

吉 祥 康 寧

第一章 緒論

中科大 自動化系 鄭志剛

2018.9



问题求助：相近颜色的布料图像 的区分

- 1. 两种米色布料，颜色相近但有差异，花纹一样；
- 2. 两种灰色布料，也是颜色相近，花纹一样；

用CCD采集布料的图像，如何区分上面的各种布料？不知道有什么好的算法或方法可以区分彼此，请大虾不吝赐教。

吉祥如意

什么是模式？

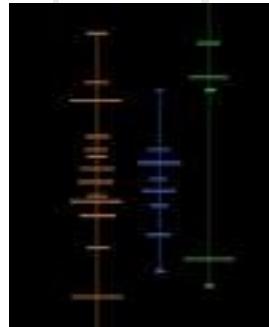
- 广义地说，存在于时间和空间中可观察的物体，如果我们可以区别它们是否相同或是否相似，都可以称之为模式。
- 模式所指的不是事物本身，而是从事物获得的信息，因此，模式往往表现为具有时间和空间分布的信息。

模式的直观特性：

- 可观察性
- 可区分性
- 相似性

吉祥慶

什么是模式（Pattern）？





模式识别的概念

- 模式识别 – 直观，无所不在，“人以类聚，物以群分”
 - 周围物体的认知：桌子、椅子
 - 人的识别：张三、李四
 - 声音的辨别：汽车、火车，狗叫、人语
 - 气味的分辨：炸带鱼、红烧肉
- 人和动物的模式识别能力是极其平常的，但对计算机来说却是非常困难的。



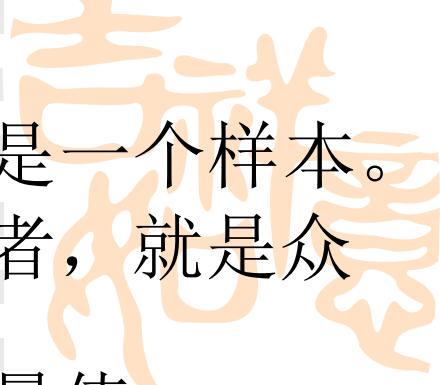


模式识别的例子（1）

- 例1：医生给一个病人看病（模式识别的完整过程）
 - 测量病人的体温和血压，化验血沉，询问临床表现；
 - 通过综合分析，抓住主要病症；
 - 医生运用自己的知识，根据主要病症，作出正确的诊断。

模式识别技术中，经常使用的术语有：样本、模式、特征、类型等。为了便于理解，可以进一步分析上述例子





- **样本**: 医院里的众多患者, 每个患者都是一个样本。
- **单一样本**: 请医生给出诊断的某一个患者, 就是众多患者中的一个样本。
- **样本的测量值**: 患者的体温、血压等测量值。
- **模式**: 样本的各测量值的综合。
- **模式样本**: 具有某种模式的样本。
- **模式采集**: 获取某样本的各测量值的过程。
- **样本特征**: 患者的主要病症。
 - **特征提取、特征选择**: 模式样本各测量值经过综合分析找出主要病症。
- **分类判决**: 医生运用自己的知识作出诊断。
- **判决准则、判决规则**: 医生的知识。
- **判决结果**: 把患者区别成某种疾病的患者。(也就是把样本(患者)区别为相应类型(疾病))。

吉
祥
如
意

模式识别的例子（2）

- 例2：选取做家具的松木。（松木、桦木）

观察：颜色、花纹、亮度、密度（样本测量值），得到样本模式

综合分析：提取、选择主要特征

根据主要区别：区分各单一样本（松木、桦木）



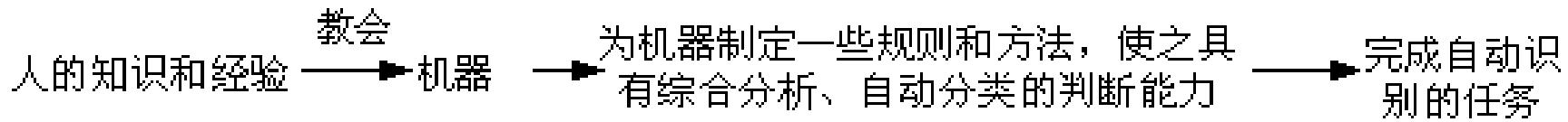
吉祥如意

模式识别的研究

- 手段：利用计算机对物理对象进行分类，在错误概率最小的条件下，使识别的结果尽量与客观物体相符合。
- $Y = F(X)$
 - X 的定义域取自特征集
 - Y 的值域为类别的标号集
 - F 是模式识别的判别方法



目的





模式识别简史（1）

- 1929年 G. Tauschek发明阅读机，能够阅读0-9的数字。
- 30年代 Fisher提出统计分类理论，奠定了统计模式识别的基础。因此，在60~70年代，统计模式识别发展很快，但由于被识别的模式愈来愈复杂，特征也愈多，就出现“维数灾难”。但由于计算机运算速度的迅猛发展，这个问题得到一定克服。统计模式识别仍是模式识别的主要理论。
- 50年代 Noam Chemskey 提出形式语言理论——傅京荪 提出句法结构模式识别。

吉祥如意

模式识别简史（2）

- 60年代 L.A.Zadeh提出了模糊集理论，模糊模式识别方法得以发展和应用。
- 80年代以Hopfield网、BP网为代表的神经网络模型导致人工神经元网络复活，并在模式识别得到较广泛的应用。
- 90年代小样本学习理论，支持向量机也受到了很大的重视。





关于模式识别的国内、国际学术组织

- 1973年 IEEE发起了第一次关于模式识别的国际会议“ICPR”，成立了国际模式识别协会---“IAPR”，每2年召开一次国际学术会议。
- 1977年 IEEE的计算机学会成立了模式分析与机器智能（PAMI）委员会，每2年召开一次模式识别与图象处理学术会议。
- 国内的组织有电子学会，通信学会，自动化协会，中文信息学会.....。



模式识别的应用

1. 字符识别：包括印刷体字符的识别；手写体字符的识别（脱机），各种OCR设备例如信函分拣、文件处理、卡片输入、支票查对、自动排版、期刊阅读、稿件输入；在线手写字符的识别（联机），各种书写输入板。
2. 医疗诊断：心电图，脑电图，染色体，癌细胞识别，疾病诊断，例如关幼波肝炎专家系统。
3. 遥感：资源卫星照片，气象卫星照片处理，数字化地球，图象分辨率可以达到1米。

吉
祥
如
意

4. 指纹识别 脸形识别
5. 检测污染分析，大气，水源，环境监测。
6. 自动检测：产品质量自动检测
7. 语声识别，机器翻译，电话号码自动查询，侦听，机器故障判断。
8. 军事应用



吉祥如意

一些工业应用

- 生物特征识别
- 人脸识别
- 车号识别
- Robocup小型组足球机器人视觉系统



吉祥如意

生物特征识别

◆ 定义

- ◆ 通过计算机与光学、声学、生物传感器和生物统计学原理等手段密切结合，利用人体固有的生理特性，（如指纹、人脸、虹膜等）和行为特征（如笔迹、声音、步态等）来进行个人身份的鉴定。

◆ 一些应用

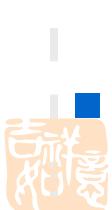
- ◆ 指纹识别
- ◆ 人脸识别
- ◆ 虹膜识别
- ◆ 步态识别



Fingerprint matching techniques

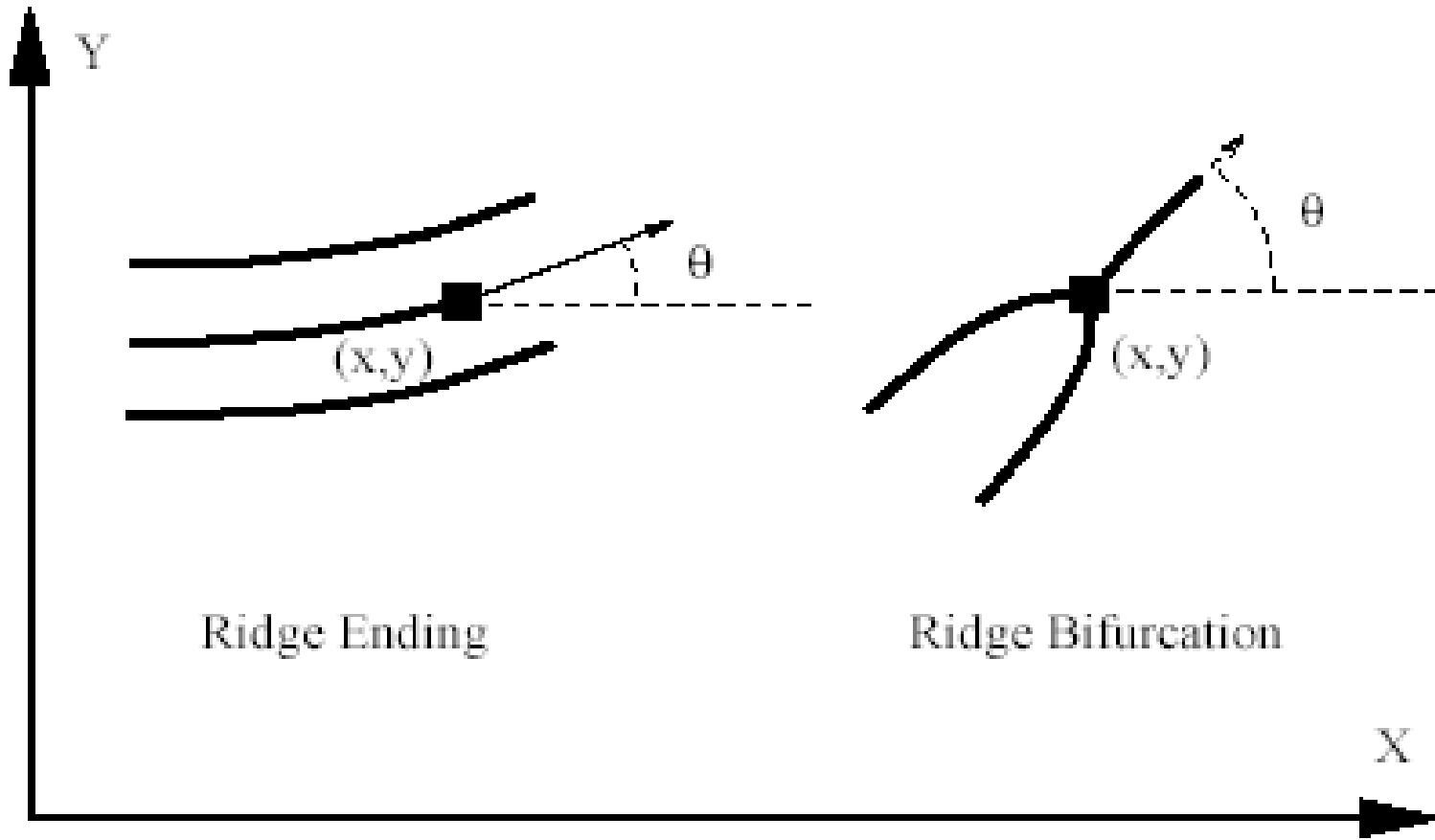


- Minutiae-based
- Correlation-based
- Ridge Feature-based

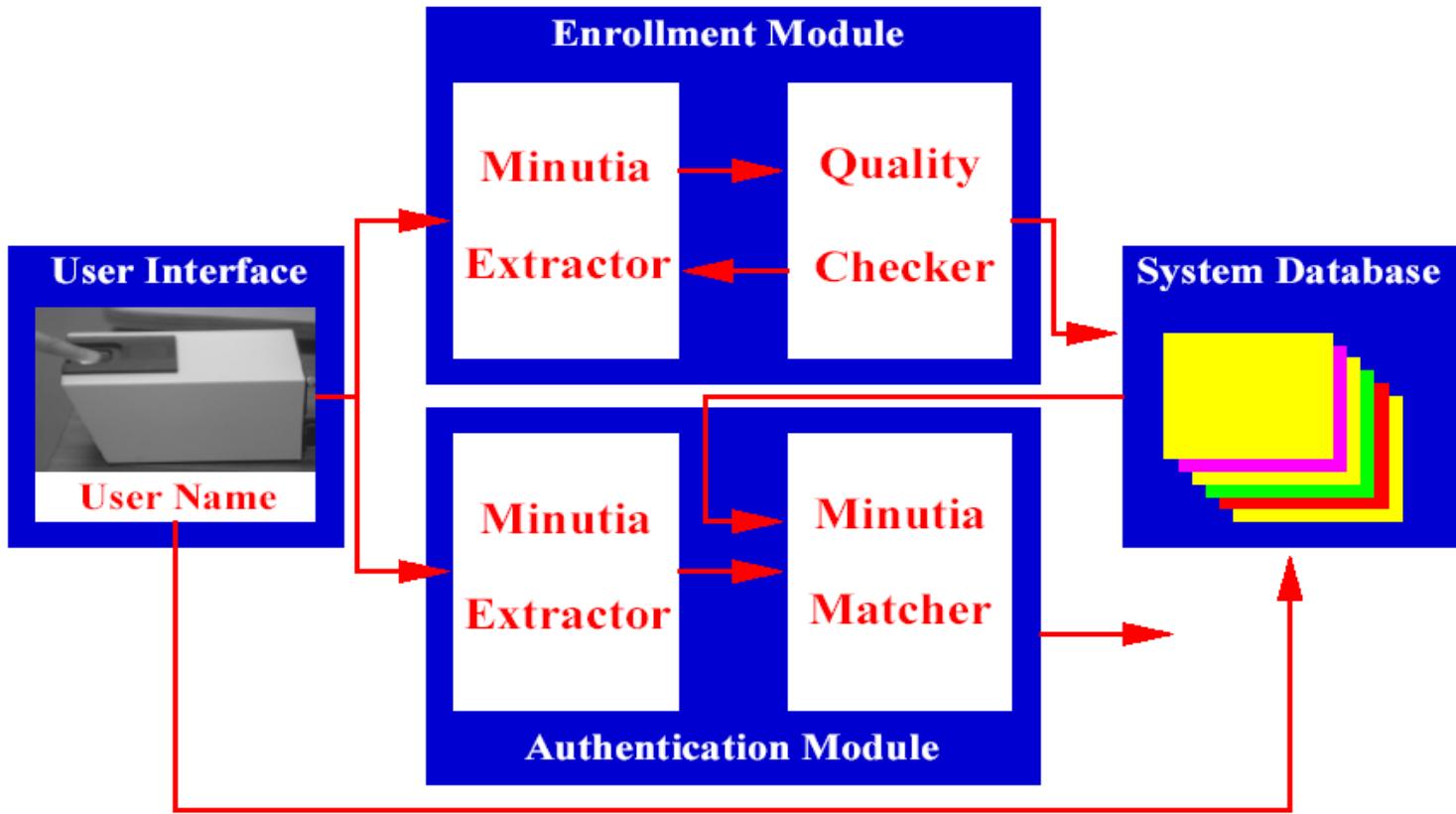




Minutiae types



Minutiae-based fingerprint verification system



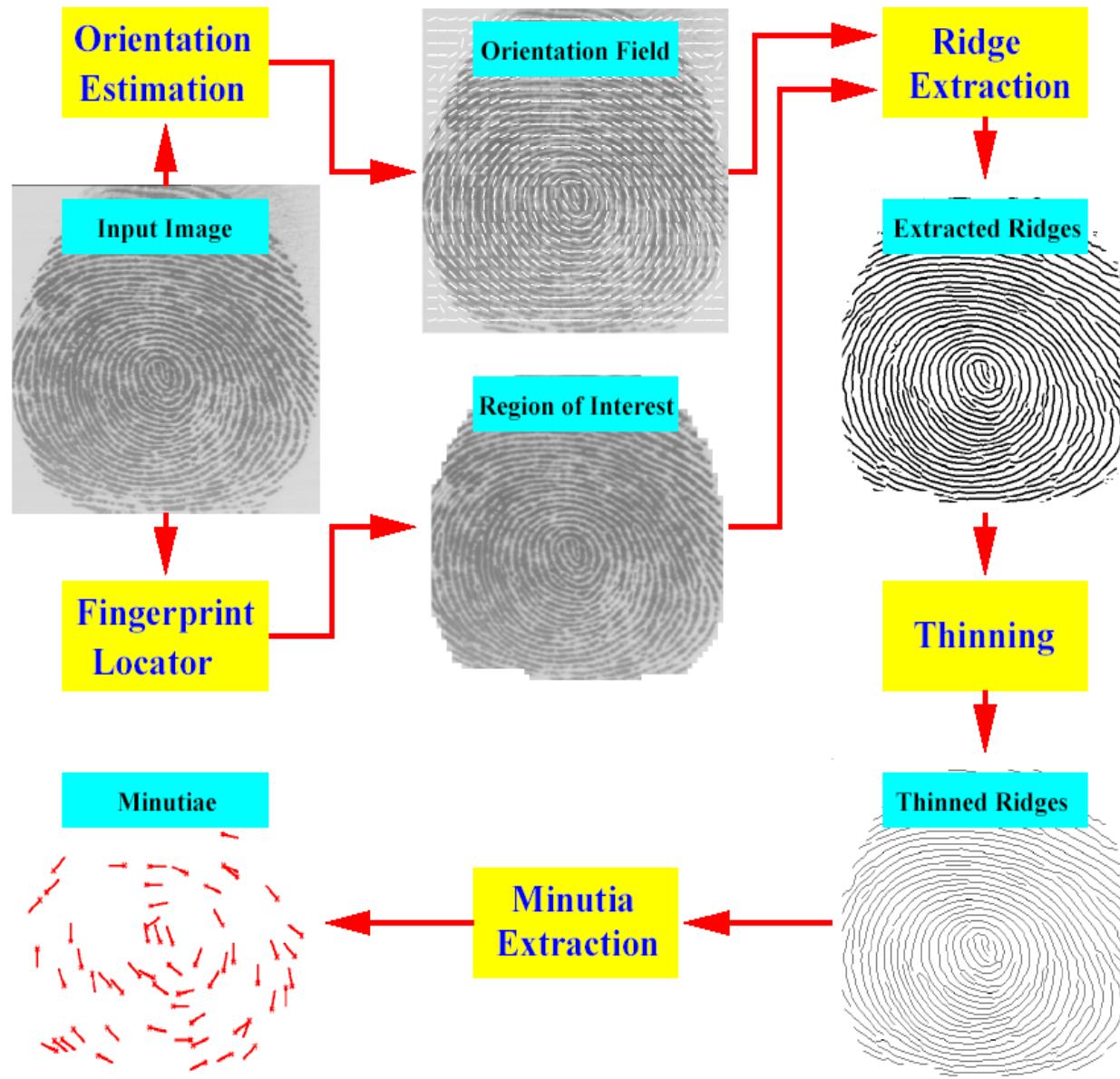
Minutiae Extraction



- Orientation field estimation
- Fingerprint area location
- Ridge extraction
- Thinning
- Minutia extraction



Minutia extraction algorithm



吉祥如意

指纹识别



原图

吉祥如意

指纹识别



分割后

吉祥如意

指纹识别



图像增强

吉祥如意

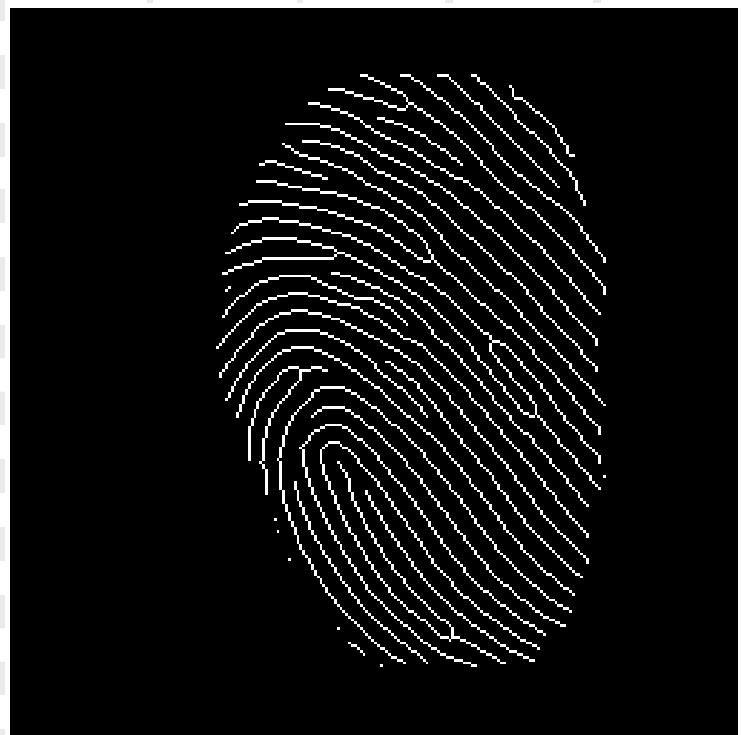
指纹识别



二值化

吉祥如意

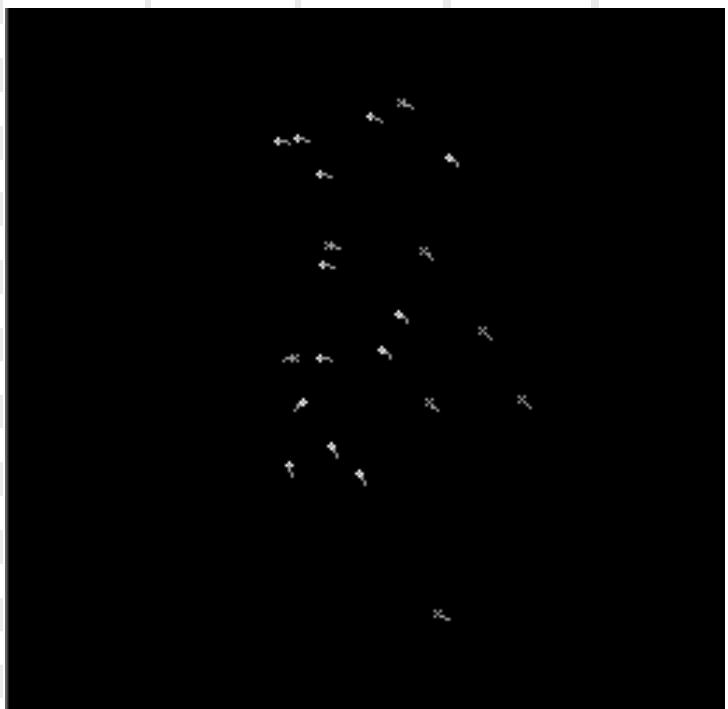
指纹识别



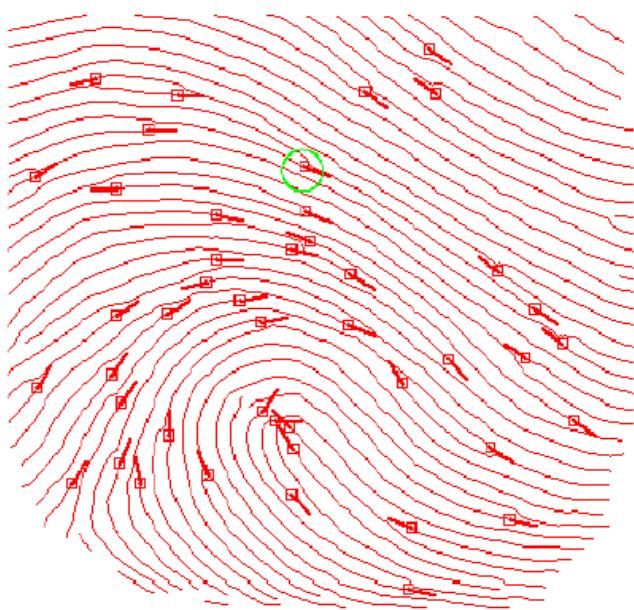
细化

吉祥如意

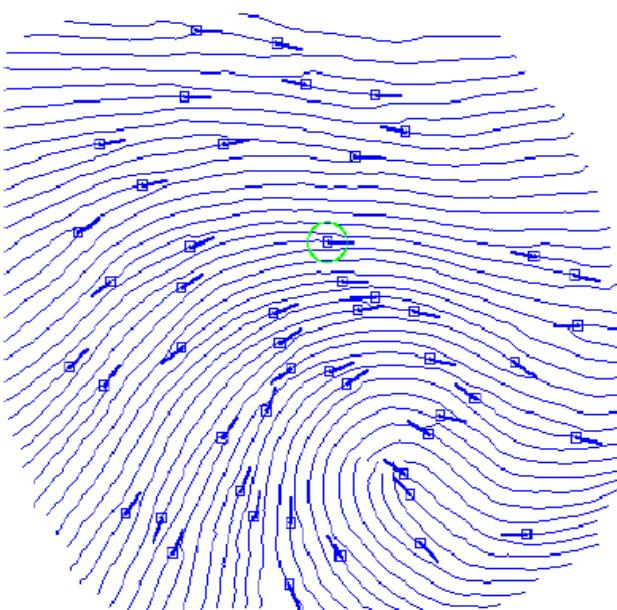
指纹识别



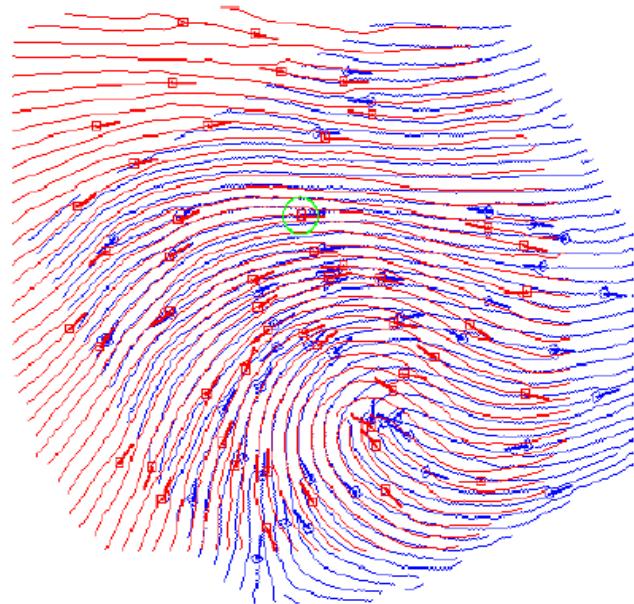
细节点，即模式识别的特征



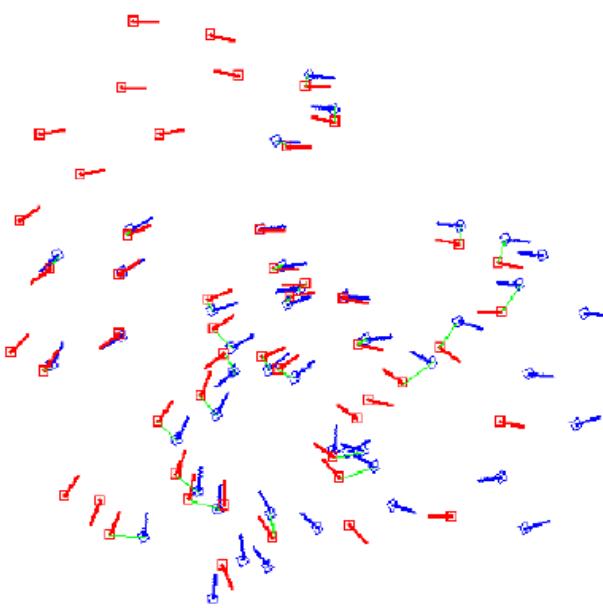
(a)



(b)



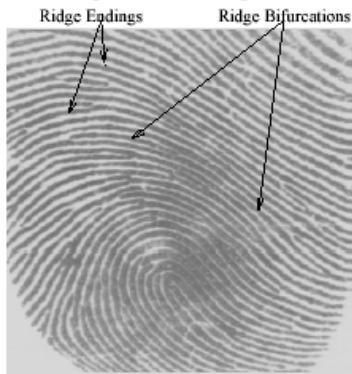
(c)



(d)



Motivation for fingerprint enhancement



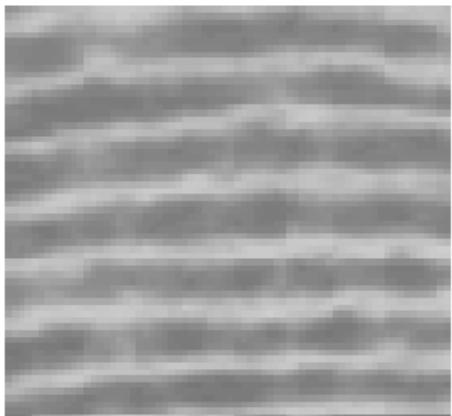
Minutia extracted from a good quality input image



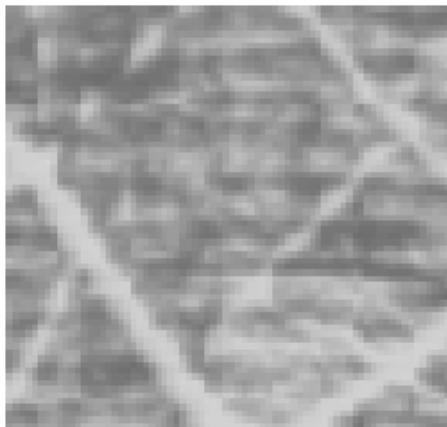
Minutia extracted from a poor quality input image

吉祥圖

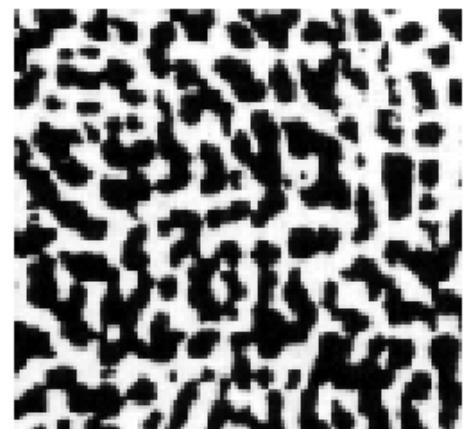
Fingerprint regions



Well defined region

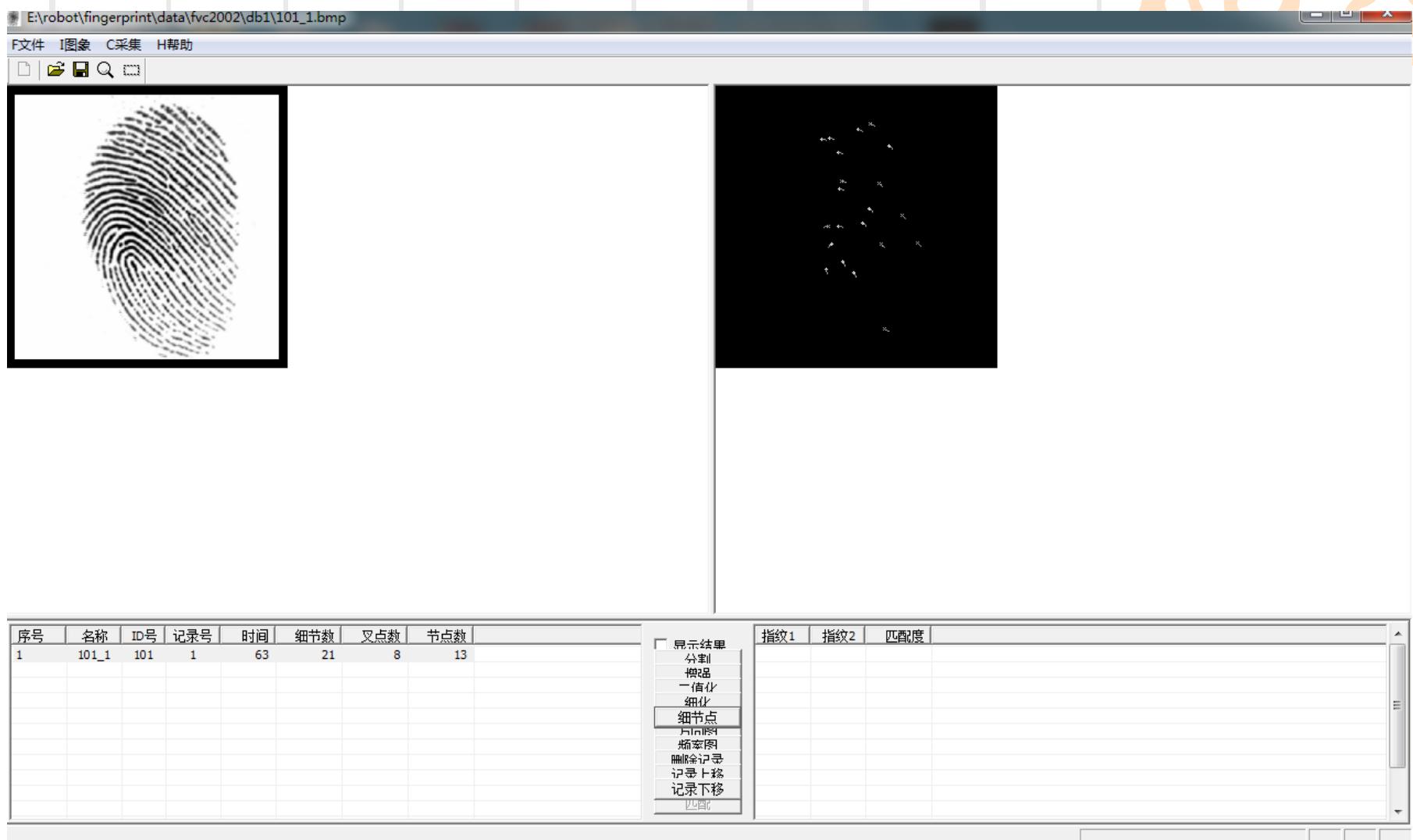


Recoverable
corrupted region

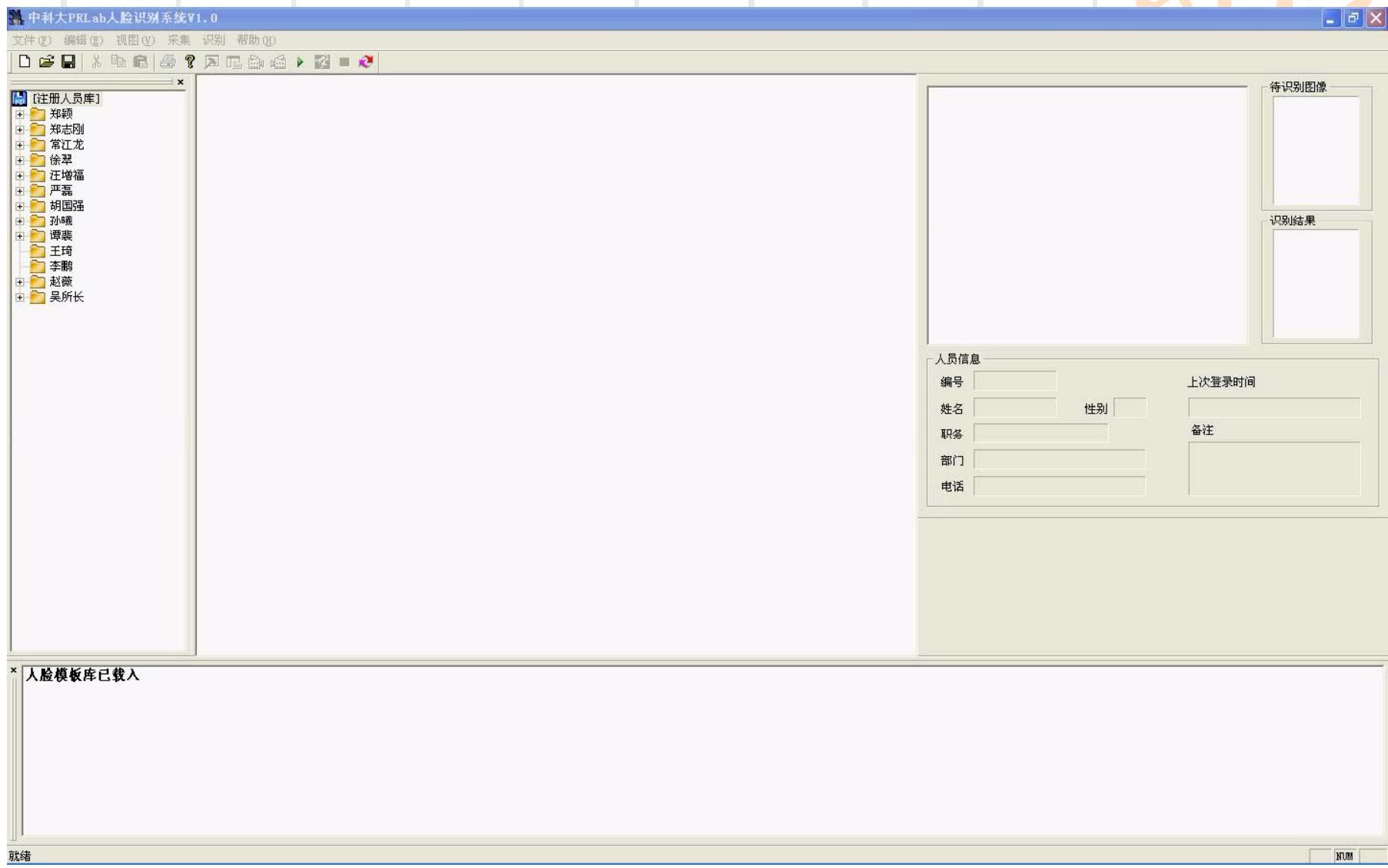


Unrecoverable
corrupted region

指纹识别系统软件界面



实验室示例-人脸识别系统



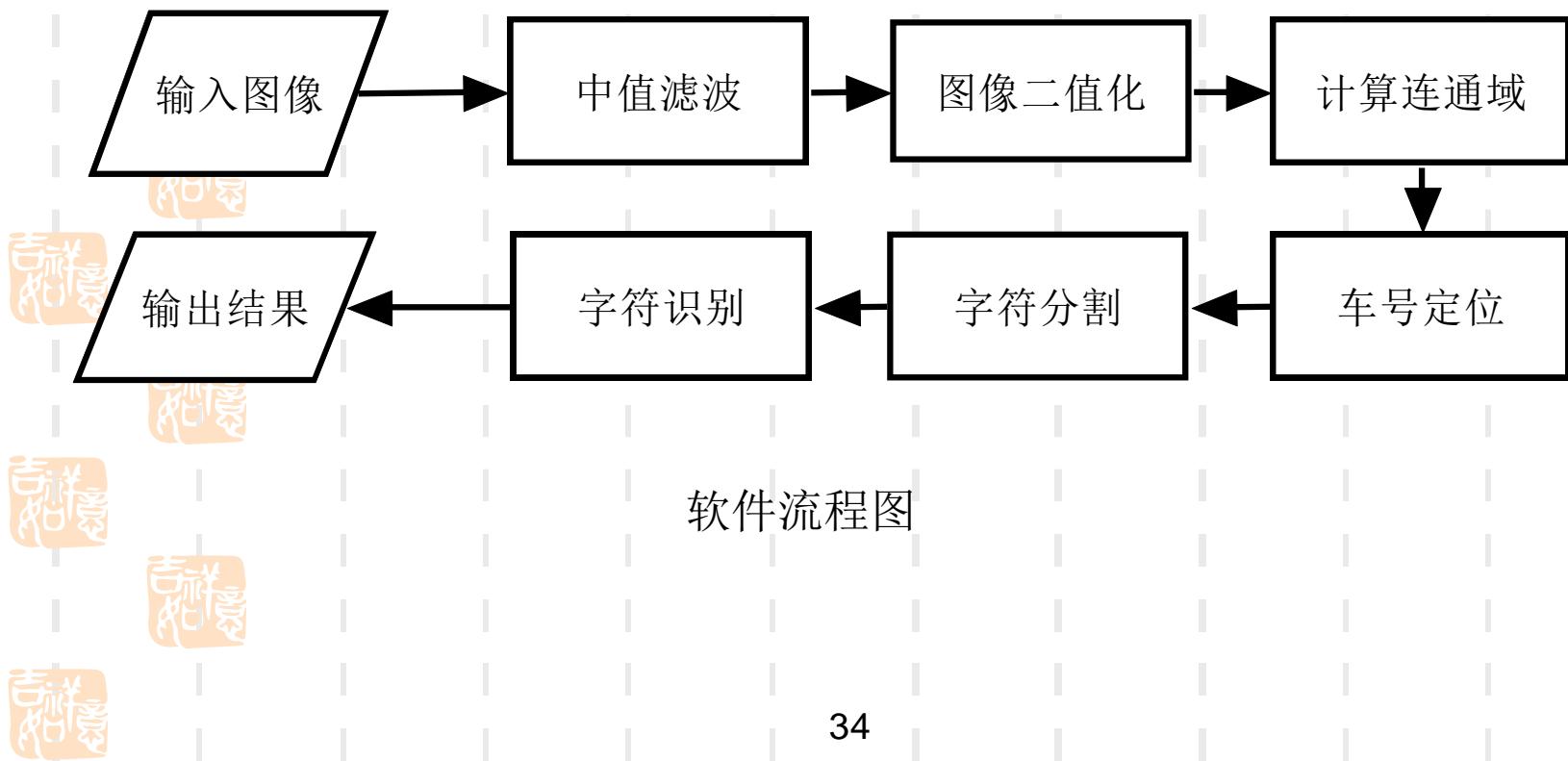
吉祥如意

列车车号识别系统



吉祥如意

列车车号识别系统



软件流程图

吉祥如意

列车车号识别系统

ZE 300601 二三七

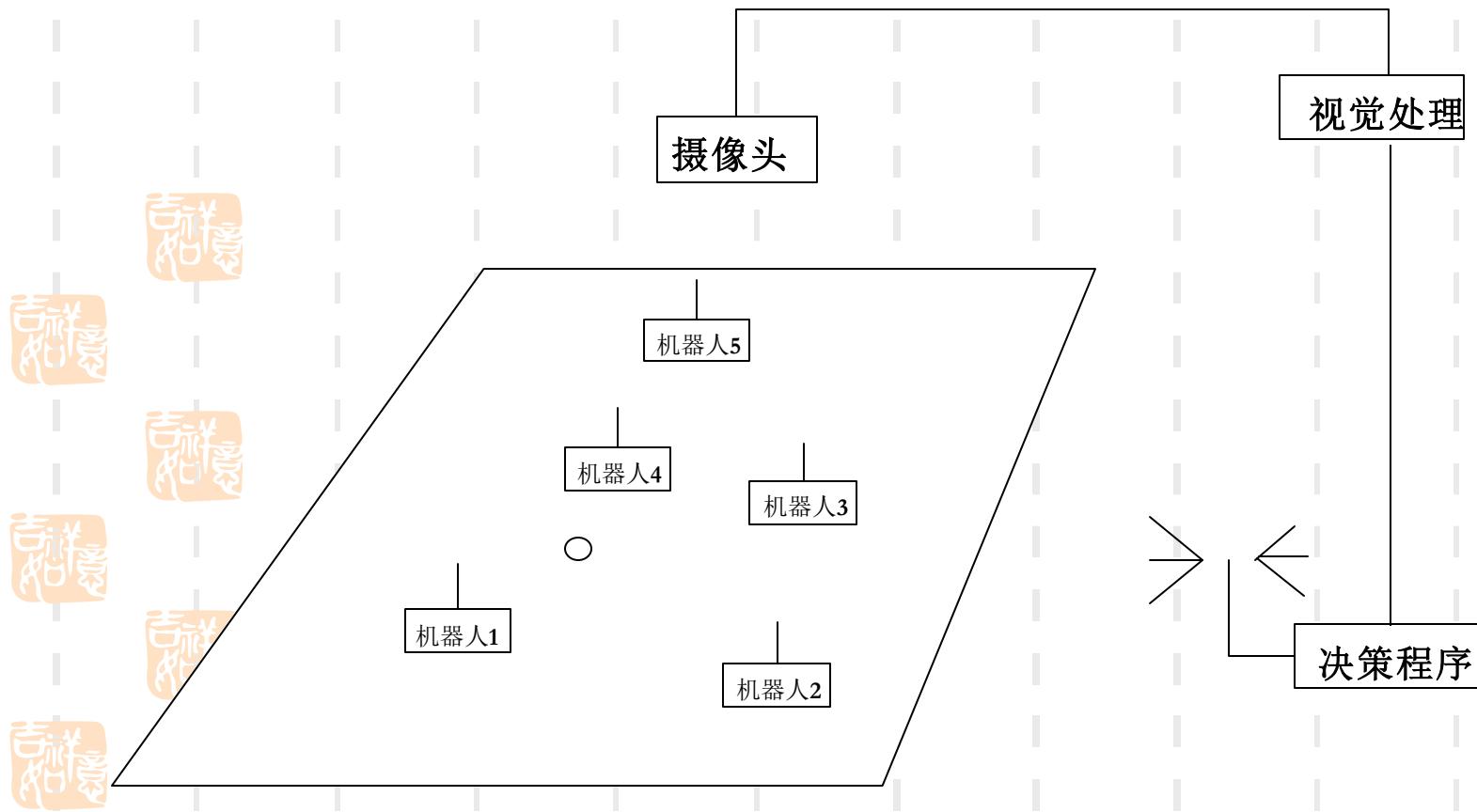


吉祥如意

列车车号识别系统

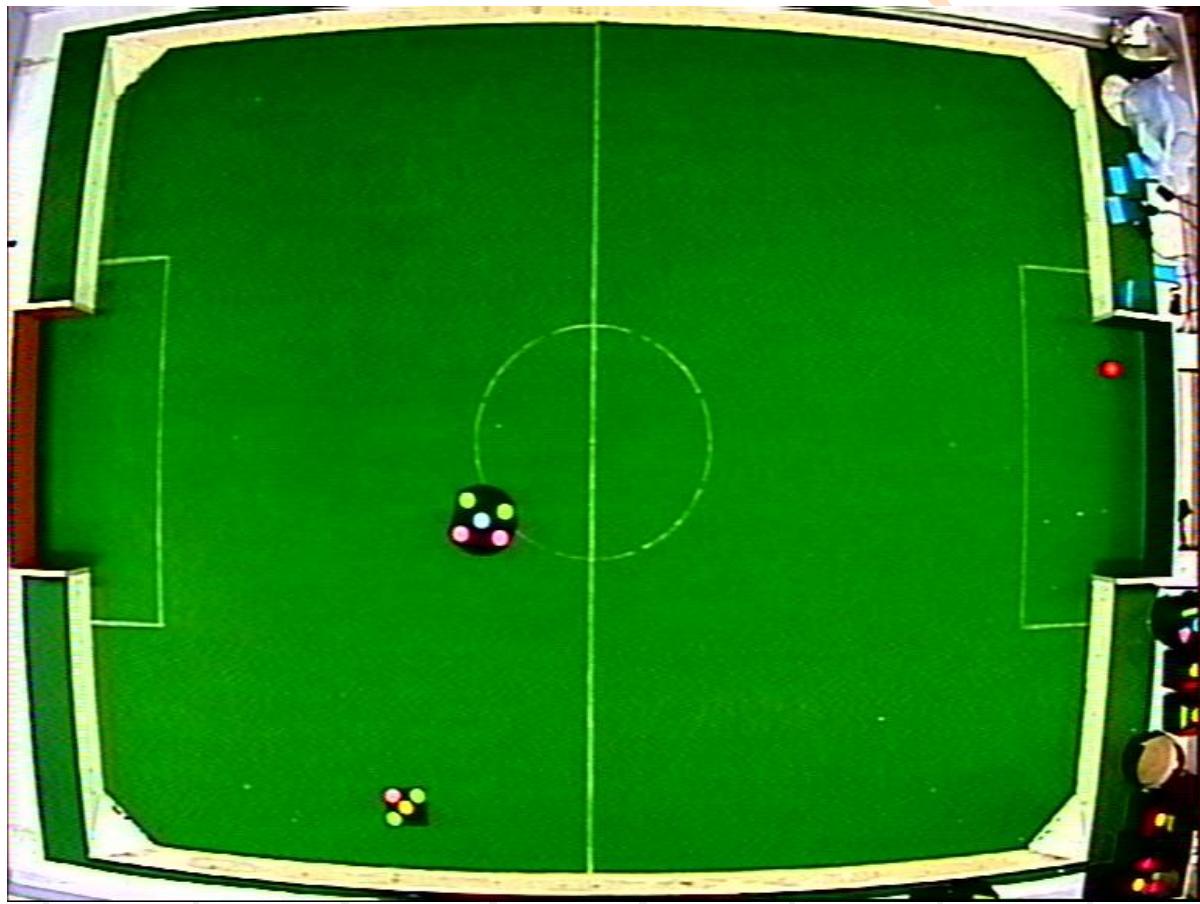
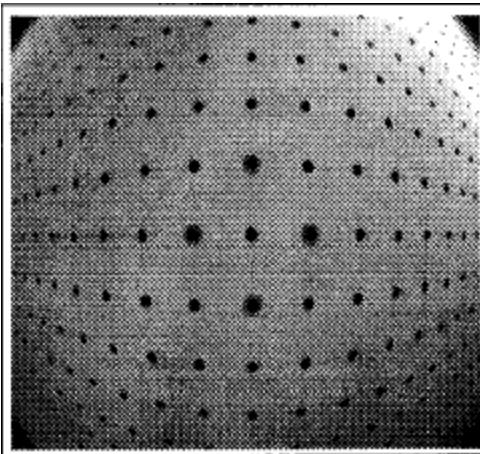


Robocup小型组足球机器人 系统总体框图



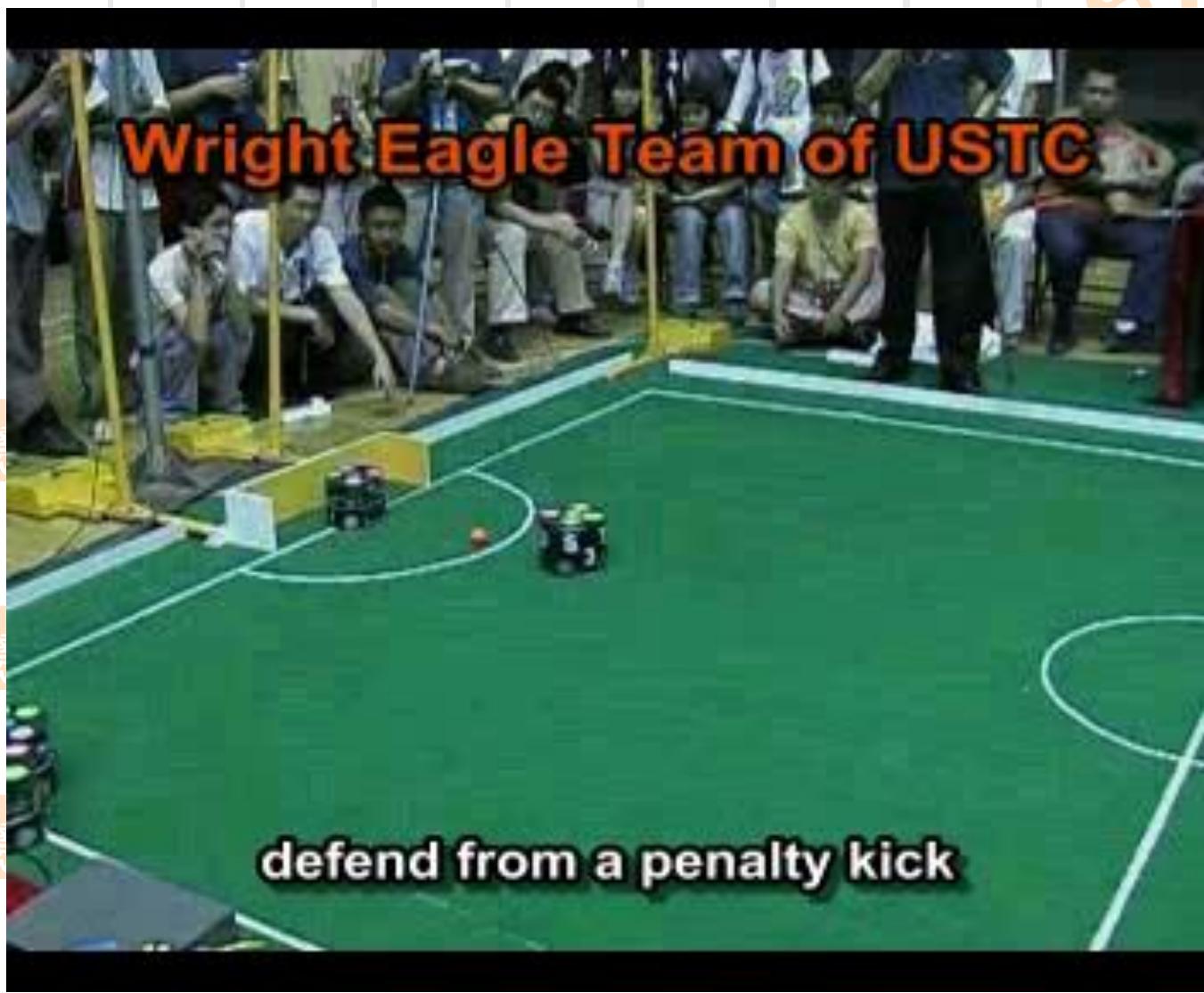
吉祥如意

摄像机鱼眼模型



吉祥如意

比赛视频



吉祥如意



模式空间、特征空间和类型空间

- **物理上可以察觉到的世界：**在客观世界里存在一些物体和事件，它们都可被适当的和足够多的函数来描述，也就是说它们在物理上是可以被测量的，他们的可测数据的集合就称为物理上觉察到的世界。显然，这些可测数据，或者说这个世界的维数是无限多的

样本：在物理上可以觉察到的世界中，所选择出的某些物体和事件

模式空间、特征空间和类型空间

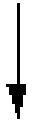


- **模式采集：**由物理上可觉察到的世界到模式空间所经历的过程
- **模式空间的维数：**与选择的样本和测量方法有关，也与特定应用有关。维数很大，但是是一个有限值。

在模式空间里，每个模式样本都是一个点，点的位置由该模式在各维上的测量值确定

模式空间、特征空间和类型空间

从模式空间 → 特征空间（特征空间的维数被大大压缩了远小于模式空间）



特征空间中每个坐标都是样本的重要特征



样本在特征空间中也是一个点，位置由样本的特征值决定

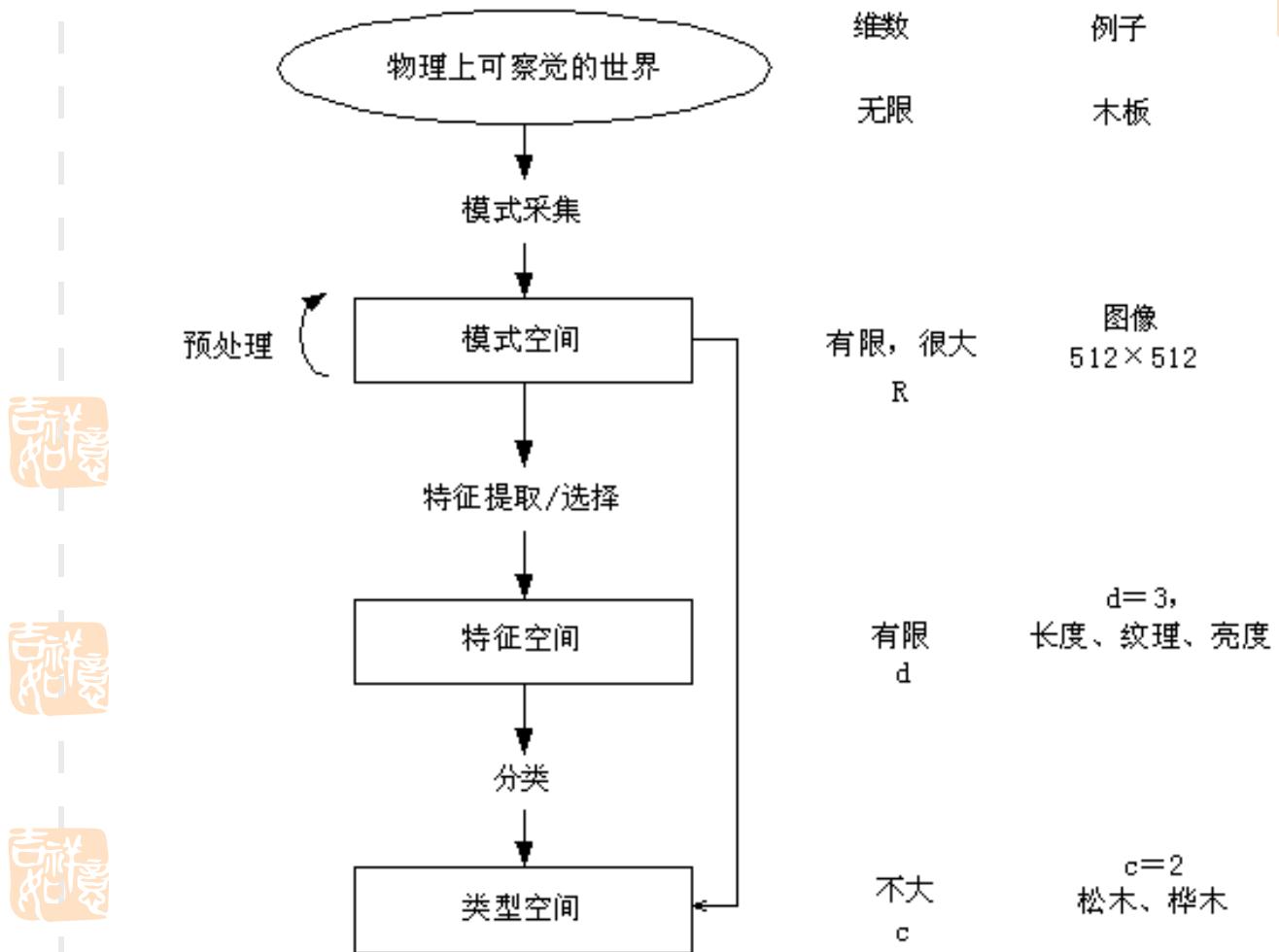
■ **特征提取和特征选择：**从模式空间到特征空间所需的变换和选择

■ **判决规则：**由某些知识和经验可以确定的分类准则根据适当的判决规则，把特征空间里的样本区分成不同的类型，从而把特征空间塑造成了——**类型空间**。

■ **决策面：**不同类型之间的分界面

吉祥如意

模式识别过程



一般来说： $c < d \ll R \ll \text{无限}$

吉祥如意

模式识别方法

- 模式识别系统的目标：在特征空间和解释空间之间找到一种映射关系，这种映射也称之为假说。
 - 特征空间：从模式得到的对分类有用的度量、属性或基元构成的空间。
 - 解释空间：将 c 个类别表示为 $\omega_i \in \Omega, i = 1, 2, \dots, c$ 其中 Ω 为所属类别的集合，称为解释空间。



假说的两种获得方法

- 监督学习、概念驱动或归纳假说：在特征空间中找到一个与解释空间的结构相对应的假说。在给定模式下假定一个解决方案，任何在训练集中接近目标的假说也都必须在“未知”的样本上得到近似的结果。
 - 依靠已知所属类别的训练样本集，按它们特征向量的分布来确定假说（通常为一个判别函数），只有在判别函数确定之后才能用它对未知的模式进行分类；
 - 对分类的模式要有足够的先验知识，通常需要采集足夠数量的具有典型性的样本进行训练。





假说的两种获得方法（续）

- 非监督学习、数据驱动或演绎假说：在解释空间中找到一个与特征空间的结构相对应的假说。这种方法试图找到一种只以特征空间中的相似关系为基础的有效假说。
 - 在没有先验知识的情况下，通常采用聚类分析方法，基于“物以类聚”的观点，用数学方法分析各特征向量之间的距离及分散情况；
 - 如果特征向量集聚若干个群，可按群间距离远近把它们划分成类；
 - 这种按各类之间的亲疏程度的划分，若事先能知道应划分成几类，则可获得更好的分类结果。



吉祥如意

模式分类的主要方法

- 数据聚类
- 统计分类
- 结构模式识别
- 神经网络



吉祥如意

数据聚类

- 目标：用某种相似性度量的方法将原始数据组织成有意义的和有用的各种数据集。
- 是一种非监督学习的方法，解决方案是数据驱动的。



吉祥如意

统计分类

- 基于概率统计模型得到各类别的特征向量的分布，以取得分类的方法。
- 特征向量分布的获得是基于一个类别已知的训练样本集。

■ 是一种监督分类的方法，分类器是概念驱动的。



吉祥如意

结构模式识别

- 该方法通过考虑识别对象的各部分之间的联系来达到识别分类的目的。
- 识别采用结构匹配的形式，通过计算一个匹配程度值（*matching score*）来评估一个未知的对象或未知对象某些部分与某种典型模式的关系如何。
- 当成功地制定出了一组可以描述对象部分之间关系的规则后，可以应用一种特殊的结构模式识别方法 – 句法模式识别，来检查一个模式基元的序列是否遵守某种规则，即句法规则或语法。

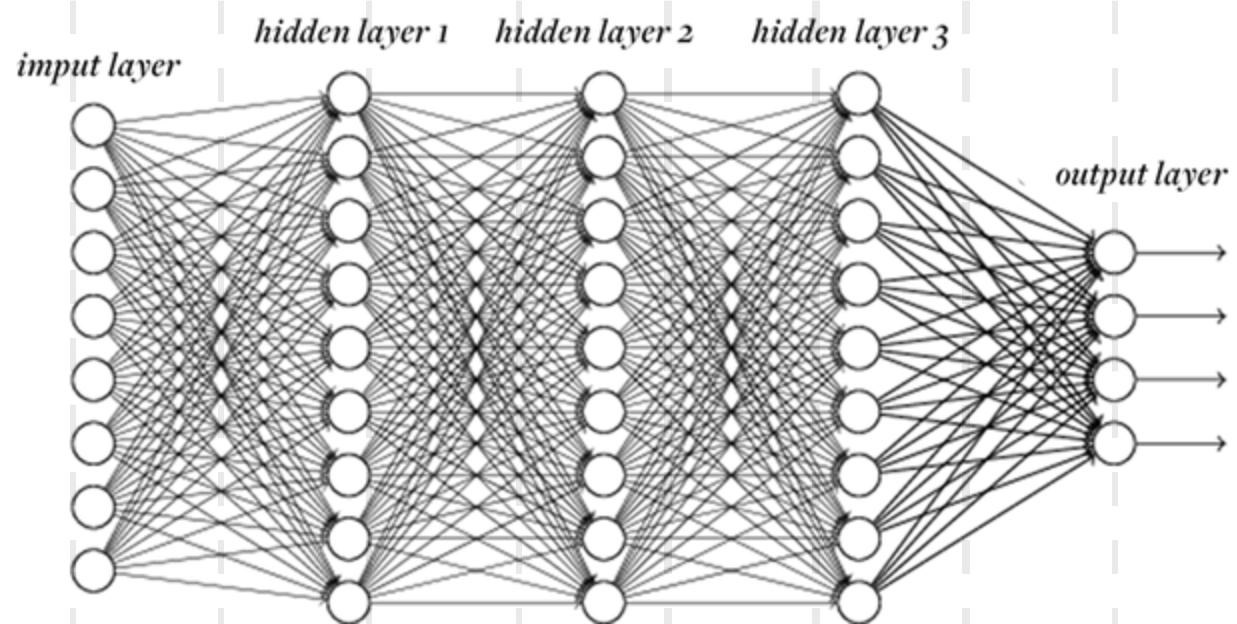
吉祥如意

神经网络

- 神经网络是受人脑组织的生理学启发而创立的。
- 由一系列互相联系的、相同的单元（神经元）组成。相互间的联系可以在不同的神经元之间传递增强或抑制信号。
- 增强或抑制是通过调整神经元相互间联系的权重系数来（weight）实现。
- 神经网络可以实现监督和非监督学习条件下的分类。

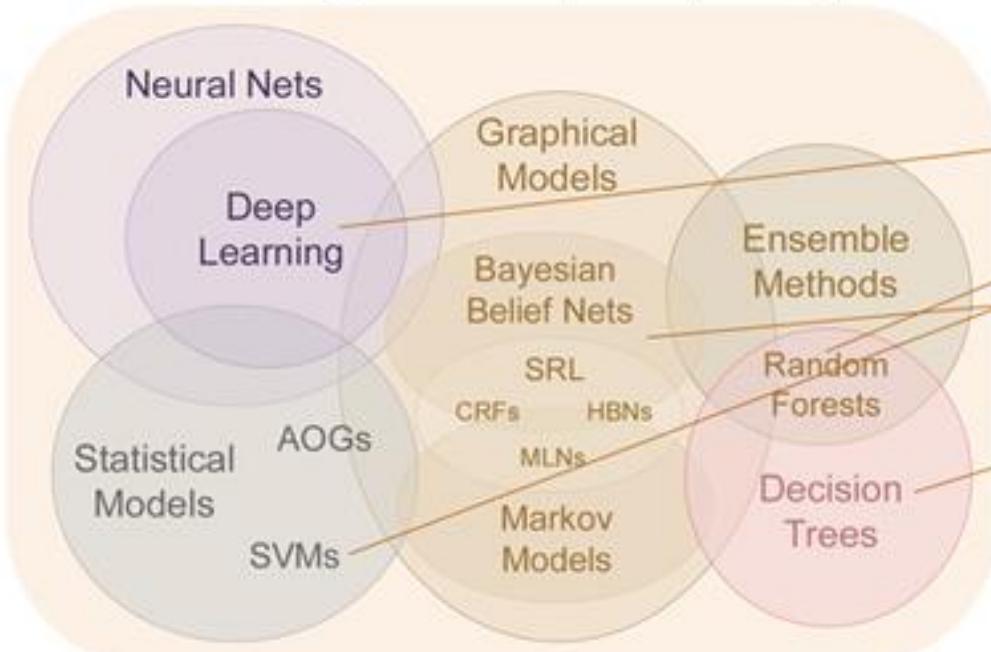


吉祥如意





Learning Techniques (today)



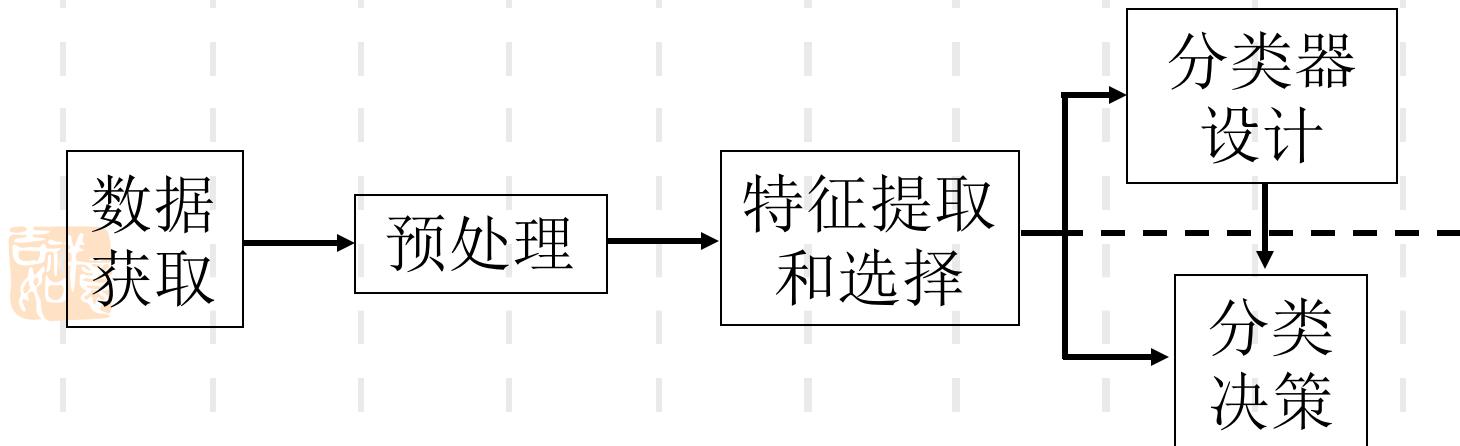
Explainability (notional)



吉祥如意

模式识别系统

- 模式识别系统的基本构成



模式识别系统组成单元

- 数据获取：用计算机可以运算的符号来表示所研究的对象
 - 二维图像：文字、指纹、地图、照片等
 - 一维波形：脑电图、心电图、季节震动波形等
 - 物理参量和逻辑值：体温、化验数据、参量正常与否的描述
- 预处理单元：去噪声，提取有用信息，并对输入测量仪器或其它因素所造成的退化现象进行复原



模式识别系统组成单元

- 特征提取和选择：对原始数据进行变换，得到最能反映分类本质的特征

➤ 目的：实现由模式空间到特征空间的转变，有效压缩维数

- 分类决策：在特征空间中用模式识别方法把被识别对象归为某一类别

➤ 基本做法：在样本训练集基础上确定某个判决规则，使得按这种规则对被识别对象进行分类所造成的错误识别率最小或引起的损失最小

吉祥如意

模式识别过程实例

- 在传送带上用光学传感器器件对鱼按品种分类



品种

鲈鱼 (Seabass)

鲑鱼 (Salmon)



吉
祥
魚

识别过程

- 数据获取：架设一个摄像机，采集一些样本图像，获取样本数据
- 预处理：去噪声，用一个分割操作把鱼和鱼之间以及鱼和背景之间分开



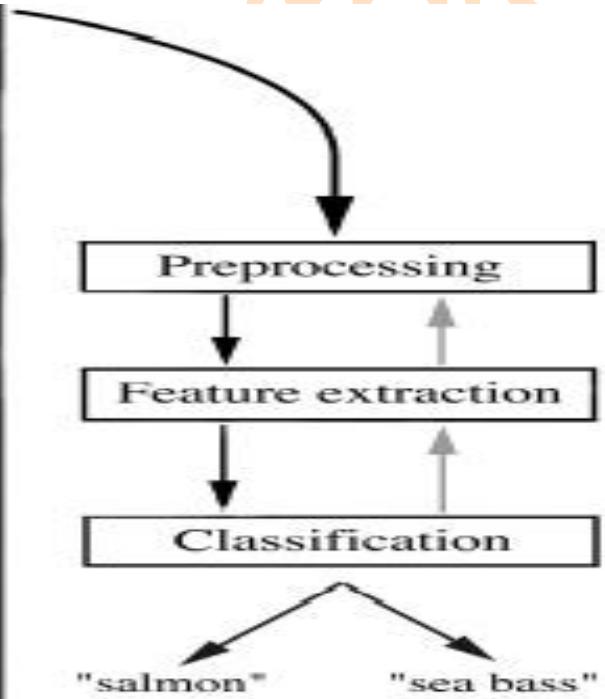
吉
祥
魚

识别过程

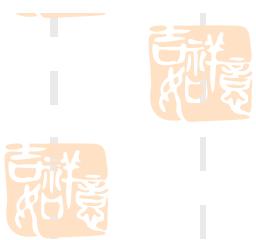
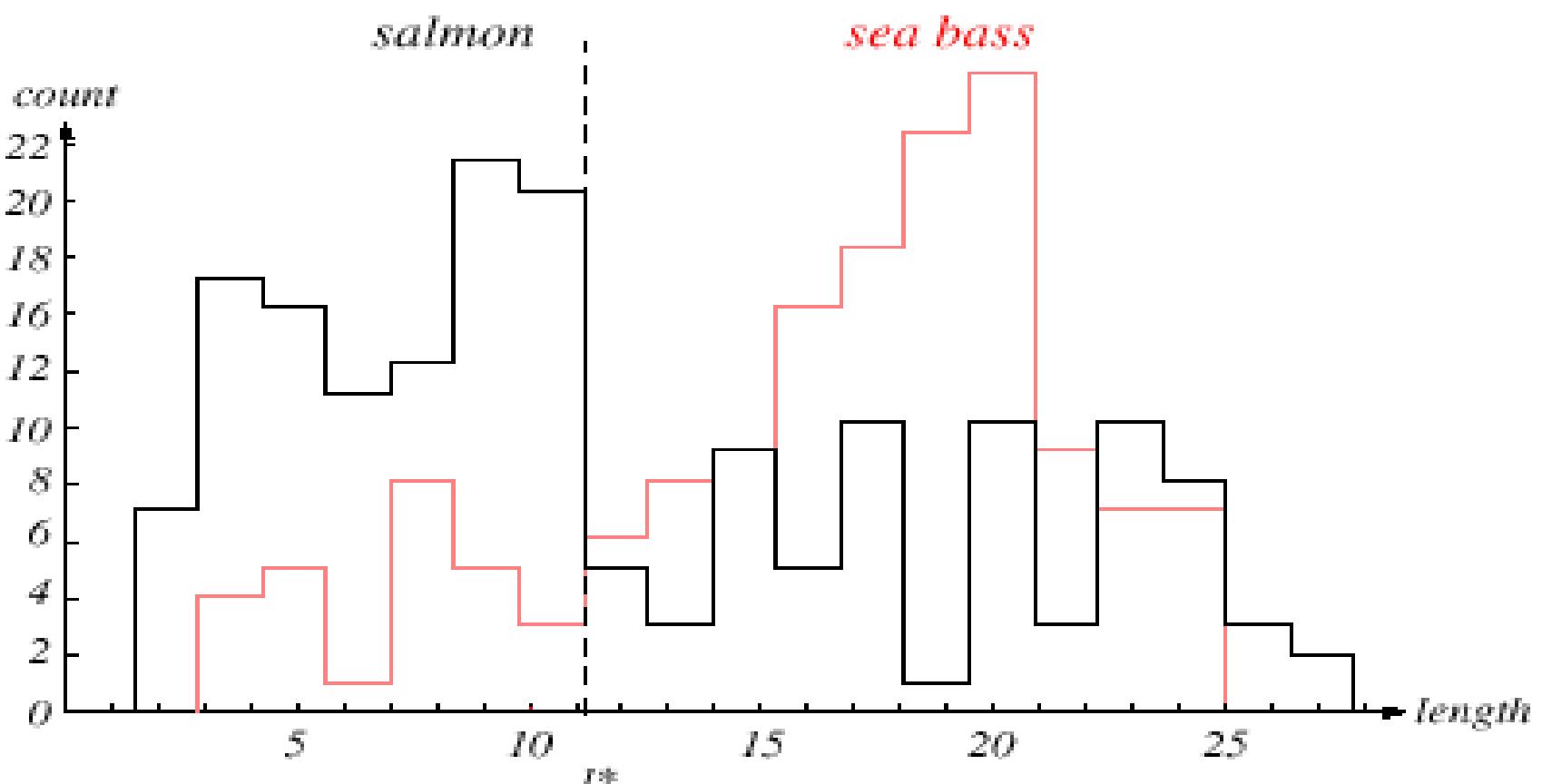
- 特征提取和选择：对单个鱼的信息进行特征选择，从而通过测量某些特征来减少信息量
 - 长度
 - 亮度
 - 宽度
 - 鱼翅的数量和形状
 - 嘴的位置，等等 ...

分类决策：把特征送入决策分类器

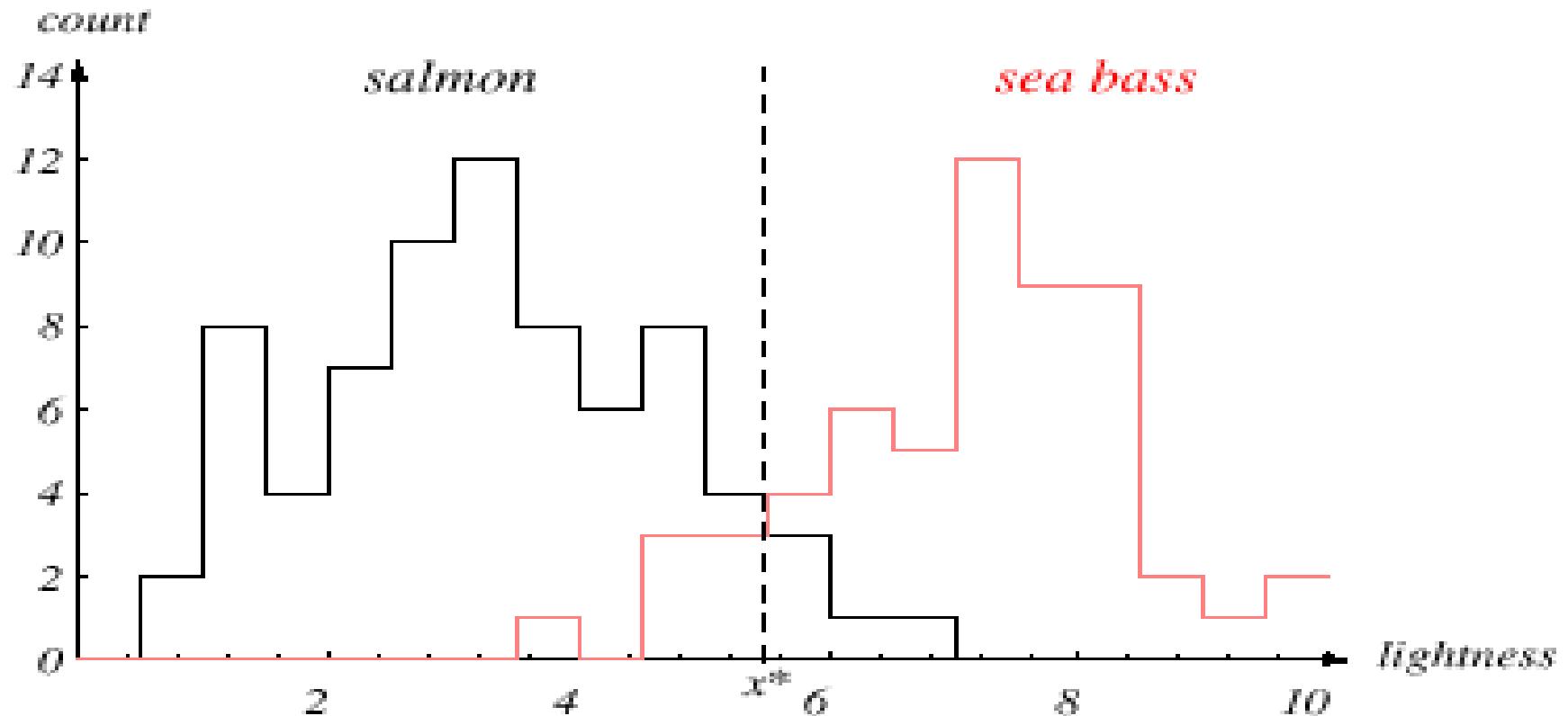
吉辭密



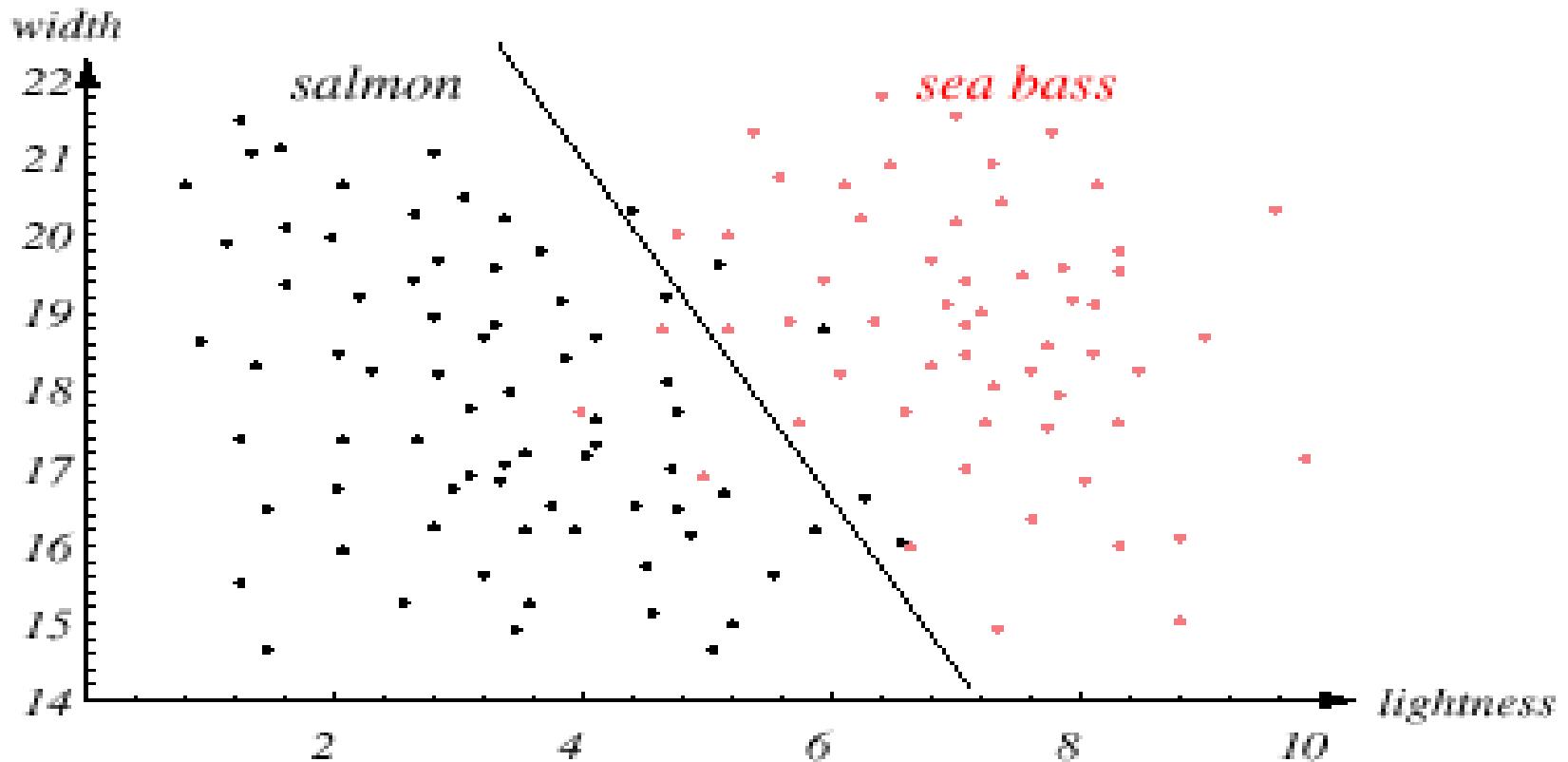
吉祥圖



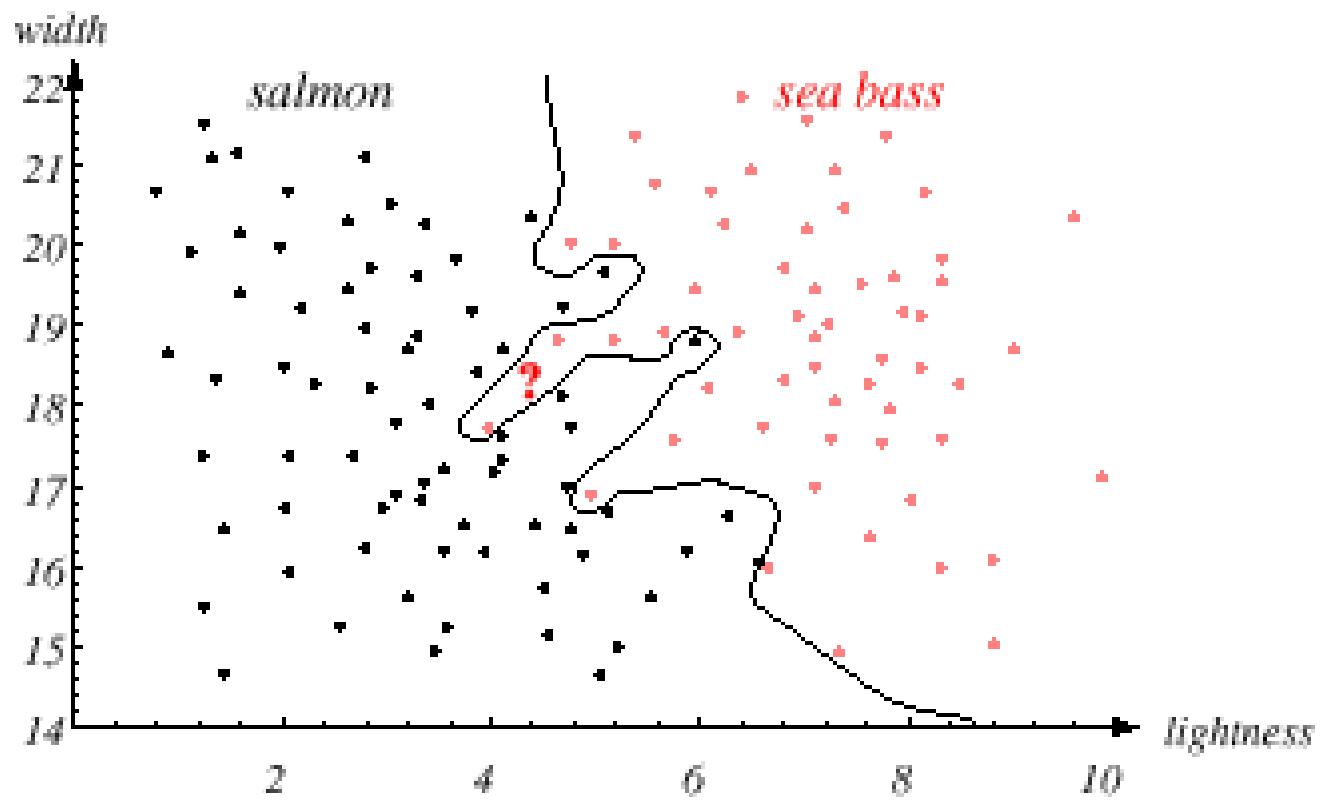
吉祥圖



吉祥圖



吉祥如意



吉祥齋

width

22
21
20
19
18
17
16
15
14

salmon

sea bass

lightness

2 4 6 8 10

吉祥齋



吉祥齋

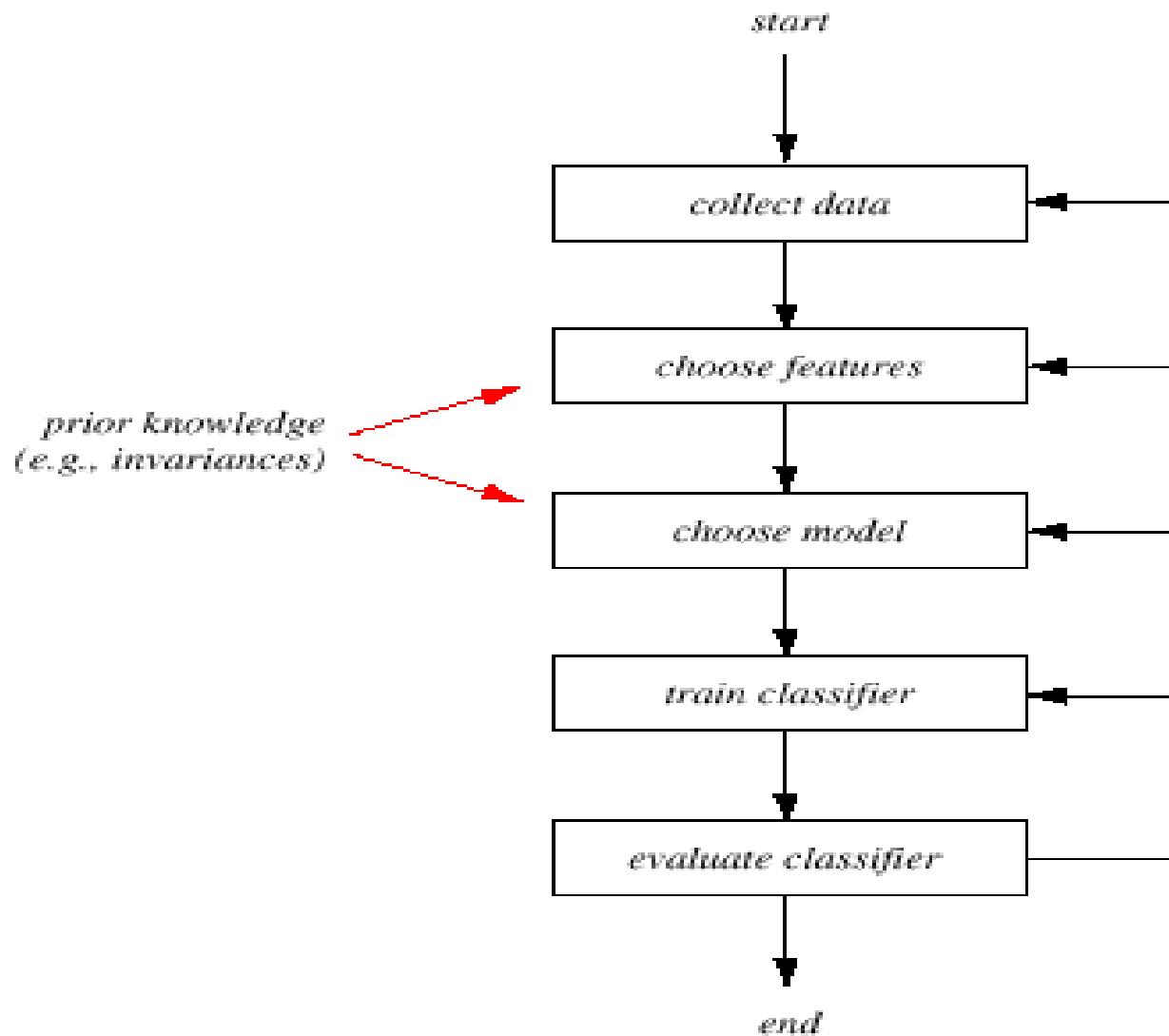
模式分类器的获取和评测过程

- 数据采集
- 特征选取
- 模型选择

■ 训练和测试

- 计算结果和复杂度分析，反馈

吉祥



吉祥如意

训练和测试

- 训练集：是一个已知样本集，在监督学习方法中，用它来开发出模式分类器。
- 测试集：在设计识别和分类系统时没有用过的独立样本集。
- 系统评价原则：为了更好地对模式识别系统性能进行评价，必须使用一组独立于训练集的测试集对系统进行测试。

吉
祥
體
驗

实例：统计模式识别

- 19名男女同学进行体检，测量了身高和体重，但事后发现其中有4人忘记填写性别，试问（在最小错误的条件下）这4人是男是女？体检数值如下：

编 号	身 高(cm)	体 重(kg)	性 别	编 号	身 高(cm)	体 重(kg)	性 别
1	170	68	男	11	140	62	男
2	130	66	女	12	150	64	女
3	180	71	男	13	120	66	女
4	190	73	男	14	150	66	男
5	160	70	女	15	130	65	男
6	150	66	男	16	140	70	?
7	190	68	男	17	150	60	?
8	210	76	男	18	145	65	?
9	100	58	女	19	160	75	?
10	170	75	男				



实例：统计模式识别（续）

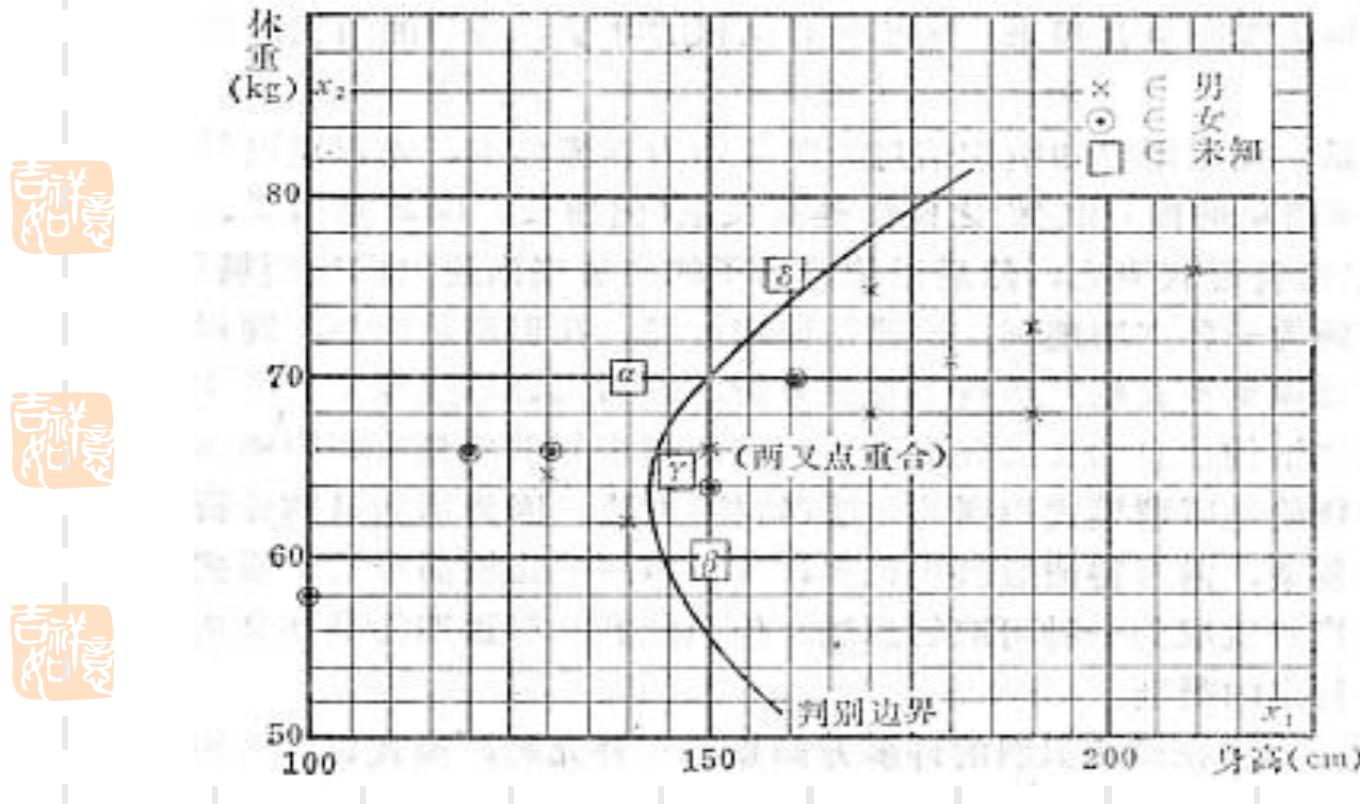
- 待识别的模式：性别（男或女）
- 测量的特征：身高和体重
- 训练样本：15名已知性别的样本特征
- 目标：希望借助于训练样本的特征建立判别函数（即数学模型）



吉祥如意

实例：统计模式识别（续）

- 由训练样本得到的特征空间分布图



吉
祥
慶

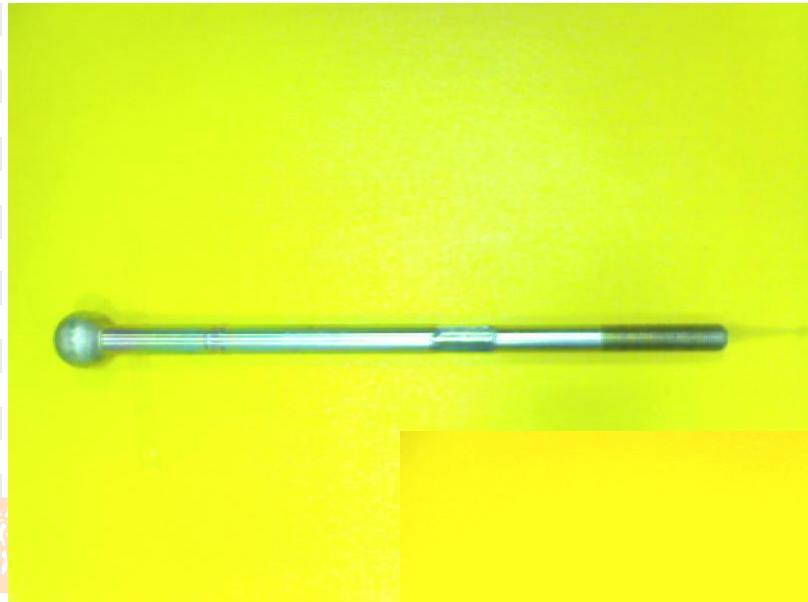
实例：统计模式识别（续）

- 从图中训练样本的分布情况，找出男、女两类特征各自的聚类特点，从而求取一个判别函数（直线或曲线）。
- 只要给出待分类的模式特征的数值，看它在特征平面上落在判别函数的哪一侧，就可以判别是男还是女了



吉祥如意

实例：统计模式识别（续）



吉祥如意

吉祥如意

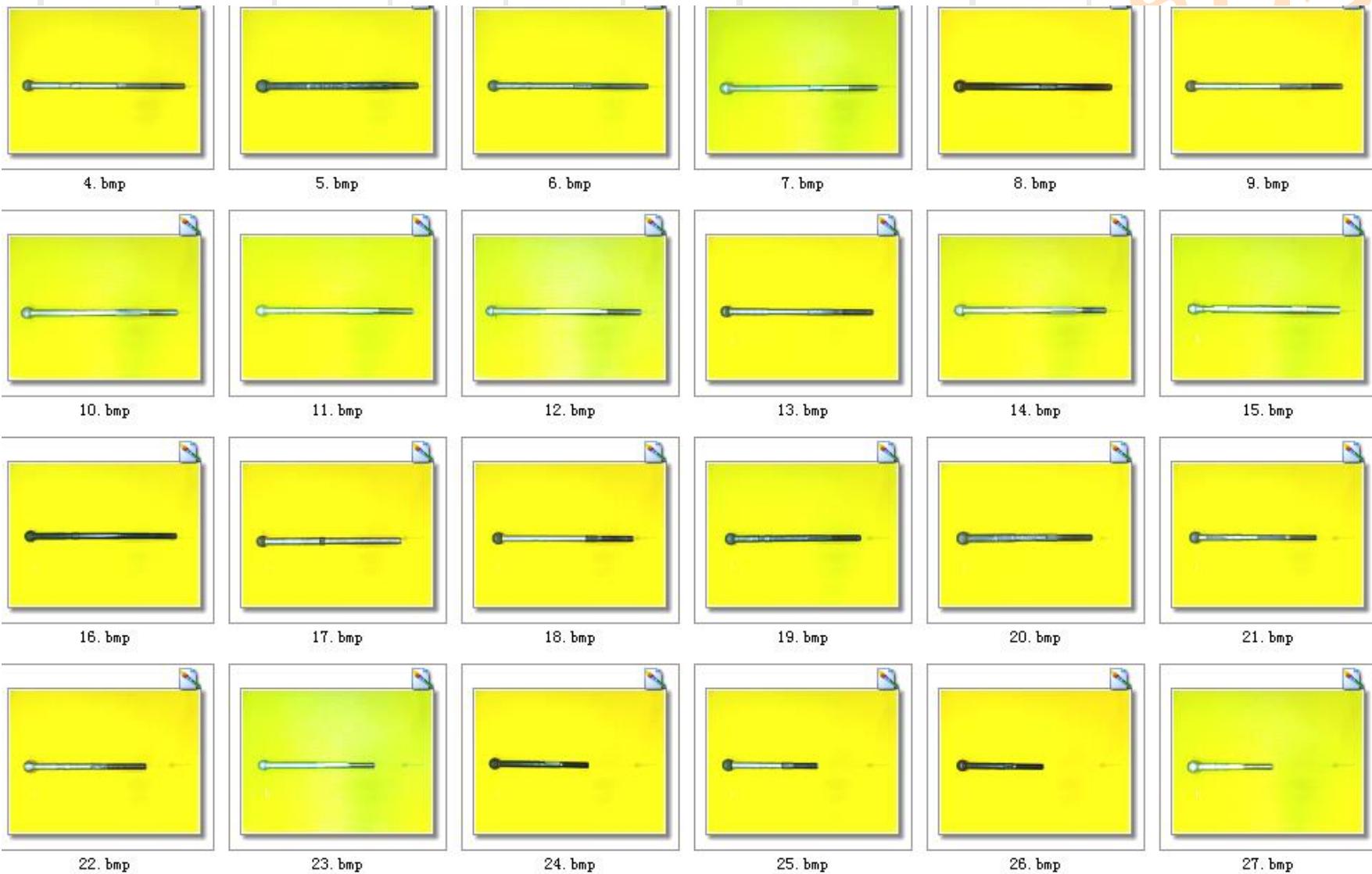
吉祥如意

吉祥如意

吉祥如意

吉祥如意

实例：统计模式识别（续）

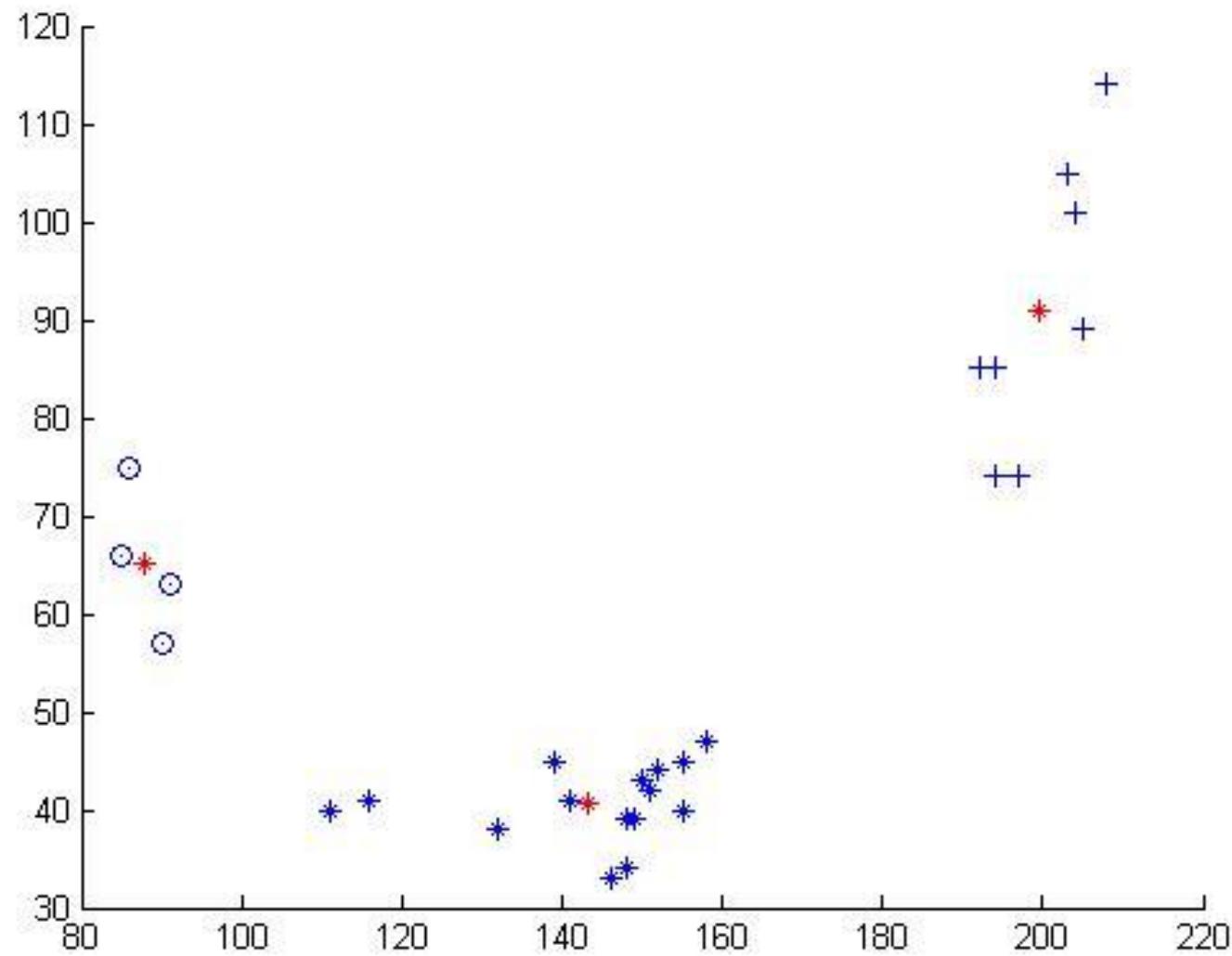


实例：统计模式识别（续）

			R	G	B	H	S	V
0	39.15000	0	37.29880	799	97	139	107	139
1	33.90000	0	33.49020	718	110	151	124	151
2	33.00000	0	32.73789	702	112	155	133	155
3	32.60000	0	32.26769	692	111	149	131	149
4	32.40000	0	32.26769	692	81	111	88	111
5	32.30000	0	31.89153	684	102	141	116	141
6	31.60000	1	31.42133	674	129	192	145	192
7	31.50000	2	31.37431	673	47	86	82	86
8	31.20000	0	31.23325	670	118	155	141	155
9	31.10000	1	31.23325	670	133	194	166	194
10	31.10000	1	31.28027	671	132	208	171	208
11	31.00000	1	30.90412	663	144	194	163	194
12	30.30000	0	30.29286	650	109	152	131	152
13	30.10000	1	30.29286	650	150	197	172	197
14	29.80000	1	30.19882	648	137	204	170	204
15	29.90000	2	30.15180	647	57	90	72	90
16	28.50000	0	28.74121	617	114	158	146	158
17	27.80000	0	28.03591	602	109	148	135	148
18	26.70000	0	26.90744	578	84	116	91	116
19	26.60000	0	27.00148	580	98	132	112	132
20	25.00000	0	25.21473	542	115	148	127	148
21	24.20000	0	24.32136	523	108	150	136	150
22	23.10000	1	23.14586	498	151	205	162	205
23	19.80000	2	19.66641	424	55	91	76	91
24	19.10000	0	18.96111	409	114	146	125	146
25	17.40000	2	17.12734	370	50	85	77	85
26	17.10000	1	16.93926	366	130	203	180	203

吉祥如意

实例：统计模式识别（续）



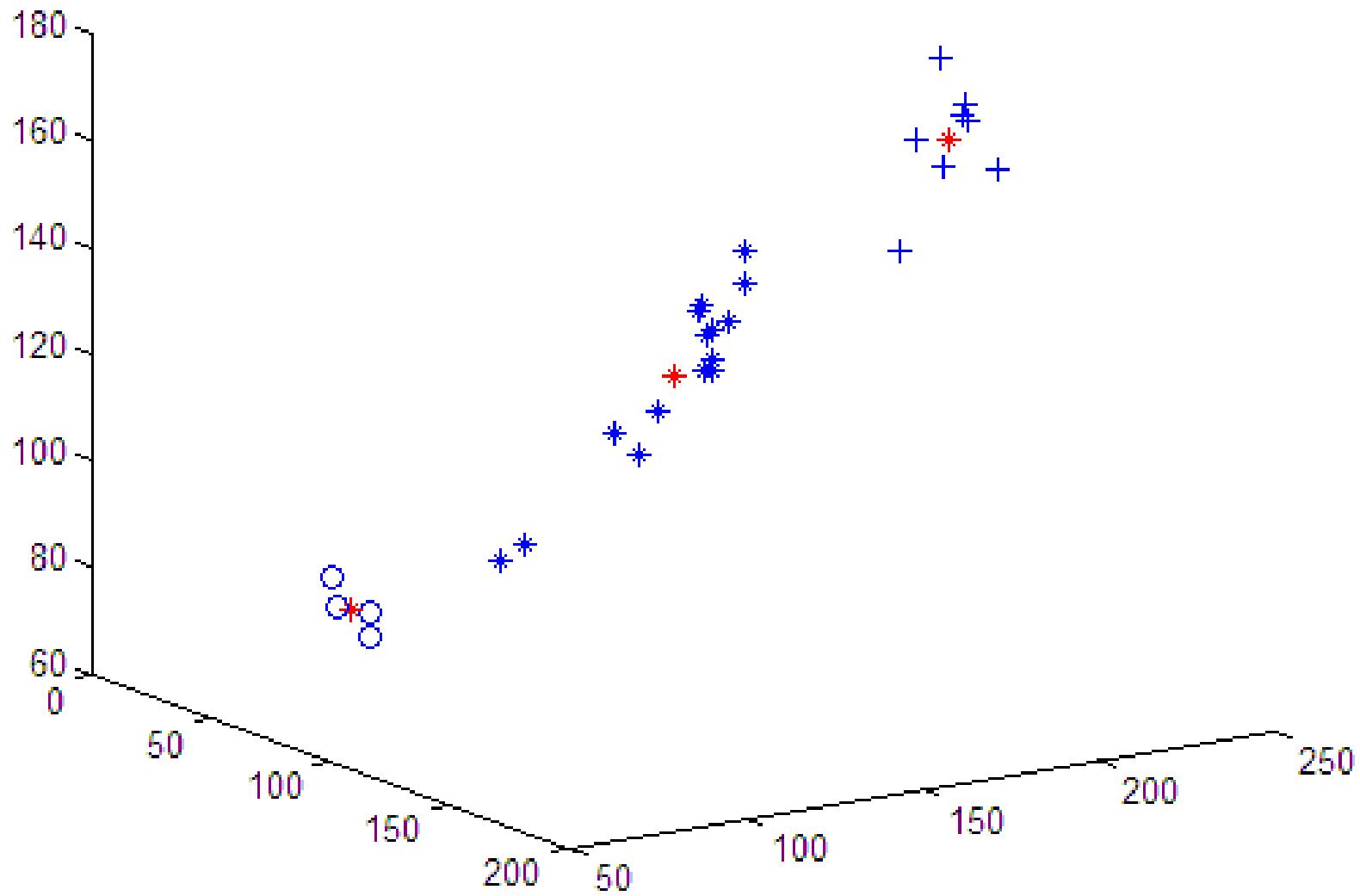
吉

祥

如



DATA



吉
祥
慶

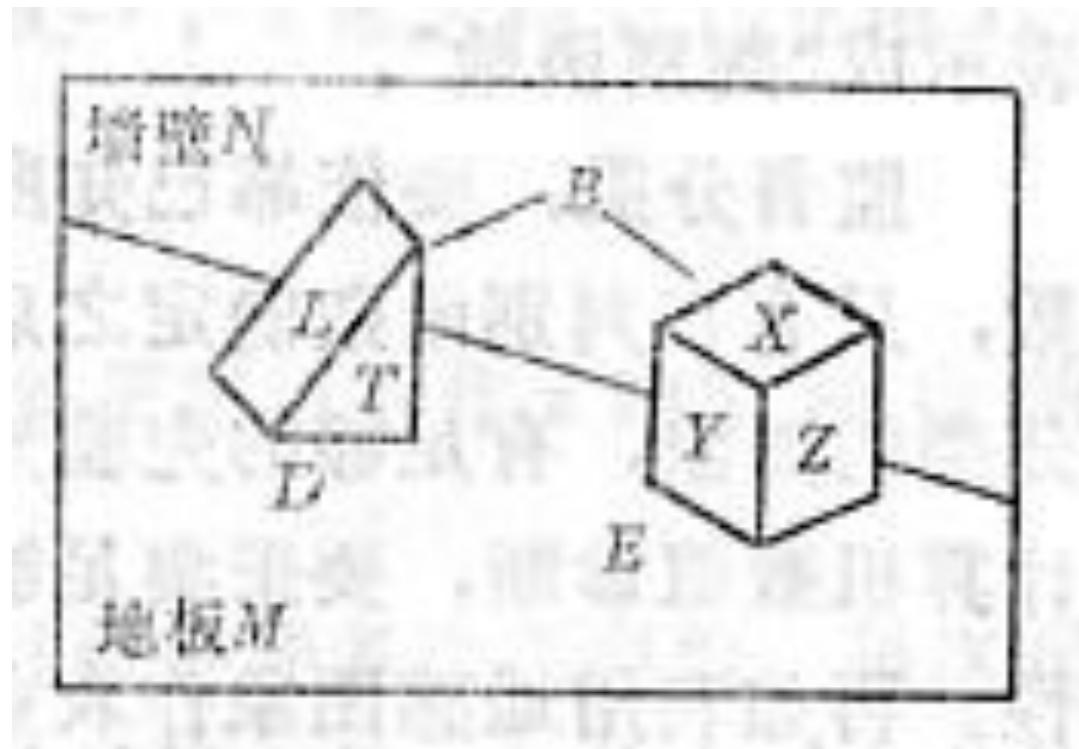
实例：工业检测的例子



吉祥如意

实例：句法模式识别

- 问题：如何利用对图像的结构信息描述，识别如下所示图片：





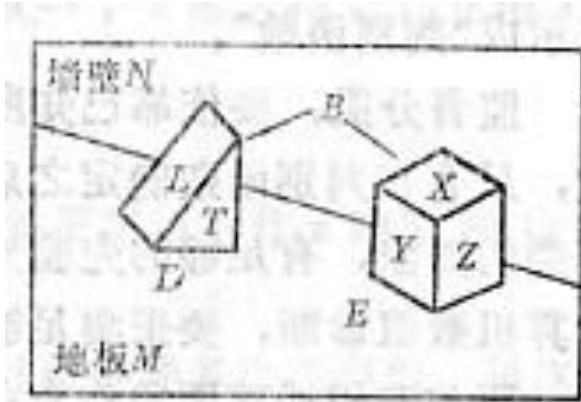
实例：句法模式识别（续）

- 将整个场景图像结构分解成一些比较简单的子图像的组合；
- 子图像又用一些更为简单的基本图像单元来表示，直至子图像达到了我们认为的最简单的图像单元（基元）；
- 所有这些基元按一定的结构关系来表示，利用多级树结构对其进行描述（这种描述可以采用形式语言理论）。

吉祥如意

实例：句法模式识别（续）

- 多级树描述结构





实例：句法模式识别（续）

- 训练过程：

- 用已知结构信息的图像作为训练样本，先识别出基元（比如场景图中的X、Y、Z等简单平面）和它们之间的连接关系（例如长方体E是由X、Y和Z三个面拼接而成），并用字母符号代表之；
- 然后用构造句子的文法来描述生成这幅场景的过程，由此推断出生成该场景的一种文法。



实例：句法模式识别（续）

■ 识别过程：

- 先对未知结构信息的图像进行基元提取及其相互结构关系的识别；
- 然后用训练过程获得的文法做句法分析；
- 如果能被已知结构信息的文法分析出来，则该幅未知图像与训练样本具有相同的结构（识别成功），否则就不是这种结构（识别失败）。



模式识别的基本问题

一. 模式(样本)表示方法

1. 向量表示：假设一个样本有n个变量(特征)

$$X = (X_1, X_2, \dots, X_n)^T$$

2. 矩阵表示：N个样本，n个变量(特征)

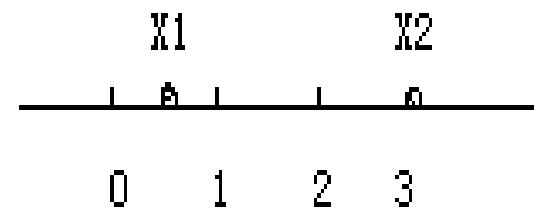
样本	X ₁	X ₂	...	X _n
X ₁	X ₁₁	X ₁₂	...	X _{1n}
X ₂	X ₂₁	X ₂₂	...	X _{2n}
...
X _N	X _{N1}	X _{N2}	...	X _{Nn}



3. 几何表示

一维表示

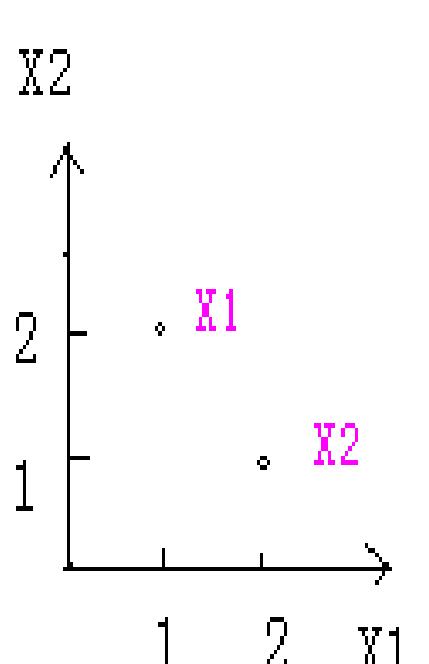
$$X_1 = 1.5 \quad X_2 = 3$$



二维表示

$$X_1 = (x_1, x_2)^\top = (1, 2)^\top$$

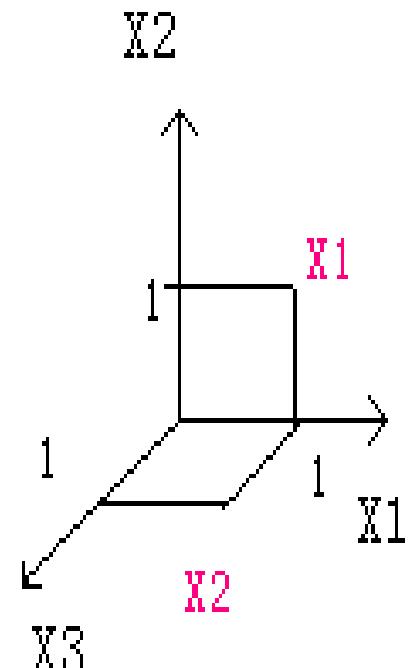
$$X_2 = (x_1, x_2)^\top = (2, 1)^\top$$



三维表示

$$X_1 = (x_1, x_2, x_3)^\top = (1, 1, 0)^\top$$

$$X_2 = (x_1, x_2, x_3)^\top = (1, 0, 1)^\top$$





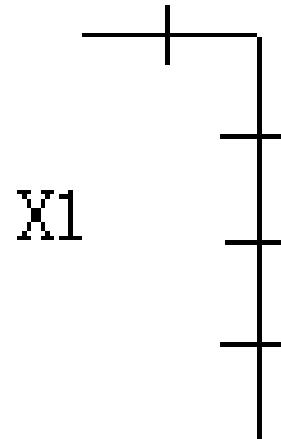
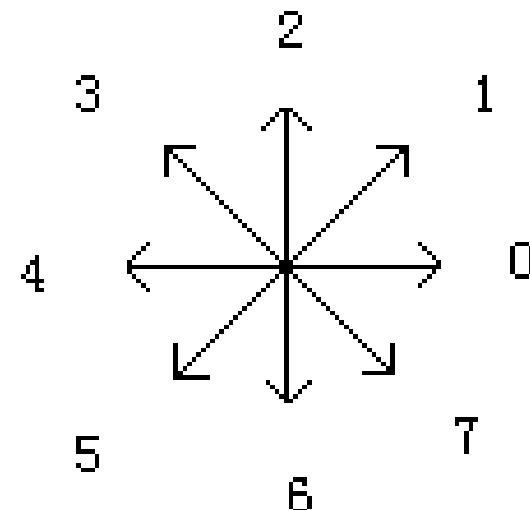
4. 基元（链码）表示：

在右侧的图中八个基元分别表示0, 1, 2, 3, 4, 5, 6, 7, 八个方向和基元线段长度。

则右侧样本可以表示为

$$X_1=006666$$

这种方法将在句法模式识别中用到。



吉
祥
圖

二. 模式类的紧致性

1. 紧致集：同一类模式类样本的分布比较集中，没有或临界样本很少，这样的模式类称紧致集。

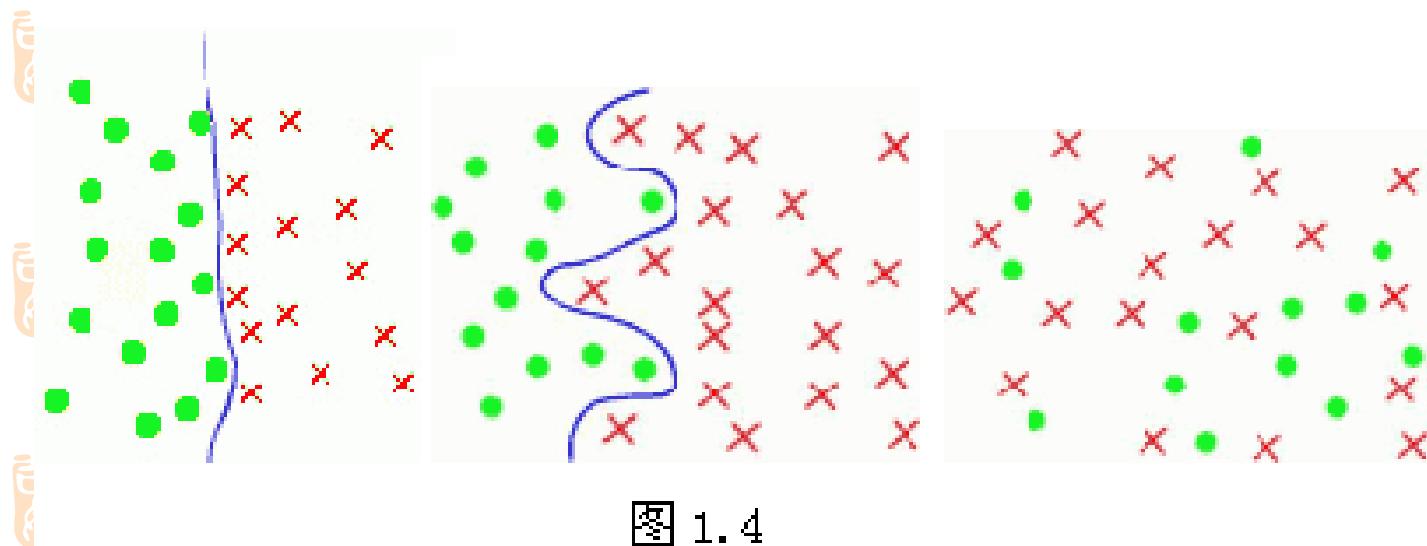


图 1.4



2. 临界点(样本): 在多类样本中, 某些样本的值有微小变化时就变成另一类样本称为临界样本(点)。

3. 紧致集的性质

① 要求临界点很少

② 集合内的任意两点的连线, 在线上的点属于同一集合

③ 集合内的每一个点都有足够大的邻域, 在邻域内只包含同一集合的点

4. 模式识别的要求: 满足紧致集, 才能很好的分类; 如果不满足紧致集, 就要采取变换的方法, 满足紧致集.



三.相似与分类

1.两个样本 x_i, x_j 之间的相似度量满足以下要求:

- ① 应为非负值
- ② 样本本身相似性度量应最大
- ③ 度量应满足对称性
- ④ 在满足紧致性的条件下, 相似性应该是点间距离的
单调函数

2.用各种距离表示相似性:

- ① 绝对值距离

已知两个样本 $x_i=(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})^T$

$x_j=(x_{j1}, x_{j2}, x_{j3}, \dots, x_{jn})^T$



$$d_{ij} = \sum_{k=1}^n |X_{ik} - X_{jk}|$$

② 欧几里德距离

$$d_{ij} = \sqrt{\sum_{k=1}^n (X_{ik} - X_{jk})^2}$$

③ 明考夫斯基距离

$$d_{ij}(q) = \left(\sum_{k=1}^n |X_{ik} - X_{jk}|^q \right)^{1/q}$$

其中当 $q=1$ 时为绝对值距离，当 $q=2$ 时为欧氏距离



④ 切比雪夫距离 $d_{ij}(\infty) = \max_{1 \leq k \leq n} |X_{ik} - X_{jk}|$

q趋向无穷大时明氏距离的极限情况

⑤ 马哈拉诺比斯距离

$$d_{ij}(M) = \sqrt{(X_i - X_j)^T \Sigma^{-1} (X_i - X_j)}$$



其中 x_i, x_j 为特征向量, Σ 为协方差。使用的条件是
样 本符合正态分布



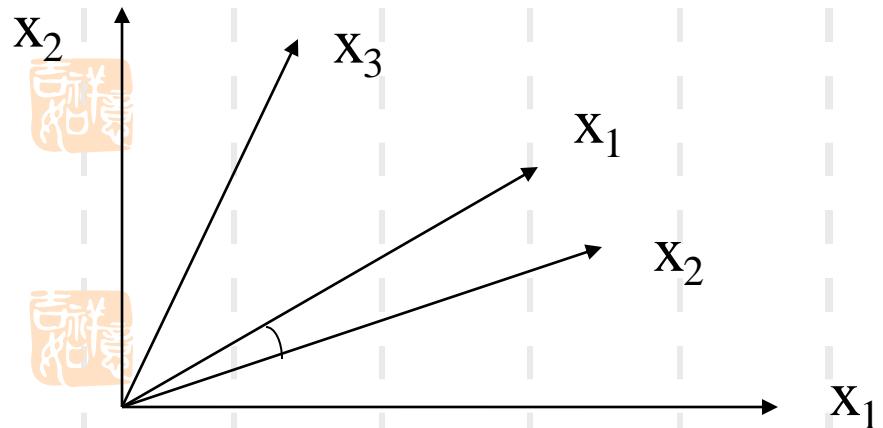
⑥ 夹角余弦

$$C_{ij} = \frac{\sum_{k=1}^n X_{ik}X_{jk}}{\sqrt{\left(\sum_{k=1}^n X_{ik}^2\right)\left(\sum_{k=1}^n X_{jk}^2\right)}}$$

吉祥如意

\bar{X}_i, \bar{X}_j 为 x_i, x_j 的均值 即样本间夹角小的为一类，具有相似性

例： x_1, x_2, x_3 的夹角如图：



因为 x_1, x_2 的夹角小，所以 x_1, x_2 最相似。



⑦ 相关系数

$$r_{ij} = \frac{\sum_{k=1}^n (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j)}{\sqrt{\sum_{k=1}^n (X_{ki} - \bar{X}_i)^2} \sum_{k=1}^n (X_{kj} - \bar{X}_j)^2}$$

\bar{X}_i, \bar{X}_j 为 x_i, x_j 的均值

注意：在求相关系数之前，要将数据标准化

3. 分类的主观性和客观性

① 分类带有主观性：目的不同，分类不同。例如：鲸鱼，牛，马从生物学的角度来讲都属于哺乳类，但是从产业角度来看鲸鱼属于水产业，牛和马属于畜牧业。

② 分类的客观性：科学性

判断分类必须有客观标准，因此分类是追求客观性的，但主观性也很难避免，这就是分类的复杂性。

吉祥如意

四. 特征的生成

1. 低层特征：

- ① 无序尺度：有明确的数量和数值。
- ② 有序尺度：有先后、好坏的次序关系，如酒分为上，中，下三个等级。
- ③ 名义尺度：无数量、无次序关系，如有红，黄两种颜色

2. 中层特征：经过计算，变换得到的特征

3. 高层特征：在中层特征的基础上有目的的经过运算形成

例如：椅子的重量=体积*比重

体积与长，宽，高有关；比重与材料，纹理，颜色有关。这里低、中、高三层特征都有了。



五.数据的标准化

1. 极差标准化，一批样本中，每个特征的最大值与最小值之差。

极差 $R_i = \max X_{ij} - \min X_{ij}$

极差标准化 $X_{ij} = \frac{(X_{ij} - \bar{X}_i)}{R_i}$

2. 方差标准化 $X_{ij} = \frac{(X_{ij} - \bar{X}_i)}{S_i}$

S_i 为方差

标准化的方法很多，原始数据是否应该标准化，应采用什么方法标准化，都要根据具体情况来定。



相关数学概念

- 随机向量及其分布

- 随机向量

- 如果一个对象的特征观察值为 $\{x_1, x_2, \dots, x_n\}$, 它可构成一个n维的特征向量值 \mathbf{x} , 即

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^T$$

式中, x_1, x_2, \dots, x_n 为特征向量 \mathbf{x} 的各个分量。

- 一个特征可以看作n维空间中的向量或点, 此空间称为模式的特征空间 R_n 。



吉祥如意

相关数学概念

- 随机向量及其分布

- 随机向量

- 在模式识别过程中，要对许多具体对象进行测量，以获得许多次观测值。
 - 每次观测值不一定相同，所以对许多对象而言，各个特征分量都是随即变量，即许多对象的特征向量在n维空间中呈随机性分布，称为随即向量。



吉祥如意

相关数学概念

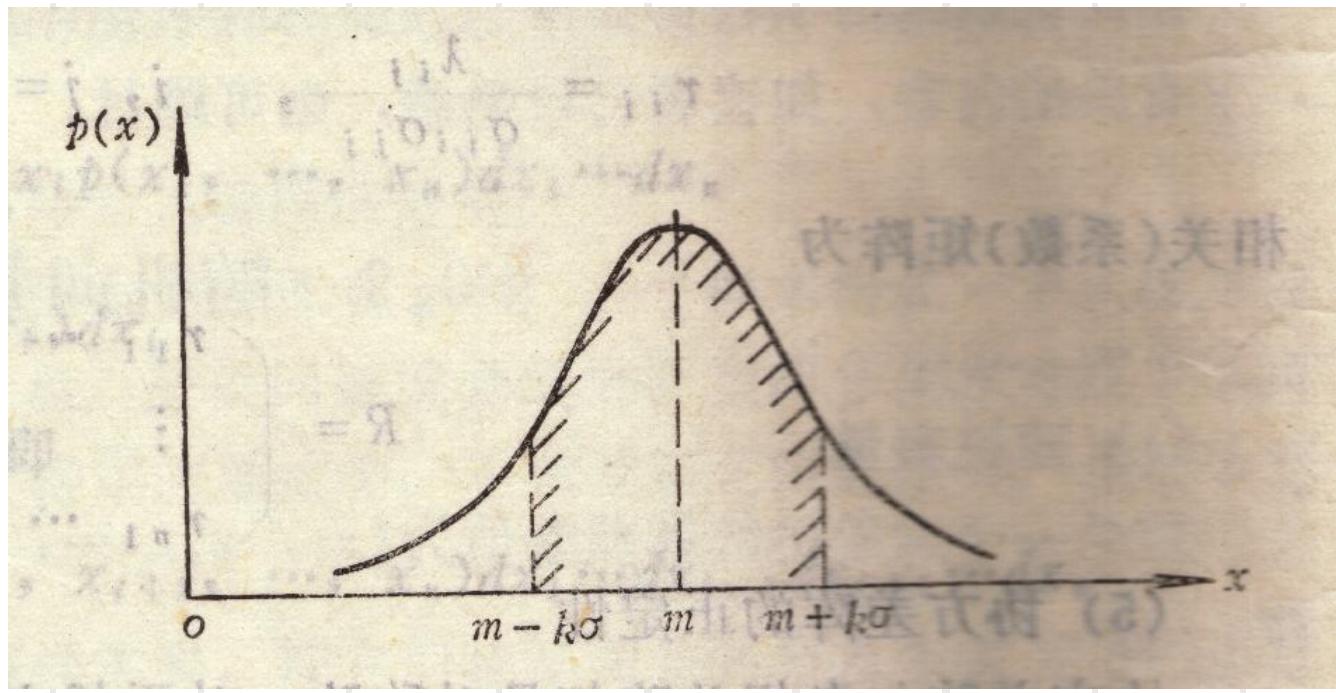
- 随机向量及其分布
 - 随机向量的参数
 - [数学期望]
 - [协方差矩阵]



吉祥如意

相关数学概念

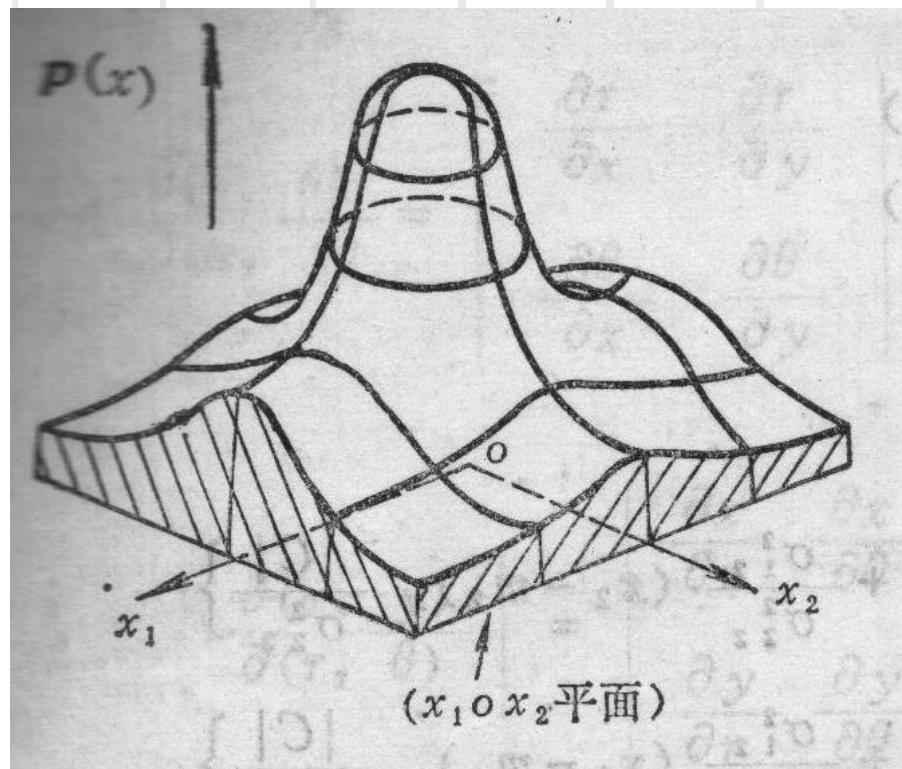
- 正态分布
 - [一维正态密度函数]



吉祥如意

相关数学概念

- 正态分布
 - [多维正态密度函数]



吉祥如意

小结

- 模式和模式识别的概念
- 模式识别的发展简史和应用
- 模式识别的主要方法
- 模式识别的系统和实例
- 模式识别的基本问题
- 几个相关的数学概念



吉祥如意

小结

