

# 数字媒体技术基础 ——数字视频技术

王晗

wanghan@bjfu.edu.cn



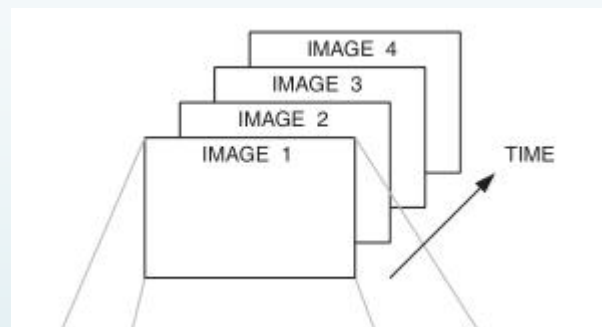
# 专题提纲

- 视频的概念
- 数字视频的获取与编辑
- 运动估计
- 视频视频编码标准

# 视频的概念

## ► 什么是视频

视频（Video）就是其内容随时间变化的一组动态图像，所以又叫运动图像或活动图像。



电影

电视



短视频



视频直播



视频监控

# 数字视频的概念

- 根据视觉暂留原理，连续的图像变化每秒超过24帧（frame）画面以上时，人眼无法辨别单幅的静态画面，看上去是平滑连续的视觉效果。
- **视觉暂留**特性，即人眼能将看到的影像在视网膜上保留一段时间。实验证明，人眼的某个视像消失后，仍可使该物像在视网膜上滞留0.1 – 0.4秒左右
- 在电影放映的过程中，电影胶片以每秒24格画面匀速转动，这就相当于每一格画面给人眼的刺激是 $1/24$ 秒（相当于0.04秒），由于人的眼睛有视觉暂留的特性，一个画面的印象还没有消失，下一个稍微有一点差别的画面又出现在银幕上，连续的影像衔接起来就组成了活动的影像。



# 数字视频的概念

## ——模拟视频和数字视频

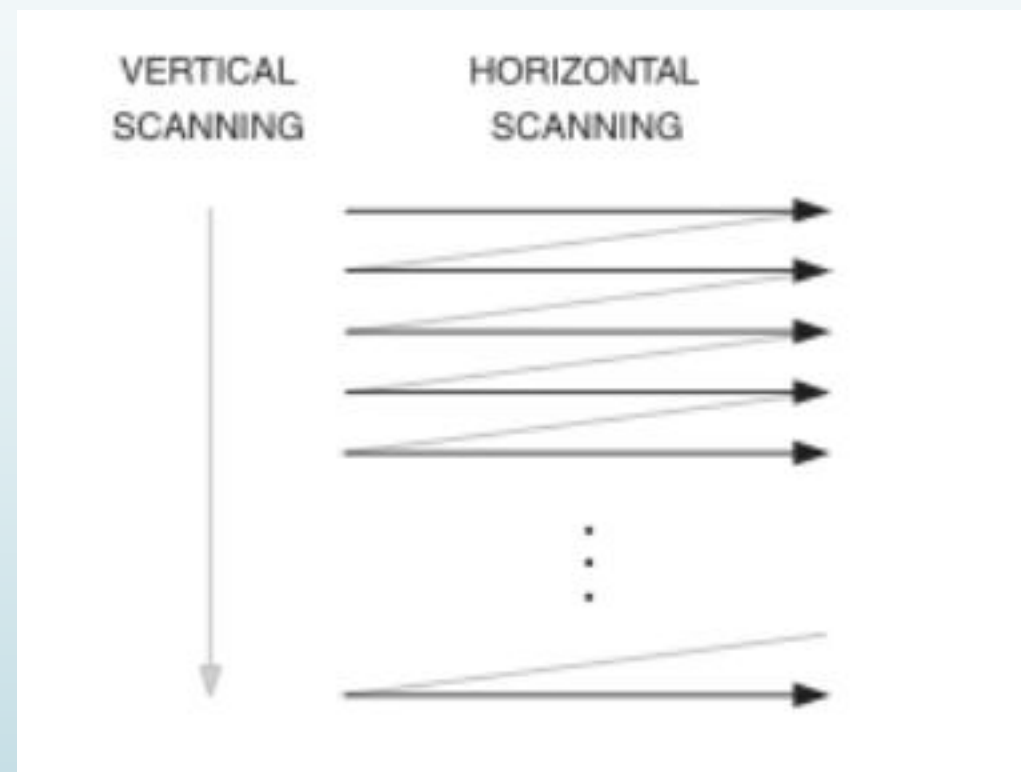
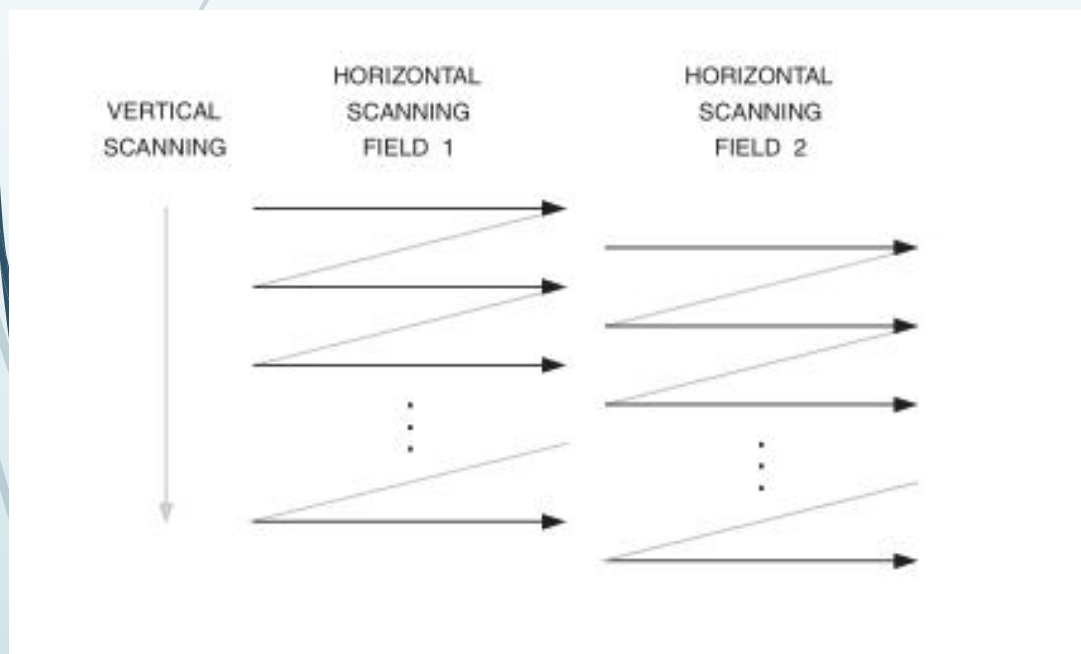
### ■ 模拟视频

- 普通广播电视信号是一种典型的模拟视频信号。
- 电视摄像机通过电子扫描将时间、空间函数所描述的景物进行光电转换后，得到单一的时间函数的电信号，其电平的高低对应于景物亮度的大小，即用一个电信号来表征景物。
- 由于光是连续的，投影后的图像在空间上是连续的，也即原始视频在空间上是连续的。

# 视频的概念

## ——隔行扫描和逐行扫描

- ➡ 隔行扫描（interlaced scanning）
- 逐行扫描（progressive scanning）



# 视频的概念

## ——隔行扫描和逐行扫描

- 视频在获取或显示时，每秒钟按逐行扫描处理的图像数称为**帧率 (FPS)**，每秒钟按隔行扫描处理的图像数称为**场率**，例如每秒25帧或50场(60场/s)。
- 逐行扫描和隔行扫描
  - 一幅图像的像素按行排列可以分为偶数行和奇数行
  - 如果偶数行和奇数行的像素是在相同时间点采样得到的，则称该帧为**逐行帧**
  - 如果偶数行和奇数行的像素是在不同的时间点采样得到的，则称该帧为**隔行帧**
  - 隔行帧中所有偶数行构成该图像的**顶场**，所有奇数行构成该图像的**底场**
  - 由逐行帧构成的视频序列称为**逐行视频**，由隔行帧构成的视频序列称为**隔行视频**
  - 在隔行视频与逐行视频具有相同数据量的情况下，采用隔行采样的方法能够增加时间采样率，使时域上的运动更加平滑，从而增强视频的数据质量

# 视频的概念

## ——模拟视频制式

- 目前国际上最常用的三大视频标准制式，它们分别定义了彩色电视机对于所接受的电视信号的解码方式、色彩处理方式和屏幕的扫描频率：

- NTSC (National Television Standards Committee)

- 水平扫描525行、帧率30帧/秒、隔行扫描、60场/s

- 北美、日本、中国台湾地区、加勒比和南美部分区域

- PAL (Phase Alternating Line)

- 水平扫描625行、帧率25帧/秒、隔行扫描、50场/s

- 中国、英国、德国、澳大利亚、亚洲国家、非洲和南美

- SECAM (Sequential Color Memory)

- 水平扫描625行、帧率25帧/秒、隔行扫描、50场/s

- 东欧和法国

不同电视标准制式间互相不兼容，某一制式的电视节目需要转换才能被另一制式的电视播放



# 数字视频的概念

## ——模拟视频和数字视频

- 1990年，美国通用仪器公司研制出高清晰度电视HDTV（High Definition Television），提出信源的视频信号及伴音信号用数字压缩编码，传输信道采用数字通信的调制和纠错技术，从此出现了信源和传输通道全数字化的真正数字电视DTV（Digital Television）。
- 数字视频（Digital Video, DV）
  - 是指用二进制数字表示的视频信号。
  - 数字视频既可直接来源于数字摄像机(例如CCD摄像机等)，也可将模拟视频信号经过数字化处理变成数字视频信号。
  - 模拟视频信号经过采样、量化和编码数字化处理后，就变成由一帧帧数字图像组成的图像序列，即数字视频信号。

# 数字视频的概念

## ——数字化

- 视频是由按一定时间间隔获取的图像序列，序列中的一幅图像也被称为一帧图像。
- 单位时间内采集到的视频图像的帧数称为**帧率**，常用的单位为“**帧/秒**”（**fps**）。帧率的倒数就是时间采样的间隔或两帧间的时间间隔。
- 视频除了包含场景的空域信息外，还包含场景的时域信息，通过在不同的时间点上采集二维的视频场景图像，可以得到视频的多个时间采样点，即时间采样。数字视频序列就是有一段时间间隔内的空间采样点构成的。因为空间采样点构成了图像，所以可以说视频序列是由运动图像构成的。
- 在一段时间间隔内，视频场景中的对象很可能发生运动，因此时间采样包含了运动信息。视频序列对场景中运动物体的描述质量受时间采样率高高低的影响，时间采样率高的视频序列会有更平滑的运动轨迹，能更好地与自然场景中的物体运动相吻合，因此视频的时间采样率的高低也决定了视频序列的视觉质量。

# 数字视频的获取与编辑

## ——视频信息的采集

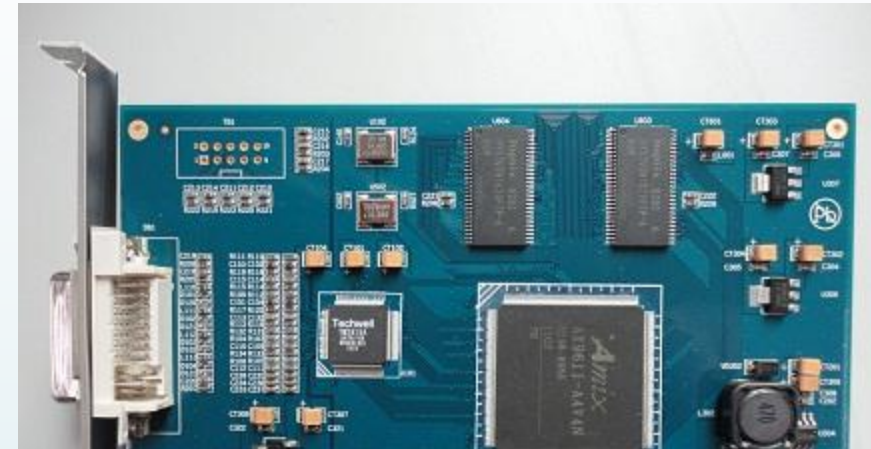
### ► 视频采集卡采集视频

视频采集卡（Video Capture Card）也叫视频捕捉卡；

视频采集卡可以接收来自视频输入端的模拟视频信号，对该信号进行采集并量化成数字信号，然后压缩编码成数字视频流。

### ► 摄像机获取数字视频

数字摄像机是获取数字视频的重要工具。镜头、CCD器件、数字信号处理（DSP）芯片、存储器和显示器件（LCD）等是数字摄像机的主要部件，尤其DSP芯片部分是数字摄像机的核心。



# 数字视频的获取与编辑

## ——视频质量评估

- ▶ 视频质量评估（Video Quality Assessment, VQA）：量化人所感知的视频质量
  - ▶ 主观视频质量评估
    - 1) 选择用于测试用的视频序列；
    - 2) 选择用于评价的测试系统；
    - 3) 选择一种测试方式；
    - 4) 邀请足够数量和类型的专家；
    - 5) 开展测试并计算每次测试的专家平均意见得分。

# 数字视频的获取与编辑

## ——视频质量评估

► 视频质量评估 (Video Quality Assessment, VQA) : 量化人所感知的视频质量

► 客观视频质量评估

根据是否提供原视频 (参考视频), 客观度量可以分为FR(Full Reference)、RR (Reduced Reference)和NR(No Reference)三种度量方式。

峰值信噪比 (Peak signal noise ratio, PSNR) ;

结构相似度 (Structure similarity Index, SSIM) ;

多尺度结构相似度 (Multi Scale Structural Similarity Index, MS-SSIM)

均方误差越小, 峰值信噪比越大, 视频的失真就越小。

# 数字视频的获取与编辑

## ——视频质量评估

峰值信噪比 (Peak signal noise ratio, PSNR) : 计算最大信号功率与均方差 (MSE) 之间的对数

$$PSNR = 10\log_{10} \frac{MaxValue^2}{MSE} = 10\log_{10} \frac{2^{bits} - 1}{MSE}$$

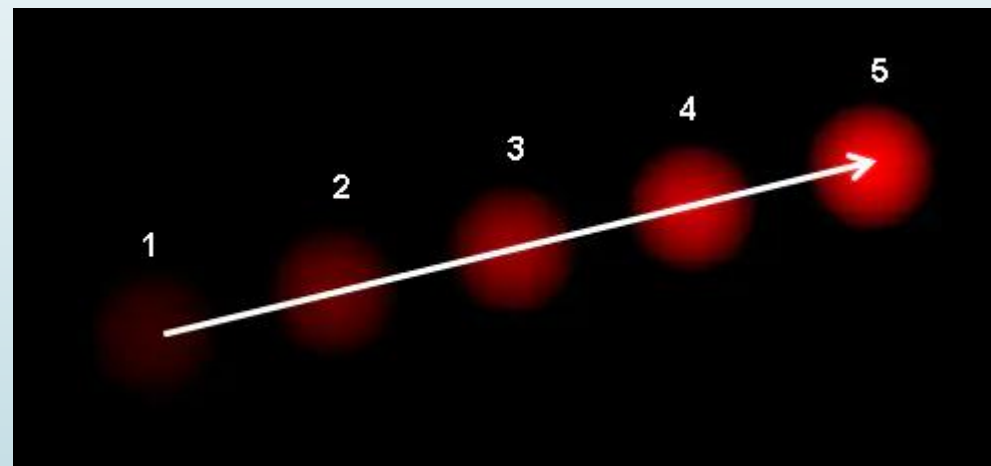
均方误差 (Mean square error, MSE)

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - Y(i, j))^2$$

# 视频运动估计

## ——光流

- 在人眼观察运动目标的过程中，目标的镜像在人眼视网膜上行程一系列连续变化的图像，这些图像所反应的光强变化。
- 光流是由对象或相机的移动引起的两个连续帧之间的图像对象的明显运动模式。是由场景中前景目标移动或相机的运动，单独或者共同产生的。
- 简单来说，所谓光流就是瞬时速率，在时间间隔很小（比如视频的连续前后两帧之间）时，也等同于目标点的位移。

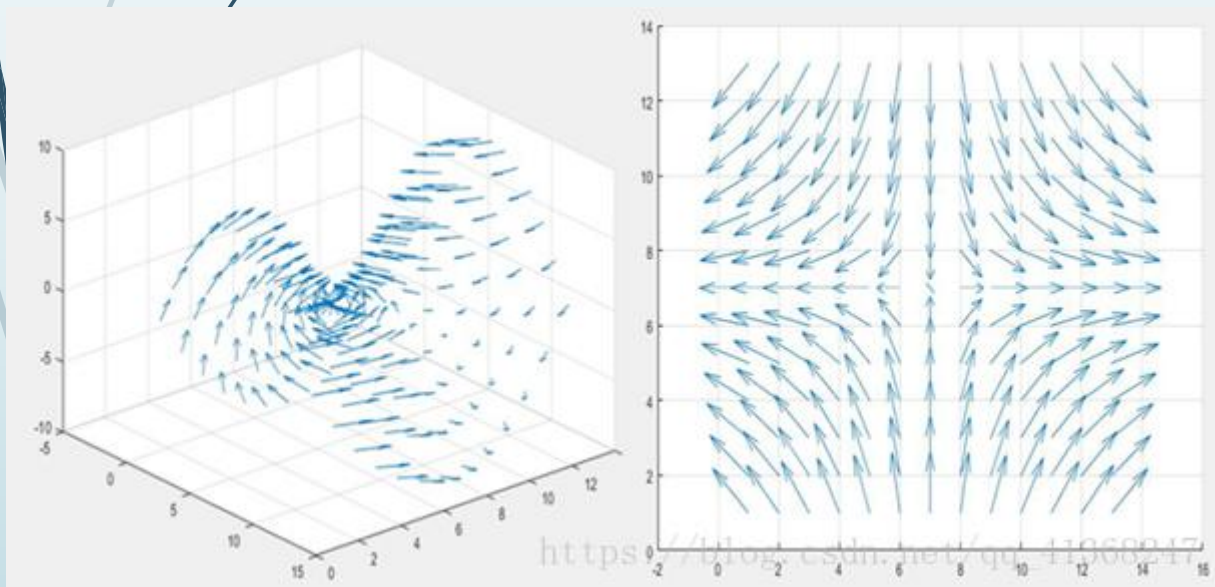




# 视频运动估计

## ——光流场

- 空间中的运动场转移到图像上就表示为光流场 (optical flow field) ;
- 光流场是一个二维矢量场, 它反映了图像上每一点灰度的变化趋势, 可看成是带有灰度的像素点在图像平面上运动而产生的瞬时速度场。





# 视频运动估计

## ——光流的计算

### ■ 光流计算依赖于三个假设：

- 亮度恒定：场景中目标图像的像素看起来在帧到帧移动是不发生改变。对于灰度图像（对于彩色图像同适用）这意味着像素的灰度值不会随着帧的跟踪改变。
- 时间持续性（微小移动）：图像上相机的移动随时间变化缓慢。实际上，这意味着时间的变化不会引起像素位置的剧烈变化，这样像素的灰度值才能对位置求对应的偏导数。
- 空间一致性：场景中相同表面的相邻点具有相似的运动，并且其投影到图像平面上的距离也比较近。

假设第一帧图像中的像素  $I(x, y, t)$  在时间  $dt$  后移动到第二帧图像的  $(x+dx, y+dy)$  处。根据上述第一条假设：灰度值不变，我们可以得到

$$I(x, y, t) = I(x + dx, y + dy, t + dt)$$

# 视频运动估计

## ——光流的计算

$$I(x, y, t) = I(x + dx, y + dy, t + dt)$$

对等号右边进行泰勒展开

$$I(x, y, t) = I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt + \varepsilon$$

$$\frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} \frac{dt}{dt} = 0$$

$$u = \frac{dx}{dt}, v = \frac{dy}{dt}$$

$$I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}, I_t = \frac{\partial I}{\partial t}$$

$$I_x u + I_y v + I_t = 0$$

# 视频运动估计

## ——光流的计算

按照理论基础与数学方法的区别把它们分成四种：基于梯度（微分）的方法、基于匹配的方法、基于能量（频率）的方法、基于相位的方法和神经动力学方法。

其中基于梯度的方法又称为微分法，利用时变图像灰度的时空微分（时空梯度函数）来计算像素的速度矢量典型的代表是**Lucas-Kanade**法，**L-K**光流法。

**K-L**光流法相关资料：

<https://www.cnblogs.com/gnuhpc/archive/2012/12/04/2802124.html>

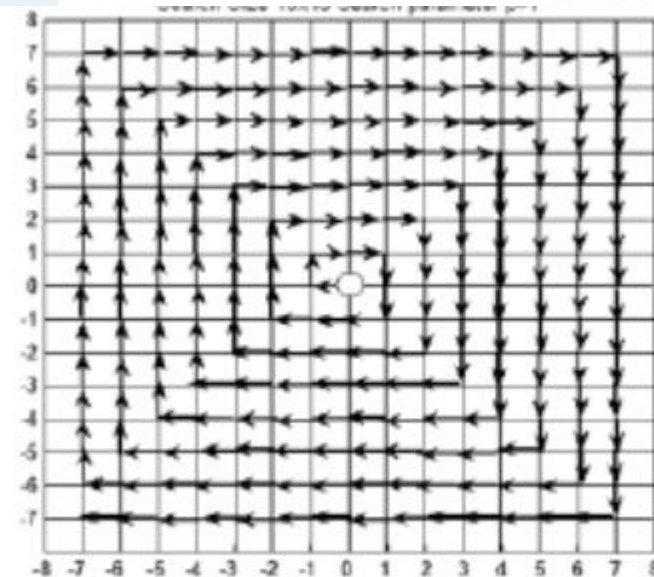
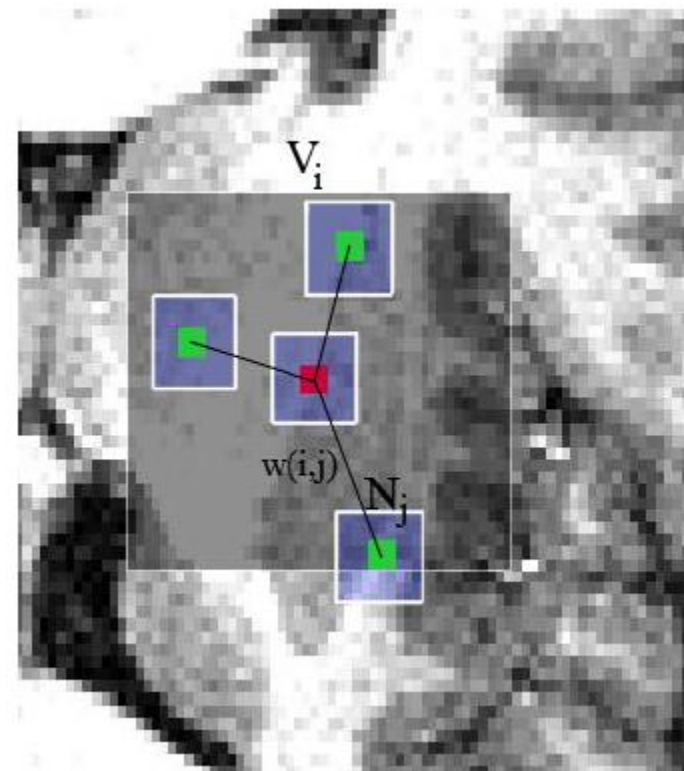
# 视频运动估计

## ——块匹配

### ► 块匹配

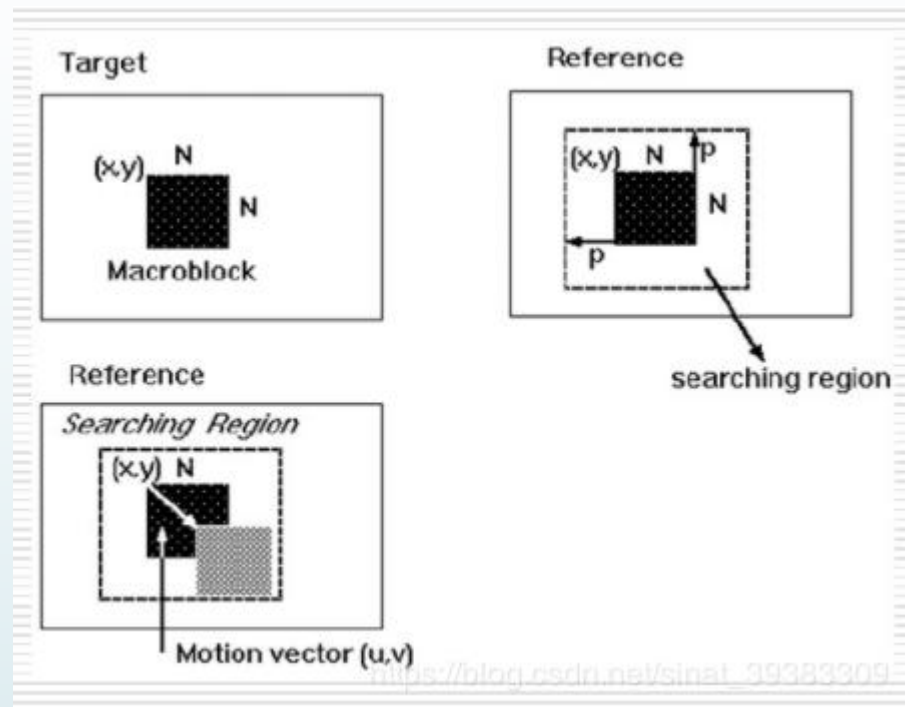
基于一些距离准则，在（过去和/或未来）参考帧中搜索与当前帧中固定大小块最匹配的相同尺寸块的位置。

块匹配是通过穷举搜索所有候选位移矢量而完成的。



# 视频运动估计

- 运动估计的基本思想是将图像序列的每一帧分成许多互不重叠的宏块，并认为宏块内所有像素的位移量都相同；
- 对每个宏块，在参考帧的某一给定搜索范围内，根据一定的匹配准则找出与当前块最相似的块，即匹配块；
- 匹配块与当前块的相对位移称为运动矢量（motion vector, MV）。



# 视频的概念

## ——视频文件类型

- AVI文件格式

AVI文件格式是Video For Windows所使用的文件格式，其扩展名为AVI。

- MOV 文件格式

MOV文件格式是QuickTime For Windows所使用的视频文件格式。

- MPG文件格式

MPG文件是最新的数字视频标准文件，也称为系统文件或隔行数据流，是采用MPEG方法进行压缩的全运动视频图像。

- DAT文件格式

DAT文件是VCD专用的视频文件格式，也是基于MPEG压缩/解压缩技术的视频文件格式。