

数字图像与视频处理-习题解答

第 1 章 图像与视频处理基础

1.9 习题

1. 说明彩色三要素的物理含义。
2. 请阐述三基色原理及其在彩色电视系统中的应用。
3. 简述 RGB 颜色空间模型、HSI 颜色空间模型是如何对颜色进行描述的。
4. 与 NTSC 制相比较, PAL 制有哪些特点?
5. TU-RBT.601 建议有哪些主要内容?有何实际意义?
6. TU-RBT.656 建议与 U-RBT.601 建议之间存在什么关系?ITU-R BT.1120 建议与 ITU-R BT.709 建议之间存在什么关系?
7. 请编写 RGB 颜色空间和 HSI 颜色空间相互转换的 MATLAB 程序。

1.9 习题解答

1. 答: 在色度学中, 任一彩色光可用亮度 (Lightness, 也称为明度)、色调(Hue)和色饱和度 (Saturation)这三个基本参量来表示, 称为彩色三要素。亮度也称为明度或明亮度, 是光作用于人眼时所引起的明亮程度的感觉, 用于表示颜色明暗的程度。色调是指颜色的类别, 通常所说的红色、绿色、蓝色等, 就是指色调。色调是决定色彩本质的基本参量, 是色彩的重要属性之一。饱和度是指彩色光所呈现色彩的深浅程度, 也称为彩度。
2. 答: 三基色原理是指自然界中常见的大部分彩色都可由三种相互独立的基色按不同的比例混得到。所谓独立, 是指其中任何一种基色都不能由另外两种基色混合得到。

三基色原理包括如下内容:

1)选择三种相互独立的颜色基色, 将这三基色按不同比例进行组合, 可获得自然界各种彩色感觉。

2)任意两种非基色的彩色相混合也可以得到一种新的彩色, 但它应该等于把两种彩色各自分解为三基色, 然后将基色分量分别相加后再相混合而得到的颜色。

3)三基色的大小决定彩色光的亮度, 混合色的亮度等于各基色分量亮度之和。

4)三基色的比例决定混合色的色调, 当三基色混合比例相同时, 色调相同。

在**彩色电视系统中的应用**: 彩色电视系统到目前为止, 在前端摄像机采集景物图像的颜色、演播室的节目制作和中间的节目传输都是采用红、绿、蓝三基色;而在终端显示部分, 只是有些企业为渲染彩色重现效果, 在电视机的信号处理电路部分分别采用“六色”或“五色”或“四色”的处理技术, 但在终端显示还是以 R、G、B 三基色相加混合重现彩色图像, 重现的彩色范围不会超过三基色相加混色限定的范围。

3. RGB 颜色空间模型:

在 RGB 颜色空间, 对任意彩色光 F, 其配色方程可写为:

$$F = r[R] + g[G] + b[B]$$

式中, r、g、b 为三色系数; r[R]、g[G]、b[B]为 F 色光的三色分量。

RGB 颜色空间模型可以用笛卡尔坐标系中的立方体来形象表示，3 个坐标轴的正方向分别是 R、G、B 三基色，用三维空间中的一个点来表示一种颜色，如下图。每个点有 3 个分量，分表代表该店颜色的红（R）、绿（G）、蓝（B）三基色的值。为了方便描述，将各基色的取值范围从 0~255 归一化到 0~1。

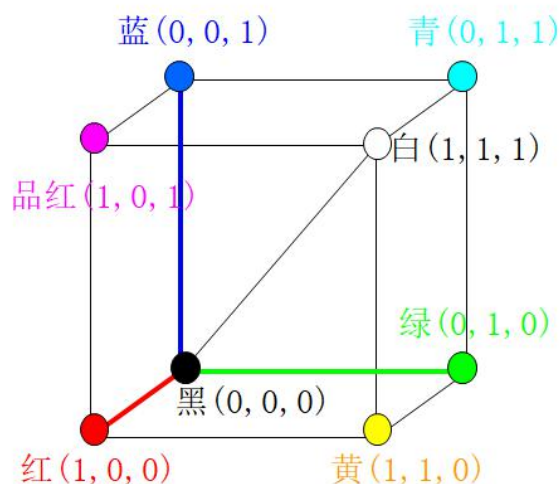


图: RGB 颜色空间模型

HSI 颜色空间模型:

HSI/HSV 颜色空间模型用色调、饱和度、亮度来描述物体的颜色，用 H、S、I/V 三参数描述颜色特性，其中 H 定义颜色的频率，称为色调；S 表示颜色的深浅程度，称为饱和度；I/V 表示强度或亮度。

可用圆锥体表示，H 以绕圆锥中心轴的角度表示，取值范围为 $[0^\circ, 360^\circ]$ 。一般假定，红色对应 $H = 0^\circ$ ，绿色 $H = 120^\circ$ ，蓝色对应 $H = 240^\circ$ 。 $0^\circ \sim 240^\circ$ 之间的色调覆盖了所有可见光谱的颜色，在 $240^\circ \sim 360^\circ$ 之间的色调为人眼可见的非光谱色（紫色）。S（色饱和度）参数有色相环的圆心（圆心）到彩色点的半径的长度表示，归一化后取值范围为 $[0, 1]$ 。V（亮度）直接用圆锥的中心轴表示，取值范围也为 $[0, 1]$ 。

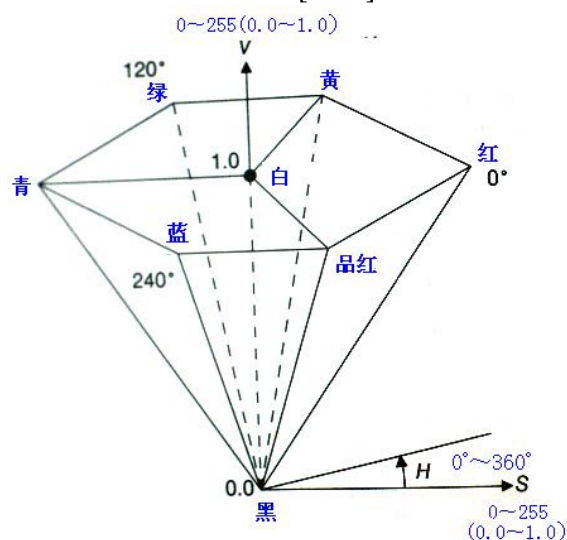


图: HSV 颜色空间模型

4. 与 NTSC 制相比较，PAL 制有下列特点：

1) 克服了 NTSC 制对相位失真敏感的缺点，使色度信号在传输过程中的相位失真对重现彩色的影响减少，因此，对传输设备和接收机的技术指标要求，PAL 制比 NTSC 制低。

- 2) 比 NTSC 制抗多径接收性能好。
- 3) PAL 制相对 NTSC 制而言，色度信号的正交失真不敏感，并且对色度信号部分抑制边带而引起的失真也不敏感。
- 4) PAL 接收机中采用梳状滤波器，可使亮度串色的幅度下降 3dB，并且可以提高彩色信噪比 3dB。
- 5) 电路、设备较 NTSC 制复杂，接收机价格较高。

5. ITU-R BT.601 建议是 1982 年 2 月在国际无线电咨询委员会第 15 次全会上通过的，确定了以分量数字编码 4:2:2 标准作为演播室彩色电视信号数字编码的国际标准。该建议考虑到现行的多种彩色电视制式，提出了一种世界范围兼容的数字编码方式，是向数字电视系统参数统一化，标准化迈出的第一步。该建议对彩色电视信号的编码方式，采样频率，采样结构都做了明确的规定。

6. ITU-R BT.601 建议是"演播室数字电视编码参数"标准,而 ITU-R BT.656 建议则是 ITU-R BT.601 建议的附件 A 中的数字接口标准,用于主要数字视频设备(包括芯片)之间采用 27Mhz/s 并口或 243Mb/s 串行接口的数字传输接口标准。

在色彩方面，ITU-R BT. 2020 建议相对于 ITU-R BT.709 建议做出了大幅度改进。首先是在色彩的比特深度方面，由的 ITU-R BT.709 标准的 8bit 提升至 10bit 或 12bit，这一提升对于整个影像在色彩层次与过渡方面的增强起到了关键作用。除了色彩比特深度的提升以外，ITU-R BT. 2020 建议定义的色彩三角形的范围远远大于 ITU-R BT.709 建议规定的范围，也就意味着超高清系统能够显示更多的色彩。

7. （1）HSI 颜色空间转换到 RGB 空间：

```
function rgb=hsi2rgb(hsi)
```

```
%将图像从 HSI 颜色空间转换到 RGB 空间
```

```
%初始化 R G B 分量矩阵
```

```
R=zeros(size(hsi,1),size(hsi,2));
```

```
G=zeros(size(hsi,1),size(hsi,2));
```

```
B=zeros(size(hsi,1),size(hsi,2));
```

```
%0<=H<2*pi/3 时
```

```
idx=find((0<=H)&(H<2*pi/3));
```

```
B(idx)=I(idx).*(1-S(idx));
```

```
R(idx)=I(idx).*(1+S(idx).*cos(H(idx))./cos(pi/3-H(idx)));
```

```
G(idx)=3*I(idx)-(R(idx)+B(idx));
```

```
% 2*pi/3<=H<4*pi/3 时
```

```
idx=find((2*pi/3<=H)&(H<4*pi/3));
```

```
R(idx)=I(idx).*(1-S(idx));
```

```
G(idx)=I(idx).*(1+S(idx).*cos(H(idx)-2*pi/3)./cos(pi-H(idx)));
```

```
B(idx)=3*I(idx)-(R(idx)+G(idx));
```

```

%4*pi/3<= H<=2*pi 时
idx=find((4*pi/3<=H)&(H<=2*pi));
G(idx)=I(idx).*(1-S(idx));
B(idx)=I(idx).*(1+S(idx).*cos(H(idx)-4*pi/3)./cos(5*pi/3-H(idx)));
R(idx)=3*I(idx)-(G(idx)+B(idx));

%合并三个通道分量，并进行归一化
rgb=cat(3,R,G,B);
rgb=max(min(rgb,1),0);

```

(2) RGB 颜色空间转换到 HSI 空间：

```

function hsi=rgb2hsi(rgb)
%将图像从 RGB 颜色空间转换到 HSI 空间

%提取 R,G,B 分量
rgb=im2double(rgb);
r=rgb(:,:,1);
g=rgb(:,:,2);
b=rgb(:,:,3);

%进行转化
num=0.5*((r-g)+(r-b));
den=sqrt((r-g).^2+(r-b).*(g-b));
theta=acos(num./(den+eps));
H=theta;
H(b>g)=2*pi-H(b>g);
H=H/(2*pi);
num=min(min(r,g),b);
den=r+g+b;
den(den==0)=eps;
S=1-3.*num./den;
H(S==0)=0;
I=(r+g+b)/3;

%合并三个分量
hsi=cat(3,H,S,I);

```

第2章 图像增强

2.10 习题

1. 图像增强的目的是什么?它包含哪些内容?
2. 灰度变换的目的是什么?有哪些实现方法?
3. 给出把灰度范围从[20, 100]扩展为[0, 250], 把灰度范围从[20, 240]压缩为[25, 150]的变换函数。
4. 什么是灰度直方图?为什么一般情况下对离散图像的直方图均衡化并不能产生完全平坦的直方图?
5. 图像平滑的目的是什么?空间域图像平滑的方法有哪些?
6. 中值滤波的原理是什么?它有哪些特点?它主要用于消除什么类型的噪声?
7. 图像锐化的目的是什么?空间域常用的图像锐化算子有哪几种?
8. 简述用于平滑滤波和锐化处理的滤波器之间的区别和联系。
9. 频率域低通滤波的原理是什么?有哪些滤波器可以利用?
10. 什么是同态滤波?简述其基本原理。
11. 什么是伪彩色图像增强?其主要目的是什么?伪彩色处理的方法有哪些?

2.10 习题解答

1. 答: 目的: 改善降质图像, 根据一定的要求将图像中感兴趣的部分加以处理或突出有用的图像特征(如边缘、轮廓、对比度等), 抑制不需要的信息, 以改善图像的主观视觉效果或便于后续的图像分析和识别。

内容: 图像增强算法按其运算处理所进行的作用域不同, 可分为空间域法和频率域法两大类: 空间域法是在空间域内直接对图像的像素值进行运算操作。空间域法又分为点运算处理法和邻域运算处理法。频率域法是先通过正交变换将图像从空间域变换到频率域, 然后在频率域中对变换系数值进行运算的操作, 增强感兴趣的频率分量, 然后再进行反变换到空间域, 得到增强后的图像。常用方法包括低通滤波, 高通滤波以及同态滤波等。此外还有彩色增强法, 常用方法包括伪彩色增强和假彩色增强。

2. 答: 灰度变换的目的是改变图像的像素灰度值, 以扩展图像的灰度值动态范围, 或增强图像的对比度, 从而使图像变得层次丰富或使图像特征变得明显。灰度变换的实现方法有: 灰度的线性变换、灰度的非线性变换、直方图修正和直方图规定化。

3. 答: 灰度范围从[20, 100]扩展为[0, 250]的变换函数为:

$$g(x,y) = \frac{250 - 0}{100 - 20} [f(x,y) - 20] + 0$$

$$\text{得: } g(x,y) = \frac{25}{8} f(x,y) - \frac{125}{2}$$

灰度范围从[20, 240]压缩为[25, 150]的变换函数为:

$$g(x,y) = \frac{150 - 25}{240 - 20} [f(x,y) - 20] + 25$$

$$\text{得: } g(x,y) = \frac{25}{44} f(x,y) + \frac{150}{11}$$

4. 答：灰度直方图是灰度级的频数，它表示图像中具有某种灰度级像素的个数，反映了图像中每种灰度级出现的概率。

由于离散图像中的灰度级是用离散的二进制整数表示的，所以均衡化的直方图不能用连续的实数表示灰度，只能用近似的整数表示，这使得对离散图像的直方图均衡化不能产生完全平坦的直方图。

5. 答：图像平滑的目的是减弱、抑制或消除噪声而改善图像质量。

空间域图像平滑的方法有邻域平均法、中值滤波和多图像平均法等。

6. 答：中值滤波的原理是选用一个含有奇数个像素的滑动窗口，将该窗口在图像上扫描，把其中所含的像素点按灰度级的升（或降）序排列，取位于中间的灰度值，来代替窗口中心点的灰度值。

中值滤波的特点：（1）对某些输入信号（如在窗口内单调增加或单调减少的序列）中值滤波具有不变性；（2）中值滤波的去噪声性能。

中值滤波对含有椒盐噪声的图像效果很好。

7. 答：图像锐化的目的是为了突出图像的边缘信息，加强图像的轮廓特征，以便于人眼的观察和机器的识别。

空间域常用的图像锐化算子有梯度算子、索贝尔(Sobel)算子、拉普拉斯(Laplacial)算子。

8. 答：平滑滤波和锐化处理的滤波器之间的区别和联系：从信号频谱角度来看，信号缓慢变化部分在频率域属于低频部分，而信号的迅速变化部分在频率域的高频部分。要在频率域消除噪声干扰的影响，就要设法减弱高频分量，可以采用低通滤波的方法来减弱高频分量，所以平滑滤波器是低通滤波器。图像中的边缘、线条等细节部分在频率域中对应于高频分量，所以采用高通滤波技术让高频分量通过，是低频分量受到抑制，就能够得到图像的边缘信息，实现图像的锐化，所以锐化滤波器是高通滤波器。

9. 答：频率域低通滤波的原理：低通滤波器输出 $G(u,v) = H(u,v)F(u,v)$ ，其中 $F(u,v)$ 是含噪声图像的傅里叶变换； $H(u,v)$ 是低通滤波器传递函数。利用 $H(u,v)$ 使 $F(u,v)$ 的高频分量得到衰减，得到 $G(u,v)$ 后再经过傅里叶反变换就得到所希望的图像 $g(x,y)$ 。

常用的频率域低通滤波器有：理想低通滤波器、巴特沃兹低通滤波器、高斯低通滤波器、梯形低通滤波器等。

10. 答：同态滤波是一种在频域中同时将图像亮度范围进行压缩和将图像对比度进行增强的方法。

同态滤波的基本原理是先对待增强的图像取对数，然后进行傅里叶变换，在频率域中进行适当的滤波，最后通过反傅里叶变换及指数变换得到增强的图像。

11. 答：伪彩色增强是针对灰度图像提出的，其目的是把离散灰度图像的不同灰度级按照线性或者非线性关系映射成不同的颜色，得到一幅彩色图像，以改善图像的视觉效果，提高图像内容的可辨识度，使得图像的细节更加突出目标更容易识别。伪彩色增强的主要方法有灰度分层法、灰度级彩色变换和频率域滤波法。

第3章 形态学图像处理

3.8 习题

1. 什么是数学形态学？其基本思想是什么？简要描述数学形态学方法的实现。
2. 数学形态学有哪几个基本运算？
3. 数学形态学方法适用于图像处理的哪些方面？写出四种以上数学形态学方法的实际应用。
4. 说明二值膨胀运算和腐蚀运算对图像处理的作用及其特点。
5. 说明二值开运算和闭运算对图像处理的作用及其特点。
6. 开运算和腐蚀运算相比有何优越性？闭运算和膨胀运算相比有何优越性？
7. 简述边缘提取算法的主要步骤。什么是内边缘、外边缘和形态学边缘？
8. 什么是图像的骨架？骨架提取有什么作用？
9. 什么是区域填充？简述区域填充的主要算法流程。
10. 什么是细化？它与腐蚀有什么区别？简述细化的主要流程。
11. 什么是粗化？它与膨胀有什么区别？简述粗化的主要流程。
12. 已经一幅灰度图像为A，结构元素为B，试写出结构元素B对A进行腐蚀运算与膨胀运算的结果。（A与B见课本，注意：A四周像素不需要计算；B按“十字形”模板，四角的“0”不参与计算）
13. 灰度开运算和灰度闭运算的定义是什么？它们的几何解释是什么？
14. 灰度的腐蚀运算、膨胀运算、开运算和闭运算分别具有什么性质？互相之间有什么关系？
15. 灰度图像的形态学梯度如何计算？有什么作用？
16. 采用一般的空间梯度算子和形态学梯度进行图像处理时有什么不同？
17. 什么是高帽（Top-hat）变换？它有什么作用？

3.8 习题解答

1. 数学形态学是一门建立在集合论基础上的学科，它是几何形态分析和描述的有力工具。基本思想是以集合论为数学工具，具有完备的数学理论基础，它的运算由集合运算完成。具体实现过程：利用具有一定形态的结构元素，即具有某种特定结构形状的基本元素作为“探针”来探测目标图像，当探针在图像中不断移动时，便可考察图像的形状和各个部分之间的相互关系，从而获取有关图像的形状结构特征的信息，进而达到对图像进行分析和识别的目的。
2. 膨胀、腐蚀、开运算、闭运算。
3. 边缘提取、骨架提取、区域填充、粗化、细化。
4. 腐蚀运算表示用某种探针（即结构元素）对一个图像进行探测，以便找出在图像内部可以放下该结构的区域。腐蚀具有缩小图像和消除图像中比结构元素小的成分的作用，可用于去除物体之间的黏连。
膨胀运算表示将图像A中每一个点 x 扩大到 $B+x$ ，其中B为结构元素，膨胀对图像有扩张作用。
5. 开运算具有平滑图像外边缘的作用。闭运算具有平滑图像内边缘的作用。
6. 腐蚀是一种消除边界点，使边界向内部收缩的过程。可以用来消除小且无意义的物体。

开运算通常是在需要去除小颗粒噪声，以及断开目标物之间粘连时使用。其主要作用与腐蚀相似，与腐蚀操作相比，具有可以基本保持目标原有大小不变的优点。

膨胀运算和闭运算相比，都具有填充图像中比结构元素小的小孔和狭窄缝隙的作用。但膨胀运算在填充图像中比结构元素小的小孔和狭窄缝隙的同时，会使图像中的目标物体极大；而闭运算在填充图像中比结构元素小的小孔和狭窄缝隙的同时，能较好地保持图像中的目标物体的大小不变。

7. 用一定的结构元素对目标图像进行形态学处理，而将处理后的结果与原图像相减。依据形态学运算的不同，可以得到二值图像的内边缘、外边缘和形态学梯度 3 种边缘。其中内边缘是用原图像减去腐蚀的结果图像得到；外边缘可用图像膨胀结果减去原图像得到；形态学梯度可用图像的膨胀结果减去图像的腐蚀结果得到。
8. 骨架是描述图像的几何形状及其拓扑性质的重要特征之一。抽取图像骨架的目的是为了表达目标形状结构，它有助于突出目标的形状特点和减少冗余的信息量。因而，骨架抽取在文字识别、工业零部件形状识别或地质构造识别等领域有着重要的作用。
9. 区域填充是指在已知区域边缘的基础上所完成的对该区域的填充操作。与边缘提取不同，区域填充是对图像背景像素进行操作，一般以图像的膨胀、求补和求交运算为基础，旨在填充图像中我们感兴趣的边界区域。

具体操作过程：首先在边界内取一初始点并标记为 1，利用 $X_k = (X_{k-1} \oplus B) \cap A^c, k=1, 2, 3 \dots$ 对图像进行区域填充，直到 $X_k = X_{k-1}$ ，最后结果为 $X_k \cup A$ 。

10. 细化就是把输入的具有一定宽度的图像轮廓用逐次去掉边缘的方法最终变为宽度仅为一个像素的估计。细化是一种保持原图像连通性的腐蚀运算。细化为了弥补腐蚀在数学形态学分析中的某种缺陷而提出来的。这种缺陷表现在：如果对一个仅有细小连接的目标图像进行腐蚀处理，当腐蚀深度达到一定的深度时，连接两部分的狭窄连接就会被腐蚀掉。原本属于同一个目标的部分就会被分解为两个独立的部分。细化就是重复地剥离二值图像的边缘像素，利用一组结构元素序列对图像进行信息化。
11. 粗化处理对于膨胀在处理邻近目标时总会将其合并的缺点做了改进。改进后的粗化算法可以用紧贴的边缘来拟合目标，从而避免了膨胀对其进行的错误合并。具体流程为，先细化所讨论集合的背景，然后对细化结果求补集，再消除不连贯的点。
12. 灰度腐蚀结果为：

1	2	3	0
2	3	0	-1
3	0	-1	0
0	-1	0	1

灰度膨胀结果：

5	6	7	8
6	7	8	7
7	8	7	4
8	7	4	5

13. 灰度开运算：某灰度结构元素对灰度输入图像先进行灰度腐蚀，然后进行灰度膨胀操作。消除所有比结构元素尺寸小的亮度细节。
- 灰度闭运算：某灰度结构元素对灰度输入图像先进行灰度膨胀，然后进行灰度腐蚀操作。

消除臂结构尺寸较小的暗细节。

14. 灰度腐蚀运算对所有元素都为正的结构元素，则输出图像比输入图像暗。当输入图像的亮细节的结构尺寸小于结构元素，则亮的效果会被削弱，削弱的程度取决于亮细节周围的灰度值和结构元素自身的形状和幅值。

灰度膨胀运算对所有元素为正的的结构元素，则输出图像比输入图像亮。当输入图像中的暗细节面积小于结构元素时，暗的效果将被削弱，削弱的程度取决于膨胀所有结构元素的形状和幅值。

灰度腐蚀和灰度膨胀之间呈对偶关系。

灰度开运算用来消除相对于结构元素尺寸较小的亮细节，灰度闭运算用来消除相对于结构元素尺寸较小的暗细节，两者也是对偶关系。

15. 将灰度膨胀和灰度腐蚀相结合可以用于计算灰度图像的形态学梯度。形态学梯度可用于提取图像的边缘。
16. 与多种空间梯度算子相比，灰度形态学梯度虽然也对噪声比较敏感，但不会加强或放大噪声，使用对称的结构元素来求图像的形态学梯度，还可以是求得的边缘受方向的影响较小。
17. 高帽变换是一种非常有效的形态学变换，因其使用类似高帽形状的结构元素进行形态学图像处理而得名。计算过程即为原图像减除灰度开运算的结果。高帽变换可以从较暗且平滑的背景中提取比较亮的细节，如增强图像阴影部分的细节，对灰度图像进行物体分割，检测灰度图像中的波峰和波谷及细长图像等。

第 4 章 图像分割

4.8 习题

1. 什么是图像分割？目前图像分割的难点主要体现在哪些方面？
2. 图像分割的依据是什么？常用的图像分割方法主要包括哪几类？分别有哪些具体方法？
3. 基于阈值的图像分割方法的基本原理是什么？什么是它的关键和难点？
4. 什么是全局阈值化分割法？基于灰度值的全局阈值化分割有哪几种常见算法？它们的算法原理分别是什么？
5. 相对于全局阈值化分割，局部阈值化分割有什么优点？其基本原理是什么？使用时需注意哪几点？
6. 动态阈值化分割有什么特点？其关键是什么？
7. 什么是图像边缘和边缘检测？
8. 请写出 Robert 算子、Prewitt 算子和 Sobel 算子的模板。它们各有什么特点？
9. 拉普拉斯算子有什么局限性和作用？高斯拉普拉斯（LoG）算子的模板要满足什么特征？请写出 LoG 算子常用模板。
10. Canny 提出的边缘检测算子应满足的 3 个判断标准是什么？Canny 算子的主要实现步骤是什么？Canny 算子有什么优缺点？
11. 什么是光栅扫描跟踪？光栅扫描跟踪方法的基本思想是什么？使用光栅扫描跟踪方法，需要遵循哪三个准则？使用光栅扫描跟踪方法实现边界跟踪的具体步骤是什么？
12. 什么是轮廓跟踪法？采用轮廓跟踪的方法，进行图像分割的具体步骤是什么？
13. 什么是基于区域的图像分割？传统的基于区域的分割方法有几种？
14. 什么是区域生长？其基于方法是什么？决定区域生长好坏的因素有哪些？
15. 什么是区域的分裂与合并？简述其基本步骤。
16. 根据轮廓曲线的不同表示方式，主动轮廓模型可以分为哪几类？Snake 模型包括哪些能量？

4.8 习题解答

1. 答：图像分割就是指把图像分成各具特性的区域，并提取出感兴趣目标的技术和过程。
图像分割的难点主要体现在：1) 绝大数分割方法都是针对具体问题提出的。实际图像中景物情况各异，需要根据实际的情况选择合适的方法。2) 没有一个统一的评价准则来判断分割结果的好坏或者正确与否，无法指导如何选择合适的分割算法。
2. 答：图像分割的依据是各区域具有不同的特性，这些特性可以是灰度、颜色、纹理等。而灰度图像分割的依据是基于相邻像素灰度值的不连续性和相似性。从分割依据的角度来看，灰度图像分割方法可以分为：
 - 1) 基于区域边界灰度不连续性的方法：其包括边缘检测、边缘跟踪、Hough 变换；
 - 2) 基于区域内部灰度相似性的方法：其包括阈值化分割、区域生长、区域分裂与合并生长。
3. 答：基于阈值的图像分割的基本原理是通过图像的灰度直方图进行数学统计，选择一个

或多个阈值将像素划分为若干类。

基于阈值的图像分割的关键和难点就是如何选取一个最佳阈值,使图像分割效果达到最好。

4. 答: 全局阈值化分割法是指在阈值化过程中只使用一个阈值,对整幅图像采取固定的阈值进行分割。

基于灰度值的全局阈值化分割包括:

1) P -分位数法: 其基本原理是根据先验知识,得到目标与背景像素的先验概率比例 P_o/P_b ,再根据此条件依次累计灰度直方图,直到累计值大于或者等于该比例数,此时的灰度值即为最佳阈值。

2) 迭代法: 其原理是根据图像目标的灰度分布情况,选取一个近似阈值作为初始阈值,一个比较好的方法就是将图像的灰度均值作为初始阈值,然后通过分割图像和修改阈值的迭代过程获得认可的最佳阈值。

3) Otsu 算法: 其原理是以目标和背景类间方差最大或类内方差最小为阈值选取准则。

5. 答: 局部阈值化分割法的基本原理是: 即不再像全局阈值一样,对整个矩阵只有一个阈值,而是针对输入矩阵的每一个位置的值都有相应的阈值,这些阈值构成了和输入矩阵同等尺寸的矩阵阈值。

相对于全局阈值化分割,局部阈值化分割的优点是: 将图像分为一个个小区域,或子图像,然后分析每个子图像,并对每个子图像选取相应的阈值,来解决场景中照明不均匀或者背景变化比较大的时候的情况。

选用局部阈值化分割法时需注意以下两点: 1) 首先截取子图像的尺寸不能太小,否则计算出的一些统计结果无意义; 2) 其次,每幅图像的分割是任意的,如果有一幅子图像正好落在目标区域或背景区域,而根据统计结果对其进行分割,可能会产生更差的分割效果。

6. 答: 动态阈值化分割的特点是: 由于在进行图像处理的过程中,当图像中目标的灰度处于不同的灰度域中,单独使用一个全局阈值,可能无法全面的提取所有目标,这时可以采用动态阈值的处理方法,将图像分成多个块,分别对每个块进行操作。

动态阈值化分割的关键: 是将图像分割成许多块,且要对每一块单独的采用迭代法进行处理。

7. 答: 图像边缘是图像最基本的特征,在图像分析中起着重要的作用。

边缘是指图像局部特性发生突变之处,主要存在于目标与目标、目标与背景、区域与区域之间(包括不同色彩)之间。

边缘检测是图像分割、图像分析和理解的重要基础。基于边缘的图像分割方法的基本思路就是先确定图像中的边缘像素,然后就可以把它们连接在一起构成所要的边界。

8. 答: 1) Robert 算子

$$\text{模板: } H_x = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad H_y = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

优缺点: Roberts 算子利用局部差分算子寻找边缘,比较粗糙,定位不是很确定,容易丢失一部分边缘,同时由于图像没经过平滑处理,因此不具备抑制噪声的能力。该算子对具有陡峭边缘且含噪声少的图像效果较好。

2) Prewitt 算子

$$\text{模板: } H_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad H_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$\text{改进模板: } H_x = \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix} \quad H_y = \begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

优缺点：对噪声具有一定的抑制能力，但不能完全排除检测结果中出现的虚假边缘。虽然算子边缘定位效果不错，但检测出的边缘容易出现多像素宽度。对灰度渐变和噪声较多的图像处理效果较好。对边缘的定位比较准确。

3) Sobel 算子的模板

$$\text{模板: } H_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad H_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

优缺点：对噪声具有一定的抑制能力，但不能完全排除检测结果中出现的虚假边缘。虽然算子边缘定位效果不错，但检测出的边缘容易出现多像素宽度。对灰度渐变和噪声较多的图像处理效果较好。对边缘的定位比较准确。

9. 答：拉普拉斯算子的局限性：1) 对噪音的敏感；2) 会产生双边效果；3) 不能检测出边的方向。拉普拉斯算子的作用：1) 检测一个像素是在边的亮的一边还是暗的一边。2) 利用二阶导数零交叉，确定边的位置。

高斯拉普拉斯 (LoG) 算子的模板要满足的特征：应用 LoG 算子，高斯函数中方差参数的选择很关键，对图像边缘检测效果有很大的影响。高斯滤波器为低通滤波器，方差越大，通频带越窄，对较高频率的噪声的抑制作用越大，避免了虚假边缘的检出，同时信号的边缘也被平滑了，造成某些边缘点的丢失。反之，方差越小，通频带越宽，可以检测到的图像更高频率的细节，但对噪声的抑制能力相对下降，容易出现虚假边缘。因此，应用 LoG 算子，为取得更佳的效果，对于不同图像应选择不同参数。

LoG 算子常用模板如下：

$$\nabla^2 = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 0 \\ -1 & -2 & 16 & -2 & -1 \\ 0 & -1 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}$$

10. 答：Canny 提出的边缘检测算子应满足的 3 个判断标准是：

1) 高的检测率，边缘检测算子应该只对边缘进行响应，检测算子不漏检任何边缘，也不应该将非边缘标记为边缘。

2) 精确定位，检测到的边缘与实际边缘之间的距离要尽可能的小。

3) 明确的响应，对每一条边缘只有一次响应，只得到一个点。

Canny 算子的主要实现步骤是：

1) 用高斯滤波器平滑图像；

- 2) 用一阶偏导的有限差分来计算平滑后图像的梯度幅值和梯度方向;
- 3) 对梯度幅值进行“非极大值抑制”;
- 4) 用双阈值算法检测和连接边缘。

Canny 算子的优缺点: 不易受噪声的干扰, 能检测到真正的弱边缘。

11. 答: 光栅扫描跟踪是一种采用电视光栅行扫描顺序对遇到的像素进行分析, 从而确定是否为边缘的跟踪方法。光栅扫描跟踪方法的基本思想是先利用检测准则确定接受对象, 然后根据被接受的对象点和跟踪准则确定新的接受对象, 最后将所有标记为 1 且相邻的对象点连接起来就得到了检测到的曲线。

使用光栅扫描跟踪方法, 需要遵循下面的 3 个准则:

- 1) 参数准则: 需要事先确定检测阈值 d 、跟踪阈值 t 、且要求 $d > t$ 。
- 2) 检测准则: 对图像进行逐行扫描, 依次将每一行中灰度值大于或等于检测阈值 d 的所有点 (称为接受对象点) 的位置记为 1。
- 3) 跟踪准则: 逐行扫描图像, 若图像中位于第 i 行的点 (i, j) 为已接受的对象点, 则在第 $i+1$ 行上找点 (i, j) 的相邻点 $(i+1, j-1)$ 、 $(i+1, j)$ 和 $(i+1, j+1)$, 将其中灰度值大于或等于跟踪阈值 t 的邻点确定为新的接受对象点, 并将相应位置记为 1。重复此过程, 直至图像中除最末一行以外的所有接受对象点扫描完为此。此时位置 1 的像素点连成的曲线即为检测到的边缘。

使用光栅扫描跟踪方法实现边界跟踪的具体步骤为:

- 1) 确定一个比较大的阈值 d , 把高于该阈值的像素作为对象点, 该阈值被称为检测阈值, 被设为 7;
- 2) 选择一个比较低的阈值 t 作为跟踪阈值, 且要求 $t < d$, 该阈值可以根据不同准则来选择;
- 3) 从第一行开始, 根据检测准则扫描图像, 并将其灰度值大于或等于检测阈值 d 的所有像素点的位置记为 1;
- 4) 确定跟踪邻域;
- 5) 从第二行起逐行扫描图像, 并按跟踪准则将灰度值大于或者等于跟踪阈值 $t=4$ 的所有像素确定为新的接受对象点, 且将其相应位置记为 1;
- 6) 对于已检测出来的某个对象点, 如果在下一行跟踪邻域中, 没有任何一个像素被接受为对象点, 那么这一条曲线的跟踪便可结束。如果同时有两个, 甚至三个相邻点均被接受为对象点, 则说明曲线发生分支, 跟踪需对各分支同时进行。如果多条曲线合并成一条曲线, 则跟踪可集中于一条曲线上进行。如果某个对象点 (在步骤 3) 产生的对象点) 在上一行的对于邻域中没有对象点, 则说明一条新的曲线可开始;
- 7) 重复步骤 5) 和 6), 直至图像中最末一行被扫描完为止。将标记为 1 (包括 1), 主要是为了便于区别检测准则和跟踪准则的结果) 的像素连接起来, 就得到了检测获得的结果曲线。

12. 答: 轮廓跟踪法是一种适用于黑白二值图像的图像分割方法, 而且轮廓跟踪改变了光栅扫描跟踪中扫描方向的单一缺点, 跟踪方向可以是任意方向, 并且有足够大的跟踪距离。显然, 轮廓跟踪是改变了邻域定义和跟踪准则的一种二值图像的光栅扫描跟踪法。

采用轮廓跟踪法进行图像分割的算法步骤如下:

- 1) 在靠近边缘出任取一起始点, 然后按照每次只前进一步, 步距为一个像素的原则开始跟踪;
- 2) 当跟踪中的某一步是由白区进入黑区时, 然后各步向左转, 直到穿出黑区为止;
- 3) 当跟踪中的某一步是由黑区进入白区时, 然后各步向右转, 直到穿出白区为止;
- 4) 当围绕目标边界循环跟踪一周回到起点时, 所跟踪的轨迹便是目标的轮廓; 否则, 应继续按步骤 2) 和步骤 3) 的原则进行跟踪。

13. 答：基于区域的图像分割是以直接寻找区域为目的的图像分割技术，其原理不同于阈值化分割和边缘检测，不需要直接利用阈值或者边界来划分图像。

传统的基于区域的图像分割方法有两种：1) 区域生长法；2) 区域分裂与合并法。

14. 答：区域生长也称为区域增长，其基本思想是根据事先定义的相似性准则，将图像中满足相似性准则的像素或子区域聚合成更大区域的过程。

其基本方法是：先对每个需要分割的区域找一个种子像素作为生长起点，然后将种子像素周围邻域中与种子像素有相同或相似性质的像素(根据某种事先确定的生长或相似准则来判定)合并到种子像素所在的区域中。将这些新像素当做新的种子像素继续进行上面的过程，直到再没有满足条件的像素可被包括进来，这样一个区域就长成了。

决定区域生长好坏的因素有：1) 初始点(种子点)的选择；2) 生长准则；3) 终止条件。

15. 答：区域分裂与合并即区域特征一致性的测度,当图像中某个区域的特征不一致时就将该区域分裂成 4 个相等的子区域,当相邻的子区域满足一致性特征时则将它们合成一个大区域,直至所有区域不再满足分裂合并的条件为止。

区域分裂与合并的具体步骤：

- 1) 对任一区域 R_i ，如果 $P(R_i)=FALSE$ ，就将其分裂成不重叠的 4 个子区域；
- 2) 对相邻的两个区域 R_i 和 R_j (它们可以大小不同，即不在同一层)，如果条件 $P(R_i \cup R_j)=TRUE$ 成立，就将它们合并为一个区域；
- 3) 重复以上各步，直到不能再进行区域分裂和合并为止。

16. 答：根据轮廓曲线的不同表示方式，主动轮廓模型大致可以分为参数主动轮廓模型和几何主动轮廓模型。

主动轮廓模型可分为经典 Snake 模型、气球模型、基于梯度矢量流模型。

Snake 模型包括的能量：

- 1) 内部能量 E_{int} 约束轮廓的连续性和光滑性；
- 2) 图像力 E_{image} 推动曲线向图像的显著特征如线、边缘和主观轮廓靠近；
- 3) 外部约束力 E_{con} 使曲线到达期望的能量局部极小值处；
- 4) E_{image} 和 E_{con} 统称外部能量 E_{ext} 。

第 5 章 数字图像与视频压缩编码原理

5.7 习题

1. 为什么要对图像数据进行压缩？其压缩原理是什么？图像压缩编码的目的是什么？目前有哪些编码方法。
2. 一个信源包含 6 个符号消息，它们的出现概率分别为 0.3、0.2、0.15、0.15、0.1、0.1，请对该信源进行哈夫曼编码，并求出码字的平均长度和编码效率。
3. 设有一个信源具有 4 个可能出现的符号 X_1 、 X_2 、 X_3 、 X_4 ，共出现的概率分别为 1/2、1/4、1/8、1/8。请以符号序列 $X_2 X_1 X_4 X_3 X_1$ 为例解释其算术编码和解码的过程。
4. 请比较算术编码和哈夫曼编码的特点。
5. 请说明预测编码的原理，并画出 DPCM 编解码器的原理框图。
6. 预测编码是无损编码还是有损编码？为什么？
7. DCT 本身能不能压缩数据？为什么？请说明 DCT 变换编码的原理。
8. 目前最常用的运动估计技术是什么？其假设的前提条件是什么？块大小的选择与运动矢量场的一致性是如何考虑的？

5.7 习题解答

1. 答：1) 图像信息在数字化后数据量巨大，为了进行有效的存储和传输，需要对图像数据进行压缩。
 2) 压缩原理在于：图像信息存在大量冗余，如空间冗余、时间冗余、统计冗余、结构冗余、知识冗余、人眼视觉冗余等，通过去除图像中的冗余信息，就可以实现压缩。
 3) 图像压缩编码的目的是：通过去除图像中的冗余信息，得到图像的有效表达，减少图像存储所需的码率。
 4) 目前的编码方法主要有：熵编码、预测编码和变换编码等。
2. 答：哈夫曼编码的过程如下，编码时概率大的赋码字 0，概率小的赋码字 1：

符号	概率	编码过程	码字	码长
x1	0.3	0.3 → 0.3 → 0.3	00	2
x2	0.2	0.2 → 0.2 → 0.3	10	2
x3	0.15	0.2 → 0.2 → 0.2	010	3
x4	0.15	0.15 → 0.15 → 0.2	011	3
x5	0.1	0.15 → 0.15 → 0.15	110	3
x6	0.1	0.1 → 0.15 → 0.15	111	3

信源熵：
$$H(X) = -\sum_{i=1}^6 p_i \log p_i = 2.471 \text{ bit/符号}$$

平均码字长度：
$$\bar{K} = \sum_{i=1}^6 p_i K_i = 2.5$$

编码效率：
$$\eta = \frac{H(X)}{\bar{K}} = 98.84\%$$

3. 答：首先根据概率可将区间[0,1]分成 4 个子区间[0,0.5), [0.5,0.75), [0.75,0.875), [0.875,1),

(1) 编码：初始时，设置当前区间左端点值 $low=0$ ，右端点值 $high=1.0$ ，当前区间长度 $length=1.0$ 。

1)对输入的 x_2 编码，有：

$$\begin{cases} low = low + length \times symbol_low = 0 + 1 \times 0.5 = 0.5 \\ high = low + length \times symbol_high = 0 + 1 \times 0.75 = 0.75 \end{cases}, \text{ length} = high - low = 0.25$$

2)对输入的 $x_2 x_1$ 编码，有：

$$\begin{cases} low = low + length \times symbol_low = 0.5 + 0.25 \times 0 = 0.5 \\ high = low + length \times symbol_high = 0.5 + 0.25 \times 0.5 = 0.625 \end{cases}, \text{ length} = 0.125$$

3)对输入的 $x_2 x_1 x_4$ 编码，有：

$$\begin{cases} low = low + length \times symbol_low = 0.5 + 0.125 \times 0.875 = 0.609375 \\ high = low + length \times symbol_high = 0.5 + 0.125 \times 1 = 0.625 \end{cases}, \text{ length} = 0.015625$$

4)对输入的 $x_2 x_1 x_4 x_3$ 编码，有：

$$\begin{cases} low = 0.609375 + 0.015625 \times 0.75 = 0.62109375 \\ high = 0.609375 + 0.015625 \times 0.875 = 0.623046875 \end{cases}, \text{ length} = 0.001953125$$

5)对输入的 $x_2 x_1 x_4 x_3 x_1$ 编码，有：

$$\begin{cases} low = 0.62109375 + 0.001953125 \times 0 = 0.62109375 \\ high = 0.62109375 + 0.001953125 \times 0.5 = 0.6220703125 \end{cases}$$

取左端点值 0.62109375 作为序列编码后的结果。

(2) 解码：

1) $(0.62109375 - 0) / 1 = 0.62109375 \in [0.5, 0.75)$ ，解码出 x_2 ；

2) $(0.62109375 - 0.5) / 0.25 = 0.484375 \in [0, 0.5)$ ，解码出 x_1 ；

3) $(0.484375 - 0) / 0.5 = 0.96875 \in [0.875, 1)$ ，解码出 x_4 ；

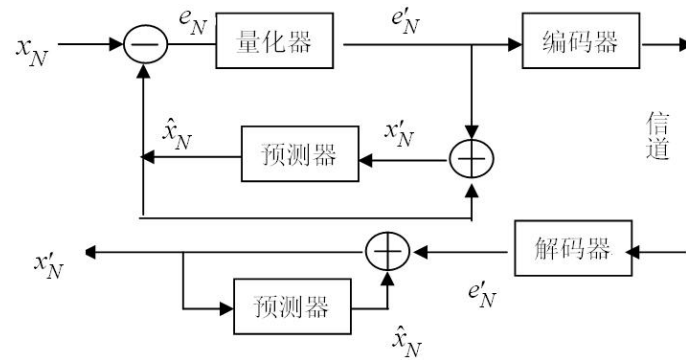
4) $(0.96875 - 0.875) / 0.125 = 0.75 \in [0.75, 0.875)$ ，解码出 x_3 ；

5) $(0.75 - 0.75) / 0.125 = 0 \in [0, 0.5)$ ，解码出 x_1 ；

4. 答：1) 两者都是利用信源符号的概率分布进行编码，都属于熵编码。2) 区别在于哈夫曼编码为每一个符号都分配一个码字，而算术编码则是针对符号序列进行编码，序列编码结果为一个小数。3) 算术编码的效率要高于哈夫曼编码，但其计算过程比哈夫曼编码更为复杂，硬件实现也更复杂。

5. 答：1) 图像预测编码的原理是利用图像数据的空间和时间相关性，用相邻的已编码传输的像素值来预测当前待编码的像素值，然后对当前待编码像素的实际值与预测值之差（预测误差）进行编码传输，而不是对当前像素本身进行编码传输，以去除图像数据中的空间和时间相关冗余。在接收端，将收到的预测误差的码字解码后再与预测值相加，得到当前像素值。

2) DCPM 编解码的原理框图如下：



6. 答：预测编码可认为是无损编码。因为编码端与解码端的预测器保持一致的情况下，如果不考虑预测误差的量化失真，则解码端可无失真地重建原始像素值。
7. 答：1) DCT 本身不能压缩数据。2) 因为 DCT 变换只是将信号从空间域变换到频率域，信号的总能量没有变化，只是能量的分布发生了变化。3) 图像 DCT 变换后，变换系数之间的相关性大大减弱，为在变换域里进行有效的压缩（如之字形描述和熵编码）创造了有利条件。
8. 答：1) 目前最常用的运动估计方法是基于块匹配的方法。2) 其假设的前提条件是：认为块内各像素的运动特性一致，即运动矢量相同，可在保持较高运动矢量精度的基础上极大地减小运动矢量估计的计算量。3) 块匹配时，块不能太大，因为块较大时，块内包含不同运动物体的概率变大，块内所有像素做相同运动的假设难以成立，运动矢量估计的精度变小；但块也不能太小，块太小其运动矢量精度会受噪声的影响，且计算量太大，传送运动矢量所需的附加比特数过多，不利用数据压缩。

第 6 章 数字图像与视频压缩编码标准

6.8 习题

1. 国际上主要有哪些数字视频编码标准？
2. 请阐述 H.264/AVC、H.265/HEVC 以及 AVS 视频编码标准中的“类”和“级”的含义。
3. 与以前的视频编码标准相比，H.264/AVC 标准引入了哪些新的技术？
4. 在 H.264/AVC 标准中采用了整数变换，与传统的 DCT 相比有什么优势？
5. 简述 H.264/AVC 标准中的帧内预测原理。
6. AVS1-P2 编码标准与 H.264 标准相比，其性能怎样？有何优势？
7. H.265/HEVC 中的波前并行处理（WPP）技术的作用是什么？

6.8 习题解答

1. 答：国际上主要的数字视频编码标准有 H.26x 系列、MPEG 系列和中国具有自主知识产权的 AVS 系列标准。H.26x 系列如 H.261、H.263、H.264 和 H.265。MPEG 系列如 MPEG-2、MPEG-4。AVS 系列如 AVS1、AVS2 等。
2. 答：“类”也称为档次（Profile），定义了一组编码工具和算法，用于产生一致性的比特流；“级”（Level）用于限定比特流的部分关键参数。
3. 答：H.264/AVC 标准引入新技术包括：4×4 块的帧内预测模式、可变块大小的帧间预测、高精度的亚像素运动估计、多参考帧的运动补偿、整数变换与量化、基于上下文自适应的熵编码等。
4. 答：1）H.264/AVC 标准中采用的整数变换是基于整数的运算，其算法中只需要加法和移位运算，因此运算速度快，并且在反变换过程中不会出现失配问题；2）整数变换基本保持了 DCT 变换的特性，可以较好地去除图像数据中的相关性。
5. 答：H.264/AVC 标准中的帧内预测主要根据空间相关性来预测，对亮度块使用 4×4 与 16×16 的预测块大小，4×4 块有 9 种帧内预测模式，适用于带有大量细节的图像块；16×16 块有 4 种帧内预测模式，适用于平坦区域的图像块。色度块常用采用 8×8 块大小，有 4 种预测模式。
6. 答：1）AVS1-P2 编码标准与 H.264 标准的编码性能相当，具有类似的编码框架。2）AVS1-P2 的主要创新在于提出了一批具体的优化技术，在较低的复杂度下，实现了与 H.264 标准相当的压缩性能，其编码复杂度相当于 H.264 的 70%，解码复杂度相当于 H.264 的 30%。
7. 答：H.265/HEVC 中的波前并行处理（WPP）技术的作用是：在熵编码时不需要打破预测的连贯性，尽可能多地利用上下文信息，可提高编码的并行处理能力，提高编码的时间效率。

第 7 章 图像和视频文件格式

7.7 习题

1. 简述 RIFF 文件格式，并画出 RIFF 文件头结构。
2. 什么是位图？位图中的像素颜色如何表示？
3. 什么是调色板？Windows 操作系统有哪几种调色板？分别起什么作用？
4. 图像文件的组成部分有哪些？分别包含什么内容？请画出图像文件结构示意图。
5. 试述 BMP 文件的结构和各部分的主要作用，描述 BMP 文件的位图文件头和位图信息头的结构。
6. GIF 文件是如何组织数据的？它有什么特点？
7. 什么是 JFIF 文件格式？JPEG 文件中有哪些常用标记？
8. 简述 PNG 文件的特性。
9. SVG 文件格式具有哪些特点？
10. 常见的动画文件格式有哪些？
11. 简要描述 SWF 文件的文件头结构。SWF 文件中的标签是如何组成和分类的？
12. 有哪些常见的视频文件格式？分别有什么特点？
13. 有哪些常见的流媒体文件格式？分别有什么特点？

7.7 习题解答

1. 答：RIFF 全称为资源互换文件格式 (Resources Interchange File Format)，RIFF 文件是 windows 环境下大部分多媒体文件遵循的一种文件结构，RIFF 文件所包含的数据类型由该文件的扩展名来标识。RIFF 文件含有一个如下图所示的文件头结构。



2. 答：(1) 静止的图像可用一个矩阵来表示，矩阵列中的各个元素用来描述构成图像的各个点（称为像素 pixel）的强度与颜色等信息，这种图像也称为位图 (Bitmap)。
- (2) 对于黑白图像，每个像素用一个字节数据来表示，而在彩色图像中，每个像素需用三个字节数据来表述。彩色图像可以分解成红(R)、绿(G)、蓝(B)三个单色图像，任何一种颜色都可以由这三种颜色混合构成。
3. 答：(1) 所谓调色板，就是在低颜色深度的模式下，在有限的像素值与 RGB 颜色之间建立对应关系的一个线性表。
- (2) 存在三种调色板，硬件调色板，逻辑调色板，系统调色板。
- (3) 硬件调色板就是显卡适配器所能实际表达的颜色深度，逻辑调色板就是通过调色板管理机制为每个窗体应用程序分配的调色板 (系统调色板只有一个，而逻辑调色板可以有多个，它的本质就是一块内存中的区域用于描述当前应用使用到的调色板，我们都

知道调色板是一个结构),所以逻辑调色板的用途在于模拟硬件调色板,以使 windows 作为操作系统可以为界面显示。

4.答:(1) BMP 文件结构可以分成位图文件头、位图信息头、颜色表(调色板数据)和图像数据四个部分。

(2) BMP 文件的数据按照从文件头开始的先后顺序分为四个部分:

位图文件头(bitmap file header): 提供文件的格式、大小等信息

位图信息头(bitmap information): 提供图像数据的尺寸、位平面数、压缩方式、颜色索引等信息

调色板(color palette): 可选,如使用索引来表示图像,调色板就是索引与其对应的颜色的映射表

位图数据(bitmap data): 图像数据区



5.答: BMP(全称 Bitmap)是 Window 操作系统中的标准图像文件格式,可以分成两类:设备相关位图(DDB)和设备无关位图(DIB),使用非常广。

位图文件头(bitmap-file header): 包含 BMP 图像文件的类型、显示内容等信息

位图信息头(bitmap-information header): 包含有 BMP 图像的宽、高、压缩方法,以及定义颜色等信息;

彩色表/调色板(color table): 这个部分是可选的,有些位图需要调色板,有些位图,比如真彩色图(24 位的 BMP)就不需要调色板;

位图数据(bitmap-data): 这部分的内容根据 BMP 位图使用的位数不同而不同,在 24 位图中直接使用 RGB,而其他的小于 24 位的使用调色板中颜色索引值。

位图文件头结构如下:

字段名	长度(bit)	描述
bfType	2	位图类别,根据不同的操作系统而不同,在 Windows 中,此字段的值总为 'BM'
bfSize	2	BMP 图像文件的大小
bfReserved1	2	总为 0
bfReserved2	2	总为 0
bfOffBits	4	BMP 图像数据的地址

位图信息头结构如下:

字段	长度(bit)	描述
biSize	4	本结构的大小,根据不同的操作系统而不同,在 Windows 中,此字段的值总为 28h 字节=40 字节
biWidth	4	BMP 图像的宽度,单位像素
biHeight	4	总为 0
biPlanes	2	总为 0

biBitCount	2	BMP 图像的色深，即一个像素用多少位表示，常见有 1、4、8、16、24 和 32，分别对应单色、16 色、256 色、16 位高彩色、24 位真彩色和 32 位增强型真彩色
biCompression	4	压缩方式，0 表示不压缩，1 表示 RLE8 压缩，2 表示 RLE4 压缩，3 表示每个像素值由指定的掩码决定
biSizeImage	4	BMP 图像数据大小，必须是 4 的倍数，图像数据大小不是 4 的倍数时用 0 填充补足
biXPelsPerMeter	4	水平分辨率，单位像素/m
biYPelsPerMeter	4	垂直分辨率，单位像素/m
biClrUsed	4	BMP 图像使用的颜色，0 表示使用全部颜色，对于 256 色位图来说，此值为 100h=256
biClrImportant	4	重要的颜色数，此值为 0 时所有颜色都重要，对于使用调色板的 BMP 图像来说，当显卡不能够显示所有颜色时，此值将辅助驱动程序显示颜色

6.答：GIF 数据块可分成 3 类，包括控制块(Control Block)，图形描绘块(Graphic-Rendering Block)和专用块(Special Purpose Block)。控制块包含有用来控制数据流(Data Stream)或者设置硬件参数的信息，其成员包括：GIF 文件头(Header)、逻辑屏幕描述块(Logical Screen Descriptor)、图形控制扩展块(Graphic Control Extension)、文件结束块(Trailer)。图形描绘块包含有用来描绘在显示设备上显示图形的信息和数据，其成员包括、图像描述块(Image Descriptor)、无格式文本扩展块(Plain Text Extension)、全局调色板、局部调色板、图像压缩数据和图像说明扩充块。特殊用途数据块包含有与图像处理无关的信息，其成员包括注释扩展块(Comment Extension)、应用扩展块(Application Extension)，除了在控制块中的逻辑屏幕描述块(Logical Screen Descriptor)和全局彩色表(Global Color Table)的作用范围是整个数据流(Data Stream)之外,所有其他控制块仅跟在它们后面的图形描绘块。

特点：体积小，像素低，色彩差，通用性强，可以做动画；

7.答：一个由独立 JPEG 小组 (Independent JPEG Group) 所建立的额外标准，称为 JFIF (JPEG File Interchange Format, JPEG 档案交换格式)

JPEG 文件中重要的标记：这里仅列出几个常用标记的标记代码、长度和表示的意义。
SOI, Start of Image, 图像开始，2 字节，固定值 0xFFD8。

8.答：PNG 格式图片特点：使用彩色查找表或者叫做调色板可支持 256 种颜色的彩色图像。
流式读/写性能：图像文件格式允许连续读出和写入图像数据。逐次逼近显示：这种特性可使在通信链路上传输图像文件的同时就在终端上显示图像,把整个轮廓显示出来之后逐步显示图像的细节。透明性：这个性能可使图像中某些部分不显示出来，用来创建一些有特色的图像。辅助信息：这个特性可用来在图像文件中存储一些文本注释信息。
独立于计算机软硬件环境。使用无损压缩。

9. 答：SVG 指可伸缩矢量图形 (Scalable Vector Graphics)，用来定义用于网络的基于矢量的图形。SVG 图像具有以下特点：1) 在放大或改变尺寸的情况下其图形质量不会有所损失；2) 可被非常多的工具读取和修改 (比如记事本)；3) SVG 与 JPEG 和 GIF 图像比起来，尺寸更小，且可压缩性更强；4) SVG 图像中的文本是可选的，同时也是可搜索的 (很适合制作地图)；5) SVG 可以与 JavaScript 技术一起运行；6) SVG 文件是纯粹的 XML；7) SVG 与其他标准 (比如 XSL 和 DOM) 相兼容。

10. 答：常见的动画文件格式有 GIF、Flic、SWF 等。

11. 答：SWF 文件头结构见下表：

字段	长度(bit)	说明
头标记	8	“F” 表示未压缩，“C”表示已压缩
头标记	8	此标识通常为 “W”
头标记	8	此标识通常为 “S”
版本	8	版本号不是 ASCII 字符，而是 8 位数字
文件长度	32	整个文件的字节长度。包括文件头整个文件的总长度
帧大小	16	表示影片的宽度和高度，存在一个 RECT 结构中，影片大小可以根据坐标（四个点的坐标）数值的变化而变化
帧速率	16	表示理想的每秒播放帧数
帧数	16	影片的总帧数

SWF 整体结构为：[Header][FileAttributes Tag][任意标签][任意标签][任意标签].....[End Tag]。SWF 标签位于头部信息之后，SWF 标签按 长度分为 2 种，一种是短标签，一种是长标签。短标签可以用 2 个字节即 UI16 就可以代表完，长标签要用 6 个字节即 UI16 + SI32 代表。FileAttributes Tag 标签固定为第一个标签 End Tag 标签固定为最后一个标签表。

12. 答：常见视频格式及特点如下：

- (1) AVI 视频格式，特点是图像质量好，可以跨多个平台使用，允许视频和音频交错在一起同步播放，支持 256 色和 RLE 压缩，但是体积过于庞大，而且为限定压缩标准。
- (2) MOV 视频格式，特点是有较高的压缩比率和完美的视频清晰度，其最大的特点是跨平台性。
- (3) MPEG 视频格式，特点是压缩率高但清晰度损失比较大
- (4) RM 视频格式，特点是可以根据网络数据传输速率的不同而采用不同的压缩比率，而且可以实现在线播放。
- (5) RMVB 视频格式，特点是及可以保证静止画面质量又大幅度提高了运动图像的画面质量。而且具有内置字幕和无需外挂插件支持的特点。
- (6) ASF 视频格式，特点是本地或网络回放、可扩充的媒体类型、部件下载以及扩展性。
- (7) FLV 视频格式，特点是可以轻松地导入到 Flash 中，同时可以通过 RTMP 协议从 Flashcom 服务器上流式播出，速度极快。

13. 答：(1) rm、rmvb 格式和 ra 音频格式主要用来在低速率的网络上实时传输活动视频影像。可以根据网络数据传输速率的不同而采用不同的压缩比率，并在数据传输过程中边下载边播放视频影像，从而实现影像数据的实时传送和播放。客户端是通过 Real Player 播放器进行播放的。rmvb 格式是由 rm 格式升级延伸而来的，vb 即 vbr，是 Variable Bit Rate(可变比特率)的英文缩写。在播放以往常见的 rm 格式电影时，可以在播放器左下角看到 225kbps 字样，这就是比特率。影片的静止画面和运动画面对压缩采样率的要求是不同的，如果始终保持固定的比特率，会对影片质量造成浪费。而 rmvb 则打破了原先 rm 格式那种平均压缩采样的方式，在保证平均压缩比的基础上，设定了一一般为平均采样率两倍的采样率值。

(2) asf 格式也是一种流行的网上流媒体格式。它是微软为了和现在的 Real Player 竞争而推出的一种视频格式，使用 MPEG-4 的压缩算法，压缩率和图像的质量都很不错。因为 asf 以“流”格式存在，所以它的图像质量比 VCD 差，但比同是视频“流”格式的 rm 格式要好。这种流式文件的使用与 Windows 操作系统是分不开的，使用的播放器是 Microsoft Media Player。wmv 的全称是 Windows Media Video，也是微软推出的一种采

用独立编码方式并且可以直接在网上实时观看视频节目的文件压缩格式，是 **asf** 格式的升级和延伸。在同等视频质量下，**wmv** 格式的体积非常小，因此很适合在网上播放和传输。**wma** 的全称是 **Windows Media Audio**，它是微软公司推出的与 **mp3** 格式齐名的一种新的音频格式。在低比特率时，效果好过 **mp3**；在高比特率时，**wma** 作用不大。因压缩率高，**wma** 文件适宜于网络下载。

（3）QuickTime 的 **qt** 格式是 **Apple** 公司开发的一种音频、视频文件格式。该格式具有先进的音频和视频功能，有包括 **Apple Mac OS**，**Microsoft Windows 95/98/NT/2000**，**Linux** 在内的所有主流计算机操作系统的支持。**QuickTime** 文件格式支持领先的集成压缩技术，提供 150 多种视频效果，并配有提供了 200 多种 **MIDI** 兼容音响和设备的声音装置。能够通过 **Internet** 提供实时的数字化信息流、工作流与文件回放功能。

（4）Flash 的 **swf** 格式是 **Macromedia** 公司的流式动画格式。由于其体积小，功能强、交互能力好等特点，越来越多地应用到网络动画中，流行于 **Internet** 上，在客户端安装 **Flash Player** 插件即可播放。

第 8 章 数字水印技术

8.7 习题

1. 什么是数字水印，数字水印主要可分成几类？
2. 举例说明数字水印的主要用途。
3. 简单描述数字水印的基本思想和基本特征，以及视频数字水印有哪些特有性质。
4. 简述数字水印的嵌入和提取过程。

8.7 习题解答

1. 答：数字水印是永久镶嵌在其他数据(宿主数据)中具有可鉴别性的数字信号或模式，而且并不影响宿主数据的可用性。数字水印的分类：1)按承载数字水印的载体可划分为：图像水印(Image watermarking)、视频水印(Video watermarking)、音频水印(Audio watermarking)、文档水印(Text watermarking)、软件水印(Software watermarking)；2)按感知特性划分：不可感知水印(不可见水印)、感知水印(可见水印)。3)按水印的抗攻击能力划分：鲁棒(Robust)水印、脆弱(Fragile)水印、半脆弱(Semi-Fragile)水印。4)按水印的嵌入域划分：时/空间域水印、变换域(频率域)水印。5)按数字水印的内容划分：有意义水印、无意义水印。6)按水印检测方式划分：非盲(Non-blind)检测水印、盲(Blind)检测水印、半盲(Semi-blind)检测水印。7)按水印的用途划分：版权保护水印、票证防伪水印、认证水印、隐蔽标识水印。
2. 答：数字水印主要应用领域：版权保护、广播监控、防止非法拷贝、数字指纹、内容验证、叛逆者追踪、元数据嵌入。
3. 答：数字水印基本思想是将一些标识信息直接嵌入数字载体当中(包括多媒体、文档、软件等)或是间接表示(修改特定区域的结构)，且不影响原载体的使用价值，也不容易被探知和再次修改。但可以被生产方识别和辨认。通过这些隐藏在载体中的信息，可以达到确认内容创建者、购买者、传送隐秘信息或者判断载体是否被篡改等目的。基本特征：不可感知性、不可感知性、水印容量、鲁棒性、可证明性、安全性。视频水印的特点：1)视频信息量大，提取水印时需要实现盲检测。2)虽然视频容量大，但视频水印经常有实时要求，要达到算法的实时性，必须降低算法复杂度。3)攻击特殊：如帧平均、帧剪切、帧重组、丢帧、帧率改变等。4)要与视频编码标准相结合。
4. 答：水印的嵌入：使用嵌入算法 E 把水印信号 W 嵌入到原始产品 I 中。水印的提取：在水印载体中精确提取水印或通过相关检测判断其是否包含某一水印。

第9章 图像和视频质量的评价方法

9.6 习题

1. 请解释对比敏感度函数和掩盖效应在图像与视频质量评价中的作用。
2. 请给出图像与视频质量主观评价方法的原理和步骤，并说明主观质量评价方法的优缺点。
3. 请给出均方误差和峰值信噪比的原理和计算公式，并说明其优缺点。
4. 请给出结构相似性测量的原理和计算公式，并说明其优缺点。多尺度结构相似性相比单尺度结构相似性有哪些好处？
5. 信号保真度与信息保真度如何区分？
6. 什么是自然场景的统计特性？可以从哪些域进行建模？
7. 视频质量评价比图像质量评价难在哪里？
8. 运动信息加权的视频质量评价方法的流程和大概步骤有哪些？
9. 请在 Matlab 中编程实现 MSE 与 PSNR 的计算。
10. 请在 Matlab 中调试 SSIM 的程序，并使用测试图像测量其准确性与计算复杂度。

9.6 习题解答

1. 答：对比敏感度函数（CSF）和掩盖效应反映了人眼观察图像与视频时的感知规律，在 IQA 与 VQA 中，若能有效地考虑 CSF 和掩盖效应对观察质量的影响，将使质量评价结果更接近人眼评价的结果。
2. 答：1) IQA 与 VQA 的主观质量评价方法主要通过人工来观察图像，并对感知到的图像质量进行打分，最后对多人打分的结果进行统计平均，得到主观评价分。目前国际上通用的主观评价方法是 BT.500-3 建议中给出的 5 级打分法，对打分人员和过程要求较为严格。如常采用的双刺激连续质量尺度法（DSCQS），是由参考图像与相应的失真图像组成图像对，观察者观看以随机顺序出现的一系列图像对（图像对中参考图像与失真图像出现的顺序是随机的，且观察者不被告知哪个是失真图像，以避免观察者打分时带有偏见），并对两者的质量都给出评价。
2) 主观评价方法的优点：评价结果最为准确、可靠，常作为基准来判断客观评价方法的准确性。但缺点是：耗费大量人力和时间，操作起来不够方便；评价会受到观察个人偏好的影响；只能做事后评价；无法自动质量评价，效率低下，无法作为质量目标来优化图像处理系统。
3. 答：均方误差（MSE）与峰值信噪比（PSNR）的原理是通过测量误差信号的强度对图像失真进行测量，其计算公式分别如下：

$$MSE = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N [f(i, j) - g(i, j)]^2,$$

$$PSNR = 10 \cdot \lg \left(\frac{L^2}{MSE} \right) = 10 \cdot \lg \left(\frac{L^2}{\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N [f(i, j) - g(i, j)]^2} \right).$$

MSE 与 PSNR 的优点在于：1) 具有明确的物理意义；2) 计算简单高效；3) 可微分，

便于数学运算和作为优化目标嵌入到图像处理系统中。

MSE 与 PSNR 的缺点在于：1) 忽视了图像信号的二维结构特性；2) 将误差信号与图像信号完全割裂开来；3) 没有反映人眼观察图像的特性，与人眼评价的一致性较低。

4. 答：1) 结构相似性 (SSIM) 的原理是根据人眼对图像中局部结构信息更加敏感的特性而提出来的图像质量评价方法，主要从亮度、对比度和结构信息上进行对比测量。其计算

公式如下：亮度 $l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$ ，对比度 $c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$ ，结构信息

$$s(x, y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \text{完整 SSIM 公式: } SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}。$$

2) 多尺度 SSIM 比单尺度 SSIM 的优点在于：模拟了不同距离对图像观察质量的影响，得到的质量评分更加稳定，与主观质量更加一致。

5. 答：信号保真度更接近信号底层表达，而较少考虑人对信号的理解；而信息保真度更接近中层及高层语义表达，更多地考虑人对信息的理解。
6. 答：1) 自然场景统计特性 (NSS) 是指自然场景中的图像与视频信号在所有可能的信号空间中仅占很小的一个子空间，这个子空间有一定的结构，可通过构建适当的模型及其统计特征进行描述。即自然无失真的图像在某些统计属性或特征参数上具有规律性，这种规律性不依赖于具体的图像内容。2) 可以从空间即像素域，或变换域（如 DCT 域、小波域等）进行建模。
7. 答：VQA 比 IQA 更加困难，主要体现在以下三点：1) 视频信息比图像信号多了一维时域信息，视频信号本身的建模更加复杂；2) 视频中既存在空域失真又存在时域失真，这些失真之间还存在相互影响；3) 人眼对视频信号的感知机理比对静态图像的感知机理更为复杂。
8. 答：运动信息加权的 VQA 方法的大概流程如下：1) 使用图像质量评价方法计算视频各帧的空域质量；2) 基于参考视频使用运动估计或光流法计算相邻帧之间的运动信息；3) 根据运动信息计算块层视觉权重；4) 使用块层运动权重计算帧层视觉质量评分；5) 根据运动信息计算帧层视觉权重，并对视频所有帧的质量评分加权得到整个视频的质量评分。
9. 答：见附件中的示例代码文件，ex9_MSE_PSNR.m，将代码文件与参考及失真图像放入 Matlab 当前目录中，直接运行即可得到 MSE 与 PSNR 的结果。
10. 答：见附件中的示例代码文件，ex10_ssim_test.m，将代码文件与参考及失真图像放入 Matlab 当前目录中，直接运行即可得到 SSIM 与可视化结果。

第 10 章 基于内容的图像和视频检索

10.5 习题

1. 什么是基于内容的检索?“内容”的含义是什么?
2. 请解释查询、索引、检索、搜索这几个术语的概念。
3. 简述基于内容检索系统的一般结构、检索过程及特点。
4. 在基于内容检索系统中为什么要采用相似性查询?精确性查询能否做到?什么样的媒体可以做到精确查询?
5. 图像的特征有哪些?请比较颜色矩、颜色直方图、颜色集在描述颜色特征上的异同点。
6. 常见的基于内容的图像检索方法有哪些?
7. 请解释帧、关键帧、镜头、场景的概念。
8. 基于内容的视频检索涉及哪些关键技术?

10.5 习题解答

1. 基于内容的检索:是指对多媒体数据(如视频、图像和音频等)所蕴含的物理的和语义的内容进行计算机分析理解,其本质是对无序的多媒体数据结构化,进而提取语义信息,这些物理特征和语义信息有助于用户从大量存储在数据库中的媒体中检索出具有相似特征的媒体数据。

内容:多媒体信息的“内容”表示含义、要旨、主题、特征、物理细节等。其在层次上可以分为:概念级内容,感知特征,逻辑关系,信号特征,特定领域的特征。

2. 查询:是用户通过设置某些查询条件,从表或其他查询中选取全部或者部分数据,以表的形式显示数据供用户浏览。

索引:对特征库的快速访问。在关系数据库中,索引是一种单独的、物理的对数据库表中一列或多列的值进行排序的一种存储结构,它是某个表中一列或若干列值的集合和相应的指向表中物理标识这些值的数据页的逻辑指针清单。

检索:是在索引支持下的快速信息获取方式。

搜索:常用于 Internet 的搜索引擎,含有搜寻的意思,又有在大规模信息库中搜寻自己所需的信息的含义。

3. 基于内容检索系统的一般结构:系统在体系结构上分为两个子系统:特征库生成子系统和查询子系统。在特征生成库子系统中,对插入的新媒体进行分割,目标标识,以便针对目标进行特征提取;接着,对用户或系统标识的媒体对象进行特征提取处理;然后,将媒体数据和插入时得到的特征数据分别存入媒体数据库和特征数据库。媒体数据库包含各种媒体数据,特征数据库包含相应媒体的特征数据;在查询子系统中,用户通过查询接口提交查询,查询接口利用检索引擎得到近似的查询结果,再通过索引/过滤器对从特征库和媒体库中检索出的结果进行过滤,最后将检索结果返回给用户。

检索过程:

- (1)、提交查询
- (2)、相似性匹配
- (3)、调整特征
- (4)、重新搜索

特点：

- (1)、直接从媒体内容中提取特征并建立索引。
- (2)、相似性检索。
- (3)、满足用户多层次的检索要求。
- (4)、大型数据库(集)的快速检索。

4. (1)、相似性查询是指将特征与特征库中的特征按照一定的匹配算法进行匹配。满足一定相似性的一组候选结果按相似度大小排列返回给用户。从通用多媒体信息管理的角度出发，基于内容检索更需要的是不限定领域知识的方法。在许多检索方法中，不限定领域的颜色、纹理和镜头等内容的检索就是如此。但这些只能做到在大规模数据库中的“粗查”，更细致的检索和查询，可以留给基于领域知识的方法，也可以留给人去做。因此，在检索过程中，相似性查询是主要的。

(2)、多媒体基于内容的分析、处理与检索的研究还处于初级阶段，尚未达到真正实用的阶段，因此目前还不可以做到精确性查询。

(3)、对于传统数据库，如数字，文本等可以做到精确性查询。

5. 特征：颜色特征、纹理特征、形状特征、空间关系特征、对象特征

	相同点	不同点
颜色矩	都不涉及位置信息。	考虑了颜色的整体分布
颜色直方图		无须对特征进行矢量化
颜色集合		同时考虑了颜色空间的选择和颜色空间的划分

6. (1) 基于颜色特征的图像检索

(2) 基于纹理特征的图像检索

(3) 基于形状特征的图像检索

(4) 基于空间关系特征的图像检索

(5) 基于对象特征的图像检索

7. 帧：帧是组成视频的最小视觉单位，是指视频中空间上独立、时间上相关的一幅独立的图像。

关键帧：为了减小数据量，提高检索效率，需要从镜头中提取一定数量的视频帧来表达该镜头的内容，这种特殊的帧称为镜头关键帧。

镜头：指摄像机从打开到关闭过程一次连续拍摄所记录的帧序列。

场景：是由一组表达同一主题、语义相关的镜头组成。

8. 镜头边界检测：将一个视频数据中的镜头边界标记出来。

关键帧的提取：在视频分割为镜头的基础上，分析镜头中图像帧的颜色、纹理等特征，根据各帧之间的相互关系，找出最能代表镜头内容的图像帧。

镜头聚类：一般基于关键帧进行，提取关键帧的特征，并把关键帧特征用对应的特征空间点表示，通过将特征空间的点聚集成簇，然后得到镜头聚类的结果。

第 11 章 图像识别

11.7 习题

1. 请画出人工神经元模型，并简述其和生物神经元是如何类比的？
2. SVM 分类方法中，核函数的作用是什么？
3. 前馈神经网络有哪些特征？
4. 简述 BP 算法的基本思想。
5. 激活函数的作用是什么？画出 Sigmoid、Tanh、ReLU 激活函数的曲线。
6. 请画出 CNN 的网络结构，并阐述各个模块的作用。
7. 连续 Hopfield 网络和离散 Hopfield 网络的激活函数有什么区别？
8. 术语“深度学习”中的“深度”指什么含义？与浅层学习相比，深度学习有哪些不同？

11.7 习题解答

1. 人工神经元模型如下：

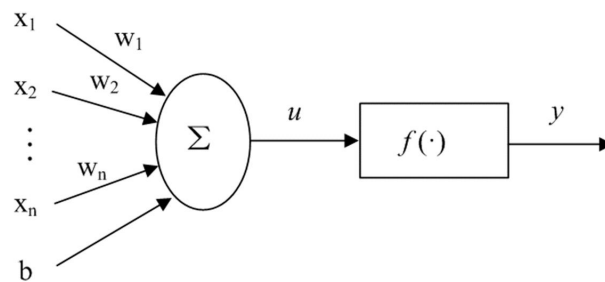


图 1 人工神经元结构模型

其中， $x_i(i=1,2,\dots,n)$ 表示来自第 i 个神经元的输入；权重值 $\omega_i(i=1,2,\dots,n)$ 表示第 i 个神经元的连接强度； b 为偏置； u 表示输入单元的整合量； f 为神经元的激活函数； y 是整个神经元的最终输出。

每一个人工神经元有如下三个基本要素：

- 1) 连接强度。用来与其他神经元的连接，模拟生物神经元的突触。
 - 2) 求和单元。计算当前神经元的所有输入信号的加权和。模拟生物神经元接收其他神经元的电位作用。
 - 3) 激励函数（传递函数）。用来将加权信号映射为输出信号。模拟生物神经元输出对输入的非线性关系。
2. 核函数将非线性的、不可区分的数据转换到一个高维空间，使数据在这个高维空间中实现分类。
 3. 前馈神经网络的每个神经元接受前一级输入，并输出到下一级，层间无反馈。

前馈神经网络结构简单，应用广泛，能够以任意精度逼近任意连续函数及平方可积函数，而且可以精确实现任意有限训练样本集。

从系统的观点看，前馈网络是一种静态非线性映射。通过简单非线性处理单元的复合映射，可获得复杂的非线性处理能力。

从计算的观点看，缺乏丰富的动力学行为。大部分前馈网络都是学习网络，其分类能力和模式识别能力一般都强于反馈网络。

4. BP 算法的基本思想是根据输出层的误差逐层反向估计隐含层的输出误差，进而调整每个隐含层的连接权重。

5. 激活函数将输入与输出的线性映射关系转变为非线性映射关系。

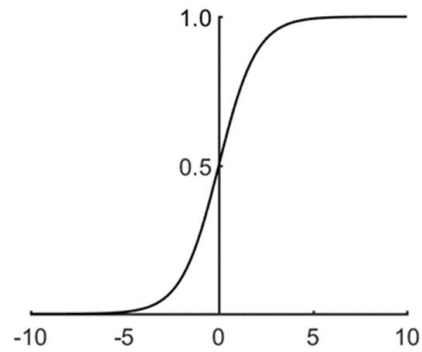


图 2 Sigmoid

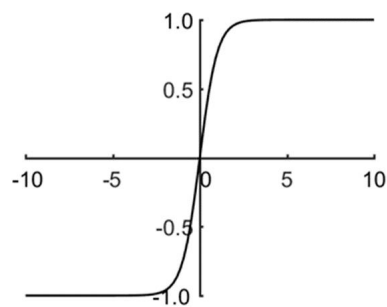


图 3 Tanh

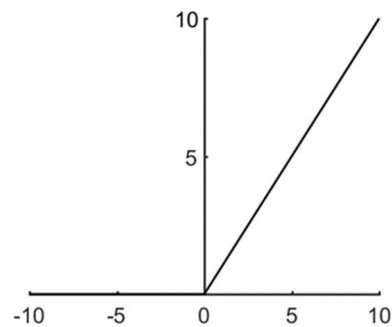


图 4 ReLU

6. CNN 网络结构如下：

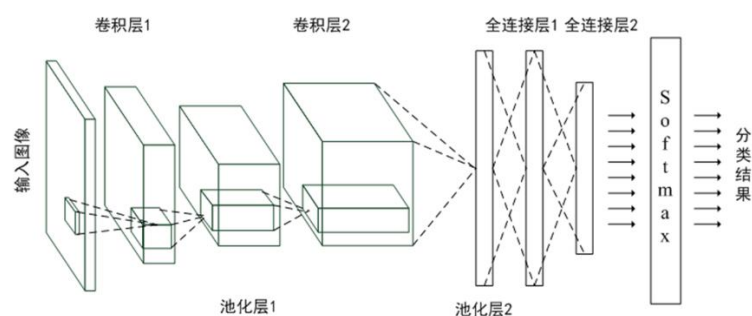


图 5 CNN 的网络结构

输入层：送入网络的输入数据。

卷积层：特征提取。

池化层：保留有用信息的基础上减少数据的处理量，加快网络训练速度，使特征更加鲁棒且具有平移不变性。

全连接层：完成分类任务。

输出层：获得输出结果（概率分布等）。

7. 连续 Hopfield 网络的激活函数是 S 型的连续函数，离散 Hopfield 网络的激活函数为阶跃函数。

8. 深度学习之所以被称为“深度”，是相对 SVM、提升算法、最大熵方法，以及只含单隐层的多层感知器等“浅层学习”方法而言，其实质是通过搭建具有多个隐层的学习模型，给其输入海量的训练数据，使其从训练数据中学习获得有用的特征，从而最终提升分类或预测的准确性。

不同于传统的浅层学习，深度学习具有以下特点：

- 1) 模型有多个隐层，一般比较深，通常有几十层，甚至成百上千层。
- 2) 模型能够从训练数据中自主提取特征。浅层学习依靠人工经验抽取样本特征，网络模型学习后获得的是没有层次结构的单层特征，模型的输入是人工已经选取好的特征，模型只用来负责分类和预测。而深度学习通过对原始数据进行逐层特征变换，将数据在原空间的特征表示变换到新的特征空间，自动地学习得到层次化的特征表示，将原始输入逐层转化为浅层特征、中层特征、高层特征直至最终的任务目标。