

# 多媒体习题整理

## 多媒体习题整理

题型

判断

选择

填空

简答

计算题

综合论述题

页码标注为中文版的第二版

## 题型

单选 2 分/题, 10 题

简答 5 分/题, 2 题 (概念、对概念的理解)

计算 15 分/题, 4 题

综合论述 10 分/题, 1 题 (结合知识点, 展开分析和讨论)

## 判断

人的可见光范围 是400nm到800nm. (×) P53 (400nm~700nm)

视杆细胞和视锥细胞相比, 视锥细胞 主要用来检测灰度信息。(×) P54 (柱状细胞)

24 比特的彩色图, 则可以有一千六百多万的可能的颜色。(√) P40

在使用 颜色查找表 的图像中, 图像里存储的值是索引值, 需要到查找表中找到从头对应的图像像素值。(√) P42

多媒体数据包括:文本、图像、视频、音频。但动画不属于多媒体数据。(×) P12

有损压缩指的是压缩率大于1的压缩方式。(×) P122, 无损压缩也会大于1

对于无损压缩来说, 压缩率并没有限制 (×)

在由字母组成的序列中, 出现频率越高的字母, 越应该由更长的码字来表示。(×) 越短

算术编码 (Arithmetic Coding), 在理论上, 可以获得比霍夫曼(赫夫曼Huffman)更大的压缩率。(√) P137

多媒体数据的压缩方式常常是有损编码与无损编码相结合。(√)

## 选择

1. 以下有可能是哈夫曼编码的结果是 **D**

- A. H的编码为01，L的编码为00，E的编码为010
- B. H的编码为01，L的编码为00，E的编码为001
- C. H的编码为0，L的编码为01，E的编码为11
- D. H的编码为01，L的编码为00，E的编码为101

**解释：**任何一个赫夫曼编码都不能作为另一个赫夫曼编码的前缀。

2. 下列属于多媒体内容范畴的是 **B**

- A. 拼写信息
- B. 图形和动画
- C. 算法与逻辑
- D. 计算机程序

3. 算术编码方法中，产生的进制编码结果的数据范围是 (**P137**)

- A.  $[0, 1)$
- B.  $[-1, 1]$
- C.  $[0, 1]$
- D.  $[-1, 0]$

**答案：A**

4. 压缩率是一个衡量压缩方法好坏的重要标准，对于串行进行编码，对于最短码字母，更应该由\_\_的码字来表示。

- A. 高
- B. 低
- C. 尽量相同
- D. 短

**答案：A**

5. 以下不可能是赫夫曼编码的结果是

- A. H 的编码为 01，L 的编码为 00，E 的编码为 001

- B. H 的编码为 01, L 的编码为 10, E 的编码为 101
- C. H 的编码为 0, L 的编码为 1, E 的编码为 01
- D. 以上都是

答案: D

解释: 任何一个赫夫曼编码都不能作为另一个赫夫曼编码的前缀。

6. 以下的方法中, 哪种方法可能会使得原有信息无法完全恢复

- A. 非均匀量化
- B. 游长编码
- C. 算术编码
- D. 赫夫曼编码

答案: A (有损压缩)

7. 以下没有使用色度的二次采样的方法是

- A. 4:4:4
- B. 4:2:0
- C. 4:1:1
- D. 4:2:2

答案: A

8. 对于音频信号进行滤波以消除不需要的频率, 一般保留\_\_的信号。 (P100)

- A. 20dB-140dB
- B. 20dB-160dB
- C. 20Hz-20kHz
- D. 50Hz-10kHz

答案: D

9. 搜索运动向量的过程中, 哪一种方法的计算复杂度最低 (P219)

- A. 顺序搜索
- B. 对数搜索
- C. 分层搜索
- D. 以上都一样

答案: C

10. 使用一个4×4的打分矩阵, 可以将图像变为原来\_\_倍, 并可以表现\_\_级亮度分辨率。

- A. 16; 17
- B. 15; 16
- C. 16; 16
- D. 15; 15

答案：A

没找到书上哪里有，是DCT和量化的知识

11. 不会考虑熵编码的数据中，各个数据的概率分布的方法是

P152~P154

- A. 非均匀量化
- B. 均匀量化
- C. 香农-凡诺算法
- D. 赫夫曼编码

答案：B

## 填空

( P93 )

人类的听觉范围为： 20Hz~20kHz

人类声音的频率范围为： 500Hz~4kHz

## 简答

请说出R、G、B的颜色模型中三个分量分别代表什么颜色，R与G的叠加是什么颜色？请说出 L\*a\*b 模型的三个分量分别是什么，有什么物理含义？

P67

**R (Red)**： 代表红色。

**G (Green)**： 代表绿色。

**B (Blue)**： 代表蓝色。

红色 (R) 与绿色 (G) 叠加会产生黄色。

P65

**L\***: 亮度 (Lightness) , 表示颜色的明暗程度。

**a\***: 红绿平衡, 正值表示偏红, 负值表示偏绿。

**b\***: 黄蓝平衡, 正值表示偏黄, 负值表示偏蓝。

请说出 additive color model和 subtractive color model 分别适用在什么情况下?

P67

#### 1. 加性颜色模型 (Additive Color Model):

- **应用场景**: 通常用于显示设备, 如计算机显示器、电视屏幕和投影仪。
- **组成颜色**: 使用RGB (红、绿、蓝) 来组成颜色。

#### 2. 减性颜色模型 (Subtractive Color Model):

- **应用场景**: 通常用于印刷和绘画。
- **组成颜色**: 使用CMY (青、品红、黄) 和CMYK (青、品红、黄、黑) 来组成颜色。

请说出两种以上的图像格式, 以及其特点。

P45

**GIF**: 图形交换格式(Graphics Interchange Format, GIF)。GIF标准仅适用于8位(256)彩色图像。该标准支持隔行扫描通过套色(four-pass)显示方法处理, 相隔的像素可以连续显示。

P47

**JPEG**: JPEG 利用了人类视觉系统一些特定的局限性, 从而获得很高的压缩率。这种压缩方式是有损的。

P48

**PNG**: 便携式网络图形(Portable Network Graphics, PNG), PNG 的独特之处是最多可支持48 位的色彩信息。

简述什么是奈奎斯特定理?

P94

为了得到正确的采样, 我们需要使用的采样频率至少是信号中最高频率的两倍。

简述频率遮掩和时间遮掩, 以及它们在音频编码中的应用?

P314

P318

**频率遮掩：**是指当两个声音的频率接近时，较强的声音会掩盖较弱的声音，使得人耳难以察觉到较弱的声音。

**时间遮掩：**是指当两个声音在时间上接近时，较强的声音会掩盖较弱的声音。这种现象分为前掩蔽和后掩蔽（。前掩蔽是指较强的声音在出现前一小段时间内会掩盖较弱的声音，后掩蔽是指较强的声音在消失后的一段时间内会掩盖较弱的声音。

**应用：**

有损的音频数据压缩算法(比如MPEG 声音或者在电影制作中常用的杜比数字(AC-3)编码)都去掉了某些被遮掩的声音，因此减少了总的信息量。

MPEG音频压缩利用它们建立一个庞大的多维查找表，通过这个表记录被频率遮掩或者时间遮掩的频率分量，以达到压缩的目的。

**简述韦伯定律的定义，和其在什么信号的编码的过程中发挥了作用，发挥了什么作用？**

P97

**定义：**声音自身的强度越大，就越需要更大的振幅来让我们感受到声音的变化。韦伯定律指出，要产生同样的感知所需要的增幅是和原来的绝对值是成比例的。

图中的公式表示：

$$\Delta \text{Response} \propto \frac{\Delta \text{Stimulus}}{\text{Stimulus}}$$

**作用：**韦伯定律利用这一感知特性设计了非均匀量化方案，并产生了 $\mu$ 律和A律编码方案，使得信噪比在输入信号范围内分布更均匀。

**简述为什么音频信号采用差分编码，及其优点？**

P111

**原因：**如果一个时间相关的信号随时间变化存在一些时间冗余，那么把当前信号值和前一个信号值相减产生的差分值得在分布图上会产生一些峰值，0附近会有最大值。因此，给差分值分配码字时，可以给出出现频率高的值分配较短的码字，给出出现频率低的值分配较长的码字。

**优点：**差分值比原始信号值小，可以使用更小的空间存储。差分信号的分布图比原始信号的分布图更加集中。

**什么是感兴趣区域编码，其实现方法的基本思想是什么？**

新的JPEG 2000 标准的一个重要特性就是可以实现感兴趣区域(ROI)编码[8]。这样，相对于图像的背景或其他部分来说，某些部分可以采取高质量的编码。一个称为MAXSHIFT 的方法，是一种可伸缩的方法，可以将ROI内的系数增加到更高层的位平面中。在这种嵌入式的编码过程中，所处理的比特将被放置到图像的非ROI部分前面。所以，给定一个缩小了的码率，ROI将在图像的其他部分之前解码和改善。由于以上这些机制的作用,ROI将会有着比背景图案更高的质量。

调研HEIF或TPG格式，与JPEG格式作对比。

特性	HEIF	TPG	JPEG
压缩效率	高，节省约50%文件大小	高，比JPEG高	较低
图像质量	高，较少压缩伪影	高，保持图像质量	一般，有压缩伪影
多功能性	支持单图像、多图像、序列、元数据	主要支持静态图像	仅支持单图像
兼容性	较差，需新硬件和软件支持	较差，推广和支持有限	高，广泛支持
解码复杂性	高，需更多处理能力	较低，设计简洁	低，处理速度快
支持动画	是	否	否

HEIF和TPG格式在压缩效率和图像质量上显著优于传统的JPEG格式，但它们的兼容性和市场普及度目前还不及JPEG。HEIF由于其多功能性和高质量，正在逐步获得更多支持，特别是在苹果设备上。TPG作为新兴格式，需要时间来推广和获得广泛支持。

为了使打印变得简单,我们设计了一个装有 CMY 传感器的相机,请问这个 CMY相机的输出能用来生成普通的 RGB 图像吗?为什么?

是的，装有 CMY（青色、品红色、黄色）传感器的相机输出可以用来生成普通的 RGB（红色、绿色、蓝色）图像。这是因为 CMY 和 RGB 是两种互补的颜色模型，且可以通过色彩转换来相互转换。

由于 CMY 和 RGB 之间是互补关系，可以通过以下公式进行转换：

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4.24}$$

其逆变换是

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} C \\ M \\ Y \end{bmatrix} \tag{4.25}$$

假设对一幅彩色图像的每一个像素，我们有24位可以利用。然而，我们发现，比起蓝色来，人们对红色和绿色更敏感。实际上，人们对红色和绿色的敏感度是蓝色的1.5倍。我们怎样才能最好地使用可以利用的位？

P54

红色和绿色的比特数： $1.5/4 \times 24 \approx 9.15/4 \times 24 \approx 9.15/4 \times 24 \approx 9$ 位

蓝色的比特数： $1/4 \times 24 \approx 6.1/4 \times 24 \approx 6.1/4 \times 24 \approx 6$ 位

普通的 TV与 HDTV 最显著的区别是什么？

P82

传统TV和HDTV之间的显著区别就是后者有更大的宽高比16:9而不是4:3

电视机与显示器通常各使用什么扫描方式?并解释这些扫描方式。

P75

电视机和显示器使用的扫描方式主要分为**隔行扫描**和**逐行扫描**。

**隔行扫描(inter-laced scanning)**。隔行扫描先扫描奇数行，然后扫描偶数行。这样就产生“奇数场”和“偶数场”，两个场组成一个帧。

**逐行扫描**，所谓逐行扫描(progressive scanning)就是按照一定的时间间隔逐行扫描完整的图像(帧)。

请分别简单解释模拟信号与数字信号?并说明它们之间最主要的区别是什么？

P75

模拟信号是一种连续变化的信号，其幅度可以在一定范围内取任意值。它通常用来表示物理量的变化，例如声音、温度、光强度等。

数字信号是一种离散的信号，其幅度只能在有限的、离散的值之间变化。数字信号通常用于计算机和数字电子设备中。

**连续性 vs 离散性：**

- 模拟信号是连续的，可以在任意时间点取任意值。
- 数字信号是离散的，只能在特定时间点取有限的值。



# 计算题

8位灰度图中有一小块图片的像素值如下:

255 200 180 144

112 78 20 0

请分别给出其最高的 Biplane(位平面)Plane7, 和最低的 Bitplane(位平面)Plane 0 的值。

P38

首先, 我们将每个像素值转换为 8 位二进制数:

- 255: 11111111
- 200: 11001000
- 180: 10110100
- 144: 10010000
- 112: 01110000
- 78: 01001110
- 20: 00010100
- 0: 00000000

Bitplane 7 (最高位平面) 包含每个像素值的第 7 位 (从右数):

```
1 1 1 1
0 0 0 0
```

Bitplane 0 (最低位平面) 包含每个像素值的第 0 位 (从右数):

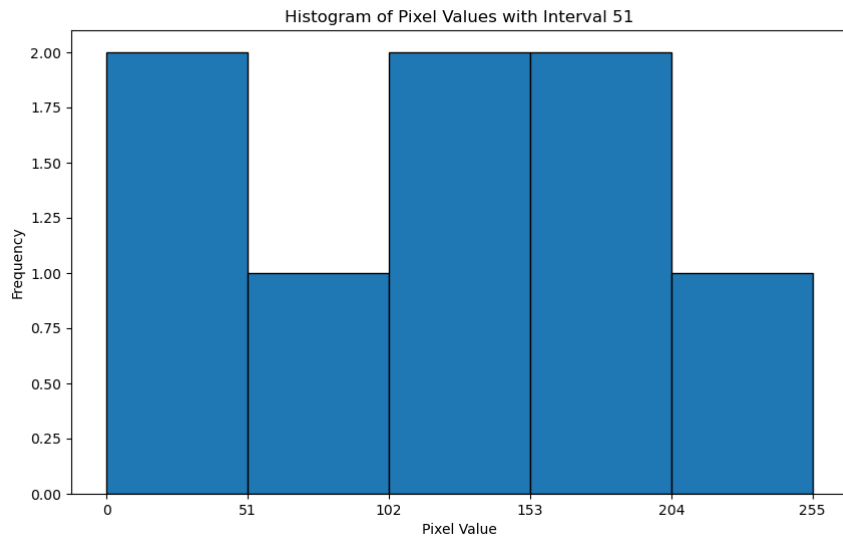
```
1 0 0 0
0 0 0 0
```

8位灰度图中有一小块图片的像素值如下:

255 200 180 144

112 78 20 0

请以51为间隔, 画出这组像素值的直方图(Histogram)表示。



现使用像素值128 作为國值，将该小块图片转化为二进制图(Binary Image) 请给出该二进制图的结果。

对于每个像素值，如果其值大于或等于 128，则设置为 1（白色）。

如果其值小于 128，则设置为 0（黑色）。

结果为：

```
1 1 1 1
0 0 0 0
```

在已知亮度量化表和色度量化表，在Y通道的DCT变换系数如下时，量化后的系数值为多少?zigzag编码之后结果为多少?如果要对其AC系数进行RLC 编码结果是什么？

640	35	44	7	5	1	-7	3
26	-3	1	5	-4	-5	3	2
-15	2	-4	3	4	3	1	-4
5	-4	4	1	-3	2	-3	0
3	5	2	-4	2	-3	-3	0
1	-4	-3	-7	-2	0	1	0
-2	6	1	1	0	0	0	0
5	-3	0	3	1	1	0	0

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

The Luminance Quantization Table

The Chrominance Quantization Table

量化过程是将每个 DCT 系数除以相应位置的量化表值并四舍五入。Y通道为亮度图

通过对 DCT 系数进行量化，我们得到了如下矩阵： [P191](#)

40	3	4	0	0	0	0	0
2	0	0	0	0	0	0	0
-1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Zigzag 编码后的线性化结果为： [P194](#)

(40, 3, 2, -1, 0, 4, 0, ..., 0)

对 AC 系数进行游程长度编码 (RLC) 后的结果为：

(0,3)(0,2)(0,-1)(1,4)(0,0)

请计算该字母表的熵(entropy)，请写出计算过程。字母表S={a, b, c, d}出现概率P=(1/8, 1/8, 1/4, 1/2)。

[P123](#)

$$\begin{aligned}
 \eta = H(s) &= - \sum_{i=1}^n P_i \log_2 P_i = - \left( \frac{1}{8} \log_2 \frac{1}{8} + \frac{1}{8} \log_2 \frac{1}{8} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{2} \log_2 \frac{1}{2} \right) \\
 &= - \left( \frac{1}{8} \times (-3) + \frac{1}{8} \times (-3) + \frac{1}{4} \times (-2) + \frac{1}{2} \times (-1) \right) \\
 &= 1.75
 \end{aligned}$$

已知原来的数据为{8 16 16 12 12 8 16 8}，经过压缩与解压缩之后得到的数据为{8 12 12 12 12 8 12 12}。请分别计算MSE、SNR和PSNR的值，请写出计算过程。

P151

**MSE:**  $\sigma^2 = \frac{1}{N} \sum_{n=1}^N (x_n - y_n)^2 = \frac{1}{8} [0 + 4^2 + 4^2 + 0^2 + 0^2 + 0^2 + 4^2 + (-4)^2] = 8$

**SNR** :

$$SNR = 10 \log_{10} \frac{\sigma_x^2}{\sigma_d^2} = 10 \log_{10} \frac{(8^2+16^2+16^2+12^2+12^2+8^2+16^2+8^2)/8}{8^2} = 10 \log_{10} \frac{1248/8}{8} = 10 \log_{10}(19.5) \approx 12.90dB$$

**PSNR:**

最大值 MAX=16    MSE = 8:

$$PSNR = 10 \log_{10} \frac{x_{peak}^2}{\sigma_d^2} = 10 \log_{10} \frac{16^2}{8} \approx 15.05dB$$

已知字母表以及对应的概率值如下，请使用Huffman编码的方法，写出字母序列"FCBEE\$"的编码结果，请写出编码过程

Symbol	Probability
A	0.2
B	0.1
C	0.2
D	0.05
E	0.3
F	0.05
\$	0.1

P127

每个字母表示一个节点，按照概率值从小到大排序：

D (0.05), F (0.05), B (0.1), \$ (0.1), A (0.2), C (0.2), E (0.3)

构建Huffman树:



根据树的结构，分配编码：

字母序列“FCBEE\$”的编码结果是：

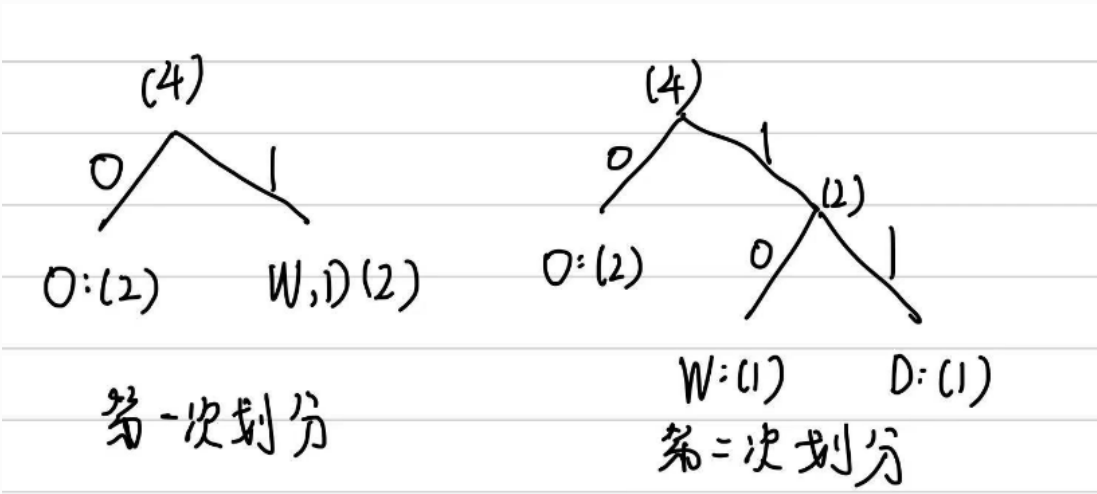
1101 01 1110 10 10 1111

使用香农-范诺算法对单词 WOOD 进行编码，并写出编码步骤。

符号	W	O	D
数目	1	2	1

P125

- 1) 根据每个符号出现的频率对符号进行排序，排序结果为 O、W、D
- 2) 递归地将这些符号分成两部分。第一次划分，出现次数为2的O，记为 O:(2)；出现总次数为2的 W、D，记录为 W、D: (3)。第二次划分产生 W:(1) 和 D:(1)



请使用算术编码方法对单词 WOOD进行编码(写出最后十进制的编码范围即可)，并写出编码步骤。

符号	W	O	D
数目	1	2	1

P138

信号	可能性	范围	高值	低值
D	0.25	[0,0.25)	0	0.25
O	0.5	[0.25,0.75)	0.25	0.75
W	0.25	[0.75,1.0)	0.75	1

请对如下字符流a b a b c b a b a b a a a a a a (注意，字符流中不含空格)进行LZW编码。请写出此 Lzw的编码过程(词典)，并给出输出码字流。(建议用列表的方式来辅助答题)(20分)

P133

s	c	output	code	string
			1	a
			2	b
			3	c
a	b	1	4	ab
b	a	2	5	ba
a	b			
ab	c	4	6	abc
c	b	3	7	cb
b	a			
ba	b	5	8	bab
b	a			
ba	b			
bab	a	8	9	baba
a	a	1	10	aa
a	a			
aa	a	10	11	aaa
a	a			
aa	a			
aaa	a	11	12	aaaa

根据上面的编码过程，输出的码字流是：

1 2 4 3 5 8 1 10 11

一张640x480的真彩色图像，在YCbCr空间采用4:2:0的色度次采样，采样后Y, Cb, Cr 三个颜色通道的分辨率分别是多少？

P80

在YCbCr色彩空间中，4:2:0的色度次采样意味着色度通道（Cb和Cr）的水平和垂直方向的分辨率都被减半。具体来说，对于一张640x480的真彩色图像，Y、Cb、Cr三个颜色通道的分辨率分别为：

1. **Y通道**：保持原始分辨率

- 宽度：640
- 高度：480
- 分辨率：640 x 480

2. **Cb通道**：水平和垂直方向的分辨率都减半

- 宽度： $640 / 2 = 320$
- 高度： $480 / 2 = 240$
- 分辨率：320 x 240

3. **Cr通道**：水平和垂直方向的分辨率都减半

- 宽度： $640 / 2 = 320$
- 高度： $480 / 2 = 240$
- 分辨率：320 x 240

因此，采样后Y, Cb, Cr 三个颜色通道的分辨率分别是：

- Y通道：640 x 480
- Cb通道：320 x 240
- Cr通道：320 x 240

一幅 YUV 彩色图像的分辨率为 720x576 像素。分别计算采用4:2:2、 4:1:1 和 4:2:0 子采样格式采样时的样本数。

1. **YUV 4:2:2 子采样格式**

在4:2:2格式下：

- Y 分量的采样率为 4，表示每个像素都有一个 Y 样本。
- U 和 V 分量的采样率为 2，表示每两个像素共享一个 U 和一个 V 样本。

对于 720x576 的图像：

- Y 分量的样本数： $720 \times 576 = 414,720$  个。
- U 分量的样本数： $\frac{720}{2} \times 576 = 360 \times 576 = 207,360$  个。
- V 分量的样本数： $\frac{720}{2} \times 576 = 360 \times 576 = 207,360$  个。

总样本数为：

$$[ 414,720 + 207,360 + 207,360 = 829,440 ]$$

## 2. YUV 4:1:1 子采样格式

在4:1:1格式下：

- Y 分量的采样率为 4，表示每个像素都有一个 Y 样本。
- U 和 V 分量的采样率为 1，表示每四个像素共享一个 U 和一个 V 样本。

对于 720x576 的图像：

- Y 分量的样本数： $720 \times 576 = 414,720$ 个。
- U 分量的样本数： $\frac{720}{4} \times 576 = 180 \times 576 = 103,680$ 个。
- V 分量的样本数： $\frac{720}{4} \times 576 = 180 \times 576 = 103,680$ 个。

总样本数为：

$$[ 414,720 + 103,680 + 103,680 = 622,080 ]$$

## 3. YUV 4:2:0 子采样格式

在4:2:0格式下：

- Y 分量的采样率为 4，表示每个像素都有一个 Y 样本。
- U 和 V 分量的采样率为 2 水平和 2 垂直方向，表示每四个像素共享一个 U 和一个 V 样本。

对于 720x576 的图像：

- Y 分量的样本数： $720 \times 576 = 414,720$  个。
- U 分量的样本数： $\frac{720}{2} \times \frac{576}{2} = 360 \times 288 = 103,680$  个。
- V 分量的样本数： $\frac{720}{2} \times \frac{576}{2} = 360 \times 288 = 103,680$  个。

总样本数为：

$$[ 414,720 + 103,680 + 103,680 = 622,080 ]$$

如果在声音信号电压为10V，噪声电压为0.1V，则信噪比(SNR)为？

P95

$$SNR = 10 \log_{10} \frac{V_{\text{信号}}^2}{V_{\text{噪声}}^2} = 20 \log_{10} \frac{V_{\text{信号}}}{V_{\text{噪声}}} = 20 \log_{10} \frac{10V}{0.1V} = 40db$$



基于参考书公式6.19-6.20对差分脉冲调制编码DPCM的介绍，量化初始误差为0，写出如下信号的DPCM编码过程

f: 140 150 130 180 190

( P114~116 )

$f$	140	150	130	180	190
$\hat{f}$	140	140	$\frac{140+148}{2} = 144$	$\frac{148+136}{2} = 142$	$\frac{136+182}{2} = 159$
$e$	0	$150 - 140 = 10$	$130 - 144 = -14$	$180 - 142 = 38$	$190 - 159 = 31$
$\tilde{e}$	0	$16 \lfloor \frac{255+10}{16} \rfloor - 248 = 8$	$16 \lfloor \frac{255-14}{16} \rfloor - 248 = -8$	$16 \lfloor \frac{255+38}{16} \rfloor - 248 = 40$	$16 \lfloor \frac{255+31}{16} \rfloor - 248 = 24$
$\tilde{f}$	140	$140 + 8 = 148$	$144 - 8 = 136$	$142 + 40 = 182$	$159 + 24 = 183$

$$\hat{f}_n = \left\lfloor \frac{\tilde{f}_{n-1} + \tilde{f}_{n-2}}{2} \right\rfloor \quad e_n = f_n - \hat{f}_n \rightarrow \text{注意中文第二版这里有问题}$$

$$\tilde{e}_n = 16 \times \left\lfloor \frac{255 + e_n}{16} \right\rfloor - 256 + 8$$

$$\tilde{f}_n = \hat{f}_n + \tilde{e}_n$$

基于参考书公式6.21对增量调制DM的介绍，量化初始误差为0，步长k=4，写出如下信号的DM编码， f:14 15 13 18 19

P116

$$\hat{f}_n = \tilde{f}_{n-1}$$

$$e_n = f_n - \hat{f}_n = f_n - \tilde{f}_{n-1}$$

$$\tilde{e}_n = \begin{cases} +k & \text{如果 } e_n > 0, \text{ 其中 } k \text{ 是常数} \\ -k & e_n \leq 0 \end{cases}$$

$$\tilde{f}_n = \hat{f}_n + \tilde{e}_n$$

	1	2	3	4	5
f	14	15	13	18	19
f	14	14	18	14	18
e	0	15-14=1	13-18=-5	18-14=4	19-18=1
e	0	4	-4	4	4
f	14	14+4=18	18-4=14	14+4=18	18+4=22

学习EZW算法, 分析如下小波系数的扫描结果, 算法初始值T0=32, 给出第一次主扫描结果

57	-37	39	-20	3	7	9	10
-29	30	17	33	8	2	1	6
14	6	15	13	9	-4	2	3
10	19	-7	9	-7	14	12	-9
12	15	33	20	-2	3	1	0
0	7	2	4	4	-1	1	1
4	1	10	3	2	0	1	0
5	6	0	0	3	1	2	1

P183

直接抄书吧, 没太看懂, 估计也不会考

什么情况下会出现假频? 如真实频率为22.05kHz, 采样频率33.075kHz, 则假频为多少?

P94

如果采样频率小于实际信号频率的两倍, 但是大于实际信号频率, 则假频等于采样频率减去真实频率。

$$22.05 < 33.075 < 22.05 * 2$$

$$f_{\text{alias}} = f_{\text{sampling}} - f_{\text{true}} = 33.075 - 22.05 = 11.025 \text{ kHz}$$

某电脑上有一块 16 位的声卡, 这里的 16 位是指什么? 其信号量化噪声比 SQNR 是多少?

在数字信号处理中，"16 位声卡"指的是其模数转换器（ADC，Analog-to-Digital Converter）的位数。这表示声卡能够将模拟声音信号转换为数字信号时，使用的是 16 位的量化器。对于 16 位的ADC，它能够将模拟信号分成  $2^{16} = 65536$  个离散级别。

信号量化噪声比（SQNR）是描述信号在量化过程中引入的噪声的指标，它与ADC的位数相关。对于一个 (N) 位的ADC，其理想的信号量化噪声比（SQNR）可以通过以下公式估算：

$$\text{SQNR} \approx 6.02 \times N + 1.76 \text{ dB}$$

对于 16 位的声卡：

$$\text{SQNR} \approx 6.02 \times 16 + 1.76 \approx 98.08 \text{ dB}$$

这意味着，在理想情况下，一个 16 位的ADC能够提供约 98.08dB 的信号量化噪声比。这个值表示了信号与量化误差之间的差异，越高的SQNR表示量化误差对原始信号的影响越小，ADC的性能越好。

输入信号为  $f(i)=[0 \ 10 \ 20 \ 30 \ 40 \ 50 \ 60 \ 70]$ ，请计算该信号的一维离散余弦变换的  $F(0)$ 。

#### P157

计算 (F(0)) 时的归一化因子  $\alpha(0)$ ：

$$\alpha(0) = \sqrt{\frac{1}{N}} = \sqrt{\frac{1}{8}} = \frac{1}{\sqrt{8}} = \frac{1}{2\sqrt{2}}$$

现在计算 (F(0))：

$$F(0) = \alpha(0) \sum_{i=0}^{N-1} f(i) \cos\left(\frac{\pi \cdot 0 \cdot (2i+1)}{2 \cdot 8}\right)$$

由于 ( $k=0$ )，余弦项  $\cos\left(\frac{\pi \cdot 0 \cdot (2i+1)}{2 \cdot 8}\right) = \cos(0) = 1$ ，所以公式简化为：

$$F(0) = \alpha(0) \sum_{i=0}^{N-1} f(i)$$

计算信号的和：

$$\sum_{i=0}^{N-1} f(i) = 0 + 10 + 20 + 30 + 40 + 50 + 60 + 70 = 280$$

所以：

$$F(0) = \frac{1}{2\sqrt{2}} \times 280$$

进一步简化：

$$F(0) = \frac{280}{2\sqrt{2}} = \frac{280}{2 \times 1.414} = \frac{280}{2.828} \approx 99.1$$

因此，输入信号 (  $f(i) = [0, 10, 20, 30, 40, 50, 60, 70]$  ) 的一维离散余弦变换的 (  $F(0)$  ) 值约为 99.1。

给定6个图像块的DC系数为130, 135, 141, 180, 182, 179，求其赫夫曼编码（按照书上的码表）。

P195

码表

L	项	二进制编码
0	0	- -
1	-1, 1	0, 1
2	-3, -2, 2, 3	00, 01, 10, 11
3	-7, -6, -5, -4, 4, 5, 6, 7	000, 001, 010, 011, 100, 101, 110, 111
4	-15, -14, ..., -8, 8, ..., -14, 15	0000, 0001, ..., 0111, 1000, ..., 1110, 1111
5	-31, -30, ..., -16, 16, ..., 30, 31	00000, 00001, ..., 01111, 10000, ..., 11110, 11111
6	-63, -62, ..., -32, 32, ..., 62, 63	000000, 000001, ..., 011111, 100000, ..., 111110, 111111
⋮	⋮	⋮

计算给定6个图像块的DC系数（130, 135, 141, 180, 182, 179）的霍夫曼编码，我们需要进行以下步骤：

1. **计算相邻DC系数的差值：** JPEG编码中，DC系数是相邻块之间的差值，因此我们首先需要计算这些差值。
2. **确定差值的位长度 (bitlen)：** 差值的位长度表示为 `bitlen`，是可以表示该差值所需的最小位数。
3. **查找霍夫曼编码：** 使用提供的亮度DC霍夫曼码表，找到 `bitlen` 对应的霍夫曼编码。
4. **计算额外位：** 根据差值的正负，计算其二进制表示（或补码表示）。
5. **组合编码：** 将霍夫曼编码和额外位组合在一起，得到最终的编码。

1. 计算相邻DC系数的差值

假设第一个块的DC系数为 `DC0 = 130`，那么差值如下：

```
DC0 = 130
Diff1 = DC1 - DC0 = 135 - 130 = 5
Diff2 = DC2 - DC1 = 141 - 135 = 6
Diff3 = DC3 - DC2 = 180 - 141 = 39
Diff4 = DC4 - DC3 = 182 - 180 = 2
Diff5 = DC5 - DC4 = 179 - 182 = -3
```

2. 确定差值的位长度 (bitlen)

```
bitlen(130) = 8 (10000010)
bitlen(5) = 3 (101)
bitlen(6) = 3 (110)
bitlen(39) = 6 (100111)
bitlen(2) = 2 (10)
bitlen(-3) = 2 (反码表示 00)
```

### 3. 查找霍夫曼编码

从提供的亮度DC霍夫曼码表中查找 `bitlen` 对应的霍夫曼编码：

```
bitlen=8 -> 霍夫曼编码 = 1000
bitlen=3 -> 霍夫曼编码 = 11
bitlen=6 -> 霍夫曼编码 = 110
bitlen=2 -> 霍夫曼编码 = 10
```

### 4. 计算额外位

计算额外位，即差值的二进制表示或补码表示：

```
DC0 = 130 -> 额外位 = 10000010
Diff1 = 5 -> 额外位 = 101
Diff2 = 6 -> 额外位 = 110
Diff3 = 39 -> 额外位 = 100111
Diff4 = 2 -> 额外位 = 10
Diff5 = -3 -> 额外位 = 00（反码表示）
```

### 5. 组合编码

将霍夫曼编码和额外位组合在一起：

```
DC0 = 130 -> 霍夫曼编码：1000 -> 最终编码：1000 10000010
Diff1 -> 霍夫曼编码：11 -> 最终编码：11 101
Diff2 -> 霍夫曼编码：11 -> 最终编码：11 110
Diff3 -> 霍夫曼编码：110 -> 最终编码：110 100111
Diff4 -> 霍夫曼编码：10 -> 最终编码：10 10
Diff5 -> 霍夫曼编码：10 -> 最终编码：10 00
```

选择采样速率为 22.05kHz 和量化位数为16 位的录音参数,在不用压缩技术的情况下计算录制2分钟的立体声需要多少MB字节的存储空间？

P100

为了计算录制2分钟的立体声音频文件需要多少MB的存储空间，可以按照以下步骤进行计算：

1. **采样速率**（Sampling Rate）：22.05 kHz，即每秒22050个样本。
2. **量化位数**（Bit Depth）：16位，即每个样本用16位表示。
3. **声道数**（Channels）：立体声有2个声道。
4. **录制时间**：2分钟，即120秒。

1. **每秒的样本数（采样速率）**：  
( 22050 ) 样本/秒

2. 每个样本的大小（量化位数）：

(16) 位 = (2) 字节 (因为1字节 = 8位)

3. 每秒的字节数：

(22050) 样本/秒 × (2) 字节/样本 = (44100) 字节/秒 (单声道)

4. 每秒的总字节数（立体声）：

(44100) 字节/秒 × (2) = (88200) 字节/秒

5. 总的字节数（2分钟）：

(88200) 字节/秒 × (120) 秒 = (10584000) 字节

6. 换算成MB：

(10584000) 字节 ÷ (1024) = (10335.9375) KB

(10335.9375) KB ÷ (1024) = (10.09375) MB

录制2分钟的立体声音频（采样速率为22.05kHz，量化位数为16位）需要大约 **10.09 MB** 的存储空间。

假设位的深度为 12 位、120fps、4:2:2 的色度二次采样方案无压缩的 4K UHDTV 视频的比特率是多少？。

1. 每个像素的Y分量位数：

每个像素有12位的Y分量。

2. 每个像素的U和V分量位数：

在4:2:2方案中，每2个像素共享一个12位的U和一个12位的V分量。因此每个像素有6位的U和6位的V分量。

3. 每个像素的总位数：

Y: 12位

U: 6位

V: 6位

总共：(12 + 6 + 6 = 24) 位

4. 每帧的总位数： **P82**

分辨率为 3840 × 2160 像素。

每帧的总位数 = 分辨率 × 每个像素的总位数

= 3840 × 2160 × 24

= (199065600) 位

5. 每秒的总位数：

帧率为 120fps。

每秒的总位数 = 每帧的总位数 × 帧率

= (199065600) 位 × 120

= (23887872000) 位

#### 6. 比特率 (bps):

1字节 = 8位。

所以每秒的比特率 (bps) = (23887872000) 位 ÷ 8  
= (2985984000) 字节  
= (2.781) GB/s

**NTSC 制式视频**，已知视频图像的宽为 400 像素，真彩色，视频时长5分钟，在不考虑视频伴音与压缩的情况下，求原始视频的文件大小

真彩色通常指的是24位色深，即每个像素由3个字节表示（RGB各8位）。 [P40](#)

NTSC制式采用4:3，帧率为 29.97 帧每秒 (fps)。 [P77](#)

$$((400 * 300) * 3 * 29.97 * (5 * 60)) / (1024 * 1024) = 3086.8\text{MB}$$

假设一个5位的灰度图像，我们需要多大的抖动矩阵来在1位的打印机上显示这幅图像？

[P39](#)

为了在1位打印机上显示5位灰度图像，需要使用适当大小的抖动矩阵进行抖动处理。以下是详细解释：

- 灰度图像的位深度：**5位灰度图像意味着每个像素可以表示 ( $2^5 = 32$ ) 种灰度级别，从0到31。
- 1位打印机的限制：**1位打印机只能显示两种颜色（通常是黑色和白色），所以我们需要将灰度图像转换成只有黑白两色的图像。
- 抖动处理：**抖动是一种将灰度图像转换为黑白图像的技术，通过在局部区域内控制黑白像素的密度来模拟灰度级别。
- 抖动矩阵的大小：**
  - 抖动矩阵大小的选择需要考虑能够表示的灰度级别。对于5位灰度图像，我们有32个灰度级别。
  - 一个  $n \times n$  的抖动矩阵可以表示  $n^2 + 1$  个灰度级别。因此，为了表示32个灰度级别，抖动矩阵的大小 (n) 需要满足  $n^2 \geq 31$ 。
  - 计算得到最小的 (n) 为6，因为  $6^2 = 36$  刚好大于31。

因此，需要一个至少为6x6大小的抖动矩阵才能在1位打印机上显示5位灰度图像。

假设有一幅 240 x 180 x 8bit 的灰度图，当用分辨率为 300dpi x300dpi 的激光打印机将其打印到12.8x9.6英寸的纸上时.每个像素的图案有多大？

打印机分辨率: dpi (dots per inch) 每英寸300点

打印纸的宽度像素数:  $12.8 \text{英寸} * 300 \text{dpi} = 3840 \text{像素}$

打印纸的高度像素数:  $9.6 \text{英寸} * 300 \text{dpi} = 2880 \text{像素}$

每个像素在打印纸上的宽度:  $3840 \text{像素} / 240 \text{像素} = 16 \text{打印像素}$

每个像素在打印纸上的高度:  $2880 \text{像素} / 180 \text{像素} = 16 \text{打印像素}$

每个打印像素的实际尺寸: 1英寸/300dpi

每个图像像素在纸上的实际尺寸:  $16 * (1/300)$

## 综合论述题

1.在 Yanny 和 aurel的声音信号中,为何有人只能听到 Yanny,为何有人只能听到 Laure,你猜测是什么原因,这属于我们通常说的什么现象?

频率遮掩现象。“Yanny”主要在较高的频率范围内,而“Laurel”主要在较低频率范围内。不同的人对高频和低频声音的敏感度不同,这取决于他们的听觉系统和年龄等因素。

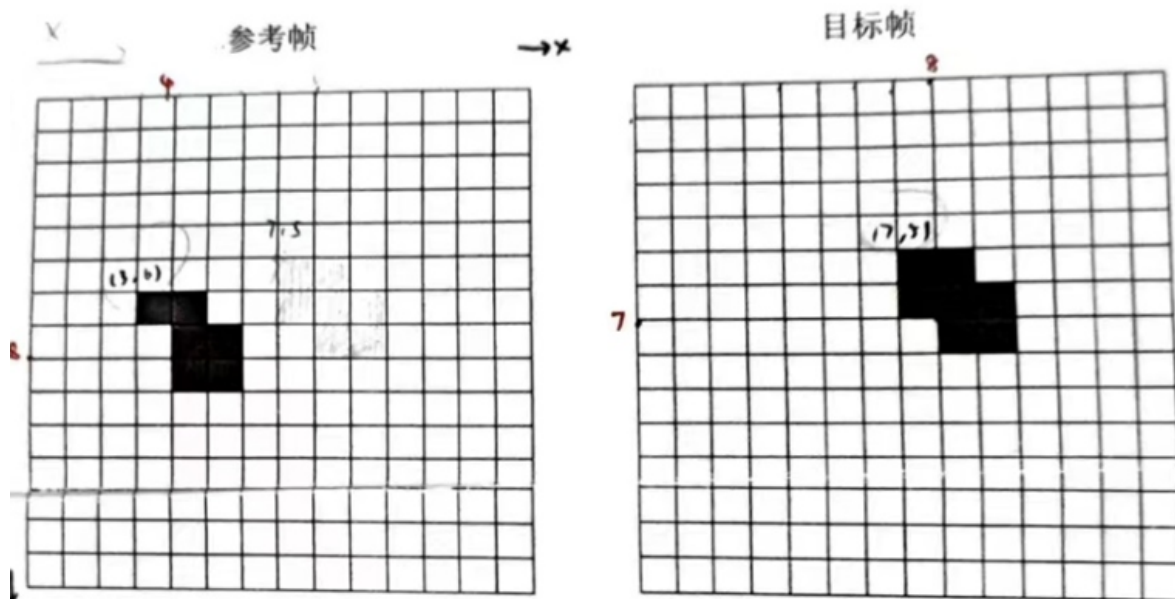
目标帧宏块大小是3x3(即九个像素点组成),运动向量  $MV(u, v); u, v \in [-p, p], p = 5$ , 帧中左上角为坐标原点, x轴向右为正方向, y轴向下为正方向。宏块的九个像素点中,红色像素的像素值为40,其余像素值为120。求最优运动向量,及对应的 MAD。

1. 帧内宏块运动估计: 帧内宏块运动估计 (帧内运动估计) 是指宏块在帧内的运动。

为了方便起见,我们用目标帧中左上角的坐标值  $(x, y)$  作为宏块的原点。设  $C(x+k, y+l)$  为目标(当前)帧的宏块中的像素,  $R(x+i+k, y+j+l)$  为参考帧的宏块中的像素,其中  $k$  和  $l$  代表宏块中的像素的索引,  $i$  和  $j$  分别为水平和垂直的位移。两个宏块的差可以用它们的平均绝对误差 (Mean Absolute Difference, MAD) 来测量,定义为:

$$MAD(i, j) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)| = (120 - 40)/9 = 80/9,$$






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

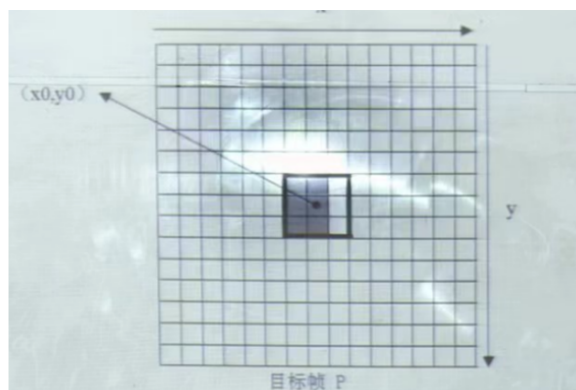
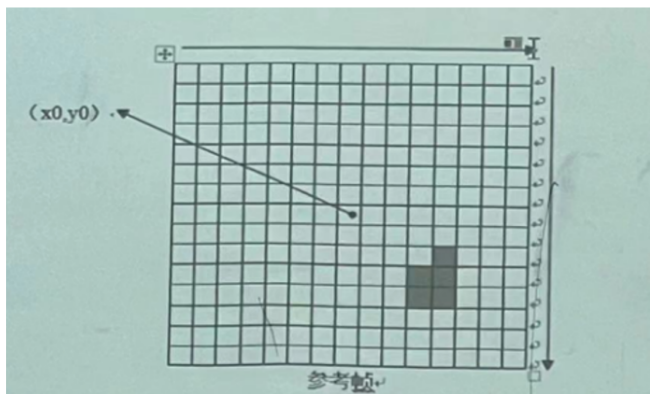
2. 目标缺块大小是  $(3 \times 3)$  (即九个像素点组成)，运动向量  $MV((v_x, v_y))$  点  $(v \in [-p, p])$ ,  $(p = 5)$ , 缺中左上角为坐标原点, x 轴向右为正方向, y 轴向下为正方向。宏块的九个像素点中, 红色像素的像素值为 40, 其余像素像素值为 120。进行 2D 对数搜索。写出搜索步骤 (用 1、2、3 画出每轮步中搜索点的位置), 并计算每一步的最优平均绝对误差 MAD。

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

P217

 3f39aa68516523d68f82faf1d3d768e

目标帧为P帧, 宏块大小是  $3 \times 3$ , 运动向量  $MV((v_x, v_y))$  点  $(v \in [-p, p])$ ,  $(p = 5)$ 。目标P帧中宏块中心坐标是  $(x_0, y_0)$ , 包含6个黑色像素(一个小单元格表示一个像素), 每个黑色像素值为50, 其余值为100。

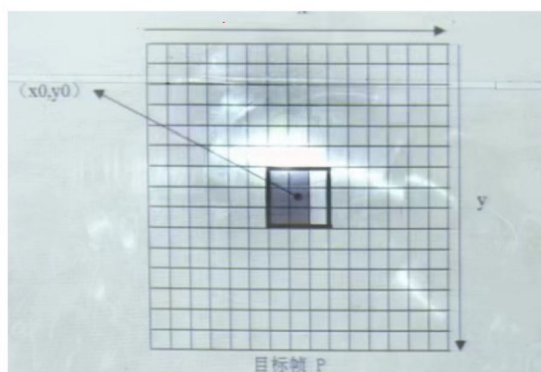
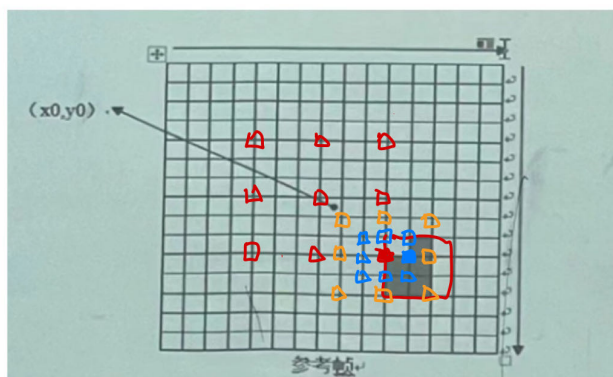


(1)求最优运动向量，及对应的 MAD;

运动向量 (4, 3)

$$MAD(i, j) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)| = (100 - 50)/9 = 50/9,$$

(2)求采用2D 对数搜索的最优匹配的宏块中心点的坐标



$$p = 5 \quad \lceil 5/2 \rceil = 3$$

(3)假设帧率为 25fps，求每秒的计算量是多少？。

P218

$$\begin{aligned} OPS\_per\_second &= (8 \cdot (\lceil \log_2 p \rceil + 1) + 1) \cdot N^2 \cdot 3 \cdot \frac{720 \times 480}{N \cdot N} \cdot 25 \\ &= (8 \cdot (\lceil \log_2 5 \rceil + 9) \times 3^2 \times 3 \times \frac{720 \times 480}{3 \times 3} \times 25. \\ &\approx 8.5536 \times 10^8 \end{aligned}$$

