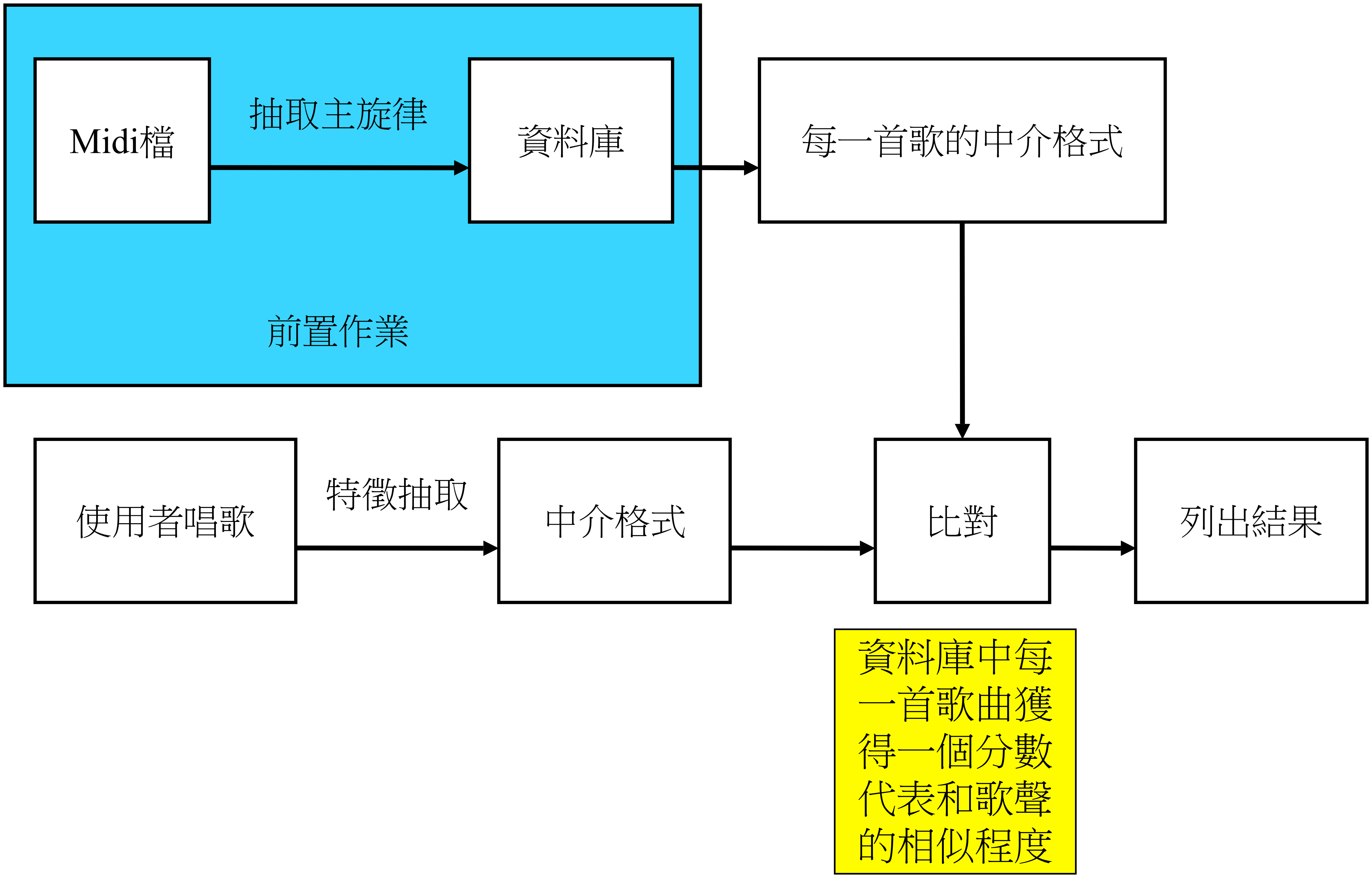


Midi檔
抽取主旋律
資料庫
前置作業
每一首歌的中介格式
使用者唱歌
特徵抽取
中介格式
比對
列出結果
每1/16秒取一個音高
（單位是semitone）

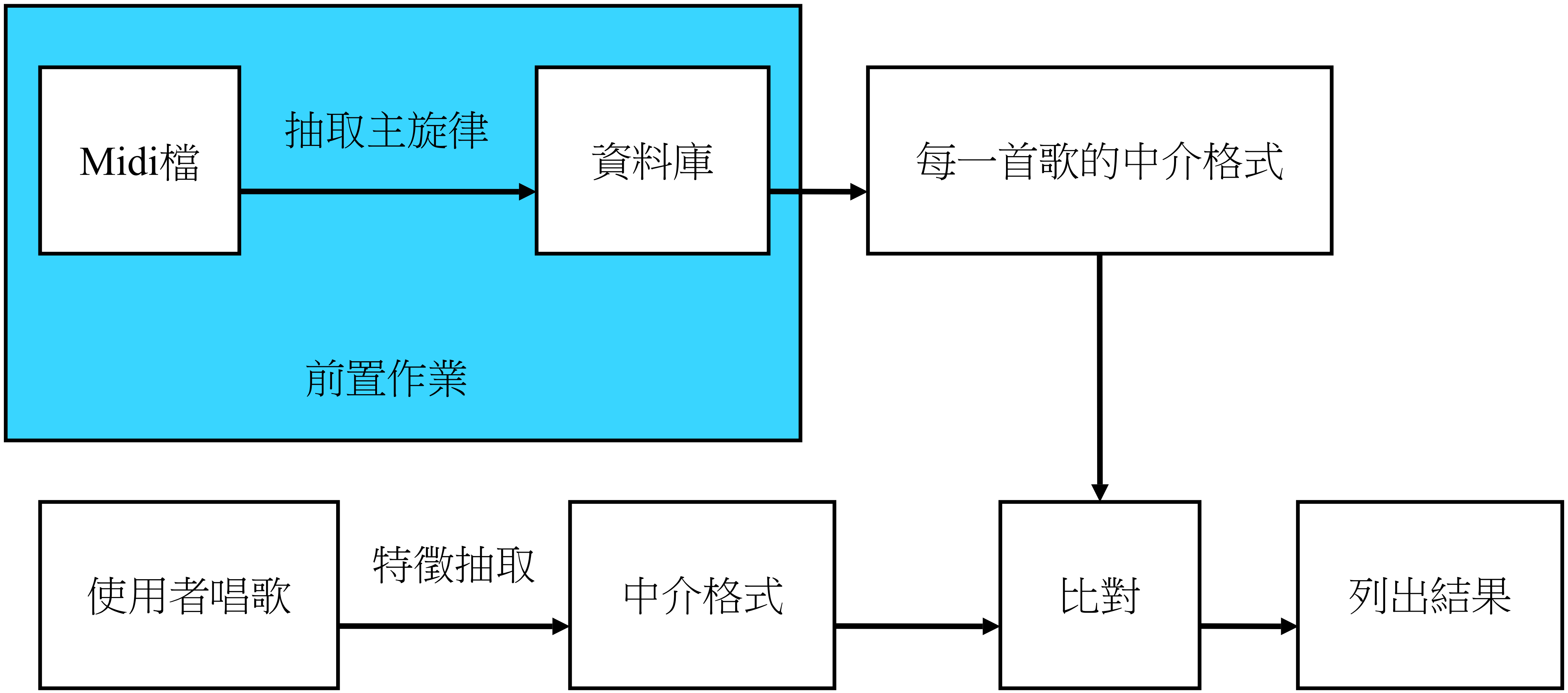
CBMR系統流程圖



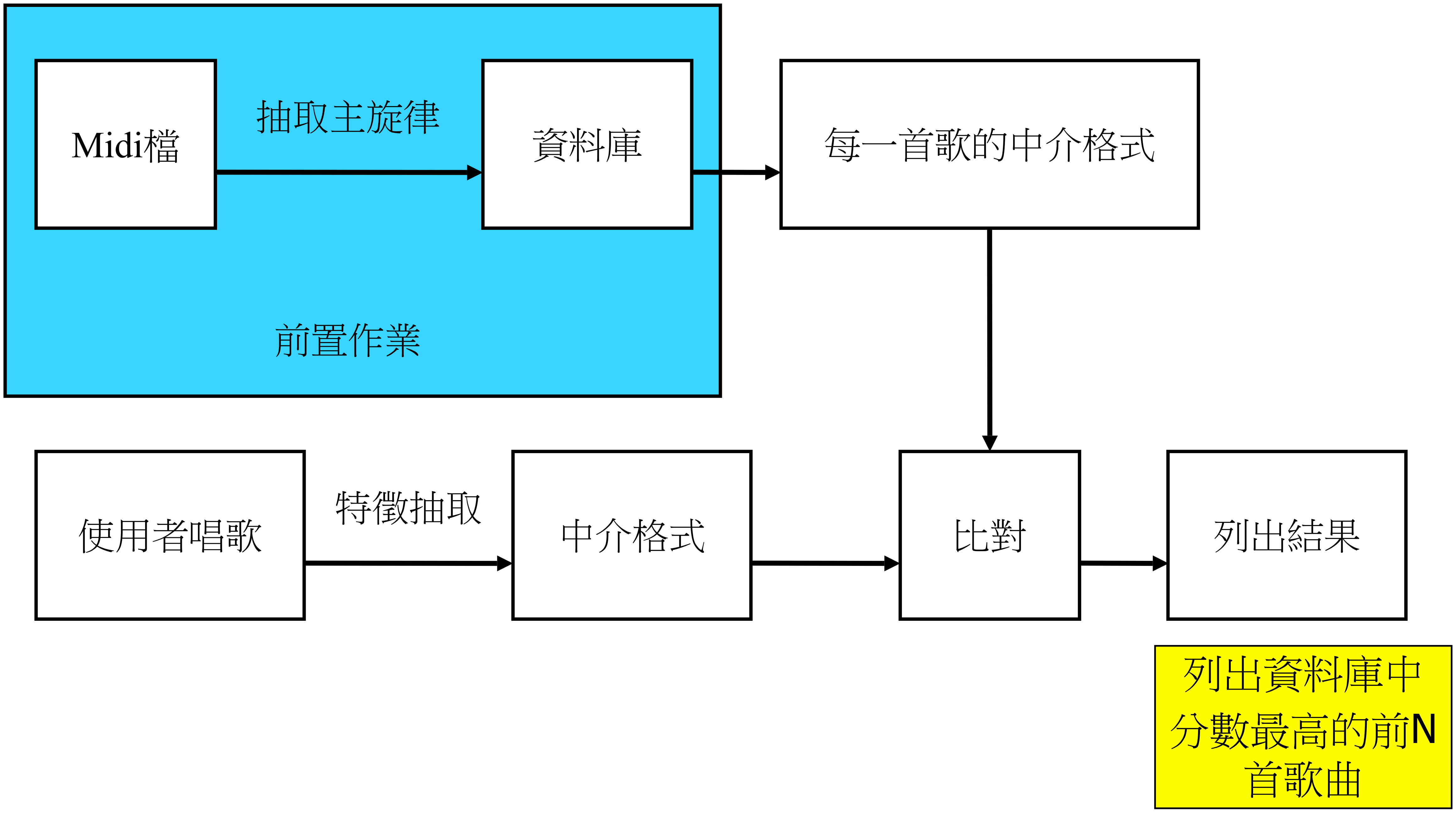
CBMR系統流程圖


Midi檔
抽取主旋律
資料庫
每一首歌的中介格式
前置作業
使用者唱歌
特徵抽取
中介格式
比對
列出結果
資料庫中每一首歌曲獲得一個分數代表和歌聲的相似程度

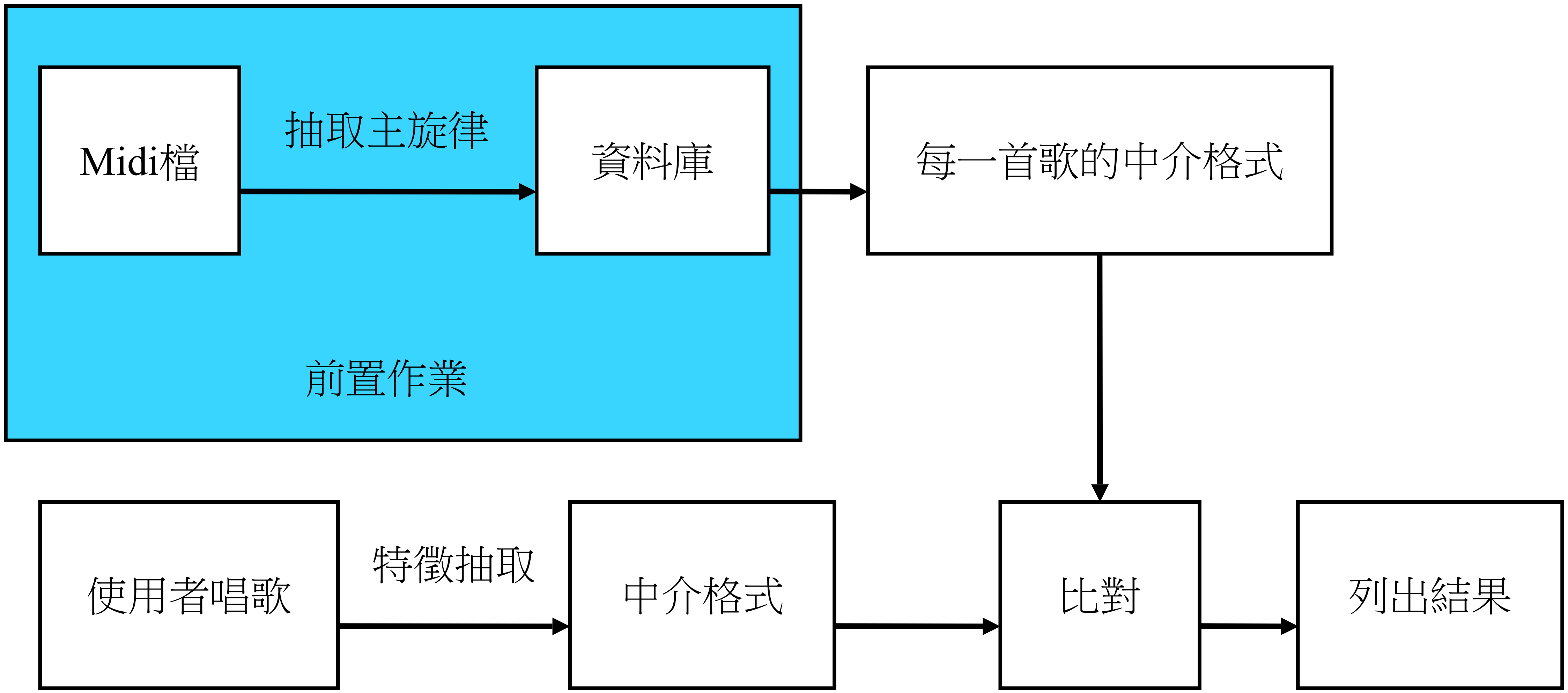
CBMR系統流程圖



CBMR系統流程圖


Midi檔
抽取主旋律
資料庫
每一首歌的中介格式
前置作業
使用者唱歌
特徵抽取
中介格式
比對
列出結果
列出資料庫中
分數最高的前N
首歌曲

CBMR系統流程圖

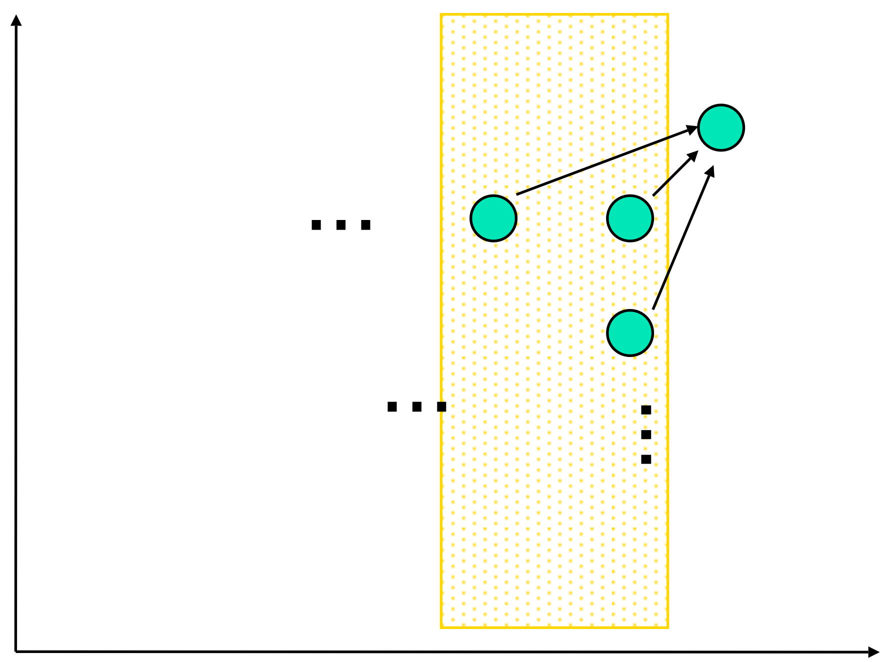


前人的方法

- 克服節奏快慢不同的問題
  - Dynamic time warping
- 克服音調高低不同的問題
  - Key transposition
    - 將使用者的歌聲和資料庫中的歌轉換後的中介格式的平均值都平移到0，做一次dtw比對
    - 將資料庫中的歌的中介格式上下平移再做四次dtw比對，以找出最短的距離
- 全曲比對費時很久且準確率低
- 使用浮點數運算

改進方法一

- 改良式dtw
  - 用dtw計算每一首歌的距離時，若發現dtw table最近兩行(column)的最小值超過之前最短距離的前k名，則停止dtw



改進方法二

- 將資料庫中每一首歌的中介格式，從每一個音符為起點切成數個長度為 D=72 的片段

中介格式 (<u>69</u>, 69, <u>67</u>, 67, 67, <u>71</u>, <u>72</u>, ……)

→ 片段1 (69, 69, 67, 67, 67, 71, 72, …)

片段2 (67, 67, 67, 71, 72, …)

片段3 (71, 72, …)

……

改進方法二

- 用兩階段的方法比對
  - 第一階段：線性伸縮比對（linear scaling）
    - 將歌聲的中介格式伸縮11次（長度為原來的0.75倍到1.25倍），分別取出前 D 點後為 T_i（$1 \leq i \leq 11$），假設資料庫中的第 j 首歌的中介格式有 n_j 個片段為 R_{jk}（1≦j≦資料庫中的歌曲數目，$1 \leq k \leq n_j$），令 T_i 和 R_{jk} 的距離為

$$dist(T_i, R_{jk}) = \sqrt{\sum_{t=1}^{D} \left(\left(T_{it} - \overline{T_i} \right) - \left(R_{jkt} - \overline{R_{jk}} \right) \right)^2}$$

改進方法二

- 令資料庫第 j 首歌的分數為 $100-\min_{\substack{1\le i\le 11\\ 1\le k\le n_j}}\frac{dist(T_i,R_{jk})}{D}$，令

$$k' = \arg\min_{\substack{1\le i\le 11\\ 1\le k\le n_j}} dist(T_i,R_{jk})$$

- 篩選出前 n=200 首分數最高者做第二階段的比對
- 缺點：每和資料庫中第 j 首歌做比對就要計算n_j*11次距離

- 第二階段：dynamic time warping
  - 將篩選後的每一首歌的最接近片段平移音調4次，總共和歌聲原本的中介格式計算5次距離，找出最小值並轉換成分數，當成資料庫該首歌最後的分數

針對全曲比對的加速方法

- 用兩階段式比對
- 在第一階段將資料庫歌曲的每一個片段看成一個個 D 度空間中的資料點，歌聲伸縮後的中介格式則是空間中的查詢點，利用快速找最近鄰居的方法找出每首歌最接近的片段
  - Vantage-point tree
  - Branch-and-bound tree
  - Equal-average hyperplane partitioning method

實驗結果

- 實驗環境
  - 電腦 PⅢ 800，256MB RAM，Windows 2000
  - 資料庫
    - 8552首，包括中文、台語、英文和日文歌曲
  - 測試歌聲wav檔
    - 從頭唱 1054 首
    - 從任意處唱 1650 首
    - 每一首長 8 秒鐘

錯誤分析

- 資料庫錯誤
  - 有前奏
  - 同一時間不只一個音符
  - 主旋律錯誤
  - 相同歌曲但不同歌名
- 測試歌聲不良
  - 有雜音
  - 氣不足、音不準
  - 唱錯
  - 拍子不準

Content based video retrieval

- CBVR can be done by:
  - Color
  - Texture
  - Shape
  - Spatial relationship
  - Semantic primitives
  - Browsing
  - Objective Attribute
  - Subjective Attribute
  - Motion
  - Text & domain concepts
- CBVR has two Approaches:
  - Attribute based
  - Object based

Differences and relations between image and video

- Images are **static**, but video are **dynamic**.
- Video stream can be viewed as sequence of image frames.

Content based video retreival

- CBVR has two phases:
  - –Database Population phase
    - Video shot boundary detection
    - Key Frames selection
    - Feature extraction
  - –Video Retrieval phase
    - Similarity measure

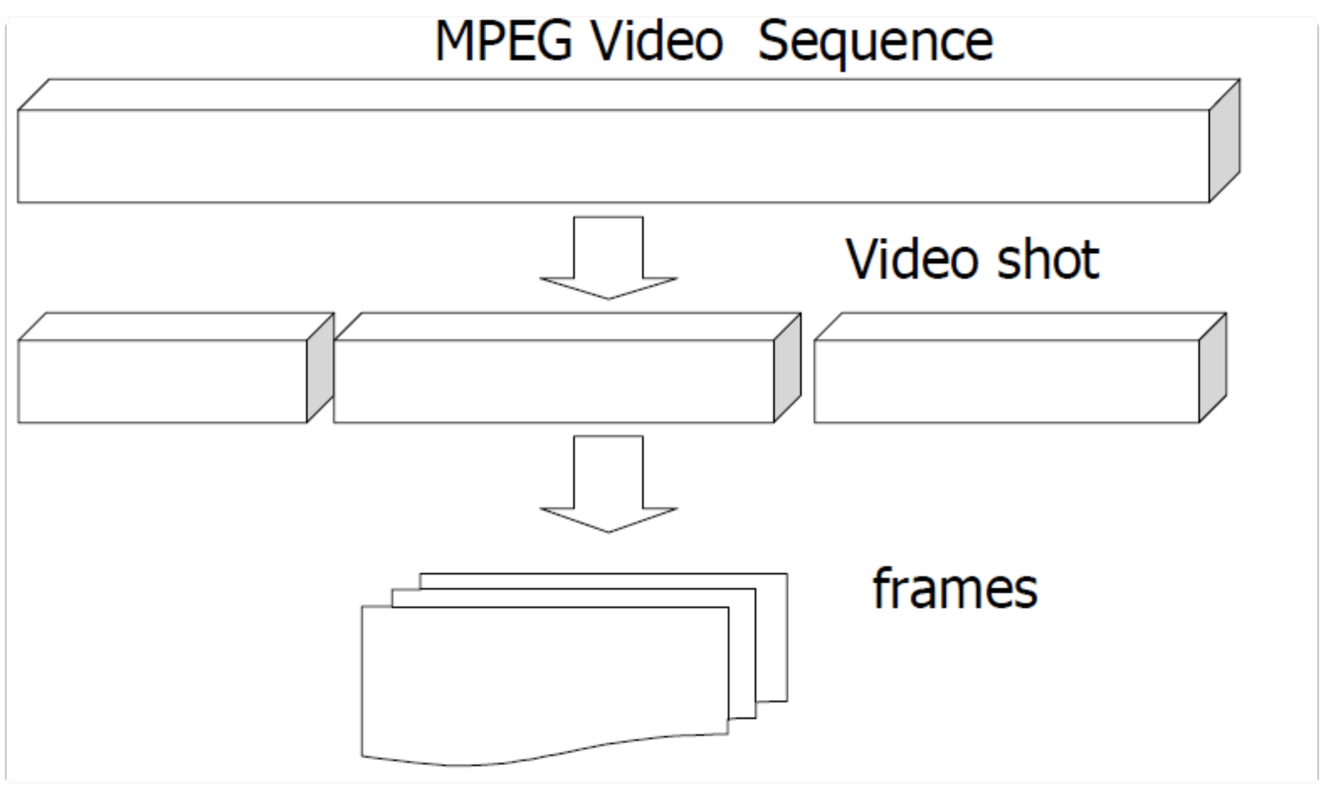
Database Population Phase

- Closest terms in the database is returned based on the similarity measure between the query images and the resulting ones.
- Cosine similarity measure is used in the vector space model.
- Cosine similarity measure on Term-by-video matrix:

$$\cos(t_1, v) = \frac{\sum_{h=1}^{k} t_1 v}{\sqrt{\sum_{h=1}^{k} t_1^{\ 2} \sum_{h=1}^{k} v^2}}$$

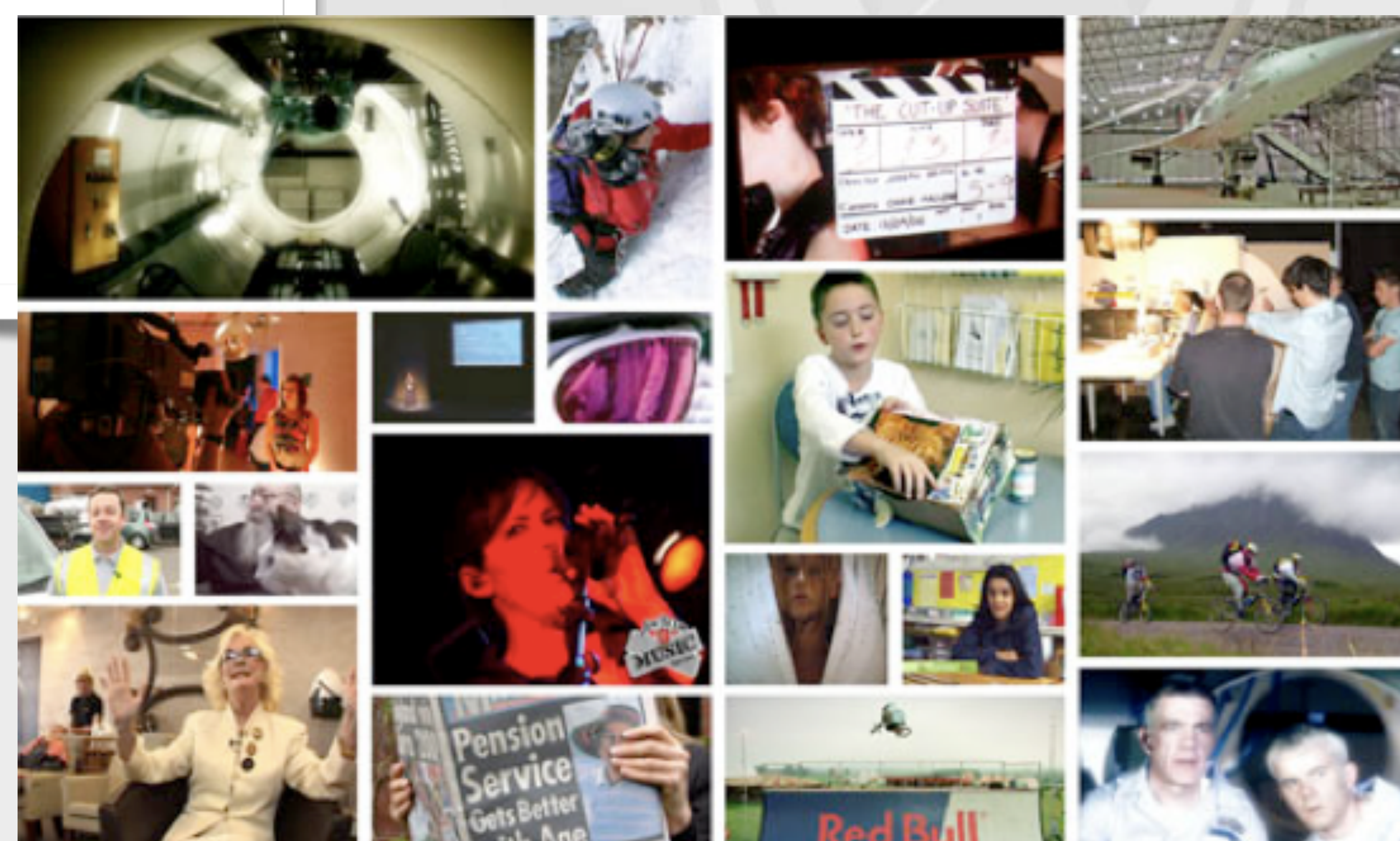
Structuralizing video data

- **semantic content layers**, e.g., scenes and shots in a video program.
  - These layers are erased when they are displayed for audience, which weakens the ability for user dealing with raw video data.



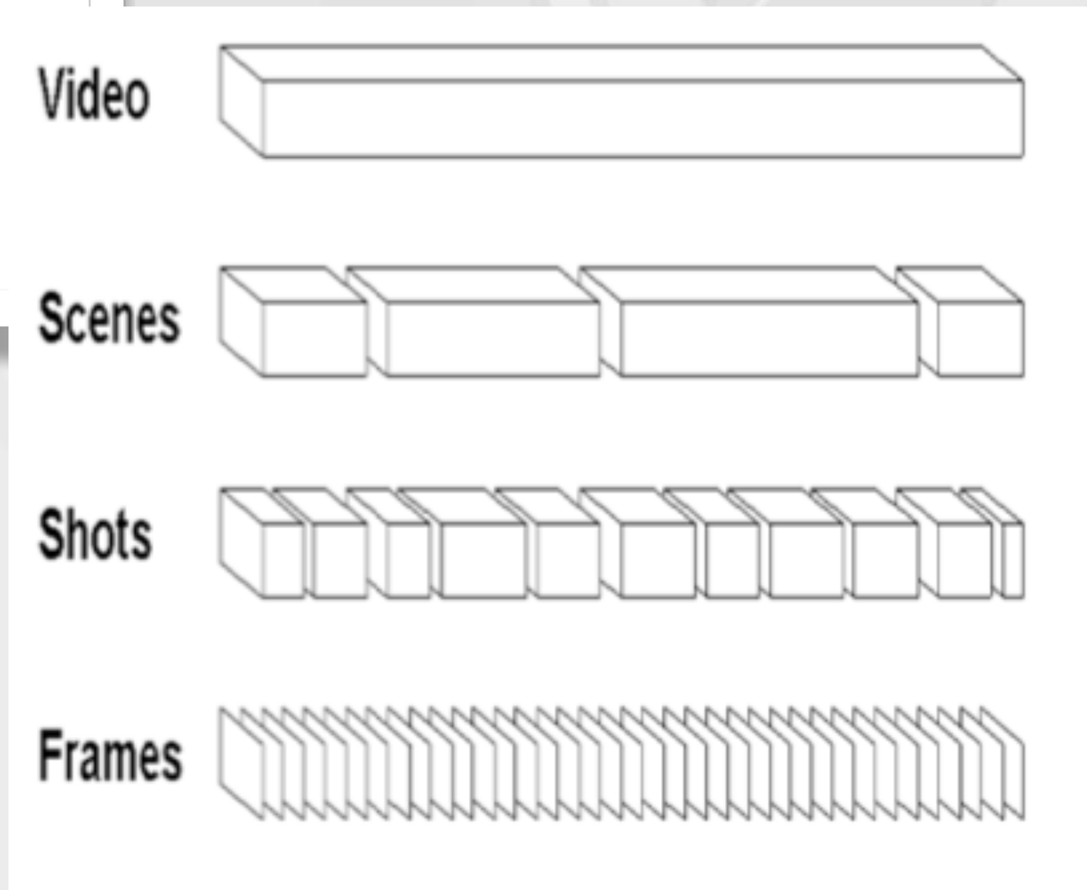
Fundamental definitions in video structurization

- Frame (帧)
- Shot (镜头)
- Key frame (关键帧)
- Scene (场景)
- Group (组)



Fundamental definitions in video structurization

- Frame (帧)
- Shot (镜头)
- Key frame (关键帧)
- Scene (场景)
- Group (组)



Basic video segmentation metrics

- Pair-wise comparison
  - Pixel-level
    - Sensitive to camera movement and motion
  - Block-level (Likelihood ratio)
    - Can tolerate small motion

$$DP_i(k,l) = \begin{cases} 1 & \text{if } |P_i(k,l) - P_{i+1}(k,l)| > t \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\sum_{k,l=1}^{M,N} DP_i(k,l)}{M * N} * 100 > T$$

$$\frac{\left[\frac{S_i+S_{i+1}}{2} + \left(\frac{m_i-m_{i+1}}{2}\right)^2\right]^2}{S_i * S_{i+1}} > t$$

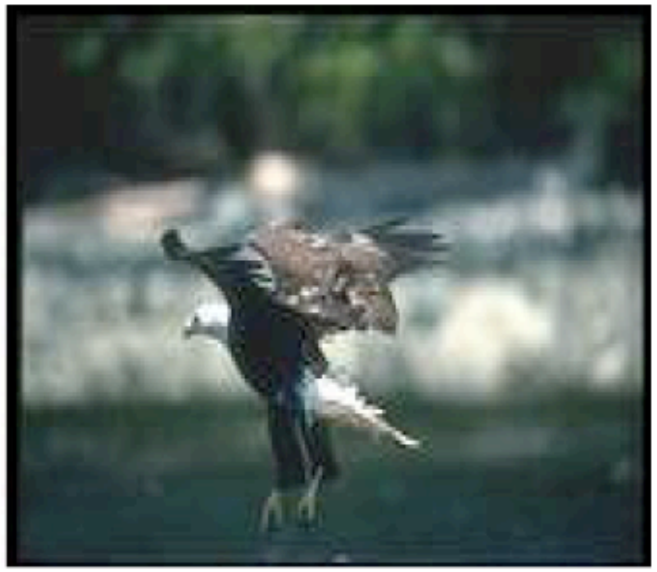
mi: mean intensity

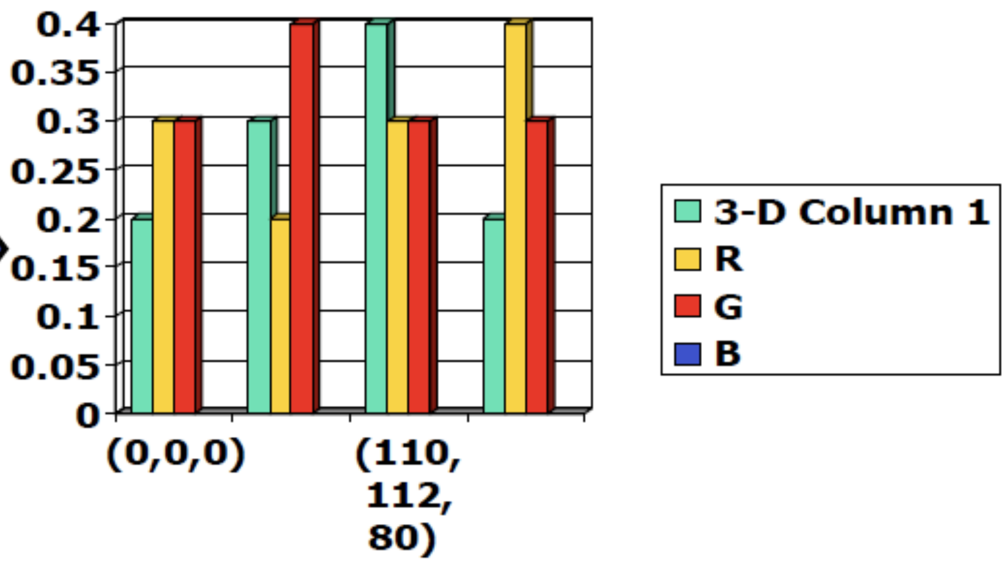
Si: corresponding variance

Basic video segmentation metrics

How to measure statistical property of video frames?

Color Histogram





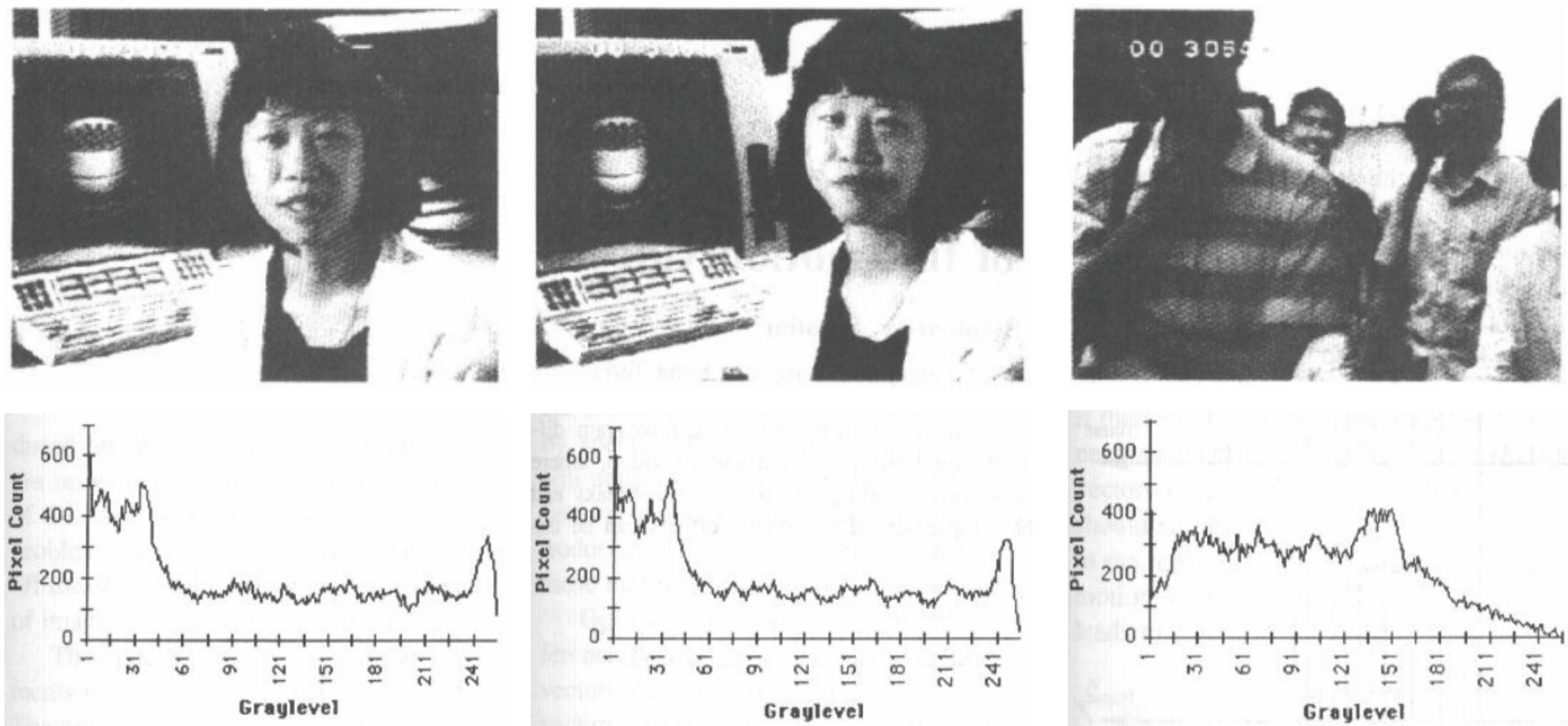
Basic video segmentation metrics

- Histogram comparison
  - Basic
    - Tolerate motion better
  - χ2-test
  - Color level can also be used but only the MSB to save the number of bins

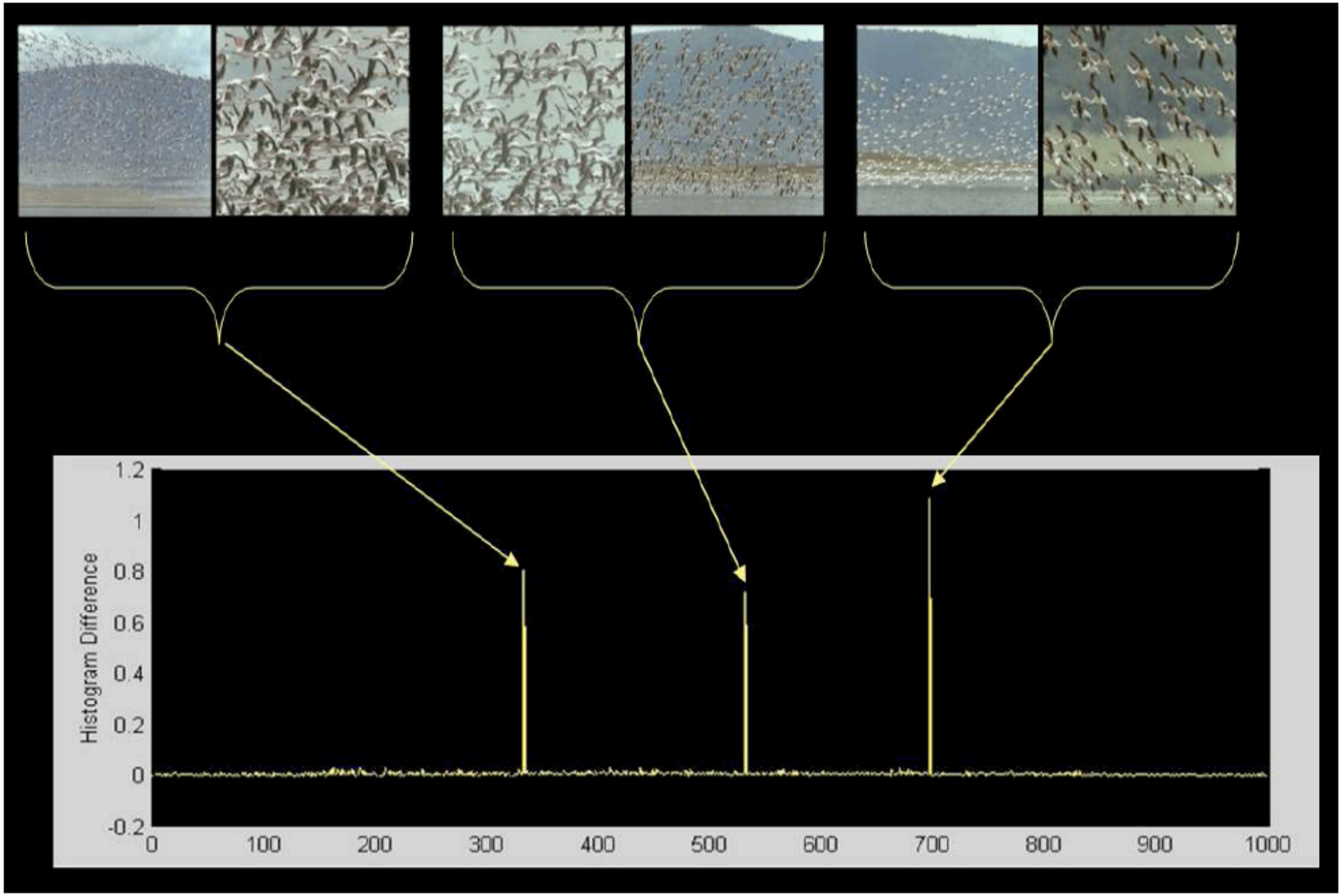
$$SD_i = \sum_{j=1}^{G} |H_i(j) - H_{i+1}(j)|$$

$$SD_i = \sum_{j=1}^{G} \frac{|H_i(j) - H_{i+1}(j)|^2}{H_{i+1}(j)}$$

Sample of using histogram

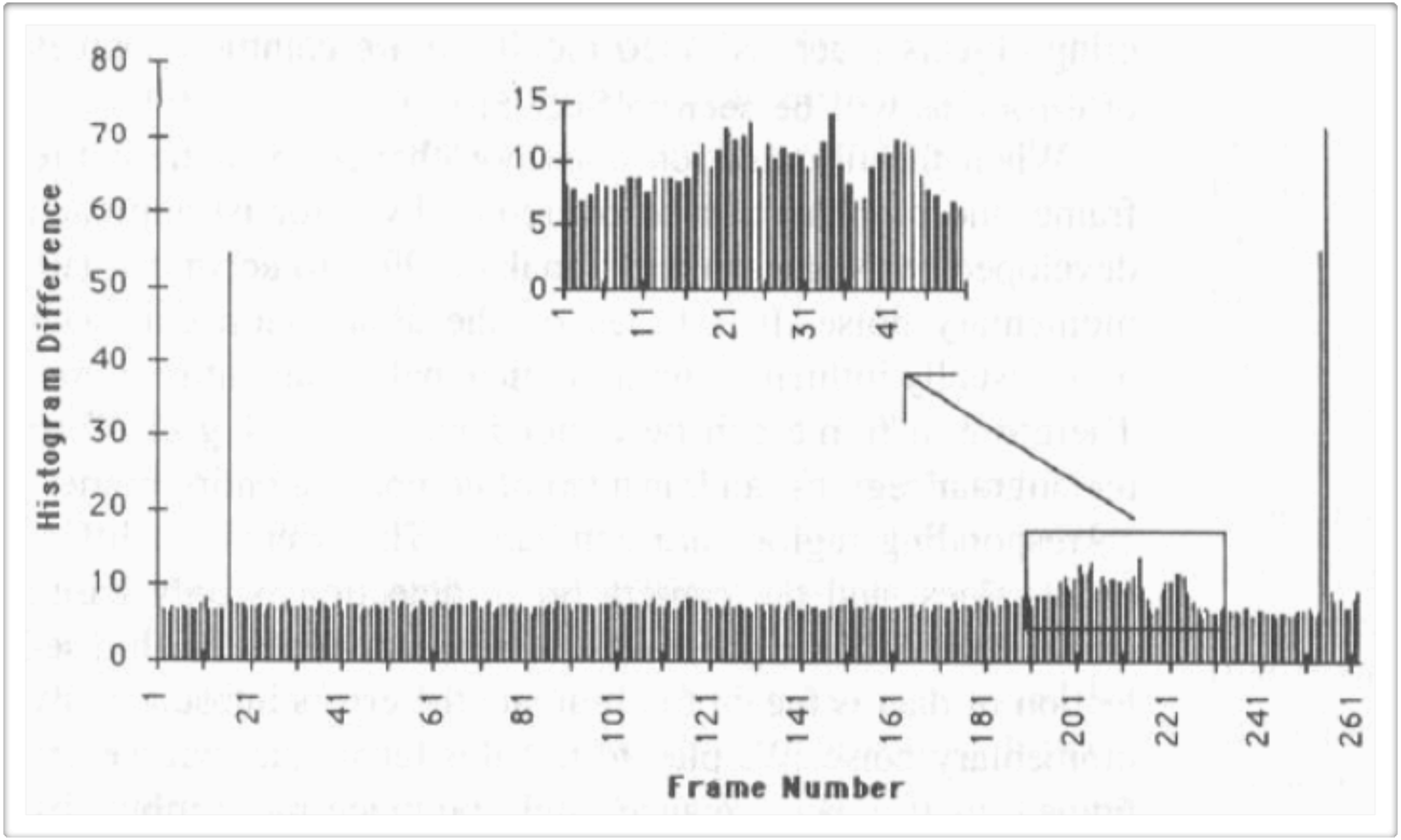


Scene Cut



Histogram Difference
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
0
100
200
300
400
500
600
700
800
900
1000

Gradual transition detection



Gradual transition detection

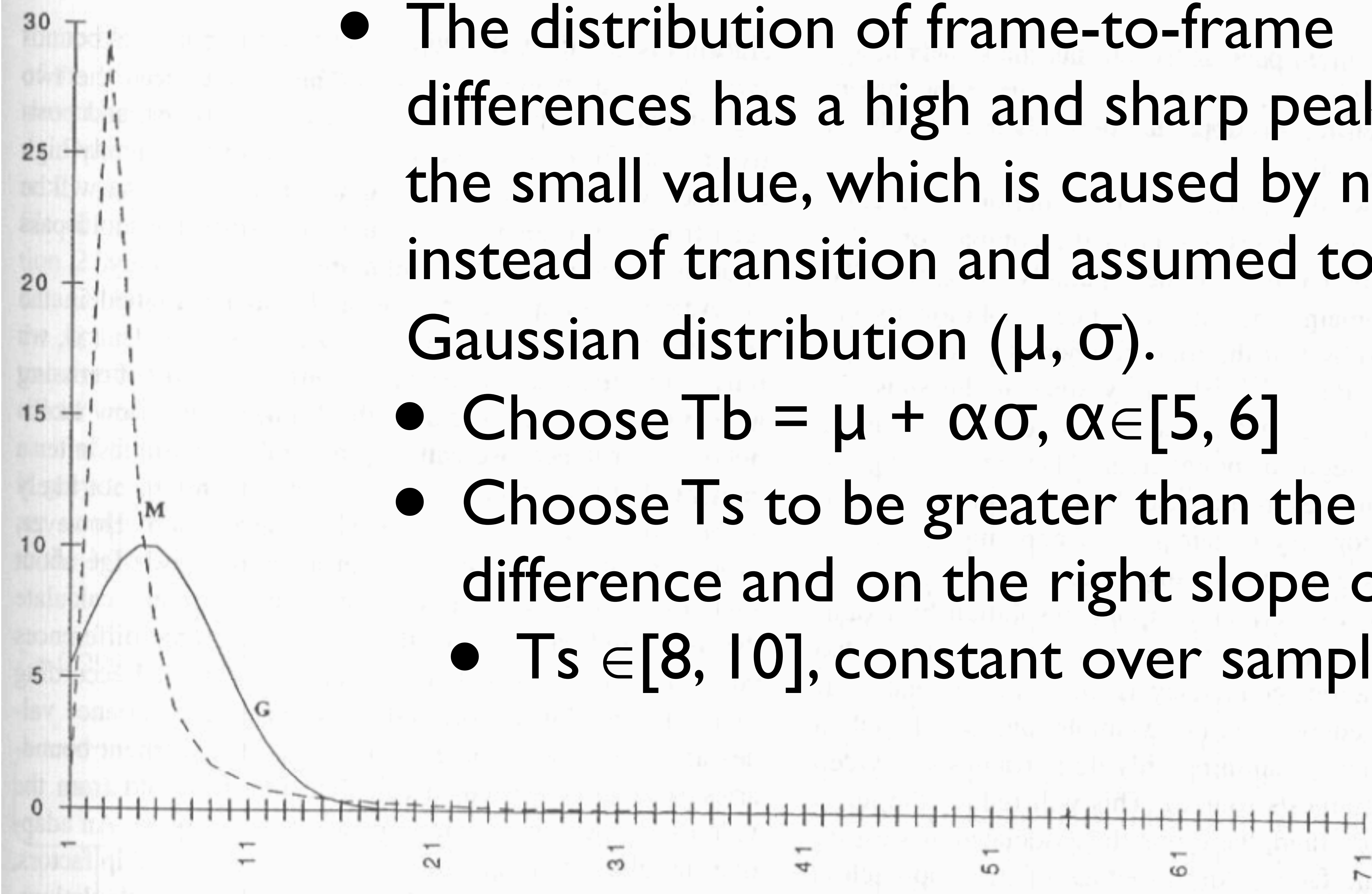
- Twin-comparison
  - Use two thresholds Tb and Ts to accommodate both **short-term** and **long-term** transitions
  - Differences of (F1, F2), (F2, F3), (F3, F4) are small, but difference of (F1, F4) is still big

1 2 3 4

- Twin-comparison
  - F_s — the potential beginning frame of the transition
  - F_e — the ending frame of the transition

scan frame
if ($Diff(F_i) \geq T_b$)
    detect as camera break
else if ($T_b > Diff(F_i) \geq T_s$)
    $F_s \leftarrow F_i$
    $i \leftarrow i + 1$
    while ($Diff(F_i) \geq T_s$)
        $i \leftarrow i + 1$
    if ($Diff(F_s, F_i) \geq T_b$)
        $F_e \leftarrow F_i$

Threshold selection (Tb, Ts)



- The distribution of frame-to-frame differences has a high and sharp peak near the small value, which is caused by noise instead of transition and assumed to follow Gaussian distribution (μ, σ).
  - Choose $Tb = \mu + \alpha\sigma$, $\alpha \in [5, 6]$
  - Choose Ts to be greater than the mean difference and on the right slope of M
    - Ts $\in [8, 10]$, constant over samples

Multi-pass approach

- Scanning all frames could be computationally hard
- Temporal skipping is more useful
  - e.g. one out of every 10 frames
  - Better for detecting gradual transition
  - May miss camera break
  - May get false detection (distance increased)
- Multi-pass approach
  - First pass, use either pair-wise or histogram with large skip factor and smaller Tb to collect the potential regions.
  - Second pass, two methods may be applied together (hybrid) to search the candidate regions while increasing the confidence.

Distinguish camera movement

- To distinguish gradual transitions from changes made by camera movements
- Basic approach— observing optical flow via motion vectors

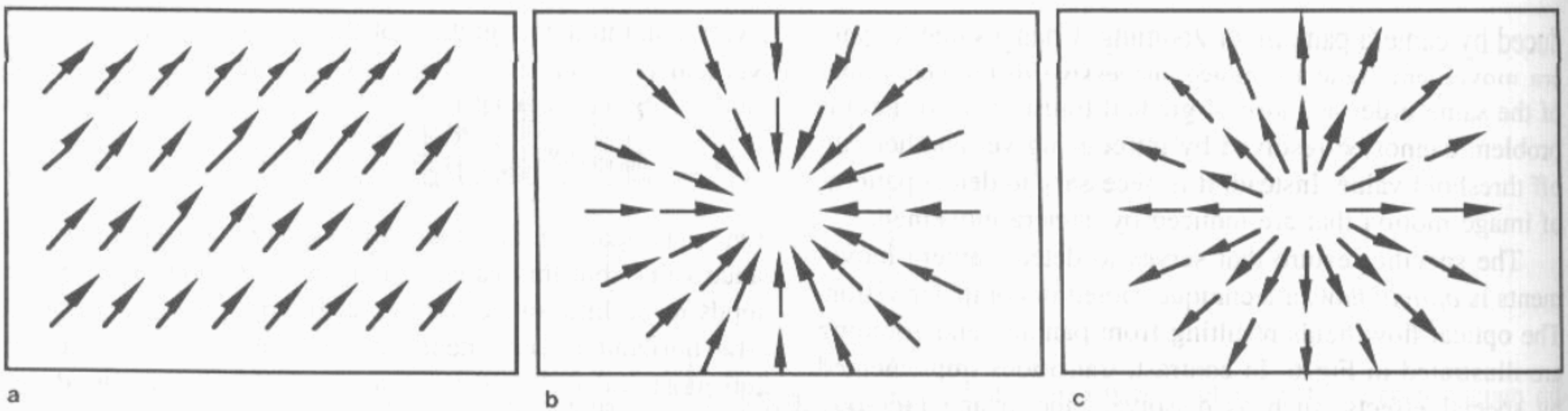


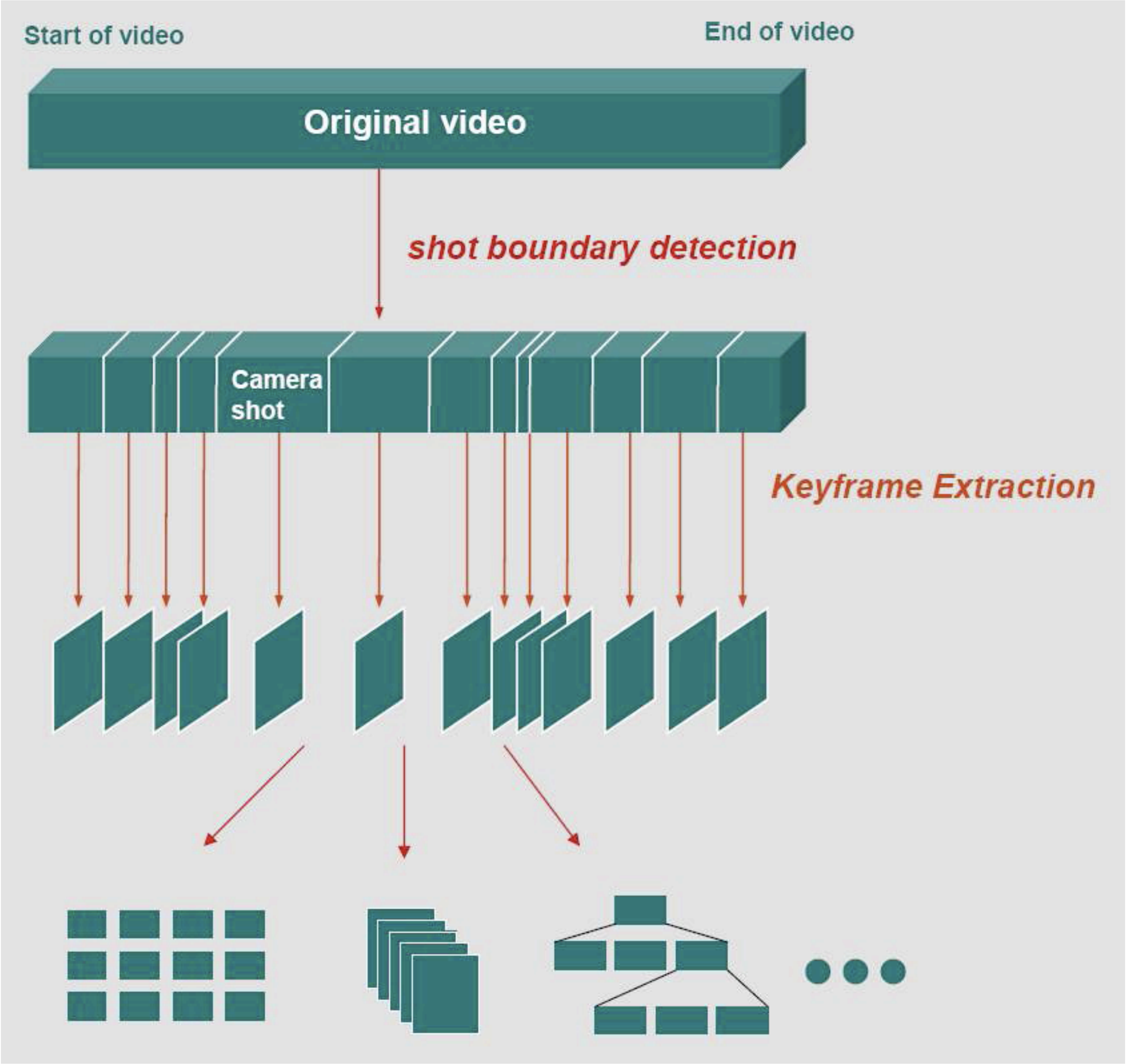
Fig. 6a–c. Motion vector patterns resulting from camera panning and zooming. **a** ⁄ Camera panning direction. **b** Camera zoom-out. **c** Camera zoom-in

Distinguish camera movement

- **Panning**
  - Distribution of motion vectors has a single modal value (θm) that corresponds to the panning direction.

$$\sum_{k}^{N} |\theta_k - \theta_m| \leq \Theta_p$$

- **Zooming**
  - The vertical components of top and bottom motion vectors have different signs.
  - Similarly for horizontal components of left and right motion vectors.


Start of video
End of video
Original video
shot boundary detection
Camera
shot
Keyframe Extraction

关键帧提取技术

- 镜头边界法
  - 选取镜头中的首帧和末帧
- 颜色特征法
  - 首帧为关键帧，其后比较与前面关键帧的颜色差异
- 运动分析法
  - 分析相机的运动
- 聚类分析法

Digital Library

形上谓道兮
形下谓器

- Main features?
- What are the major difference comparing with other applications?

What is a Library?

- A place in which literary, musical, artistic, or reference materials are kept for use but not for sale
- A collection of such materials
  - books, manuscripts, recordings, or films

http://www.m-w.com/cgi-bin/dictionary?book=Dictionary&va=library&x=0&y=0

What is a Digital Library (DL)?

"...a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network" (Arms, p. 2)

What is a Digital Library

- Library ++ (library+archive+museum+…)
- Distributed information system + organization + effective interface
- User community + collection + services
- Digital objects, repositories, IPR management, handles, indexes, federated search, hyper-base, annotation

Digital archives

- Archives differ from libraries
  - Containing primary sources of information
    - typically letters and papers directly produced by an individual or organization
    - rather than the secondary sources found in a library (books, etc);
  - Having their contents organized in groups rather than individual items.
  - Having unique contents

Architecture of Digital Library
Kahn/Wilensky Framework(KWF)

Digital library objects are more than collections of bits

objects = metadata + data

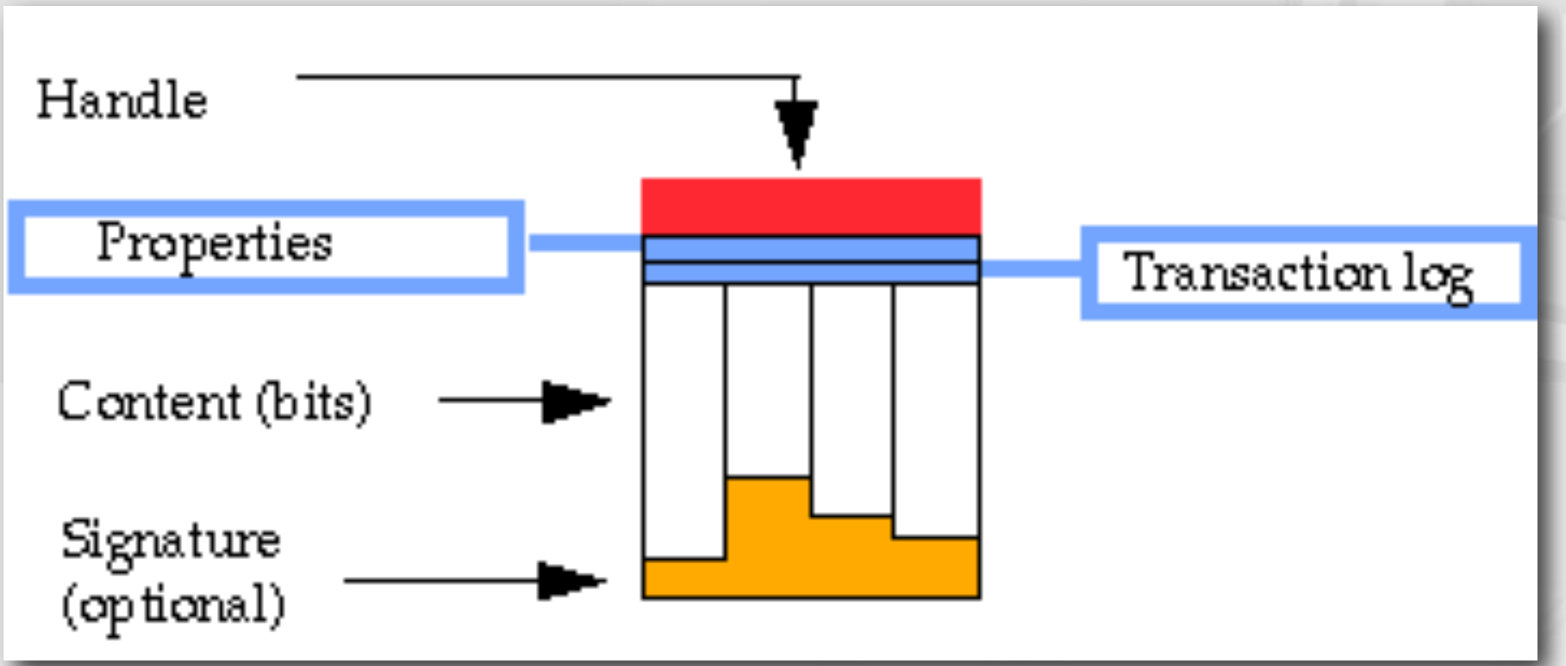


figure 2 in http://www.dlib.org/dlib/July95/07arms.html

Users want intellectual works, but not only digital objects

- The **D**igital **L**ibrary architect's needs should not inconvenience the users' needs
- recombination of objects
  - –what is an object in your world view?

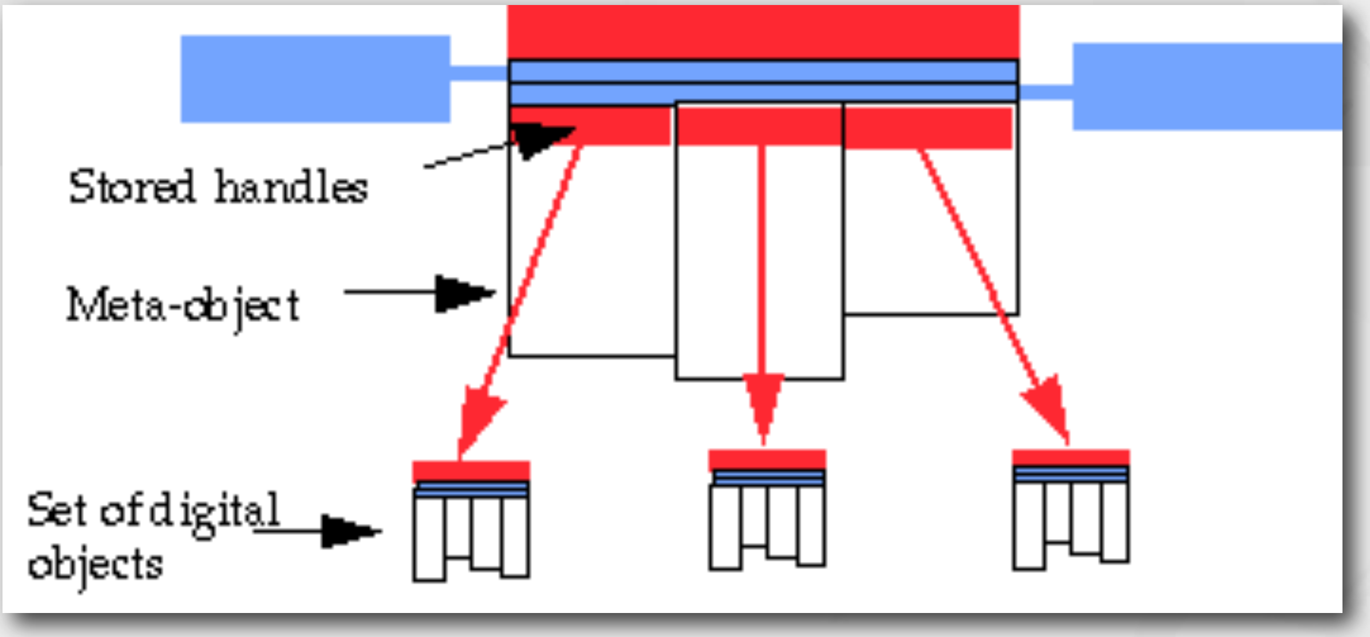
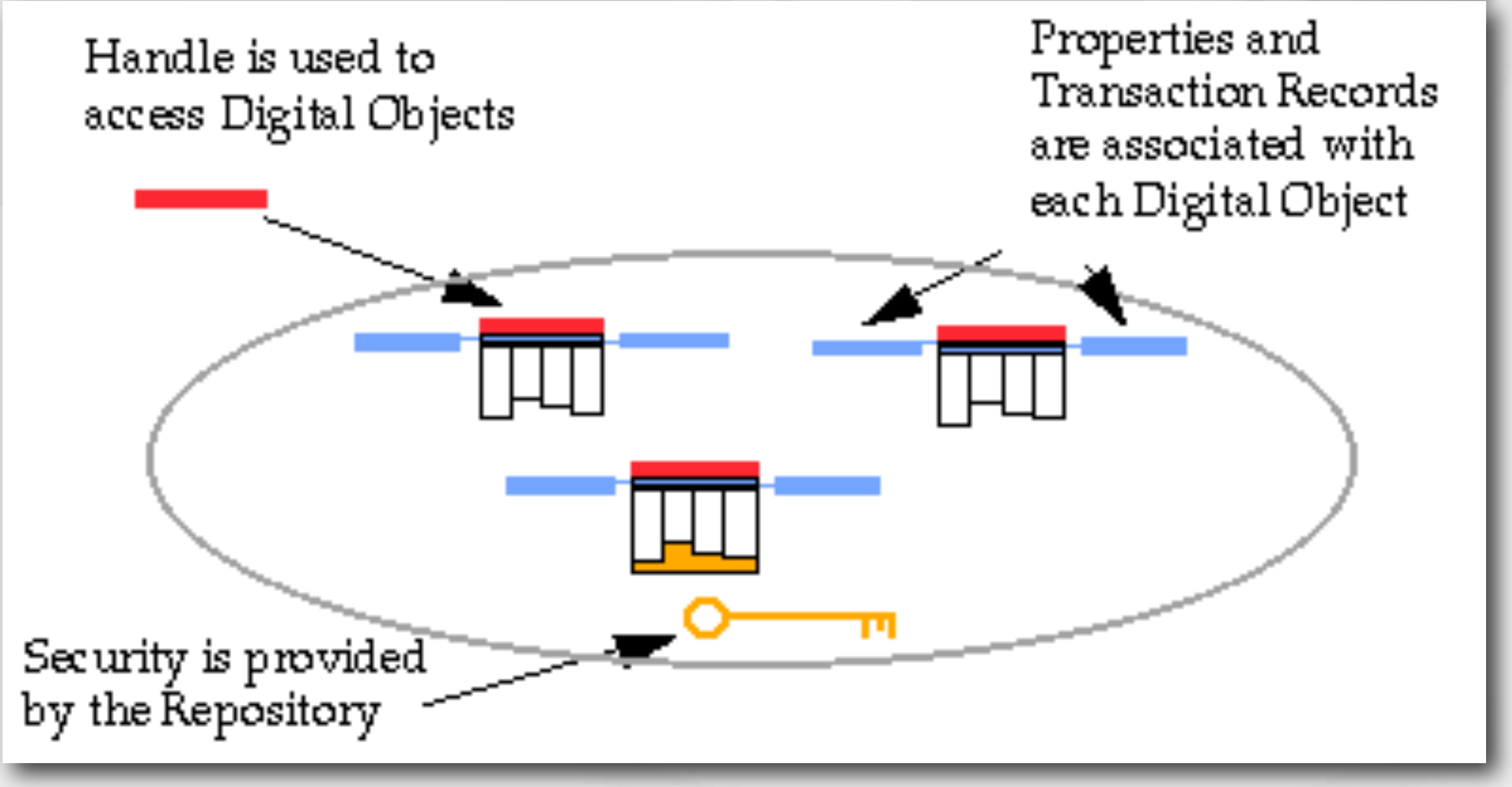


figure 4 in http://www.dlib.org/dlib/July95/07arms.html

Repositories must look after the information they hold

- "Repository Access Protocol"
  - –Kahn Wilensky Framework
    - http://www.cnri.reston.va.us/home/cstr/arch/k-w.html

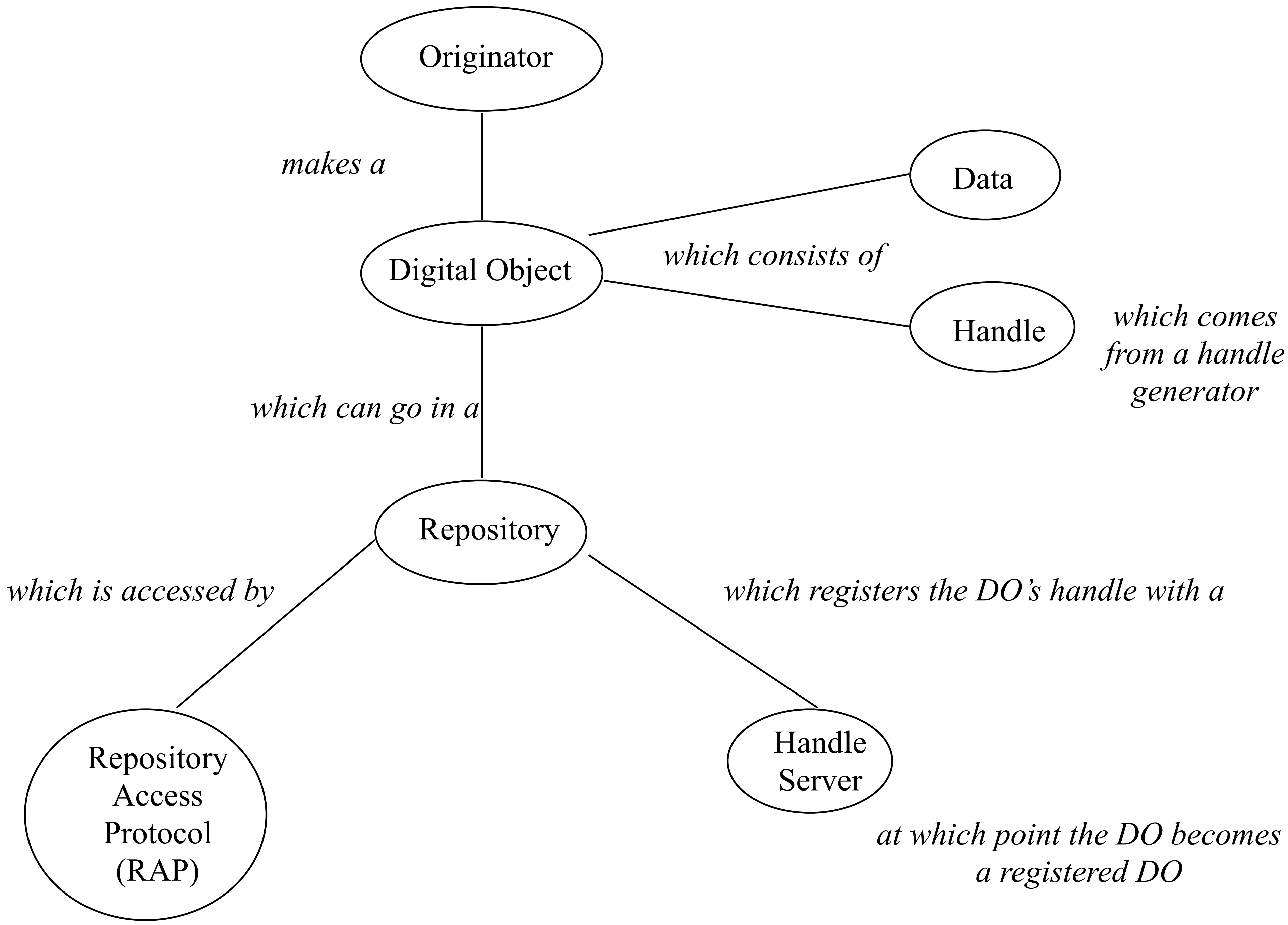
http://www.dlib.org/dlib/July95/07arms.html



Key KWF Terms

- digital objects (DOs)
  - a unit of exchange for the DL with a particular data structure and characteristics
- repository
  - the place where DOs live
- handles
  - a unique, persistent name for a DO

Kahn/Wilensky Framework(KWF)



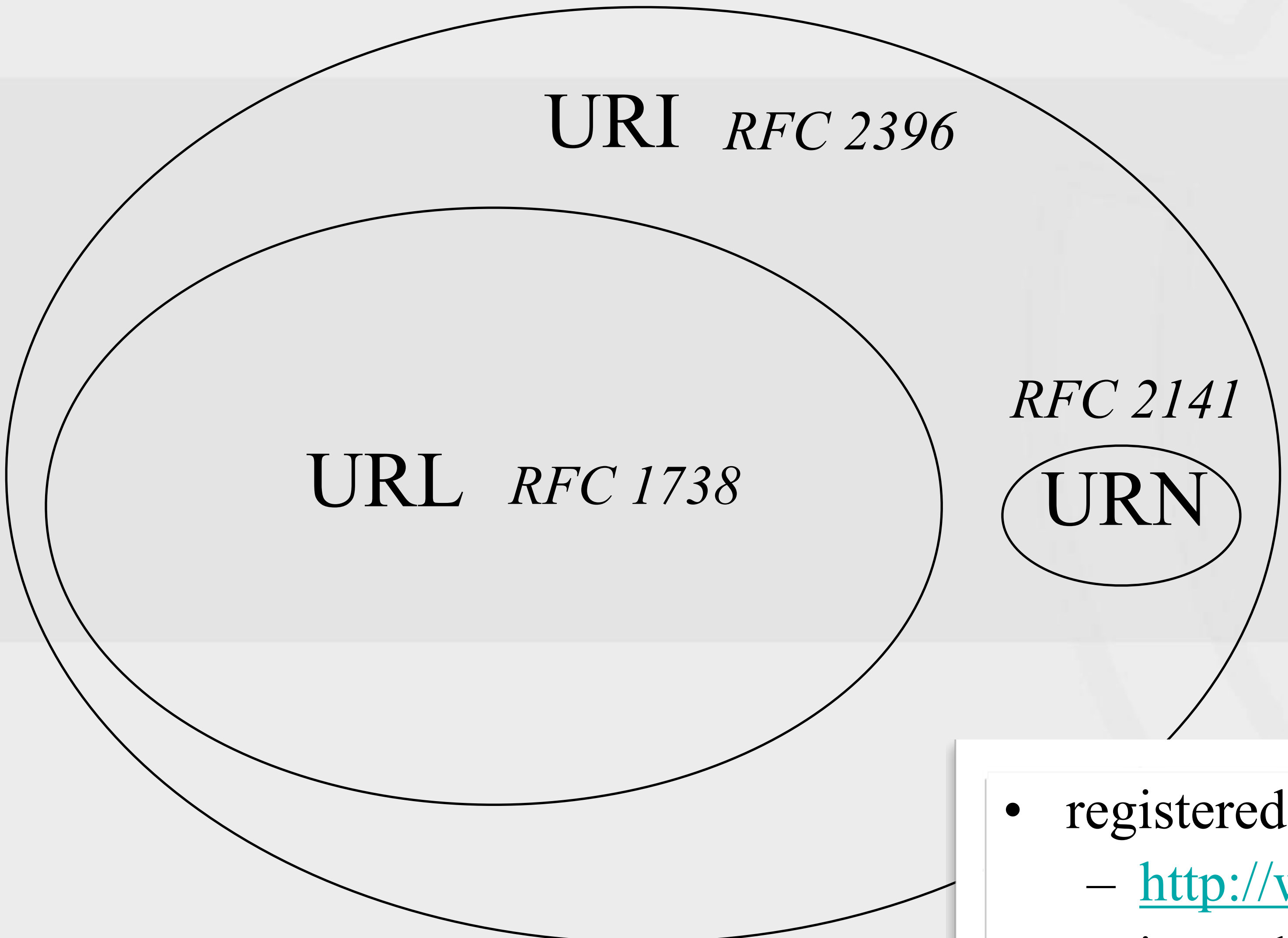
Digital Objects

- Digital object = data + key-metadata
  - data is classified; core classes include:
    - bit-sequence / set-of-bit-sequences
    - digital-object / set-of-digital-objects
    - handle / set-of-handles
  - other types can be defined, and registered with a global type registry
    - definition and registration left undefined
    - similar to MIME?
  - key-metadata includes handle, possibly other metadata (left undefined in KWF)

Digital Objects

- Typed data; example from KWF:
  - a DO subtype: computer-science-tech-report
  - with metadata: author, institution, series, etc.
- Composite DOs:
  - a DO with data of type digital-object
  - non-composite DOs are *elemental* DOs
  - composite DOs can be used to collect similar works together
    - composite DO that contains a DO for each work of Shakespeare...

Uniform Resource Identifiers



- registered URI schemes
  - http://www.iana.org/assignments/uri-schemes
- registered URN namespaces
  - http://www.iana.org/assignments/urn-namespaces

Handles

- Handles can be thought of as a Uniform Resource Name (URN) implementation
  - historical comparison of efforts
    - http://www.dlib.org/dlib/february96/02arms.html
- the handle system (http://www.handle.net/)
  - persistence
  - location independence
  - multiple instances

Repositories

- “A network accessible storage system in which digital objects may be stored for possible subsequent access or retrieval” (KWF)
- A *stored* DO is a DO that resides in a repository
- A *registered* DO is a DO that the repository has registered with a handle server
  - storing and registering can be the same or different processes

Repositories

- A repository keeps a *properties record* for each DO
  - contains key-metadata and any other metadata the repository chooses to keep
- A *repository of record* (ROR) is the first repository that a DO is placed in
  - ROR authorizes additional instances of the DO
- A *dissemination* is the result of an access service request

Repository Access Protocol (RAP)

- “Protocol” may be misleading, its really just the skeleton for a protocol
- RAP is designed to be simple
  - –repositories themselves should be simple
- KWF defines 3 basic operation classes:
  - –ACCESS_DO
  - –DEPOSIT_DO
  - –ACCESS_REF
    - this is the catch-all operation for all meta-services...

RAP

- RAP is fleshed out more in Cornell CS 95-TR1540
- Where KWF suggested that the operations would take "metadata", "key-metadata", and "digital object" as arguments, TR1540 splits some of those into separate operations
- RAP could be implemented as a subset of a more sophisticated protocol (Dienst, Z39.50, etc.)
  - prelude to the Open Archives Initiative (OAI) metadata harvesting protocol

RAP

Operation	Origin	Description
ACCESS_DO	KWF	requests a dissemination of a DO
VERIFY_DO	TR1540	verify that a DO is in the repository
ACCESS_META	TR1540	access a metadata element of a DO
MUTATE_DO	TR1540	modify data for a DO
MUTATE_META	TR1540	modify metadata for a DO (not key-metadata!)
DEPOSIT_DO	KWF	put a DO in a repository
REPLICATE_DO	TR1540	copy a DO to another repository (new handle)
REINSTANTIATE_DO	TR1540	copy a DO in the same repository (still gets a new handle)
DELETE_DO	TR1540	delete DO from repository and handle from handle server
ACCESS_REF	KWF	return references to servers that perform operations on this repository

New Digital Library Architecture

- Open Architecture
  - functionality partitioned into set of **well-defined services**
  - services accessible via **well-defined protocol**
- Modularization
  - promotes interoperability
  - scalable to different clientele (research library, informal web)
- Federation
  - enable aggregations into logical collections
- Distribution
  - of content (collections) and services
  - of administration and management of DL

metadata

- Standard and how to use it?

天下来同

- Game development:
  - Open questions ...

Metadata vs. Data

- **Data** refers to digital objects or digital representations of objects
- **Metadata** is information about the objects (e.g. title, author, etc.): descriptive, interpretive, administrative, …
- Many digital library efforts, including the Open Archives Initiative (OAI), focus on metadata, with the implicit understanding that metadata usually contains useful links to the source digital objects
- Purists would argue metadata is just data

Metadata

- “data about data” is about as good as the definition gets...
- As a DL grows, metadata becomes more important
- Lack of metadata has different consequences
  - documentation: metadata can be regenerated automatically, or by hand
  - datasets, pictures: once lost, can be impossible to regenerate
    - LaRC windtunnel example

Types of Metadata

- Descriptive See http://www.loc.gov/standards/metadata.html
  - Discovery / description of objects
    - Title, author, abstract, etc.
- Structural
  - Storage & presentation of objects
    - 1 pdf file, 1 ppt file, 1 LaTeX file, etc.
- Administrative
  - Managing and preservation of objects
    - Access control lists, terms and conditions, format descriptions, “meta-metadata”

Digital Objects

- Digital Objects = digital information + metadata + handle
  - XML markup, digitized radio programs, computer programs…
- Rules and conventions for each category of digital objects
  - Grouping several digital objects for complex works
    - Reports = several chapters
  - Representing relation between works
    - Versions, translation, …
  - Naming
  - Need a user interface aware of rules and conventions…

Metadata Formats

- MARC is very rich
  - good candidate for an “archival” metadata format, from which simpler formats can be derived
- Dublin Core designed to be simple enough for the average author to generate by hand
  - only 15 core fields defined
- Other formats defined for specific purposes:
  - BibTeX: TeX/LaTeX publishing
  - refer: troff/nroff
  - RFC-1807: email exchange

Interesting Formats

- Library science
  - Machine Readable Catalogue (MARC): huge, extensive, all purpose, one size fits all format
    - pro: does everything
    - con: kids, don't try this at home!
- Computer science
  - application-specific formats: refer, BibTeX, RFC-1807, etc.
- DC - common ground?

Background and Primary Goal of DC

- Came out of a 1995 joint OCLC/NCSA workshop in Dublin, Ohio
- An attempt to improve resource discovery on the Web
  - resource discovery/description/evaluation
- Now adopted by many resource description communities
  - Museums, libraries, government agencies, and commercial organizations
- Building an interdisciplinary consensus about a core element set for resource discovery
  - Simplicity of creation and maintenance
  - Semantic Interoperability
  - International Consensus
  - Flexible extensibility
  - Metadata modularity on the Web

Dublin Core Element Sets

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type

- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

- 15 elements of descriptive metadata
- All elements are optional and repeatable
- Dublin Core is extensible
  - Offering a starting point for semantically richer descriptions

Dublin Core and RDF/XML

- Dublin Core is about semantics
  - What we are trying to say about resources
- RDF is about structure
  - Conventions for encoding the assertions about a resource that uses DC semantics
- XML
  - Syntax for encoding assertions in RDF
  - RDF-encoded DC metadata

MARC

from:
http://m27-5.bgsu.edu/nwoetf/marc/phillippe.html

Leader: : 01663ngm 22002771 4500:
005: : 19950927090218. 0 :
007: :vducgaiuu:
008: : 950927s1993 mau--- d vlfre d :

Ctrl Numb 001 200312310
Cntl Iden 003 OBgNWOET
ISBN Numb 020 -- a 0300056958
Catl Orig 040 __ a OBgNWOET
Tran c OBgNWOET
Lang Summ 041 -- b fre
Titl Main 245 00 a A la recontre de Philippe
GMD h [videorecording] /
Resp c Massachusetts Institute of Technology ; written by Gilberte Furstenburg ; directed by Janet H. Murray ; software programmed by Stuart
A. Malone.
Pubn City 260 __ a Cambridge, MA :
Publ b dist. by Annenberg/CPB.,
Date c 1993.
Desc Extn 300 __ a 1 laserdisc (CAV) :
Othr b sd., col. :
Dimn c 12 in. +
Accm e Teacher guide + 3 computer disks.
Note Genl 500 __ a Issued as videodisc.
Note Genl 500 __ a Title from cover.
Note Summ 520 __ a Provides an engaging way to sharpen comprehension skills. Students navigate through Paris neighborhoods and shops,dealing with friends, tradespeople, telephones and answering machines with the goal of finding an apartment for the hapless Philippe. Includes many helpful tools,such as self-testing exercises and an electronic glossary, visual and audio resources, including maps, telephones and newspapers which help students function within the story. Teachers can customize the program according to their students levels and abilities.
Note Targ 521 2_ a Senior high and college.
Note Targ 521 2_ a 09-adult.
Note Tech 538 -- a Macintosh computer ; system 6.0 or later ; 2 MB of
RAM ; 3.5 MB of hard disk space ; videodisc player ; video monitor.
Subj Topc 650 _0 a Languages, Modern.
Subj Topc 650 _0 a Language and languages.
Subj Topc 658 _7 a Foreign languages, French.
Srce 2 nwoet
Locn Coll 852 1_ a OBgNWOET
SubA b Northwest Ohio Media Center
Clas h 200312310
BarC p 200312310

Dublin Core, pre-XML

```
<META NAME="DC.title" CONTENT="Metadata: Enabling the Internet">
<META NAME="DC.subject" CONTENT="(SCHEME=keyword) Metadata, Dublin Core, PICS, Resource Discovery">
<META NAME="DC.author" CONTENT="(TYPE=name) Renato Iannella">
<META NAME="DC.author" CONTENT="(TYPE=email) renato@dstc.edu.au">
<META NAME="DC.author" CONTENT="(TYPE=affiliation) DSTC Pty Ltd">
<META NAME="DC.author" CONTENT="(TYPE=name) Andrew Waugh">
<META NAME="DC.author" CONTENT="(TYPE=email) a.waugh@cmis.csiro.au">
<META NAME="DC.author" CONTENT="(TYPE=affiliation) CSIRO">
<META NAME="DC.publisher" CONTENT="(TYPE=name) DSTC Pty Ltd">
<META NAME="DC.date" CONTENT="(TYPE=creation) (SCHEME=ISO31) 1997-01-20">
<META NAME="DC.date" CONTENT="(TYPE=current) (SCHEME=ISO31) 1997-01-20">
<META NAME="DC.form" CONTENT="(SCHEME=imt) text/html">
<META NAME="DC.identifier" CONTENT="(TYPE=url) <http://www.dstc.edu.au/RDU/reports/CAUSE97/>">
<META NAME="DC.language" CONTENT="(SCHEME=iso639) en">
```

see Internet RFC-2731

from:
http://www.dstc.edu.au/RDU/reports/CAUSE97/

Dublin Core, XML-encoded

```xml
<?xml version="1.0"?>
	<!DOCTYPE rdf:RDF SYSTEM "http://purl.org/dc/schemas/dcmes-xml-20000714.dtd">
	<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
		xmlns:dc ="http://purl.org/dc/elements/1.1/">
	 <rdf:Description about="http://foo.edu/dl/report-1">
	   <dc:title>Perpetual Motion Machine</dc:title>
	   <dc:description>This report redefines physics.</dc:description>
	   <dc:date>1998-10-10</dc:date>
	   <dc:format>text/html</dc:format>
	   <dc:language>en</dc:language>
	   <dc:contributor>Kant, B. Reproduced</dc:contributor>
	 </rdf:Description>
	</rdf:RDF>
```

example adapted from: http://www.purl.org/dc/documents/wd/dcmes-xml-20000714.htm

期末考试

- 目的:
  - 知识整理，加深印象
- 方式:
  - 开卷有益
- 时间:
  - 11月7日

考试题型

- 100分 vs. 120分钟
  - 名词解释：20%
  - 辨析题：20%
  - 计算：30%
  - 综合设计：30%