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

视频压缩技术 II
Basic Video Compression Techniques

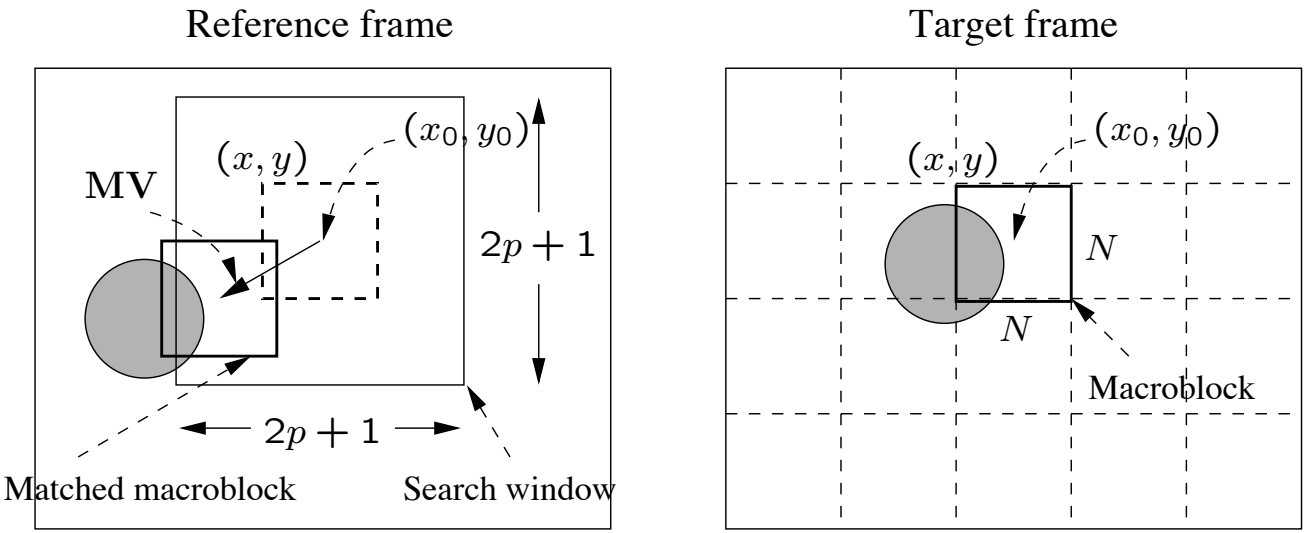
知识点回顾

- Introduction to Video Compression
- Video Compression with Motion Compensation
- Motion Compensation
- Sequential Search
- Logarithmic Search
- Hierarchical Search

Video Compression with Motion Compensation (运动补偿)

- Temporal redundancy (时间冗余) exists between consecutive (连续的) frames
  - Not every frame of the video needs to be coded independently (独立的) as a new image
  - The difference between the current frame and other frame(s) in the sequence will be coded.
- Steps of video compression based motion compensation (MC)
  - Motion estimation
  - MC-based prediction
  - Derivation of the prediction error

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 $(2p + 1) \times (2p + 1)$.

Search for Motion Vectors

- The difference between two macroblocks can be measured by Mean Absolute Difference (MAD)， 平均绝对误差

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

N – size of the macroblock,

k and l – indices for pixels in the macroblock,

i and j – horizontal and vertical displacements,

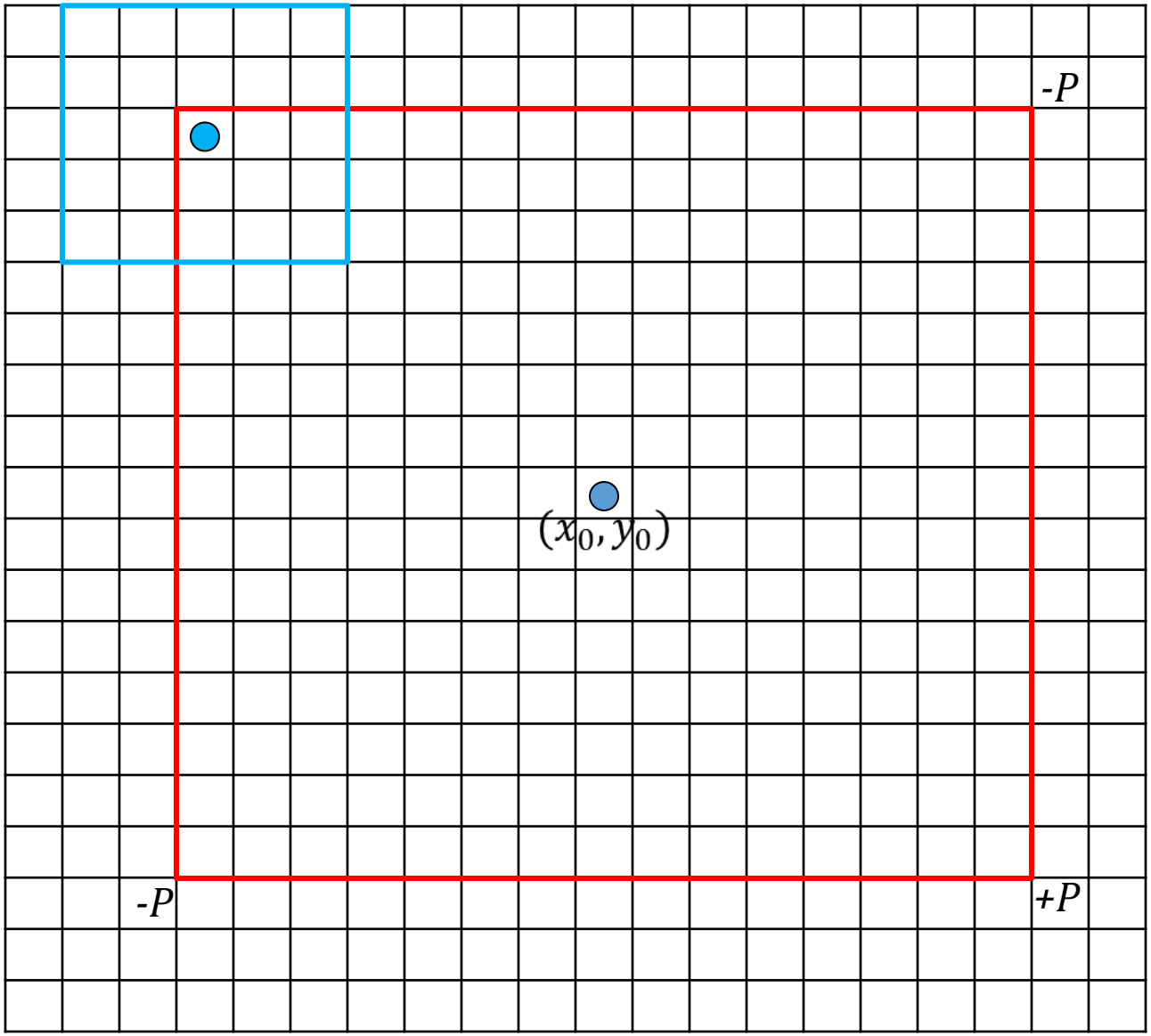
$C(x+k, y+l)$ – pixels in macroblock in Target frame,

$R(x+i+k, y+j+l)$ – pixels in macroblock in Reference frame.

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

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

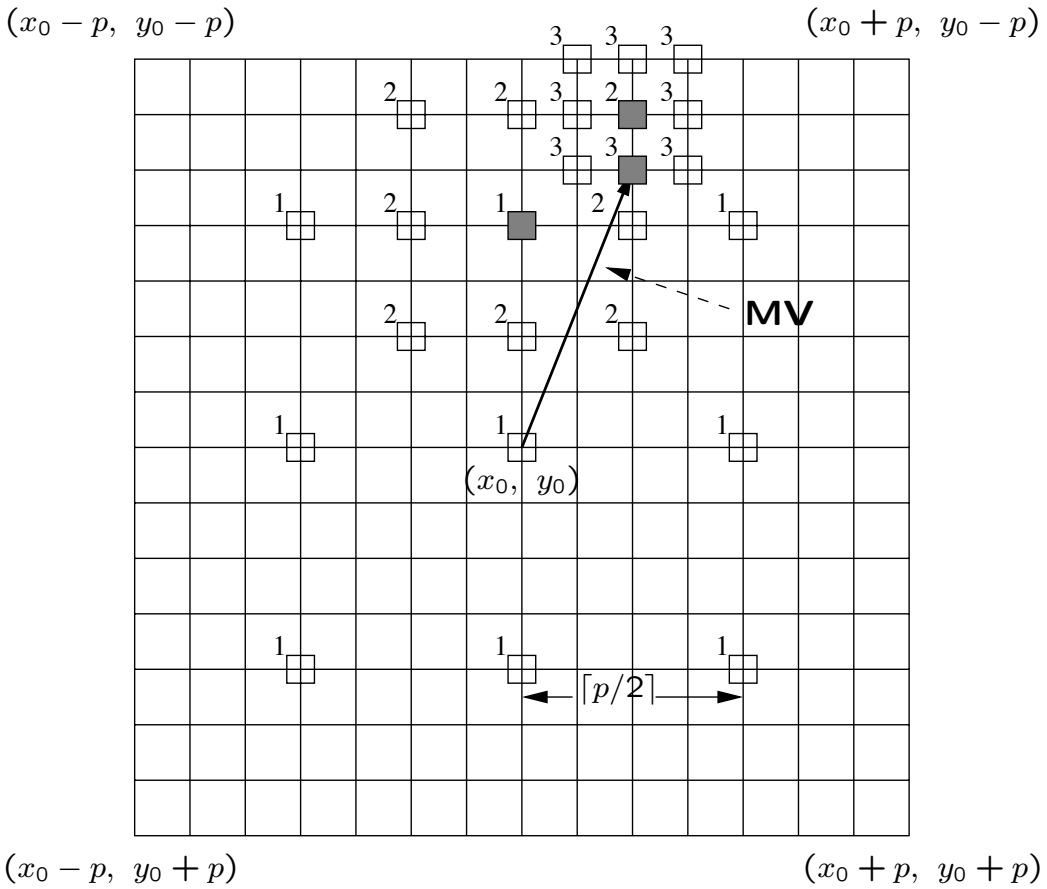
Sequential Search（顺序搜索）



$P=7$

$N=5$

2D Logarithmic Search for Motion Vectors



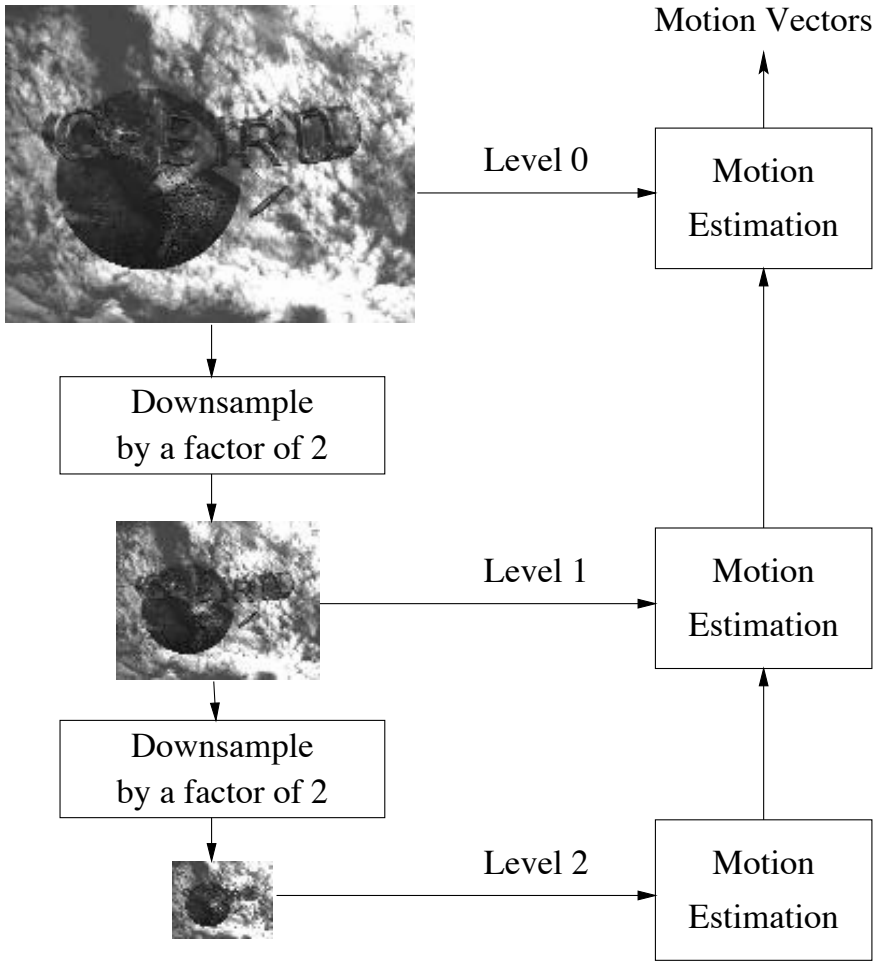
The total operations per second is dropped to: $O(\log p^* N^2)$

Save a lot cost than sequential search $O(p^2 N^2)$

A Three-level Hierarchical Search for Motion Vectors

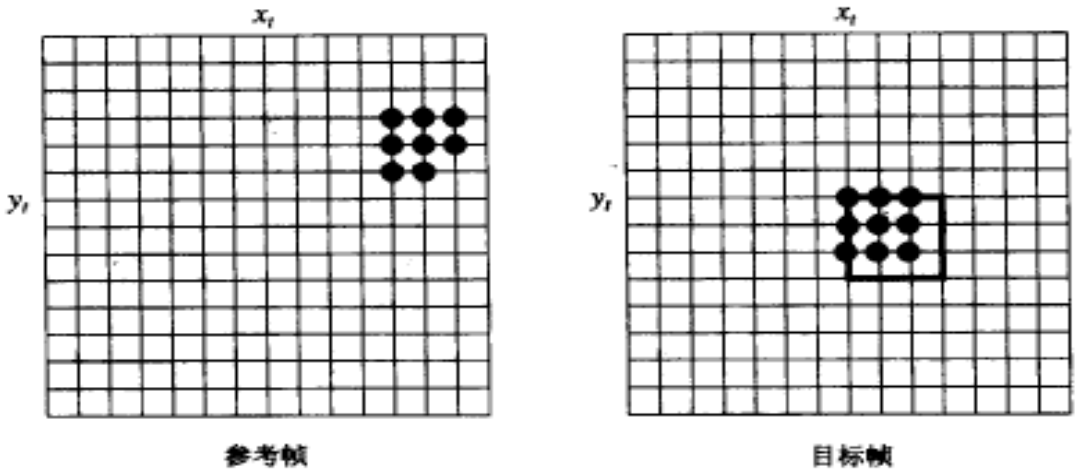
Original image is at Level 0, images at Levels 1 and 2 are obtained by down-sampling from the previous levels by a factor of 2, and the initial search is conducted at Level 2.

Since the size of the macroblock is smaller and p can also be proportionally reduced, the number of operations required is greatly reduced.



练习题

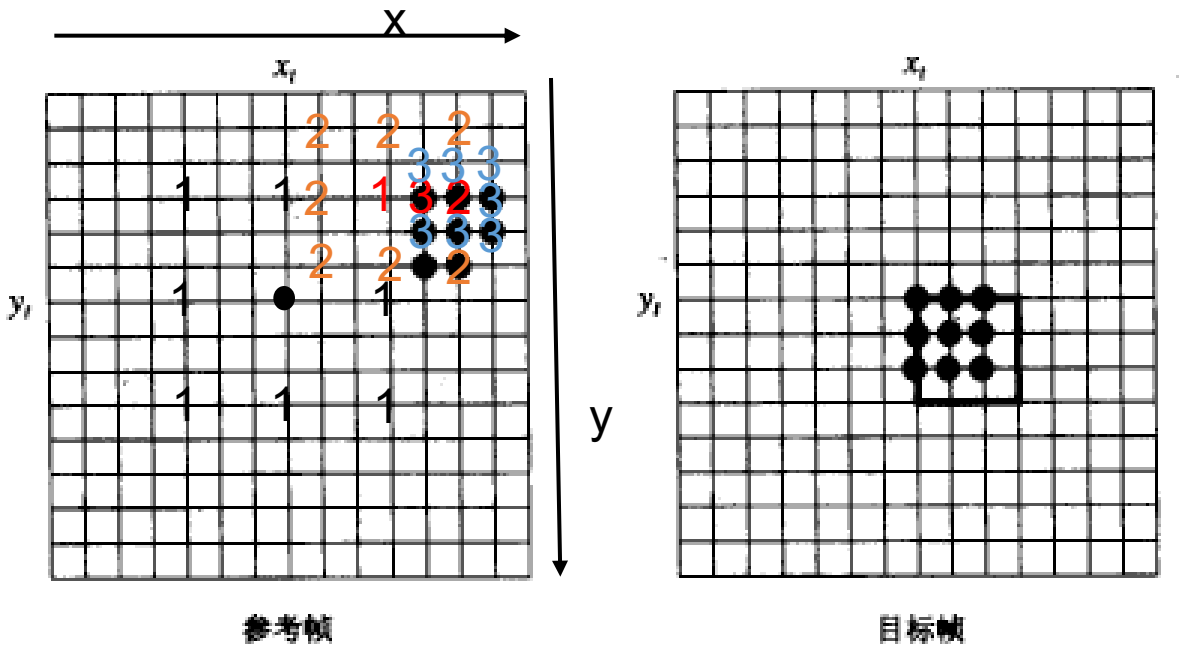
4. 回答下面用于运动向量的 2D 对数搜索的问题（见图 10-14）。
目标（当前）帧为 P 帧。宏块的大小为 4×4。运动向量是 $MV(\Delta x, \Delta y)$，其中 $\Delta x \in [-p, p]$，$\Delta y \in [-p, p]$。在这个问题中，假设 $p \equiv 5$。
帧中黑色的宏块左上角的坐标是 (x_t, y_t)。它包含 9 个黑色的像素，每个像素的亮度值为 10；其余 7 个像素点是背景的一部分，统一亮度值为 100。参考帧（前一帧）有 8 个黑色像素点。
（a）求 Δx、Δy 的最优值，宏块的平均绝对误差（MAE）是多少？
（b）一步步地说明如何进行 2D 对数搜索，包括搜索的位置和通道以及 Δx、Δy 和 MAE 的所有中间值。



● 亮度值为 10 的像素
其他是背景（未标识）像素，亮度为 100

图 10-14　运动向量的 2D 对数搜索

练习题



第一步搜索，步长为$\left\lceil \frac{5}{2} \right\rceil$=3，最优MV(3,-3)，MAD为5*|10-100|/16

第二步搜索，步长为$\left\lceil \frac{3}{2} \right\rceil$=2，最优MV(5,-3)，MAD为4*|10-100|/16

第三步搜索，步长为$\left\lceil \frac{2}{2} \right\rceil$=1，最优MV(4,-3)，MAD为|10-100|/16

编码标准

- H.261编码标准介绍; H.261里面的I FRAME，和P FRAME 概念和编码过程。

- 介绍MPEG编码，MPEG-1编码中的运动补偿，与H.261的主要区别

H.261 Frame Sequence

- **H.261**: An earlier digital video compression standard, its principle of MC-based compression is retained in all later video compression standards.

  –The standard was designed for videophone, video conferencing and other audiovisual services over ISDN(综合业务数字网).

  –The video codec supports bit-rates of $p \times 64$ kbps, where p ranges from 1 to 30 (Hence also known as $p * 64$).

  –Require that the delay of the video encoder be less than 150 msec so that the video can be used for real-time bi- directional video conferencing.

H.261 Frame Sequence

- Two types of image frames: Intra-frames (I-frames，帧内编码) and Inter-frames (P-frames，帧间编码)
  - I-frames are treated as independent (spatial redundancy removal)
  - P-frames are not independent: coded by a forward predictive coding method (prediction from a previous I-frame or P-frame)
  - P-frame coding includes spatial redundancy removal and temporal redundancy removal, whereas I-frame coding includes spatial redundancy removal
  - p = 15 in motion vectors of H.261

H.261 Frame Sequence

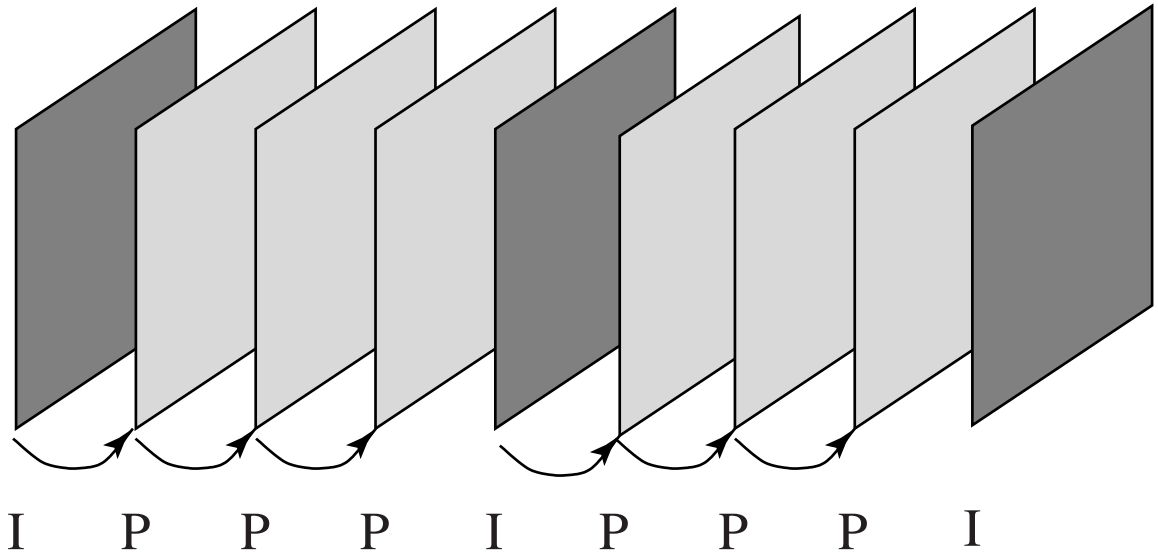


Fig. 10.4: H.261 Frame Sequence.

Intra-frame (I-frame) Coding

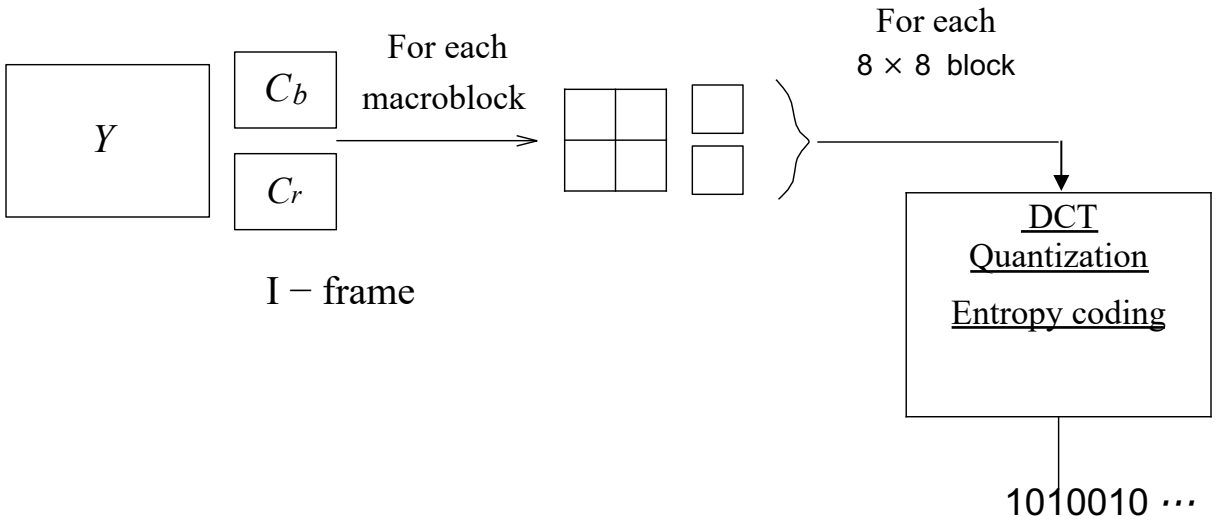
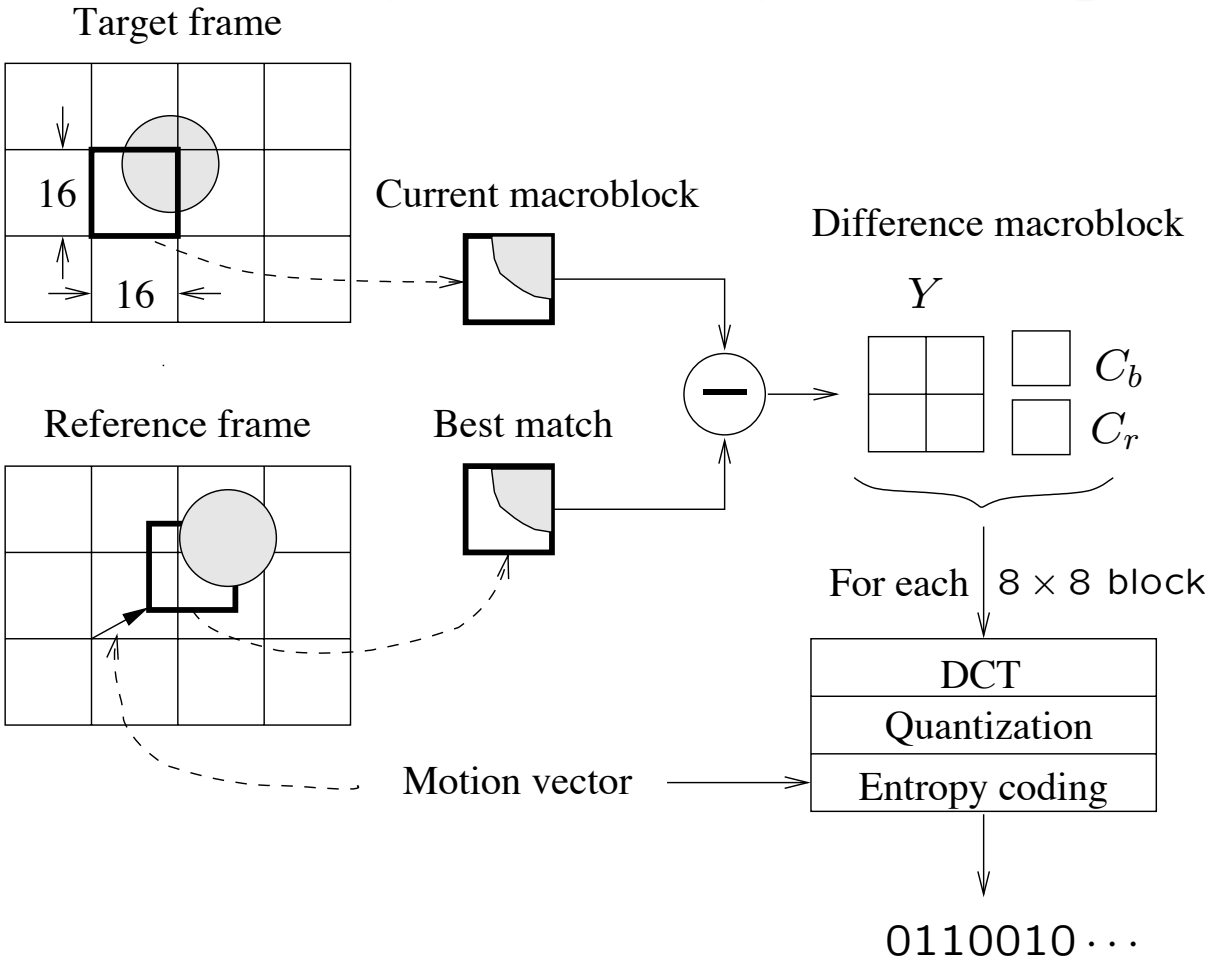


Fig. 10.5: I-frame Coding.

- **Macroblocks** are of size 16 × 16 pixels for the Y frame, and 8 × 8 for Cb and Cr frames, since 4:2:0 chroma subsampling is employed. A macroblock consists of four Y, one Cb, and one Cr 8 × 8 blocks.
- For each 8 × 8 block a DCT transform is applied, the DCT coefficients then go through quantization zigzag scan and entropy coding.

Inter-frame (P-frame) Coding



–For each macroblock in the Target frame, a motion vector is allocated by one of the search methods discussed earlier.

–A *difference macroblock* is derived to measure the *prediction error*.

–Each of these 8 × 8 blocks go through DCT, quantization, zigzag scan and entropy coding procedures.

Quantization in H.261

- The quantization in H.261 uses a constant step_size, for all DCT coefficients within a macroblock

for DC coefficients

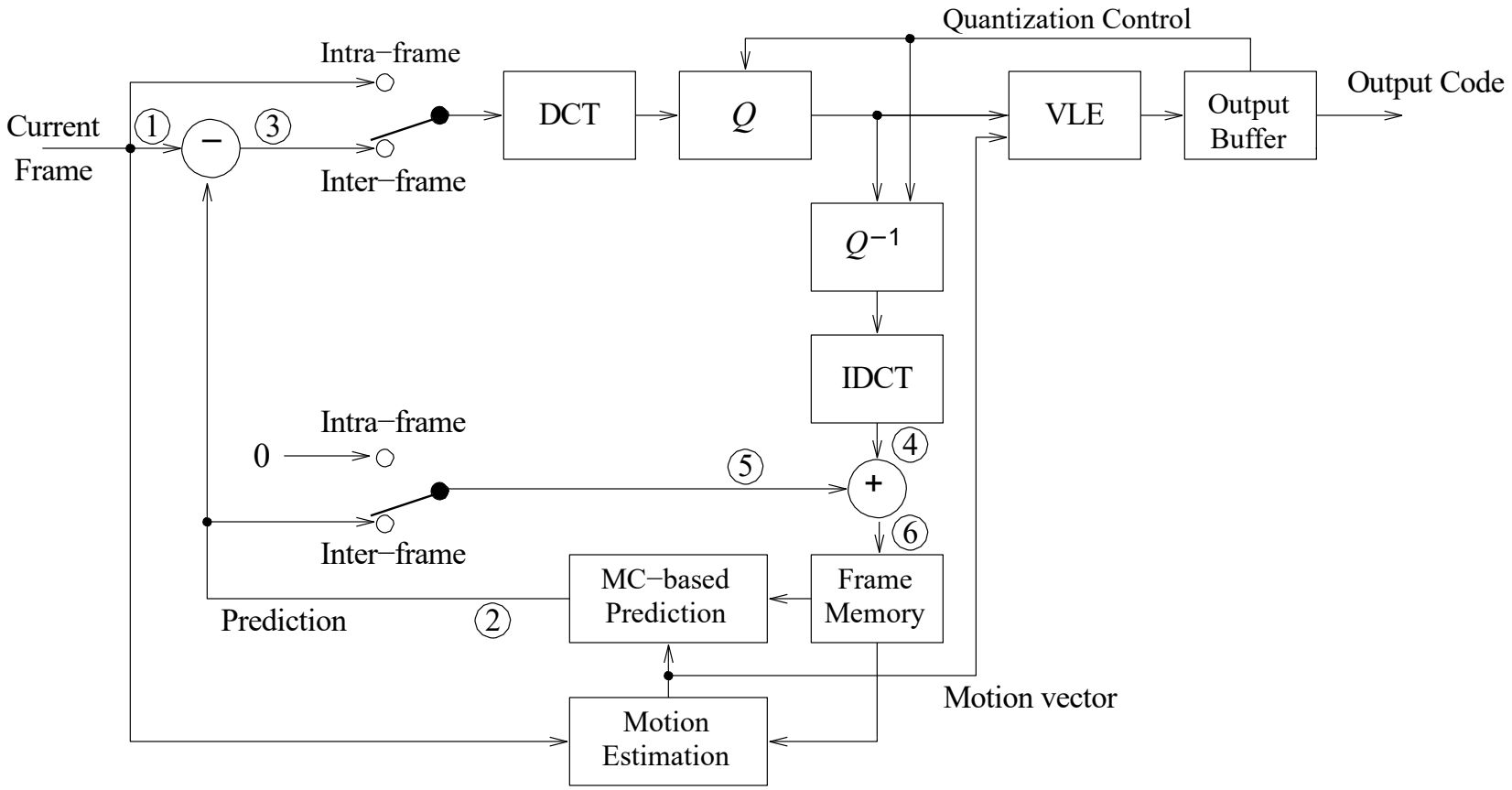
$$QDCT = round\left(\frac{DCT}{step\_size}\right) = round\left(\frac{DCT}{8}\right) \qquad (10.4)$$

for all other coefficients:

$$QDCT = \left\lfloor \frac{DCT}{step\_size} \right\rfloor = \left\lfloor \frac{DCT}{2 * scale} \right\rfloor \qquad (10.5)$$

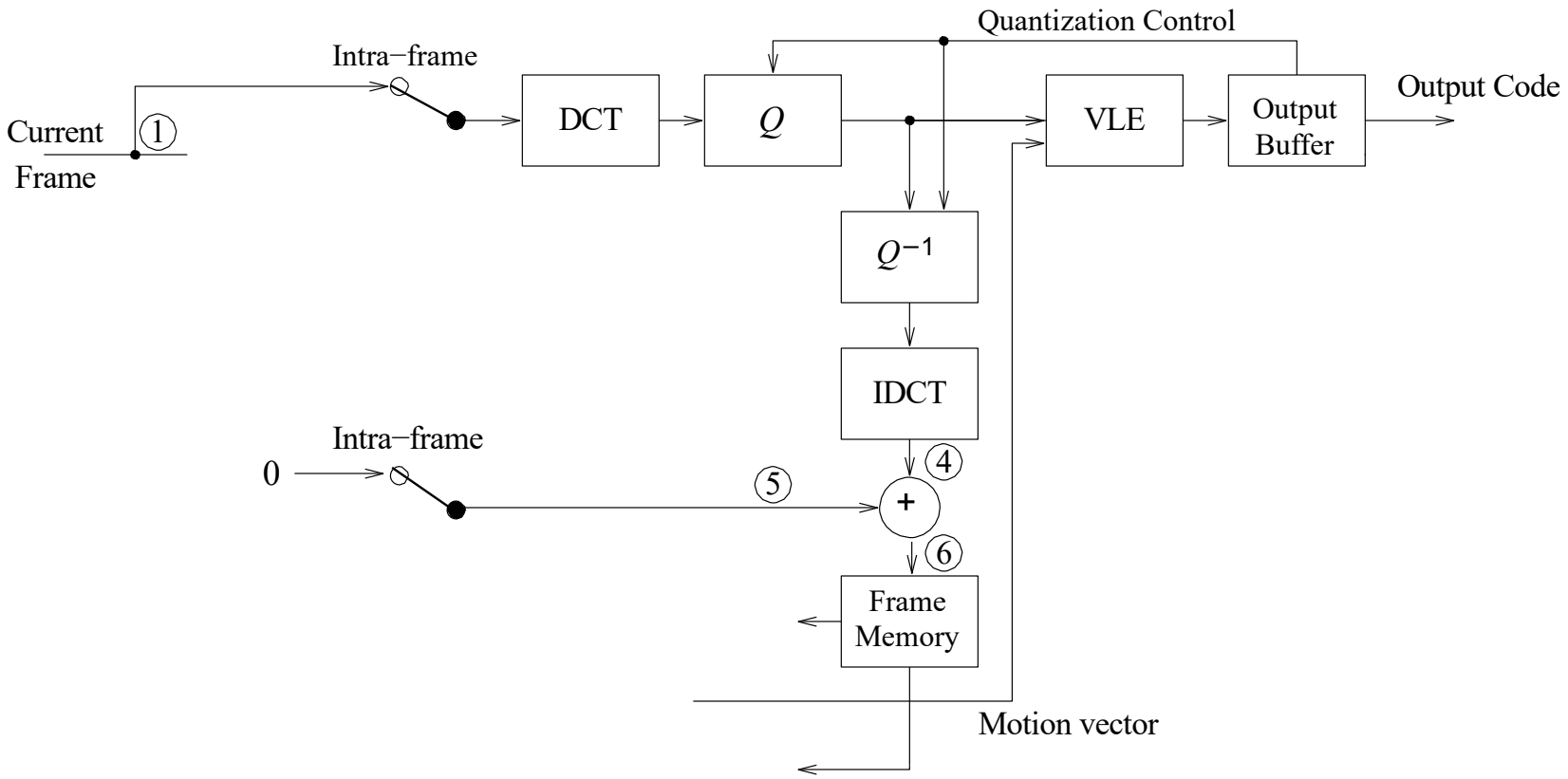
$scale$ — an integer in the range of [1, 31].

H.261 Encoder



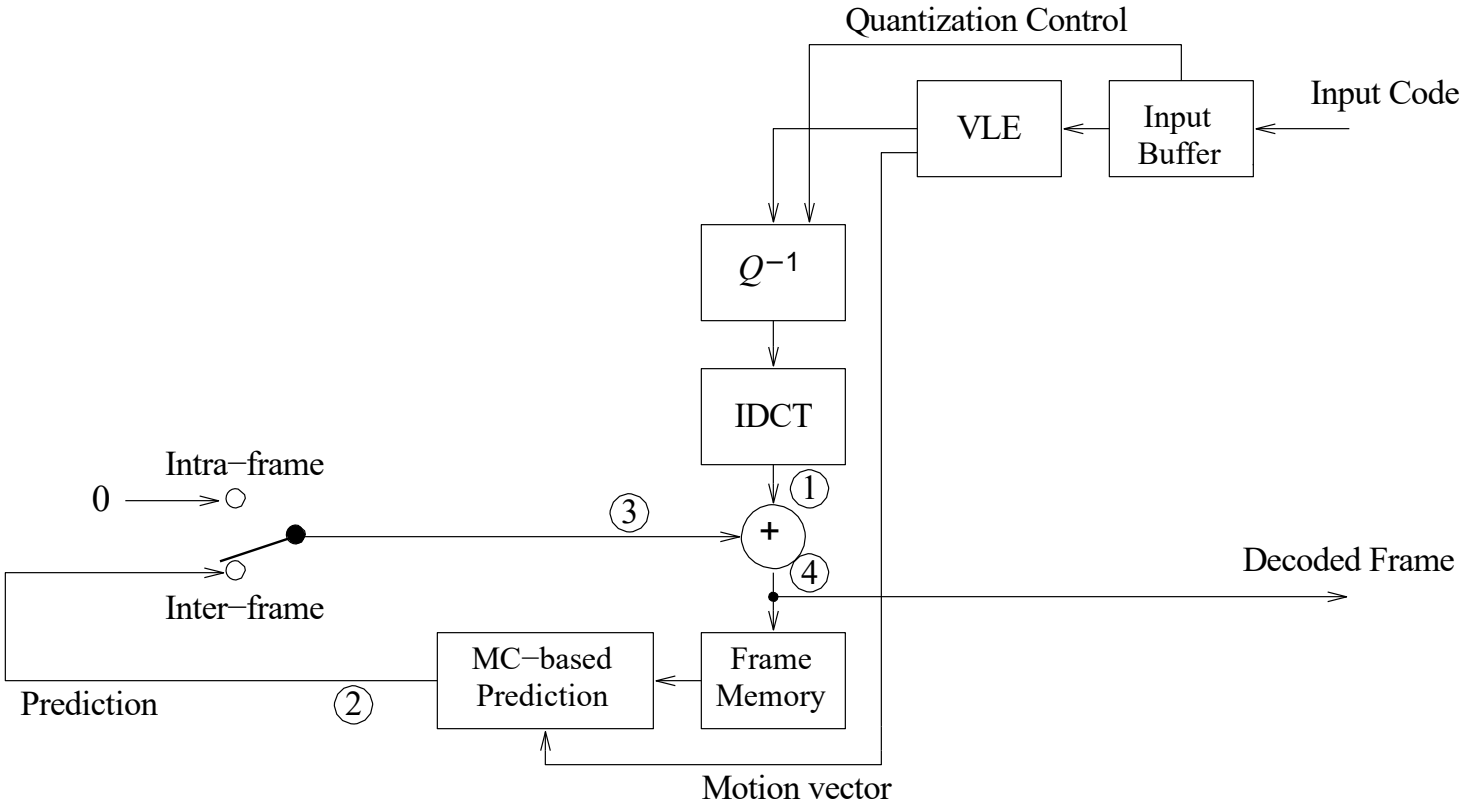
(a) Encoder

H.261 Encoder



(a) Encoder

H.261 Decoder



(b) Decoder

编码标准

- H.261编码标准介绍; H.261里面的I FRAME，和P FRAME 概念和编码过程。

- 介绍MPEG编码，MPEG-1编码中的运动补偿，与H.261的主要区别

MPEG Overview

- MPEG: Moving Pictures Expert Group(运动图像专家组), 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
  - The compression algorithms, and thus the encoders, are completely up to the manufacturers
  - MPEG标准主要有以下五个，MPEG-1、MPEG-2、MPEG-4、MPEG-7及MPEG-21等

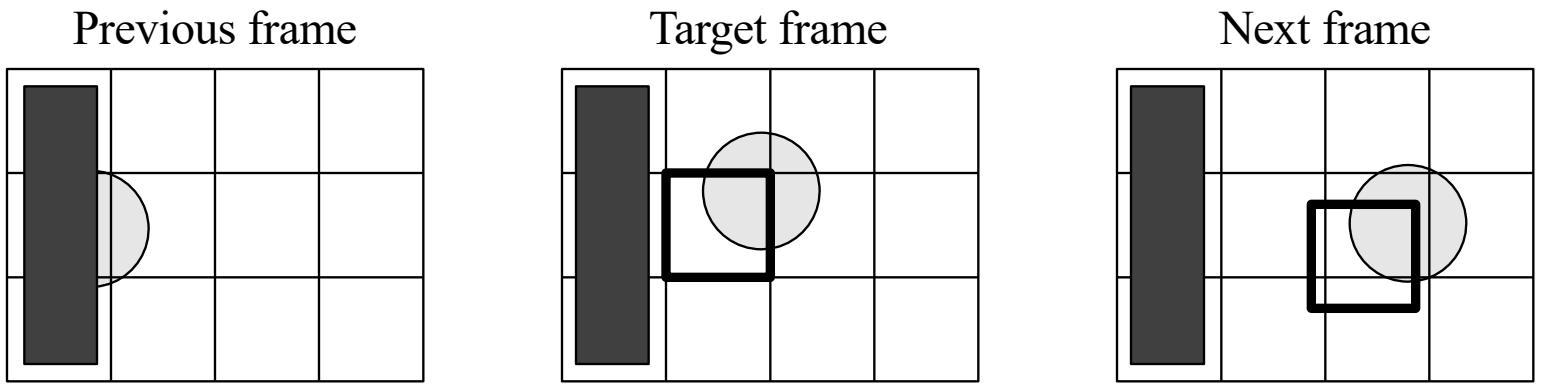
MPEG-1

- The MPEG-1 was approved by the (ISO/IEC) MPEG group in November 1991
- Coding of Moving Pictures and Associated Audio for Digital Storage Media， CDs， VCDs
- at up to about 1.5 Mbit/s
- MPEG-1 adopts the CCIR601 digital TV format also known as SIF (Source Input Format)
- MPEG-1 supports only non-interlaced (非隔行) video. It uses 4:2:0 chroma subsampling
- The MPEG-1 standard is also referred to as ISO/IEC 11172. It has five parts: 11172-1 Systems, 11172-2 Video, 11172-3 Audio, 11172-4 Conformance, and 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:** The difference between the MB and its matching MB, sent to DCT and its subsequent encoding steps.
  - The prediction is from a previous frame — **forward prediction**.

The Need for Bidirectional Search



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.

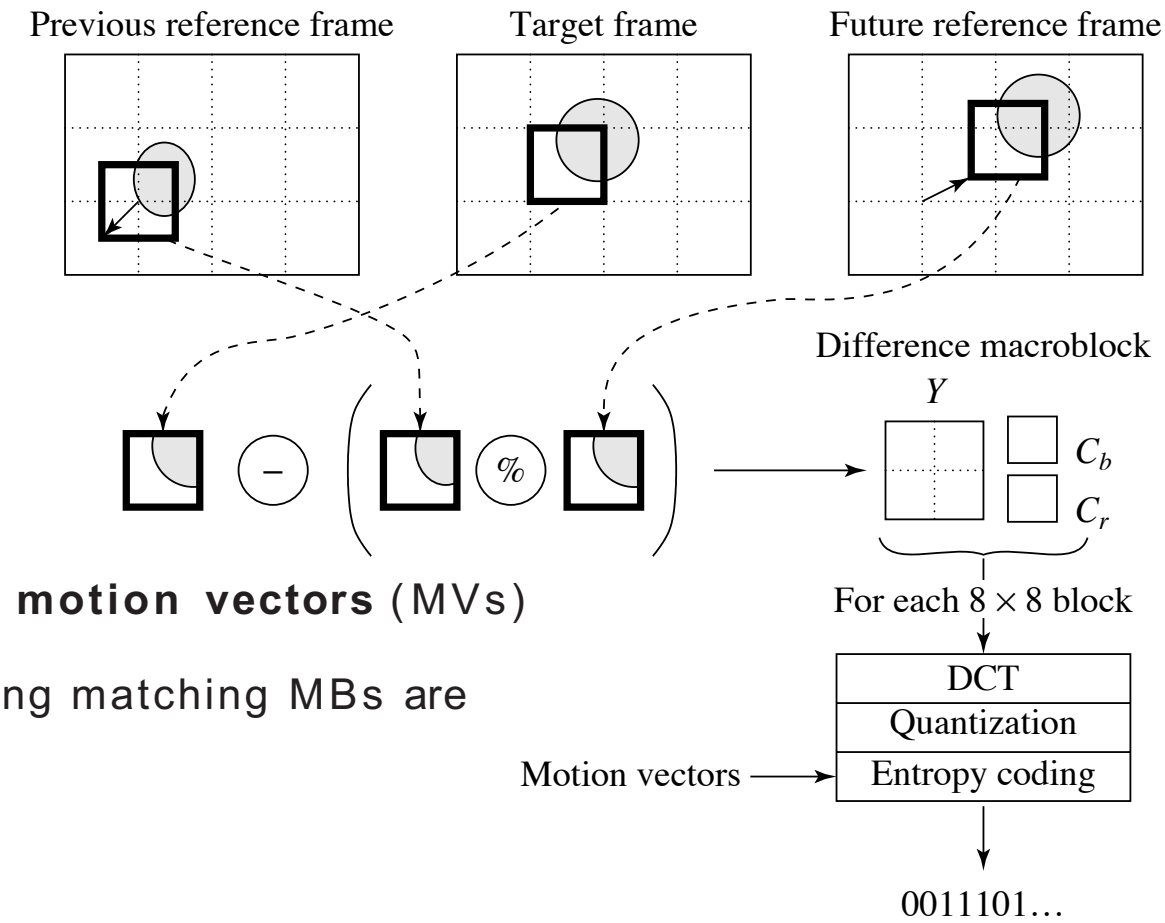
- The MC-based B-frame coding idea is :

  – Each MB from a B-frame will have up to ***two* motion vectors** (MVs) (one from the forward and one from the backward prediction).

  – If matching in both directions is successful, then two MVs will be sent and the two corresponding matching MBs are **averaged** (indicated by '%' in the figure) before comparing to the Target MB for generating the prediction error.

  – If an acceptable match can be found in only one of the reference frames, then only one MV and its corresponding MB will be used from either the forward or backward prediction.

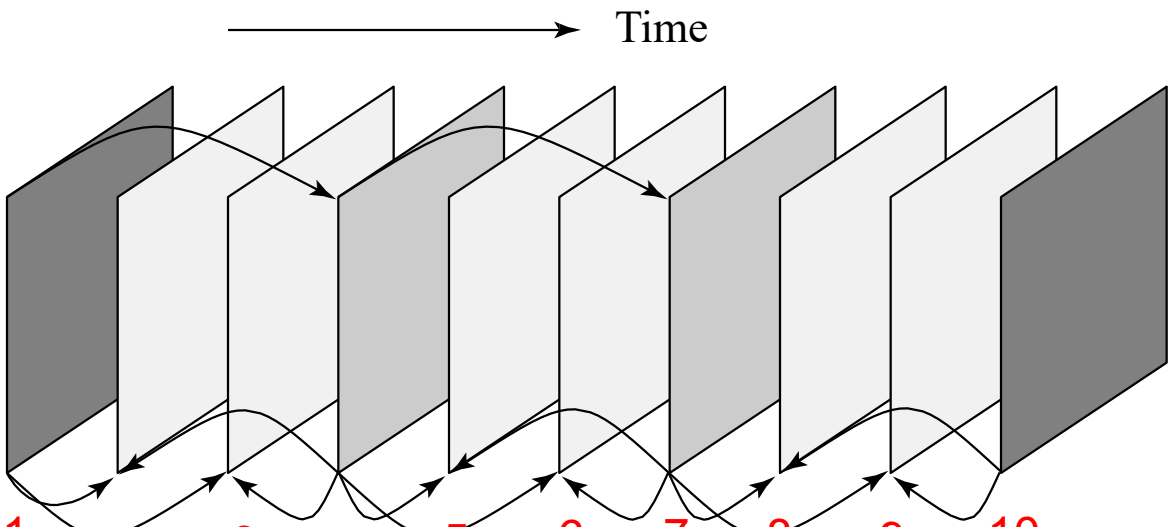
B-frame Coding Based on Bidirectional Motion Compensation



have up to ***two* motion vectors** (MVs)

two corresponding matching MBs are *averaged*

B-frame Coding Based on Bidirectional Motion Compensation



MPEG Frame Sequence.

Typical Sizes of MPEG-1 Frames

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

Table 11.4:Typical Compression Performance of MPEG-1 Frames

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

Typical Sizes of MPEG-1 Frames

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

If a video sequence has 2 I frame each second, 8 P frame each second and 20 B frame each second, how about the average compression rate?

$$\frac{2+8+20}{2\times\frac{1}{7}+8\times\frac{1}{20}+20\times\frac{1}{50}}$$

编码标准

- H.261编码标准介绍; H.261里面的I FRAME，和P FRAME 概念和编码过程，H.261的量化过程。

- 介绍MPEG编码，MPEG-1编码中的运动补偿，与H.261的主要区别

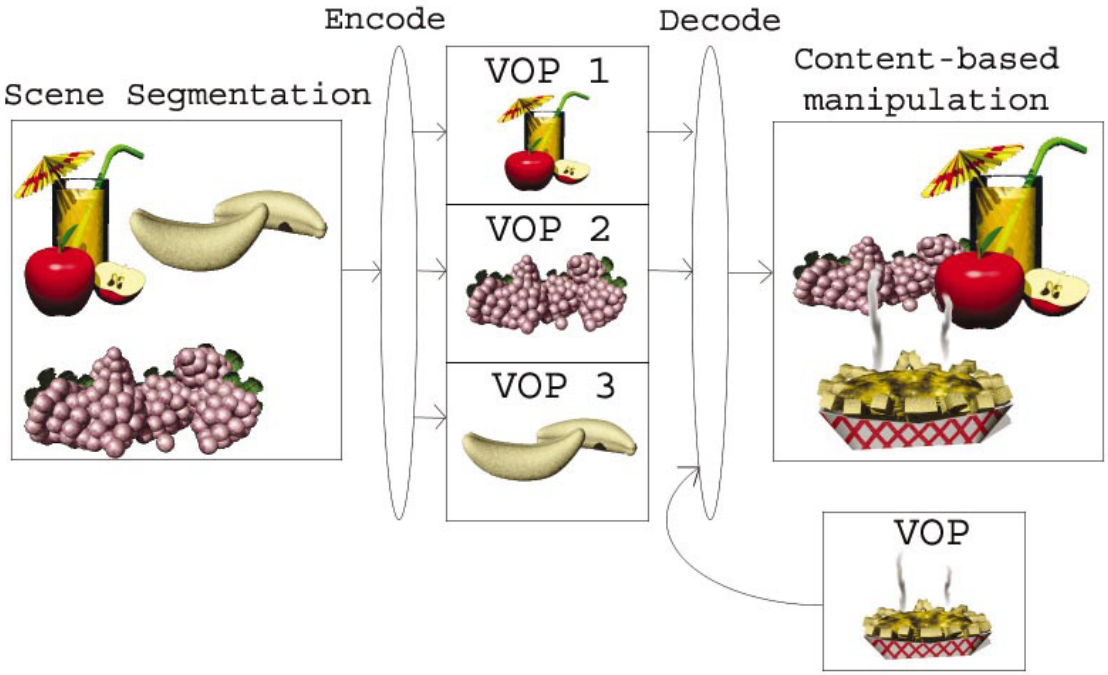
思考题

- MPEG-1标准中引入了B帧，给编码带来了很大的好处，如在低码率的情况下增加了SNR而且节省带宽，那么B帧有哪些缺点？

- 答案：
- 编码复杂度提高：相对P帧，两倍的运动向量搜索和运动补偿的计算量；
- 编解码依赖性：需要被参考的I/P帧编/解码完成后，方能完成相应的B帧编/解码。不适用于低时延要求的应用。

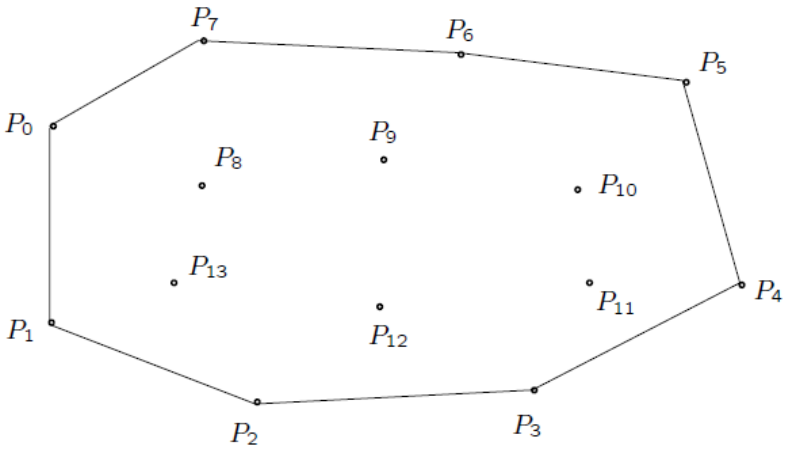
Overview of MPEG-4

- MPEG-4: a newer standard. Besides compression, pays great attention to issues about user interactivities.
- MPEG-4 departs from its predecessors in adopting a new object-based coding.
- The bit-rate for MPEG-4 video now covers a large range between 5 kbps to 10 Mbps.



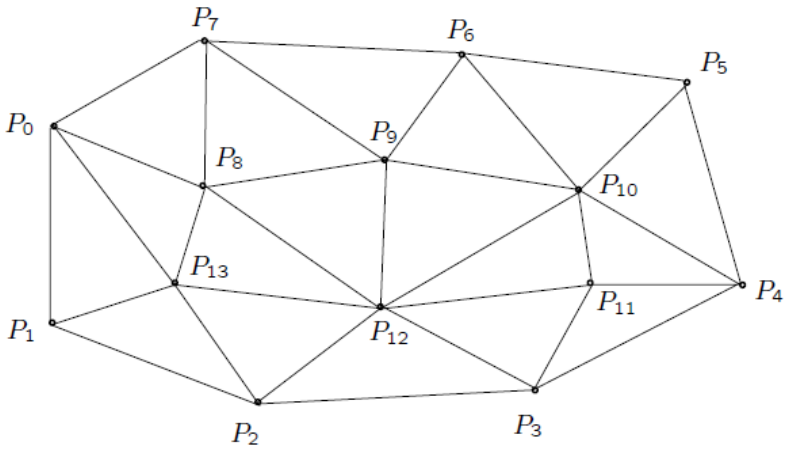
2D Mesh Object Coding

- The vertices of the polygons are referred to as nodes of the mesh.
- The most popular meshes are triangular meshes where all polygons are triangles.
- The MPEG-4 standard makes use of two types of 2D mesh: uniform mesh and Delaunay mesh



(a)

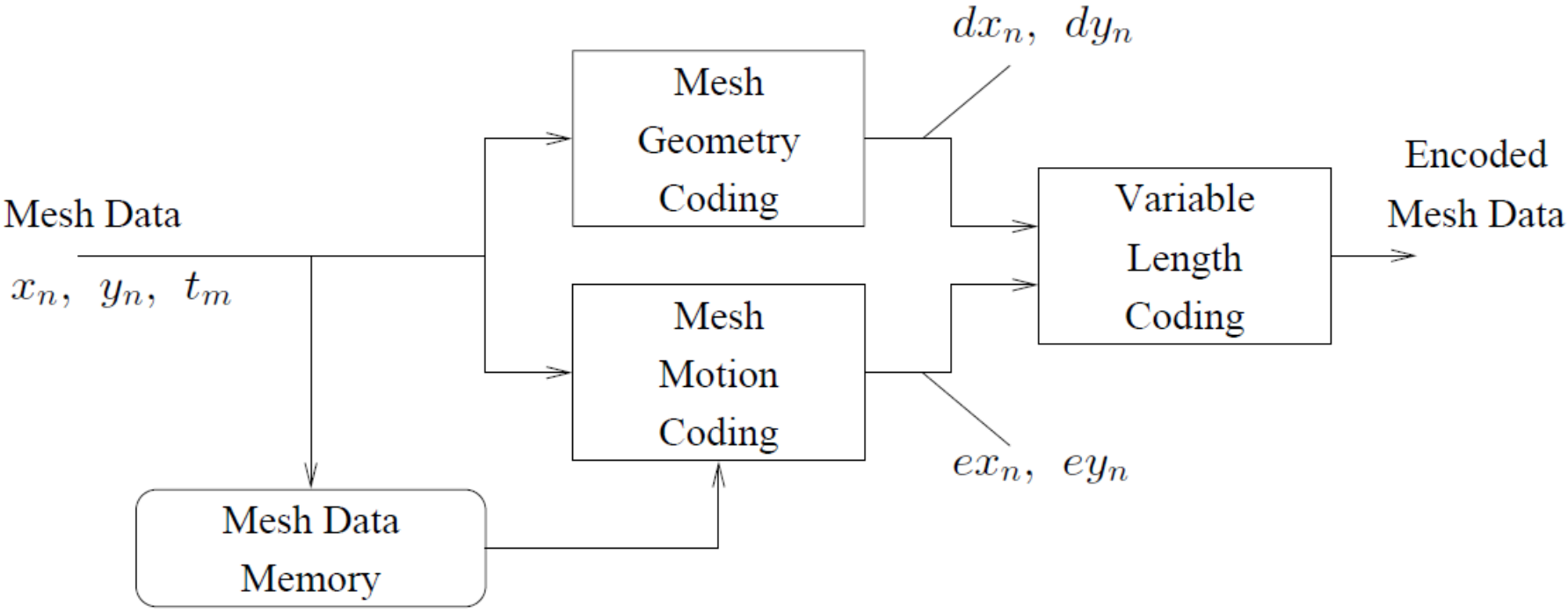
Nodes



(b)

Delaunay Mesh

2D Mesh Object Coding

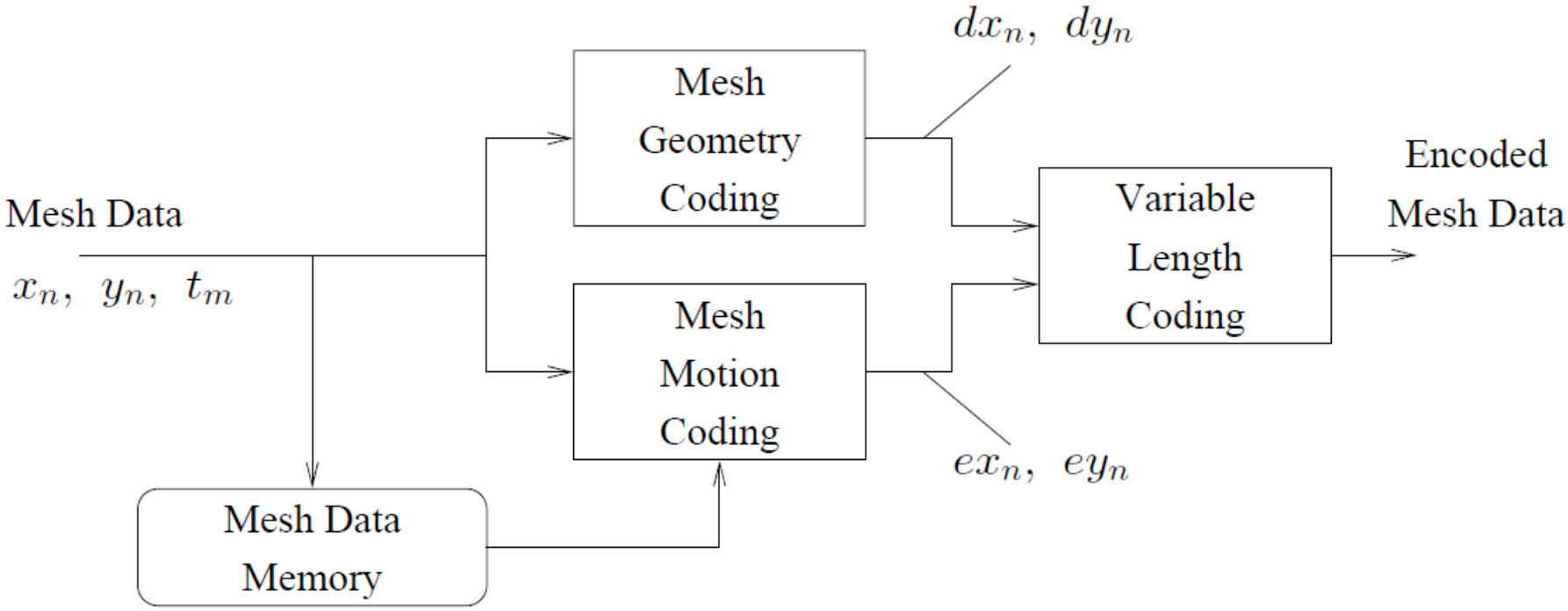


- Mesh geometry coding:

Except for the first location (x0; y0), all subsequent coordinates are coded deferentially—— that is, for n > 1,

$$dx_n = x_n - x_{n-1}, \qquad dy_n = y_n - y_{n-1},$$

2D Mesh Object Coding



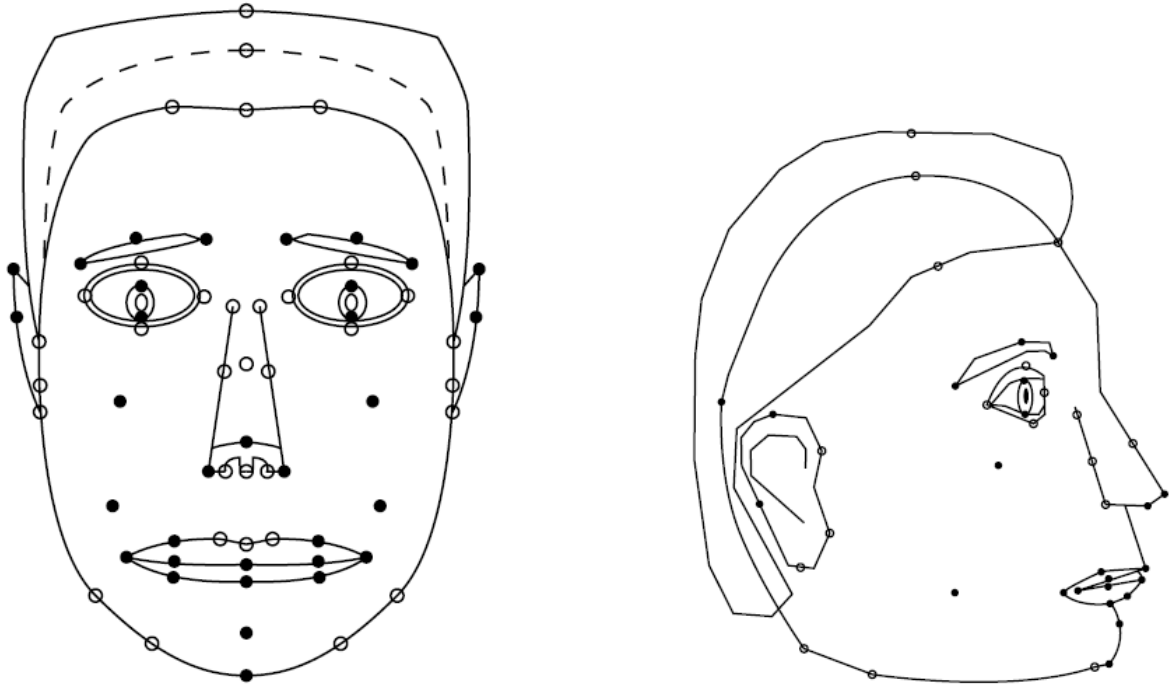
- Mesh motion coding:

For any MOP triangle (*Pi; Pj; Pk*), if the motion vectors for *Pi* and *Pj* are known to be $\mathbf{MV_i}$ and $\mathbf{MV_j}$, $\mathbf{Pred_k = 0.5(MV_i + MV_j)}$

The prediction error $\mathbf{e_k}$ is coded as, $ek = \mathbf{MV_k - Pred_k}$

3D Model-Based Coding

- MPEG-4 has adopted a generic default face model, which was developed by VRML Consortium.
- Face Animation Parameters (FAPs) can be specified to achieve desirable animations，deviations from the original neutral face.
- For compression, the FAPs are coded using predictive coding.



扩展

- 5G网络时代视频压缩

  https://v.qq.com/x/page/a0559imqg4y.html

- **只有5G而没有视频压缩，那么多媒体传输一切都是0**
- **视频流量一直是占领网络流量的先驱，在5G趋势下会只增不减**

扩展

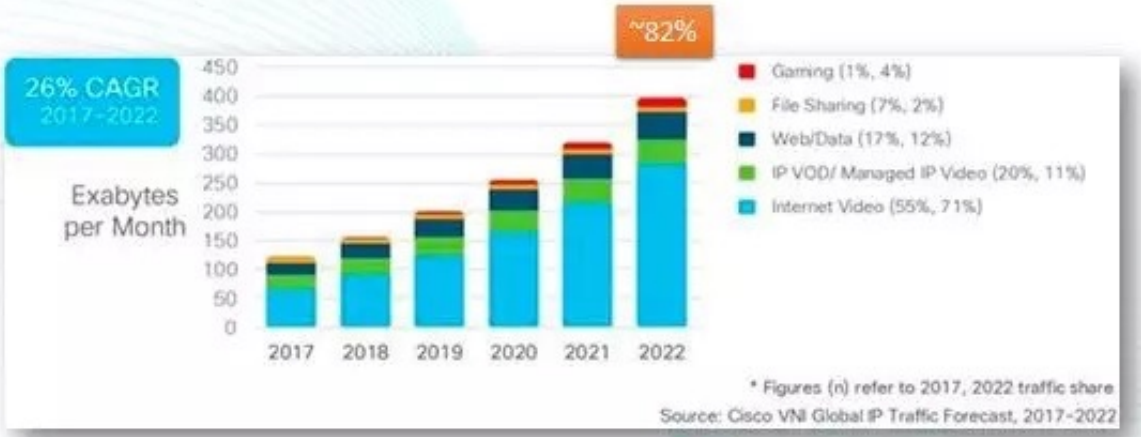
5G时代下的媒体应用挑战

Media application Challenges under 5G

Tencent Media Lab

- 视频流量主导未来网络发展
  - 4K，8K，沉浸式媒体
  - 超低延迟的需求：
    - 云游戏
    - 自动驾驶
  - IoT， OTT

- 现实：
  - 5G带宽仍满足不了需求
  - 单个应用带宽需求增长
  - 接入设备增量巨大
  - 双向交互的需要 – 上传压力大



出现新型视频压缩标准：比如**VVC（H266）**视频压缩标准