

# 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

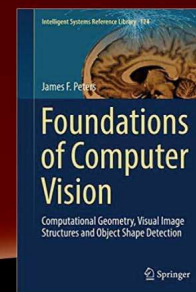
清华计算机图书·译丛

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8章 洛韦关键点、最大核聚类、轮廓和形状

继续使用叠加在数字图像上的沃罗诺伊网格作为揭示图像几何形状的手段

图像几何是一种对图像的解析几何观点

几何结构包括图像邻域、图像聚类、图像片段、图像镶嵌、最接近特定图像点的点集、聚集在一起的近邻图像区域集、相邻图像区域、以及多边形图像区域的几何形状

图像拓扑研究像素到像素集合的接近程度



## 8.1 图像分析

图像分析聚焦于各种数字图像测量，如像素尺寸、像素邻接、像素特征值、像素邻域隶属度、像素梯度方向、像素梯度幅度、像素强度、像素强度分布、图像邻域的接近度、直方条化、图像尺寸和图像分辨率

图像分析的另一个重要部分是对图像缩放、平移、旋转不变和对照明变化和仿射或3-D投影部分不变的像素特征的检测



## 8.1 图像分析

图像分析中有许多重要的基于区域的方法：等数据阈值化、Otsu方法、分水岭分割、最大沃罗诺伊网格核（识别具有最大边数的网格多边形）和非最大抑制（通过抑制比其周围像素更不可能的所有像素来找到局部最大值）

在图像分析中，目标和背景像素以不同的邻接关系（邻域）相关联，有两种基本类型的邻域，即邻接邻域和拓扑邻域



## 8.2 场景分析

场景分析聚焦数字图像的结构

图像场景是摄像机瞬间看到的快照

图像结构包括图像目标的形状、主要图像形状和图像几何形状

视觉场景是视场中捕捉人们注意力的物体集合

在人类视觉中，视场是可以看到的物体的总面积。正常视场大约为每只眼睛的垂直子午线左右大约 $60^\circ$ ，水平子午线以上大约 $60^\circ$ 、以下大约 $75^\circ$



## 8.2 场景分析

场景分析的基础建立在罗森菲尔德开创性的数字拓扑工作（后来称为数字几何）

数字拓扑的工作与计算几何的引入是并行的，并建立在沃罗诺伊和其他的空间镶嵌工作基础之上

要分析和理解图像场景，有必要识别场景中的目标。这些目标可以在几何上被视为连接边缘的集合或被视为像素集的图像区域，这些像素在某种意义上是彼此接近或者接近一个固定点的点集



## 8.3 像素边缘增强

像素边缘强度也称为像素梯度幅度

像素\text{Img}(x, y)的边缘强度由E(x, y)表示

$$E(x, y) = \sqrt{\left(\frac{\partial \text{Img}(x, y)}{\partial x}\right)^2 + \left(\frac{\partial \text{Img}(x, y)}{\partial y}\right)^2} = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (\text{像素边缘强度})$$

图中围绕头部的大圆半径的长度表示  
该圆中心像素的边缘强度。该半径的角度  
(约 $75^\circ$ )表示中心像素的梯度方向

每个圆心被称为关键点 (兴趣点)

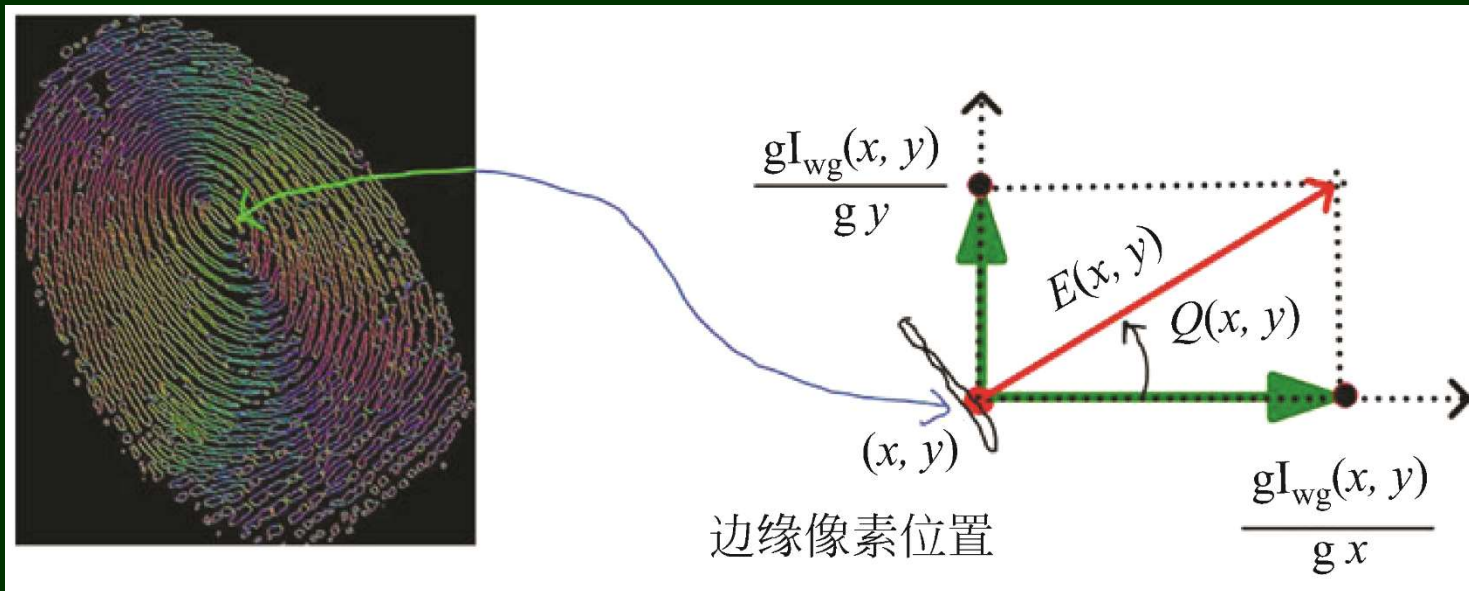






## 8.3 像素边缘增强

像素边缘强度可由斜边的长度表示，这是图中所示的图像几何形状的一部分，以沿着指纹轮廓的边缘像素来说明



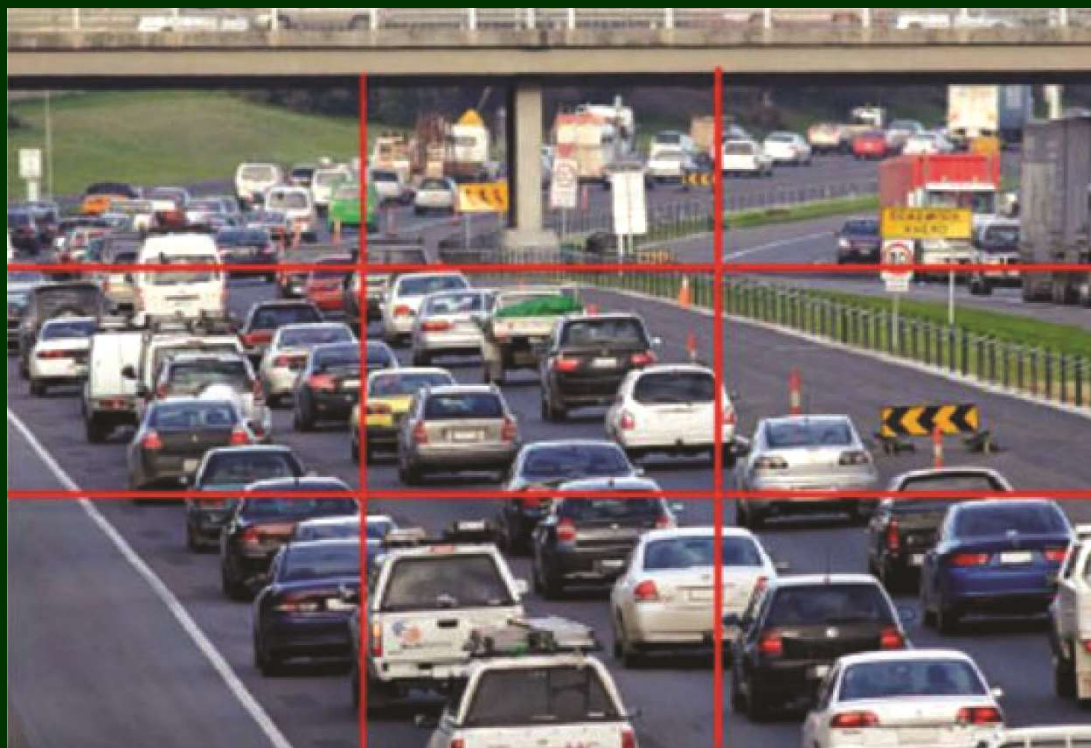




## 8.4 数字图像的裁剪和稀疏表示

对于复杂的视频帧（例如交通视频帧），有必要裁剪每个帧，然后仅选择裁剪帧的一部分进行镶嵌

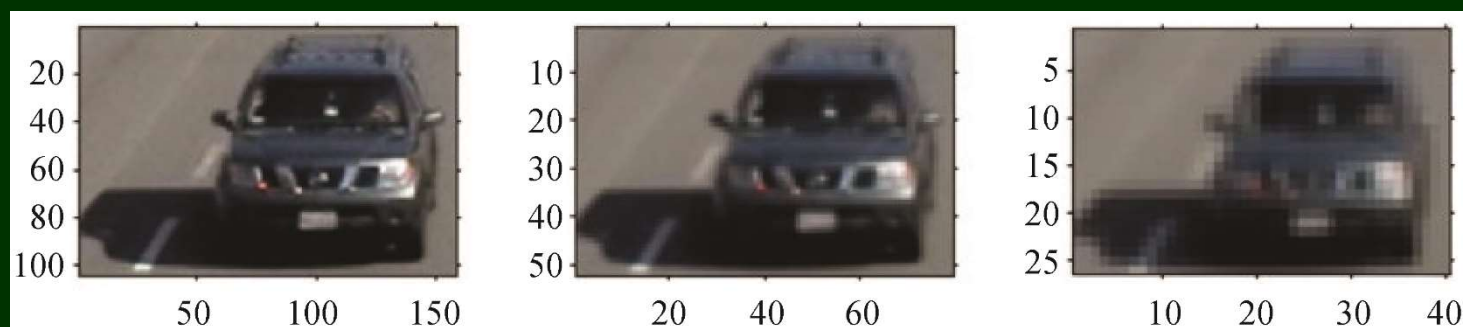
裁剪图像的意思是移除图像的外部部分以隔离和放大感兴趣的区域





## 8.4 数字图像的裁剪和稀疏表示

数字图像的稀疏表示是一个很有前景的研究领域  
在裁剪后的图像上尝试收缩和扩展金字塔方案



(a) 金字塔收缩



(b) 金字塔扩展



## 8.4 数字图像的裁剪和稀疏表示

基于小波的稀疏  
表达金字塔方案

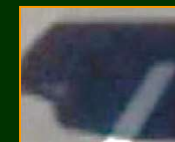




## 8.5 形状理论和2-D图像目标的形状： 面向图像目标形状检测

对图像中目标形状检测和目标类别识别在计算机视觉中得到极大关注

基本形状特征可以由边界片段表示，形状外观可以通过诸如交通视频帧中的自动阴影形状的图片片来表示



对视频帧中图像目标形状检测的一种有希望的方法是跟踪图像目标轮廓（形状）的改变并最小化组合了区域、边界和形状信息的能量函数



## 8.5 形状理论和2-D图像目标的形状： 面向图像目标形状检测

### 图像目标形状检测的基本步骤

在研究特定图像中的目标形状时，一种好的做法是将每个目标形状视为一类形状的成员

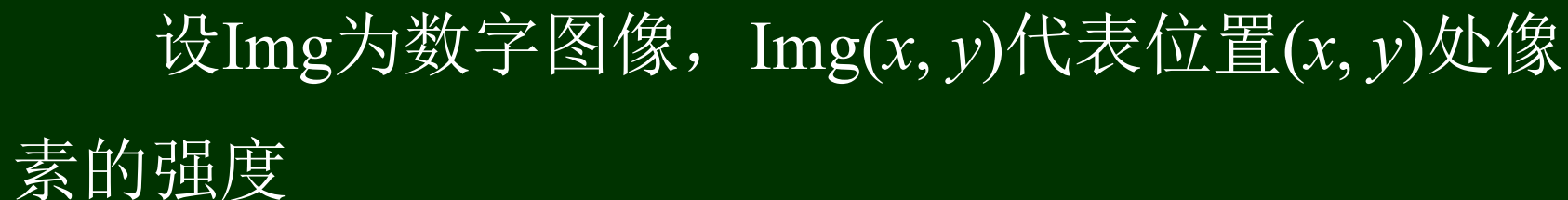
一个形状类是一组具有匹配特征的形状。这样，对图像目标形状的检测简化为检查特定图像中目标形状的特征是否与已知形状类的代表性特征的匹配问题

这里的重点是特定形状类的形状隶属度





## 8.6 图像像素梯度的朝向和强度

设为数字图像， $\text{img}(x, y)$ 代表位置 $(x, y)$ 处像素的强度

由于是两个变量 $x$ 和 $y$ 的函数，所以要计算 $\partial \text{img}(x, y) / \partial x$ 和 $\partial \text{img}(x, y) / \partial y$

**2-D**像素强度的罗森菲尔德8-邻域

$$\text{Nbhd}(f(x, y)) = \begin{bmatrix} f(x-1, y+1) & f(x, y+1) & f(x+1, y+1) \\ f(x-1, y) & f(x, y) & f(x+1, y) \\ f(x-1, y-1) & f(x, y-1) & f(x+1, y-1) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$



## 8.6 图像像素梯度的朝向和强度

使用Chen的计算 $f(x, y)$ 离散偏导数的方法

$$\frac{\partial f(x, y)}{\partial x} = f(x+1, y) - f(x, y) = 1 - 1 = 0$$

$$\frac{\partial f(x, y)}{\partial y} = f(x, y+1) - f(x, y) = 0 - 1 = -1$$

Chen方法的替代方案是Sobel方法

$$\begin{aligned} \frac{\partial f(x, y)}{\partial x} &= \frac{2}{4} f(x+1, y) - f(x-1, y) + \frac{1}{4} f(x+1, y+1) - f(x-1, y+1) \\ &\quad + \frac{1}{4} f(x+1, y-1) - f(x-1, y-1) = -\frac{2}{4} - \frac{2}{4} = -1 \end{aligned}$$

$$\begin{aligned} \frac{\partial f(x, y)}{\partial y} &= \frac{2}{4} f(x, y+1) - f(x, y-1) + \frac{1}{4} f(x+1, y+1) - f(x+1, y-1) \\ &\quad + \frac{1}{4} f(x-1, y+1) - f(x-1, y-1) = \frac{2}{4} - \frac{2}{4} = 0 \end{aligned}$$





## 8.7 高斯差

高斯差分（DoG）函数由将高斯与两个不同尺度级的图像卷积并计算该对卷积图像之间的差异来定义

设 $\text{Img}(x, y)$ 为强度图像， $G(x, y, s)$ 为由变量尺度 $s$ 定义的高斯函数

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

设 $k$ 为比例因子，高斯差分图像（记为 $D(x, y, s)$ ）

$$D(x, y, \sigma) = G(x, y, k\sigma) * \text{Img}(x, y) - G(x, y, \sigma) * \text{Img}(x, y)$$

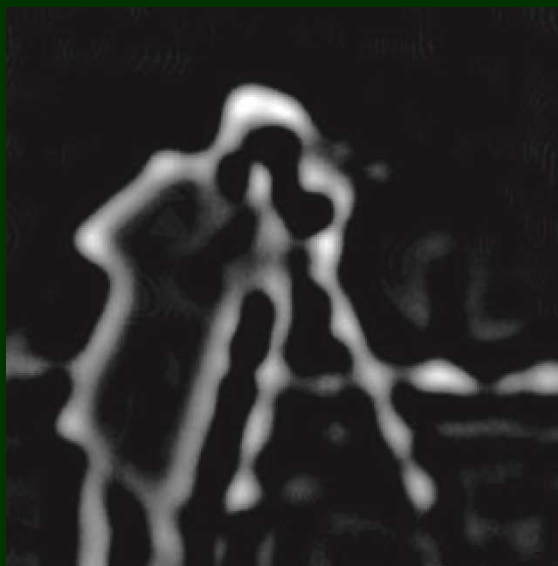


## 8.7 高斯差

高斯差（DoG）中使用较小的标准差可获得更好的高斯差图像



(a) 原始图像



(b)  $k=1.5, \sigma=5.55$



(c)  $k=1.5, \sigma=0.98$



## 8.8 图像关键点：洛韦SIFT方法

尺度不变特征变换（SIFT）是解决物体识别以及物体跟踪问题的重要支柱

**SIFT.1:** 使用高斯差函数来识别对尺度和朝向不变的潜在兴趣点

**SIFT.2:** 根据稳定性的度量选择关键点

**SIFT.3:** 可基于梯度朝向（方向）区分关键点

**SIFT.4:** 在 $x$ 和 $y$ 方向上的局部像素梯度幅度值可用于计算像素边缘强度



## 8.8 图像关键点：洛韦SIFT方法

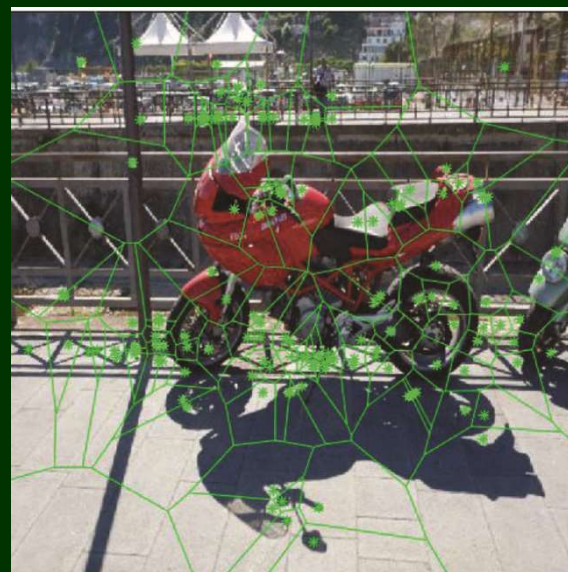
使用关键点作为沃罗诺伊或德劳内图像镶嵌的生成器，144个关键点进行摩托车图像的沃罗诺伊镶嵌



(a) 摩托车图像



(b) 144个关键点



(c) 摩托车图像的沃罗诺伊网格



## 8.9 应用：图像网格核的关键点边界

沿图像网格核边界在多边形中找到的关键点的实际应用

在最大核聚类（MNC）中围绕核的至少有四种不同类型的轮廓边缘集

网状神经是所谓的MNC辐条的集合，每个辐条都是MNC核和相邻多边形的组合

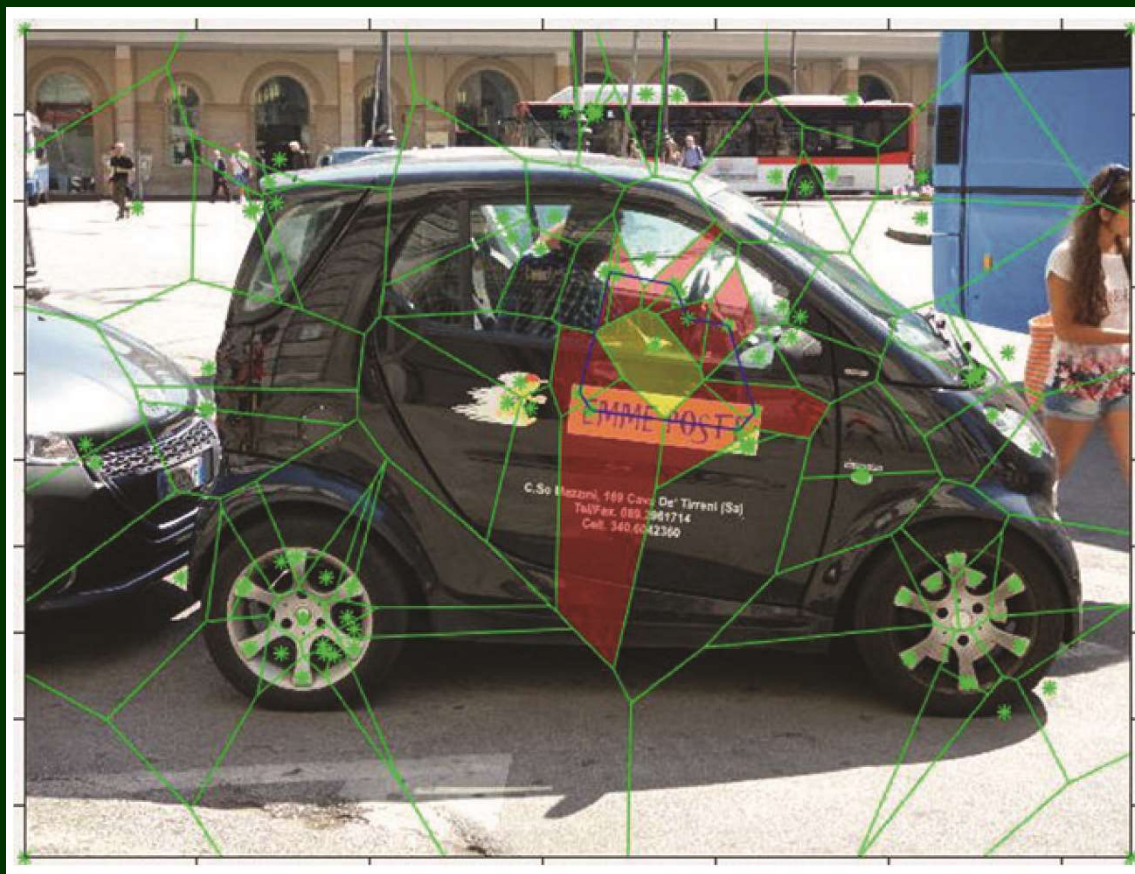
网状神经是以网状核为中心的轮辐状投影的集合





## 8.9 应用：图像网格核的关键点边界

从MNC核辐射出来的辐条集合





## 8.9 应用：图像网格核的关键点边界

**MNC**轮廓周边的种类

**IP**边缘集：精细周边（起点——称为内部周边或简称为IP）

**S1P**：第1级 粗糙周边1（起点——称为超1级周边或简称为S1P）

**S2P**：第2级 粗糙周边2（超2级周边或简称为S2P）

**S3P**：第3级 粗糙周边3（超3级周边或简称为S3P）

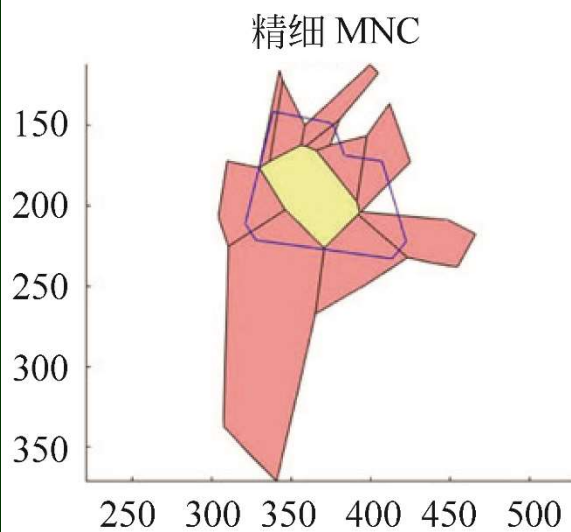




## 8.9 应用：图像网格核的关键点边界

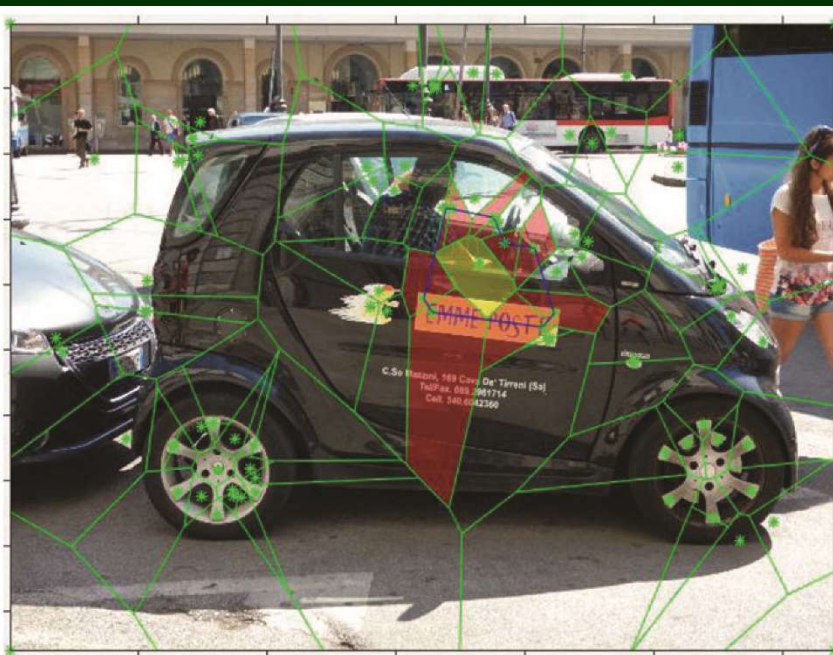
IP（内部周边）边缘  
集由连接核周围关键

点的蓝线段—•—构成



(a) IP边缘集几何结构

沃罗诺伊网格由89  
个关键点构成



(b) 图像IP



## 8.10 超（外）核轮廓

粗糙核轮廓可用于检测大图像目标的形状

通过连接沃罗诺伊网格多边形内部的关键点可找到粗糙核轮廓，这些关键点沿着MNC核的精细周边多边形的边界

在MNC中，粗糙轮廓也称为超轮廓或外轮廓

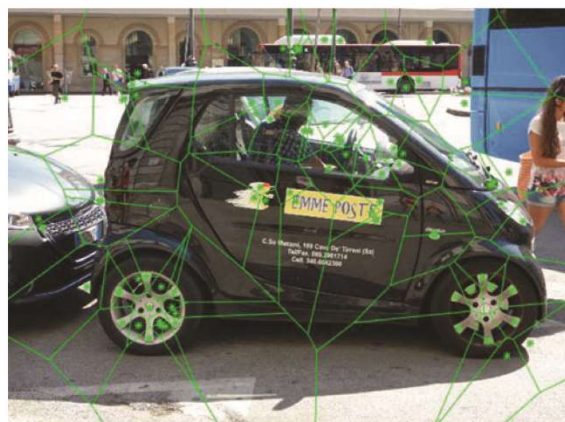
粗糙轮廓的长度（核周长）追踪由MNC覆盖的中型目标的形状

S1P（超1级周边）是最里面的MNC粗糙轮廓



## 8.10 超（外）核轮廓

通过MNC  
SIP 粗糙  
轮廓边缘  
集显示的  
图像几何  
结构



(a) 沃罗诺伊网格



(b) 网格上的SIP边缘集



(c) 网格上的SIP边缘集



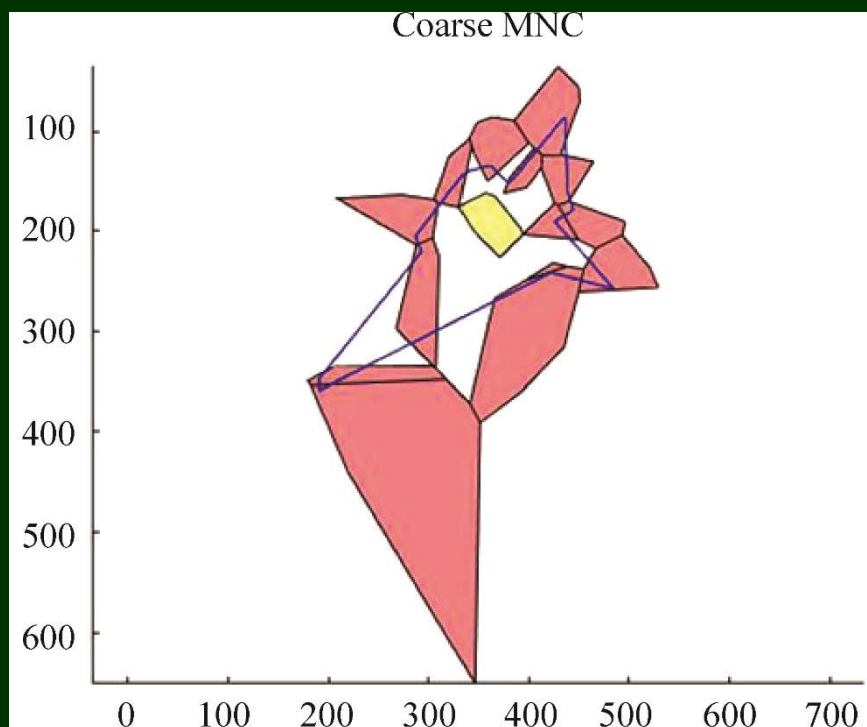
(d) SIP边缘集自身



## 8.10 超（外）核轮廓

最大核聚类 (MNC) 中的  
S1P 粗糙边缘集几何结构

S1P核的特写视图







## 8.11 最大核聚类轮廓形状的质量

MNC轮廓形状的质量取决于所选择的目标形状

在目标识别的设置中，目标形状是希望与单个图像或视频图像帧序列中的样本形状进行比较的

在目标形状的周长接近样本MNC轮廓周长的情况下，MNC轮廓形状的质量高

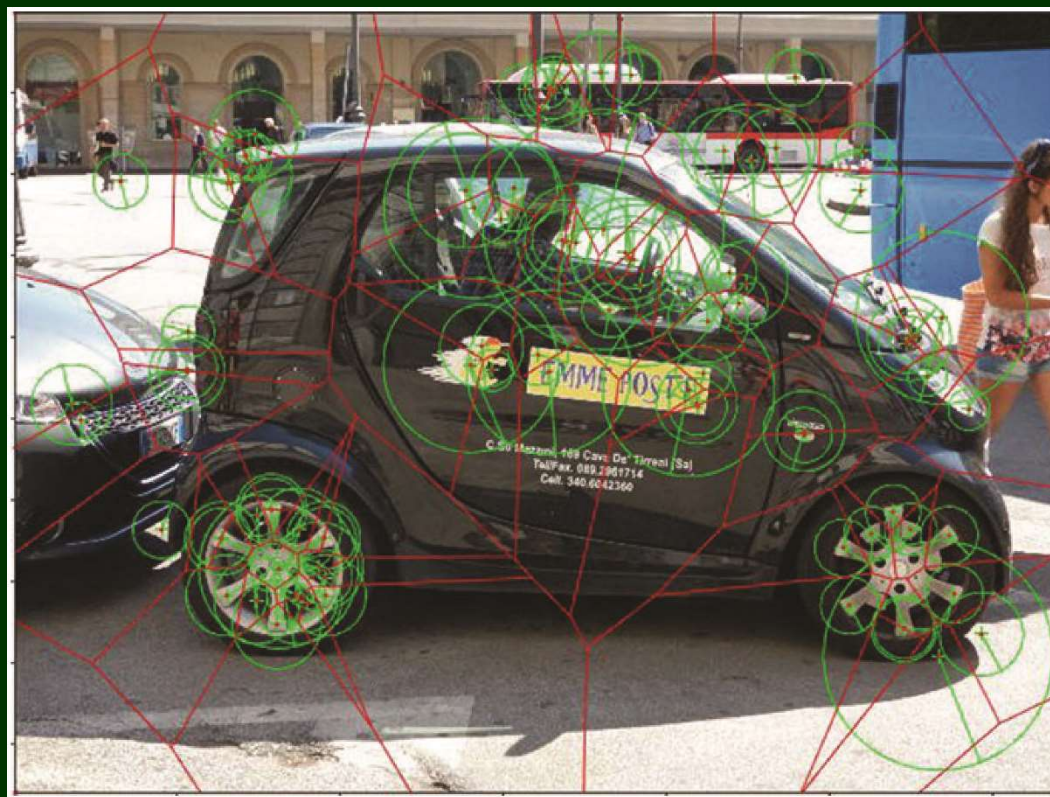
换句话说，MNC轮廓形状的质量与目标轮廓形状与样本轮廓形状的接近度成比例



## 8.12 粗糙S2P和S3P（2级和3级） 最大核聚类轮廓

S2P轮廓通常紧密地结合在图像上的MNC聚类中的S1P轮廓周围

关键点聚集在  
车辆上的驾驶员和  
图标以及车轮周围  
可以期望在汽车  
中部找到最大的  
核聚类





## 8.12 粗糙S2P和S3P（2级和3级） 最大核聚类轮廓

紧密结合的S1P和S2P轮廓

S2P 轮廓  
边缘集在图中  
以白色显示

S2P 轮廓  
围绕 S1P 轮廓  
紧密结合





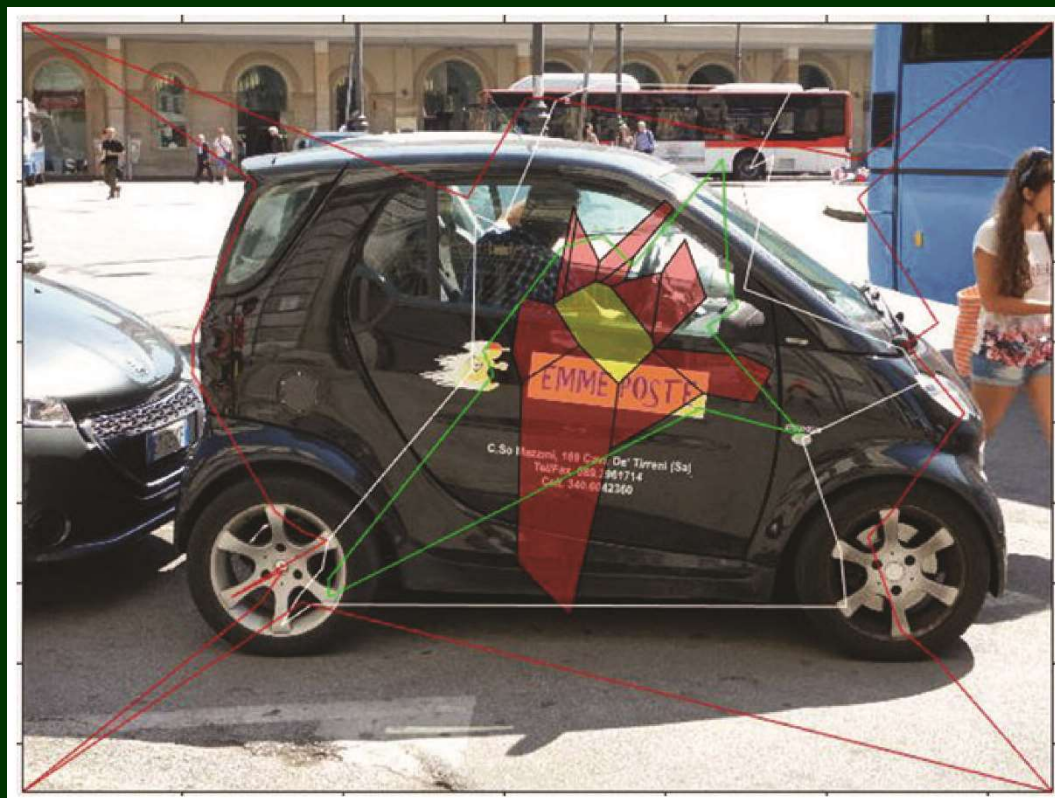


## 8.12 粗糙S2P和S3P（2级和3级） 最大核聚类轮廓

紧密结合的S1P、S2P和S3P轮廓

将 S3P 轮廓  
显示为一系列连  
接的红线段 ●—●

S3P轮廓中的  
线段通常不围绕  
MNC核周围的内  
轮廓紧密地结合





## 8.13 关键点数量的实验

先前仅考虑了包含一个最大核聚类的镶嵌图像  
通过改变生成点的数量，有可能改变镶嵌图像中的MNC的数量

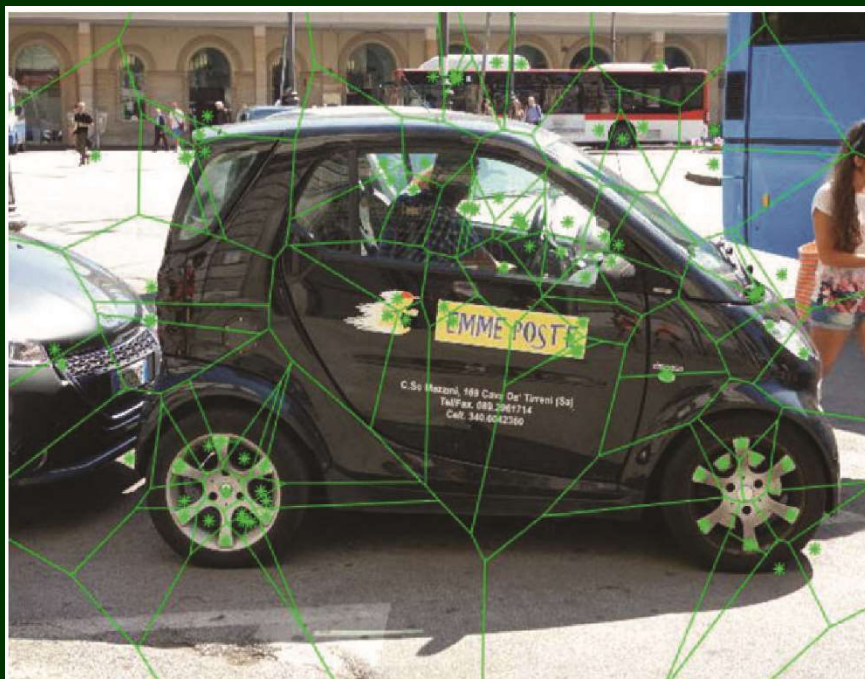
这里的目标是构建包含相邻或重叠MNC的图像  
网格，它可用作图像目标的标记

邻接的**MNC**是最大核聚类，其中一个MNC中的多边形与另一个MNC中的多边形有共享边；只要整个多边形对于两个MNC都是公共的，重叠的**MNC**就会产生

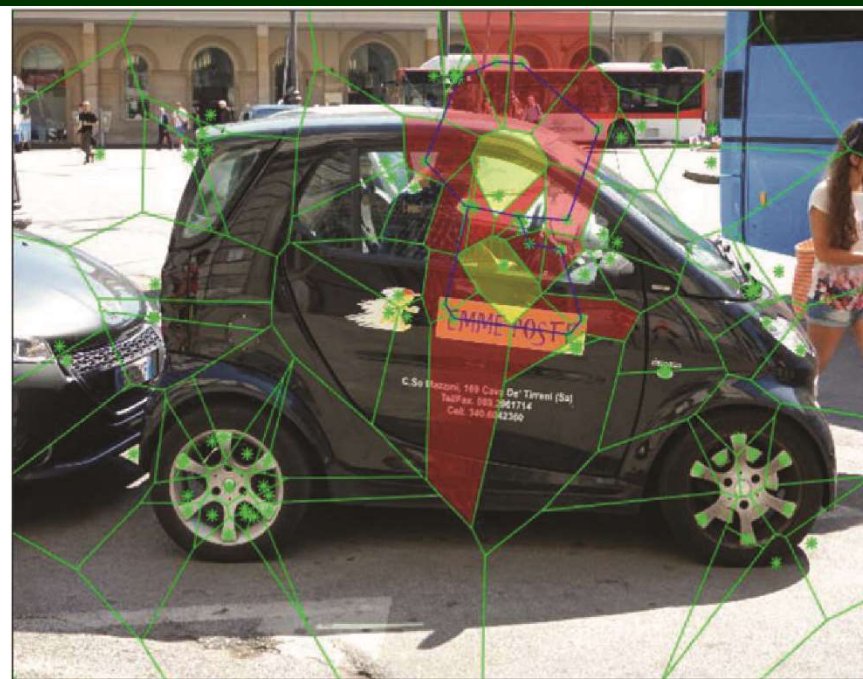


## 8.13 关键点数量的实验

通过双重和重叠的MNC来可视化图像几何结构



(a) 沃罗诺伊网格



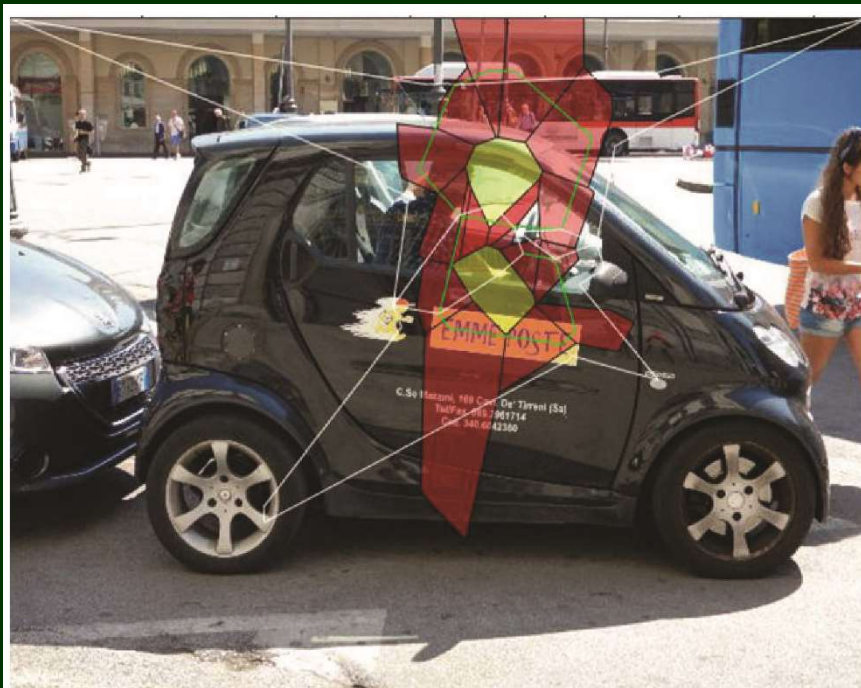
(b) 双重精细轮廓IP边缘集



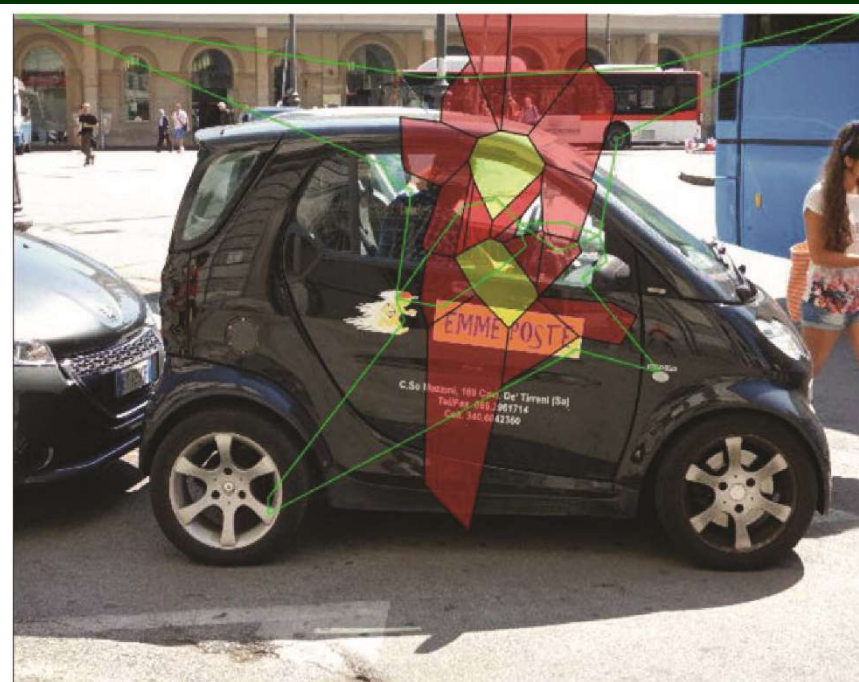


## 8.13 关键点数量的实验

通过双重MNC粗糙轮廓来可视化图像几何结构



(a) 双重粗糙SIP和精细IP边缘集



(b) 双重粗糙SIP边缘集自身



## 8.13 关键点数量的实验

在为选定数量的关键点获得具有多个MNC的沃罗诺伊网格之后，这些MNC可以分离（覆盖图像的不同部分）或重叠

接下来，考虑对用于搜索具有紧密结合的多个重叠MNC的网格中的关键点数量进行变化实验

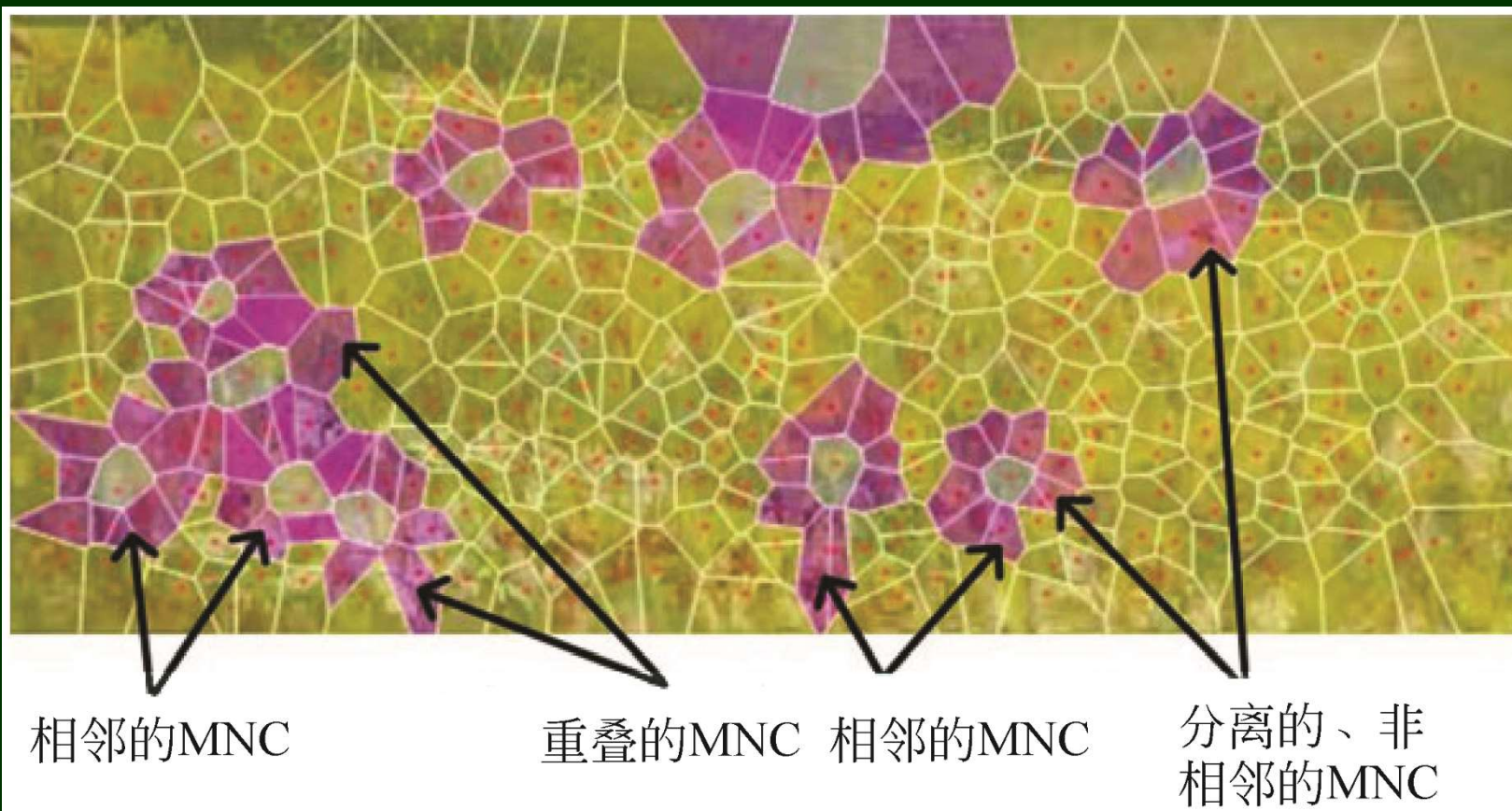
理想的情况是找到重叠的MNC，以使S1P和S2P轮廓长度的差异很小。设 $\varepsilon$ 为一个正数，令S1P<sub>c</sub>、S2P<sub>c</sub>为长度（以像素为单位）：

$$|S1P_c - S2P_c| < \varepsilon$$



## 8.13 关键点数量的实验

### 相邻的和非相邻的MNC







## 8.13 关键点数量的实验

双重、重叠MNC上的边缘集形状

围绕双重MNC核显示了与一对S1P边缘集重叠的一对S2P（粗糙周边）边缘集（白色）。这些边缘集由连接的绿线段—（S1P粗糙轮廓）和连接的白色S2P粗糙轮廓构成







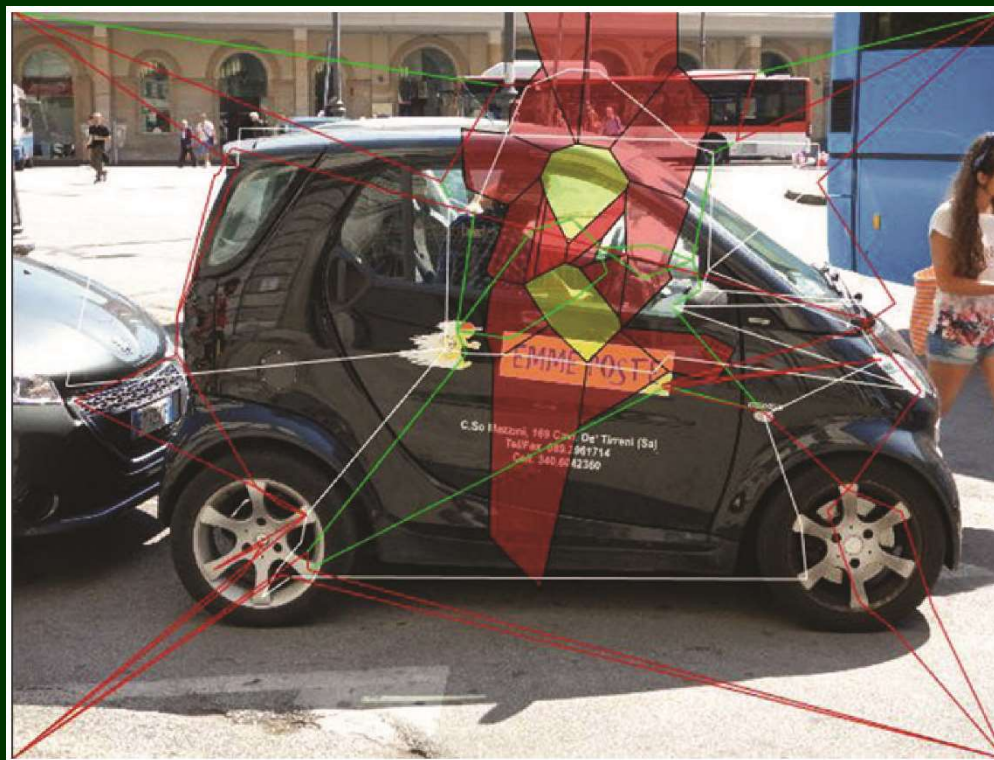
## 8.13 关键点数量的实验

双重、重叠MNC上的边缘集形状

围绕S2P轮廓的是由连接的红线段•—•构成的一对

S3P边缘集（S3P粗糙轮廓）

双重S1P、S2P  
和 S3P 轮廓重叠在  
MNC上





## 8.14 双重最大核聚类上的粗糙周边

因为希望看到S1P周边覆盖了子图像的哪个部分，所以S1P多边形在图中被抑制。现在，仅S1P周边被显示为一系列连接的绿线段 $\bullet - \bullet$ ，其中使用关键点作为S1P中每个线段的端点。显然，S1P轮廓形状比精细IP轮廓形状能包围车辆的更多中间部分

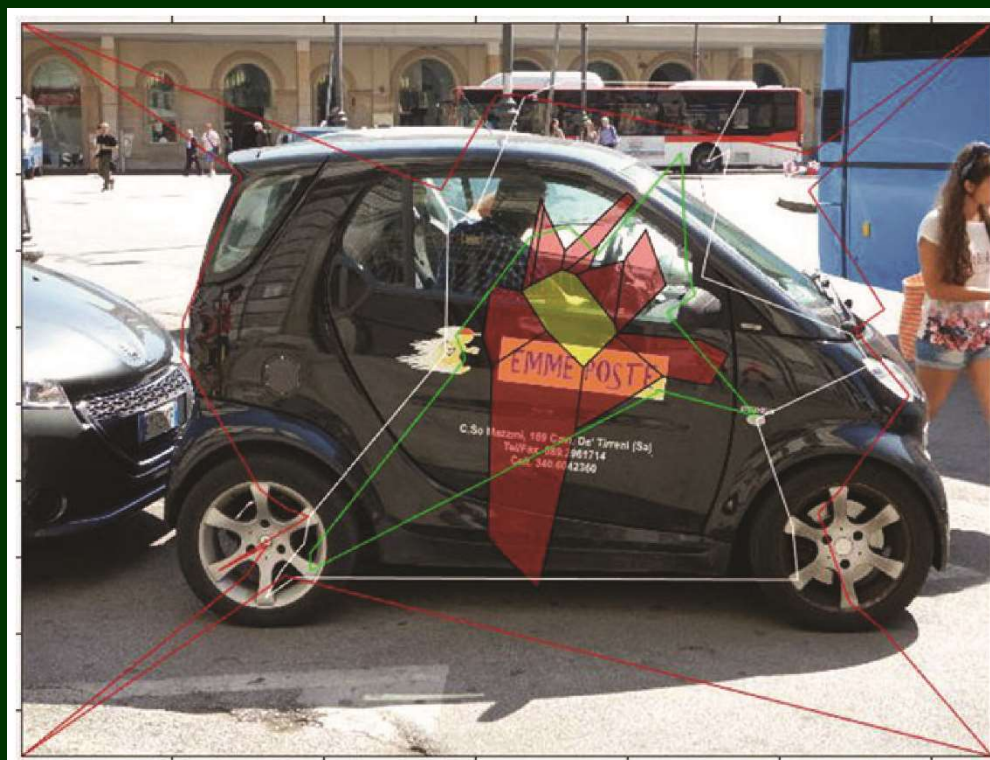




## 8.14 双重最大核聚类上的粗糙周边

将S3P轮廓显示为一系列连接的红线段 $\bullet\text{---}\bullet$ ，它们使用关键点作为S3P中每个线段的端点。每个S3P线段在沿着S2P多边形边界的一对相邻多边形中的关键点之间绘制

S3P轮廓中的线段通常不围绕MNC核周围的内轮廓紧密地结合





## 8.15 图像最大核聚类区域的莱利熵

具有高信息水平的图像区域

图像的MNC熵高于周围所围绕的非MNC熵

莱利熵对应于一组数据的信息水平

设 $p(x_1), \dots, p(x_i), \dots, p(x_n)$ 是事件序列 $x_1, \dots, x_i, \dots, x_n$ 的概率, 令 $b \geq 1$ 。那么, 一组事件 $X$ 的莱利熵 $H_b(X)$ 定义为:

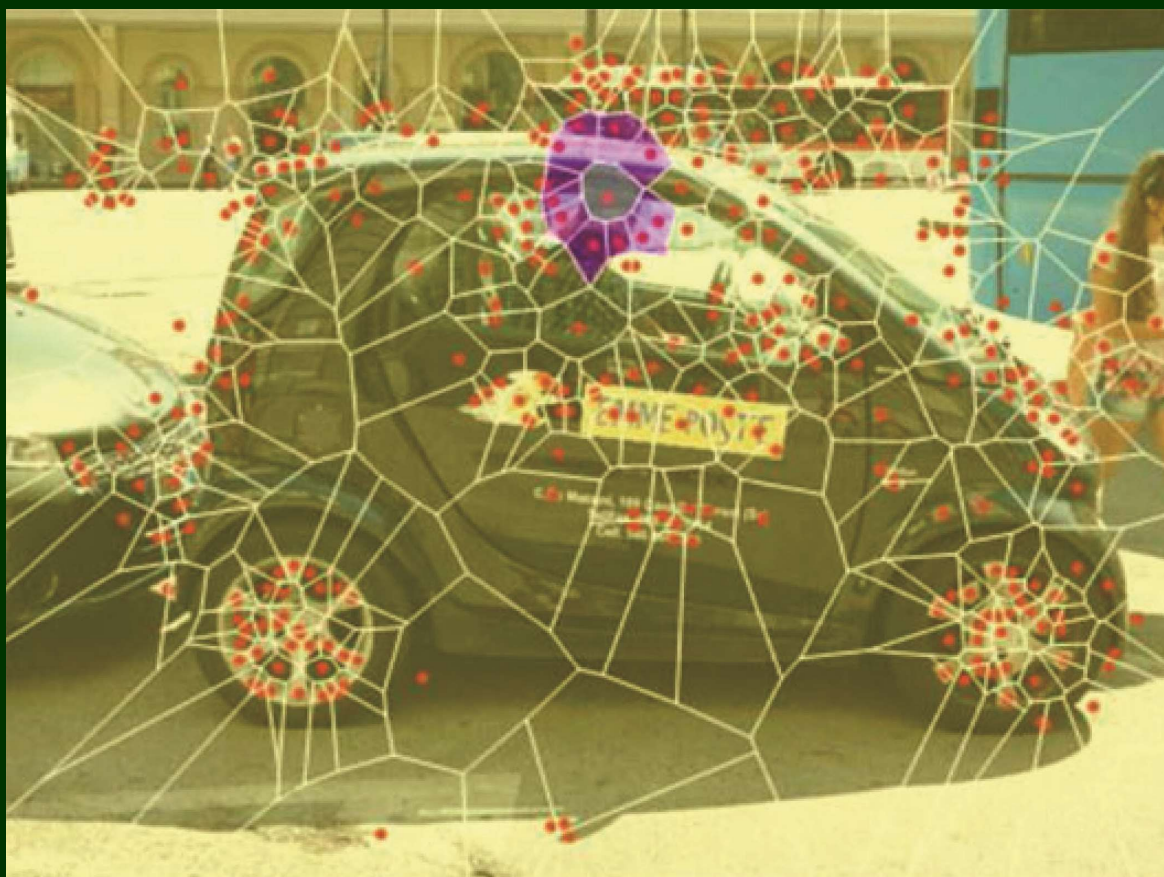
$$H_\beta(X) = \frac{1}{1-\beta} \ln \sum_{i=1}^n p^\beta(x_i) \quad (\text{莱利熵})$$





## 8.15 图像最大核聚类区域的莱利熵

具有376个关键点的沃罗诺伊网格中的单个MNC







## 8.15 图像最大核聚类区域的莱利熵

具有145个关键点的沃罗诺伊网格中的双重MNC





## 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](mailto:zhang-yj@tsinghua.edu.cn)
- ✎ 个人主页：[oa.ee.tsinghua.edu.cn/~zhangyujin/](http://oa.ee.tsinghua.edu.cn/~zhangyujin/)  
(下载更新的讲稿和教材修改表)