

计算机视觉基础

章毓晋

清华计算机图书·译丛

Foundations of Computer Vision

计算机视觉基础

本书介绍计算机视觉的基础内容，比较侧重计算几何和目标检测方面。本书对图像网格的构建和叠加、德劳内三角剖分和沃罗诺伊镶嵌、多边形拼贴、图像拓扑等都有比较全面深入的介绍，并对图像结构给出了直观可视的描述。书中提供了大量相应的Matlab程序，可结合原理学习进行实验，以进一步加深理解并解决实际问题。

本书可作为信号与信息处理、通信与信息系统、电子与通信工程、模式识别与智能系统、计算机科学等学科大学高年级本科生或研究生专业课教材和教学参考书，还可供涉及图像和机器视觉技术的应用行业（如生物医学、电视广播、工业自动化、文档识别、机器人、电子医疗设备、遥感测绘、增强现实、智能交通和军事侦察等）的科技工作者和从业者参考和自学。

课件下载·样书申请



书圈

清华社官方微信信号



扫 我 有 惊 喜



定价：98.00元

清华计算机图书·译丛

计算机视觉基础

清华大学出版社



Springer

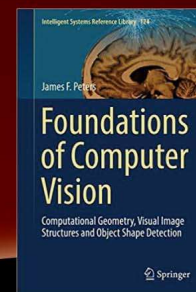
清华计算机图书·译丛

Foundations of Computer Vision

计算机视觉基础

[加] 詹姆斯·彼得斯 (James F. Peters) 著

章毓晋 译



清华大学出版社



第1章 通往机器视觉的基础知识

计算机视觉是利用计算机来实现人类视觉系统功能的一门学科

计算机视觉作为一门学科，与数学、物理学、生理学、感知心理学、神经科学以及计算机科学等都有密切的联系

机器视觉在很多情况下被看作计算机视觉的同义词，但机器视觉或机器人视觉更关注图像获取、系统构造和算法实现，侧重计算机处理图像数据时，利用电子设备和光学感知技术自动获取和解释场景的图像以控制机器的过程



1.1 什么是计算机视觉？

计算机视觉的主要目的是基于数码相机拍摄的图像内容重建和解释自然场景

基于光学传感器阵列的自然场景或者用相机获取的单幅图像，或者用摄像机获取的视频帧图像

每个图像场景的基本内容包括像素，边缘，角度，图像几何，颜色，形状，和纹理

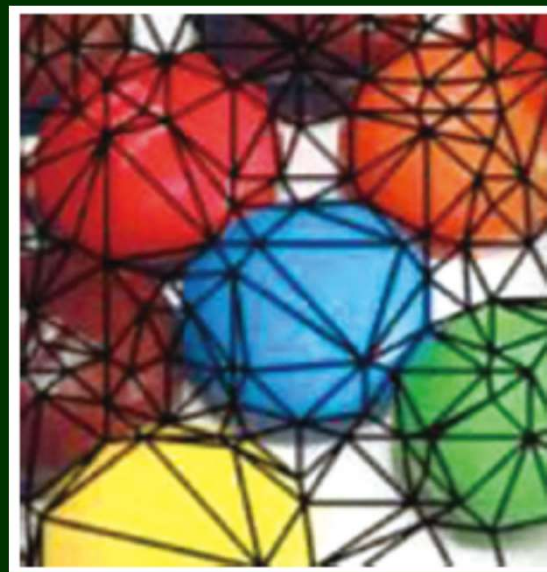


1.2 分而治之的方法

通过用已知几何形状，如三角形（德劳内三角剖分方法）和多边形（沃罗诺伊图方法），来拼贴（镶嵌）场景图像更容易重建和解释自然场景

例子：形状检测

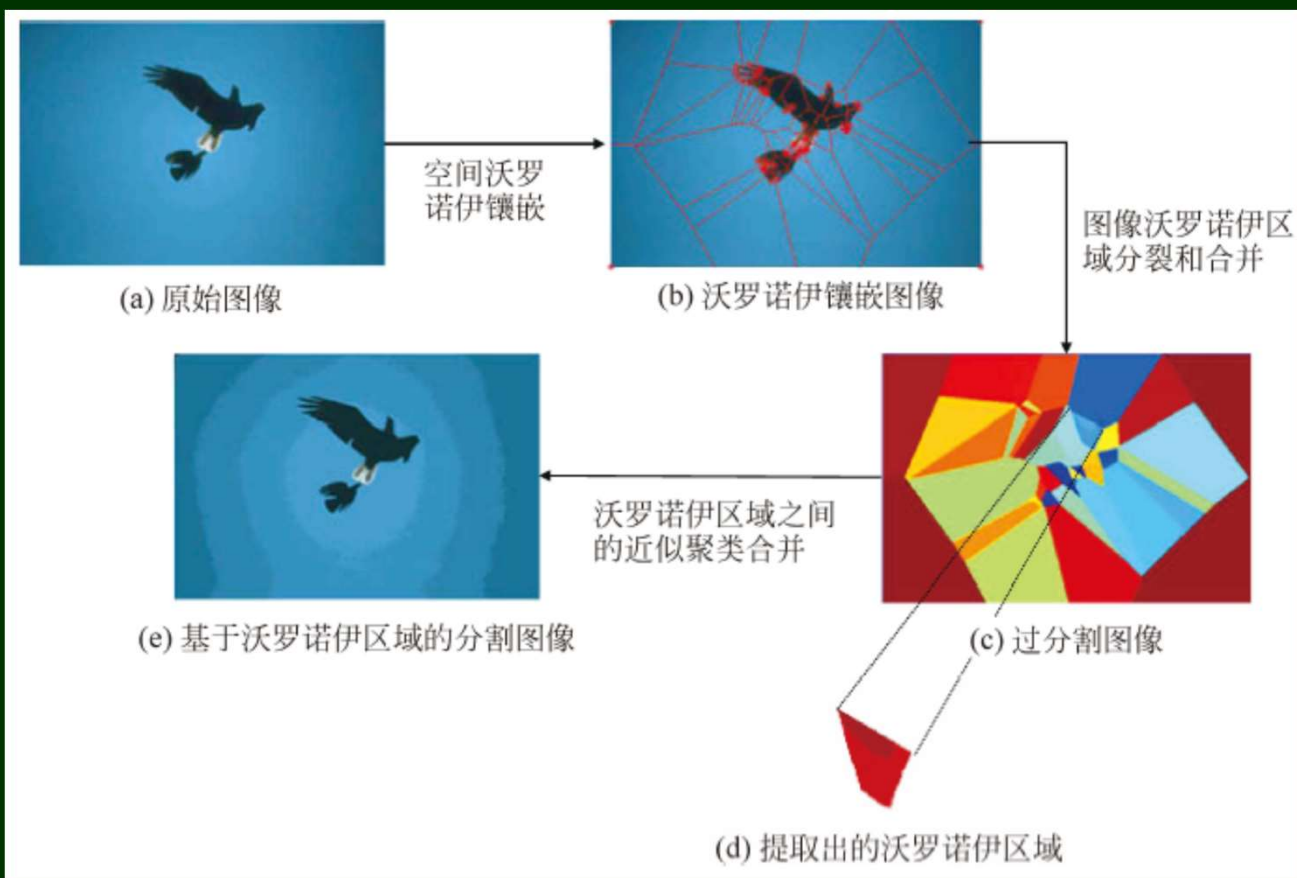
使用德劳内三角剖分的
视频帧形状检测





1.2 分而治之的方法

例子：视频点画





1.3 覆盖在图像上的沃罗诺伊图

设 S 是数字图像中的任何一组选定的像素，并且让 $p \in S$ 。 S 中的像素称为网点（或生成点）

对 $p \in S$ 的沃罗诺伊区域：

$$V_p = \{x \in E : \|x - p\| \leq \|x - q\| \text{ 对所有 } q \in S\}$$

S 中的每个网点只属于一个沃罗诺伊区域

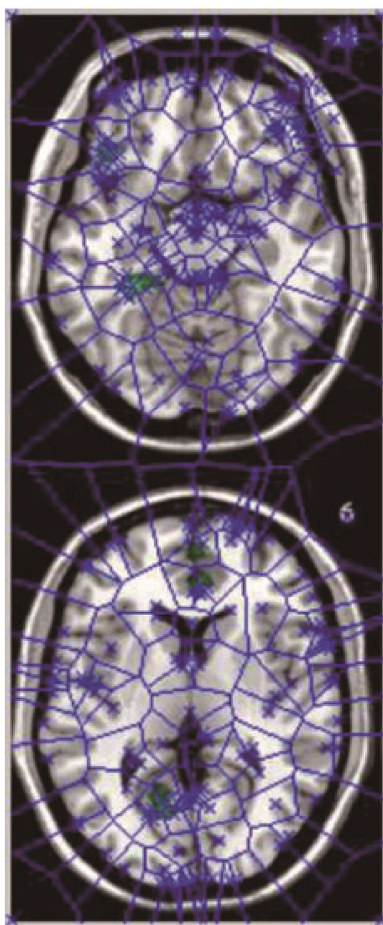
被沃罗诺伊区域覆盖的数字图像称为镶嵌图像

覆盖一幅图像的所有沃罗诺伊区域称为沃罗诺伊图或沃罗诺伊网格

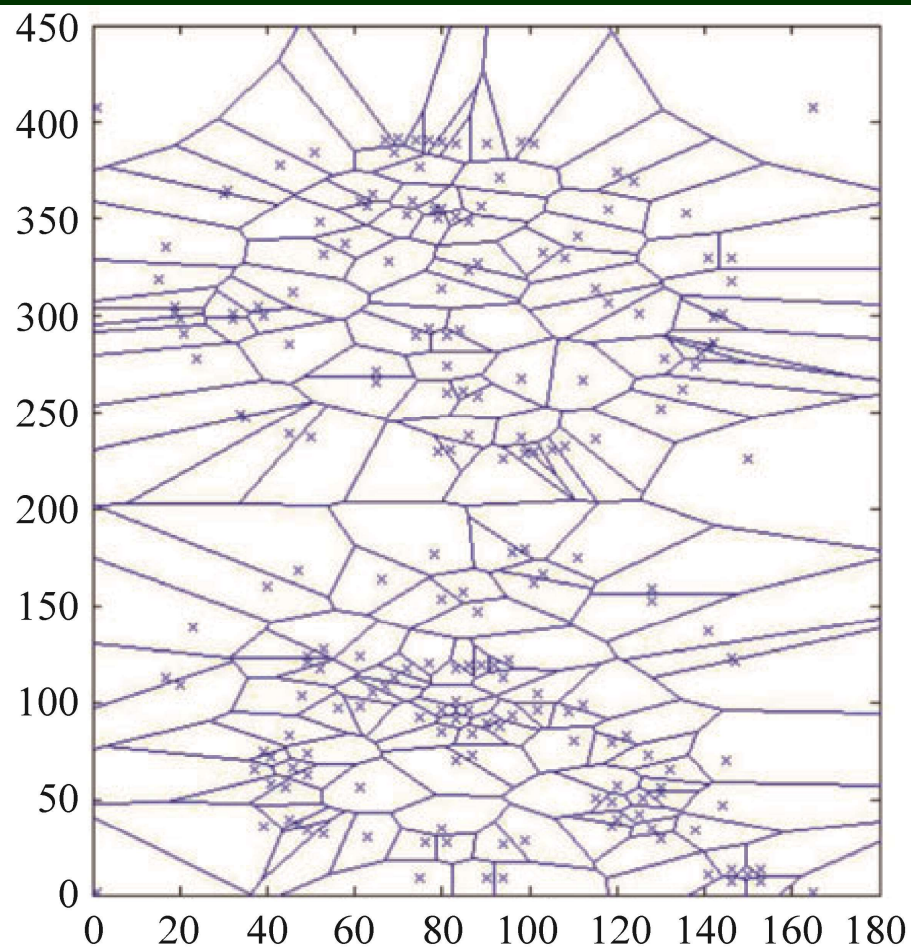


1.3 覆盖在图像上的沃罗诺伊图

图像
结构的
沃罗诺伊
几何
视图



(a) 功能磁共振沃罗诺伊

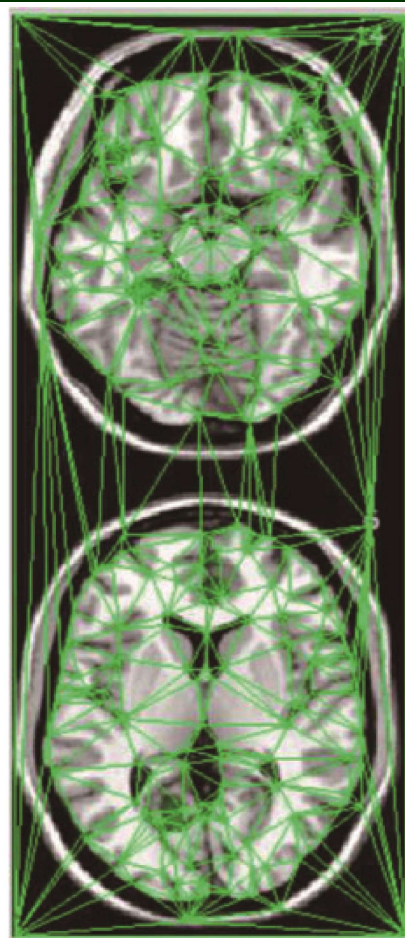


(b) 功能磁共振沃罗诺伊网格

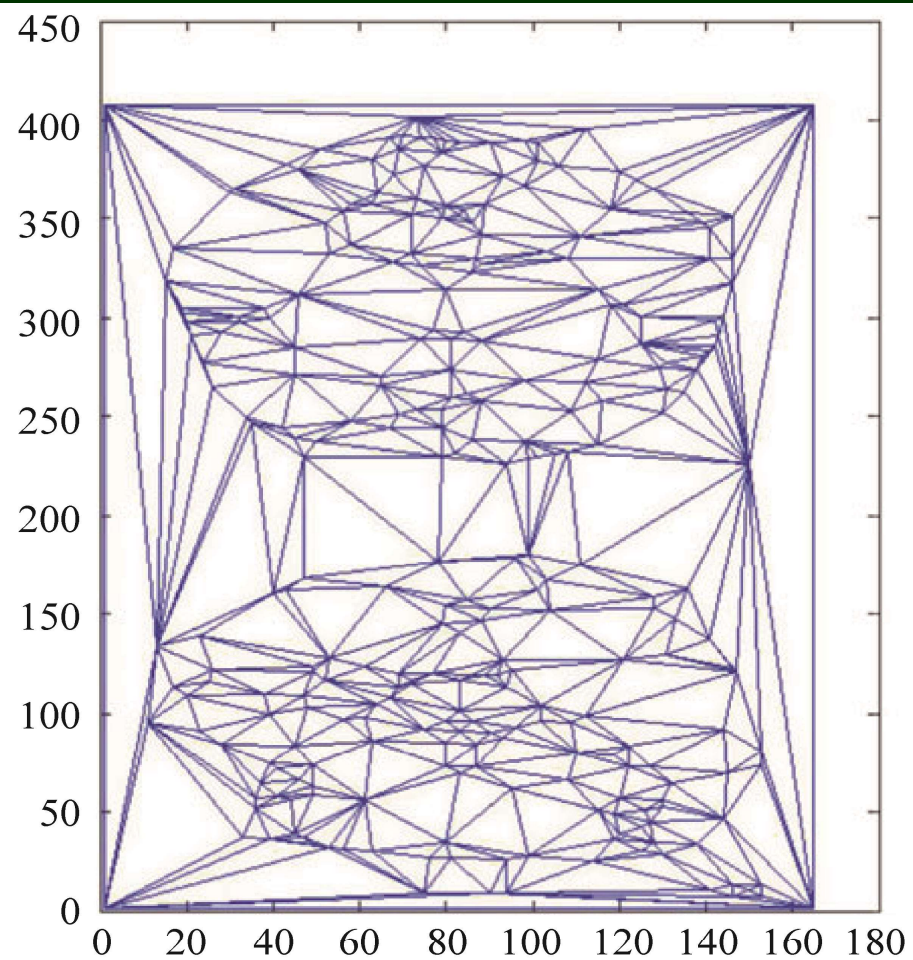


1.3 覆盖在图像上的沃罗诺伊图

图像
结构的
德劳内
几何
视图



(a) 功能磁共振德劳内



(b) 功能磁共振德劳内网格



1.5 数字图像的框架

数字图像包含了空间和强度信息

灰度数字图像由2-D光强度函数 $I(x, y)$ 表示

一个像素是光栅图像中的物理点

视场或视野是在特定时刻可见世界（摄像机前场景的一部分）的范围

人眼可以识别每度视弧120个像素，即如果2个点之间的夹角小于 $1/120$ 度，那么眼睛就分不清楚

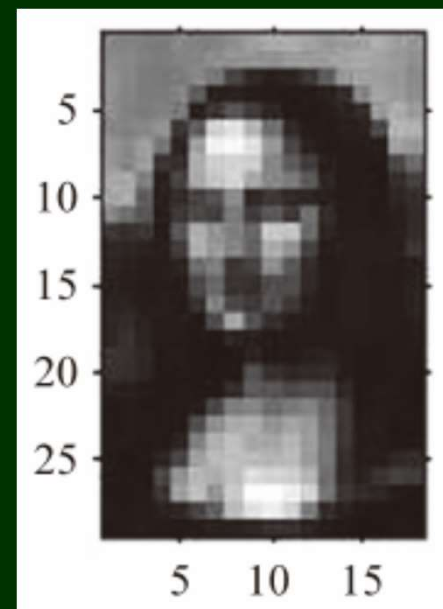


1.7 创建你自己的图像

任何在 $[0, n]$, $n \in N$ (自然数) 范围内的2-D自然数数组都可以被视为灰度数字图像。每个自然数指定一个像素强度。强度范围的上限 n 通常为255

Matlab在内部用微小的子图像代表一个单一的强度 (子图像中的每个像素具有相同的强度)

列表1.4 生成图1.14的
Matlab程序readTxt.m

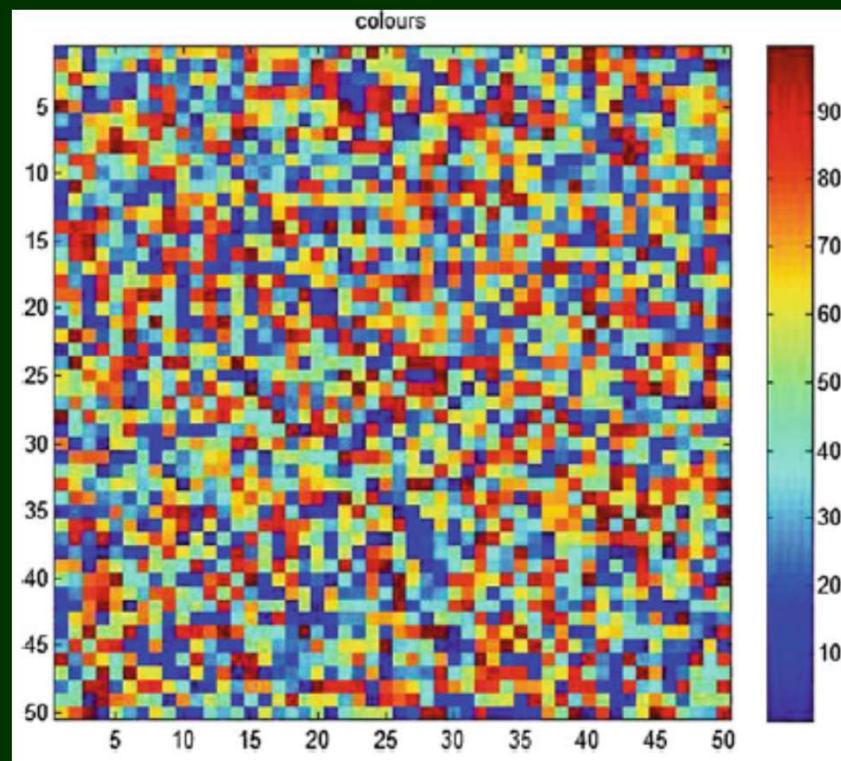




1.8 随机生成图像

使用随机数生成数字图像。这可用Matlab中的函数rand实现

函数image将矩阵
 I 显示为图像，矩阵 I
的每个元素指定一个
矩形块的彩色

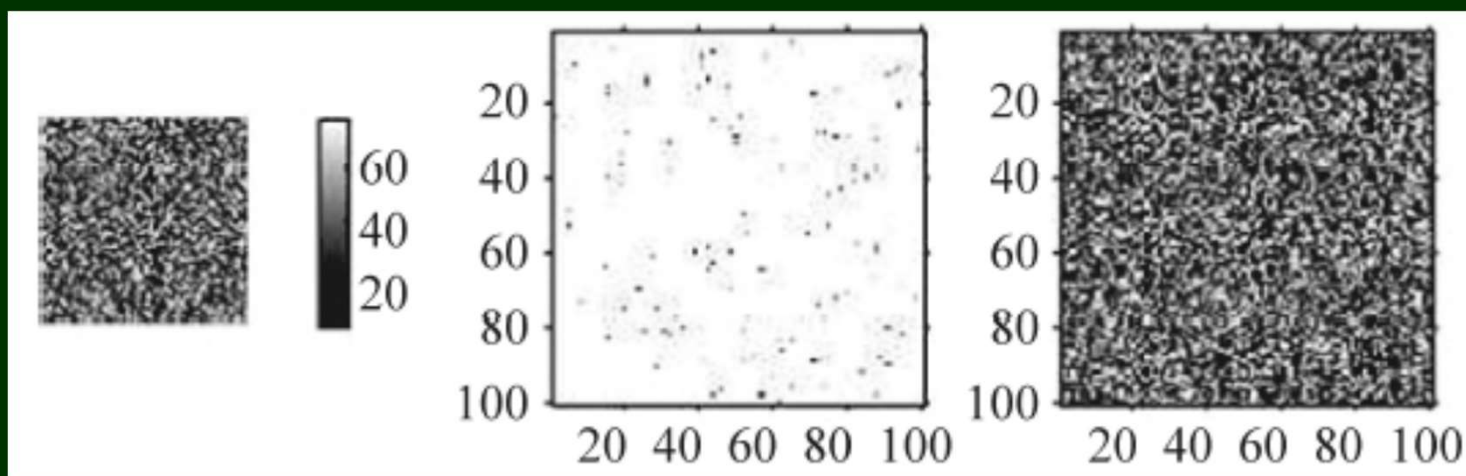




1.9 显示图像的方法

通过控制或不控制图像的缩放，可获得许多不同的图像显示方法

使用subplot(r, c, i)在*c*列中显示（或绘制）*r*行图像，其中 $1 \leq i \leq r * c$





1.10 数字图像的格式

- (1) .bmp (位图图片)
- (2) .gif (图形交换格式)
- (3) .jpg, .jpeg (联合摄影专家组)
- (4) .png (便携式网络图形)
- (5) .svg (可缩放矢量图形)
- (6) .tif, .tiff (标记图像文件格式)



1.11 图像数据类型

- (1) 二值（逻辑）图像
黑色对应0，白色对应1
- (2) 强度（灰度）图像：0到255的8比特数
- (3) 真彩色图像：每个像素分配3个数值
- (4) 假彩色图像：≠ 伪彩色
- (5) 浮点图像：试一下列表1.10的代码

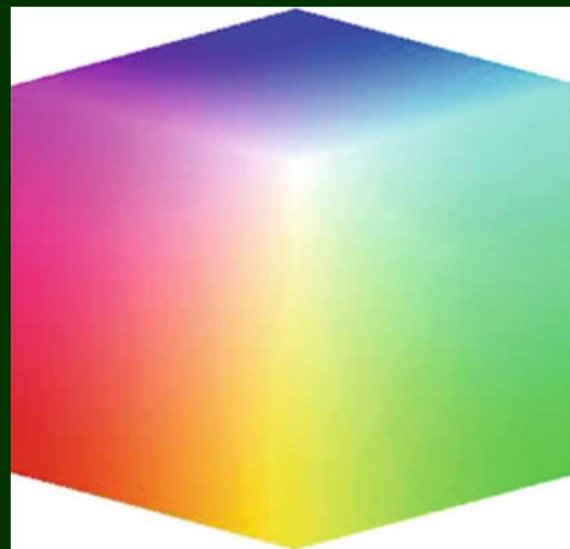


1.12 彩色图像

HSB（色调、饱和度、亮度）彩色空间是HSV（色调、饱和度、值）彩色空间的变型

彩色图像的彩色通道由三个不同的2-D数组表示，每种彩色对应一个数组

用列表1.11的Matlab代码生成的RGB彩色立方体





1.13 彩色查找表

彩色转换表或彩色查找表（LUT）将像素强度值（0到255）关联到彩色值

在Matlab中，LUT称为色彩图

基于色调、饱和度和值模型（HSV）或色调、饱和度和亮度（HSL）模型的彩色转换表有可能具有实数而不是整数的条目

一般总自动提供一个标准的或缺省的彩色表



1.14 图像几何初步

以下技巧用于查看单个像素，修改像素值和显示图像

技巧1：使用%表示注解

技巧2：Matlab = %赋值操作符

技巧3：Matlab `imread(image)` %将图像移入工作区

技巧4：Matlab `imtool(image)` %交互查看图像

技巧5：`I(x, y)` %在位置 (x, y) 显示像素值

技巧6：`I(x, y, :)` %显示RGB图像中的彩色通道值



1.14 图像几何初步

技巧7: $I(x, y)=255$ %将最大强度值赋给在位置
(x, y)的像素

技巧8: Matlab imshow()%显示图像

技巧9: $I(25, 50)$ %显示像素的彩色通道值

技巧10: $I(x, y, :) = 255$ %将各个彩色通道赋成白色

技巧11: Matlab line%在当前图像中画一条线段

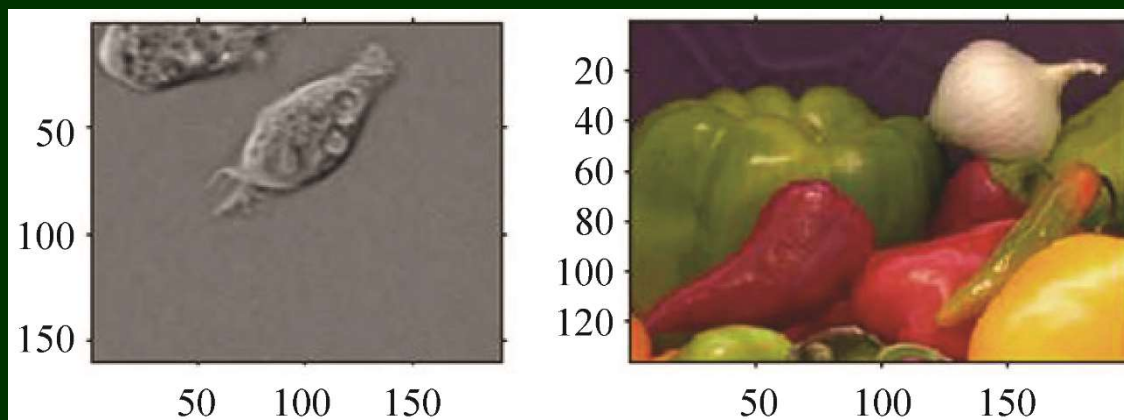
技巧12: Matlab improfile %沿图像中一条线段或
多条路径计算强度值



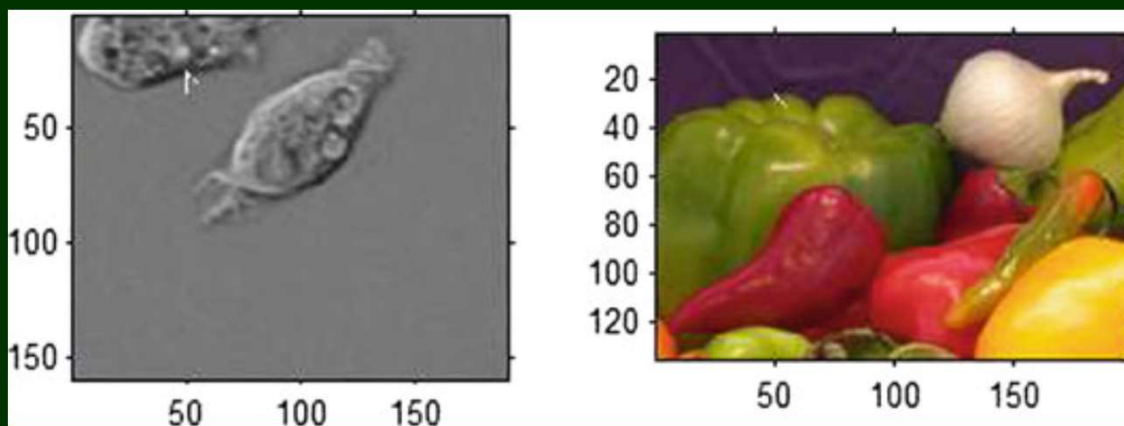
1.15 访问和修改图像像素值

使用列表1.15中的代码

改变像素值前的
图像显示



改变像素值后的
图像显示

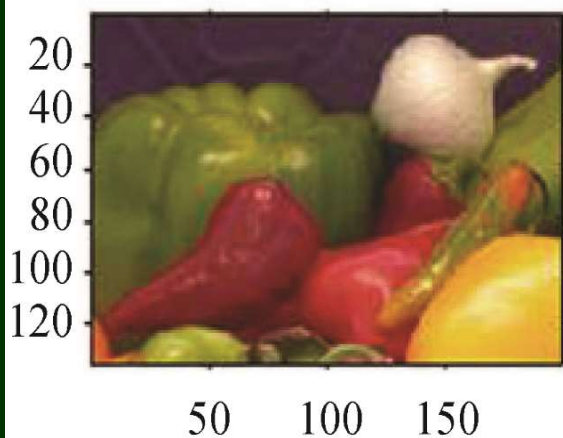




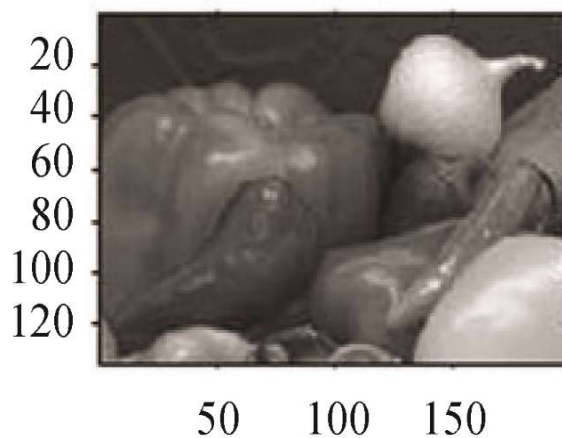
1.16 彩色、灰度和二值图像

使用列表1.16可生成

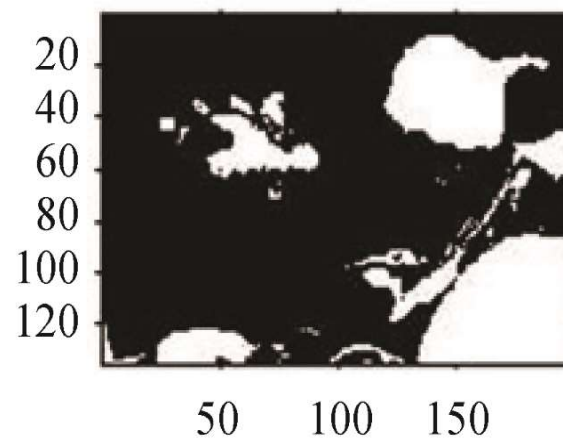
png(RGB)图像



灰度图像



二值图像



转换图像成为灰度图像和二值图像



1.17 像素的罗森菲尔德8-邻域

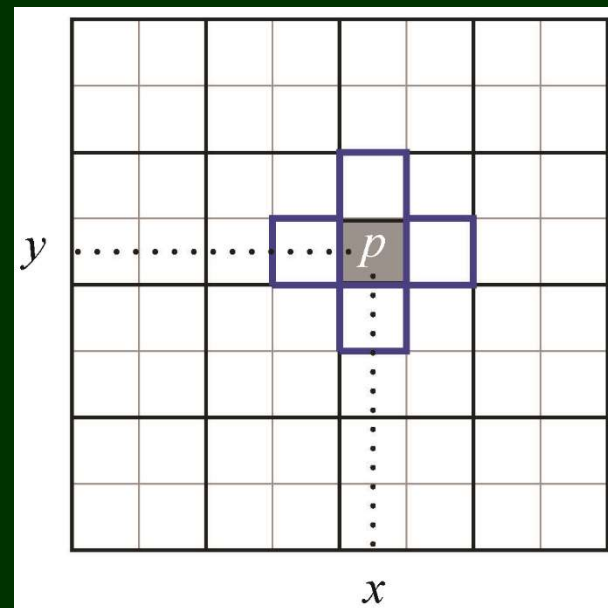
像素 p 的4-邻域, $N_4(p)$, 是与一个像素相邻的4个像素的集合。设 p 在图像 I 中的坐标为 (x, y)

$N_4(p)$ 是像素 p 的十字形邻域

$$N_4(p) = \{p(x, y), p(x-1, y), p(x+1, y), p(x, y-1), p(x, y+1)\}$$

设 B 由图中的角点像素构成

$$B = \{p(x-1, y-1), p(x+1, y+1), p(x+1, y-1), p(x-1, y+1)\}$$

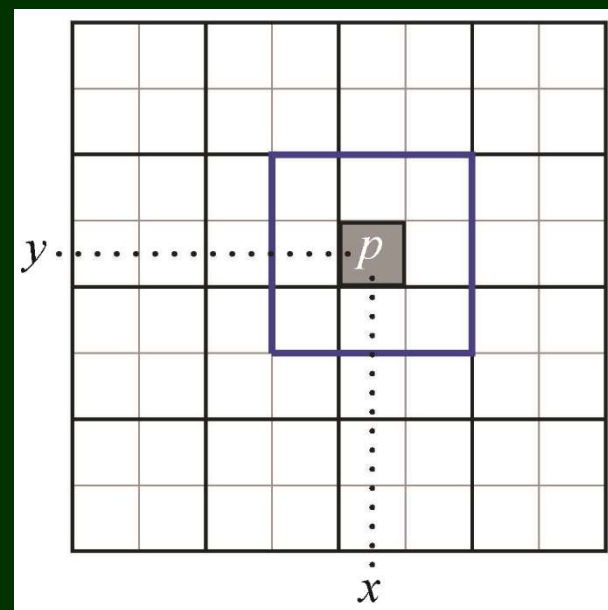




1.17 像素的罗森菲尔德8-邻域

像素 p 的罗森菲尔德**8-邻域**（记为 $N_8(p)$ ）定义为
 $N_4(p)$ 和 B 的并集

$$\begin{aligned} N_8(p) &= N_4(p) \cup B \\ &= N_4(p) \cup \{p(x-1, y-1), \\ &\quad p(x+1, y+1), p(x+1, y-1), \\ &\quad p(x-1, y+1)\} \\ &= \{p(x, y), p(x-1, y), p(x+1, y), \\ &\quad p(x, y-1), p(x, y+1)\} \cup \\ &\quad \{p(x-1, y-1), p(x+1, y+1), \\ &\quad p(x+1, y-1), p(x-1, y+1)\} \end{aligned}$$





1.18 距离：欧几里得和出租车测度

欧几里得距离度量

令点 $x, y \in \mathbb{R}^n$ 都有 n 个坐标，那么 $\mathbf{x} = (x_1, \dots, x_n)$ ， $\mathbf{y} = (y_1, \dots, y_n)$ 。 $\mathbf{x} \in \mathbb{R}^n$ 的范数（记为 $\|\mathbf{x}\|$ ）是：

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \quad (\text{从原点出发的矢量长度})$$

矢量 \mathbf{x} 和 \mathbf{y} 之间的距离是 $\mathbf{x} - \mathbf{y}$ 的范数（记为 $\|\mathbf{x} - \mathbf{y}\|$ ）

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^n (x_i^2 - y_i^2)} \quad (\text{欧几里得距离})$$



1.18 距离：欧几里得和出租车测度

曼哈顿距离度量（也称为出租车测度 d_{taxi} ）

出租车测度是沿着平面网格的垂直轴和水平轴使用点之间差异的绝对值来计算的：

$$d_{\text{taxi}} = |x_1 - x_2| + |y_1 - y_2| \quad (\mathbb{R}^2 \text{中两个矢量之间的出租车距离})$$

\mathbb{R}^n 中两点之间的出租车距离模仿了乘出租车沿着一条街走向另一条街，直至到达目的地所记录的距离

$$d_{\text{taxi}} = \sum_{i=1}^n |x_i - y_i| \quad (\mathbb{R}^n \text{中的出租车距离})$$



1.19 假彩色：点彩派绘画

点彩派绘画的方法使用假彩色修改数字图像中模式的外观

点彩派是一种绘画形式，其中将纯色小点应用于画布，使其在观察者的眼中混合

使用纯色小点的点彩绘画方法对数字图像中隐藏图案的所选像素进行假彩色着色



1.19 假彩色：点彩派绘画

假彩色RGB图像模式

- (1) 确定一幅数字图像 (**RGB**图像)
- (2) 选择要用假彩色绘制的图像模式
- (3) 确定一个假彩色绘制像素的方法
- (4) 确定方法中参数的值，如初始像素坐标
- (5) 使用步骤 (2) 中的方法
- (6) 假彩色步骤。如果彩色像素 q 的强度满足模式规则，那么最大化 q 的强度
- (7) 显示具有假彩色的图像
- (8) 重复步骤 (2) 显示不同的图像模式



1.19 假彩色：点彩派绘画

假彩色灰度图像模式

- (1) 确定一幅数字图像
- (2) 选择要用假彩色绘制的图像模式
- (3) 使用模式规则确定一个假彩色绘制像素的方法
- (4) 伪彩色图像生成步骤。灰度图像 \Rightarrow 伪彩色图像
- (5) 确定方法中参数的值，如初始像素坐标
- (6) 使用步骤(2)中的模式规则
- (7) 假彩色步骤。如果灰度像素 q 的强度满足模式规则，那么最大化 q 的伪彩色通道
- (8) 显示具有假彩色的图像
- (9) 重复步骤 (2) 显示不同的图像模式



1.20 数字图像上的矢量空间

矢量空间是一组可以加在一起并与数字相乘的对象或元素（任一操作的结果都是空间的元素）

点积

给定一对矢量 \mathbf{x} 和 \mathbf{y} ，点积（记为 $\mathbf{x} \bullet \mathbf{y}$ ）等于将 \mathbf{x} 投影到 \mathbf{y} 上的长度

$$\mathbf{x} \bullet \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta \quad (\text{点积})$$

矢量之间夹角

$$\theta = \arccos \left[\frac{\mathbf{x} \bullet \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \right] \quad (\text{矢量之间的夹角})$$



1.20 数字图像上的矢量空间

图像梯度

矢量（像素强度的位置）的梯度是矢量的斜率

f 在 (x, y) 位置处的梯度 (∇f) 被定义为2-D列矢量

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (f \text{ 在 } (x, y) \text{ 的梯度})$$

梯度 ∇f 指向在位置 (x, y) 处 f 的最大变化方向



1.21 相机看见什么：智能系统视图

智能意味着将拍照设备的能力与可用的传感器信息结合以便于捕获最佳图片

如果相机能选择场景中的一个或多个类似于脸部的物体，则相机可实现人脸检测

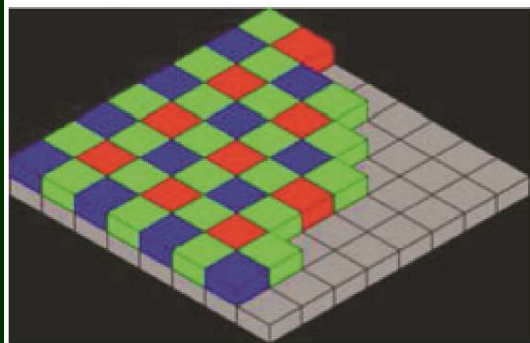
在人眼中，中央凹（视网膜的中央区域）提供高质量的视力，而眼睛的周边区域提供不太详细的成像。这与鱼眼镜头类似



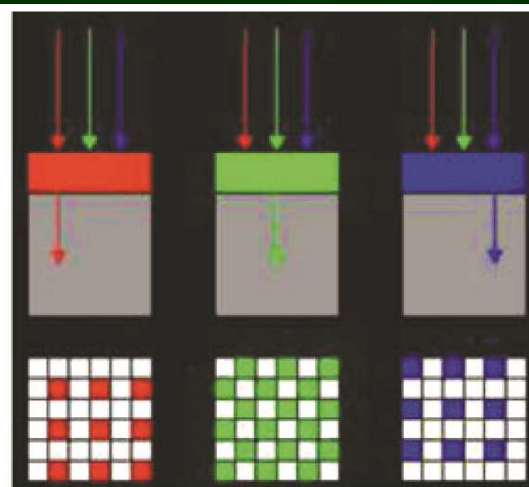
1.21 相机看见什么：智能系统视图

拜尔滤波器

绿色滤光片的数量约是蓝色和红色滤光片的两倍
图像传感器中的彩色排列



(a) 彩色滤波器



(b) 拜尔传感器阵列



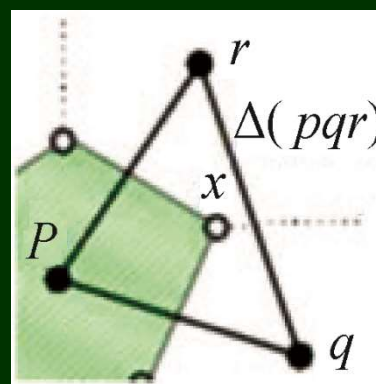
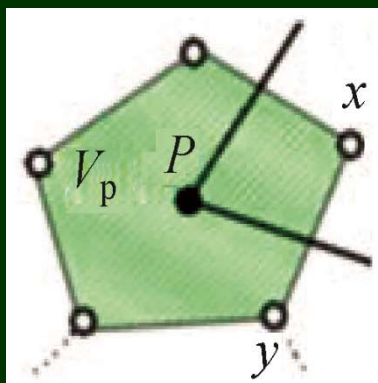
1.22 图像几何：图像上的沃罗诺伊和德劳内网格

创建覆盖在数字图像上的沃罗诺伊网格

设 $V(p)$ 为角点生成点 p 的图像沃罗诺伊区域

$V(p)$ 内部和沿着边缘的点都更接近生成点 p ，而不是任何其他生成点

沃罗诺伊区域



德劳内三角形



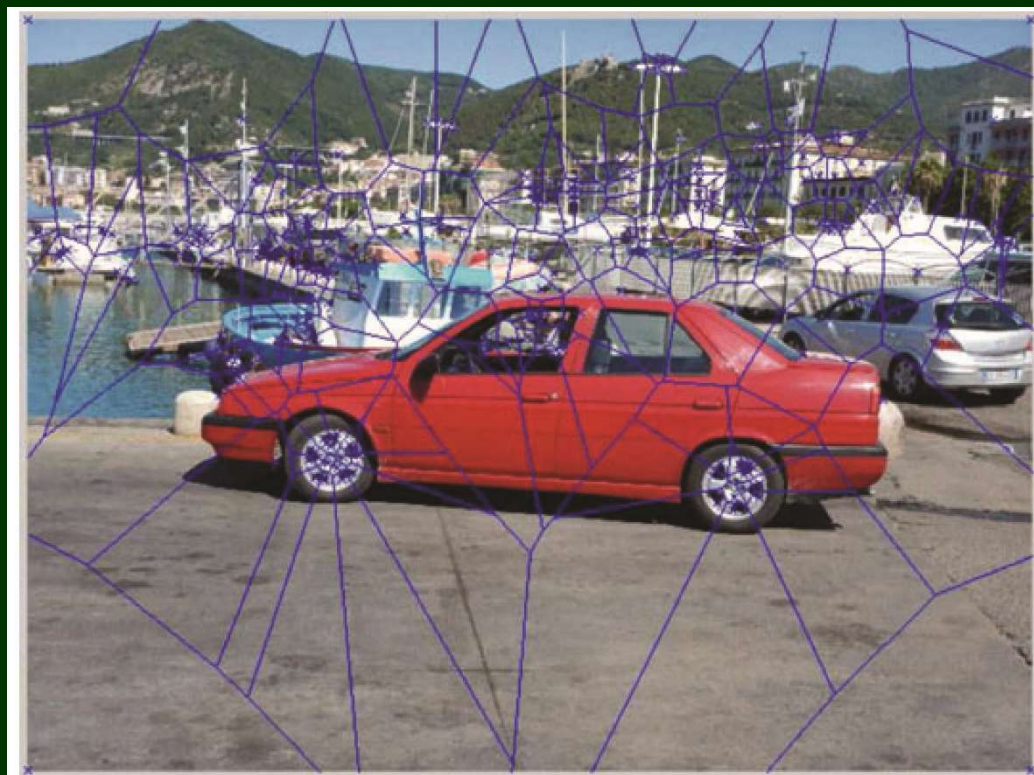
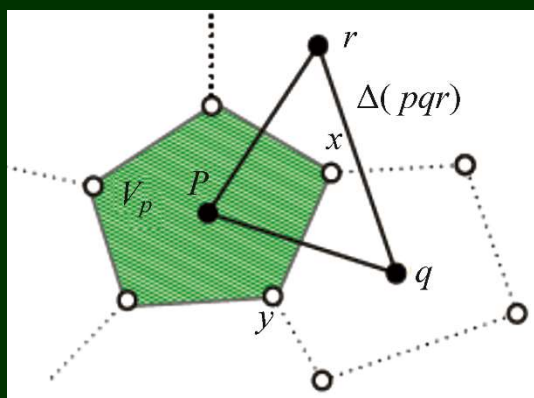
1.22 图像几何：图像上的沃罗诺伊和德劳内网格

基于角点的沃罗诺伊网格图像覆盖格

$p, q \in S,$

$\Delta(pqr) =$

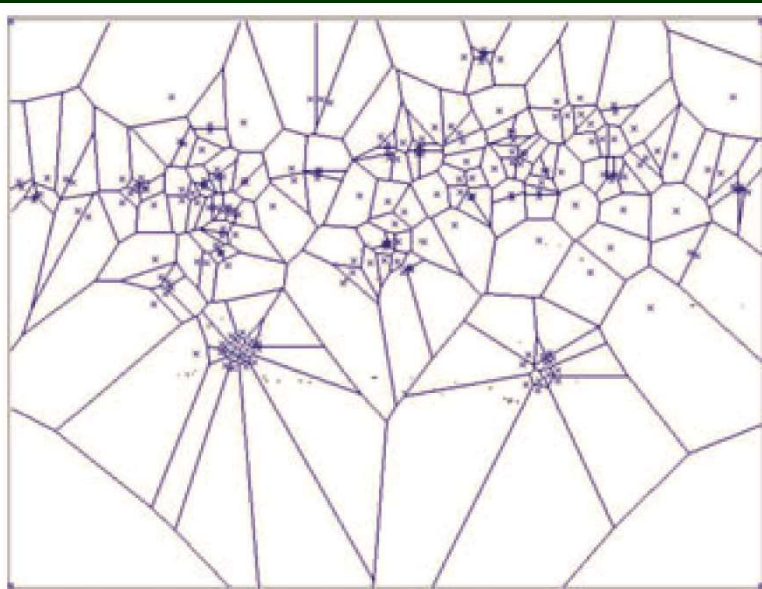
德劳内三角形



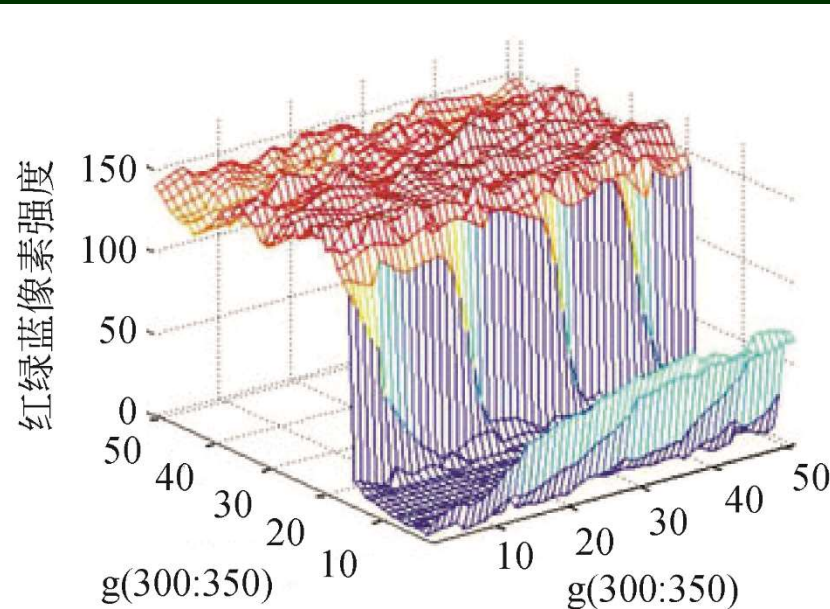


1.22 图像几何：图像上的沃罗诺伊和德劳内网格

从镶嵌的数字图像中提取的沃罗诺伊网格倾向于揭示图像几何和图像目标的存在



(a) 汽车图像沃罗诺伊网格



(b) 汽车子图像强度



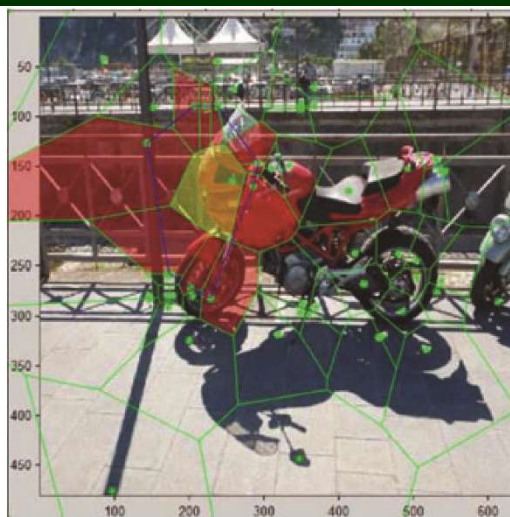
1.23 神经结构

四种神经结构

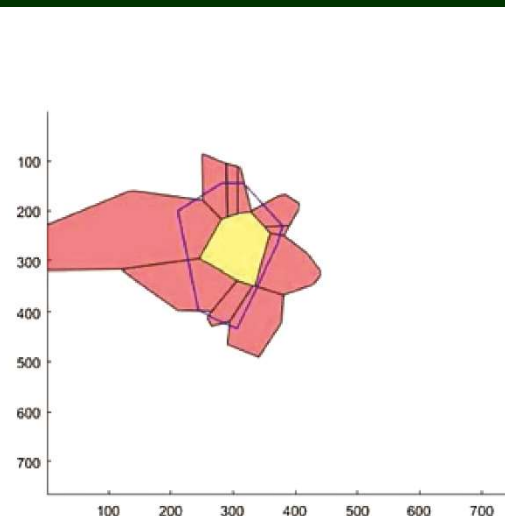
(1) 非图像的纯粹几何神经是围绕中心多边形（神经核）的多边形集合



(a) 摩托车图像



(b) 图像网格神经



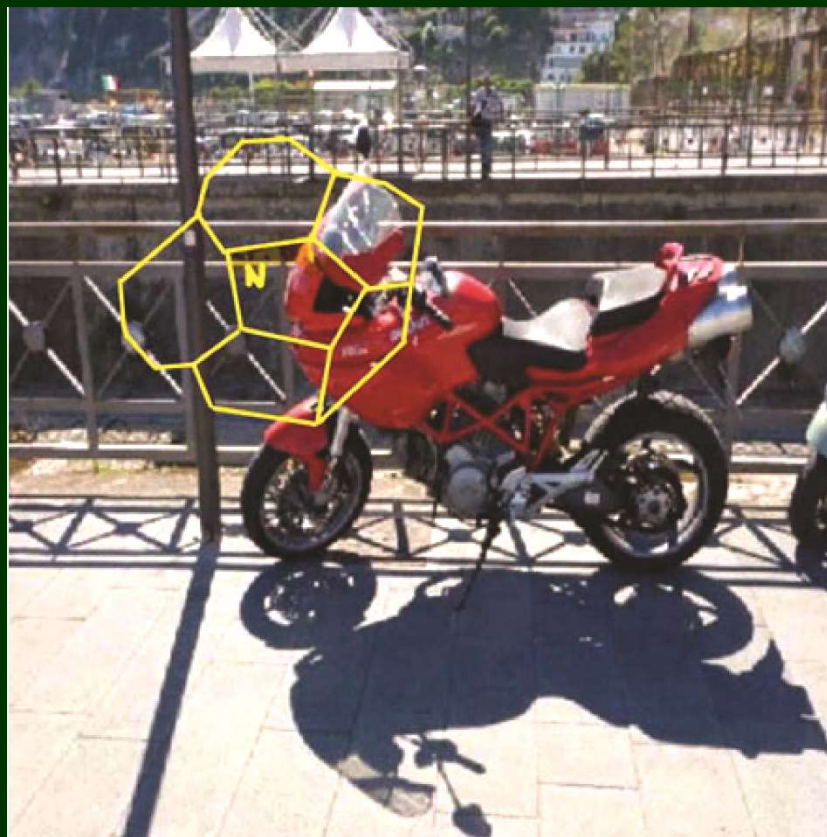
(c) 网格神经凸包



1.23 神经结构

四种神经结构

(2) 图像几何神经是源自数字图像并围绕中心多边形(神经核)的多边形集合

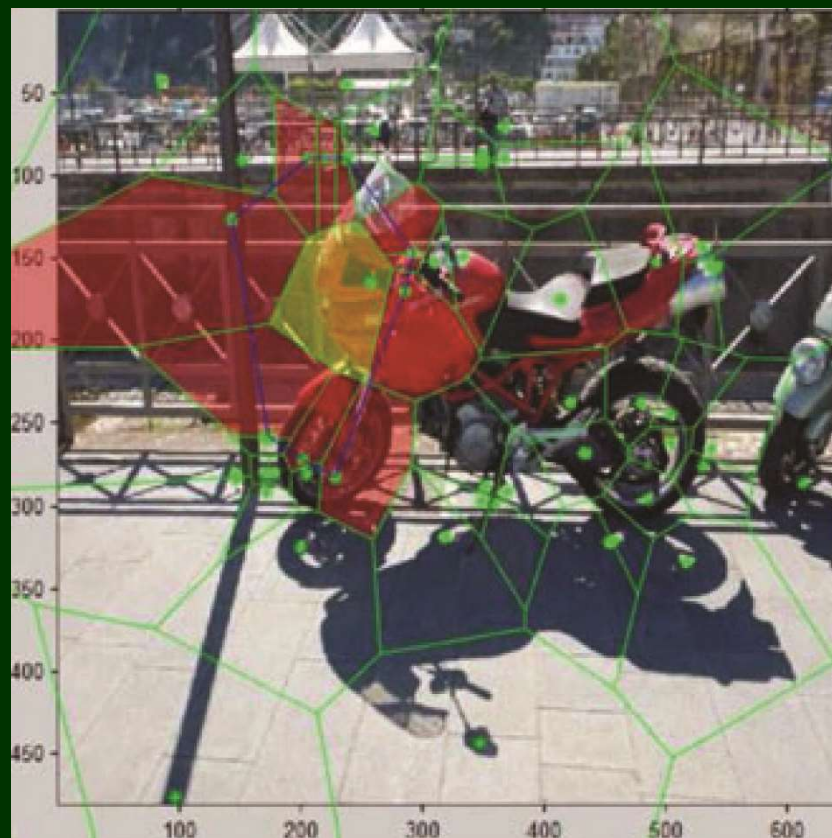




1.23 神经结构

四种神经结构

(3) 沃罗诺伊网格神经是多边形沃罗诺伊区域的集合，源自在数字图像上的生成点并围绕中心多边形（神经核）





1.23 神经结构

四种神经结构

(4) 德劳内网格神经
是德劳内三角形的集合，其顶点在图像上的生成点并围绕中心三角形（神经核）





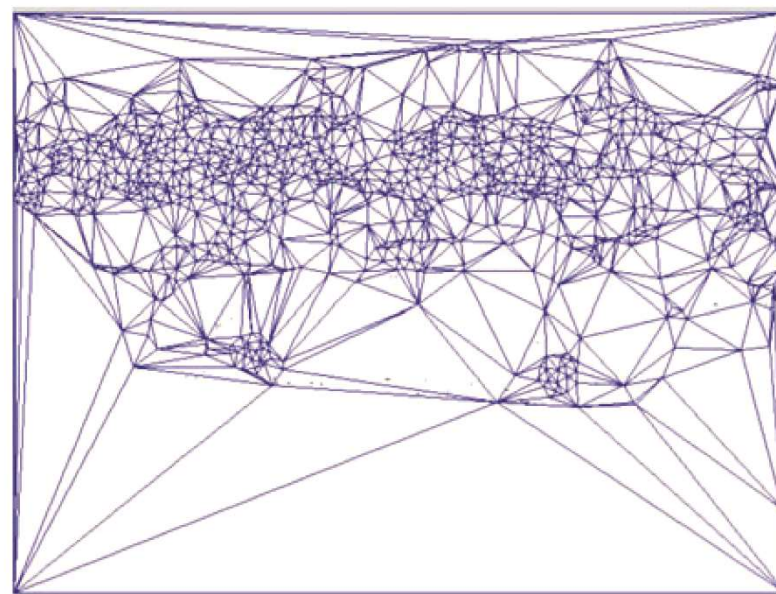
1.23 神经结构

创建德劳内网格覆盖

获得由德劳内网格提供的更简单、均匀的三角形



(a) 德劳内网格覆盖

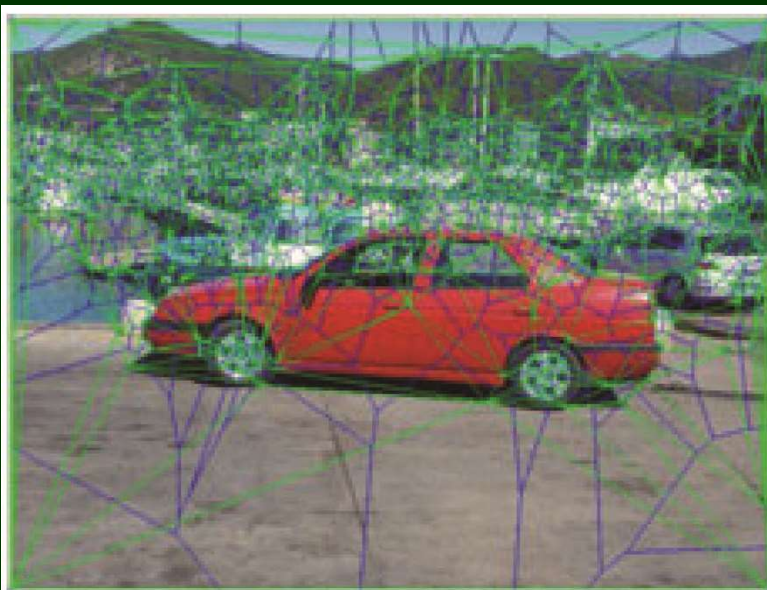


(b) 德劳内网格

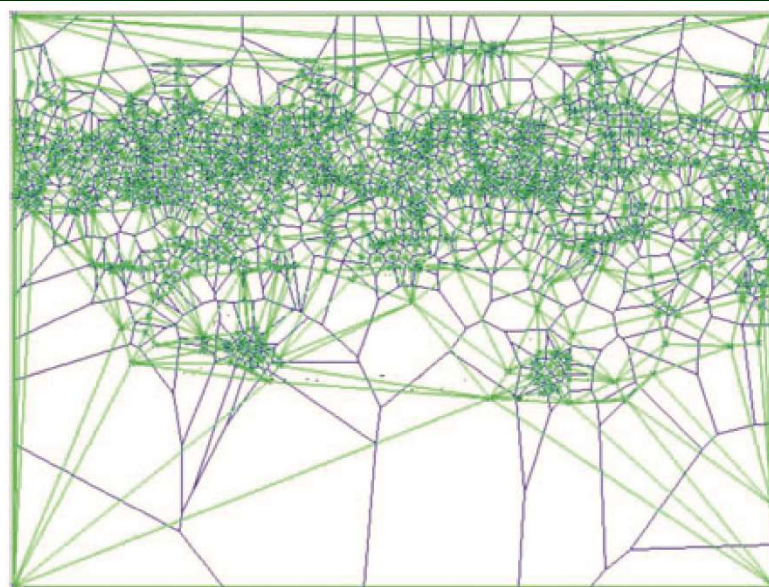


1.23 神经结构

结合沃罗诺伊网格和德劳内网格进行覆盖
三角形区域越均匀，则覆盖的子图像的质量越高



(a) 沃罗诺伊+德劳内



(b) 沃罗诺伊+德劳内网格



1.24 视频帧网格覆盖

通过用（在附近）围绕图像目标的网格多边形覆盖每个视频帧图像来检测数字图像中目标的几何形状

图像网格多边形偏重于揭示目标的形状和身份

种子点（也称为网点或生成器）为生成狄利克雷镶嵌（也称为沃罗诺伊图）和德劳内三角化的点集提供了基础

平面拼贴是平面图形的平面填充排列，覆盖平面而没有间隙或重叠



1.24 视频帧网格覆盖

实时视频处理的基本步骤

- (1) 开始对变化场景的视频采集
- (2) 从视频中选取当前帧
- (3) 从当前视频帧中提取信息，例如，通过用沃罗诺伊网格覆盖视频帧来发现几何信息
- (4) 重复步骤（2）直到视频完整

算法1.6示出了特定形式的视频处理。该算法生成实时视频帧沃罗诺伊网格



译者（章毓晋）联系信息

- ✎ 通信地址：北京清华大学电子工程系
- ✎ 邮政编码：100084
- ✎ 办公地址：清华大学罗姆楼，6层305室
- ✎ 办公电话：(010) 62798540
- ✎ 传真号码：(010) 62770317
- ✎ 电子邮件：zhang-yj@tsinghua.edu.cn
- ✎ 个人主页：oa.ee.tsinghua.edu.cn/~zhangyujin/
(下载更新的讲稿和教材修改表)