



Research Institute for Future Media Computing    Institute of Computer Vision  
未来媒体技术与研究所    计算机视觉研究所



# 多媒体系统导论

## Fundamentals of Multimedia System

授课教师：张小燕

Email: [xyzhang15@szu.edu.cn](mailto:xyzhang15@szu.edu.cn)

# 第一讲

# Introduction

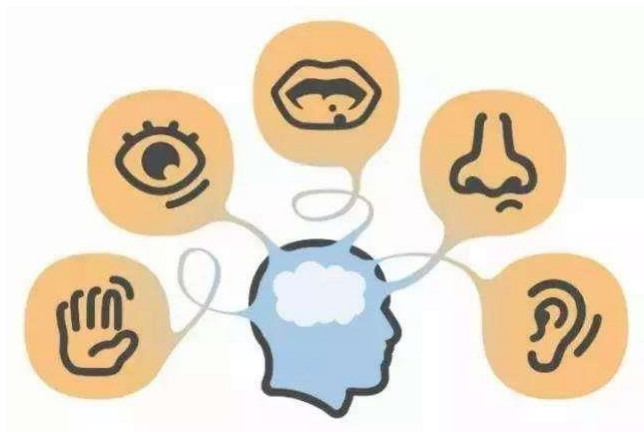
第1-2章

# Outline

- ❖ Definition of Multimedia 什么是多媒体
- ❖ Why Study Multimedia 为什么学习多媒体
- ❖ Multimedia System 多媒体系统
- ❖ Multimedia Computing 多媒体技术
- ❖ Multimedia Authoring and Tools 多媒体编著工具
- ❖ Related Journals and Conferences 相关的期刊和会议
- ❖ 思考题

# Definition of Multimedia

- ◆ Media 媒体：是指传播信息的媒介
- ◆ How do humans accept information?  
人是怎么接受信息的？



视觉、听觉、味觉、嗅觉、触觉

- ◆ 表示信息的媒介？
  - Text 文本, audio 声音, graphic 图形, image 图像, animation 动画, and video 视频

# Definition of Multimedia

- ◆ “Multi” more than one.
- ◆ Multimedia in computer science
  - Computer-based processing of the data includes **at least two** of the following elements:
    - Text
    - Audio and speech
    - Images
    - Graphics
    - Video
    - Animation

# Text, Image, Graphic

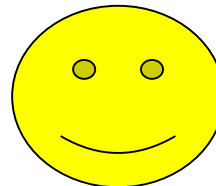
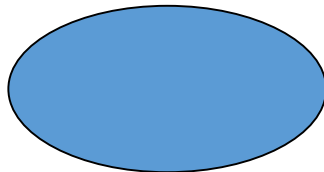
- ◆ Text: A text (in the sense of literary theory) is any object that can be read---wikipedia

Example: 多媒体 (国标码) , *multimedia* (ASCII)

- ◆ Image: An image depicts visual perception, such as a photograph or other two-dimensional picture, particularly one that resembles a subject (usually a physical object). In the context of signal processing, an image is a distributed amplitude of color(s)[1]

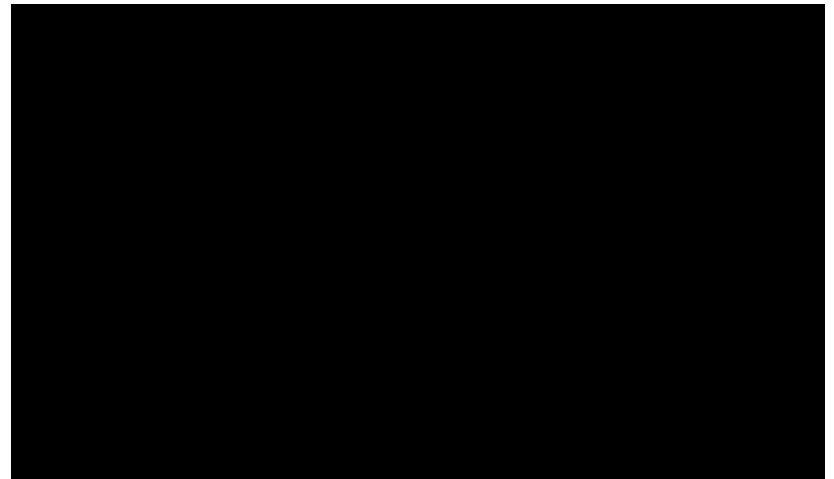
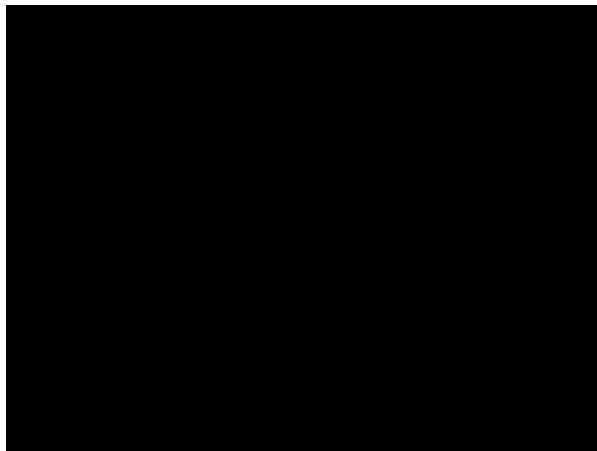


- ◆ Graphics (图形): are visual images or designs on some surface, such as a wall, canvas, screen, paper, or stone to inform, illustrate, or entertain.



# Video, Animation

- ◆ Video: for the recording, copying, playback, broadcasting, and display of **moving visual media**. Video was first developed for television systems.
- ◆ Animation 动画: Animation is a method in which pictures are manipulated to appear as **moving images**. In traditional animation, images are drawn or painted by hand on transparent celluloid sheets to be photographed and exhibited on film. Today, most animations are made with computer-generated imagery (CGI).



# Audio

- ◆ Audio: most commonly refers to sound, as it is transmitted in signal form.



Speech (语音) is a small range of audio



# Why Study Multimedia

- ◆ Consistent with human's perception
  - We perceive the world through all the senses we have at once
  - The integration of media is natural
- ◆ Multimedia technology intends to imitate (模仿) the human's communication with the world
  - Drama is an typical example of multimedia
  - The separation of media is artificial and may often unsatisfactory
- ◆ Multimedia is not only the reversion of the real life but also the creation

# Why Study Multimedia

- ◆ 人工智能简称 AI (Artificial Intelligence)。通常情况下我们对于人工智能的定义是：人工智能是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。
- ◆ 人的智能体现在那几个方面？
- ◆ 思考，学习，运动，听，说，看，情感

# Why Study Multimedia

- ◆ 人工智能简称 AI (Artificial Intelligence)。通常情况下我们对于人工智能的定义是：人工智能是研究、开发用于**模拟、延伸和扩展人的智能**的理论、方法、技术及应用系统的一门新的技术科学。**人工智能是计算机科学的一个分支，它企图了解智能的实质，并生产出一种新的以人类智能相似的方式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。**



# Why Study Multimedia

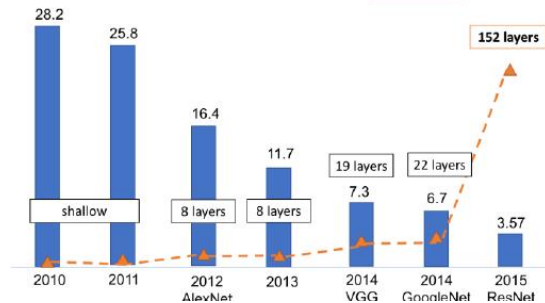


Simple but fathomless Go game

- $10^{170}$  possible layouts
- More than number of atoms in the universe

2015

- ResNet + Image



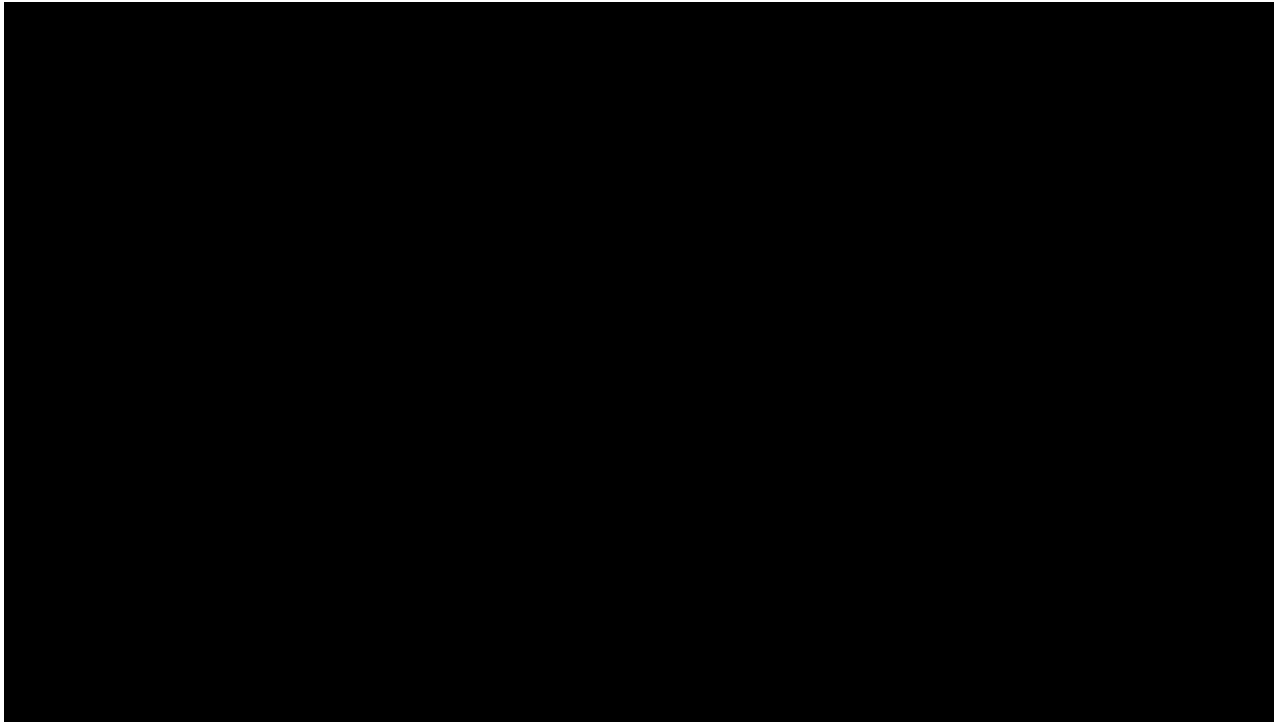
2022

- Best AI image generator

# Why Study Multimedia

## ➞ Robot

- <https://www.youtube.com/watch?v=tIJJME8-au8>



# Why Study Multimedia

## ➞ Photo maker

Reference Face



Instagram photo of S\*

# Why Study Multimedia

## → Sora

We're sharing our research progress early to get feedback from people outside of OpenAI and to give people a sense of what AI capabilities are on the horizon.

We will be taking several important safety steps before this research becomes available in any of our products.

Sora is a new AI model that can create realistic and imaginative scenes from text prompts.

# Multimedia System

- ◆ Multimedia involves more than simple addition of new data types
- ◆ It integrates a wide range of symbol modes simultaneously into a coherent (一致的) framework
- ◆ The framework is usually denoted as a multimedia system
- ◆ 腾讯课堂是不是一个多媒体系统？



# Examples of Multimedia Systems

*Real Estate*

real estate: 房地产

The screenshot displays a 'House Flip Book' interface with a blue background. At the top, it shows 'Page 2 of 6' and a 'Close' button. Navigation buttons include 'Next', 'Previous', 'First', 'Last', and 'Load...'. The main area features four property listings, each with a color image, a title, a price, and a description. The listings are for houses in Cambridge, with prices ranging from \$990,000 to \$1,050,000. To the right of the listings is a 'Video Conference' window showing a man with glasses and a headset, with an 'Out' button. Below the video conference is a 'Mortgage Calculator' window with the text '(values are only approximations)'. It contains a table with two columns, A and B, and rows for various mortgage calculations.

	A	B
0	Home Mortgage Calculations	
1		
2		
3	House Price	120000.00
4	Down Payment	20000.00
5	Interest Rate	10.00
6	Years of loan	30.00
7		
8	Monthly Payments	877.57
9	Total Payments	335925.77
10	Total Interest	215925.77

This prototype real estate application was developed with the AthenaMuse software at GTE Laboratories. House descriptions and color images are retrieved from a multimedia database and placed in a customized listing booklet. The client and Realtor can discuss candidate homes and financing options via desktop videoconferencing and a shared document facility.

Interface:  
Russell Sasnett, at  
GTE Laboratories.

*Real Estate* in MIT's Project Athena



# Examples of Multimedia Systems

- ➡ 网易伏羲全明星街球派对AI竞赛

<https://fuxi.163.com/competition/DunkCity>



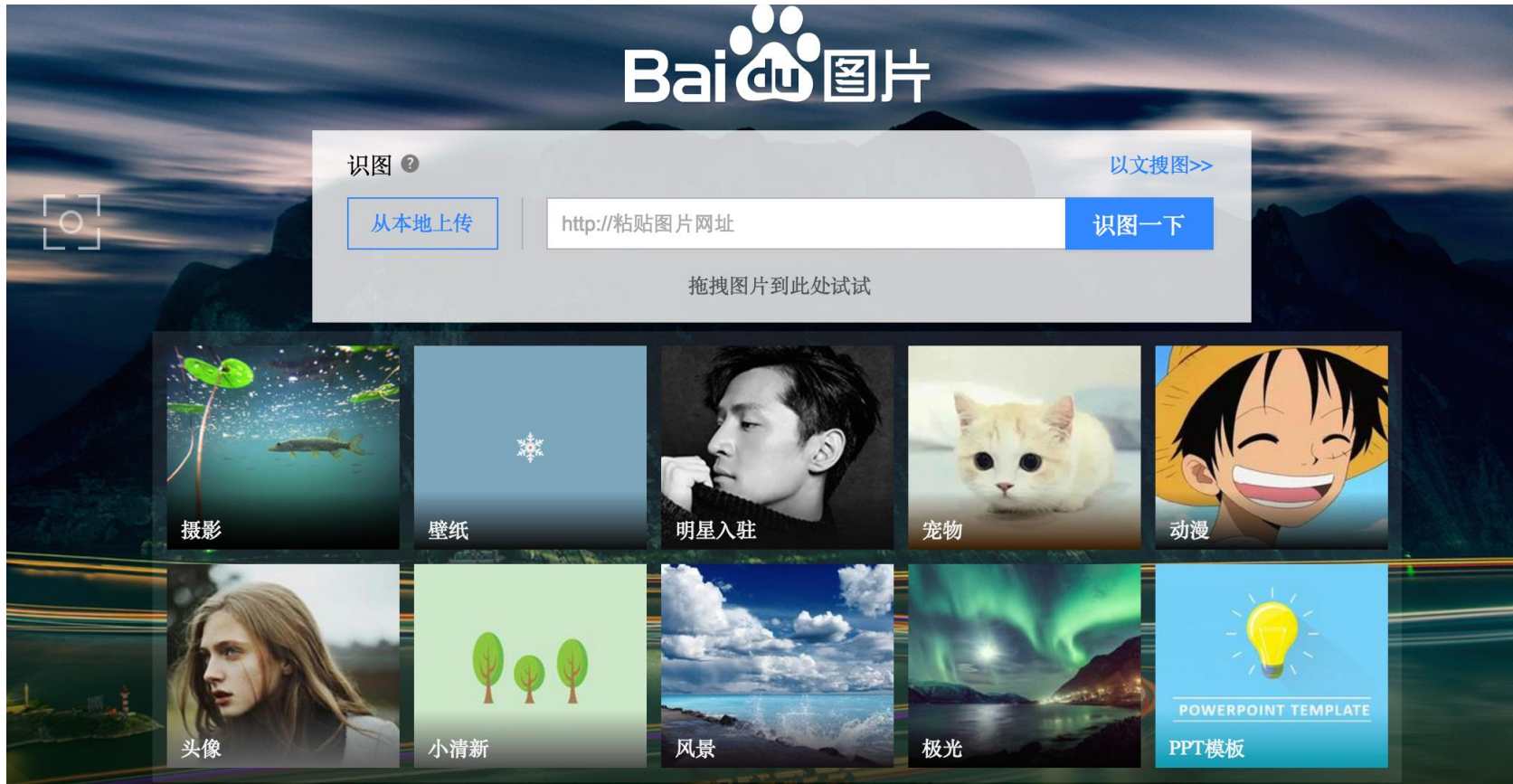
- ➡ 腾讯 开悟

<https://aiarena.tencent.com/aiarena/zh>



# Examples of Multimedia Systems

Baidu's image-keyed image retrieval



<http://image.baidu.com/?fr=shitu>

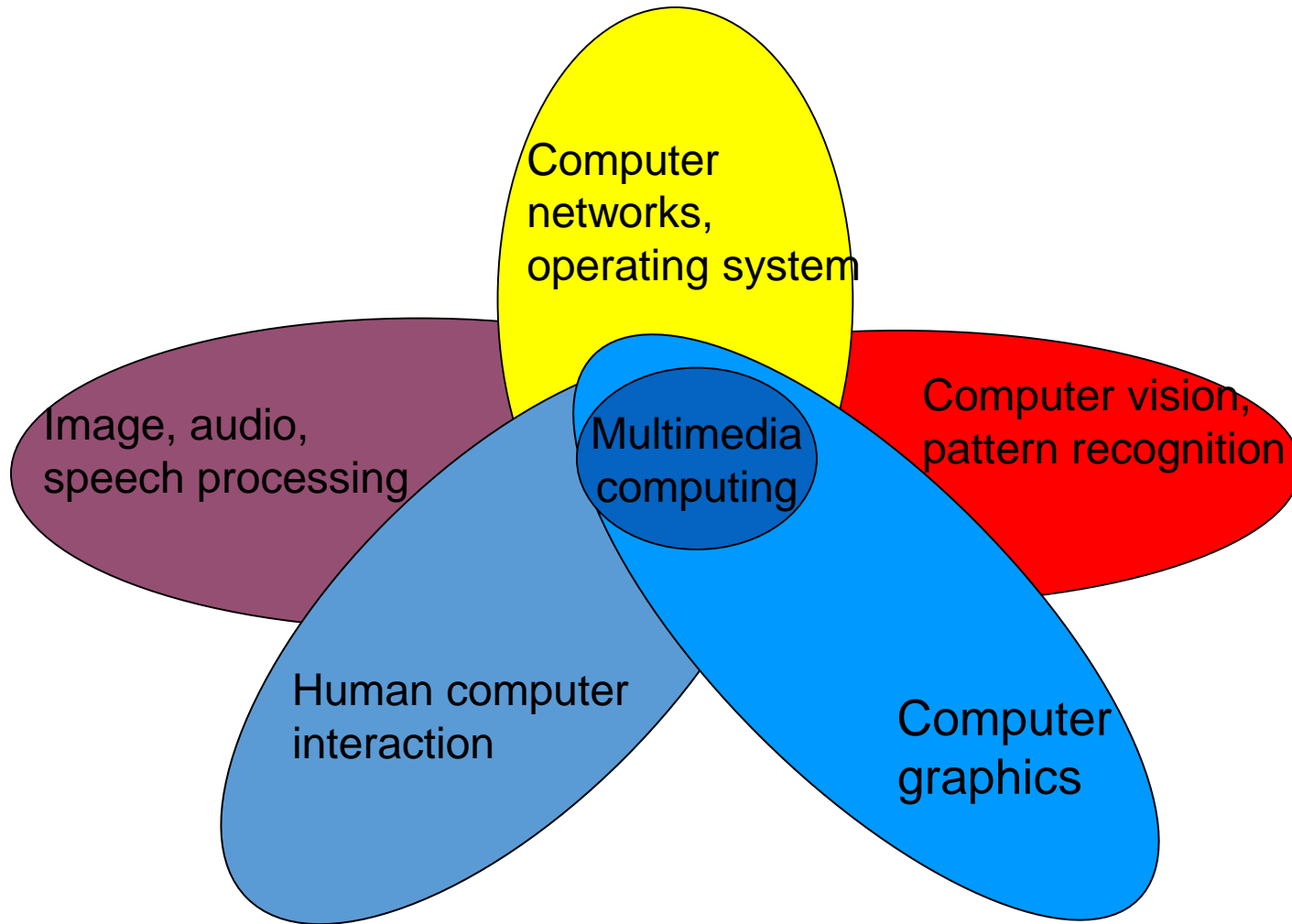
# Multimedia Computing

- ◆ Multimedia systems involve some basic enabling techniques
  - Multimedia data representation and compression
  - Multimedia data processing and analysis
  - Transmitting （传播） multimedia data through communication networks
  - Multimedia database, indexing （索引） and retrieval （检索）

# ■ Challenges of Multimedia Computing

- ◆ Developing a successful multimedia system is hard
  - Continuous media types such as video need a lot of **space** to **store** and very high **bandwidth** to **transmit**
  - They also have tight **timing constraints**
  - Automatically **analyzing, indexing and organizing** information in audio, image and video is much harder than from text
  - Multimedia involves many different research areas and needs more complex and more efficient algorithms and hardware platforms

# Multimedia is Multidisciplinary





# Data Compression



Raw image takes about 6M bytes  
(without header information)



24k bytes with jpeg, Q=50

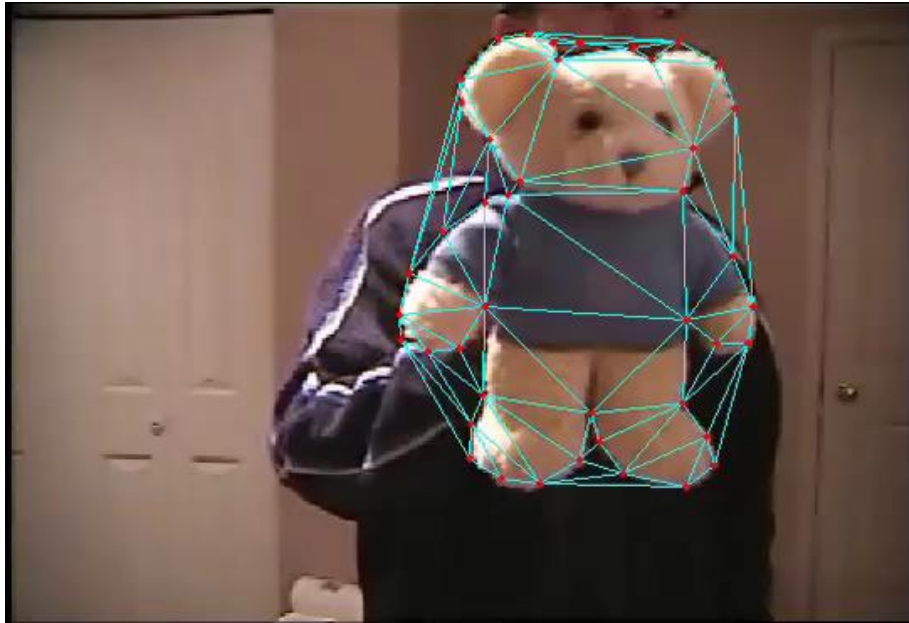
# Multimedia Processing and Analysis

- ◆ Automatic data analysis has to be done to extract semantic meanings from audios, images and videos
- ◆ Try to bridge the semantic gap (语义鸿沟) between low level features and high level semantic meaning
- ◆ Some media processing applications such as:
  - Object tracking (face, eyes)
  - Object recognition
  - Gesture (姿态) recognition, etc.

Based on them, we can build more effective human computer interfaces.



# Object Tracking



## Articulated Multi-person Tracking in the Wild

Eldar Insafutdinov, Mykhaylo Andriluka, Leonid Pishchulin,  
Siyu Tang, Evgeny Levinkov, Bjoern Andres, Bernt Schiele

Max Planck Institute for Informatics

# MOT Demo by Kalman Filter

## ❖ Link

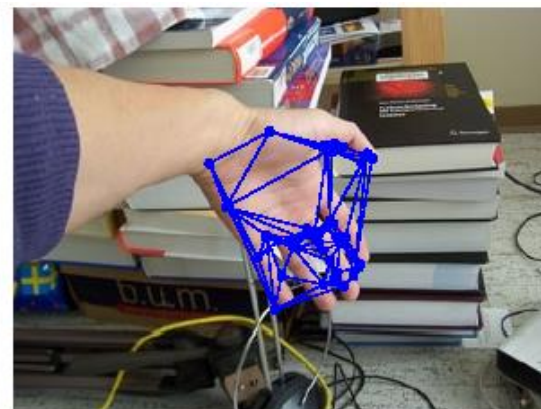
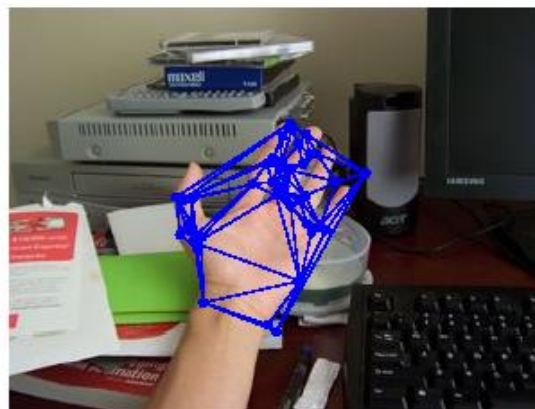
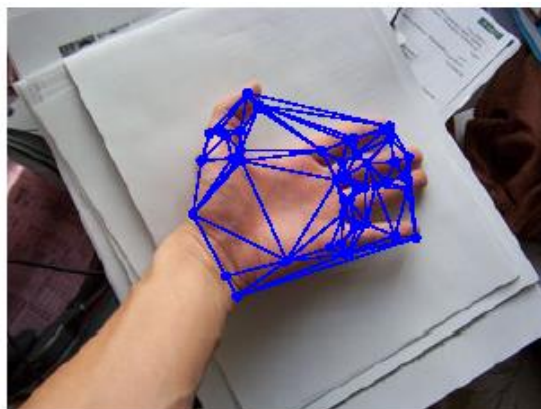
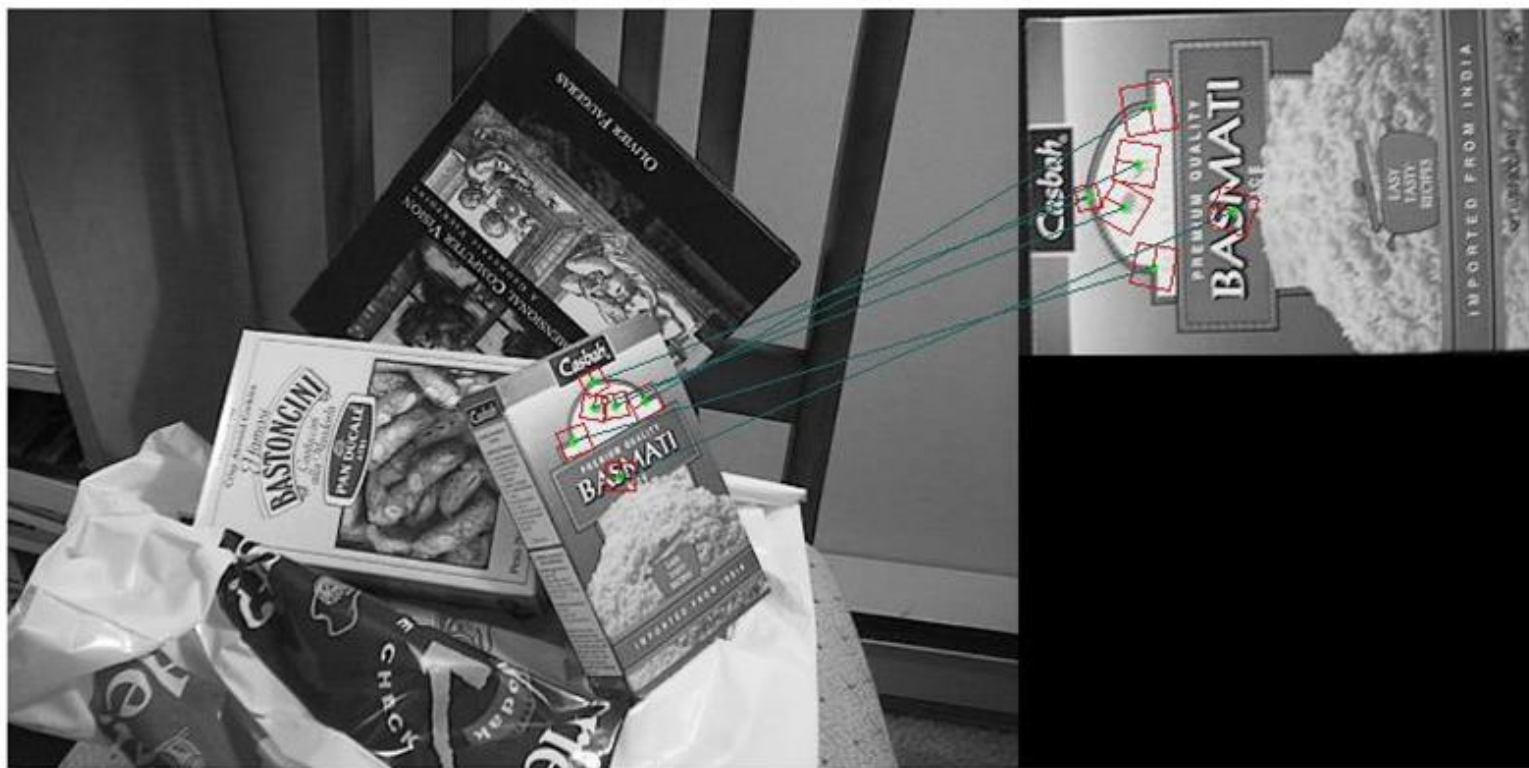
- <http://cn.mathworks.com/help/vision/examples/motion-based-multiple-object-tracking.html?requestedDomain=www.mathworks.com>

## ❖ Click: open this example

- Enter Matlab command:
- `openExample('vision/MotionBasedMultiObjectTracking Example')`

# Finding Objects in Clutter

Corresponding FP between image1 and image2



# Handwritten Recognition

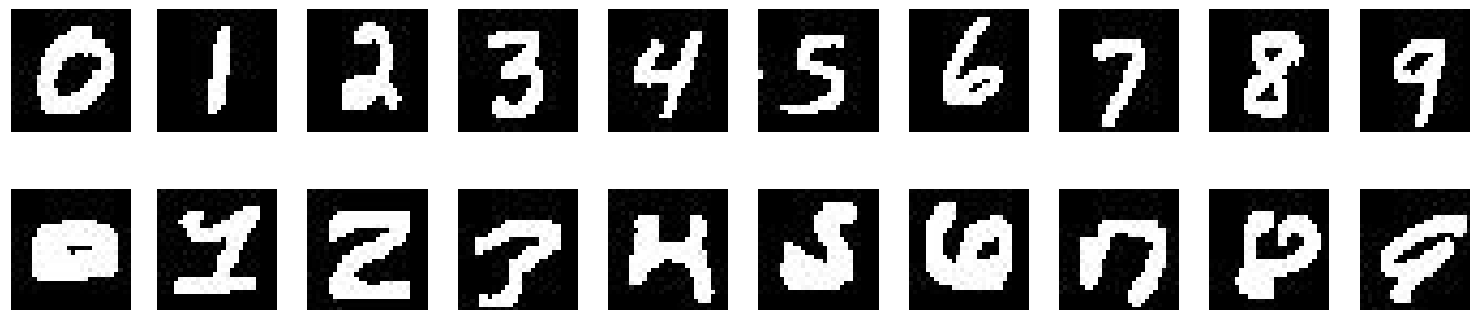


Fig. Examples of prototype images (the top row) and non-typical images (the bottom row) from MNIST.

## ◆ Link

- <http://yann.lecun.com/exdb/lenet/index.html>

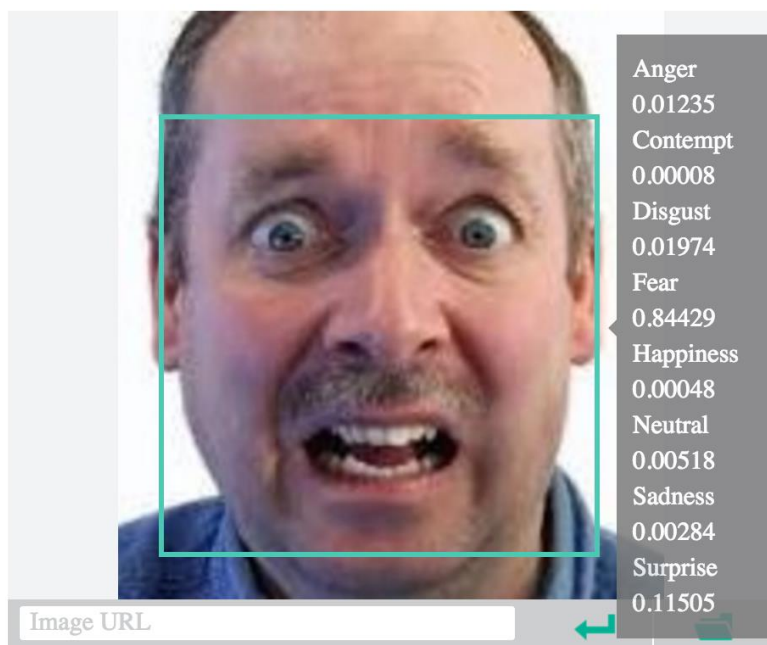
# Face Detection and Emotion Recognition



# Face Detection and Emotion Recognition

## ◆ Microsoft face API

<https://azure.microsoft.com/en-us/services/cognitive-services/face/#demo>



Anger	0.01235
Contempt	0.00008
Disgust	0.01974
Fear	0.84429
Happiness	0.00048
Neutral	0.00518
Sadness	0.00284
Surprise	0.11505

```
{
  "faceRectangle": {
    "left": 12,
    "top": 31,
    "width": 132,
    "height": 132
  },
  "scores": {
    "anger": 0.0123501224,
    "contempt": 0.00007971443,
    "disgust": 0.019741632,
    "fear": 0.844286442,
    "happiness": 0.000475379871,
    "neutral": 0.00517949,
    "sadness": 0.00283718458,
    "surprise": 0.11505001
  }
}
```

Help Us Improve



# Action Detection



Jogging: 慢跑

Boxing: 拳击

Hand clapping: 鼓掌

# Action Detection

- ◆ Dataset

<http://www.nada.kth.se/cvap/actions/>

- ◆ Code

[http://cn.mathworks.com/matlabcentral/fileexchange/47020-pictorial-suspicious-action-detection?s\\_tid=srchtitle](http://cn.mathworks.com/matlabcentral/fileexchange/47020-pictorial-suspicious-action-detection?s_tid=srchtitle)



# Virtual Reality (虚拟现实)



# Media Delivery

- ◆ Transmitting multimedia data across the network is another topic in multimedia computing
- ◆ Different issues when we transmit multimedia data through packet network
  - Quality of service
  - Synchronization (同步)
  - Error and congestion control (拥塞控制)
  - ...

# Multimedia Indexing and Retrieval

- ❖ Multimedia data is big data
- ❖ Multimedia data needs new data structures, indexing and searching methods
- ❖ Content-based multimedia retrieval is still an ongoing research topic

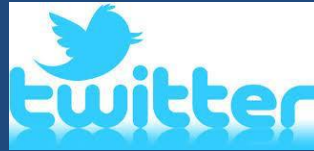
# Big Multimedia Data on Web

facebook®

142 Billion Images  
6 Billion added  
monthly



6 Billion Images



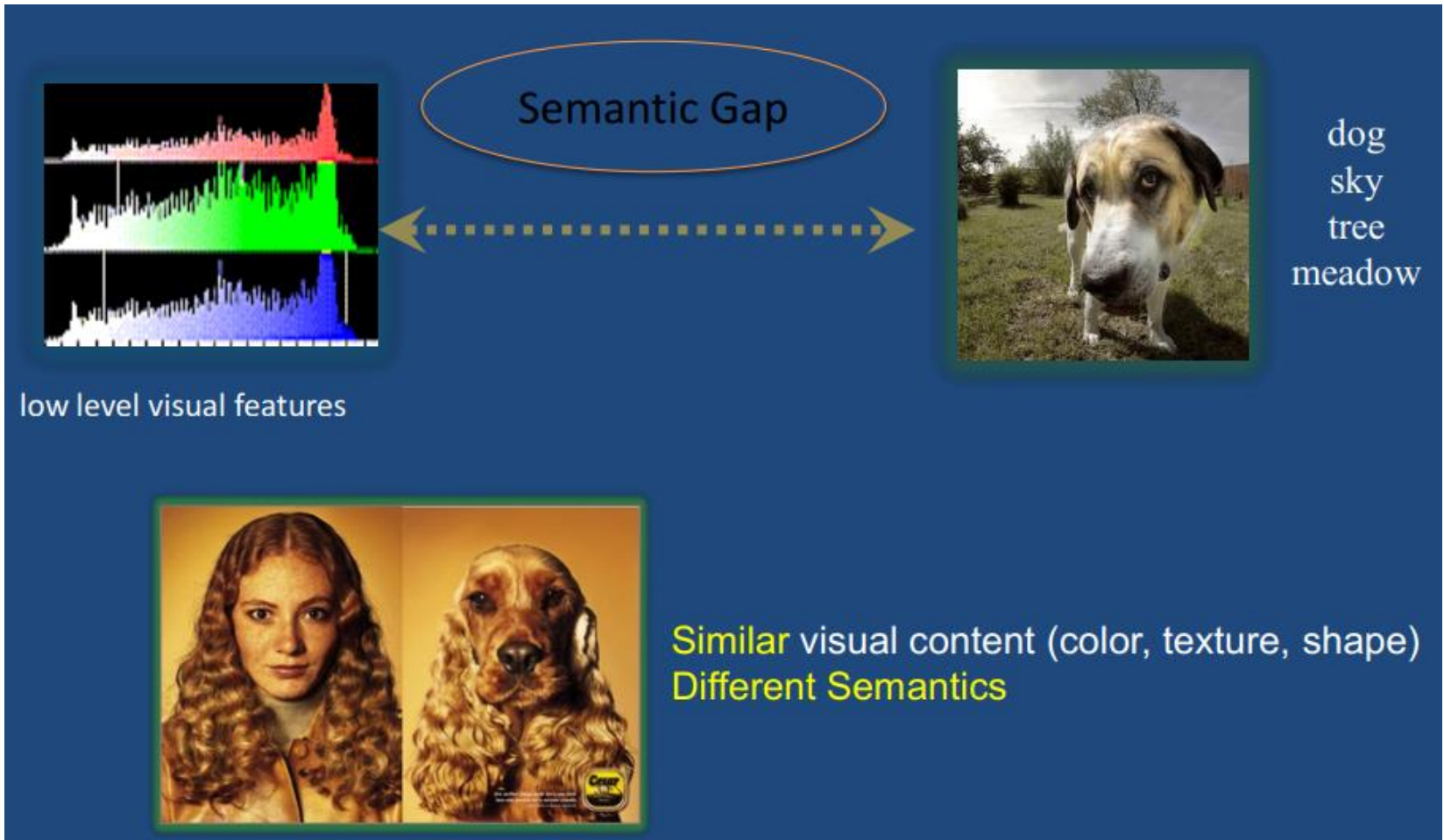
Instagram



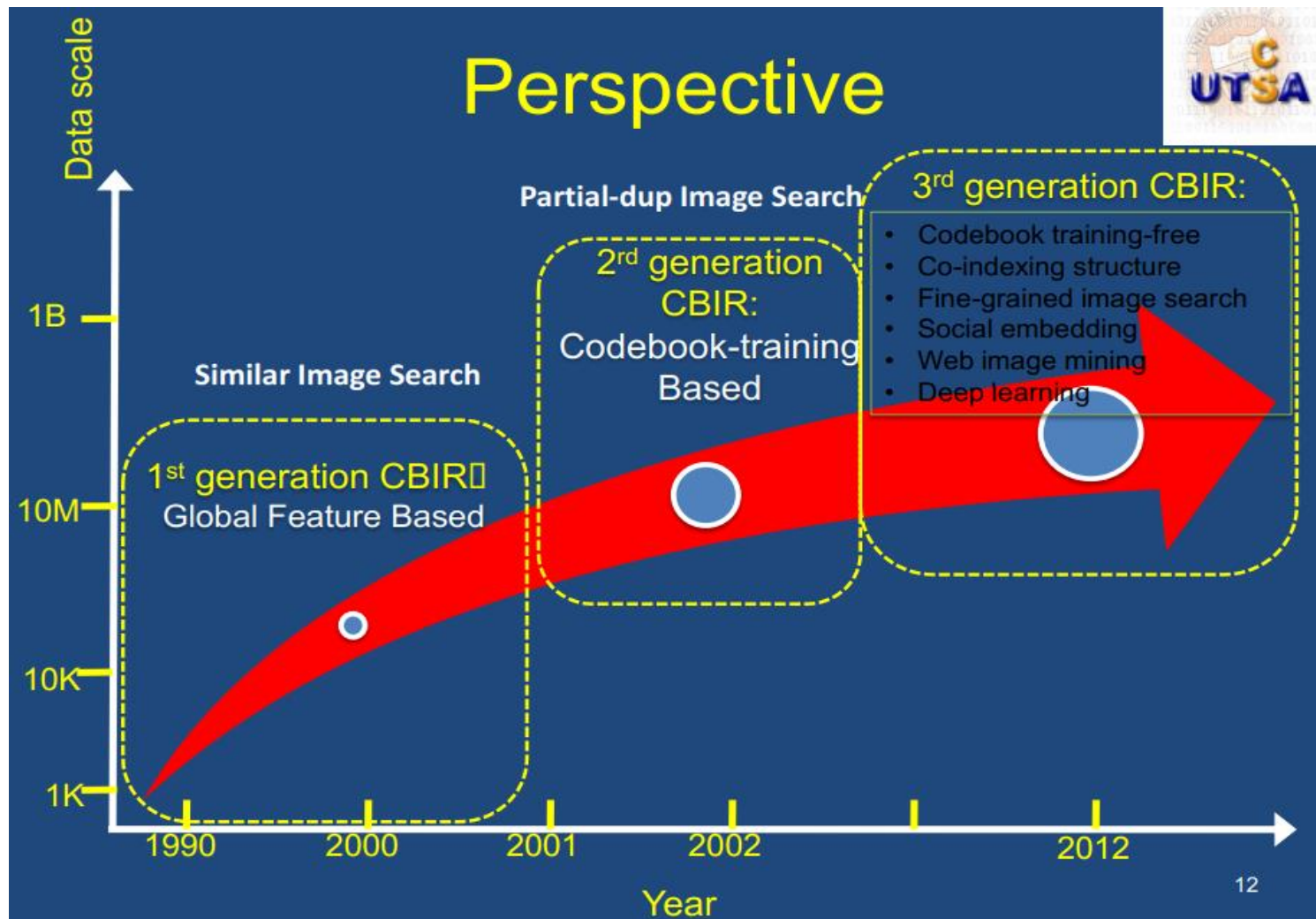
90% of Web data is visual!



# The Challenge of Semantic Gap



# Generations of CBIR





# Multimedia Authoring and Tools



Photoshop



Premiere



Audition



Flash



# Related Journals and Conferences

## ❖ Journals

- IEEE Transaction on Multimedia (TMM)
- IEEE Transaction on Image Processing (TIP)
- IEEE Transaction on Pattern Recognition and Machine Intelligence (TPAMI)
- ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)
- IEEE Multimedia
- Multimedia Tools and Applications (MTAP)

## ❖ Conferences

- ACM Annual International Conference on Multimedia (ACM MM)
- ACM Annual International Conference on Multimedia Retrieval (ICMR)
- International Conference on Multimedia and Expo (ICME)
- International Conference on Computer Vision (ICCV)
- IEEE Computer Vision and Pattern Recognition (CVPR)



## 思考题

- ❖ 远程教学系统是不是多媒体系统？
- ❖ 涉及哪些技术问题？
- ❖ 说出三种较有新意的多媒体应用。并给出你认为有新意的理由。