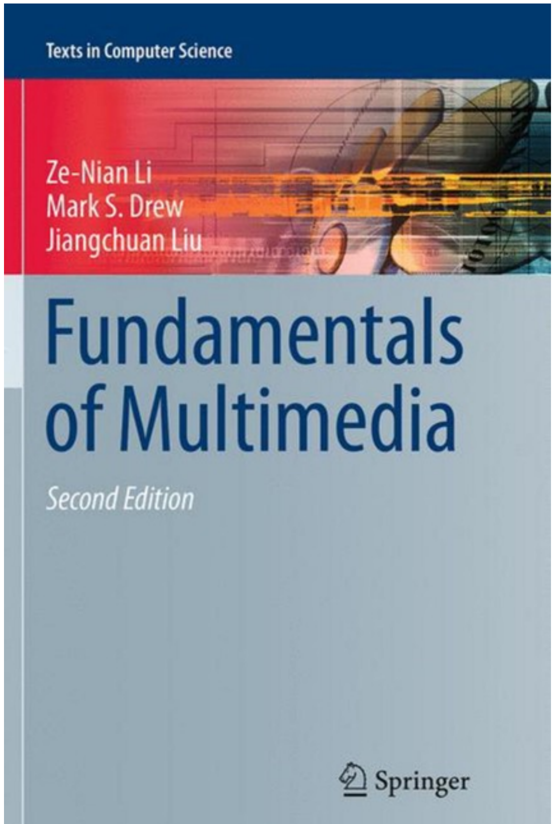
Background Photo:
Opera Garnier Ceiling

MULTIMEDIA

Dr Siba HAIDAR • INFO430 • 2019-2020

Textbook: Fundamentals of Multimedia • Z.-N. Li et al.

Course Outline



1. Introduction to multimedia
2. Digital representation of graphics and images
3. Colors in images and video
4. Fundamental Concepts in Video
5. Lossless compression algorithms
6. Lossy compression algorithms (JPEG)
7. **Video Coding (MPEG)**
8. Introduction to Image Processing

Background Photo:
The Starry Night

VIDEO CODING STANDARDS

Chapter Outline

- Video Coding Standards
- chapters 10, 11 and 12 in textbook

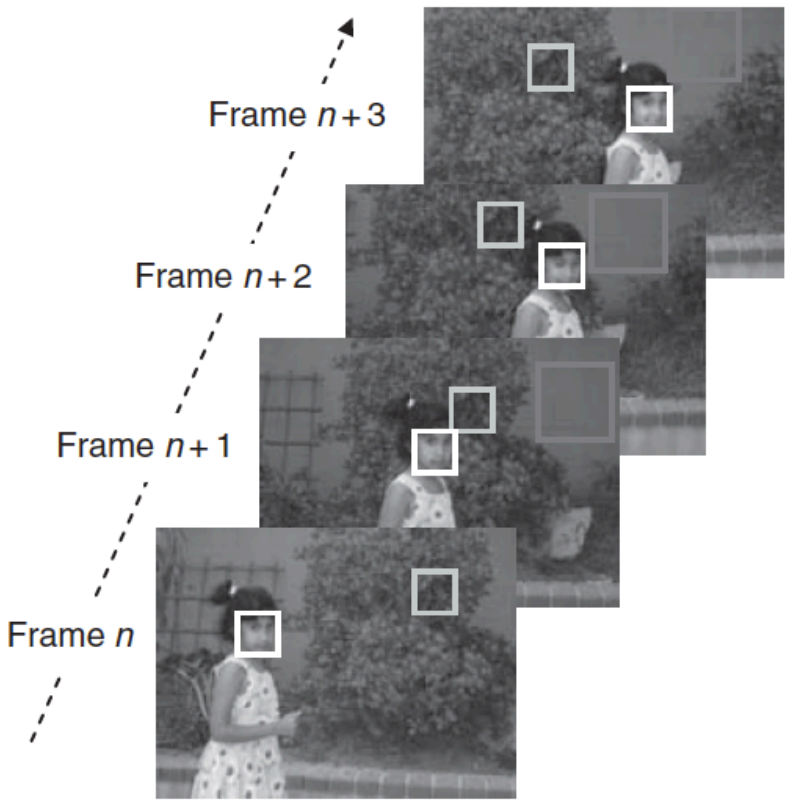
1. Basic Video Compression Techniques
2. Video Compression with Motion Compensation
3. Search for Motion Vectors
4. Compression Standard Committees & Compression Standards
5. H.261 & H.263
6. MPEG Video Coding — MPEG-1, 2, 4 & 7
7. New Video Coding Standards: H.264 & H.265

Background Photo:
The Milk Maid

Basic Video Compression Techniques

Introduction to Video Compression

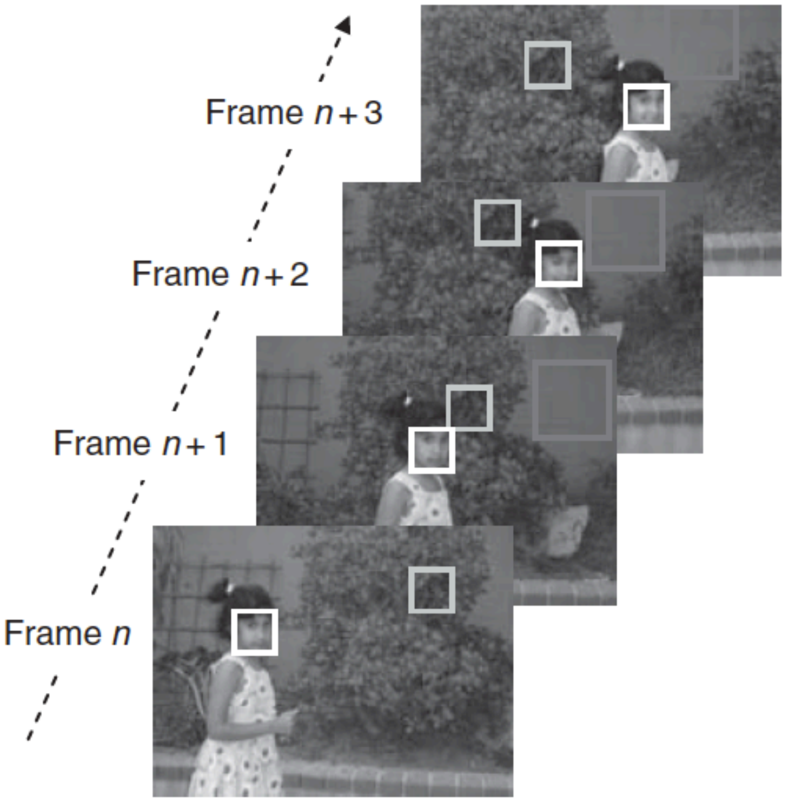
- a video consists of
  - a time-ordered sequence of frames
- obvious solution to video compression
  - predictive coding based on previous frames
- compression proceeds by subtracting images
  - subtract in time order and code residual error
- can be done even better by
  - searching for just the right parts of the image to subtract from the previous frame



Textbook: Fundamentals of Multimedia • Z.-N. Li et al.

Video Compression with Motion Compensation

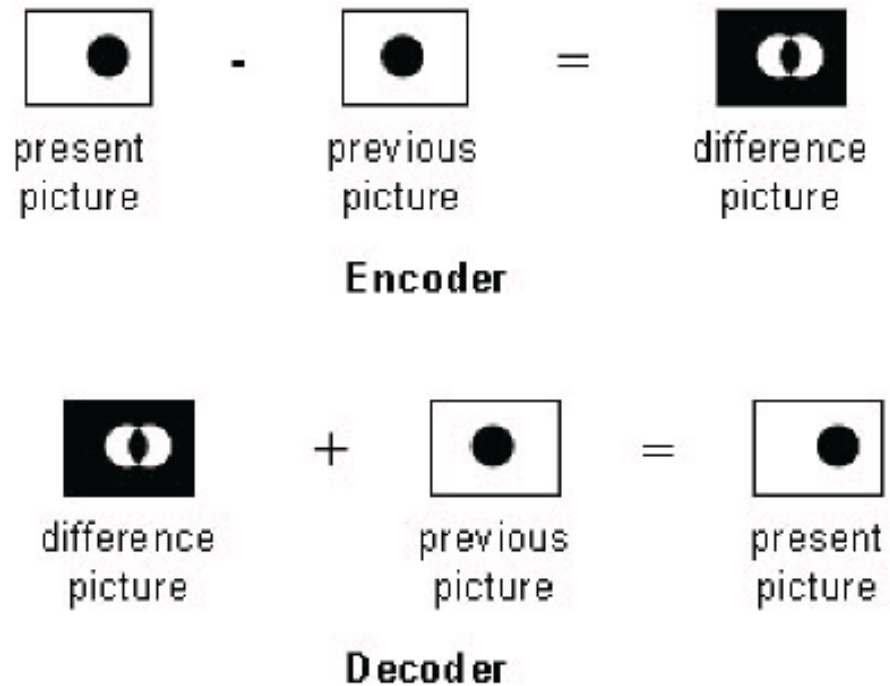
- consecutive frames in a video are similar
  - temporal redundancy
  - not every frame needs to be coded independently as a new image
- code the difference between current frame & other frame(s)
  - small values and low entropy
  - good for compression
- steps of Video compression based on Motion Compensation (MC):
  - 1. Motion Estimation ME
    - Motion Vector MV search
  - 2. MC-based Prediction
  - 3. derive prediction error
    - difference



Textbook: Fundamentals of Multimedia • Z.-N. Li et al.

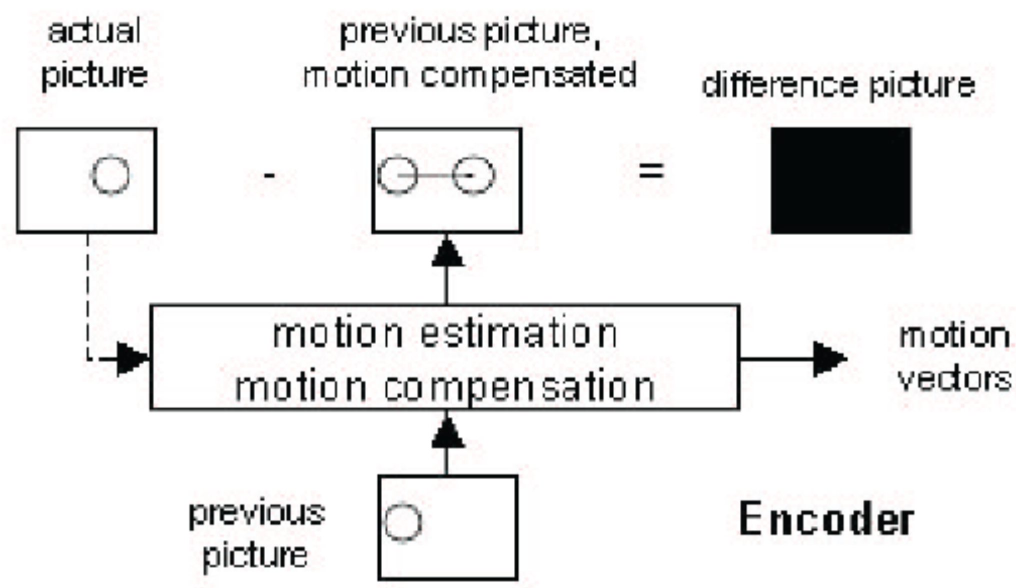
Simple Motion Example

- Consider a simple image (block) of a moving circle.
- Let's just consider the difference between 2 frames.
- It is simple to encode / decode:



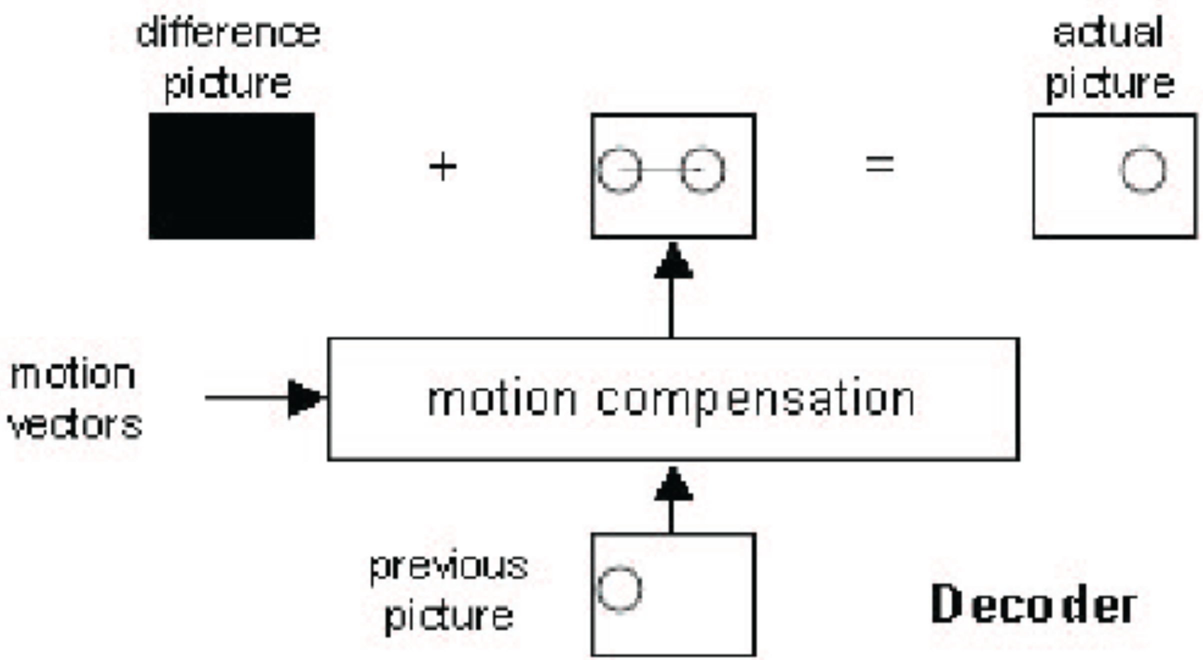
Simple Motion Example

- We will examine the methods of estimating motion vectors in due course.



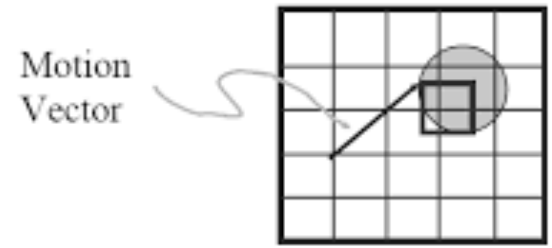
Simple Motion Example

- Decoding Motion of blocks



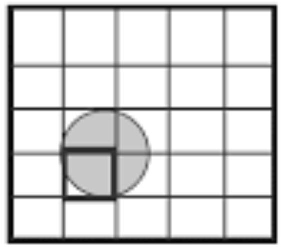
Motion Compensation

- image divided into macroblocks of size N x N
  - N = 16 for luminance by default
  - N = 8 for chrominance if chroma subsampling 4:2:0
- motion compensation MC → at macroblock level
  - Target Frame (TF) or Target Image (TI): current image
  - Reference frame(s)) (RF): previous and/or future frame(s)
  - MC → a match is sought between the macroblock in TF and the most similar macroblock in RF
  - motion vector MV → displacement of the reference macroblock to the target macroblock
- forward prediction: RF is a previous frame
- backward prediction: RF is a future frame



Reference Image

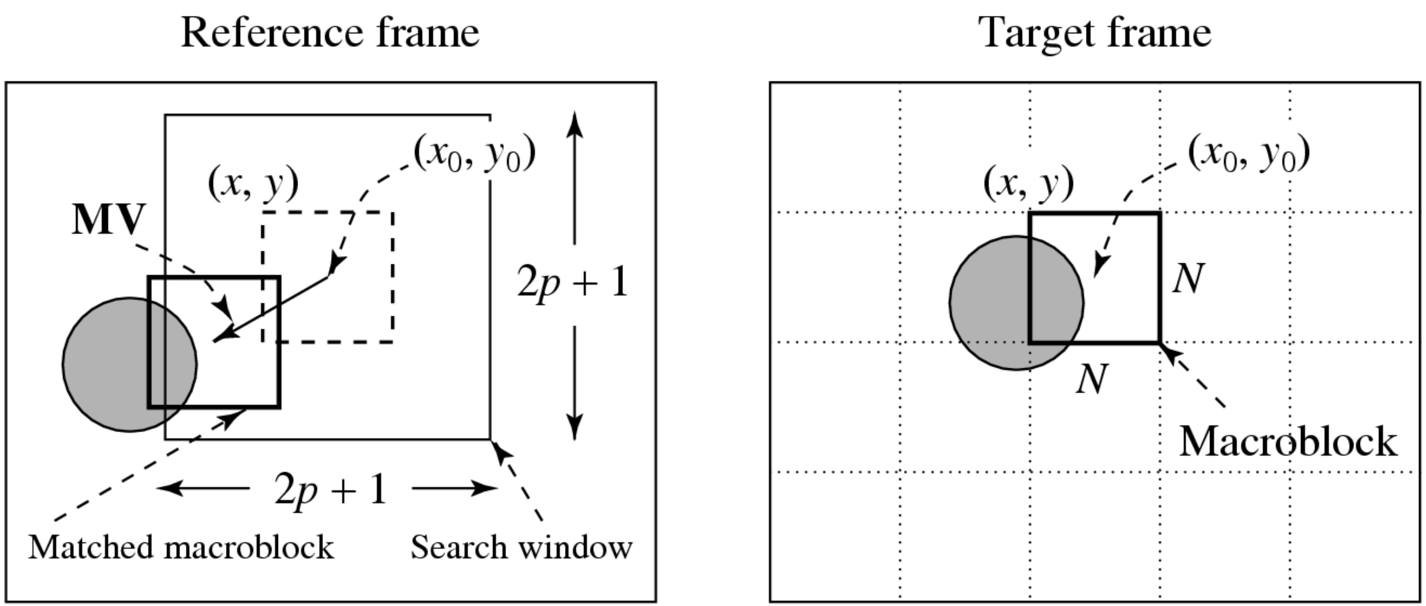
Target Image



Textbook: Fundamentals of Multimedia • Z.-N. Li et al.

Macroblocks and Motion Vector in Video Compression

- MV search is usually limited to a small immediate neighborhood
  - both horizontal and vertical displacements in the range [−p, p]
  - this makes a search window of size (2p + 1) x (2p + 1)



Textbook: Fundamentals of Multimedia • Z.-N. Li et al.

Background Photo:
The Milk Maid

Search for Motion Vectors

Search for Motion Vectors

- the difference between two macroblocks can then be measured by their Mean Absolute Difference (MAD):

$$MAD(i,j) = \frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\left|C(x+k,y+l) - R(x+i+k,y+j+l)\right|$$

- the goal of the search is to find
  - a vector (i, j) as the motion vector MV = (u, v) /
  - MAD(i, j) is minimum:

$$(u,v) = \left[\ (i,j)\ |\ MAD(i,j)\ is\ minimum,\ i \in [-p,p], j \in [-p,p]\right]$$

N	size of MB
k & l	indices for pixels in MB
i & j	hor. & vert. displacements
C(x+k,y+l)	pixels in MB in TF
R(x+i+k,y+j+l)	pixels in MB in RF

Three methods can be used

1. Sequential Search
2. 2D Logarithmic Search
3. Hierarchical Search

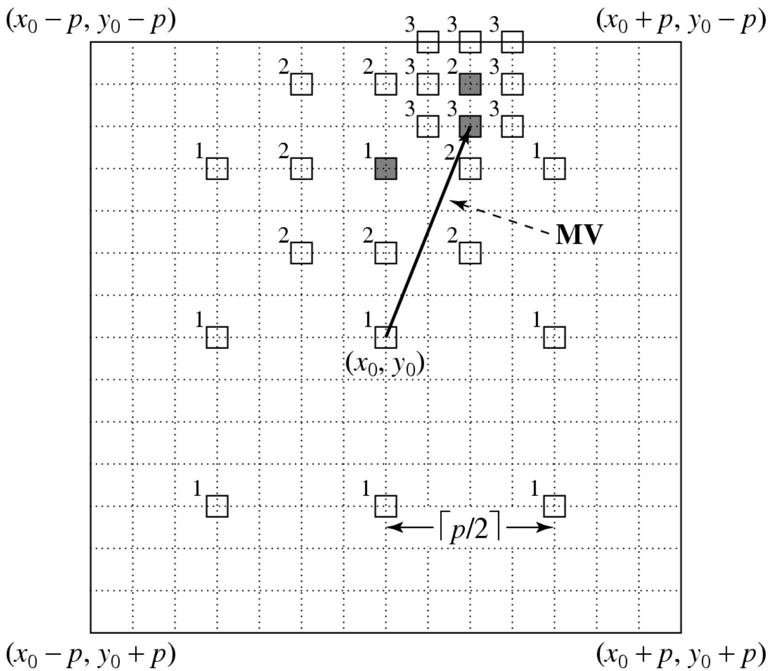
Sequential Search

- sequentially search whole window in RF
  - (2p + 1) x (2p + 1) MAD calculation
- → Full search
  - a macroblock centered at each of the positions within the window is compared to the macroblock in the TF using MAD
  - the vector (i, j) that offers the least MAD is designated as the MV (u, v) for the macroblock in the TF
- very costly

```
begin
    min_MAD = LARGE NUMBER;      /* Init */
    for i = −p to p
        for j = −p to p {
                cur_MAD = MAD(i, j);
                if cur_MAD  <  min_MAD {
                        min_MAD = cur_MAD;
                        /* Get coord. for MV. */
                        u = i;
                        v = j;
                }
        }
end
```

2D Logarithmic Search

- cheaper version, suboptimal yet effective
- several iterations akin to a binary search:
  - initially only 9 locations in the search window are used as seeds for a MAD-based search; marked as ‘1’.
  - the one that yields the min MAD is located
    - move center of new search region
    - reduce step-size (“offset”) to half
  - in the next iteration
    - the 9 new locations are marked as ‘2’
  - and so on…

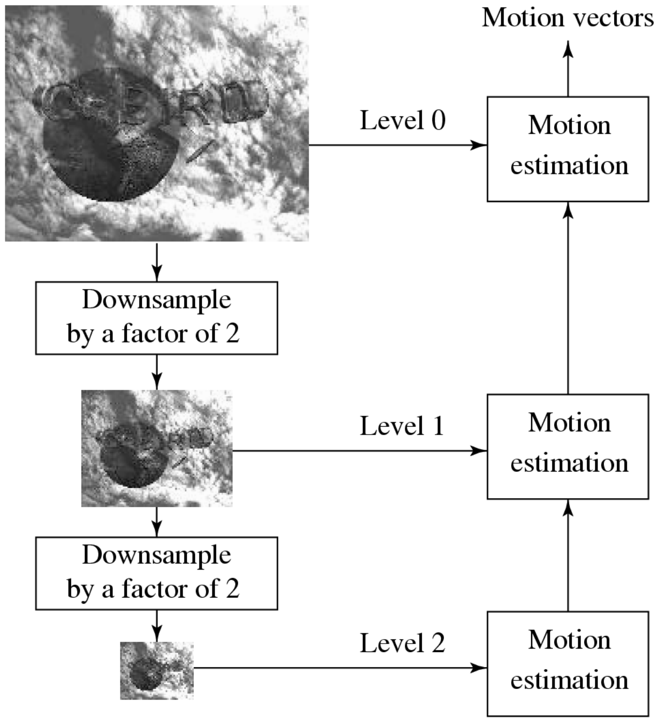


2D Logarithmic Search for Motion Vectors

Textbook: Fundamentals of Multimedia • Z.-N. Li et al.

Hierarchical Search

- benefit from a hierarchical (multiresolution) approach
  - initial estimation of the MV obtained from images with reduced resolution
- example:
  - a three-level hierarchical search
  - original image is at Level 0
  - Levels 1 and 2 down-sampled by a factor of 2
  - initial search → conducted at Level 2
- size of macroblock is smaller
  - p proportionally reduced
  - number of ops greatly reduced



A Three-level Hierarchical Search for Motion Vectors

Textbook: Fundamentals of Multimedia • Z.-N. Li et al.

Comparison of Computational Cost of Motion Vector Search based on examples

Search Method	OPS_per_second for 720×480 at 30 fps	
	$p = 15$	$p = 7$
Sequential search	29.89×10^9	7.00×10^9
2D Logarithmic search	1.25×10^9	0.78×10^9
3-level Hierarchical search	0.51×10^9	0.40×10^9

Background Photo:
[The Milk Maid]

Compression Standard Committees & Compression Standards

Compression Standard Committees & Compression Standards

- image, video & audio compression standards have been specified and released by two main groups since 1985:
  - ISO: International Standards Organisation: JPEG, MPEG
  - ITU: International Telecommunications Union: H.261 -- 264
    - CCITT: Comité Consultatif International Téléphonique et Télégraphique whose parent company is ITU
- whilst in many cases one of the groups have specified separate standards, there is some crossover between the groups
- JCT-VC: Joint Collaborative Team on Video Coding from the groups of ITU-T VCEG (Video Coding Experts Group) and ISO/IEC MPEG

ISO	ITU
JPEG → 1989	*adopted as T.81*
MPEG → 1991	
	1993 ← H.261 *based on CCITT 1990 draft*
	1994 ← H.262 *known as MPEG-2*
	1996 ← H.263 *extended as H.263+, H.263++*
MPEG 4 → 1998	
	2002 ← H.264 *for DVD quality now part of MPEG 4 (Part 10)*
HEVC *High Efficiency Video Coding* MPEG-H → 2013 ← H.265	

Background Photo:
The Milk Maid

H.261 & H.263

H.261

- earlier digital video compression standard
  - its principle of MC-based compression is retained in all later video compression standards
- designed for videophone, video conferencing and other audiovisual services over ISDN
- video codec supports bit-rates of p x 64 kbps
  - where p ranges from 1 to 30
  - require delay < 150 msec so that the video can be used for real-time bidirectional video conferencing

- H.261 belongs to the following set of ITU recommendations for visual telephony systems:

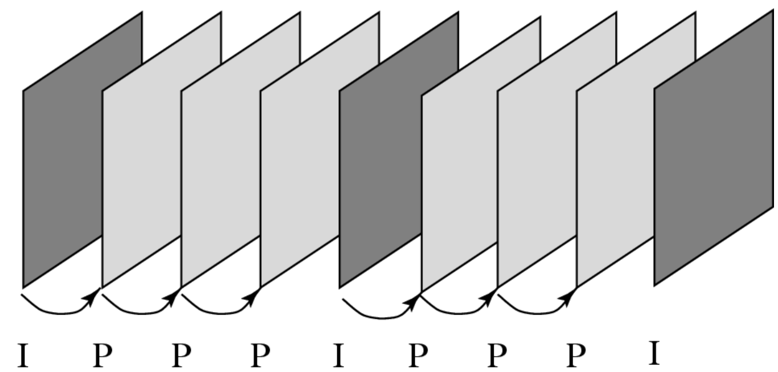
H.221	Frame structure for an audiovisual channel supporting 64 to 1,920 kbps.
H.230	Frame control signals for audiovisual systems.
H.242	Audiovisual communication protocols.
H.261	Video encoder/decoder for audiovisual services at p x 64 kbps.
H.320	Narrow-band audiovisual terminal equipment for p x 64 kbps transmission.

Video Formats Supported by H.261

Video format	Luminance image resolution	Chrominance image resolution	Bit-rate (Mbps) (if 30 fps and uncompressed)	H.261 support
QCIF	176×144	88×72	9.1	required
CIF	352×288	176×144	36.5	optional

H.261 Frame Sequence

- two types of image frames are defined:
  - Intra-frames (I-frames)
    - treated as independent images
    - transform coding method similar to JPEG
    - spatial redundancy removal
  - Inter-frames (P-frames)
    - P-frames are not independent:
    - forward predictive coding method (from a previous P or I-frame)
    - temporal redundancy removal

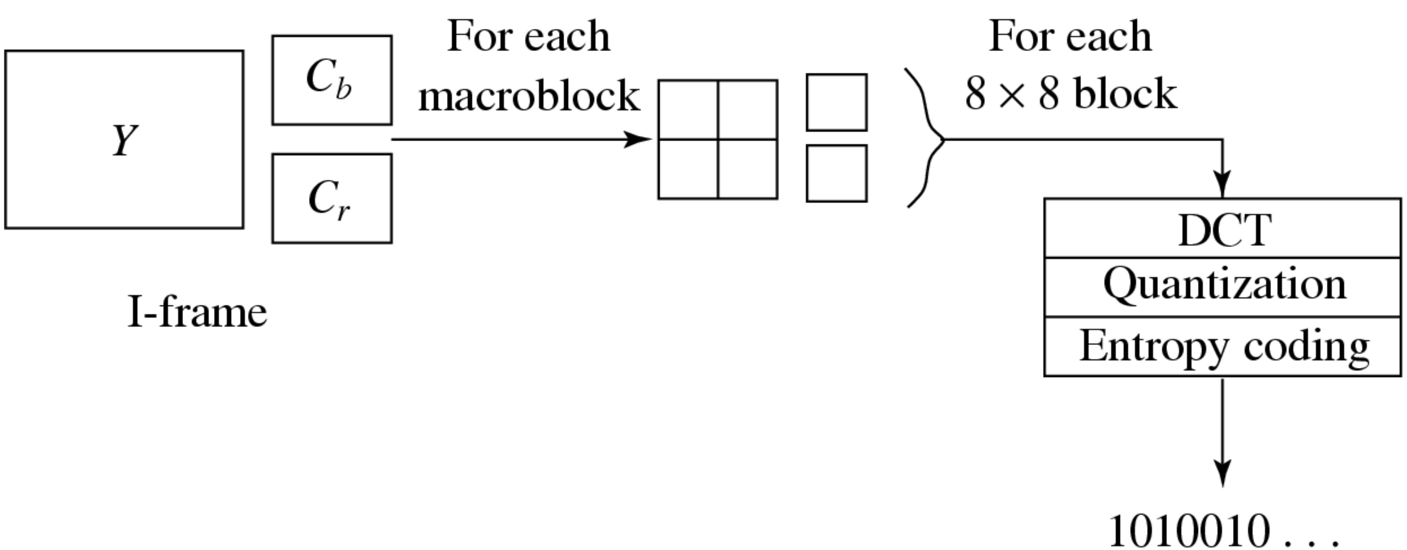


- to avoid propagation of coding errors
  - an I-frame is usually sent a couple of times in each second of the video

- MV in H.261 measured in units of full pixel with limit ± 15 pixels

Intra-frame (I-frame) Coding

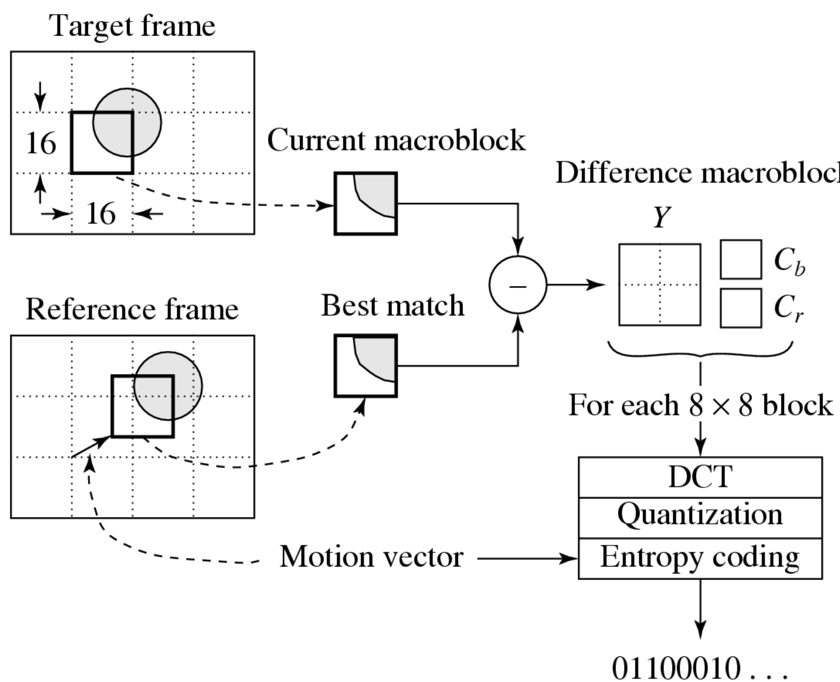
- macroblocks 16 x 16 for Y, 8 x 8 for Cb & Cr – since 4:2:0
- a macroblock consists of : (4 Y + 1 Cb + 1 Cr) 8 x 8 blocks
- for each 8 x 8 block: DCT → quantization → zigzag scan → entropy coding



Textbook: Fundamentals of Multimedia • Z.-N. Li et al.

Inter-frame (P-frame) Predictive Coding

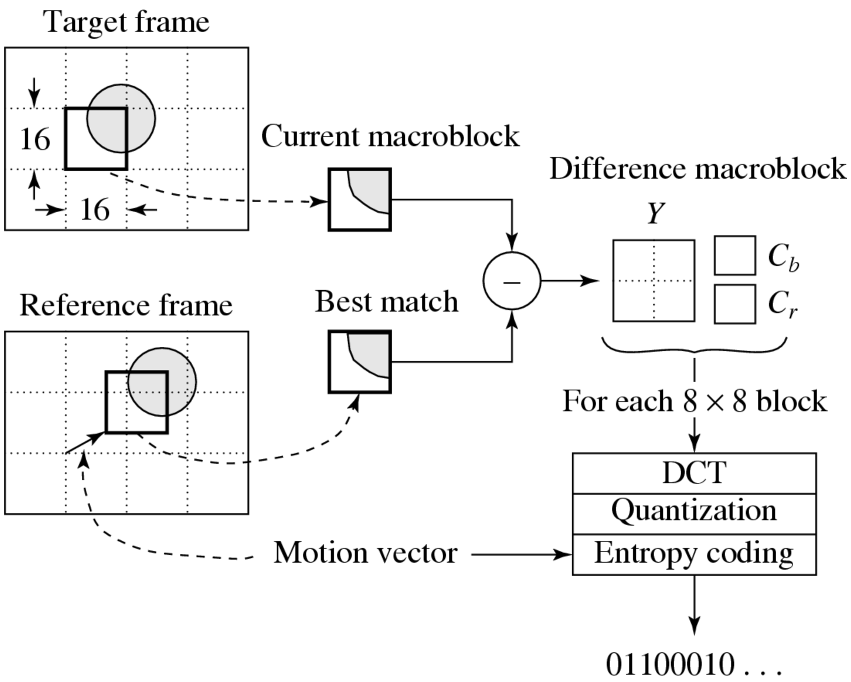
- for each macroblock in TF, a MV is allocated by one of the search methods discussed earlier
- then a difference macroblock is derived to measure prediction error
- each of these 8 x 8 blocks go through DCT → quantization → zigzag scan → entropy coding procedures



H.261 P-frame Coding Based on MC

Inter-frame (P-frame) Predictive Coding

- P-frame coding encodes ***difference macroblock*** MVD (not the Target macroblock itself)
- sometimes, a good match cannot be found, i.e., the prediction error exceeds a certain acceptable level
  - MB itself is then encoded (treated as an Intra MB)
  - termed a non-MC MB
- for a MV, the difference MVD is sent for entropy coding:
  - MVD = MVPreceding − MVCurrent



H.261 P-frame Coding Based on MC

H.263

- H.263 is an improved video coding standard for video conferencing and other audiovisual services transmitted on Public Switched Telephone Networks (PSTN)
  - aims at low bit-rate communications at bit-rates of less than 64 kbps
  - uses
    - predictive coding for inter-frames to reduce temporal redundancy and
    - transform coding for the remaining signal to reduce spatial redundancy (for both Intra-frames and inter-frame prediction)

Video Formats Supported by H.263

Video format	Luminance image resolution	Chrominance image resolution	Bit-rate (Mbps) (if 30 fps and uncompressed)	Bit-rate (kbps) BPPmaxKb (compressed)
sub-QCIF	128×96	64×48	4.4	64
QCIF	176×144	88×72	9.1	64
CIF	352×288	176×144	36.5	256
4CIF	704×576	352×288	146.0	512
16CIF	$1,408 \times 1,152$	704×576	583.9	1024

Background Photo: The Milk Maid

MPEG Video Coding — MPEG-1, 2, 4 & 7

Overview

- MPEG:
  - Moving Pictures Experts Group
  - established in 1988 for the development of digital video
- it is appropriately recognized that proprietary interests need to be maintained within the family of MPEG standards:
  - accomplished by defining only a compressed bitstream that implicitly defines the decoder
  - compression algorithms, and thus encoders, completely up to the manufacturers

MPEG-1

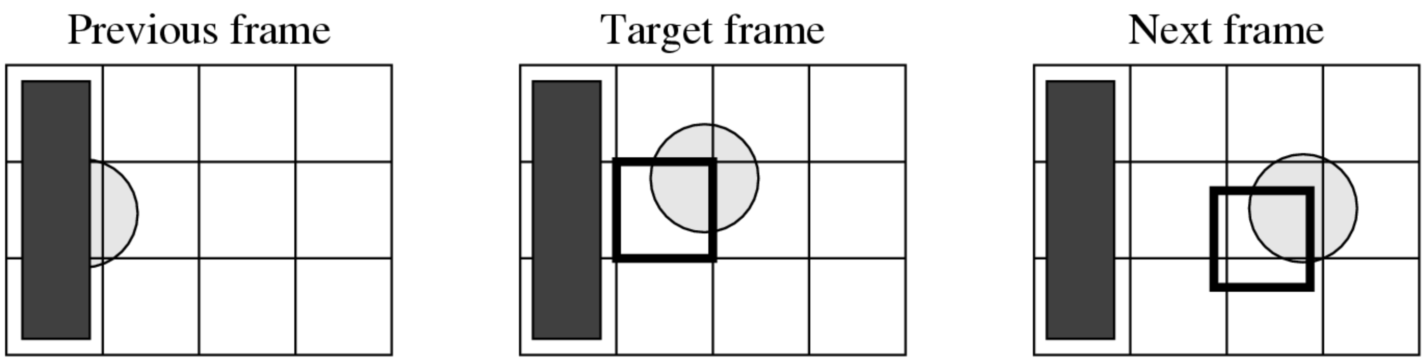
- adopts the CCIR601 digital TV format known as SIF
  - SIF: Source Input Format
- supports only non-interlaced video
- picture resolution
  - 352 × 240 for NTSC video at 30 fps
  - 352 × 288 for PAL video at 25 fps
- uses 4:2:0 chroma subsampling
- also referred to as ISO/IEC 11172
- it has five parts:

11172-1	Systems
11172-2	Video
11172-3	Audio
11172-4	Conformance
11172-5	Software

Motion Compensation in MPEG-1

- Motion Compensation (MC) based video encoding in H.261 works as follows:
- in Motion Estimation (ME),
  - each macroblock (MB) of the Target P-frame is assigned a best matching MB from the previously coded I or P frame - **prediction**
- **prediction error**:
  - difference between the MB and its matching MB
  - sent to DCT and its subsequent encoding steps
- **forward prediction** → prediction is from a previous frame

The Need for Bidirectional Search

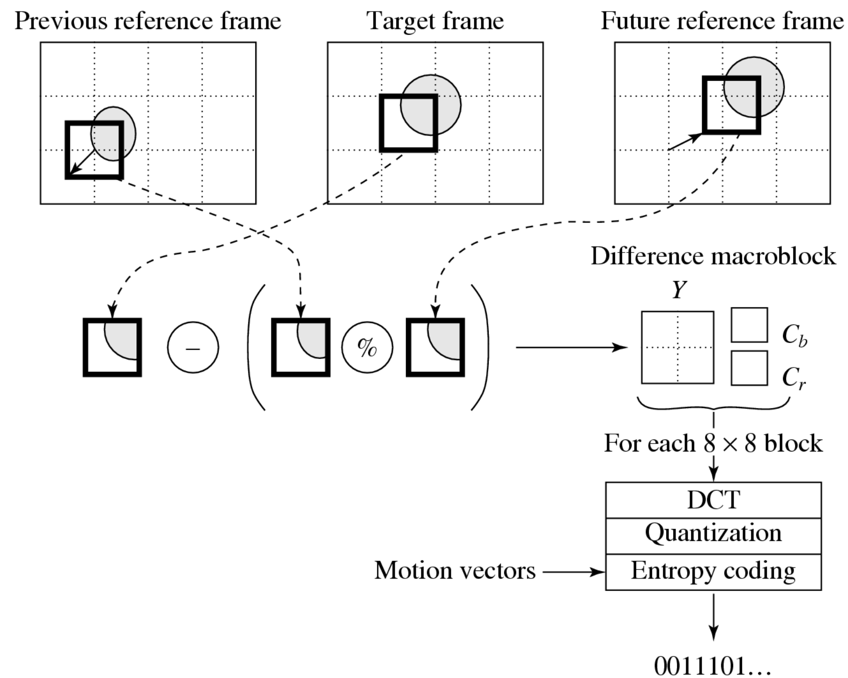


- The MB containing part of a ball in the Target frame cannot find a good matching MB in the previous frame because half of the ball was occluded by another object.
- A match however can readily be obtained from the next frame.

Motion Compensation in MPEG-1

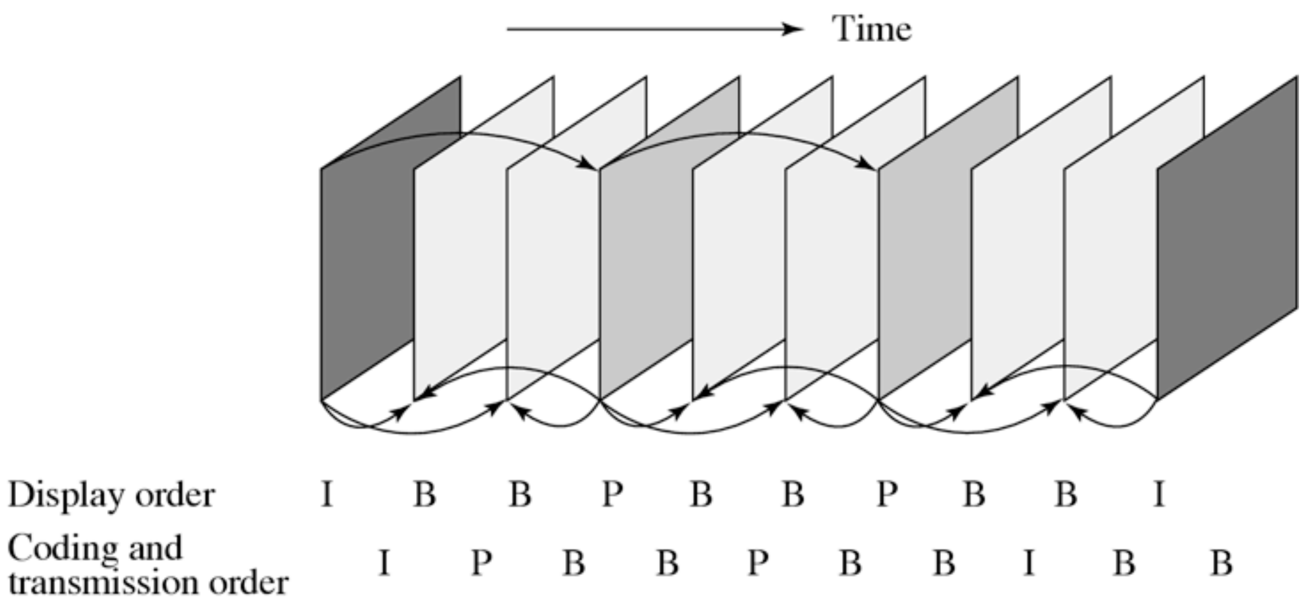
- MPEG introduces a third frame type: B-frames, and its accompanying bi-directional motion compensation

- each MB from a B-frame will have up to 2 MVs (1 from forward + 1 from backward prediction)
  - if matching in both directions is successful
    - 2 MVs sent → 2 averaged MBs → generate prediction error
  - if an acceptable match can be found in only 1 of the reference frames
    - only 1 MV and its corresponding MB used



B-frame Coding Based on Bidirectional Motion Compensation

MPEG Frame Sequence

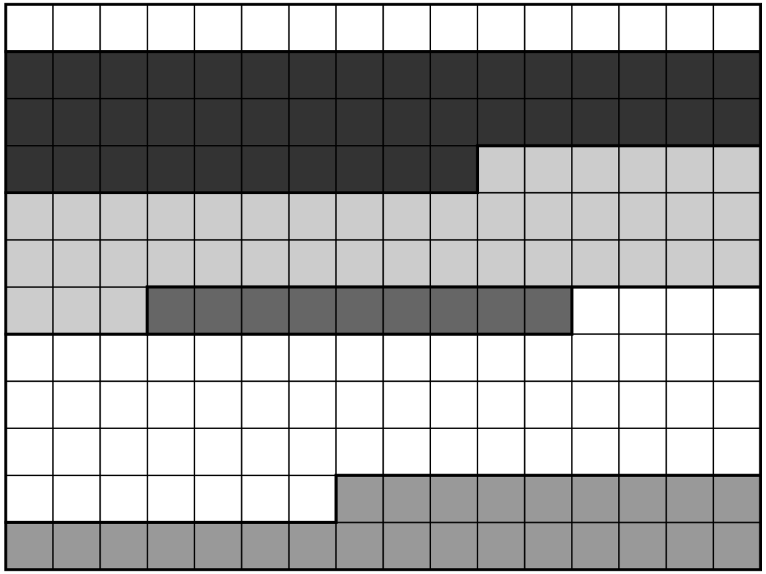


The MPEG-1 Constrained Parameter Set

Parameter	Value
Horizontal size of picture	≤ 768
Vertical size of picture	≤ 576
No. of MBs / picture	≤ 396
No. of MBs / second	≤ 9,900
Frame rate	≤ 30 fps
Bit-rate	≤ 1,856 kbps

Major Differences from H.261

- instead of GOBs as in H.261, an MPEG-1 picture can be divided into one or more slices:
  - may contain variable numbers of macroblocks in a single picture
  - may also start and end anywhere as long as they fill the whole picture
  - each slice is coded independently — additional flexibility in bit-rate control
  - slice concept is important for error recovery



Slices in an MPEG-1 Picture

Major Differences from H.261

- quantization: MPEG-1 quantization uses different quantization tables for its Intra and Inter coding
- for DCT coefficients in Intra mode:

$$QDCT[i,j] = round\left(\frac{8 \times DCT[i,j]}{step\_size[i,j]}\right) = round\left(\frac{8 \times DCT[i,j]}{Q_1[i,j] * scale}\right)$$

- for DCT coefficients in Inter mode:

$$QDCT[i,j] = \left\lfloor \frac{8 \times DCT[i,j]}{step\_size[i,j]} \right\rfloor = \left\lfloor \frac{8 \times DCT[i,j]}{Q_2[i,j] * scale} \right\rfloor$$

Default Quantization Table (Q1) for Intra-Coding

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

Default Quantization Table (Q2) for Inter-Coding

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

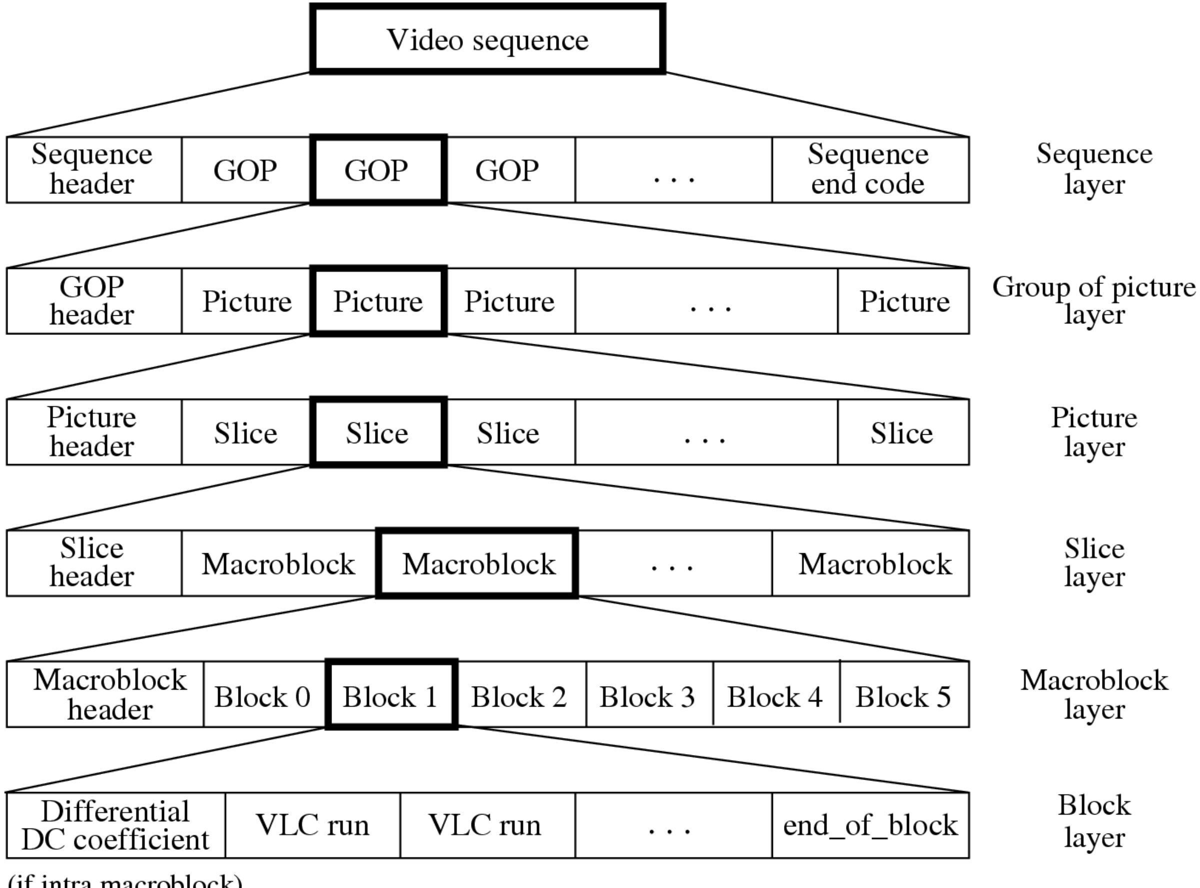
Typical Sizes of MPEG-1 Frames

- typical size of compressed P-frames is significantly smaller than that of I-frames
  - because temporal redundancy is exploited in inter-frame compression
- B-frames are even smaller than P-frames because of
  - (a) advantage of bi-directional prediction
  - (b) lowest priority given to B-frames

- Typical Compression Performance of MPEG-1 Frames

Type	Size	Compression
I	18kB	7:1
P	6kB	20:1
B	2.5kB	50:1
Avg	4.8kB	27:1

Layers of MPEG-1 Video Bitstream



MPEG-2

- for higher quality video at a bit-rate of more than 4 Mbps
- 7 profiles aimed at different applications:
  1. Simple
  2. Main
  3. SNR scalable
  4. Spatially scalable
  5. High
  6. 4:2:2
  7. Multiview
- within each profile
  - up to 4 levels are defined
- DVD video specification allows only four display resolutions:
  - 720×480
  - 704×480
  - 352×480
  - 352×240 → a restricted form of the MPEG-2 Main profile at the Main and Low levels

Profiles and Levels in MPEG-2

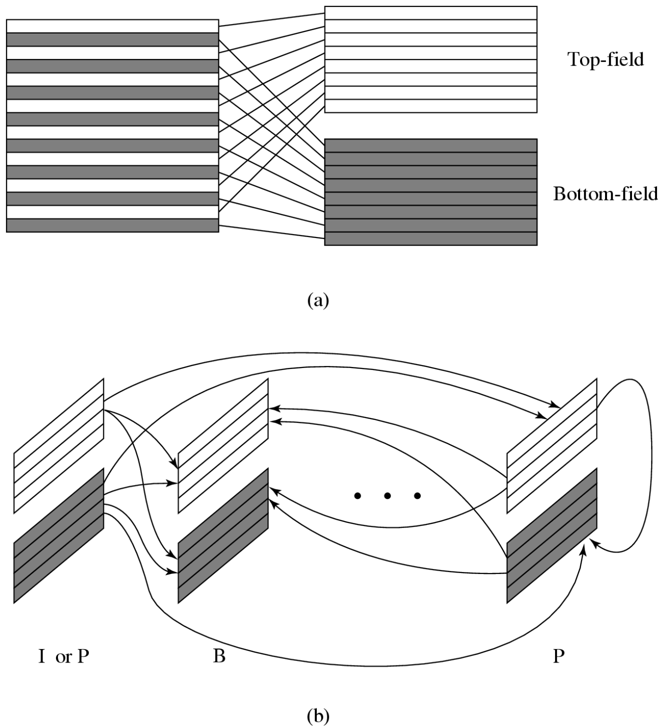
Level	Simple profile	Main profile	SNR Scalable profile	Spatially Scalable profile	High Profile	4:2:2 Profile	Multiview Profile
High		*			*		
High 1440		*		*	*		
Main	*	*	*		*	*	*
Low		*	*				

Four Levels in the Main Profile of MPEG-2

Level	Max. Resolution	Max fps	Max pixels /sec	Max coded Data Rate (Mbps)	Application
High	1,920 × 1,152	60	62.7×10^6	80	film production
High 1440	1,440 × 1,152	60	47.0×10^6	60	consumer HDTV
Main	720 × 576	30	10.4×10^6	15	studio TV
Low	352 × 288	30	3.0×10^6	4	consumer tape equiv.

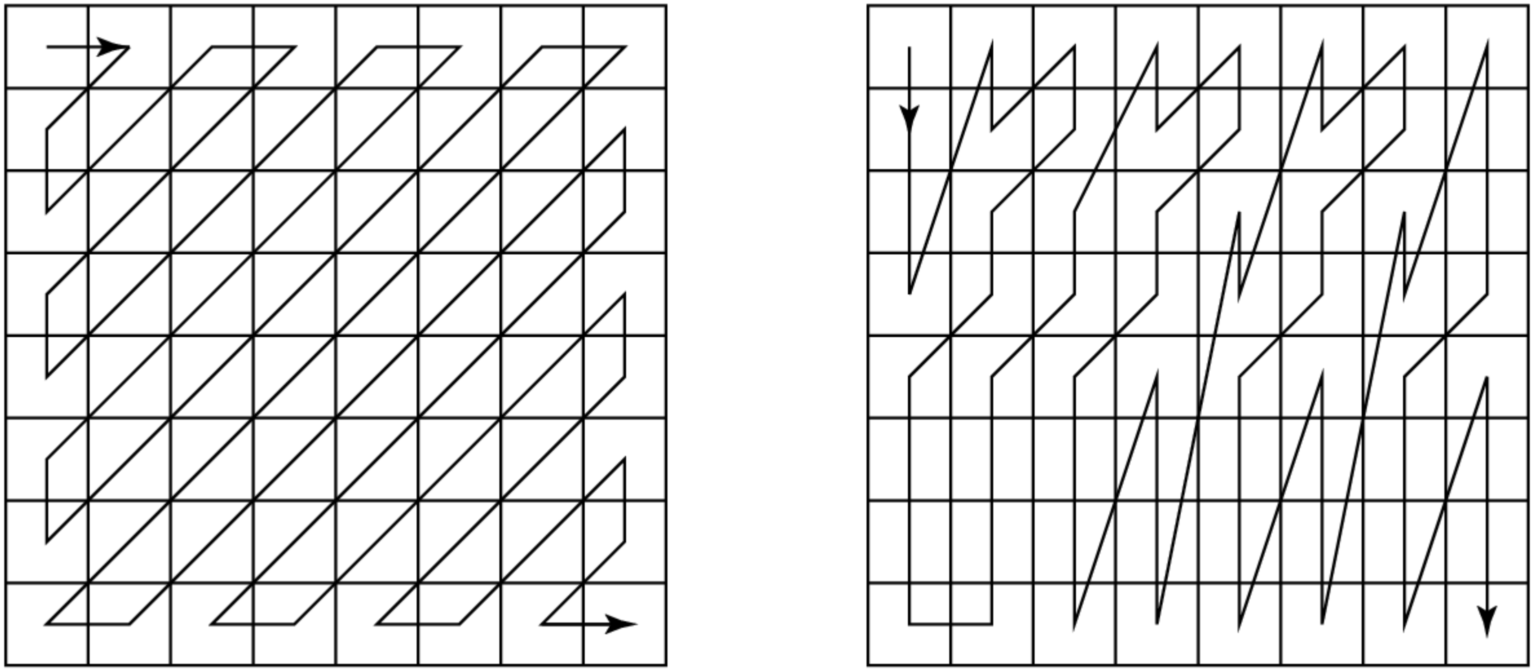
Supporting Interlaced Video

- MPEG-2 must support interlaced video as well since this is one of the options for digital broadcast TV and HDTV
- in interlaced video each frame consists of two fields, referred to as the *top-field* and the *bottom-field*
- *Frame-picture:* all scanlines from both fields are interleaved to form a single frame, then divided into 16×16 macroblocks and coded using MC
- *Field-picture:* each field is treated as a separate picture



Field pictures and Field-prediction for Field-pictures in MPEG-2
(a) Frame−picture vs. Field−pictures
(b) Field Prediction for Field−pictures

Zigzag and Alternate Scans of DCT Coefficients for Progressive and Interlaced Videos in MPEG-2



(a) (b)

Some Major Differences from MPEG-1

- better resilience to bit-errors:
  - in addition to Program Stream, a Transport Stream is added to MPEG-2 bit streams
- support of 4:2:2 and 4:4:4 chroma subsampling
- more restricted slice structure:
  - MPEG-2 slices must start and end in the same macroblock row
  - → left edge of a picture always starts a new slice & the longest slice in MPEG-2 can have only one row of macroblocks
- more flexible video formats:
  - it supports various picture resolutions as defined by DVD, ATV and HDTV

Other Major Differences from MPEG-1

- nonlinear quantization — two types of scales are allowed:
  1. scale same as in MPEG-1 → integer in the range of [1, 31] and $scale_i = i$
  2. a nonlinear relationship exists, $scale_i \neq I$
     1. ith scale value can be looked up from Table 11.7

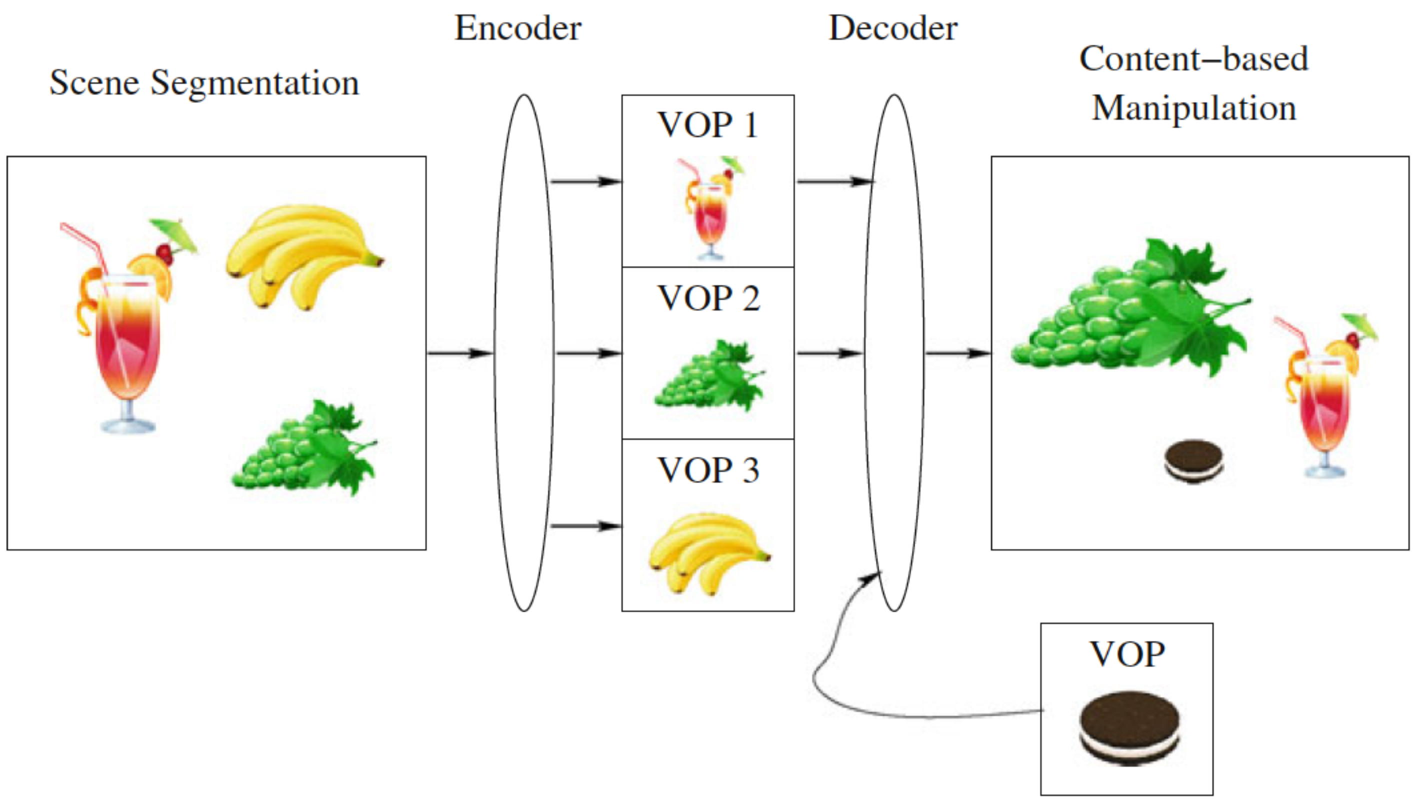
Table 11.7: Possible Nonlinear Scale in MPEG-2

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$scale_i$	1	2	3	4	5	6	7	8	10	12	14	16	18	20	22	24
i	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
$scale_i$	28	32	36	40	44	48	52	56	64	72	80	88	96	104	112	

MPEG-4

- MPEG-1 and -2 → frame-based coding techniques
  - macroblocks→ block-based coding
  - main concern → high compression ratio and satisfactory quality of video
- MPEG-4 had a very different emphasis
  - besides compression
  - great attention to user interactivity
- users++ create & communicate their multimedia presentations and applications on new infrastructures
- adopts new object-based coding approach
  - media objects either natural or synthetic
- object-based coding
  - higher compression ratio
  - beneficial for digital video composition, manipulation, indexing, and retrieval
- version 1 → standard in 1999
- MPEG-4 Part 2
  - targeted at low-bitrate communication
  - 4.8–64 kbps for mobile apps
  - up to 2 Mbps for other apps
- bitrate now
  - between 5 kbps and 10 Mbps

Composition and manipulation of MPEG-4 videos (VOP = Video object plane)



Textbook: Fundamentals of Multimedia • Z.-N. Li et al.

MPEG-7

- effective and efficient audiovisual content-based retrieval
- in 1996 → became an international standard in 2001
- common ground between MPEG-4 and MPEG-7 is the focus on audiovisual objects
- applicable to any multimedia applications
  - generation (content creation)
  - usage (content consumption)
- → Multimedia Content Description Interface
- same first seven parts
- parts 8 to 12: focus on various profiles and query format

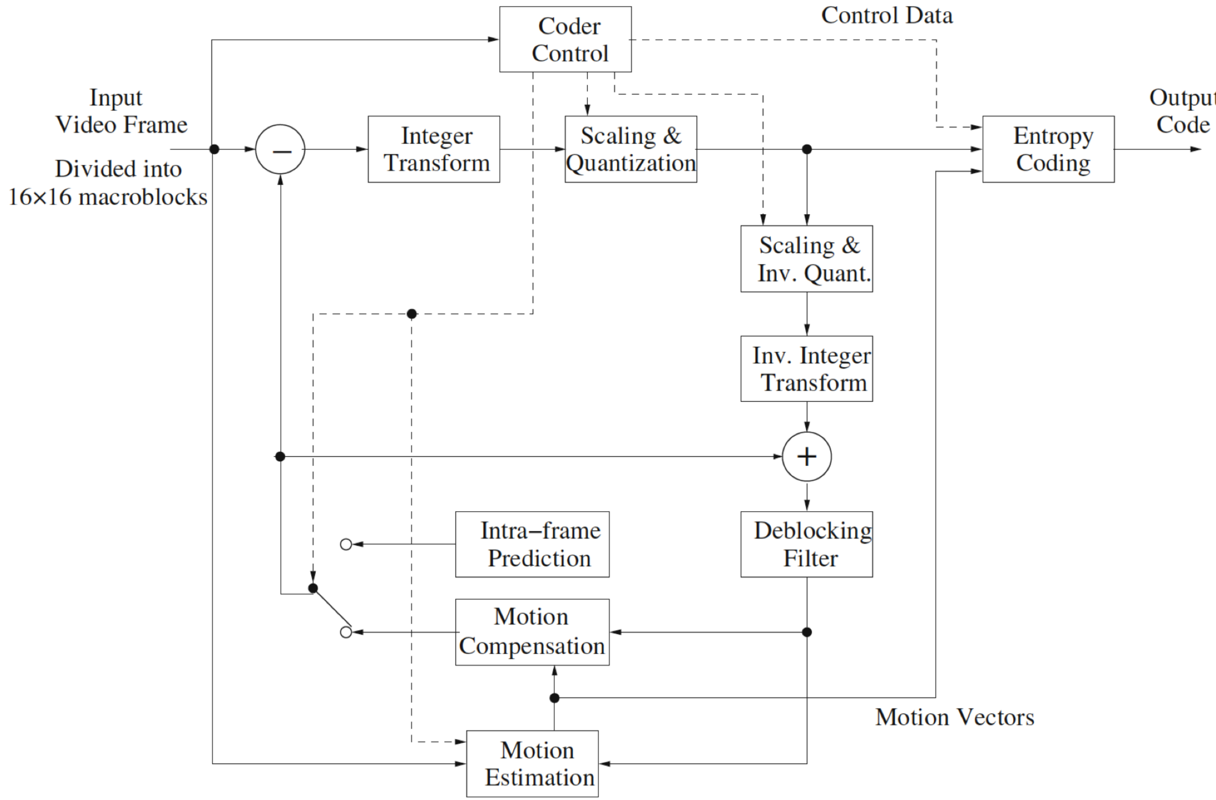
Background Photo:
[The Milk Maid]

New Video Coding Standards: H.264 & H.265

H.264

- By the Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (Video Coding Experts Group) in May 2003
- known as MPEG-4 Part 10, AVC (Advanced Video Coding)
- provides a higher video coding efficiency
  - up to 50% better compression than MPEG-2
  - up to 30% better than H.263+ and MPEG-4 Advanced Simple Profile
  - same quality of the compressed video
- covers a broad range of applications, from high bitrate to very low bitrate
- improved core features + new coding tools → significant improvement in compression ratio, error resiliency, and subjective quality over existing ITU-T & MPEG standards
- default standard for various applications
  - Blu-ray discs - HDTV broadcasts -streaming video on the Internet - web software such as Flash and Silverlight - apps on mobile and portable devices …
- similar to previous, block-based hybrid coding scheme + MC + transform coding
- each picture separated into macroblocks (16x16 blocks)
  - arbitrary sized slices can group multiple macroblocks into self-contained units

Basic encoder for H.264/AVC



Main features of H.264/AVC

- integer transform in 4x4 blocks
  - low complexity, no drifting
- variable block-size motion compensation
  - from 16x16 to 4x4 in luma images
- quarter-pixel accuracy in MV
  - accomplished by interpolations
- multiple reference picture MC
  - more than just P or B frames for ME
- directional spatial prediction for intra frames
- in-loop deblocking filtering
- CAVLC & CABAC
  - Context-Adaptive Variable Length Coding(CAVLC) and Context-Adaptive Binary Arithmetic Coding (CABAC)
- more robust to data errors and data losses
  - more flexible in synchronization and switching of video streams produced by different decoders
- decoder has 5 major blocks:
  1. entropy decoding
  2. inverse quantization and transform of residual pixels
  3. MC or intra-prediction
  4. Reconstruction
  5. In-loop deblocking filter on reconstructed pixels

H.265

- HEVC (High Efficiency Video Coding) was the latest standard jointly developed by the Joint Collaborative Team on Video Coding (JCT-VC) from the groups of ITU-T VCEG (Video Coding Experts Group) and ISO/IEC MPEG
- final draft In January 2013
  - ISO/IEC → HEVC = MPEG-H Part 2 (ISO/IEC 23008-2)
  - ITU-T → HEVC = Recommendation H.265
- development motivated by two factors:
  - (a) increasing video resolution (e.g., up to 8k x 4k in UHDTV)
  - (b) speed up by exploiting parallel processing
- initial goal was a further 50% reduction of the size of the compressed video (with the same visual quality) from H.264 → exceeded
- H.264 and H.265 are currently the leading candidates to carry a whole range of video contents on many potential applications
- default format for color video in H.265 is YCbCr
- in main profiles, chroma subsampling is 4:2:0

Main features of H.265

- variable block-size MC
  - from 4x4 up to 64x64 in luma images
- MB structure replaced by a quadtree structure of coding blocks at various levels and sizes
- exploration of parallel processing
- integer transform in various sizes
  - from 4x4, 8x8, 16x16 to 32x32
- improved interpolation methods for the quarter-pixel accuracy in MV
- expanded directional spatial prediction for intra coding
  - 33 angular directions
- potential use of DST in luma intra coding
  - Discrete Sine Transform
- in-loop filters including deblocking-filtering and SAO
  - Sample Adaptive Offset
- only CABAC, no more ~~CAVLC~~