Lecture 1: Introduction to Multimedia Systems: Research & Applications

Dr. Jian Zhang
NICTA & CSE UNSW
COMP9519 Multimedia Systems
S2 2007
jzhang@cse.unsw.edu.au
1.1 Introduction

- Our team’s profile – Multimedia and Visual Communication Research Group (MVC); NICTA

- The lectures of this course will be given by Jian Zhang, Jing Chen and Reji Mathew

- This course covers three parts:
  - Image/video coding technology,
  - Streaming multimedia and
  - Multimedia content description, analysis and retrieval
1.1 Introduction
Who I am-- Dr. Jian Zhang’s profile

- Principal Researcher and team leader of MMVC in NICTA
  - Lead three NICTA research projects
- MSc. from School of CS, Flinders University of South Australia, PhD from Dept. of EE, University College, UNSW@ADFA.
  - Senior research engineer, later Principal research engineer and the team manager of Visual Communications Research team in Motorola Labs in Sydney.
  - Co-author of more than ten patents filed in the US, UK, Japan, and Australia,
    - Six issued patents from the US patent office,
    - One issued patent from the UK, and Japan and Australian patent offices respectively
- Many journal and conference papers on my webpage
  - [http://www.cse.unsw.edu.au/~jzhang](http://www.cse.unsw.edu.au/~jzhang)
- Jian’s research interest include: Multimedia content management; video understanding; QoS based joint source and channel coding for video transmission over wireless network systems
1.1 Introduction
Who I am-- Dr. Jing Chen

- Jing Chen graduated in Information Engineering from Beijing University of Post & Telecommunication in 1990.
- Jing Chen got his master degree in Telecommunication Engineering from Shanghai Jiaotong University in 1993.
- Jing Chen received his Ph.D degree in Electrical Engineering from University of Sydney in 2003.
- Jing Chen was a senior system engineer with China Telecomm during 1994 and 1998.
- Jing Chen was a senior software engineer with Motorola Australia Software Sydney Center during 1999 and 2003.
- Jing Chen joined NICTA as a senior research engineer since 2004.
- Jing Chen's research interests include video process and video communication, video QoS enhancement over wireless networks, and video representation and understanding using machine learning technologies. Jing has had more than 10 papers published in the relevant areas.
1.1 Introduction
Who I am– Reji Mathew’s profile

- Reji Mathew graduated in Electrical Engineering from University of Western Australia in 1990.
- Reji completed a Master of Engineering, by research, from UNSW@ADFA in 1996.
- Reji worked as a senior research engineer in Visual Information Processing Lab, Motorola Labs, in Sydney, from 1997 to 2003.
- Reji joined NICTA in 2004 as senior research engineer and is currently pursuing a PhD at the school of Electrical Engineering and Telecommunication at UNSW.
- Reji’s research interests include digital video processing, video coding and streaming technologies. Reji has published a number of papers and holds two international patents in the area of video processing and coding.
1.2 Course Scope & Arrangement

- The Scope of this Course:
  - Provide fundamentals of state-of-art multimedia technologies
    - Concepts
    - Principles of these technologies and,
    - Their applications
  - Provide a base of introduction to multimedia system
    - Digital audio and image/video signal coding and compression;
    - Multimedia streaming and multimedia presentation
    - Multimedia content description
    - Video structure analysis; video summarization and representation
    - Multimedia database indexing, browsing and retrieval;
1.2 Course Scope & Arrangement – Subject Outline

- Objectives:
  - On successful completion of this subject, students will:
    - understand fundamental concepts, theory and techniques of:
      - digital audio and image/video signal coding and compression;
      - multimedia streaming and multimedia presentation
      - multimedia content description
      - video structure analysis; video summarization and representation
      - multimedia database indexing, browsing and retrieval;
    - be familiar with applications of multimedia systems and their implementations;
    - be able to apply the techniques in real applications
    - gain skills and knowledge beneficial to future work and post-graduate study in multimedia area
1.2 Course Scope & Arrangement

- Lecture 1 -- Introduction (Jian Zhang)
- Lecture 2 -- Image processing -- Fundamentals (J. Zhang)
- Lecture 3 -- Image and video coding techniques (I) (J. Zhang)
- Lecture 4 -- Image and video coding techniques (II) (J. Zhang)
- Assignment 1
- Lecture 5 -- Video compression standards (part 1) (J. Zhang)
- Lecture 6 -- Video compression standards (part 2) (J. Zhang)
- Lecture 7 -- Internet Streaming Media (R. Mathew)
1.2 Course Scope & Arrangement

- Lecture 9 -- Multimedia Content Description (part 1) (R. Mathew)
- Lecture 10 -- Multimedia Content Description (part 2) (J. Chen)
- Lecture 11 -- Video Representation and Summarization (J. Chen)
  Assignment 2
- Lecture 12 -- Fundamentals of Content Based Image and Video Retrieval (J. Chen)
- Lecture 13 -- Video mining, Multimedia content retrieval system, High-dimension Multimedia database, Dimension reduction, Indexing scheme, and Top-k algorithm (J. Chen)
- Lecture 14 -- Course Overview (J. Zhang, J. Chen and R. Mathew)
1.2 Course Scope & Arrangement

- Consultation room & time:
  - Level 4 NICTA L5 Building. Every Wednesday 3-4 PM from Week 2.
1.2 Assessment

- Assignment 1 (30%)
- Assignment 2 (30%)
- Final Exam (40%)
  - Understand basic concepts
  - Describe concepts for problem solving
1.3 Multimedia Applications -- Digital Video

- Video conference and telephony
- Multimedia communications
- Digital video Camera
- DVD/VCD
- HDTV and SDTV
- Video surveillance and security
- Video/image database
- Interactive multimedia
- Multimedia data storage and management
- Digital terrestrial and satellite TV
1.3 Multimedia Applications -- Streaming Video (MPEG-4 standard)

- **Content Creation**
  - Live MPEG-4
  - Wireline Multicast (e.g., daycare, security)
  - Professional MPEG-4 Encoder (Content Provider)
  - Live MPEG-4 Wireless Camera
- **Delivery**
  - Error resilient encoder
  - MPEG-4 Server
  - Internet Content
  - Live MPEG-4 Wireline Multicast
- **Consumption**
  - Home entertainment unit
  - Embedded MPEG-4 Players
  - Error resilient decoder

Networks:
- IP Network
- Wired Networks
- GPRS
- EDGE
- 3G Network
- Home RF Network

Network Protocols:
- 3G
- EDGE
- GPRS

Devices:
- M'Phone
- PC
- Home RF

Applications:
- Streaming Video (MPEG-4 standard)
1.3 Multimedia Applications-- Content Management Demonstration Platform

- Client / Server platform demonstrating content based search using MPEG-7 visual descriptors
  - Content can be searched using methods “query by specification” or “query by example”.
  - For “query by example”, the Analysis Engine at the server extracts visual features from images
  - the Search Engine searches for archived images that have similar features to those of the example region.

Query by Example
1.3 Multimedia Applications-- Content Management Demonstration Platform

- Real Time MPEG-7 metadata and coordinated display at the client

  The metadata stream refers to locations of faces identified by a face detection algorithm. The server performs Stream Management allowing multiple clients to request and receive real-time streams of video and metadata. A coordinated display of both streams is performed by the client.
1.3 Multimedia Applications -- Surveillance System
1.4 Introduction to Multimedia Research

- Why we need video compression?
  - There is a big gap of digital bandwidth demand between users' expectation for multimedia application.
    - To handle huge uncompressed digital video
    - The state-of-the-art of computer and telecommunication technologies
      - storage capacities including access speed
      - transmission capacities
  - Money saving to pay the cost for video data in transmission and storage applications
- The race has been in between:
  - Advanced video applications
  - Computer power
  - Increase of transmission and storage bandwidth and capacities
1.4 Introduction to Multimedia Research

- Comparison figures: Uncompressed ITU-R Rec. 601:
  - 720 x 576 pixels at 25 frames per second (4:2:2);
  - 8-bit per component, leading to
    Total bit rate = 25 frames/s x (720+360+360) x 576 component pixels/frame x 8 bits per component pixel
    = 166 Mbps,
    and a 90 minute movie requires over 100 GBytes.
  - 10-bit per component, leading to
    Total bit rate = 25 frames/s x (720+360+360) component pixels/frame x 10 bits per component pixel
    = 207 Mbps,

Digital Storage Capacities:

CD/VCD: 640~700 MB, typically at 1.2Mbps, (142:1 compression required for storage and 138:1 compression required for bandwidth);

DVD: 4.6 GB, at 4.5 to 6 Mbps (21:1 compression required for storage, 28:1 for bandwidth);

PC Hard Disk: 40 to 80 GB with PCI bus bandwidth of ~132 MBps

Ref: H. Wu
## 1.4 Introduction to Multimedia Research

<table>
<thead>
<tr>
<th>Connection technology</th>
<th>No. of TDM voice channels</th>
<th>Transmission rate</th>
<th>Compression required for video</th>
</tr>
</thead>
<tbody>
<tr>
<td>V.32bis/V.34/V42.bis Modems</td>
<td>1/2/6/24/30</td>
<td>14.4/28.8/56 Kbps</td>
<td>~11528/5764/2964:1</td>
</tr>
<tr>
<td>T1</td>
<td>1</td>
<td>1.544 Mbps</td>
<td>~108:1</td>
</tr>
<tr>
<td>HDSL</td>
<td>1</td>
<td>1.5 Mbps</td>
<td>~110:1</td>
</tr>
<tr>
<td>ADSL 3</td>
<td>1</td>
<td>6.312 Mbps</td>
<td>~26:1</td>
</tr>
<tr>
<td>Cable modem</td>
<td>1</td>
<td>10 Mbps</td>
<td>~16.6:1</td>
</tr>
<tr>
<td>ISDN Channel Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>64 kbps</td>
<td>~2600:1</td>
</tr>
<tr>
<td>2B</td>
<td>2</td>
<td>128 kbps</td>
<td>~1297:1</td>
</tr>
<tr>
<td>H0</td>
<td>6</td>
<td>384 kbps</td>
<td>~432:1</td>
</tr>
<tr>
<td>H11</td>
<td>24</td>
<td>1536 kbps</td>
<td>~108:1</td>
</tr>
<tr>
<td>H12</td>
<td>30</td>
<td>1920 kbps</td>
<td>~87:1</td>
</tr>
<tr>
<td>H4</td>
<td>2344</td>
<td>Up to 150 Mbps</td>
<td>~1.1:1</td>
</tr>
<tr>
<td>B-ISDN (ATM SONET)</td>
<td></td>
<td>100 Mbps up</td>
<td>&lt;1.6:1</td>
</tr>
</tbody>
</table>
1.4 Multimedia Content Management

- Why we need multimedia content management?
  - There is a strong need to effectively manage the vast amount of resulting multimedia content.
    - It is necessary to develop forms of audiovisual information representation that go beyond the simple formats.
    - It is necessary to provide a common multimedia content description interface (defined by MPEG-7 standard).
    - It is necessary to develop a rich set of standardized tools to describe multimedia content.
1.4 Multimedia Content Management

- Why we need multimedia content management?
  - It is necessary to develop a platform to handle the following cases

  - where audiovisual information is created, exchanged, retrieved, and re-used by computational systems

  - where information retrieval is required for quickly and efficiently searching for various types of multimedia documents of interests of the users

  - where a stream of audiovisual content description for users to receive only those multimedia data items which satisfy their preference
1.4 Introduction to Multimedia Research

- Abstract representation of possible applications using MPEG-7

Ref: J. Martinez

How we can achieve the auto. feature extraction???
1.4 Introduction to Multimedia Research

- Example of a Hierarchical Summary of a video of a soccer game -- a multiple level key-frame hierarchy

The Hierarchical Summary denotes the fidelity (i.e., $f_0$, $f_1$) of each key-frame with respect to the video segment referred to by the key-frames at the next lower level.
1.4 Introduction to Multimedia Research

- The Space and Frequency Graph describes the decomposition of an audio or visual signal in space (time) and frequency

Ref: J. Martinez
1.4 Multimedia Content Management

- Given that the strong need for multimedia content management, there are some key challenges:
  - Majority of existing techniques for content management are based on low-level features
  - There is a significant gap between:
    - low-level feature extraction and users expectation on high level understanding (semantic level)
  - Video analysis and understanding technologies (tools) serves the key enabling technology towards semantic content description and representation
  - **This field of research presents significant challenges and enormous opportunities !!!**
1.4 Introduction to Multimedia Research

- Demos:
  - Video Content Understanding
    - Search by event
    - Search by example
  - Video Content Analysis
    - Object tracking under heavy collusions
    - Object classification
    - Tracking trajectory
  - Video coding and Communication
    - MPEG-4 coding and transmission
    - Scalable coding
    - Error resilience coding
    - Two-way video communication on PDA or 3G Mobile phone
5.3 Layered Multimedia Computing Research

- A three-layer of Multimedia Computing Research
  - Video Content Understanding (Demo)
    - Aims to achieve semantic and structural representation (Ontology) of video content to enable meaningful content search and retrieval
    - Topics include
      - Video summarization towards table of video content generation
        - E.g: video shot with semantic description and scene (story) generation
      - Automatic and semi-automatic annotation of image/video
        - E.g: supervised learning to build statistical model for video sequence annotation – indoor/outdoor, car, sky etc
      - Semantic video representation
        - E.g: different modalities video to key frame plus text, video to synthetic video with audio explanation.
5.3 Layered Multimedia Computing Research

- Video Content Analysis (Demo)

- This research investigates the provision of content features and object information to enable meaningful video content presentation.

- Topics include
  - Key frame extraction and video shot segmentation
    - The basic element to build video scene (story) segmentation
  - Object segmentation & tracking (multiple objects and occlusion)
  - Object classification
    - Classify human being, car, trucks, in indoor/outdoor area
1.4 Introduction to Multimedia Research

- Video Coding and Communication (Demo)
  - This research aims to develop enabling coding technology for video archiving, adaptation and transmission to a variety of networked received.
  - Topic include:
    - Advanced scalable video coding technology (e.g: heterogeneous networks)
    - Low bitrate video coding technology (e.g: mobile video telephony)
    - Real time streaming video coding technology (e.g. online multimedia delivery)
    - Video coding adaptation to deliver personalized video content to users with different terminals and different usage scenarios
      - MMS applications
1.5 Basic Concepts in Multimedia Signal Processing

- 1.5.1 Audio data rates
- 1.5.2 Image and Video Sequence
- 1.5.3 Pixel Representation
- 1.5.4 Chrominance sub-sampling
- 1.5.5 Digital Video Formats
- 1.5.6 The Need for Video Compression
1.5.1 Audio Data Rates

- Audio sampling & total bitrates for:
  1) Compact Disk (CD), 2) Digital Audio Tape (DAT), 3) Digital Compact Cassette (DCC) and, 4) MiniDisk (MD)

<table>
<thead>
<tr>
<th>Digital storage Medium</th>
<th>Audio Bandwidth (kHz)</th>
<th>Sampling Rate (kHz)</th>
<th>Stereo Audio Bitrate (kb/s)</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>20</td>
<td>44.1</td>
<td>1411.2</td>
<td>PCM</td>
</tr>
<tr>
<td>DAT</td>
<td>16</td>
<td>32.0</td>
<td>1024.0</td>
<td>PCM</td>
</tr>
<tr>
<td>DCC</td>
<td>20</td>
<td>44.1</td>
<td>384</td>
<td>PASC</td>
</tr>
<tr>
<td>MD</td>
<td>22</td>
<td>44.1</td>
<td>292</td>
<td>ATRAC</td>
</tr>
</tbody>
</table>

PCM: Plus Code Modulation, PASC: Precision Adaptive Subband coding
ATRAC: Adaptive Transform Acoustic Coding
Uncompressed audio typically comprises 16 bits per sample and two channels of stereo for a total bitrate of 32 times the sampling rate.
1.5.2. Image and Video Sequence

- **Digital Image & Video**
  - The basic unit to build an image is called pixel (pel)
  - Image resolution is calculated by pixels in horizontal and vertical coordinates
  - Video sequences consist of a number of motion pictures
  - One of the video sequence formats is well known as Interlaced structure that follows the analogue PAL format

**Scanning pattern of TV**

- Odd field
- Even field

Frame
1.5.2. Video Image and Sequence

- Interlaced Video Format
  - Standard frame rates (25 or 30 frames per second) are high enough to provide smooth motion, but are not high enough to prevent flickering.
  - To prevent perceptible flicker in a bright image, the refresh rate must be at least 50 frames per second. With 2-to-1 interlacing, the odd numbered lines of a frame are displayed first (field 1), followed by the even lines (field 2).
  - A 25 frame per second sequence is displayed at 50 fields per second.
  - The eye does not readily perceive flickering objects that are small, the 25 per second repetition rate of any one scan line is not seen as flickering, but the entire picture appears to be refreshed 50 times per second.
1.5.2. Image and Video Sequence

- Image & Video Sequence

(a) Frame or picture

(b) Video sequence
1.5.3 Pixel Representation

- **Y,U,V Colour Space**
  - The Human Visual System (HVS) is sensitive to three colour components. Colour can be represented by Red, Green and Blue components (RGB).
  - Transform to YUV or YCbCr with less correlated representation:

\[
Y = 0.299R + 0.587G + 0.114B
\]

\[
\begin{bmatrix}
Y \\
U_t \\
V_t
\end{bmatrix} = \begin{bmatrix}
0.299 & 0.587 & 0.114 \\
-0.147 & -0.289 & 0.436 \\
0.615 & -0.515 & -0.100
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]

Note:
The two chrominance components (U,V) contain considerably less information than the luminance component. For this reason, chrominance is often sub-sampled.
1.5.3 Pixel Representation

- \( Y_d C_b C_r \) Colour Space

For digital component signal (CCIR Rec 601), 8-bit digital variables are used, however:

1. Full digital range is not used to give working margins for coding and filtering.
2. RGB to \( Y_d C_b C_r \) conversion is given by

\[
\begin{align*}
\begin{bmatrix}
Y_d \\
C_b \\
C_r
\end{bmatrix} &= \begin{bmatrix}
0.257 & 0.504 & 0.098 \\
-0.148 & -0.291 & 0.439 \\
0.439 & -0.368 & -0.071
\end{bmatrix} \begin{bmatrix}
R_d \\
G_d \\
B_d
\end{bmatrix} + \begin{bmatrix}
16 \\
128 \\
128
\end{bmatrix} \\
\begin{bmatrix}
R_d \\
G_d \\
B_d
\end{bmatrix} &= \begin{bmatrix}
1.164 & 0.000 & 1.596 \\
1.164 & -0.392 & -0.813 \\
1.164 & 2.017 & 0.000
\end{bmatrix} \begin{bmatrix}
Y_d - 16 \\
C_b - 128 \\
C_r - 128
\end{bmatrix}
\end{align*}
\]

The positive/negative values of U and V are scaled and zero shifted in a transformation to the Cb and Cr coordinates.

where digital luminance, \( Y_d \), has a range of (16-235) with 220 levels starting at 16, and digital chrominance difference signals, Cb and Cr, have a range of (16-240) with 225 levels centered at 128.
1.5.3 Pixel Representation

- Yd, Cb, Cr Colour Space

(a) Red
(b) Green
(c) Blue
(d) Yd
(e) Cb
(f) Cr
1.5.4 Chrominance sub-sampling

- Human vision is relatively insensitive to chrominance. For this reason, chrominance is often sub-sampled.
- Chrominance sub-sampling is specified as a three-element ratio.
1.5.4 Chrominance sub-sampling

- In 4:4:4 format: Y, Cr & Cb – 720 x 576 pixels per frame
- In 4:2:2 format: Y – 720 x 576 and Cr & Cb – 360 x 576 pixels per frame
- In 4:2:0 format: Y – 720 x 576 and Cr & Cb – 360 x 288 pixels per frame
  - A commonly used format is 4:2:0 which is obtained by sub-sampling each colour component of 4:2:2 source vertically to reduce the number of lines to 288;
- 4:1:1 format is obtained by sub-sampling 4:4:4 source by 4:1 horizontally only. Y – 720x576 and Cr&Cb – 144x576
1.5.5 Digital Video Formats

- International Consultative Committee for Radio (CCIR) Rec. 601:
  - Two display rates:
    - **50Hz**: 720x576 pixels at 50 fields per second.
    - **60Hz**: 720x480 pixels at 60 fields per second.
  - Both rates are 2:1 interlaced and 4:2:2 chrominance sampling (with optional 4:4:4).

**NOTE:** Figures referred from Y. Wang et al, 2002.
1.5.5 Digital Video Formats

- **Common Intermediate Format (CIF):**
  - This format was defined by CCITT (TSS) for H.261 coding standard (teleconferencing and videophone).
  - Several size formats:
    - SQCIF: 88x72 pixels.
    - QCIF: 176x144 pixels.
    - CIF: 352x288 pixels.
    - 4CIF: 704x576 pixels.

- Non-interlaced (progressive), and chrominance sub-sampling using 4:2:0.

- Frame rates up to 25 frames/sec
1.5.5 Digital Video Format

- Source Input Format (SIF):
  - Utilized in MPEG as a compromise with Rec. 601.
  - Two size formats (similar to CIF):
    - **QSIF**: 180x120 or 176x144 pixels at 30 or 25 fps
    - **SIF**: 360x240 or 352x288 pixels at 30 or 25 fps
  - Non-interlaced (progressive), and chrominance sub-sampling using 4:2:0.
  - It is assumed that SIF is derived from a Rec.601.

- High Definition Television (HDTV):
  - 1080x720 pixels.
  - 1920x1080 pixels.

- JPEG format
  - ITIF

- Audio Format
1.5.5 Digital Video Format

Interlaced format conversion
1.6 Summary

- Today’s topics
  - Introduction
    - Multimedia applications;
    - Introduction to multimedia research;
    - Basic concepts in multimedia signal processing

- Question/Answer