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
Basic Video Compression Techniques

Outline of Lecture 10

- Introduction to Video Compression
- Video Compression with Motion Compensation
- Motion Compensation
- Sequential Search
- Logarithmic Search
- Hierarchical Search
- Reference reading
  - Chapter 10

Introduction to Video Compression

- A video consists of a time-ordered sequence of frames → images

- An obvious solution to video compression would be predictive coding based on previous frames
  - Compression proceeds by subtracting images: subtract in time order and code the residual error

Introduction to Video Compression

- An obvious solution to video compression would be predictive coding based on previous frames

Introduction to Video Compression

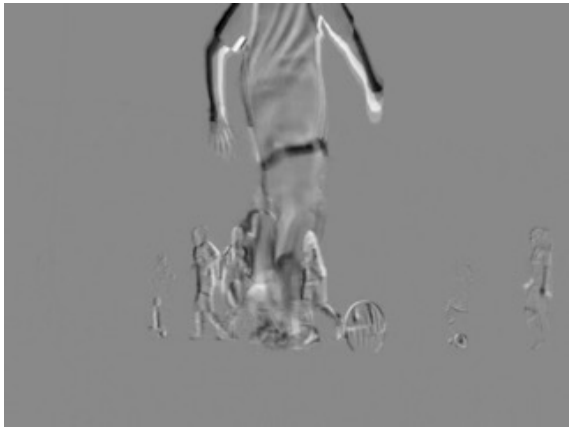
- An obvious solution to video compression would be predictive coding based on previous frames

Introduction to Video Compression

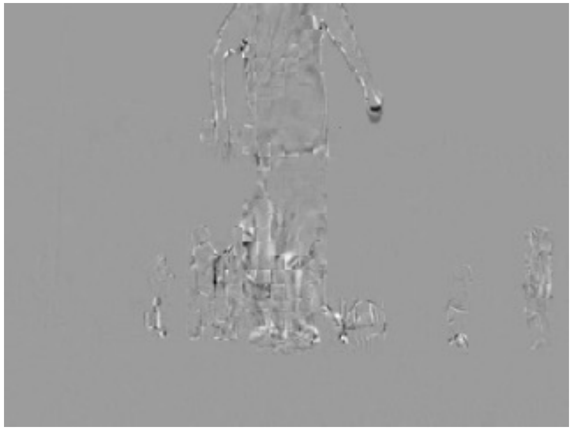
- It can be done even better by searching for just the right parts of the image to subtract from the previous frame

Reference

Target



prediction without MC



MC prediction

Outline of Lecture 10

- Introduction to Video Compression
- Video Compression with Motion Compensation
- Motion Compensation
- Sequential Search
- Logarithmic Search
- Hierarchical Search
- Reference reading
  - Chapter 10

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

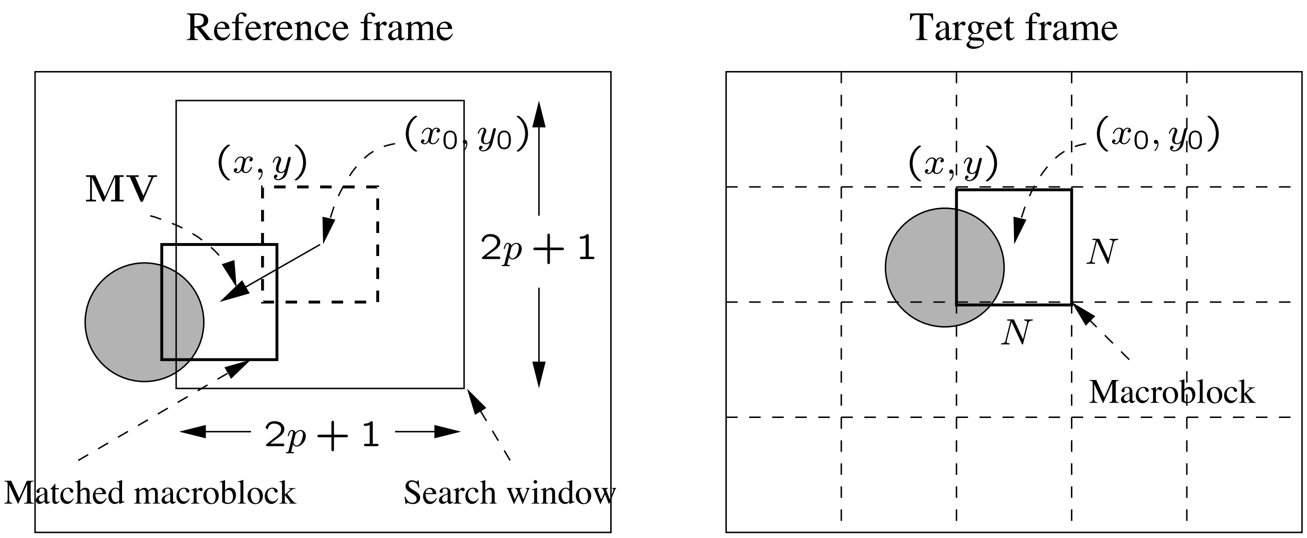
Outline of Lecture 10

- Introduction to Video Compression
- Video Compression with Motion Compensation
- Motion Compensation
- Sequential Search
- Logarithmic Search
- Hierarchical Search
- Reference reading
  - Chapter 10

Motion Compensation

- Each image is divided into macroblocks（宏块）of size N × N .
  - By default, N = 16 for luminance images.
  - For chrominance images, N = 8 if 4:2:0 chroma subsampling is adopted.
- Motion compensation is performed at the macroblock level.
  - The current image frame is referred to as ***Target Frame***.
  - A match is sought between the macroblock in the Target Frame and the most similar macroblock in previous and/or future frame(s) (referred to as ***Reference frame***(*s*)).
  - The displacement of the reference macroblock to the target macroblock is called a *motion vector* **MV**.

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

- Calculating the displacement of the reference macroblock to the target macroblock

Reference image

Target image

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,

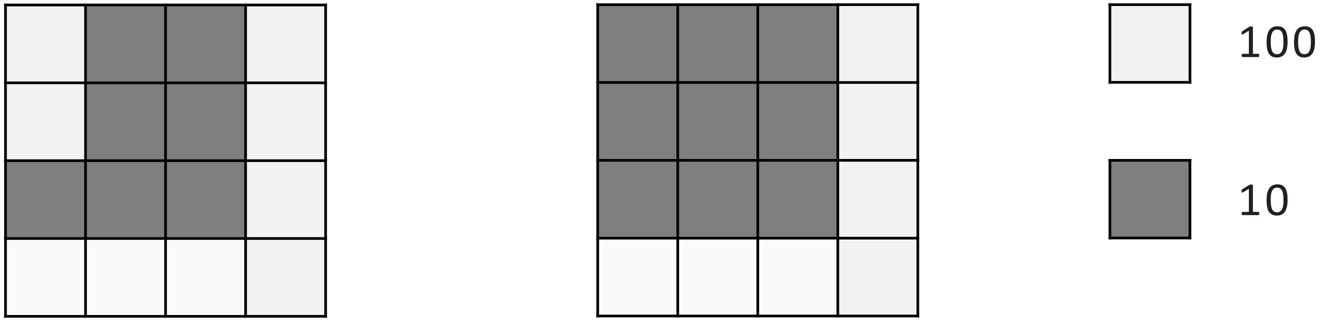
$C(x+k, y+l)$ – pixels in macroblock in Target frame,

$R(x+i+k, y+j+l)$ – 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

$$(u, v) = [\ (i, j)\ \mid MAD(i, j) \text{ is minimum}, i \in [-p, p],\ j \in [-p, p]\ ]$$

Mean Absolute Difference (MAD)



Reference

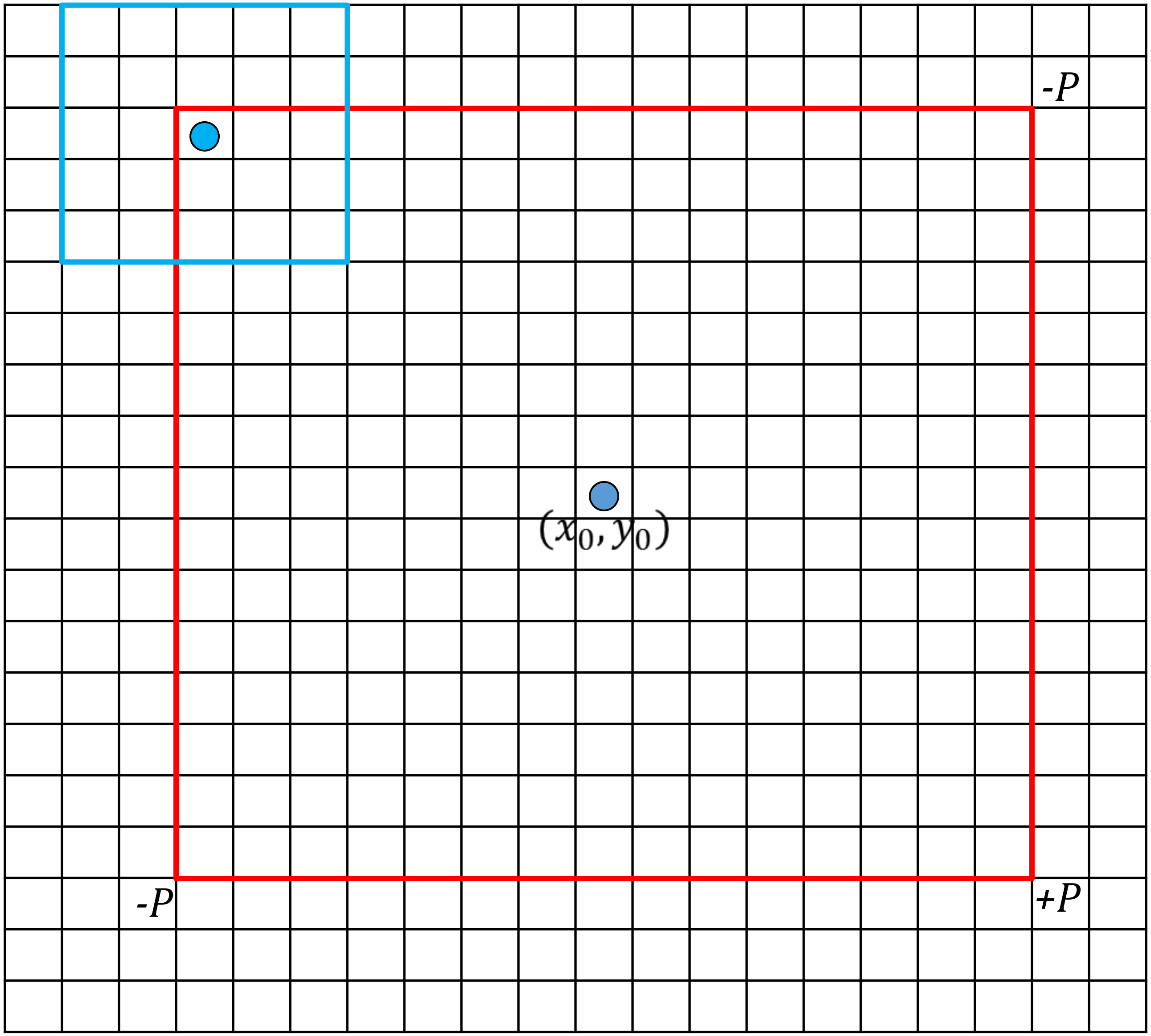
Target

Calculate MAD：
MAD=(|10-100|+|10-100|+7* |10-10|+7* |100-100|)/16=180/16

Outline of Lecture 10

- Introduction to Video Compression
- Video Compression with Motion Compensation
- Motion Compensation
- Sequential Search
- Logarithmic Search
- Hierarchical Search
- Reference reading
  - Chapter 10

Sequential Search（顺序搜索）



P=7

N=5

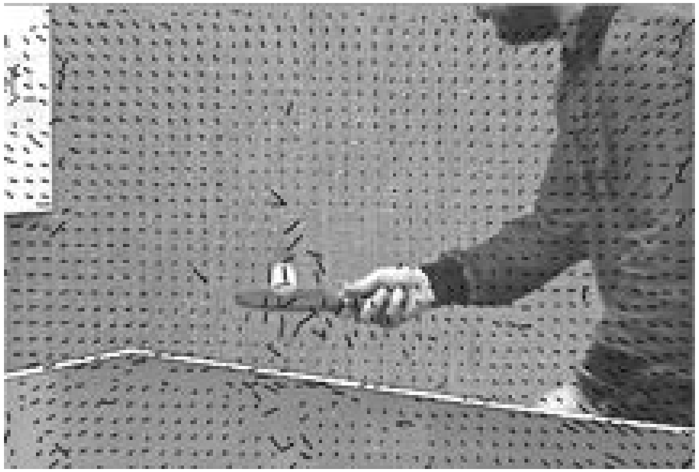
Sequential Search（顺序搜索）

- Sequential search: sequentially search the whole $(2p+1) \times (2p+1)$ window in the Reference frame

  - The vector (i, j) that offers the least M AD is designated as the MV (u, v) for the macroblock in the Target frame.
  - This method is costly — assuming each pixel comparison requires three operations (subtraction, absolute value, addition), the cost for obtaining a motion vector for a single macroblock is

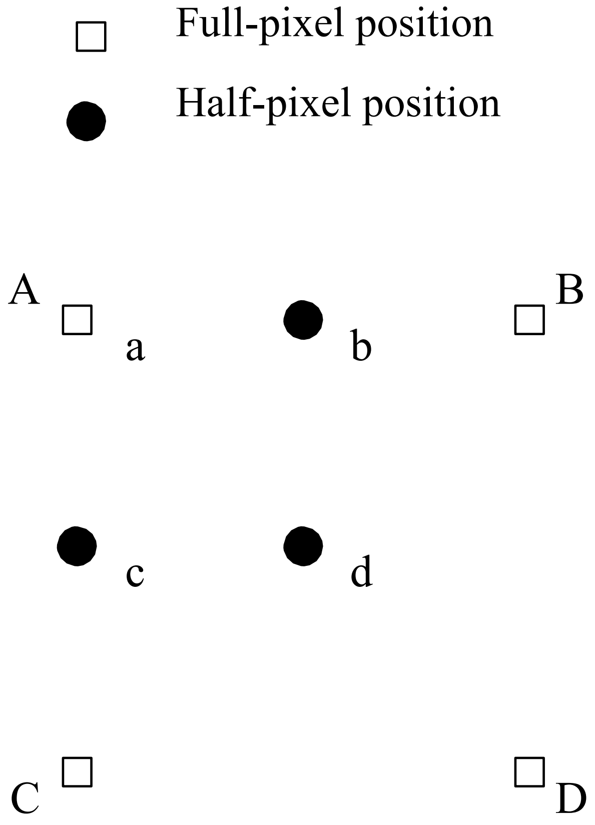
$$(2p+1)\cdot(2p+1)\cdot N^2\cdot 3 \Rightarrow \mathrm{O}(p^2N^2)$$

Sequential Search（顺序搜索）



- See demo ***predictionMC.m***

思考：MV（u，v）只能为整数吗？


□ Full-pixel position

● Half-pixel position

Half-pixel Prediction by Bilinear Interpolation in H.263

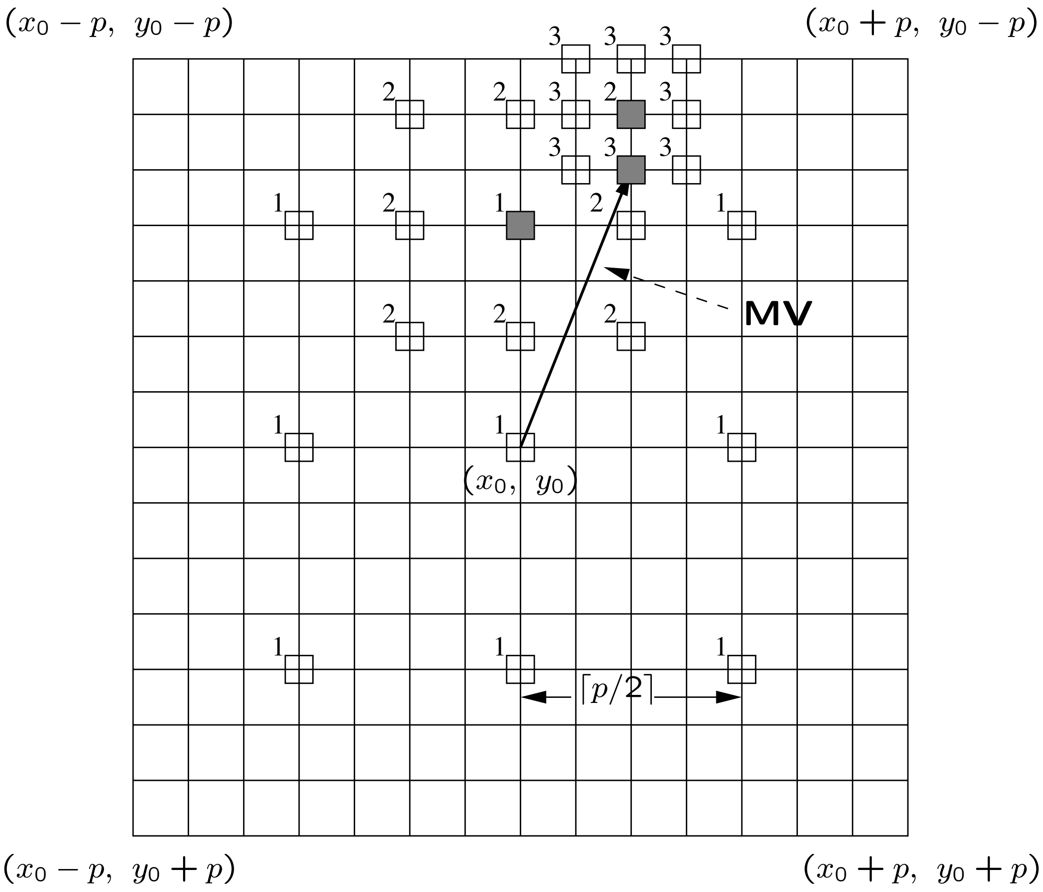
Outline of Lecture 12

- Introduction to Video Compression
- Video Compression with Motion Compensation
- Motion Compensation
- Sequential Search
- Logarithmic Search
- Hierarchical Search
- Reference reading
  - Chapter 10

2D Logarithmic Search

- Logarithmic search: a cheaper version, which is suboptimal but usually effective
- The procedure for 2D Logarithmic search takes several iterations (迭代) and is akin to a binary search
  - Initially only nine locations in the search window are used as seeds for a MAD-based search
  - After the one that yields the minimum *MAD* is located, the center of the new search region is moved to it and the step-size ("offset") is reduced to half
  - In the next iteration, the nine new locations are marked as '2', and so on.

2D Logarithmic Search for Motion Vectors

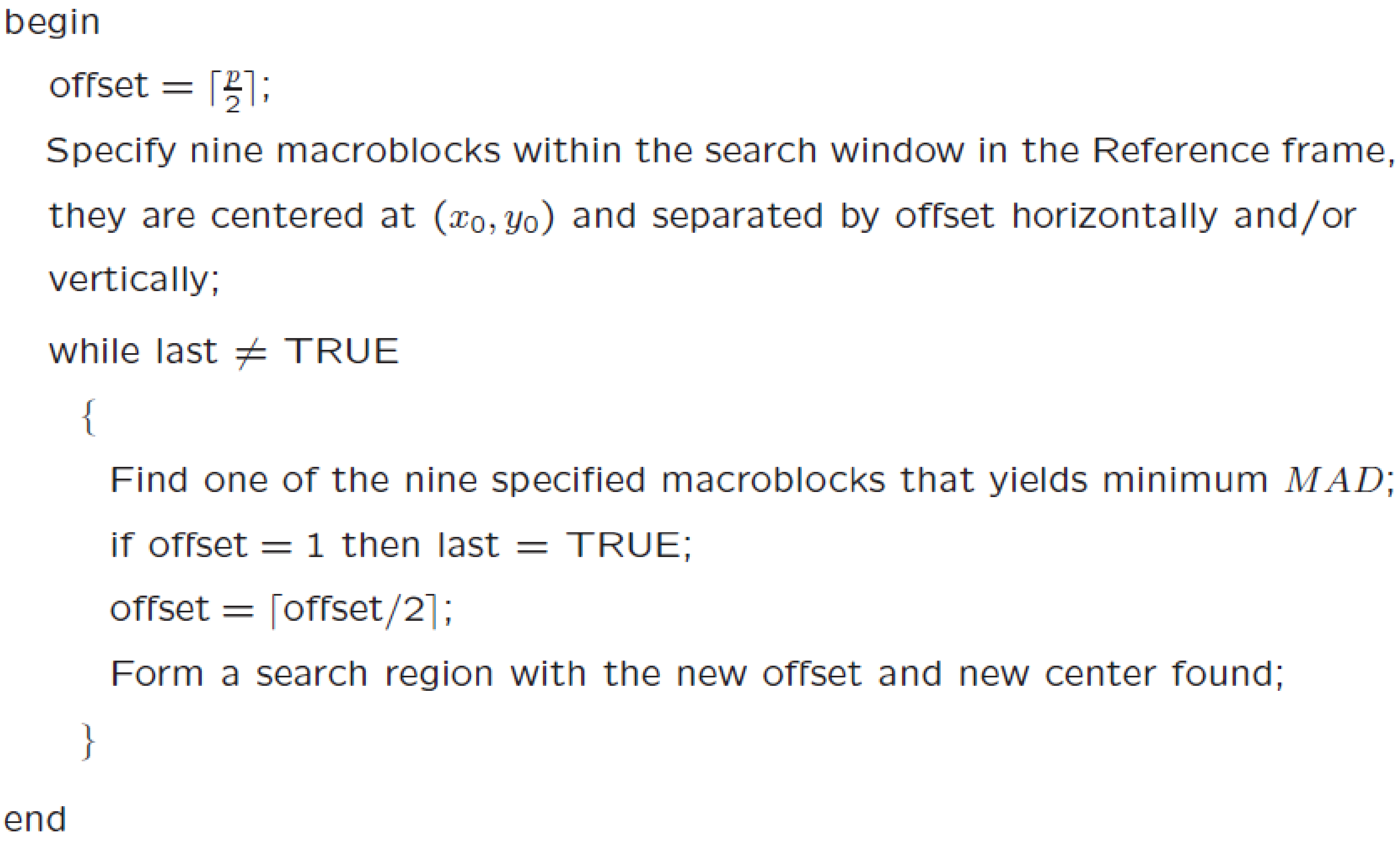


The total operations per second is dropped to:

$$O(\log p^{*} N^{2})$$

每一步搜索：步长=$\left\lceil \frac{p}{2} \right\rceil$, 最优MV，MAD

Algorithm of 2D-Logarithmic Search



```
begin
    offset = $\lceil \frac{p}{2} \rceil$;
    Specify nine macroblocks within the search window in the Reference frame,
    they are centered at $(x_0, y_0)$ and separated by offset horizontally and/or
    vertically;
    while last $\neq$ TRUE
      {
        Find one of the nine specified macroblocks that yields minimum $MAD$;
        if offset = 1 then last = TRUE;
        offset = $\lceil$offset/2$\rceil$;
        Form a search region with the new offset and new center found;
      }
end
```

Outline of Lecture 12

- Introduction to Video Compression
- Video Compression with Motion Compensation
- Motion Compensation
- Sequential Search
- Logarithmic Search
- Hierarchical Search
- Reference reading
  - Chapter 10

Hierarchical Search （分层搜索）

- Sequential search method is very costly

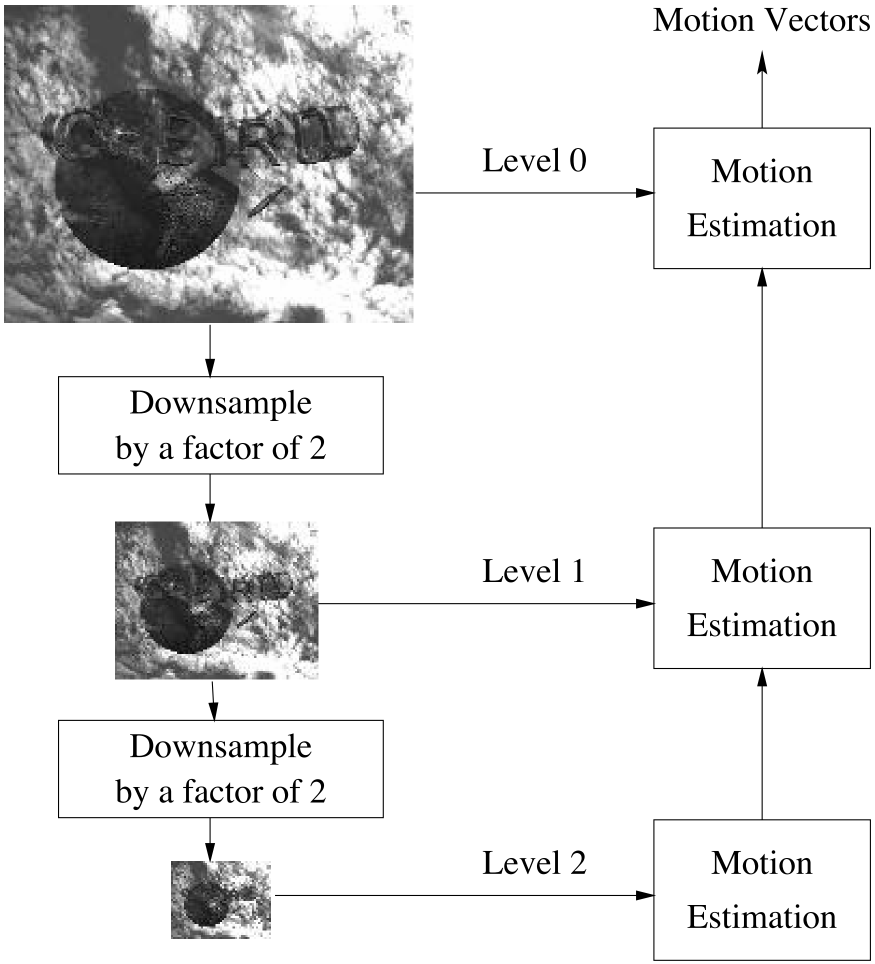
$$(2p+1)\cdot(2p+1)\cdot N^2\cdot 3 \Rightarrow O(p^2N^2)$$

- Reduce p and N, could reduce the cost.
- The search can benefit from a hierarchical (multiresolution) approach in which initial estimation of the motion vector can be obtained from images with a significantly reduced resolution.

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.



Comparison of Computational Cost

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

练习题

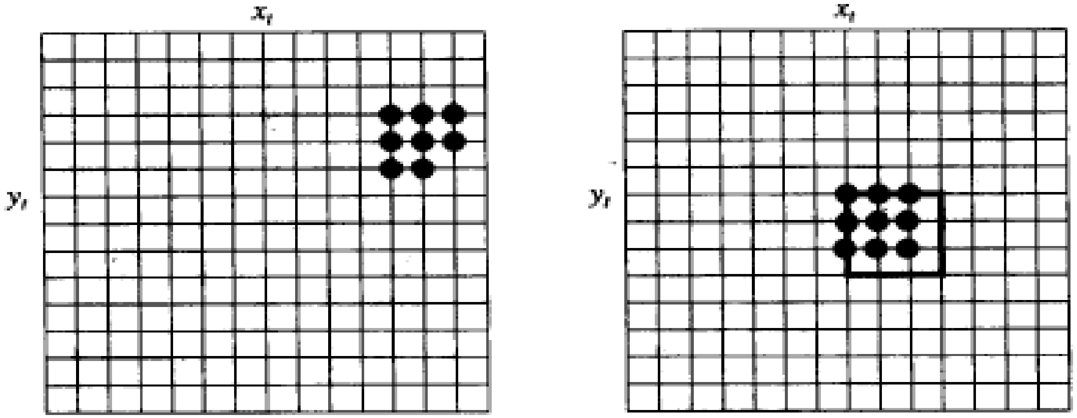
4. 回答下面用于运动向量的 2D 对数搜索的问题（见图 10-14）。

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 $p = 5$。

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。



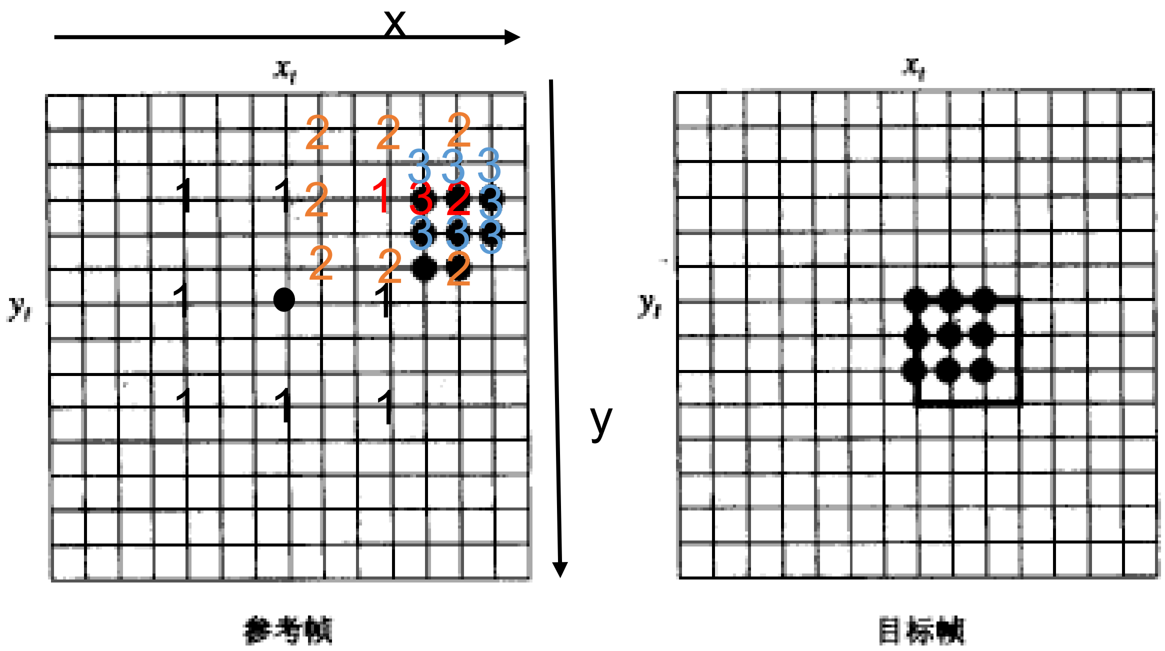
参考帧　　　　目标帧

● 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

思考题

- 在基于块的视频编码中，压缩或者解压缩哪个耗费代价高？简要说明原因。

- 答：压缩环节耗费代价高，因为运动向量的获取，需要多次比较计算当前帧中的预测块与参考帧中的待匹配块之间的MAD值；而解码时仅需直接从码流中根据运动向量的值，进行运动补偿即可。