Lecture 12
Large-Scale Multimedia Analysis: MPEG Video and Visual Search

Guest Speaker: Wen-Hsiao Peng
National Chiao Tung University (NCTU), Taiwan

Ching-Yung Lin, Ph.D.
Adjunct Professor, Dept. of Electrical Engineering and Computer Science
IBM Chief Scientist, Graph Computing

Wen-Hsiao (Wen) Peng

- 2015 -- : **Visiting Scholar**, IBM T. J. Watson, New York, US
- 2006 -- : **Associate Professor**, Nat'l Chiao Tung Univ., CS Dept.
- 2005 : **Ph.D. in EE**, Nat'l Chiao Tung Univ., Taiwan
- 2013 -- : **IEEE Senior Member**
- 2009 -- : **Technical Committee Member**, IEEE CASS Visual Signal Processing and Communications (VSPC) & Multimedia Systems and Applications (MSA)
- 2003 -- : **ISO/IEC MPEG Delegate**, Taiwan Team Coordinator
- 2015 -- : **Lead Guest Editor**, IEEE J. Emerg. Sel. Topics in Circuits and Systems
- 2006 -- : **TPC Co-Chair/Member/Area Chair** for IEEE VCIP, ISCAS, ICME, etc.
- 2000 -- 2001: **Intel Microprocessor Research Lab**, Santa Clara, US

Outline

- **Part I – ISO/IEC Moving Picture Experts Group (MPEG)**
  - Background
  - Recent Milestones
  - Future Video, Machine Learning, and Media Internet-of-Things
- **Part II – MPEG Compact Descriptor for Visual Search (CDVS)**
  - Large-scale Image Retrieval
  - Local Image Descriptors
  - Global Image Descriptors
  - Image Matching
  - Use Case: Mobile Indoor Navigation
- **Part III – Cross-domain Data Retrieval**
  - Canonical Correlation Analysis
  - Deep Boltzmann Machine

Outline

- **Part I – ISO/IEC Moving Picture Experts Group (MPEG)**
  - Background
  - Recent Milestones
  - Future Video, Machine Learning, and Media Internet-of-Things
- **Part II – MPEG Compact Descriptor for Visual Search (CDVS)**
  - Large-scale Image Retrieval
  - Local Image Descriptors
  - Global Image Descriptors
  - Image Matching
  - Use Case: Mobile Indoor Navigation
- **Part III – Cross-domain Data Retrieval**
  - Canonical Correlation Analysis
  - Deep Boltzmann Machine

What would you do on a self-driving car?

https://youtu.be/hvqeLjVLcAc

Standards Organizations

- ISO/IEC **Moving Picture Experts Group** (**MPEG**)
  - MPEG-1, MPEG-2, MPEG-4, MPEG-4 AVC, MPEG-H HEVC
- ITU-T **Video Coding Experts Group** (**VCEG**)
  - H.261, H.263, H.264, H.265
- Joint Collaborative Team on Video Coding (**JCT-VC**)

ISO – International Standardization Organization

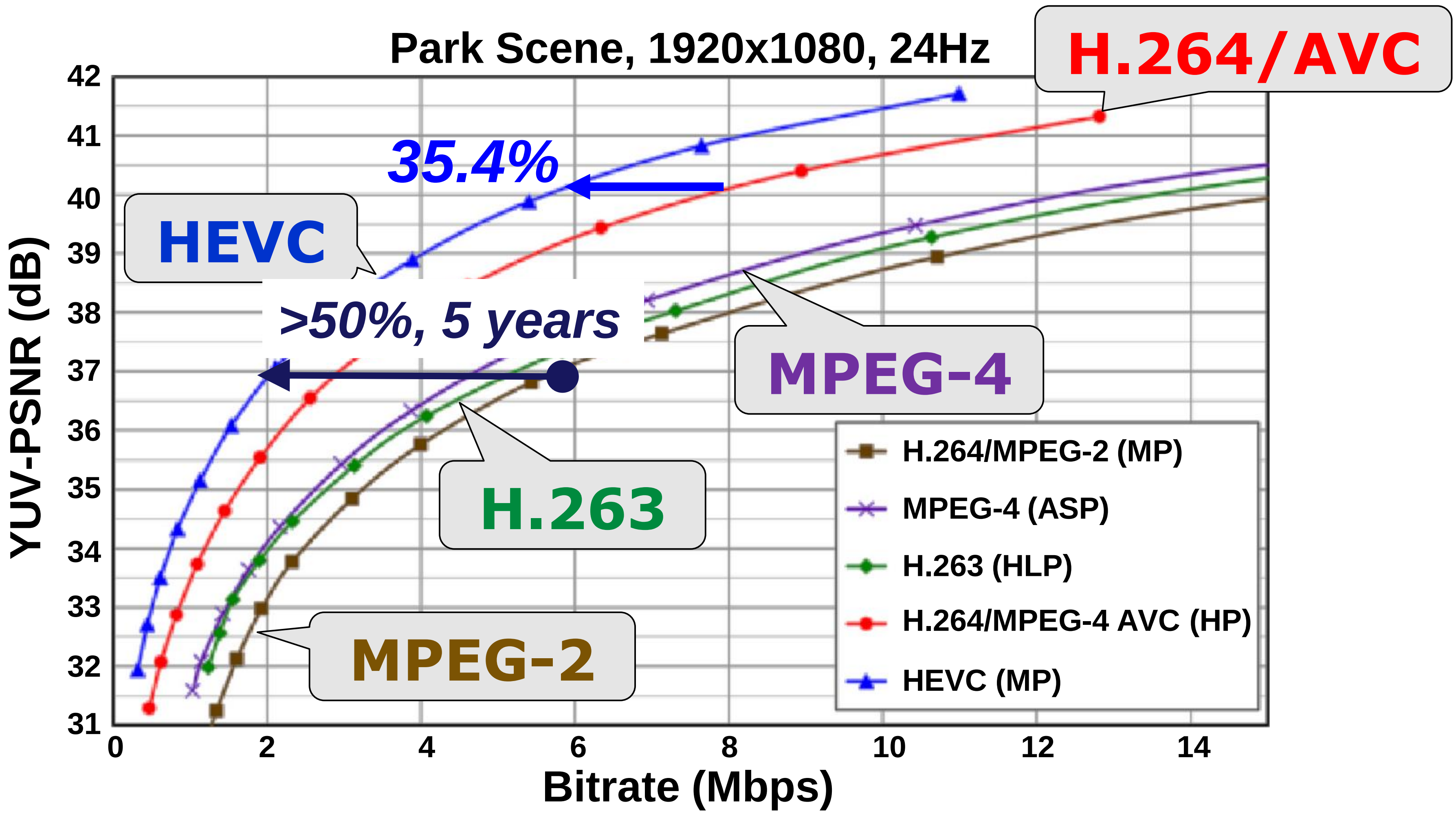
IEC – International Electrotechnical Commission

ITU – International Telecommunication Union

Progress of Video Coding Standards

- H.261 (CCITT/ITU;1984, 88, 90) – videoconf.
- MPEG-1 (1988 -- **92**) – VCD
- MPEG-2 (1990 -- **94**) – DVD, DTV
- MPEG-4 Part 2 (1992 -- 99) – Internet, WL
- H.263 (1993 -- 95; ver.3: 2000) – WL
- AVC/H.264 (1998 -- **03**) – WL, HD-DVD
- AVC Amd. (2003 -- 2007) – Scalable Video Coding
- AVC Amd. (-- 2008) – Multiview Video Coding
- HEVC/H.265 (2010 - **13**) – Ultra-HD Video
- HEVC Amd. (2014 - **16**) – Screen Content Coding
- **Next-Generation Video Coding (2016 – 2020??) – HQ-OTT, VR, …**

Coding Efficiency Improvement



J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, “Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)”, IEEE Trans. CSVT, Dec., 2012

Moving Picture Experts Group (MPEG)

A working group of ISO/IEC founded in **1988** with the mission to develop **standards for coded representation of digital audio and video** and related data

- MPEG-1 (VCD)
- MPEG-2 (DVD, DVB)
- MPEG-4 AVC (Blu-ray, DVB, Smartphone)
- MPEG-H HEVC (also known as H.265)
- MPEG-A, B, C, V, …

Convener: **Dr. Leonardo Chiariglione**

- 1996 Emmy for Technical Excellence
- 2008 ATAS Primetime Emmy Award
- 2009 NATAS Tech & Eng Emmy Award

MPEG Basics

- **4** meetings (**5-10 days**) per year
- **350+** experts; **200+** companies; **20+** countries
- 200+ (sometimes, 1000+) input documents
- Meetings divided into subgroups with plenaries on Mon/Wed/Fri

WG 11
MPEG Committee

COLUMBIA UNIVERSITY

Brief History

- 1st meeting, 1988

  Hiroshi Yasuda &

  Leonardo Chiariglione

  (Ottawa, Canada)

… 25 years …

- 100th meeting, 2012

  Leonardo Chiariglione

  (Geneva, Switzerland)

Proc. IEEE, April 2012

INVITED PAPER

Multimedia Standards: Interfaces to Innovation

A history of the Motion Picture Experts Group is provided and its probable future activities are discussed, including understanding 3-D audio–video, machine design, and creating best practices and models.

By Leonardo Chiariglione

ABSTRACT | Standardization is concerned with interfaces; industry is concerned with systems. This papers brings the evidence brought by the MPEG standardization group to show how, through the proper management of interface evolution, the constituent industries have been able to achieve product and service interoperability, room for differentiation, and opportunities for innovation in the context of the tectonics shift also known as convergence.

KEYWORDS | Compression; convergence; digital media; standards

same "standard" if C is ever to be able to have bolts that screw into the nut. In the case of nuts and bolts, the standard is the "thread" which represents the interface between the nut and the bolt.

Standards play a fundamental role in enabling a diversified industry. Once a standard has been published—and the industry has adopted it—independent manufacturers can build products conforming to it that can immediately reach a potentially global market. Users can choose products that are more convenient for their needs from different suppliers.

There is also a prevailing view that standards choke

Products with MPEG Technologies

The MPEG Process

1. **Exploration**
   Search for new technology
2. **Requirements**
   Establish work scope
   Call for Proposals (CfP)
3. **Competitive phase**
   Do Homework
   Response to CfP
   Initial technology selection
4. **Collaborative phase**
   Core Experiments
   Working Drafts
5. **Standardization**
   Ballots
   National Body Comments
6. **Amendment**
   Adding new technology
7. **Corrigenda**
   Corrective actions
8. **New subdivisions**
   Add new non-compatible technology

Standardization Scope: Decoder

(1) Bitstream structure

(2) Syntax
(e.g. slice_type)

(3) Semantics

Table 7-7 – Name association to slice_type

slice_type	Name of slice_type
0	B (B slice)
1	P (P slice)
2	I (I slice)

(4) Decoding process
(e.g. tasks to perform when slice_type=0)

slice_segment_header( ) {
first_slice_segment_in_pic_flag
if(nal_unit_type >= BLA_W_LP && nal_unit_type <= RSV_IRAP_VCL23)
no_output_of_prior_pics_flag
slice_pic_parameter_set_id
if(!first_slice_segment_in_pic_flag) {
if(dependent_slice_segments_enabled_flag)
dependent_slice_segment_flag
slice_segment_address
}
if(!dependent_slice_segment_flag) {
for(i = 0; i < num_extra_slice_header_bits; i++)
slice_reserved_flag[i]
slice_type
if(output_flag_present_flag)
pic_output_flag
if(separate_colour_plane_flag = = 1)
colour_plane_id
if(nal_unit_type != IDR_W_RADL && nal_unit_type != IDR_N_LP) {
slice_pic_order_cnt_lsb
short_term_ref_pic_set_sps_flag
if(!short_term_ref_pic_set_sps_flag)
short_term_ref_pic_set(num_short_term_ref_pic_sets)
else if(num_short_term_ref_pic_sets > 1)
short_term_ref_pic_set_idx

Outline

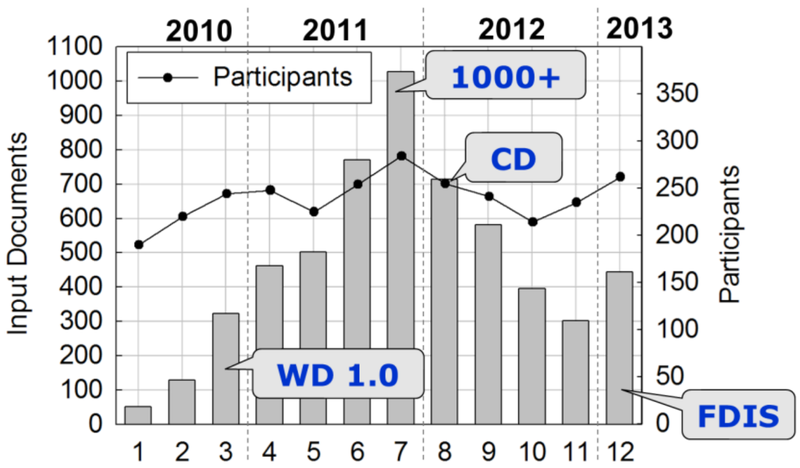
- **Part I – ISO/IEC Moving Picture Experts Group (MPEG)**
  - Background
  - Recent Milestones
  - Future Video, Machine Learning, and Media Internet-of-Things
- **Part II – MPEG Compact Descriptor for Visual Search (CDVS)**
  - Large-scale Image Retrieval
  - Local Image Descriptors
  - Global Image Descriptors
  - Image Matching
  - Use Case: Mobile Indoor Navigation
- **Part III – Cross-domain Data Retrieval**
  - Canonical Correlation Analysis
  - Deep Boltzmann Machine

High Efficiency Video Coding (HEVC)

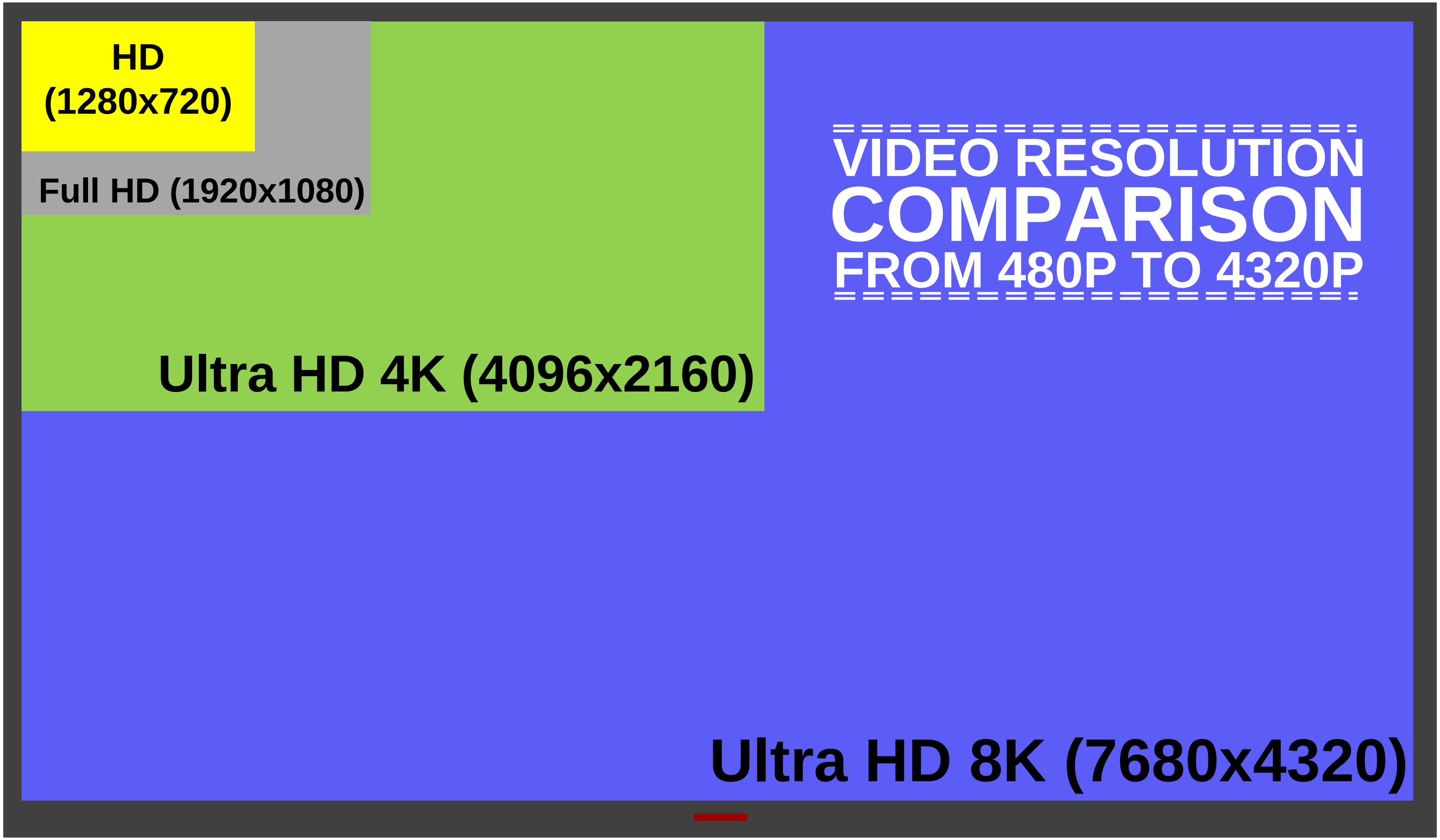
- The latest video coding standard developed by a joint team of experts from MPEG and VCEG

- **Objective:** Offer substantially better performance than the state-of-the-art AVC/H.264, especially in coding **HD and Ultra-HD** video (e.g. 4k or 8k video)

- International Standard, Apr. 2013
  - Exploration started in 2005
  - Call for Proposals issued in 2010
  - 27 proposals received
  - 5600+ contributions in 3 years



Ultra HD vs. HD



8kx4k 10-bit Sequences

Provided by NHK, Japan

Nebuta Festival
300 frames, 60 fps

Steam Locomotive Train
300 frames, 60 fps

Raw Data ~ 30Gb/s

(SSD Read ~ 4.4Gb/s; HDMI ~ 18Gb/s)

Participants of HEVC's CfP

Columbia University

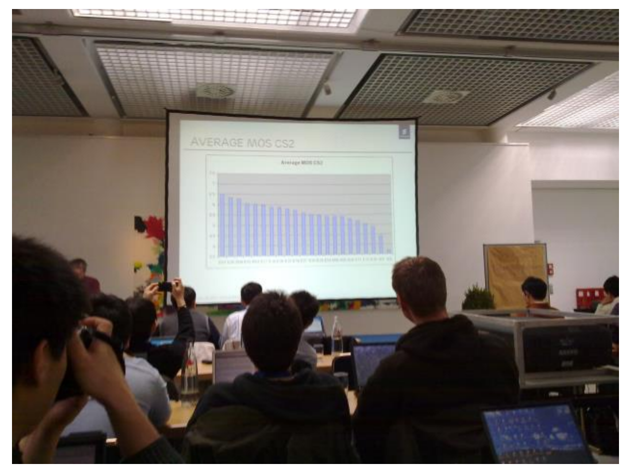
TI & MIT
Hitachi
Sony
NEC
Sharp
Intel
Mitsubishi
JVC
MediaTek
LG
Huawei & Hisilicon

RWTH Aachen
SK telecom, Sejong Univ. & Sungkyunkwan Univ.
France Telecom, NTT, NTT DOCOMO, Panasonic & Technicolor
Fujitsu
Fraunhofer HHI
Toshiba
Microsoft
Tandberg, Ericsson & Nokia
RIM
Qualcomm

NHK & Mitsubishi
NCTU
Samsung & BBC
BBC & Samsung
Renesas
ETRI

HEVC's CfP by the Numbers

- **27** proposals (the highest in MPEG history)
- **145** test cases (taking 2 weeks to complete 1 round of simulation using 120 CPU cores)
- **800** human subjects to rate all proposals
- **$10K** per proposal for subjective evaluation



1st JCT-VC Meeting in Dresden, Apr. 2010

HEVC Version 1

- **Major milestone in MPEG video history**

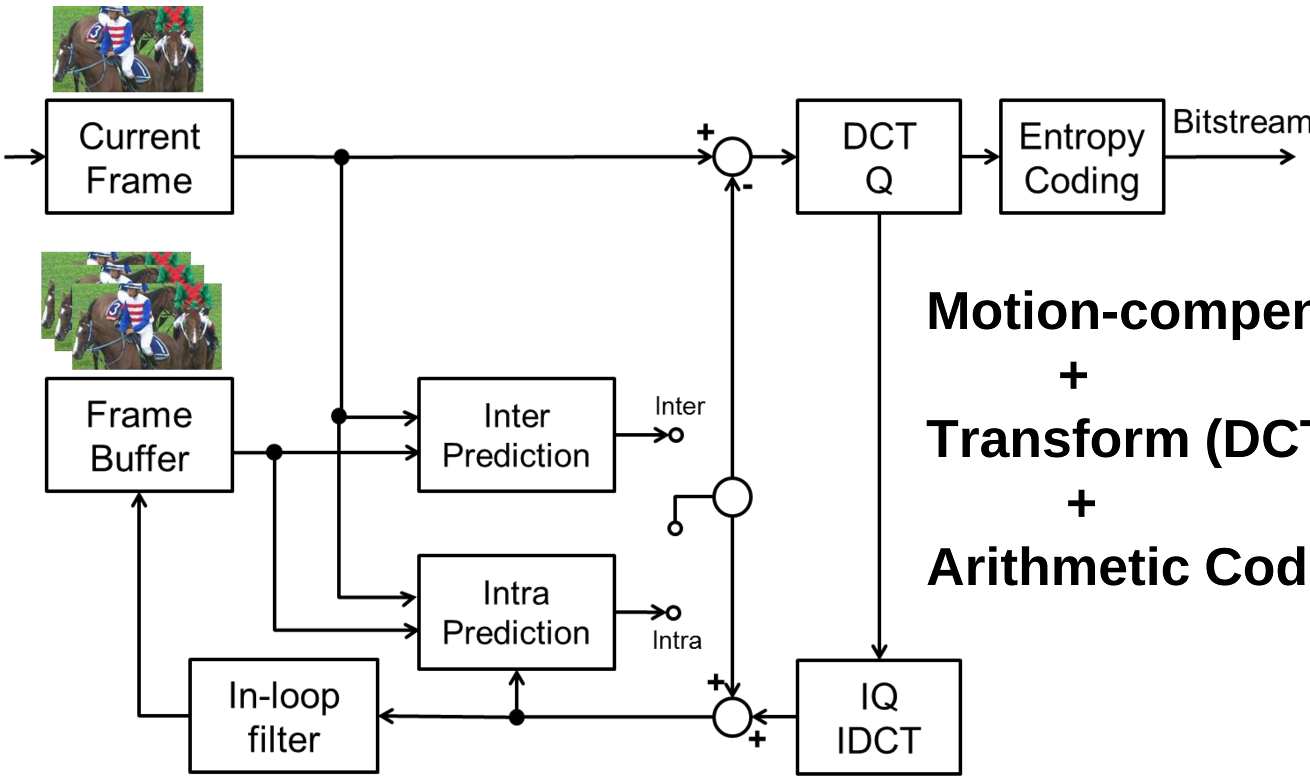
  Press letter of 103rd Geneva Meeting (N13253) –

  "ISO/IEC JTC1/SC29/WG11 MPEG is proud to announce the completion of the new High Efficiency Video Coding (HEVC) standard which has been promoted to Final Draft International Standard (FDIS) status at the 103rd MPEG meeting."

- **Officially International Standard (IS) since Apr. 2013**

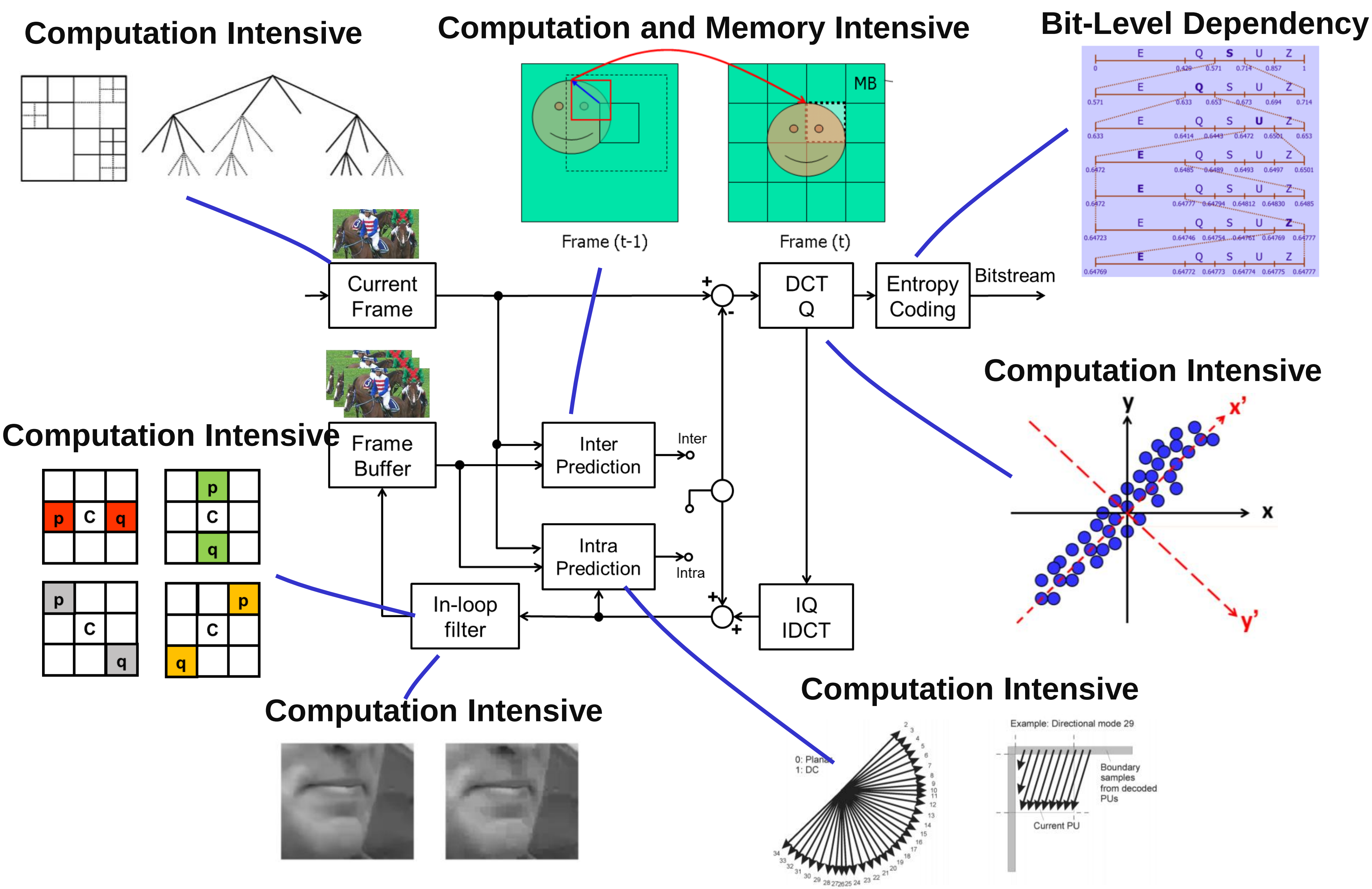
Technology – Block-based Hybrid Coding

- Same architecture/pipeline as prior standards, yet with
  - New elements proposed
  - Existing elements and building blocks re-designed
  - Careful considerations given to **parallel processing**



Motion-compensated Prediction
+
Transform (DCT)
+
Arithmetic Coding

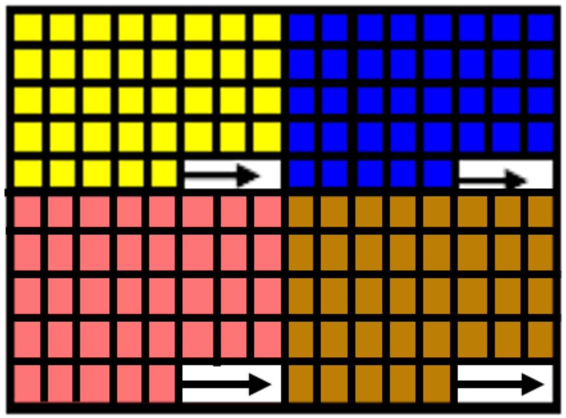
Video Coding Concept



Parallel Processing

Tiles:

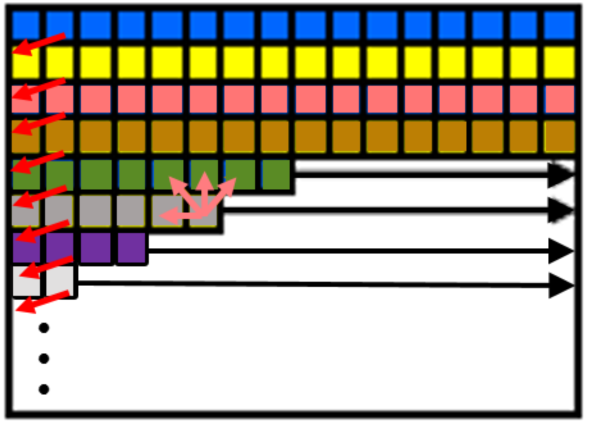
Independently decodable regions



[4 Tiles in a Slice/Picture]

Wavefront:

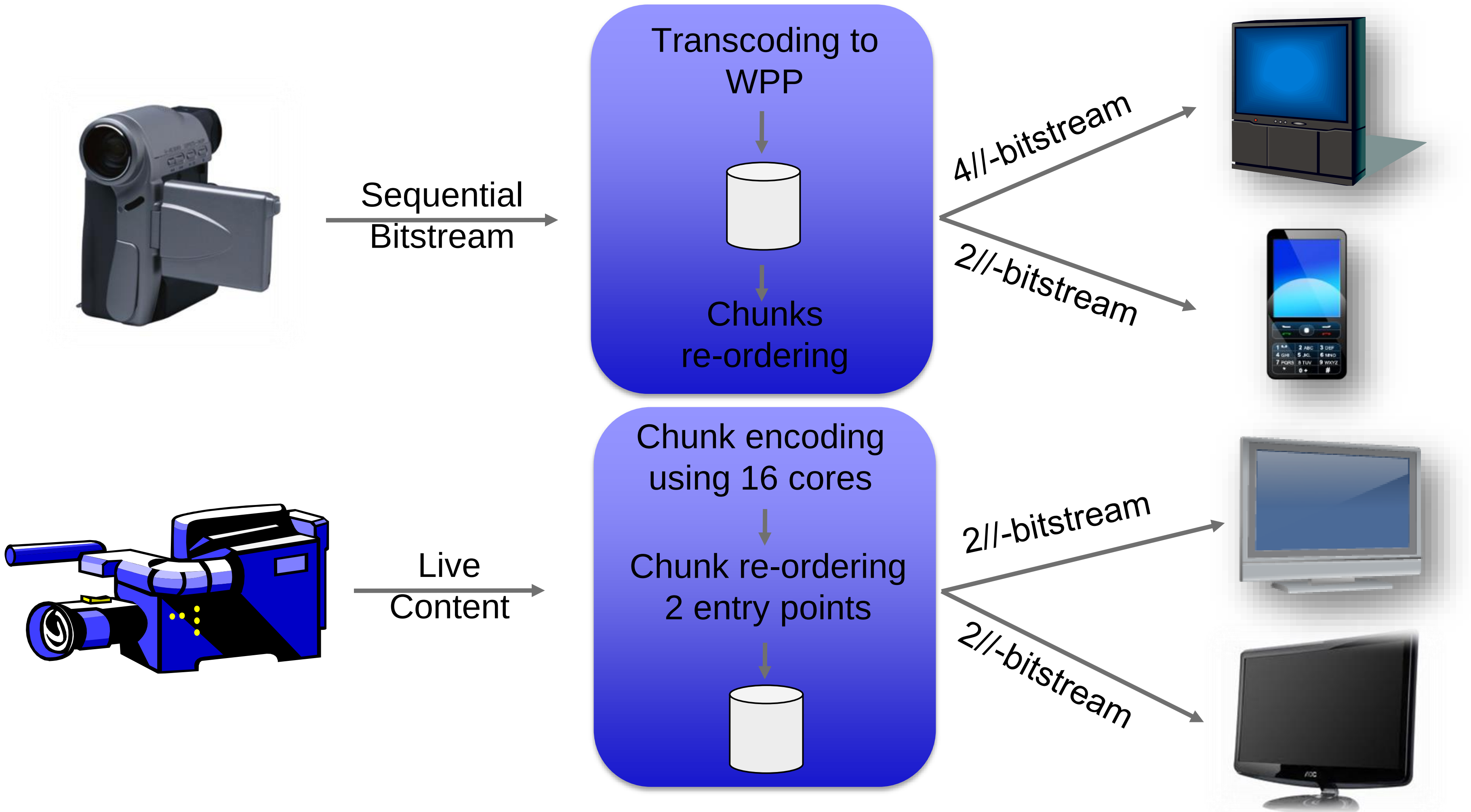
Parallel CTU rows processing



[n Waves in a Slice]

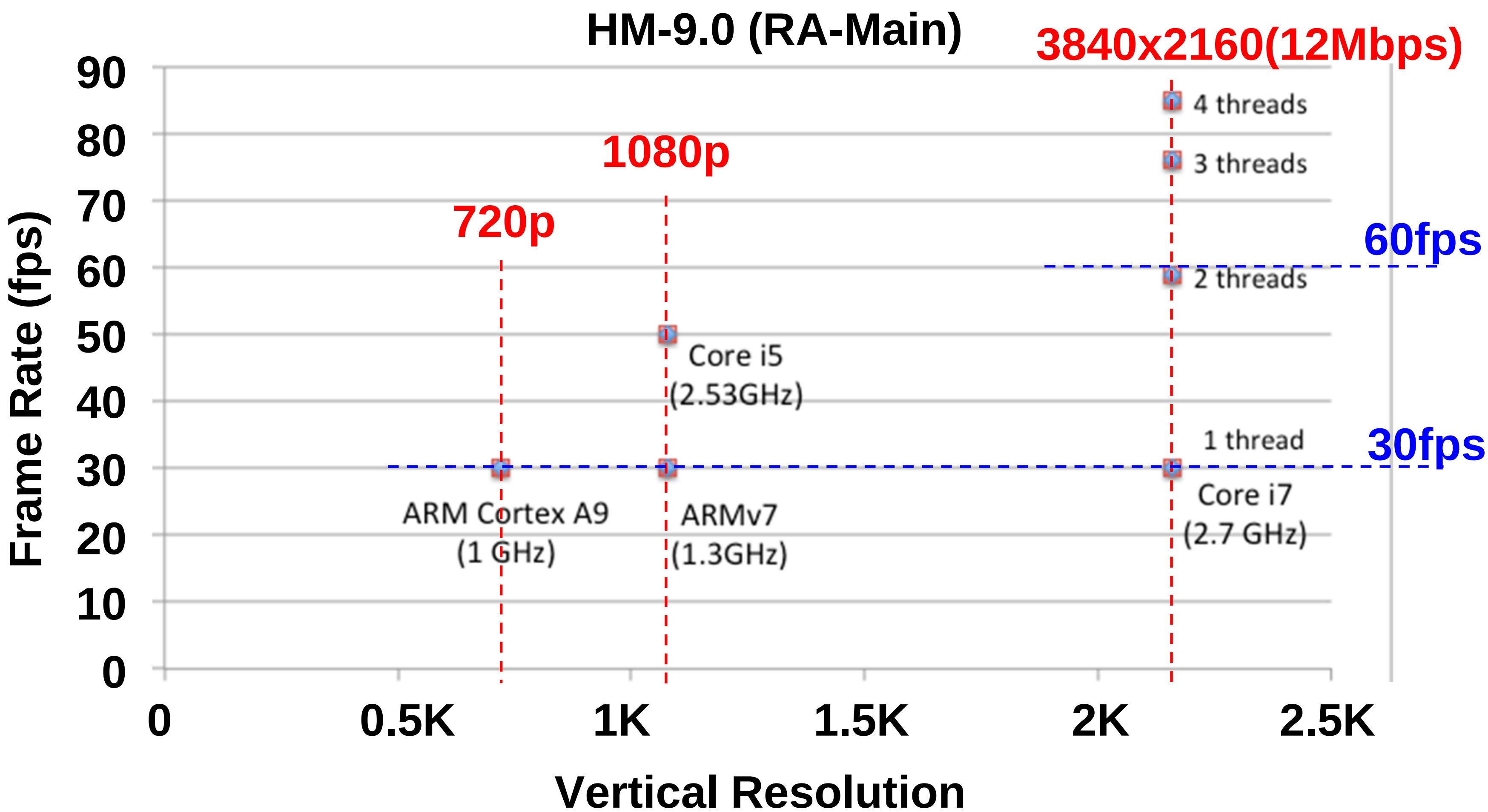
	Tiles	Wavefront
Parsing	Independent	Dependent
Reconstruction	Independent	Dependent
Granularity	Coarse (Regions)	Fine (CTU Rows)

Parallel-friendly Design



G. Clare, F. Henry, and S. Pateux,"Wavefront and Cabac Flush: Different Degrees of Parallelism Without Transcoding", JCTVC-F275, Torino, IT, July, 2011.

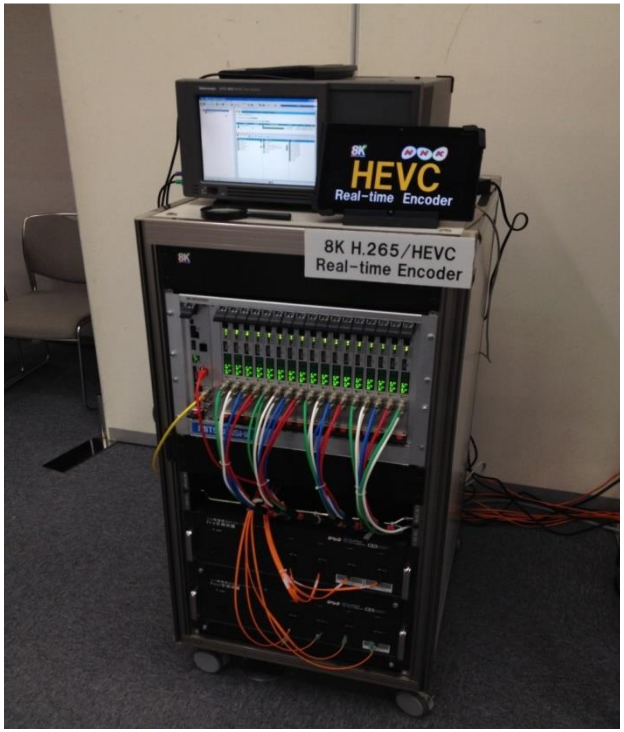
Multi-thread Decoding

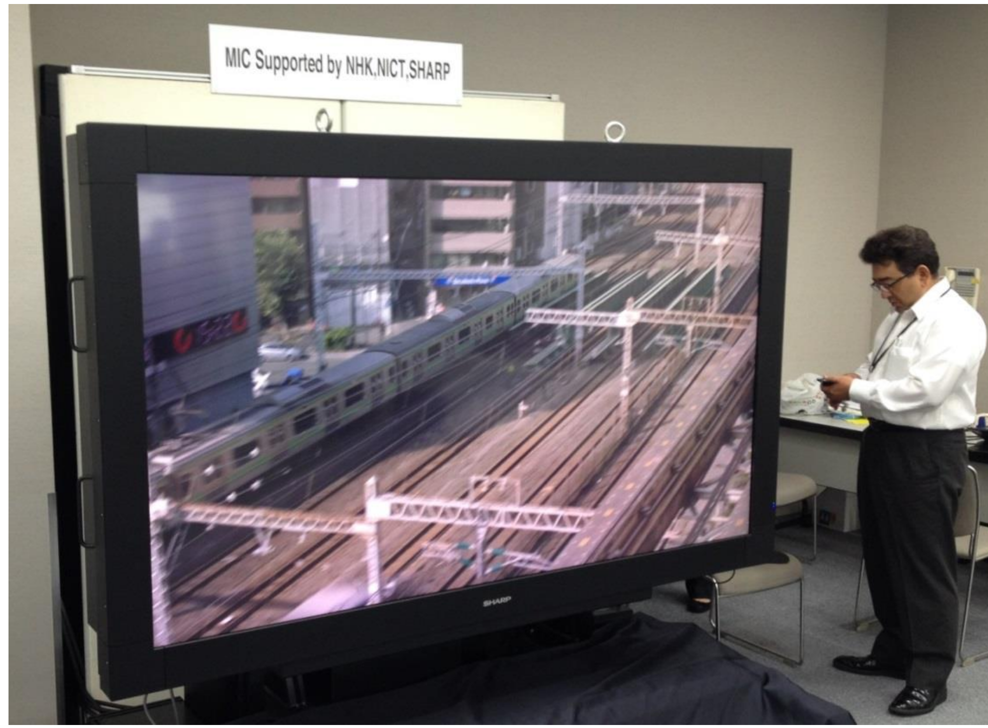


T. Tan, Y. Suzuki, and F. Bossen, "On software complexity: decoding 4K60p content on a laptop," JCTVC-L0098, Geneva, CH, Jan., 2013.

Experimental 8K HEVC Real-time Encoder

- Real-time 8K 10-bit video encoding at 60fps & 85Mbps

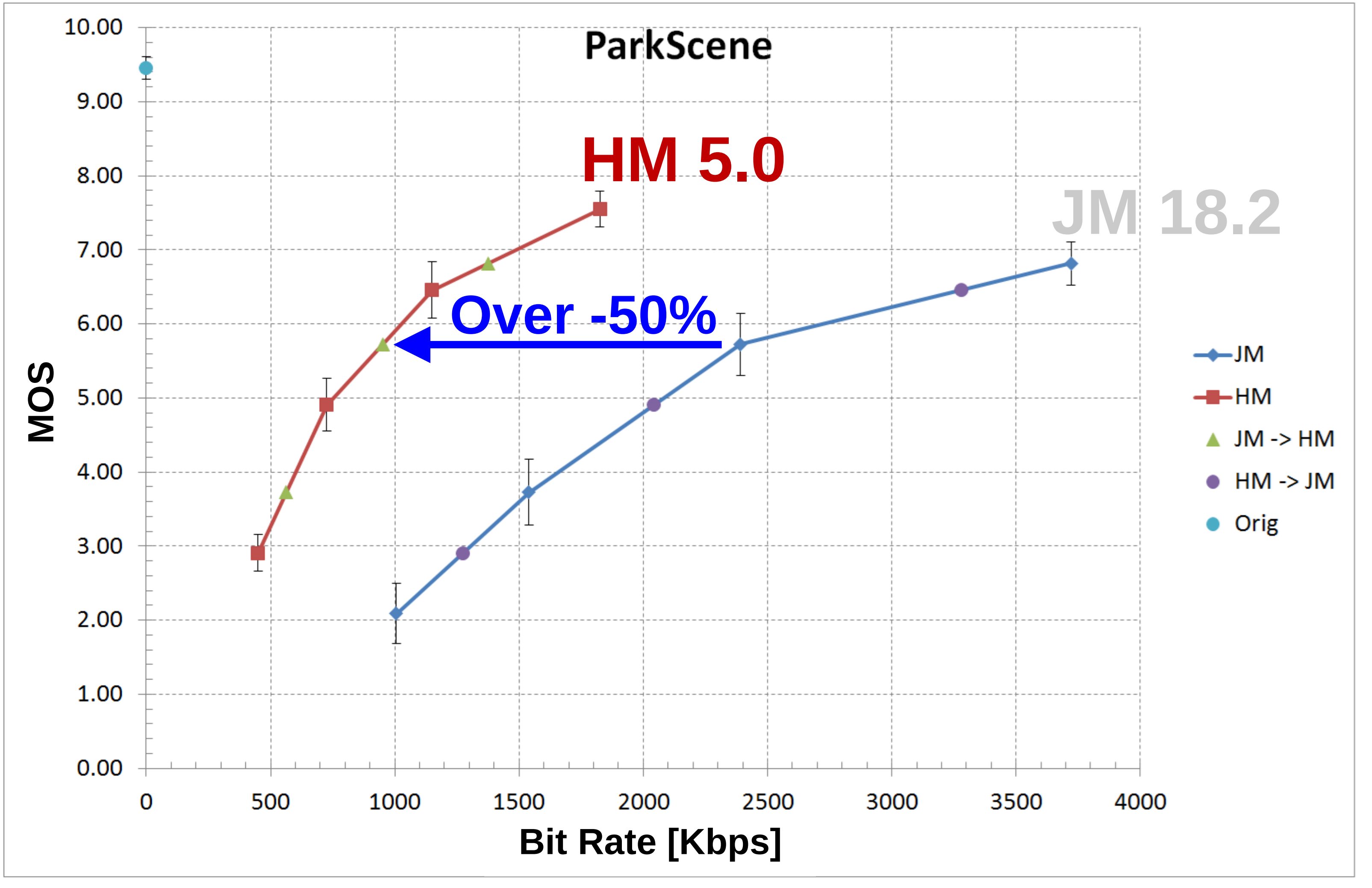




(Live Demo at 109th MPEG, Sapporo, June 2014)

Subjective Assessment

Columbia University



J. R. Ohm, G. J. Sullivan, F. Bossen, T. Wiegand, V. Baroncini, M. Wien, and J. Xu,"JCT-VC AHG report: HM subjective quality investigation (AHG22)", JCTVC-H0022, San José, CA, Feb., 2012.

Seeing is believing …

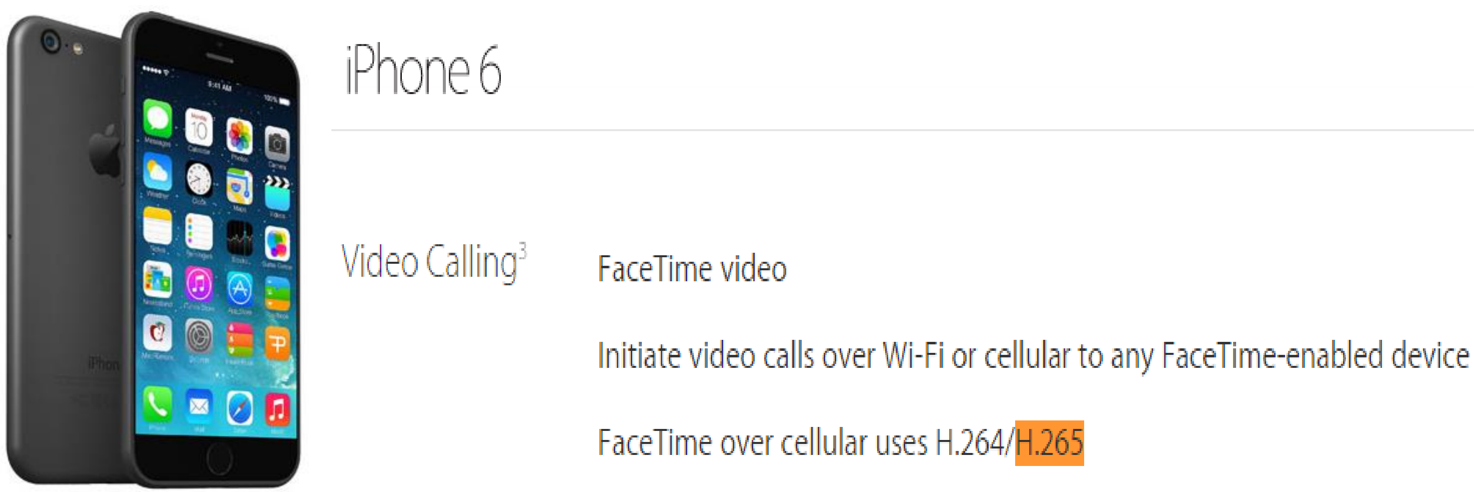
AVC/H.264

HEVC/H.265

Basketball Drive: 832x480_30Hz @ 1Mbps
Compression ratio ~144

Mass Adoption?

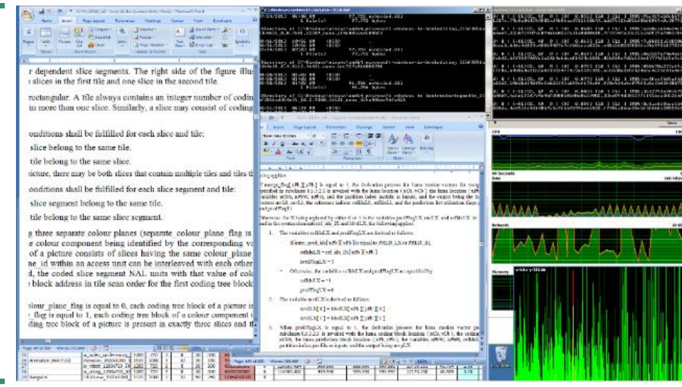
- Appeared on **few devices** and in **trial services**
- **Mass adoption has yet to occur** -- waiting for content providers to switch over

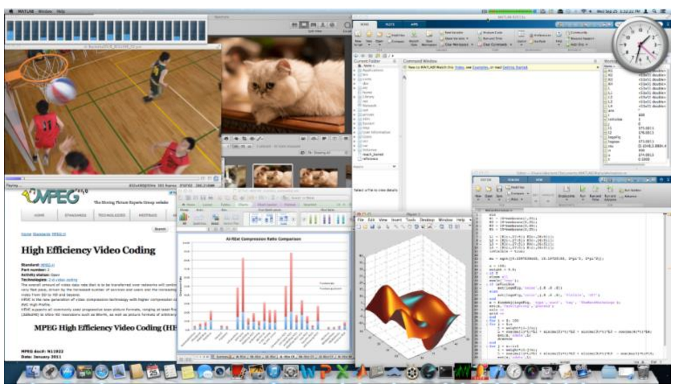


Screen Content Coding Extensions

- **Goal:** Enhance HEVC's capability in coding **screen content**
- Screen content coding — encoding screen visuals in the form of video and treating **text/graphics as pixel data**

Computer graphics and text with motion





Mixture of natural video and graphics/text

Computer-generated animation content

- Wireless display, **cloud gaming**, desktop sharing and collaboration, PC-over-IP, etc.

Desktop Collaboration

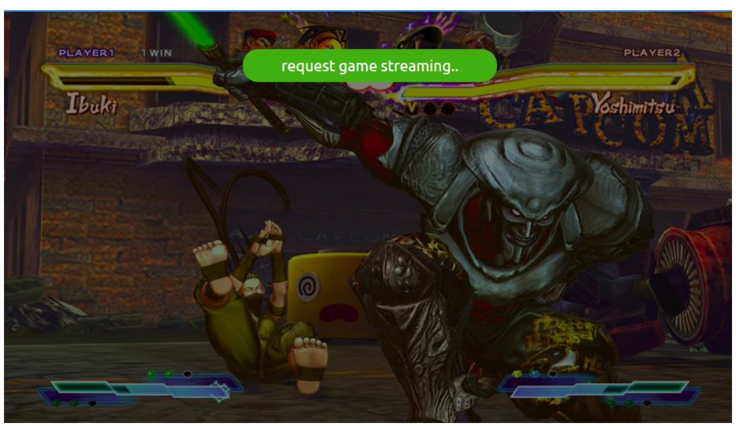
Cloud Gaming

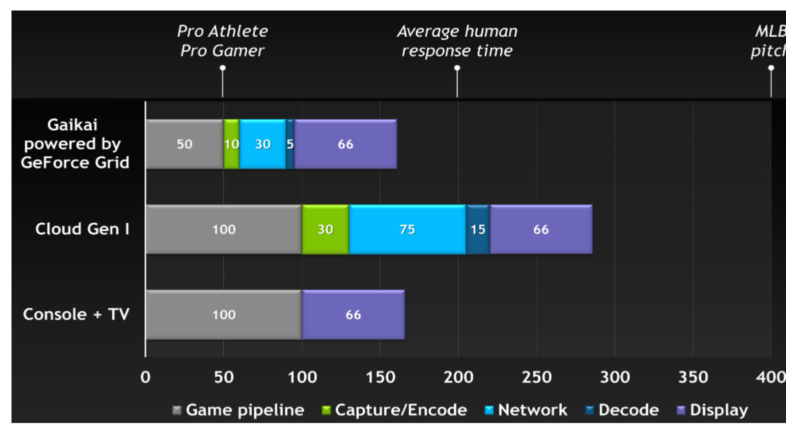
Second Screen

Screen Sharing

Challenges

- Mixture of contents with distinct attributes
  - **Text/graphics:** noise-free, discrete-tone, sharp edges
  - **Natural images:** noisy, continuous-tone, complex texture
- Varied level of distortion sensitivity in different types of content
  - Compression **artifacts in synthetic areas** easily visible
- Usually stringent low-delay requirements **(<300ms round trip**)
  - **Cloud gaming**, screen sharing, etc.

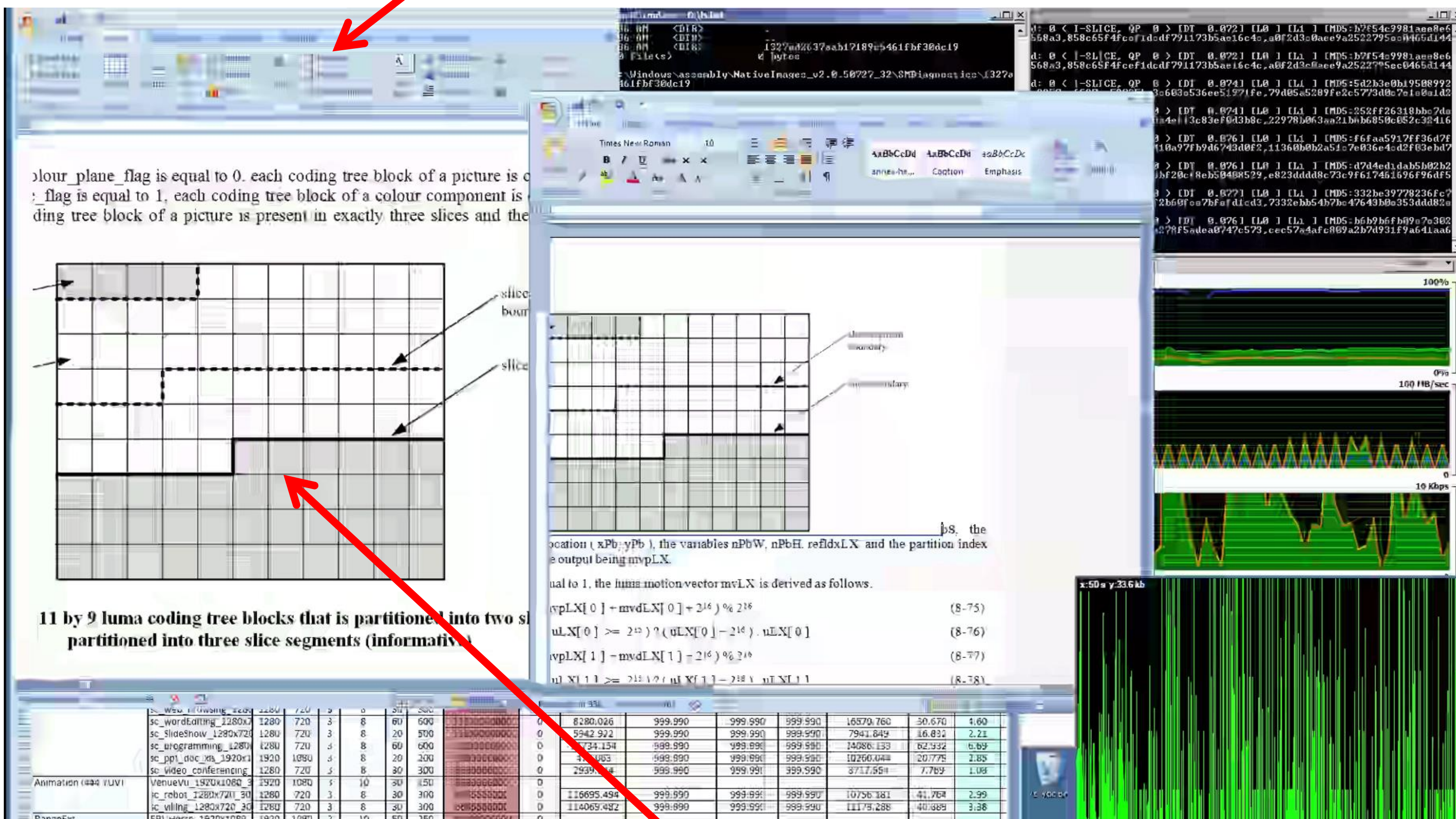




Firefox, http://www.ugamenow.com/

Result with hybrid-based codecs

Illegible details



Ringing artifacts

1920x1080@60Hz: 43Mbps (All Intra)
PSNR-Y= 28dB

SCC Call-for-Proposals

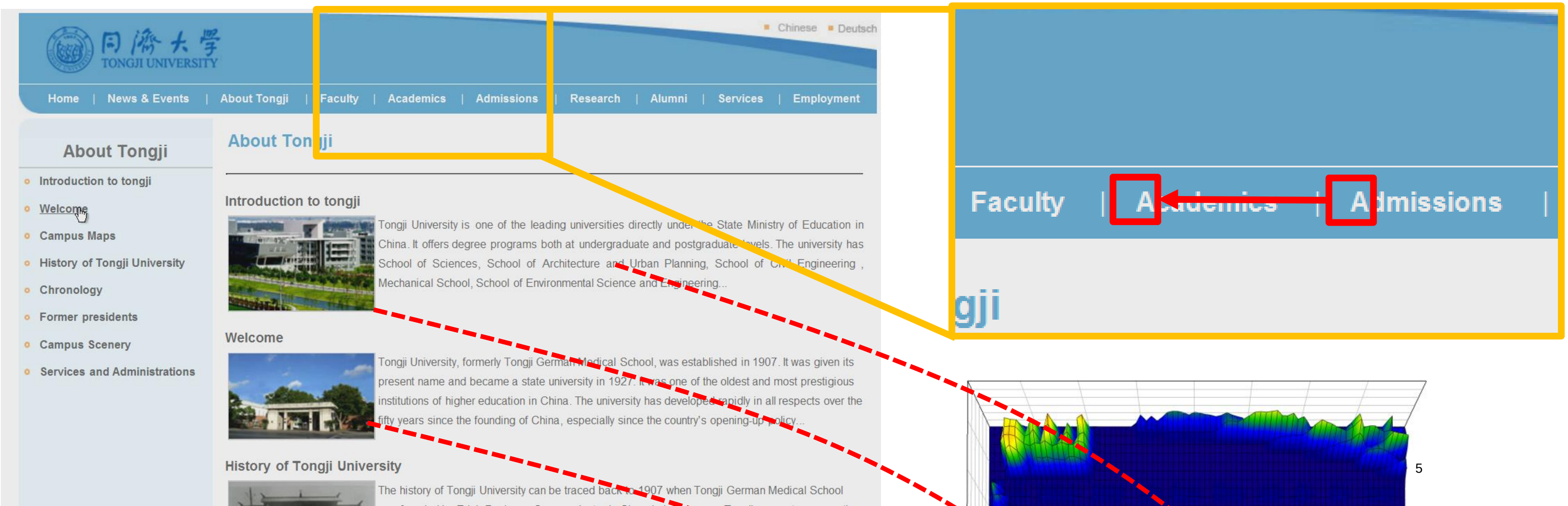
- MPEG 108 @ Valencia, Spain (**Apr. 2014**)

- Participants (7) — Qualcomm, Microsoft, MediaTek, InterDigital, Huawei, Mitsubishi, and **NCTU**

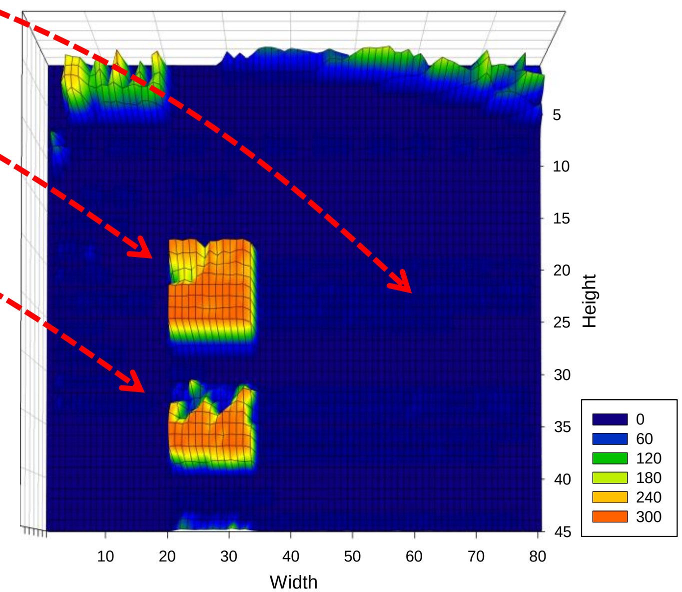
- 25-35% rate reductions over HEVC
  - Intra block/line copy
  - Palette mode
  - String matching
  - Adaptive in-loop color transform
  - Encoding optimizations

Principles behind Screen Content Coding

(1) Some patterns often **repeat** themselves



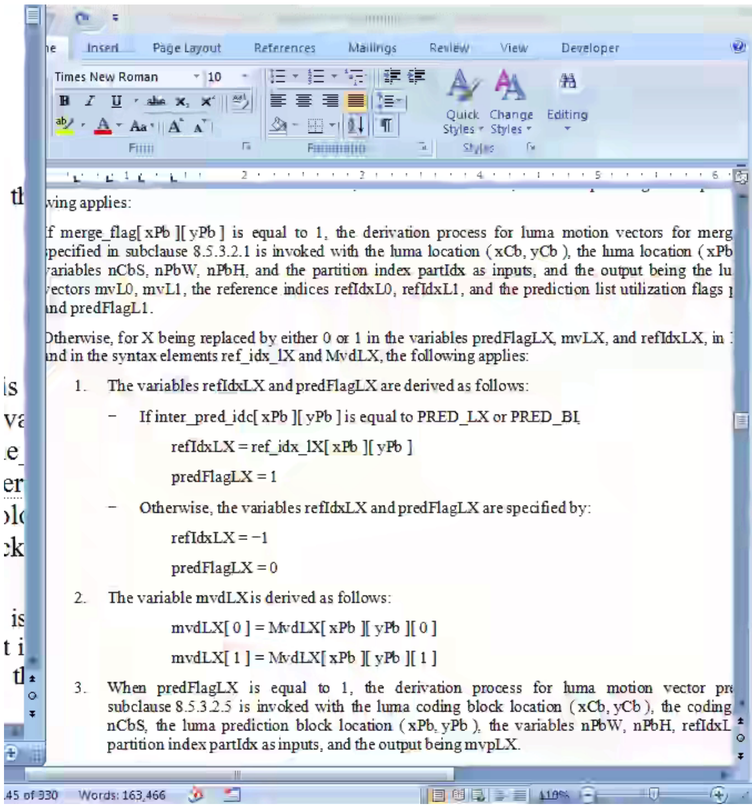
(2) **Only few colors** in a local patch



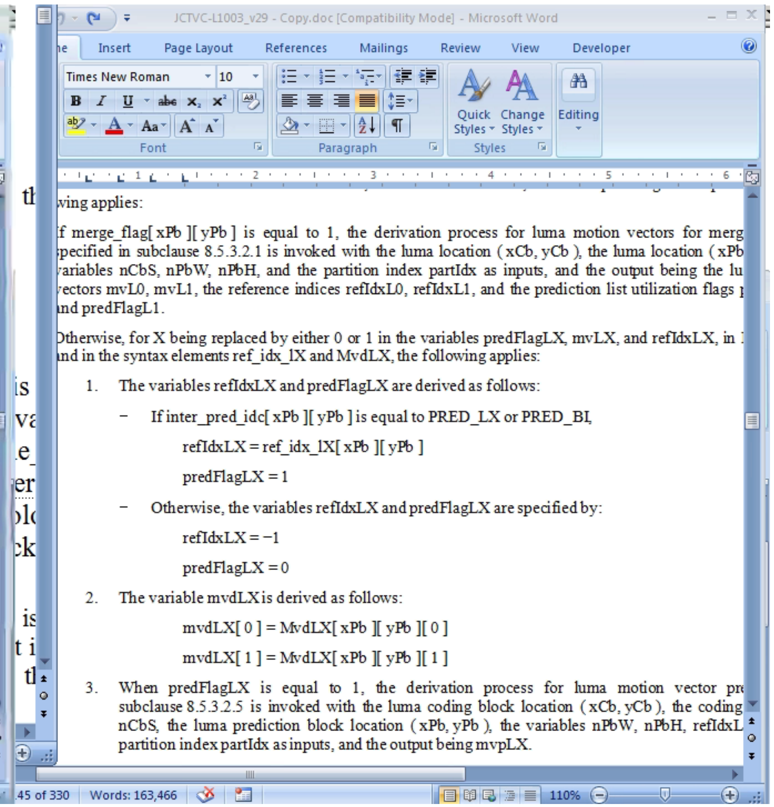
Heat map showing the distribution of color numbers

Where we are now …

HEVC



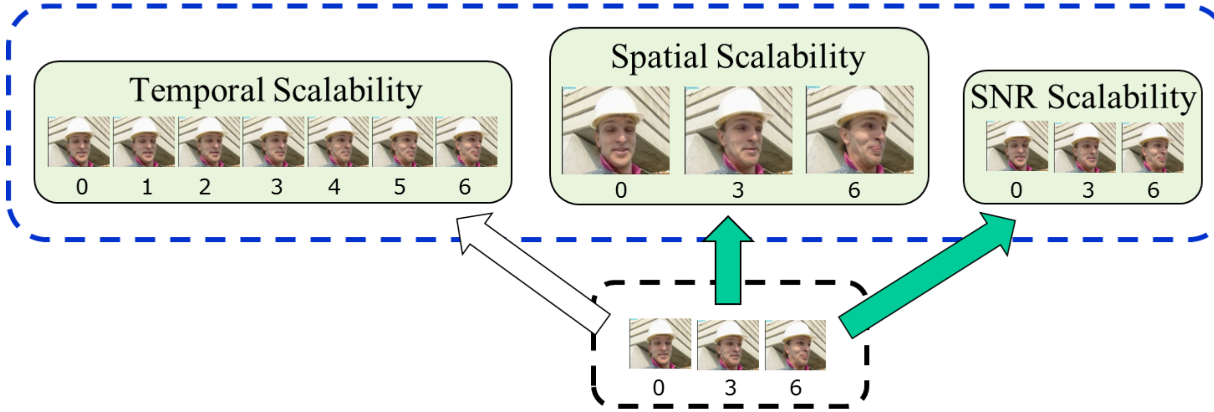
SCC



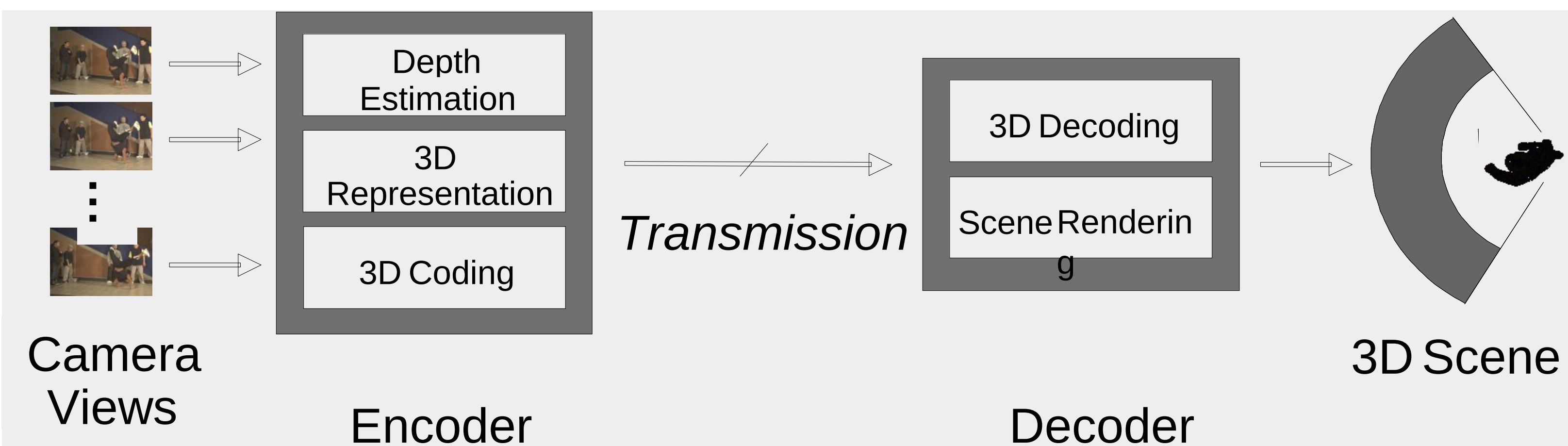
Desktop: 1920x1080_60Hz (All Intra)
Compression ratio ~70

Scalability and 3D Extensions

- **Scalability Extension:** single encoding process to generate a multi-resolution bitstream that can be partially decoded



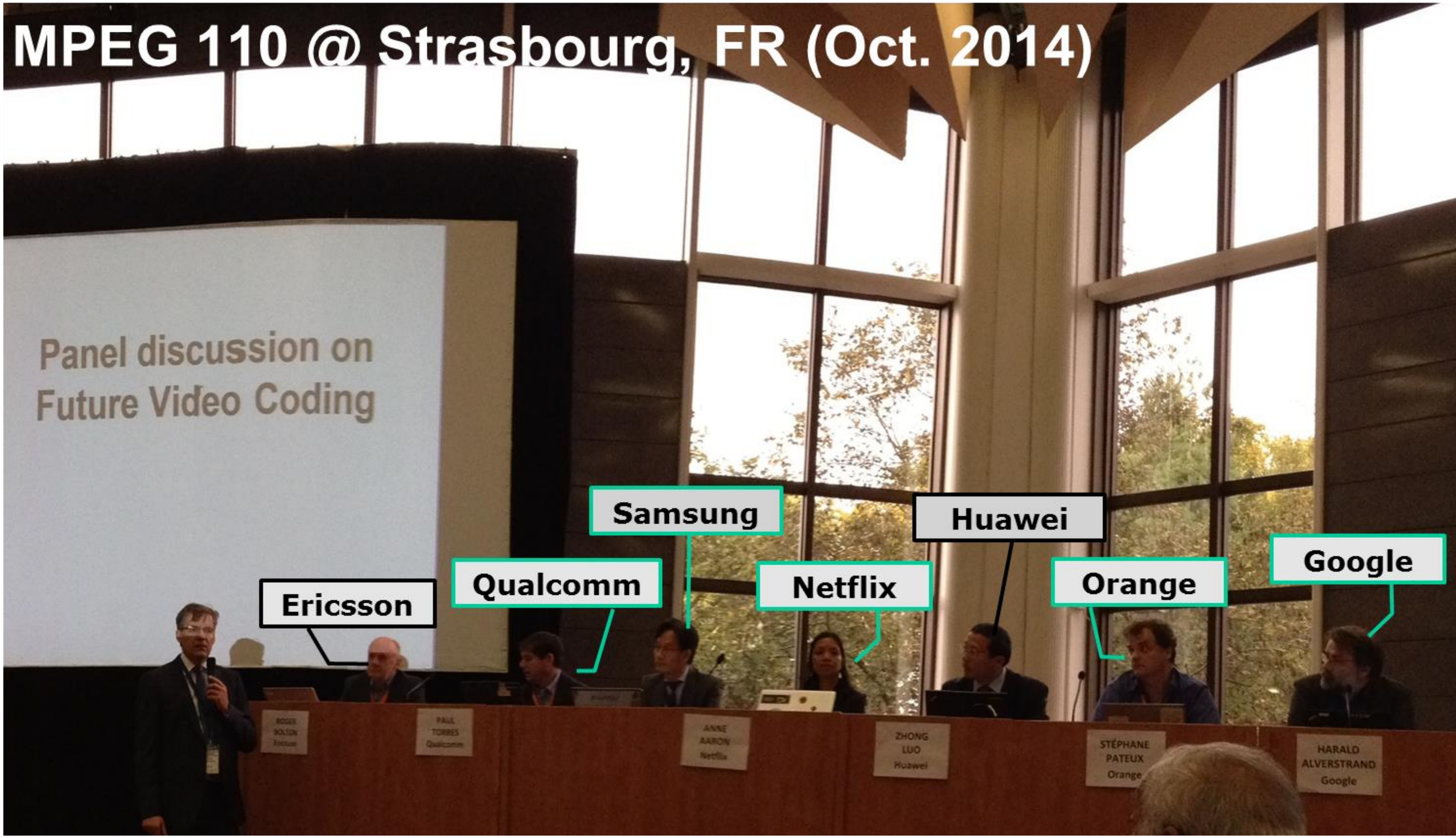
- **3-D Video Extension:** “efficient compression” and “view reconstruction” for a number of dense views



Outline

- **Part I – ISO/IEC Moving Picture Experts Group (MPEG)**
  - Background
  - Recent Milestones
  - Future Video, Machine Learning, and Media Internet-of-Things
- **Part II – MPEG Compact Descriptor for Visual Search (CDVS)**
  - Large-scale Image Retrieval
  - Local Image Descriptors
  - Global Image Descriptors
  - Image Matching
  - Use Case: Mobile Indoor Navigation
- **Part III – Cross-domain Data Retrieval**
  - Canonical Correlation Analysis
  - Deep Boltzmann Machine

What's Next?


MPEG 110 @ Strasbourg, FR (Oct. 2014)
Panel discussion on
Future Video Coding
Ericsson
Qualcomm
Samsung
Netflix
Huawei
Orange
Google
PAUL
TORRES
Qualcomm
ANNE
AARON
Netflix
ZHONG
LUO
Huawei
STÉPHANE
PATEUX
Orange
HARALD
ALVERSTRAND
Google

Some Notes from Brainstorming

- More video traffic coming up
  - **Spherical/360 video, VR video**, **cloud gaming**, social media, OTT applications, mobile video, (live) user-generated content, surveillance video, etc.
  - High dynamic range, wide color gamut, high frame rate
- Cost per video bit still too high
  - 2x performance gain preferred
- Adaption of media processing to **new network architectures**
  - Mobile Edge Computing, 5G, IoT
- **Video analytics**, video quality assessment, etc.

Immersive Video Camera

http://www.airpano.com/360-videos.php

Network-distributed FutureVideo Requirements

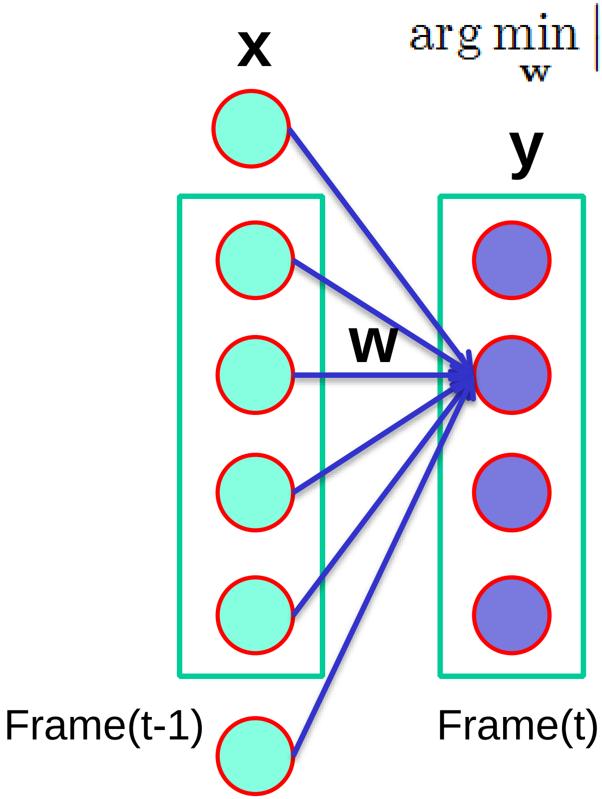
Ad-hoc Group Mandates (March 2016):

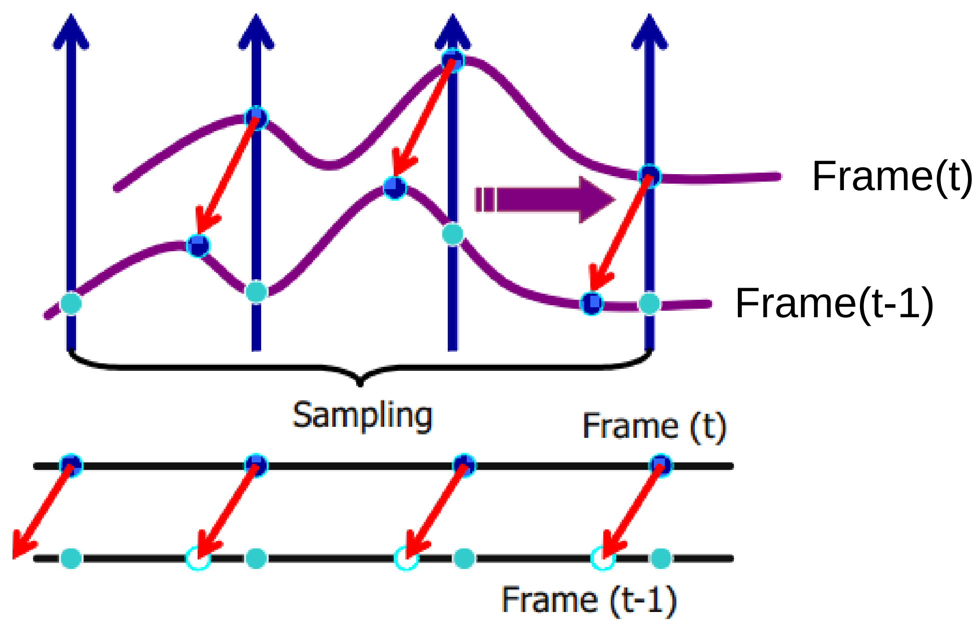
- Develop requirements for encoders whose components are distributed across a media delivery network
- Develop requirements for decoders whose components are distributed across a media delivery network
- Explore what role (if any) scalable video coding has in FutureVideo
- **Consider the role of pre-processing, post-processing, and machine learning in codec standards development and testing**
- Estimate capabilities of media delivery networks (head end, edge, gateway, client), computing resources, consumer display interfaces, and codecs for the year 2020

Use of Machine Learning in Video Coding

- Only few attempts in literature, due mainly to
  - **High complexity** of learning algorithms (even for testing)
  - Stringent requirements on **real-time** processing
- **Case I (Linear Regression):** Fractional-pel Interpolation Filter

$$\arg\min_{\mathbf{w}} \left\|\mathbf{y} - \mathbf{w}^T\mathbf{x}\right\|^2$$

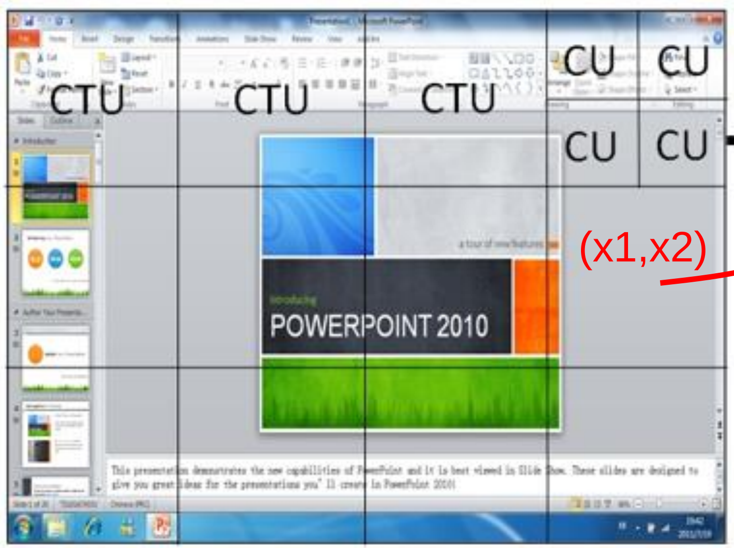


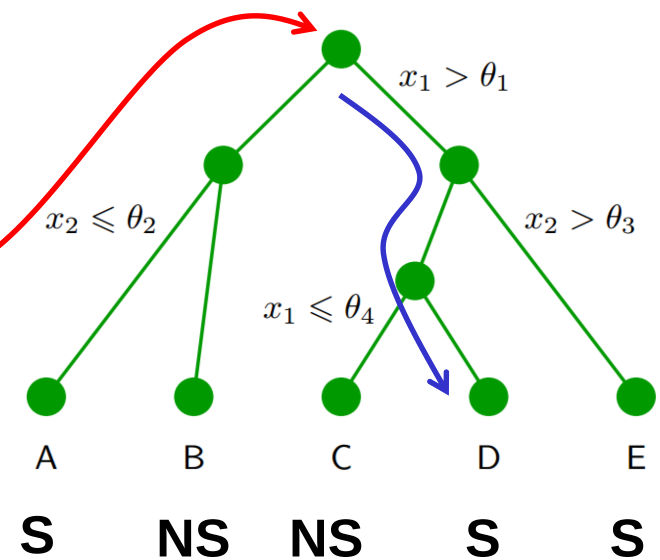


Case II: Fast Mode Decision Using Decision Trees

- **Task:** Decide whether an image block should be split or not
  - (x1, x2): handcrafted features extracted from an image block
  - Labeled training data obtained by brute-force method
- Decision trees are relatively **less complex** than other classifiers

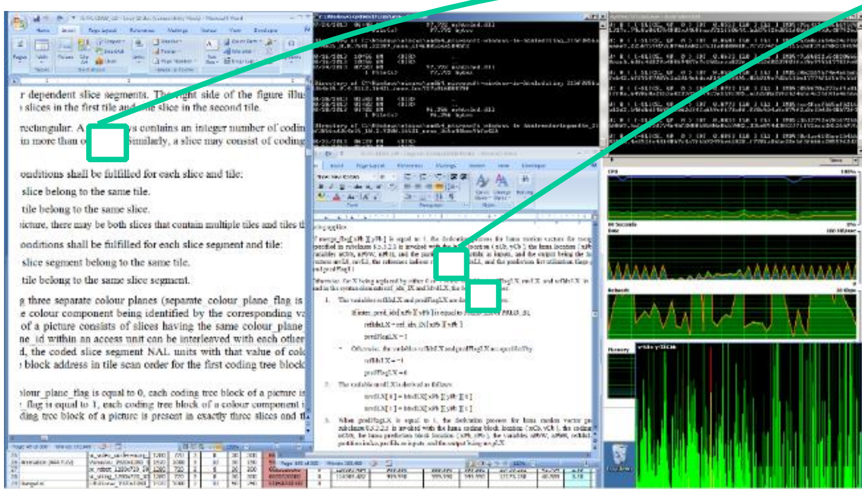
Split (**S**) vs. Not Split (**NS**)





- **Task:** Learn an over-complete dictionary **D** to sparsely represent textual blocks **Y**
- **Training** (D and C both unknown):

$$\min \|\mathbf{Y} - \mathbf{DC}\|_F^2 \ \text{s.t.} \ \|\mathbf{c}_i\|_0 \le T, \|\mathbf{d}_i\|_2^2 = 1, i = 1, 2, ..., N$$



$$\mathbf{Y} = \underbrace{\begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_N \end{bmatrix}_{n \times N}}_{\text{Textual Blocks}}$$

$$\mathbf{C} = \underbrace{\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_N \end{bmatrix}_{K \times N}}_{\text{Sparse Representations}}$$

$$\mathbf{D} = \underbrace{\begin{bmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_K \end{bmatrix}_{n \times K}}_{\text{Over-complete Dictionary, } K >> n}$$

- **Testing**: Once we have **D**, the sparse representation **c** for any input textual block **y** can be obtained using **matching pursuit**

Case IV: Image Quality Assessment

- **Task:** Fuse multiple objective metrics (e.g. PSNR, SSIM, VIF) to better predict human perception of image quality
- <u>Support Vector Regression</u>
  - **Input:** scores computed with different metrics w.r.t an image
  - **Target output:** human rating w.r.t. the same image

$$\underbrace{\mathrm{x}_i = [m_{i,1}, m_{i,2}, ..., m_{i,k}]^T}_{\text{Input}} \rightarrow \underbrace{y_i}_{\text{Target output}}$$

  - **Objective:** compute a predictor of the form

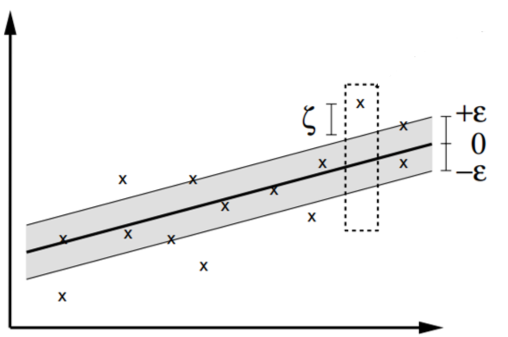
$$f(\mathrm{x}_i) = \mathbf{w}^T \mathrm{x}_i + b$$

such that

$$\min \frac{1}{2} \|\mathbf{w}\|_2^2 \text{ s.t. } |f(\mathrm{x}_i) - y_i| < \varepsilon, i = 1, 2, ..., N$$

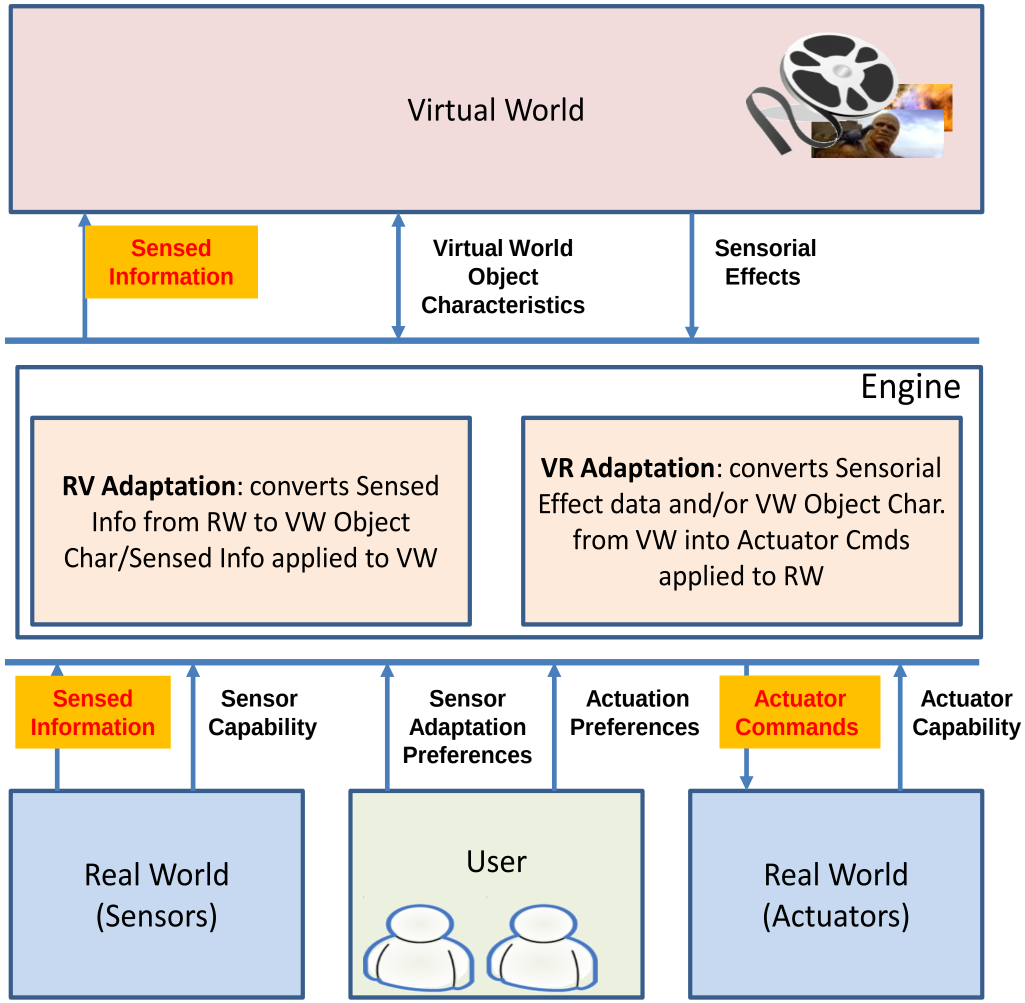
$\Downarrow$

Can be relaxed by introducing slack variables



MPEG-V: Media Context and Control

- Architecture and associated information representations to enable interoperability (1) between virtual worlds, and (2) between real and virtual worlds

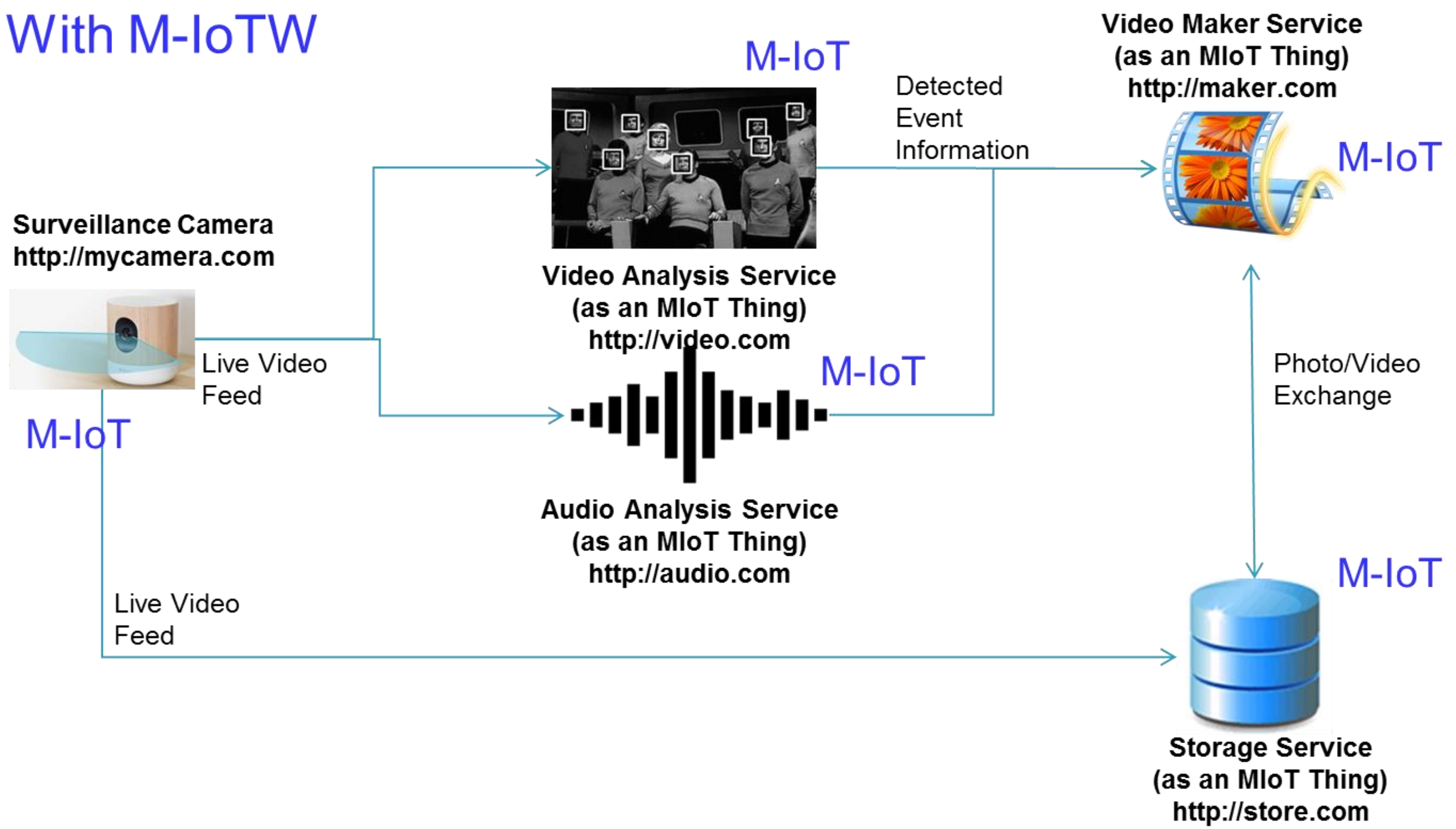


Source: ISO/IEC JTC1/SC29/WG11 N14187

Use Cases:

- Avatar control in virtual word based on sensorial data captured in real world
- 4D film: 3D + physical effects

- **Entity:** Physical/virtual object sensed by and/or acted on by Things
- **Thing:** Things that can communicate, and may sense/act on Entities
- **Media Thing:** Things with audio/visual sensing/actuating capabilities



Source: MPEG 3DG Plenary Report @ 114 Meeting, San Diego

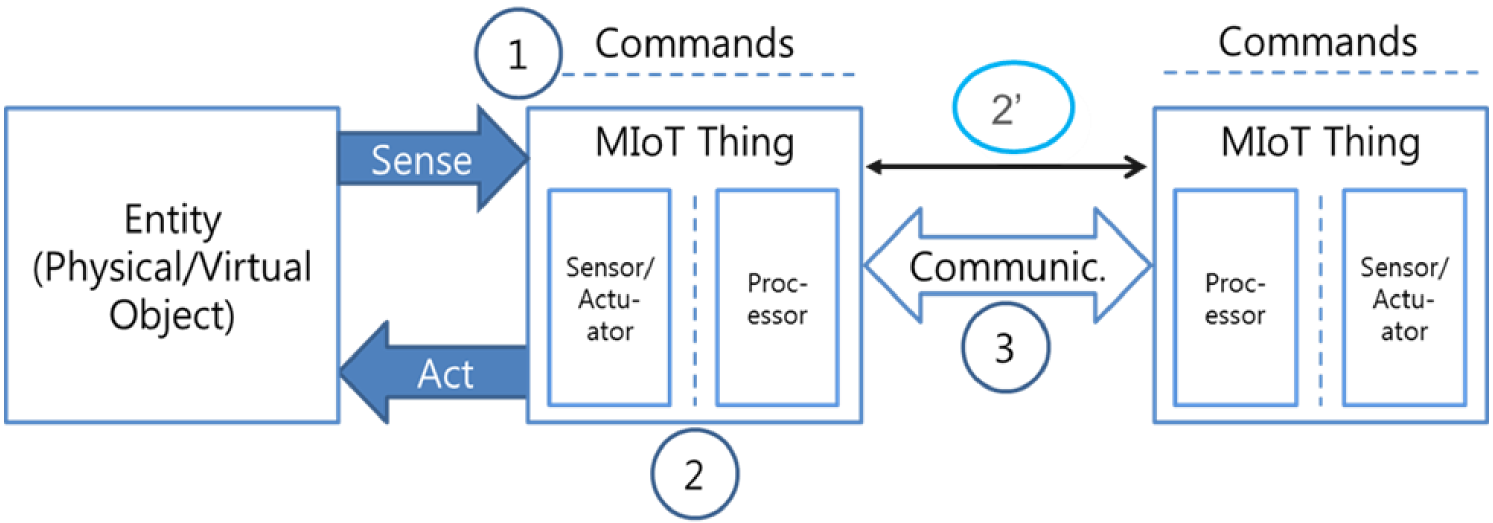
MIoTW: Interfaces to be Standardized

1: data provided by the system designer (set-up, commands, services)

2: raw or high level (descriptions) AV data within Mthing

2’: a wrapped and transmission friendly version of Interface 2

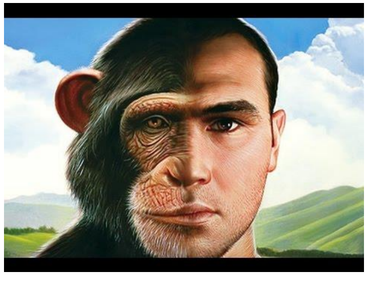
3: M-IoT capabilities, discovery, configuration data



Source: ISO/IEC JTC1/SC29/WG11 N15727

Many Others

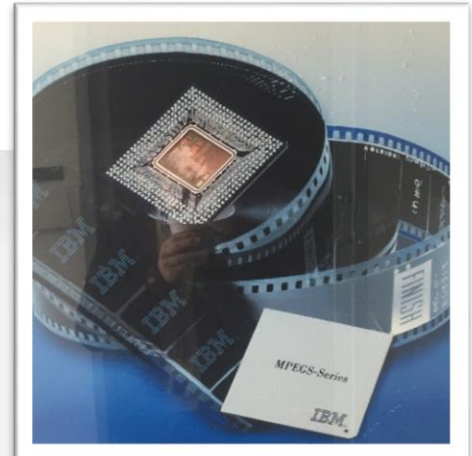
- **MPEG-ARAF: Augmented Reality Application Format**
  - Architecture and representation to create media-rich AR apps
- **MPEG UD: User Descriptor**
  - Standardized user, context, service, recommendation descriptors to help recommendation engines deliver better, personalized, and relevant choices to users
- **MPEG-7 CDVS: Compact Descriptor for Visual Search**
  - Compact descriptors suitable for visual search
- **MPEG-7 CDVA: Compact Descriptor for Video Analysis**
- **Green MPEG: Energy-efficient Media Technologies**
- **“Genome” compression**
- **Point cloud compression**
- **Big Media …**



Visit http://mpeg.chiariglione.org/about

Follow @MPEGgroup

The Moving Picture Experts Group



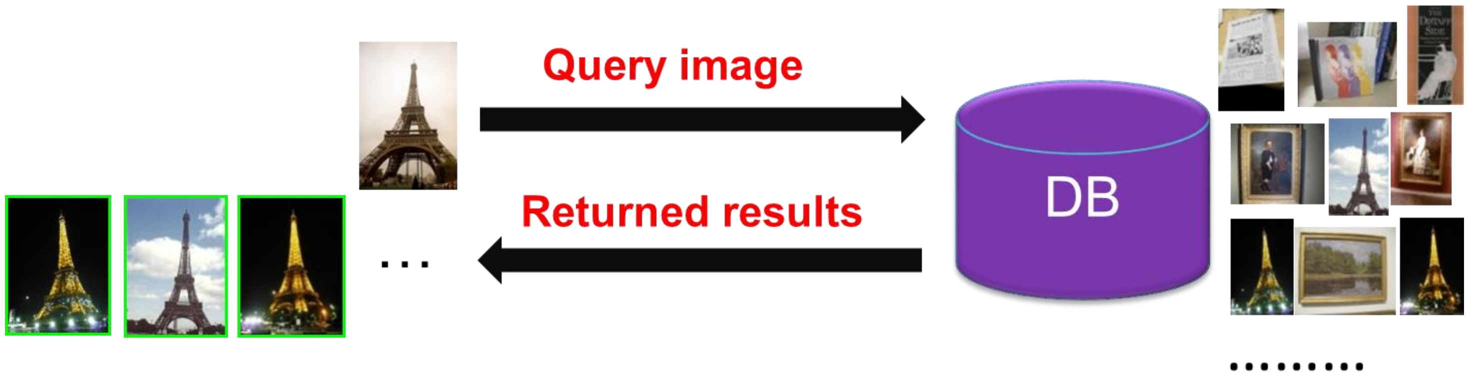
HOME | STANDARDS | TECHNOLOGIES | MEETINGS | ABOUT MPEG

Outline

- **Part I – ISO/IEC Moving Picture Experts Group (MPEG)**
  - Background
  - Recent Milestones
  - Future Video, Machine Learning, and Media Internet-of-Things
- **Part II – MPEG Compact Descriptor for Visual Search (CDVS)**
  - Large-scale Image Retrieval
  - Local Image Descriptors
  - Global Image Descriptors
  - Image Matching
  - Use Case: Mobile Indoor Navigation
- **Part III – Cross-domain Data Retrieval**
  - Canonical Correlation Analysis
  - Deep Boltzmann Machine

"Content-based" Image Retrieval

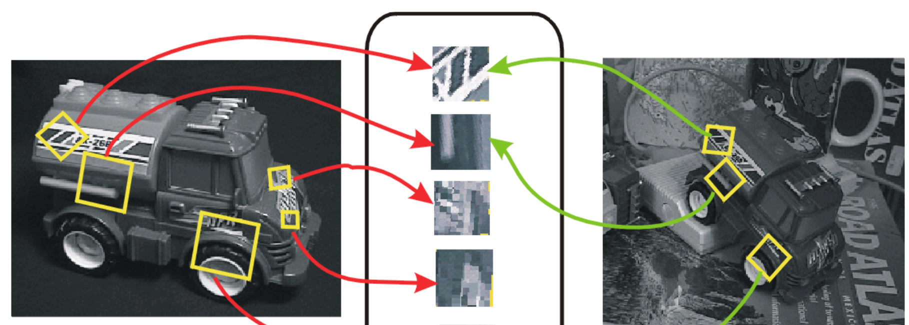
- Classical and well-studied problem



General approach:

(1) Extract **appearance features** from images

(2) Search images with **similar appearance features**



Credit: David Lowe

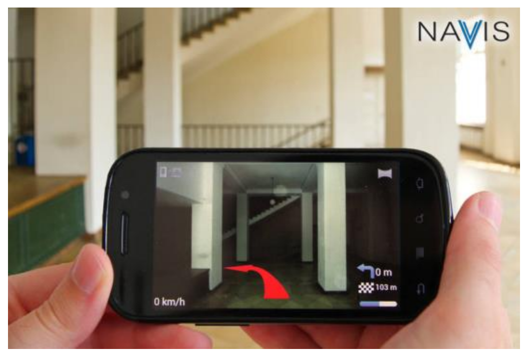
Visual Search Applications

- Find information by taking images as **information proxy**

Mobile Visual Search

Augmented Reality Applications

Indoor Navigation

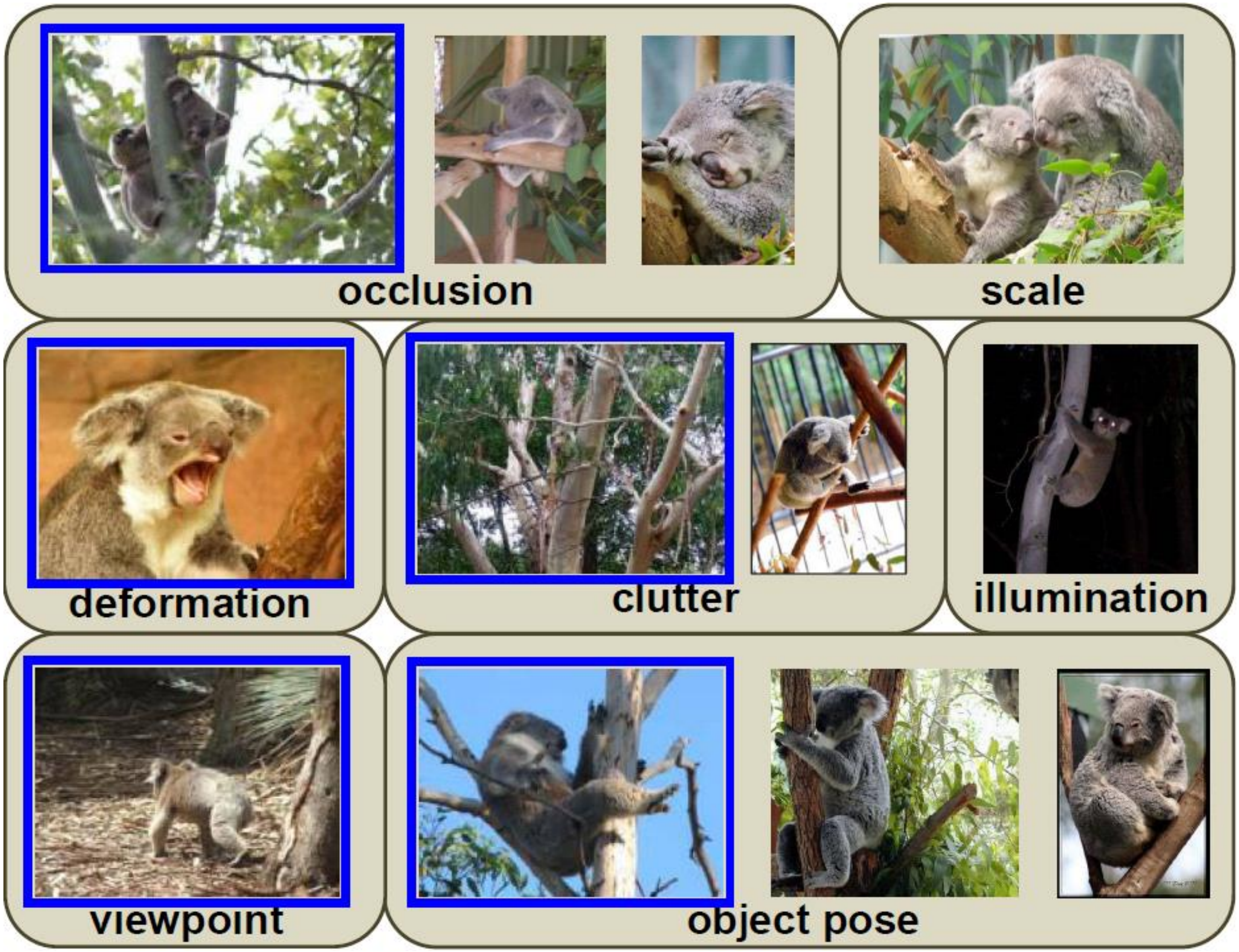


Automotive Applications

TV/IPTV Applications

“Compact Descriptors for Visual Search: Applications and Use Scenarios,” ISO/IEC MPEG 93rd me
http://www.tum.de/en/about-tum/news/press-releases/long/article/30040/

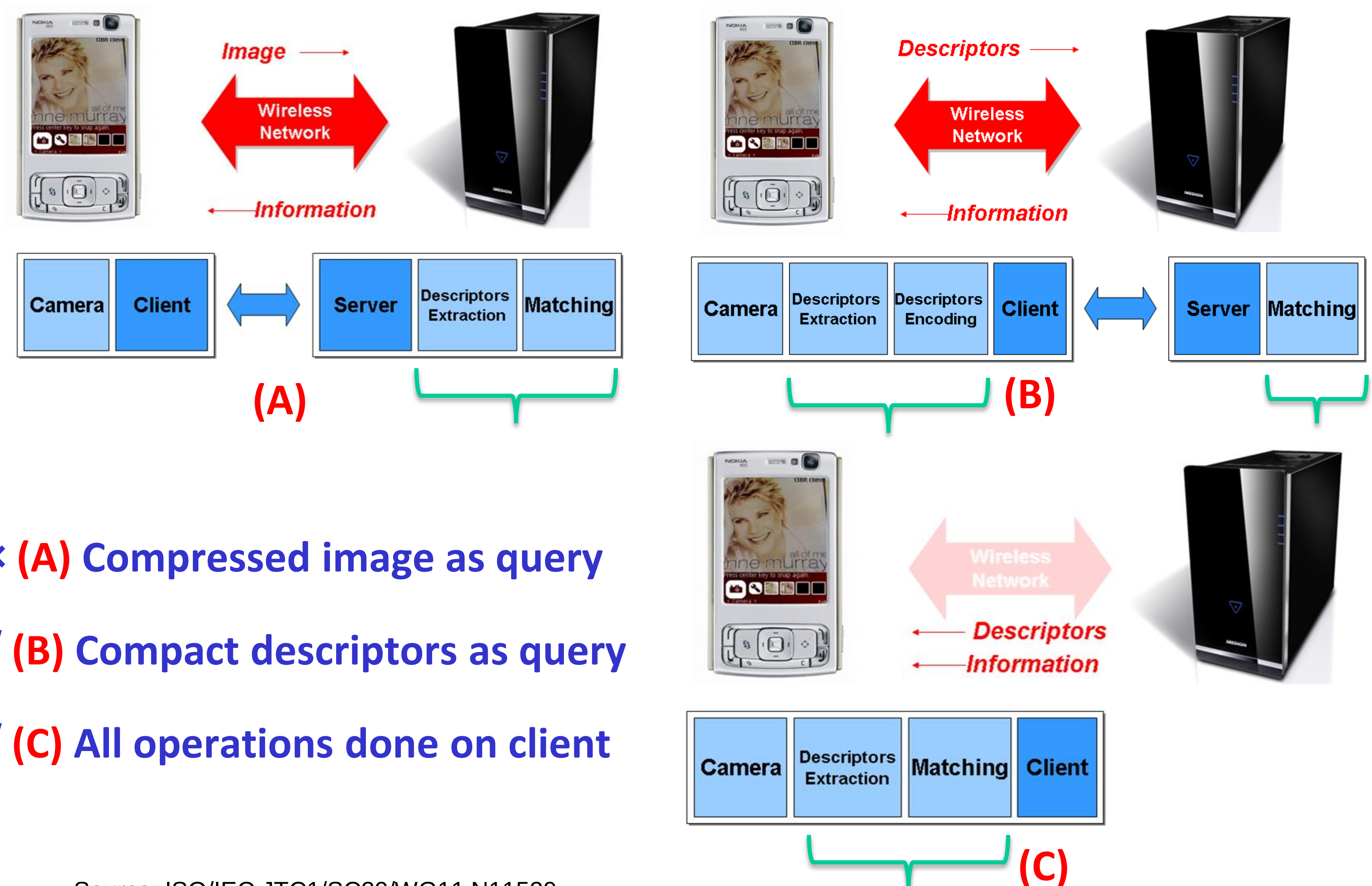
Challenges (1/2) – Are these similar images?



Challenges (2/2) – Large-scale Databases

- Accurate search has to be done in **split second**, even with database containing **billions of images**

System Architectures



× **(A) Compressed image as query**

√ **(B) Compact descriptors as query**

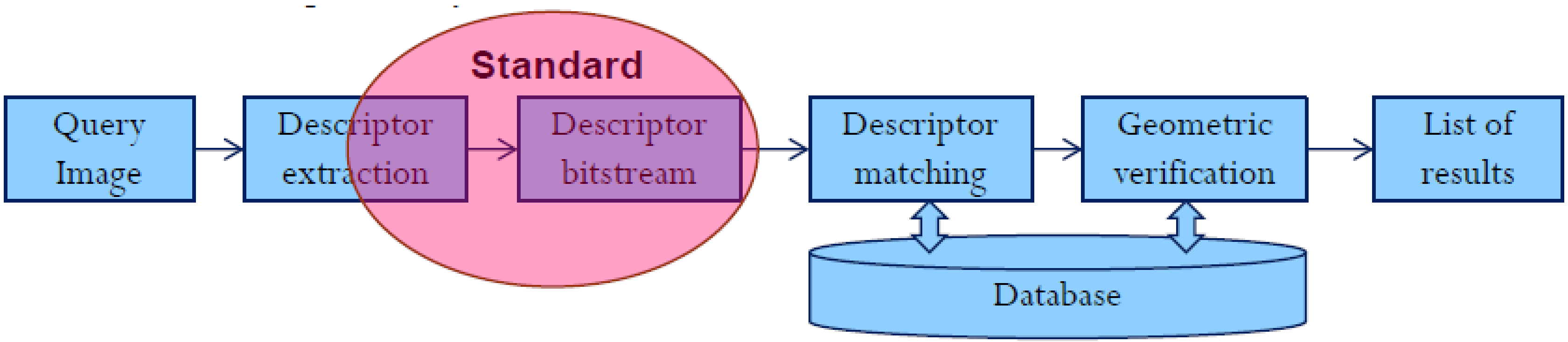
√ **(C) All operations done on client**

Source: ISO/IEC JTC1/SC29/WG11 N11529

MPEG-7 Part 13 Compact Descriptors for Visual Search

- Compact descriptors that allow for efficient **content-based search** of images in databases

- <u>Scope of Standardization:</u>
  - Bitstream of descriptors
  - Part of descriptor extraction process for interoperability (e.g. Key point detection)



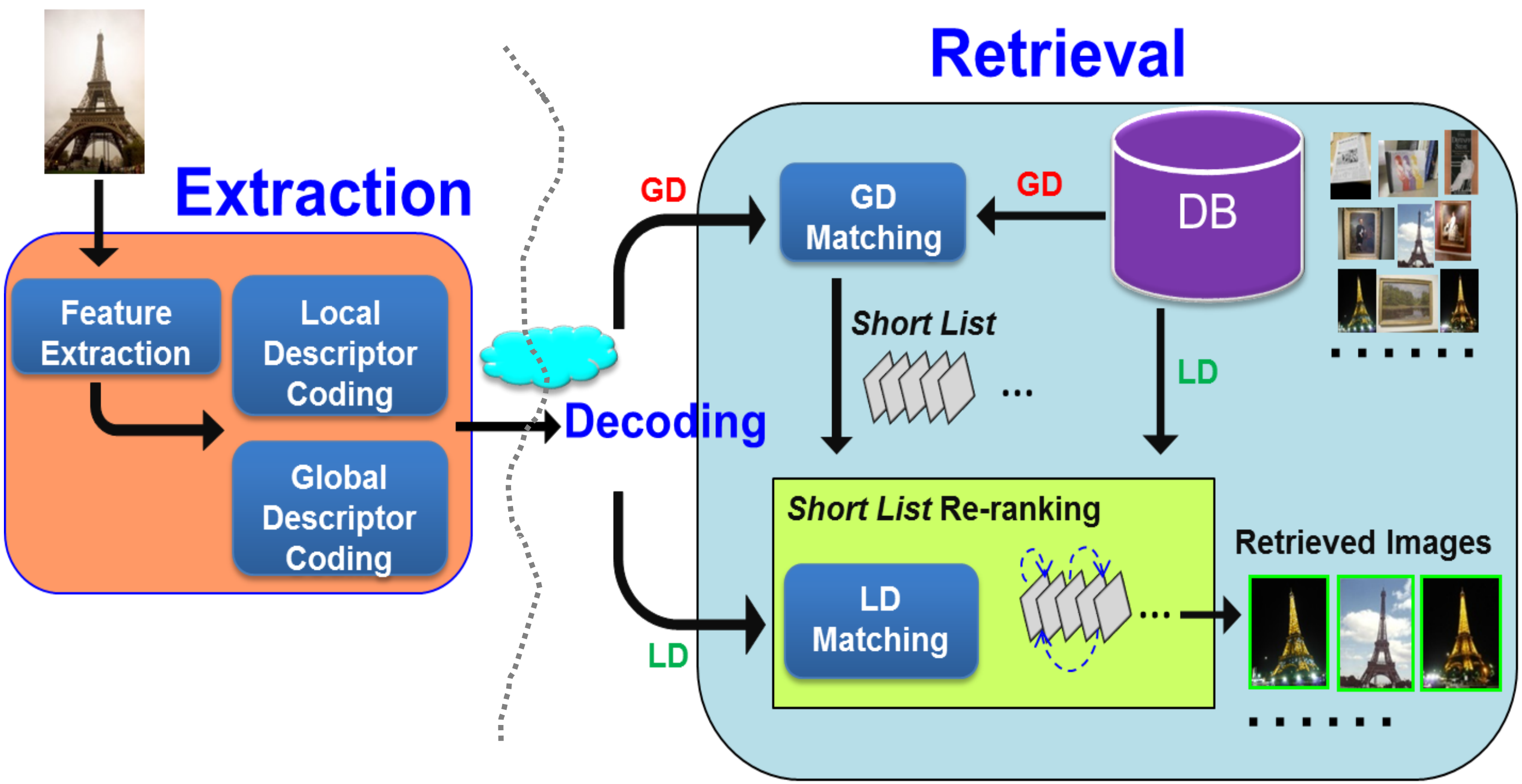
- **International Standard since 2014/10**

Requirements

- Robustness
  - High matching accuracy subject to **image changes** (e.g. in vantage point, camera parameters, lighting conditions, partial occlusions)
- Compactness
  - **Minimized lengths** of image descriptors
- Scalability
  - **Adaptation of descriptor lengths** to support the required performance level and database size
  - Support of **web-scale visual search** applications and databases
- Localization
  - **Identification and localization** of matching regions
  - Estimation of **geometric transformation**
- Low complexity for extraction and matching
  - **Memory and computation requirements**

CDVS Retrieval Pipeline

- Extraction of local and global descriptors
  - **Local descriptors (LD):** appearances of (key) local image patches
  - **Global descriptor (GD):** aggregation of LD's into a single vector
- Generation of **shortlist by GD** and **re-ranking by LD's matching**



Normative **Non-normative (other ways possible)**

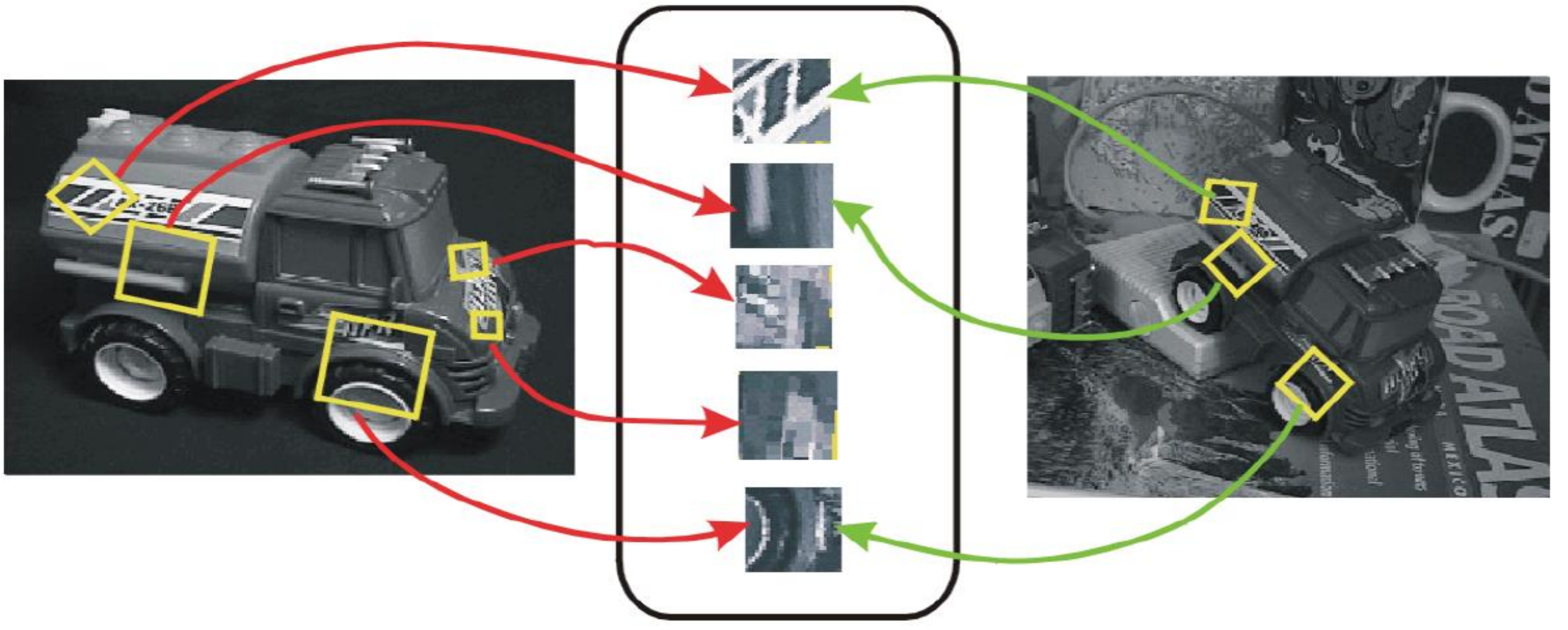
Outline

- **Part I – ISO/IEC Moving Picture Experts Group (MPEG)**
  - Background
  - Recent Milestones
  - Future Video, Machine Learning, and Media Internet-of-Things
- **Part II – MPEG Compact Descriptor for Visual Search (CDVS)**
  - Large-scale Image Retrieval
  - Local Image Descriptors
  - Global Image Descriptors
  - Image Matching
  - Use Case: Mobile Indoor Navigation
- **Part III – Cross-domain Data Retrieval**
  - Canonical Correlation Analysis
  - Deep Boltzmann Machine

Matching with Local Features

- Local features facilitate robust matching over **geometric and photometric transformations**

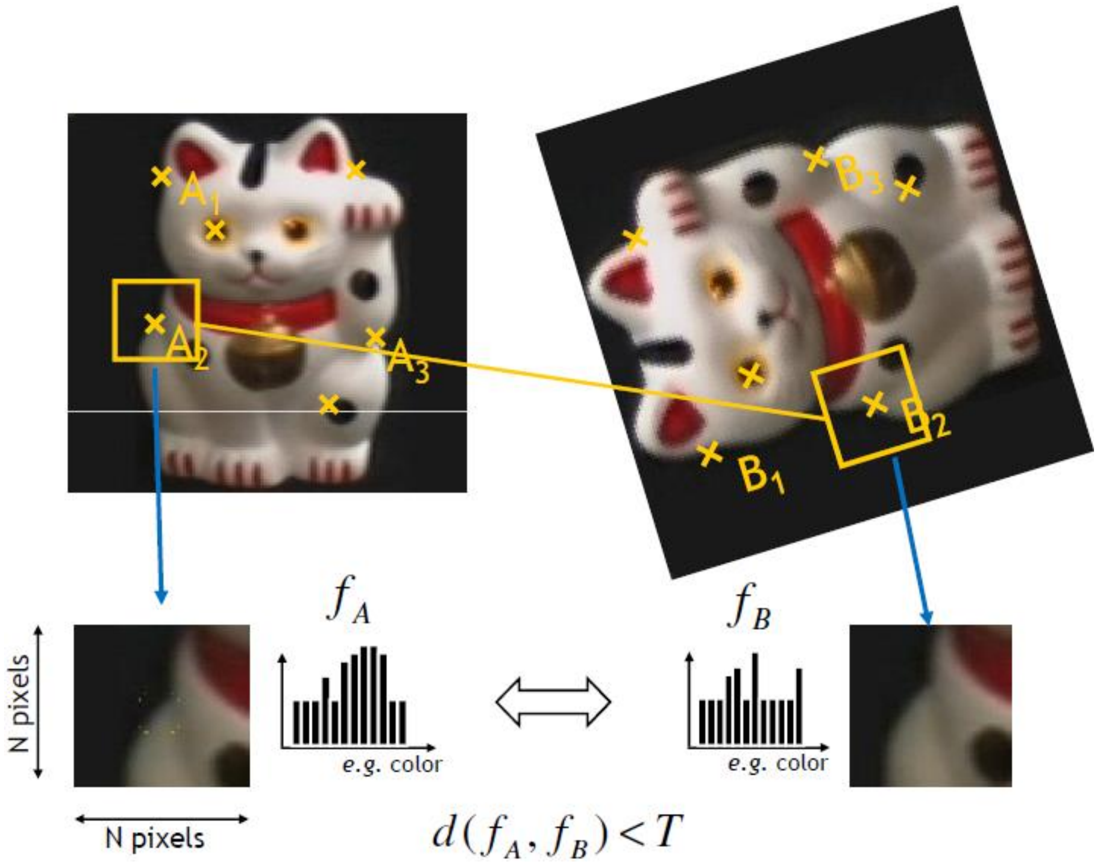
  → More robust but usually time-consuming



Credit: David Lowe

General Procedure

- Find a set of distinctive **keypoints**
- Define a region around each keypoint
- Extract and normalize the region content
- Compute a **local descriptor** from the normalized region
- Match local descriptors



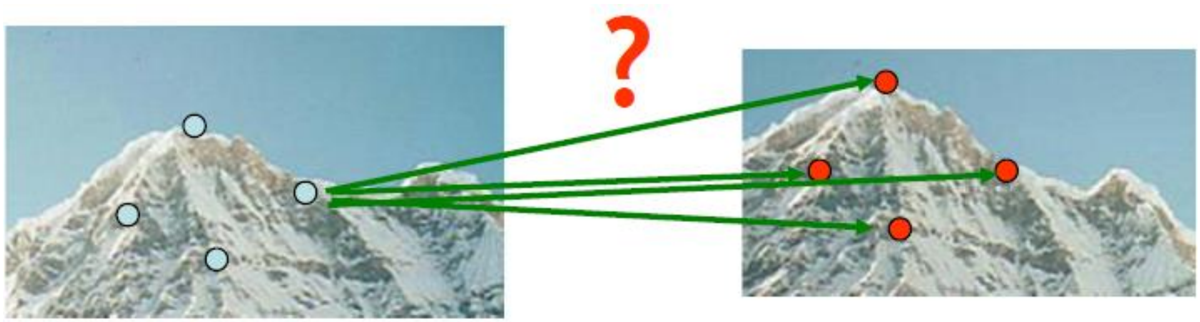
Credit: Kristen Grauman, Bastian Leibe

Desirable Properties

- **Repeatable keypoint detection** vs. image transformations

No chance to match!

- **Discriminative description** of local image patches



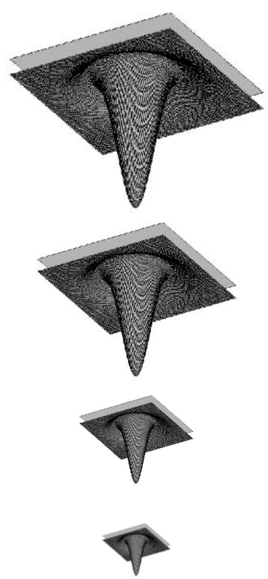
Laplacian-of-Gaussian (LoG): Finding Blobs

- Input → $\underbrace{\text{Gaussian filtering}}$ → $\underbrace{\text{Laplacian}}$ = Input → <u>(Blob detector)</u>

<u>Scaling</u> <u>Edge Detector (second derivatives)</u>

- **Blob:** Points or regions in the image that are brighter or darker than the surrounding

filter
scale σ



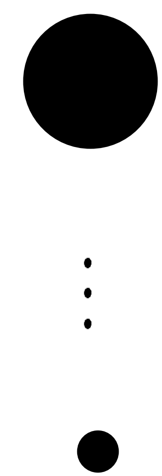
discrete

$$\begin{pmatrix} 0 & 0 & 3 & 2 & 2 & 2 & 3 & 0 & 0 \\ 0 & 2 & 3 & 5 & 5 & 5 & 3 & 2 & 0 \\ 3 & 3 & 5 & 3 & 0 & 3 & 5 & 3 & 3 \\ 2 & 5 & 3 & -12 & -23 & -12 & 3 & 5 & 2 \\ 2 & 5 & 0 & -23 & -40 & -23 & 0 & 5 & 2 \\ 2 & 5 & 3 & -12 & -23 & -12 & 3 & 5 & 2 \\ 3 & 3 & 5 & 3 & 0 & 3 & 5 & 3 & 3 \\ 0 & 2 & 3 & 5 & 5 & 5 & 3 & 2 & 0 \\ 0 & 0 & 3 & 2 & 2 & 2 & 3 & 0 & 0 \end{pmatrix}$$

$\vdots$

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & -16 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

difference
scale



Credit: Bastian Leibe

Blobs detected by LoG

Credit: Lana Lazebnik

Other Detectors

- Hessian & Harris — [Beaudet '78], [Harris '88]
- Laplacian, DoG — [Lindeberg '98], [Lowe '99]
- Harris-/Hessian-Laplace — [Mikolajczyk & Schmid '01]
- Harris-/Hessian-Affine — [Mikolajczyk & Schmid '04]
- EBR and IBR — [Tuytelaars & Van Gool '04]
- MSER — [Matas '02]
- Salient Regions — [Kadir & Brady '01]
- Others…

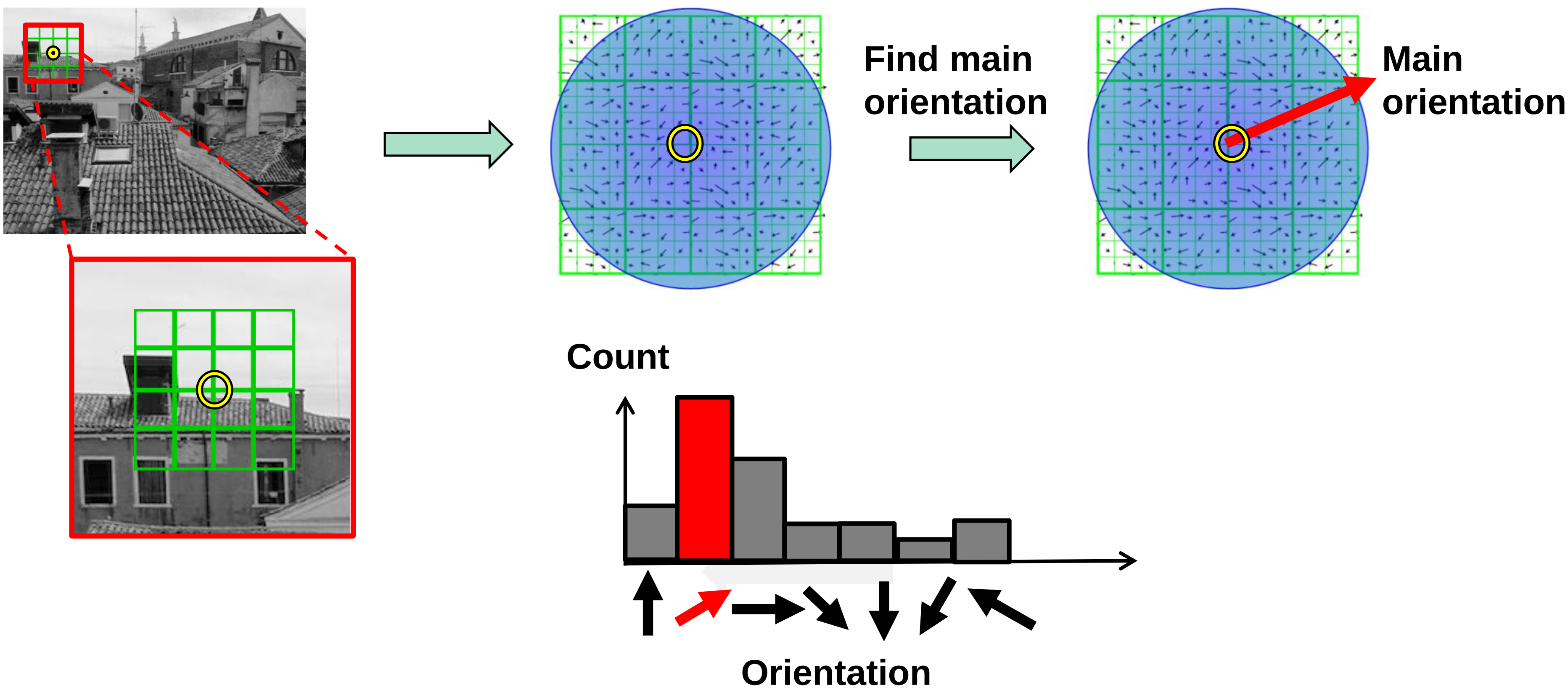
Those detectors have become a ***basic building block*** *for many recent applications in Computer Vision.*

Credit: Kristen Grauman, Bastian Leibe

- LoG (Keypoint) + **Gradient Histogram (Descriptor)**

  (1) Compute gradient histogram (16x16) around keypoint

  (2) Determine main orientation (peak value)

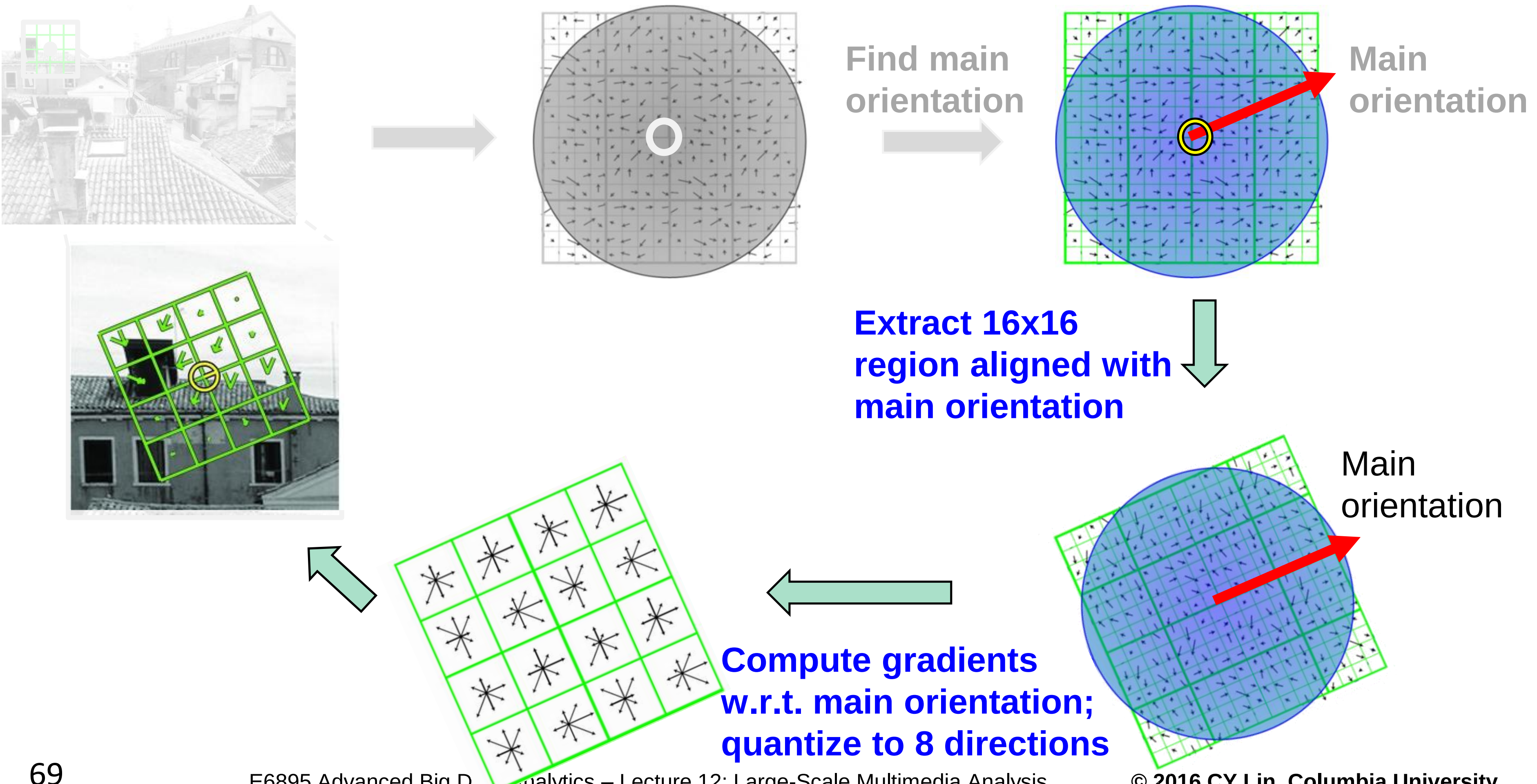


Scale-Invariant Feature Transform (SIFT) – 2/2

- LoG (Keypoint) + **Gradient Histogram (Descriptor)**

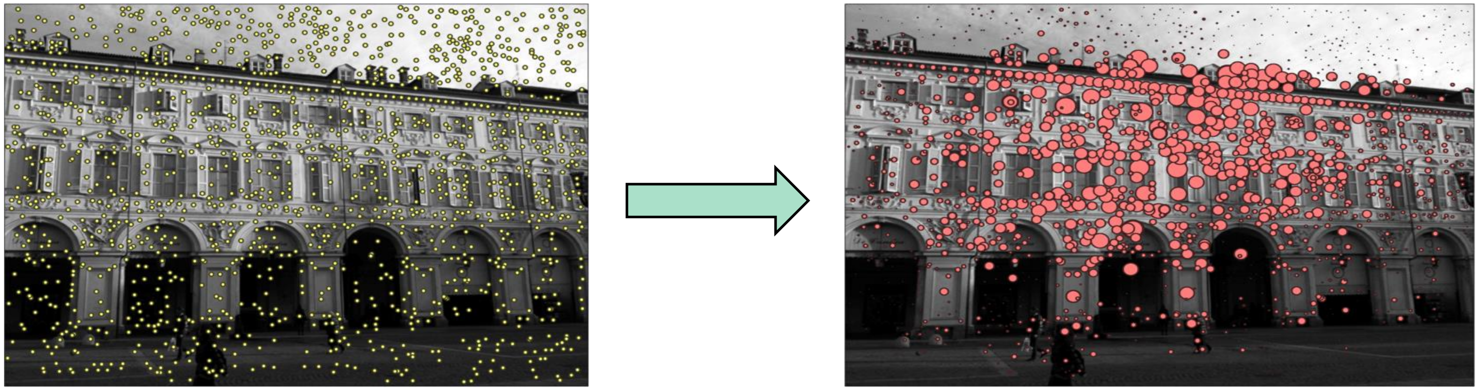
  (3) Extract 16x16 region aligned with main orientation

  (4) Compute gradients **w.r.t. main orientation**

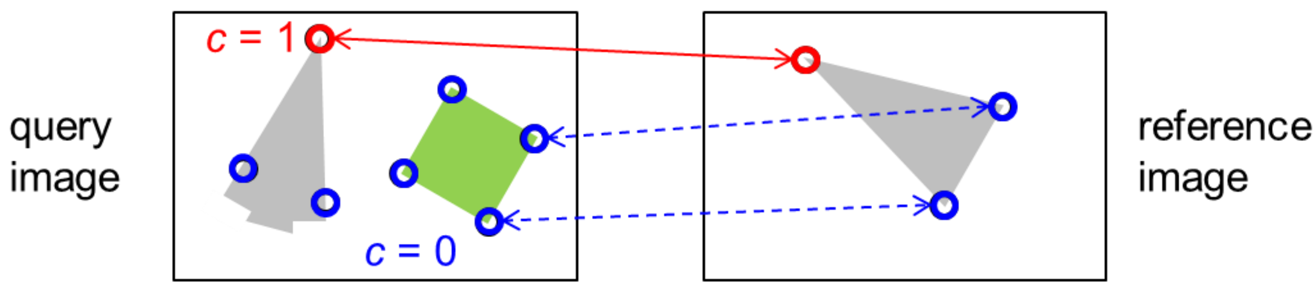


Keypoint Selection

Problem: There could be excessive keypoints/descirptors



Solution: Keep those with higher probability of correct match c = 1 given their attributes **σ** (scale), **θ** (orientation), **l** (LoG value), **d** (distance to image center)



$$\text{Score}(\sigma, \theta, l, d) = p(c = 1 | \sigma \in \Sigma_i) \times p(c = 1 | \theta \in \Theta_j) \times p(c = 1 | l \in L_k) \times p(c = 1 | d \in D_h)$$

Local Descriptor Coding (1/2)

- **Transform** + Quantization + Variable Length Coding

SIFT $\boxed{x_0 \mid x_1 \mid \dots \mid x_{127}}$

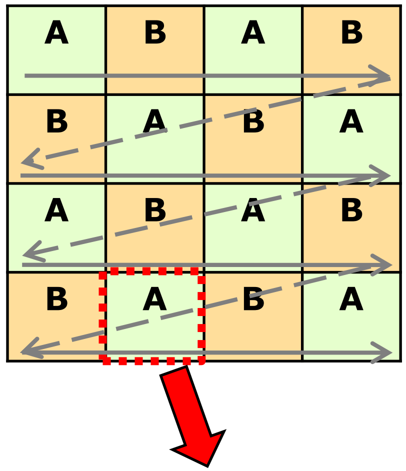
(128 bytes = 1024 bits)

Transform

$\boxed{x'_0 \mid x'_1 \mid \dots \mid x'_{127}}$

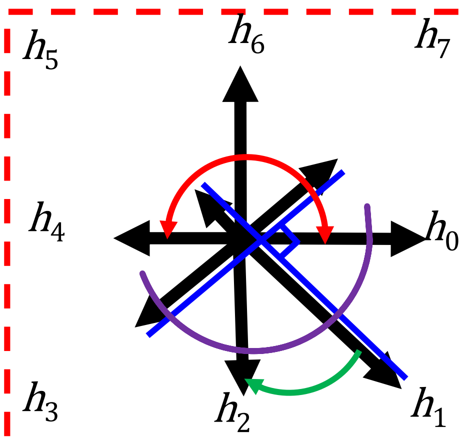
Scalar Quantization

$\boxed{s_0 \mid s_1 \mid \dots \mid s_{127}}$



Transform A

Transform B



16 transform units

Opposite (180º)

$v_0 = (h_0 - h_4)/2$
$v_1 = (h_1 - h_5)/2$
$v_2 = (h_2 - h_6)/2$
$v_3 = (h_3 - h_7)/2$

Neighbouring (45º)

$v_4 = (h_1 - h_2)/2$
$v_5 = (h_3 - h_4)/2$
$v_6 = (h_5 - h_6)/2$
$v_7 = (h_7 - h_0)/2$
$v_8 = (h_0 - h_1)/2$
$v_9 = (h_2 - h_3)/2$
$v_{10} = (h_4 - h_5)/2$
$v_{11} = (h_6 - h_7)/2$

Line (perpendicular)

$v_{12} = ((h_1 + h_5) - (h_3 + h_7))/4$
$v_{13} = ((h_0 + h_4) - (h_2 + h_6))/4$

Halve

$v_{14} = ((h_0 + h_1 + h_2 + h_3) - (h_4 + h_5 + h_6 + h_7))/8$
$v_{15} = ((h_0 + h_2 + h_4 + h_6) - (h_1 + h_3 + h_5 + h_7))/8$

Info. about shape of SIFT

Local Descriptor Coding (2/2)

- Transform + **Quantization** + **Variable Length Coding**

SIFT

x_0	x_1	…	x_{127}

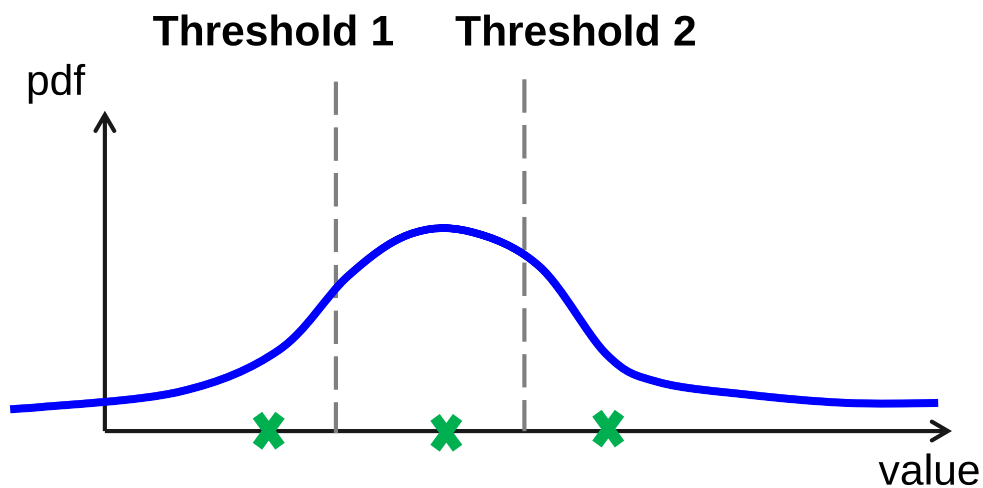
(128 bytes = 1024 bits)

Transform

x'_0	x'_1	…	x'_{127}

Scalar Quantization

s_0	s_1	…	s_{127}

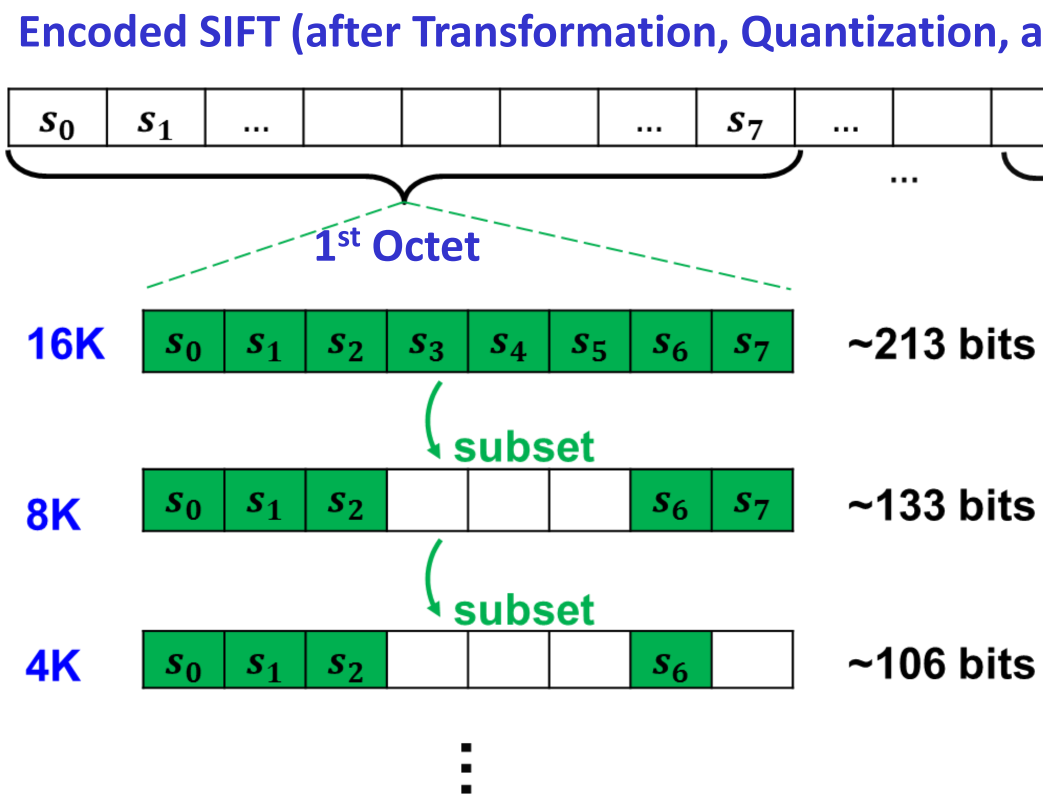


Ternary Value -1 , 0 , 1

Codeword 10 , 0 , 11

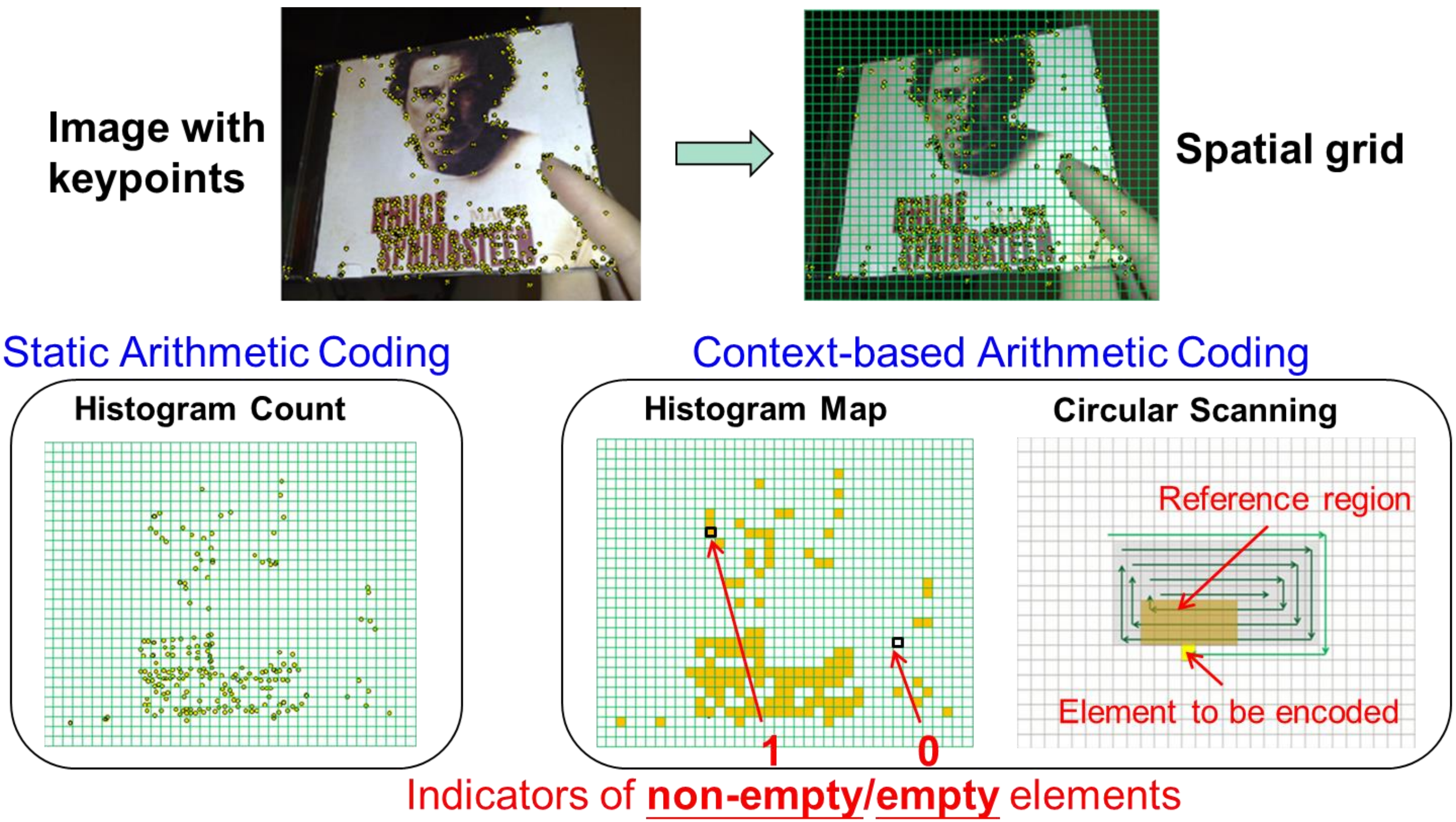
Scalability

- **Scalability:** adaptation of descriptor length
- **Embedded** bitstream allows matching between local descriptors of different sizes (e.g. query 4k vs. reference 16k)



Coordinates Coding

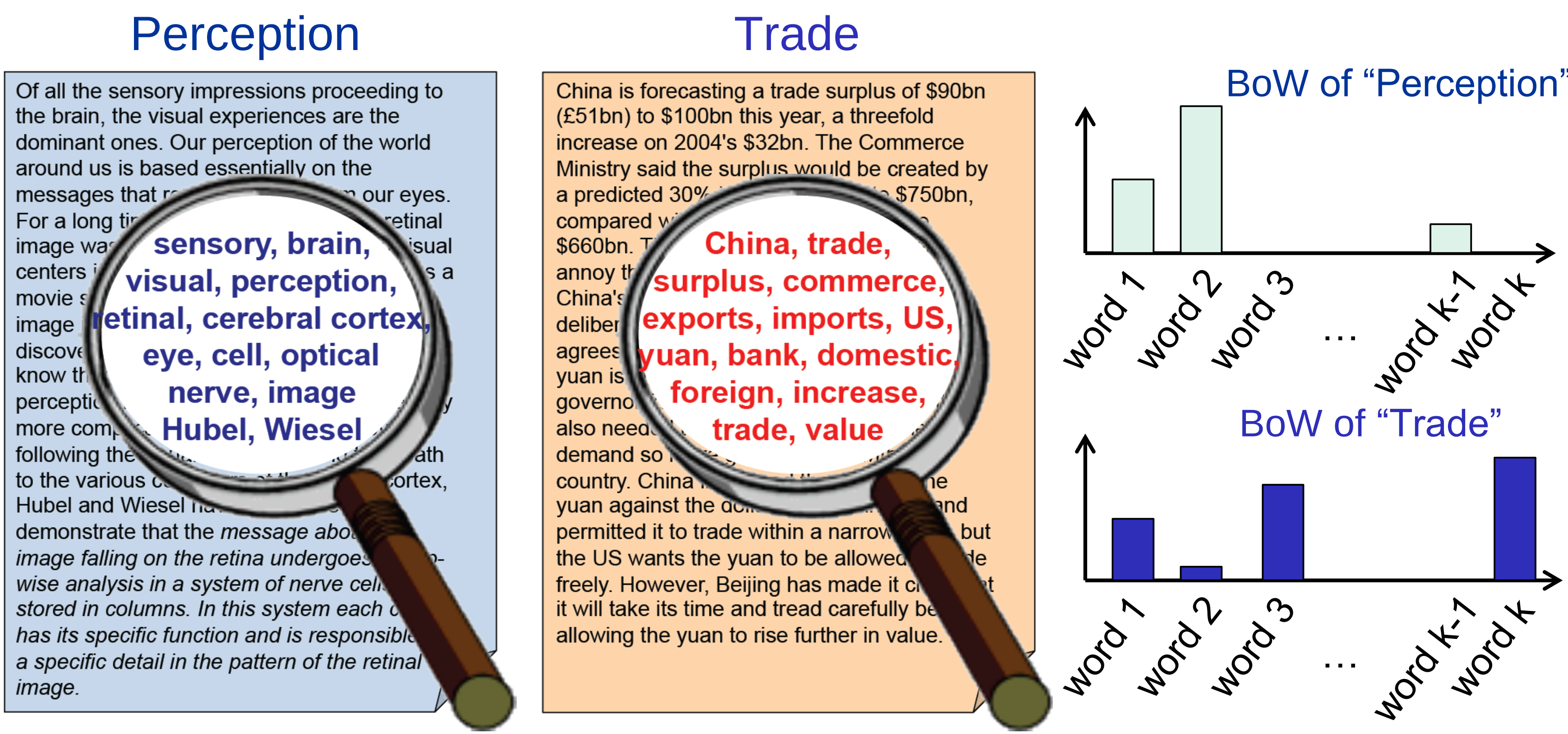
- Compression of **coordinates of selected keypoints**
- Coordinates = histogram count + histogram map



Outline

- **Part I – ISO/IEC Moving Picture Experts Group (MPEG)**
  - Background
  - Recent Milestones
  - Future Video, Machine Learning, and Media Internet-of-Things
- **Part II – MPEG Compact Descriptor for Visual Search (CDVS)**
  - Large-scale Image Retrieval
  - Local Image Descriptors
  - Global Image Descriptors
  - Image Matching
  - Use Case: Mobile Indoor Navigation
- **Part III – Cross-domain Data Retrieval**
  - Canonical Correlation Analysis
  - Deep Boltzmann Machine

- **<u>Bag-of-Words (BoW):</u>** Document as **composition of words**
- Methods for efficient document retrieval are mature and effective enough to deal with millions at once

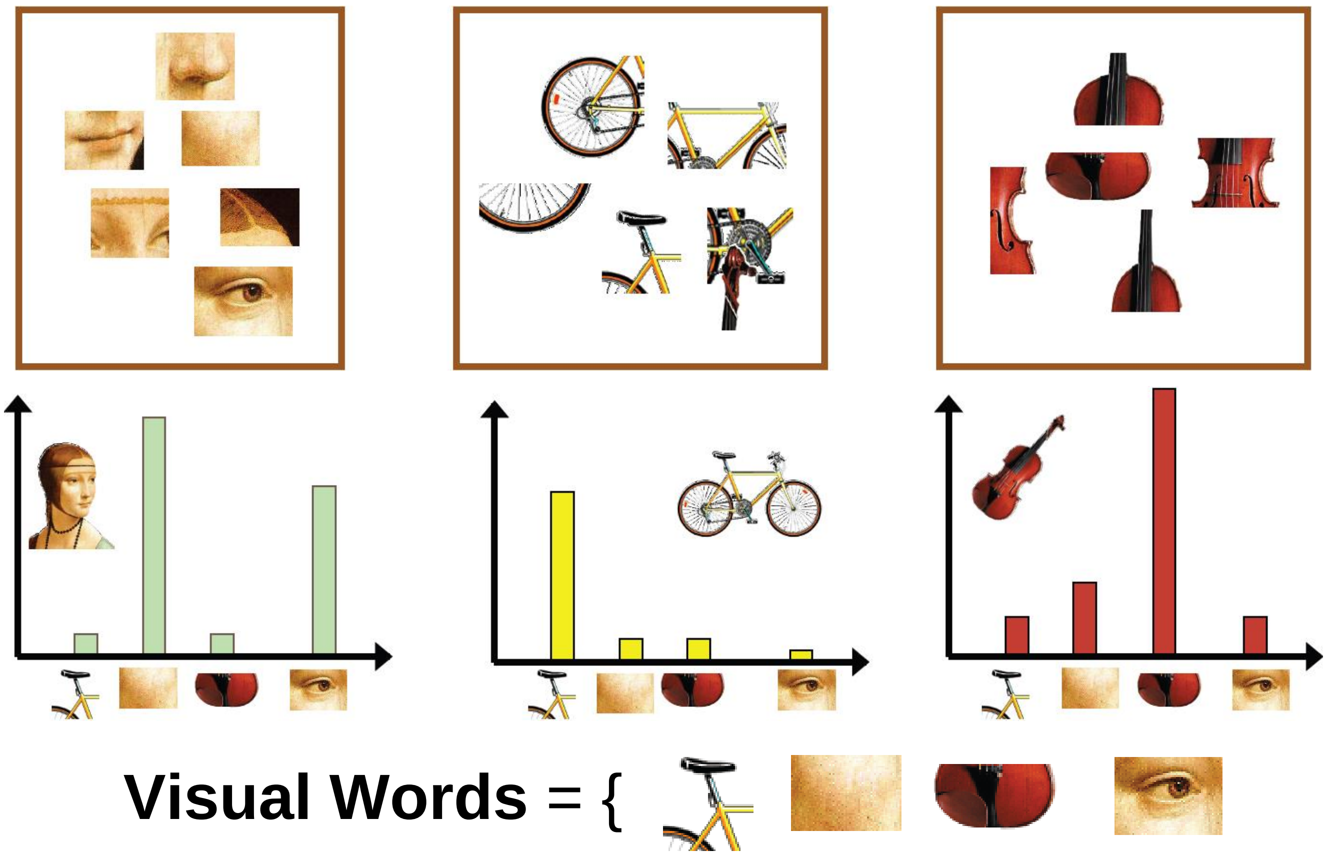


Credit: Fei-Fei Li

BoW Model for "Image Retrieval"

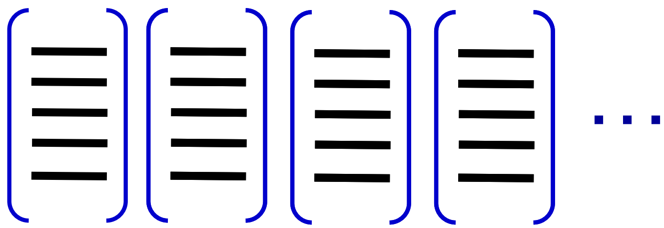
- Image as **composition of visual features (e.g. SIFTs)**

  → Words are discrete, but **visual features are real-valued**

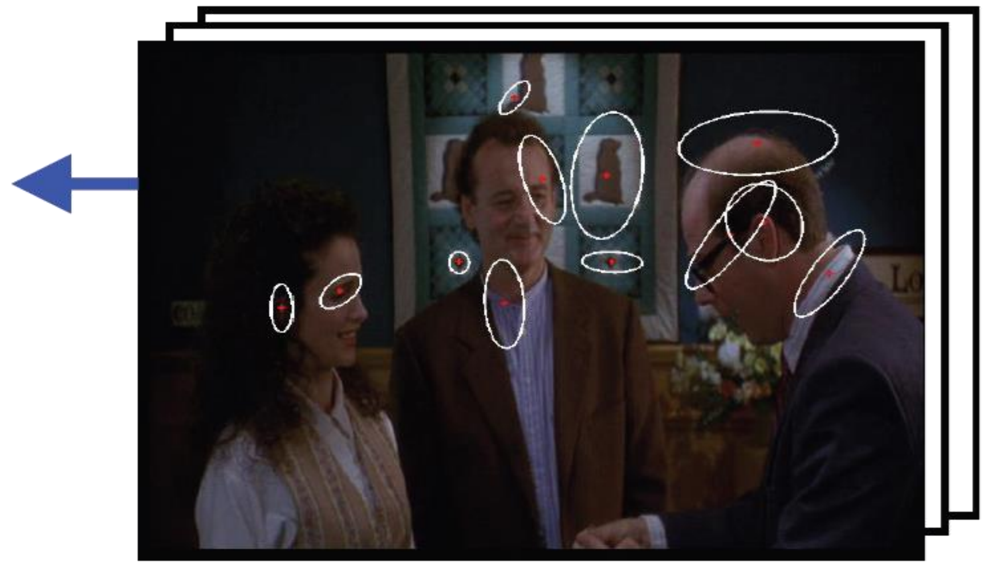


Credit: Fei-Fei Li

Visual Word Learning (1/3)



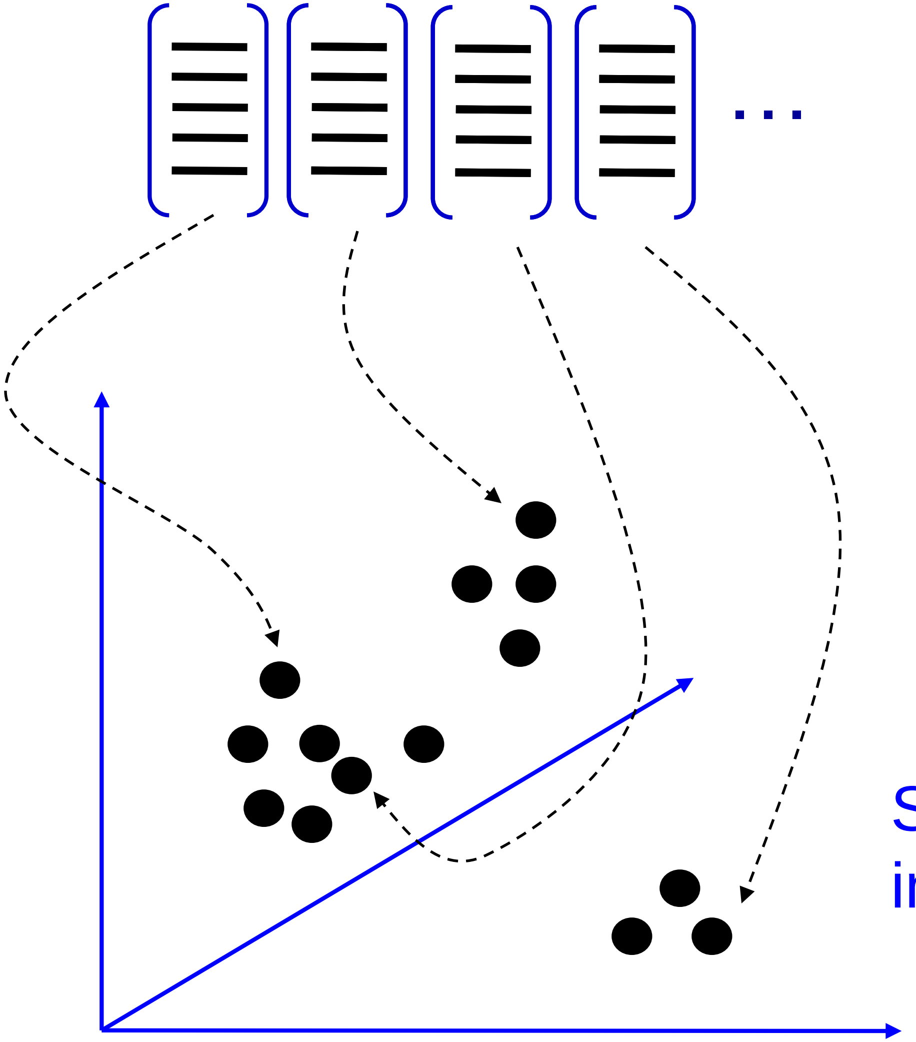
Compute
SIFT
descriptors



Detect keypoints from training images

Credit: Fei-Fei Li

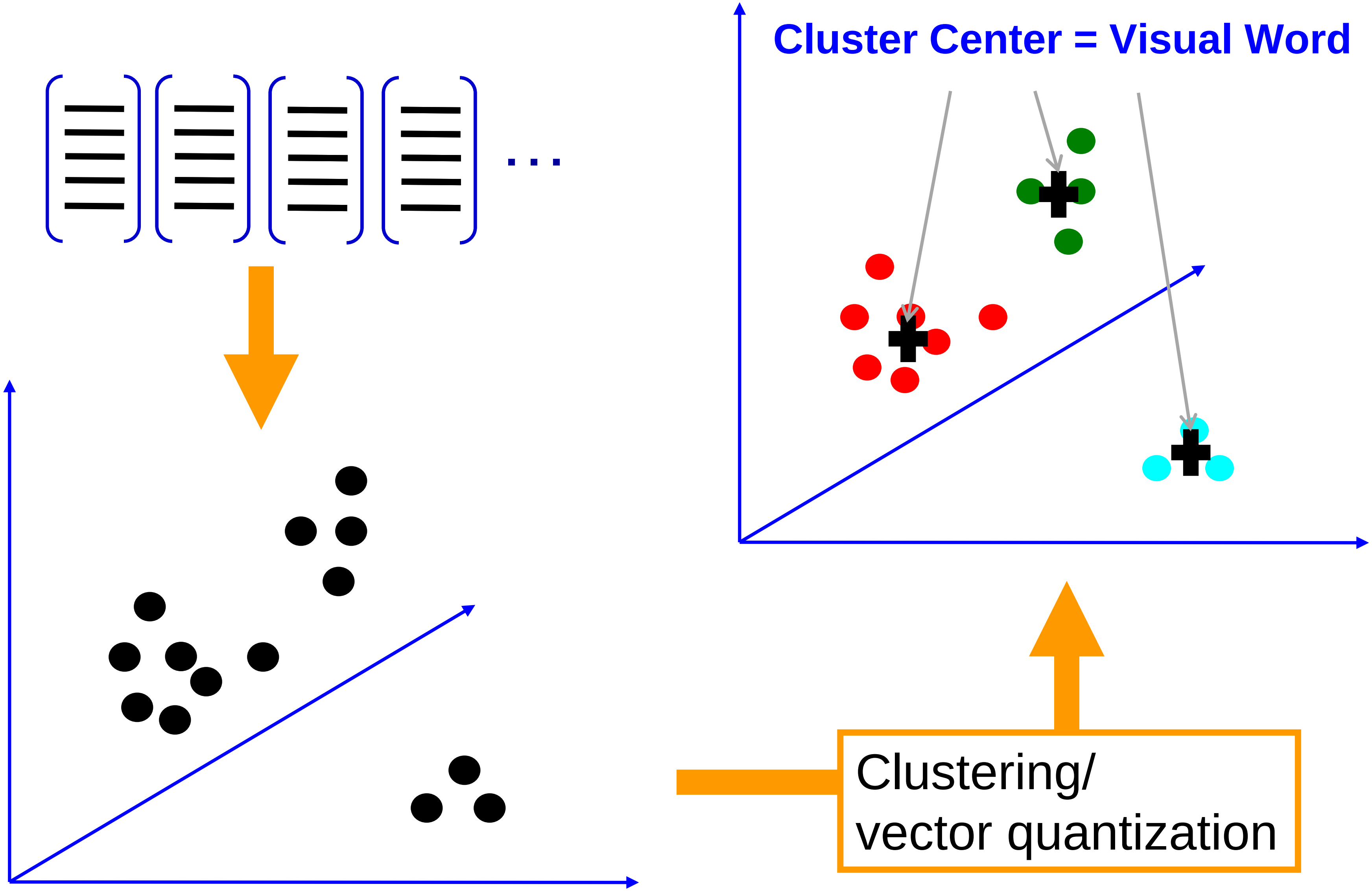
Visual Word Learning (2/3)



SIFT descriptors of training images in **feature space**

Credit: Fei-Fei Li

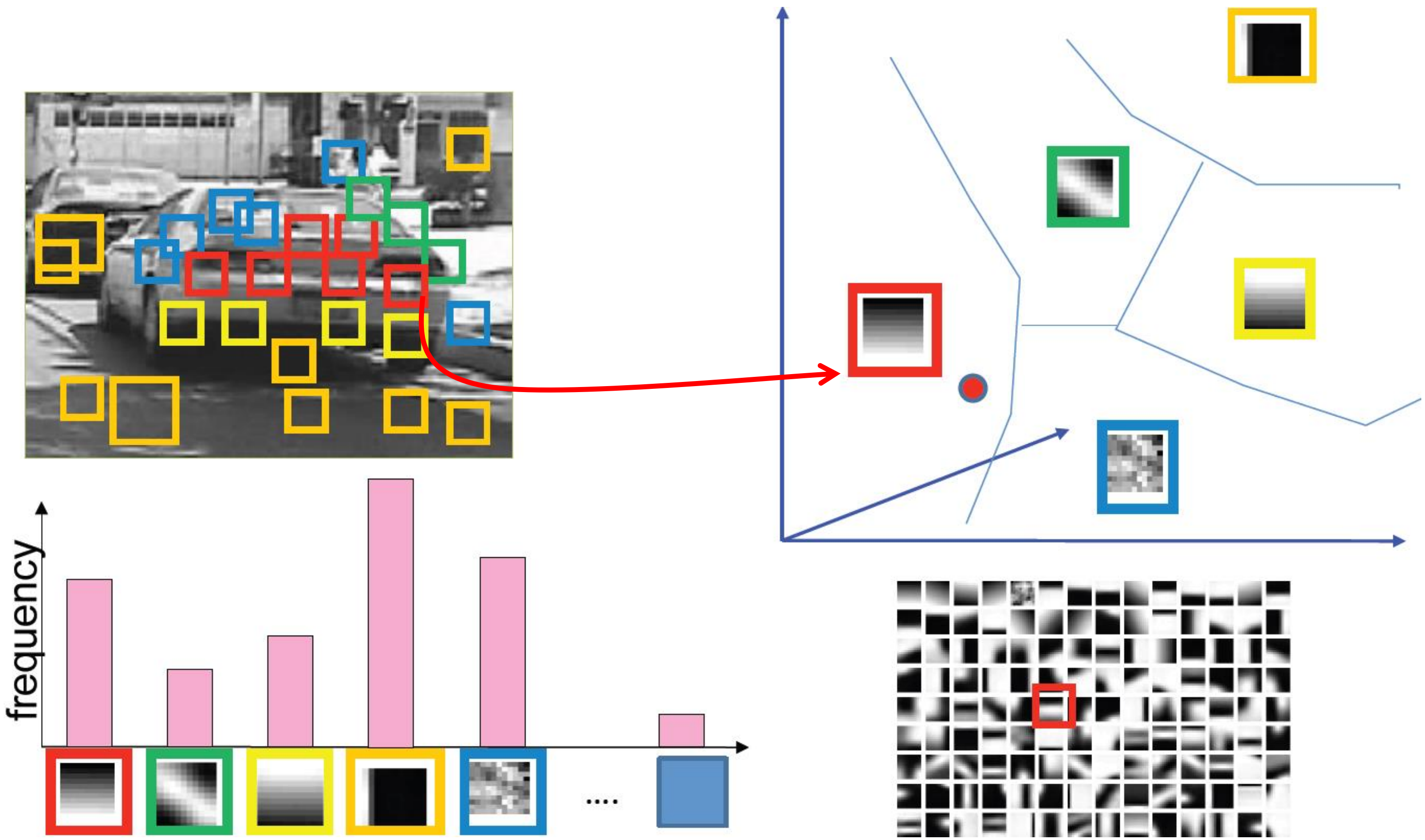
Visual Word Learning (3/3)



Credit: Fei-Fei Li

BoW: Coding & Pooling

- **Coding --** Quantize local features (e.g. SIFT) into visual words
- **Pooling --** Summarize visual words into a single vector

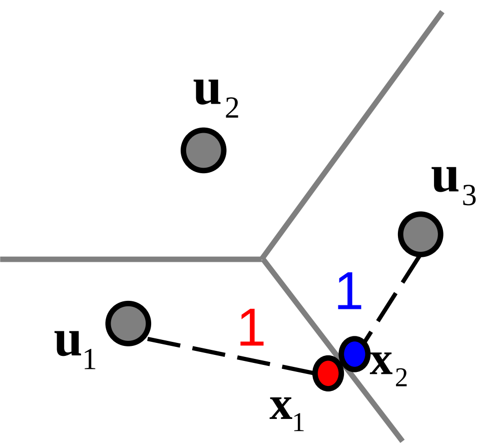


Credit: Fei-Fei Li

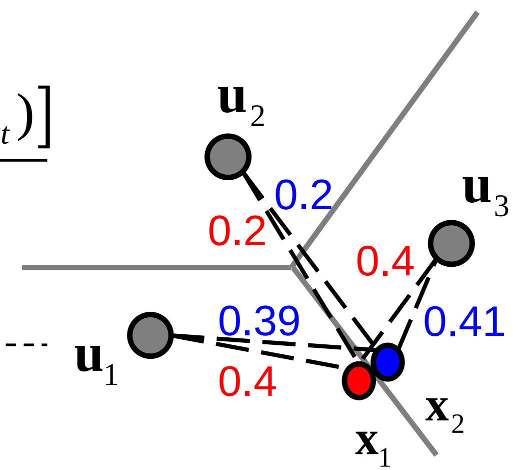
Generalization on Coding & Pooling

COLUMBIA UNIVERSITY

- **Hard assignment (HA)**
  - Assign a local feature to its nearest visual word
- **Soft assignment (SA)**
  - Assign a local feature to multiple visual words with weights (which may depend on their distances to the feature)



Coding	$\mathbf{x}_t \rightarrow \mathbf{r}(\mathbf{x}_t) = [r_1(\mathbf{x}_t),\ r_2(\mathbf{x}_t),\ r_3(\mathbf{x}_t)]$
HA	$\mathbf{x}_1 \rightarrow \mathbf{r}(\mathbf{x}_1) = [\ 1\ ,\ 0\ ,\ 0\ ]$
	$\mathbf{x}_2 \rightarrow \mathbf{r}(\mathbf{x}_2) = [\ 0\ ,\ 0\ ,\ 1\ ]$
SA	$\mathbf{x}_1 \rightarrow \mathbf{r}(\mathbf{x}_1) = [\ 0.4\ ,\ 0.2,\ 0.4\ ]$
	$\mathbf{x}_2 \rightarrow \mathbf{r}(\mathbf{x}_2) = \underbrace{[0.39,\ 0.2,\ 0.41]}$



Pooling: element-wise max/min vs. simple averaging

CDVS: Probabilistic Generative Model of SIFT

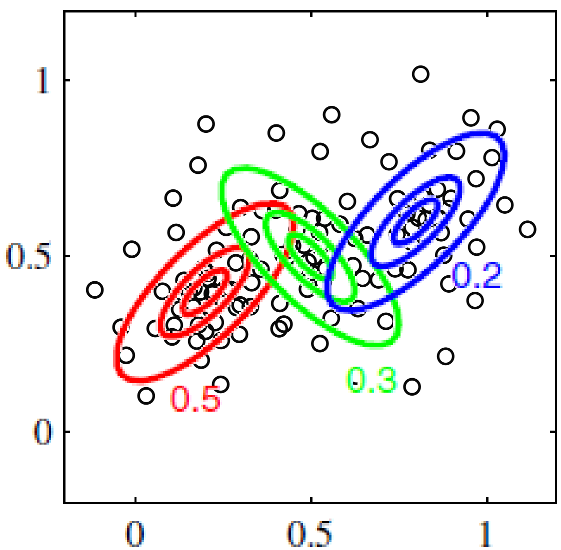
- Local features (SIFT) follow **Gaussian Mixture Model (GMM)**

GMM: Linear combinations of multiple (K) Gaussian distributions

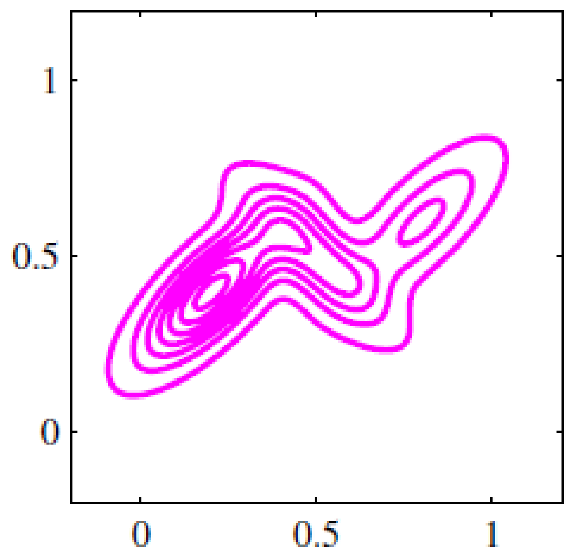
$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{i=1}^{K} \pi_i \mathcal{N}(\mathbf{x}|\mu_i, \Sigma_i), \boldsymbol{\theta} = \{\boldsymbol{\pi}, \mu, \boldsymbol{\Sigma}\}$$

(mean vectors u_i = visual words)

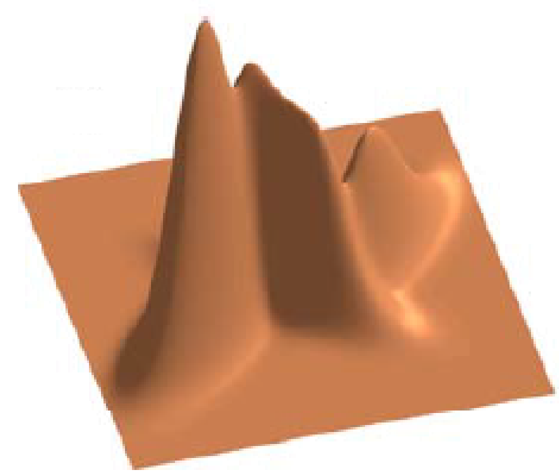
Contour of each Gaussian



Contour of $p(\mathbf{x})$



MoG distribution $p(\mathbf{x})$



CDVS: Fisher Vector (1/2)

- **Definition:** gradient w.r.t. parameter vector

$$\mathrm{g}(\mathrm{x},\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} p(\mathrm{x}|\boldsymbol{\theta})$$

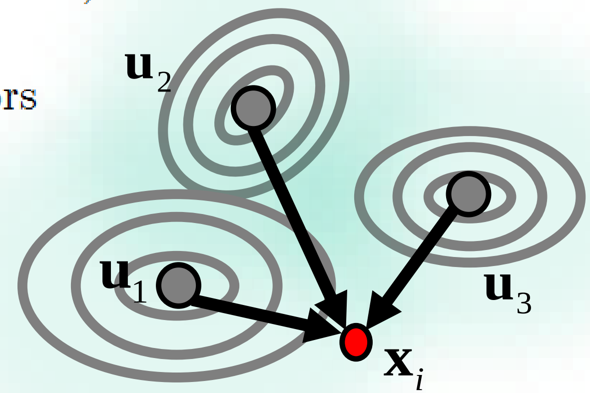
- **Result** (w.r.t mean vectors only):

$$\mathbf{x}_t \rightarrow \mathbf{r}(\mathbf{x}_t) = \left[\frac{r_1(\mathbf{x}_t)}{\sqrt{\pi_1}} \left(\frac{\mathbf{x}_t - \mathbf{u}_1}{\boldsymbol{\sigma}_1} \right), \frac{r_2(\mathbf{x}_t)}{\sqrt{\pi_2}} \left(\frac{\mathbf{x}_t - \mathbf{u}_2}{\boldsymbol{\sigma}_2} \right), \frac{r_3(\mathbf{x}_t)}{\sqrt{\pi_3}} \left(\frac{\mathbf{x}_t - \mathbf{u}_3}{\boldsymbol{\sigma}_3} \right) \right]$$

$r_i(x_t)$: Posterior prob. of visual word i given x_t (like SA)

$\frac{1}{\sqrt{\pi_i}}$: Discounting frequent visual words (like IDF)

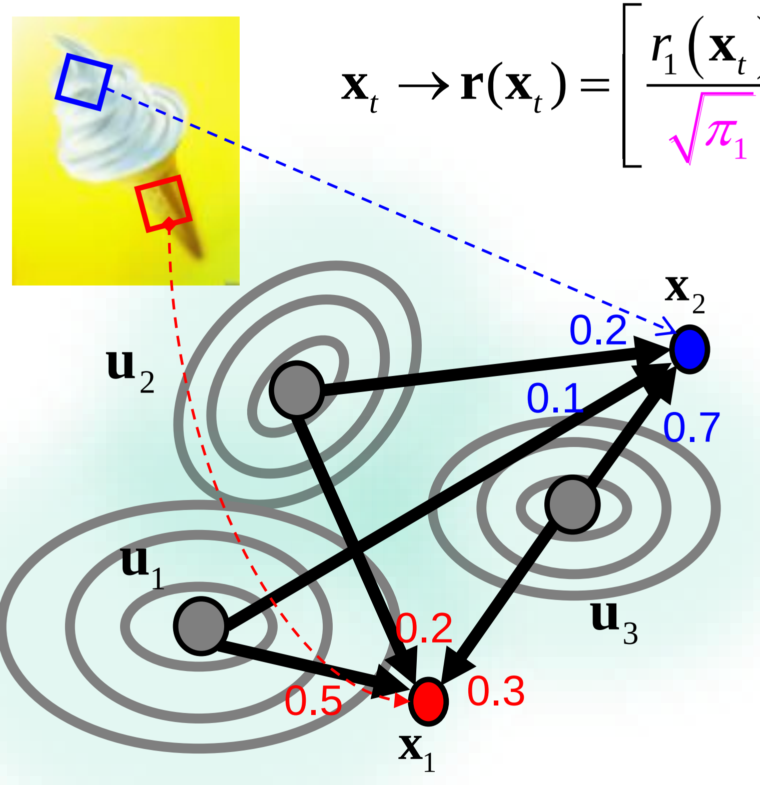
$\frac{x_t - \mu_i}{\sigma_i}$: Normalized deviations from mean vectors



CDVS: Fisher Vector (2/2)

- **Example:** local features (x1, x2) encoded & pooled together

$$\mathbf{x}_t \rightarrow \mathbf{r}(\mathbf{x}_t) = \left[\frac{r_1(\mathbf{x}_t)}{\sqrt{\pi_1}} \left(\frac{\mathbf{x}_t - \mathbf{u}_1}{\boldsymbol{\sigma}_1} \right), \frac{r_2(\mathbf{x}_t)}{\sqrt{\pi_2}} \left(\frac{\mathbf{x}_t - \mathbf{u}_2}{\boldsymbol{\sigma}_2} \right), \frac{r_3(\mathbf{x}_t)}{\sqrt{\pi_3}} \left(\frac{\mathbf{x}_t - \mathbf{u}_3}{\boldsymbol{\sigma}_3} \right) \right]$$



Step 1: Fisher Vector Encoding

$$\mathbf{x}_1 \rightarrow \mathbf{r}(\mathbf{x}_1) = \left[0.5 \cdot (\rightarrow),\ 0.2 \cdot (\searrow),\ 0.3 \cdot (\swarrow) \right]$$

$$\mathbf{x}_2 \rightarrow \mathbf{r}(\mathbf{x}_2) = \left[0.1 \cdot (\rightarrow),\ 0.2 \cdot (\nearrow),\ 0.7 \cdot (\nearrow) \right]$$

Step 2: Average Pooling

$$\mathbf{g} = [\mathbf{g}_1,\ \mathbf{g}_2,\ \mathbf{g}_3] = [(\nearrow),\ (\rightarrow),\ (\swarrow)]$$

Step 3: Sign Binarizaiton & Encoding

$$[[1, 0, 0, 1]\ [1, 1, 0, 1]\ [1, 1, 1, 0]]$$

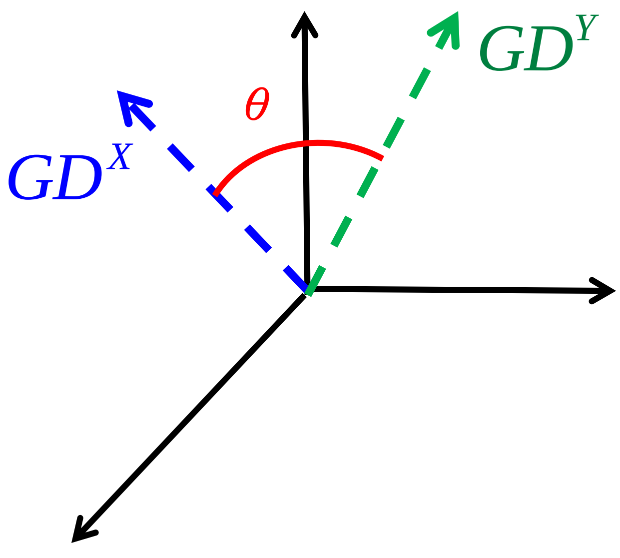
Outline

- **Part I – ISO/IEC Moving Picture Experts Group (MPEG)**
  - Background
  - Recent Milestones
  - Future Video, Machine Learning, and Media Internet-of-Things
- **Part II – MPEG Compact Descriptor for Visual Search (CDVS)**
  - Large-scale Image Retrieval
  - Local Image Descriptors
  - Global Image Descriptors
  - Image Matching
  - Use Case: Mobile Indoor Navigation
- **Part III – Cross-domain Data Retrieval**
  - Canonical Correlation Analysis
  - Deep Boltzmann Machine

- CDVS measures image similarity by **cosine similarity**
  [Note: Not part of the standard specification]

$$GD^X = \left[[1,-1,-1,1] \ [1,1,-1,1] \ [1,1,1,-1] \right]$$
$$GD^Y = \left[[-1,1,1,-1] \ [1,-1,1,1] \ [-1,1,1,1] \right]$$

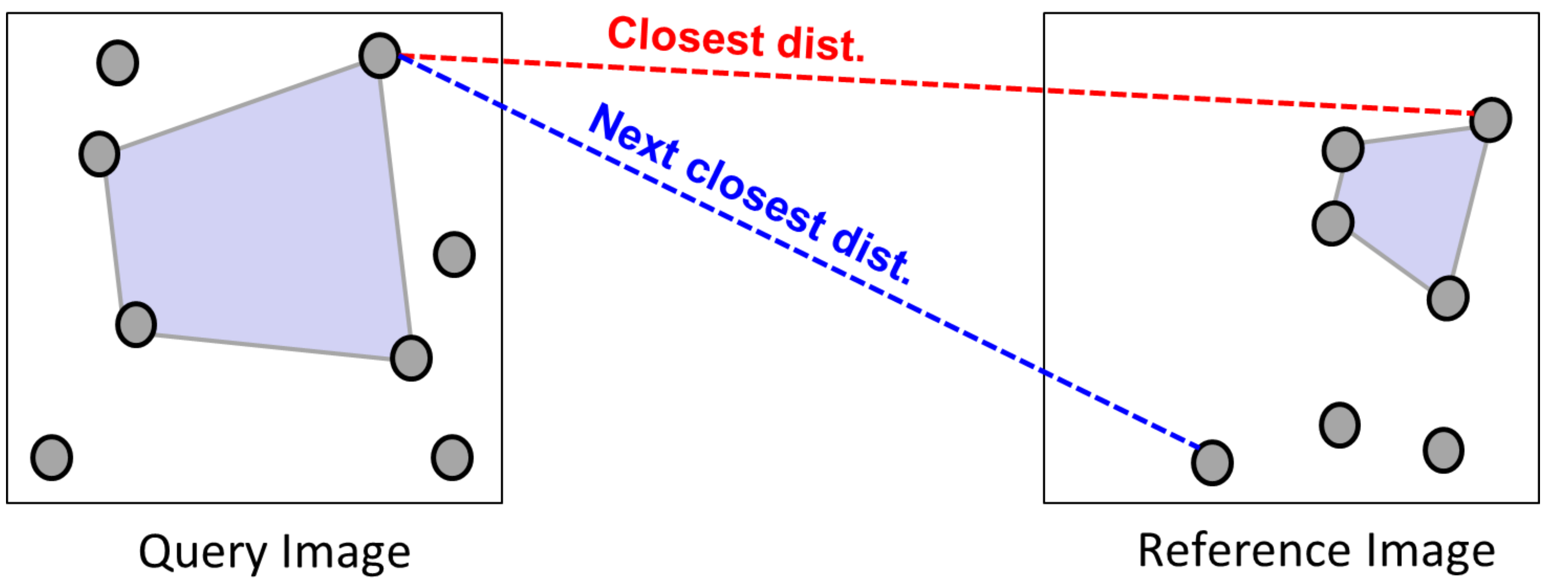
$$\cos\theta = \frac{\left(GD^X\right)^T GD^Y}{\left\|GD^X\right\|_2 \left\|GD^Y\right\|_2}$$



- Other metrics are possible -- Lp distance, Hamming distance, etc

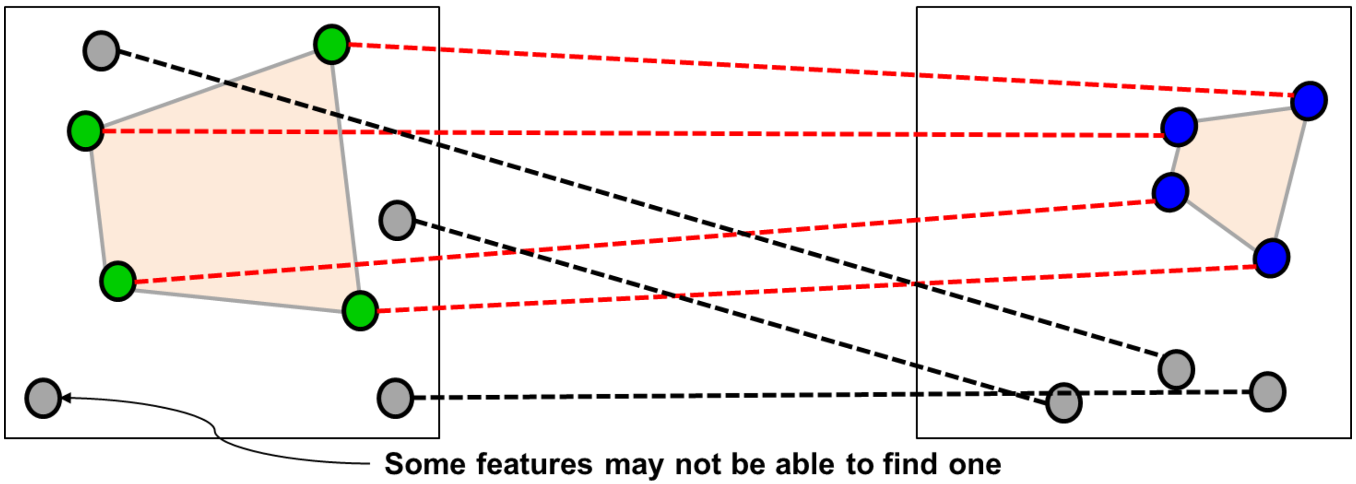
- (One-to-One Correspondence) For each LD, find its "reliable" correspondence by checking **Feature-space Distance Ratio**

$$distance\ ratio = \frac{\textbf{Closest dist.}}{\textbf{Next closest dist.}} < \textbf{threshold}$$

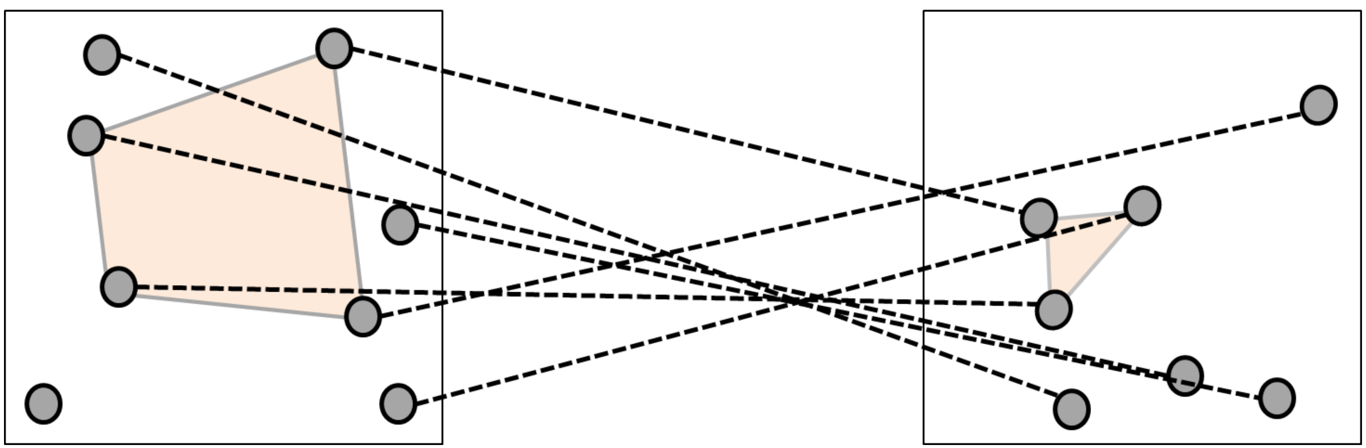


Local Descriptor (LD) Matching (2/5)

Case 1: An object shows up in two images, and usually shows a certain level of geometry relation between them



Case 2: Two images contain different objects

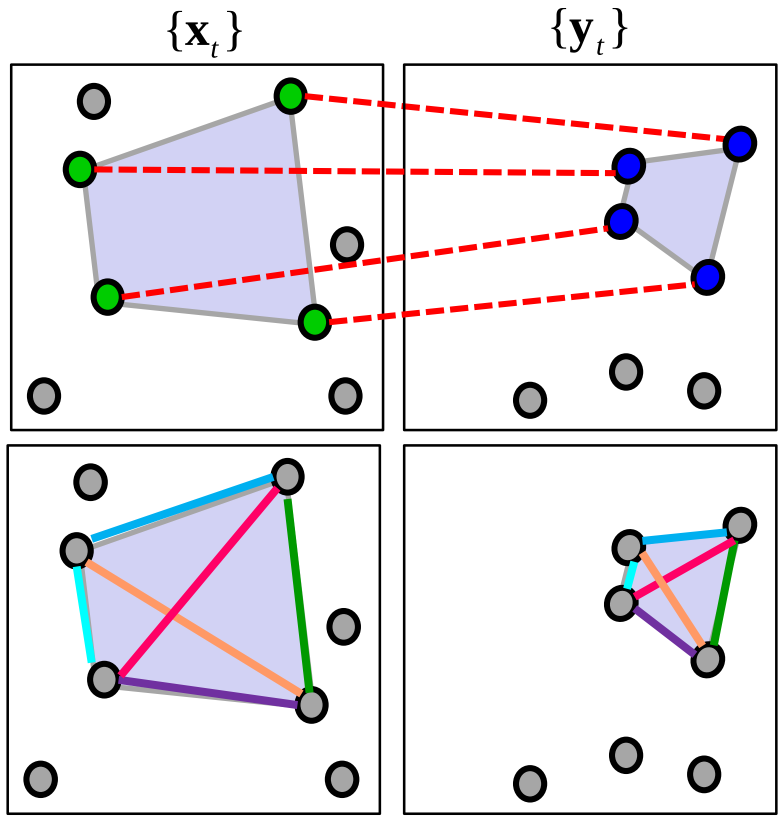


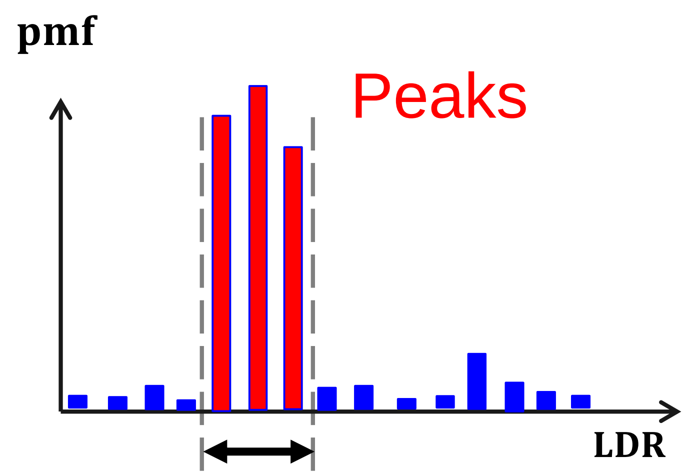
Local Descriptor (LD) Matching (3/5)

Log Distance Ratio (LDR)

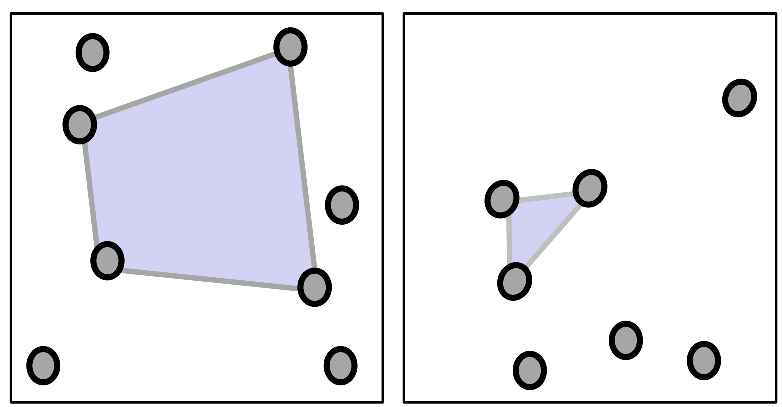
$$z_{ij} = \ln \frac{\left\| \mathbf{x}_i - \mathbf{x}_j \right\|_2}{\left\| \mathbf{y}_i - \mathbf{y}_j \right\|_2}$$

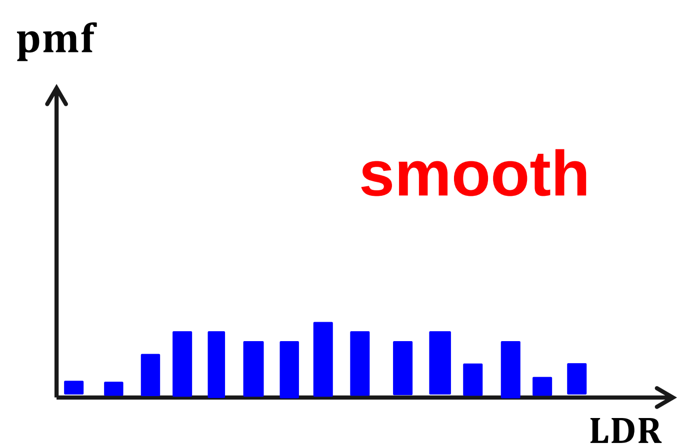
Match





Non-match





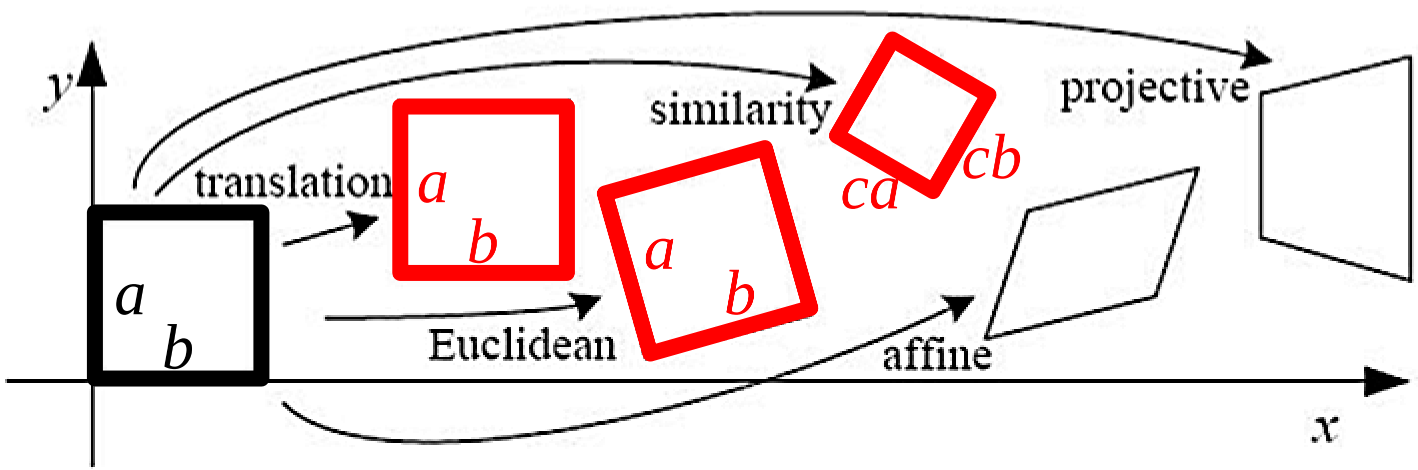
Local Descriptor (LD) Matching (4/5)

Caveat:

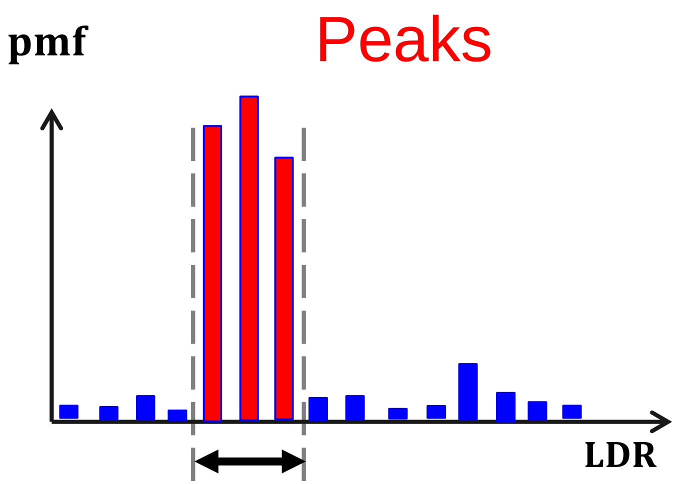
- Log Distance Ratio works for
  - Translation
  - Rotation
  - Scale change

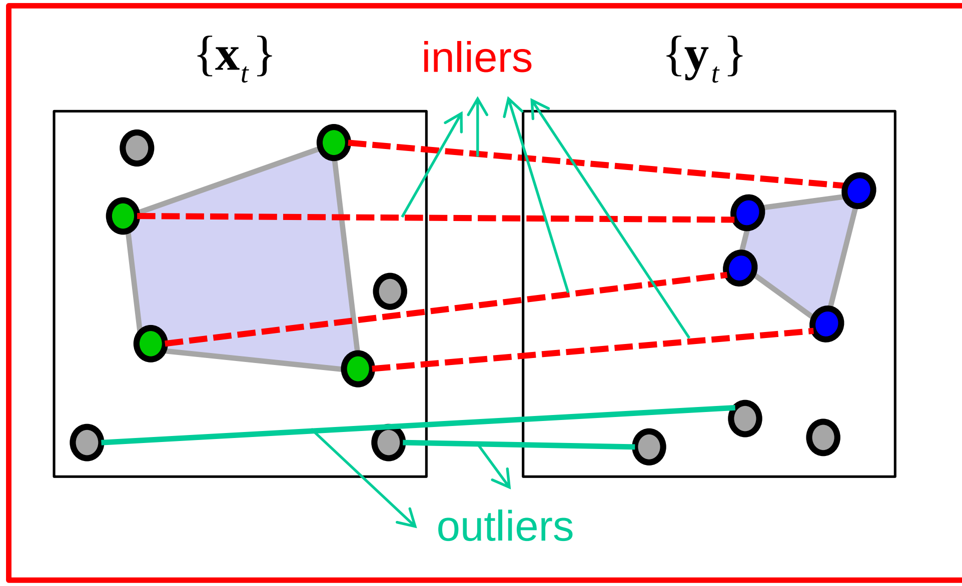
"Spatial-space Distance Ratio" is consistent, i.e., $\frac{a}{ca} = \frac{b}{cb} = \frac{1}{c}$

- But, NOT invariant to Affine & Projective transformations



Local Descriptor (LD) Matching (5/5)





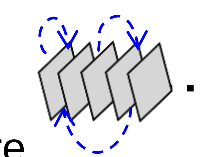
Select the pairs within the peaks

$$\begin{array}{cc} & \begin{array}{ccc} y_1 & \cdots & y_n \end{array} \\ \begin{array}{c} x_1 \\ \vdots \\ \vdots \\ x_n \end{array} & \begin{bmatrix} w_{11} & \cdots & \cdots & \cdots \\ \vdots & & & \\ \vdots & & & \\ \vdots & & & \end{bmatrix} \end{array}$$

Geometry check (eigenvalue and eigenvector)

Threshold (inliers and outliers)

Re-rank Score = sum of weights inliers



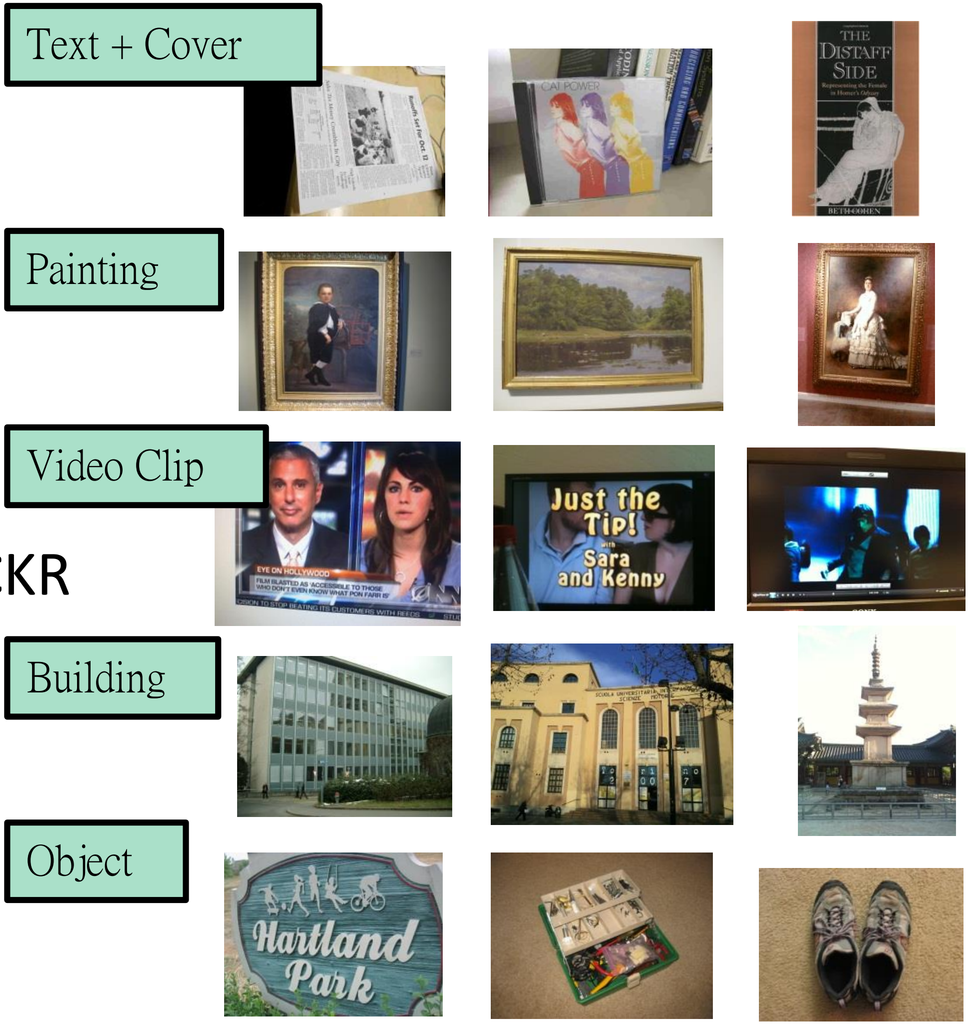
CDVS Test Conditions

Dataset

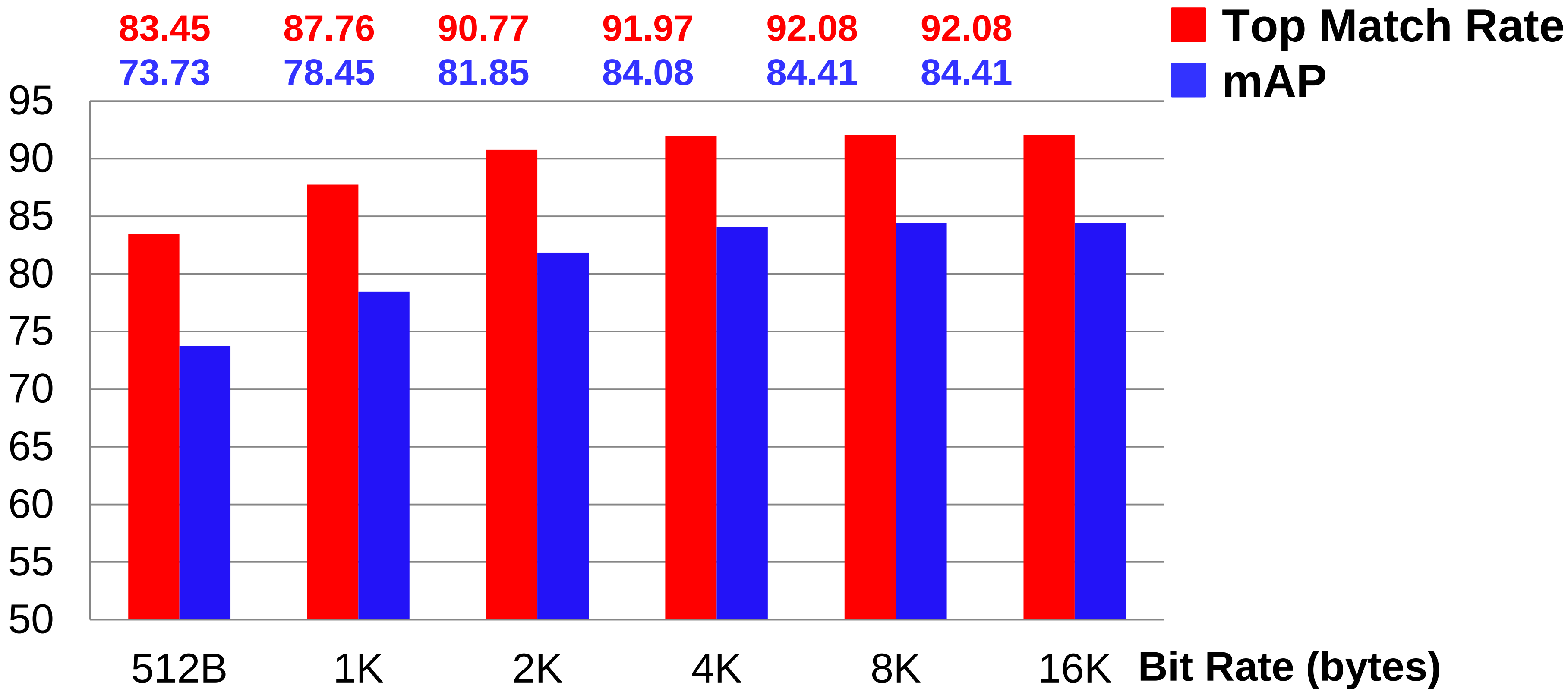
- 30K annotated images
  - Viewpoint change
  - Camera parameter
  - Lighting condition
  - Occlusion, clutter
- 1M distractor images from FLICKR

Descriptor Lengths (GD+LD's)

512, 1K, 2K, 4K, 8K, 16K bytes



CDVS TM11.0 Performance & Time Complexity



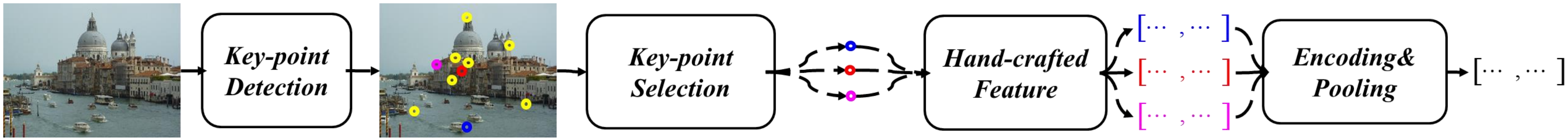
	512B	1K	2K	4K	8K	16K
Extraction (s)	0.33	0.33	0.33	0.35	0.40	0.43
Retrieval (s)	2.09	2.19	2.41	2.70	2.85	2.85

Windows 7
Intel i5-2410 3.2GHz

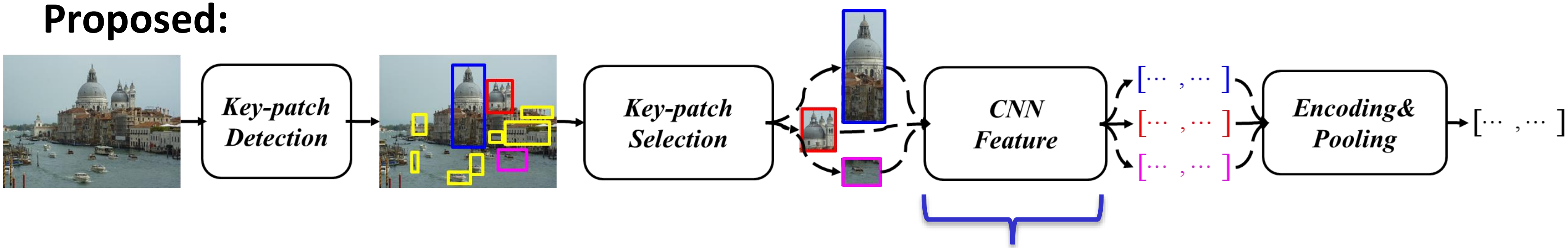
Remarks

- Easy to find better performing technology in terms of accuracy
  - SIFT → Convolutional Neural Network (CNN)
  - Keypoints → Object Proposals
  - Fisher Vector → Topic Modeling with Latent Dirichlet Allocation

MPEG CDVS:



Proposed:



Can be computationally more intensive than some video encoders

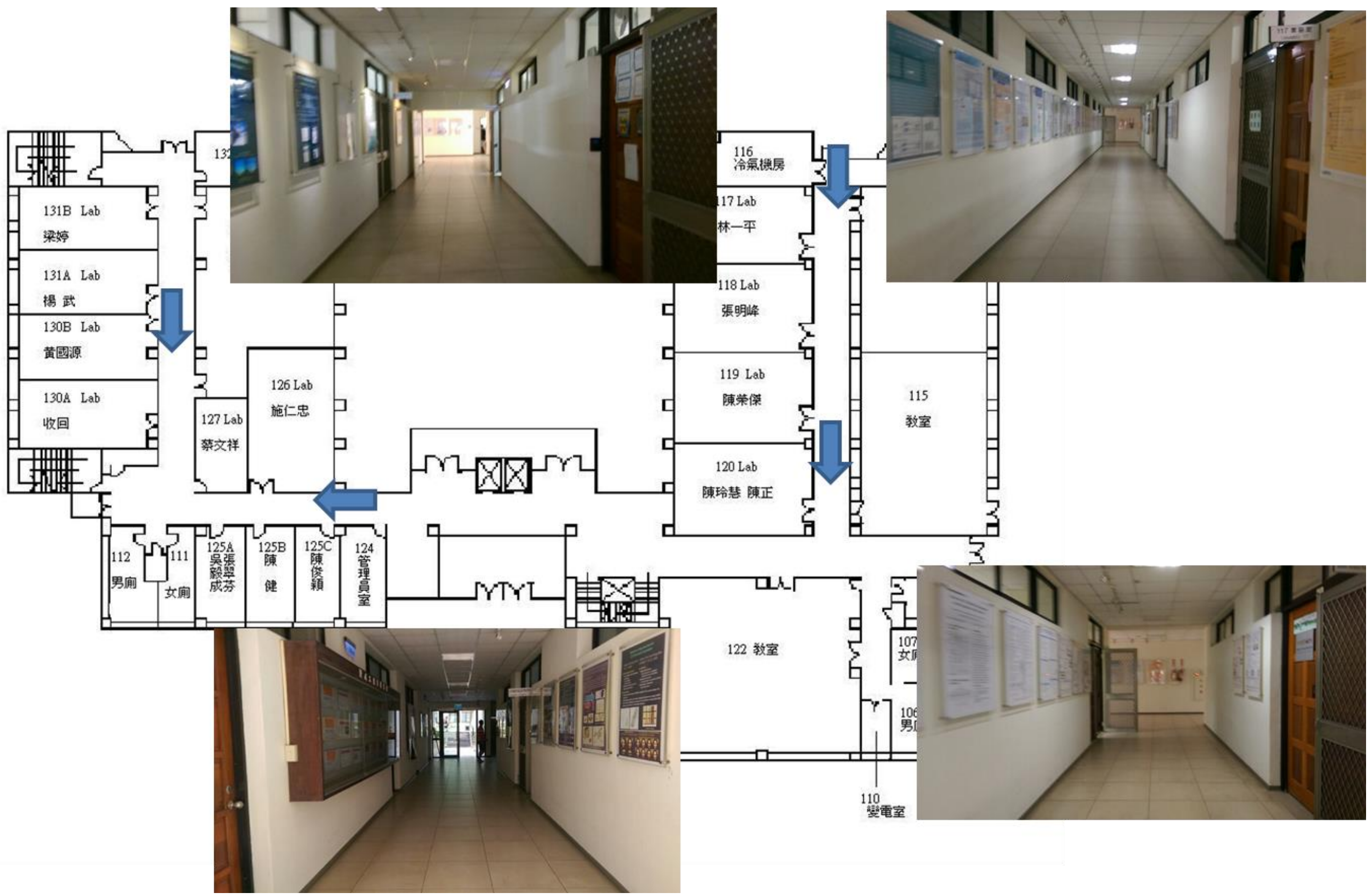
- All about trade-off between performance and complexity

Outline

- **Part I – ISO/IEC Moving Picture Experts Group (MPEG)**
  - Background
  - Recent Milestones
  - Future Video, Machine Learning, and Media Internet-of-Things
- **Part II – MPEG Compact Descriptor for Visual Search (CDVS)**
  - Large-scale Image Retrieval
  - Local Image Descriptors
  - Global Image Descriptors
  - Image Matching
  - Use Case: Mobile Indoor Navigation
- **Part III – Cross-domain Data Retrieval**
  - Canonical Correlation Analysis
  - Deep Boltzmann Machine

Use Case: Mobile Indoor Navigation

- Indoor navigation using **visual search** and **mobile sensor data**
- **<u>Database:</u>** images of indoor scenes with locations as metadata

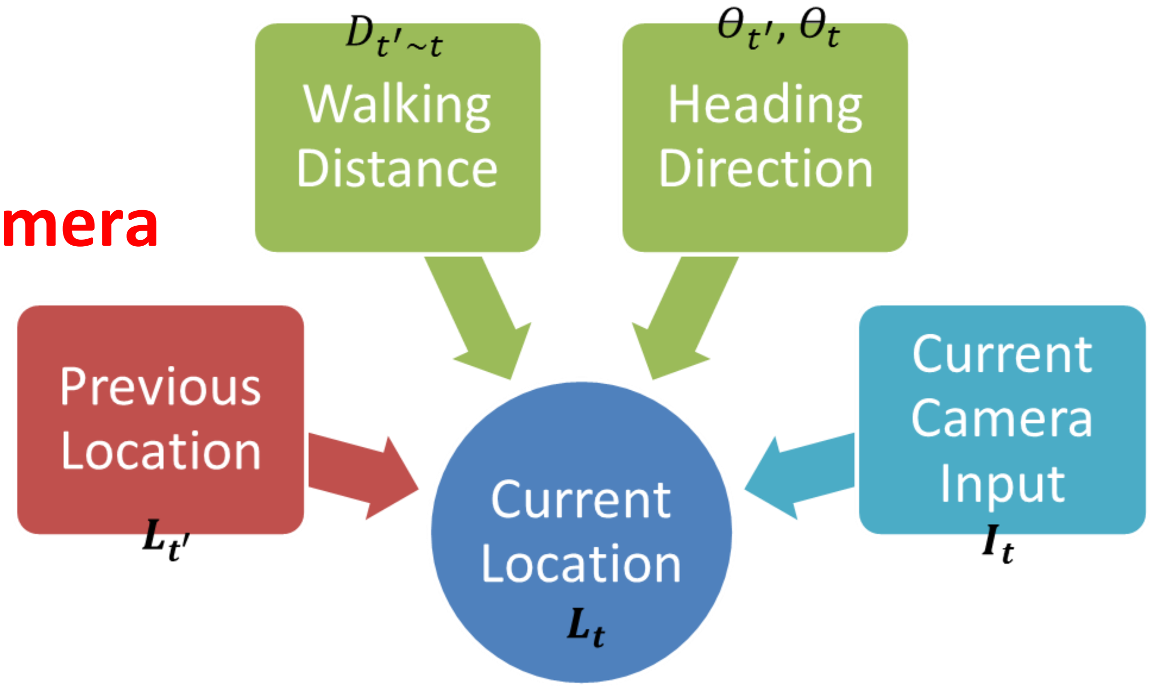


Credit: Wendy Tseng

Main Ideas (1/2)

$$P(\boldsymbol{L_t} =?\,|\boldsymbol{L_{t'}}, D_{t'\sim t}, \theta_{t'}, \theta_t, \boldsymbol{I_t})$$

Inertial Sensors & Camera



Credit: Wendy Tseng

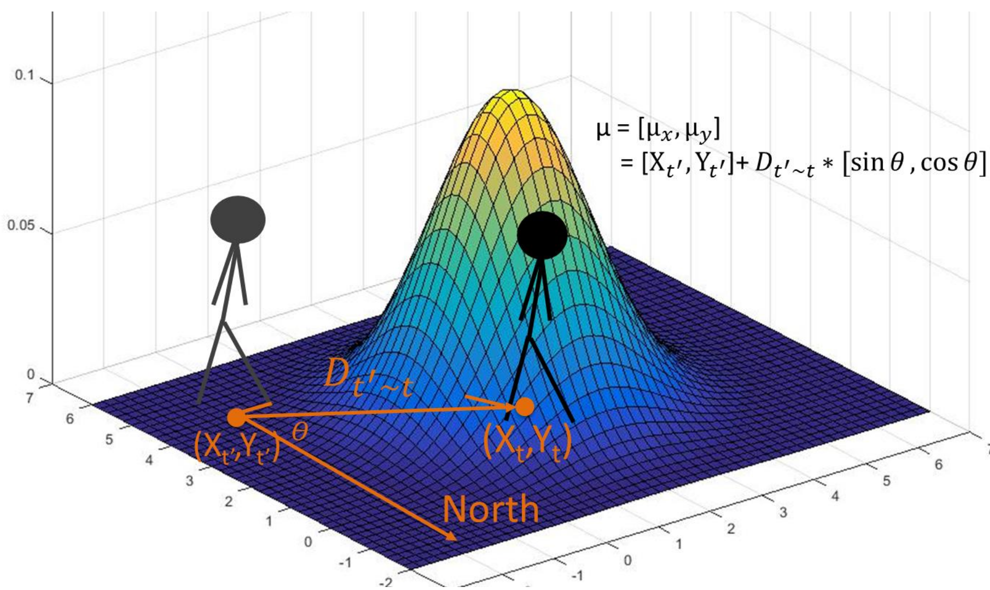
Main Ideas (2/2)

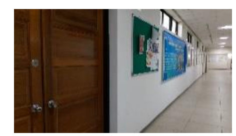
$$P(\boldsymbol{L_t} = ?\,|\,\boldsymbol{L_{t'}}, D_{t'\sim t}, \theta_{t'}, \theta_t, \boldsymbol{I_t})$$

$$\propto P(\boldsymbol{L_t} = ?\,|\,\boldsymbol{L_{t'}}, D_{t'\sim t}, \theta_{t'}, \theta_t)\; \mathrm{P}(\boldsymbol{I_t}|\boldsymbol{L_t} = ?)$$

Probability conditioned on previous location and sensors data

Likelihood function given query image





Location A

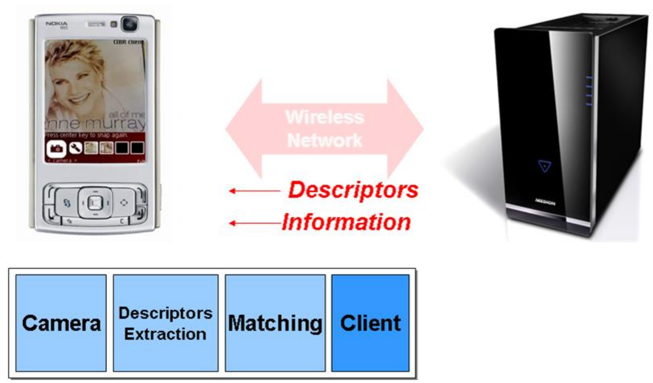
Location B

Location C

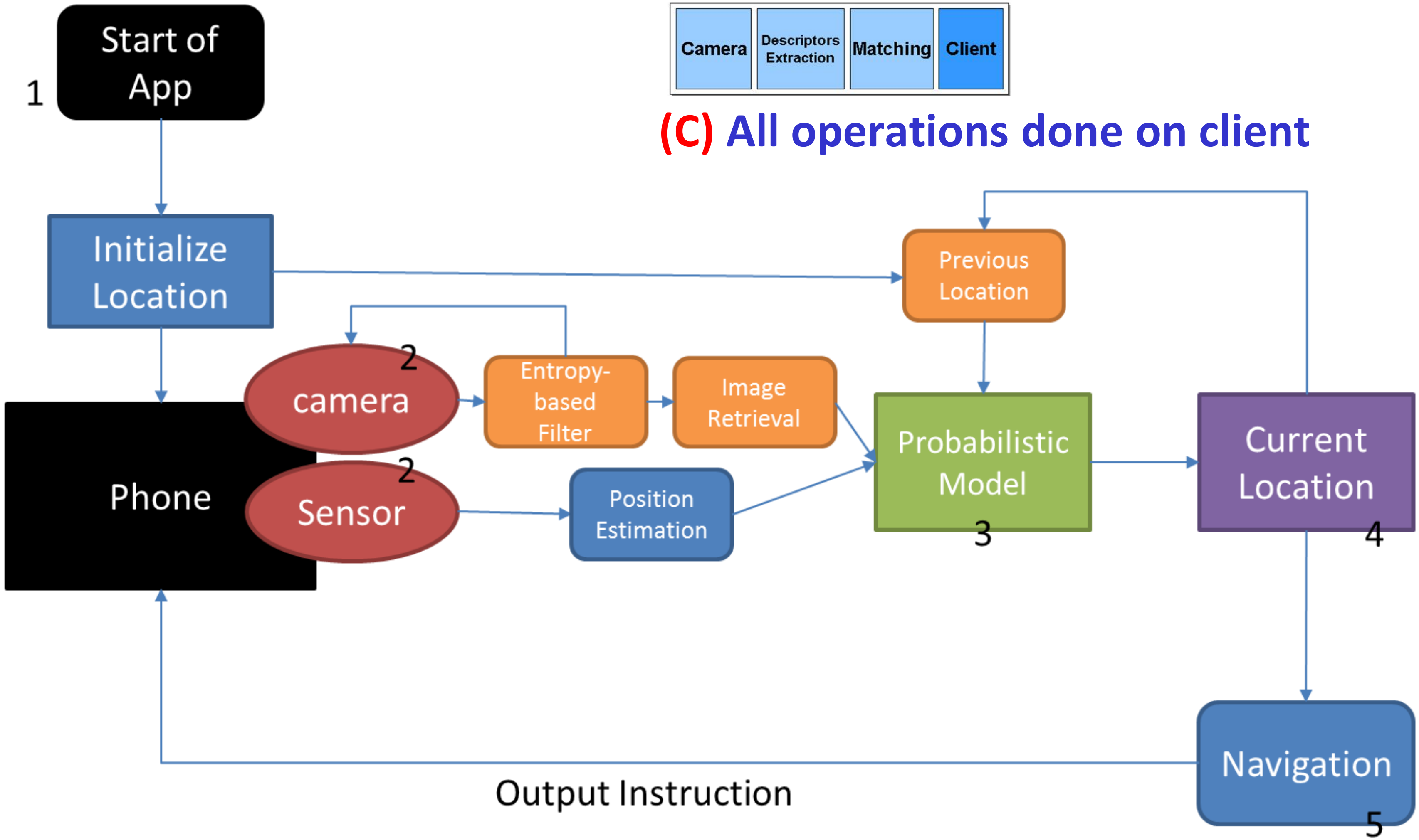
Location A

Credit: Wendy Tseng

Flow Chart



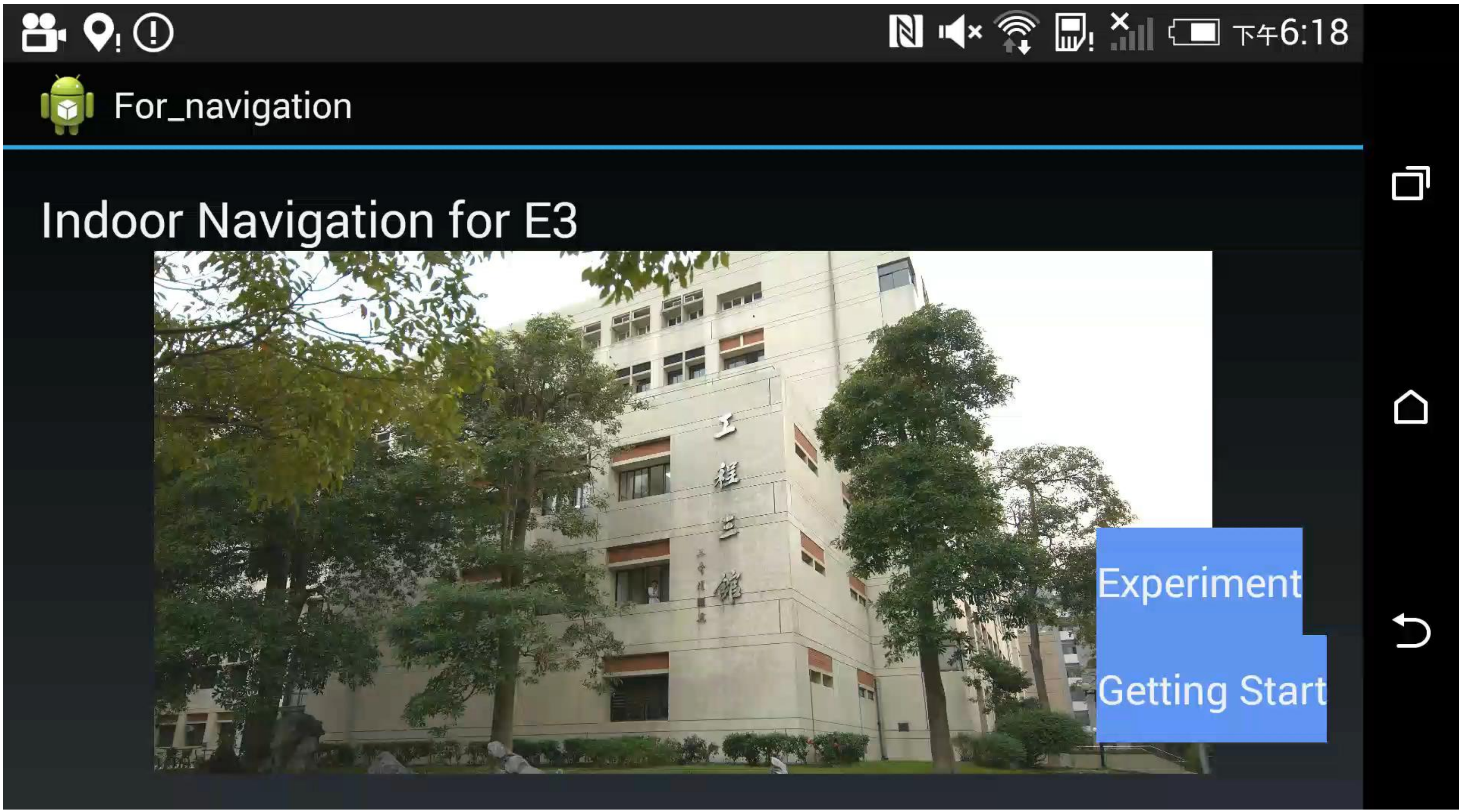
(C) All operations done on client



Credit: Wendy Tseng

Video Demo

- Floorplan divided into squares of size 1m x 1m (locations)
- Pictures taken from 8 directions at each location



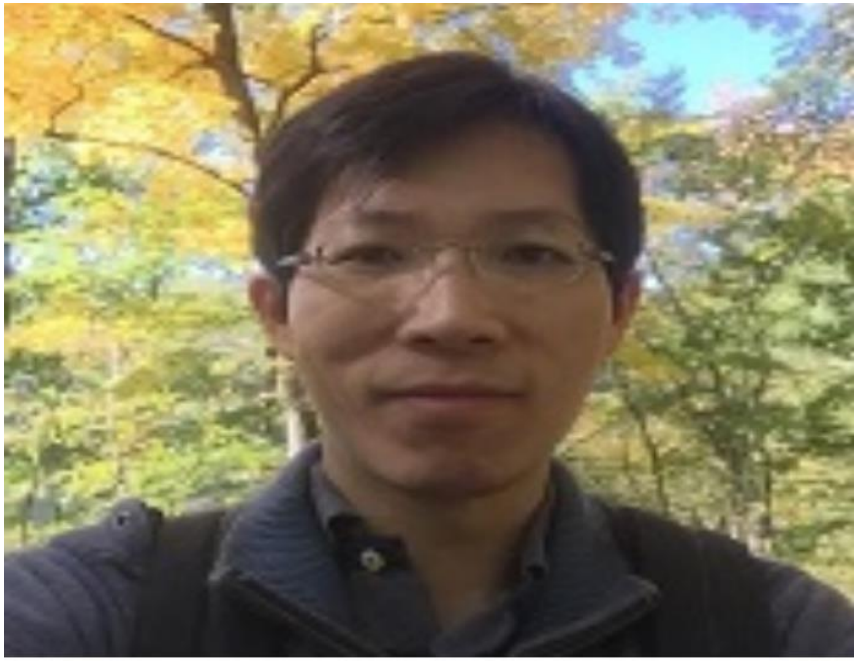
Credit: Wendy Tseng

Outline

- **Part I – ISO/IEC Moving Picture Experts Group (MPEG)**
  - Background
  - Recent Milestones
  - Future Video, Machine Learning, and Media Internet-of-Things
- **Part II – MPEG Compact Descriptor for Visual Search (CDVS)**
  - Large-scale Image Retrieval
  - Local Image Descriptors
  - Global Image Descriptors
  - Image Matching
  - Use Case: Mobile Indoor Navigation
- **Part III – Cross-domain Data Retrieval**
  - Canonical Correlation Analysis
  - Deep Boltzmann Machine

Cross-domain Data Retrieval: Image to Text

- Example: Image auto-captioning



1. a man in a shirt and tie standing in front of a tree -0.812122
2. a man in a black shirt and tie standing in front of a tree -0.874686
3. a man in a shirt and tie standing in front of a forest -0.877506
4. a man in a shirt and tie standing in front of trees -0.900125
5. a man in a black shirt and tie standing in front of a forest -0.922222

1. a black bear walking across a dirt road -0.813335
2. a couple of animals that are standing in the dirt -0.843556
3. a couple of animals walking across a dirt road -0.855182
4. a couple of animals that are standing in the grass -0.856973
5. a black bear walking across a lush green field -0.857406

Video auto-captioning also possible!!

Credit: Youssef Mroueh

Challenging Case??

1. a bunch of different types of scissors on a table -1.177987
2. a room with a lot of different types of items on it -1.222625
3. a bunch of different items are on a table -1.227481
4. a room with a lot of different types of items on the wall -1.325266
5. a room with a lot of different types of items -1.330639

Ground truth?

Credit: Youssef Mroueh

The Magic: Canonical Correlation Analysis

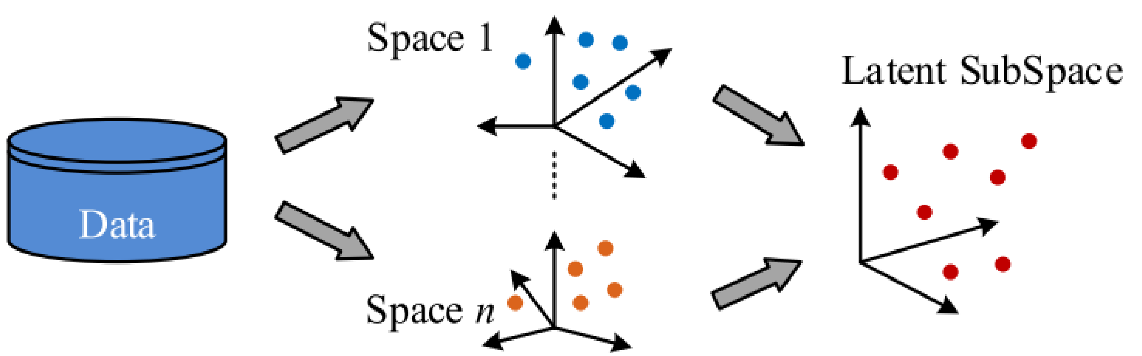
- Given **matching** images **x** and sentences **y** (viewed as vectors)

- **Objective:** Project **x** and **y** onto a latent subspace k

$$\begin{aligned}\widetilde{\mathbf{x}}_{k\times 1} &= \mathbf{U}_{k\times m}\mathbf{x}_{m\times 1} \\ \widetilde{\mathbf{y}}_{k\times 1} &= \mathbf{V}_{k\times n}\mathbf{y}_{n\times 1}\end{aligned}\quad \text{, where } k = \min(m, n)$$

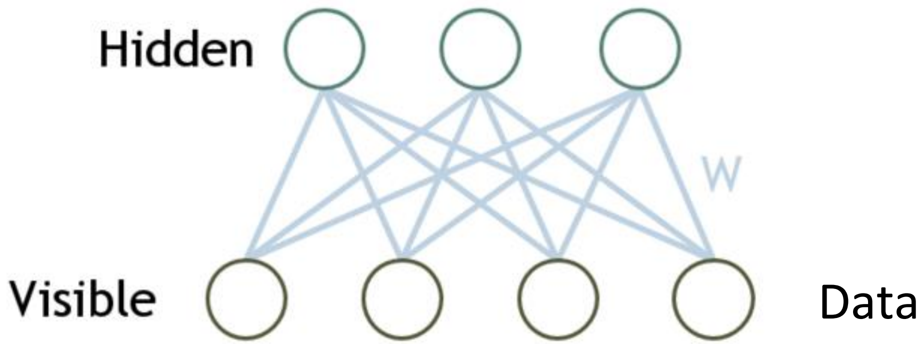
in which **their correlation is maximized**

$$\max_{\{\mathbf{U},\mathbf{V}\}} E\left(\widetilde{\mathbf{x}}^T\widetilde{\mathbf{y}}\right) \text{ s.t. } E(\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}^T) = \mathbf{I}_{k\times k} \text{ and } E(\widetilde{\mathbf{y}}\widetilde{\mathbf{y}}^T) = \mathbf{I}_{k\times k}$$



http://arxiv.org/pdf/1511.06267.pdf

- **Restricted Boltzmann Machine:** Generative Probabilistic Model (Markov Network)



Potential Function and Joint Distribution

$$E(v,h) = -\sum_{i \in visible} a_i v_i - \sum_{j \in hidden} b_j h_j - \sum_{i,j} v_i h_j w_{ij}$$

bias term

weight

$$\rightarrow P(v,h) = \frac{1}{Z} e^{-E(v,h)} \qquad for\ Z = \sum_{v,h} e^{-E(v,h)}$$

Posterior Probabilities

$$P(h_j = 1|v) = \sigma(b_j + \sum_i v_i w_{ij})$$

$$P(v_i = 1|h) = \sigma(a_i + \sum_j h_j w_{ij})$$

Sigmoid **Linear Form**

There are learning algorithms for estimating a, b, w from visible data

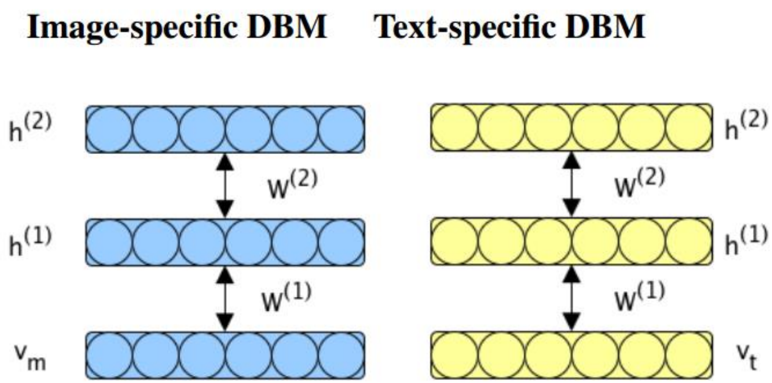
Source: N Srivastava, et al., "Multimodal Learning with Deep Boltzmann Machines".

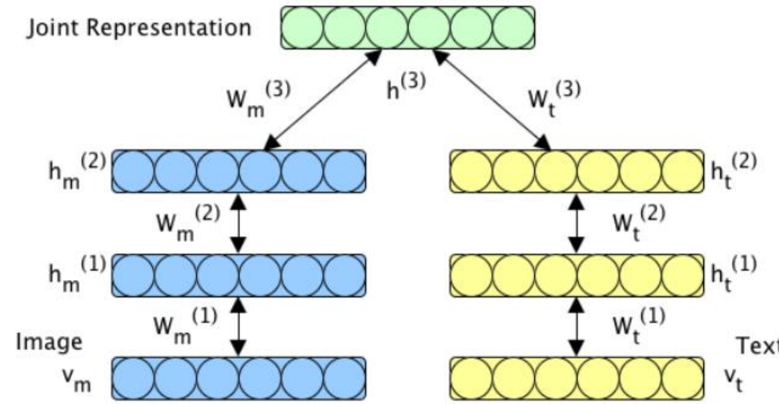
Multimodal Deep Boltzmann Machine (2/2)

Columbia University

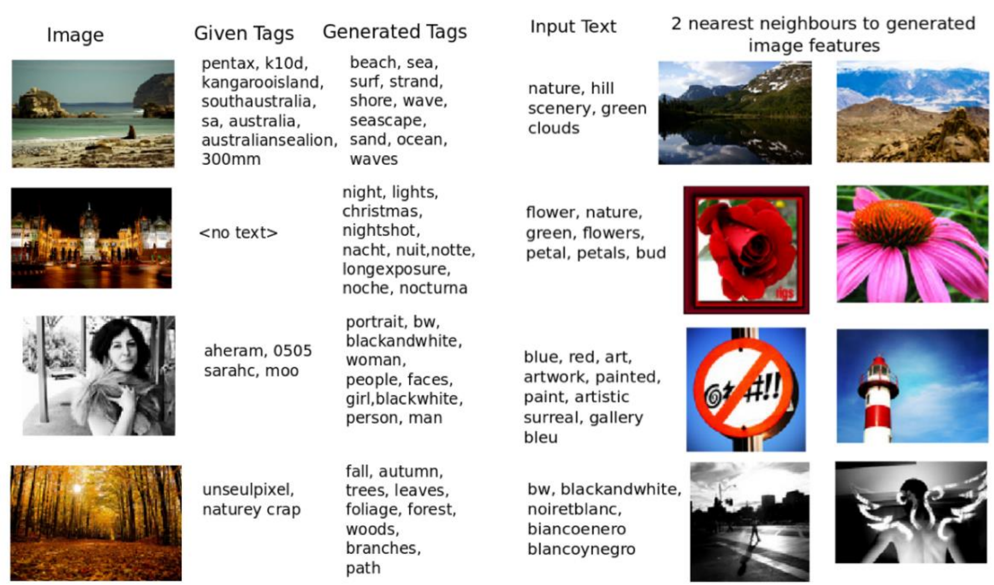
(1) Extension to multiple layers by layer-wise unsupervised learning

(2) Fine tuning network with supervised learning





Inference:



Source: N Srivastava, et al., “Multimodal Learning with Deep Boltzmann Machines”.

Thank You

Guest Speaker: Wen-Hsiao Peng

National Chiao Tung University (NCTU), Taiwan

Ching-Yung Lin, Ph.D.

Adjunct Professor, Dept. of Electrical Engineering and Computer Science

IBM Chief Scientist, Graph Computing